"*I Apologize For Not Understanding Your Policy*": Exploring the Specification and Evaluation of User-Managed Access Control Policies by AI Virtual Assistants

Jennifer Mondragon
jmondragon6@islander.tamucc.edu
Texas A&M University- Corpus Christi
Corpus Christi, Texas, USA

Gael Cruz
cruzg29@gator.uhd.edu
University of Houston - Downtown
Houston, Texas, USA

Dvijesh Shastri
shastrid@uhd.edu
University of Houston - Downtown
Houston, Texas, USA

Carlos Rubio-Medrano
carlos.rubiomedrano@tamucc.edu
Texas A&M University- Corpus Christi
Corpus Christi, Texas, USA

ABSTRACT

The rapid evolution of Artificial Intelligence (AI)-based Virtual Assistants (VAs), e.g., Google Gemini, ChatGPT, Microsoft Copilot, and High-Flyer Deepseek, has turned them into convenient interfaces for managing emerging technologies such as Smart Homes, Smart Cars, and Electronic Health Records. By leveraging explicit commands, e.g., prompts, which can be even launched via voice, VAs provide a very natural interface for end-users. However, the proper specification and evaluation of User-Managed Access Control Policies (U-MAPs), the rules issued and managed by end-users to govern access to sensitive data and device functionality within these VAs, presents significant challenges as this process is crucial for preventing security vulnerabilities and privacy leaks without impacting user experience. This work-in-progress study provides an initial exploratory investigation on whether current publicly-available VAs can manage U-MAPs effectively across differing scenarios. By conducting unstructured to structured tests, we evaluated the comprehension of such VAs, revealing a lack of understanding in varying U-MAP approaches. Our research not only identifies key limitations, but offers valuable insights into how VAs can be further improved to manage complex authorization rules and adapt to dynamic changes.

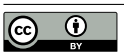
ACM Reference Format:
Jennifer Mondragon, Gael Cruz, Dvijesh Shastri, and Carlos Rubio-Medrano. 2025. "*I Apologize For Not Understanding Your Policy*": Exploring the Specification and Evaluation of User-Managed Access Control Policies by AI Virtual Assistants . In *Proceedings of the 2025 Workshop on Human-Centered AI Privacy and Security (HAIPS '25), October 13–17, 2025, Taipei, Taiwan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3733816.3760753

1 INTRODUCTION

Virtual Assistants (VAs), powered by Large Language Models (LLMs), are increasingly integrated into everyday technologies due to their intuitive, natural language interfaces [3, 17, 20]. They can play roles in Smart Homes, in which they aid in controlling smart devices (TVs, Lights, Locks) [9], Smart Cars, in which they can be utilized for giving directions or controlling functionality via voice commands [13], and in Electronic Health Records (EHR), as they can provide patient information in an expedite and efficient manner during mission-critical operations such as first-response emergencies [18]. Not surprisingly, several different commercial products are already in the market, either in a *hardware-based* mode, i.e., Amazon Alexa and Google Nest, as well as in an *online-based* approach, i.e., Apple Siri, ChatGPT, Google Gemini, Microsoft Copilot, etc. As of today, several other companies are actively working towards providing efficient VAs for a variety of application domains [1].

In such a context, the management of User-Managed Access Control Policies (U-MAPs) [6], i.e., the rules governing access to *sensitive* data and functionality within computer systems, may be required in VA scenarios [11]. For instance, the correct specification, evaluation, and enforcement of U-MAPs may be crucial to mediate *who* is allowed to control devices (Smart Homes), *who* can give directions and change car settings (Smart Cars), and to mediate *who* can access private patient information (EHRs). In these scenarios, not handling U-MAPs correctly can have serious consequences, e.g., thieves controlling a Smart Lock (Smart Homes), kids altering the course of action of a car and causing an accident (Smart Cars), or a surgeon missing important allergy information on a patient (EHRs). However, despite the exciting possibilities of VAs for improving human-computer interactions, and the various solutions that are becoming commercially available, it is not clear if publicly-available, general-purpose VAs can effectively and efficiently handle U-MAPs. More specifically, it is still unclear if the management of U-MAPs via VAs correctly assigns authorization privileges, a.k.a., *permissions*, to protected resources, e.g., Smart Home devices and Smart Car functionality, thus potentially avoiding the introduction of security vulnerabilities otherwise, which could be exploited by malicious third parties to compromise the security of such systems. In addition, it remains unclear if VAs can manage U-MAPs in a user-friendly way, such as offering clear and timely responses without adding interaction burdens.

In order to address these concerns, this paper presents an exploratory investigation of four widely-used and publicly accessible



HAIPS '25, October 13–17, 2025, Taipei, Taiwan

ACM ISBN 979-8-4007-1905-9/2025/10
https://doi.org/10.1145/3733816.3760753

VAs, including ChatGPT (Version GPT-4o), Google Gemini (Version 2024.09.04), Microsoft Copilot (Version 10.28), and High-Flyer Deepseek (Version 2025.01.20). Our study examines their ability to manage U-MAPs across differing scenarios, and provides the following contributions:

- As a part of Sec. 4, we introduce a range of policy specification formats, including plain natural language and increasingly structured representations, designed to communicate U-MAPs more effectively to VAs. We detail their syntactic and semantic contents, alongside the theoretical backgrounds.
- In Sec. 5, we present the results from an exploratory study assessing the performance of four publicly-available VAs when handling U-MAPs. Results reveal varying levels of proficiency, some excelling with structured formats, while some perform better in unstructured formats, highlighting key strengths and limitations.
- Finally, in Sec. 5, we outline practical recommendations for improving VAs policy handling, such as enhancing contextual understanding of unstructured inputs through more robust natural language processing techniques.

This paper is organized as follows: we start in Sec. 2 by reviewing some relevant background information, followed by the problem statement in Sec. 3. We then move on to present the methodology for the study in Sec. 4, and presents its results and subsequent recommendations in Sec. 5, and limitations in Sec. 6. Finally, Sec. 7 concludes this paper with some interesting topics for future work.

2 BACKGROUND AND RELATED WORK

We begin by providing essential background on key topics discussed throughout the paper. Section 2.1 covers the rise of VAs and recent advancements in LLM-related security, while Section 2.2 introduces U-MAPs, defining their role and importance in managing emerging technologies.

