Complex Systems Informatics and Modeling Quarterly (CSIMQ)
eISSN: 2255-9922
Published online by RTU Press, https://csimq-journals.rtu.lv
Article 257, Issue 46, March/April 2026, Pages 45–66
https://doi.org/10.7250/csimq.2026-46.03

Advancing a Speech Act-Based Model to Improve Future Quality of Information Security Policies Using Large Language Models

Fredrik Karlsson[1*], Shang Gao[1], John Krogstie[2], and Leila Aro-Sati[1]

[1]Department of Informatics, Örebro University, Örebro, 701 82 Örebro, Sweden
[2]Department of Computer Science, Norwegian University of Science and Technology, Trondheim, NO-7491, Norway

fredrik.karlsson@oru.se, shang.gao@oru.se, john.krogstie@ntnu.no, leila.aro-sati@oru.se

Abstract. Employee compliance with Information Security Policies (ISPs) depends on communicating clear and comprehensible content. However, existing research has shown that many ISPs are of poor communicative quality. Large Language Models (LLMs) could enhance ISPs if fine-tuned on high-quality data, but to do such fine-tuning requires a conceptual model for classifying the data and evaluating the resulting text. Therefore, as a step in this direction, the aim of this article is to develop a conceptual model of ISPs using Speech Act Theory as a theoretical lens to enable assessments of the communicative quality of ISPs. We used conceptual modeling and document analysis to develop the model based on 600 ISP statements from ten British National Health Service ISPs. We used selected parts from the SEQUAL framework to evaluate the model. The evaluation pointed to potential areas for improving the model's semantic, empirical, physical, and deontic qualities. By incorporating these improvements, the final class diagram contains 21 classes, six of which address ISP statement quality as speech acts.
Keywords: Information Security Policy, Cybersecurity Policy, Speech Act, Large Language Model.

1 Introduction

Organizations today face a wide range of information security risks from both internal and external sources, potentially leading to security breaches. Such breaches can severely impact an organization's reputation and finances [1] while also harming individuals [2], for instance, through leaked personal data on the Darknet. As a result, information security – the safeguarding of an organization's information assets [3] – is essential for maintaining trust and stability.

* Corresponding author

Reference: F. Karlsson, S. Gao, J. Krogstie, and L. Aro-Sati, "Advancing a Speech Act-Based Model to Improve Future Quality of Information Security Policies Using Large Language Models," Complex Systems Informatics and Modeling Quarterly, CSIMQ, no. 46, pp. 45–66, 2026. Available: https://doi.org/10.7250/csimq.2026-46.03

Additional information. Author ORCID iD: F. Karlsson – https://orcid.org/0000-0002-3265-7627, S. Gao – https://orcid.org/0000-0002-3722-6797, J. Krogstie – https://orcid.org/0000-0003-4830-1876, and L. Aro-Sati – https://orcid.org/0009-0001-9207-3236.
PII S225599222600257X. Received: 13 December 2025. Accepted: 9 March 2026. Available online: 30 April 2026.

Organizations can implement controls, i.e., measures addressing risks [4], within their business processes to strengthen information security. However, despite advancements in technical controls, they alone cannot guarantee information security. Both industry reports [5], [6] and research [7] show the critical role of human behavior. Employees with access to sensitive data can pose risks through intentional leaks, accidental disclosures, or actions enabling external threats, such as clicking on malicious links. For over three decades, human behavior has been a top information security threat [8], [9]. To mitigate this, organizations implement Information Security Policies (ISPs).

An ISP is commonly understood to be a "direction-driven document" that protects an organization's information assets [10]. The term ISP is used with differing meanings in the literature [11], and ISPs are typically structured across three levels [12], [13], [14]. At the highest level, strategic ISPs articulate top management's overall vision, scope, and guiding principles for the organization's information security efforts [15]. At the middle level, operational ISPs specify the procedures and rules that employees are expected to follow every day [16]. At the lowest level, technical ISPs address the security architecture and technical controls of information systems [14]. In this article, we focus on operational ISPs.

Even though ISPs contain procedures and rules that employees are to follow, many organizations struggle with poor ISP compliance among employees [17]. Of course, it is close at hand to blame employees for not following procedures, but Rostami, et al. [18] argue that ISP non-compliance often results from both employee-related factors and ISP design. Poorly designed ISPs, can hinder compliance, meaning employees should not always be held responsible. Studies stress the importance of designing clear, comprehensible ISPs [19], [20], [21], yet many still fall short for instance, when it comes to pinpointing actionable advice in a clear manner [22] and use vague or contradictory wordings [21], [23]. Consequently, there is a need for the content of ISPs to be formulated in a clearer way. While the ISO 27000 standard-series [24] and EU directives [25] provide guidance, most focus on high-level design aspects, such as relevant topics to include in policies [24]. However, there is still limited support on how to improve the messages conveyed by ISPs.

At the same time, advancements in natural language processing have been remarkable, introducing powerful new techniques. One such technology is Large Language Models (LLMs) [26], with the chatbot ChatGPT being a notable example [27]. This form of Artificial Intelligence (AI) is trained on vast amounts of textual data, enabling them to identify and analyze language patterns. As a result, LLMs can assist with tasks such as answering questions, summarizing text, and generating content, potentially improving text quality. However, as argued in a shorter version of this work presented at the 24th International Conference on Perspectives in Business Informatics Research [28], general LLMs are not specifically designed to enhance ISP content; they require fine-tuning or developing a Retrieval-Augmented Generation (RAG) solution using a high-quality dataset of ISP statements.

To this end, Speech Act Theory (SAT) [29] may offer a useful theoretical framework. This theory provides support for analyzing poorly written sentences by focusing on the intended function of an utterance such as instructing an employee rather than just its wording. When language is, for instance, vague or unclear, analyzing what the speaker is trying to do with the sentence can clarify its meaning and purpose, and enhance the possibilities for an LLM to improve it. SAT includes five types of speech acts (assertive, commissive, declarative, directive, expressive) that have different communicative functions, which means that the high-quality dataset of ISP statements should distinguish between them and include an evaluation of these statements based on quality criteria that are relevant to each type of speech act.

Against this backdrop, the aim of this article is to develop a conceptual model of ISPs using SAT as a theoretical lens to enable assessments of the communicative quality of ISPs. By structuring ISP statements based on their communicative functions, the model offers a systematic framework for assessing communicative quality and can act as a domain model. The ultimate goal is to create a high-quality dataset of ISP statements to refine LLMs, enhancing ISP content and

improving its effectiveness in guiding employees' information security behavior. We therefore pose the following research question: How can SAT be used to develop a conceptual model of ISPs to assess their communicative quality?

The remainder of the article is structured as follows. Section 2 presents related research in four parts: first, existing studies on ISP design as background; second, LLMs and ISPs; third, an introduction to SAT, which informed our modeling; and fourth, the SEQUAL framework, used to evaluate our conceptual model. Section 3 describes our research method. Section 4 presents our results, i.e., our conceptual model and evaluation findings. In Section 5, we discuss our findings, the limitations of the study and future research. Finally, the article ends with a short conclusion in Section 6.

As an extension of Karlsson, et al. [28], this article shares the same core results when it comes to the development and evaluation of the conceptual model presented in Section 4. The extension made related to this part is that we have integrated the results from the evaluation into a revised conceptual model, which means that the modeling work has taken a step forward compared to Karlsson, et al. [28]. Other extensions made are to the Related research, Research method and Discussion sections. When it comes to related research in Section 2, we have added more details to our description of existing research on ISP design (Section 2.1) and on LLMs and ISP design (Section 2.2). We also provide a more detailed description of the SAT in Section 2.3. We have extended the research method description in Section 3, where the method is now presented in four clear methodological steps. This clearer method makes it easier to assess the current study, but also to conduct similar studies in the future. Finally, we have expanded the Discussion section (Section 5), where we discuss more in-depth how our results contribute to existing research, the methodological decisions made during the study and the limitations of the results. This was made possible by providing the additional details about related research and the research method in Sections 2 and 3 respectively.

2 Related Work

2.1 Information Security Policy Design Research

ISP design can refer to both the product and the process, which are closely interconnected [21]. The intended design of an ISP influences the process, while the activities within the process impact the final ISP. For instance, an ISP design process that begins with a risk assessment [15] produces policy content focused on the organization's actual threats. However, our focus is on developing a conceptual model to classify ISP content, meaning we approach ISP design as a product. Several studies have examined ISP design quality, particularly the importance of clear and consistent instructions for employees [11], [20], [21], [23], [30]–[32]. However, few studies provide comprehensive conceptual models or frameworks for structuring and evaluating ISP statements.

Rostami et al. [32] proposed a conceptual model for tailoring ISPs to different organizational target groups using policy components. These components include three types of ISP statements: actionable advice, educational content, and general content. Of these three types, actionable advice is the most interesting because the concept is similar to directives in SAT (see Section 2.3 about the directive): "a demarcated part of an ISP, that instructs someone on a task to execute or not to execute regarding information security, and, in case of execution, how to carry out the task" [33]. However, their categorization of ISP statements is not based on SAT. In Rostami et al. [34], they outline three quality criteria for actionable advice: (1) it should use employees' work-related language, (2) specify clear responsibilities and consequences, and (3) be based on well-defined concepts. Rostami and Karlsson [22] have also introduced the Keyword Loss of Specificity metric to measure how consistently specific keywords are used to signal actionable advice in an ISP. This metric builds on Diver's [35] idea that certain keywords should be reserved for guiding employees' information security behavior. Taken together, this conceptual model and its related research do not offer a comprehensive framework with quality criteria for classifying ISP statements.

