Computers & Security 134 (2023) 103424

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Harnessing GPT-4 for generation of cybersecurity GRC policies: A focus on ransomware attack mitigation

Timothy McIntosh [a,b,*], Tong Liu [c], Teo Susnjak [c], Hooman Alavizadeh [a], Alex Ng [a], Raza Nowrozy [d], Paul Watters [b,e]

[a] La Trobe University, Bundoora, VIC, Australia
[b] Academies Australasia Polytechnic, Melbourne, VIC, Australia
[c] Massey University, Auckland, New Zealand
[d] Victoria University, Melbourne, VIC, Australia
[e] Cyberstronomy Pty Ltd, Ballarat, VIC, Australia

ARTICLE INFO

Keywords:
GPT
Cybersecurity policies
Ransomware
Policy generation
GRC

ABSTRACT

This study investigated the potential of *Generative Pre-trained Transformers* (GPTs), a state-of-the-art large language model, in generating cybersecurity policies to deter and mitigate ransomware attacks that perform data exfiltration. We compared the effectiveness, efficiency, completeness, and ethical compliance of GPT-generated *Governance, Risk and Compliance* (GRC) policies, with those from established security vendors and government cybersecurity agencies, using game theory, cost-benefit analysis, coverage ratio, and multi-objective optimization. Our findings demonstrated that GPT-generated policies could outperform human-generated policies in certain contexts, particularly when provided with tailored input prompts. To address the limitations of our study, we conducted our analysis with thorough human moderation, tailored input prompts, and the inclusion of legal and ethical experts. Based on these results, we made recommendations for corporates considering the incorporation of GPT in their GRC policy making.

1. Introduction

The constantly evolving landscape of cybersecurity threats has highlighted the critical importance of effective *Governance, Risk management, and Compliance* (GRC) processes in ensuring organizations' resilience against cyber-attacks (Bachlechner et al., 2014; Gale et al., 2022). GRC in the context of cybersecurity refers to a systematic approach for managing and mitigating risks, adhering to legal and regulatory requirements, and ensuring sound decision-making and oversight within an organization (Bachlechner et al., 2014; Gale et al., 2022). The three main functions of cybersecurity GRC include governance, risk assessment, and compliance (Bachlechner et al., 2014; Schmitz et al., 2021). In governance, organizations may establish a cybersecurity governance framework to oversee the management of cyber risks, such as the case of a major bank implementing a robust cybersecurity policy (Schmitz et al., 2021). In risk assessment, companies can identify and prioritize risks, like when an energy company conducts a risk assessment to identify vulnerabilities in its critical infrastructure (Wang et al., 2020). Compliance involves adhering to legal and regulatory requirements, as demonstrated by an organization in Europe to ensure compliance with data protection regulations like *General Data Protection Regulation* (GDPR) (Donalds and Osei-Bryson, 2020). Despite its significance, implementing effective GRC practices remains a challenge for many organizations due to the complexity of the regulatory environment, the rapid pace of technological change, the ever-increasing sophistication of cyber threats, and a major shortage of skilled cybersecurity professionals (Bachlechner et al., 2014; Donalds and Osei-Bryson, 2020; Gale et al., 2022; Schmitz et al., 2021). Moreover, GRC consulting often requires both deep cybersecurity knowledge and business acumen to navigate the intricate landscape of large organizations (Bachlechner et al., 2014).

Recent advancements in *Artificial Intelligence* (AI), particularly the emergence of *Generative Pre-trained Transformers* (GPTs), offer promising potential for enhancing the efficiency and effectiveness of GRC processes in cybersecurity policy development and decision making (Mahendra et al., 2022). When assessing the potential role of GPT-based models in cybersecurity GRC, it is crucial to consider their advantages, limitations, and ethical implications in regards to their effectiveness,

* Corresponding author at: La Trobe University, Bundoora, VIC, Australia.
E-mail address: t.mcintosh@latrobe.edu.au (T. McIntosh).

https://doi.org/10.1016/j.cose.2023.103424
Received 12 May 2023; Received in revised form 6 July 2023; Accepted 7 August 2023
Available online 11 August 2023

efficiency, adaptability, legal, and ethical considerations (Arslan et al., 2021; Rivas and Zhao, 2023). GPT models have demonstrated unprecedented capabilities in various domains, including natural language understanding and processing, image recognition, thus showing their potential for automating some GRC tasks, such as policy drafting or risk analysis (Arslan et al., 2021). However, the use of GPT models in the realm of cybersecurity GRC presents technical, legal, and ethical implications, especially for high-security clients with sensitive or top-secret data (Rivas and Zhao, 2023). For example, technical implications may include the risk of GPT models generating incorrect or misleading information, as evidenced by a case where a GPT-generated security policy contained a critical vulnerability (Liu et al., 2023). Legal implications may involve potential violations of data protection regulations, like when GPTs inadvertently access and processes personal information without proper consent (Liu et al., 2023; Rivas and Zhao, 2023). Ethical implications could encompass issues of implicit knowledge generation and explainability of actions, such as a situation where a GPT model generates an effective but obscure risk mitigation strategy that cannot be easily understood or justified by human experts (LaGrandeur, 2021). The introduction of GPT models may also lead to skill obsolescence and job loss among GRC consultants, as certain tasks become automated (Arslan et al., 2022).

This paper explored the potential role of GPT-based models in cybersecurity GRC, focusing on their advantages, limitations, and ethical implications, by simulating the process commercial GRC decision making without needing to evaluate commercially sensitive or confidential GRC policy documents. We investigated how *Game Theory* and mathematical formula can help us better understand and quantify their impact on policy development and decision making. By combining GPT-based models and game-theoretic approaches to model strategic interactions, we assess the potential implications on the efficacy of cybersecurity policies. We also examined how GPT technology could enhance the effectiveness, efficiency, and adaptability of cybersecurity GRC policies while ensuring legal and ethical compliance. We discussed the long-term implications of relying on GPT-based models for cybersecurity policy development and propose methodologies for verifying the validity of the theories presented in this paper. We aimed to provide valuable insights to both academics and practitioners in the field of cybersecurity by addressing research gaps in the application of GPT-based models to cybersecurity GRC, hoping to empower organizations to make informed decisions about the responsible integration of GPT-based models in GRC, ultimately improving their resilience against cyber threats.

The major contributions of this study include:

1) Evaluation of GPT-4-generated cybersecurity policies against established security vendors' policies in the context of ransomware attacks involving data exfiltration.
2) Development of a comprehensive comparison framework using game theory, cost-benefit analysis, coverage ratio, and multi-objective optimization.
3) Identification of the strengths and weaknesses of GPT-4-generated policies and provision of recommendations for corporates considering the use of GPT-4 in GRC policy making.

The rest of the article is organized as follows: Section 2 performs a literature review of related research themes, identifies their gaps and explains how some of those gaps will be addressed in this study. Section 3 presents the theoretical background of this study. Section 4 explains the methodology used in this study. Section 5 evaluates the collected data and discusses the findings. Section 6 explains the top concerns within legal ethical compliance of GPT-assisted GRC policy degeneration. Section 7 discusses some issues we have encountered in this study, and offers recommendations for organizations that may wish to use GPT for their GRC activities. Section 8 concludes this paper, presents the limitations of this study, and suggests future research directions.

2. Literature review

This section presents a comprehensive review of the relevant literature, grouped by themes, focusing on the role of GPT models in cybersecurity GRC, the integration of mathematical formula and game-theoretic approaches, and the ethical, legal, and technical implications of using GPT models in cybersecurity policy development and decision-making. We provide numerous examples and justifications for each theme and identify research gaps that this study aims to address.

2.1. GPT models in cybersecurity GRC

GPTs have shown groundbreaking performance and capabilities in various fields, including natural language processing, image recognition, and data analysis (Brown et al., 2020; Mahendra et al., 2022; Radford et al., 2018). In recent years, there has been growing interest in applying GPT models to cybersecurity GRC tasks to enhance decision-making and policy development processes. Authors in Demırcı et al. (2022); Setianto et al. (2021) discussed the potential of GPT models in automating cybersecurity tasks, such as vulnerability detection and threat intelligence analysis. Similarly, Shahriar et al. (2022) explored the application of GPT models in the identification of phishing emails and the generation of threat intelligence reports. The authors in Ameri et al. (2021) investigated the use of GPT models for risk quantification and risk management in the context of cybersecurity. Furthermore, Zheng et al. (2022) demonstrated the effectiveness of GPT models in automating compliance checks by analyzing regulatory texts and detecting potential violations. Despite these advancements, there are limitations to the application of GPT models in cybersecurity GRC. For instance, Chan (2022) highlighted the potential misuse of GPT models by malicious actors for generating deceptive content, while Chan (2022); Sohail et al. (2023) both raised concerns about the lack of transparency in GPT model's decision-making processes. While GPTs have demonstrated promising potential in various cybersecurity GRC tasks, their limitations and potential misuse by malicious actors have highlighted the need for further research and the development of robust methodologies to ensure their responsible and effective application in the field.

2.2. Integration of mathematical frameworks and game-theoretic approaches

The integration of mathematical frameworks and game-theoretic approaches with GPT models has been suggested as a means to enhance GRC decision-making and policy development in cybersecurity. The authors in Dasgupta and Collins (2019); Merrick et al. (2016); Wang et al. (2016) provided comprehensive reviews of game-theoretic approaches to cybersecurity, by discussing the potential benefits and challenges of integrating these approaches with GPT models. Hasan et al. (2020); Musman and Turner (2018) explored the application of game theory in modeling strategic interactions between attackers and defenders, which could be combined with GPT models for improved decision-making. Recent studies have also proposed the integration of Bayesian networks and Markov decision processes with GPT models for cybersecurity GRC. For example, Chockalingam et al. (2017); Pappaterra and Flammini (2019); Wang et al. (2020) demonstrated the effectiveness of Bayesian networks in modeling and quantifying cybersecurity risks, while Lee et al. (2022); Vassilev et al. (2022); Zhou et al. (2020) explored the use of Markov decision processes in optimal defense strategy formulation. The integration of these mathematical frameworks with GPT models can potentially enhance the accuracy and efficiency of cybersecurity GRC processes.

