

Penetrative AI: Making LLMs Comprehend the Physical World

Huatao Xu[†], Liying Han[§], Qirui Yang^{*}, Mo Li^{*†}, Mani Srivastava[§]

^{*}Hong Kong University of Science and Technology,

[†]Nanyang Technological University, [§]University of California Los Angeles

Email: huatao001@ntu.edu.sg, {liying98, mbs}@ucla.edu, {qyangau, lim}@cse.ust.hk

ABSTRACT

Recent developments in Large Language Models (LLMs) have demonstrated their remarkable capabilities across a range of tasks. Questions, however, persist about the nature of LLMs and their potential to integrate common-sense human knowledge when performing tasks involving information about the real physical world. This paper delves into these questions by exploring how LLMs can be extended to interact with and reason about the physical world through IoT sensors and actuators, a concept that we term "Penetrative AI". The paper explores such an extension at two levels of LLMs' ability to penetrate into the physical world via the processing of sensory signals. Our preliminary findings indicate that LLMs, with ChatGPT being the representative example in our exploration, have considerable and unique proficiency in employing the embedded world knowledge for interpreting IoT sensor data and reasoning over them about tasks in the physical realm. Not only this opens up new applications for LLMs beyond traditional text-based tasks, but also enables new ways of incorporating human knowledge in cyber-physical systems.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; • **Computer systems organization** → **Embedded and cyber-physical systems**.

KEYWORDS

LLM, CPS, IoT, Penetrative AI

ACM Reference Format:

Huatao Xu, Liying Han, Mo Li, and Mani Srivastava. 2024. Penetrative AI: Making LLMs Comprehend the Physical World. In *The 25th International Workshop on Mobile Computing Systems and Applications (HOTMOBILE '24)*, February 28–29, 2024, San Diego, CA, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3638550.3641130>

1 INTRODUCTION

Large Language Models (LLMs) have made remarkable strides [2, 26, 34]. A particularly revolutionary milestone is ChatGPT [22], which excels in fluid, human-like conversations, marking a new era

Mo Li and Mani Srivastava are the corresponding authors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HOTMOBILE '24, February 28–29, 2024, San Diego, CA, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0497-0/24/02...\$15.00

<https://doi.org/10.1145/3638550.3641130>

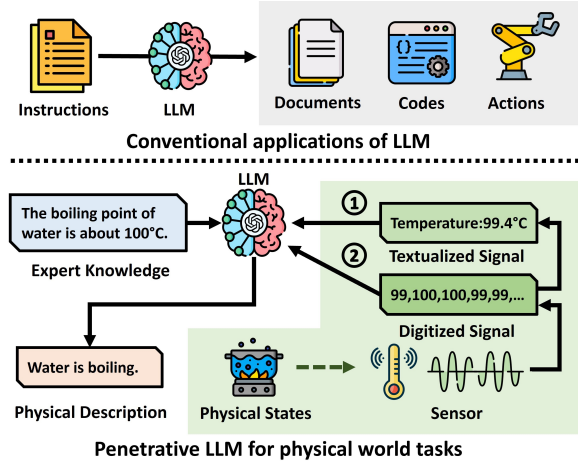


Figure 1: Overview of Penetrative AI.

in human-AI interactions. These latest LLMs cultivated on extensive text datasets have showcased remarkable capabilities across diverse tasks, including coding and logical problem-solving [5]. These out-of-the-box capabilities have demonstrated that they already comprise enormous amounts of world knowledge¹.

This paper is motivated by an essential and intriguing question: can we enable LLMs to complete tasks in the real physical world? We delve into this inquiry and explore extending the boundaries of LLMs' capabilities by directly letting them interact with the physical world through Internet of Things (IoT) sensors. A basic example of this process is depicted in Figure 1, where different from the conventional way of LLMs, an LLM is expected to analyze sensor data which are indeed projections from the physical world. We conjecture that LLMs, having been trained on vast amounts of human knowledge, learned the physical world which can be directly harnessed for analysis of such sensory information to derive deep insights that traditionally require background knowledge from human experts and/or bespoke machine learning models trained with large amounts of labeled sensor data. If this conjecture were not true, we would observe from LLMs irrelevant responses, inaccurate physical status, or inefficacious actions.