Beyond existing conceptual models, previous research on ISP design includes different criteria and metrics to evaluate ISPs [11], [21], [23]. Stahl et al. [23], based on an analysis of British National Health Service (NHS) ISPs, recommended using accessible language, clear terminology, and providing employees with specific, actionable guidance. Similarly, based on empirical insights from analyzing ISPs, Karlsson et al. [21] suggested eight quality criteria for ISPs. ISPs should (1) "contain congruent guidelines for actions that are well adapted to the current work practice", (2) "have a clear and congruent conceptual framework adapted to the current work practice", (3) "have a clear structure", (4) "have clear communicative objectives", (5) be free from goal conflicts, (6) "have clear and uniformed targets groups", (7) external policies should be translated to current work practices, (8) "be constitutively clear, clarifying responsibilities, social commitments and expectations". However, these high-level recommendations targeting ISPs as documents are not easily transferable to individual ISP statements.

Goel and Chengalur-Smith [11] established quantitative metrics to evaluate the communicative effectiveness of ISPs, focusing on three key characteristics: brevity, breadth, and clarity. Brevity assesses word repetitiveness within a document, with the premise that lower redundancy reduces unnecessary jargon. Breadth measures an ISP's comprehensiveness, aligning with the recommendation that policies should be as thorough as possible [11]. To quantify this, they compared the frequency of information security-related terms in ISPs against a master glossary, where a higher match rate indicated greater comprehensiveness. Lastly, clarity pertains to readability, for which they applied three established text analysis metrics: the Flesch Reading Ease Score, the Flesch-Kincaid Grade Level, and the Gunning Fog Index. However, they noted that these indices do not fully account for a reader's challenges in understanding the content. Although valuable, the metrics suggested by Goel and Chengalur-Smith [11] are not applicable to individual ISP statements.

In addition to existing research, there is practitioner-oriented literature that guides the design of ISPs, e.g., [24], [35]–[38]. This type of literature focuses primarily on the structure of ISPs and the topics they should cover [24]. For instance, Diver [35] provides guidelines on how to phrase directives, with the aim of making them easy for employees to understand. However, this literature does not provide a conceptual model of ISP content or a set of quality criteria for ISP statements.

2.2 Large Language Models and Information Security Policies

In recent years, LLMs have emerged as a powerful tool for analyzing and producing human-like text across various fields [39]. LLMs are trained by analyzing vast amounts of text, from books and articles to online content, and identifying complex statistical patterns in how words, sentences, and concepts relate to one another across different contexts [40]. This training process involves multiple stages, such as large-scale pretraining and task-specific fine-tuning, which enable the model to capture linguistic structure, contextual meaning, and probabilistic relationships rather than simply memorizing text. During training, the model repeatedly predicts missing or next words in example texts and adjusts its internal parameters to reduce errors. Over time, this process enables the model to capture relationships, structures, and regularities in language, allowing it to generate and interpret text in a way that appears meaningful and coherent. The ability to capture relationships, structures, and regularities makes LLMs very suitable tools for assessing and supporting the design of ISP statements.

Research has explored their application in government policy communication [41], access control policies [42], and digital asset privacy policies [43]. However, to our knowledge, no studies have specifically examined their use in designing or assessing ISP statements. Despite their potential, LLMs are susceptible to bias, inaccuracies, and misinformation [44], largely because they learn directly from the data used to train them. Shortcomings in the training corpus, such as errors, inconsistencies, or unrepresentative examples, can be reflected in the LLM's outputs. This poses a particular challenge for information security, where precision, clarity, and correctness are critical [11], [21]. To reduce the likelihood of generating or reinforcing such inaccuracies [45], it

is therefore pivotal to train LLMs on well-structured, high-quality ISP datasets. Curated and reliable ISP datasets help ensure that the model learns to identify different types of ISP statements and what is considered as high quality ISP statements, ultimately leading to more trustworthy and effective analyses.

2.3 Speech Act Theory

ISPs are not passive documents in organizations. They are used to perform communicative actions by the organizations. For instance, when an ISP says that "Your password must contain at least 10 characters and include a combination of uppercase letters, lowercase letters, numbers, and special symbols", it is not describing existing employee behavior; instead, it is creating a rule and imposing an obligation. Consequently, ISPs function as performative language that creates, for instance, obligations, permissions, and organizational structures. When designing an ISP, Chief Information Security Officers (CISOs), or those in comparable roles, speak on the organization's behalf. Although a human agent communicates the ISP, the act simultaneously counts as an action of the organization as a whole. In this sense, individuals operate through organizational roles, and any action taken by an employee carries a dual character: it is both a personal act and an institutional act [46].

SAT, first introduced by Austin [47] and later revised by Searle [48], is ideal for analyzing language from a performative perspective. The core idea of SAT is that communication constitutes a form of action, both when speaking and writing. Searle [48] divides speech acts into four interrelated sub-acts: (1) the utterance act, (2) the propositional act, (3) the illocutionary act, and (4) the perlocutionary act. The utterance act means expressing a sequence of words that form a meaningful ISP statement, for instance, the sentence above about the password design. The propositional act represents what the sentence is about. Here, it depicts a world in which certain password requirements exist. The illocutionary act reflects the speaker's intention: here, the organization is giving employees a directive, where these password requirements must be adhered to. The perlocutionary act concerns the impact the speaker (e.g., CISO) aims to have on the audience, namely encouraging employees to create strong passwords. It means that the illocutionary act comes with expectations on the employee, making them fundamental units of human communication [49]. Research suggests that the way speech acts are performed, including their level of directness, affects comprehension, as listeners must infer the intended meaning [50].

Searle [51] identifies five types of illocutionary acts:

- Assertives: committing the speaker to a statement's truth, i.e., the speaker asserts how things are in the world. For instance, "Phishing emails are a major threat to our organization".
- Commissives: committing the speaker to future actions, i.e., the speaker takes on some obligation. For instance, "Any unauthorized software will be removed from the laptop".
- Declaratives: changing reality through speech, i.e., bringing about a change in the external situation simply by being uttered (given that the speaker has the authority to execute said change). For instance, "This policy supersedes all previous information security directives".
- Directives: urging the listener to act, i.e., expressing the speaker's desire or intention to influence behavior. For instance, "You must not click on links or open attachments from unknown sources".
- Expressives: conveying the speaker's attitude about a situation, i.e., reflecting what is important to the speaker. For instance, "We value transparency in reporting security concerns".

The use of SAT in policy research is not new. Stahl et al. [23] used the speech act concept in their critical analysis of ISPs. Furthermore, SAT has been applied in areas such as educational policy analysis [52] and, more recently, in studies on communication intent and AI authorship, exploring the combined use of LLMs and SAT [53]. However, to our knowledge, no research has specifically applied LLMs to analyzing ISP statements using SAT categories.

2.4 SEQUAL – A Framework for Evaluating Conceptual Models

Model quality is an essential aspect of modeling, and this also applies to our conceptual model. The quality of models has been on the research agenda since the early nineties [54]. SEQUAL [55], [56] is a comprehensive framework for evaluating different types of conceptual models. The current version also takes heed of the fact that, these days, models can be developed both by humans and machines, or in a combination, and through pipelines and networks of models. Referring to Figure 1, the set of model-creating actors is denoted ***C***. It can be a combination of technical model creators (***TC***) (including AI tools, such as an LLM suggested in this article) and social model creators (***SC***) (humans and organizations) that collaborate in the development of the externalized model ***M***.

In SEQUAL, quality has been defined referring to the correspondence between statements belonging to the following sets:

- ***G***, the goal(s) of the modeling task.
- ***L***, the language extension, i.e., the set of all statements that are possible to make according to the rules of the modeling languages used.
- ***D***, the domain, i.e., the set of all statements that can be stated about the situation.
- ***M***, the externalized model itself expressed in a modeling language/notation.
- ***A***, the part of the model that can be accessed by one or more actors, actors being either persons or tools (including AI tools). The collected set of actors is called the audience of the model.
- ***K***, the explicit knowledge relevant to the domain of the audience.
- ***SI***, the social actor interpretation, i.e., the set of all statements that the people in the audience interpret that an externalized model consists of.
- ***TI***, the technical actor interpretation, i.e., the statements in the model as 'interpreted' by modeling tools and AI tools.

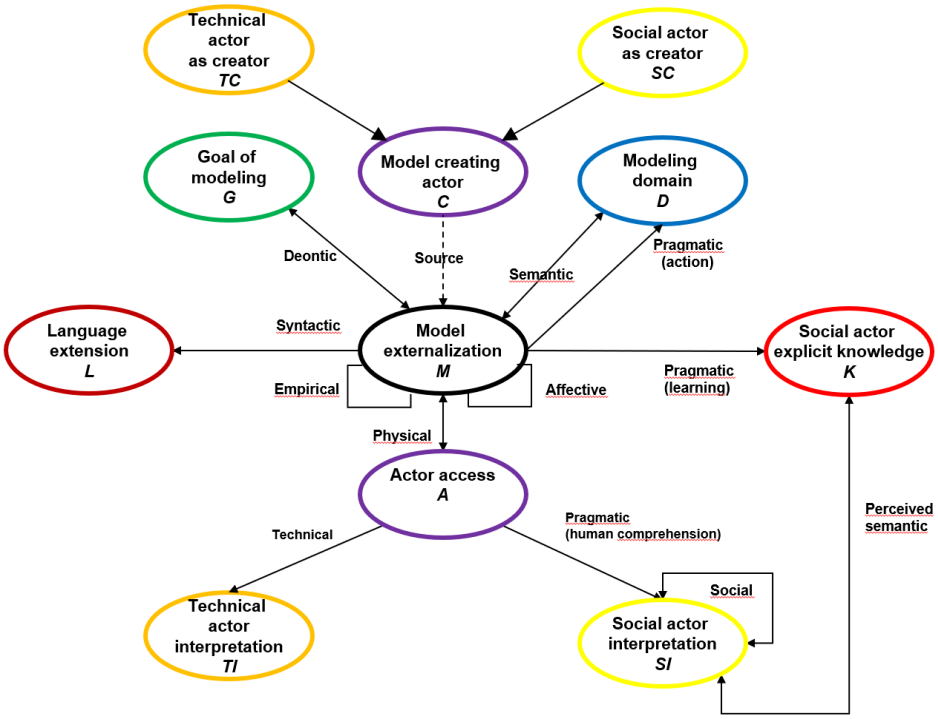

Figure 1. SEQUAL Framework

Central quality types, as illustrated in Figure 1, are:

- Physical quality addresses the basic quality goal that the externalized model M is available to the relevant actors. The available subset is denoted A.

