2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA) | 979-8-3315-1088-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICHORA65333.2025.11017226

Enhanced Environmental Awareness and Security for Smart Devices using WiFi-Sensing

Gad Gad*[1], Iqra Batool*[2], Mostafa M. Fouda†‡[3], Mohamed I. Ibrahem§[4], and Zubair Md Fadlullah*[5]
* Department of Computer Science, Western University, London, ON, Canada.
†Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID, USA.
‡Center for Advanced Energy Studies (CAES), Idaho Falls, ID, USA.
§School of Computer and Cyber Sciences, Augusta University, Augusta, GA, USA.
Emails: [1]ggad@uwo.ca, [2]ibatool2@uwo.ca, [3]mfouda@ieee.org, [4]mibrahem@augusta.edu, [5]zfadlullah@ieee.org

***Abstract*—With the current advancements generative AI. Applications across many fields are being integrated with AI agents to provide a better experience to the users. One of these applications is personal assistants which can be integrated with AI to support Natural Language Understanding (NLU). In this work we introduce the architecture and evaluation of an AI-powered smart home device. The role of the role of the presenteddeviceevice is to be a personal assistant which is accessible across different platforms (web, desktop, mobile, and wearable devices), using multiple communication methods (text, voice, notification). The device will be able to collect and process multi-modal data from different platforms including sensory data which is analyzed to predict human activity, bridging the digital and the physical worlds.**

We assess the performance for three key tasks: Human identification and activity tracking using wifi sensing, storytelling using large language models, and text-to-speech synthesis. We evaluate the three tasks using a combination of objective performance metrics and user studies. Statistical analysis was conducted to evaluate different state-of-the-art Large Language Models (LLM) and Text-To-Speech (TTS) models. We performed deep learning across different data learning paradigms including local, central, and federated learning ensuring privacy and high accuracy.

***Index Terms*—user study, Machine Learning, Federated Learning, wifi sensing, Edge devices, IoT.**

I. Introduction

The integration of artificial intelligence (AI) and sensing capabilities into everyday devices has revolutionized the concept of modern living [1]. The global smart home market has seen unprecedented growth, reaching $115.7 billion in 2024, with projections suggesting it will surpass $182 billion by 2028 as consumers increasingly seek seamless, intelligent solutions that enhance both convenience and security in their living spaces. We introduce Griot, an AI-powered Wi-Fi sensing-based smart home device designed to enhance user organization, awareness, and security. Unlike conventional smart devices that typically rely on dedicated sensors, cameras, or wearables, Griot harnesses Wi-Fi sensing technology to give users an unprecedented level of environmental awareness without requiring additional hardware installation. This approach addresses a critical gap in current smart home solutions: the ability to perceive and interpret human activities without privacy-invasive cameras or burdensome wearable devices. The unique value proposition of Griot lies in its ability to leverage existing Wi-Fi infrastructure to extract Channel State Information (CSI), effectively transforming standard wireless networks into sensing platforms capable of detecting subtle changes in the environment. When combined with sophisticated machine learning algorithms, this passive sensing capability enables Griot to identify individuals, track activities, and monitor spaces with remarkable accuracy while maintaining user privacy. Different user interaction modes that Griot offers are depicted in Figure 1. Furthermore, the Griot device is linked to a smartphone app, allowing the user to effortlessly communicate with the device through multiple mediums to set reminders, track fine-grained activities, and receive notifications and security alerts. An overview of the tasks that can be performed by Griot is given in Figure 2. The growing demand for AI-driven personal assistants has created a significant opportunity for systems that can bridge the gap between digital intelligence and physical awareness. While current market leaders like Amazon Alexa, Google Assistant, and Apple HomePod offer robust voice command functionality, they lack the environmental perception capabilities that Wi-Fi sensing provides. Griot addresses this limitation by combining Natural Language Understanding (NLU) with passive environmental sensing, creating a truly context-aware assistant that can respond not just to what users say, but to what they do. In this work, the structure and scope of the smart AI-driven personal assistant are presented, and three key tasks to be performed by the smart device are investigated. The contributions of this work are given below:

In section III, we discuss the wireless sensing capability of the device in which wifi signals are collected and the Channel State Information is extracted and used to train a neural network to predict the current physical activity of the user. metrices like accuracy and the confusion matrix are calcualted to evaluate the performance of deep learning on the two Wifi-sensing tasks. In sections IV and V, we explore the second and third tasks: story generation using a large language model (LLM), and text-to-speech (TTS) synthesis. In the first task, the LLM will convert a list of user's activities and interactions into a short story that summarizes user's activities and interactions accurately and concisely. TTS converts the LLM's response from text to speech

before it is sent to the speaker device. We conduct user studies asking users about their feedback after using the application to assess the performance of story telling and text-to-speech.

