



# **BRAININFO 2024**

The Ninth International Conference on Neuroscience and Cognitive Brain  
Information

ISBN: 978-1-68558-127-5

March 10th –14th, 2024

Athens, Greece

## **BRAININFO 2024 Editors**

Rory Lewis, University of Colorado, Colorado Springs, USA

Clement Leung, Chinese University of Hong Kong, Shenzhen, China

# BRAININFO 2024

## Forward

The Ninth International Conference on Neuroscience and Cognitive Brain Information (BRAININFO 2024), held between March 10<sup>th</sup> and March 14<sup>th</sup>, 2024, continued a series of international events dedicated to evaluating current achievements and identifying potential ways of making use of the acquired knowledge, covering the neuroscience, brain connectivity, brain intelligence paradigms, cognitive information, and specific applications.

The complexity of the human brain and its cognitive actions stimulated many researches for decades. Most of the findings were adapted in virtual/artificial systems with the idea of brain-like modeling them and using them in human-centered medical cures, especially for neurotechnology. Information representation, retrieval, and internal data connections still constitute a domain where solutions are either missing or in a very early stage.

We take here the opportunity to warmly thank all the members of the BRAININFO 2024 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to BRAININFO 2024. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the BRAININFO 2024 organizing committee for their help in handling the logistics of this event.

We hope that BRAININFO 2024 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of neuroscience and cognitive brain information.

### **BRAININFO 2024 Chairs**

#### **BRAININFO 2024 Steering Committee**

Erwin Lemche, Institute of Psychiatry, Psychology & Neuroscience | King's College School of Medicine and Dentistry, UK

Yu-Dong (Eugene) Zhang, University of Leicester, UK

Chih Lai, University of St. Thomas Minnesota, USA

William Speier, UCLA David Geffen School of Medicine, USA

Sule Yildirim-Yayilgan, Norwegian University of Science and Technology, Norway

Ricardo Ron-Angevin, University of Málaga, Spain

Rory Lewis, University of Colorado at Colorado Springs, USA

#### **BRAININFO 2024 Publicity Chairs**

José Miguel Jiménez, Universitat Politècnica de Valencia, Spain

Sandra Viciano Tudela, Universitat Politècnica de Valencia, Spain

## **BRAININFO 2024 Committee**

### **BRAININFO 2024 Steering Committee**

Erwin Lemche, Institute of Psychiatry, Psychology & Neuroscience | King's College School of Medicine and Dentistry, UK

Yu-Dong (Eugene) Zhang, University of Leicester, UK

Chih Lai, University of St. Thomas Minnesota, USA

William Speier, UCLA David Geffen School of Medicine, USA

Sule Yildirim-Yayilgan, Norwegian University of Science and Technology, Norway

Ricardo Ron-Angevin, University of Málaga, Spain

Rory Lewis, University of Colorado at Colorado Springs, USA

### **BRAININFO 2024 Publicity Chairs**

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

### **BRAININFO 2024 Technical Program Committee**

Aftab Ahmad, John Jay, The City University of New York (CUNY), USA

Mariano Alcañiz Raya, Institute of Research and Innovation in Bioengineering (I3B) - Universidad Politécnica Valencia, Spain

Saad Alqithami, Albaha University, Saudi Arabia

Panagiotis Bamidis, Aristotle University of Thessaloniki, Greece

Gian Carlo Cardarilli, University of Rome Tor Vergata, Italy

German Castellanos-Dominguez, Universidad Nacional de Colombia, Sede Manizales, Colombia

Srijata Chakravorti, Vanderbilt University, USA

Ahana Roy Choudhury, Florida State University, USA

Shiaofen Fang, Indiana University Purdue University Indianapolis, USA

Mai Gamal, German University in Cairo, Egypt

Faizal Hajamohideen, University of Technology and Applied Sciences, Oman

Monte Hancock, 4Digital Inc., USA

Cosimo Ieracitano, University Mediterranea of Reggio Calabria, Italy

Kavikumar Jacob, Universiti Tun Hussein Onn Malaysia, Malaysia

M. Shamim Kaiser, Jahangirnagar University, Bangladesh

Stamatis Karlos, University of Patras, Greece

Peter Kieseberg, St. Poelten University of Applied Sciences, Austria

Hongzhi Kuai, Maebashi Institute of Technology, Japan

Chih Lai, University of St. Thomas Minnesota, USA

Erwin Lemche, Institute of Psychiatry, Psychology & Neuroscience | King's College School of Medicine / Dentistry Department of Psychosis Studies | Section of Cognitive Neuropsychiatry, UK

Clement Leung, Chinese University of Hong Kong, China

Rory Lewis, University of Colorado at Colorado Springs, USA

Junhua Li, University of Essex, UK

Pablo Martínez Cañada, University of Granada, Spain

Vasilis Megalooikonomou, University of Patras, Greece  
Marta Molinas, NTNU, Norway  
Rafael Morales Herrera, Universidad de Castilla-La Mancha, Spain  
Shahryar Noei, Fondazione Bruno Kessler, Italy  
Ricardo Nuno Vigário, Nova School of Science and Technology, Lisbon, Portugal  
Lauren Reinerman-Jones, UCF Institute for Simulation and Training, USA  
Izabela Rejer, West Pomeranian University of Szczecin, Poland  
Ricardo Ron Angevin, University of Málaga, Spain  
Alessia Sarica, Magna Graecia University of Catanzaro, Italy  
Md. Shahriare Satu, Noakhali Science and Technology University, Bangladesh  
Noushath Shaffi, College of Applied Sciences, Oman  
Linlin Shen, Shenzhen University, China / University of Nottingham, UK / University of Macau, China  
Bo Song, University of Southern Queensland, Australia  
Suraj Sood, University of West Georgia / Polytechnique Inc. / LB Tutoring Academy LLC, USA  
William Speier, UCLA David Geffen School of Medicine, USA  
Ryszard Tadeusiewicz, AGH University of Science and Technology, Poland  
Giorgio Terracina, University of Calabria, Italy  
Héctor Fabio Torres Cardona, Universidad de Caldas, Colombia  
Christos Troussas, University of West Attica, Greece  
Boris M. Velichkovsky, Kurchatov Institute and the Russian State University for the Humanities, Moscow, Russia  
Sule Yildirim Yayilgan, NTNU, Norway  
Yu-Dong Zhang, University of Leicester, UK

## Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

## Table of Contents

Neurophysiological Changes Underlying Inhibitory Control in Mild Age-Related Hearing Loss <i>Shraddha Shende and Raksha Mudar</i>	1
Emotional Recognition and Classification Using Large Language Models <i>Clement Leung and Zhifei Xu</i>	4
A Self-Learning Neuromorphic System <i>Rory Lewis, Michael Bihn, Daniel Barbotko, and Zhenqi Liu</i>	11
A Comparative Analysis of Episodic Memory between Humans and AI Agents with Context Correlation <i>Shweta Singh, Anant Singh, and Vedant Ghatnekar</i>	17

# Neurophysiological Changes Underlying Inhibitory Control in Mild Age-Related Hearing Loss

Shraddha A. Shende

Department of Communication Sciences and Disorders  
Illinois State University  
Normal, IL, USA  
Email: sashend@ilstu.edu

Raksha A. Mudar

Department of Speech and Hearing Science  
University of Illinois Urbana-Champaign  
Champaign, IL, USA  
Email: raksha@illinois.edu

**Abstract**— Emerging evidence suggests behavioral alterations in inhibitory control in older adults with mild Age-Related Hearing Loss (ARHL). Whether there are underlying alterations in the neurophysiological mechanisms linked to these behavioral changes remains unexplored. The current study examined Event-Related Potentials (ERPs) and Event-Related Spectral Perturbations (ERSPs) linked to two Go/NoGo tasks (Single-Car/Object-Animal) in 17 older adults with unaided mild ARHL and 25 normal hearing controls. Group differences in N2 and P3 (ERPs) latency and amplitude and theta and alpha (ERSPs) power were examined in addition to their association with speech-in-noise recognition. Findings revealed differences in ERPs and ERSPs for the NoGo versus Go trials in the two groups. The mild ARHL group showed longer NoGo N2 latency relative to Go N2 latency on the Single-Car task, but similar findings were not observed within the control group. The control group showed higher P3 amplitude and greater alpha desynchronization for NoGo versus Go trials on the Object-Animal task, but this differentiation was lacking in the hearing loss group. These findings suggest alterations in neurophysiological mechanisms underlying inhibitory control in unaided mild ARHL. Additionally, poorer recognition of speech-in-noise in the hearing loss group was related to higher P3 amplitude for Go trials on the Object-Animal task, with a similar trend observed for NoGo trials, suggesting that those with worse central hearing exert greater neural effort on inhibitory control tasks. The study findings add to the literature on the impact of ARHL on cognition and its association to changes in complex listening functions.

**Keywords**—age-related hearing loss; inhibitory control; event-related potentials; event-related spectral perturbations; Go/NoGo tasks.

## I. INTRODUCTION

Inhibitory control is a cognitive control process, which allows us to suppress irrelevant information/responses in order to attend to relevant information [1]. It is often used in common listening situations, such as understanding Speech in Noise (SiN). For instance, in busy restaurants, inhibitory control allows one to suppress background noise and focus on their relevant conversation. Age-Related Hearing Loss (ARHL), a globally prevalent condition, affects various listening functions, including understanding SiN [2][3].

Growing evidence shows behavioral alterations in inhibitory control in older adults with ARHL relative to age- and education-matched Normal Hearing (NH) controls [4][5],

even in those with mild severity of hearing loss on common inhibitory control tasks such as Stroop and Go/NoGo [4]. Theoretical postulations linking hearing loss and cognition have long suggested underlying neural changes in individuals with ARHL [6]. It is plausible that neurophysiological changes underlie the overt inhibitory control changes in older adults with mild ARHL; however, this has not been examined. Event-Related Electroencephalography (EEG), which taps into real-time neural processing linked to cognitive processes, would be useful in this context. Both Event-Related Potentials (ERPs) and Event-Related Spectral Perturbations (ERSPs)/event-related oscillations measures, which are derived from EEG, might offer valuable insights. ERPs capture temporal aspects of the EEG signal, whereas ERSPs delineate the spectral and temporal aspects. On inhibitory control tasks, ERP studies have typically examined N2 (negative deflection at ~200 ms) and P3 (positive deflection at ~300 ms) components, while ERSP studies have examined theta (4-7 Hz) and alpha (8-12 Hz) power [7][8].

This study primarily examined differences in N2 and P3 amplitude and latency and alpha and theta power corresponding to two Go/NoGo tasks between older adults with unaided mild ARHL relative to NH controls. Our secondary aim was to examine associations between SiN recognition and ERPs/ERSPs corresponding to Go/NoGo tasks. Findings from this study will establish whether neurocognitive alterations in inhibitory control occur in mild ARHL. This is critical given that ARHL has been considered one of only 12 modifiable risk factors for dementia [9]. Markers that can assist in the identification of neurocognitive alterations will be instrumental in early detection and timely intervention for these individuals.

The rest of this paper is described as follows. Section II describes the methods. Section III describes the main results and discussion. Section IV concludes the article. The acknowledgment closes the article.

## II. METHODS

### A. Participants

Participants included 17 older adults with unaided mild ARHL and 25 NH controls with comparable age and education. Those with a history of neurological and psychological disorders, and other known etiologies of hearing loss were excluded.

### B. Tasks and Procedure

All participants completed a comprehensive audiological examination, including the Quick Speech-in-Noise (QuickSIN) test to examine SiN recognition. Inhibitory control was examined using two in-house developed Go/NoGo tasks, Single-Car, and Object-Animal tasks [10]. The simpler Single-Car task was a basic categorization task, where participants were shown line drawings of a car (160 trials) and a dog (40 trials) and were required to push a button to the stimuli of cars (Go trials) but withhold button push to stimuli of dogs (NoGo trials). For the more complex Object-Animal task, a superordinate categorization task, participants saw multiple exemplars of objects (160 trials) and animals (40 trials) and were required to push a button to stimuli of objects but withhold to stimuli of animals. Reaction time and accuracy were obtained.

### C. EEG Data Collection and Processing

EEG was collected while participants performed the two Go/NoGo tasks using a 64-electrode Neuroscan QuikCap. Collected data were pre-processed offline with noisy data and poorly functioning electrodes removed. Subsequently data were epoched from -500 to 0 ms. For ERP analyses, baseline correction was done from -500 to 0 ms, and ERP averages were created separately for trial type (Go/NoGo) and task (Single-Car/Object-Animal). Guided by previous research and visual inspection, N2 component between 150-300 ms was extracted across an average of frontal (F1, Fz, F2) and frontocentral (FC1, FCz, FC2) electrodes [7]. P3 was extracted between 250-600 ms at an average of frontocentral (Fc1, Fcz, Fc2), central (C1, Cz, C2), and centroparietal (CP1, CPz, CP2) [7]. Latency and mean amplitudes were used as measures. For ERSP analyses, EEGLAB toolbox with *newtimef.m* function [11] was used and baseline correction was conducted using a gain model [12]. Theta and alpha power were obtained across five electrode clusters: frontal (F1, Fz, F2), frontocentral (FC1, FCz, FC2), central (C1, Cz, C2), centroparietal (CP1, CPz, CP2), and parietal (P1, Pz, P2).

### D. Statistical Analyses

All data were analyzed using IBM SPSS Statistics (Version 26). General Linear Models (GLMs) for analyses, with group (ARHL/NH) as a between-subject factor and trial type (Go/NoGo) as a within-subject factor. Alpha was fixed at 0.05, and in the case of significant group-by-trial interactions, post hoc comparisons were carried out. Bonferroni corrections were used to correct for multiple comparisons. Given the small sample size, separate analyses were conducted for the simpler Single-Car task and the complex Object-Animal task.

## III. MAIN RESULTS AND DISCUSSION

Behavioral data showed evidence of changes in inhibitory control in individuals with mild ARHL. This was observed on post-hoc comparisons, with lower accuracy on NoGo versus

Go trials within the ARHL group on the Single-Car task ( $p < 0.001$ ), although similar differences were not observed within the NH group ( $p > 0.05$ ). These findings suggest that the mild ARHL group experienced challenges in withholding a prepotent response. Furthermore, this was noted during perceptual processing, since the findings were observed on the basic categorization task, Single-Car, which mainly consisted of perceptual stimuli.

EEG findings also revealed differential processing of NoGo versus Go trials between the two groups. These findings were observed on post-hoc comparisons. Particularly, longer N2 latency was noted on the NoGo versus Go trials within the ARHL group for the Single-Car task ( $p = 0.006$ ), but similar patterns were not seen within the NH group ( $p > 0.05$ ). This finding suggests individuals with mild ARHL had prolonged neural processing times early on (150-300 ms) for the NoGo (inhibition trials) versus Go trials, but this differential processing was not seen within NH controls. However, the control group showed differential neural processing at later time points. Higher P3 amplitude was noted for NoGo versus Go trials within the control group for the Object-Animal task ( $p = 0.033$ ) between 250-600 ms. The control group also showed more negative alpha power for the NoGo versus Go trials on both tasks ( $p < 0.001$ ) between 300-650 ms, but the same pattern was not observed within the ARHL group. Given that P3 ERP and alpha band have been linked to neural effort [13], and that the reaction time for both tasks was within 450 ms, it seems that the NH group devotes more neural resources/effort to evaluate the stimuli of the inhibition trials at later time points, likely after making a response, but this is not done by the ARHL group. On analyses for our secondary aim, we found a positive relationship between QuickSIN score and P3 amplitude on Go trials on the Object-Animal task ( $p = 0.021$ ) in the mild ARHL group. A similar trend was observed with NoGo trials ( $p = 0.062$ ). These findings suggest that individuals with ARHL who had poorer SiN recognition scores used greater neural effort/resources for performing an inhibitory control task that involved superordinate categorization.

