NeptuneX: A Privacy-Focused Voice Assistant Architecture

Madhumitha Santhanakrishnan
University of California, Davis
msan@ucdavis.edu

Vikraman Senthil Kumar
University of California, Davis
viksenthil@ucdavis.edu

ABSTRACT

This paper introduces a novel privacy-centric voice assistant developed using OpenAI's GPT-3.5 Turbo, designed to enhance user privacy in digital interactions. Our system enables users to control their ad experience through personalized settings, ensuring relevance and privacy in content delivery. The assistant employs real-time voice recognition and semantic analysis to categorize user queries into 15 predefined sensitive ad categories, thereby offering a tailored and secure user experience. Initial testing with a small user group demonstrated high satisfaction in privacy and usability, underscoring the potential of privacy-focused voice assistants in modern digital ecosystems.

1 INTRODUCTION

In today's digital age, the explosion in the popularity of smart speakers in the United States alone has been nothing short of remarkable. From a modest installed base of 5 million units in June 2017, the landscape has dramatically transformed, with an astonishing 46.5 million units recorded by January 2020 [1]. This spectacular rise is a testament to the allure of seamless, voice-activated technology in enhancing everyday convenience. However, this technological renaissance also brings to the fore a critical conundrum—the imperative to balance such convenience with the sacrosanct principle of user privacy. As these devices permeate an ever-increasing number of living rooms, kitchens, and bedrooms, they not only listen and learn but also amass vast amounts of data, weaving intricate digital trails that sketch the intimate contours of our lives. Amid the growing number of these devices, there's increasing unease about how convenient they seem, as they leave behind data trails that could be misused.

The poignant experience shared by Gillian Brockell [2] serves as a lens through which to examine the profound and often overlooked gaps in the edifice of digital privacy and targeted advertising. Brockell's distressing encounter with the unyielding deluge of maternity and baby-related advertisements, following the devastating loss of her baby, lays bare a critical shortcoming in the algorithms that underpin digital advertising. These algorithms, with their relentless pursuit of relevance and engagement, reveal an acute inability to navigate the complexities of human emotion and life experiences. Brockell's ordeal, while deeply personal, underscores a universal vulnerability—our collective exposure to technologies that, in their current form, fail to discern or respect the most sensitive and private moments of our lives. Her story is a clear call for technology companies to revisit and refine their data processing paradigms, ensuring that their algorithms are equipped not only to detect but also to respond with sensitivity to the nuances of human experience, thereby safeguarding individuals from further emotional harm.

At this pivotal moment in technology, the rapid growth of smart devices presents an opportunity to pursue the development of technologies that prioritize privacy. Our project is aiming to establish a harmony between the convenience offered by technology and the preservation of personal privacy. The creation of a privacy-centric voice assistant is a testament to our dedication to protecting individual privacy and dignity within the digital realm. This work represents a call to action for integrating respect for digital privacy into the core of technological design and usage, advocating for privacy to be recognized not as a privilege, but as a fundamental right.

2 RESEARCH QUESTION

At the core of our study lies a question that bridges technology, economy and ethics: How can a voice assistant provide personalized ad experiences in an economically sustainable way while enhancing user privacy? This question goes beyond simply balancing privacy and personalization; it challenges us to expand what we believe is achievable with voice assistants. It encourages us to think outside the usual trade-offs, aiming for a world where personalization and privacy can thrive together, without sacrificing one for the other. The impact of exploring this question could fundamentally change how we approach privacy, personalization, and economic sustainability in the world of digital technology.

3 HYPOTHESIS

Implementing advanced privacy-enhancing technologies within voice assistants can lead to a more user-centric advertising model that not only respects ethical standards of privacy but also proves to be economically beneficial by fostering greater user trust and engagement.

4 RELATED WORK

The academic works surrounding voice-assisted technologies and privacy concerns has been both broad and nuanced, offering a rich foundation for our research. The field opens with a broad overview on the adoption and utility of smart devices, set against the evolving concerns around user privacy. This polarity sets the stage for a deeper research into how these technologies intersect with the everyday lives of users, revealing a complex interplay of convenience, privacy apprehensions, and the desire for greater control over personal data.