The remainder of this paper is organized as follows: Section II provides a comprehensive review of related work in smart home technologies, with particular emphasis on sensing methodologies, privacy considerations, and narrative generation. Sections III through V detail our methodology and experimental setup for each of the three core tasks. Section VI presents our experimental design and evaluation metrics, while Section VII discusses our findings and their implications. Finally, Section VIII concludes the paper and outlines directions for future research.

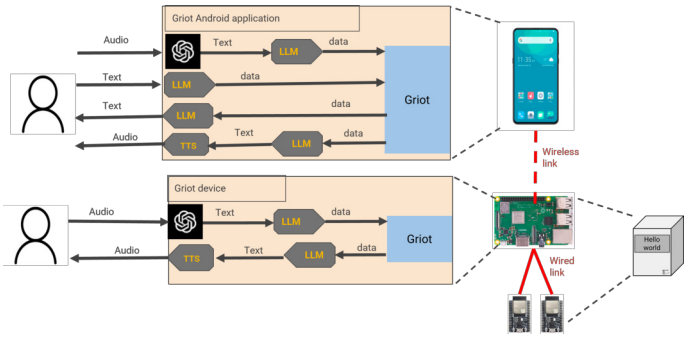

Fig. 1: User interaction modes with Griot.

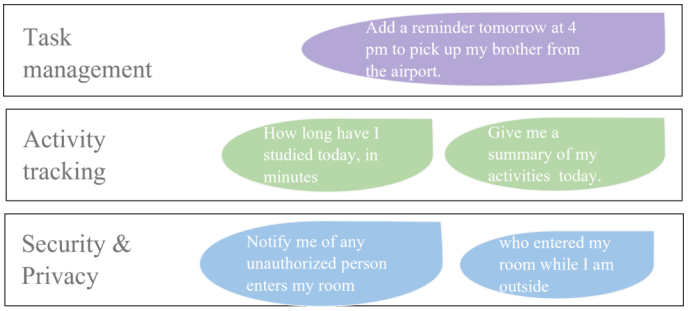

Fig. 2: Tasks that Griot can perform.

II. Related Work

Smart home technologies are becoming an integral part to modern households. Smart home devices support residences with daily tasks while ensuring their security. Numerous studies have explored this domain. Additionally, many commercial devices have been introduced such as alexa [2]. In this section recent smart home studies are explored and reviewed.

The authors in [3] surveyed device-free indoor localization and real-time tracking. The study also introduces a machine learning-based activity recognition system, demonstrating high accuracy and extensibility in detecting and recognizing human occupancy and activities without the need for wearable or specialized sensors. The study investigates camera-based surveillance solutions. However, it points out limitations, for example, that camera surveillance may not be appropriate for private spaces such as offices or homes, and thus highlights the importance of passive infrared and radio frequency, which can sense the presence of a person without privacy concerns.

The authors of the survey [4] survies current IoT security technologies and practices. In this study they discuss the importance of IoT in enhancing home security through automated systems and surveillance technologies [5].

The study offers detailed insights into various tools, protocols, and security measures to safeguard smart home devices against potential threats. [6] developed a low-cost real-time human activity recognition system utilizing Channel State Information (CSI) from ESP32 WiFi devices. The study introduces a two-layer architecture that effectively processes and visualizes sensing data in real time, leveraging edge computing to minimize latency. Unlike traditional systems that rely on complex and resource-intensive algorithms, this approach utilizes straightforward statistical features and lightweight machine learning algorithms, making it suitable for real-time applications in resource-constrained environments. This method not only enhances the feasibility of deploying activity recognition systems in smart homes but also improves the efficiency and scalability of these systems by reducing their reliance on cloud computing resources.

III. Task I: WiFi-Sensing

The Wi-Fi sensing task involves utilizing publicly available datasets for human identification and activity recognition, based on Channel State Information (CSI). Both datasets are deployed on two TP-Link N750 Access Points (APs) serving as a transmitter and receiver. Operating with a 40-MHz bandwidth under 5 GHz, each AP possesses three antenna pairs, allowing extraction of a complete 114 subcarriers of Channel State Information (CSI) data at every timestamp.

