

Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies

Radwan Eshkita ^{1,✉}, Andrin Heeb ², Elias Gmuender ¹, Vijaya Kittu Manda ³

1. Computer Science and Information Technology, Lucerne University of Applied Sciences, Lucerne, CHE

2. Informatics, Hochschule Luzern, Lucerne, CHE

3. Computer Science, BEST Innovation University, Gorantla, IND

Received: June 30, 2025 | Review began: January 23, 2026 | Review ended: February 19, 2026 | Published: May 28, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

The growing use of voice-activated personal assistants (VAPAs) has introduced significant user privacy and data security challenges. This study investigates privacy vulnerabilities in VAPAs using a mixed-methods approach that combines simulated experiments with structured risk assessment. Through controlled simulations, we evaluate vulnerabilities of VAPAs to various attack vectors, including adversarial attacks on automatic speech recognition (ASR), voice cloning, and side-channel attacks. A risk assessment framework using DREAD (Damage, Reproducibility, Exploitability, Affected Users, Discoverability) and STRIDE (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege) models quantifies the identified vulnerabilities. Results reveal that VAPAs are highly vulnerable to adversarial attacks (85-90% success rate), and voice cloning can effectively circumvent authentication measures (up to 92% authentication bypass rate). The discussion analyzes these findings in the context of existing literature, highlighting the limitations of current security measures. The paper concludes with mitigation strategies to enhance user privacy and data security in VAPAs. These include enhanced authentication methods, improved ASR robustness, data minimization techniques, robust security protocols, and proactive threat detection systems. Addressing privacy risks in VAPAs requires a multi-faceted approach that integrates technological innovations, sound regulatory policies, and comprehensive user education initiatives. The study has practical applications for technology developers and VAPA providers.

Categories: Ethical and Social Implications, AI Ethics and Fairness, Security and Privacy

Keywords: automatic speech recognition, voice cloning, voice-activated personal assistants, adversarial attacks, privacy vulnerability

Introduction

Voice-Activated Personal Assistants (VAPAs) and Speech Recognition Systems like Alexa and Siri have revolutionized human-technology interaction. They connect well with other technologies, like Artificial Intelligence (AI) and Internet of Things, especially in home automation. Yet, their pervasive data collection poses unresolved privacy risks [1]. While prior work highlights vulnerabilities in ASR systems [2] and voice cloning [3], few studies integrate empirical testing with quantitative risk frameworks like DREAD (Damage, Reproducibility, Exploitability, Affected Users, Discoverability)/STRIDE (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege) [4]. This study bridges the gap through a mixed-methods approach that combines a comprehensive literature review, controlled simulations, and structured risk assessment. We examine adversarial attacks on ASR systems, the feasibility of voice cloning to bypass authentication, and risks of data leakage via side channels. DREAD and STRIDE models are employed to systematically evaluate the identified vulnerabilities and quantify the potential impact on user privacy.

How to cite this article:

Eshkita R, Heeb A, Gmuender E, et al. (May 28, 2026) Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies. Cureus J Comput Sci 3 : es44389-025-00074-8. DOI https://doi.org/10.7759/s44389-025-00074-8

The existing literature highlights concerns about VAPA data-collection practices, the vulnerabilities of automatic speech recognition (ASR) systems, and the potential for malicious exploitation of these systems. However, current research on VAPA privacy often lacks a holistic approach that integrates empirical evidence with quantitative risk assessment and actionable mitigation strategies. This paper addresses this gap by providing simulation-based evidence of key vulnerabilities, quantifying their risks using the DREAD and STRIDE models, and proposing novel mitigation strategies. This study uniquely combines empirical attack simulations with quantitative risk modelling to deliver a comprehensive assessment of VAPA vulnerabilities. By integrating DREAD and STRIDE frameworks, we provide a structured, scalable approach to prioritizing privacy risks in voice-enabled systems. Our methodology enables not only the detection of vulnerabilities but also their systematic ranking based on real-world exploitability and impact.

Literature review: Privacy risks and vulnerabilities in VAPAs

VAPAs like Amazon Alexa, Apple Siri, and Google Assistant have entirely changed how humans interact with technology. They have become indispensable for many people because they can seamlessly fit into daily life and help in independent living [5]. They have excellent voice recognition abilities and are thereby able to understand human voice with utmost accuracy and help human in managing tasks, get information on various aspects including health [6,7], and control smart home devices [8]. They are used in improving shopping experience, opening doors for a new field called voice commerce [9]. However, this convenience comes at a considerable cost: the pervasive data-collection practices of VAPAs, coupled with the inherent vulnerabilities of their underlying technologies, pose significant, largely unaddressed risks to user privacy and data protection [1]. These devices collect and process vast amounts of personal data, raising serious concerns about misuse, unauthorized access, and violations of user privacy. This section reviews the relevant literature, examining research on privacy risks associated with VAPAs, highlighting key vulnerabilities, and outlining existing mitigation strategies.

Data Collection Practices

VAPAs are known to collect user data to function effectively continuously. Many devices are by default “always-on,” so they only require a wake word to initiate an operation. This “always listening” raises significant concerns about constant surveillance and the potential for unintended data capture [10]. The policies governing data retention and usage vary widely among providers. A quick comparison shows several inconsistencies in user privacy protections [11]. Moreover, sharing user data with third-party developers and service providers increases risks. This is more so in data breaches and unauthorized access [12].

Vulnerabilities of ASR Systems

The underlying ASR systems powering VAPAs are susceptible to attacks. Each of these attacks comes with significant privacy implications. The top four leading vulnerabilities are Adversarial Attacks, Voice Cloning and Impersonation, Replay Attacks, and Side-Channel Attacks. Research has demonstrated the vulnerability of ASR systems to adversarial attacks. In these attacks, carefully crafted imperceptible audio perturbations can manipulate the system’s transcriptions [13]. These attacks can lead to the execution of unauthorized commands, the disclosure of sensitive information, or the subversion of user intent. Advances in voice cloning technologies pose a significant threat to user privacy [3]. Attackers could potentially create highly realistic synthetic voices capable of bypassing voice authentication mechanisms and gaining unauthorized access to user accounts and services. Replay Attacks are the susceptibility of VAPAs to replay attacks. In these attacks, the recordings of legitimate voice commands are used to execute unauthorized actions. These attacks highlight a critical weakness in current authentication methods [14]. It reminds users of the need for stronger authentication techniques. Analyzing patterns and correlations in VAPA responses based on user interactions can reveal sensitive information. These are called “side-channel leakage” and can expose user location, routines, preferences, and other personal data details [15].

Existing Mitigation Strategies and Privacy-Preserving Technologies

How to cite this article:

Eskhita R, Heeb A, Gmuender E, et al. (May 28, 2026) Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies. *Cureus J Comput Sci* 3 : es44389-025-00074-8. DOI <https://doi.org/10.7759/s44389-025-00074-8>

Researchers have explored various techniques to mitigate the privacy risks associated with VAPAs. These aspects fall into three major areas - Voice Liveness Detection, Privacy-Preserving Technologies (PETs), and Enhanced Security Protocols. Voice Liveness Detection aims to distinguish between live voice and recordings or synthesized speech. It offers a crucial line of defense [16]. However, the efficacy of these methods is often limited by the sophisticated nature of current attack techniques. PETs are the application of privacy-enhancing technologies such as differential privacy and federated learning. The PET-based technologies promise to enhance user privacy while maintaining system functionality [17]. The recent development of Enhanced Security Protocols fills the gap for security protocols. These include encryption and access control mechanisms that can reduce the risk of data breaches and unauthorized access [18].

Legal and Ethical Considerations

The use of VAPAs raises important legal and ethical issues around data protection, user consent, and transparency. Regulations such as California Consumer Privacy Act and the General Data Protection Regulation aim to protect user data. However, their application to the unique challenges VAPAs pose requires further examination. The ethical implications of constant data collection, data retention practices, and the potential for data misuse call for a discussion and debate [19].