- Empirical quality deals with comprehensibility when a model M is read by different (primarily social) actors. Patterns and readability (e.g., according to the Gunning Fog Index) in the labels used are discussed under empirical lexical quality.
- Syntactic quality is the correspondence between the model M and the language extension L. For syntactic quality, we can differentiate between adhering to the rules of the visual notation (syntactic notational quality), and complying of textual parts with the rules of the natural language used for writing the labels, comments, etc. (syntactic lexical quality).
- Semantic quality is the correspondence between the model M and the domain D. This includes both validity and completeness of the model.
- Perceived semantic quality is the similar correspondence between the social actor interpretation SI of a model M and the actors' current knowledge K of the domain D.
- Main aspect of pragmatic quality is the correspondence between the available part of the model M (i.e., A) and the actor interpretation (SI) of it. One differentiates between social pragmatic quality (to what extent people understand and can learn from and act based on the model) and technical pragmatic quality (to what extent tools can be made that can interpret and potentially act based on the model). Learning and action, either within the existing domain or resulting in domain change, are thus also part of pragmatic quality in addition to comprehension.
- Affective quality deals with the positive or negative impression that the model instills in social actors when interpreted.
- Social quality focuses on agreement among social actor's interpretations (I). Aspects of trust in the source of the model, that can be based on the variety of sources behind the model, influence social quality.
- The deontic quality of the model relates to the fact that all statements in the model M contribute to fulfilling the goals of modeling G, and that all the goals of modeling G are addressed through the model M using available resources (time, cost, competence etc.) available for the modeling task.

3 Research Method

The aim of this article is to develop a conceptual model of ISPs using SAT as a theoretical lens to enable assessments of the communicative quality of ISPs. To this end, we applied conceptual modeling [57] via Unified Modeling Language (UML) class diagrams and document analysis [58] of real-world ISPs. Document analysis is a natural choice since ISPs are often published in the form of documents. Although these two activities are distinct, and our presentation below is essentially sequential, we used them in an integrated manner to develop the conceptual model. The results from our conceptual modeling activity were evaluated using the SEQUAL framework.

3.1 Data Collection

In order to empirically ground the conceptual model, it was important that our document analysis and conceptual modeling work were based on real-world ISPs. Using authentic ISPs ensures that the resulting model reflects the complexity, variation, and practical communication patterns present in organizational information security documents. Following the approach taken by Stahl et al. [23], we selected the British NHS as our source context. To assemble our dataset, we executed a web search using the terms "information security policy" and "UK NHS". These searches led us to a range of publicly available ISP documents published by different NHS trusts and healthcare entities. We downloaded these documents and screened them to ensure they were operational ISPs. We then randomly selected ten ISPs for inclusion in our document analysis.

3.2 Document Analysis

Our document analysis focused on identifying different types of speech acts in our data. First, we read all the selected documents to ensure that they were ISPs and that all statements can be considered as ISP statements. The reading did not result in the need to remove any documents. Second, we used an LLM (ChatGPT 4o) to elicit ISP statements, i.e., sentences, from these ISPs. These statements could potentially represent different types of speech acts. However, at this stage, the generated ISP statements were not expected to reflect any predefined speech act categories. The primary goal was to produce a broad sample of ISP statements that could later be manually classified by the authors according to the various speech act categories. We used the following prompt, which is in four parts: "1) An information security policy statement is an utterance or a performative utterance about information security within an organization. 2) Elicit all information security policy statements in the uploaded document. 3) Quote the elicited information security policy statements in an Excel spreadsheet. 4) List one information security policy statement per row". For quality assurance, the prompt was tested on a small section of one of the ISPs to ensure that the results were consistent with sentences found in the actual policy.

In total, we elicited 3,765 ISP statements and added them to an Excel spreadsheet. Modeling text documents is resource intensive. Therefore, we needed to balance the available resources for modeling while ensuring the development of a stable conceptual model. We developed a macro in Excel that randomly selected 600 unique ISP statements for our analysis[†]. The random selection allowed us to get a spread of which potential speech acts they were among these ISP statements.

Third, we analyzed the selected subset of ISP statements using SAT by classifying them according to specific speech act types. To carry out this analysis, the first, second, and fourth authors each independently classified 200 of the randomly selected ISP statements. The classification followed Searle's taxonomy of speech acts [29], i.e., distinguishing between assertives, commissives, declaratives, directives, and expressives.

3.3 Conceptual Modeling

For our conceptual modeling, we chose to use UML class diagrams. Our modeling work was carried out during two workshops, in which the first, second, and fourth authors participated. We used our different and complementary skills to decrease potential bias and prevent the holistic fallacy in our study [59]. The protocol used during the workshops included (1) reviewing and resolving uncertainties in initial speech act classifications, (2) identifying all the speech act categories represented in the dataset, (3) mapping ISP statements to relevant organizational actors, (4) deriving quality criteria for each speech act category, and (5) developing a conceptual model using UML class diagrams.

During the modeling workshops, the three researchers began by reviewing the initial classifications of ISP statements. Any uncertainties or disagreements in the classifications were discussed in detail and resolved through consensus, ensuring a shared understanding of how each statement aligned with Searle's taxonomy [29]. As shown in the Result section, we identified all five types of speech acts defined by Searle [29] within the selected ISP documents. Building on this foundation, we analyzed how each type of ISP statement, interpreted as a speech act, relates to different actors within an organization. For instance, directives primarily target employees, while sanctions (a subclass of commissives, see Figure 2) involve commitments made by management. This actor-oriented analysis helped us understand the organizational roles and responsibilities implied by each speech act. We then applied SAT to explore the strengths and weaknesses associated with the various speech acts identified in the ISP statements. This analysis enabled us to identify communication issues (such as ambiguity and lack of actionability). From these insights, we developed our quality criteria.

[†] The ISPAI-CSIMQ dataset is publicly available: https://www.oru.se/english/research/research-environments/ent/ceris/ispai-csimq-dataset/

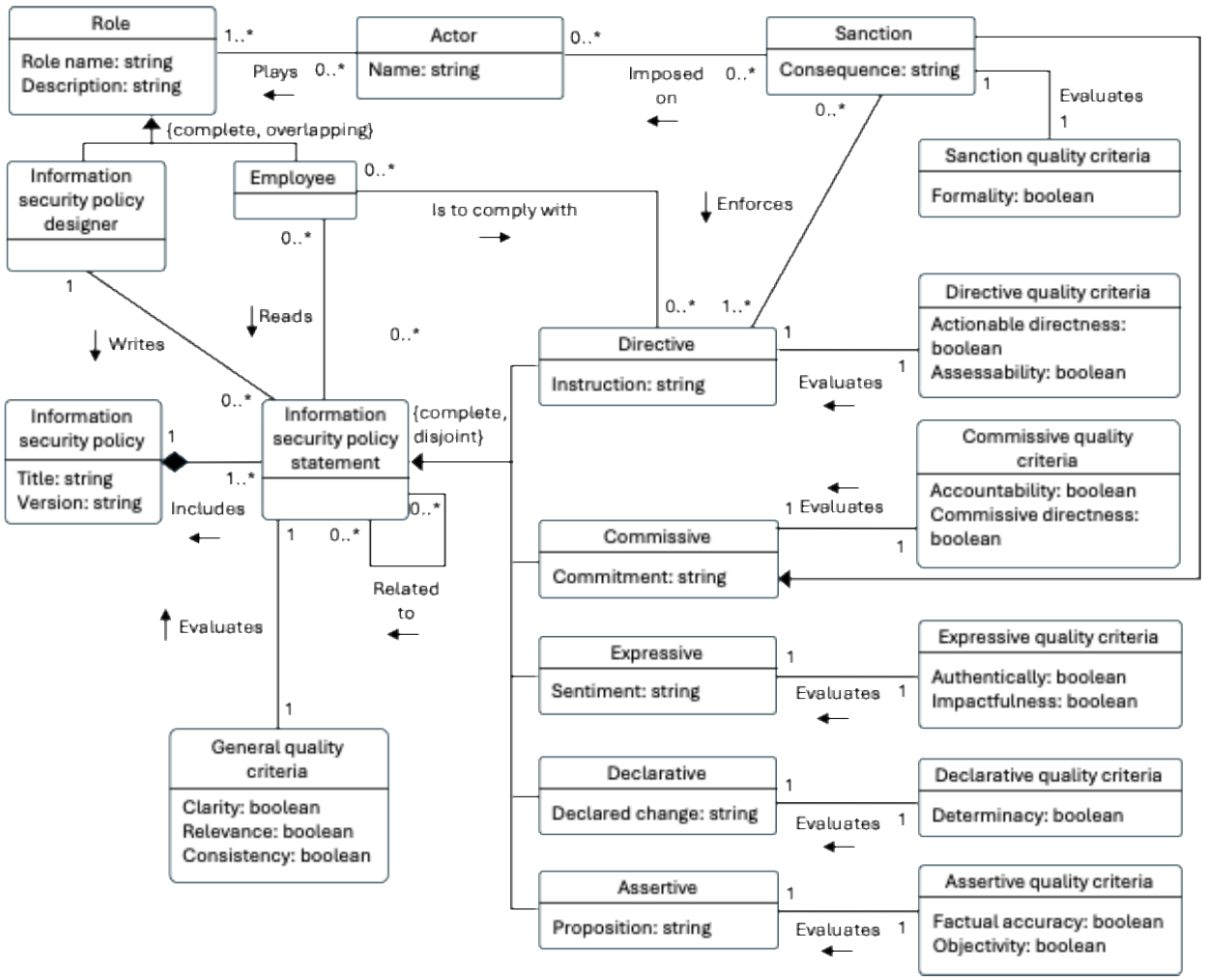

Figure 2. Information Security Policy Conceptual Model version 1.0 (presented first in Karlson, et al., 2025 [28])

Finally, we examined the scope of these criteria to determine whether each criterion applied specifically to a single speech act type or whether it could be generalized across multiple types. This step allowed us to distinguish between criteria that are speech-act-specific (e.g., actionable directiveness for directives in Figure 2) and those that have broader relevance (e.g., clarity, consistency, relevance).

