Wi-Chat: Large Language Model-powered Wi-Fi-based Human Activity Recognition

Yili Ren*
University of South Florida
Tampa, FL, USA
yiliren@usf.edu

Haopeng Zhang*
University of Hawaii at Manoa
Honolulu, HI, USA
haopengz@hawaii.edu

Haohan Yuan
University of Hawaii at Manoa
Honolulu, HI, USA
haohany@hawaii.edu

Jingzhe Zhang
University of South Florida
Tampa, FL, USA
jingzhe@usf.edu

Yitong Shen
University of South Florida
Tampa, FL, USA
shen202@usf.edu

Abstract

Recent advancements in LLMs have demonstrated remarkable capabilities across diverse tasks. However, their potential to integrate physical model knowledge for real-world signal interpretation remains largely unexplored. In this work, we introduce Wi-Chat, the first LLM-powered Wi-Fi-based human activity recognition system. We demonstrate that LLMs can process Wi-Fi signals and infer human activities by incorporating the physical model of Wi-Fi sensing into prompts. Specifically, our approach leverages physical model insights to guide LLMs in interpreting Channel State Information (CSI) data without traditional signal processing techniques or labor-intensive training. The experiments show that LLMs exhibit strong reasoning capabilities, achieving zero-shot activity recognition. These findings highlight a new paradigm for Wi-Fi sensing and expand LLM applications beyond conventional language tasks.

CCS Concepts

• Human-centered computing → Ubiquitous and mobile computing systems and tools; **• Computing methodologies → Artificial intelligence**.

*Both authors contributed equally to this research.

EnvSys '25, Anaheim, CA, USA

ACM ISBN 979-8-4007-1986-8/25/06
https://doi.org/10.1145/3742460.3742979

Keywords

Large Language Model, Wi-Fi Sensing, Channel State Information, Human Activity Recognition

ACM Reference Format:
Yili Ren, Haopeng Zhang, Haohan Yuan, Jingzhe Zhang, and Yitong Shen. 2025. Wi-Chat: Large Language Model-powered Wi-Fi-based Human Activity Recognition. In *International Workshop on Environmental Sensing Systems for Smart Cities (EnvSys '25), June 23–27, 2025, Anaheim, CA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3742460.3742979

1 Introduction

Large Language Models (LLMs) have achieved significant advancements in generating human-like conversations, marking a transformative shift in human-AI interactions [11]. Moreover, the latest LLMs demonstrated remarkable reasoning capabilities and exceptional generalization skills. However, their reliance on training with collections of textual content from the Internet leaves them considerably distant from achieving a profound understanding of the physical world. Meanwhile, the ubiquitous Wi-Fi devices and the extensive coverage of Wi-Fi networks present an opportunity to expand Wi-Fi capabilities beyond communication, particularly in sensing the physical world [12]. As Wi-Fi signals traverse the physical environment, they interact with surrounding people and objects, causing reflection, diffraction, scattering, etc. Consequently, the received signals can carry a substantial amount of information about both people and objects. Conventional Wi-Fi-based sensing systems can achieve various sensing tasks in the physical environment, such as human activity recognition and localization [7]. However, these systems typically rely on complex signal processing techniques and the labor-intensive training of machine learning or deep learning models. This raises a fundamental and compelling question: *Can we integrate LLMs with Wi-Fi sensing to interpret the physical world without complex signal processing and in a zero-shot manner?*

We investigate this question by exploring the capabilities of LLMs to understand Wi-Fi signals and incorporating physical model knowledge of Wi-Fi sensing. Specifically, we introduce Wi-Chat, an LLM-powered Wi-Fi sensing system for human activity recognition. Unlike existing LLMs that primarily analyze traditional textual and visual data, Wi-Chat can understand Wi-Fi signals which are real-world projections of human activity. We demonstrate that LLMs, having been trained on extensive human knowledge, when integrated with the physical models of Wi-Fi sensing, can be directly leveraged for Wi-Fi signal analysis. This approach can derive deep insights that traditionally require complex signal processing and machine learning or deep learning models trained on large volumes of labeled data.