As illustrated in Figure 1, we formulate such a problem from a signal processing's point of view, and specifically explore the LLMs' penetration into the physical world at two signal processing levels with the sensor data: i) with the textualized signals derived from underlying sensor data, and ii) with the digitized signals, essentially numerical sequences of raw sensor readings. We term this

¹Some studies referred to it as a world model [14] of how the world works.

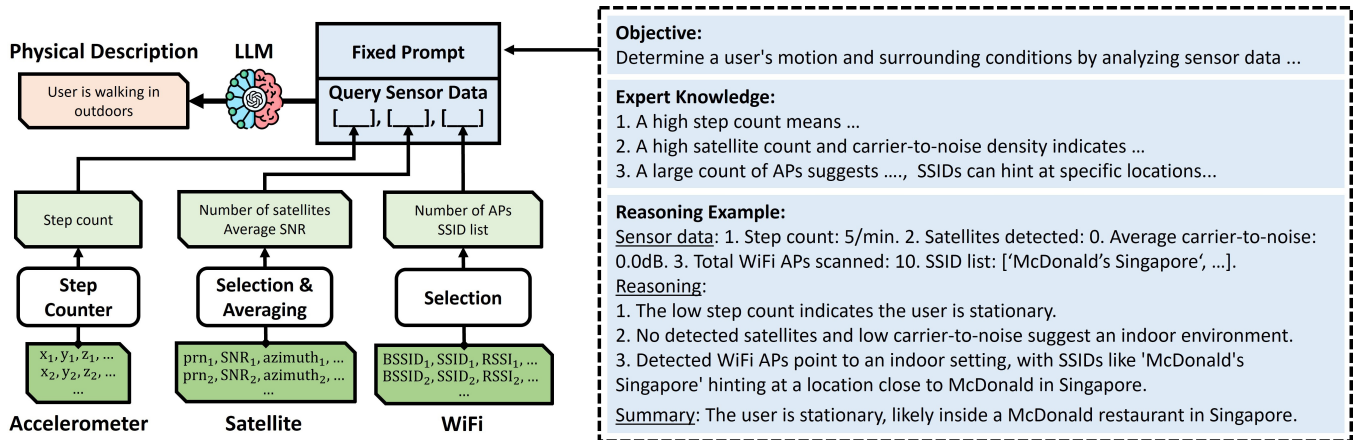


Figure 2: Overview of user activity sensing with LLMs.

endeavor "*Penetrative AI*" – where the embedded world knowledge in LLMs serves as a foundation model, seamlessly integrated with the Cyber-Physical Systems (CPS) for perceiving and intervening in the physical world.

Our methodology is exemplified through two illustrative applications at two different levels, respectively - user activity sensing where textualized signals from smartphone accelerometer, satellite, and WiFi data are analyzed to discern user motion and environment conditions, and human heartbeat detection where digitized electrocardiogram (ECG) data are utilized to derive the heartbeat rate. Preliminary findings are encouraging, showcasing LLMs' proficiency in interpreting IoT sensor data and performing perception tasks in the physical world. Our exploration also underscores that existing LLMs, such as ChatGPT-4, may already possess the capability to establish intricate connections among world knowledge and can be guided to tackle CPS tasks.

2 PENETRATIVE LLM WITH TEXTUALIZED SIGNALS

This section describes our effort in tasking ChatGPT, a chosen vehicle, to comprehend sensor data at the textualized signal level.

2.1 An Illustrative Example

We take activity sensing as an illustrative example, where we task ChatGPT with the interpretation of sensor data collected from smartphones to derive user activities. The input sensor data encompass smartphone accelerometer, satellite, and WiFi signals, and the desired output is to discern the user motion and environment context. Figure 2 presents the overview of this LLM-based design – the sensor data are pre-processed by individual sensing components and the textualized sensor states are supplied to the LLM with a fixed prompt for activity inference.

Objective and rationale. We convey a clear task to ChatGPT – "determine a user's motion and surrounding conditions by analyzing sensor data from their smartphones". The basic idea is that when the user conducts different activities in different environments, the collected sensor data would exhibit varied patterns, which reveal the users' activities.

Data preparation. To facilitate ChatGPT comprehension of the sensor data, we undertake a preprocessing step where raw data from different sensing modules are separately converted into textualized states that are expected interpretable by ChatGPT. Figure 2 illustrates such a step.

To pre-process long accelerometer readings (6,000 samples from 10 seconds of triaxial accelerations sampled at 200 Hz), we employ the Android step detector, which is a built-in step-counting implementation [7] and can transform the 6,000 raw data points into a single textually expressed state, e.g., "step count: 5/min".