2.1 LLM-based Virtual Assistants

Due to their remarkable performance, Large Language Models (LLMs) are rapidly gaining traction across a multitude of domains, from finance and health care to education and software development [3, 5, 17, 20, 25, 30]. In particular, Virtual Assistants (VAs) leverage LLMs as their foundational model for the processing and retrieval of general and domain-specific knowledge, and can be further augmented with additional data processing and storage capabilities by pairing their LLM-based back-end with traditional databases, ontologies, knowledge graphs, etc. [8]. This makes VAs an ideal and convenient user interface for emerging smart technologies such as Smart Homes, Smart Cars, and EHRs, allowing them to be operated naturally and intuitively.

However, despite the growing interest and emerging applications, the use of LLMs as a back-end module for VAs can pose security vulnerabilities, leading to undesirable outcomes for their users. Prior studies have highlighted the vulnerabilities of LLM-powered systems, including hardware, software, and network level attacks [31], insecure code generation (with 40% of GitHub Copilot outputs in one study introducing vulnerabilities) [22], and risks associated with third-party plugin integration, such as malicious prompt injection and unauthorized data access [14].

Although prior research has evaluated VAs for security vulnerabilities and data privacy, little attention has been given to how these VAs handle the interpretation of formal policies, which is a necessity for safe and effective operations of smart technologies. This gap motivates our study, which systematically examines how VAs apply U-MAPs and evaluates their reliability in security-sensitive contexts.

2.2 User-Managed Access Control Policies

For the purposes of this paper, we define User-Managed Access Control Policies (U-MAPs) to be the set of access control or authorization policies whose *life-cycle*, i.e., their specification, update, and removal over time, is handled by *end-users*, who may have not received formal training in access control, resorting to a combination of domain-specific knowledge and common sense instead [24].

The correct management of U-MAPs may become central to the development and adoption of modern emerging technologies, which are fading away from expert-managed, *centralized* approaches, to more *distributed* approaches in which end-users are given full control of the way they interact with a given technology, restricting access to functionality and data at will. In such a scenario, U-MAPs are expected to provide a balance between *domain-specific* settings, e.g., conveniently restricting access to functionality, and *security-specific* settings, e.g., the collection and retrieval of data.

Table 1 provides an example of a U-MAP handling a Smart Home environment consisting of a Smart Lock, a Smart TV, and a set of Smart Lights, all of them interconnected via an Internet of Things (IoT) setting, and controlled, e.g., access mediated, by means of a VA. In such a scenario, the U-MAP, stated in Natural Language, e.g., English, restricts access to each smart device to only a subset of users, identified by their *role*, e.g., *Homeowner*, *Partner*, etc., thus following an approach inspired by the well-known methodology, Role-Based Access Control (RBAC) [23].

For the purposes of this paper, we will consider a subset of the eXtensible Access Control Markup Language (XACML) [27], the *de facto* standard language for access control, in an attempt to provide a consistent foundation for the U-MAPs considered in our study, as it will be further described in Sec. 4. For instance, the U-MAP shown in Table 1 is expressed using a simplified subset of XACML. In this subset, a U-MAP consists of a series of *rules*, and each rule contains a *target*, which identified the resource being accessed, a *condition*, which specifies when access should be allowed or denied, and a *rule decision*, which determines whether access is allowed or denied if the condition is satisfied. This simplified XACML subset supports a one rule-combining algorithm, i.e., *First Applicable*, where the U-MAP is evaluated from top to bottom, resulting in an policy determination from the first rule whose target *equals* the requested resource and whose condition evaluates to true.

3 PROBLEM STATEMENT

As discussed in Sec. 2, VAs offer a natural interface for users to interact with increasingly complex systems. This ease of use is particularly valuable in domains like Smart Homes, Smart Cars, and Healthcare, where quick, hands-free access can be essential. However, these environments often involve sensitive information and critical operations, making access control necessary [19].

While VAs streamline interactions, it remains unclear whether they can correctly interpret and enforce U-MAPs. Effective policy interpretation and evaluation requires the VA to grant or deny access reliably to prevent security risks, while also maintaining usability for non-technical users. Achieving both *effectiveness* and *efficiency* poses a key challenge, and is a driving force behind this study.

This paper evaluates how well current VAs can handle U-MAPs in realistic scenarios, identifies their limitations, analyzes how policy format influences performance, and offers design recommendations for safer and more usable VAs. Our study is guided by the following research questions:

- **RQ1**: *Can VAs evaluate U-MAPs effectively from a security perspective?*
- **RQ2**: *What recommendations can be given to future VAs to effectively and efficiently handle U-MAPs?*

To highlight the stakes, consider a U-MAP shown in Table 1, which governs access in a smart home. A request such as *"Can kids watch TV?"* should be answered accurately based on the defined rules. While a misinterpretation might merely cause inconvenience in this case, other queries, such as *"Can visitors manipulate the Smart Lock?"*, have clear security implications. These examples, though simple, underscore the need to evaluate how VAs process and enforce U-MAPs.

4 METHODOLOGY

To address our research questions, we conducted an exploratory study assessing the capabilities of four publicly available, general-purpose VAs: OpenAI ChatGPT [1] (Version GPT-4o), Google Gemini[2] (Version 2024.09.04), Microsoft Copilot[3] (Version 10.28), and High-Flyer Deepseek[4] (Version 2025.01.20) on tasks related to the evaluation of a series of U-MAPs. The study began by selecting the VAs previously mentioned, due to their accessibility and relevance. Next, we established the evaluation domains, provided rationale for selection, and developed a series of U-MAPs utilizing various formats. Finally, we outline the methods for conducting interactive VA sessions, distinguishing between *Contextual* and *Non-Contextual* methods for prompting.

4.1 Selecting VAs

ChatGPT, Gemini, Copilot, and Deepseek were chosen due to their broad user-base and frequent use, indicating that each VAs has likely accumulated knowledge from diverse interactions, including security-related queries. This exposure makes them ideal for evaluating VAs' handling of U-MAPs, despite none being tailored exclusively to security.

4.1.1 OpenAI ChatGPT. ChatGPT, developed by OpenAI, is one of the most widely used conversational AI models. The latest version, GPT-4o, handles a broad range of tasks and has been trained on a vast dataset, enabling it to generate detailed and coherent responses across various domains. It is refined with Reinforcement Learning from Human Feedback, which enhances the structure of responses utilizing developer ideals [4]. While ChatGPT does not learn from real-time interactions, its extensive pre-trained knowledge base includes access control which is necessary for this study, though challenges may occur from incorrect yet plausible responses [4].