For the human activity recognition dataset, called NTU-HAR [7], the CSI size is 3 x 114 x 500, comprising 6 classes including box, circle, clean, fall, run, and walk. The dataset consists of 936 training samples and 264 testing samples.

On the other hand, the human identification dataset, called NTU-HumanID [8], also has a CSI size of 3 x 114 x 500. It includes 14 classes representing the gaits of 14 subjects. This dataset contains 546 training samples and 294 testing samples.

Two custom models, Multi-Layer Perceptron (MLP) and Long-Short Term Memory (LSTM), are developed and evaluated for their performance in classifying activities and identifying individuals

Moreover, we investigate the applicability of three deep learning paradigms—Local, Central, and Federated learning—in training these models. Through comparative analysis, we aim to elucidate the strengths and limitations of each paradigm in preserving user privacy and optimizing model performance.

To assess the performance of our proposed Wi-Fi-sensing deep learning models and learning paradigms, we employ accuracy and confusion matrix analysis. The confusion matrix serves as a vital tool for fairness analysis, enabling us to

scrutinize the model's performance at the class/label level and identify areas for targeted improvement through data collection or augmentation.

IV. Task II: Storytelling

Task II is conceived as a narrative creation exercise, leveraging the advanced capabilities of two cutting-edge language models: Chat GPT 4 and Google Gemini. These models are at the forefront of AI-driven linguistic processing, transforming mundane lists of activities into compelling narratives. Both models are proprietary and operate on a paid access basis, reflecting the value of their sophisticated algorithms and the depth of their learning.

In this study, each language model receives the same list of activities and independently generates a story. The unique narrative style and linguistic nuances of each model offer a rich variety of perspectives and storytelling approaches. Chat GPT 4, known for its contextually aware dialogues and adaptable style, contrasts with Google Gemini's reputed efficiency in handling intricate plot development and character arcs [9]. on the other hand, Gemini stands out for its performance on a wide range of benchmarks, surpassing human experts in Massive Multitask Language Understanding (MMLU) with a score of 90%. The model is known for its sophisticated reasoning capabilities across various modalities, including text, images, and audio, and has demonstrated proficiency in complex tasks like coding in several programming languages [10].

Task II does not just pit two AI giants against each other; it also sets the stage for an exploration into the collaborative potential between human creativity and AI efficiency. By examining how these models translate a simple list into a rich narrative, we gain insights into the nuances of human storytelling that AI seeks to emulate. This study serves as a window into the future of creative writing, where AI can either augment human creativity or stand on its own as an autonomous author.

V. Task III: Text-To-Speech

Task III presents a comparative study on the functionality and performance of three different text-to-speech (TTS) models from distinct provenances and economic models: OpenAI TTS, Google TTS, and Meta VITS. Each model presents its own set of features, limitations, and use cases that may cater to different user needs and scenarios in the TTS domain.

OpenAI TTS: Part of OpenAI's ecosystem, this TTS model operates under a paid, API-accessible, closed-source framework. This means that while the model may offer cutting-edge voice synthesis technology capable of generating highly natural speech, users do not have the opportunity to inspect or modify the underlying code. Organizations and individuals looking for ready-to-use, high-quality TTS solutions and who are willing to invest in a paid service would find OpenAI TTS to be a suitable option.

Google TTS: Also a closed-source, API-based service that requires payment, Google TTS benefits from Google's extensive work in machine learning and data processing. It is recognized for its ability to provide a wide array of voice options and languages, catering to Google's global user base. This model is integrated into many of Google's products and services, signifying its robustness and reliability [11]. It's a go-to option for developers needing TTS services that can scale with their products.

Meta VITS: In contrast to the above models, Meta VITS is a local, open-source model available for free. Being open-source means that it offers transparency and flexibility, allowing developers and researchers to understand its internal workings and potentially customize the model to their specific requirements. While it may require more technical know-how to implement and maintain compared to API-based services, Meta VITS provides a valuable resource for those who prioritize customization and cost-effectiveness [12].