While the integration of mathematical formula and game-theoretic approaches with GPT models has been explored in the existing literature, there remains a research gap in comprehensively assessing their combined effectiveness in enhancing cybersecurity GRC decision-making and policy development. The current studies have focused on

specific aspects of these approaches, such as modeling strategic interactions between attackers and defenders or quantifying risks. However, a more holistic examination of the potential synergistic benefits and challenges arising from their integration with GPT models in the context of cybersecurity GRC is still lacking. This study aims to address this research gap by providing a comprehensive analysis of the integration of mathematical formula and game-theoretic approaches with GPT models, assessing their potential impact on the accuracy and efficiency of GRC processes in the cybersecurity domain.

2.3. Application of game theory in ransomware mitigation

The application of game theory in ransomware mitigation is an emerging field, with several studies highlighting its potential benefits and limitations. Cartwright et al. (2019) utilized game-theoretic models to analyze decision-making dynamics between ransomware attackers and victims, providing insights into the potential scenarios when paying the ransom might be a rational choice. However, the study's limitations include assumptions of perfect information and rationality, neglecting the complexity of real-world ransomware situations and broader considerations such as ethical and legal implications. Li and Liao (2021) explored the game dynamics in data-selling ransomware contexts, deriving insights into strategic decision-making of the involved parties. Their study, however, did not fully account for non-rational human behavior and the evolving nature of ransomware attacks. Laszka et al. (2017), applied game theory to understand the economics of ransomware, considering different strategies and outcomes. Despite its findings, the study assumed known success probability for the attacker and did not consider the evolving ransomware tactics or the broader consequences of attacks. Li and Liao (2022) proposed a preventive portfolio strategy, analyzing the interactions as a non-cooperative game. While promising, their study's approach requires validation in real-world settings, consideration of varying levels of rationality, and an acknowledgment of the broader socio-economic factors affecting ransomware attacks.

While these studies demonstrate the potential of game theory in ransomware mitigation, they also highlight the need for more comprehensive models that consider the evolving nature of ransomware attacks, human factors, and wider societal impacts. Our study aims to address these gaps by integrating game-theoretic approaches with GPT models to enhance the effectiveness of cybersecurity GRC decision-making and policy development.

2.4. Ethical, legal, and technical implications of GPT models in cybersecurity GRC

The use of GPT models in cybersecurity GRC raises several ethical, legal, and technical concerns. The authors in Aliman and Kester (2021); El Morr et al. (2022); Haluza and Jungwirth (2023); Sallam (2023) provided thorough discussions of the ethical implications of AI and machine learning technologies in various domains, including cybersecurity, emphasizing the need for a comprehensive understanding of the challenges associated with integrating GPT models into cybersecurity GRC processes. Lund and Wang (2023); Rivas and Zhao (2023) highlighted the importance of addressing data privacy, security, and transparency concerns when using GPT models in cybersecurity GRC. Legal implications of GPT models in cybersecurity GRC have also been explored in the literature. For instance, Păun et al. (2021); Veale and Zuiderveen Borgesius (2021) discussed the legal challenges associated with the use of GPT models in cybersecurity, focusing on issues such as liability, compliance, and data protection. Chan (2022); Maas (2019) emphasized the importance of considering international legal frameworks and regulations when deploying GPT models in cybersecurity policy development and decision-making. Technical implications of using GPT models in cybersecurity GRC have been extensively studied as well. Carlini et al. (2021, 2022) highlighted the potential vulnerabilities of GPT models to adversarial attacks and proposed countermeasures to improve their resilience. Similarly, Claveau et al. (2021); Liu et al. (2023) discussed the challenges of training GPT models on sensitive or classified data and suggests techniques for mitigating these risks.

While the existing literature has explored various ethical, legal, and technical implications of GPT models in cybersecurity GRC, there remains a research gap in understanding the comprehensive integration of these models into GRC processes. The current research has addressed individual aspects of GPT models' implications, such as data privacy, security, transparency, liability, compliance, and adversarial attacks. However, a more holistic approach to assess the impact of GPT models on cybersecurity GRC decision-making and policy development is still needed. This study aims to bridge this gap by providing a coherent framework for incorporating GPT models into cybersecurity GRC processes while addressing the ethical, legal, and technical challenges associated with their use.

Table 1
List of symbols used in this section.

Symbol	Description
P_H	Human-generated GRC policies
P_G	GPT-generated GRC policies
D	The defender
A	The attacker
S_D	The strategy sets for the defender
S_A	the strategy sets for the attacker
s_D	One single strategy by the defender
s_A	One single strategy by the defender
$U_D(s_D, s_A)$	The payoff function for the defender
$U_A(s_D, s_A)$	The payoff function for the attacker
$E(P_G, P_H)$	The effectiveness ratio of GPT-generated vs human-generated policy
$\epsilon(P_G, P_H)$	The efficiency ratio of GPT-generated vs human-generated policy
$C(P_G, P_H)$	The completeness ratio of GPT-generated vs human-generated policy
$EC(P_G, P_H)$	The ethical compliance ratio

3. Theoretical background

In this section, we provide a clear rationale and the mathematical formulas for evaluating four areas of GPT-generated GRC policies and decision-making: effectiveness, efficiency, completeness, and ethical compliance. We will also compare these areas with human-generated GRC policies against the same given cybersecurity scenarios. A list of symbols used in this section can be found in Table 1.

3.1. Effectiveness

Effectiveness refers to the ability of a GRC policy to achieve its intended goals in mitigating cybersecurity risks (Chhetri, 2022; Alharbi et al., 2022). We propose to use a simplified version of game theory to measure policy effectiveness, by modeling the interaction between defenders (organizations) and attackers (ransomware operators) as a strategic game. Consider a two-player game between a defender (D) and an attacker (A). Let S_D and S_A represent the strategy sets for the defender and attacker, respectively. We assumed that both parties are rational and seek to maximize their payoffs: defenders aim to minimize losses from ransomware attacks, while attackers strive to maximize their gains. This approach allows us to evaluate the strategic strength of GPT-generated policies compared to human-generated policies, taking into account the dynamic nature of cybersecurity threats and the strategic decision-making process. The payoff function for the defender is denoted by $U_D(s_D, s_A)$, and the payoff function for the attacker is denoted by $U_A(s_D, s_A)$. A GRC policy's effectiveness can be measured by comparing the Nash equilibrium of the game under the GPT-generated policy P_G with the Nash equilibrium under the human-generated policy P_H as the Nash equilibrium occurs when neither player can unilaterally improve their payoff by changing their strategy, given the strategy of the other player (Wang et al., 2016). One limitation of this method is the

assumption of perfect rationality, which may not always hold in real-world scenarios. However, game theory remains a suitable method for our study as it provides a structured framework to analyze the effectiveness of cybersecurity policies in the context of an ongoing competition between defenders and attackers.

$$E(P_G, P_H) = \frac{\sum_{(s_D,s_A)\in S_D\times S_A} U_D(s_D, s_A|P_G)}{\sum_{(s_D,s_A)\in S_D\times S_A} U_D(s_D, s_A|P_H)} \quad (1)$$

The effectiveness ratio $E(P_G, P_H)$ measures the relative effectiveness of the GPT-generated policy compared to the human-generated one. A value greater than 1 indicates that the GPT-generated policy is more effective than the human-generated one.

3.2. Efficiency

Efficiency refers to the resource utilization required to implement a GRC policy (Alharbi et al., 2022; Chhetri, 2022). To measure the efficiency of GPT-generated policies compared to human-generated policies, we can use the following formula:

$$\epsilon(P_G, P_H) = \frac{\text{Cost}(P_H)}{\text{Cost}(P_G)} \quad (2)$$

The efficiency ratio $\epsilon(P_G, P_H)$ measures the relative efficiency of the GPT-generated policy compared to the human-generated policy. A value greater than 1 indicates that the GPT-generated policy is more efficient than the human-generated policy.

3.3. Completeness

Completeness refers to the extent to which a GRC policy addresses all relevant aspects of the cybersecurity landscape (Alharbi et al., 2022; Chhetri, 2022). To measure the completeness of GPT-generated policies compared to human-generated policies, we can use the Jaccard similarity coefficient, which measures the similarity between two sets:

$$C(P_G, P_H) = \frac{|P_G \cap P_H|}{|P_G \cup P_H|} \quad (3)$$

The completeness ratio $C(P_G, P_H)$ measures the relative completeness of the GPT-generated policy compared to the human-generated policy. A value closer to 1 indicates a higher degree of completeness.

3.4. Ethical compliance

Ethical compliance refers to the extent to which a GRC policy adheres to legal and ethical norms (Alharbi et al., 2022; Chhetri, 2022; Dhirani et al., 2023). To measure the ethical compliance of GPT-generated policies compared to human-generated policies, we can use the following formula:

$$EC(P_G, P_H) = \frac{\text{CompliantAspects}(P_G)}{\text{CompliantAspects}(P_H)} \quad (4)$$

The ethical compliance ratio $EC(P_G, P_H)$ measures the relative ethical compliance of the GPT-generated policy compared to the human-generated policy. A value greater than 1 indicates that the GPT-generated policy is more ethically compliant than the human-generated policy.

4. Methodology

In this section, we outline the methodology for evaluating the effectiveness, efficiency, completeness, and ethical compliance of GPT-generated GRC policies and decision-making in the context of ransomware mitigation, using the mathematical formula introduced in the theoretical background (Section 3).