## IV. CONCLUSION AND FUTURE WORK

Our study shows that neural processing related to inhibitory control in those with mild ARHL is different from normal hearing controls. This differentiation was evident in visual tasks, suggesting modality-independent changes in inhibitory control in those with untreated and mildest degree of ARHL, which constitutes the largest percentage of older adults with this condition [13]. Furthermore, these inhibitory control changes are related to complex listening functions such as SiN recognition. In summary, our work significantly advances the knowledge of neural changes underlying cognitive alterations in older adults with ARHL. However, the current work has some limitations. While our groups were not significantly different in age, they were unequal in number. Larger sample sizes with equal groups are needed to validate the findings. Additionally, future work examining visual inhibitory control is needed to examine the replicability of the current findings.



ACKNOWLEDGMENT

The authors thank the ASH Foundation and the Health, Aging, and Disability Center at the University of Illinois for funding support.

REFERENCES

- [1] L. Hasher and R. T. Zacks, "Working Memory, Comprehension, and Aging: A Review and a New View," *Psychology of Learning and Motivation - Advances in Research and Theory*, vol. 22, no. C, pp. 193–225, 1988, doi: 10.1016/S0079-7421(08)60041-9.
- [2] CHABA, "Speech understanding and aging. Working Group on Speech Understanding and Aging. Committee on Hearing, Bioacoustics, and Biomechanics, Commission on Behavioral and Social Sciences and Education, National Research Council," *J Acoust Soc Am*, vol. 83, no. 3, pp. 859–895, 1988.
- [3] National Institute on Deafness and Other Communication, "Age-related hearing loss," *NIDCD Fact Sheet*, pp. 1–4, Feb. 2023. [Retrieved: December, 2023]. <https://www.nidcd.nih.gov/health/age-related-hearing-loss>
- [4] S. A. Shende and R. A. Mudar, "Cognitive control in age-related hearing loss: A narrative review," *Hearing Research*, vol. 436. Elsevier B.V., Sep. 01, 2023. doi: 10.1016/j.heares.2023.108814.
- [5] D. M. P. Jayakody, P. L. Friedland, R. H. Eikelboom, R. N. Martins, and H. R. Sohrabi, "A novel study on association between untreated hearing loss and cognitive functions of older adults: Baseline non-verbal cognitive assessment results," *Clinical Otolaryngology*, vol. 43, no. 1, pp. 182–191, 2018, doi: 10.1111/coa.12937.
- [6] U. Lindenberger and P. Baltes, "Sensory Functioning and Intelligence in Old Age: A Strong Connection," *Psychol Aging*, vol. 9, pp. 339–355, 1994, doi: 10.1037/0882-7974.9.3.339.
- [7] M. Falkenstein, J. Hoormann, and J. Hohnsbein, "ERP components in Go/Nogo tasks and their relation to inhibition," *Acta Psychol (Amst)*, vol. 101, no. 2, pp. 267–291, 1999, doi: [https://doi.org/10.1016/S0001-6918\(99\)00008-6](https://doi.org/10.1016/S0001-6918(99)00008-6).
- [8] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Res Rev*, vol. 29, no. 2, pp. 169–195, 1999, doi: [https://doi.org/10.1016/S0165-0173\(98\)00056-3](https://doi.org/10.1016/S0165-0173(98)00056-3).
- [9] G. Livingston *et al.*, "Dementia prevention, intervention, and care: 2020 report of the Lancet Commission," *The Lancet*, vol. 396, no. 10248. Lancet Publishing Group, pp. 413–446, Aug. 08, 2020. doi: 10.1016/S0140-6736(20)30367-6.
- [10] M. Maguire *et al.*, "The influence of perceptual and semantic categorization on inhibitory processing as measured by the N2-P3 response," *Brain Cogn*, vol. 71, no. 3, pp. 196–203, 2009, doi: 10.1016/j.bandc.2009.08.018.
- [11] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J Neurosci Methods*, vol. 134, no. 1, pp. 9–21, 2004, doi: 10.1016/j.jneumeth.2003.10.009.
- [12] R. Grandchamp and A. Delorme, "Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials.," *Front Psychol*, vol. 2, p. 236, 2011, doi: 10.3389/fpsyg.2011.00236.
- [13] S. K. Mamo, C. L. Nieman, and F. R. Lin, "Prevalence of untreated hearing loss by income among older adults in the United States," *J Health Care Poor Underserved*, vol. 27, no. 4, pp. 1812–1818, Nov. 2016, doi: 10.1353/hpu.2016.0164

# Emotional Recognition and Classification Using Large Language Models

Clement Leung

*School of Science and Engineering and  
Guangdong Provincial Key Laboratory of  
Future Networks of Intelligence  
Chinese University of Hong Kong  
Shenzhen, China  
clementleung@cuhk.edu.cn*

Zhifei Xu

*School of Science and Engineering  
Chinese University of Hong Kong  
Shenzhen, China  
zhifeixu@link.cuhk.edu.cn*

**Abstract**—Many tasks, particularly safety-critical ones, require the associated human performers to be in the right emotional states. Correct emotion state recognition frequently becomes an important concern and mainstream methods often use Pre-trained Language Models (PLMs) as the backbone to incorporate emotional information. The latest Large Language Models (LLMs), such as ChatGPT have demonstrated strong capabilities in various natural language processing tasks. However, existing research on ChatGPT zero-shot has received insufficient evaluation of the performance of image emotion recognition and analysis. In this paper, we study emotion classification and prediction based on positive and negative emotional states and evaluate the emotion recognition capabilities of ChatGPT4 focusing primarily on images. We empirically analyze the impact of labeled emotion recognition and interpretability of different datasets. Experimental results show that, while ChatGPT4 can make some useful predictions of emotions based on images, there is still a substantial gap in prediction results and accuracy. Qualitative analysis shows its potential compared to state-of-the-art methods, but it also suffers from limitations in robustness and accurate inferences.

**Index Terms**—*image emotion recognition, large language model, zero-shot, ChatGPT4.*

## I. INTRODUCTION

Emotion recognition and prediction have been recognized to be a significant factor affecting human safety and have been widely studied [1][2][3][4][5][6]. There exist in general multiple ways for people to express their emotions or feelings naturally, such as voice, text, video, facial expressions, and physical behaviors. Moreover, since the ChatGPT [7] and Instruct-GPT [8] are currently believed to be a powerful and usable tool in different applications, we wish to investigate how they can be leveraged to assist in performing effective emotion recognition. Since emotional support is currently a key capability for many people in a wide variety of conversational scenarios, such as inter-social actions, mental health support, and customer chat services, we investigate the usefulness and competence of ChatGPT4 [9] to classify emotions based on facial expressions.

In fact, in today's society, people are under more and more pressure, such as being criticized by leaders, unfair

experiences, relationship break-ups, and so on. Once in a stressful situation, people may lose control of their emotions [10][11]. In this case, they may act irrationally to hurt themselves or others. Examples of incidents linked to emotional problems are rife: suicidal thoughts under the stress of school or work; frequent school shootings in the United States; and fatal crashes of vehicles involving angry drivers. In some special jobs, the emotion of the employee plays a particularly important role, such as a surgeon, pilot, truck driver, and so on (e.g., a recent incident of a pilot who attempted to cut off plane engines in mid-air was found to suffer from depression [12]). These highlight why emotion recognition is so important in our everyday life. So, what precisely is emotion recognition? Emotion recognition is a subfield of artificial intelligence that focuses on identifying and analyzing human emotions based on various inputs. The main goal of emotion recognition is to discern the emotional state of an individual or a group of individuals.

In recent years, since emerging large language model technologies and their rapid iterative development have produced many human-computer interaction robots, which have brought a new technological revolution to the field of dialogue, represented by ChatGPT4. At the same time, they demonstrate strong general language processing capabilities and also bring unprecedented semantic understanding and response generation capabilities to humans. Since their emergence has greatly improved the interactive experience with human users, the question of whether it shows emotion in the conversation has not yet been explored, and we are interested in the development of emotional dialogue technology in ChatGPT4. At the same time, we hope to explore the multi-modal tasks of ChatGPT4 in the field of emotion recognition and analyze its advantages and disadvantages [13][14][15].

In the next section, we discuss what others have done and why existing solutions are not enough. In the third section, we describe how to recognize and predict emotions. In the fourth section, experimental analysis is carried out for different categories of emotion. Finally, the conclusion of our study is drawn and summarized.

## II. RELATED WORK

Two main models have emerged to represent and explain emotion recognition: categorical and continuous. Categorical models, also known as discrete emotion models, assume the existence of a certain number of primary or basic emotions that are universally recognized and experienced by humans, regardless of cultural or individual differences. The main approach of categories in emotion recognition which is one of the most well-known categorical models was proposed by Ekman, who identified six basic emotions: happiness, sadness, fear, anger, surprise, and disgust. These emotions are considered fundamental and universally recognizable. The other is Plutchik's extension [16] of Ekman's model [17] by including eight primary emotions arranged in a wheel. These include joy, trust, fear, surprise, sadness, disgust, anger, and anticipation. On the other hand, continuous models represent emotions in a multidimensional space, usually visualized as a spectrum or continuum [18] [19]. These dimensions are used to represent a person's emotional state in a more granular manner than discrete categories. *Valence* refers to the positivity or negativity of an emotion. *Arousal* refers to the degree of excitement or calm. *Dominance* refers to the degree of control or influence a person feels in a situation.

Emotion recognition has been studied primarily in terms of single modes. However, people express emotions through voice, text, video, facial expressions, and physical behavior; therefore, it is difficult to accurately judge emotions through a single mode alone. Since multimodal emotion classification involves the integration of multiple information sources, such as facial expressions, intonation, and physiological signals, we shall make use of images of facial expressions and texts using multimodal data sets, and Convolutional Neural Networks (CNN) [20] or transform classification models to identify and classify emotions. In many application scenarios, in addition to the current classification, it is necessary to also predict the evolution of emotion states.

## III. EMOTION TRANSITION AND CLASSIFICATION

In the real world, people's emotions are usually continuous, transitioning from one emotional state to another [21]. In practical situations, it is often necessary to predict the emotional state of the relevant personnel, e.g., allocating work rosters and scheduling hospital operations. As indicated earlier, people may act out emotionally when they are unfair or tired. In safety-critical jobs, we want the person doing the work to be in a sound emotional state [12]. In other words, people who work in high-risk industries can endanger the safety of others if they have emotional problems.

We shall focus on the categorical models and make use of both Plutchik's model, as well as Ekman's model. For Plutchik's categories using the eight primary emotions of joy, trust, fear, surprise, sadness, disgust, anger, and anticipation, we group these into positive and negative emotions, so that an individual is regarded as emotionally competent if s/he is a positive emotion state, and incompetent otherwise. We group

the states of joy, trust, surprise, and anticipation as positive (+1) emotions and group fear, surprise, sadness, disgust, and anger as negative (-1) emotions. We consider an individual to be emotionally competent if s/he is in a positive emotional state.

For Ekman's categories using the six primary emotions of happiness, sadness, fear, anger, surprise, and disgust, we also group these into positive and negative emotions. We group the states happiness and surprise as positive (+1) emotions and group sadness, fear, anger, and disgust as negative (-1) emotions. As a variation, for safety-critical jobs, we may wish to be extra safe and more strict concerning positive emotion, and we may take surprise out from the positive emotion category, and place it in the negative emotion category. Here, though, we place surprise in the positive emotion category.

We represent the emotional state at time  $t$  by  $S(t)$ ;  $t$  is time, and  $S(t)$  is the person's emotional change with time. As indicated,  $S(t)$  can be take on the values  $S(t) = 1$  or  $S(t) = -1$ , which corresponds, respectively, to positive (+1) and negative (-1) emotions. Since humans are continually bombarded by various external happenings, mood changes are often caused by events outside their control, which may be due to a variety of factors. Such factors may be related to changing conditions of financial situation, relationships, health, work, stock market, and family, and the combination of these may cause a transition from a positive emotion state to a negative emotion state and vice versa.

First, let  $S(0) = 1$  then, we represent the transition time points (from +1 to -1, or from -1 to +1) by a Poisson Process. Now,  $S(t) = 1$  if the number of transitions in the time interval  $(0, t)$  is even, and  $S(t) = -1$  if this number is odd. Therefore,

$$P[S(t) = 1 | S(0) = 1] = p_0 + p_2 + p_4 + \dots + \dots, \quad (1)$$

where  $p_k$  is the number of Poisson points in  $(0, t)$  with parameter  $\lambda$ . That is,

$$\begin{aligned} P[S(t) = 1 | S(0) = 1] &= e^{-\lambda t} \left[ 1 + \frac{(\lambda t)^2}{2!} + \frac{(\lambda t)^4}{4!} \dots + \dots \right] \\ &= e^{-\lambda t} \cosh \lambda t \end{aligned} \quad (2)$$

Now,  $S(t) = -1$  if the number of points in the time interval  $(0, t)$  is odd; that is,

$$\begin{aligned} P[S(t) = -1 | S(0) = 1] &= e^{-\lambda t} \left[ 1 + \frac{(\lambda t)^3}{3!} + \frac{(\lambda t)^5}{5!} \dots + \dots \right] \\ &= e^{-\lambda t} \sinh \lambda t \end{aligned} \quad (3)$$

Equation (2) represents the probability that the emotion is still positive at time  $t$  given that it was positive at time 0. Equation (3) gives the probability that the emotion is positive at time  $t$  given that it was negative at time 0. The parameter  $\lambda$  in both expressions represents a rate at which emotions change or decay over time. A larger value of  $\lambda$  would mean emotions

change more rapidly, while a smaller value would mean they change more slowly. Thus

$$E[S(t)|S(0) = 1] = e^{-\lambda t} [\cosh \lambda t - \sinh \lambda t] = e^{-2\lambda t} \quad (4)$$

#### IV. EXPERIMENTATION

##### A. Single Emotion Recognition

Emotion Recognition in Conversations [22][23] (ERC) is widely used in various conversation environments, including emotional analysis of comment areas on social media and supervision of various high-pressure industry personnel. At the same time, conversational emotion recognition can be implemented in chatbots to assess the user's emotional state and promote emotion-driven responses. As mentioned earlier, ChatGPT4 is a form of conversational bot, and we are interested in analyzing whether it can recognize emotions and sentiments.

1) *Dataset and Evaluation Graph:* We using three different datasets from Kaggle, Facial Expressions Training Data, Emotion Detection, and Natural Human Face Images for Emotion Recognition.

**Emotion Detection** This dataset consists of 35,685 examples of 48x48 pixel grayscale images, which contain two folders, one is trained, and the other one is tested. The folders contain different categories of emotional images. In addition, the images have been labeled by the authors for different types of emotions, including anger, disgust, fear, happiness, neutral, sad, and surprise.

**Facial Expressions Training Data** AffectNet [24] is a large database of faces marked with "impact" (the psychological term for facial expressions). In order to accommodate common memory limitations in this dataset, the authors reduce the resolution to 96x96 for the neural network processing, which indicates that all images are 96x96 pixels. Meanwhile, using Singular Value Decomposition, each image's Principal Component Analysis is calculated. The threshold for the Percentage of the First Component (index 0) in the principal components (in short the PFC%) was set to lower than 90%. This means that most if not all of the monochromatic images were filtered out. Finally, the dataset is based on Affectnet-HQ, using a state-of-the-art Facial Expression Recognition (FER) model that refines the AffectNet original label to re-label its dataset, which contains eight emotional categories - anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise.

**Natural Human Face Images for Emotion Recognition** Since facial expression recognition is usually performed using standard datasets, such as the Facial Expression Recognition dataset (FER), Extended Cohn-Kanade dataset (CK+) and Karolinska Directed Emotional Faces dataset (KDEF) for machine learning, however, this dataset was collected from the internet and manually annotated to provide additional data on real faces, with over 5,500 + images with 8 emotions categories: anger, contempt, disgust, fear, happiness, neutrality, sadness and surprise. All images contain grayscale human faces (or sketches). Each image is 224 x 224

pixel grayscale in Portable Network Graphics (PNG) format. Images are sourced from the internet where they are freely available for download e.g., Google, Unsplash, Flickr, etc.