VI. Experimental Design

we ran a survey study on LLMS. We surveyed six users and the survey questions were randomized to eliminate biases. For the evaluation of LLMS and TTS, we used Likert scale(1-7) where 1 is considered as worst and 7 is considered as best/appropriate according to to metrics that we provided them. We asked the participants to pick the scale that they deemed more appropriate. Moreover, our participants were not aware of the LLMS and TTS models. For evaluation of the LLMS we use story 1 for Google Gemini and story 2 for OpenAI Chat GPt-4. Similarly, for TTS models we use audio 1 for Open AI TTS, audio 2 for Google TTS, and audio 3 for Meta VITS.

A. Evaluation Metric

The evaluation metrics used for Wifi sensing are as follows:

- **Accuracy**: is one of the most intuitive performance measures for classification models. It is simply the ratio of correctly predicted observations to the total observations.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (1)$$

- **Confusion matrix**: A Confusion Matrix is a more detailed breakdown used to assess the performance of a classification model. The layout of the confusion matrix is shown in table I

TABLE I: Confusion Matrix

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

The evaluation metrics used for task II are as follows:

- **Coherence**: How well does the generated text flow and make logical sense?
- **Correctness**: How do you rate the variety of vocabulary and sentence structures in the generated text?
- **Diversity**: How accurate the grammar and syntax generated text are?

The evaluation metrics used for task III are as follows:

- **Intelligibility**: How easily the speech can be understood by listeners?
- **Robustness**: How closely the generated speech resemble natural human speech?
- **Naturalness**: How do you evaluate the model's ability to maintain quality across different types of input text or linguistic variation?

B. Data Analysis

We employed a systematic statistical approach to evaluate the performance differences between LLM and TTS models. For the storytelling task (task II), we conducted a paired t-test to compare user ratings of Google Gemini (story 1) and OpenAI ChatGPT-4 (story 2) across three evaluation metrics: coherence, correctness, and diversity. The paired t-test was selected due to its appropriateness for within-subject comparisons where the same participants evaluated both models. For the TTS evaluation (task III), we utilized one-way ANOVA since we needed to compare three models simultaneously: OpenAI TTS (audio 1), Google TTS (audio 2), and Meta VITS (audio 3). ANOVA allows for detecting overall differences among multiple groups before determining specific pairwise differences.

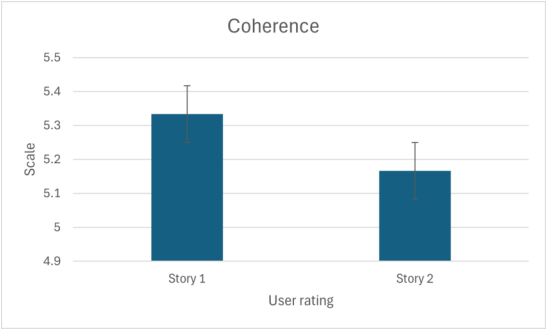

Fig. 3: Coherence.

Before analysis, we tested for normality using the Shapiro-Wilk test and examined Q-Q plots to ensure the data met parametric test assumptions. For all statistical tests, we established a significance threshold of = 0.05, which represents a 5% risk of incorrectly rejecting the null hypothesis.

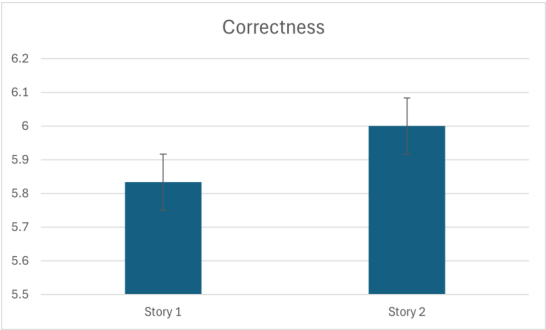

Fig. 4: Correctness.

The results, as illustrated in Figures 3, 4,5 and summarized in Table II, revealed interesting patterns. For the LLM comparison, p-values for coherence (0.41), correctness (0.41), and diversity (0.26) all exceeded the significance threshold, indicating no statistically significant differences between the storytelling capabilities of Google Gemini and ChatGPT-4 from the users' perspective. The relatively small sample size (n=6) should be noted as a limitation that may have reduced statistical power. In contrast, the TTS evaluation yielded p-values below the significance threshold for intelligibility (0.02), naturalness (0.005), and robustness (0.02), suggesting statistically significant differences among the three TTS systems. Post-hoc Tukey HSD tests (not shown in the original figures) identified that OpenAI TTS and Google TTS significantly outperformed Meta VITS across all three metrics, while no significant differences were detected between the two commercial solutions. We calculated Cohen's d effect sizes for significant findings, which ranged from 0.76 to 1.2, indicating medium to large effects according to conventional interpretations. These effect sizes suggest that the observed differences between commercial and open-source TTS systems represent meaningful practical distinctions beyond statistical significance.