Wi-Chat directly inputs smoothed raw Wi-Fi signals into well-known LLMs, such as ChatGPT, DeepSeek, and LLaMA, for human activity recognition. Additionally, we integrate physical models of Wi-Fi sensing into LLMs via prompting, enabling a deeper understanding of human activities in the physical world through Wi-Fi signals. To evaluate Wi-Chat, we conduct experiments using a self-collected human activity dataset. We compare its performance against conventional Wi-Fi-based human activity recognition systems and basic machine learning models. Our results demonstrate that LLMs can achieve zero-shot human activity recognition directly from Wi-Fi signals, attaining an accuracy of 90%.

2 Related Work

Wi-Fi Sensing. Wi-Fi sensing has been widely studied for applications like human activity recognition due to its non-contact nature and low cost [6, 9, 16]. E-eyes [14] first used Wi-Fi signals for daily activity recognition, while CARM [13] leveraged a hidden Markov model for temporal feature extraction. Yang et al. [18] applied CNNs and RNNs to extract activity-related features. Despite strong performance, these methods rely on multi-stage signal processing and require large datasets for training.

Large Language Model Applications. LLMs have transformed NLP research with their ability to understand, analyze, and generate text using vast pre-trained knowledge [20, 21]. Beyond NLP, they also advance fields like healthcare, law, and finance [3, 4, 19]. Researchers also explored LLMs for sensing. Penetrative AI[17] integrates LLMs with the physical world for sensor data analysis, while HARGPT[5] applies LLMs to human activity recognition using IMU data. These highlight LLMs' growing role in physical-world sensing.

3 Background

3.1 Wi-Fi Sensing

Wi-Fi signals travel through line-of-sight (LoS) and reflected paths, bouncing off objects and humans before reaching the receiver. We use Channel State Information (CSI) to capture Wi-Fi signal changes caused by a target. CSI characterizes how signals propagate through space, traveling via a LoS path and multiple reflections from objects and people. It represents the superposition of signals from all paths and can be expressed as: $H(f,t) = \sum_{i=1}^{N} a_i e^{-j2\pi \frac{d_i(t)}{\lambda}}$, where a_i is the attenuation and $d_i(t)$ is the length of the i^{th} path, N is the number of paths, λ is the wavelength, and f is the frequency. CSI can be further decomposed into static and dynamic components. The static component consists of the LoS signals and reflections from stationary objects. In contrast, the dynamic component arises from reflections caused by the moving target, such as a person. For simplicity, we assume that there is only a single signal reflection from the target. Thus, the CSI can be denoted as: $H(f,t) = H_s(f,t) + H_d(f,t) = H_s(f,t) + a(f,t)e^{-j2\pi \frac{d(t)}{\lambda}}$, where $H_s(f,t)$ is the static component, $a(f,t)$, $e^{-j2\pi \frac{d(t)}{\lambda}}$, and $d(t)$ are the complex attenuation, phase shift and path length of dynamic component $H_d(f,t)$, respectively.

3.2 Approaches to Wi-Fi-based Human Activity Recognition

Figure 1 illustrates different approaches to Wi-Fi-based human activity recognition. A Wi-Fi transmitter emits signals, which are received and processed to capture human activities using Wi-Fi CSI extracted from network interface controllers. This work examines the following distinct approaches.

