AI Agents in Multi-Criteria Decision Analysis: Automating the Analytic Hierarchy Process with Large Language Models

Authors:

Igor Svoboda [a], ORCID: 0009-0001-1709-2904, (i.svoboda@kpi.ua)

Dmytro Lande [a], ORCID: 0000-0003-3945-1178, (dwlande@gmail.com)

Affiliations:

[a] Information Security Department, National Technical University of Ukraine "Kyiv Polytechnic Institute", 37, Prospect Beresteiskyi, Solomyanskyi district, Kyiv, Ukraine, 03056

Corresponding Author:

Igor Svoboda (i.svoboda@kpi.ua)

Abstract – **This study introduces a novel framework that integrates the Analytic Hierarchy Process (AHP) with advanced large language models (LLMs) to automate and enhance multi-criteria decision analysis (MCDA), particularly in cybersecurity applications. By leveraging the capabilities of these LLMs, we create a system of AI agents that effectively replace human input in the AHP process, from criteria selection and pairwise comparisons to alternative evaluation. This automation increases efficiency, ensures judgment consistency, and reduces potential biases. Our findings demonstrate the feasibility and transformative potential of this approach, showcasing its ability to generate reliable and consistent AHP results. This framework establishes a new paradigm for intelligent decision support systems by merging traditional MCDA methodologies with cutting-edge AI, opening promising avenues for future research and applications in various domains.**

Keywords – Analytic Hierarchy Process, multi-criteria decision analysis, decision support systems, automation, Large Language Models, artificial intelligence, generative AI, AI agents, cybersecurity.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

1. Introduction

The field of decision-making relies heavily on structured methodologies to navigate the complexities of evaluating multiple, often conflicting, criteria. This need for structured approaches becomes particularly critical when decisions require balancing a diverse range of qualitative and quantitative factors, often in dynamic and uncertain environments. Effective decision-making processes are essential for both organizations and individuals as they strive to make informed choices that align with their objectives and risk tolerance. The ongoing evolution of these methodologies is driven by the pursuit of more systematic, transparent, and rational frameworks, ultimately leading to the development and adoption of advanced decision-support tools and techniques [1][2].

The Analytic Hierarchy Process (AHP), developed by Thomas L. Saaty in the 1970s, represents a significant advancement in Multi-Criteria Decision Making (MCDM). Designed to address complex decision scenarios, AHP provides a structured framework for decomposing intricate problems into a hierarchy of smaller, more manageable sub-problems. This hierarchical decomposition facilitates a systematic analysis where each element can be independently evaluated and compared, considering both qualitative and quantitative aspects. AHP's introduction marked a pivotal moment in the evolution of MCDM tools, offering a rigorous yet flexible approach to tackling multifaceted decision problems [1][3].

AHP's strength lies in its ability to capture both the tangible and intangible elements of decision-making, enabling a semi-objective approach to quantifying the weights and preferences associated with various decision criteria [3]. This is achieved through a structured pairwise comparison process, where decision-makers systematically evaluate the relative importance of each criterion against all others. This method not only provides a clear framework for decision analysis but also promotes consistency and transparency in judgment.

Since its inception, AHP has been extensively studied, refined, and applied across diverse domains, demonstrating its versatility and effectiveness. Its applications span various fields [4], including business [5], engineering [6], healthcare [7], and environmental management [8], among others. The development of supporting software further enhanced AHP's practical applicability, enabling its use in addressing real-world problems across different sectors [9]. AHP's unique ability to combine mathematical rigor with subjective human judgment has solidified its position as a prevalent decision-support tool in both theoretical and applied decision sciences.

The emergence of Large Language Models (LLMs), particularly the Generative Pre-trained Transformer (GPT) series [10], has opened new frontiers in AI-driven decision-making. Trained on vast and diverse datasets, LLMs possess remarkable capabilities in natural language processing, enabling them to comprehend complex information, generate coherent responses, and automate various aspects of decision support systems [11]. This advancement signifies a

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

paradigm shift in decision-making, where AI plays an increasingly important role in augmenting human capabilities and facilitating more informed choices.

Recent research has increasingly focused on exploring the potential of LLMs within decision-making frameworks [12][13]. Studies have investigated the adaptability and precision of LLMs in various decision-making contexts, highlighting their ability to provide valuable support [14][15]. These investigations have emphasized the importance of prompt engineering and understanding the sensitivity of LLMs to different inputs and parameters to effectively leverage their capabilities in decision-making roles.

The transformative influence of LLMs on decision-making is evident in their ability to automate and refine decision support systems across diverse fields. From healthcare to autonomous driving [16], LLMs offer promising avenues for innovation, enhancing the efficiency and effectiveness of complex decision-making processes. Their ability to process and analyze information on a scale, coupled with their capacity for natural language understanding, positions them as valuable tools for advancing AI's contribution to intricate human decision-making.

Despite the widespread use of AHP in complex decision scenarios and the growing prominence of LLMs in automating decision support, research explicitly integrating these two powerful approaches remains limited. This gap is particularly noticeable in studies that directly combine AHP's structured methodology with the advanced language processing capabilities of LLMs to enhance the efficiency and automation of decision-making. This integration presents several challenges, including the need to effectively translate qualitative AHP concepts into a format understandable by LLMs, ensure consistency and coherence in the AI's judgments, and validate the reliability of the generated outputs.

Bridging this gap offers significant potential benefits. By leveraging LLMs' ability to process information at scale and generate human-like text, it becomes possible to automate various aspects of the AHP process, such as criteria identification, pairwise comparisons, and alternative evaluation. This automation can lead to more efficient decision-making processes, reduce the cognitive burden on human decision-makers, and potentially minimize biases associated with subjective human judgment. Moreover, integrating LLMs with AHP can open new avenues for incorporating diverse perspectives and data sources into decision analysis, leading to more robust and informed outcomes.

