Early Prediction of Cybersickness in Virtual Reality Using a Large Language Model for Multimodal Time Series Data

Yoonseon Choi
Hanyang University
Department of Data Science
Seoul, Republic of Korea
yoonseonchoi@hanyang.ac.kr

Dayoung Jeong
Hanyang University
Department of Artificial Intelligence
Seoul, Republic of Korea
dayoungjeong@hanyang.ac.kr

Bogoan Kim
Hanyang University
Department of Data Science
Seoul, Republic of Korea
bogoankim@hanyang.ac.kr

Kyungsik Han*
Hanyang University
Department of Data Science
Seoul, Republic of Korea
kyungsikhan@hanyang.ac.kr

ABSTRACT

Cybersickness in virtual reality (VR) significantly disrupts user immersion. Although recent studies have proposed cybersickness prediction models, existing models have considered the moment of cybersickness onset, limiting their applicability in proactive detection. To address this limitation, we used long-term time series forecasting (LTSF) models based on multimodal sensor data collected from the head-mounted display (HMD). We used a pre-trained large language model (LLM) to effectively learn the salient features (e.g., seasonality) of multimodal sensor data by understanding the nuanced context within the data. The results of our experiment demonstrated that our model achieved comparable performance to the baseline models, with an MAE of 0.971 and an RMSE of 1.696. This indicates the potential for early prediction of cybersickness by employing LLM- and LTSF-based models with multimodal sensor data, suggesting a new direction in model development.

CCS CONCEPTS

• Human-centered computing → Virtual reality; **• Computing methodologies → Modeling methodologies**.

KEYWORDS

Cybersickness; Time-LLM; Early prediction; Multimodal sensor data

ACM Reference Format:
Yoonseon Choi, Dayoung Jeong, Bogoan Kim, and Kyungsik Han. 2024. Early Prediction of Cybersickness in Virtual Reality Using a Large Language Model for Multimodal Time Series Data. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing Pervasive and Ubiquitous Computing (UbiComp Companion '24), October 5–9, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3675094.3677578

*Corresponding author

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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).
UbiComp Companion '24, October 5–9, 2024, Melbourne, VIC, Australia
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1058-2/24/10
https://doi.org/10.1145/3675094.3677578

1 INTRODUCTION

As VR technology becomes more accessible, its applications are gradually expanding in various fields (e.g., education, healthcare, entertainment) [13, 14, 19]. However, the occurrence of cybersickness in VR (VR sickness), characterized by symptoms such as eye strain, nausea, and dizziness, poses a significant challenge as these symptoms disrupt user immersion and undermine the usability of VR. To measure cybersickness, research has commonly used the Simulator Sickness Questionnaire (SSQ) [11] or the Fast Motion Sickness Scale (FMS) [12]. Recently, there has been a growing preference [25] for the FMS over the SSQ due to concerns about the consistency of SSQ results (e.g., varying standards of interpretation). This trend reflects a shift towards seeking more reliable and standardized measures in VR research.

Recent studies on cybersickness have proposed prediction models using electroencephalography (EEG), video, and head-mounted display (HMD) data to enhance the VR user experience [7, 10]. In particular, EEG data is effective in capturing users' cognitive activities, but the invasive nature of additional measurement devices beyond the HMD can lead to user discomfort and a diminished VR experience. As a result, data collected in somewhat intrusive environments may lack the reliability and generalizability compared to the model with data collected in environments more comfortable for the users [23]. Therefore, the use of HMD data, which is less intrusive, can be effective for cybersickness prediction modeling.

Most cybersickness prediction models use point-wise sensor data at the current moment to predict cybersickness at a single point in time, and have demonstrated high performance [3–6, 16]. Recently, the importance of not only developing cybersickness prediction models but also applying them to VR content has been highlighted [10]. However, when models developed based on point-wise data are integrated into VR content, mitigation measures can only be taken after the user experiences cybersickness. To ensure a more natural VR user experience, it is important to take mitigation measures before the user feels cybersickness; therefore, the early prediction of cybersickness is important.