3.4 Evaluation

The conceptual model was evaluated by the third author, who had *not* participated in the document analysis or the modeling workshops. Conducting the evaluation independently helped reduce the risk of bias and provided a more impartial, external assessment of the model. The model was evaluated by specifying the sets included in SEQUAL (see Section 2.4) and the following quality types: Physical, Empirical, Semantic, and Deontic. It is important to emphasize that no conceptual model can be optimized across all SEQUAL quality dimensions simultaneously. Trade-offs are inevitable. For instance, semantic completeness often means including edge cases, variant relationships, conditional logic, and domain-specific constraints. In contrast, empirical quality benefits from simpler, more abstract, and readable models. As in many modeling efforts, finding a balance between completeness and readability is essential.

For the evaluation, the third author used the conceptual model presented in Figure 2 along with a detailed list of definitions for all classes, attributes, and associations included in the model. This documentation is similar to what a future model user would have access to when using the model

to classify ISP statements. Finally, we used the evaluation results to revise the conceptual model. The revised version is presented in Figure 3 and Section 4.3.

4 Result

Our result is in two parts. The first part is the conceptual model, which is the result of modeling the content of ten ISPs using SAT as a lens. The second part is the evaluation of the conceptual model using the SEQUAL framework.

4.1 Information Security Policy Conceptual Model

The conceptual model of ISPs is shown in Figure 2 as a UML class diagram. The diagram consists of 19 classes, and between these classes, we found several named associations. Below, we do a structured walk-through of these classes and provide empirical examples of how they are grounded in our empirical data, i.e., the ISPs we used when modeling. In addition, we relate to the SAT and existing research on ISP design to show the theoretical grounding of the model.

4.1.1 Information Security Policy Core Parts

We start in the left center part of the figure, with the Information Security Policy class itself. As discussed in the Introduction, an ISP is supposed to guide employees on how to act securely in their daily work [16], and may be presented as documents or on organizations' web pages. In our case, they are represented by the ten ISP documents from the British NHS. From a SAT perspective, these policies include a set of ISP statements (speech acts), i.e., utterances that perform actions by expressing the speaker's intention [51], thereby trying to affect the reader's understanding and behavior. In our case, the utterances are made by ISP designer(s), a collective term for those involved in shaping the policy; an example is a CISO [60]. The ISP statements are intended for employees, such as physicians, nurses, and assistant nurses in the British NHS, who are supposed to be guided by reading them. Thus, ISP designers and employees are roles, i.e., functions played by actors, in the organization [32], where the actors are individuals who perform work tasks in the organization.

Drawing on SAT, as already described, Searle [29] provides five different types of speech acts with different so-called illocutionary points: directive, commissive, expressive, declarative, and assertive (synonym of representative [29]). The first type of speech act in our model is Directive. In an ISP, a directive is a statement that seeks to compel a particular kind of agent to take a particular action regarding information security. During our modeling work, we identified directives such as "You must include digits (0–9) in your password" (ISP#1), "Line Managers MUST ensure all IT accounts are removed or disabled when a staff member leaves the organization" (ISP#6), and "Use a fax cover sheet that contains a confidentiality statement" (ISP#3). In total, we identified 226 directives among the ISP statements analyzed.

The second type of speech act is Commissive, which, in an ISP, commits the speaker to do something in the future related to information security and/or the organization. Examples encountered during our modeling are: "Centrally stored information will be accessible from any healthcare provider location in England" (ISP#1), "Only authorized personnel who have an identified need are given access to restricted areas containing information systems such as the server room or a file storeroom" (ISP#2) and "New operational software will be quality assured" (ISP#3). We identified 123 commissive speech acts in our sample of ISP statements. A commissive speech act can further be refined as a sanction – a threat [50] – when it commits the speaker to imposing a consequence in the future if an employee does not comply with a directive. Thus, focusing on the forward-looking aspect, i.e., the commitment to a sanction is distinct from the situations where a sanction is carried out (e.g., an employee getting its credential withdrawn). In the latter case, an executed sanction has the force of a declarative, changing the social or legal

reality. In our sample, the following are examples of sanctions: “Offenders are liable to disciplinary action and civil and criminal prosecution” (ISP#6) and “Using Trust or NHS information systems for personal gain shall be deemed a disciplinary offence” (ISP#10).

The third type of speech act is Expressive. An expressive speech act in an ISP conveys the organization’s attitudes regarding information security practices and expectations. In the sample of speech acts we analyzed, we found three ISP statements that were expressive: “Our philosophy and commitment to care goes above and beyond our legal duty to enable us to provide high-quality services” (ISP#8), “It is essential that all of the CCG’s information systems are protected to an adequate level from business risks” (ISP#5) and “Each employee is responsible for the security of the information they use and maintaining its confidentiality, integrity and availability to the highest standard” (ISP#3).

The fourth type of speech act is Declarative, in which an ISP enacts or brings about a change in the status, rules, or conditions of information security by the very act of their declaration. In other words, when a declarative statement is issued in an ISP, it does not merely describe something about information security, but also creates, modifies, or formalizes a security condition by virtue of its declaration. In our sample of ISP statements, we only identified one declarative speech act: “The status of this document is FINAL” (ISP#6).

The fifth and final type of speech act is Assertive. In an ISP, an assertive speech act involves making statements that convey information and describe situations involving information security, the organization, or the world. For instance, among the analyzed ISP statements we found the following: “2FA stands for Two-Factor Authentication, which is the use of a second form of authentication, such as a password and a smart card” (ISP#6), “The current supplier is [name of the company] – the Trust will manage this contract centrally” (ISP#1), and “Where an overnight stay for work purposes is required the same principles apply” (ISP#5). In total, we classified 247 ISP statements as assertive in our analysis.

4.1.2 Quality Criteria for Information Security Policy Statements

The quality of the five types of ISP statements can be evaluated using a set of quality criteria, some of which are shared across all ISP statements and some of which are unique to specific types of speech acts. Starting with the ones shared across all ISP statements, they fall under the General quality criteria class. All ISP statements can be evaluated in relation to their clarity, relevance, and consistency. Clarity means that each ISP statement clearly conveys its intended message without ambiguity. Drawing on SAT, the illocutionary force [29], [51] – the speaker’s intention – should be explicit, leaving no room for misinterpretation. Relevance means each speech act must be pertinent to the specific context of the organization’s operations and the roles of the addressees. This ensures that the locutionary act [51], i.e., the actual content of the ISP statement, is meaningful and applicable to the addressees [22], [23]. Finally, consistency ensures that each ISP statement is consistent with other existing ISP statements. This internal congruence of the ISP prevents confusion across ISP statements [21], which could otherwise leave room for interpretations.

Turning our attention to the quality criteria specific to different types of speech acts, we start with the Directive quality criteria class, which contains two criteria: actionable directness and assessability. Actionable directness evaluates the precision in the directive, i.e., how well it guides the employee towards an intended outcome. Directives function through illocutionary force, attempting to influence, in our case, the employees to establish a world-to-word fit [29], i.e., changing reality to align with what is the desired state [50]. A directive with actionable directness ensures that it is clear how to achieve the desired state [32]; for instance, “Log off the computer when leaving by pressing Ctrl+Alt+Del”. Assessability in information security directives refers to the ability to evaluate whether a directive has been successfully executed. Drawing on Searle [29], directives inherently involve an expectation that the employee will carry out a specified action. Furthermore, the directive’s degree of imposition may vary [51]. Without the ability to evaluate, the essential condition (i.e., the employee’s obligation) is weakened because it is unclear whether

compliance has been achieved. For instance, a non-assessable directive is “Avoid using weak passwords”, if there are no measurable standards for determining compliance.

The commissive speech acts are evaluated using the Commissive quality criteria: accountability and commissive directness. Accountability in commissive speech acts means that each commitment must clearly define who is responsible for carrying out the promised action. Thus, we draw on the aspect that, through a commissive, the ISP designer binds someone to future action [29], [51]. For instance, “The Information Security Manager will determine what action is appropriate regarding system vulnerabilities” (ISP#1) links the commissive to an institutional role, in this case, the information security manager. Commissive directness focuses on the precision of the actions that the actor is committed to and the timeframe for execution, i.e., clarifying responsibilities [21]. In the example above, the commissive directness is weak since it does not commit the information security manager to how to determine if an action is appropriate, or a timeframe when it should be determined. The subset of commissive speech acts that are sanctions is evaluated using the Sanction evaluation criteria class, which includes one criterion: Formality. Formality is evaluated based on whether it is explicitly stated that the sanction is legally mandated or internally enforced. A formal sanction is derived from laws or external regulations (e.g., GDPR, HIPAA, or SOX), while an informal sanction does not carry legal weight but may result in internal penalties.