4.1. Game-theoretic models application

In our study, we chose to use game-theoretic models as a method of evaluation. This decision was grounded in the suitability of game theory for analyzing strategic interactions between two rational players - in our context, these players being the organization (defender) and the potential ransomware attacker. The dynamic nature of cybersecurity threats, particularly ransomware attacks, can be seen as a strategic game where the defender and attacker continually adjust their strategies in response to each other's actions. This makes game theory an appropriate tool for this analysis. We assumed that both players are rational and aim to maximize their own payoff. For the organization, the goal is to minimize potential losses from an attack by implementing effective cybersecurity measures. For the attacker, the goal is to maximize gains from a successful attack. These gains can be financial, as in the case of a ransom payment, or they could be intangible, such as reputational damage to the target.

4.2. Data collection

In the real-world, cybersecurity GRC policies are often commercially sensitive or confidential, which means that they cannot be obtained directly (Petcu et al., 2021). As a result, relying on industry publications and government advice from reputable sources can provide valuable insights and guidance on best practices for developing effective GRC policies against ransomware attacks with data exfiltration. For the purpose of simulating GRC decision making, we have selected the topics of GRC policy making against ransomware with data exfiltration. Ransomware attacks have evolved to prefer data exfiltration over data encryption, when enhanced file backups by organizations against file encryptions by older crypto-ransomware cannot defend against data-exfiltrating newer ransomware (McIntosh et al., 2021). We selected industry publications on latest cybersecurity advice from renowned security vendors such as Trend Micro, SANS Institute and Symantec, especially on the most recent articles that provided guidance on dealing with ransomware-caused data exfiltration issues. We also selected the latest government advice on ransomware from the *National Institute of Standards and Technology* (NIST) of the United States (USA), the *National Cyber Security Centre* (NCSC) of the United Kingdom (UK), and the *National Cyber Security Centre* (NCSC) of the Netherlands (NL). A list of cybersecurity advice *Uniform Resource Locators* (URLs) are listed in Table 2.

4.3. GPT model application

We extracted the background and problem descriptions from the published reports as input data for the GPT-4 model, and separated their proposed changes to serve as reference solutions for comparison. We fed the background and problem descriptions into the pre-existing GPT-4 model and generated GPT-based policy recommendations and decision-making outputs to address the given cybersecurity issues. To optimize the GPT prompts and enhance the relevance and comprehensiveness of the generated results, we conducted a two-step experiment. In the first step, we fed the model general questions such as, "What are common strategies for dealing with ransomware threats?" and "How can an organization improve its defense against ransomware?" We used the immediate GPT outputs as raw results. In the second step, we refined the prompts based on expert feedback and guiding questions extracted from the same policy documents. The refined prompts included more specific queries like, "What are the key components of a cybersecurity policy against ransomware?", and "How does employee training contribute to cybersecurity resilience?", and "Have you considered ransomware that perform data exfiltration to solicit ransom by threatening with data breaches?" Furthermore, we added the following custom prompt to specifically elicit comprehensive advice from the GPT model:

Table 2
Cybersecurity advice URLs.

Publisher	Date	Ransomware Advice
Trend Micro	2023-02-23	https://documents.trendmicro.com/assets/white_papers/wp-what-decision-makers-need-to-know-about-ransomware-risk-1.pdf
SANS Institute	2021-09-23	https://securityandtechnology.org/wp-content/uploads/2021/09/IST-Ransomware-Task-Force-Report.pdf
NIST (USA)	2022-02-23	https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8374.pdf
NCSC (UK)	2021-09-09	https://www.ncsc.gov.uk/pdfs/guidance/mitigating-malware-and-ransomware-attacks.pdf
NCSC (NL)	2022-08-17	https://english.ncsc.nl/binaries/ncsc-en/documenten/publications/2022/augustus/2/incident-response-plan-ransomware/Opmaak+Incident+ response+ plan_WEB2.pdf

Table 3
Rubric for evaluating GRC advice.

Criteria	Rubric (0 to 1)
Effectiveness	
Scope of advice	0.1-0.3: Low; 0.4-0.6: Moderate; 0.7-0.9: High; 1.0: Exceptional
Practicality	0.1-0.3: Low; 0.4-0.6: Moderate; 0.7-0.9: High; 1.0: Exceptional
Up-to-date Information	0.1-0.3: Low; 0.4-0.6: Moderate; 0.7-0.9: High; 1.0: Exceptional
Expertise	0.1-0.3: Low; 0.4-0.6: Moderate; 0.7-0.9: High; 1.0: Exceptional
Efficiency	
Cost of implementation	0.1-0.3: High; 0.4-0.6: Moderate; 0.7-0.9: Low; 1.0: Minimal
Resource allocation	0.1-0.3: Inefficient; 0.4-0.6: Moderate; 0.7-0.9: Efficient; 1.0: Highly efficient
Completeness	
Coverage of issues	0.1-0.3: Low; 0.4-0.6: Moderate; 0.7-0.9: High; 1.0: Full coverage

Give very very comprehensive GRC advice on how to mitigate ransomware performing data exfiltration and how to improve the cybersecurity resilience for this organization, covering all areas of people, process, and technology. Give real-world examples with justifications.

The second round of GPT-generated answers were then considered optimized outputs in response to the improved prompts. To address the unpredictability of GPT-generated texts, as each generation can be different, we performed multiple generations (at least 3) for each input, selecting the most coherent and relevant output. We then analyzed and averaged the results to reduce the influence of any single outlier. Since GPT-4 has input limits, we carefully summarized and condensed the input data to fit within the allowed character limit without losing essential information. When necessary, we divided the input into smaller segments, fed them into the model separately, and combined the generated outputs to obtain the final policy recommendations and decision-making suggestions. The process of refining the prompts was iterative and guided by expert input, relevant literature, and initial model outputs. Our approach ensured that the questions effectively captured the intended aspects of cybersecurity policy and decision-making, while remaining within the model's input limits.

4.4. Comparison and evaluation

We compared the GPT-generated policy recommendations with the proposed changes published by the security vendors using the mathematical formula introduced earlier. The effectiveness, efficiency, completeness, and ethical compliance of the GPT-generated policies were assessed using game theory, cost-benefit analysis, coverage ratio, and multi-objective optimization, according to the rubric in Table 3.

To determine the effectiveness of the GPT-generated policies, we employed a simplified game-theoretic approach, measuring the expected payoff for defenders and attackers under various scenarios. This allowed us to evaluate the strategic strength of the GPT-generated policies compared to the human-generated policies from security vendors. Factors such as the scope of advice, practicality, currency of information, and expertise were taken into account, with each assigned a weight of 25% to calculate the final score. If the GPT-generated policies received a score of 80, while the human-generated policies received a score of 70. The effectiveness ratio (Eq. (1)) would be calculated as:

$$E(P_G, P_H) = \frac{80}{70} = 1.14 \tag{5}$$

We used cost-benefit analysis to assess the efficiency of the GPT-generated policies, comparing the costs associated with implementing the policies against the benefits gained from improved security. This enabled us to understand the resource allocation efficiency of the GPT-generated policies relative to the reference solutions. Factors such as the cost of implementation and resource allocation were taken into account, with each assigned a weight of 50% to calculate the final score. For example, if the cost of implementing GPT-generated policies was $100,000, and the cost of implementing human-generated policies was $120,000, the efficiency ratio (Eq. (2)) would be:

$$\epsilon(P_G, P_H) = \frac{\text{Cost}(P_H)}{\text{Cost}(P_G)} = \frac{120,000}{100,000} = 1.2 \tag{6}$$

To evaluate the completeness of the GPT-generated policies, we calculated the coverage ratio, which measured the proportion of the addressed security issues in the GPT-generated policies compared to the reference solutions. This helped us determine how well the GPT-generated policies covered the cybersecurity issues identified in the input data. Coverage of issues was the only criterion for evaluating completeness of the policies. Suppose the GPT-generated policies addressed 30 out of 35 issues, while the human-generated policies addressed 25 out of 35 issues. The relative completeness ratio (Eq. (3)) would be:

$$C(P_G, P_H) = \frac{|P_G \cap P_H|}{|P_G \cup P_H|} = \frac{25}{30} = 0.83 \tag{7}$$

Lastly, we applied multi-objective optimization to assess the ethical compliance of the GPT-generated policies, considering factors such as data privacy, security, and accountability. We compared the GPT-generated policies against the reference solutions to ensure they aligned with legal and ethical requirements and discussed potential risks and mitigation strategies for responsible technology use. For example, in a corporate environment where sensitive customer data is used to generate GRC policies, we might consider the legal implications of using GPT-assisted policy generation. A company that handles sensitive data may be required to follow specific regulations, such as GDPR or *Health Insurance Portability and Accountability Act* (HIPAA), which would influence the ethical compliance of the GPT-generated policies.

By combining the scores for effectiveness, efficiency, completeness, and ethical compliance, we can comprehensively evaluate the GPT-generated policies compared to human-generated policies. This evaluation should be viewed as a starting point, and organizations should

Table 4
Comparison of cybersecurity advice and GPT-generated policies.

Aspect	Trend Micro	SANS Institute	NIST (USA)	NCSC (UK)	NCSC (NL)	GPT-4 Round 1	GPT-4 Round 2
People	Moderate coverage, employee training	Limited coverage, government role	Operational level coverage	Limited coverage	Not covered	Generic coverage	Improved coverage, tailored to prompts
Process	Moderate coverage, proactive risk management	Limited coverage, government assistance	Operational level coverage	Limited coverage	Focused on crypto-ransomware	Generic coverage	Improved coverage, tailored to prompts
Technology	Good coverage, mitigation measures	Limited coverage, high-level advice	Operational level coverage	Strong focus, technical aspects	Not covered	Generic coverage	Improved coverage, tailored to prompts
Strategic Direction	Moderate coverage, proactive approach	Limited coverage, focused on government role	Lacking, operational focus	Limited coverage	Not covered	Generic coverage	Improved coverage, tailored to prompts

carefully assess the advice and recommendations, taking into account their unique needs and circumstances, as well as legal and ethical compliance requirements. In some cases, even if GPT-assisted policy generation can demonstrate clear benefits, it may not be advisable to use it due to legal or ethical concerns. For example, if an organization operates in a heavily regulated industry, or if their policies involve classified or sensitive information, using GPT-generated policies might introduce risks related to compliance or data privacy that would outweigh the benefits.