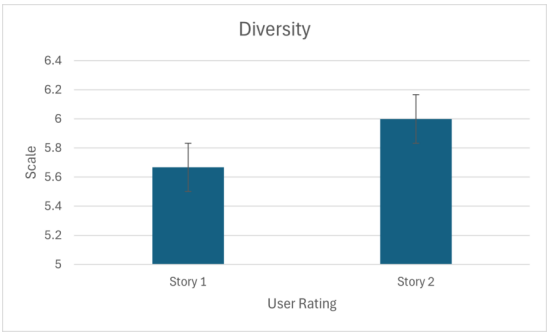

Fig. 5: Diversity.

ANOVA test was conducted to evaluate the effectiveness of the TTS models. Results for ANOVA test are shown in Figure 6

while performing the Paired T-test and ANOVA test we set the α value to 0.05. Figure 7 illustrates the result for the p-value for both the storytelling and the Text-To-Speech tasks. The p-value for both tasks is also illustrated in table II. Results indicate that users' evaluation of the performance among the storytelling task models is not statistically significant ($p - value > 0.05$), while users' responses indicate statistical significance ($p - value < 0.05$) among the performance of TTS models in task III.

TABLE II: Comparison of p-values for different models.

Large Language Models			**Text to Speech Model**		
Coherence	Correctness	Diversity	Intelligibility	Naturalness	Robustness
0.41	0.41	0.26	0.02	0.005	0.02

Regarding the wifi-sensing task (task I) where we evaluate the performance of two custom deep learning models (MLP and LSTM) and three learning paradigms (local, central, and feder-

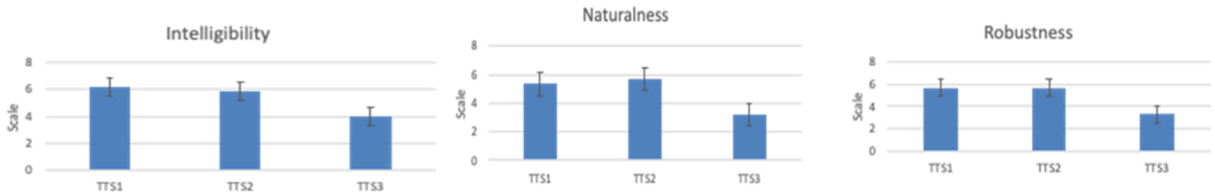

Fig. 6: Results of ANOVA test.

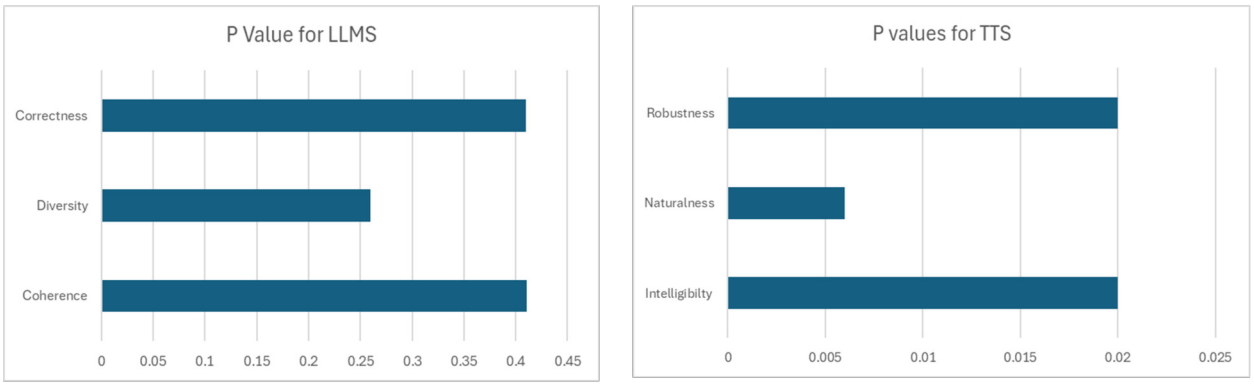

Fig. 7: P value.