Delving into specific challenges, the seminal work by Alepis and Patsakis [3] sheds light on the inherent security vulnerabilities of voice assistants. Their thorough examination of potential threats not only highlights the critical need for robust security measures but also raises questions about the ethical implications of data handling by these devices. This exploration serves as a great foundation, underpinning the importance of designing voice assistants with an impenetrable security framework.

Transitioning from security to user perception, Lau, Zimmerman, and Schaub [4] provide a detailed analysis of privacy perceptions and concerns associated with smart speakers. Their detailed study captures the intricacies of user interactions with these devices, highlighting a palpable tension between the allure of convenience and privacy invasions. This body of work emphasizes the need for voice assistants that not only understand user commands but also respect and protect user privacy.

Building on the understanding of user perceptions, Malkin et al. [5] offer a granular view of privacy attitudes among smart speaker users. Their research, rooted in empirical data, maps the varied landscape of user comfort and concern, advocating for a personalized approach to privacy that can adapt to the diverse expectations of users. This insight directs our focus towards creating a voice assistant that is not only responsive but also respectful of individual privacy preferences.

Further narrowing the scope, Schroeder, Haug, and Gewald [6] explore the unique privacy concerns of mature adults engaging with smart technologies. Their findings reveal the nuanced anxieties of this demographic, particularly in relation to health data (mHealth apps) and smart devices. This perspective supplements our approach, guiding us to consider a broader spectrum of user needs and to ensure that our voice assistant is inclusive, accessible and protective of users at every stage of life.

As we approach to the narrower end, we find the work of Iqbal et al. [7] which delves into the complexities of ad targeting within smart speaker systems, shedding light on the need for enhanced transparency and user autonomy in ad personalization. Their insights resonate with our project's objectives, which focus on providing users with clear and meaningful choices about the advertisements they receive, ensuring that these ads complement rather than compromise user privacy.

Finally, focusing on the specific and often overlooked issue of unintentional device activation, Dubois et al. [8] draw attention to the privacy implications of misactivations in voice assistants. Their work fits well with the goals of our project, which not only recognizes but also addresses the need to minimize accidental data collection, a crucial aspect of maintaining user privacy.

Drawing from these critical contributions to the field, our project seeks to chart a new course in voice assistant development, one that places a premium on safeguarding user privacy. By integrating and building upon the findings of these studies, we aim to create a voice assistant that sets new benchmarks in both functionality and privacy protection, without compromising its economic viability. Our approach underscores the conviction that technology should be designed with the utmost regard for user privacy, establishing a foundation for future innovations in the sector.

5 THEORETICAL CONTRIBUTION

This paper's theoretical contribution spans several key areas, intertwining the advancement of artificial intelligence (AI) with the imperative of privacy protection. It presents a novel exploration of how privacy can be integrated into AI, specifically through the development of a voice assistant that prioritizes user privacy without compromising on functionality. This challenges the prevailing belief that technological convenience necessitates a sacrifice in personal privacy, proposing instead that AI can be ethically designed to respect user autonomy.

Moreover, our project extends the privacy-by-design principles to the burgeoning field of voice-assisted technologies. It illustrates how privacy considerations can be embedded from the ground up in smart devices, embodying the ideals of data minimization, encryption, and explicit user consent in a new context. This not only advances the theoretical framework of privacy-by-design but also sets a precedent for its application in emerging technological domains.

Additionally, through a study with initial feedback, this research deepens the understanding of privacy expectations in the digital age. It unveils nuanced insights into how individuals perceive privacy when interacting with smart technologies, contributing to a more comprehensive theoretical understanding of privacy that encompasses both data protection and respect for personal boundaries.

Lastly, the project envisages a re-calibrated social contract between technology and society, one that prioritizes

human values and ethical considerations. It advocates for a technological landscape where devices and algorithms serve societal needs and uphold principles of dignity, respect, and privacy. This re conceptualization suggests a future where the development and deployment of technology are aligned more closely with the common good, challenging stakeholders to rethink the role of technology in our lives.

Together, these contributions signal a shift towards a more ethical, privacy-conscious technological future, urging a reevaluation of how AI is integrated into daily life and how it can be developed in harmony with human values and societal norms.

6 METHOD

In the development of our privacy-centric voice assistant, a methodological approach was centered around addressing the nuanced and often sensitive nature of user interactions with digital advertisements. To this end, we identified 15 broad categories of content that are considered to be the most invasive and potentially triggering to users. These categories, meticulously selected for their relevance and potential impact on privacy and user experience, are included in the Table 1 below.