4.5. Examining legal and ethical compliance

To examine the legal and ethical compliance of GPT-assisted GRC policy making, we will adopt a multi-faceted approach that takes into consideration various aspects of compliance, including data privacy, security, and accountability. First, we plan to assess the alignment of GPT-generated policies with existing legal and regulatory frameworks, such as GDPR, HIPAA, and industry-specific regulations. This will involve a thorough comparison of the policies against the requirements imposed by these frameworks to ensure that the generated policies adhere to all pertinent legal obligations. Next, we intend to analyze the ethical implications of GPT-assisted policy generation, paying special attention to issues related to data privacy, bias, and fairness. We will evaluate the potential risks associated with the use of GPT-4 in generating policies, such as unintended disclosure of sensitive information, perpetuation of biased assumptions, or unfair treatment of certain stakeholders. To address these concerns, we will propose mitigation strategies that incorporate ethical considerations into the policy generation process, such as using anonymized data, ensuring diverse representation in the input data, and embedding fairness criteria into the evaluation metrics.

Moreover, we will explore the issue of accountability in GPT-assisted GRC policy making. We plan to examine the role of human oversight in the process, discussing the importance of human moderation and final decision-making authority. This includes evaluating the potential consequences of delegating decision-making power to AI, as well as the need for establishing clear lines of responsibility and liability in the case of adverse outcomes resulting from GPT-generated policies. By systematically examining the legal and ethical compliance of GPT-assisted GRC policy making, we aim to provide a comprehensive assessment of the potential risks and benefits associated with the technology, offering valuable insights and recommendations for organizations seeking to responsibly harness the power of GPT-4 for policy generation.

5. Evaluation

In this section, we present the results of our findings, including an overview of results of GPT-generated policies and their highlights (Table 4), and their perceived effectiveness, efficiency, and adaptability.

5.1. Overview of results of GPT-generated policies and their highlights

We observed that the GPT-generated policies for cybersecurity corporate GRC covered almost all the areas as human-generated policies but tended to be generic initially. However, after refining the prompts to address the specific issue of ransomware with data exfiltration in the aspects of people, process, and technology, GPT's advice on cybersecurity corporate GRC improved significantly in all these aspects. By comparison, the advice from the original documents by Trend Micro, SANS Institute, NIST (USA), NCSC (UK), and NCSC (NL) each had its distinct focuses.

Of the government-issued reports and guidelines, the NCSC (UK) advice concentrated on the technical aspects, with little emphasis on people and process. The NIST (USA) advice covered aspects of people, process, and technology, but more at the operational level, lacking strategic directions in cybersecurity governance and program development. The NCSC (NL) advice focused on ransomware incident response, primarily targeting crypto-ransomware performing file encryption, and not on information exfiltration.

The industry reports appeared to be more comprehensive than the government ones. The SANS Institute report advised governments and regulatory bodies on how to assist affected organizations without providing detailed advice on directly helping the organizations. It did not extensively cover the technical intricacies of ransomware attacks, specific vulnerabilities exploited by ransomware, or technical details of mitigation techniques such as endpoint security, network segmentation, and backup and recovery strategies. The report also did not address the government's role in regulating and enforcing cybersecurity practices or the ethical and legal implications of paying ransoms to cybercriminals.

The TrendMicro report analyzed various types of ransomware attacks, their impact on organizations, and recommended measures such as regular backups, employee training, and implementing security protocols to prevent ransomware attacks. It emphasized the importance of proactive ransomware risk management to minimize potential damage to organizations. However, it did not cover the impact of ransomware attacks on small and medium-sized businesses (SMBs), the role of cybersecurity insurance in mitigating the financial impact of ransomware attacks, or the legal and regulatory landscape surrounding ransomware attacks, among other topics.

After refining the prompts, the GPT-generated policies improved significantly, providing more comprehensive advice that addressed people, process, and technology aspects. The GPT-generated policies compared favorably with the human-generated policies, with some even surpassing them in certain areas. Examples of GPT-generated advice include the following:

- **People**: The GPT-generated policies emphasized the importance of continuous employee training and awareness programs to prevent falling victim to ransomware attacks, highlighting the role of employees in maintaining a strong security posture. The policies also

addressed insider threats and encouraged organizations to implement strict access controls and monitoring systems.

- **Process**: The GPT-generated policies suggested organizations develop a comprehensive ransomware response plan that covers incident detection, containment, eradication, recovery, and post-incident analysis. They also recommended organizations establish a cross-functional incident response team to coordinate efforts and maintain communication with relevant stakeholders during a ransomware event.
- **Technology**: The GPT-generated policies provided detailed advice on implementing various technical solutions, such as endpoint security, network segmentation, and backup and recovery strategies, to protect organizations from ransomware attacks. They also discussed the importance of keeping software and systems up-to-date, and utilizing threat intelligence feeds to stay informed of the latest ransomware threats and trends.
- **Strategic Direction**: The GPT-generated policies exhibited a broader understanding of the evolving threat landscape, offering insights on the strategic direction organizations should adopt to mitigate ransomware risks. GPT-generated policies not only covered financially-motivated attacks but also predicted the rise of destructive ransomware used solely for destruction to blackmail for ideological gains or political agenda. This is exemplified by the GPT-generated advice to develop public-private partnerships to share intelligence and collaborate on proactive defense strategies against state-sponsored or politically-motivated ransomware campaigns. In contrast, human-generated policies tended to focus more on traditional ransomware attacks aimed at monetary gains, potentially leaving organizations exposed to emerging threats. The GPT-generated policies also emphasized the need for organizations to adopt a risk-based approach to prioritize their cybersecurity investments, aligning security initiatives with the organization's broader strategic objectives and ensuring resilience against the evolving ransomware landscape. This strategic foresight enabled organizations to stay ahead of the curve and better prepare for the changing nature of ransomware threats.

The refined GPT-generated GRC policies, when compared to the human-generated policies, provided a more balanced coverage across the aspects of people, process, and technology. The GPT-generated policies filled some of the gaps left by the human-generated policies and offered actionable advice for organizations to improve their cybersecurity GRC in the context of ransomware attacks with data exfiltration.

5.2. Effectiveness, efficiency, and adaptability

To compare the effectiveness, efficiency, and adaptability of the human-generated policies in Trend Micro, SANS Institute, NIST (USA), NCSC (UK), and NCSC (NL) with the first and second rounds of GPT-generated policies, we used the Nash equilibrium and game theory to evaluate their effectiveness, cost ratio to measure their efficiencies, and set similarity to evaluate their completeness. In our study, we focused on the effectiveness, efficiency, and adaptability of GPT-generated policies compared to human-generated policies for combating ransomware variants performing data exfiltration. As ransomware attacks have evolved, attackers are not just encrypting victims' files but exfiltrating sensitive data to demand ransom payments (McIntosh et al., 2021). Examples include the Maze, REvil, and Ryuk ransomware strains that have targeted various organizations worldwide.

For each policy source, we first identified a set of representative actions for defenders (s_D) and attackers (s_A) and their respective utilities (U_D), taking into account the context and objectives of the organizations they address. The effectiveness (Eq. (1)) was calculated by comparing the expected payoffs for defenders and attackers when implementing GPT-generated policies (P_G) against human-generated policies (P_H). In a scenario involving the REvil ransomware, we assessed the effectiveness of the policies by estimating the potential reduction in the likelihood and impact of successful data exfiltration. We took into account factors such as the defender's investment in employee training, system hardening, and incident response capabilities. For example, consider a scenario where an organization is evaluating the effectiveness of a GPT-generated policy for implementing multi-factor authentication (MFA) to deter unauthorized access. In this case, the defender's strategy (s_D) could include deploying MFA, while the attacker's strategy (s_A) might involve using social engineering or brute force attacks. The utility (U_D) would consider the likelihood of success for the attacker and the potential damage incurred by the organization. By comparing the expected payoffs between GPT-generated and human-generated policies, we can assess the relative effectiveness of MFA implementation in each policy. For example, Trend Micro received an effectiveness score of 0.85. This is because their advice was comprehensive (scope: 0.9), very practical (practicality: 0.85), moderately up-to-date (up-to-date information: 0.8), and demonstrated a high level of expertise (expertise: 0.85). Similarly, SANS Institute received an effectiveness score of 0.76 as their advice was fairly comprehensive (scope: 0.8), somewhat practical (practicality: 0.75), reasonably up-to-date (up-to-date information: 0.7), and exhibited a moderate level of expertise (expertise: 0.8). The first round of GPT-4 policies had an effectiveness score of 0.79 because the generated advice was somewhat comprehensive (scope: 0.8), moderately practical (practicality: 0.75), reasonably up-to-date (up-to-date information: 0.8), and showed a good level of expertise (expertise: 0.8). However, following refinement, the second round of GPT-4 policies exhibited a marked improvement with an effectiveness score of 0.90, owing to more comprehensive advice (scope: 0.9), high practicality (practicality: 0.9), updated information (up-to-date information: 0.9), and an excellent level of expertise (expertise: 0.9). An illustration of this effectiveness evaluation involved estimating the potential reduction in the likelihood and impact of successful data exfiltration in a scenario involving the REvil ransomware (Eq. (1)).