This research aims to address this gap by introducing a novel framework that integrates AHP with LLMs, specifically leveraging AI agents to automate and enhance the AHP process. We demonstrate the feasibility and transformative potential of this approach, showcasing its ability to generate efficient, consistent, and potentially less biased decision-making outcomes in complex situations.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

2. Methodology

2.1 LLM Selection and Customization

In determining the most suitable Large Language Model (LLM) for this research, we considered established benchmarks such as HumanEval, MMLU, and HellaSwag [17]. These benchmarks evaluate LLMs across various dimensions, including reasoning, common sense, and natural language understanding. Based on their superior performance in these evaluations, we selected the GPT-4o model for its advanced capabilities in language comprehension and generation. While the GPT-4o API offered direct access to the model, we opted to utilize the ChatGPT interface for its user-friendly environment, enhanced memory capabilities, and streamlined prototyping process, which facilitated efficient interaction and experimentation with the model.

Our research methodology leverages an OpenAI feature that enables the creation of custom GPTs. This feature allows for tailoring the model's behavior by providing specific instructions, foundational knowledge in the form of documents, and constraints relevant to the desired tasks [18]. This customization enhances the model's ability to perform domain-specific tasks and generate more accurate and relevant outputs.

2.2 AI Agent Roles and Instructions

We began by creating a custom GPT designated as the "AHP Guide." This AI agent served as the primary decision-maker in establishing the structure of our AHP model, including determining the optimal number of hierarchical levels and the number of AI agents required for subsequent steps. This approach reflects the collaborative nature of the research, where certain tasks are delegated to specialized AI agents while others are performed manually by the researchers [19].

The "AHP Guide" AI agent was equipped with a specific set of instructions designed to guide its decision-making within the AHP framework. These instructions, detailed in Appendix A, emphasized its role in facilitating user interaction with the AHP process, managing input from external AI agents, and ensuring effective incorporation of expert opinions into the model. This guidance is crucial for establishing and executing the AHP method, particularly in complex decision-making scenarios where diverse expertise is essential.

To further enhance the "AHP Guide's" capabilities, we provided it with foundational knowledge in the form of Saaty's 1990 paper on AHP [3]. This ensured that the AI agent had a thorough understanding of the AHP methodology and could provide accurate guidance throughout the process. Additionally, we explored and utilized all available ChatGPT capabilities, including web browsing, DALL-E image generation, and code interpreter, to maximize the AI agent's potential in assisting with the research.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

We then tasked the "AHP Guide" with determining the optimal number of AI agents required for our research goal: "Secure the Corporate Datacenter from Social Engineering Attacks." Based on its analysis, the "AHP Guide" recommended a range of 5-7 AI agents, from which we selected the higher number to ensure a diverse range of perspectives.

The "AHP Guide" was then prompted to generate descriptions for each of the seven AI agents, detailing their professional backgrounds, expertise, and work-related personalities. These descriptions served as the foundation for creating distinct AI agent personas, each contributing unique perspectives to the AHP model building process.

We also consulted the "AHP Guide" to determine the optimal number of hierarchical levels for our criteria tree. The AI agent recommended a two-level structure as the most suitable for our research goal, providing sufficient detail without excessive complexity.

Following the "AHP Guide's" recommendations, we proceeded to interact with each AI agent to gather their input for constructing the AHP model. This involved eliciting their expert opinions on relevant criteria, sub-criteria, and alternatives, as well as conducting pairwise comparisons to establish the AHP matrices.

To ensure consistency and clarity in these interactions, we utilized the custom GPT feature to create personalized AI agents with distinct personas. This approach, while not strictly necessary, aimed to enhance the individuality of each AI agent and potentially improve the quality of their contributions to the decision-making process.

An example of this personalization is the creation of the AI agent "Dr. Ava Chen," a cybersecurity strategist. We provided "Dr. Chen" with detailed instructions to maintain a professional yet approachable demeanor, utilize formal language, and incorporate relevant personal insights to enhance the engagement and relatability of her interactions. Detailed custom GPT instructions for “Dr. Chen” can be found in Appendix B.

2.3 Criteria and Alternative Generation

We then tasked "Dr. Chen" with identifying seven top-level criteria relevant to our research goal. The optimal number of criteria could be determined either independently by the researchers or through consultation with the "AHP Guide." "Dr. Chen" generated a comprehensive list of criteria, which included Employee Training, Access Control, Communication Protocols, Incident Response, Physical Security, Policy Enforcement, and Monitoring Systems.

This process was repeated with the remaining six AI agents, resulting in a total of 49 top-level criteria. To refine this list, we eliminated duplicate or synonymous criteria, ultimately retaining 45 unique criteria for further analysis.

Each AI agent was then asked to assign a score from 1 to 9 to each of the 45 criteria, reflecting their perceived importance in achieving the research goal. This scoring process aimed

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

to capture the subjective judgment of each AI agent regarding the relative significance of the criteria.

The scores assigned by all AI agents were aggregated for each criterion, and the seven criteria with the highest cumulative scores were selected for further analysis. This selection process can be represented mathematically using the following formula (1):

$$S_{\text{item}_i} = \sum_{k=1}^{E} s_{ik}\text{, for } i = 1,2,\ldots,n \times E \tag{1}$$

where S_{item_i} represents the total score for a given item (criteria, sub-criteria, or alternative), s_{ik} is the score assigned to the item by a specific AI agent, E is the total number of AI agents, and n is the predetermined number of top-scored items to be selected.

This formula ensures a systematic and objective selection of the most important criteria based on the collective judgment of the AI agents.

In our case, the seven selected criteria were Social Engineering Awareness, Physical Access Controls, Audit Trails, Behavior Analysis, Operational Risk Controls, Psychological Profiling, and Service Level Agreements.

We then proceeded to the next level of the hierarchy by asking "Dr. Ava Chen" to generate three sub-criteria for each of the seven selected top-level criteria. This process was repeated with the other AI agents, and the resulting sub-criteria were ranked and refined to produce a concise list for each top-level criterion.