Table 1: List of Categories

Political Campaigns	Pharmaceuticals and Supplements
Financial Services	Self-defense and Home Security
Religious Content	Weight Loss and Cosmetic Surgery
Parenting and Fertility Services	Dietary Choices and Veganism
Alcohol and Tobacco	Dating Services
Legal Services	Adult Content and Products
Gambling and Betting	Environmental and Animal Rights
Mental Health Services	Uncategorized

Our goal was to empower users with the ability to control their digital environment actively. By integrating a feature that allows users to block any advertisements related to these categories, we provide a level of customization and control previously unavailable in standard voice assistant offerings. Furthermore, recognizing the deeply personal nature of these topics, our system was designed to offer users the option not to store any data related to their inquiries or interactions with content within these categories.

This approach underscores our commitment to transparency and user empowerment. At all times, users are informed of the data being collected, if any, and are given the full autonomy to manage their privacy settings, including the option to block or allow advertisements and content from the specified categories. This method not only enhances the user's control over their personal data but also significantly reduces the likelihood of encountering unwanted or potentially harmful content, aligning with our overarching aim of creating a safer, more respectful digital ecosystem.

7 SYSTEM DESIGN

This section outlines the architectural evolution of our privacy-centric voice assistant, tracing its development from an initial design phase that utilized a Python-based simulator with SpaCy for natural language processing, to its final iteration, which harnesses OpenAI's API. This progression culminated in the creation of a full-stack application built with Streamlit, showcasing the project's technological advancement.

7.1 Initial Design

The initial design phase of our project was predicated on utilizing SpaCy, a well-regarded open-source natural language processing library. SpaCy was chosen for its comprehensive suite of NLP features, including tokenization, part-of-speech tagging, named entity recognition, and dependency parsing, which we anticipated would be instrumental in accurately categorizing user queries into predefined advertisement categories. Our aim was to create a system that could understand and process user commands with a high degree of accuracy while ensuring privacy.

During this phase, we built a prototype that integrated SpaCy's NLP capabilities to analyze and categorize user queries. The design was structured to be minimalistic, focusing primarily on the backend processing of voice commands. However, as testing progressed, it became evident that while SpaCy offered robust tools for general NLP tasks, its performance in accurately predicting our specific set of advertisement categories from voice queries was insufficient. The accuracy of categorization was below the 40% threshold, a result that led us to reassess our approach to achieving the desired level of precision and efficiency required for our project.

7.2 Final Design

Acknowledging the limitations encountered with SpaCy in our initial design, our project transitioned to a more sophisticated natural language processing (NLP) solution by integrating OpenAI's GPT-3.5 API. This strategic pivot was crucial in overcoming the challenges of accurately categorizing voice queries, a task at which SpaCy's performance had fallen short. GPT's advanced language understanding and generation capabilities were instrumental in significantly enhancing the accuracy of our system's voice query categorization. This integration represented a pivotal moment

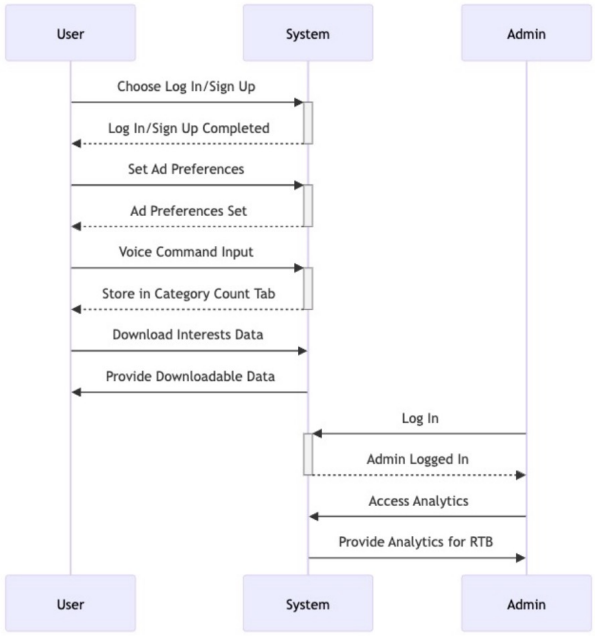

Figure 1: Sequence Diagram illustrating the user-system interaction and administrative data flow in a privacy-centric voice assistant platform

in our project, allowing us to better align our system's capabilities with our core privacy-centric objectives. Figure 1, explains the sequence of work flow within the application.