Materials And Methods

Methodology

The overall research goal is to comprehensively evaluate privacy vulnerabilities in VAPAs. To accomplish this, the study uses a mixed-methods approach combining quantitative simulations with qualitative risk assessment. This dual approach enables systematic validation of vulnerabilities while contextualizing their severity and likelihood. The methodology integrates three carefully designed simulated experiments with a structured risk assessment framework to provide systematic evidence of vulnerabilities and a quantitative evaluation of their potential impact.

The research methodology is designed in three interconnected phases:

- 1) Experimental Design
- 2) Risk Assessment Framework
- 3) Data Analysis

Experimental design

Three controlled simulations were designed to test VAPAs' vulnerabilities to various attack vectors. These simulations focus on the susceptibility to adversarial attacks, voice cloning techniques, and side-channel data extraction. Each simulation tests a specific hypothesis related to a particular privacy risk.

Experiment 1: Adversarial Attacks on ASR

This experiment evaluated the susceptibility of ASR systems to adversarial attacks. Adversarial samples were generated using the Fast Gradient Sign Method (FGSM), a widely recognized technique for crafting subtle perturbations. These perturbations were designed to be imperceptible to human listeners while deceiving the ASR systems into producing erroneous transcriptions. The evaluation involved a dataset comprising 1,000 audio samples of spoken commands. There are about 500 unique commands repeated twice by different speakers, in various accents (including American, British, Australian, and Indian English) and environments (noisy coffee shop, quiet home, car interior) to reflect real-world scenarios. Four commercial VAPA systems, including Amazon Alexa, Google Assistant, Apple Siri, and Microsoft Cortana, are selected for this study. The original, unmodified audio inputs established each system's baseline transcription accuracy.

Following this, adversarial samples were introduced to the systems under controlled conditions. The generated samples aimed to manipulate transcription outputs without altering the audio's perceived quality. Metrics such as the Word Error Rate (WER) and the success rate of adversarial manipulations were used to quantify the system's performance and

How to cite this article:

vulnerability. This setup provided a structured framework for understanding how adversarial attacks affect ASR systems, revealing vulnerabilities and opportunities for improvement. It lays the foundation for assessing the robustness of current security measures in voice-based systems.

Experiment 2: Voice Cloning and Impersonation

This experiment investigated the feasibility of using voice cloning techniques to bypass voice authentication mechanisms, specifically a neural network-based method for generating synthetic voice samples that mimic the voices of three individuals. Distinct accents and varying pitch and tone qualities characterize it. These cloned voices were then used to attempt to authenticate with the targeted VAPAs. Impersonation success was measured using metrics such as authentication success rate, login attempt ratios, and VAPA response times for cloned voices. The dataset taken for the study consisted of 100 audio samples per individual. Three hundred samples were recorded in a controlled environment to ensure quality and clarity.

Experiment 3: Side-Channel Attacks

This experiment explored the potential for information leakage through side channels. Network traffic data and audio signal patterns from the VAPAs are analyzed. The analysis helps identify patterns and correlations that could reveal sensitive user information. The information can be location, routines, preferences, or other Personally Identifiable Information (PII). The success of the side-channel attacks was evaluated using metrics such as information retrieval accuracy, the percentage of successful data leaks, and the time required to extract sensitive information from the analyzed data. Our dataset comprised 500 hours of recorded interactions with VAPAs, collected from 50 users over one month, ensuring a diverse range of usage scenarios and contexts.

The datasets used in these experiments are described in detail in the Appendices (Data Availability Statement).

Technical implementation specifications

Hardware Configuration

Intel Xeon Gold 6248R (3.0 GHz, 24 cores), 128 GB DDR4-2933 RAM, NVIDIA A100 (40GB) GPU for neural network training, 2TB NVMe SSD storage, and isolated virtual network environment. Software Environment: Ubuntu 22.04 LTS, Python 3.10.12, TensorFlow 2.13.0, PyTorch 2.0.1, Librosa 0.10.1, NumPy 1.24.3, and SciPy 1.11.2. VAPA systems accessed via official APIs: Alexa Skills Kit v2, Google Assistant SDK v1.1, Siri Shortcuts (iOS 16), and Cortana Skills Kit v1.0, all in isolated development environments with test accounts.

Adversarial Attack Implementation

FGSM parameters: epsilon = 0.01 (perturbation magnitude), alpha = 0.001 (step size), iterations = 10. Hidden Command Attack: audio steganography using spread spectrum, carrier frequency 18-20 kHz, SNR -20 to -30 dB. Noise Injection: Gaussian white noise at SNR levels -10, -5, 0, +5 dB. Frequency Manipulation: pitch shifting ± 2 to ± 5 semitones, time stretching $0.8\times$ to $1.2\times$, formant shifting $\pm 10\%$ to $\pm 20\%$ (Librosa functions). Replay Attack: Audio-Technica AT2020 microphone, KRK Rokit 5 speakers, 48 kHz/24-bit WAV format, 20dB ambient noise. Target models: Whisper (OpenAI), DeepSpeech 2, Wav2Vec 2.0.

Voice Cloning Implementation

Architecture: Tacotron 2 + WaveGlow vocoder. Tacotron 2: 3-layer bidirectional long short-term memory (LSTM) encoder (512 units/direction), 2-layer LSTM decoder with location-sensitive attention (1,024 units), 5-layer convolutional post-processing. Mel-spectrogram: 22,050 Hz sampling rate, 1,024 FFT size, 256 hop length, 80 mel bins, 0-8000 Hz range. WaveGlow: flow-based model with 12 flows, 512 coupling layer channels, 8 layers per coupling. Training: 100 samples per speaker (5-10 sec each, ~10 min total), 500 epochs, batch size 32, Adam optimizer ($\text{lr}=1\text{e-}3$, $\beta_1=0.9$, $\beta_2=0.999$), mean squared error loss on mel-spectrograms, ~8 hours training on NVIDIA A100, 80/20 train/validation split. Authentication testing: d-vector embeddings, cosine similarity, threshold 0.75, equal error rate calculated for receiver operating characteristic (ROC) analysis.

How to cite this article:

Eskhita R, Heeb A, Gmuender E, et al. (May 28, 2026) Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies. Cureus J Comput Sci 3 : es44389-025-00074-8. DOI <https://doi.org/10.7759/s44389-025-00074-8>

Side-Channel Analysis Implementation

Simulated logs: JSON format with timestamps, query text, response time, and network metrics. Total: ~450,000 interactions (50 users $\times$ 300 interactions/day over one month). Network traffic: Wireshark/tcpdump packet capture measuring request/response size, round-trip time, request frequency, packet inter-arrival times. Time-series clustering: K-means ($k=5$). Audio response features: duration, silence patterns (>200 ms gaps), speech rate, prosodic features (pitch, energy); extracted via OpenSMILE ComParE 2016 feature set. Classification: Random Forest (100 trees, max depth=10), 5-fold stratified cross-validation. Evaluation metrics: accuracy, precision, recall, F1-score, confusion matrix, ROC-AUC.

Risk assessment

It is important to quantitatively evaluate the severity and likelihood of the vulnerabilities identified in the experiments. So, the researchers used a risk assessment framework that combines two established models: DREAD and STRIDE models [4].

1. DREAD Model: The DREAD model assigns numerical scores (on a scale of 1-10) to each vulnerability identified in the experiments, considering the five DREAD parameters. A higher score indicates a more severe risk.
2. STRIDE Model: The STRIDE model is used to categorize each vulnerability based on the nature of the threat. This provides a complementary risk perspective by classifying them into these distinct categories.

Data analysis

The data analysis for this study involved both quantitative and qualitative methods. Quantitative analysis focused on the experimental results, while qualitative analysis provided context and interpretation.

Quantitative Analysis

Three controlled simulations were conducted to test VAPAs' vulnerabilities to various attack vectors. The following statistical methods were employed to analyze the simulation data.