4.1.2 Google Gemini. Gemini, developed by Google DeepMind, integrates advanced reasoning with internet retrieval, providing an advantage when handling up-to-date queries. It leverages Google's powerful machine learning models, which have been trained on diverse datasets across various domains [12]. Its processing capabilities enable it to analyze and adapt to various formats, including

Table 1: Comparison of different U-MAP encoding approaches (formal, informal, semi-formal) for Smart Homes, including policy variants.

Approach	U-MAP
XACML **(Formal)**	<policy, CA=First-Applicable> <rule result=Allow> <target>Lock</target> <cond.>role=Homeowner</cond.> </rule> <rule result=Deny> ...</rule> </policy>
Informal **(INF)**	Only homeowners and partners are allowed to use the Lock, everybody is allowed to use the TV, everybody is allowed to use the Lights. Everything else is denied.
Modified Informal **(Mod-INF)**	Partners cannot use the lock, but visitors can use the lock now.
Semi-Formal **(SEMI-UNROLL)**	1. Homeowner can access Lock, 2. Homeowners can access Lights, ...10. Deny access to everything else.
Modified Semi-Formal **(Mod-SEMI-UNROLL)**	Remove Rule: Partner can access Lock. Create New Rule: Visitor can access Lock.
Semi-Formal-Rule-Based **(SEMI-RULE)**	1. if role = homeowner, Lock Access = allowed 2. if role = partner, Lock Access = allowed ... 5. Deny access to everything else.
Modified Semi-Formal-Rule-Based **(Mod-SEMI-RULE)**	if role = visitor: lock access = allowed, if role = partner: lock access = denied.

[1] https://openai.com/chatgpt/
[2] https://gemini.google.com/app
[3] https://copilot.microsoft.com/
[4] https://www.deepseek.com

those in this study. However, its reliance on web-sourced data introduces potential inconsistencies, as it may provide responses based on publicly available but unverified information [12].

4.1.3 Microsoft Copilot. Copilot, developed by Microsoft, is a VA designed to assist with daily tasks across various platforms, though this study utilizes the general-use interface [7]. Built on the GPT-4o model, Copilot is optimized for productivity and contains integration with Microsoft's security and compliance tools, making it relevant for U-MAP interpretation [21]. Although primarily focused on productivity, Copilot was included to evaluate security-handling capabilities, offering insight into possibilities beyond typical tasks.

4.1.4 High-Flyer Deepseek. Deepseek, developed by High-Flyer, is an emerging VA with a focus on reasoning and structured data-processing, using a Mixture-of-Experts model [26]. While it is less documented compared to the other VAs in this study, its emphasis on logic makes it an interesting candidate for U-MAP interpretation and evaluation [26]. Deepseek may handle structured policies better than general-purpose VAs, but its effectiveness with natural language policies and security practices remains unclear.

Each of these VAs represent a differing design priority among a spectrum of models. ChatGPT and Gemini emphasize general-purpose reasoning, though Gemini integrates real-time retrieval, which may affect the reliability of security-related responses. Copilot's focus on productivity and enterprise integration may make the queries related to access control particularly relevant. Deepseek, in contrast, highlights structured reasoning, but remains relatively unexplored. These distinctions may influence how each VA interprets and applies U-MAPs due to their underlying differences in training data, reasoning strategies, and integrations, as well as how each VA interprets format structures (i.e. Informal).

4.2 Determining Application Domains

We selected three application domains where VAs could significantly enhance user experience through natural language interactions. These include Smart Homes, where VAs can manage IoT devices for seamless control [9], Smart Cars, where VAs can assist with navigation and comfort settings [2, 13], and EHRs, where VAs can support quick and accurate information retrieval in critical situations [10, 18].

4.2.1 Smart Homes. Smart Home technologies are becoming increasingly common, often utilizing Graphical User Interfaces (GUIs) for device management and policies [9]. While accuracy in controlling these devices is integral, minor errors are typically not disruptive, unless they relate to security devices (e.g., cameras, locks). Slow responses, however, can reduce the overall effectiveness of the system, as users expect seamless executions of commands. While management may not be as crucial, poor implementation will cause a smart home to quickly lose effectiveness, reliability, and security. Similarly to EHRs, RBAC is also a model that could be implemented, though issues with complexities may arise due to the scalability of users (i.e., family members) and devices (e.g. thermostats, lights). Therefore, the smart home domain was selected due to the growing popularity and potential streamlining that VAs can bring to the domain.

4.2.2 Smart Cars. Connected Automated Vehicles (CAVs) [13] are an emerging technology that rely on voice control for tasks, such as adjusting in-car settings (e.g. AC, radio). While some cars may have visual systems (i.e., multimedia receivers), the use of GUIs within the CAV context is not practical, making hands-free options crucial. Accuracy and response time are critical, as errors in vehicle functions or delays could lead to safety risks and serious consequences to those within and outside the CAV. These components need to be effectively handled by a robust management system, ensuring that security protocols are followed to keep the system reliable and safe. Though RBAC is a possible solution, it may not be ideal due to the necessary flexibility that stems from the various users (e.g. driver, passenger) and dynamic contexts (e.g., autonomous vs. manual driving). The addition of the smart car domain was valuable for this study as the domain is still in development stages, allowing an evaluation on the VAs ability to handle U-MAPs with little background knowledge.

4.2.3 Electronic Health Records. EHR systems are well-established and have become a fundamental part of modern healthcare practices as they are primarily used by healthcare professionals to store, retrieve, and update patient records [18]. These tasks make access control an integral component to ensure the protection of private information, as well as accuracy, as incorrect policy application could have severe or life-threatening consequences. While response times can be important, it may not be as urgent unless an emergency situation arises, where it will then become crucial for immediate access. Beyond that, EHR systems need to maintain an effective management system, to not only protect privacy and maintain regulations, but to improve the overall efficiency and reliability of healthcare delivery. Previous solutions include RBAC as it ensures that different roles (e.g. doctor, nurse) have appropriate levels of access to sensitive medical data, making this domain suitable to explore the capabilities of VAs [18].

4.3 Establishing U-MAPs

To evaluate VA capabilities across formats, this study employs a method that generates different instructions from a single U-MAP in three primary styles: informal (natural language), semi-formal (concise statements), and semi-formal-rule-based or formal (code-like). Table 1, as well as Tables 3 and 4 (Appendix A), provide a full listing of the U-MAPs.

4.3.1 Informal (INF). Developed as an initial format, it is intended to exercise the capabilities of VAs for handling moderately descriptive U-MAPs in natural language. It contains different statements explaining the U-MAP rules, which are composed of descriptions of roles, protected resources, and quantity pronouns, i.e., everybody.

