Sara Kim
CSCI 598B SP TPS: IOT Privacy and Security
Fall 2023
Final Paper

"Hey Siri, Who's Listening?"
Investigating Privacy Concerns in Voice-Controlled Systems

Abstract

This research project delves into the voice functionality of Siri on iMac, with a primary focus on unintended activations and associated implications for IoT security and privacy. Through meticulous testing, the study uncovers a significant disparity between Siri activations in on-state and off-state scenarios, revealing potential vulnerabilities and raising privacy concerns. Unexpected triggers, particularly during multi-person conversations, underscore the need for algorithmic refinement and enhanced user awareness. The study's contributions lie in unraveling the complexities of Siri's behavior, emphasizing the importance of addressing privacy and security challenges in the ever-expanding landscape of voice-controlled IoT devices. As voice interfaces become integral to daily life, this research advances our understanding of Siri's behavior, laying the groundwork for responsible development and deployment of secure and privacy-preserving voice-controlled systems.

Introduction

The proliferation of Internet of Things (IoT) devices has reached unprecedented levels, introducing challenges in privacy and security [1]. The surge in connected devices magnifies the potential risks associated with unauthorized data access and surveillance [2]. Siri, Apple's voice-controlled assistant embedded in iMac systems, stands as a quintessential IoT device. It serves as an interface for users to interact with their computers through voice commands that allow a spectrum of functionalities, from executing commands to enabling hands-free operation [3]. Nevertheless, the possibility of Siri inadvertently activating, especially when the iMac is seemingly turned off or inactive, raises substantial concerns regarding privacy and security. This behavior suggests that users might unwittingly expose their conversations or sensitive information, as Siri could activate, listen, and potentially transmit data without explicit user consent.

The primary objective of this project is to conduct an in-depth investigation into the voice functionality of Siri on iMac, with a specific focus on instances where Siri appears to activate and listen without user initiation. This inquiry is particularly pertinent when the iMac appears to be asleep. Understanding the mechanisms behind unintended activations is crucial for mitigating privacy and security risks associated with voice-controlled IoT devices. This research endeavor seeks to mitigate these risks by investigating, understanding, and delineating the implications of Siri's voice functionality on iMac concerning unintended activations.

Literature Review

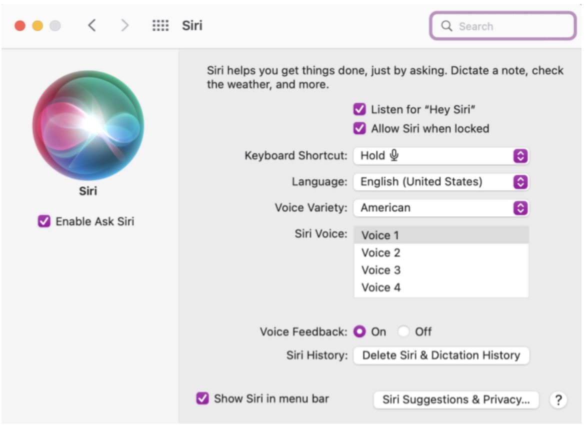
The existing body of literature extensively delineates Siri's designed functionality. Operated through user-initiated "Hey Siri" commands, the system employs a small speech recognizer that consistently remains active, listening for the trigger words. To optimize accuracy and reduce false triggers, users are prompted to enroll in a brief session, articulating "Hey Siri" phrases that subsequently contribute to training a Deep Neural Network (DNN). The anticipated response involves the prompt transmission of a cancellation signal if a phrase other than "Hey Siri" is detected [3].

Research addressing the privacy and security concerns associated with Siri's functionality spans various dimensions. Unintended activations present a critical issue, potentially leading to unauthorized recordings. Opaque data practices, marked by a lack of transparency in audio data collection, processing, and usage by Virtual Cognitive Digital Assistants (VCDA) manufacturers, give rise to significant privacy concerns. The processing of unnecessary personal information by Siri alongside voice commands, coupled with limited user control over recorded data, introduces additional privacy risks. The richness of audio data enables the inference of personal details, exposing users to profiling and targeted advertisements. Additionally, there are overarching security concerns, including vulnerabilities related to network traffic [4].

In response to these pressing privacy and security concerns, VoiceGuard emerges as a promising solution, presenting an architecture for privacy-preserving speech processing on untrusted systems. This versatile solution accommodates various deployment scenarios, including third-party, user, or vendor deployment as service providers [5]. However, it is worth noting that the widespread accessibility of VoiceGuard is lacking.