To complement the backend capabilities provided by GPT, we developed a full-stack application using Streamlit, which served as the primary interface for user interaction with the voice assistant. Streamlit's framework enabled the rapid development of a user-friendly application, incorporating essential features such as login/signup functionality, session management, and secure password storage through SHA-256 hashing. The application was thoughtfully designed to ensure an intuitive and accessible user interface, while the backend efficiently handled the complexities of processing and categorizing voice queries.

A key feature of the final design was its emphasis on data privacy and security. We implemented measures to anonymize user data and limit data retention, reflecting our commitment to a privacy-first approach. Additionally, the system included an admin access page, providing a comprehensive overview of user interactions. This feature was designed with privacy and security in mind, allowing for the monitoring of system usage while maintaining strict data protection standards. Administrators could access detailed user interaction data, supporting system improvements and user experience enhancements, with the capability to download this information for deeper analysis.

The evolution from the initial design, with its reliance on SpaCy, to the final, more refined system leveraging GPT and Streamlit, illustrates our project's journey of technological adaptation and improvement. This progression enabled us to surmount early obstacles and achieve a system that upholds our stringent requirements for privacy, accuracy, and user experience. The final design not only meets these criteria but also sets a new benchmark for the development of privacy-centric voice assistants, balancing advanced functionality with a deep commitment to user privacy.

8 MEASURES

To assess the impact and user reception of our privacy-centric voice assistant, we utilized survey feedback as a pivotal evaluation tool. After engaging with our system via the Streamlit-developed full-stack application, participants provided feedback through a survey, which played a crucial role in understanding the system's strengths and areas for improvement. This feedback covered the system's voice command recognition accuracy, application usability, and the users' views on privacy and security enhancements. The insights gained from this feedback were invaluable, offering both quantitative and qualitative data that illuminated how the system was perceived and the extent to which it met user expectations.

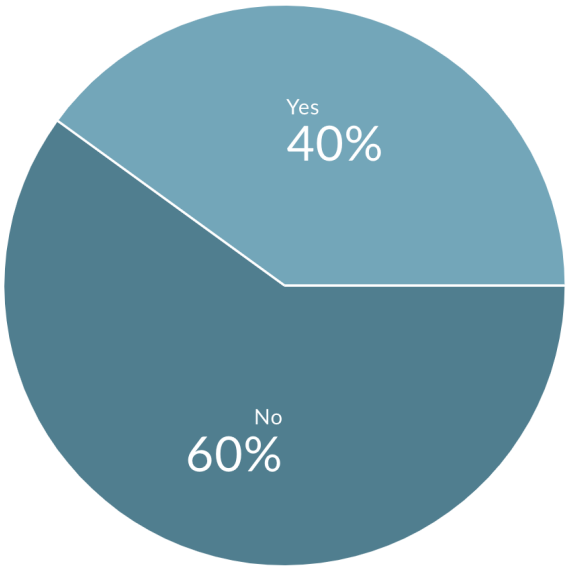

Figure 2: Has your mental state even been affected because of an ad?

Two notable findings were identified from the survey, casting spotlight on user attitudes towards digital advertising

and the potential impact of ads on mental well-being. When participants were asked if they would consider disabling their ad-blockers if our system's approach to less intrusive advertising was implemented, a significant majority, 80% responded negatively, indicating a prevailing reluctance to expose themselves to online advertisements, even if they were less invasive. However, 20% of respondents expressed a willingness to suspend their ad-blocking tools under these conditions, as shown in Figure 3. This minority group's openness suggests that there is a segment of users who could be receptive to a new advertising paradigm, provided it significantly enhances privacy and reduces intrusiveness.

Furthermore, inquiring about the impact of advertisements on users' mental states revealed that while 60% of respondents reported no adverse effects, a substantial 40% acknowledged that their mental well-being had been negatively influenced by ads, as illustrated in . This significant percentage underscores the necessity of addressing the content and delivery of digital advertisements, as it is clear that a considerable portion of users are affected by them. The fact that nearly half of the survey participants have experienced some form of negative impact due to advertisements is compelling evidence that the conversation around digital advertising practices, privacy, and mental health deserves further exploration and action.