1) Conventional Wi-Fi-based Sensing Systems: Conventional Wi-Fi sensing begins with signal denoising, including phase offset removal and outlier filtering [1]. Signal transformation methods like FFT, STFT, DWT are then applied for time-frequency analysis. Next, PCA, ICA, and SVD are commonly used for feature extraction, signal separation, and dimensionality reduction. Finally, machine learning models are trained to map Wi-Fi signals to activity labels for activity recognition. *2) Machine Learning Models with Raw Signals:* An alternative approach feeds raw Wi-Fi signals directly into machine learning models like CNNs, RNNs, and SVMs, with only basic signal smoothing. However, similar to conventional systems, they still require extensive manual labeling and training. *3) Wi-Chat: Zero-shot Inference with LLM:* In Wi-Chat, the input data consists of raw Wi-Fi signals, processed with simple signal smoothing. We explicitly instruct LLMs to recognize a person's activity by analyzing Wi-Fi signals. By integrating physical models of Wi-Fi sensing into prompts, we provide physical model guidance to LLMs.

4 System Design

4.1 Physical Model Knowledge of Wi-Fi Sensing

In this section, we analyze the physical models of Wi-Fi sensing in terms of human walking, falling, breathing, and

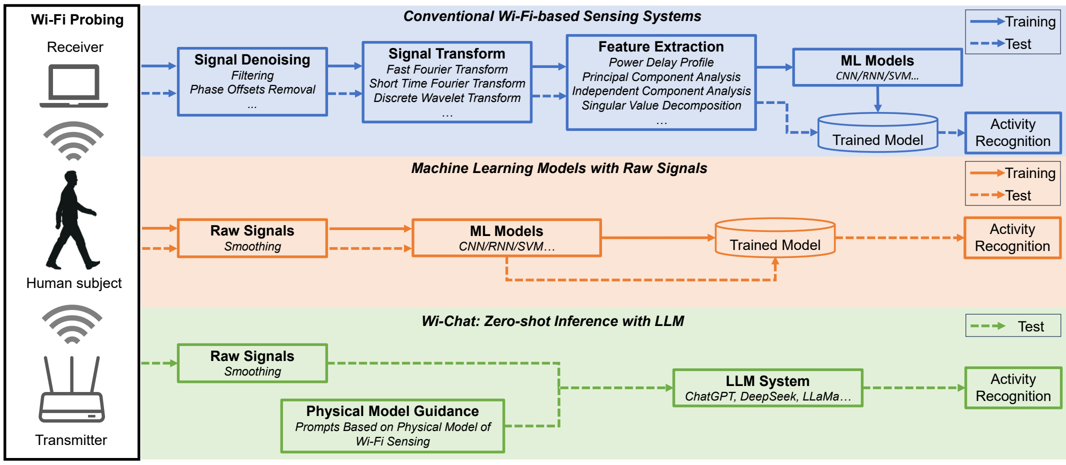

Figure 1: The system flows of different approaches to Wi-Fi-based human activity recognition.

no-event scenarios. Then we derive the LLM prompting guidance according to such knowledge.

Walking. As shown in Figure 2(a), a person is walking, with the Wi-Fi transmitter and receiver placed at fixed locations. The Wi-Fi transmitter emits signals that propagate through the multipath environment. The static components (i.e., solid red lines) include the LoS signal and signals reflected off the wall. The dynamic components (i.e., dotted blue lines) consist of signals reflected off the human body. As the person moves from location P_A to location P_N, the path length of the dynamic component changes. The duration of this movement is denoted as ΔT.