Long-term time series forecasting (LTSF) models have been found to be effective in predicting future data by learning the characteristics of time series data [27]. To further enhance the

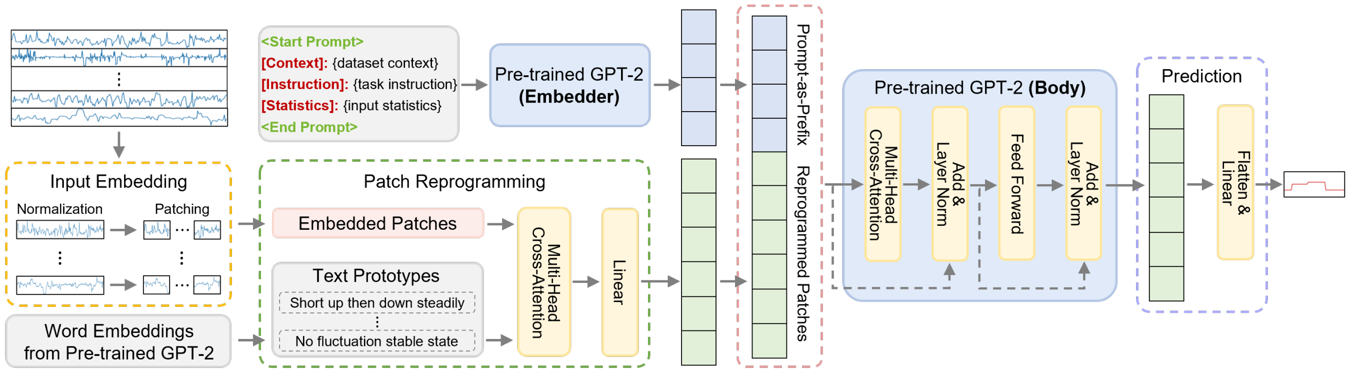

Figure 1: Time-LLM architecture based on multimodal sensor data.

prediction performance, there has been active research on time series forecasting that employed pre-trained large language models (LLMs) [2, 8, 9, 30]. LLMs are effective in predicting time series data because they can identify hidden trends and seasonality based on their pre-trained knowledge of the domain and the characteristics of time series data through a significant amount of data [8]. However, with multimodal sensor data measured over time, the high dimensionality, non-linear relationships, and continuous nature of the data pose challenges for embedding it into LLMs [15]. It is necessary to learn not only individual data points, but also dynamic patterns over time [15]. Because of these challenges, research on modeling multimodal sensor data using LLMs remains insufficient.

This paper proposes a model for the early prediction of cybersickness using Time-LLM [9] based on multimodal sensor data collected from an HMD. Time-LLM is an LTSF model that uses the pre-trained LLM as its backbone. Our model processes multivariate data highly correlated with cybersickness (e.g., eye and head movements) as univariate data and embeds a prompt with descriptions of the input data along with the multimodal sensor data to enable effective learning. Experimental results using HMD data from 45 participants show that our model achieves comparable performance to baseline LTSF models (MAE=0.971, RMSE=1.696). These results indicate that the use of the LLM with multimodal sensor data can help capture data flows that are difficult to identify using LTSF models alone. By incorporating comprehensive descriptions of the input data into the prompt, it is possible to achieve more accurate learning of the characteristics of multimodal sensor data. Furthermore, this research demonstrates that cybersickness can be predicted using only HMD data and highlights the significance of early prediction in the cybersickness domain by using LLMs and LTSF models.

2 METHODOLOGY

We adopted Time-LLM [9] as the model for early prediction of cybersickness. Figure 1 represents the architecture of Time-LLM, based on multimodal sensor data. Time-LLM is a generalized model designed across various domains (e.g., traffic, finance), and the pre-trained LLM used in Time-LLM is applied without domain-specific fine-tuning. To efficiently train on the multimodal sensor data with numerous features, we used pre-trained GPT-2[1] [20] as the backbone, given its relatively lightweight nature as an LLM.

[1]We used pre-trained GPT-2 due to our computational limitation.

Time-LLM, functioning as an LTSF model, sequentially takes input data of length L $(x_1, \ldots, x_L)$ and predicts future data of length T $(x_{L+1}, \ldots, x_{L+T})$. The model consists of four key components: (1) input embedding, (2) patch reprogramming, (3) prompt-as-prefix, and (4) prediction.