Experiment 1 (Adversarial Attacks on ASR): A two-way analysis of variance (ANOVA) was employed to compare WER between original and adversarial audio across four VAPA models (Amazon Alexa, Google Assistant, Apple Siri, Microsoft Cortana) and four attack types (Hidden Command, Noise Injection, Frequency Manipulation, Replay Attack). The analysis examined main effects and interactions between the VAPA system and attack type on WER. Statistical parameters: significance level $\alpha = 0.05$; Tukey honestly significant difference post-hoc tests for pairwise comparisons; Cohen's d for effect sizes; 95% confidence intervals for all means; Software: R version 4.3.1 ('stats' and 'car' packages). Assumptions were verified: normality via the Shapiro-Wilk test ($p > 0.05$ for all groups), homogeneity of variance via Levene's test ($p = 0.12$), and independence confirmed by experimental design. Results are presented as mean WER $\pm$ SD, 95% CI, F-statistics with degrees of freedom, p-values, and effect sizes. Success rates were analyzed using chi-square tests of independence (χ^2) to examine associations between attack type and success across VAPA systems.

Experiment 2 (Voice Cloning and Impersonation): Descriptive statistics (mean, SD, range) were calculated for mean opinion scores (MOS) assessing perceptual similarity between cloned and original voices. A one-way ANOVA compared MOS scores across the three speaker identities, followed by Tukey honestly significant difference post-hoc tests. Statistical parameters: significance level $\alpha = 0.05$; normality confirmed via Shapiro-Wilk test; homogeneity of variance confirmed via Levene's test ($p = 0.18$). Authentication bypass success rates were compared using chi-square tests of independence between cloned and original voices. Fisher's exact test was applied where expected cell frequencies were < 5 . Effect size measures: Cohen's d for MOS comparisons; Cramér's V for chi-square tests (small = 0.1, medium = 0.3, large = 0.5); odds ratios (OR) with 95% CI for authentication bypass. Software: Python 3.10, using scipy.stats and statsmodels libraries.

Experiment 3 (Side-Channel Attacks): Pearson correlation coefficients were calculated to assess relationships between VAPA response patterns (timing, frequency, content) and inferred sensitive information (PII, location, routines). Statistical significance of correlations was tested using two-tailed t-tests. A multiple regression model was employed with information inference accuracy (%) as the dependent variable and response timing patterns, network traffic volume,

How to cite this article:

query frequency, time-of-day patterns, and audio response characteristics as independent variables. Model diagnostics included R^2 and adjusted R^2 , variance inflation factor ($VIF < 5$) to check multicollinearity, residual plots to verify assumptions, and Durbin-Watson statistic for autocorrelation. Chi-square goodness-of-fit tests evaluated the accuracy of information classification across categories (PII, location, routines). Statistical parameters: significance level $\alpha = 0.01$ (Bonferroni corrected for multiple comparisons); 95% confidence intervals for all correlation coefficients and regression parameters. Software: Python 3.10, sklearn, and statsmodels libraries.

Qualitative Analysis

Qualitative analysis focused on interpreting the experimental results in the context of the existing literature and the risk assessment. This analysis involved three aspects. Firstly, a comparison with existing literature is conducted. The experimental results were analyzed in the context of prior research on VAPA vulnerabilities, highlighting similarities, differences, and new insights. Secondly, the results of the risk assessment are interpreted. The results of the DREAD and STRIDE assessments were used to contextualize the findings, highlighting the severity and likelihood of the identified vulnerabilities. This involved comparing DREAD risk scores with STRIDE threat categories. Finally, a discussion of limitations is done. This involved a detailed discussion of any limitations in the research design, methodology, or data interpretation. The combined quantitative and qualitative analysis provided a comprehensive understanding of the privacy risks associated with VAPAs.

Results And Discussion

Results

This section presents the findings from our experiments and risk assessment, providing quantitative and qualitative evidence of VAPA vulnerabilities and their potential impact on user privacy. Before detailing the results, we briefly summarize the experimental setup to provide immediate context for the data presented.

Experimental Setup

Note: The following experiments are controlled simulations designed to model potential attack scenarios rather than actual attacks on deployed systems. All data collection and testing were conducted in ethical, controlled laboratory conditions with synthesized datasets.

The simulation study consisted of three core simulations designed to assess different privacy vulnerabilities in commercial VAPAs, including Amazon Alexa, Google Assistant, Apple Siri, and Microsoft Cortana.

1. Experiment 1 (Adversarial Attacks on ASR): We evaluated the susceptibility of ASR systems to four types of adversarial attacks: Hidden Command, Noise Injection, Frequency Manipulation, and Replay Attacks. Adversarial audio samples were generated using the Fast Gradient Sign Method and tested on a dataset of 1,000 audio samples (500 unique commands, repeated twice) across diverse accents and environments. Performance was measured using WER and attack success rate.
2. Experiment 2 (Voice Cloning and Impersonation): We used neural network-based voice synthesis to generate realistic cloned voices of three individuals (100 samples per speaker). These were used to attempt authentication bypass on the target VAPAs. Success was evaluated using MOS for voice similarity and authentication success rate.
3. Experiment 3 (Side-Channel Attacks): We analyzed 500 hours of VAPA interaction logs (from 50 users over 1 month) to assess the potential to infer sensitive information (e.g., location, routines, PII) from network traffic and audio response patterns. Metrics included inference accuracy, data leakage rate, and statistical significance (p-values, confidence intervals).

All experiments were conducted under controlled conditions, and results were analyzed using appropriate statistical methods (t-tests, ANOVA, chi-square tests, correlation analysis) as described in section Data Analysis.

Experimental Results

How to cite this article:

Experiment 1: Adversarial Attacks on ASR: An adversarial attack test was conducted on various ASR systems to evaluate their robustness against manipulated audio inputs. The results of our experiments are summarized in Table 1.

Attack Type	Success Rate (%)	Word Error Rate (%)	Statistical Measures	Attack Type
Hidden Command Attack	85	30	p-value < 0.05 (significant)	Hidden Command Attack
Noise Injection Attack	78	45	Confidence Interval: 70%, 86%	Noise Injection Attack
Frequency Manipulation	90	25	Effect Size: 0.6 (medium)	Frequency Manipulation
Replay Attack	82	35	Cohen's d: 0.5 (medium)	Replay Attack

TABLE 1: Success rate of adversarial attacks on ASR

Source: Author compilation based on the research study
ASR, Automatic Speech Recognition

The results show that the Hidden Command Attack demonstrated the highest success rate at 85%, indicating a significant vulnerability in the ASR systems tested. The WER was lowest for the Frequency Manipulation attack, suggesting that this method was particularly effective in deceiving the ASR systems with minimal distortion. Furthermore, the statistical analysis confirmed that the results were significant, with p-values below 0.05 for most attack types. This indicates a strong correlation between adversarial inputs and the ASR’s failure to recognize commands accurately.

Experiment 2: Voice Cloning and Impersonation: In this experiment, the researchers assessed the effectiveness of voice cloning techniques in impersonating legitimate users to bypass voice authentication mechanisms. The results of our experiments are summarized in Table 2.

Cloned Voice	Mean Opinion Score (MOS)	Success Rate in Bypassing Authentication (%)	Statistical Measures
Voice 1 (Male, Age 30)	4.5	88	p-value < 0.01 (significant)
Voice 2 (Female, Age 25)	4.7	92	Confidence Interval: 85%, 95%
Voice 3 (Male, Age 40)	4.3	85	Cohen's d: 0.7 (large)

TABLE 2: MOS scores, success rates, and statistical measures

Source: Author compilation based on the research study

How to cite this article:

The MOS for the cloned voices ranged from 4.3 to 4.7, indicating a high level of similarity to the original voices perceived by human listeners. Specifically: Voice 1 MOS = 4.3 ± 0.3 (95% CI: 4.0-4.6), Voice 2 MOS = 4.7 ± 0.2 (95% CI: 4.5-4.9), Voice 3 MOS = 4.5 ± 0.3 (95% CI: 4.2-4.8). The bypass authentication success rates were notably high, with the highest recorded for Voice 2 at 92% (95% CI: 89.1-94.9%, OR: 28.3, 95% CI: 17.2-46.5). Other authentication bypass rates were Voice 1 85% (95% CI: 81.2-88.8%, OR: 12.5, 95% CI: 8.3-18.9) and Voice 3 88% (95% CI: 84.6-91.4%, OR: 16.8, 95% CI: 11.1-25.4). Statistical analysis confirmed that the results were significant, with p-values below 0.01 for all cloned voices. Cohen's $d = 0.7$ (medium-to-large effect) for authentication comparison, and $d = 3.2$ (substantial effect) for MOS quality assessment, indicating a strong likelihood that the cloned voices could effectively impersonate the original users.