We can plot the signals in the in-phase quadrature (IQ) plane [13] as illustrated in Figure 2(b). The static component vector $\vec{H}_s$ remains fixed, while the dynamic component vector $\vec{H}_d$ can change and rotate. The overall CSI $\vec{H}$ is the sum of vectors $\vec{H}_s$ and $\vec{H}_d$. When the dynamic and static component vectors are aligned in the same direction (e.g., at P_N), they add constructively, resulting in the maximum CSI amplitude (i.e., $|\vec{H}| = |\vec{H}_s|+|\vec{H}_d|$). Conversely, when they are in opposite directions (e.g., at P_B), they add destructively, minimizing the CSI amplitude (i.e., $|\vec{H}| = |\vec{H}_s|-|\vec{H}_d|$). Note that when the path length of the dynamic component changes by one wavelength (e.g., about 6 *cm* for 5 GHz Wi-Fi), its phase rotates by 2π [10]. Since walking is a large-scale activity, each step can cause changes of many wavelengths in the propagation path, resulting in multiple phase rotations in the dynamic components. This leads to multiple peaks ($|\vec{H}_s| + |\vec{H}_d|$) and troughs ($|\vec{H}_s| - |\vec{H}_d|$) for CSI amplitude ($|\vec{H}|$) as shown in Figure 2(c). We note that walking is a continuous activity with a relatively long duration ΔT. We summarize the physical model knowledge of walking as follows: *"Walking is a large-scale activity that induces significant changes in the Wi-Fi CSI amplitude over time, characterized by the presence of numerous peaks and troughs."*

Falling. Similar to walking, falling is a large-scale activity in which a person moves from P_A to P_N, as depicted in Figure 3(a). This movement causes the overall CSI amplitude ($|\vec{H}|$) to reach both maximum and minimum values, corresponding to $|\vec{H}_s|+|\vec{H}_d|$ and $|\vec{H}_s|-|\vec{H}_d|$, respectively, as shown in Figures 3(b) and (c). However, the duration ΔT of a fall can be very short (e.g., about 0.5 seconds). As a result, the peaks and troughs caused by the fall are concentrated within a brief period. After signal smoothing, these rapid fluctuations can be approximated as a single significant peak or trough. Following the fall, the person could remain motionless, leading to a static period after the fall. Therefore, we characterize the physical model knowledge of falling as follows:: *"Falling is a large-scale and sudden activity that induces a single significant peak/trough in the Wi-Fi CSI amplitude, followed by a relatively stable period."*

Breathing. Human breathing is a small-scale activity, as the typical range of chest expansion and contraction during a breath is only a few centimeters (from P_A to P_B, as shown in Figure 4(a)). This leads to dynamic path length changes that are typically very small. Thus, the overall CSI amplitude may not reach its maximum or minimum values, meaning that $(|\vec{H}_s| - |\vec{H}_d|) < |\vec{H}| < (|\vec{H}_s| + |\vec{H}_d|)$, as illustrated in Figures 4(b) and (c). Furthermore, breathing is a continuous and smooth activity, meaning its duration (ΔT) is long. We formulate the physical model knowledge of breathing as follows: *"Breathing is a small-scale and smooth activity that causes slow and gradual changes in Wi-Fi CSI amplitude over time, with a moderate variation range."*

No-event Scenario. In this scenario, only static signal components exist, such as the LoS signals and the signals reflected by stationary objects, as depicted in Figures 5(a) and (b). Since no movement occurs, no dynamic component is introduced into the Wi-Fi signal propagation. As a result, the overall CSI amplitude is determined solely by the amplitude

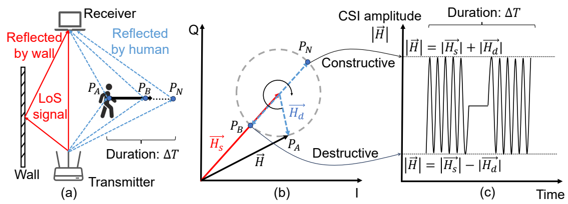

Figure 2: Modeling the signals of walking.

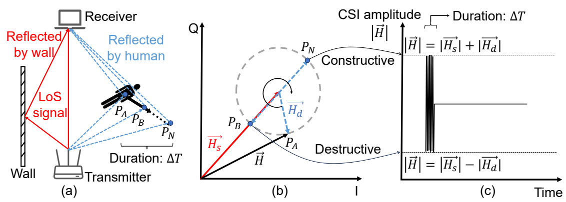

Figure 3: Modeling the signals of falling.

of these static signal components and remains nearly constant over time (i.e., ($|\vec{H}| = |\vec{H}_s|$), as illustrated in Figure 5(c). The physical model knowledge of the no-event scenario is: *"In the no-event scenario, the time-series CSI amplitude remains stable, meaning the variation range is very small."*

We further illustrate examples of real Wi-Fi signals corresponding to different human activities in Figure 6. These signal patterns closely align with our physical model for Wi-Fi sensing and can be described using the aforementioned physical model knowledge.