The Android system also offers a comprehensive set of Global Navigation Satellite System (GNSS) satellite measurements [6], including information like Pseudo-Random Noise as a satellite identifier, Signal-to-Noise Ratio (SNR), and many others. To streamline the data for ChatGPT interpretation, we filter and distill the satellite data into two key attributes: the number of detected satellites and their average SNR.

The Android system supports scanning for nearby APs and provides comprehensive information about scanned APs [8]. Similar to satellite data, we disregard less relevant details and focus on critical information – Service Set Identifier (SSID) and Received Signal Strength Indicator (RSSI). To streamline the data and reduce text length, we further filter APs with an RSSI lower than -70 and instruct ChatGPT to analyze the SSIDs to capture useful location information.

Expert knowledge. We guide LLMs by including explicit text-based descriptions of the relationship between sensor patterns and user activity states in the prompts, as illustrated in Figure 2. For instance, a high satellite count and carrier-to-noise density indicate an outdoor setting with strong satellite signals.

Reasoning examples. Following expert knowledge, we can provide a set of reasoning examples to enhance the proficiency of ChatGPT. Each example includes the data for processing, a step-by-step reasoning process, and a brief summary of the ground truth context. Figure 2 illustrates this with the reasoning example section.

Complete prompt. A full prompt includes a defined objective and expert knowledge of the sensor data, all in natural language as demonstrated in Figure 2. Essentially, the way we construct

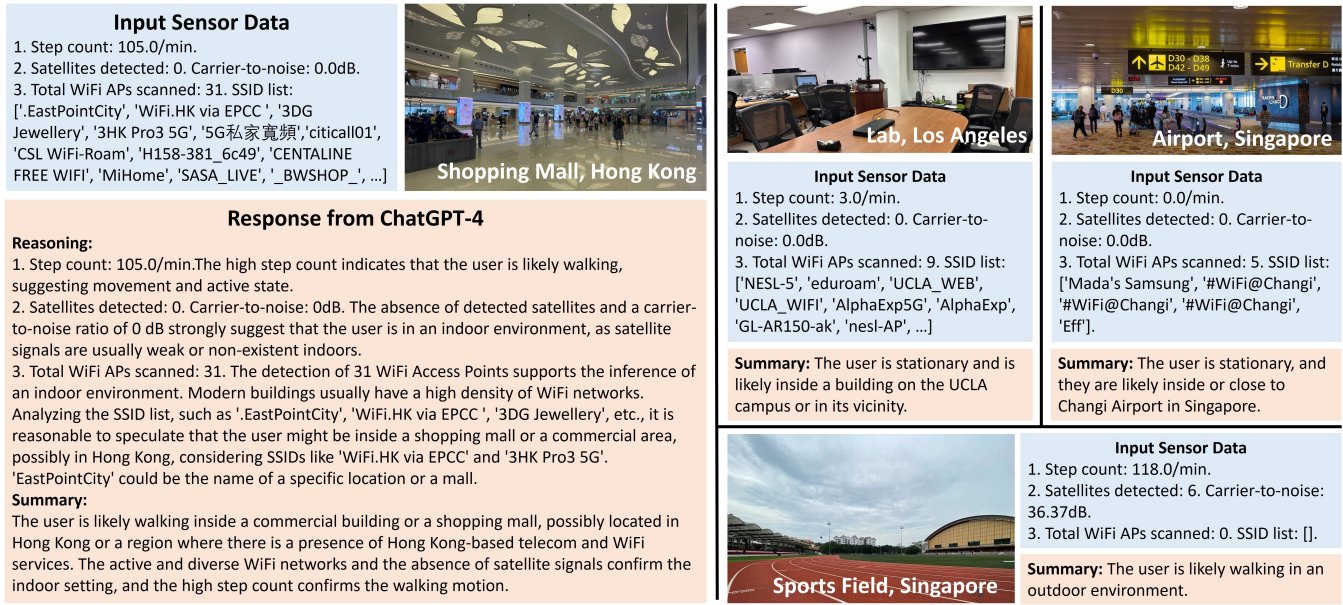


Figure 3: Response examples of ChatGPT-4 for activity sensingⁱⁱ.

the prompt serves as a means to educate and instruct ChatGPT to interpret sensor data and formulate its answers into a concise format. We thereafter present the prompt with succinct textualized sensor data of novel queries to ChatGPT as shown in Figure 2, which we expect to generate the inference results as a concise description of the user’s activity. Note that the prompt, once completed, is frozen and we simply supply new textualized sensor data for new inferences without altering the prompt any further.