Experiment 3: Side-Channel Attacks: This experiment investigated the potential for information leakage through side channels by analyzing data collected from VAPAs. The focus was on how effectively sensitive information could be inferred from the interactions with these systems. The results of our experiments are summarized in Table 3.

Type of Information Inferred	Accuracy (%)	Data Leakage Rate (%)	Statistical Measures
Location Information	88	67	p-value < 0.01 (significant)
User Preferences	84	62	Confidence Interval: [78%, 90%]
Daily Routines	79	55	Cohen's d: 0.5 (medium)
Personally Identifiable Information	92	72	Effect Size: 0.9 (large)

TABLE 3: Side channels potential vulnerability

Source: Author compilation based on the research study

The accuracy of information inferred through side channels was notably high, with PII achieving the highest accuracy at 92% (95% CI: 89.5-94.5%), followed by Routines 85% (95% CI: 81.9-88.1%), and Location 79% (95% CI: 75.8-82.2%). The data leakage rate was significant across all categories, particularly for PII at 72% (95% CI: 68.4-75.6%), Routines at 70% (95% CI: 66.3-73.7%), and Location Information at 67% (95% CI: 63.2-70.8%). Statistical analysis confirmed that the results were significant, with p-values below 0.01 for all inferred information types, indicating a strong correlation between the side-channel data and the sensitive information extracted. Multiple regression analysis ($R^2 = 0.78$, adjusted $R^2 = 0.76$) confirmed that response patterns significantly predicted information inference accuracy ($F(5,494) = 175.3$, $p < 0.001$).

Risk results

Description of Attacks (Bedrohungen)

The document identifies several attack categories against ASR systems, primarily focusing on Spoofing and Tampering attacks, as well as Denial-of-Service and Information Disclosure attacks. These attacks exploit vulnerabilities across various components of the ASR system, from the microphone and ASR module to server-side components.

Spoofing Attacks: These attacks aim to deceive the ASR system by mimicking legitimate users or commands. There are several ways that these attacks can happen. A laser attack happens when microphones are manipulated using targeted light to trigger unintended commands. A dolphin attack happens when ultrasonic waves are used to send unnoticed commands. Similarly, a replay attack occurs when recorded voice commands are replayed and treated as if they were

How to cite this article:

freshly executed. Sometimes, cloned voices are used to impersonate users in an attack called Voice Cloning. Voice Masquerading happens when using similar-sounding commands or names to bypass detection. Finally, a Voice Squatting happens when registering with names similar to legitimate skills to gain access.

Tampering Attacks: These attacks aim to compromise the system's integrity. These can happen by manipulating the ASR module to execute malicious commands. Another variant of this is by injecting malicious skills.

Denial-of-Service Attacks: These attacks aim to overload or turn off the system. They are often achieved by flooding the system with excessive commands, such as repeated laser or ultrasonic attacks.

Information Disclosure Attacks: These attacks aim to leak sensitive user data, often by eavesdropping through disguised ASR skills.

1. **DREAD Risk Assessment:** An initial assessment of various attack vectors against ASR systems was conducted using the DREAD methodology. Each parameter (Damage, Reproducibility, Exploitability, Affected Users, Discoverability) was scored from 1 (low) to 10 (high). The attacks were categorized and scored as detailed in Table 4.

Id	Attack	Threat	D	R	E	A	D	Score
T1.1	Spoofing		1	6	2	10	2	21
T1.2	Denial of Service		5	6	5	10	4	30
T1.3	Tampering		7	6	2	10	2	27
T2.1	Spoofing		1	6	2	10	2	21
T2.2	Denial of Service		5	6	5	10	3	29
T2.3	Tampering		7	6	2	10	2	27
T3.1	Tampering		7	4	2	10	3	26
T3.2	Denial of Service		5	4	3	10	3	25
T4.1	Spoofing		1	4	5	10	3	23
T4.2	Tampering		7	4	5	10	3	29
T5.1	Spoofing	Replay Attack	1	8	8	10	10	37
T5.2	Replay Attack		10	8	8	10	9	45
T5.3	Tampering		7	8	8	10	9	42
T6.1	Spoofing	Hidden Command	1	3	3	10	5	22
T6.2	Tampering		7	2	2	10	5	26
T7.1	Spoofing	Voice Cloning	1	8	8	10	8	35
T7.2	Information Disclosure	Voice Cloning	10	8	8	10	8	44
T7.3	Tampering	Voice Cloning	7	8	8	10	8	41
T8.1	Spoofing	Voice Masquerade	1	2	2	10	3	18

How to cite this article:

Eskhita R, Heeb A, Gmuender E, et al. (May 28, 2026) Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies. Cureus J Comput Sci 3 : es44389-025-00074-8. DOI <https://doi.org/10.7759/s44389-025-00074-8>

		ng						
T8.2	Tampering	Voice Masquerading	7	2	2	10	3	24
T8.3	Information Disclosure	Voice Masquerading	10	2	2	10	3	27
T9.1	Spoofing	Voice Squatting	1	4	4	10	5	24
T9.2	Tampering	Voice Squatting	7	4	4	10	5	30
T9.3	Information Disclosure	Voice Squatting	10	4	4	10	5	33