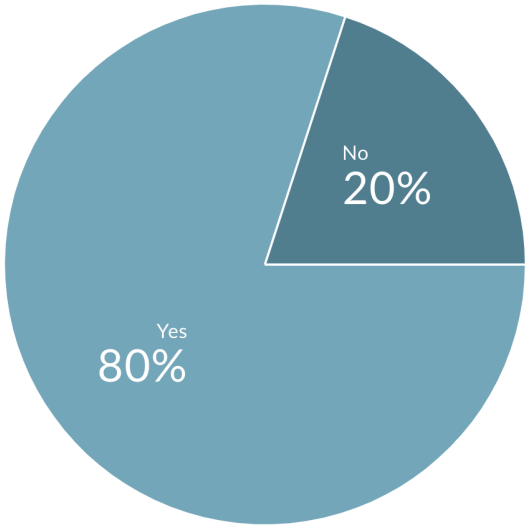

Figure 3: Would you continue to use an ad-blocker if our method were implemented?

These survey results highlight the complexity of user attitudes towards digital advertising and the potential for privacy-centric technologies to reshape these perceptions. While the reluctance to disable ad-blockers remains strong, there is an evident demand for more respectful and less intrusive advertising approaches. Moreover, the acknowledged impact of ads on mental well-being by a significant fraction of users strengthens the case for continued discussion and innovation in how digital advertisements are crafted and targeted, emphasizing the need for a balance between commercial interests and user welfare.

9 RISKS/LIMITATIONS

A primary limitation of our project is its adaptability and the potential market size of users who prioritize privacy in voice-assisted technology. While there is a growing awareness and concern over digital privacy, the extent to which users are willing to switch from established providers to new, privacy-focused alternatives remains uncertain. This was seen with Mycroft, a privacy-focused voice assistant, but it failed to gather enough attention to sustain their development costs. Additionally, NeptuneX's reliance on OpenAI's GPT API introduces a dependency on external API limitations and costs, which could affect scalability and long-term sustainability. Another risk is the potential for the system to misinterpret user commands, especially in complex or noisy environments, which could impact user satisfaction and the overall effectiveness of the assistant.

10 EXECUTION

The execution of our project involved a multifaceted system design process, incorporating the careful selection of technologies and development tools. Initially, we experimented with SpaCy for natural language processing capabilities. However, recognizing the need for enhanced performance in voice query understanding and processing, we transitioned to utilizing OpenAI's API. This strategic shift was instrumental in significantly improving the system's accuracy in categorizing voice commands and generating relevant responses. In parallel with these developments, we crafted a full-stack application using Streamlit to serve as the primary interface for user interaction with the voice assistant. This application was designed to be user-friendly, incorporating secure login/signup functionalities and SHA-256 password hashing to ensure a secure user experience. Adding to our comprehensive system design, Figure 4 showcases the application's functionality for users to set their ad preferences in an intuitive and user-friendly manner. This feature empowers users to have direct control over the types of advertisements they wish to see or block, aligning with our commitment to privacy and user autonomy. Furthermore, Figure 5 provides a glimpse into our model's capabilities by illustrating how a user's voice input, such as "Where can I bet?", is accurately classified into the gambling category. This example underscores the effectiveness of our system