4.2 Prompting Strategies for Wi-Fi-Based Activity Recognition

This section outlines different prompting strategies for leveraging LLMs in Wi-Fi-based human activity recognition. We aim to explore how LLMs can interpret Wi-Fi signals and improve activity recognition without extensive model training or complex signal processing.

1) Base: For the base setting, we provide the LLM with raw CSI amplitude data, represented as a time series, and prompt it to recognize human activity directly. *2) In-context Learning (ICL):* Recent studies have demonstrated that LLMs exhibit strong few-shot learning capabilities across various tasks, a phenomenon known as ICL [2]. By learning from exemplars and physical model knowledge of Wi-Fi sensing, LLM can recognize patterns in the signals and improve its classification accuracy without additional fine-tuning. *3) Chain-of-Thought (CoT):* Beyond simple input-output mappings, CoT reasoning into prompts can further enhance the model's interpretability [8, 15]. By including explicit intermediate steps and physical model knowledge-based reasoning, CoT prompting helps the model better capture the relationships between signal patterns and human activities. By exploring these prompting strategies, we aim to assess the feasibility of LLMs for Wi-Fi-based activity recognition and understand how different types of input representations influence their performance. Examples of prompts are shown in Figure 7.

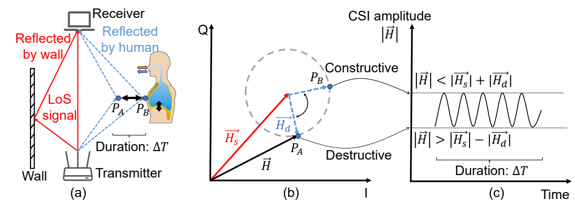

Figure 4: Modeling the signals of breathing.

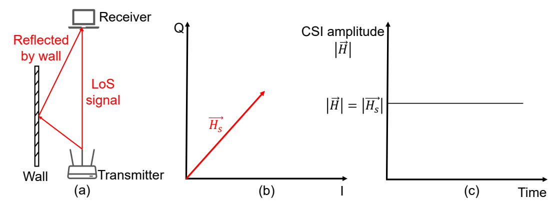

Figure 5: Modeling the signals of no-event scenario.

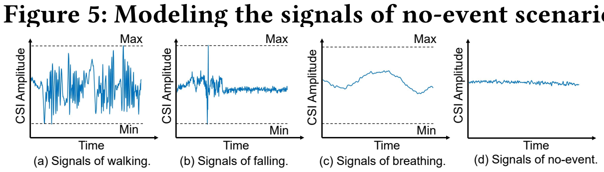

Figure 6: Signals of different human activities.

5 Experiment

5.1 Data Collection

We conducted experiments using Dell LATITUDE laptops as both transmitters and receivers, each with three antennas. The Wi-Fi channel operated at 5.32 GHz with a 40 MHz bandwidth and a transmission rate of 1,000 packets per second. We used the Linux 802.11 CSI tool to extract Channel State Information (CSI) from 30 OFDM subcarriers per packet. The dataset includes over 1,965,000 CSI packets collected from five participants (2 females, 3 males) with varying heights (160–185 cm), weights (50–80 kg), and ages (22–35 years). Each participant performed four activities—walking, falling, breathing, and staying still—in three environments: a bedroom, kitchen, and living room. The study was approved by the IRB of the authors' institution.

Objective: You are given a time series of Wi-Fi CSI amplitude values captured in an environment. Your task is to recognize human activity based on Wi-Fi CSI.
Input Data: {signal}.