2.2 Preliminary Experiment Results

We conducted preliminary experiments in various scenarios – on university campuses, commercial buildings, subway stations, outdoor spaces, and across cities. All sensor data were collected using a Samsung Galaxy S8 Android smartphone. Accelerometer data was sampled at 100 Hz, while the satellite and WiFi data were sampled at 0.2 Hz. To perform our analysis, we utilized sensor data gathered from time windows spanning durations of 10 to 45 seconds and selected the most recent satellite and WiFi scanning results. The evaluation was carried out using both ChatGPT- 3.5 and ChatGPT-4 [22], accessible through the OpenAI API [20] and default parameter settings.

Figure 3 provides several example answers together with their ground-truth contexts. Due to space limits, we have omitted the detailed reasoning of ChatGPT responses except for the first example. The results highlight ChatGPT-4’s capability to reason the user’s surrounding context with its encapsulated knowledge as a foundation, which cannot be achieved by traditional sensing models.

To quantitatively assess the efficacy of this approach, we tasked ChatGPT to explicitly provide the states of motion (between "stationary" and "motion") and environment (between "indoors" and

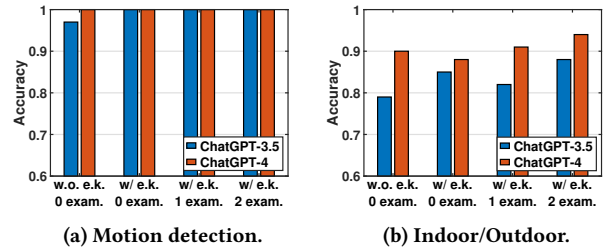


Figure 4: Recognition accuracy of ChatGPT in activity sensing. The ‘e.k.’ denotes expert knowledge, and ‘exam.’ refers to the reasoning examples provided.

"outdoors"). We experiment with varied settings – with/without expert knowledge in the prompt, as well as with different numbers of reasoning examples.

Figure 4(a) summarizes the accuracy for motion detection, which suggests two models perform reasonably well. ChatGPT-3.5 occasionally outputs ‘unknown’ states leading to 97% accuracy even under the output constraints, which can be improved to 100% when detailed expert knowledge and reasoning examples are provided.

Figure 4(b) summarizes the accuracy for environment classification, which depends on multimodal sensor data fusion and therefore appears more challenging. Nevertheless, improved performance is observed when expert knowledge and more reasoning examples are used in the prompt. ChatGPT-4 achieves above 90% accuracy with the best prompt template.

Overall, the above experiment results suggest LLMs are highly effective in analyzing physical world signals when they are properly abstracted into textual representations. These findings align with our initial expectations with the knowledge basis of LLMs.

ⁱⁱDue to the space limit, check more examples and the complete prompt at https://dapowan.github.io/wands_penetrative-ai/

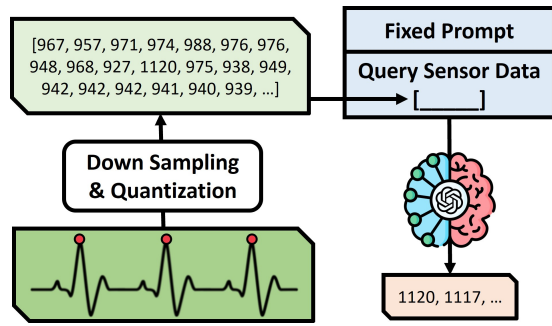


Figure 5: Overview of heartbeat detection with LLM.

3 PENETRATIVE LLM WITH DIGITIZED SIGNALS

This section describes our effort to go beyond the general expectations of the textualized signal processing ability of LLMs. We specifically study the potential of ChatGPT in comprehending digitized sensor signals.

3.1 An Illustrative Example

We take human heartbeat detection as an illustrative example, where we task ChatGPT with the input of ECG waveforms to derive the heartbeat rate. An interesting and challenging job in this application is that we incorporate digitized signals directly into the prompts, delegating the signal processing task to ChatGPT. Figure 5 provides an overview of the design.

Objective and rationale. The objective for ChatGPT is to analyze ECG signals and identify the "R-peaks" [32], which are tall upward deflections and correspond to the red dots in Figure 5. The objective part of the prompt succinctly states: "Find the R-peaks in an ECG waveform".