2) *Task Definition of Single Emotion:* We are given the three data sets and select six types of emotions in the data set: anger, disgust, happiness, neutral, sadness, and surprise. Table I shows some examples of comparison between annotation and ChatGPT4's prediction, where red highlights the discrepancy. In each data set, 50 images of 6 types of emotions are randomly selected and put into ChatGPT4 for judgment. At the same time, since ChatGPT4 was released in 2023, the above experiments are all conducted using ChatGPT4. We use supervised learning and evaluate ChatGPT4's performance in zero-shot prompting settings for the above task. After the judgment of ChatGPT4, if the result is the same as our cognitive result, it will be recorded as 1, if the result is different, it will be recorded as 0, and the emotion will be recorded as positive, negative or neutral according to the description of ChatGPT4. Additionally, a Receiver Operating Characteristic (ROC) [25] curve is generated based on our recorded results. In the ROC curve, if it is a positive emotion, such as happiness, neutral, or surprise; we mark the fact result as 1. On the contrary, if it is a negative emotion, such as anger, disgust, or sadness; we mark the fact result as 0. The prediction result of ChatGPT4, in the positive emotion, is recorded as 1 if it is consistent with the actual result, otherwise, it is recorded as 0. In the same way, if it is a negative emotion if it is consistent with the fact, it will be recorded as 0, and if it is opposite, it will be recorded as 1. The evaluation index is divided into 1-3 points, 1 point means low confidence, 2 points means moderate confidence, and 3 points means high confidence.

3) *Result of Single emotion:* For tabulated data, TPR is True Positive Rate also known as Sensitivity, which measures the proportion of actual positives that are correctly identified by the model. FPR is False Positive Rate also known as 1-Specificity, it is the ratio of negative instances that are incorrectly classified as positive. Observed Operating Points are points on the ROC curve that correspond to specific thresholds used in the classifier. Each point represents the balance between TPR and FPR for a specific threshold. For example, a high threshold may result in low FPR but also low TPR, while a low threshold may increase both TPR and FPR. These points help evaluate the performance of the model and select the best threshold for the classification task. They demonstrate the trade-off between capturing as many positive results as possible (higher TPR) and avoiding false positives (lower FPR).

Table II shows the results of ChatGPT4's prediction based on a single emotion. For the surprise positive emotion, we see that the accuracy of ChatGPT4's prediction results is around 70%; for the happiness positive emotion, we see that the corresponding accuracy is around 78%, indicating highly discriminative discerning of positive emotions. Grouping the two positive emotions, a good degree of accuracy is obtained.

TABLE I. EXAMPLE OF CHATGPT4'S PREDICTION ON ERC TASK WITH IMAGES.

Image Content	Question	Annotation	Prediction
	What is the emotion of this person?	anger	surprise/shock/fear
	What is the emotion of this person?	happiness	happiness
	What is the emotion of this person?	happiness	happiness/joy
	What is the emotion of this person?	anger	frustration/concern/disapproval
	What is the emotion of this person?	sadness	sadness/crying
	What is the emotion of this person?	surprise	surprise

For negative emotions, the accuracy of ChatGPT4’s predictions from low to high is disgust, anger, and sadness. In the actual test, we find that zero-shot ChatGPT4 can predict negative emotions, but it cannot accurately determine whether it is disgust or anger. At the same time, because the individual expressions of disgust emotions are inconsistent, the prediction results are the lowest. We can find that GPT4 has six types of emotion recognition accuracy from high to low: happiness, surprise, neutral, fear, anger, and disgust. For surprise images, although GPT4 can identify most of the images as surprise or astonishment, it cannot accurately judge whether surprise is a positive emotion or a negative emotion, so it thinks that the emotion of surprise is mainly neutral. This is why the result is very similar to the neutral result.

As mentioned above, in order to avoid the harm caused by negative emotions to people in high-risk industries or high-risk groups, we mainly look at the three categories of emotions: anger, disgust, and sadness. We observe that the FPR of sadness is 0.3267, the FPR of anger is 0.4800, and the FPR of disgust is 0.6467. According to the above explanation of the FPR index, it means that the emotion of disgust is the least accurate to identify, and the emotion of the disgust category is the most difficult to judge among the six categories of emotion. In addition, the accuracy of negative emotion recognition is too low, and more prompt words may be needed to help GPT4 make judgments because according to the current zero-shot, GPT4 can determine that people have negative emotions, but cannot accurately identify disgust, contempt, or anger.

TABLE II. RESULT OF CHATGPT4’S PREDICTION ON SINGLE EMOTION RECOGNITION TASK WITH IMAGES

Emotion	Accuracy
anger	30%
disgust	19.30%
happiness	78%
neutral	69.34%
sadness	44.30%
surprise	70%

4) *Analysis and Discussion:* During the training process, it is inevitable that the images in some data sets are inconsistent with our cognition in real life. Since people have different feelings about images, there may be biases in partial image emotion recognition. For this part of the image, we use our cognition as the final judgment and compare it with the results of GPT4.

Additionally, we discover another issue: an inconsistency between ChatGPT4 and the dataset guide. Examining these actual prediction samples shows that the main challenge of ChatGPT4 is the bias between its norm and the norm of the dataset. Although dataset annotations may follow specific guidelines for determining corresponding sentiments, for specific cases, ChatGPT4 has its own interpretations and standards. For example, the dataset annotation classifies emotions when the person in the image is described as angry, while ChatGPT4 considers it as sad or lost. The difference

cannot be attributed to one being right and the other wrong, but rather emphasizes the use of different criteria, both of which are negative emotions. Upon further discussion, this misaligned criterion may not be due to the functionality of ChatGPT4 but may be attributed to under-posting tips. As prompt word guides become more complex, it becomes unreasonable to cover them with only a small amount of content. This insight can speculate on possible future directions: if the goal is not to strictly adhere to a specific guideline, then enhancements based on a few prompt settings (e.g., describing people in images) are feasible. However, evaluation using dataset labels may not be appropriate and may require extensive manual evaluation. Conversely, if the goal is to strictly adhere to specific guidelines, then several prompt settings may not be the best option, and supervised fine-tuning of the model is still a better option.

*B. Different Categories Emotion Recognition in Different Dataset*

1) *Task Definition of Emotion Dataset:* First, we use three data sets: emotion detection, facial expressions training data, and natural human faces. Since each dataset has different label classifications, each dataset randomly selects 50 images from 6 images of the same category (anger, disgust, happiness, neutral, sadness, surprise), for a total of 300 images. Next, we put them into GPT4 for inspection and record the results, which are shown in Table III.

2) *Result of Emotion Dataset:* Since the concept of the partial definition has already been explained previously, here we only discuss the Fitted ROC Area and Empiric ROC Area. Fitted ROC Area refers to the area under the ROC curve that uses some form of parametric or semi-parametric model to fit the data. Here we use the maximum likelihood fit of a binormal model to calculate and draw the ROC curve. Empirical ROC Area, often referred to simply as Area Under the Curve (AUC), is a measure based on an empirical ROC curve constructed directly from data. The curve is created by plotting the True Positive Rate (TPR), versus the False Positive Rate (FPR), at different threshold settings. The AUC of an empirical ROC curve provides a measure of a model’s ability to differentiate between two classes (positive and negative) at all possible thresholds. The larger the AUC, the better the model performance. An AUC of 0.5 indicates no discrimination (equivalent to a random guess), whereas an AUC of 1.0 indicates perfect discrimination.

In the Emotion Detection data set, because the two Observed Operating Points of TPR are both 0.8133, the Fitted ROC Area is Degenerate. In addition, we can find that the prediction results of Emotion Detection are the best regardless of the accuracy or empirical ROC Area, which shows that using the Emotion Detection data set for ChatGPT has the highest zero-shot prediction. Next is Natural Human, Facial Expression. Facial Expression is an RGB image data set, and the other two are black-and-white image data sets. Therefore, we find that the accuracy of RGB images is not necessarily higher

TABLE III. COMPARISON DIFFERENT DATASET OF CHATGPT4'S PREDICTION FOR EMOTION RECOGNITION TASK WITH IMAGES

Dataset	Accuracy	Sensitivity	Specificity	Fitted ROC Area	Empiric ROC Area
<b>Emotion Detection</b>	75.30%	81.30%	69.30%	Degenerate	0.74
<b>Facial Expression</b>	66.00%	61.30%	70.70%	0.665	0.634
<b>Natural Human</b>	70.30%	74.70%	66.00%	0.752	0.681

than that of black and white images, which means that color has little impact on the prediction process in emotional image recognition.

The ordinate of the ROC curve represents sensitivity. The higher the index, the higher the diagnostic accuracy. The abscissa represents 1-specificity. The lower the index, the lower the false positive rate. So in general, the closer the point is to the upper left corner of the ROC space, the better the diagnostic effect is. This means that the closer the sensitivity is to 1, the higher the prediction accuracy of the model. We can find that the sensitivities of the three data sets are 81.3% (emotion detection), 61.3% (facial expression), and 74.7% (natural human), respectively. From the specificity, it would appear that GPT4's emotion detection may be considered to be acceptable, especially for the first and last datasets. From the Accuracy and Specificity columns of Table III, the figures are somewhat comparable to the sensitivity, although marginally less acceptable.

V. CONCLUSION AND FUTURE WORK

In this paper, we study the zero-shot ability of ChatGPT4 in emotional reasoning and judgment based on images. The experimental results show that ChatGPT's predictive ability is limited, but it has the potential to improve via mental health analysis and some humanistic inputs. We target the analysis for limitations, such as unstable predictions and inaccurate inferences. Overall, our study shows that subjective tasks, such as mental health analysis and image conversational emotion reasoning remain challenging for ChatGPT. With more refined prompt engineering and contextual example selection, we believe greater future efforts are needed to improve the performance of ChatGPT and address its limitations in order to enable it to be practically applied to real-world mental health and related situations.

REFERENCES

[1] C. H. C. Leung, J. J. Deng, and Y. Li, "Enhanced Human-Machine Interactive Learning for Multimodal Emotion Recognition in Dialogue System," Proceedings of the 5th International Conference on Algorithms, Computing and Artificial Intelligence, pp. 1-7, 2022.

[2] J. J. Deng, and C. H. C. Leung, "Towards Learning a Joint Representation from Transformer in Multimodal Emotion Recognition," Brain Informatics: 14th International Conference, BI 2021, Virtual Event, September 17-19, 2021, Proceedings 14, pp. 179-188, 2021, Springer.

[3] J. J. Deng, C. H. C. Leung, and Y. Li, "Multimodal emotion recognition using transfer learning on audio and text data," Computational Science and Its Applications-ICCSA 2021: 21st

International Conference, Cagliari, Italy, September 13-16, 2021, Proceedings, Part III 21, pp. 552-563, 2021, Springer.

[4] J. J. Deng, and C. H. C. Leung, "Deep Convolutional and Recurrent Neural Networks for Emotion Recognition from Human Behaviors," Computational Science and Its Applications-ICCSA 2020: 20th International Conference, Cagliari, Italy, July 1-4, 2020, Proceedings, Part II 20, pp. 550-561, 2020, Springer.

[5] J. J. Deng, and C. H. C. Leung, "Dynamic time warping for music retrieval using time series modeling of musical emotions," IEEE transactions on affective computing, vol. 6, no. 2, pp. 137-151, 2015.

[6] J. J. Deng, C. H. C. Leung, M. Alfredo, and L. Chen, "Emotional states associated with music: Classification, prediction of changes, and consideration in recommendation," ACM Transactions on Interactive Intelligent Systems (TiiS), vol. 5, no. 1, pp. 1-36, 2015, ACM New York, NY, USA.

[7] T. Brown et al., "Language models are few-shot learners," Advances in neural information processing systems, vol. 33, pp. 1877-1901, 2020.

[8] L. Ouyang et al., "Training language models to follow instructions with human feedback," Advances in Neural Information Processing Systems, vol. 35, pp. 27730-27744, 2022.

[9] Open AI, ChatGPT-4, "https://openai.com/gpt-4".

[10] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, "Natural language processing applied to mental illness detection: a narrative review," NPJ digital medicine, vol. 5, no. 1, pp. 46, 2022, Nature Publishing Group UK London.

[11] D. Ciralo, A. Celesti, M. Fazio, M. Bonanno, M. Villari, and R. S. Calabrò, "Emotional Artificial Intelligence Enabled Facial Expression Recognition for Tele-Rehabilitation: A Preliminary Study," 2023 IEEE Symposium on Computers and Communications (ISCC), pp. 1-6, 2023.

[12] R. Lewis, and J. Rose, 'I'm not okay,' off-duty Alaska pilot allegedly said before trying to cut the engines, 'https://www.npr.org/2023/10/24/1208244311/alaska-airlines-off-duty-pilot-switch-off-engines', Oct. 2023.

[13] K. Yang, S. Ji, T. Zhang, Q. Xie, and S. Ananiadou, "On the evaluations of chatgpt and emotion-enhanced prompting for mental health analysis," arXiv preprint arXiv:2304.03347, 2023.

[14] W. Zhao, Y. Zhao, X. Lu, S. Wang, Y. Tong, and B. Qin, "Is ChatGPT Equipped with Emotional Dialogue Capabilities?" arXiv preprint arXiv:2304.09582, 2023.

[15] H. D. Le, G. S. Lee, S. H. Kim, S. Kim, and H. J. Yang, "Multi-Label Multimodal Emotion Recognition With Transformer-Based Fusion and Emotion-Level Representation Learning," IEEE Access, vol. 11, pp. 14742-14751, 2023.

[16] P. Robert, "Emotion: Theory, research, and experience. vol. 1:

- Theories of emotion,” 1980, Academic Press: Cambridge, MA, USA.
- [17] P. Ekman, ”Facial expressions of emotion: New findings, new questions,” 1992, SAGE Publications Sage CA: Los Angeles, CA.
- [18] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, ”Emotion recognition in context,” Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1667–1675, 2017.
- [19] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, ”Context based emotion recognition using emotic dataset,” IEEE transactions on pattern analysis and machine intelligence, vol. 42, no. 11, pp. 2755–2766, 2019.
- [20] W. Zhang, X. He, and W. Lu, ”Exploring discriminative representations for image emotion recognition with CNNs,” IEEE Transactions on Multimedia, vol. 22, no. 2, pp. 515–523, 2019.
- [21] A. Metallinou, and S. Narayanan, ”Annotation and processing of continuous emotional attributes: Challenges and opportunities,” 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), pp. 1–8, 2013.
- [22] S. Poria, N. Majumder, R. Mihalcea, and E. Hovy, ”Emotion recognition in conversation: Research challenges, datasets, and recent advances,” IEEE Access, vol. 7, pp. 100943–100953, 2019.
- [23] S. Poria et al., ”Recognizing emotion cause in conversations,” Cognitive Computation, vol. 13, pp. 1317–1332, 2021, Springer.
- [24] A. Mollahosseini, B. Hasani, and M. H. Mahoor, ”Affectnet: A database for facial expression, valence, and arousal computing in the wild,” IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18–31, 2017.
- [25] T. Fawcett, ”An introduction to ROC analysis,” Pattern recognition letters, vol. 27, no. 8, pp. 861–874, 2006, Elsevier.



# A Self-Learning Neuromorphic System

Rory Lewis\*, Michael Bihn†, Zhenqi Liu‡, Daniel S. Barbotko§  
 Department of Computer Science, University of Colorado at Colorado Springs  
 Colorado Springs, Colorado, USA  
 \*rlewis5@uccs.edu, †mbihn@uccs.edu, ‡zliu3@uccs.edu  
 §dbarbotk@uccs.edu

**Abstract**—In the continuing research to implement a plurality of self-wiring synapses comprised of Field Programmable Gate Arrays (FPGAs) on a Complementary Metal Oxide Semiconductor (CMOS) system to accommodate Artificial Intelligence (AI) on a microprocessor, we delve into how a system can emulate not just a self-wiring CMOS system, but also how it can emulate brain growth at the connectome level. The enigma of contemporary advancements in AI and chip manufacturing diverging from bio-inspired systems is fascinating, especially given that AI and microprocessor engineers readily acknowledge the superior capabilities of biological brains. This paper introduces a bio-inspired device made of steel, plastic, and silica, which autonomously rewires itself, evolving and enhancing its intelligence without human intervention. The research will delve into the intricacies of the FPGA prototype’s functionality, shedding light on both its technical aspects and the broader social and technological implications associated with the development of this neuromorphic chip. Next, we introduce the theoretical ability for the CMOS to grow its connectivity to FPGAs as does a human baby. Herein, we introduce the uniqueness of applying the logistical growth function to the curve fitting of the multidimensional measures of brain growth, on a CMOS system.