Base:
Determine the person's activity.
The activity belongs to one of the following categories: {activity_list}.

In-context Learning (ICL):
Walking: Large and multiple changes in amplitude over time. Example_1: {signal_1, activity: walking}.
Falling: A single significant peak or trough followed by a relatively stable period. Example_2: {signal_2, activity: falling}.
Breathing: Smooth and moderate changes in amplitude over time. Example_3: {signal_3, activity: breathing}.
No-event: The amplitude remains mostly stable with minimal fluctuations. Example_4: {signal_4, activity: no-event}.
Determine the person's activity. The activity belongs to one of the following categories: {activity_list}.

Chain-of-Thought (CoT):
Let's think step by step.
(1) What is the variation range of the data? If the variation range is very small, does it suggest a no-event scenario?
(2) Does the data change smoothly over time with a moderate variation range? If so, could this indicate breathing?
(3) Identify large peaks or troughs characterized by a significant increase or decrease, which then return to the overall range.
(4) How many large peaks or troughs are present in the data?
(5) If the data contains only one significant peak or trough (dramatically larger than others), does it transition into a relatively stable period with minor variations? If so, does this indicate a falling event?
(6) If multiple large peaks and troughs are occurring regularly, does this suggest a walking activity?
Based on the above step-by-step analysis, determine the person's activity. The activity belongs to one of the following categories: {activity_list}.

Figure 7: Prompts based on physical model knowledge of Wi-Fi sensing.

5.2 Baselines

We compare Wi-Chat with the following baseline systems. Conventional Wi-Fi-based sensing systems follow a multi-step pipeline as described in Section 3.2. Specifically, we reproduce two well-known systems: *1) CARM* [13]: It utilizes a PCA-based method for signal denoising, applies DWT for feature extraction, and employs a Hidden Markov Model for activity recognition. *2) E-eyes* [14]: It filters out data outliers and then builds activity classifiers using Earth Mover's Distance. We also evaluate the performance of machine learning models, including *3) CNN*, *4) RNN*, and *5) SVM*. These models take the smoothed CSI amplitude as input and are trained in a supervised manner using labeled datasets.

5.3 Experimental Settings

LLMs. In the zero-shot setting, we directly input the time-series smoothed CSI amplitude into the LLMs without any prior examples. For the few-shot setting, we provide four examples, each representing a different activity.

Baselines. For all supervised baselines, we randomly split the dataset into 70% training and 30% testing. CNN and RNN models are trained on an NVIDIA RTX 4090 with a 0.001 learning rate and the Adam optimizer. The SVM model uses a Radial Basis Function kernel. For CARM and E-eyes, we follow their original pipelines, including denoising, feature extraction, and model construction. Zero-shot evaluations use untrained models to assess their performance. We evaluate all methods using accuracy, precision, recall, and F1-score.

5.4 Overall Performance

Table 1 summarizes the results of human activity recognition across different methods and systems. In the zero-shot setting, conventional Wi-Fi-based sensing systems and machine learning models exhibit relatively low recognition performance due to their inability to handle unseen data. In contrast, the LLM model GPT-4o achieves an accuracy of 0.47, surpassing both conventional Wi-Fi-based systems and machine learning models. The ICL approach significantly enhances performance, demonstrating the advantages of in-context learning combined with physical model knowledge of Wi-Fi sensing, achieving an accuracy of 0.77. Furthermore, GPT-4o, augmented with physical model knowledge-based CoT reasoning, attains the highest accuracy of 0.90 in the zero-shot setting with unseen Wi-Fi data. This result highlights the effectiveness of advanced prompting techniques and the integration of physical model knowledge in improving LLM-powered Wi-Fi-based human activity recognition.