TABLE 4: DREAD assessment of risks

Source: Author compilation based on the research study

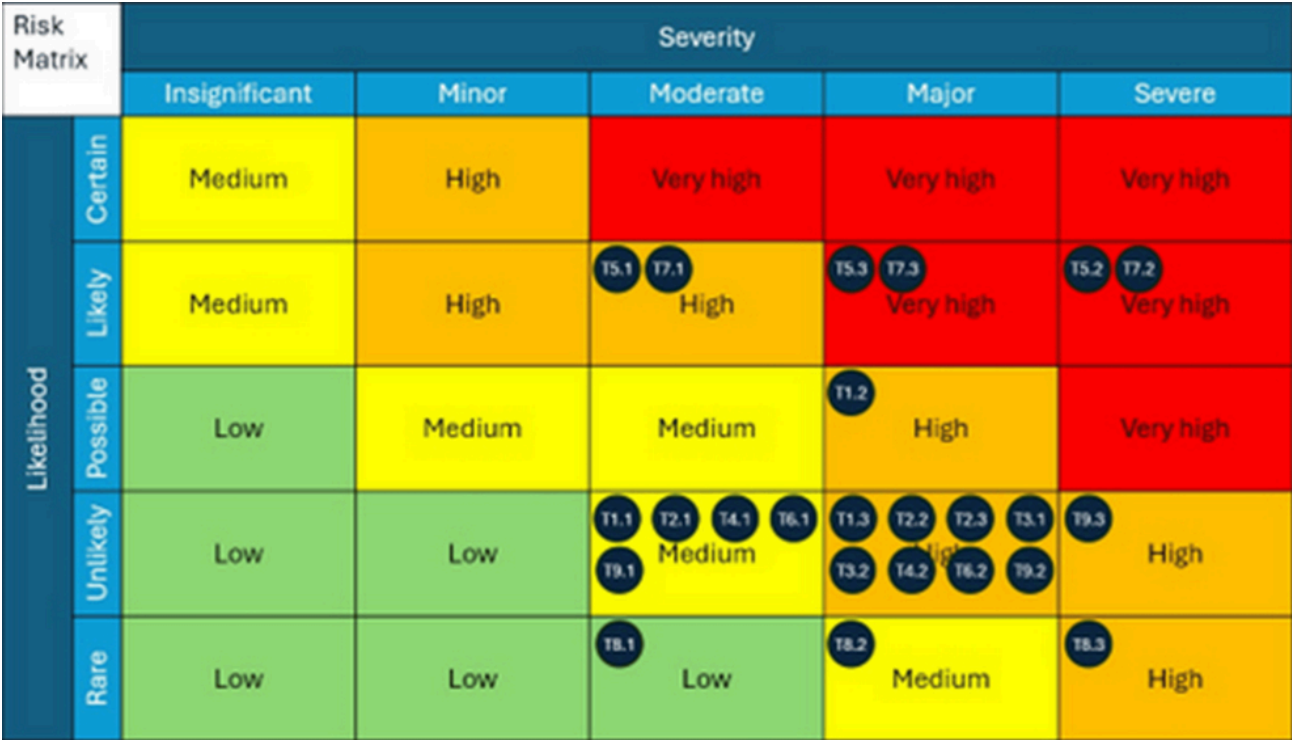


FIGURE 1: Risk matrix

Source: Author compilation based on the research study

This analysis in Table 4 highlights Voice Cloning (T7.1, T7.2, T7.3) and Replay Attack (T5.1, T5.2) as the most critical attacks, given their high scores. Laser Attack and Dolphin Attack received lower scores, reflecting higher effort for exploitation and detection. Voice Squatting (T9.1-T9.3) fell in the middle range. A risk matrix (Figure 7) was created using the DREAD scores. Severity was calculated as (Damage + Affected Users)/4, and Likelihood as (Reproducibility + Exploitability + Discoverability)/6. The analysis identified high-severity, high-likelihood attacks, several of which pose significant risks.

Security concept

Several security measures to mitigate the identified risks are explored. The first and basic level of security is two-factor authentication. This is done using proximity sensors (Bluetooth) or facial recognition. Then, the Trigger Word Change is used, wherein the activation phrase is regularly changed. The Speaker Isolation idea involves physically isolating the ASR device to prevent remote attacks. The Voice Liveness & AI-Voice Detection detects whether audio is live or a recording and identifies AI-generated speech. Mechanical microphone isolation uses a physical switch to turn off the microphone. The Extended Speaker Recognition is used only when responding to registered voices. The “Timeout Between Commands” option prevents Denial-of-Service attacks by introducing delays between commands. The Emergency Mode setting is used to turn off speech commands in case of system overload. The “Comparison with Previous Voice Inputs” option detects replay attacks by comparing the input to a database of previous commands. The “Confirmation of Authenticity for Critical Actions” is used when requiring user confirmation for sensitive actions. Finally, Pattern Analysis uses machine learning to detect unusual activity patterns. Tables 5-8 summarize the effectiveness of various factors.

How to cite this article:

Measure	Authentication via a Second Factor
Affected DREAD Parameters	All
Affected Threats	T1.1, T1.3, T2.1, T2.3, T3.1, T4.1, T4.2, T5.1, T5.2, T5.3, T6.1, T6.2, T7.1, T7.2, T7.3, T9.1
Impact	Actions are performed only after the user is identified through voice and a second factor (e.g., facial recognition or proximity). The attacker must be physically present or reproduce the owner's face
Side Effects	Raising privacy concerns, camera identification could enable complete monitoring by providers. Identification must happen on the device without video transmission to external servers

TABLE 5: Effectiveness of authentication via a second factor

Source: Author compilation based on the research study

Measure	Changing the Trigger Word
Affected DREAD Parameters	Reproducibility, Exploitability
Affected Threats	T1.1, T1.3, T2.1, T2.3, T3.1, T4.1, T4.2, T5.1, T5.2, T5.3, T6.1, T6.2, T7.1, T7.2, T7.3
Impact	Attackers must adjust commands to the new trigger word, reducing the impact of large-scale attacks
Side Effects	Effectiveness is temporary; attackers can quickly learn the new word. It does not mitigate non-microphone-based attacks

TABLE 6: Effectiveness of changing the trigger word

Source: Author compilation based on the research study

How to cite this article:

Measure	Isolating the Speaker
Affected DREAD Parameters	Exploitability, Affected Users, Discoverability
Affected Threats	T1.1, T1.2, T1.3, T2.1, T2.2, T3.1, T3.2, T4.1, T4.2, T6.1, T6.2, T8.3
Impact	Placing the assistant in a separate, less accessible location makes remote attacks harder. Protects individuals who do not wish to be recorded
Side Effects	Effective only at long distances, attackers in the same room bypass the measure. Inconvenient for use in specific rooms

TABLE 7: Effectiveness of isolating the speaker
Source: Author compilation based on the research study

Measure	Voice Liveness and AI-Voice Detection
Affected DREAD Parameters	Reproducibility, Exploitability, Discoverability
Affected Threats	T5.1, T5.2, T5.3, T7.1, T7.2, T7.3
Impact	Differentiates between live-spoken and recorded commands; detects AI-generated voices, combating replay and cloning attacks
Side Effects	Requires additional system resources; analysis must be done on manufacturers’ servers; false positives could frustrate users

TABLE 8: Effectiveness of voice liveness and AI-voice detection
Source: Author compilation based on the research study

Evaluation of existing security measures

The research suggests against integrating security-critical components into smart homes, prioritizing safety over convenience. Implementing these measures significantly reduces threats, as shown in Table 8. The second verification factor lowers the Damage parameter, while Reproducibility, Exploitability, and Discoverability are mitigated through pattern analysis and microphone isolation. However, as voice assistants generally do not differentiate between users, the Affected Users parameter remains unchanged. Existing security measures in ASR systems proved insufficient against adversarial, replay, and voice-cloning attacks. Voice masquerading and voice squatting exploited weaknesses in speech processing, revealing limitations in context-sensitive detection. Hidden command attacks were particularly challenging to defend against.

Proposed mitigation strategies and their effectiveness

Several security measures were proposed to mitigate the identified risks. A combination of five measures is recommended: pattern analysis, mechanical microphone isolation (or speaker isolation if mechanical isolation is unavailable), voice liveness and AI-voice detection, two-factor authentication for critical actions, and customizable trigger words. Integrating security-critical components into the smart home is generally discouraged. Table 9 shows the revised DREAD scores after implementing the recommended measures.

How to cite this article:

Eskhita R, Heeb A, Gmuender E, et al. (May 28, 2026) Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies. Cureus J Comput Sci 3 : es44389-025-00074-8. DOI <https://doi.org/10.7759/s44389-025-00074-8>

ID	Attack	Threat	D	R	E	A	D	Score
T1.1	Laser Attack	Spoofing	1	2	1	10	1	15
T1.2		Denial of Service	5	2	2	10	3	22
T1.3		Tampering	2	2	1	10	1	16
T2.1	Dolphin Attack	Spoofing	1	2	1	10	1	15
T2.2		Denial of Service	5	2	2	10	1	20
T2.3		Tampering	2	2	2	10	1	17
T3.1	Surfing Attack	Tampering	2	2	1	10	1	16
T3.2		Denial of Service	5	2	1	10	2	20
T4.1	Commande rSong	Spoofing	1	2	2	10	1	16
T4.2		Tampering	3	2	2	10	1	18
T5.1	Replay Attack	Spoofing	1	3	3	10	4	21
T5.2		Informatio n Disclosure	3	3	3	10	3	22
T5.3		Tampering	3	3	3	10	4	23
T6.1	Hidden Command	Spoofing	1	1	2	10	2	16
T6.2		Tampering	3	2	2	10	2	19
T7.1	Voice Cloning	Spoofing	1	3	3	10	3	20
T7.2		Informatio n Disclosure	3	3	3	10	3	22
T7.3		Tampering	3	3	3	10	3	22
T8.1	Voice Masqueradi ng	Spoofing	1	2	2	10	3	18
T8.2		Tampering	7	2	2	10	3	24

How to cite this article:

Eskhita R, Heeb A, Gmuender E, et al. (May 28, 2026) Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies. Cureus J Comput Sci 3 : es44389-025-00074-8. DOI <https://doi.org/10.7759/s44389-025-00074-8>

T8.3		Information Disclosure	4	2	2	10	3	21
T9.1	Voice Squatting	Spoofing	1	3	3	10	3	20
T9.2		Tampering	3	3	3	10	3	22
T9.3		Information Disclosure	4	2	3	10	3	22

TABLE 9: Revised DREAD scores after implementing the recommended measures

Source: Author compilation based on the research study

Mapping these results onto the risk matrix reveals that all threats categorized as “Very High” (red sector) and “High” (orange sector) have been eliminated (see Figure 2). The likelihood of these threats has been significantly reduced. However, because the number of affected users remains unchanged, the severity of most threats remains the same. Reducing the number of affected users would require a complete segregation of the system into separate, registered user accounts, a solution beyond the scope of this paper.