**Index Terms**—Bio-Inspired; Neuromorphic; AI.

## I. INTRODUCTION

The motivation for this research effort is that, despite significant advances in neuromorphic systems, the AI systems based on them are still far from their biological counterparts [1]. Such gaps exist because, while the world lauds the progress of ChatGPT and other high-end AI systems, the engineers are reticent to reveal that these systems consume approximately 200 terawatt hours of energy per year [2], shown in Fig. 1a. The issue is that there is no financial impetus for chipmakers to stop reaping the financial benefits of this explosion in processor requirements and take on high-risk bio-inspired chips. We will illustrate how state-of-the-art systems such as TrueNorth, Loihi, SpiNNaker, BrainScaleS, and NeuronFlow [3] have been unable to synthesize biological neurons onto a solid-state-device because: i) their neuromorphic hardware systems are based on existing CMOS technology, and CMOS devices can only numerically simulate biological neural networks [1], and ii) neuroscientists do not understand exactly how neurons function. Specifically, we do not understand how neurogenesis, differentiation, and synaptogenesis work [4]. Yes, we know that: i) neurons send and receive neurotransmitters, chemicals that carry information between brain cells [5], and ii) depending on where a neuron is located, it can perform the job of a sensory neuron, a motor neuron, or an interneuron, so there

is no single process that explicitly synchronizes the work of all neurons [4]. We can show this flaw by considering the following abridged neurological developmental scenario.

**EXAMPLE 1** *First Event*: In Fig. 1b, the child is playing with his red ball. He releases it and notices that the ball dropped downwards onto the floor. We represent this by simplifying the synaptic-dendrite connections and its neuron with a green dot. We note that, for argument’s sake, he also receives 100,000 sensory items with the first six being: i) he is on a soft carpet, ii) in the living room, iii) Mom is happy, vi) outside the sky is bright blue, v) birds are singing, and vi) it is nice and warm. Similarly, the last three of the 100,000 sensory item neurons are labeled 99,998 99,999 and 100,000. *Second Event*: Two weeks pass. In Fig. 1c, Mom and Dad visit a friend while the child sleeps. At the end of the evening, while Dad carries him outside and fastens him into the car seat, the child awakens and again drops his red ball. He notices that, just like two weeks ago, it dropped downwards again onto the floor of the car. It did not go up. Again, his brain receives 100,000 sensory pulses - the first six being: i) he’s in a car, ii) Dad is talking on the phone, iii) Mom is not happy. iv) it is dark outside, v) it is raining and vi) it is cold. As before, we label the last three of the 100,000 sensory item neurons, 99,998 99,999 and 100,000. Note that the orange arrow in Fig. 2b represents synaptic connectivity between the *First Event* and the *Second Event*. This illustrates synaptogenesis, which is the formation of synapses between neurons in the nervous system.

### A. Synaptogenesis

The phenomena of synaptogenesis have been difficult for neuroscientists to study. Consider what happens when synaptogenesis fails: Fig. 3a represents that moment of time right before the orange arrow correctly connects synapses together. We represent this moment and illustrate failed synaptogenesis by randomly connecting *First Event* neurons with *Second Event*

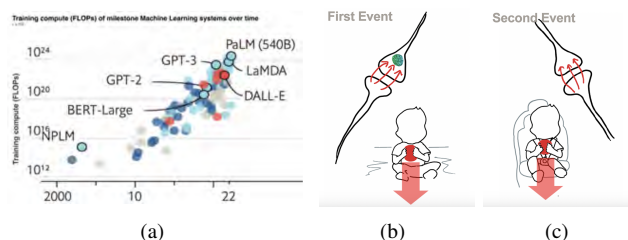


Fig. 1. (a) Increase in computing power demands petaFLOPS [6]. (b) Baby lets go of his red ball. (c) Baby lets go of his red ball again.

synapses. Here, the blue synaptic connection has incorrectly connected the Living Room to the Car. If one believed this was correct and asked the child; "Since your ball dropped when you were in the Living Room it obviously will not drop when you're in the Car, it may go to the ceiling, right?" The child would think that was absurd. Similarly, this paper shows why present-day AI systems on both von Neuman and Neuromorphic chips are NOT autonomous. In other words, both systems still need humans to train the AI to learn. For example, Living Rooms and Cars have nothing to do with the ball dropping. It is important to remember that humans, coding the AI, would have to train the AI to learn that the purple line in Fig. 2a connecting Carpet with Mom Not Happy is wrong. Meaning, we would have to train our AI that Birds Singing with Sound of Rain and 100,000 with 99,998 have nothing to do with the ball dropping.

**EXAMPLE 2.** Consider a video on YouTube called "Donkeys laughing at a Dog that Electrocutes Himself". Fig. 4b shows that when the dog goes up to the donkeys, his nose touches an electric fence, and he is shocked. Intuitively, we know that the next time the dog passes the fence, he will not think: "It's dark now, or there's no donkeys now, or it's raining, so now I can touch the fence!" As absurd as this sounds, we humans have to code even the best AI systems to ignore millions of these unrelated states and synapses.

**B. IBM's bump**

When TrueNorth engineers and neuroscientists from International Business Machines Corporation (IBM) studied how to emulate synaptogenesis, as illustrated in Fig. 2b, they assumed that the voltage inside the orange arrow was that of a typical sine wave, the fundamental waveform they'd seen many times in electroencephalograms (EEGs) and electrocardiograms (EKGs), from which other waveforms such as Gaussian curves may be generated. Here, they developed a modular approach to map bio-inspired excitatory and inhibitory conductance-based neural elements onto hardware [7] as illustrated in Fig. 3 where we see how our orange arrow connects two neurons. Here, an incoming spike signal arrives from the red horizontal axons, and is collected at the end of the orange arrow by the red vertical dendrites.

However, while Defense Advanced Research Projects Agency (DARPA) and IBM celebrated TrueNorth's ability to run at a very low rate of power, these Axon-Hillock neurons inside the orange arrow were not connecting correctly [8]. Schmidt & Avitabile found that TrueNorth's sinusoidal wave's orange arrow was randomly connecting, as illustrated in Fig.

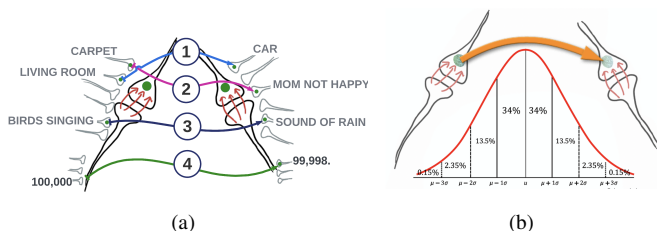


Fig. 2. (a) FAILED Synaptogenesis. Neurons randomly connect with Second Event synapses. (b) IBM's assumed signal over the neuron.

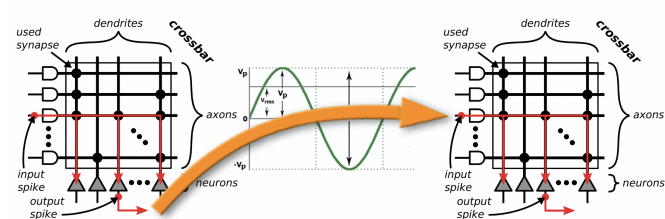


Fig. 3. Assumed signal onto spiking neuromorphic TrueNorth hardware [11].

2a, to any neuron in the fully connected layer [9]. This explains why IBM's engineers were either: i) using a huge number of cores as splitters to implement this fanout as shown in Fig. 4a or ii) adding additional hardware resources to rearrange the 3D convolutional layers [10].

**C. Houston, IBM Has a Problem**

The authors were intrigued, and went back to Fig. 1 and asked themselves: "How does the neuronal path of the red ball neuron, illustrated by the orange arrow, know that it is going to disregard: carpet, living room, Mom's mood, etc. and only connect the correct dendrites?" Surely the answer must lie somewhere in the information on that neuron's sinusoidal wave, that guides it to the correct dendrite.

Going back to 1934, the studies of Hodgkin & Huxley (H&H) seemed like a great place to start reexamining the conceptual framework to understand neuromorphic spike propagation in axons and presynaptic inhibition on spike propagation. Right from the beginning, when H&H clamped an oscilloscope onto a giant squid's neurons, which are about 100 times larger than a human's neuron since they have no skull [13]–[15], the H&H research proved that dendrites are equipped with not one, but many voltages from the Ca<sup>2+</sup> dendritic and axosomatic channels [12]. These channels proved to give rise to local spikes in dendrites and dendritic spines, as illustrated by Larkum *et al*, in Fig. 4c [13], where they examined the timing and cause of a burst from a single Na<sup>+</sup> action potential. Here, H&H observed a biocytin-filled L5 neocortical pyramidal neuron of a rat brain using four electrodes, visible as silhouettes in Fig. 4c [16].

In 1940, Curtis & Cole [15] continued H&H's experiments

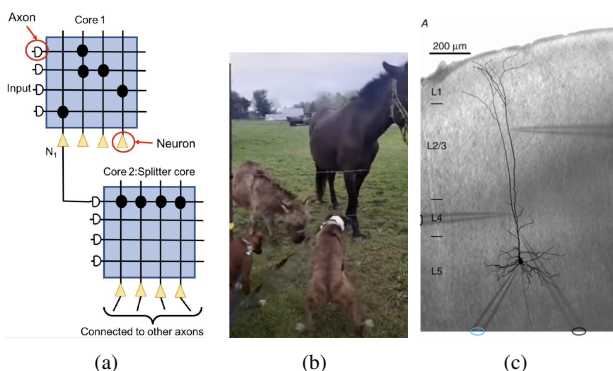


Fig. 4. (a) Splitters on TrueNorth for increasing a neuron's fan-out [9]. (b) Shocked. (c) Neocortical pyramidal neuron of rat brain [12].

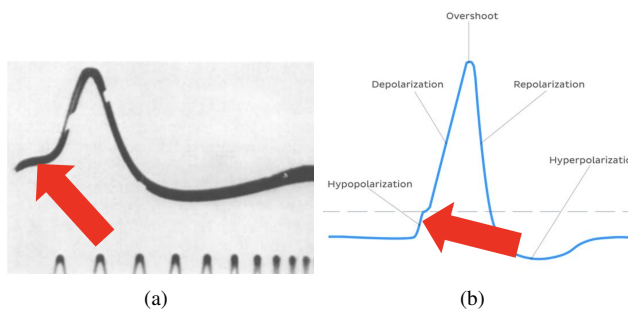


Fig. 5. (a) Membrane potentials of the squid axon from a capillary electrode. (b) hypopolarization.

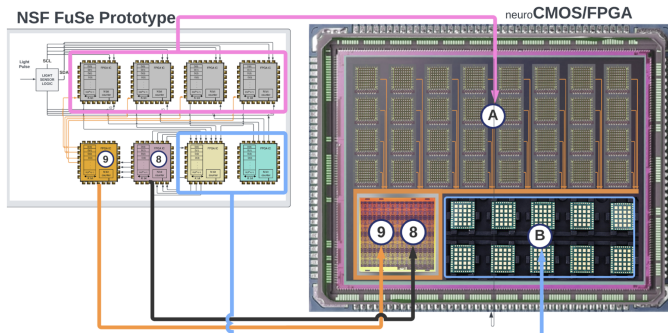


Fig. 6. Comparison of the architecture of the prototype and neuroCMOS/FPGA.

of measuring the voltage across a neuron by inserting a small electrode through the membrane of a giant squid’s axon, and then into the squid’s axoplasm. They found that this yielded a relatively small amount of injury to the axon. Then, they measured the potential between the inside and outside electrode, yielding the result shown in Fig. 5a. It became clear that during the three phases of the action potential (depolarization, overshoot, repolarization) we do indeed witness a sinusoidal wave. However, as shown in Fig. 5b, there is a short ‘hidden state’ [17] of hypopolarization, which precedes the depolarization, that forms a very small bump. This, in a sense, piggy-backs on the trailing edge of the sinusoidal-biological wave, as indicated by the red arrows in Fig. 5a & b. Additionally, we note that in Fig. 7a, this addition of the same small bump ① onto the sinusoidal wave on our bio-inspired chip ②, corrects the sinusoidal wave in Fig. 2b, and was most likely overlooked by IBM’s engineers.

**PRELIMINARIES** Fig. 6 illustrates how our prototype on the left side, called NSF FuSe Prototype, is being converted into our neuroCMOS/FPGA architecture on the right. The prototype’s Grey State FPGAs, highlighted by the pink box, are replaced by section A in the neuroCMOS/FPGA architecture that is comprised of 32-memristors. This means the chip can autonomously re-wire neurons (FPGAs) between 32 states, whereas our prototype only had four states. Additionally, the neuroCMOS/FPGA architecture is comprised of neuromorphic controllers based off the Linearized Hodgkin-Huxley circuits designed in 1934 [18]. Note that the FPGAs highlighted by the blue box, execute the trivial function of only allowing an input from a sensor to move to the next available memristor,

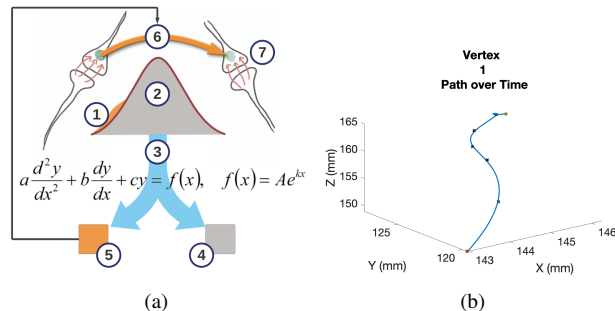


Fig. 7. (a) 2nd Order Diff Eq conversion. (b) 3D graphic shows the path the first vertex takes through the brain from 1 month (pentagon) to 24 months(hexagon)

and then order it in an  $n \times n$  sequence in the memristor itself, located in A. The realization of this small bump, shown as ① in Fig. 7a, forms the basis for our hypothesis for this research effort. Taking the 2nd order Differential Equation ③ of both ① and ② we split the combined area under the curve into grey area ④ that constitutes the power necessary to project the neuron’s signal ⑥ across the orange arrow to its destination ⑦, while the orange area ⑤ is what we believe, carries the neuron’s ROAD MAP that guides it to its destination ⑦.

**THE ISSUE FOR THIS PAPER** is that we do not know: i) how many FPGAs we will need to sort and order the data in each memristor and ii) how we need to design the neuroCMOS/FPGA architecture to expand its knowledge and neuronal connection in a way that will mimic how brains in nature grow and expand. Additionally, to complicate the issue, the ‘FPGA ⑧ & ⑨’ in the neuroCMOS/FPGA architecture is not a square unit but rather a series of FPGA gates. These gates do both the matrix mathematics and the synaptogenesis of rewiring the connectors in the neuroCMOS/FPGA chip.

## II. EXPERIMENTS

### A. Hypothesis

Our hypothesis is that the optimal means to design the autonomous addition of synapses (FPGAs) is to leverage our research measuring how the human brain expands and grows connectomes in the infantile brain.

### B. Mathematically Defining Connectome Growth

To code how our neuroCMOS/FPGA will autonomously expand, we first need to mathematically define connectome growth in the infantile brain. To accomplish this, we continue our research from our last article in the BrainInfo2023 [19]. We bring in the MatLab poly5 fittings for the X, Y and Z longitudinal values of the first vertex over 1, 3, 6, 9, 12, 18 and 24 months. Using the fitted curves for the coordinates over time yields a very nice path as seen in the Fig. 7b. We use these functions to build the unit tangent vector as the directional vector for the directional derivative. Usually, the directional derivative is built with the maximized gradient on the surface, but we do not have a surface with the data from the baby connectome data. The best we can build is the unit tangent vector which we substitute in for the maximized gradient. The functions for the smoothed curves of X from "poly5" are fifth degree polynomials with coefficients:  $p1 =$

1.742e-05, p2 = -0.001163, p3 = 0.02857, p4 = -0.3194, p5 = 1.702, p6 = 141.5. This yields a function for the x variable of the first vertex position in eq.(1).

$$\begin{aligned} X(t) &= p1 * t^5 + p2 * t^4 + p3 * t^3 + p4 * t^2 + p5 * t + p6 \\ X(t) &= 1.742e - 05 * t^5 + -0.001163 * t^4 + 0.02857 * t^3 \\ &\quad + -0.3194 * t^2 + 1.702 * t + 141. \end{aligned} \quad (1)$$

The functions for the smoothed curves of Y from "poly5" are fifth degrees polynomials with the following coefficients: p1 = -9.346e-07, p2 = 7.466e-05, p3 = -0.0008573, p4 = -0.05254, p5 = 1.396, p6 = 118.2.