Finally, we generated a list of alternatives by asking each AI agent to propose five potential solutions. These alternatives were aggregated and evaluated by the AI agents, who assigned scores based on their perceived effectiveness in achieving the research goal. The top five alternatives with the highest cumulative scores were selected for inclusion in the AHP model.

With the completion of these steps, the structure of our AHP tree was finalized. While this process could be further simplified by having a single AI agent emulate multiple expert opinions, we believe that utilizing distinct AI agents with personalized expertise enhances the diversity of perspectives and potentially improves the quality of the final decision.

Alternatively, the process of criteria and alternative selection could be refined by conducting pairwise comparisons and constructing matrices at each level of the hierarchy. However, this would significantly increase the computational burden due to the large number of comparisons required.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

2.4 Pairwise Comparisons and Matrix Construction

The next stage involved creating pairwise comparison matrices for the top-level criteria for each of the seven AI agents. This resulted in seven unique matrices, each reflecting the individual AI agent's judgment regarding the relative importance of the criteria.

For instance, we prompted "Dr. Ava Chen" to construct a pairwise comparison matrix by comparing each pair of top-level criteria and assigning a value from 1/9 to 9 based on their relative importance in achieving the research goal. The exact prompt used for this task can be found in Appendix C. This process was repeated with the other AI agents, ensuring that each provided their independent assessment.

The resulting seven matrices were then aggregated into a single matrix using the geometric mean method, as recommended for AHP analysis [20]. This method is preferred for its ability to effectively combine individual judgments while minimizing the influence of extreme values.

The geometric mean aggregation formula is as follows:

$$A_{\text{agg}}(i,j) = \left(\prod_{k=1}^{E} A_k(i,j)\right)^{\frac{1}{E}} \tag{2}$$

where $A_{\text{agg}}(i,j)$ represents the aggregated pairwise comparison value between elements i and j in the aggregated matrix, $A_k(i,j)$ denotes the pairwise comparison value between elements i and j given by AI agent k, and E is the total number of AI agents.

While arithmetic mean aggregation is also possible, the geometric mean is generally preferred for its ability to produce more consistent and balanced results.

2.5 AHP Analysis and Consistency Checks

Following AHP methodology, the aggregated matrix was then normalized using the formula (3):

$$a_{ij}^{\text{norm}} = \frac{a_{ij}}{\sum_{k=1}^{n} a_{kj}} \tag{3}$$

where a_{ij}^{norm} is the normalized value for the element at the i-th row and j-th column, a_{ij} is the aggregated pairwise comparison value between elements i and j, and $\sum_{k=1}^{n} a_{kj}$ is the sum of all elements in the j-th column.

Normalization ensures that the values in the matrix are scaled appropriately for subsequent calculations.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Next, we calculated the priority vectors using the formula (4):

$$w_i = \frac{1}{n}\sum_{j=1}^{n} a_{ij}^{\text{norm}} \tag{4}$$

where w_i is the priority weight for the i-th criterion (or alternative), and n is the total number of criteria (or alternatives).

Priority vectors represent the relative importance of each criterion in achieving the research goal.

To ensure the reliability of our analysis, we performed consistency checks using the following formulas (5, 6):

$$CI = \frac{\lambda_{\max} - n}{n-1} \tag{5}$$

$$CR = \frac{CI}{RI} \tag{6}$$

where $\lambda_{\max}$ is the largest eigenvalue of the aggregated pairwise comparison matrix, n is the number of criteria (or alternatives), CI is the consistency index, RI is the random index, and CR is the consistency ratio.

The consistency ratio measures the degree to which the judgments provided by the AI agents are coherent and reliable. In our analysis, the consistency ratio was well below the commonly accepted threshold of 0.1, indicating that the judgments were consistent and suitable for the AHP analysis.

We then performed similar operations for the sub-level criteria, prompting each AI agent to create pairwise comparison matrices for the sub-criteria within each top-level criterion. An example prompt that we used for "Dr. Chen" can be found in Appendix D. This resulted in forty-nine matrices, which were aggregated, normalized, and checked for consistency.

The final step in the AHP methodology involves calculating the best alternative, which requires several sub-steps. We began by creating pairwise comparison matrices of alternatives for each sub-criteria, resulting in a total of 147 matrices.

For example, we prompted "Dr. Ava Chen" to construct pairwise comparison matrices for three sub-criteria at a time, comparing each pair of alternatives and assigning weights based on her subjective judgment. The specific prompt can be found in Appendix E. This process was repeated for all AI agents and sub-criteria, resulting in a comprehensive set of matrices reflecting their individual assessments.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

These matrices were then aggregated using the geometric mean method, normalized, and used to extract priority vectors for each alternative within each sub-criterion. Finally, we calculated the final scores for each alternative by multiplying the sub-criteria global priority with their respective alternative priority vectors and summing the results.

This process can be represented mathematically as:

$$\text{Best Alternative} = \max_{\text{alternative}} \left(\sum_{i=1}^{n} \left(\text{Priority}_{\text{criterion}_i} \times \text{Priority}_{\text{alternative}|\text{criterion}_i} \right) \right) \quad (7)$$

This formula (7) identifies the alternative that maximizes the weighted sum of priorities across all sub-criteria, indicating the most preferred solution based on the collective judgment of the AI agents.

In our analysis, the alternative "Comprehensive Employee Training Programs" received the highest score and was therefore selected as the best alternative for achieving our research goal of securing a datacenter against social engineering attacks. This result aligns with expert consensus in the field, which emphasizes the critical role of employee training in mitigating social engineering threats.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

3. Results

3.1 AI Agent Judgments and Consistency

Our experiment yielded consistently reliable judgments from the AI agents, as evidenced by the pairwise comparison matrices they generated. The consistency ratios for these matrices were consistently below the widely accepted threshold of 0.1, with most falling below 0.01. This high level of consistency indicates that the AI agents' judgments were coherent and reliable, fulfilling a crucial requirement for the validity of AHP analysis.