2.1 Input Embedding

To enhance the memory efficiency during training Time-LLM, we performed instance normalization and patching during the input embedding process.

2.1.1 Instance normalization. To preserve the characteristics of each data feature (channel) independently without mixing channels, we decomposed it into univariate series and performed instance normalization for each channel separately, rather than normalizing the multivariate data all at once.

2.1.2 Patching. Capturing patterns that appear over time by preserving the locality of the time-series data is crucial for improving prediction performance. Therefore, we implemented a patching process to handle the multimodal sensor data in a series-wise manner instead of a point-wise approach.

$$N = \left\lfloor \frac{L - P}{S} \right\rfloor + 2 \tag{1}$$

where N is the number of patches, P is the length of a patch, and S is the length of a slide. N represents the number of input tokens. The use of patches reduces the number of input tokens compared to point-wise processing, lowering the spatiotemporal complexity.

2.2 Patch Reprogramming

To train the model with multimodal sensor data and natural language together, an embedding process that aligns the two modalities (eye and head movements) is essential. Through the patch reprogramming process, we retained text prototypes related to the characteristics of the time-series data (e.g., up, stable, periodic) by linearly probing word embeddings from the pre-trained GPT-2. These text prototypes learn connecting language cues (e.g., short up, stable state) to represent specific patches' information and combine these cues (e.g., short up then stable state). Using patches and text prototypes, the time-series information of patches is learned as vectors through multi-head cross-attention. Then, by linearly projecting these vectors, we can obtain the reprogrammed patches, which are referred to embedded multimodal sensor data.

<Start Prompt>
[Dataset context]: The "Sickness" dataset consists of sensor data on eye and head movements per second. This dataset …
[Task instruction]: Forecast the next **{T}** steps given the previous **{L}** steps information.
[Input statistics]: The input has a minimum of **{minimum}**, a maximum of **{maximum}**, and a median of **{median}**. The overall trend is **{upward or downward}**. The top five lags are **{lags}**.
<End Prompt>

Figure 2: Example of a prompt including key information of the input data.

2.3 Prompt-as-Prefix

Directly transforming multimodal sensor data, consisting of numerical sequences, into natural language poses a challenge in generating prompts without degrading the performance of LLMs [28]. Therefore, instead of directly embedding only the multimodal sensor data into the LLM (GPT-2), we embed it together with prompts (Figure 2) that describe key information about the input data. To address this, prompts are vectorized by tokenizing them through the pre-trained GPT-2 embedder. Then, the embedded prompt is prepended to the reprogrammed patches. This embedded prompt is referred to as a prompt-as-prefix.

The prompt includes the dataset context, task instruction, and input statistics. The dataset context provides background information to help understand the cybersickness dataset. The task instruction provides a specific guideline for the model's training task by specifying the length of the input data (L) and the output prediction data (T). The input statistics include the minimum, maximum, and median values, which are basic statistical information for time series data. To capture the characteristics of the multimodal sensor data, the input statistics also include trends (e.g., up/downward) and the top-5 lags.

2.4 Prediction

The packed vector resulting from concatenating the prompt-as-prefix and reprogrammed patches is embedded into the pre-trained GPT-2, and passed through a Vanilla Transformer encoder. The prompt-as-prefix is removed from the packed vector, and the reprogrammed patches are obtained as output representations. Through flattening and linear projection processes, the model predicts the cybersickness (FMS) with a length of T.

3 EXPERIMENT

3.1 Data Setup

We used an MSCVR dataset [5] obtained from a user study involving 45 users who watched 20 cybersickness-inducing videos, each 45 seconds long, using the HTC VIVE Pro Eye HMD. This dataset includes responses from each participant, who rated their level of cybersickness every 15 seconds using the FMS, allowing for a quantitative assessment of their cybersickness levels. The FMS score ranges from 0 to 20, with 0 indicating "no sickness" and 20 indicating "severe sickness." Additionally, we collected sensor data with two modalities via only the HMD. The sensor data comprises 23 signals related to eye movements (e.g., gaze direction of both eyes, pupil position) and 6 signals related to head movements (e.g., head position and rotation), measured at a frequency of 90Hz.