4.3.2 Formal (XACML). Developed from the INF format, it leverages the subset of XACML in an effort to provide a formalization of each U-MAP from which other formats can be developed from. This format was not used directly in the procedures described here.

4.3.3 Modified Informal (Mod-INF). Developed from the INF format, the modified informal structure is intended to exercise the capabilities of VAs for handling specific updated U-MAPs in natural language.

4.3.4 Semi-Formal (SEMI-UNROLL). Developed from the XACML format by syntactically *unrolling* each XACML rule into natural language one, this format is an effort to assess the capabilities of VAs for understanding U-MAP that are described as concise statements.

4.3.5 Modified Semi-Formal(Mod-SEMI-UNROLL). Developed from the SEMI-UNROLL format, this modifies certain U-MAP rules from the original set, in an effort to assess the dynamic capabilities of VAs for U-MAPs.

4.3.6 Semi-Formal-Rule-Based (SEMI-RULE) (Formal). Developed from the XACML format, this format provides a shorter version of the unrolling with respect to SEMI-UNROLL using a stricter natural language syntax in a rule form, in an effort to provide a more concise description of U-MAPs for evaluation purposes.

4.3.7 Modified Formal (Mod-SEMI-RULE). Finally, this format was developed from the SEMI-RULE format by *compressing* several rules to assess the capabilities of VAs for handling dynamic U-MAPs.

4.4 Constructing Inquiries

To evaluate the VAs ability to interpret and apply U-MAPs, a set of structured questions focusing on retrieval accuracy, which measures how well the VA extracts and applies explicit rules, and logical consistency, which considers whether the VA provides reason rather than arbitrary responses, were developed. The questions were straightforward based on the given U-MAPs in each domain, for example, *"Can partners manipulate the Smart Lock?"* or *"Can visitors watch TV?"* tested direct rule application, while *"Can nurses access PII-Data?"* evaluates implicit reasoning when the policy states *"Only Admin Staff can access PII-Data"*. Responses were recorded as *true* or *false*, with notes on unclear or unsupported justifications. Analysis considered both correctness against ground truth and the quality of explainability.

4.5 Conducting Interactive Sessions

To assess how well VAs interpret and apply U-MAPs, two evaluation methods were used: *Contextual* and *Non-Contextual*. The Contextual Method simulated scenarios where VAs have domain-specific knowledge before policy evaluation, while the Non-Contextual Method assesses how well a VA performs without any prior interaction. By comparing these methods, we aim to determine the extent to which contextual grounding affects policy comprehension and accuracy. The interactive sessions were conducted during September 2024 and March 2025, where moderately trained team members independently engaged each VA using the structure in Figure 1, while two experts, one in U-MAPs, one in VAs, reviewed outcomes to derive the recommendations in Sec. 5.

4.5.1 Contextual Method. In the Contextual method, VAs were prompted with domain-specific concepts through structured prompts before being asked U-MAP related questions. Domain knowledge was provided in the form of queries and definitions (e.g. "Explain RBAC and its key components."), followed by domain-specific applications (e.g. "What about RBAC within Smart Home domains?"). After this grounding, the VA was instructed to generate an RBAC-style U-MAP based on the provided context, which includes predefined U-MAPs. The interactive session included two rounds, a

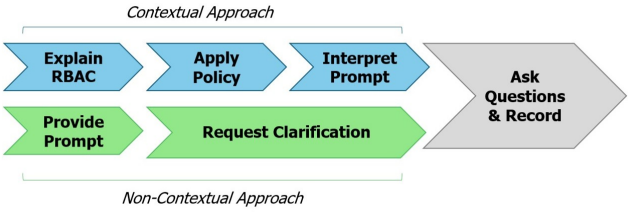


Figure 1: Contextual and non-contextual VA interaction flows, highlighting policy and interpretation differences. See Sec. 4.5.1 and Sec. 4.5.2 for detailed steps.

pre-modification using the baseline policy and questions, and a post-modification, where the U-MAP was updated and the questions were reiterated. The session follows this process:

(1) Ask the VA to explain RBAC, including definitions, within a new chat session.
(2) Query the VA on domain applications (i.e., Smart Cars).
(3) Instruct the VA to generate an RBAC-style U-MAP based on domain-specific prompts.
(4) Evaluate the U-MAP utilizing questions from Table 2.
(5) Record all responses, noting assumptions, inconsistencies, or deviations.
(6) Provide modified rules and repeat the process.
(7) Close the session and clear memory (if possible).

4.5.2 Non-Contextual Method. In the Non-Contextual Method, the VA was presented with the U-MAP directly, without prior interaction or explanatory prompts. For example, a session would begin with direct instructions (e.g. "Here is an RBAC policy.), followed by the pre-defined U-MAPs. This method utilized the same two-round structure of pre-modification and post-modification.

(1) Provide the U-MAP to the VA within in a new chat session.
(2) Evaluate the U-MAP using questions from Table 2.
(3) Record responses and in some cases, VAs responded with apologies when questioned about the incorrect responses [29].
(4) Provide modified rules and repeat the process.
(5) Close the session and clear memory (if possible).

5 RESULTS

The evaluation of Gemini, ChatGPT, Copilot, and Deepseek revealed distinct performance characteristics across various domains. The VAs were assessed in five key areas: **Accuracy** (Sec. 5.1), **Consistency** (Sec 5.2), **Perception** (Sec. 5.3), **Explainability** (Sec. 5.4), and **Interaction Performance** (Sec. 5.5). Sections 5.1 to 5.5 are then focused on providing answers to **RQ1**, whereas Sec. 5.6 provides a set of interesting **Recommendations** to answer **RQ2**.

5.1 Accuracy

Accuracy assesses how each VA applies U-MAPs against ground truth. As access decisions directly impact system security and user privacy, accuracy is a core requirement for the deployment of VAs in real-world scenarios. Accuracy from all domains, session types, and formats are visible in Figure 2 and Figure 3.

Across all VAs, two key patterns emerge. Firstly, ChatGPT and Deepseek consistently demonstrate higher accuracy across diverse

Table 2: User access questions across Smart Homes, Smart Cars, and EHRs based on U-MAP scenarios.