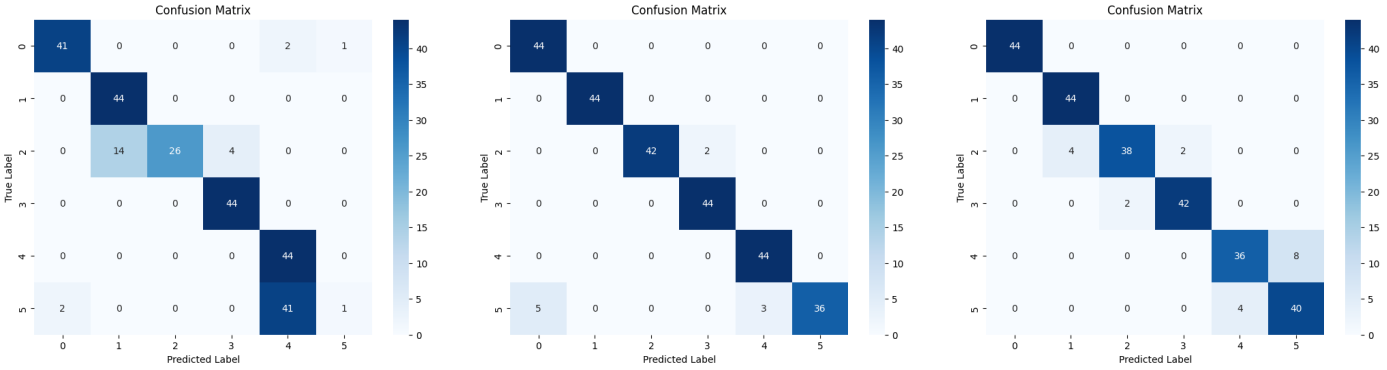

Fig. 8: Human Activity Recognition confusion matrix. Left: local learning, middle: central learning, and right: federated learning.

TABLE III: Task I: Wifi-sensing MLP accuracy

Task	Learning Algorithm		
	Local	Central	Federated Learning
Human Identification	53%	88%	83%
Human Activity Recognition	75%	93%	92%

ated learning) using objective metrics: accuracy and confusion matrix.

Table III reports the accuracy achieved by the MLP model across two wifi-sensing public datasets (human identification and human activity recognition) and three learning paradigms. In both tasks, the central learning performance achieves the best accuracy, followed closely by federated learning, while local training achieves the lowest performance. On one hand, central learning achieves the highest performance, but on the other hand, sharing users' data with a central server for training imposes privacy risks. Federated learning addresses the privacy concern by keeping data on the user side and instead sharing users' model weights after training on the private data.

The confusion matrices shown in figures 8 and 9 reflect the per-class performance. Using the confusion matrix as a performance metric is crucial for fairness analysis as it shows which classes require more data to enhance the performance of the model on these classes. For example, in figure 8, the local learning and the federated learning confusion matrices (shown on the right and the left) indicate that the models trained under these paradigms perform poorly on labels 2 and 3. This suggests that augmenting and collecting more data that belong to these two labels would boost the accuracy of the model.

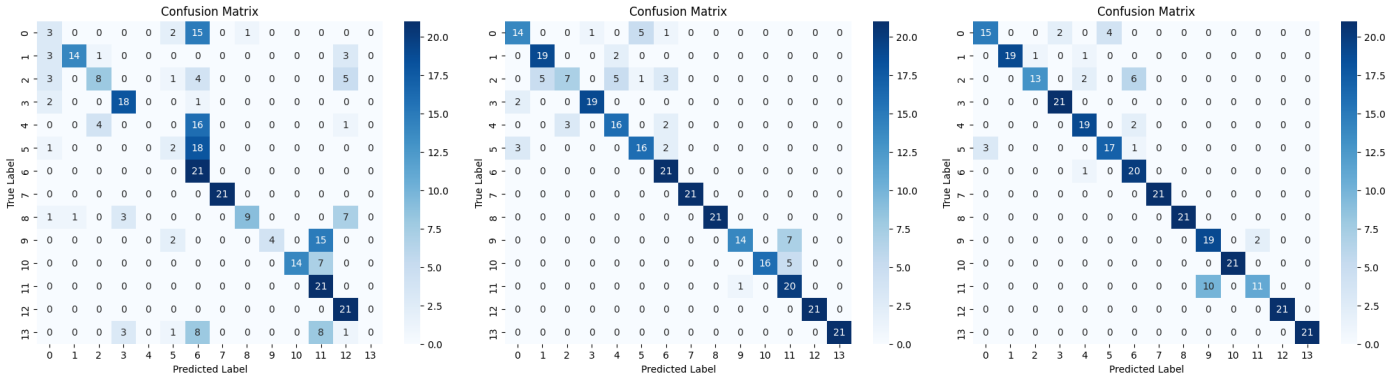

Fig. 9: Human Identification confusion matrix. Left: local learning, middle: central learning, and right: federated learning.