The quality of expressive speech acts is assessed using the quality criteria in the Expressive quality criteria class: authentically and impactfulness. Authentically means that the expressed attitude reflects the organization’s genuine attitude toward information security. This criterion focuses on the fact that there may be a difference between the psychological state [51] expressed by the ISP designer and what is communicated through actual actions. For an expressive to be valid, it must genuinely reflect the organization’s actual stance. Impactfulness means the speech act should effectively convey an attitude that motivates and engages the employee in active participation in information security practices. Thus, impactfulness relates to the illocutionary force of the speech acts [29], [51] – the strength with which a speech act conveys meaning to the employee.

Declarative speech acts are evaluated using the Declarative quality criteria class: determinacy. Determinacy focuses on whether the declaration makes the moment of change explicit. This relates to Searle’s concept of institutional declarations [29], where the act’s success depends on the authority of the speaker and the institutional context. For instance, “Effective March 1, 2025, access to company networks requires multi-factor authentication” is a determinate declaration because it clearly signals when the change takes effect, leaving no room for interpretation.

Finally, assertive speech acts are evaluated using the Assertive quality criteria class: factual accuracy and objectivity. Both criteria relate to the truth of the expressed proposition [29], [51], where statements must match reality. Factual accuracy assesses whether the assertive speech act accurately reflects the organization’s information security posture, the organization, and/or the world. Thus, this type of ISP statement is possible to assess as true or false [29]. Inaccurate descriptions of security measures mislead employees and undermine trust. For instance, “All company data is encrypted using industry-standard protocols” is accurate only if the organization implements such encryption. Objectivity means the speech act is to be presented impartially, without bias or subjective opinions. Thus, even if a statement in an ISP is factually correct, it lacks objectivity if it is framed in a way that introduces bias or persuasion.

4.2 Evaluation of the Conceptual Model

As stated in the Research Method section, the further evaluation of the Information Security Policy Conceptual Model version 1 was carried out by the third author. This author was not involved in the development of the model. SEQUAL has been used to structure the evaluation, starting with a description of the relevant sets. Below, we present the evaluation of the semantic, empirical, physical, and deontic quality types.

The third author's interpretation of the modeling work, which formed the backdrop for the evaluation, was as follows. The goal ***G1*** is: The model is to be used for fine-tuning LLMs or supporting a RAG-solution for producing better ISP statements using a general LLM. In addition, we have ***G2***; the model needs to act as a domain model that experts on ISPs and other social actors (***SA)*** agree upon, that represents good ISPs. The focus in this paper is ***G2,*** although we also look at some specifics related to what needs to be done with the model in the next step (when it is to be used for fine-tuning or to support a RAG solution, i.e., ***G1***). The (data)-model ***M*** in focus is the UML model for relevant aspects of ISPs, accompanied by a set of ISP statements from British NHS ISPs. These parts are detailed in the previous section, which is represented so as to adhere to this model. Thus, ***L*** is all the possible statements that can be expressed in UML. The actors and stakeholders are social actors creating the UML model ***(SC),*** but interpreters are both social actors ***(SI)*** (for refining and agreeing on the model), and technical actors ***(TI)*** in connection to interpret the model to achieve ***G2***. In the case of ***G2***, the technical actor is a basic modeling tool that can store the data model in a format accessible and interpretable by the fine-tuning mechanism.

4.2.1 Semantic Quality

The model should cover all types of operational ISPs and should not include aspects not relevant for ISPs, such as documents for building the context model, or ISPs that are no longer in effect. The model is based on currently available and up-to-date British NHS ISPs, which means the model has good correspondence with current documented information security practices in this industry. All 600 elicited ISP statements have been used to create the model, leaving no outstanding statements in the sampled data. The model addresses the quality of the ISPs through quality criteria that build on SAT. Thus, the quality criteria do not explicitly include SAT terminology. Speech acts include aspects such as "direction of fit" and illocutionary force, containing the elements degree of strength of illocutionary point, mode of achievement, propositional content, preparatory condition, sincerity conditions, and degree of strength of sincerity conditions. Representing these aspects would make the relation to SAT more explicit and make a more complete model, but one that can be harder to comprehend.

The model is more detailed regarding the associations between Directives, Sanctions, and Employees, compared to how the other speech acts are represented. The model could be extended by capturing associations between Employees and Commissives and Directives. In a commissive speech act, someone commits to do something (although it might be everyone). In a declarative speech act, there must be an actor in a specific role who makes the declaration. Currently, the model does not capture the time-period of validity of speech acts. This might be linked to versions of the ISP, but then information on from-to validity of the version should be represented. Usually, a decision date is given for when the ISP will be applied, and that it is valid until further notice, which would make it reasonable to add such an attribute to the Information Security Policy class. The organization being subject to the ISP should also be represented. Finally, from the point of view of completeness of the domain model, external laws and regulations that form the basis for some of the directives and sanctions can be represented [cf., 32], including their period of applicability. At the same time, these classes are considered to be peripheral when it comes to analyzing and improving the ISP content. These classes are therefore also considered to be of less relevance to fine-tuning an LLM ***(TI)*** for improving content, which mainly consists of ISP statements.

4.2.2 Empirical Quality

For ***G2***, the model should follow guidelines for graph-layout in standard UML-notation. For the conceptual model, most lines are straight and with short distance, which means the model overall performs well. A long association is found, which is a layout compromise to avoid crossing lines. The two bends on associations, Employee "is to comply with" Directive and Sanction "enforces"

Directive, could be avoided. The latter association could have been a straight line with a somewhat different positioning of the Sanction class. In general, it is easy to map labels to relationships, thus having good syntactic disjointness.

4.2.3 Physical Quality

The model must be available in a format understandable by ***C***. Currently, the model is available for ***SC*** and ***SA*** using SmartDraw native file format (and the available export formats), which is sufficient for ***G2*** (to act as a domain model that experts on ISPs and SA agree is a good representation of ISPs). Versioning is supported by keeping separate model files. To address ***G1*** (to be used for fine-tuning of LLMs or to support a RAG solution for producing better ISP statements) in the future, the model has to be made available in a format understandable to the fine-tuning or RAG solution mechanism in an LLM.

4.2.4 Deontic Quality

The conceptual model has two goals: ***G1*** (to be used for fine-tuning of LLMs or to support a RAG solution for producing better ISP statements) and ***G2*** (to act as a domain model that experts on ISPs and SA agree represents good ISPs). Goal ***G2*** is broader than goal ***G1***, which affects the boundaries of what should be included in the conceptual model. This is evident in the evaluation of semantic quality above. For instance, including external laws and directives in the model would bring the model closer to ***G2***. In the current version, the model is closer to fulfilling ***G1***.

4.3 Revised Conceptual Model

Figure 3 presents the revised conceptual model after implementing the evaluator's suggestions on how to improve the semantic and empirical qualities of the model.

First, we have added two more classes. The first class, Organization, captures the fact that ISPs exist in the context of organizations, although it is not the case that all organizations have an ISP. The second class, External regulations, captures the aspect that external laws, directives, and international and national standards may form the basis for some of the ISP statements that are presented in an ISP. It is important to stress that it is about the connections that the ISP designer has expressed in the ISP. They do not necessarily have to match the connections between, for instance, a law and an ISP statement that actually exist.

We have also modified the class Information Security Policy by adding the requested attribute Valid from, to show from when an ISP has become valid.

We have added a new association between Commissive and Role to capture that someone has been committed to do something. From a design perspective, this required adding another long association with intersecting lines to the model. Thus, this becomes a trade-off between semantic and empirical quality, where we have prioritized semantic completeness.

Regarding the need to add an association between Role and Declarative to capture the role of who declares, there would be an overlap with the role that writes an ISP statement; therefore, we have chosen not to include an additional association to capture this.

Finally, to improve the model's empirical quality, we have removed the bends on the two associations: Employee is to comply with Directive and Sanction enforces Directive.