Next, we used the unit tangent to find the directional derivative over 1 month to 24 months in one tenth of a month interval. This directional derivative is a better representation of the growth rate than the previously presented derivatives in each of the axis (X, Y, Z). With the first data collection at one month, the first growth rate we observe is approximately 3.46 mm. The growth rate then spikes back up to 0.482478 mm/month at 24 months of age. This same analysis can be performed on any of the 163,842 vertices tracked in the diffusion tensor magnetic resonance imaging (DT/MRI). This directional derivative could then be curve fitted to provide coefficients of growth to represent a particular vertex for a particular patient with specific traits. Using the baby connectome data from the human connectome project we have devised a method to provide coefficients of growth. These coefficients could then be used to find correlation between growth and behavioral traits. There is most likely some correlation between the coefficients and DNA. Of course, we will need long term DT/MRI data from thousands of patients over the first two years of life to obtain the big data necessary for machine learning.

### C. Logistical Growth Function Solved for $\rho$

From differential equations [20] we have the logistical growth model. Let us first review exponential growth. The rate of change of a population's growth is dependent on the current population. Here  $y$  is the population and  $dy/dt$  is the rate of population change over time as seen in (2). Replacing the  $f(y)$  with  $r$  representing the rate of growth that is proportional to the population, it yields (3).

$$\frac{\partial y}{\partial t} = f(y) \quad (2) \quad \frac{\partial y}{\partial t} = ry \quad (3)$$

Note that for infantile brain growth we will only examine positive  $r$ , rate of growth. Divide both sides by  $y$  and multiple both sides by  $\partial t$  yields (4), then after integrating both sides we get (5) to which we take the exponential of both sides and yield (6).

$$\frac{\partial y}{y} = \frac{r}{\partial t} \quad \ln(y) = rt + c \quad y = e^{rt+c} = e^c e^{rt} \quad (4) \quad (5) \quad (6)$$

Going to the logistic growth, we replace the  $r$  with a function of  $y$ ,  $h(y) = r$ . We need to choose  $h$  such that when  $y$  is either small or large,  $h(y) > 0$  it reflects the start of growth,

and limits factors of starvation. Here,  $h(y) = (r - ay)$  satisfies these conditions. Applying applicable algebra in (7 & 8).

$$h(y) = r(1 - \frac{ay}{r}) \quad (7) \quad \frac{\partial y}{\partial t} = r(1 - \frac{ay}{r})y \quad (8)$$

We let  $k = \frac{a}{r}$  where  $r$  is intrinsic growth rate and  $K$  becomes the equilibrium for a sustained population greater than zero. Note that we will not attain  $K$  as we know the brain continues to grow after two years of age as shown in (9), and with zero growth in (10).

$$\frac{\partial y}{\partial t} = r(1 - \frac{y}{K})y \quad (9) \quad 0 = r(1 - \frac{y}{K})y \quad (10)$$

This is satisfied when either  $y = 0$  or when  $(1 - \frac{y}{K})$  is zero, hence  $y = K$ . We know that at birth, the brain has some volume of neurons, therefore we will not consider  $y = 0$  or a population of zero. Dividing both side by  $(1 - \frac{y}{K})y$  and multiplying both sides by  $\partial t$ :

$$\frac{\partial y}{(1 - \frac{y}{K})y} = r\partial t \quad (11)$$

Perform Partial Fractions

$$\begin{aligned} \frac{1}{(1 - \frac{y}{K})y} &= \frac{A}{y} + \frac{B}{(1 - \frac{y}{K})} \quad (12) \\ \frac{1}{(1 - \frac{y}{K})y} &= \frac{A}{y} * (\frac{1 - \frac{y}{k}}{1 - \frac{y}{k}}) + \frac{B}{(1 - \frac{y}{K})} * \frac{y}{y} \\ A * (1 - \frac{y}{k}) + B * y &= 1 \end{aligned}$$

letting  $\frac{y}{k} = 1$  or  $y = k$  we get  $B * y = 1$ , and then dividing by  $y$  and substituting  $k$  for  $y$  we get  $B = \frac{1}{k}$  and then substituting back into (13 & 14),

$$A * (1 - \frac{y}{k}) + \frac{y}{k} = 1 \quad (13) \quad A = \frac{(1 - \frac{y}{k})}{(1 - \frac{y}{k})} = 1 \quad (14)$$

It yields (15) and now we integrate both sides by breaking the left side apart as seen in (16)

$$(\frac{1}{y} + \frac{\frac{1}{k}}{(1 - \frac{y}{k})})\partial y = r\partial t \quad (15) \quad \int \frac{1}{y} \partial y = \ln|y| \quad (16)$$

for (17), then using a  $u$  substitution we get (18).

$$\int \frac{1}{(1 - \frac{y}{k})} \partial y \quad (17) \quad \partial u = -\frac{1}{k} \partial y \quad (18)$$

by the chain rule, with  $g(x) = uf(u) = \frac{1}{u}$

$$\int f(g(x))g'(x) = \int f(u)\partial u = \ln|u| = \ln|1 - \frac{y}{K}|$$

therefore

$$\int (\frac{1}{y} + \frac{\frac{1}{k}}{(1 - \frac{y}{k})})\partial y = \int r\partial t \quad (19)$$

becomes

$$\ln|y| - \ln|1 - \frac{y}{K}| = rt + c \quad (20)$$

otherwise known as the logistical growth model.

Note that the neuron growth does not depend on the number of neurons as population growth does. Rather, neuron growth depends on the number of radial glial cells which are dependent on the number of neuroepithelial cells [21]. The number of these different cell types are unavailable at this time. Thus, our challenge is to build a logistical brain growth model that shows the dependencies on the different cell populations and to do this we move from the logistical growth function to the Reaction Diffusion Equation.

From the Mathematical Biology text [22] and Konukoglu *et al.* [23] we have:

$$\frac{\partial u}{\partial t} = \nabla \cdot (D\nabla u) + \rho u(1 - u) \quad (21)$$

with  $D\nabla u \cdot \vec{n}_{\rho\Omega} = 0$  Where  $u$  being density,  $D$  is the Diffusion tensor,  $\rho$  is the proliferation rate,  $\Omega$  the brain domain, and  $\rho\Omega$  the brain boundaries [23] "The traveling wave solution of Equation 21 has the form

$$u(x, t) = u(x - vt) = u(\mathcal{E}) \quad (22)$$

where  $\mathcal{E}$  is the moving frame and  $v$  is the asymptotic speed of this frame, the wavefront. When this solution is plugged into the reaction-diffusion equation 21 we obtain the ordinary differential equation" known as the Eikonal Equation.

$$\mathbf{n}' D \mathbf{n} \frac{\partial^2 u}{\partial \mathcal{E}^2} + v \frac{\partial u}{\partial \mathcal{E}} + \rho u(1 - u) = 0 \quad (23)$$

The Eikonal equation has been used to describe brain tumor growth [23]. Konukoglu *et al.* discerned the growth rate difference of brain tumors in white matter vs. grey matter. Since DT-MRI do not provide density Konukoglu *et al.* switched their parameter to the moving front of the tumor cell. With the data from the Baby Connectome Project (BCP), we have polygons (triangles) of area. We also have no diffusion data. And we see that the area of the triangles are increasing. We make the presumption that there is no diffusion. Eliminating the diffusion term from the reaction diffusion equation yields the logistical growth function (24).

$$\frac{\partial A}{\partial t} = -\rho A \left(1 - \frac{A}{K}\right) \quad (24)$$

The same result can be obtained by using the *Fisher-Kolmogoroff equation* from [24] [22] and setting the diffusion term to zero. With  $K$ (capacity) being the max surface area attained. For this work, we are only looking at the growth from 1 month to two years. The max surface area will be the surface area at two years of age. We attain  $\frac{\partial A}{\partial t}$  by curve fitting the surface area over time and taking the derivative. Then we can attain  $\rho$ , the growth function for the area of the entity being investigated. Dividing both sides by  $A \left(1 - \frac{A}{K}\right)$

$$\frac{\frac{\partial A}{\partial t}}{A \left(1 - \frac{A}{K}\right)} = -\rho \quad (25) \quad \rho = -\frac{\frac{\partial A}{\partial t}}{A \left(1 - \frac{A}{K}\right)} \quad (26)$$

Thus, we have produced a differential equation  $\rho(t)$  equal to the first derivative of brain surface area over time divided

by a function of surface area over time. Several neuroscientists have noted that different lobes and white matter pathways develop at different times and rates. With this methodology we can construct the logistical growth functions,  $\rho_i(t)$ , to reflect those differences by constructing a sum of  $\rho_i(t)$  where  $t$  is all the triangles of a given lobe. Brain growth is reflected by several measures, those being increasing surface area, increasing volume and vertex movement through the skull which is also growing. The grey matter has 14 layers which comprise the grey matter lobes. The 42 white matter pathways grow underneath the gray matter layers. For each of these measures, we define the characteristic to describe the individual lobe/pathway to be time dependent, hence the characteristics to be solved for are:

$$C_i(t) = \rho_i(t), \forall i = \text{lobes, pathways} \quad (27)$$

where  $i$  represents the brain lobe/pathway under consideration.

#### D. Logistical Growth Function Applied to Brain Surface Area Growth

The polygons are actually triangles. We found the polygons are stable over time, and they have the same vertices at the seven times utilized in this work. The numbered polygons, 327,680 of them, have the same vertices, in the same order, for all seven times, 1, 3, 6, 9, 12, 18 and 24 months of age. Since the brain is growing, we expect the area of the polygons will grow over time. We use the three-dimensional distance formula to obtain the length of each of the three sides of the triangle in (28). We then use Heron's Formula [25] (28).

$$l = \sqrt{(X1 - X2)^2 + (Y1 - Y2)^2 + (Z1 - Z2)^2} \quad (28)$$

Rather than calculating the base and the height of the triangle, Heron's formula is simplistic in its three subtractions, three multiplications and one square root in (29).

$$\text{area} = \sqrt{s * (s - a) * (s - b) * (s - c)} \quad (29)$$

We examined the first polygon in the Visualization Toolkit (VTK) files. It consists of vertices 1, 40965, and 40963. Over time we found the areas were 0.0312, 0.0426, 0.0555, 0.0606, 0.0648, 0.0665, 0.0730. And yes, the area of the polygon is increasing over time. We apply curve fitting methodology for polygon 1 area. Polygon 1 consists of vertices 1, 40965, and 40963. These positions are subsequently used for the monthly area calculations. The curve fitting of the Area over time produced a piecewise solution of six equations. The first equation is used from one month to three months. The second equation is used from three months to six months and so forth. Third equation, six to nine months, forth, nine to twelve months, fifth, twelve to eighteen months, sixth, eighteen to twenty-four months. These equations were used to plot the line in Fig. 8 where each equation is of the form  $a * t^3 + b * t^2 + c * t + d$  yielding the coefficients for the six equations for the resulting coefficients for the piece-wise solution to the curve fitting of actual polygon 1 area at 1, 3, 6, 9, 12, 18, and 24 months, as seen in Table I.

TABLE I  
COEFFICIENTS FOR THE SIX EQUATIONS

eq #	a	b	c	d	months
1	-3.2754e-05	0	0.0052	0.0321	1-3
2	-1.9036e-05	-1.9653e-04	0.0048	0.0422	3-6
3	2.3455e-05	-3.6785e-04	0.0031	0.0544	6-9
4	3.3306e-06	-1.5676e-04	0.0016	0.0612	9-12
5	1.2741e-05	-1.2678e-04	7.0962e-04	0.0645	12-18
6	-5.6974e-06	1.0255e-04	5.6424e-04	0.0670	18-24

To build the six equations for area growth we start with the first derivative yielding growth rate. Each equation is of the form:  $a*t^2 + b*t + c$  yielding coefficients for the six equations as shown in Table II.

TABLE II  
FIRST DERIVATIVE YIELDING GROWTH RATE

eq #	a	b	c	months
1	3*-3.2754e-05	0	0.0052	1-3
2	3*-1.9036e-05	2*-1.9653e-04	0.0048	3-6
3	3*2.3455e-05	2*-3.6785e-04	0.0031	6-9
4	3*3.3306e-06	2*-1.5676e-04	0.0016	9-12
5	3*1.2741e-05	2*-1.2678e-04	7.0962e-04	12-18
6	3*-5.6974e-06	2*1.0255e-04	5.6424e-04	18=24

With the function for Area over time and the first derivative for Area over time, we can now solve for  $\rho$ , the logistical growth function of Area over time. Using the definition of  $\rho$  from the previous section we have:

$$\rho = -\frac{\frac{\partial A}{\partial t}}{A(1 - \frac{A}{K})} \quad (30)$$

where  $K$  is the area of the polygon at max growth in Fig. 8a.

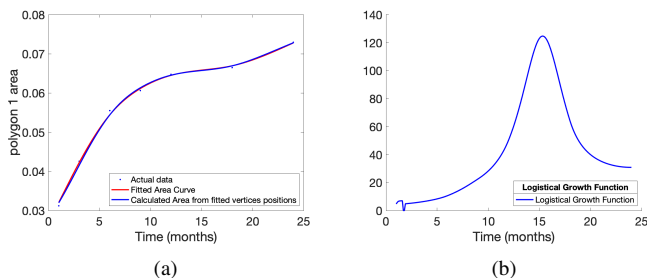


Fig. 8. (a) The difference between the actual areas (points) and fitted curve for area and the calculated area from the fitted vertices over time. (b) The Logistical growth function for the polygon 1 area of the brain from one month to twenty-four months of age.

### III. CONCLUSION & FUTURE WORK

Because our neuroCMOS/FPGA is in an "infantile" state, we barely have to go beyond the infantile state for a human. Here we have concluded that, with this statement in mind, we can set  $K$  to a number greater than the area obtained at two years of age, avoiding the divide by zero anomaly. We currently set  $K$  to twice the polygon 1 area at two years of age and plot this logistical growth function from one month to twenty-four months as shown in Fig. 8b. Next, we model the aforementioned onto a simulated software version of our neuroCMOS/FPGA, that will show we can procure a chip that rewires itself and grows like an infantile's brain grows.

### REFERENCES

[1] D. Ivanov, A. Chezhegov, M. Kiselev, A. Grunin, and D. Larionov, "Neuromorphic artificial intelligence systems," *Frontiers in Neuroscience*, vol. 16, p. 1513, 2022.

[2] A. Mehonic and A. J. Kenyon, "Brain-inspired computing needs a master plan," *Nature*, vol. 604, no. 7905, pp. 255–260, 2022.

[3] P. Pietrzak, S. Szczęśny, D. Huderek, and Ł. Przyborowski, "Overview of spiking neural network learning approaches and their computational complexities," *Sensors*, vol. 23, no. 6, p. 3037, 2023.

[4] "Brain basics: The life and death of a neuron," 2019, [Online; accessed 2024-01-24]. [Online]. Available: <https://www.ninds.nih.gov/health-information/public-education/brain-basics/brain-basics-life-and-death-neuron>

[5] N. Banerjee, "Neurotransmitters in alcoholism: A review of neurobiological and genetic studies," *Indian journal of human genetics*, vol. 20, no. 1, p. 20, 2014.

[6] C. Szegedy *et al.*, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.

[7] M. P. Löhr, C. Jarvers, and H. Neumann, "Complex neuron dynamics on the ibm trueneuroth neurosynaptic system," in *2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. IEEE, 2020, pp. 113–117.

[8] R. A. Nawrocki, R. M. Voyles, and S. E. Shaheen, "A mini review of neuromorphic architectures and implementations," *IEEE Transactions on Electron Devices*, vol. 63, no. 10, pp. 3819–3829, 2016.