Despite these collective efforts, literature addressing unexpected activations remains limited. A study on voice-control interaction in blind participants briefly touches upon the phenomenon, but fails to go into detail or address it directly [6]. The scarcity of research in this area underscores a substantial gap in understanding unintended Siri activations, especially considering their potential privacy implications.

The primary contributions of this project, then, revolve around its specific focus on unintended Siri activations. By delving into the specifics of Siri's unintended activations, the project aims to deepen our understanding of potential privacy implications in this domain.



Siri helps you get things done, just by asking. Dictate a note, check the weather, and more.

☑ Enable Ask Siri

☑ Listen for "Hey Siri"

☑ Allow Siri when locked

Keyboard Shortcut: Hold 🎙

Language: English (United States)

Voice Variety: American

Siri Voice: Voice 1
Voice 2
Voice 3
Voice 4

Voice Feedback: ● On ○ Off

Siri History: Delete Siri & Dictation History

☑ Show Siri in menu bar

Siri Suggestions & Privacy... ?

Figure 1. Siri configuration

Methodology

The iMac model employed in this study is a 24-inch 2021 iMac with an Apple M1 chip running macOS Monterey Version 12.1. The specific configuration of Siri settings is pictured in Figure 1. The technical attributes of the iMac and

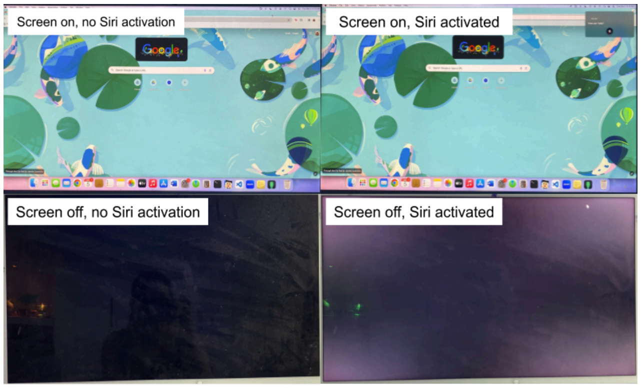
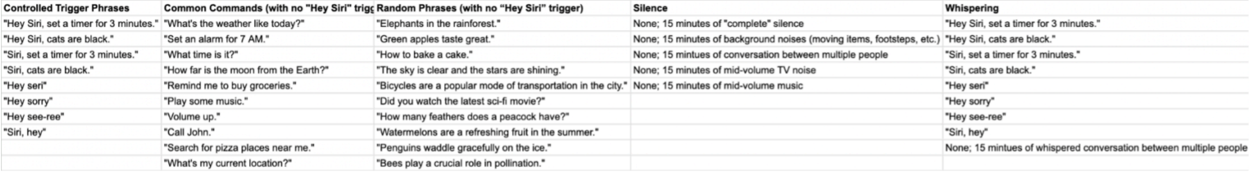

Screen on, no Siri activation
Screen on, Siri activated
Screen off, no Siri activation
Screen off, Siri activated

Figure 2. Screen states

nuanced Siri settings can significantly impact Siri responsiveness and behavior, warranting detailed attention during the investigative phase. When the iMac screen is active, Siri activation manifests through a distinctive pop-up notification located in the upper right corner. This visual cue serves as an indicator of Siri's engagement and responsiveness. In contrast, when the iMac screen is inactive or turned off, Siri activation is discerned through a nuanced behavior. The screen becomes backlit, deviating from complete darkness, as shown in Figure 2. This subtle change can easily be overlooked, which warrants some suspicion as to the purpose of such a design. This variability in activation indication introduces an additional layer of complexity, prompting a thorough examination of Siri's behavior in different states of desktop activity. Data collection is based on observation of changes in these screen states in response to audio input.

The study encompasses a comprehensive analysis of Siri's voice functionality on a personal iMac. Controlled experiments have been meticulously devised to observe and record instances of Siri activation under specific conditions, particularly when the desktop screen is inactive. See Figure 3 for a complete list of tests. Each test scenario was executed in both the screen active (on) and screen inactive (off) states. This dual testing approach aims to discern potential variations in Siri's behavior based on the active or inactive status of the desktop. To ensure the reliability and repeatability of the experiments, each test was run three times in both of the specified desktop states. This iterative process aids in capturing potential nuances or variations in Siri's activation patterns, contributing to a robust and detailed analysis of its behavior. The methodological framework outlined above establishes a systematic approach for investigating Siri's voice functionality on the iMac, offering a structured foundation for the subsequent data collection and analysis phases of this research endeavor.