We preprocessed the MSCVR dataset to fit the LTSF model structure. We downsampled the data from 90Hz to 1Hz using the mean value for each second [4]. While traditional LTSF models are regression models that predict continuous values, the FMS is interval scale data with values ranging from 0 to 20. Thus, each FMS value measured every 15 seconds was flattened across 15 data points. Additionally, we split the MSCVR dataset into three parts for model training: the first 70% of each 45-second video as the training dataset, the next 10% as the test dataset, and the remaining 20% as the validation dataset.

3.2 Model Setup

During model training, we ensured that data from different videos were not mixed when input to the model. Hyperparameters necessary for training were set according to the data length of 45 seconds per video. The length of the input data L was set to 25, the length of the prediction data T was set to 20, and the batch size was set to 45 to match the number of data points per video. In the input embedding process, the patch length P was set to 16, and the slide length S was set to 8. Among other hyperparameters necessary for training, the learning rate, was set to 0.0001, the optimizer was Adam, and the loss function was MSE to optimize the model.

3.3 Baseline Models

To compare the performance with existing well-known LTSF models, we selected DLinear [29], Vanilla Transformer [26], and PatchTST [18] as baseline models. Vanilla Transformer was chosen because Time-LLM uses the Vanilla Transformer encoder structure when prompting the pre-trained GPT-2 with a pair of the embedded prompt and reprogrammed patches. PatchTST is an appropriate model to validate the effectiveness of using LLM on multimodal sensor data because the input embedding component of Time-LLM, which uses instance normalization and patching method, is based on PatchTST. For comparison with Transformer-based LTSF models, we selected DLinear, a one-layer linear model. Furthermore, to evaluate the effectiveness of the patch reprogramming and prompt-as-prefix, we conducted experiments by removing each module of Time-LLM.

Table 1: Results of LTSF models with the best result in bold and the second best underlined. "Time-LLM" represents the original model, while "Time-LLM-NR" denotes the version with only the prompt-as-prefix, and "Time-LLM-NP" denotes the version with only the reprogrammed patches.

Models	MAE	RMSE
DLinear	1.328	1.910
Transformer	2.144	3.044
PatchTST	**0.931**	**1.678**
Time-LLM-NR	1.766	2.743
Time-LLM-NP	1.038	1.751
Time-LLM	0.971	1.696

3.4 The Performance by LTSF Models

To evaluate the prediction performance of the regression model, the MAE and RMSE metrics were used. Our results are shown in Table 1. Time-LLM achieved a performance of 0.971 for MAE and 1.696 for RMSE. Compared to Vanilla Transformer, Time-LLM reduced MAE by 55% and RMSE by 45%. Compared to DLinear, Time-LLM reduced MAE by 27% and RMSE by 21%. In addition, compared to Time-LLM, Time-LLM-NR showed a 45% decrease in MAE and a 38% decrease in RMSE, while Time-LLM-NP showed a 7% decrease in MAE and a 3% decrease in RMSE. However, compared to PatchTST, Time-LLM showed a slight decrease in performance, with a 4% increase in MAE and a 1% increase in RMSE.

4 DISCUSSION

4.1 Performance Implications

As shown in Section 3.4, the results suggest the effective application of Time-LLM to multimodal sensor data. Our experimental results with Time-LLM-NR and Time-LLM-NP demonstrated that each module has a significant impact on model performance. Through the Time-LLM-NR experiment, we confirmed that patch reprogramming may help the model understand patterns and trends within data by using pre-trained word embeddings. The results of Time-LLM-NP suggest the importance of well-designed prompt-as-prefix. Multimodal sensor data is collected by participant or content in VR environments. Therefore, in an LTSF model where data is input sequentially, the model must be designed to segment the multimodal sensor data by participant or content for training and prediction. For more accurate prediction, it is essential for the model to understand overall context of the multimodal sensor data by using prompt-as-prefix.