In the supervised learning setting, models achieve accuracies exceeding 0.94. This is expected as supervised models are explicitly trained on labeled data, enabling them to learn precise decision boundaries. Additionally, conventional Wi-Fi-based sensing systems, which integrate signal processing techniques with supervised models, achieve accuracies above 0.98. However, despite their high accuracy, supervised methods typically require extensive labeled datasets and complex signal processing. Notably, the best performance of Wi-Chat is already comparable to that of conventional Wi-Fi-based activity recognition systems and machine learning models. This finding suggests that LLMs demonstrate strong performance in zero-shot and few-shot settings for Wi-Fi-based human activity recognition, making them particularly valuable in real-world scenarios with limited annotated data.

Table 1: Performance comparison of different methods and systems.

Method	Accuracy	Precision	Recall	F1-score
Zero/Few-shot				
E-eyes (zero-shot)	0.26	0.26	0.27	0.26
CARM (zero-shot)	0.24	0.24	0.24	0.24
SVM (zero-shot)	0.27	0.28	0.28	0.27
CNN (zero-shot)	0.23	0.24	0.23	0.23
RNN (zero-shot)	0.26	0.26	0.26	0.26
Wi-Chat: GPT-4o: base (zero-shot)	0.47	0.62	0.47	0.53
Wi-Chat: GPT-4o: ICL (4-shot)	0.77	0.84	0.77	0.80
Wi-Chat: GPT-4o: CoT (zero-shot)	**0.90**	**0.91**	**0.90**	**0.90**
Supervised				
SVM	0.94	0.92	0.91	0.91
CNN	0.98	0.98	0.97	0.97
RNN	0.99	0.99	0.99	0.99
Signal Processing + Supervised				
CARM	0.98	0.98	0.98	0.98
E-eyes	1.00	1.00	1.00	1.00

5.5 Comparison of Different LLMs

Figure 8 presents a comparative analysis of the performance of various LLMs under zero-shot base settings. The figure highlights notable differences in accuracy among the models. GPT-4o and DeepSeek demonstrate the highest performance, achieving accuracy rates exceeding 54%, whereas models such as Gemma2, Mistral, and LLaMA exhibit significantly lower accuracy, falling below 32%. These results suggest that larger models, such as GPT-4o, benefit from enhanced reasoning capabilities, likely due to more extensive training data and sophisticated architectures. In contrast, smaller models show varying degrees of effectiveness, potentially due to limitations in their parameter sizes or architectural constraints. This performance gap underscores the impact of model scale and design on zero-shot generalization.

6 Discussion

Our work is the first to leverage LLMs for interpreting Wi-Fi signals, but it has limitations:

Simple Sensing Task. We focus on human activity recognition involving four activities. Extending it to complex activity recognition may require integrating prompt-based understanding with Wi-Fi signal embeddings within LLMs.

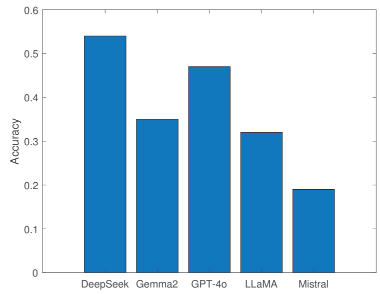

Figure 8: Comparison of LLMs under the zero-shot (base) settings.

Robustness against Interference. While effective across environments and individuals, our system remains sensitive to interference from other people and objects. Improving robustness in dynamic settings is an open challenge.

Privacy Concerns. Wi-Fi-based sensing raises privacy issues, which may be further amplified by the use of LLM-powered sensing. Future research could explore privacy-preserving approaches to LLM-based sensing.

7 Conclusion

In this paper, we introduced Wi-Chat, the first LLM-powered Wi-Fi-based human activity recognition system that combines the reasoning capabilities of LLMs with the sensing capabilities of wireless signals. Our approach directly inputs raw Wi-Fi signals into LLMs while incorporating physical model knowledge of Wi-Fi sensing. Experimental results demonstrate the strong potential of LLMs in enabling zero-shot Wi-Fi sensing. These findings suggest a novel paradigm for human activity recognition that eliminates the need for extensive labeled data or complex signal processing.