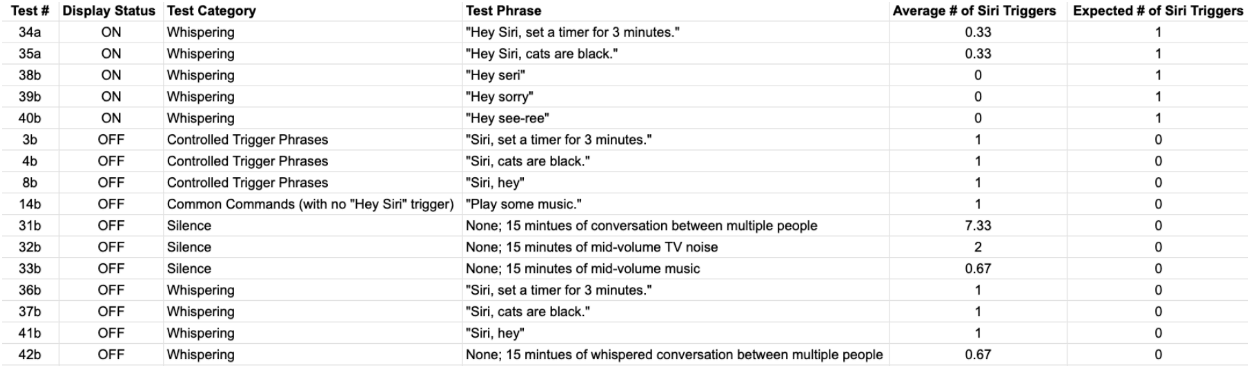


Controlled Trigger Phrases	Common Commands (with no "Hey Siri" trigg	Random Phrases (with no "Hey Siri" trigger)	Silence	Whispering
"Hey Siri, set a timer for 3 minutes."	"What's the weather like today?"	"Elephants in the rainforest."	None; 15 minutes of "complete" silence	"Hey Siri, set a timer for 3 minutes."
"Hey Siri, cats are black."	"Set an alarm for 7 AM."	"Green apples taste great."	None; 15 minutes of background noises (moving items, footsteps, etc.)	"Hey Siri, cats are black."
"Siri, set a timer for 3 minutes."	"What time is it?"	"How to bake a cake."	None; 15 mintues of conversation between multiple people	"Siri, set a timer for 3 minutes."
"Siri, cats are black."	"How far is the moon from the Earth?"	"The sky is clear and the stars are shining."	None; 15 minutes of mid-volume TV noise	"Siri, cats are black."
"Hey seri"	"Remind me to buy groceries."	"Bicycles are a popular mode of transportation in the city."	None; 15 minutes of mid-volume music	"Hey seri"
"Hey sorry"	"Play some music."	"Did you watch the latest sci-fi movie?"		"Hey sorry"
"Hey see-ree"	"Volume up."	"How many feathers does a peacock have?"		"Hey see-ree"
"Siri, hey"	"Call John."	"Watermelons are a refreshing fruit in the summer."		"Siri, hey"
	"Search for pizza places near me."	"Penguins waddle gracefully on the ice."		None; 15 mintues of whispered conversation between multiple people
	"What's my current location?"	"Bees play a crucial role in pollination."		

Figure 3. All tests

Data Collection and Analysis

The data collection and analysis phase of this investigation rigorously scrutinized Siri's responsiveness across a spectrum of scenarios, encompassing both anticipated and unforeseen triggering conditions. A conspicuous incongruity surfaced in the instances of unexpected triggers, revealing a pronounced dissimilarity between the off-state (14 occurrences) and the on-state (2 occurrences) scenarios, indicating a heightened susceptibility to unintended activations when the iMac was ostensibly inactive. A detailed breakdown of unexpected