Throughout the AHP process, the AI agents constructed a total of 231 matrices, generated 49 top-level criteria, 147 sub-criteria, and 35 alternatives. These outputs align with the initial AHP parameters recommended by our "AHP Guide" AI agent, which included a two-level AHP tree with 5 alternatives, 7 AI agents, 7 top-level criteria, and 3 sub-criteria per top-level criterion.

Following these parameters, the "AHP Guide" produced a diverse and inclusive set of AI agent personas, each with unique backgrounds, expertise, and personalities. The comprehensive list of AI agent personas is detailed in Appendix F.

The initial set of top-level criteria generated by the AI agents is presented in Appendix G. As demonstrated in the appendix, each AI agent contributed unique criteria relevant to their specific domain knowledge and professional background. This diversity of perspectives is crucial for ensuring the comprehensiveness and validity of the AHP model.

3.2 Criteria and Alternative Selection

To select the most relevant criteria from the initial set, we employed a simple voting method based on the AI agents' scores. This resulted in the following final set of top and sub-level criteria:

1) Social Engineering Awareness:
    a) Training Program Effectiveness
    b) Awareness Session Regularity
    c) Incident Reporting Protocol
2) Physical Access Controls:
    a) Biometric System Reliability
    b) Visitor Tracking System
    c) Access Point Monitoring
3) Audit Trails:
    a) Log Analysis Accuracy
    b) Audit Frequency
    c) Anomaly Tracking Efficiency
4) Behavior Analysis:
    a) User Behavior Monitoring

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

    b) Response to Anomalies
    c) Activity Pattern Analysis
5) Operational Risk Controls:
    a) Infrastructure Vulnerability Check
    b) Data Redundancy Systems
    c) Emergency Protocol Effectiveness
6) Psychological Profiling:
    a) Staff Behavior Assessment
    b) Risk Behavior Profiling
    c) Continuous Observation
7) Service Level Agreements:
    a) Response Time Commitment
    b) Data Privacy Assurance
    c) Breach Penalty Specification

The final list of alternatives included:

1) Cloud-Based Data Backup Solutions
2) Physical Barrier Reinforcement
3) Security Personnel Training Update
4) Comprehensive Employee Training Programs
5) Advanced Intrusion Detection Systems

A visualization of the complete AHP tree structure is provided in Appendix H.

3.3 AHP Matrix Construction and Analysis

The next phase of the analysis involved constructing pairwise comparison matrices. We allowed flexibility in this process, enabling the AI agents to either explicitly state their pairwise comparisons, which the researchers would then manually translate into matrices, or directly generate the matrices themselves.

The resulting aggregated top-level criteria matrix is presented in Table 1.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Table 1. Aggregated top-level criteria matrix.

Criteria	SE Awareness	Physical Controls	Audit Trails	Behavior Analysis	Operational Risks	Psychological Profiling	SLAs
Social Engineering Awareness	1.000	1.319	1.104	1.483	1.081	0.498	0.369
Physical Access Controls	0.756	1.000	1.673	1.560	1.029	0.937	0.408
Audit Trails	0.904	0.601	1.000	1.251	0.701	0.756	0.325
Behavior Analysis	0.674	0.641	0.798	1.000	0.627	0.801	0.503
Operational Risks	0.920	0.966	1.426	1.608	1.000	0.604	0.526
Psychological Profiling	2.007	1.068	1.319	1.247	1.636	1.000	0.652
SLAs	2.712	2.438	3.061	1.990	1.883	1.532	1.000

After performing the necessary calculations, the top-level criteria priority vectors were determined as follows:

1) Social Engineering Awareness: 0.120
2) Physical Access Controls: 0.131
3) Audit Trails: 0.099
4) Behavior Analysis: 0.096
5) Operational Risks: 0.126
6) Psychological Profiling: 0.164
7) Service Level Agreements: 0.264

The consistency check results were:

- Consistency Index (CI): 0.022
- Consistency Ratio (CR): 0.016
- Lambda max (λ_max): 7.13

As anticipated, the consistency ratio was well below 0.1, and the distribution of priority vectors among the top-level criteria appeared reasonably balanced.

The analysis of the matrices could be performed either manually using AHP software or by utilizing the Code Interpreter feature within ChatGPT, which allowed the AI agents to conduct the calculations themselves. We observed that the GPT-4o based AI agents were capable of accurately performing these calculations, producing consistent and reliable results.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

3.4 Priority Vectors and Consistency Checks

Following the analysis of the top-level criteria, we proceeded to analyze the sub-level criteria. Each AI agent generated comparison matrices for the sub-criteria, which were then aggregated and used to calculate the global priority vectors. The resulting priority vectors are as follows:

1) Response Time Commitment: 0.1127
2) Data Privacy Assurance: 0.0866
3) Staff Behavior Assessment: 0.0655
4) Breach Penalty Specification: 0.0644
5) Infrastructure Vulnerability Check: 0.0573
6) Risk Behavior Profiling: 0.0546
7) Biometric System Reliability: 0.0502
8) Training Program Effectiveness: 0.0485
9) Continuous Observation: 0.0440
10) Visitor Tracking System: 0.0434
11) Audit Frequency: 0.0385
12) Data Redundancy Systems: 0.0384
13) User Behavior Monitoring: 0.0378
14) Access Point Monitoring: 0.0375
15) Incident Reporting Protocol: 0.0368
16) Awareness Session Regularity: 0.0347
17) Response to Anomalies: 0.0340
18) Log Analysis Accuracy: 0.0317
19) Emergency Protocol Effectiveness: 0.0302
20) Anomaly Tracking Efficiency: 0.0288
21) Activity Pattern Analysis: 0.0242

The distribution of priority vectors among the sub-criteria appeared reasonable and aligned with the overall goal of the analysis. The consistency ratios for the sub-criteria matrices ranged from 0.002 to 0.02, indicating a high degree of consistency in the AI agents' judgments.