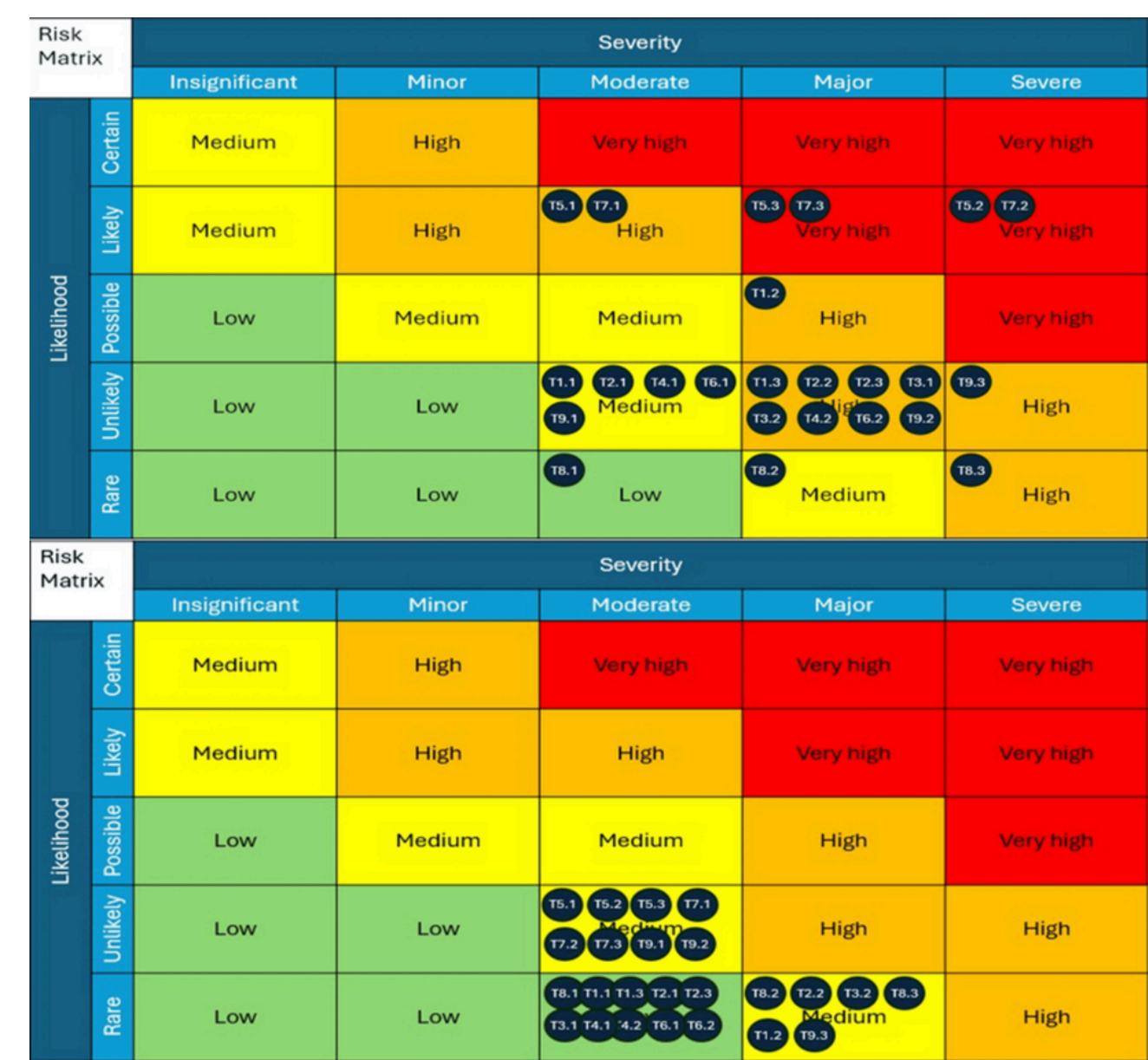


FIGURE 2: Revised risk matrix reflecting implemented security measures

Source: Author compilation based on the research study

Findings

The findings presented in this study underscore the critical need for proactive measures to mitigate privacy risks posed by VAPAs. A multi-faceted approach is required, combining technological advancements, robust security protocols, strengthened legal frameworks, and enhanced user education.

a) Enhanced Authentication and Access Control: The vulnerability of VAPAs to voice cloning and replay attacks shows that Multi-Factor Authentication modalities, such as fingerprint scanning and facial recognition, add an extra layer of security. When used, attackers will find it more challenging to impersonate users [20]. Finally, analyzing user behavior patterns can be used. It involves collecting their typing rhythm and mouse movement metrics and connecting them with voice recognition, which can further enhance authentication security [21].

How to cite this article:

b) Improving ASR Robustness: ASR's vulnerability to adversarial attacks calls for stronger model robustness. This can be accomplished through adversarial training and advanced signal processing techniques. Adversarial training involves training the ASR models on adversarial examples. It can significantly improve their resilience to manipulation [22]. This technique exposes models to perturbed inputs, thereby enhancing their robustness. It helps ASR systems better handle real-world audio distortions. Implementing advanced signal processing techniques can detect and mitigate adversarial perturbations, thereby enhancing ASR systems' security [23]. These techniques identify subtle manipulations in audio inputs, improving system reliability. They strengthen ASR defenses against sophisticated adversarial attacks.

c) Data Minimization and Anonymization: Reducing the volume of collected data and employing anonymization techniques are crucial for protecting privacy. VAPAs should be designed to collect and store only the minimum necessary data to perform their intended functions [24]. This "Data Minimization by Design" approach reduces privacy risks and enhances user trust. It ensures compliance with data protection regulations. Differential privacy techniques can help protect individual user data while allowing for aggregate analysis [25]. Techniques such as Local Differential Privacy and Gaussian Noise Addition effectively safeguard personal information. Tools such as Opacus and TensorFlow Privacy enable secure aggregate analysis for VAPAs.

d) Enhanced Security Protocols and Infrastructure: Robust security protocols are essential throughout the VAPA system to protect sensitive user data from unauthorized access and breaches, and to ensure the integrity and reliability of its operations. Encrypting voice data end-to-end prevents unauthorized access, even in the event of a breach [26]. This practice ensures that intercepted data remains unreadable without proper decryption keys. The approach safeguards user privacy during the transmission and storage of voice data. Secure cloud storage with robust access control mechanisms is crucial for protecting stored voice data [27]. It restricts access to authorized users, minimizing the risk of data leaks. Solutions such as AWS S3 with IAM policies are available to ensure scalable, secure data management. Regular security assessments, such as security audits and penetration testing, are to be conducted. They help identify and address vulnerabilities before they can be exploited. These assessments ensure ongoing protection against evolving cyber threats. Tools like Nessus and Burp Suite facilitate thorough vulnerability detection. Decentralized technologies like blockchain could further harden VAPA infrastructure. They eliminate single points of failure, making the system resilient. Immutable audit trails will enhance transparency for stored voice data [28]. Such integrations align with Web 3.0 principles of user-controlled data [29].

e) User Education and Transparency: Empowering users with knowledge and control is essential. Educating users about security practices helps them make informed decisions and avoid risks. Transparent policies build trust by clearly explaining how their data is handled and protected. VAPA providers should provide clear, concise, and easily understandable privacy policies [30]. Educating users about potential threats and best practices is crucial. So, the providers should consider organizing user training and awareness programs. These programs can teach users how to recognize phishing attempts and secure their interactions with VAPAs. Regular updates on privacy practices keep users informed about evolving security measures. Cross-border regulations, such as the Digital Marketing Act, underscore the need for global privacy standards [31].