Although Time-LLM for multimodal sensor data achieved comparable performance to the baseline models, Time-LLM showed a slight decrease in performance compared to PatchTST. Performance is likely to be affected by the type of LLM, as shown experimentally by existing LLM-based LTSF models [17]. For example, we have confirmed that using LLaMA-7B [24], a more recent LLM than the GPT-2 we used, for an LTSF model may result in better performance [9]. Prior research has also indicated that PatchTST slightly outperforms the GPT-2-based model in long-term forecasting tasks, among various time series forecasting methods such as short-term or few-shot forecasting [30]. Based on this prior research, it is possible that using a more advanced LLM than GPT-2 could improve performance in the early prediction of cybersickness.

Nevertheless, the use of pre-trained LLMs can be effective for cybersickness prediction modeling. Many studies have used the modality-specific method to minimize information loss about the characteristics of each modality that are highly correlated with cybersickness when training models [3–5]. Since cybersickness occurs from the sensory conflict between various body signals [1, 21, 22], embedding detailed descriptions of the characteristics of each body signal and the correlations between them into a Time-LLM can effectively capture features in multimodal sensor data. This approach, through the use of pre-trained LLMs, enables integrated analysis and training of modalities, which can help simplify the training process for each modality, and the performance evaluation results of Time-LLM are encouraging.

4.2 Limitation and Future Work

Although the use of LLMs for early prediction of cybersickness has shown effectiveness, we acknowledge two important limitations.

First, through the experiments, we confirmed that the training times of Time-LLM were longer compared to the existing LTSF models. Training LLMs with multimodal sensor data can lead to increased training time due to the excessive number of parameters. The primary goal of early prediction of cybersickness is to enhance the VR user experience by proactively mitigating cybersickness based on model predictions. Thus, improving the training efficiency of the model is important for quickly predicting future cybersickness occurrences. To address this, future research will focus on optimizing patch reprogramming. Specifically, improving the linearly probing method used in the patch reprogramming component to better extract words related to time series data will enable more selective identification of relevant terms from multimodal sensor data. This refinement is expected to reduce the number of parameters from pre-trained word embeddings.

Second, while the performance of the cybersickness prediction model by using Time-LLM was comparable to the baseline model, there is still room for improvement. To enhance the prediction performance of Time-LLM, it is important to develop a prompt that can effectively explain the multimodal sensor data. The current prompt provides a general description of the input data. To improve the existing prompt, we propose to perform feature extraction to identify the variables most closely associated with cybersickness within the modalities of eye and head movements. By embedding detailed descriptions of these features into the prompt, we aim to enhance the interpretability of the modalities and, consequently, improve the performance of the model.

5 CONCLUSION

In this paper, we proposed a Time-LLM-based model for early prediction of cybersickness using multimodal sensor data. We demonstrated that cybersickness can be predicted using only the multimodal sensor data collected from HMDs, and that it is possible to predict cybersickness early using an LLM-based LTSF model. In particular, our experiments confirmed the effectiveness of patch reprogramming and using prompt-as-prefix to learn from multimodal sensor data. This study suggests that Time-LLM can be effectively used for cybersickness prediction. We hope to provide valuable insights into the field of cybersickness prediction modeling.

ACKNOWLEDGMENTS

This research was supported by the National Research Foundation of Korea (NRF) (2021M3A9E4080780, RS-2024-00427828).

REFERENCES

[1] Willem Bles, Jelte E Bos, Bernd De Graaf, Eric Groen, and Alexander H Wertheim. 1998. Motion sickness: only one provocative conflict? *Brain research bulletin* 47, 5 (1998), 481–487.
[2] Ching Chang, Wei-Yao Wang, Wen-Chih Peng, and Tien-Fu Chen. 2024. Llm4ts: Aligning pre-trained llms as data-efficient time-series forecasters. (2024).
[3] Rifatul Islam, Kevin Desai, and John Quarles. 2021. Cybersickness prediction from integrated hmd's sensors: A multimodal deep fusion approach using eye-tracking and head-tracking data. In *2021 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE, 31–40.
[4] Rifatul Islam, Kevin Desai, and John Quarles. 2022. Towards forecasting the onset of cybersickness by fusing physiological, head-tracking and eye-tracking with

multimodal deep fusion network. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 121–130.