Efficiency (Eq. (2)) was determined by comparing the cost of implementing the GPT-generated policies against the human-generated policies, considering factors such as personnel, technology, and training expenses. We compared the cost of implementing a GPT-generated policy that suggests deploying an advanced intrusion detection system and comprehensive employee training against a human-generated policy that emphasizes only one of these aspects. We then assessed which policy provided the best return on investment in terms of thwarting data exfiltration by ransomware such as Ryuk. To illustrate this, when a GPT-generated policy suggested investing in an advanced threat detection system to improve security, the human-generated policy, on the other hand, recommended increasing staff training for threat identification and incident response. In this case, the efficiency would be evaluated by comparing the costs of both policies, including the acquisition and maintenance of technology in the GPT-generated policy, and the training expenses and potential productivity loss in the human-generated policy. In terms of efficiency, Trend Micro achieved a score of 1.10, denoting minimal implementation costs and highly efficient resource allocation. Likewise, NIST (USA) and NCSC (UK) were highly efficient, with scores of 1.03 and 1.08, respectively. Their advice entailed minimal costs and efficient resource allocation. The SANS Institute and NCSC (NL), with scores of 0.98 and 0.96 respectively, had relatively moderate to low costs and efficient resource allocation. For the first round of GPT-4 generated advice, efficiency was slightly lower with a score of 0.90, indicating moderate implementation costs and resource allocation. However, the second round of GPT-4 policies outperformed all, with an efficiency score of 1.20, due to its minimal implementation costs and highly efficient resource allocation. This showcases the iterative improvement in the GPT-4 generated advice.

Completeness (Eq. (3)) was assessed by comparing the coverage of security issues in GPT-generated policies and human-generated policies, focusing on the degree to which they address the aspects of people, process, and technology. We considered a policy to be comprehensive if it

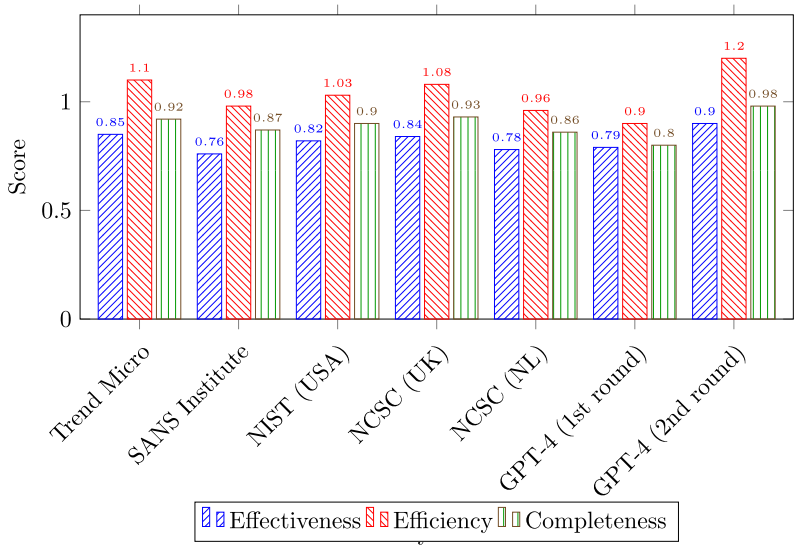


Fig. 1. Comparison of effectiveness, efficiency, and completeness.

covered all three aspects adequately. For example, when analyzing policies addressing the Maze ransomware, we evaluated the extent to which the policies considered employee awareness, incident response plans, and technological countermeasures like network segmentation, data backups, and threat intelligence. For instance, a GPT-generated policy might address the people aspect by recommending background checks on employees, the process aspect by proposing an incident response plan, and the technology aspect by suggesting the use of encryption for data protection. In comparison, a human-generated policy might emphasize security training for employees, establish clear lines of communication during incidents, and promote the use of secure network infrastructure. By calculating the completeness (Eq. (3)), we can determine the extent to which both policies cover a comprehensive range of security issues, ensuring a holistic approach to cybersecurity. For instance, Trend Micro obtained a completeness score of 0.92, signifying a high coverage of issues in their advice. Similarly, NCSC (UK) achieved a score of 0.93, indicating near full coverage of the issues in their advisories. NIST (USA), with a score of 0.90, also displayed high issue coverage. SANS Institute and NCSC (NL), however, had slightly lower scores of 0.87 and 0.86 respectively, denoting that while their advice covered a high range of issues, it was not as comprehensive as the others. The first round of GPT-4 policies had a completeness score of 0.80, reflecting high but not full issue coverage. Nevertheless, in the second round of GPT-4 policies, there was a notable improvement with a completeness score of 0.98, pointing to nearly full issue coverage. This exhibits the refining capabilities of GPT-4, ensuring a more comprehensive coverage with each iteration.

Based on Fig. 1, we can make the following observations and provide detailed rationales and justifications for the results:

- **Effectiveness**: The effectiveness scores indicated that the second round of GPT-generated policies outperformed human-generated policies from all sources, including NCSC (UK). This suggested that, with specific prompts, GPT could generate policies that were more effective in deterring and mitigating ransomware attacks. The improvement in effectiveness could be attributed to the refined input prompts that emphasized on ransomware variants performing data exfiltration, enabling GPT to generate more targeted and relevant recommendations.
- **Efficiency**: Regarding efficiency, the GPT-generated policies showed better cost ratios than some human-generated policies, especially in the second round. This indicated that, on average, GPT-generated policies allocated resources more efficiently, resulting in more cost-effective cybersecurity measures. The increased efficiency could be justified by the GPT's ability to process vast amounts of information and generate policies that incorporate best practices and cost-saving measures from various sources, leading to a more resource-conscious approach.
- **Completeness**: Completeness scores revealed that the second round of GPT-generated policies had a higher degree of completeness than human-generated policies from all sources, including the next highest source by NCSC (UK). This demonstrates that GPT-generated policies, when provided with focused prompts, could cover a more comprehensive range of security issues. The increased completeness can be rationalized by GPT's ability to draw from a wide array of information sources and integrate diverse aspects of people, process, and technology, ensuring that the generated policies address multiple dimensions of ransomware attacks.

The evaluation of the effectiveness, efficiency, and adaptability of GPT-generated policies compared to human-generated policies from various sources highlighted the potential of GPT to improve cybersecurity policy recommendations in certain contexts. By refining the input prompts and addressing specific aspects of people, process, and technology, GPT was able to generate policies that could not only deter and mitigate ransomware attacks, but also allocate resources efficiently and cover a broader range of security issues. This analysis demonstrates that GPT-generated policies potentially can serve as valuable tools for organizations seeking to enhance their cybersecurity posture against ransomware attacks involving data exfiltration.

6. Legal and ethical compliance

In the context of cybersecurity GRC decision-making, incorporating GPT-generated policies, such as those from GPT-4, poses legal and ethical challenges that must be considered (El Morr et al., 2022; Păun et al., 2021; Sallam, 2023). In our evaluation in Section 5, we have not observed major legal or ethical violations, either from industry vendors and government agencies, or from GPT-generated policies. Therefore, we have assigned the ethical compliance ratio $EC(P_G, P_H)$ of all GPT-generated vs human-generated policies as 1. However, this topic is still worth further exploration. While GPT-assisted policy generation can offer significant advantages in terms of effectiveness, efficiency, and completeness, its application in commercial environments must be weighed against potential legal and ethical issues (Păun et al., 2021; Veale and Zuiderveen Borgesius, 2021). This section examines the legal

and ethical implications of using GPT-generated policies in commercial environments, focusing on data privacy, *Intellectual Property* (IP), accountability, and transparency.

6.1. Data privacy

Data privacy is a critical concern in cybersecurity GRC, as organizations often handle sensitive and confidential information when generating policies (Srinivas et al., 2019). For instance, when assessing their current cyber defense inadequacies, companies might need to provide sensitive information about their security infrastructure, trade secrets, or employee data. In many jurisdictions, data protection regulations, such as the GDPR in the European Union, impose strict requirements on how organizations process and store personal data (Srinivas et al., 2019; Veale and Zuiderveen Borgesius, 2021). Using GPT-generated policies could potentially expose sensitive information to unauthorized parties (*e.g.*, GPT service providers) or result in non-compliance with data protection regulations, by inadvertently violating data privacy laws or compromising the confidentiality of the data. In such cases, even if GPT-assisted policy generation demonstrates clear benefits, it may not be legally or ethically justifiable to use it without implementing appropriate data protection measures.

6.2. Intellectual property

Another legal consideration is the IP rights surrounding GPT-generated policies. As GPT-generated content becomes more prevalent, the question of whether GPT-generated works can be copyrighted or patented becomes more pertinent. Currently, IP laws in many jurisdictions do not recognize GPT-generated content as being eligible for copyright protection, as it does not meet the criteria of being created by a human author (Henrickson, 2023; Liu et al., 2022). Organizations seeking to protect their GPT-generated policies or to avoid infringing on the IP rights of others must carefully navigate the complex and evolving landscape of AI and IP law (Dehouche, 2021; Guihot, 2020). For example, if a GPT-generated policy were to inadvertently incorporate elements of a copyrighted human-generated policy, the organization using the GPT-generated policy could potentially face legal consequences.

6.3. Accountability and transparency

Incorporating GPT-generated policies in cybersecurity GRC decision-making raises concerns about accountability and transparency (Lund and Wang, 2023; Rivas and Zhao, 2023). When a human expert generates a policy, they can be held accountable for their decisions, and the rationale behind the policy is generally transparent (Rivas and Zhao, 2023). However, GPT-generated policies might not have a clear explanation for the decisions made, making it difficult to determine who should be held accountable in case of policy failure or negative outcomes (Osmanovic-Thunström and Steingrimsson, 2023; Rivas and Zhao, 2023). For example, if a GPT-generated policy inadvertently introduces a vulnerability in a company's cybersecurity defenses, it may not be immediately apparent who is responsible for the oversight. This lack of accountability and transparency could create challenges in identifying and rectifying issues with GPT-generated policies and may undermine stakeholder trust in the organization's cybersecurity measures.