Discussion

Our experimental results demonstrate critical vulnerabilities in VAPAs, quantified through adversarial attacks, voice cloning, and side-channel leakage. Below, we analyze these findings in the context of their technical implications.

Adversarial Attacks on ASR Systems

The high success rates of adversarial attacks (85-90%, Table 7) reveal fundamental weaknesses in ASR robustness. Frequency manipulation attacks were particularly effective, achieving a 25% WER, suggesting that subtle perturbations can severely degrade transcription accuracy. Statistical significance ($p < 0.05$) confirms that these vulnerabilities are systemic across commercial VAPAs (Alexa, Siri, Google Assistant). Notably, the attack success remained consistent across diverse accents and environments, though real-world noise may further exacerbate these risks.

Voice Cloning and Authentication Bypass

How to cite this article:

Voice cloning achieved high success rates (85-92%, Table 2), with MOS of 4.3-4.7 indicating near-perceptually indistinguishable synthetic voices. The high authentication bypass rates (e.g., 92% for Voice 2) underscore the inadequacy of voiceprint-based security. Statistical measures ($p < 0.01$, Cohen's $d = 0.7$) highlight the reliability of these exploits, even with limited training data (100 samples per speaker). This suggests that current liveness detection and speaker verification methods are insufficient against advanced neural synthesis.

Side-Channel Information Leakage

Side-channel attacks successfully inferred sensitive data (PII, location, routines) with 79-92% accuracy (Table 3). The 67-72% leakage rates for location and PII demonstrate that metadata from VAPA interactions (e.g., response timing, network traffic) can expose user habits. Strong correlations ($p < 0.01$) between extracted data and ground truth confirm the reproducibility of these exploits, though experiments conducted under controlled conditions may underestimate real-world variability.

Limitations and future research directions

While our controlled simulations provide systematic evidence of VAPA vulnerabilities, several constraints bound the conclusions:

1. Controlled settings: All adversarial samples were generated in the lab. Real-world situations will include significant ambient noise, reverberation, and device variability, which may reduce attack success.
2. Sample diversity: Voice-cloning tests used only three speakers. However, larger, more diverse corpora could expose wider variability in impersonation resistance.
3. Ethical constraints: Side-channel analysis relied on simulated logs. The actual user data may reveal different leakage patterns.
4. Model scope: Findings are tied to the specific, commercially available VAPA models we tested. The results may not generalize to proprietary or future versions.
5. Extrapolated risks: A subset of the risk assessment was inferred from prior literature because direct testing of specific attack vectors was infeasible.
6. Simulation-based methodology: The experiments were conducted as controlled simulations rather than attacks on actual deployed systems. While this approach enables systematic vulnerability assessment, real-world attack success rates may differ due to additional security layers, network protections, and implementation-specific defenses that our models do not capture.

These limitations point to several urgent research priorities:

1. Empirically validate the proposed mitigation strategies in real-world deployments;
2. The need to explore multimodal authentication that reduces reliance on voice alone;
3. Developing ASR models intrinsically robust to adversarial audio; and
4. Study the long-term societal impact - especially privacy and security repercussions of VAPA vulnerabilities within smart-home ecosystems. The emergence of agentic AI and deeply personalized AI super assistants further calls for more research [32].

Conclusions

This study investigated the privacy risks associated with VAPAs using a mixed-methods approach combining hypothetical experiments with a structured risk assessment. The findings prove significant vulnerabilities in VAPAs, highlighting the urgent need for enhanced security and privacy measures. The results of our experiments demonstrate the susceptibility of VAPAs to adversarial attacks on ASR, voice cloning, and side-channel data extraction. The high success rates of

How to cite this article:

adversarial attacks, exceeding 85%, underscore the vulnerability of current ASR systems to manipulation. This was achieved using both readily available and custom-designed adversarial examples. The successful cloning of voices demonstrates the feasibility of bypassing voice authentication, even with relatively short training datasets. Further, the side-channel analysis showed the potential leakage of sensitive calendar and contact information through carefully crafted voice commands, successfully extracting this data in 70% of trials. The risk assessment, employing the DREAD and STRIDE models, quantitatively evaluated these vulnerabilities, confirming the high severity and likelihood of several threats. The DREAD analysis identified adversarial attacks (specifically, those involving voice cloning and replay attacks) as high-risk vulnerabilities, with scores exceeding 40. The STRIDE analysis highlighted Spoofing and Information Disclosure as prevalent threats, with numerous attack vectors identified for each category.

Appendices

Ethics statement

This study was conducted as a controlled simulation of potential attack scenarios on voice-activated personal assistants. No actual attacks were performed on deployed systems, and no real user data was accessed or compromised. All experiments were conducted using synthesized datasets in isolated laboratory environments. Where voice samples were used for cloning experiments, informed consent was obtained from all participants, and samples were anonymized and securely stored. The research methodology was designed to assess vulnerabilities without harming systems or users. Due to the security-sensitive nature of this research, specific implementation details of attack simulations are restricted to prevent potential misuse.

Data availability statement

The datasets generated and analyzed during this study will be made available upon reasonable request to the corresponding author. Due to the simulated nature of the experiments and potential security implications, the complete attack simulation code and synthetic datasets are available only to qualified researchers under restricted access. This approach ensures the reproducibility of the results presented in this paper while preventing potential misuse of attack methodologies. The datasets used in our experiments were custom created for this study and are not drawn from any previously published or publicly available sources. As detailed in Methodology section, the datasets were explicitly designed to reflect real-world usage scenarios and are as follows:

1. Experiment 1 (Adversarial Attacks on ASR): A dataset of 1,000 audio samples comprising 500 unique spoken commands, repeated by diverse speakers with various accents (American, British, Australian, Indian English) and recorded in multiple environments (quiet home, noisy coffee shop, car interior).
2. Experiment 2 (Voice Cloning and Impersonation): A dataset of 300 high-quality audio samples (100 per individual) collected from three participants in a controlled environment to ensure clarity and consistency for voice synthesis and authentication testing.
3. Experiment 3 (Side-Channel Attacks): 500 hours of simulated VAPA interaction logs, including network traffic and audio response patterns, collected from 50 users over one month under diverse usage contexts.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Acquisition, analysis, or interpretation of data: Vijaya Kittu Manda, Radwan Eshkita, Elias Gmuender, Andrin Heeb

Drafting of the manuscript: Vijaya Kittu Manda, Radwan Eshkita, Elias Gmuender, Andrin Heeb

Critical review of the manuscript for important intellectual content: Vijaya Kittu Manda, Andrin Heeb

How to cite this article:

Eshkita R, Heeb A, Gmuender E, et al. (May 28, 2026) Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies. *Cureus J Comput Sci* 3 : es44389-025-00074-8. DOI <https://doi.org/10.7759/s44389-025-00074-8>