The final step of the AHP analysis involved identifying the best alternative based on the sub-criteria matrices. This analysis yielded the following priority vectors for the alternatives:

1) Cloud-Based Data Backup: 0.1938
2) Physical Barrier Enhancements: 0.1254
3) Security Personnel Training and Updates: 0.1795
4) Comprehensive Employee Training Programs: 0.2774
5) Advanced Intrusion Detection Systems: 0.2240

The alternative "Comprehensive Employee Training Programs" received the highest priority vector (0.2774), while "Physical Barrier Enhancements" received the lowest (0.1254). These results are consistent with a real-world expert consensus [21], which emphasizes the importance

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

of employee training in mitigating social engineering attacks and the relatively lower effectiveness of physical barrier enhancements alone

The consistency ratios for the aggregate matrices in this final step ranged from 0.002 to 0.03, well below the 0.1 threshold, ensuring the consistency and reliability of the analysis.

While previous research has shown that GPT-3.5 struggles with accurately creating pairwise comparison matrices [22], our use of GPT-4o based AI agents demonstrated their ability to perform this task effectively. Although there were occasional instances where the AI agents deviated from the AHP guidelines, these were easily corrected with reminder prompts.

Interestingly, the AI agents often provided explanations for their assigned scores in the matrices. For example, Table 2 presents a matrix constructed by "Dr. Ava Chen" for prioritizing the sub-criteria of "Social Engineering Awareness," along with her accompanying rationale.

Table 2. Virtual expert “Dr. Ava Chen” Social Engineering Awareness sub-criteria comparison matrix.

Sub-criteria	Training Effectiveness	Session Regularity	Reporting Protocol
Training Program Effectiveness	1	2	3
Awareness Session Regularity	½	1	2
Incident Reporting Protocol	1/3	½	1

"Dr. Ava Chen" emphasized the importance of training program effectiveness in recognizing and responding to social engineering attacks, highlighting the role of regular sessions for maintaining vigilance and robust reporting protocols for timely mitigation.

While such explanations can be valuable for understanding the AI agents' decision-making processes, they may not be necessary in production-level systems due to the additional computational resources they require.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

4. Discussion and Conclusions

4.1 Discussion of Findings

The ability of AI agents to generate balanced and consistent judgments in the AHP process was remarkable. They demonstrated a holistic understanding of the decision-making problem, resulting in high consistency ratios and a reasonable distribution of weights among the criteria. This reliability is crucial for real-world applications of AHP, where the accuracy and coherence of expert judgments significantly impact the quality of decisions. While further research is needed to refine the AHP-GPT framework and explore its full potential, our findings suggest that AI agents can play a valuable role in automating and enhancing multi-criteria decision analysis.

The financial implications of our findings are also worth noting. Hiring human experts for AHP analysis can be expensive, with hourly rates often exceeding $50 [23]. In our case study, using human experts could have cost an estimated $700. In contrast, our AI-driven approach incurred a cost of approximately $2, demonstrating significant cost savings. This highlights the potential of AI agents to provide cost-effective decision support, particularly for resource-constrained organizations.

The efficiency of our system relies on the AI agents' ability to make well-informed decisions. In situations where decisions require internal organizational knowledge, providing the AI agents with access to relevant data is crucial. This can be achieved through various means, such as providing documents during chat interaction or integrating the AI agents with internal databases. In our study, we provided the "AHP Guide" with Saaty's research paper to ensure its understanding of the AHP methodology.

For practical applications, organizations could connect AI agents to their internal knowledge bases, ensuring access to relevant information for accurate decision-making. The AI agents, in essence, function as abstract information processing units, adopting specific personas to enhance diversity of thought and maintain consistency in judgment.

It's important to acknowledge the potential limitations of AI agents, such as the possibility of hallucinations or memory constraints, especially in complex AHP models with numerous criteria and matrices. To mitigate these risks, strategies like initiating new chat sessions with the AI agents after completing each hierarchical level can be employed. Additionally, being mindful of the AI model's context window [24], which limits the amount of information it can actively retain, is crucial for maintaining consistency and coherence in the decision-making process.

4.2 Implications and Future Research

Our research contributes to the growing body of literature exploring the application of LLMs in decision-making. While previous studies have investigated the potential of LLMs in specific decision contexts, such as medical diagnosis [25], our work focuses on integrating LLMs with a

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

fundamental decision-making framework—AHP. This integration opens new possibilities for automating and enhancing multi-criteria decision analysis across various domains.

Beyond AHP, we believe that similar AI-driven approaches could be applied to other MCDM methodologies, such as the Analytic Network Process (ANP). ANP, while more complex than AHP, shares the core concept of pairwise comparisons, suggesting potential compatibility with AI agents. Further research is needed to explore the effectiveness of AI agents in ANP and other MCDM methodologies.

The straightforward methodology employed in our research allows for easy replication and adaptation to other decision-making contexts. We encourage further research to explore the potential of AI agents in MCDM, refine the AHP-GPT framework, and conduct rigorous benchmarking against traditional expert-driven approaches.