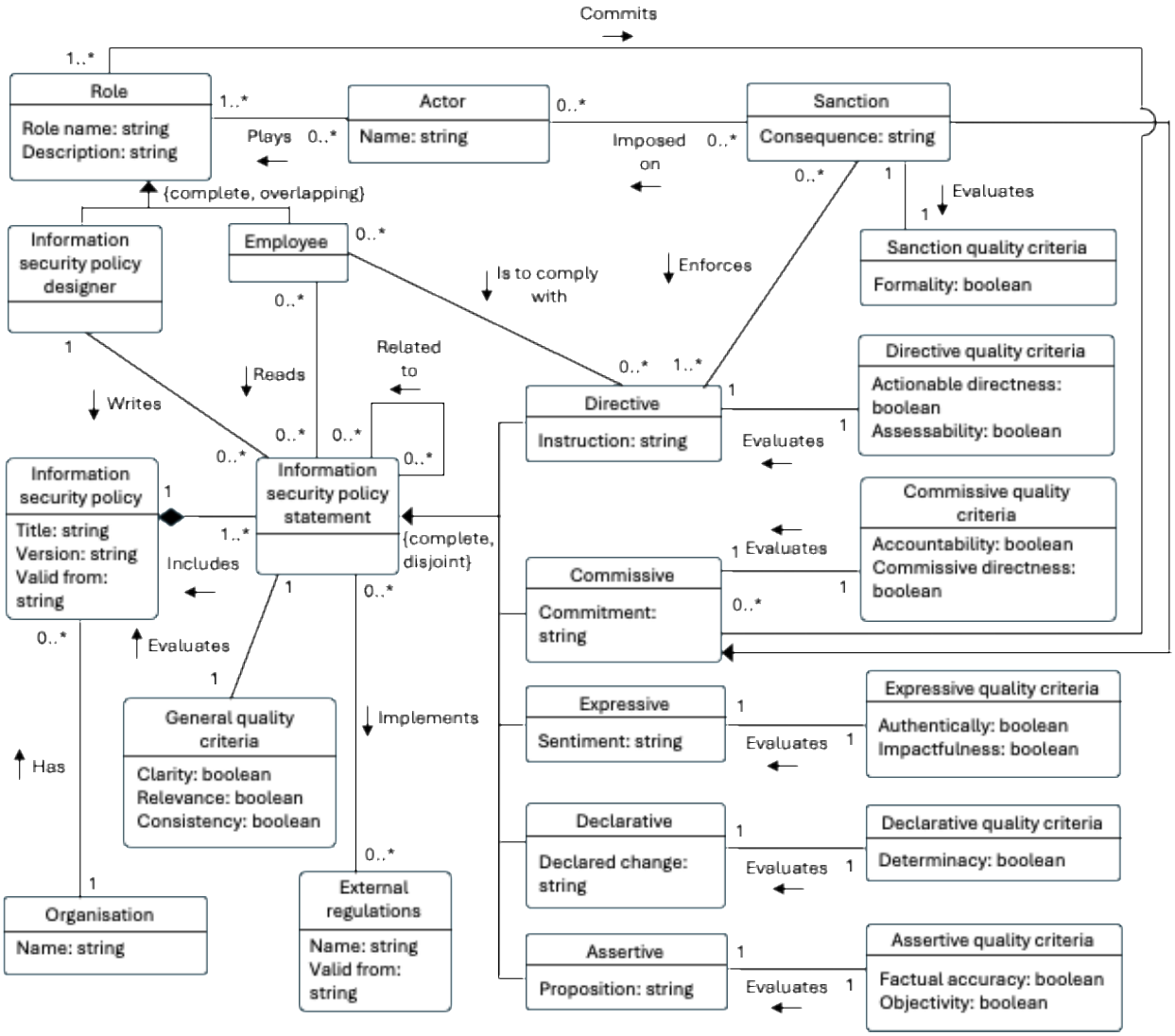

Figure 3. Information Security Policy Conceptual Model version 1.1

5 Discussion

In this study, we developed a SAT-based conceptual model of ISPs to assess the communicative quality of ISP statements. In the long term, this model is intended to support the creation of a high-quality ISP dataset for training LLMs. Below, we discuss the contributions and the implications of our findings both for research and practice. We end this section by addressing the limitations and future research.

5.1 Contributions to Research

The first contribution of this study is a conceptual model. As shown in Section 2, there are only a few conceptual model studies focusing on the structure and quality of ISP statements. For instance, previous research proposed one conceptual model for tailoring ISPs to specific target groups in organizations [32]. Our classification of ISP content differs from the existing model, as ours is more detailed regarding ISP content. We organize the content into six types of ISP statements. Within this structure, sanction is included as a subclass of commissive, which is also associated with directive ISP statements. This can be compared to the previous model's three classes to divide ISP content. Thus, our study advances the literature on ISP design, and could pave the way for future refinement of the policy component model [32] used in ISP tailoring [18].

In particular, our model makes a contribution in the context of ISP quality [11], [21]–[23]. More specifically, our conceptual model introduces quality criteria – applicable for the information security context – for all five types of speech acts in Searle's taxonomy [48]. Some criteria, such as *clarity*, *relevance*, and *consistency*, apply universally across all ISP statements, regardless of speech act type. In addition to these general criteria, the model also specifies criteria unique to particular types of speech act. For instance, *actionable directiveness* and *assessability* are proposed as additional quality requirements for statements functioning as directives. By establishing these criteria, our model provides a structured and theoretically grounded approach for evaluating the communicative quality of ISP statements, with the goal of contributing to the design of more effective and well-defined ISPs. Thus, we contribute with quality assessment of ISPs at a more detailed level than general quality metrics such as brevity, breadth, and clarity [11] that focus on the overall ISP. At the same time, our criteria offer more nuances in the evaluation than the previous metric Keyword loss of specificity [22], which evaluates the use of designated keywords in ISP statements.

Second, the SAT has been applied to classify different types of statements in ISPs demonstrating the relevance of this theory within ISP management. Our analysis of ten ISPs from the British NHS identified all five types of speech acts in Searle's taxonomy [29] – directive, commissive, expressive, declarative, and assertive. The presence of all five speech act types within real-world ISPs shows both the relevance and applicability of SAT for ISP analysis and design. Although many prior studies have examined how ISPs are written or how their content can be improved [11], [20], [21], [23], [30]–[32], to our knowledge, this study is among the first to examine ISPs through the theoretical lens of SAT. Thus, this study contributes to the existing literature [23] on its application in this domain.

One reflection of our speech act classification process is that it was not straightforward in a mechanical sense. Searle's [48] speech act categories are meant to classify the primary illocutionary force of an utterance, so in principle, an ISP statement performs one main type of speech act at a time. However, in practice, the same sentence can have multiple possible interpretations depending on the context, or it may perform a direct act and an indirect act simultaneously. For instance, the following (fictive) example, "All employees must complete information security awareness training by the end of the year", may be understood as a declarative or as a directive. As a declarative, it creates a rule that makes the awareness training mandatory. As a directive, it pushes employees and managers to act. This shows that an ISP statement itself is not inherently bound to only one category; its classification depends on the speaker's intention and context, which means that our classifications were not always clear-cut. In terms of our conceptual model, the condition in the specialization (complete, disjoint) of ISP statements can therefore be seen as a simplification. An alternative could be to allow multi-label classification, i.e., an ISP statement could belong to more than one speech act. Considering the ultimate goal of helping CISOs improve ISP content, this could mean that the LLM provides more than one recommendation on how to improve an ISP statement, depending on the CISO's intention with the ISP statement. If only one recommendation is given, it is based on the classification being in line with the CISO's communicative intentions, and if the classification is not correct, the recommendation risks being less helpful. However, this is an aspect that needs to be left open for future design solutions so that CISOs' preferences regarding the number of recommendations can be taken into account.

Third, our findings also reveal patterns in the distribution of speech acts in ISPs from the British NHS. This means that we contribute to existing knowledge about how ISPs are structured from a communicative perspective [21], [23]. Assertive and directive speech acts overwhelmingly dominate ISP content: among the 600 statements analyzed, 247 were classified as assertives and 226 as directives. This aligns with the functional role of ISPs [10], which describes organizational expectations (assertives) and instructs employees on required behaviors (directives). In contrast, expressive and declarative speech acts are extremely rare, with only three expressive and one

declarative statement identified. This distribution offers empirical insight into how ISPs communicate intent, set expectations, and attempt to shape employee actions.

Fourth, the proposed conceptual model has been evaluated using the SEQUAL framework, which provides a systematic structure for assessing the quality of models and modeling languages. SEQUAL has been used to evaluate models, for instance, in BPMN [55]. This study contributes to the existing literature by demonstrating the use of a quality framework in evaluating UML Class diagrams. Our findings suggest that the proposed conceptual model serves as a foundational framework for improving ISP statements, which can potentially be used as an input for LLMs to enhance ISP content. Additionally, external laws and regulations, which form the basis for sanctions, can be incorporated into the model as well as specifying proposed associations between the Actor class and ISP statements that are Commissive and Declarative (see Section 4.2 and semantic quality). Adding these aspects to the conceptual model would bring it closer to an ISP domain model.

5.2 Contributions to Practice

Our study also has some practical implications. First, LLMs have the potential to enhance ISP content quality for fine-tuned on high-quality ISP datasets. To perform such fine-tuning, a conceptual model is needed to classify the data and evaluate the resulting text. The proposed conceptual model serves as an essential first step by providing the foundation to enhance the quality of LLM-generated ISP statements. Importantly, the process of building such a high-quality dataset does not have to be limited to research. Practitioners can also apply the conceptual model to develop structured, labelled corpora of ISP statements suitable for LLM fine-tuning or to build a RAG solution. By systematically applying the quality criteria defined for each speech act type, organizations can curate datasets in which every statement is clearly formulated, contextually appropriate, and aligned with its intended communicative purpose. This allows for the creation of structured, high-fidelity datasets organized by speech act category. When developing such a dataset, practitioners have to decide whether an ISP statement can belong to the speech act category or if multi-label classification should be possible. Such curated datasets might differ between, for instance, countries depending on linguistic and cultural differences, or language differences that are based on sector-specific regulations or risk environments. In this way, the conceptual model can support the development of software systems that generate clearer, more actionable, and more consistent ISP content. Ultimately, this helps organizations produce ISPs that better guide employee information security behavior.

Second, the conceptual model can support both the creation of new ISP statements and the evaluation of existing ones. When developing new policies, ISP designers such as CISOs can use the model as a reference to determine which speech act type is most appropriate for conveying a particular message or enforcing a specific behavior. This helps ensure that each statement aligns with its intended communicative purpose, whether that involves instructing employees, describing organizational requirements, or committing the organization to certain actions. In addition, the model provides a structured basis for evaluating the quality of statements in existing ISPs. By applying the proposed quality criteria, ISP designers can assess whether statements are clear, relevant, consistent, actionable, or appropriately supported by sanctions where necessary. This makes it possible to identify ambiguous or contradictory statements that may hinder employee understanding.

Third and finally, the conceptual model serves as a reference model for ISP designers, providing an overview of the speech acts used in ISPs. The model enables designers to reflect on questions about the structure and balance of their own ISPs; for instance, “Which types of speech acts are emphasized in the current ISP?”. A policy dominated by directives may appear authoritative but risks being overly prescriptive, while one with too many assertives may lack actionable guidance. ISP designers may also consider are the selected speech acts appropriate for their organization’s context and should the ISP statements be refined to enhance the ISP’s overall effectiveness. Thus,

an overview of the speech acts used by ISPs can help ISP designers identify areas for refinement in their policies.

5.3 Rigor and Limitations

The trustworthiness of qualitative research like ours can be assessed using four criteria [61], [62]: credibility, transferability, dependability, and confirmability. Although we used these criteria to strengthen the rigor of our study during the research process, our study still has certain limitations in each area.