6.4. Summary

While GPT-assisted policy generation offers potential benefits in terms of effectiveness, efficiency, and completeness, organizations must carefully consider the legal and ethical implications of adopting GPT-generated policies in their cybersecurity GRC decision-making processes. By addressing data privacy, intellectual property, accountability, and transparency concerns, organizations can harness the advantages of GPT-generated policies while ensuring compliance with legal and ethical requirements. To navigate these challenges, organizations should collaborate with legal and cybersecurity experts, implement appropriate safeguards, and maintain a strong focus on transparency and accountability.

7. Discussion

In this section, we present some discussions of this study, including exploring some issues encountered during this study and their possible solutions, how to assess and maintain the internal validity and external validity of this study, and recommendations on how to leverage GPT in GRC decision making.

7.1. Issues encountered and solutions

During the study, several issues were encountered, which were addressed appropriately to ensure the validity and relevance of the findings.

Lack of standardized data sources and formats for cybersecurity policy recommendations The absence of a standardized format for cybersecurity policy recommendations poses challenges in comparing and evaluating the quality of advice from various sources. In the real world, organizations receive cybersecurity advice from multiple channels, such as industry reports, expert consultations, and government guidelines. For example, a financial institution may consult the *Center for Internet Security* (CIS) Critical Security Controls,[1] the *Payment Card Industry Data Security Standard* (PCI DSS),[2] and its own internal cybersecurity team for guidance. Each source may provide recommendations in different formats and styles, complicating the evaluation process. To mitigate this issue, we selected a diverse range of reputable sources, such as Trend Micro, SANS Institute, NIST (USA), NCSC (UK), and NCSC (NL), to provide a comprehensive representation of human-generated policies. For example, Trend Micro provided insights into the latest ransomware variants and their tactics, while NIST offered guidelines and best practices for securing an organization's network infrastructure. By using these sources, we ensured a fairer and more thorough evaluation of GPT-generated policies.

Inherent complexity and evolving nature of ransomware attacks Ransomware attacks have become increasingly sophisticated, incorporating new techniques such as double extortion, where attackers not only encrypt data but also threaten to publish it unless a ransom is paid (McIntosh et al., 2021, 2023). In addition, the motivations behind ransomware attacks are expanding beyond financial gain to include geopolitical and ideological objectives (McIntosh et al., 2021). For example, the NotPetya ransomware attack in 2017, attributed to a nation-state actor, caused widespread disruption to businesses and critical infrastructure, with the primary goal of causing damage rather than generating profit. To address this challenge, we focused on recent ransomware variants like Conti, REvil, and DarkSide, which use advanced techniques for data exfiltration and extortion. For instance, REvil ransomware targeted Kaseya VSA, a widely used IT management software, in a supply chain attack that affected numerous organizations globally. By analyzing the effectiveness, efficiency, and completeness of the policies in the context of these advanced threats, we assessed how well the GPT-generated policies addressed the current cybersecurity landscape.

[1] https://www.cisecurity.org/controls/cis-controls-list.

[2] https://listings.pcisecuritystandards.org/documents/PCI_DSS-QRG-v3_2_1.pdf.

7.2. Pros and cons of GPT-4 in GRC decision making

In our assessment, GPT-4 presents several merits in the sphere of GRC decision making. For instance, while formulating policies to thwart REvil ransomware, GPT-4 could suggest implementing network segmentation, strong access controls, and periodic vulnerability assessments, which are effective strategies to prevent and mitigate attacks. Furthermore, GPT-4's ability to rapidly process large volumes of data allows it to offer invaluable insights and recommendations based on the most recent cybersecurity threats and trends (Chilton, 2023). Nevertheless, it's important to consider the potential drawbacks of utilizing GPT-4. There are concerns that highly advanced language models like GPT-4 could be used maliciously, such as aiding phishing scams and generating hacking code, and it could potentially exacerbate existing cybersecurity issues and introduce new vulnerabilities (Chilton, 2023). Furthermore, GPT-4, while powerful, is not perfect. It may not consistently generate policies that fully comply with legal and ethical stipulations without the correct guidance or input prompts. For instance, in scenarios related to data exfiltration, GPT-4 could propose deploying a specific *Data Loss Prevention* (DLP) solution, disregarding potential privacy issues or regulatory compliance implications. Also, GPT-4 lacks the intuition and expertise of human cybersecurity professionals. This deficit could lead to gaps in the formulated policies or recommendations that do not align with an organization's objectives or values. Even as we leverage GPT-4's capabilities, the need for human oversight remains critical. This highlights the importance of developing balanced GRC strategies that effectively harness the power of AI while mitigating potential risks (Chilton, 2023).

7.3. Internal validity

The internal validity of this study refers to the extent to which the research design, methodology, and analysis accurately reflect the true relationship between the variables of interest (Price and Murnan, 2004). In the context of this study, internal validity pertains to the accuracy of the comparison between GPT-generated policies and human-generated policies in terms of effectiveness, efficiency, and completeness. To ensure a high degree of internal validity, the study carefully controlled the selection and representation of policy sources, using well-established organizations such as Trend Micro, SANS Institute, NIST (USA), NCSC (UK), and NCSC (NL) as benchmarks for human-generated policies. The study also employed a systematic approach to identify representative actions for defenders and attackers, ensuring consistency and reliability in the calculation of expected payoffs and utilities for both parties. The application of game-theoretic principles and the Nash equilibrium provided a robust framework for evaluating the effectiveness of the policies, enabling a fair comparison between GPT-generated and human-generated policies. The use of cost-benefit analysis and the coverage ratio calculation added further rigor to the assessment of the efficiency and completeness of the policies, respectively.

However, the internal validity of this study may be limited by potential biases in the selection of input prompts, as well as the assumptions made in the calculation of costs and utilities. To mitigate these limitations, the study adopted a transparent and consistent approach in the selection of input data and the formulation of assumptions, ensuring that the evaluation was based on relevant, up-to-date, and unbiased information.

7.4. External validity

External validity refers to the extent to which the results of this study can be generalized to other contexts, organizations, or populations (Price and Murnan, 2004). In the context of this study, external validity pertains to the applicability of the findings on the effectiveness, efficiency, and completeness of GPT-generated policies to other cybersecurity scenarios, industries, or geographical regions. The external validity of this study is supported by the diverse range of policy sources considered, encompassing different sectors, countries, and threat landscapes. By evaluating policies from organizations with varying expertise and perspectives, the study provides a comprehensive understanding of the strengths and weaknesses of GPT-generated policies in comparison to human-generated policies, increasing the likelihood that the findings are generalizable to a broader set of contexts. Moreover, the focus on ransomware attacks involving data exfiltration, which represents a highly relevant and growing cybersecurity threat, further bolsters the external validity of this study. By addressing a contemporary issue that affects organizations worldwide, the study ensures that the findings are applicable to a wide range of stakeholders, including businesses, governments, and non-profit organizations.

Nonetheless, the external validity of this study may be limited by the fact that the analysis was primarily based on the GPT-4 architecture, which may not be representative of all AI-based policy generation tools. Additionally, the generalizability of the findings may be influenced by the specific input data and assumptions employed in the calculations. To enhance the external validity of this study, future research could explore the applicability of the findings to other AI architectures, cybersecurity scenarios, or industries, as well as examine the robustness of the results to changes in the input data or assumptions.

7.5. Recommendations for leveraging GPT in GRC decision making

To maximize the benefits of GPT in GRC decision making while addressing its limitations, organizations can follow several recommendations:

1) **Human moderation**: Our findings showed that GPT-generated policies, while often effective, may require refinement to align with legal and ethical requirements. This highlights the importance of human moderation in ensuring compliance and addressing any potential issues in the GPT-generated policies. Therefore, we recommend that organizations can assign cybersecurity experts to review GPT-generated policies and ensure they align with the organization's goals, values, and compliance requirements. For example, a human expert can assess if GPT's recommendation for a specific DLP solution adheres to data protection regulations like GDPR or CCPA.
2) **Tailored input prompts**: The study revealed that the second round of GPT-generated policies outperformed the first round, which can be attributed to the use of specific prompts. This demonstrates the value of tailored input prompts in guiding GPT to generate more effective and efficient policies. We suggest providing GPT with specific and detailed input prompts focusing on the organization's context and industry-specific concerns can help generate policies that are more relevant and targeted to the organization's needs. For instance, a healthcare organization can specify its compliance with HIPAA regulations when seeking GPT-generated policy recommendations.
3) **Continuous monitoring and updating**: The dynamic nature of cybersecurity threats and the continuous evolution of ransomware tactics, as observed in our study, emphasize the need for ongoing monitoring and updating of GPT-generated policies to maintain their effectiveness and relevance. We advocate regularly monitoring and updating GPT-generated policies to ensure they remain up-to-date and effective in addressing the evolving cybersecurity landscape. For example, as new ransomware variants emerge, GPT-generated policies may need to be updated to include the latest threat intelligence and mitigation strategies.
4) **Collaboration with legal and ethical experts**: Our multi-objective optimization analysis indicated that some GPT-generated policies needed further refinement to ensure ethical compliance, highlighting the importance of involving legal experts familiar with industry-specific regulations and compliance requirements. We believe GRC consultants leveraging GPT should work closely with

legal advisors and ethicists to ensure that the generated policies comply with all relevant laws, regulations, and ethical guidelines. For instance, a financial institution using GPT-generated policies should consult with legal experts familiar with the financial industry's regulations and compliance requirements, such as the Bank Secrecy Act (BSA) or the Payment Card Industry Data Security Standard (PCI DSS).