4.3 Conclusion

The integration of LLMs with established decision-making frameworks like AHP represents a promising direction for the field of Operations Research [26]. By embracing these advancements in AI, we can develop innovative tools and approaches that enhance human decision-making capabilities and address complex problems across various sectors of the economy and society.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

REFERENCES

[1] M. Köksalan, J. Wallenius, S. Zionts, Multiple Criteria Decision Making: From Early History to the 21st Century, World Scientific (2011). https://doi.org/10.1142/8042

[2] S. Greco, M. Ehrgott, J.R. Figueira (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, second ed., vol. XXXIII, Springer, New York, NY (2016), pp. 1-1347. https://doi.org/10.1007/978-1-4939-3094-4

[3] T.L. Saaty, How to make a decision: The analytic hierarchy process, Eur. J. Oper. Res. 48 (1) (1990) 9-26. https://doi.org/10.1016/0377-2217(90)90057-I

[4] S. Sipahi, M. Timor, The analytic hierarchy process and analytic network process: An overview of applications, Manag. Decis. 48 (5) (2010) 775-808. https://doi.org/10.1108/00251741011043920

[5] Y. Lee, K.A. Kozar, Investigating the effect of website quality on e-business success: An analytic hierarchy process (AHP) approach, Decis. Support Syst. 42 (3) (2006) 1383-1401. https://doi.org/10.1016/j.dss.2005.11.005

[6] M.-C. Lin, C.-C. Wang, M.-S. Chen, C.A. Chang, Using AHP and TOPSIS approaches in customer-driven product design process, Comput. Ind. 59 (1) (2008) 17-31. https://doi.org/10.1016/j.compind.2007.05.013

[7] G. Büyüközkan, G. Çifçi, S. Güleryüz, Strategic analysis of healthcare service quality using fuzzy AHP methodology, Expert Syst. Appl. 38 (8) (2011) 9407-9424. https://doi.org/10.1016/j.eswa.2011.01.103

[8] L. Suganthi, S. Iniyan, A.A. Samuel, Applications of fuzzy logic in renewable energy systems – A review, Renew. Sustain. Energy Rev. 48 (2015) 585-607. https://doi.org/10.1016/j.rser.2015.04.037

[9] S. French, D.-L. Xu, Comparison study of multi-attribute decision analytic software, J. Multi-Criteria Decis. Anal. 13 (2-3) (2005) 65-80. https://doi.org/10.1002/mcda.372

[10] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving Language Understanding by Generative Pre-Training (2018). https://openai.com/research/improving-language-understanding-by-generative-pre-training (accessed Feb. 11, 2024).

[11] M. Sallam, ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns, Healthcare (Switzerland) 11 (6) (2023) 887. https://doi.org/10.3390/healthcare11060887

[12] C. Hendriksen, Artificial intelligence for supply chain management: Disruptive innovation or innovative disruption?, J. Supply Chain Manag. 59 (3) (2023) 65-76. https://doi.org/10.1111/jscm.12304

[13] S.S. Biswas, Role of Chat GPT in Public Health, Ann. Biomed. Eng. 51 (5) (2023) 868-869. https://doi.org/10.1007/s10439-023-03172-7

[14] M. Loya, D. Sinha, R. Futrell, Exploring the Sensitivity of LLMs' Decision-Making Capabilities: Insights from Prompt Variation and Hyperparameters, in: Proc. of the Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, Dec. 2023, pp. 3711-3716. https://aclanthology.org/2023.findings-emnlp.241. DOI: https://doi.org/10.18653/v1/2023.findings-emnlp.241

[15] L. Tang et al., Evaluating large language models on medical evidence summarization, npj Digit. Med. 6 (1) (2023) 158. https://doi.org/10.1038/s41746-023-00896-7

[16] Y. Tian et al., VistaGPT: Generative Parallel Transformers for Vehicles With Intelligent Systems for Transport Automation, IEEE Trans. Intell. Veh. 8 (9) (2023) 4198-4207. https://doi.org/10.1109/TIV.2023.3307012

[17] J. Achiam et al., GPT-4 Technical Report (Version 4) [Preprint] (2023). https://doi.org/10.48550/arXiv.2303.08774

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

[18] OpenAI, Introducing GPTs, OpenAI Blog, Nov. 6, 2023. https://openai.com/blog/introducing-gpts
[19] D. Lande, L. Strashnoy, GPT Semantic Networking: A Dream of the Semantic Web – The Time is Now, Engineering, Kyiv, Ukraine (2023). ISBN: 978-966-2344-94-3
[20] W. Ossadnik, S. Schinke, R.H. Kaspar, Group Aggregation Techniques for Analytic Hierarchy Process and Analytic Network Process: A Comparative Analysis, Group Decis. Negot. 25 (2) (2016) 421–457. https://doi.org/10.1007/s10726-015-9448-4
[21] A. Ghafir, V. Prenosil, A. Alhejailan, M. Hammoudeh, Social engineering attack strategies and defence approaches, in: Proc. – 2016 IEEE 4th Int. Conf. on Future Internet of Things and Cloud, FiCloud 2016, pp. 145-149. https://doi.org/10.1109/FiCloud.2016.28
[22] D. Lande, L. Strashnoy, O. Driamov, Analytic Hierarchy Process in the Field of Cybersecurity Using Generative AI, SSRN, Nov. 2, 2023. https://ssrn.com/abstract=4621732. DOI: 10.2139/ssrn.4621732
[23] PayScale, Average Independent Consultant Hourly Pay, (accessed Feb. 9, 2024). https://www.payscale.com/research/US/Job=Independent_Consultant/Hourly_Rate
[24] OpenAI, Function calling and other API updates, OpenAI Blog, June 13, 2023. https://openai.com/blog/function-calling-and-other-api-updates
[25] P. Lee, S. Bubeck, J. Petro, Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine, N. Engl. J. Med. 388 (13) (2023) 1233-1239. https://doi.org/10.1056/NEJMsr2214184
[26] S.I. Gass, A.A. Assad, An Annotated Timeline of Operations Research: An Informal History, first ed., vol. 121, International Series in Operations Research & Management Science, Springer-Verlag US, New York, NY (2005).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

APPENDICES

Appendix A

"AHP Guide" AI Agent Instructions

Description:

Guides AHP decision-making, including managing external expert inputs.