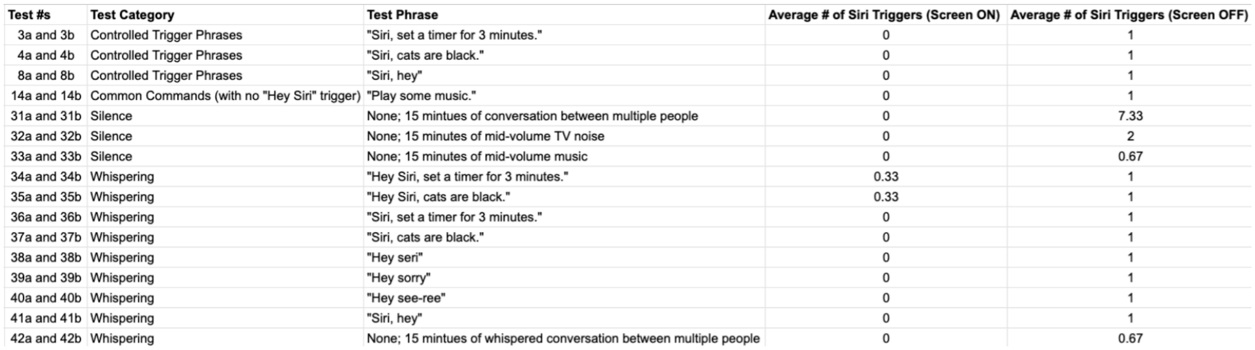


Test #	Display Status	Test Category	Test Phrase	Average # of Siri Triggers	Expected # of Siri Triggers
34a	ON	Whispering	"Hey Siri, set a timer for 3 minutes."	0.33	1
35a	ON	Whispering	"Hey Siri, cats are black."	0.33	1
38b	ON	Whispering	"Hey seri"	0	1
39b	ON	Whispering	"Hey sorry"	0	1
40b	ON	Whispering	"Hey see-ree"	0	1
3b	OFF	Controlled Trigger Phrases	"Siri, set a timer for 3 minutes."	1	0
4b	OFF	Controlled Trigger Phrases	"Siri, cats are black."	1	0
8b	OFF	Controlled Trigger Phrases	"Siri, hey"	1	0
14b	OFF	Common Commands (with no "Hey Siri" trigger)	"Play some music."	1	0
31b	OFF	Silence	None; 15 mintues of conversation between multiple people	7.33	0
32b	OFF	Silence	None; 15 minutes of mid-volume TV noise	2	0
33b	OFF	Silence	None; 15 minutes of mid-volume music	0.67	0
36b	OFF	Whispering	"Siri, set a timer for 3 minutes."	1	0
37b	OFF	Whispering	"Siri, cats are black."	1	0
41b	OFF	Whispering	"Siri, hey"	1	0
42b	OFF	Whispering	None; 15 mintues of whispered conversation between multiple people	0.67	0

Figure 4. Discrepancies in Expected Number of Triggers

behaviors is available in Figure 4. Intriguingly, on-state disparities exclusively materialized in situations where triggers were expected but yielded fewer activations. It was specifically the whispering tests, such as the whispered "Hey Siri" commands and phonetically similar variations, that consistently fell short of expectations, registering averages of 0.33 and 0 triggers, contrary to the anticipated 1 trigger. This unexpected outcome prompts an inquiry into the reliability of these ostensibly dependable trigger scenarios. Conversely, off-state discrepancies solely manifested in instances where triggers were unanticipated, resulting in more activations than expected. Notably, commands commencing with "Siri" consistently induced activations, both when spoken and whispered, contrary to the expected 0 triggers. Furthermore, unforeseen triggers occurred during 15-minute intervals of conversation (average of 7.33 triggers), TV activity (average of 2 triggers), music playback (average of 0.67 triggers), and whispered conversation (average of 0.67 triggers). The command "Play some music" also unexpectedly triggered, averaging 1 activation when none was anticipated.

Significant disparities between on-state and off-state results, detailed in Figure 5, further underscored the sensitivity discrepancies between these two states. Notably, the off-state exhibited heightened responsiveness to all whispering tests, consistently triggering Siri, with an average number of activations of 1, except for the 15-minute whispered conversation, which averaged 0.67. In contrast, on-state whisper tests infrequently triggered Siri, with an average number of activations of 0 for all cases except phrases starting with "Hey Siri," which averaged 0.33. In all instances, the off-state displayed greater sensitivity to whispered commands. Additional disparities were observed in phrases commencing with "Siri" (as opposed to “Hey