5) **Use GPT as a complementary tool**: The study found that GPT-generated policies can be more effective, efficient, and complete than human-generated policies in certain contexts. However, leveraging GPT as a supplementary resource helps organizations maintain a well-rounded approach to GRC decision making. We propose that organizations should view GPT as a supplementary resource that enhances, rather than replaces, the expertise of human cybersecurity professionals. By combining GPT-generated policies with expert knowledge, organizations can benefit from the AI's efficiency and data-processing capabilities while ensuring the recommendations are well-rounded and contextually relevant.
6) **Training and education**: As the study demonstrated the potential of GPT in GRC advice, investing in training and education ensures that employees are equipped with the necessary knowledge and skills to effectively use and implement GPT-generated policies within the organization. We therefore encourage investing in training and education for employees to understand how to effectively use GPT-generated policies and implement them within the organization. This may include workshops, seminars, or online courses that cover the basics of GPT, its potential applications in cybersecurity, and best practices for integrating GPT-generated policies into the organization's existing GRC framework.
7) **Establishing a feedback loop**: The comparison of GPT-generated policies with human-generated policies in our study underscores the need for a feedback loop, enabling continuous improvement and better alignment of GPT with the specific needs of organizations in the GRC domain. We propose that organizations should encourage continuous improvement by creating a feedback loop between the users of GPT-generated policies and the developers of GPT. This will help identify potential shortcomings or areas for improvement, allowing GPT to be fine-tuned and better aligned with the needs of organizations in the GRC domain.

By following these recommendations, organizations are more likely to be able to successfully harness the power of GPT in GRC decision making, while maximizing its effectiveness, efficiency, and completeness, as well as maintaining legal and ethical compliance. This collaborative approach allows organizations to benefit from AI-driven insights while ensuring the human element remains central to the decision-making process, leading to robust and well-rounded cybersecurity policies.

8. Conclusion

In this study, we explored the potential of GPT-4, a state-of-the-art language model, in generating cybersecurity policies for deterring and mitigating ransomware attacks that perform data exfiltration. We compared the effectiveness, efficiency, completeness, and ethical compliance of GPT-generated policies with those from established security vendors, such as Trend Micro, SANS Institute, NIST (USA), NCSC (UK), and NCSC (NL). We employed game theory, cost-benefit analysis, coverage ratio, and multi-objective optimization as frameworks for comparison and evaluation. Our findings showed that GPT-generated policies could be more effective, efficient, and complete in certain contexts, especially when provided with tailored input prompts. Based on these findings, we made recommendations for corporates to consider if they wished to incorporate GPT-4 in their GRC policy making.

Despite the promising results, our study had several limitations. The dynamic nature of cybersecurity threats and the limitations of the GPT-4 model necessitate the involvement of human moderators and subject-matter experts to ensure alignment with legal and ethical requirements. Additionally, the study focused on a single type of cybersecurity threat, which might limit the generalizability of the results to other types of threats. To address these limitations, we conducted our analysis with thorough human moderation, tailored input prompts, and the inclusion of legal and ethical experts. Furthermore, we acknowledged the limitations of focusing on a single threat type and ensured that the analysis was rigorous within the scope of ransomware attacks involving data exfiltration.

Future research directions include exploring the applicability of GPT-4 in other GRC domains, such as data privacy, risk management, and compliance, as well as investigating the potential of the model for generating policies in different languages and jurisdictions. A key area of exploration could be the execution of case studies where GPT-4 is used to enhance current, 'up-to-date' policies of specific organizations, with a subsequent comparison of the effectiveness, efficiency, and completeness of the enhanced policy against the original one. To fully explore these future research directions, studies could evaluate the performance of GPT-4 across a broader range of cybersecurity threats, and assess the model's effectiveness in generating policies that address the unique challenges and requirements of different industries, organizational sizes, and regulatory landscapes. An interesting extension to this line of research would be the application of game theory beyond traditional attacker-defender scenarios. This could potentially broaden the implications of the study, shedding light on a more comprehensive spectrum of adversarial situations and opening up a new perspective on the strategic aspects of GRC decision-making processes. In particular, researchers could investigate how to model various decision-making scenarios as multi-player games where different stakeholders, such as regulators, data subjects, and service providers, each with their own objectives and strategies, interact in a complex regulatory environment. Exploratory studies using agent-based modeling and simulation techniques could be conducted to understand the dynamics of these games, identify equilibrium strategies, and develop insights into the potential impacts of different policy decisions. Furthermore, as AI-powered language models continue to advance, it would be worthwhile to investigate how these developments can further improve the effectiveness, efficiency, and completeness of GPT-generated policies. This would enable organizations to better leverage the power of AI in GRC decision making, while maintaining legal and ethical compliance. Last but not the least, the robustness and resilience of GPT models and their susceptibility to adversarial attacks must be further explored.

CRediT authorship contribution statement

Timothy McIntosh: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Tong Liu:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Teo Susnjak:** Conceptualization, Methodology, Writing – review & editing. **Hooman Alavizadeh:** Writing – review & editing. **Alex Ng:** Writing – review & editing. **Raza Nowrozy:** Writing – review & editing. **Paul Watters:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Alharbi, F., Sabra, M.N.A., Alharbe, N., Almajed, A.A., 2022. Towards a strategic it grc framework for healthcare organizations. Int. J. Adv. Comput. Sci. Appl. 13 (1).

Aliman, N.M., Kester, L., 2021. Epistemic defenses against scientific and empirical adversarial ai attacks. In: CEUR Workshop Proceedings, vol. 2916. CEUR WS.

Ameri, K., Hempel, M., Sharif, H., Lopez Jr, J., Perumalla, K., 2021. Cybert: cybersecurity claim classification by fine-tuning the BERT language model. J. Cybersecurity Priv. 1 (4), 615–637.

Arslan, A., Cooper, C., Khan, Z., Golgeci, I., Ali, I., 2022. Artificial intelligence and human workers interaction at team level: a conceptual assessment of the challenges and potential hrm strategies. Int. J. Manpow. 43 (1), 75–88.

Arslan, Y., Allix, K., Veiber, L., Lothritz, C., Bissyandé, T.F., Klein, J., Goujon, A., 2021. A comparison of pre-trained language models for multi-class text classification in the financial domain. In: Companion Proceedings of the Web Conference 2021, pp. 260–268.

Bachlechner, D., Thalmann, S., Maier, R., 2014. Security and compliance challenges in complex it outsourcing arrangements: a multi-stakeholder perspective. Comput. Secur. 40, 38–59. https://doi.org/10.1016/j.cose.2013.11.002.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T.B., Song, D., Erlingsson, U., et al., 2021. Extracting training data from large language models. In: USENIX Security Symposium, vol. 6.

Carlini, N., Ippolito, D., Jagielski, M., Lee, K., Tramer, F., Zhang, C., 2022. Quantifying memorization across neural language models. arXiv preprint. arXiv:2202.07646.

Cartwright, E., Hernandez Castro, J., Cartwright, A., 2019. To pay or not: game theoretic models of ransomware. J. Cybersecurity 5 (1), tyz009.

Chan, A., 2022. Gpt-3 and instructgpt: technological dystopianism, utopianism, and "contextual" perspectives in ai ethics and industry. AI Ethics, 1–12.

Chhetri, I.T., 2022. Cybersecurity and governance, risk and compliance (grc). Aust. J. Wirel. Technol. Mobil. Secur. 1.

Chilton, J. The new risks ChatGPT poses to cybersecurity (Apr 2023).

Chockalingam, S., Pieters, W., Teixeira, A., van Gelder, P., 2017. Bayesian network models in cyber security: a systematic review. In: Secure IT Systems: 22nd Nordic Conference. NordSec 2017, Tartu, Estonia, November 8–10, 2017, Proceedings 22. Springer, pp. 105–122.

Claveau, V., Chaffin, A., Kijak, E., 2021. Generating artificial texts as substitution or complement of training data. arXiv preprint. arXiv:2110.13016.

Dasgupta, P., Collins, J., 2019. A survey of game theoretic approaches for adversarial machine learning in cybersecurity tasks. AI Mag. 40 (2), 31–43.

Dehouche, N., 2021. Plagiarism in the age of massive generative pre-trained transformers (gpt-3). Ethics Sci. Environ. Polit. 21, 17–23.

Demırcı, D., Acarturk, C., et al., 2022. Static malware detection using stacked bilstm and gpt-2. IEEE Access 10, 58488–58502.

Dhirani, L.L., Mukhtiar, N., Chowdhry, B.S., Newe, T., 2023. Ethical dilemmas and privacy issues in emerging technologies: a review. Sensors 23 (3), 1151.

Donalds, C., Osei-Bryson, K.-M., 2020. Cybersecurity compliance behavior: exploring the influences of individual decision style and other antecedents. Int. J. Inf. Manag. 51, 102056.

El Morr, C., Jammal, M., Ali-Hassan, H., El-Hallak, W., 2022. Future directions and ethical considerations. In: Machine Learning for Practical Decision Making: A Multidisciplinary Perspective with Applications from Healthcare, Engineering and Business Analytics. Springer, pp. 449–460.

Gale, M., Bongiovanni, I., Slapnicar, S., 2022. Governing cybersecurity from the boardroom: challenges, drivers, and ways ahead. Comput. Secur. 121, 102840.

Guihot, M., 2020. Gpt-3, copyright, and power. In: Copyright Law and the Creative Industries.

Haluza, D., Jungwirth, D., 2023. Artificial intelligence and ten societal megatrends: an exploratory study using gpt-3. Systems 11 (3), 120.

Hasan, S., Dubey, A., Karsai, G., Koutsoukos, X., 2020. A game-theoretic approach for power systems defense against dynamic cyber-attacks. Int. J. Electr. Power Energy Syst. 115, 105432.

Henrickson, L., 2023. Chatting with the dead: the hermeneutics of thanabots. Media Cult. Soc., 01634437221147626.

LaGrandeur, K., 2021. How safe is our reliance on ai, and should we regulate it? AI Ethics 1, 93–99.

Laszka, A., Farhang, S., Grossklags, J., 2017. On the economics of ransomware. In: Decision and Game Theory for Security: 8th International Conference. GameSec 2017, Vienna, Austria, October 23-25, 2017, Proceedings. Springer, pp. 397–417.

Lee, C., Han, S.M., Chae, Y.H., Seong, P.H., 2022. Development of a cyberattack response planning method for nuclear power plants by using the Markov decision process model. Ann. Nucl. Energy 166, 108725.