Smart Homes		Smart Cars		EHRs	
ID	**Question**	**ID**	**Question**	**ID**	**Question**
Q-SM-1	*Can Visitors manipulate the Smart Lock?*	Q-SC-1	*Can Drivers give Directions?*	Q-EHR-1	*Can Physicians access Medical Data?*
Q-SM-2	*Can Visitors watch TV?*	Q-SC-2	*Can Co-Pilots give Directions?*	Q-EHR-2	*Can Staff access Medical Data?*
Q-SM-3	*Can Homeowner turn on the Lights?*	Q-SC-3	*Can Passengers set the Radio?*	Q-EHR-3	*Can Nurses access PII-Data?*
Q-SM-4	*Can Partner manipulate the Smart Lock?*	Q-SC-4	*Can Drivers set the Radio?*	Q-EHR-4	*Can Staff access PII-Data?*
Q-SM-5	*Can Homeowner access a Meter?*	Q-SC-5	*Can Kids give Directions?*	Q-EHR-5	*Can First Responders access PII-Data?*

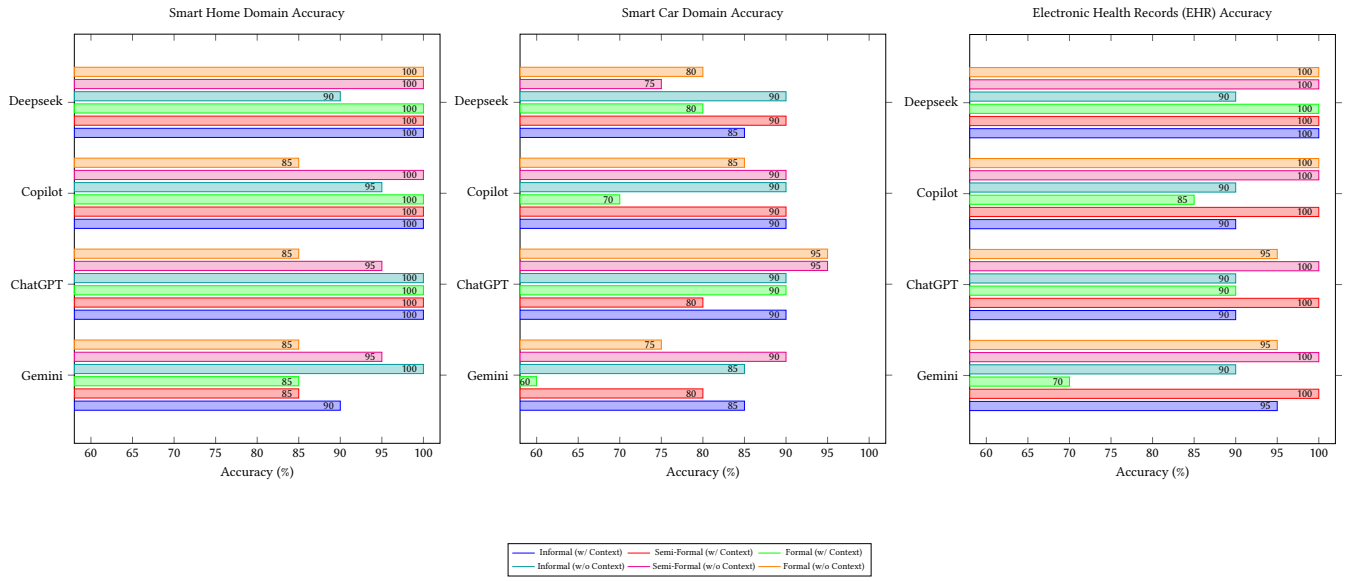


Figure 2: Session accuracy across Smart Home, Smart Car, and EHR domains evaluated using informal, semi-formal, and formal U-MAPs. Results are shown with and without contextual information to highlight impact of context on performance.

formats, with scores clustering above 90%, indicating strong policy interpretation. Secondly, the Smart Car domain stands out as a consistent source of reduced accuracy across all models, suggesting that complexity or ambiguity in this domain may pose unique challenges for VAs, and may call for semantic reasoning capabilities.

These findings imply that current VAs are capable of reasonably high accuracy when handling natural or semi-structured policies, but tend to falter in domains where decision logic is more context-sensitive.

5.2 Consistency

Consistency evaluates how well a VA recalls and applies U-MAPs across repeated interactions, which is crucial for trust and security. Our study shows longer sessions improve consistency, as restating or adjusting U-MAPs aid retrieval. For instance, ChatGPT's accuracy rose from around 93% pre-modification, to over 96% after modification, while Gemini improved the most, from 82% to 91%. Copilot and Deepseek saw gains of 3%-4%, with Deepseek starting as low as 88%, suggesting enhancements through iterative dialogue and indicating that some VAs benefit from sustained engagement.

These findings highlight that initial accuracy may not fully represent VAs reliability, as sustained interactions and clarifications were critical to achieving consistent enforcement. Designing VAs that support such iterative refinement can enhance the robustness of these systems.

5.3 Perception

Perception, in the context of this study, refers to how well each VA interprets U-MAPs across varying formats, which provides insight into semantic understanding, pre-training scope, and adaptability.

All evaluated VAs performed well in the *Informal* format, achieving accuracy between 87% and 100%, suggesting conversational