[9] H. Schmidt and D. Avitabile, "Bumps and oscillons in networks of spiking neurons," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 3, 2020.

[10] R. Shukla, M. Lipasti, B. Van Essen, A. Moody, and N. Maruyama, "Remodel: Rethinking deep cnn models to detect and count on a neurosynaptic system," *Frontiers in neuroscience*, vol. 13, p. 4, 2019.

[11] P. R. Huttenlocher and A. S. Dabholkar, "Regional differences in synaptogenesis in human cerebral cortex," *Journal of comparative Neurology*, vol. 387, no. 2, pp. 167–178, 1997.

[12] Y. Manor, J. Rinzel, I. Segev, and Y. Yarom, "Low-amplitude oscillations in the inferior olive: a model based on electrical coupling of neurons with heterogeneous channel densities," *Journal of neurophysiology*, vol. 77, no. 5, pp. 2736–2752, 1997.

[13] C. Meunier and I. Segev, "Playing the devil's advocate: Is the hodgkin-huxley model useful?" *TRENDS in Neurosciences*, vol. 25, no. 11, pp. 558–563, 2002.

[14] B. L. d'Incamps, C. Meunier, D. Zytnecki, and L. Jami, "Flexible processing of sensory information induced by axo-axonic synapses on afferent fibers," *Journal of Physiology-Paris*, vol. 93, no. 4, pp. 369–377, 1999.

[15] H. J. Curtis and K. S. Cole, "Membrane action potentials from the squid giant axon," *Journal of Cellular and Comparative Physiology*, vol. 15, no. 2, pp. 147–157, 1940.

[16] M. E. Larkum, J. J. Zhu, and B. Sakmann, "Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons," *The Journal of physiology*, vol. 533, no. 2, pp. 447–466, 2001.

[17] "Action potential," [Online; accessed 2024-01-24]. [Online]. Available: <https://www.kenhub.com/en/library/anatomy/action-potential>

[18] L. Ribar and R. Sepulchre, "Neuromorphic control: Designing multiscale mixed-feedback systems," *IEEE Control Systems Magazine*, vol. 41, no. 6, pp. 34–63, 2021.

[19] M. Bihn and R. Lewis, "A postulate: Connectome development is the driving factor of brain growth," pp. 12–15, 2023, [Online; accessed 2024-01-24]. [Online]. Available: <https://shorturl.at/IOWZ7>

[20] W. E. Boyce and R. C. DiPrima, *Elementary differential equations and boundary value problems*. Wiley, 2020.

[21] M. Götz and W. B. Huttner, "The cell biology of neurogenesis," *Nature reviews Molecular cell biology*, vol. 6, no. 10, pp. 777–788, 2005.

[22] J. D. Murray, "Vignettes from the field of mathematical biology: the application of mathematics to biology and medicine," *Interface Focus*, vol. 2, no. 4, pp. 397–406, 2012.

[23] E. Konukoglu *et al.*, "Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 77–95, 2009.

[24] J. D. Murray, "Mathematical biology ii: Spatial models and biomedical applications," *Monographs on Applied and Computational Mathematics*, vol. 3, pp. 399–400, 1973.

[25] R. B. Nelsen, "Heron's formula via proofs without words," *The College Mathematics Journal*, vol. 32, no. 4, p. 290, 2001.

# A Comparative Analysis of Episodic Memory between Humans and AI Agents with Context Correlation

Shweta Singh

IIIT

Hyderabad, India

shweta.singh@research.iiit.ac.in

Anant Pratap Singh

VIT

Chennai, India

anantpratap.singh2021@vitstudent.ac.in

Vedant Ghatnekar

MIT WPU

Pune, India

1032190997@mitwpu.edu.in

**Abstract**—This study delves into a comparative analysis of episodic memory, examining both human cognition and Artificial Intelligence (AI) agents. Through an in-depth exploration, the research focuses on the nuanced aspects of episodic memory encoding, retrieval, and associative capabilities in humans and AI systems. The investigation incorporates Electroencephalography (EEG) as a fundamental tool to comprehend and compare the underlying neural mechanisms associated with episodic memory in humans while drawing parallels to memory processes in AI agents. The findings illuminate similarities and disparities, shedding light on the cognitive frameworks and technological advancements shaping episodic memory across biological and artificial entities. This exploration provides valuable insights into the convergence and divergence of memory mechanisms, potentially influencing future AI developments and understanding human cognition.

**Index Terms**—EEG-Electroencephalography, Episodic Memory, Human Cognition, Artificial Intelligence (AI) Agents, Comparative Analysis, Memory Retrieval, Context.

## I. INTRODUCTION

Episodic memory stands as a pivotal facet of human cognition, representing the ability to recall specific past events, experiences, and their contextual details within a personal timeline. It encompasses the richness of autobiographical memory, allowing individuals to mentally travel back in time and relive moments while integrating sensory perceptions, emotions, and spatial-temporal context. This unique cognitive ability enables humans to navigate daily life, learn from past experiences, and project themselves into the future, forming the cornerstone of our identity and decision-making processes.

Concomitant with advancements in AI, the emergence of AI agents equipped with memory systems presents a paradigm shift in technological capabilities. These agents, ranging from sophisticated chatbots to complex neural networks, are designed to mimic cognitive processes, including memory encoding, retrieval, and learning. AI memory frameworks, though algorithmically driven and fundamentally distinct from human cognition, are pivotal in enabling these agents to retain and utilize information, make decisions, and perform tasks across various domains.

This paper aims to undertake a comparative analysis of episodic memory, focusing on the influence of contextual fac-

tors on memory encoding, retrieval, and associative processes in both humans and AI agents. It delineates the impact of context on memory mechanisms, leveraging EEG as a tool to probe neural correlates associated with episodic memory in humans and explore parallels or distinctions in AI memory frameworks. The paper is structured to first delve into the nuances of episodic memory in humans, subsequently transitioning to the emerging landscape of AI memory systems. Through a comparative lens, it examines context-mediated memory effects and EEG correlations, ultimately aiming to elucidate the convergence and divergence between human cognition and AI memory mechanisms.

## II. EPISODIC MEMORY IN HUMANS

### A. Memory Encoding and Context Effects

Memory encoding in humans involves the initial processing of sensory information into a form that can be stored and later retrieved. This process occurs through various stages, including attention, perception, and consolidation, where information is integrated into existing memory networks.

Contextual cues, encompassing environmental, emotional, and situational factors, play a pivotal role in memory formation. The encoding specificity principle posits that retrieval of information is most effective when the context at encoding matches the context at retrieval. This principle underscores the significance of contextual congruence in memory formation and recall.

Several studies employing EEG have revealed insights into context effects on human episodic memory. For instance, research [3], showcased increased neural synchrony in specific brain regions during memory encoding when contextual cues were present, emphasizing the influence of context on neural patterns associated with memory formation.

### B. Memory Retrieval and Contextual Influences

Memory retrieval involves accessing stored information from memory networks. Context plays a pivotal role in triggering recall by acting as retrieval cues [1], [2], facilitating the retrieval of associated memories when the context at recall aligns with the context at encoding.

Studies investigating context-dependent memory retrieval in humans have consistently demonstrated the impact of context on recall. EEG studies revealed distinct neural signatures during context-induced memory retrieval, highlighting the role of neural oscillations and synchronization in retrieving context-linked memories.

### C. Associative Memory and Contextual Linkages

Associative memory involves forming connections between different pieces of information. Contextual information acts as a binding factor, strengthening associations between items encoded within a similar context.

EEG studies exploring neural correlates of associative memory in context-rich environments have unveiled patterns of neural activation in specific brain regions, elucidating the neural mechanisms underlying the influence of context on associative memory processes. For instance, research has revealed increased coherence between brain regions associated with contextual processing and memory association tasks.

## III. EPISODIC MEMORY IN AI AGENTS

### A. Memory Encoding Mechanisms

Memory Encoding mechanisms in AI agents primarily rely on structured databases and algorithms. Databases store information in a structured format, enabling efficient retrieval and manipulation. Algorithms manage the Encoding, organization, and retrieval of data, utilizing various techniques such as indexing, hashing, and neural network architectures for memory representation.

Integrating contextual information into AI memory frameworks poses significant challenges. AI agents traditionally process information based on predefined patterns and lack the inherent contextual understanding characteristic of human cognition. Challenges include contextual ambiguity, dynamic context changes, and the computational complexity of incorporating multifaceted contextual cues. However, integrating contextual information offers potential benefits, enhancing the adaptability, relevance, and decision-making capabilities of AI systems.

### B. Memory Retrieval and Contextual Integration

The AI agents retrieve information from stored data using algorithms tailored for efficient search and retrieval. Context plays a crucial role in retrieval algorithms, aiding in narrowing down search results or providing relevant cues for retrieving associated information. Contextual integration involves algorithms that utilize contextual cues to refine retrieval processes, akin to humans using context as retrieval cues.

Ongoing research and developments aim to imbue AI memory systems with contextual awareness. For instance, advancements in Natural Language Processing (NLP) incorporate contextual embeddings or attention mechanisms, allowing AI models to consider contextual information in text-based tasks. Additionally, research in machine vision explores contextual understanding in image recognition tasks by leveraging spatial

and semantic context to improve object recognition and scene understanding in AI systems.

For our experimentation, we use a TransformerXL backbone by [6] and modify its sequential memory buffer with Automatic Chunking [7] to enable the transformer to apply attention to only relevant parts of memory depending on the current context that might not always be sequential. The model architecture is described in Section VI.

## IV. EEG DATA COLLECTION AND ANALYSIS

EEG data was collected using an Emotiv FLEX EEG cap, which features 32 channels for recording neural activity during encoding and retrieval phases in human participants. This cap was equipped with monopolar gel-based electrodes strategically positioned across the scalp to capture electrical signals emanating from various regions of the brain. The electrode positions were determined according to the 10-20 international system for EEG electrode placement (Figure 1), ensuring standardized and precise positioning for data acquisition.

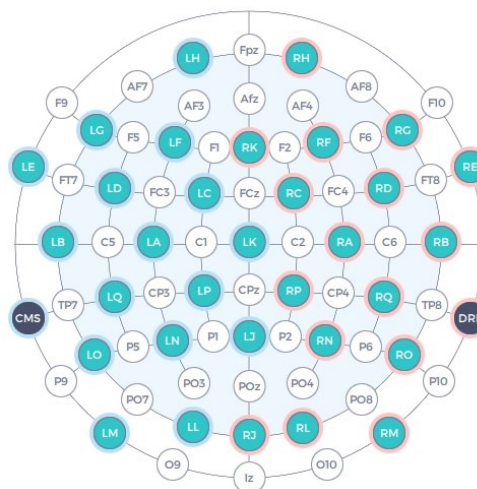


Fig. 1: EEG Electrode Placement.

### A. Captured EEG Data

The EEG data was captured using the CyKit software [11] during the encoding and retrieval phases of the experiment. To provide a comprehensive understanding of the analyzed brain activity, this section details crucial information regarding the data acquisition process. The recording duration for encoding phase was 60-70 seconds while for decoding phase the subjects were not bound for time intervals, allowing for analysis of the temporal dynamics of brain activity. Additionally, a high-pass filter with a cutoff frequency of around 0.16 Hz and notch filters at 50 Hz and 60 Hz to remove power line noise interference was applied to the data using CyKIT, focusing on the specific frequency band of interest. Moreover, the data is downsampled to 128 Hz before transmission. Finally, the analysis incorporates data from a single trial from the encoding and retrieval phases, enhancing the generalizability of the findings.



### B. Topographical Brain Mapping

The topographical brain activity maps were generated for EEG data obtained from the participants using resources from [10] as a baseline. This code repository offered tools and utilities specifically designed for EEG data processing in Python, facilitating the creation of detailed and informative brain activity maps. The generated maps serve as valuable tools for analyzing and interpreting the complex patterns of brain activity observed during the experiment, enabling us to gain deeper insights into memory-related cognitive processes. The graphs were generated such that the top data points correspond with the front of the scalp. Please refer to Appendix A and B for the EEG maps plotted during all our experimentation. The graph was rendered to provide a top-down perspective of the head, where the upper regions of the graph represent the front of the head, the left sections correspond to the left-hand side, and the right sections depict the right-hand side. This approach ensured that the spatial orientation of the depicted neural activity aligned appropriately with anatomical references, facilitating a clear and intuitive interpretation of the topographical brain mapping results.

### C. Encoding Phase Analysis

During the encoding phase, EEG data analysis involved assessing neural correlates linked to the processing of contextual cues (such as wall colours) and memory encoding. The topographical brain maps derived from this phase showcased neural activation patterns specific to encoding information within distinct contextual contexts.

### D. Retrieval Phase Analysis

Similarly, during the retrieval phase, EEG data analysis focused on discerning neural signatures associated with memory retrieval and decision-making while navigating the game. The topographical brain maps generated during retrieval indicated neural activity patterns corresponding to successful memory recall and decision-making processes influenced by contextual cues.

### E. Integration of Topographical Maps

The topographical brain maps were generated using the Akima interpolation method, which was chosen due to its effectiveness in facilitating a smoother visualization of spatial distribution. No parameter tuning was performed as default settings were deemed sufficient for the analysis. The time window used for calculating brain activity represented in each topographical map corresponds to the encoding or retrieval phase of the memory task. Specifically, the observation interval spans from the onset of the memory task to the offset of the task period. The topographical brain maps provided visual representations of the neural activation patterns across different scalp regions. Areas exhibiting heightened or suppressed electrical activity were depicted, aiding in the identification of brain regions implicated in context-mediated memory encoding and retrieval [4] [5]. Variations in neural activity across scalp regions were indicative of the brain's

response to contextual cues during memory-related tasks. The transformation from 2D matrices of channels by samples to 2D spatial maps involved several key steps in topographical brain mapping. Initially, each channel within the EEG data corresponded to a specific electrode position on the scalp, known from a standardized electrode montage such as the 10-20 system. Subsequently, contour plotting techniques were applied to visualize the spatial distribution of this interpolated activity, creating a 2D map where different colors or shading indicated varying levels of neural activity across scalp regions. This process enabled us to gain insights into the spatial dynamics of brain function during cognitive tasks or experimental conditions.

### F. Integration with Behavioral Performance

These EEG-derived topographical brain maps were correlated with participants' behavioral performance during the game-based task. The association between neural activation patterns depicted in the maps and the accuracy/speed of memory-related decisions offered insights into the neural mechanisms underlying context-induced effects on episodic memory.

## V. EXPERIMENTS: CONTEXT EFFECTS ON EPISODIC MEMORY IN HUMANS AND AI AGENTS

### A. Experimental Design 1

a) *Objective:* The objective of this experiment was to investigate and compare the impact of contextual cues on episodic memory encoding and retrieval in both human participants and an AI agent model. The study aimed to explore EEG correlations to identify neural signatures associated with context-mediated memory processes in humans and simulate analogous processes within an AI system.

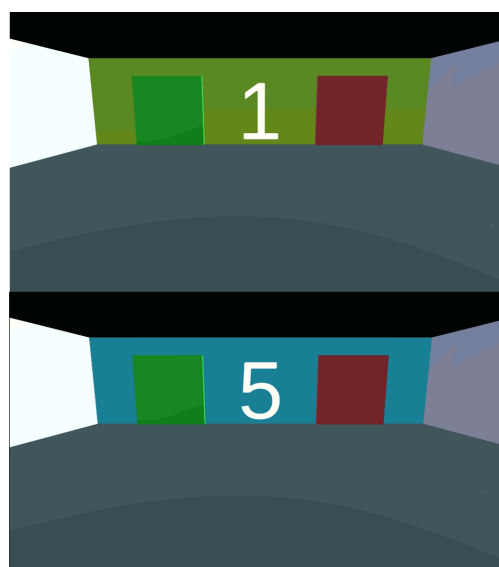


Fig. 2: Experiment 1 Game environment.

b) *Methodology*: Four individuals (aged 20-30) without any neurological disorders participated in this experiment. We utilized a neural network-based AI model that simulates memory processes akin to episodic memory for comparison with human subjects.

c) *Task design*: Participants engaged in a game involving 20 rooms, each containing numbers (1-20) and distinct wall colours (see Figure 2). In each room, there was a red and a green door. Participants had to select one door, aiming to choose the correct door to progress to the next room. Correct door selection allowed advancement to the next room, while an incorrect choice led to a shift back to the previous room.

d) *Encoding Phase*: Participants were explicitly instructed to remember the correct door in each room during the game as part of the encoding phase. They were told the correct door and played through the environment a few times till they felt they had memorized all the doors. The EEG plots of participant A and B were collected to see which regions of the brain would should high activation during encoding (see Figure 9).

e) *Retrieval Phase*: During the retrieval phase, participants replayed the game without explicit instructions, relying on their memory for choosing the correct door in each room. The EEG plots of participant A and B were also collected to see which regions of the brain would should high activation during decoding (see Figure 10).

f) *Contextual Manipulation*: The distinct wall colours in each room served as contextual cues.