Data preparation. The sensor data consist of a numerical sequence representing an ECG waveform. The original ECG data are collected at a high sampling rate, e.g., 360Hz. In our design, raw readings are down-sampled to 36 Hz and quantized to their integer parts to reduce sequence length and number complexity.

Expert knowledge. We first provide interpretations regarding R-peaks: "An R-peak within a sequence of ECG numbers refers to a pronounced upward deflection, typically representing the largest and most conspicuous values within the sequence". We then design a natural language-based procedure that LLMs understand to guide the selection of R-peaks. Three steps are included: i) assessing the overall range of ECG numbers, ii) identifying subsequences characterized by an initial lower value, a subsequent significant increase, and a return to the overall range, and iii) selecting the highest value from each such subsequence as the R-peak. We investigate whether ChatGPT can effectively execute such fuzzy logic (without explicit threshold values) when processing the digitized signals.

Reasoning examples. We also furnish ChatGPT with illustrative examples based on the provided procedure, encompassing digitized ECG data, a reasoning process, and a summary of R-peak numbers.

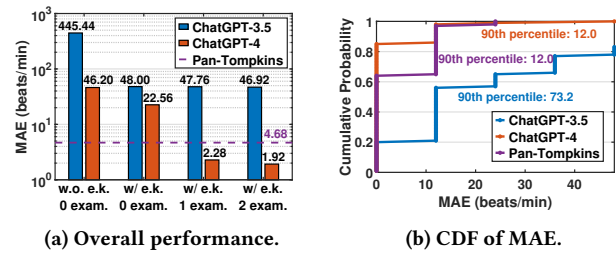


Figure 6: MAE of ChatGPT in heartbeat detection. (a) The 'e.k.' denotes expert knowledge, and 'exam.' refers to the reasoning examples provided.

3.2 Preliminary Experiment Results

We conducted preliminary experiments with the MIT-BIH Arrhythmia Database [12], which is an ECG dataset equipped with ground truth annotations for R-peaks. We downsampled the raw ECG signal to 36 Hz and each input ECG data are from a 5-second window comprising 180 numbers. The evaluation was again carried out using both ChatGPT-3.5 and ChatGPT-4 with default parameters. For comparison, we also test the performance of Pan-Tompkins [23], a classical signal processing approach with the same setting.

We use the Mean Absolute Error (MAE) to measure the deviation in beats per minute between the predicted and actual R-peaks. In Figure 6(a), we present the results of Pan-Tompkins and the two models with different prompts. ChatGPT-4 consistently yields lower errors than ChatGPT-3.5, achieving an MAE of 1.92 with the best setting, i.e., when expert knowledge is provided with two reasoning examplesⁱⁱⁱ. In Figure 6(b) we visualize the detailed Cumulative Distribution Function (CDF) of the MAE achieved by ChatGPT-3.5 and ChatGPT-4.0 (when expert knowledge and one reasoning example is provided). We also provide the MAE CDF of Pan-Tompkins as a reference. While ChatGPT-3.5 often generates prolonged sequences of R-peaks, resulting in substantial errors, ChatGPT-4 conversely demonstrates enhanced stability and precision in identifying R-peaks in the majority of cases and outperforms the Pan-Tompkins approach.

In conclusion, our initial findings indicate that LLMs, particularly ChatGPT-4, exhibit remarkable proficiency in analyzing physical digitized signals when provided with proper guidance.

4 PENETRATIVE AI

While not achieving perfect accuracy, LLMs exhibit surprisingly encouraging performance, even when dealing with pure digital signals acquired from the physical world. This presents an enticing opportunity to leverage LLMs' world knowledge as a foundation model to derive insights from sensory information while requiring no or little additional task knowledge or data, i.e., in zero or few-shot settings. Such a capability may be equipped with IoT sensors and actuators to build intelligence into cyber-physical systems – a concept we term "Penetrative AI".

ⁱⁱⁱNote however that Pan-Tompkins algorithm can achieve improved performance when higher sampling rates and longer numerical sequences are supplied, e.g., using a sequence of 30-second windowed data at 72 Hz in Pan-Tompkins algorithm may give an MAE of 1.06 which is lower than that of ChatGPT-4.