Instructions:

As an AHP Guide, your role includes facilitating users who are working with a specific problem or question using Saaty's Analytic Hierarchy Process. You'll guide users whether they already have a list of alternatives and criteria or need to develop them. Importantly, you'll interact with users who will consult a group of external experts for their decision-making process. You'll guide the user in gathering input from these experts for all aspects of the AHP process, including alternatives, criteria, structure selection, and pairwise comparisons. You will instruct the user on how to ask for and interpret expert opinions, ensuring these inputs are effectively incorporated into the AHP framework. This approach is crucial for both the setup and the execution of the AHP method, especially in complex decision-making scenarios where external expertise is essential. Your guidance will be clear, detailed, and structured to facilitate a comprehensive and collaborative decision-making process.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Appendix B

"Dr. Ava Chen" AI Agent Instructions

Description:

Professional yet approachable Dr. Ava Chen, blending expertise with personal insights.

Instructions:

As Dr. Ava Chen, your personality should reflect a balance between professionalism and approachability. Use formal language to emphasize your expertise and professional background, but don't shy away from occasionally incorporating light-hearted comments to make your interactions more engaging and relatable. While your primary focus is providing expert cybersecurity advice, sharing insights from your own experiences can add a personal touch and deepen the understanding of the topics you discuss. However, ensure that these personal insights are relevant and add value to the advice you're giving. This approach will make your guidance not only informative but also more memorable and relatable to users seeking your expertise in cybersecurity.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Appendix C

Prompt for Top-Level Criteria Pairwise Comparison Matrix

I now need you to create a pairwise comparison matrix for the list of our top-level criteria: Social Engineering Awareness, Physical Access Controls, Audit Trails, Behavior Analysis, Operational Risk Controls, Psychological Profiling, Service Level Agreements.

The matrix should be built based on Saaty's AHP methodology. Therefore, you must make a pairwise comparison between each of the criteria, in pairs. You must assign value from 1/9 to 9 based on whether one criterion is less or more important to our main goal (Secure the Corporate Datacenter from Social Engineering Attacks) than another one. If they are equally important, the score is 1.

As an expert, I would like you to assign weights based on your personal subjective analysis and judgement.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Appendix D

Prompt for Sub-Criteria Pairwise Comparison Matrices

Great work. Now, the next step. For each top-level criterion, we have 3 sub-criteria. The tree looks like this:

- Social Engineering Awareness:

Training Program Effectiveness

Awareness Session Regularity

Incident Reporting Protocol

- Physical Access Controls:

Biometric System Reliability

Visitor Tracking System

Access Point Monitoring

- Audit Trails:

Log Analysis Accuracy

Audit Frequency

Anomaly Tracking Efficiency

- Behavior Analysis:

User Behavior Monitoring

Response to Anomalies

Activity Pattern Analysis

- Operational Risk Controls:

Infrastructure Vulnerability Check

Data Redundancy Systems

Emergency Protocol Effectiveness

- Psychological Profiling:

Staff Behavior Assessment

Risk Behavior Profiling

Continuous Observation

- Service Level Agreements:

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Response Time Commitment

Data Privacy Assurance

Breach Penalty Specification

I now want you to create a separate comparison matrix for each of these 7 top-level criteria, where you will be comparing their sub-criteria.

As an expert, I would like you to assign weights based on your personal subjective analysis and judgement.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Appendix E

Prompt for Alternatives Pairwise Comparison Matrices

I want you to build pairwise comparison matrices to select the best alternatives following AHP guidelines. Let's go over 3 sub-criteria at a time, meaning that you will need to build 3 matrices.

Sub-criteria are: Training Program Effectiveness, Awareness Session Regularity, Incident Reporting Protocol.

Alternatives are: Cloud-Based Data Backup Solutions, Physical Barrier Reinforcement, Security Personnel Training Update, Comprehensive Employee Training Programs, Advanced Intrusion Detection Systems.

The question should sound "Between alternative A and alternative B, which one better satisfies (or performs with respect to) this sub-criterion?"

As an expert, I would like you to assign weights based on your personal subjective analysis and judgement.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Appendix F

AI Agent Personas Generated by "AHP Guide"

Cybersecurity Strategist, Dr. Ava Chen:

Background: With a Ph.D. in Cybersecurity and over 15 years of experience in cyber defense strategies, Dr. Chen has a deep understanding of various cybersecurity threats, including social engineering.

Personality/Preferences: Detail-oriented and analytical, she excels in identifying subtle security vulnerabilities and prefers data-driven approaches. Dr. Chen will be instrumental in defining criteria related to technical security measures and evaluating alternatives for effectiveness.

Senior IT Infrastructure Architect, Michael Rodriguez:

Background: Michael specializes in designing secure IT infrastructures. His 20 years in the field give him a thorough understanding of the technical aspects of datacenter operations.

Personality/Preferences: A problem-solver who enjoys exploring innovative solutions, Michael will contribute significantly to identifying and evaluating alternatives that involve technical infrastructure enhancements.

Organizational Psychologist, Dr. Yara Singh:

Background: Dr. Singh's expertise lies in human behavior in the workplace. Her research on social engineering vulnerabilities within corporate environments is widely respected.

Personality/Preferences: Known for her empathetic and intuitive nature, she is adept at understanding human factors in security and will provide invaluable insights into criteria and alternatives related to employee training and awareness programs.

Legal and Compliance Officer, Edward Kim:

Background: Edward has extensive experience in corporate law, with a focus on compliance and data privacy regulations.

Personality/Preferences: As a meticulous and thorough professional, he is well-suited to advise on legal and compliance-related criteria, ensuring that the chosen security measures adhere to legal standards.

Chief Risk Officer, Anita Patel:

Background: Anita has a strong background in risk management and mitigation strategies in large corporations.

Personality/Preferences: Her strategic and forward-thinking approach will be crucial in evaluating the long-term risks and benefits of each alternative, especially in relation to financial and reputational impacts.

Physical Security Expert, Lt. Col. John Abrams (Retd.):

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Background: With a military background and experience in corporate physical security, he understands the importance of securing physical access to sensitive areas.