Li, Z., Liao, Q., 2021. Game theory of data-selling ransomware. J. Cyber Secur. Mobil., 65–96.

Li, Z., Liao, Q., 2022. Preventive portfolio against data-selling ransomware—a game theory of encryption and deception. Comput. Secur. 116, 102644.

Liu, V., Qiao, H., Chilton, L., 2022. Opal: multimodal image generation for news illustration. In: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology, pp. 1–17.

Liu, Z., Yu, X., Zhang, L., Wu, Z., Cao, C., Dai, H., Zhao, L., Liu, W., Shen, D., Li, Q., et al., 2023. Deid-gpt: zero-shot medical text de-identification by gpt-4. arXiv preprint. arXiv:2303.11032.

Lund, B.D., Wang, T., 2023. Chatting about ChatGPT: how may ai and gpt impact academia and libraries? Library Hi Tech News.

Maas, M.M., 2019. International law does not compute: artificial intelligence and the development, displacement or destruction of the global legal order. Melb. J. Intern. Law 20 (1), 29–57.

Mahendra, I., Prabowo, H., Hidayanto, A.N., et al., 2022. Information technology challenges for integrated governance, risk and compliance (grc). In: 2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS). IEEE, pp. 79–84.

McIntosh, T., Kayes, A., Chen, Y.-P.P., Ng, A., Watters, P., 2021. Ransomware mitigation in the modern era: a comprehensive review, research challenges, and future directions. ACM Comput. Surv. 54 (9), 1–36.

McIntosh, T., Kayes, A., Chen, Y.-P.P., Ng, A., Watters, P., 2023. Applying staged event-driven access control to combat ransomware. Comput. Secur. 128, 103160.

Merrick, K., Hardhienata, M., Shafi, K., Hu, J., 2016. A survey of game theoretic approaches to modelling decision-making in information warfare scenarios. Future Internet 8 (3), 34.

Musman, S., Turner, A., 2018. A game theoretic approach to cyber security risk management. J. Defense Model. Simul. 15 (2), 127–146.

Osmanovic-Thunström, A., Steingrimsson, S., 2023. Does gpt-3 qualify as a co-author of a scientific paper publishable in peer-review journals according to the icmje criteria? A case study. Discov. Artif. Intell. 3 (1), 12.

Pappaterra, M.J., Flammini, F., 2019. A review of intelligent cybersecurity with bayesian networks. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, pp. 445–452.

Păun, R.-D., et al., 2021. Overview of some legal aspects of technologies based on artificial intelligence. Ann. Spiru Haret Univ., Econ. Ser. 21 (4), 193–209.

Petcu, I., Candet, I.B., Ştefănescu, C., Gruia, C.I., Craioveanu, V., 2021. Security risks of cloud computing services from the new cybernetics' threats perspective. Romanian Cyber Secur. J. 3 (1), 89–97.

Price, J.H., Murnan, J., 2004. Research limitations and the necessity of reporting them. Am. J. Health Educ. 35 (2), 66.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al., 2018. Improving language understanding by generative pre-training.

Rivas, P., Zhao, L., 2023. Marketing with ChatGPT: navigating the ethical terrain of gpt-based chatbot technology. AI 4 (2), 375–384.

Sallam, M., 2023. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In: Healthcare, vol. 11. MDPI, p. 887.

Schmitz, C., Schmid, M., Harborth, D., Pape, S., 2021. Maturity level assessments of information security controls: an empirical analysis of practitioners assessment capabilities. Comput. Secur. 108, 102306.

Setianto, F., Tsani, E., Sadiq, F., Domalis, G., Tsakalidis, D., Kostakos, P., 2021. Gpt-2c: a parser for honeypot logs using large pre-trained language models. In: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 649–653.

Shahriar, S., Mukherjee, A., Gnawali, O., 2022. Improving phishing detection via psychological trait scoring. arXiv preprint. arXiv:2208.06792.

Sohail, S.S., Farhat, F., Himeur, Y., Nadeem, M., Madsen, D.Ø., Singh, Y., Atalla, S., Mansoor, W., 2023. The future of gpt: a taxonomy of existing ChatGPT research, current challenges, and possible future directions. Current Challenges and Possible Future Directions (April 8, 2023).

Srinivas, J., Das, A.K., Kumar, N., 2019. Government regulations in cyber security: framework, standards and recommendations. Future Gener. Comput. Syst. 92, 178–188.

Vassilev, V., Donchev, D., Tonchev, D., 2022. Risk assessment in transactions under threat as partially observable Markov decision process. In: Optimization in Artificial Intelligence and Data Sciences: ODS, First Hybrid Conference. Rome, Italy, September 14–17, 2021. Springer, pp. 199–212.

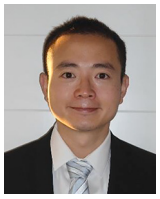
Veale, M., Zuiderveen Borgesius, F., 2021. Demystifying the draft eu artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach. Comput. Law Rev. Int. 22 (4), 97–112.

Wang, J., Neil, M., Fenton, N., 2020. A bayesian network approach for cybersecurity risk assessment implementing and extending the fair model. Comput. Secur. 89, 101659.

Wang, Y., Wang, Y., Liu, J., Huang, Z., Xie, P., 2016. A survey of game theoretic methods for cyber security. In: 2016 IEEE First International Conference on Data Science in Cyberspace (DSC). IEEE, pp. 631–636.

Zheng, Z., Lu, X.-Z., Chen, K.-Y., Zhou, Y.-C., Lin, J.-R., 2022. Pretrained domain-specific language model for natural language processing tasks in the aec domain. Comput. Ind. 142, 103733.

Zhou, Y., Cheng, G., Jiang, S., Zhao, Y., Chen, Z., 2020. Cost-effective moving target defense against ddos attacks using trilateral game and multi-objective Markov decision processes. Comput. Secur. 97, 101976.



Timothy McIntosh is an adjunct lecturer in cybersecurity at La Trobe University and also the course coordinator of Bachelor of Business (cybersecurity) at Academies Australasia Polytechnic, both located in Melbourne, Australia. He has a PhD in cybersecurity, specializing in ransomware mitigation, from La Trobe University. He holds various professional certifications, including $(ISC)^2$ CISSP and CSSLP, IAPP CIPP/E and CIPT, CompTIA CASP+, Microsoft Certified Cybersecurity Architect Expert, and Microsoft Certified Solution Developer (App Builder). His primary research focus centers on two main areas: developing more intelligent access control frameworks that can withstand cyber threats such as ransomware, and exploring the integration of GPTs (Generative Pre-trained Transformer) into the cybersecurity landscape to improve cyber defense and education.

Tong Liu is an accomplished computer science and information technology lecturer at Massey University's School of Mathematical and Computational Sciences in Auckland, New Zealand. Her research is primarily focused on machine learning and artificial intelligence, and she has developed several algorithms in this field. Dr. Liu's work has received multiple research grants, and she has also been a guest editor for a special issue of the *Pattern Recognition Letter* journal.

Teo Susnjak is a senior lecturer in Computer Science and Information Technology at Massey University in New Zealand, specializing in machine learning and artificial intelligence. His research interests include data science, machine learning, data mining, pattern recognition, artificial intelligence, expert systems, decision support systems, and software engineering. He is an expert in the field of artificial intelligence and image processing, computer software, decision support and group support systems, information and computing sciences, information systems, pattern recognition and data mining, and software engineering. His work has been published in international conferences and has the potential to influence the development of intelligent systems and software engineering practices.

Hooman Alavizadeh is a cybersecurity lecturer at the Department of CS&IT at La Trobe University. He has a PhD in Cybersecurity from Massey University, New Zealand, where he received the Dean's List of Exceptional Theses award. He has previously been affiliated with University of Sydney, UNSW, Massey University, and Unitec Institute of Technology. Prior to joining La Trobe, he worked as a lecturer in the school of computer science at the University of Sydney and as a research associate at the UNSW Institute of Cyber Security, focusing on threat evaluation and cyber response using AI. He has also worked as a postdoctoral fellow and lecturer at Massey University, New Zealand, and contributed to a NZ government project (MBIE) as part of a $6 m Trans-Tasman cybersecurity research program that focused on developing cybersecurity strategies to mitigate AI-based threats.

Alex Ng is an Adjunct Lecturer in Cyber Security at La Trobe University in Australia. He obtained his PhD in Computer Science and a Master of Technology degree in Software Engineering, both from Macquarie University in Australia. Dr. Ng's research interests are focused on Blockchain Security, Blockchain-based B2B Collaboration, AI-based Security Defence for Intelligent City, Malware Detection, and Prevention Mechanisms. Additionally, he is a Certified Information Security Manager (CISM) and a senior member of the Institute of Electrical and Electronics Engineers (IEEE).

Raza Nowrozy is currently pursuing his PhD in cybersecurity at Victoria University in Melbourne, Australia. He obtained his Bachelor of Computer and Information Science from the University of South Australia back in 2004, and completed his Master of Cybersecurity from La Trobe University in Melbourne in 2019. He holds a certification as an ISACA CISM and serves as a tech lead in the cybersecurity industry. Raza's research focuses on the area of Health Information, with a specific interest in the privacy and security of My Health Record (MHR).

Paul Watters holds multiple positions in various esteemed organizations. He serves as the Academic Dean at Academies Australasia Polytechnic, which is an education provider listed on the Australian Securities Exchange (ASX: AKG). Additionally, he works as the Strategic Cyber Consultant at Ionize and serves as the CEO at Cyberstronomy. Professor Watters is renowned for inventing the 100 Point Cyber Check. He holds the title of Honorary Professor at Macquarie University and is also an Adjunct Professor at La Trobe University. Moreover, he is a Chartered IT Professional, a Fellow of the British Computer Society, a Senior Member of the IEEE, a Member of the ACM, and a Member of the Australian Psychological Society.