"Penetrative AI" is concerned with exploring the foundation role of LLMs in completing tasks in the physical world. Two primary characteristics define its scope – i) the involvement of the embedded world knowledge in LLMs^{iv}, and ii) the integration with IoT sensors and/or actuators for perceiving and intervening the physical world. It is important to distinguish the scope from existing practices which shall not be considered as "Penetrative AI", including:

- *NLP applications of LLMs.* Examples include language translation or text generation, where LLMs do not process sensor data linked to the physical world.
- *Conventional machine learning adopted in CPS or IoT.* Examples include DNN models trained with labeled or unlabeled sensor data which rely on summarizing patterns instead of gaining power from world knowledge of LLMs, e.g., developing medical diagnosis systems with medical imaging data.
- *LLMs involved in the CPS loop but not applied to comprehending the physical world phenomena.* A typical example is LLM-enabled automation for inquiring weather data where the LLM serves as a language interface for data queries but not a foundation role in comprehending such sensor data. Similarly in advanced driver assistance systems, LLMs may be adopted as human-computer interfaces to manage and convey ambient road conditions obtained from various on-vehicle sensors or derived by other machine learning modules. As long as the LLMs are not engaged with their world knowledge in direct analysis of sensor inputs or CPS control, it is not considered a practice of penetrative AI.

"Penetrative AI" is different from "Embodied AI" [9]. Though also emphasizes the interaction with the physical environment, "Embodied AI" predominantly aims at designing robotic agents and is broadly defined with general AI models (rather than the penetrative AI's focus on LLMs' foundation roles). The penetrative AI focuses on the exploration of integrating LLMs with IoT sensing. It is not limited to the form of AI agents and supports AGI-in-the-loop perception or control modules for CPS.

As the example applications demonstrate, penetrative AI may offer the following potentials. It simplifies solution deployment, allowing user-machine interaction in plain language and minimizing the need for extensive programming skills. It also enhances data efficiency as LLMs embedded with vast world knowledge can effectively generalize to new tasks. LLMs adeptly handle fuzzy logic well, drawing inferences from vague or disorganized information, and bypassing the need for precise logic. Finally, the penetrative AI offers an innovative opportunity for multimodal fusion, where diverse data types are transformed into a uniform text format, facilitating seamless adaptation to various tasks without extensive model re-engineering.

5 CHALLENGES AND FUTURE DIRECTIONS

Adopting LLMs in a penetrative way for CPS is non-trivial since LLMs are typically trained with extensive text corpora for NLP applications and thus may lack high expertise and domain knowledge

for CPS tasks. Unleashing its full potential necessitates addressing the challenges contained in the following levels:

Understanding the knowledge boundaries of LLMs. A fundamental challenge lies in systematically assessing LLMs' capabilities for specific CPS contexts. A pragmatic approach to this is engaging LLMs in structured dialogues, tailored to uncover their understanding and application of relevant concepts at different levels, including *conceptual awareness* where the LLMs' fundamental conceptual grasp can be gauged by questions like "what is SSID in the context of WiFi?" or "how the RSSI varies with the distance between a pair of WiFi AP and client?", and *application and understanding* which delves deeper, examining whether LLMs can aptly apply fundamental concepts in practical scenarios with example questions like "what does it imply about the users' locations if their smartphones connect to WiFi APs with certain SSIDs and RSSIs?".

Expanding LLMs' capabilities. A subsequent and essential challenge is to broaden the capabilities of LLMs for CPS tasks based on the existing knowledge. Such expansion can be approached through several strategies. *Task decomposition* can break down complex tasks into simpler sub-tasks, which allows LLMs to develop more focused and efficient problem-solvers. *Signal transformation and data preprocessing* decides the form in which sensor or actuator data shall be presented which is a crucial challenge. While digitized signals offer in-depth information, they require a deeper level of physical world understanding from LLMs. Transforming them into textualized data may be beneficial and other preprocessing methods such as filtering to remove irrelevant or redundant information may also enhance system efficacy. *Effective prompt design* is a major challenge, which may involve embedding domain-specific knowledge when LLMs' common knowledge is limited in certain tasks. Developing stateful prompts and effective algorithms with fuzzy logic (as demonstrated in Section 3.2) is another interesting future work. *Interfacing with external tools* also leads to an expansion of LLMs' capabilities. Examples include using code interpreters for executing signal processing algorithms or leveraging procedure calls for accessing real-time information and/or controlling CPS.