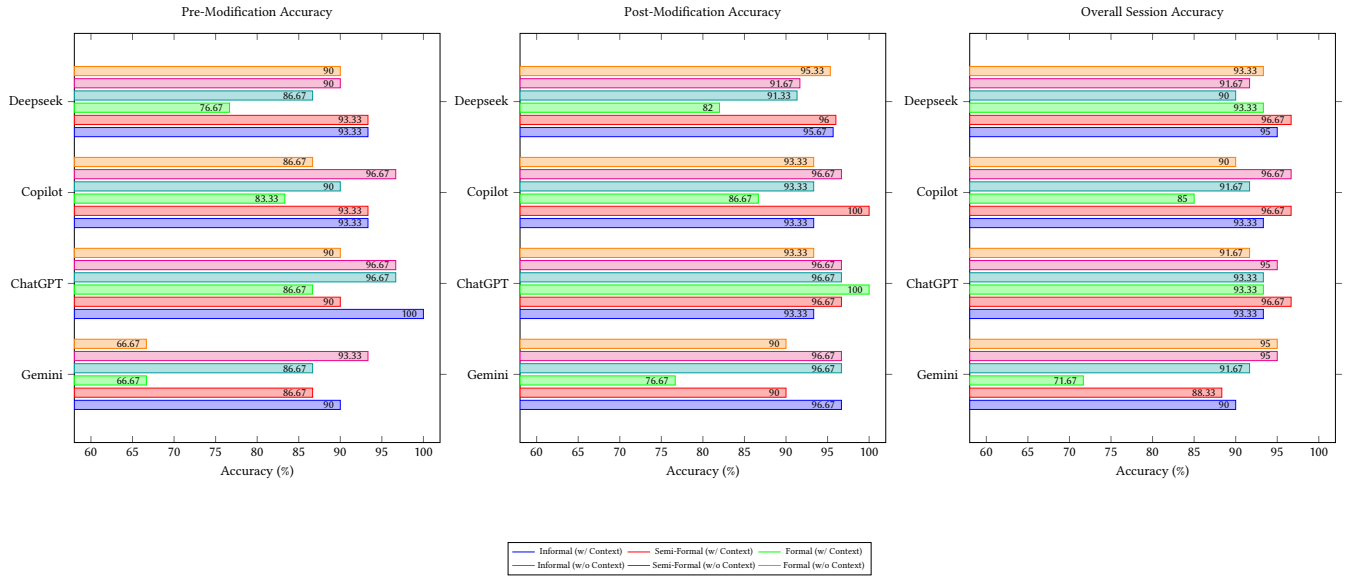


Figure 3: Overall, pre-modification, and post-modification session accuracy for all domains and U-MAP types. VAs were evaluated with and without contextual information to asses how policy modifications affect VA performance.

language aligns with current capabilities. Accuracy increased with the *Semi-Formal* format, with scores ranging from 88% to 100%, indicating improvement with structure. In contrast, the *Formal* format presented a challenge, as accuracy dropped as low as 67%, likely due to the code-like format.

Overall, natural language is a viable entry point for policy specification, and introducing lightweight structure can enhance interpretability. However, fully formal formats has limitations, highlighting a need for improved reasoning capabilities, and extended training for future VAs.

5.4 Explainability

Explainability reflects how clearly VAs communicate decision-making when applying U-MAPs in an interpretable and consistent way. In the context of this study, explainability was assessed utilizing the VAs justification and clarity of reasoning, allowing insight into rationale and understandability of the VA response.

Most VAs demonstrated structured explanations, though justification and clarity varied. Deepseek, with built-in reasoning, provided some of the most thorough justification, while ChatGPT and Copilot tended to follow a simpler pattern, typically offering an explicit answer (e.g., *"No, the Homeowner cannot access a Meter"*) paired with a minimal rationale (e.g., *"because everything else is denied"*). In contrast, Gemini extended its justifications and clarity by appending summaries of the U-MAP, while Deepseek went further by suggesting possible improvements to the U-MAP. Though this is not always necessary, these responses demonstrate that VAs may have potential for more user-aligned and adaptive explainability.

Inference-based explainability proved particularly challenging, as multiple VAs failed to justify access restrictions from general rules, producing vague or incorrect explanations. Some responses even included an apology, such as *"I apologize for not understanding your policy"*, which inspired this paper's title.

5.5 Interaction Performance

Interaction performance, in this study, evaluates the ease and efficiency of interaction with a VA, encompassing response latency, interaction flow, and friction points. Across the VAs, response times were generally quick, averaging between 2.88 seconds and 3.21 seconds. Copilot had the quickest average response time (2.88s), followed by ChatGPT (3.15s), Gemini (3.21s), and Deepseek (14s).

Message limits varied across systems, impacting conversational flow. ChatGPT and Copilot enforced caps of 60 messages per hour, with Copilot restricting sessions to 30 messages. Gemini allowed for longer sessions, supporting roughly 100 messages per hour before blocking, while Deepseek applied rigid restrictions by throttling after approximately 20 messages, then blocking for the remainder of the hour. These barriers, often communicated through vague error messages (e.g., *"temporary server error"*), created notable friction.

While most VAs maintained high responsiveness and conversational fluency, the presence of hard caps and delayed feedback mechanisms may hinder real-time U-MAPs, as interaction performance in this context requires speed and interaction continuity.

5.6 Recommendations

Each of the previously discussed metrics revealed both strengths and critical limitations across current VAs. A central issue identified was the inability to perform robust inference and apply logical structure to U-MAPs and the queries. These challenges were compounded by limitations in domain-specific pre-training and an inability to adapt policies dynamically. Below are the key recommendations for improving the integration of VAs.

5.6.1 Embedding Explainability. Our evaluation highlights a clear trade-off between explainability quality and interaction performance. Systems such as Deepseek demonstrate stronger logical inference capabilities, but incur significantly longer response times, often adding approximately four-to-six seconds [28]. To address this, future VAs could adopt structured reasoning, where it is invoked selectively for ambiguous or policy-critical prompts. Some systems, such as the Modular Reasoning, Knowledge, and Language (MRKL) offer a potential solution by selectively activating reasoning components when needed [16]. Though MRKL focuses on mathematical problems, the architecture suggests a scalable way to balance clarity and efficiency in natural language settings.

5.6.2 Fine-Tuning for Domains. The quality of VA responses heavily depend on the breadth and relevance of their pre-training data. Domains like EHR benefit from extensive, well-document access control policies, while Smart Homes, and especially Smart Car domains, suffer from limited and less mature documentation. To bridge this gap, generic pre-trained models should be fine-tuned or adapted with domain-specific data and knowledge sources, allowing them to improve their understanding of specialized concepts, policies, and emerging technologies, while ultimately enhancing the accuracy and usability in these complex environments.

5.6.3 Real-Time Adaptability. Currently, VAs lack the ability to learn and update knowledge during live interactions, often relying on static pre-training data. This limitation reduces their effectiveness in evolving domains such as Smart Cars, where policies and technologies frequently update. Future VAs should combine real-time data access with controlled online learning to update their understanding based on user feedback and new information. Implementing such adaptive learning requires strong validation to avoid incorporating incorrect information or harmful inputs, striking a balance between responsiveness and reliability in dynamic environments [15].

5.6.4 Response Consistency. The varied response styles across all the VAs, ranging from concise to detailed, highlight the need for greater consistency and clarity in interactions. In some cases, responses included long justifications that combined answers, reasoning, additional context, and code. In others, justifications were short, only containing the answer and minimal reasoning. Instead of extensive explanations, clear and structured responses with focused justifications can make the reasoning process easier to follow. Future VAs should balance thoroughness with clarity by providing enough context to explain decisions without unnecessary detail, which will support transparency and consistency.

6 LIMITATIONS

While this study offers meaningful insights into how VAs handle U-MAPs, it remains a work-in-progress, resulting in some limitations stemming from the use of simplified policies, limited perspectives, and a narrow selection of publicly-available VAs.