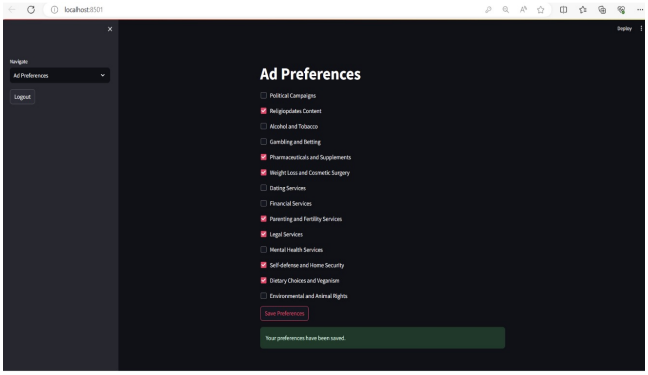

Figure 4: UI for Customizing Ad Preferences

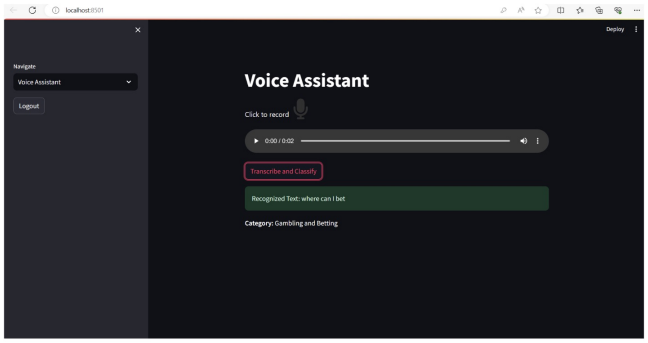

Figure 5: Voice Assistant UI Demonstrating Transcription and Category Classification

in interpreting and categorizing user queries, showcasing the practical application of our advanced NLP solution in action. Together, these figures highlight key aspects of our system's design and functionality, demonstrating the user-centric approach and technological sophistication that define our project.

11 RESULTS

We analyzed feedback from pilot study participants, highlighting significant enhancements in voice command recognition accuracy with the integration of OpenAI's API–a clear-cut improvement over initial tests using SpaCy. This upgrade underscored the potential of advanced AI to refine user interactions while prioritizing privacy. The Streamlit-based application was commended for its ease of use and intuitive design, with privacy and security measures, particularly around data handling and password protection, being recognized as key strengths.

Key insights emerged from Figures 6 and 7, relating to technical performance and policy implications. Figure 6 illustrated a 143% performance increase with GPT-3.5-Turbo over SpaCy, showcasing the advanced AI's capability in processing human language more accurately. Figure 7 revealed that 85% of respondents supported implementing our project's architecture as a policy, reflecting a societal demand for technologies that honor privacy and ethical standards.

These findings highlight our project's technical success and the broader societal call for privacy-focused technological solutions, indicating a pathway for future developments that balance technological innovation with ethical considerations.

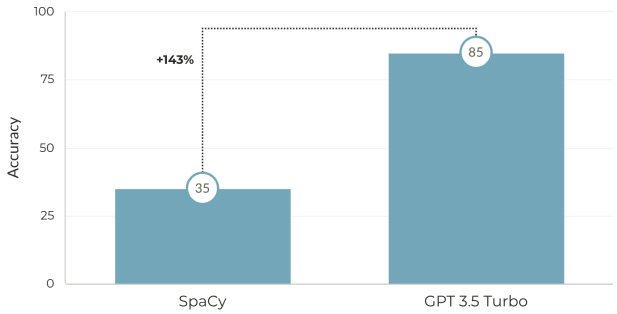

Figure 6: Comparative Analysis of NLP Accuracy: GPT-3.5 Turbo vs. SpaCy

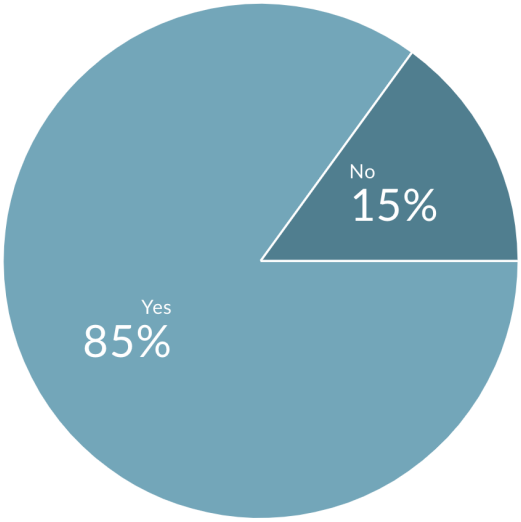

Figure 7: Should NeptuneX Architecture be implemented as a policy?

12 ANALYSIS AND DISCUSSION

The future of privacy-centric voice assistants, as exemplified by our project, invites a nuanced analysis within the broader context of digital privacy, technological advancements, and

user expectations. The increasing awareness and concern over digital privacy suggest a promising market for devices and applications that prioritize user privacy. However, this is contingent upon overcoming challenges related to adaptability, market acceptance, and the continuous evolution of privacy laws and norms. Our project's journey from initial design considerations, through the integration of advanced NLP solutions, to the development of a user-friendly application, underscores the potential for such technologies to address pressing privacy concerns while maintaining high usability standards.