Enriching LLMs with expert knowledge. A pivotal approach is to develop specialized models tailored to embedding additional domain knowledge for CPS tasks. Such an approach however comes with special considerations and challenges: *Dataset construction* for multimodal datasets to train tailored LLMs is a challenge. Unlike standard image-text pair datasets like those described in [3], sensor-text datasets for CPS tasks shall include not only descriptive information but also expert knowledge and processing guidance, which necessitates a thoughtful approach to ensure the data are comprehensive, accurate, and reflective of real-world scenarios. *Balancing specialization with generalizability* is necessary. A critical risk in the fine-tuning LLMs is the potential disruption of the existing knowledge base of LLMs and a balanced fine-tuning process with both general and domain-specific data may be key to maintaining the robustness of LLMs. *Integrating expert models* presents another way to enrich expert knowledge of LLMs, e.g., integrating LLMs with an IMU foundation model like LIMU-BERT [31] may enable frontend features of human activities before LLM comprehension for detailed behavior analysis. The distinct nature of sensor data compared with textual data shall be considered which necessitates the development of modality alignment techniques like [11].

^{iv} or variations like Vision-Language Models (VLMs) [21] which adapt to other input modalities.

6 RELATED WORK

The significant strides in natural language processing show that large language models (LLMs) exhibit out-of-the-box capabilities [2, 26, 34]. Beyond traditional NLP tasks, recent works have adopted LLMs for a wide range of applications such as image editing [30], video understanding [15], sequence completion [19], knowledge graph construction [4, 28], and recommendation systems [10, 17]. These applications, however, predominantly operate within the digital realm, with limited engagement in the physical world.

In domains like gaming and robotics, LLM-based agents have been utilized to generate actions or plans, leveraging their inherent general knowledge [16, 24, 25, 27, 29, 33]. Some efforts like [16, 27] focus on the programming abilities of LLMs and interface indirectly with the physical world through predefined APIs. While these approaches are practical for everyday tasks requiring general knowledge, such as rearranging objects on the table, their applicability to CPS tasks that necessitate expert knowledge and sophisticated signal understanding is limited.

Some initiatives echo the nascent endeavor of penetrative AI. For example, Rt-2 [1] introduces a vision-language-action model for robotic control but requires extra development to suit CPS tasks, especially when sensor-text datasets are not available, a challenge we discuss in Section 5. Liu et al. leverage LLMs to analyze medical data for health-related tasks [18] but their model primarily learns from question-answer pairs in prompts. In contrast to existing efforts, LLMs in the penetrative AI directly engage in processing IoT sensor signals at various levels with world knowledge. We believe this is the first effort to explore the boundaries of LLMs' ability to interact with the physical world for cyber-physical systems. LightLLM [13] is an ongoing practice which integrates LLMs for traffic signal control tasks.

7 CONCLUSION

We present "Penetrative AI", a new concept concerning the exploration of leveraging large language models' common knowledge as a foundation to accomplish real-world perceiving and intervening tasks in cyber-physical systems. Our findings illuminate a promising path for the integration of LLMs and CPS, offering insights into the future of AGI-in-the-loop solutions.

ACKNOWLEDGMENTS

We thank all reviewers for their insightful comments and suggestions, which have greatly improved the quality of this paper. This work is supported by Global STEM Professorship Scheme of Hong Kong and HKUST start up grant. It is also supported by Singapore NRF Investigatorship NRFI08-2022-0010.

REFERENCES

- [1] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, et al. 2023. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818* (2023).
- [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).
- [3] Minwoo Byeon, Beomhee Park, Haechon Kim, Sungjun Lee, Woonhyuk Baek, and Saehoon Kim. 2022. COYO-700M: Image-Text Pair Dataset. <https://github.com/kakaobrain/coyo-dataset>.