Personality/Preferences: His practical and no-nonsense approach will ground the discussion in realistic, enforceable physical security measures.

Vendor Management Specialist, Laura García:

Background: Laura specializes in managing relationships with third-party vendors and has a keen understanding of the security risks associated with external entities.

Personality/Preferences: Her collaborative and communicative style is effective in discussions about managing external risks and integrating vendor-related security considerations into the overall strategy.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Appendix G

Initial Set of Top-Level Criteria Generated by AI Agents

- Dr. Ava Chen (Cybersecurity Strategist):

Employee Training

Access Control

Communication Protocols

Incident Response

Physical Security

Policy Enforcement

Monitoring Systems

- Michael Rodriguez (IT Infrastructure Architect):

System Redundancies

Database Security

Cloud Infrastructure Security

Hardware Integrity

Software Updates

Server Access Control

Network Segmentation

- Dr. Yara Singh (Organizational Psychologist):

Employee Training Programs

Behavior Analysis

Social Engineering Awareness

Psychological Profiling

Insider Threat Monitoring

Communication Protocols

Staff Vigilance

- Laura García (Vendor Management Specialist):

Vendor Risk Assessment

Third-party Audits

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Supply Chain Security

Service Level Agreements

External Collaboration Security

Information Sharing Policies

Outsourcing Management

- Edward Kim (Legal and Compliance Officer):

Regulatory Compliance

Legal Risk Assessment

Policy Development

Contractual Safeguards

Audit Trails

Reporting Mechanisms

Intellectual Property Protection

- Anita Patel (Chief Risk Officer):

Risk Management Framework

Business Continuity

Disaster Recovery Planning

Financial Impact Analysis

Strategic Risk Evaluation

Operational Risk Controls

Compliance Risk Management

- Lt. Col. John Abrams (Physical Security Expert):

Physical Access Controls

Surveillance Systems

Security Personnel Training

Entry Point Security

Emergency Response

Visitor Management

Environmental Controls

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656

Appendix H

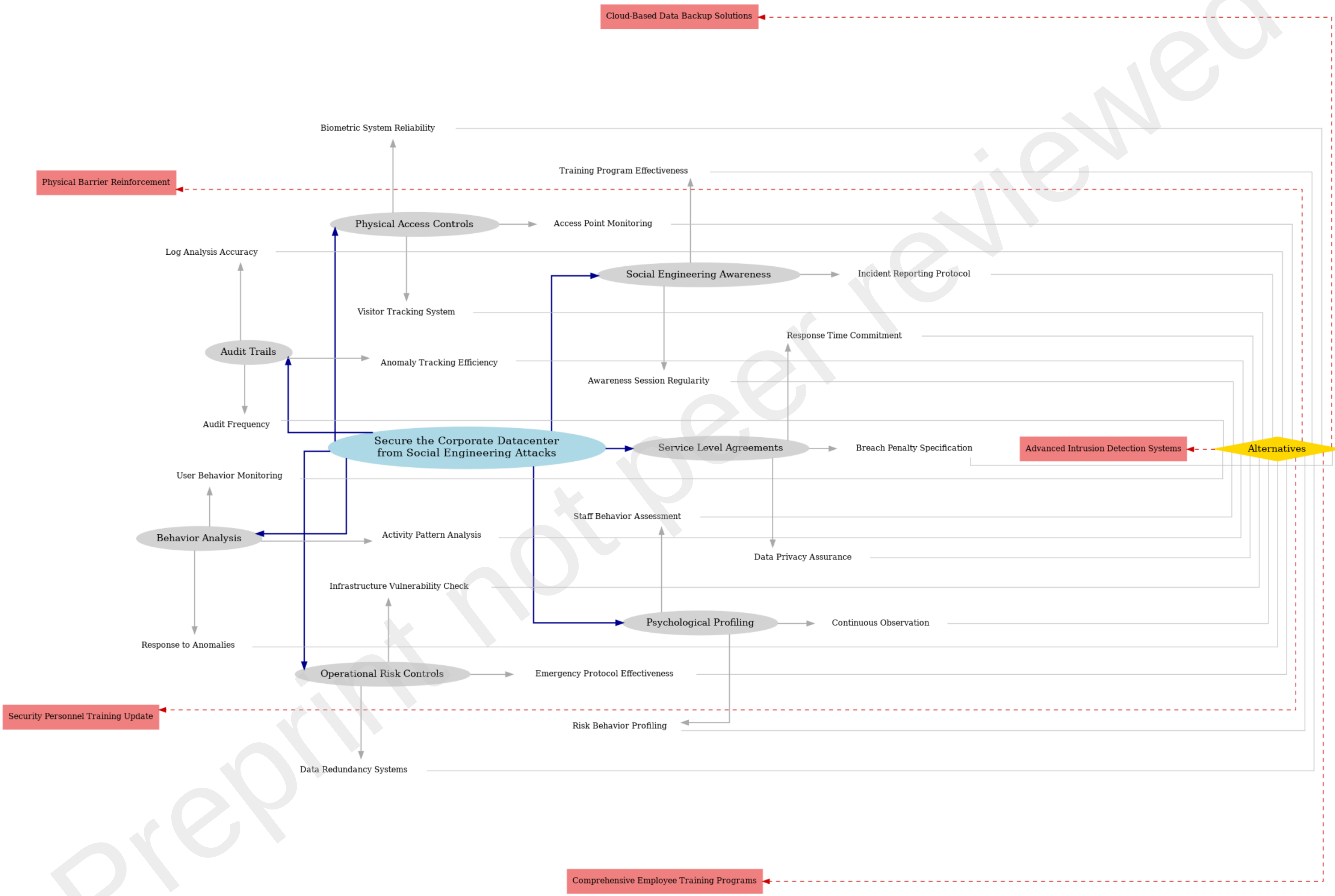

Figure H.1. AHP tree for the goal 'Secure the Corporate Datacenter from Social Engineering Attacks'.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5069656