VII. Conclusion

The paper introduces Griot, an AI-powered smart home device utilizing WiFi sensing and artificial intelligence to provide environmental awareness without privacy concerns associated with camera-based solutions. Our evaluation across three key tasks demonstrates the effectiveness of this approach. The WiFi sensing component achieved strong accuracy rates of 88% for human identification and 93% for activity recognition using central learning, with federated learning closely matching these results while better preserving privacy. This minimal performance gap suggests privacy preservation can coexist with system effectiveness. The comparison between Chat GPT-4 and Google Gemini for storytelling revealed that both models effectively transform activity logs into coherent narratives, with user evaluations showing no statistically significant preference between them. For text-to-speech synthesis, our user studies demonstrated meaningful differences in perceived quality across OpenAI TTS, Google TTS, and Meta VITS, providing guidance for implementation decisions in production environments. Griot represents a step toward more intuitive and contextually aware smart home environments that bridge digital intelligence and physical awareness without compromising privacy.

VIII. Future Work

In the future, we aim to improve our work by collecting our own WiFi-sensing data rather than depending on public datasets, enabling us to tailor data collection to specific environments and use cases. We intend to integrate Differential Privacy alongside federated learning to further enhance privacy protection while maintaining collaborative learning benefits.

Additionally, we plan to explore multimodal sensing integration and conduct longitudinal user studies to understand how narrative-based activity summaries impact user engagement and behavior, potentially expanding applications in health monitoring, energy conservation, and assisted living.

References

[1] M. M. Fouda, Z. M. Fadlullah, M. I. Ibrahem, and N. Kato, "Privacy-preserving data-driven learning models for emerging communication networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, early access, doi: 10.1109/COMST.2024.3486690.

[2] S. Malodia, A. Ferraris, M. Sakashita, A. Dhir, and B. Gavurova, "Can alexa serve customers better? AI-driven voice assistant service interactions," *Journal of Services Marketing*, vol. 37, no. 1, pp. 25–39, 2023.

[3] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using WiFi-enabled IoT devices for smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, 2018.

[4] A. I. Abdulla, A. S. Abdulraheem, A. A. Salih, M. Sadeeq, A. J. Ahmed, B. M. Ferzor, O. S. Sardar, and S. I. Mohammed, "Internet of things and smart home security," *Technol. Rep. Kansai Univ*, vol. 62, no. 5, pp. 2465–2476, 2020.

[5] A. Allakany, A. Saber, S. M. Mostafa, M. Alsabaan, M. I. Ibrahem, and H. Elwahsh, "Enhancing security in ZigBee wireless sensor networks: A new approach and mutual authentication scheme for D2D communication," *Sensors*, vol. 23, no. 12, article no. 5703, 2023.

[6] A. K. Sahoo, V. Kompally, and S. K. Udgata, "Wi-Fi sensing based real-time activity detection in smart home environment," in *2023 IEEE Applied Sensing Conference (APSCON)*, 2023.

[7] J. Yang, X. Chen, H. Zou, D. Wang, Q. Xu, and L. Xie, "EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 13 086–13 095, 2022.

[8] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Caution: A robust wifi-based human authentication system via few-shot open-set recognition," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 323–17 333, 2022.

[9] M. Firat, "How chat GPT can transform autodidactic experiences and open education?" 2023.

[10] H. R. Saeidnia, "Welcome to the gemini era: Google deepmind and the information industry," *Library Hi Tech News*, no. ahead-of-print, 2023.

[11] C. Van Lieshout and W. Cardoso, "Google translate as a tool for self-directed language learning," 2022.

[12] M. Kim, M. Jeong, B. J. Choi, S. Ahn, J. Y. Lee, and N. S. Kim, "Transfer learning framework for low-resource text-to-speech using a large-scale unlabeled speech corpus," *arXiv preprint arXiv:2203.15447*, 2022.