Further, the role of AI in enabling these privacy-centric solutions while ensuring accuracy and relevance in user interactions points towards a future where technological innovation does not come at the expense of personal privacy. Nonetheless, the discussion also acknowledges limitations, such as the project's scalability and the broader adaptability of privacy-focused technologies in a market dominated by established players with different priorities.

The implementation of privacy-centric architectures by companies hinges not only on the technological feasibility but also on economic benefits or a significant enough impact on users affected by sensitive ads. While our project has laid the groundwork for such technologies, it is clear that further research is required to explore the economic and social incentives for companies to adopt these privacy-first approaches. The case of Gillian Brockell serves as a poignant example that there is not only room but a pressing need for discussion and action in this domain. Although our project does not provide definitive answers to these complex questions, it underscores the importance of continuing the conversation and research into how digital platforms can be reimagined to better serve and protect user privacy. Brockell's experience highlights the critical need for a shift in how digital advertising and user data are handled, suggesting that a future where technology respects personal boundaries is not only desirable but essential.

13 FUTURE WORKS

To broaden the impact and applicability of our privacy-centric approach, one immediate avenue for future work involves the integration of our system within online websites. This expansion aims to empower users with the ability to block ads related to certain sensitive topics directly through their web browsers. By implementing a plugin or browser extension that leverages the underlying technology of our voice assistant, users can gain an easy-to-access tool for enhancing their online privacy. This tool would analyze the content of the ads and the context in which they are served, offering users the option to filter out ads based on their personal preferences and sensitivities. The development of such a feature would not only enhance user control over their digital environment but also contribute to the broader ecosystem of privacy-focused tools, making digital spaces safer and more respectful of individual privacy concerns.

Additionally, to address concerns regarding data privacy and the reliance on external APIs, another critical area of future work involves the development of a transformer-based model that can be deployed locally on a user's machine. This model would emulate the functionality of our current system, processing and categorizing queries without the need to transmit data to a remote server. Building a local, transformer-based NLP model presents a significant step towards ensuring user data remains private and secure, as all processing would be done on the user's own hardware. Such an approach not only aligns with the overarching goal of enhancing privacy but also opens up possibilities for personalized model training, where the system learns and adapts to the individual's unique patterns of interaction and inquiry. By pursuing this direction, we aim to further empower users, giving them complete control over their data and the functionality, thereby strengthening the foundation of trust and privacy that our project seeks to build.

14 ACKNOWLEDGEMENTS

We are particularly thankful to Prof. Alexander Gamero-Garrido for his pivotal role in shaping the direction of this project. His course inspired us to explore and develop policies aimed at enhancing digital privacy, guiding our efforts towards meaningful outcomes. Additionally, we wish to express our appreciation to our friends and colleagues who generously participated in our feedback survey, providing insights that were crucial for refining our project. Their engagement and support have been invaluable to our research. We would also like to acknowledge the use of ChatGPT in designing the sequence diagram and for assisting in the refinement of some textual elements of our work, contributing to the clarity of our work.

REFERENCES

[1] Statista. Smart speaker devices installed base in the united states from 2017 to 2020, 2022.
[2] The Washington Post. Dear tech companies, i don't want to see pregnancy ads after my child was stillborn, 2018.
[3] Efthimios Alepis and Constantinos Patsakis. Monkey says, monkey does: security and privacy on voice assistants. *IEEE Access*, 5:17841–17851, 2017.
[4] Joseph Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. 2(CSCW):Article 102, 2018.
[5] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies*, 2019(4), 2019.

[6] Torben Schroeder, Maximilian Haug, and Heiko Gewald. Data privacy concerns using mhealth apps and smart speakers: Comparative interview study among mature adults. *JMIR Formative Research*, 6(6):e28025, 2022.

[7] Usman Iqbal, Parinaz Nikkhah Bahrami, Riza Trimananda, Hang Cui, Antonio Gamero-Garrido, Daniel J Dubois, David Choffnes, Athina Markopoulou, Franziska Roesner, and Zubair Shafiq. Tracking, profiling, and ad targeting in the alexa echo smart speaker ecosystem. 2023.

[8] Daniel Dubois, Roman Kolcun, Anna Mandalari, Muhammad Paracha, David Choffnes, and Hamed Haddadi. When speakers are all ears: Characterizing misactivations of iot smart speakers. *Proceedings on Privacy Enhancing Technologies*, 2020:255–276, 2020.