References

[1] Kamran Ali, Alex X Liu, Wei Wang, and Muhammad Shahzad. 2017. Recognizing keystrokes using WiFi devices. *IEEE Journal on Selected Areas in Communications* 35, 5 (2017), 1175–1190.

[2] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[3] Zhiyu Zoey Chen, Jing Ma, Xinlu Zhang, Nan Hao, An Yan, Armineh Nourbakhsh, Xianjun Yang, Julian McAuley, Linda Petzold, and William Yang Wang. 2024. A Survey on Large Language Models for Critical Societal Domains: Finance, Healthcare, and Law. *arXiv preprint arXiv:2405.01769* (2024).

[4] Kai He, Rui Mao, Qika Lin, Yucheng Ruan, Xiang Lan, Mengling Feng, and Erik Cambria. 2025. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. *Information Fusion* (2025), 102963.

[5] Sijie Ji, Xinzhe Zheng, and Chenshu Wu. 2024. HARGPT: Are LLMs Zero-Shot Human Activity Recognizers? *arXiv preprint arXiv:2403.02727* (2024).

[6] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun, and Lu Su. 2020. Towards 3D human pose construction using WiFi. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

[7] Yongsen Ma, Gang Zhou, and Shuangquan Wang. 2019. WiFi sensing with channel state information: A survey. *ACM Computing Surveys (CSUR)* 52, 3 (2019), 1–36.

[8] Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, et al. 2021. Show your work: Scratchpads for intermediate computation with language models. *arXiv preprint arXiv:2112.00114* (2021).

[9] Yili Ren, Sheng Tan, Linghan Zhang, Zi Wang, Zhi Wang, and Jie Yang. 2020. Liquid level sensing using commodity wifi in a smart home environment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–30.

[10] Yili Ren, Zi Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2021. Winect: 3d human pose tracking for free-form activity using commodity wifi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–29.

[11] Alaa Saleh, Praveen Kumar Donta, Roberto Morabito, Naser Hossein Motlagh, Sasu Tarkoma, and Lauri Lovén. 2025. Follow-me ai: Energy-efficient user interaction with smart environments. *IEEE Pervasive Computing* (2025).

[12] Sheng Tan, Yili Ren, Jie Yang, and Yingying Chen. 2022. Commodity WiFi sensing in ten years: Status, challenges, and opportunities. *IEEE Internet of Things Journal* 9, 18 (2022), 17832–17843.

[13] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st annual international conference on mobile computing and networking*. 65–76.

[14] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. 617–628.

[15] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903* (2022).

[16] Yaxiong Xie, Jie Xiong, Mo Li, and Kyle Jamieson. 2019. mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[17] Huatao Xu, Liying Han, Qirui Yang, Mo Li, and Mani Srivastava. 2024. Penetrative ai: Making llms comprehend the physical world. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*. 1–7.

[18] Jianfei Yang, Han Zou, Yuxun Zhou, and Lihua Xie. 2019. Learning gestures from WiFi: A siamese recurrent convolutional architecture. *IEEE Internet of Things Journal* 6, 6 (2019), 10763–10772.

[19] Haohan Yuan, Siu Cheung Hui, and Haopeng Zhang. 2024. A Structure-aware Generative Model for Biomedical Event Extraction. *arXiv preprint arXiv:2408.06583* (2024).

[20] Haopeng Zhang, Xiao Liu, and Jiawei Zhang. 2023. SummIt: Iterative Text Summarization via ChatGPT. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. 10644–10657.

[21] Haopeng Zhang, Philip S Yu, and Jiawei Zhang. 2024. A Systematic Survey of Text Summarization: From Statistical Methods to Large Language Models. *arXiv preprint arXiv:2406.11289* (2024).