Credibility refers to the level of confidence that can be placed in the "truth value" of our conceptual model. Thus, the credibility of our results depends on whether the classes, relationships, and attributes in the model genuinely emerge from the ISP statements and the speech act analysis. Therefore, it has been important to provide an extensive account in Section 4.1 of how the elements of the class diagrams in Figure 2 and Figure 3 were derived in order to enhance credibility. Still, it is a challenge to fully make any interpretative frame explicit. Furthermore, the conceptual model has been evaluated by a researcher not involved in the modeling activities, in order not to introduce bias. The evaluation was done using the SEQUAL framework, but limited to four of its quality types.

Transferability refers to whether the findings of a study are relevant in other contexts, and it is the reader who must make this judgement [61]. We modeled a randomly selected subset of ISP statements from ten ISPs from the British NHS. Thus, we have achieved randomness at the statement-level. However, because we did not randomly select the ISPs themselves, and they were all from English-language documents and all from UK healthcare organizations, there is a systematic constraint on the sample. By limiting the ISPs to English-language text, we exclude ISP statements written in other languages. This means linguistic and cultural differences in non-English context are not reflected. Our focus on healthcare organizations means excluding ISPs from other sectors where the nature of information security, threats, governance structures, and regulatory frameworks may differ. Consequently, the findings may not be transferable to other countries or organizational contexts. Therefore, we do not claim that the conceptual model can be transferable beyond this context. Furthermore, the model has been developed for ISPs, which means we do not claim that the model or quality criteria are transferable to other types of policies.

Dependability concerns the consistency and stability of our results [61]. Given the interpretive nature of our study, we do not claim that other researchers would come to exactly the same conceptual model. SAT is fundamentally concerned with the performative aspect of language, rather than analyzing stated facts. Thus, analyzing ISP statements using SAT means making interpretations of the language used in ISPs. Consequently, researchers may interpret the same ISP statement differently depending on their prior experiences with ISPs, their familiarity with the SAT, and their judgments about context (see our discussion in Section 5.1). Therefore, we cannot claim that other researchers would have classified individual statements in exactly the same way. When it comes to developing the conceptual model, the patterns we found in the use of speech acts are more interesting than the classification of each individual ISP statement. However, as Collingridge and Gantt [63] point out, dependability (they use the reliability concept) in qualitative research focuses on ensuring consistent quality in the findings, rather than expecting identical outcomes.

Confirmability refers to the degree to which our results are shaped by the empirical data rather than researcher bias, motivation, or interest [61]. In conceptual modeling, the researchers play an active role in deciding on which parts of the empirical data are relevant to the model and how to represent these parts. This means that the conceptual model is not expected to be free from researcher influence. While our understanding of ISPs informed our modeling, we followed established procedures for developing UML diagrams to reduce the risk of systematic errors. Furthermore, having a researcher evaluate the model without participating in the modeling activity helped us to critically examine these interpretations.

5.4 Future Research

The contributions and limitations discussed above open new avenues for future research. In order to assess and ensure the actual transferability of the conceptual model, further research must extend beyond the current sample of ISPs. This expansion should involve ISPs from a broader range of organizational sectors and countries. Validating the conceptual model against ISPs from organizational sectors that may feature other security threats, governance structures and regulatory pressures is important. Exploring ISPs from other countries potentially introduces variation in legal frameworks and language practices. The latter is important, because language is not simply a medium of translation. Language carries structure, nuance, modality and performative elements that may differ significantly across linguistic contexts. For instance, modal verbs, directive phrasing, and the use of obligation versus permission language may differ between languages.

The conceptual model also needs further evaluation in real-world settings, for instance, through a case study involving a new set of ISPs. One way to do this is by applying additional quality types from the SEQUAL framework. For instance, this involves addressing social quality and pragmatic quality. Regarding the latter, it includes social pragmatic quality (to what extent people understand the model) and technical pragmatic quality (to what extent tools can be made that can interpret and potentially act based on the model). Further research should investigate the technical pragmatic quality of the conceptual model by assessing its use in developing a dataset in which ISP statements are classified according to the model, and their communicative quality is evaluated using the quality criteria. The latter could be achieved, for instance, by operationalizing the quality criteria into a scoring system, where each ISP statement in the dataset is given a quality score. The technical pragmatic quality is particularly important, given our ultimate goal of creating high-quality ISP datasets to refine LLMs.

6 Conclusion

In this study, we developed a conceptual model of information security policies (ISPs) using speech act theory as a theoretical lens. The long-term idea is that it will be used to fine-tune large language models (LLMs) or develop a RAG solution to improve ISP content. To develop the model, we applied conceptual modeling and document analysis to 600 ISP statements from ten ISP documents within the British National Health Service. The quality of the developed model was evaluated using the SEQUAL framework, focusing on physical, empirical, deontic, and semantic qualities. The evaluation identified potential areas for further refinement. For instance, the semantic quality of the conceptual model can be enhanced by adding associations between Employees and Commissives, and Directives ISP statements. The final conceptual model, which incorporates these improvements, comprises 21 classes, with six specifically addressing the quality of ISP statements as speech acts. This model has the potential to serve as a valuable input for LLMs to enhance ISP content, ultimately improving its effectiveness in guiding employee security behavior.

References

[1] B. Kör and B. Metin, "Understanding human aspects for an effective information security management implementation," *International Journal of Applied Decision Sciences,* vol. 14, no. 2, pp. 105–122, 2021. Available: https://doi.org/10.1504/IJADS.2021.113532

[2] M. J. Culnan and C. C. Williams, "How Ethics Can Enhance Organizational Privacy: Lessons from the Choicepoint and TJX Data Breaches," *MIS Quarterly,* vol. 33, no. 4, pp. 673–687, 2009. Available: https://doi.org/10.2307/20650322

[3] E. Kolkowska, F. Karlsson, and K. Hedström, "Towards analysing the rationale of information security noncompliance: Devising a Value-Based Compliance analysis method," *Journal of Strategic Information Systems,* vol. 26, no. 1, pp. 39–57, 2017. Available: https://doi.org/10.1016/j.jsis.2016.08.005

[4] G. Dhillon, *Information Security – Text & Cases*. Edition 2.0 ed. Burlington, USA: Prospect Press, 2017.

[5] Truesec, "Threat Intelligence Report 2023," *Truesec*, Stockholm, Sweden, 2023.

[6] Crowdstrike, "2025 Global Threat Report," *Crowdstrike Inc*, 2025.

[7] S. Chatterjee, X. Gao, S. Sarkar, and C. Uzmanoglu, "Reacting to the scope of a data breach: The differential role of fear and anger," *Journal of Business Research,* vol. 101, pp. 183–193, 2019. Available: https://doi.org/10.1016/j.jbusres.2019.04.024

[8] K. D. Loch, H. H. Carr, and M. E. Warkentin, "Threats to information systems: today's reality, yesterday's understanding," *MIS Quarterly,* vol. 16, no. 2, pp. 173–186, 1992. Available: https://doi.org/10.2307/249574

[9] N. H. Chowdhury, M. T. Adam, and G. Skinner, "The impact of time pressure on cybersecurity behaviour: a systematic literature review," *Behav. Inf. Technol.,* vol. 38, no. 12, pp. 1290–1308, 2019. Available: https://doi.org/10.1080/0144929X.2019.1583769

[10] K. Höne and J. H. P. Eloff, "Information security policy – what do international information security standards say?" *Computers & Security,* vol. 21, no. 5, pp. 402–409, 2002. Available: https://doi.org/10.1016/S0167-4048(02)00504-7

[11] S. Goel and I. N. Chengalur-Smith, "Metrics for characterizing the form of security policies," *Journal of Strategic Information Systems,* vol. 19, no. 4, pp. 281–295, 2010. Available: https://doi.org/10.1016/j.jsis.2010.10.002

[12] R. Baskerville and M. Siponen, "An information security meta-policy for emergent organizations," *Logistics Information Management,* vol. 15, no. 5/6, pp. 337–346, 2002. Available: https://doi.org/10.1108/09576050210447019

[13] W. A. Cram, J. G. Proudfoot, and J. D'Arcy, "Organizational information security policies: a review and research framework," *European Journal of Information Systems,* vol. 26, no. 6, pp. 605–641, 2017. Available: https://doi.org/10.1057/s41303-017-0059-9

[14] M. E. Whitman, "Security Policy – From Design to Maintenance," in *Information Security – Policy, Processes, and Practices*, 2008, pp. 123–151.

[15] S. V. Flowerday and T. Tuyikeze, "Information security policy development and implementation: The what, how and who," *Computers & Security,* vol. 61, pp. 169–183, 2016. Available: https://doi.org/10.1016/j.cose.2016.06.002

[16] M. Siponen and A. Vance, "Neutralization: New Insights into the Problem of Employee Information Systems Security Policy Violations," *MIS Quarterly,* vol. 34, no. 3, pp. 487–502, 2010. Available: https://doi.org/10.2307/25750688

[17] Ponemon, "Cost of insider threats global report," Ponemon Institute, North Traverse City, 2020. Available: https://www.ibm.com/downloads/documents/us-en/107a02e94cc8f836

[18] E. Rostami, F. Karlsson, E. Kolkowska, and S. Gao, "Towards software for tailoring information security policies to organisations' different target groups," *Computers & Security,* vol. 159, Article 104687, 2025. Available: https://doi.org/10.1016/j.cose.2025.104687

[19] K. Höne and J. H. P. Eloff, "What makes an effective information security policy?" *Network Security,* vol. 6, no. 1, pp. 14–16, 2002. Available: https://doi.org/10.1016/S1353-4858(02)06011-7

[20] I. Lopes and P. Oliveira, "Applying Action Research in the Formulation of Information Security Policies," in *New Contributions in Information Systems and Technologies. Advances in Intelligent Systems and Computing,* vol. 353, pp. 513–522, 2015. Available: https://doi.org/10.1007/978-3-319-16486-1_50

[21] F. Karlsson, K. Hedström, and G. Goldkuhl, "Practice-based discourse analysis of information security policies," *Computers & Security,* vol. 67, pp. 267–279, 2017. Available: https://doi.org/10.1016/j.cose.2016.12.012

[22] E. Rostami and F. Karlsson, "Qualitative Content Analysis of Actionable Advice in Information Security Policies – Introducing the Keyword Loss of Specificity Metric," *Information & Computer Security,* vol. 32, no. 4, pp. 492–508, 2024. Available: https://doi.org/10.1108/ICS-10-2023-0187

[23] B. C. Stahl, N. F. Doherty, and M. Shaw, "Information security policies in the UK healthcare sector: a critical evaluation," *Information Systems Journal,* vol. 22, pp. 77–94, 2012. Available: https://doi.org/10.1111/j.1365-2575.2011.00378.x

[24] ISO, "ISO/IEC 27002:2022 Information security, cybersecurity and privacy protection – Information security controls," International Organization for Standardization (ISO), 2022.