6.0.1 Simplified Policy Structures. The U-MAPs used in this study were intentionally scoped to reflect simple cases commonly encountered in smart environments. This allowed the isolation and examination of foundational challenges in VAs without the effects of highly complex logic. While these U-MAPs do not cover advanced constructs such as compounded conditions, they remain highly relevant to everyday use. More importantly, the baseline scenarios revealed significant limitations in current VA performance. Future work will build upon these findings by incorporating complex U-MAPs with multi-step decision-making to further examine the capabilities of VAs.

6.0.2 Limited User Perspectives. This study involved a small team of researchers with varying levels of experience in access control, enabling initial insights into how expertise influences VA interactions. However, real-world end-users often have little to no technical backgrounds. To broaden the impact and usability of future systems, subsequent studies will prioritize participation from non-expert users, focusing on how effectively VAs communicate policy decisions and support understanding in everyday contexts.

6.0.3 Scope of VAs. The VAs evaluated were general-purpose, publicly available systems selected for their widespread use and accessibility. While not specialized in security, these models provided a baseline for identifying current capabilities. Building on these insights, future work will involve developing and evaluating a fine-tuned VA tailored to access control scenarios, with the goal of improving consistency, accuracy, and domain-specific reasoning.

7 CONCLUSIONS AND FUTURE WORK

This paper presented an exploratory study evaluating the ability of publicly available VAs to mange U-MAPs across multiple domain scenarios. Our findings highlight that while current VAs exhibit some promising capabilities, particularly when U-MAPs are expressed in natural or structured formats, they also face significant challenges in dynamic settings.

Notably, we observed that each VA demonstrated competence with at least one policy format (i.e. INF), yet struggled with others (i.e. SEMI-RULE), underscoring the need for further customization to ensure reliable U-MAP evaluation without sacrificing interaction performance. Importantly, our analysis utilized policies based on the RBAC model, but for future work, we aim to expand this investigation to include alternative access control paradigms, such as Attribute-Based Access Control. In addition, we plan to implement our proposed recommendations within custom VAs to explore how targeted improvements can enhance security and user experience.

By extending this line of research, we aim to better understand the role of LLM-powered VAs in enforcing access control, and to chart a path towards more trustworthy, adaptable, and user-centered smart technology ecosystems.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation under Grants Nos. 2232911 and 2131263, and by the CAHSI-Google Institutional Research Program, sponsored by Google and the Computing Alliance of Hispanic Serving Institutions (CAHSI), a National INCLUDES Alliance. The authors also thank Joseph Uzoigwe for the valuable contributions he provided towards the development of this work.

REFERENCES

[1] Android Authority. 2024. Siri vs Alexa vs Google Assistant vs Bixby: Which one reigns supreme? https://www.androidauthority.com/siri-vs-alexa-vs-google-assistant-vs-bixby-3192996/. [Online; accessed September-12-2024].

[2] Fabio Arena, Giovanni Pau, and Alessandro Severino. 2020. An overview on the current status and future perspectives of smart cars. *Infrastructures* 5, 7 (2020), 53.

[3] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2024. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology* 15, 3 (2024), 1–45.

[4] OpenAI ChatGPT. 2022. Introducing ChatGPT. https://openai.com/index/chatgpt/

[5] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde De Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. 2021. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374* (2021).

[6] Chung, David Ferraiolo, and David Kuhn. 2006. Assessment of Access Control Systems. https://doi.org/10.6028/NIST.IR.7316.

[7] Microsoft Copilot. 2025. Empower your organization with Copilot. https://www.microsoft.com/en-us/microsoft-copilot/organizations

[8] Xin Luna Dong, Seungwhan Moon, Yifan Ethan Xu, Kshitiz Malik, and Zhou Yu. 2023. Towards next-generation intelligent assistants leveraging llm techniques. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5792–5793.

[9] Jide S Edu, Jose M Such, and Guillermo Suarez-Tangil. 2020. Smart home personal assistants: a security and privacy review. *ACM Computing Surveys (CSUR)* 53, 6 (2020), 1–36.

[10] R Scott Evans. 2016. Electronic health records: then, now, and in the future. *Yearbook of medical informatics* 25, S 01 (2016), S48–S61.

[11] Fortune. 2024. Apple, Google, and Amazon May Have Violated Your Privacy by Reviewing Digital Assistant Commands. https://fortune.com/2019/08/05/google-apple-amazon-digital-assistants/. [Online; accessed September-12-2024].

[12] Google Gemini. 2023. Introducing Gemini: our largest and most capable AI model. https://blog.google/technology/ai/google-gemini-ai/#sundar-note

[13] Jacopo Guanetti, Yeojun Kim, and Francesco Borrelli. 2018. Control of connected and automated vehicles: State of the art and future challenges. *Annual reviews in control* 45 (2018), 18–40.

[14] Umar Iqbal, Tadayoshi Kohno, and Franziska Roesner. 2023. LLM Platform Security: Applying a Systematic Evaluation Framework to OpenAI's ChatGPT Plugins. *arXiv preprint arXiv:2309.10254* (2023).

[15] Arun Iyengar, Dhaval Patel, Shrey Shrivastava, Nianjun Zhou, and Anuradha Bhamidipaty. 2020. Real-Time Data Quality Analysis. In *2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI)*. 101–108. https://doi.org/10.1109/CogMI50398.2020.00022

[16] Ehud Karpas, Omri Abend, Yonatan Belinkov, Barak Lenz, Opher Lieber, Nir Ratner, Yoav Shoham, Hofit Bata, Yoav Levine, Kevin Leyton-Brown, Dor Muhlgay, Noam Rozen, Erez Schwartz, Gal Shachaf, Shai Shalev-Shwartz, Amnon Shashua, and Moshe Tenenholtz. 2022. MRKL Systems: A modular, neuro-symbolic architecture that combines large language models, external knowledge sources and discrete reasoning. arXiv:2205.00445 [cs.CL] https://arxiv.org/abs/2205.00445

[17] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (2023), 102274.

[18] Yaa A Kumah-Crystal, Claude J Pirtle, Harrison M Whyte, Edward S Goode, Shilo H Anders, and Christoph U Lehmann. 2018. Electronic health record interactions through voice: a review. *Applied clinical informatics* 9, 03 (2018), 541–552.