Test #s	Test Category	Test Phrase	Average # of Siri Triggers (Screen ON)	Average # of Siri Triggers (Screen OFF)
3a and 3b	Controlled Trigger Phrases	"Siri, set a timer for 3 minutes."	0	1
4a and 4b	Controlled Trigger Phrases	"Siri, cats are black."	0	1
8a and 8b	Controlled Trigger Phrases	"Siri, hey"	0	1
14a and 14b	Common Commands (with no "Hey Siri" trigger)	"Play some music."	0	1
31a and 31b	Silence	None; 15 mintues of conversation between multiple people	0	7.33
32a and 32b	Silence	None; 15 minutes of mid-volume TV noise	0	2
33a and 33b	Silence	None; 15 minutes of mid-volume music	0	0.67
34a and 34b	Whispering	"Hey Siri, set a timer for 3 minutes."	0.33	1
35a and 35b	Whispering	"Hey Siri, cats are black."	0.33	1
36a and 36b	Whispering	"Siri, set a timer for 3 minutes."	0	1
37a and 37b	Whispering	"Siri, cats are black."	0	1
38a and 38b	Whispering	"Hey seri"	0	1
39a and 39b	Whispering	"Hey sorry"	0	1
40a and 40b	Whispering	"Hey see-ree"	0	1
41a and 41b	Whispering	"Siri, hey"	0	1
42a and 42b	Whispering	None; 15 mintues of whispered conversation between multiple people	0	0.67

Figure 5. Discrepancies between On-state and Off-state

Siri"); off-state "Siri" tests—both spoken and whispered—unfailingly triggered, while the on-state never exhibited activations. Some specific triggers were exclusively observed in the off-state, including during 15-minute intervals of TV noise (average of 2 triggers), 15-minute intervals of music noise (average of 0.67 triggers), and the command "Play some music" (average of 1 trigger). A substantial red flag surfaced in the comparison of triggers during a 15-minute conversation between multiple people; the on-state showed no triggers (0 on average), while the off-state displayed a substantial average of 7.33 triggers, indicating a potential vulnerability in multi-person conversational contexts. Collectively, the results reveal evident disparities between on-state and off-state triggering behaviors, prompting concerns regarding privacy and security.

Another notable intriguing behavior manifested in phrases with similar sounds (e.g., "Hey seri," "Hey sorry," "Hey see-ree"): Siri consistently responded to these spoken (not whispered) phrases, irrespective of the on or off state. While this behavior aligns with expectations, it is noteworthy as it underscores Siri's accessibility across diverse accents and pronunciation nuances.

Privacy and Security Assessment

The unexpected behavior observed in Siri triggers may find partial explanation in Siri's consistent response to phrases bearing similarity to the designated wake word, such as "Hey seri," "Hey sorry," or "Hey see-ree." It is conceivable that unforeseen activations, particularly in the context of a 15-minute conversation, might be attributable to the use of phrases with similar sounds during speech. This observation implies a potential avenue for algorithmic improvement in Siri's wake word detection mechanism. An enhanced algorithm could augment Siri's capacity to accurately discern and respond to a wider array of voice inputs, encompassing variations in pronunciation.

However, the persistent dissimilarities in Siri's behavior between the on and off states remain elusive in their explanation. While the advantages in accessibility and algorithmic refinement may contribute to Siri's consistent response to certain phrases, the explicit distinctions in activation patterns contingent upon the desktop state evoke compelling queries. Delving into potential explanations for this disparity necessitates a nuanced investigation, considering factors such as background processes, power management, or inadvertent audio cues that might influence Siri's responsiveness.

Notwithstanding these uncertainties, unintended Siri activations raise critical privacy concerns, as they bear the potential of capturing and transmitting user conversations or sensitive information without explicit consent. The inadvertent interception of audio data during unintended activations poses a tangible risk of unauthorized access, thereby engendering potential privacy breaches. Users may inadvertently disclose personal or confidential information during unintended interactions with Siri, underscoring the imperative need for robust privacy safeguards. Furthermore, the security risks associated with Siri's functioning in unintended situations extend beyond privacy concerns. Unintended activations could render users susceptible to malicious exploitation, with the possibility of unauthorized eavesdropping or interception of sensitive information. A comprehensive understanding and effective mitigation of these security risks are indispensable to prevent potential abuse of voice-controlled interfaces and safeguard user information from unauthorized access or malicious activities.

This privacy and security assessment underscores the multifaceted nature of unintended Siri activations, integrating considerations of accessibility advantages, algorithmic refinement, on/off state disparities, and the associated risks. A thorough examination of these aspects is paramount for advancing the development of secure and privacy-preserving voice-controlled systems within the realm of IoT devices.