[25] C. Sundt, “Information security and the law,” *Information Security Technical Report,* vol. 11, no. 1, pp. 2–9, 2006. Available: https://doi.org/10.1016/j.istr.2005.11.003

[26] I. Trummer, “From BERT to GPT-3 codex: harnessing the potential of very large language models for data management,” in *Proceedings of the VLDB Endowment*, vol. 15, no. 12, 2022, pp. 3770–3773. Available: https://doi.org/10.14778/3554821.3554896

[27] M. Abdullah, A. Madain, and Y. Jararweh, “ChatGPT: Fundamentals, Applications and Social Impacts,” *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2022, pp. 1–8. Available: https://doi.org/10.1109/SNAMS58071.2022.10062688

[28] F. Karlsson, S. Gao, J. Krogstie, and L. Aro-Sati, “Towards a Speech Act-Based Model to Enable Future Quality Improvements of Information Security Policies Using Large Language Models,” *Perspectives in Business Informatics Research. BIR 2025. Lecture Notes in Business Information Processing*, vol. 562, 2025, pp. 349–364. Available: https://doi.org/10.1007/978-3-032-04375-7_22

[29] J. R. Searle, “A Classification of Illocutionary Acts,” *Language in Society,* vol. 5, no. 1, pp. 1–23, 1976. Available: https://doi.org/10.1017/S0047404500006837

[30] M. Alshaikh, S. B. Maynard, A. Ahmad, and S. Chang, “Information Security Policy: A Management Practice Perspective,” *Australasian Conference on Information Systems,* 2015.

[31] N. Doherty and H. Fulford, “Aligning the information security policy with the strategic information systems plan,” *Computer & Security,* vol. 25, no. 1, pp. 55–63, 2006. Available: https://doi.org/10.1016/j.cose.2005.09.009

[32] E. Rostami, F. Karlsson, and G. Shang, “Policy components – a conceptual model for modularizing and tailoring of information security policies,” *Information & Computer Security,* vol. 31, no. 3, pp. 331–352, 2023. Available: https://doi.org/10.1108/ICS-10-2022-0160

[33] E. Rostami, M. Hanif, F. Karlsson, and S. Gao, “Defining Actionable Advice in Information Security Policies - Guiding Employees to Strengthen Digital Sovereignty of Organizations,” *Procedia Computer Science,* vol. 254, pp. 30–38, 2025. Available: https://doi.org/10.1016/j.procs.2025.02.061

[34] E. Rostami, F. Karlsson, and S. Gao, “Requirements for computerized tools to design information security policies,” *Computers & Security,* vol. 99, Article 102063, 2020. Available: https://doi.org/10.1016/j.cose.2020.102063

[35] S. Diver, “Information Security Policy – A Development Guide for Large and Small Companies,” SANS Institute, 2021.

[36] NIST, “Information Security Handbook: A Guide for Managers,” National Institute of Standards and Technology, Gaithersburg, USA, 2006.

[37] T. R. Peltier, *Information Security Policies and Procedures – A Practitioner’s Reference*. Second Edition, Boca Raton, 2004. Available: https://doi.org/10.1201/9780203488737

[38] C. R. Smith, “The Definitive Guide to Writing Effective Information Security Policies and Procedures,” *Createspace*, 2010.

[39] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun, and Y. Zhang, “A survey on large language model (LLM) security and privacy: The Good, The Bad, and The Ugly,” *High-Confidence Computing,* vol. 4, no. 2, Article 100211, 2024. Available: https://doi.org/10.1016/j.hcc.2024.100211

[40] J. Yang et al., “Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond,” *ACM Transactions on Knowledge Discovery from Data,* vol. 18, no. 6, pp. 1–32, 2024. Available: https://doi.org/10.1145/3649506

[41] L. Yun, S. Yun, and H. Xue, “Improving citizen-government interactions with generative artificial intelligence: Novel human-computer interaction strategies for policy understanding through large language models,” *PLoS ONE,* vol. 19, no. 12, 2024. Available: https://doi.org/10.1371/journal.pone.0311410

[42] S. Lawal, X. Zhao, A. Rios, R. Krishnan, and D. Ferraiolo, “Translating Natural Language Specifications into Access Control Policies by Leveraging Large Language Models,” *2024 IEEE 6th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA)*, pp. 361–370, 2024. Available: https://doi.org/10.1109/TPS-ISA62245.2024.00048

[43] E. Quevedo *et al.*, “Creation and Analysis of a Natural Language Understanding Dataset for DoD Cybersecurity Policies (CSIAC-DoDIN V1. 0),” *2023 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 91–98, 2023. Available: https://doi.org/10.1109/CSCI62032.2023.00021

[44] S. Deldari *et al.*, "AuditNet: A Conversational AI-based Security Assistant," *Adjunct Proceedings of the 26th International Conference on Mobile Human-Computer Interaction (MobileHCI '24 Adjunct)*, pp. 1–4, 2024. Available: https://doi.org/10.1145/3640471.3680444

[45] D. Najafali, J. M. Camacho, E. Reiche, L. G. Galbraith, S. D. Morrison, and A. H. Dorafshar, "Truth or lies? The pitfalls and limitations of ChatGPT in systematic review creation," *Aesthetic Surgery Journal,* vol. 43, no. 8, pp. NP654-NP655, 2023. Available: https://doi.org/10.1093/asj/sjad108

[46] G. Goldkuhl and E. Braf, "Organisational Ability: Constituents and Congruencies," in *Knowledge Management in the SocioTechnical World*, pp. 30–42, 2002. Available: https://doi.org/10.1007/978-1-4471-0187-1_4

[47] J. L. Austin, *How to do Things with Words*. Cambridge: Oxford University Press, 1962.

[48] J. R. Searle, *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press, 1969. Available: https://doi.org/10.1017/CBO9781139173438

[49] J. R. Searle and D. Vanderveken, "Speech acts and illocutionary logic," in *Logic, Thought and Action*. *Logic, Epistemology, and the Unity of Science*, vol. 2, pp. 109–132, 1985. Available: https://doi.org/10.1007/1-4020-3167-X_5

[50] T. Holtgraves, "The production and perception of implicit performatives," *Journal of Pragmatics,* vol. 37, no. 12, pp. 2024–2043, 2005. Available: https://doi.org/10.1016/j.pragma.2005.03.005

[51] J. R. Searle, *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press, 1979. Available: https://doi.org/10.1017/CBO9780511609213

[52] R. Gasparatou, "How to do things with words: Speech acts in education," *Educational Philosophy and Theory,* vol. 50, no. 5, pp. 510–518, 2018. Available: https://doi.org/10.1080/00131857.2017.1382353

[53] J. V. Schmidt, "Can Artificial Agents be Authors?" *Philosophy & Technology,* vol. 38, no. 1, pp. 1–25, 2025. Available: https://doi.org/10.1007/s13347-025-00839-y

[54] O. I. Lindland, G. Sindre, and A. Solvberg, "Understanding quality in conceptual modeling," *IEEE Software,* vol. 11, no. 2, pp. 42–49, 1994. Available: https://doi.org/10.1109/52.268955

[55] J. Krogstie, *Quality in Business Process Modeling*. Springer, 2016. Available: https://doi.org/10.1007/978-3-319-42512-2

[56] J. Krogstie, *Model-Based Development and Evolution of Information Systems: A Quality Approach*. Springer, 2012. Available: https://doi.org/10.1007/978-1-4471-2936-3

[57] B. Thalheim, "The Science of Conceptual Modelling," *Database and Expert Systems Applications. DEXA 2011. Lecture Notes in Computer Science*, vol. 6860, 2011, pp. 12–26. Available: https://doi.org/10.1007/978-3-642-23088-2_2

[58] G. A. Bowen, "Document Analysis as a Qualitative Research Method," *Qualitative Research Journal,* vol. 9, no. 2, pp. 27–40, 2009. Available: https://doi.org/10.3316/QRJ0902027

[59] M. E. Duffy, "Methodological triangulation: a vehicle for merging quantitative and qualitative research methods," *Image: The Journal of Nursing Scholarship,* vol. 19, no. 3, pp. 130–133, 1987. Available: https://doi.org/10.1111/j.1547-5069.1987.tb00609.x

[60] E. Rostami, "Tailoring information security policies - computerized tool and a design theory," Ph.D. dissertation, Department of Informatics, Örebro University, Örebro, 2023.

[61] E. G. Guba and Y. S. Lincoln, *Fourth Generation Evaluation*. SAGE Publications, 1989.

[62] Y. S. Lincoln and E. G. Guba, *Naturalistic Inquiry*. Sage Publications, 1985.

[63] D. S. Collingridge and E. E. Gantt, "The Quality of Qualitative Research," *American Journal of Medical Quality,* vol. 23, no. 5, pp. 389–395, 2008, Available: https://doi.org/10.1177/1062860608320646