B. Experimental Design 2

a) *Objective*: The objective of the second experiment was to explore the influence of diverse contextual cues on context-dependent memory retrieval and episodic memory association. Participants engaged in a game-based task involving various contextual environments to examine the influence of these contexts on memory recall and associative processes.

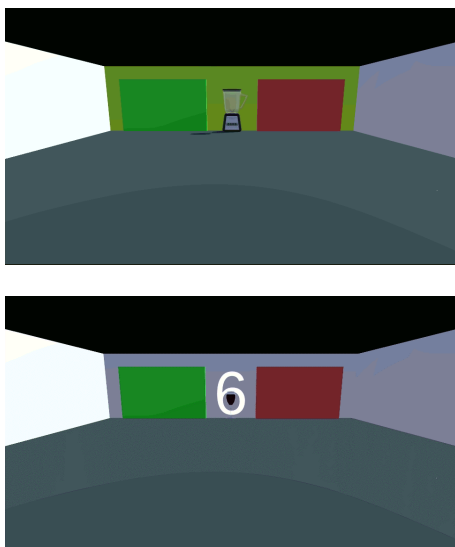


Fig. 3: Game Environment for Experiment 1 containing objects.

b) *Methodology*: The experiment involved two individuals familiar with the game environment from a prior session.

c) *Task Design with Diverse Contextual cues*: Participants navigated through 30 rooms similar to Experiment 1, each designed with specific contextual cues:

- 10 rooms with numbers (1-10).
- 10 rooms with distinct wall colours.
- 10 rooms with attached scenery (no wall colours).

Each room contained 20 random objects and two doors (red and green), requiring participants to choose one to progress (see Figure 3).

d) *Encoding Phase*: During the encoding phase, participants played the game twice while being exposed to varied contextual environments, once with rooms in normal series and once with all rooms shuffled. The objective was to encourage the association of contextual cues with the correct door choice in each room. The EEG plots of both subjects were plotted during encoding (see Figure 11).

e) *Retrieval Phase*: In the retrieval phase, participants were presented with rooms lacking contextual cues (no wall colours, numbers, or scenery). Participants were tasked with recalling the context associated with each room and selecting the correct door choice based solely on their episodic memory. Recall was tested and EEG data measured right after encoding (see Figure 12), 6 hours after encoding (see Figure 13) and 24 hours after encoding (see Figure 14).

VI. MODEL ARCHITECTURE

A. TransformerXL

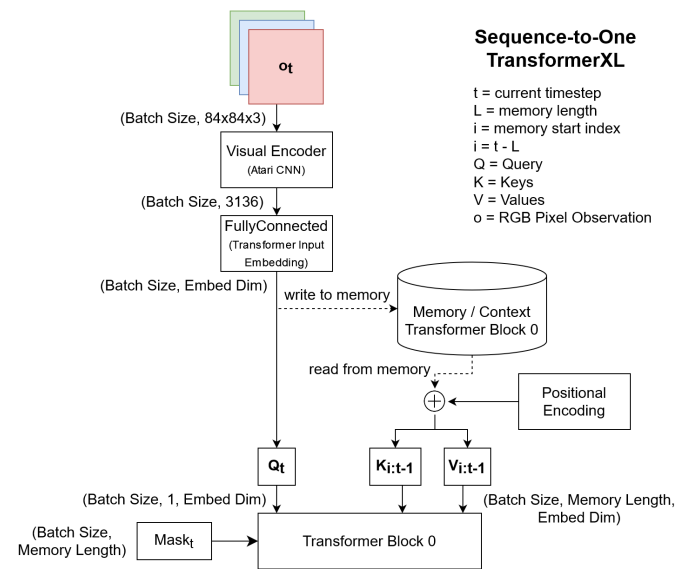


Fig. 4: TransformerXL Architecture [6].

We use a modified decoder-only TransformerXL architecture for testing the performance of AI agents with episodic memory. Ego-centric visual observations are used which are encoded into an embedded representation using a 3-layer Convolutional encoder. The encoded observation is saved in

the memory buffer and used as the query in the Transformer decoder where attention is performed inside the encoder between the memory buffer and the query. A categorical action probability is calculated for each action by applying a linear layer to the output of the decoder. The model weights are trained using Proximal Policy Optimization 2 (PPO2) based on [12] and mini-batch training and the model learns to predict a suitable action given the current observation and the past context. Adam optimizer from [12] is used to decay the learning rate and other PPO2 parameters. Multiple instances of the environments are used to create larger batched data for training.

The memory buffer used in TransformerXL is a simple sequential buffer that stores the encoded observations every timestep. Using all the memories in the buffer during attention in the layer is computationally intensive and scales with the buffer length. All past observations might also not be appropriate in the current context. To combat this, we use Automatic Chunking before the memory buffer is passed to the decoder. This is similar to how human episodic memory is chunked based on certain groups of events that are correlated as discussed previously.

### B. Automatic Chunking Mechanism

Automatic Chunking works on the memory buffer and divides the buffer into chunks of constant size. A summary value is calculated for each chunk using mean pooling. Top-level attention is performed between the summary values and the current observation, and the Top-k chunks are chosen with the highest correlation values. These chunks are then concatenated and used as the summarised memory buffer which is used by the decoder. The summarised memory buffer includes memories that are most relevant in the current context thus leading to better action calculation by the transformer decoder. With Automatic chunking, the TransformerXL memory buffer is modified to act like an Episodic memory buffer due to the fact that it includes egocentric sequences of past events and can recall the most appropriate sequences from it. We can thus compare the TransformerXL model with Automatic chunking with the episodic memory of human subjects.

### C. Tasks

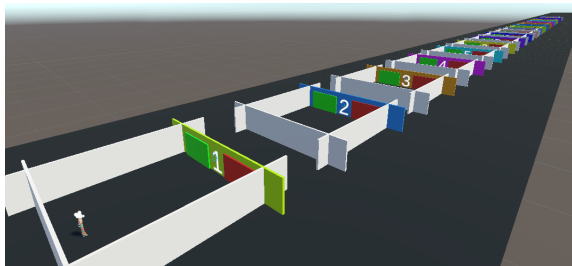


Fig. 5: Game Environment for Experiment 1.

The game environments used in our experimentation are created using Unity MLAgents [10]. Our reinforcement learning models are trained using the Gym API [11] for MLAgents

while the games can be directly played by the human subjects. The AI models were trained until they achieved the maximum reward in the task.

We created two variations of the Experiment 1 game environment used for human testing and AI testing as described later. The observation space of the agent includes the agent's visual observations of size 40x40x3, the agent's position and the current room number of the agent. The rewards were set such that the agent got a positive reward proportional to the room number in each room. If it made a mistake in a room, a negative reward proportional to the room number was given and the agent was teleported back two rooms.

The two variations of the task were:

- The Unshuffled variant: Here, the rooms were in numerical order from 1 to 20 and the correct door colours were fixed (see Figure 5). This variant tested the agent's and player's ability to remember long sequences of information.
- The Shuffled variant: Here, the order of rooms was shuffled during training and testing for every episode. The correct doors for each individual room were fixed. Thus, here, the agent and player needed to learn the correlation between the context i.e., the number of the room and wall colour and the correct door and recall this information during testing non-sequentially.

We trained TransformerXL with and without Automatic Chunking using the same parameters on both variants of the task. The memory length was set to the length of each episode at 500 while the Automatic Chunking parameters used were 10 chunks of size 30. We used default PPO2 parameters except for a changed initial learning rate of  $5 \times 10^{-5}$ . The parameters for Automatic Chunking were decided based on the experimentation by [7] where it was found that chunking the memory and only using around 60% of the memory in the transformer gave the best results in most tasks. Keeping a small chunk size helps in this task as well as the model can access smaller sequences in further apart sections of the memory which is required in this task due to the amount of time spent in a single room is short compared to the total episode length.

The trained models were tested on the same environments for 50 episodes to test whether the model had learnt the proper sequences or mapping in the tasks.

### D. Training Results

a) *Unshuffled Variant*: In Figure 6, we plot the average rewards over the multiple instances of the environments versus the episode number for both models. Both models achieve the maximum reward of 100 in the environment in 400 episodes.

b) *Shuffled Variant*: In Figure 7, we plot the average rewards over the multiple instances of the environments versus the episode number for both models. Both models achieve the maximum reward of 100 in the environment in around 800 episodes.

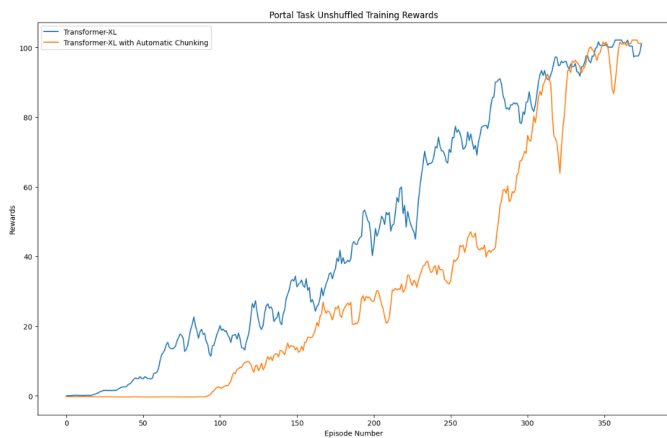


Fig. 6: Average Training Rewards v/s Episode Number for Unshuffled Task.

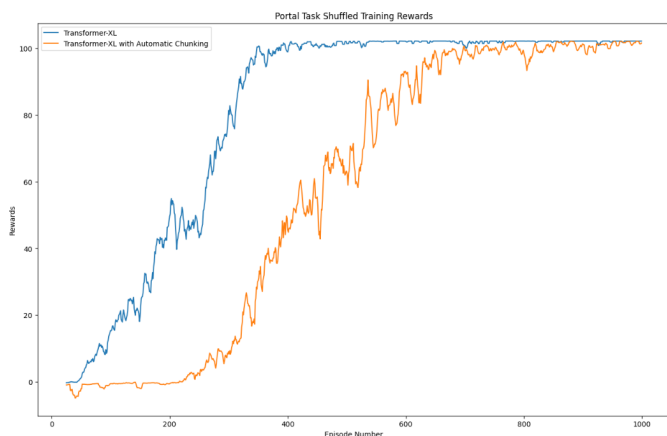


Fig. 7: Average Training Rewards v/s Episode Number for Shuffled Task.

E. Testing Results

Both models successfully reached the final room and completed the task in the Unshuffled variant with a success rate of 98% This proved that the model had learnt the long sequence well and achieved a higher success rate than the human subjects. Both models successfully completed the shuffled variant as well with a success rate of 100% This proved that the models learned the correlation between the context and the

TABLE I: TESTING RESULTS OF MODEL.

Task	Model	Success rate	Failure rate
Unshuffled Portal Task	TransformerXL	49/50 (98%)	1/50 (2%)
	TransformerXL with Automatic Chunking	49/50 (98%)	1/50 (2%)
Shuffled Portal Task	TransformerXL	50/50 (100%)	0/50 (0%)
	TransformerXL with Automatic Chunking	50/50 (100%)	0/50 (0%)

goal correctly. The models also proved their generalizability by learning the random environment with shuffling which proved difficult for the human subjects.

VII. RESULTS

A. Experiment 1: Context Effects on Game Performance

The findings from the first experiment revealed variations in in-game performance and completion among participants:

a) *Subject A Performance:* They successfully completed the game with 8 mistakes. They demonstrated efficient memory recall and decision-making, navigating through the rooms and completing the task.

b) *Subject B Performance:* They experienced difficulty progressing through the game, halting at the 17th room. They made 12 mistakes, indicating challenges in memory recall or decision-making during the task.

These outcomes suggest individual differences in memory retrieval and game performance, highlighting varying abilities to recall contextual cues and make accurate decisions during the game.

B. Experiment 2: Contextual Recall and Episodic Memory Association

In the second experiment, participants’ performance in recalling context and associating it with the correct door choice was analyzed:

a) *Subject C Contextual Recall:* They recalled only 6 correct contexts associated with the rooms. They made 3 mistakes, indicating limitations in episodic memory recall and association with contextual cues.

b) *Subject D Contextual Recall:* They successfully recalled 15 correct contexts associated with the rooms. They made 3 mistakes during the recall phase, showcasing robust episodic memory association with contextual cues.

These results indicate significant differences in participants’ abilities to recall and associate contextual cues with correct door choices, highlighting varying levels of episodic memory recall between individuals.

C. Overall Insights

The results from both experiments underscore the impact of contextual cues on memory recall and decision-making during the game-based tasks. Individual variations in memory retrieval abilities and the influence of contextual cues on episodic memory association were evident, showcasing the significance of vivid contextual cues in enhancing memory recall and performance.

VIII. CONCLUSION

A. Human Memory Limitations and AI Advantages

The limitations of human memory capacity, as observed in the experiments, underscore the potential advantages of AI systems in memory-related tasks. While humans exhibited varying degrees of memory recall and performance limitations,

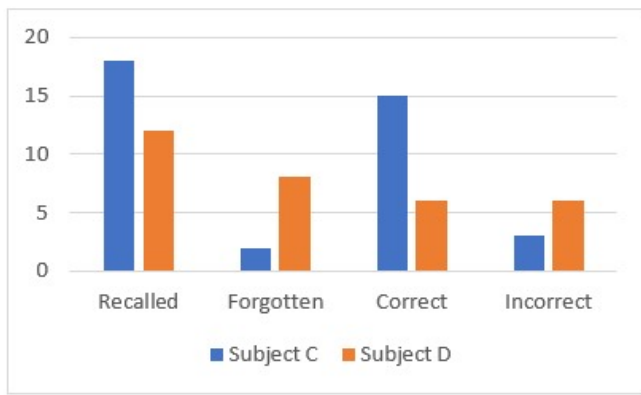


Fig. 8: Subject performance in Experiment 2.

	Recalled	Forgotten	Correct	Incorrect
Subject C	18	2	15	3
Subject D	12	8	6	6

AI models showcased consistent memory retrieval capabilities. This suggests that AI systems, being devoid of cognitive constraints, possess the ability to store and retrieve vast amounts of information more reliably than human memory.

*B. Neural Activation Patterns During Encoding and Retrieval*

The observed neural activation patterns during the encoding and retrieval phases provide insights into the underlying neural mechanisms associated with episodic memory processes. Activation in the prefrontal lobe during encoding aligns with previous research highlighting its role in memory encoding and organization of information. Contrastingly, the predominant activation in the right hemisphere, particularly in the temporal and prefrontal lobes, during retrieval resonates with studies emphasizing the involvement of these brain regions in memory retrieval and associative processes.

*C. Time-Dependent Memory Fading and Contextual Complexity*

The experiments revealed nuances regarding the influence of time and contextual complexity on memory retention. Human memory demonstrated susceptibility to memory fading over time, impacting the accuracy and completeness of memory recall. Moreover, the complexity of contextual cues played a pivotal role in memory association and retrieval. Clear and distinct contextual cues facilitated better memory recall and association, while vague or complex contexts led to limitations in memory retrieval and decision-making, underscoring the importance of context clarity in enhancing memory performance.

*D. Implications and Future Directions*

Understanding the interplay between human memory limitations, neural activation patterns, temporal effects on memory, and contextual complexity holds implications for both cognitive research and AI development. Further investigations could

delve into strategies to optimize human memory recall, leveraging insights from AI memory frameworks to enhance human cognitive processes. Additionally, refining AI memory systems to mimic or adapt to human-like memory constraints in varied contexts could revolutionize AI applications in memory-intensive tasks. Adding abilities such as Future imagination and forgetting could be a step towards emulating human-like cognition in robots and we are exploring these in our future work.