[19] Song Liao, Christin Wilson, Long Cheng, Hongxin Hu, and Huixing Deng. 2020. Measuring the Effectiveness of Privacy Policies for Voice Assistant Applications. http://arxiv.org/abs/2007.14570

[20] Amarachi B Mbakwe, Ismini Lourentzou, Leo Anthony Celi, Oren J Mechanic, and Alon Dagan. 2023. ChatGPT passing USMLE shines a spotlight on the flaws of medical education. , e0000205 pages.

[21] Microsoft. 2024. Microsoft Trustworthy AI: Unlocking human potential starts with trust. https://blogs.microsoft.com/blog/2024/09/24/microsoft-trustworthy-ai-unlocking-human-potential-starts-with-trust/

[22] Hammond Pearce, Baleegh Ahmad, Benjamin Tan, Brendan Dolan-Gavitt, and Ramesh Karri. 2022. Asleep at the keyboard? assessing the security of github copilot's code contributions. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 754–768.

[23] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. 1996. Role-Based Access Control Models. *Computer* 29, 2 (Feb. 1996), 38–47.

[24] D. K. Smetters and Nathan Good. 2009. How users use access control. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (Mountain View, California, USA) *(SOUPS '09)*. Association for Computing Machinery, New York, NY, USA, Article 15, 12 pages. https://doi.org/10.1145/1572532.1572552

[25] Zhongxiang Sun. 2023. A short survey of viewing large language models in legal aspect. *arXiv preprint arXiv:2303.09136* (2023).

[26] DeepSeek-AI Research Team. 2025. DeepSeek-V3 Technical Report. (February 2025). https://arxiv.org/pdf/2412.19437

[27] Fatih Turkmen, Jerry den Hartog, Silvio Ranise, and Nicola Zannone. 2017. Formal analysis of XACML policies using SMT. *Computers & Security* 66 (2017), 185–203. https://doi.org/10.1016/j.cose.2017.01.009

[28] Chris Varner. 2025. Understanding Different ChatGPT Models: Key Details to Consider. https://teamai.com/blog/large-language-models-llms/understanding-different-chatgpt-models/

[29] Joel Wester, Tim Schrills, Henning Pohl, and Niels van Berkel. 2024. "As an AI language model, I cannot": Investigating LLM Denials of User Requests. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.

[30] Binwei Yao, Ming Jiang, Diyi Yang, and Junjie Hu. 2023. Benchmarking llm-based machine translation on cultural awareness. *arXiv preprint arXiv:2305.14328* (2023).

[31] Yifan Yao, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Zhibo Sun, and Yue Zhang. 2024. A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing* (2024), 100211.

8 APPENDIX A: USER-MANAGED ACCESS CONTROL POLICIES AND SAMPLE RESULTS

Table 3: Comparison of different U-MAP encoding approaches (formal, informal, semi-formal) for Smart Cars, including policy variants.

Approach	U-MAP
XACML (Formal)	<policy, CA=First-Applicable> <rule result=Deny> <target>Directions</target> <cond.>role=Co-Pilot</cond.> </rule> </policy>
Informal **(INF)**	*Only drivers should be allowed to give driving directions to the smart car. Only drivers and co-pilots should be allowed to modify smart car settings. Other passengers should be only allowed to change the radio settings.*
Modified Informal **(Mod-INF)**	*Kids can't give directions, but co-pilots can.*
Semi-Formal **(SEMI-UNROLL)**	1. Co-Pilots cannot give Directions, 2. Passengers cannot give Directions, ... 4. Everything else is Allowed.
Modified Semi-Formal **(Mod-SEMI-UNROLL)**	Create new Rule: Kids cannot give Directions; Remove Rule: Co-Pilots cannot give Directions

Semi-Formal-Rule-Based **(SEMI-RULE)**	if Role = Co-Pilot:
	Directions = denied,
	if Role = Passenger:
	Settings = denied
	& Directions = denied,
	Default Radio
	& Directions
	& Settings = approved

Modified Semi-Formal-Rule-Based **(Mod-SEMI-RULE)**	if Role = Kids:
	Directions = denied,
	if Role = Co-Pilot:
	Directions = approved

Table 4: Comparison of different U-MAP encoding approaches (formal, informal, semi-formal) for Electronic Health Records (EHRs), including policy variants.

Approach	U-MAP
XACML (Formal)	<policy, CA=First-Applicable> <rule result=Allow> <target>Medical</target> <cond.>role=Physician</cond.> </rule> <rule result=Allow>...</rule> </policy>
Informal **(INF)**	*Only Physicians, Nurses, and First Responders are allowed to access Medical Data.* *Only Admin Staff is allowed to access Personal Identifiable Information (PII) Data.*
Modified Informal **(Mod-INF)**	*First Responders can now access PII-Data and Staff can now access Medical Data.*
Semi-Formal **(SEMI-UNROLL)**	1. Physicians can access Medical Data, 2. Nurses can access Medical Data, ... 9. Everything else is denied.
Modified Semi-Formal **(Mod-SEMI-UNROLL)**	Create new Rule: "First Responders can access PII- Data"; Remove Rule: "First Responders cannot access PII-Data"; Create new Rule: "Staff can access Medical Data"
Semi-Formal-Rule-Based **(SEMI-RULE)**	if role = Staff: PII-Data = Approve; if role = Physician, Nurse, First Responders: Medical Data = Approved; Default Access = Denied
Modified Semi-Formal-Rule-Based **(Mod-SEMI-RULE)**	if role = Staff, First Responders: PII-Data = Approve; if role = Physician, Nurse, First Responders, Staff: Medical Data = Approved; Default Access = Denied

Table 5: VA responses to Smart Home access questions using various U-MAP encodings (original and modified), with and without context.

		ChatGPT				Google Gemini				Microsoft Copilot				Deepseek			
		Context		No Context		Context		No Context		Context		No Context		Context		No Context	
Question	GT	R-1	R- 2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2
INF Format																	
Q-SM-1	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Q-SM-2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-5	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Modified INF Format																	
Q-SM-1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓
Q-SM-2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-4	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗
Q-SM-5	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
SEMI-UNROLL Format																	
Q-SM-1	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Q-SM-2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-4	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-5	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Modified SEMI-UNROLL Format																	
Q-SM-1	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-4	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Q-SM-5	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
SEMI-RULE Format																	
Q-SM-1	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
Q-SM-2	✓	✓	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✗	✗	✗	✓	✓
Q-SM-3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓
Q-SM-4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-5	✗	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗
Modified SEMI-RULE Format																	
Q-SM-1	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Q-SM-2	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓
Q-SM-3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓
Q-SM-4	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Q-SM-5	✗	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