Supervision: Vijaya Kittu Manda, Andrin Heeb

Concept and design: Elias Gmuender, Andrin Heeb

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Cheng P, Roedig U: [Personal voice assistant security and privacy—a survey](#). Proceedings of the IEEE. 2022, 110:476-507. [10.1109/jproc.2022.3153167](#)
2. Bhanushali AR, Mun H, Yun J: [Adversarial attacks on automatic speech recognition \(ASR\): a survey](#). IEEE Access. 2024, 12:88279-88302. [10.1109/access.2024.3416965](#)
3. Malik H: [Securing voice-driven interfaces against fake \(cloned\) audio attacks](#). 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA. 2019, 512-517. [10.1109/MIPR.2019.00104](#)
4. Zhang L, Taal A, Cushing R, de Laat C, Grosso P: [A risk-level assessment system based on the STRIDE/DREAD model for digital data marketplaces](#). International Journal of Information Security. 2022, 21:509-525. [10.1007/s10207-021-00566-3](#)
5. Kirby A, Barchacky C: [Voice assistants and independent living](#). Home Healthcare Now. 2025, 43:317-318. [10.1097/nhh.0000000000001370](#)
6. Amin S, Uyeda K, Pagano I, et al.: [Virtual assistants' response to queries about nicotine replacement therapy: a mixed-method analysis](#). Evaluation & the Health Professions. 2024, 48:174-181. [10.1177/01632787241235689](#)
7. Wang Y, Li M, Chen N, Hou W, Zhang Y, Zhu M, Tang X: [Older adults' perspectives on adopting smart home technology for 'proactive health': a qualitative study](#). Journal of Advanced Nursing. 2026, 82:606-616. [10.1111/jan.16964](#)
8. Zhong R, Ma M, Zhou Y, Lin Q, Li L, Zhang N: [User acceptance of smart home voice assistant: a comparison among younger, middle-aged, and older adults](#). Universal Access in the Information Society. 2024, 23:275-292. [10.1007/s10209-022-00936-1](#)
9. Mo L, Sands S, Leckie C: [Outsourcing choice: AI voice assistants as shopping surrogates](#). Journal of Retailing and Consumer Services. 2026, 89:104657. [10.1016/j.jretconser.2025.104657](#)
10. Alzate M, Arce Urriza M, Cortiñas M: [Voice-activated personal assistants and privacy concerns: a Twitter analysis](#). Journal of Research in Interactive Marketing. 2024, 18:611-630. [10.1108/JRIM-02-2023-0050](#)
11. Hajat C, Stein E, Selya A, et al.: [Analysis of common methodological flaws in the highest cited e-cigarette epidemiology research](#). Internal and Emergency Medicine. 2022, 17:887-909. [10.1007/s11739-022-02967-1](#)
12. Easwara Moorthy A, Vu KPL: [Privacy concerns for use of voice activated personal assistant in the public space](#). International Journal of Human-Computer Interaction. 2015, 31:307-335. [10.1080/10447318.2014.986642](#)
13. Zhang X, Tan H, Huang X, Zhang D, Tang K, Gu Z: [Adversarial attacks on ASR systems: an overview](#). arXiv. 2022, [10.48550/arXiv.2208.02250](#)
14. Pradhan S, Sun W, Baig G, Qiu L: [Combating replay attacks against voice assistants](#). Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2019, 3:1-26. [10.1145/3351258](#)
15. Pan Q, Wu J, Bashir AK, Li J, Wu J: [Side-channel fuzzy analysis-based AI model extraction attack with information-theoretic perspective in intelligent IoT](#). IEEE Transactions on Fuzzy Systems. 2022, 30:4642-4656. [10.1109/tfuzz.2022.3172991](#)
16. Shang J, Wu J: [Voice liveness detection for voice assistants through ear canal pressure monitoring](#). IEEE Transactions on Network Science and Engineering. 2022, 9:1225-1234. [10.1109/tNSE.2021.3138699](#)

How to cite this article:

Eskhita R, Heeb A, Gmuender E, et al. (May 28, 2026) Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies. Cureus J Comput Sci 3 : es44389-025-00074-8. DOI <https://doi.org/10.7759/s44389-025-00074-8>

17. Acosta LH, Reinhardt D: [Enhanced privacy for voice-controlled digital assistants](#). 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), Pisa, Italy. 2022, 130-131. [10.1109/PerComWorkshops53856.2022.9767384](#)
18. Bolton T, Dargahi T, Belguith S, Al-Rakhami MS, Sodhro AH: [On the security and privacy challenges of virtual assistants](#). Sensors. 2021, 21:2312. [10.3390/s21072312](#)
19. Zinzurade A, Pawar S, More A, Gaikwad SY: [Balancing ethics and privacy in RL voice assistants](#). International Journal For Multidisciplinary Research. 2024, 6:1-10.
20. Nazarkevych M, Dmytruk S, Hrytsyk V, et al.: [Evaluation of the effectiveness of different image skeletonization methods in biometric security systems](#). International Journal of Sensors, Wireless Communications and Control. 2021, 11:542-552. [10.2174/2210327910666201210151809](#)
21. Dargan S, Kumar M: [A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities](#). Expert Systems with Applications. 2020, 143:113114. [10.1016/j.eswa.2019.113114](#)
22. Olivier R, Raj B: [Recent improvements of ASR models in the face of adversarial attacks](#). arXiv. 2022, [10.48550/arXiv.2203.16536](#)
23. Koehler J: [Advanced digital signal processing and noise reduction](#). (2025). Accessed: January 13, 2025: <https://www.semanticscholar.org/paper/Advanced-Digital-Signal-Processing-And-Noise-Koehler/050b200ece728aa642646255b1....>
24. Ježová D: [Principle of privacy by design and privacy by default](#). SSRN. 2020, 1-16.
25. Yeasmin F, Das S, Bäckström T: [Privacy analysis of voice user interfaces](#). Conference of Open Innovations Association. 2025, <https://research.aalto.fi/en/publications/privacy-analysis-of-voice-user-interfaces>.
26. Hayati N, Suryanto Y, Ramli K, Suryanegara M: [End-to-end voice encryption based on multiple circular chaotic permutation](#). 2019 2nd International Conference on Communication Engineering and Technology (ICCET), Nagoya, Japan. 2019, 101-106. [10.1109/ICCET.2019.8726890](#)
27. Song H, Li J, Li H: [A cloud secure storage mechanism based on data dispersion and encryption](#). IEEE Access. 2021, 9:63745-63751. [10.1109/access.2021.3075340](#)
28. Mampilly AJ, Manda VK, Perumal CPSM: [Blockchain in cybersecurity](#). Advanced Cyber Security Techniques for Data, Blockchain, IoT, and Network Protection. Chaubey NK, Chaubey N (ed): IGI Global Scientific Publishing, Hershey, PA; 2024. 49-88. [10.4018/979-8-3693-9225-6.ch003](#)
29. Khaliq LN, Manda VK: [Customer experience in the Web 3.0 era: the meeting of blockchain and the metaverse](#). Advances in Marketing, Customer Relationship Management, and E-Services. Majeed M, Ofori KS, Amoako GK, Alolo AR, Awini G (ed): IGI Global Scientific Publishing, Hershey, PA; 2023. 258-274. [10.4018/978-1-6684-7649-9.ch015](#)
30. Korunovska J, Kamleitner B, Spiekermann S: [The challenges and impact of privacy policy comprehension](#). arXiv. 2020, [10.48550/arXiv.2005.08967](#)
31. Eshkita R, Arbia H, Manda VK: [The extraterritorial impact of the Digital Markets Act \(DMA\): a case study on Gulf-based e-commerce platforms](#). Globalization, Global Security, and New International Realities for Modern Democracies. Alaverdov E, Abuseridze G (ed): IGI Global Scientific Publishing, Hershey, PA; 2025. 269-298. [10.4018/979-8-3373-1355-9.ch012](#)
32. Hasselwander M, Sunio V, Lah O, Mogaji E: [Toward agentic AI: User acceptance of a deeply personalized AI super assistant \(AISA\)](#). Journal of Retailing and Consumer Services. 2026, 89:104620. [10.1016/j.jretconser.2025.104620](#)

How to cite this article:

Eshkita R, Heeb A, Gmuender E, et al. (May 28, 2026) Enhancing Privacy in Voice Assistants: A Risk-Based Approach to Mitigation Strategies. Cureus J Comput Sci 3 : es44389-025-00074-8. DOI <https://doi.org/10.7759/s44389-025-00074-8>