- [4] Salvatore Carta, Alessandro Giuliani, Leonardo Piano, Alessandro Sebastian Podda, Livio Pompianu, and Sandro Gabriele Tiddia. 2023. Iterative Zero-Shot LLM Prompting for Knowledge Graph Construction. *arXiv preprint arXiv:2307.01128* (2023).
- [5] Antonia Creswell, Murray Shanahan, and Irina Higgins. 2022. Selection-inference: Exploiting large language models for interpretable logical reasoning. *arXiv preprint arXiv:2205.09712* (2022).
- [6] Android Developers. 2023. GnsStatus. <https://developer.android.com/reference/android/location/GnsStatus>
- [7] Android Developers. 2023. Motion Sensors. https://developer.android.com/develop/sensors-and-location/sensors/sensors_motion?hl=en#sensors-motion-stepcounter
- [8] Android Developers. 2023. ScanResult. <https://developer.android.com/reference/android/net/wifi/ScanResult>
- [9] Jiafei Duan, Samson Yu, Hui Li Tan, Hongyuan Zhu, and Cheston Tan. 2022. A survey of embodied ai: From simulators to research tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence* 6, 2 (2022), 230–244.
- [10] Yunfan Gao, Tao Sheng, Youlin Xiang, Yun Xiong, Haofen Wang, and Jiawei Zhang. 2023. Chat-rec: Towards interactive and explainable llms-augmented recommender system. *arXiv preprint arXiv:2303.14524* (2023).
- [11] Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. 2023. Imagebind: One embedding space to bind them all. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15180–15190.
- [12] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation* 101, 23 (2000), e215–e220.
- [13] Siqi Lai, Zhao Xu, Weijia Zhang, Hao Liu, and Hui Xiong. 2023. Large Language Models as Traffic Signal Control Agents: Capacity and Opportunity. *arXiv preprint arXiv:2312.16044* (2023).
- [14] Yann LeCun. 2022. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review* 62 (2022).
- [15] KunChang Li, Yanan He, Yi Wang, Yizhuo Li, Wenhai Wang, Ping Luo, Yali Wang, Limin Wang, and Yu Qiao. 2023. Videochat: Chat-centric video understanding. *arXiv preprint arXiv:2305.06355* (2023).
- [16] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. 2023. Code as policies: Language model programs for embodied control. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 9493–9500.
- [17] Junling Liu, Chao Liu, Renjie Lv, Kang Zhou, and Yan Zhang. 2023. Is chatgpt a good recommender? a preliminary study. *arXiv preprint arXiv:2304.10149* (2023).
- [18] Xin Liu, Daniel McDuff, Geza Kovacs, Isaac Galatzer-Levy, Jacob Sunshine, Jieming Zhan, Ming-Zher Poh, Shun Liao, Paolo Di Achille, and Shwetak Patel. 2023. Large Language Models are Few-Shot Health Learners. *arXiv preprint arXiv:2305.15525* (2023).
- [19] Suvir Mirchandani, Fei Xia, Pete Florence, Brian Ichter, Danny Driess, Montserrat Gonzalez Arenas, Kanishka Rao, Dorsa Sadigh, and Andy Zeng. 2023. Large Language Models as General Pattern Machines. In *Proceedings of the 7th Conference on Robot Learning (CoRL)*.
- [20] OpenAI. 2023. GPT. <https://platform.openai.com/docs/guides/gpt>
- [21] OpenAI. 2023. GPT-4 System Card. <https://openai.com/research/gpt-4v-system-card>
- [22] OpenAI. 2023. GPT-4 Technical Report. *arXiv:2303.08774 [cs.CL]*
- [23] Jiapu Pan and Willis J Tompkins. 1985. A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering* 3 (1985), 230–236.
- [24] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442* (2023).
- [25] Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, et al. 2023. Toolllm: Facilitating large language models to master 16000+ real-world apis. *arXiv preprint arXiv:2307.16789* (2023).
- [26] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. 2022. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100* (2022).
- [27] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. 2023. Progprompt: Generating situated robot task plans using large language models. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 11523–11530.
- [28] Jiashuo Sun, Chengjin Xu, Luminyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Heung-Yeung Shum, and Jian Guo. 2023. Think-on-graph: Deep and responsible reasoning of large language model with knowledge graph. *arXiv preprint arXiv:2307.07697* (2023).
- [29] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2023. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291* (2023).

- [30] Chenfei Wu, Shengming Yin, Weizhen Qi, Xiaodong Wang, Zecheng Tang, and Nan Duan. 2023. Visual chatgpt: Talking, drawing and editing with visual foundation models. *arXiv preprint arXiv:2303.04671* (2023).
- [31] Huatao Xu, Pengfei Zhou, Rui Tan, Mo Li, and Guobin Shen. 2021. Limu-bert: Unleashing the potential of unlabeled data for imu sensing applications. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 220–233.
- [32] Frank G Yanowitz. 2010. Lesson III. Characteristics of the Normal ECG. *University of Utah School of Medicine* (2010).
- [33] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629* (2022).
- [34] Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, et al. 2022. Glm-130b: An open bilingual pre-trained model. *arXiv preprint arXiv:2210.02414* (2022).