Recommendations and Future Work

Given the potential privacy and security concerns identified in this study, a series of recommendations are proposed to mitigate risks and bolster privacy, along with suggestions for advancing research in IoT security. First and foremost, there appears to be a need for the enhancement of Siri's wake word detection algorithm to mitigate unintended activations, fostering a more discerning response to user-initiated commands. Continuous refinement of the algorithm, guided by user feedback and exposure to diverse linguistic inputs, is of paramount importance. Additionally, the implementation of mechanisms designed to augment user awareness and control over Siri's activation behavior could prove beneficial. Empowering users with granular settings to customize activation parameters, coupled with educational initiatives elucidating potential privacy implications, can facilitate informed decision-making. Furthermore, a comprehensive privacy impact assessment may be beneficial to delineate potential risks and implement robust measures for data encryption and secure transmission during voice interactions. This includes the incorporation of stringent anonymization techniques to safeguard user identities during data processing.

Future research endeavors in this area could diversify test scenarios to encompass a broader range of real-world situations, incorporating variables such as varying background noise levels, accents, and environmental conditions. This expansion aims to contribute to a more comprehensive understanding of Siri's behavior and enhance its adaptability in diverse contexts. Moreover, future work might leverage advanced network traffic monitoring tools to scrutinize data exchanges between the iMac and external servers during Siri interactions. This analysis would elucidate the nature and extent of data transmitted, facilitating a more nuanced comprehension of potential security vulnerabilities. A detailed examination of the data recorded during Siri activations could provide valuable insights by identifying the types of information stored, the duration of storage, and the purposes for which the data is utilized. Such scrutiny is pivotal in enhancing transparency and ensuring adherence to privacy standards. Additionally, further tests should extend to diverse devices, both within and beyond the Apple ecosystem. Assessing Siri's behavior on other Apple products and exploring potential variations in activation patterns on non-Apple voice recognition devices would offer valuable insights into Siri's interoperability and security across platforms. These recommendations and avenues for future research collectively contribute to the ongoing efforts to enhance the privacy and security of voice-controlled interfaces within the realm of IoT devices.

Conclusion

This study has systematically investigated the voice functionality of Siri on iMac, shedding light on nuanced behaviors and unexpected triggers that hold implications for both privacy and security in the realm of voice-controlled IoT devices. The key findings revealed a notable disparity between on-state and off-state Siri activations, prompting concerns about unintended and potentially unauthorized interactions in specific scenarios. Despite Siri's consistent response to phrases sounding similar to its wake word, the unexplained disparities raise questions about the underlying mechanisms and necessitate a more nuanced investigation. Unintended Siri activations, particularly in off-state scenarios, pose critical privacy concerns as they could lead to the inadvertent interception and transmission of user conversations or sensitive information without explicit consent. Beyond privacy implications, the security risks associated with unintended activations extend to the potential exploitation of voice-controlled interfaces, emphasizing the need for comprehensive safeguards.

This research contributes to the understanding of Siri's voice functionality on iMac, highlighting the need for algorithmic refinement, enhanced user awareness, and privacy safeguards. The multifaceted nature of unintended activations underscores the complexities inherent in voice-controlled systems, emphasizing the ongoing importance of addressing privacy and security challenges in the rapidly evolving domain of IoT devices. As voice-controlled interfaces become increasingly integrated into daily life, safeguarding user privacy and security remains paramount for the responsible development and deployment of such technologies.

References

[1] D. Kumar *et al.*, “All things considered: an analysis of IoT devices on home networks,” *SEC’19: Proceedings of the 28th USENIX Conference on Security Symposium*, pp. 1169–1185, Apr. 2019.

[2] L. Tawalbeh, F. Muheidat, M. Tawalbeh, and M. Quwaider, “IOT privacy and security: Challenges and solutions,” *Applied Sciences*, vol. 10, no. 12, Jun. 2020. doi:10.3390/app10124102

[3] Siri Team, “Hey Siri: An on-device DNN-powered voice trigger for Apple’s personal assistant,” Apple Machine Learning Research, https://machinelearning.apple.com/research/hey-siri (accessed Dec. 5, 2023).

[4] L. Hernández Acosta and D. Reinhardt, “A survey on privacy issues and solutions for voice-controlled Digital assistants,” *Pervasive and Mobile Computing*, vol. 80, Feb. 2022. doi:10.1016/j.pmcj.2021.101523

[5] F. Brasser *et al.*, “Voiceguard: Secure and private speech processing,” *Interspeech 2018*, pp. 1303–1307, Sep. 2018. doi:10.21437/interspeech.2018-2032

[6] A. Abdolrahmani, R. Kuber, and S. M. Branham, “‘Siri talks at you’: An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind,” *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 249–258, Oct. 2018. doi:10.1145/3234695.3236344