AI surpasses human memory in several aspects, primarily in Encoding capacity, retrieval speed, and consistency. Unlike human memory prone to forgetting and capacity limitations, AI systems retain vast amounts of information without degradation or inaccuracies. They retrieve data rapidly and consistently, handling multiple tasks simultaneously, a feat challenging for human memory. AI’s adaptability, immunity to cognitive biases, and continual learning surpass human memory’s limitations, making it resilient, precise, and constantly improving. Its applications across diverse domains further underscore its potential to revolutionize memory-intensive tasks, offering unparalleled advantages over human memory capabilities.

Since our current experimentation was a preliminary study, we plan to increase the number of subjects in future trials. By expanding our sample size, we aim to enhance the robustness and generalizability of our findings by sampling larger variations in brain activity patterns and responses.

REFERENCES

- [1] A. Bornstein and K. Norman, "Reinstated episodic context guides sampling-based decisions for reward", *Nat Neurosci* 20, pp. 997–1003, 2017, DOI: <https://doi.org/10.1038/nn.4573>.
- [2] N. Herweg, A. Sharan, M. Sperling, A. Brandt, A. Schulze-Bonhage and M. Kahana, *Journal of Neuroscience* 4 March 2020, 40 (10), pp. 2119-2128, 2020, DOI: [10.1523/JNEUROSCI.1640-19.2019](https://doi.org/10.1523/JNEUROSCI.1640-19.2019).
- [3] G. Waldhauser, V. Braun and S. Hanslmayr, *Journal of Neuroscience* 6 January 2016, 36 (1), pp. 251-260, 2016, DOI: [10.1523/JNEUROSCI.2101-15.2016](https://doi.org/10.1523/JNEUROSCI.2101-15.2016).
- [4] Z. Koles and R. Paranjape, "Topographic mapping of the EEG: an examination of accuracy and precision", *Brain Topogr.* 1988 Winter 1(2), pp. 87-95, 1988, DOI: [10.1007/BF01129173](https://doi.org/10.1007/BF01129173), PMID: 3275120.
- [5] L. Hooi, H. Nisar and Y. V. Voon, "Tracking of EEG activity using topographic maps", 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, pp. 287-291, 2015, DOI: [10.1109/ICSIPA.2015.7412206](https://doi.org/10.1109/ICSIPA.2015.7412206).
- [6] M. Pleines, "TransformerXL as Episodic Memory in Proximal Policy Optimization", Github Repository, 2023, Available at: <https://github.com/MarcoMeter/episodic-transformer-memory-ppo>.
- [7] S. Singh, V. Ghatnekar and S. Katti, "Long Horizon Episodic Decision Making for Cognitively Inspired Robots", 2024 Available at SSRN: <https://dx.doi.org/10.2139/ssrn.4660876>, unpublished.
- [8] A. Juliani et al., "Unity: A general platform for intelligent agents", arXiv preprint arXiv:1809.02627. Available at: <https://arxiv.org/pdf/1809.02627.pdf>.
- [9] G. Brockman et al., "OpenAI Gym", arXiv Eprint arXiv:1606.01540, 2016 Available at: <http://arxiv.org/abs/1606.01540>.
- [10] I. Jayarathne, "EEG-processing-python", 2020, Available at: <https://github.com/ijmax/EEG-processing-python>, [retrieved: Feb 2024].
- [11] CymatiCorp, "CyKit", 2020, Available at: <https://github.com/CymatiCorp/CyKit>, [retrieved: Feb 2024].
- [12] OpenAI, "OpenAI Baselines", 2017, Available at: <https://github.com/openai/baselines>, [retrieved: Feb 2024].

APPENDIX

A. Experiment 1 EEG Plots

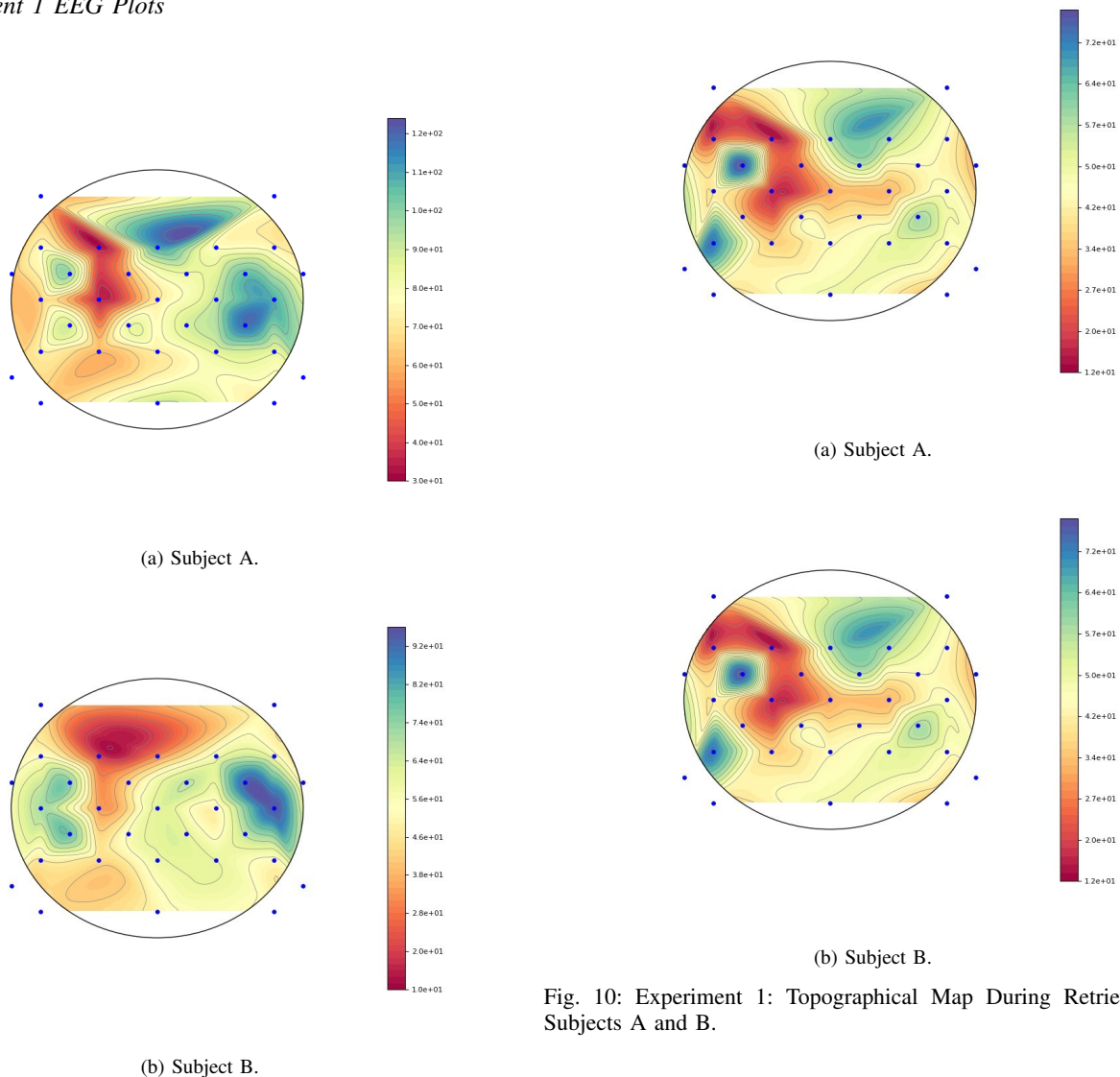
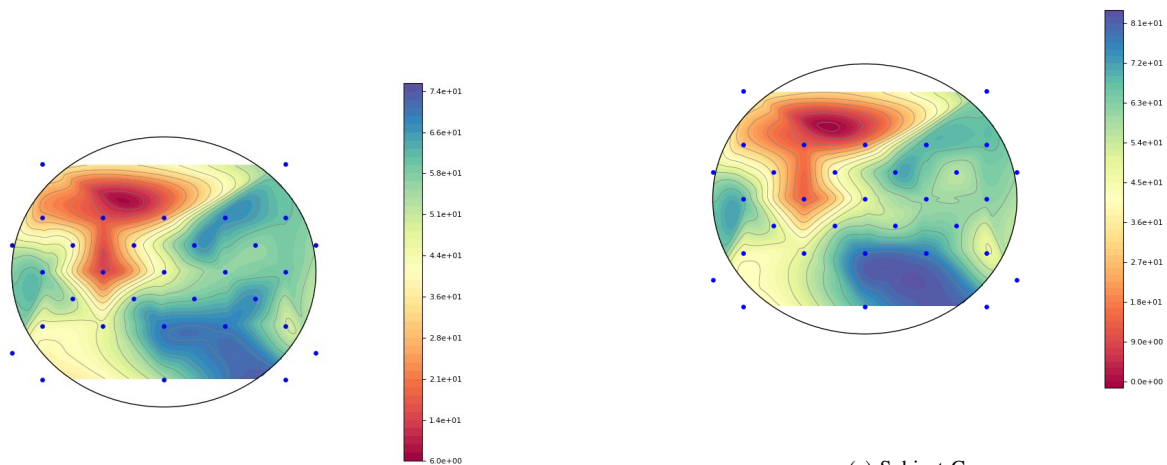


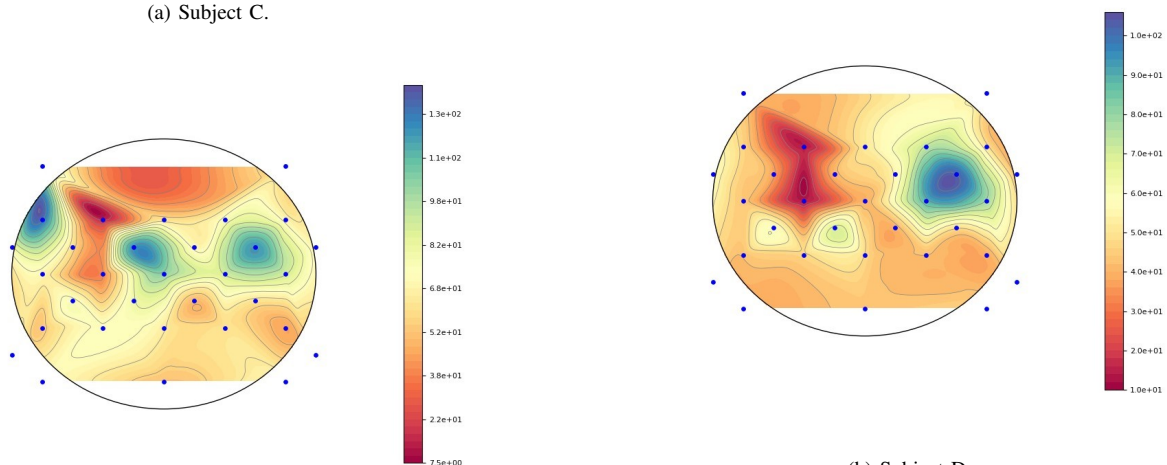
Fig. 10: Experiment 1: Topographical Map During Retrieval for Subjects A and B.

Fig. 9: Experiment 1: Topographical Map During Encoding for Subjects A and B.

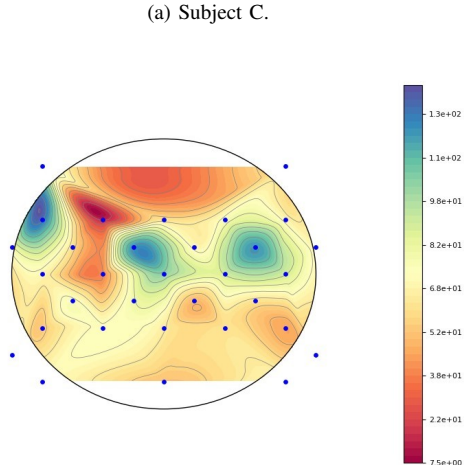
B. Experiment 2 EEG Plots



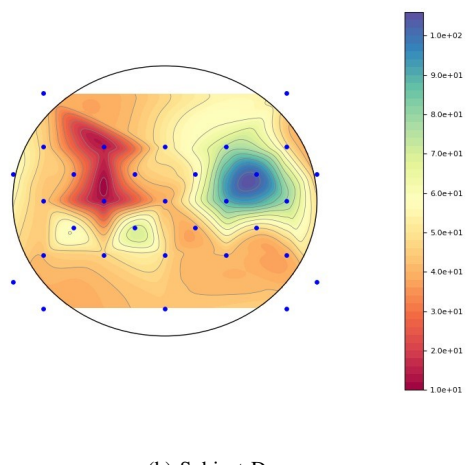
(a) Subject C.



(a) Subject C.



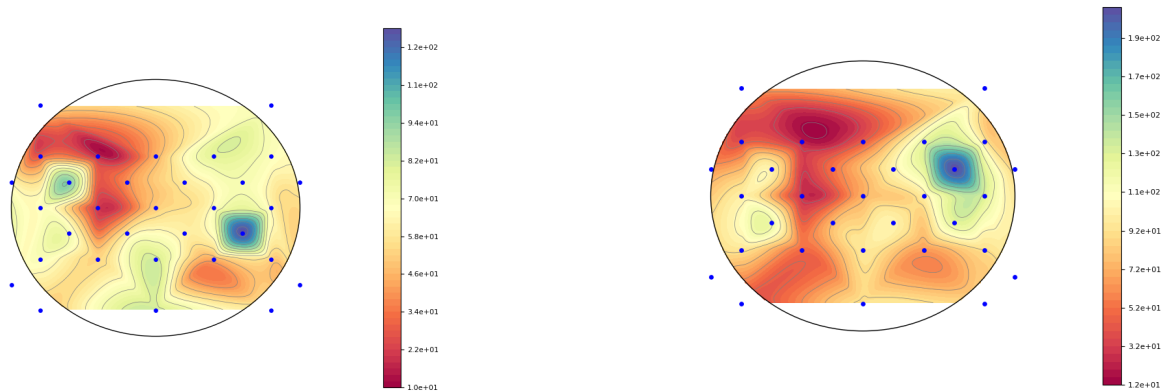
(b) Subject D.



(b) Subject D.

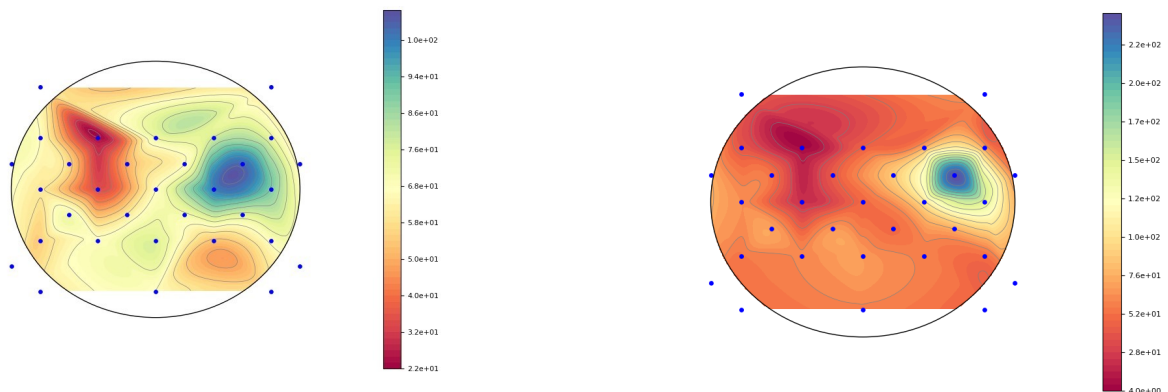
Fig. 11: Experiment 2: Topographical Map During Encoding for Subjects C and D.

Fig. 12: Experiment 2: Topographical Map During Retrieval for Subjects C and D (after encoding).



(a) Subject C.

(a) Subject C.



(b) Subject D.

(b) Subject D.

Fig. 13: Experiment 2: Topographical Map During Retrieval for Subjects C and D (after 6 hours).

Fig. 14: Experiment 2: Topographical Map During Retrieval for Subjects C and D (after 24 hours).