[5] Dayoung Jeong and Kyungsik Han. 2024. PRECYSE: Predicting Cybersickness using Transformer for Multimodal Time-Series Sensor Data. *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–24.

[6] Dayoung Jeong, Seungwon Paik, YoungTae Noh, and Kyungsik Han. 2023. MAC: multimodal, attention-based cybersickness prediction modeling in virtual reality. *Virtual Reality* 27, 3 (2023), 2315–2330.

[7] Daekyo Jeong, Sangbong Yoo, and Jang Yun. 2019. Cybersickness analysis with eeg using deep learning algorithms. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 827–835.

[8] Mingyu Jin, Hua Tang, Chong Zhang, Qinkai Yu, Chengzhi Liu, Suiyuan Zhu, Yongfeng Zhang, and Mengnan Du. 2024. Time Series Forecasting with LLMs: Understanding and Enhancing Model Capabilities. *arXiv preprint arXiv:2402.10835* (2024).

[9] Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. 2023. Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728* (2023).

[10] Weina Jin, Jianyu Fan, Diane Gromala, and Philippe Pasquier. 2018. Automatic prediction of cybersickness for virtual reality games. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE, 1–9.

[11] Robert Kennedy, Norman Lane, Kevin Berbaum, and Michael Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.

[12] Behrang Keshavarz and Heiko Hecht. 2011. Validating an efficient method to quantify motion sickness. *Human factors* 53, 4 (2011), 415–426.

[13] Bogoan Kim, Dayoung Jeong, Mingon Jeong, Taehyung Noh, Sung-In Kim, Taewan Kim, So-Youn Jang, Hee Jeong Yoo, Jennifer Kim, Hwajung Hong, et al. 2022. Vista: User-centered vr training system for effectively deriving characteristics of people with autism spectrum disorder. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*. 1–12.

[14] Bogoan Kim, Dayoung Jeong, Jennifer G Kim, Hwajung Hong, and Kyungsik Han. 2023. V-DAT (Virtual Reality Data Analysis Tool): Supporting Self-Awareness for Autistic People from Multimodal VR Sensor Data. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[15] Yubin Kim, Xuhai Xu, Daniel McDuff, Cynthia Breazeal, and Hae Won Park. 2024. Health-llm: Large language models for health prediction via wearable sensor data. *arXiv preprint arXiv:2401.06866* (2024).

[16] Ripan Kumar Kundu, Rifatul Islam, John Quarles, and Khaza Anuarul Hoque. 2023. LiteVR: Interpretable and Lightweight Cybersickness Detection using Explainable AI. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 609–619.

[17] Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. 2024. AutoTimes: Autoregressive Time Series Forecasters via Large Language Models. *arXiv preprint arXiv:2402.02370* (2024).

[18] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730* (2022).

[19] Pablo E Paredes, Stephanie Balters, Kyle Qian, Elizabeth L Murnane, Francisco Ordóñez, Wendy Ju, and James A Landay. 2018. Driving with the fishes: Towards calming and mindful virtual reality experiences for the car. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–21.

[20] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[21] James T Reason. 1978. Motion sickness adaptation: a neural mismatch model. *Journal of the royal society of medicine* 71, 11 (1978), 819–829.

[22] James T Reason and Joseph John Brand. 1975. *Motion sickness.* Academic press.

[23] Mahsa Soufineyestani, Dale Dowling, and Arshia Khan. 2020. Electroencephalography (EEG) technology applications and available devices. *Applied Sciences* 10, 21 (2020), 7453.

[24] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[25] Thomas van Gemert, Niels Christian Nilsson, Teresa Hirzle, and Joanna Bergström. 2024. Sicknificant Steps: A Systematic Review and Meta-analysis of VR Sickness in Walking-based Locomotion for Virtual Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–36.

[26] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[27] Neo Wu, Bradley Green, Xue Ben, and Shawn O'Banion. 2020. Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:2001.08317* (2020).

[28] Hao Xue and Flora D Salim. 2022. Prompt-based time series forecasting: A new task and dataset. *arXiv preprint arXiv:2210.08964* (2022).

[29] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2023. Are transformers effective for time series forecasting?. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 11121–11128.

[30] Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. 2023. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems* 36 (2023), 43322–43355.