Inside the Mind of Corruption: Towards Detecting Bribery Tendency with Deep Learning on EEG

Samira Ayman Elsamad
School of Computer Science,
Universiti Sains Malaysia
Penang, Malaysia
samiraelsamad@student.usm.my

Syaheerah Lebai Lutfi
School of Computer Science,
Universiti Sains Malaysia
Penang, Malaysia
syaheerah@usm.my
AND College of Medicine
and Health Sciences,
Sultan Qaboos University, Oman

***Abstract*— Corruption remains a pervasive issue globally, and understanding the underlying cognitive processes involved in bribery decision-making is crucial to develop training addressing this problem. This proposed project will investigate the neural mechanisms associated with bribery tendencies using Electroencephalography (EEG) data. By focusing on government officers and laypersons as participants, the research will explore how different individuals respond to bribery stimuli under varying conditions such as risk levels, moral framing, and social pressure. The EEG signals are to be analyzed using a Hybrid CNN-LSTM architecture, combining Convolutional Neural Networks (CNN) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for capturing temporal dynamics. This is an interdisciplinary project with the goal of building a Machine Learning (ML) model that can classify bribery-related decision-making. When complete, this project has the potential to inform interventions aimed at reducing unethical behavior and to be integrated into recruitment processes in certain professions.**

Keywords—Bribery, Corruption, Unethical Decision Making, EEG, Neuro-Imaging, Deep Learning.

I. Introduction

Corruption is a persistent global challenge, affecting governments and institutions across all regions. In Malaysia, discussions of corruption have dominated public discourse for years. A recent example that sparked significant outrage was the partial pardon of former Prime Minister Najib Razak in February 2024 [1]. Najib Razak, pictured in Fig. 1. arriving at a court house in Kuala Lumpur where he was found guilty of committing seven counts of corruption, was convicted in one of the world's largest financial fraud cases just four years earlier. Despite the magnitude of the crimes, his sentence was reduced by half, from 12 to 6 years, prompting Malaysian citizens to express their dismay on social media platforms like X (formerly Twitter), where the pardon became a trending topic. His case exemplifies kleptocracy, a form of corruption where a political leader exploits their influence for personal gain.

Cases like these are not isolated and tend to shape attitudes towards corrupt behavior. Research suggests that exposure to high-profile corruption cases normalizes unethical conduct among individuals in positions of authority [2]. Bribery is a form of corrupt behavior in which individuals in positions of power provide favorable outcomes in exchange for a favor, monetary or otherwise, often to the detriment of the public [4]. Despite widespread attention to corruption, and more specifically bribery, as a social and legal issue, there remains a critical gap in understanding the neural mechanisms that underlie bribery-related decisions Identifying the brain signals that differentiate honest from corrupt decision-making can support the development of scientific tools for early detection, training, or intervention.

Fig. 1. Former Malaysian Prime Minister Najib Razak, arriving at the court house in Kuala Lumpur, Malaysia, Tuesday, July 28, 2020 [3].

This project is work in progress which aims to investigate the neural mechanisms that occur during bribery-related decision-making, focusing on government officers and laypersons as study participants using Electroencephalography (EEG) and Machine Learning (ML). These neural mechanisms we aim to identify can manifest in specific frequency bands, cortical activation patterns, and temporal markers [5]. They can reflect underlying cognitive conflict and moral disengagement and hold the promise of differentiating between honest and corrupt decision making. It has already been established that distinct brain activation patterns can be linked to specific moral judgments as Heraishi et. al [6] demonstrated that left temporal region activation is stronger during morally bad judgments than during morally good judgments.

EEG was chosen as the neural modality in this work and it is a non-invasive and cost-effective neuroimaging technique that captures electric activity in the brain through electrodes placed on the scalp [7]. EEG is well-suited to studying decision-making, emotional regulation, and other

Funded by USM RUTeam grant R502-KR-RUT001-0000000472-K134.

cognitive processes that may require participants to perform tasks during testing, due to its high temporal resolution and the freedom of movement it affords.

Traditionally, EEG analysis has been conducted through statistical methods such as ANOVA, t-tests and linear discriminant analysis [8], but with the recent advancements in computer processing capabilities and the advent of ML models that can efficiently analyze large and rich volumes of EEG signals, there has been a shift in research towards artificial intelligence approaches in EEG analysis. Some popular applications of ML in this field are seizure detection [9], [10] deception detection [11], [12] and emotion recognition [13], [14]. The potential of applying ML models on EEG signals to classify decision-making processes has been successfully explored [15], showing promise in using this approach to differentiate bribery decision-making in our project.

This is a proposed interdisciplinary study involving criminology, neuroscience, and computer science departments aiming to understand the neural mechanisms behind bribery-related decision-making and to shed light on the cognitive processes underlying corrupt behavior. This research also holds promise to inform practical applications in the Malaysian context. This technology can potentially be adapted into recruitment screening tools for high-integrity roles, especially within the public sector and the insights can be used to design targeted training programs aimed at strengthening ethical decision-making. As a locally grounded initiative, this project reflects Malaysia's commitment to addressing corruption through interdisciplinary innovation, drawing from neuroscience, data science, and behavioral research.

II. Literature review

Table 1 presents an overview of some applications of ML on EEG data. All of these studies gathered their data through social experiments that predominantly used visual stimuli.

Deception detection is an application of neural decoding that has gained significant attention as an alternative to traditional lie detection methods. Chen et al. [16] introduced a novel EEG dataset to examine the neural markers of deceptive behavior in a competitive two-player card game. Their study integrated event-related potential (ERP) analysis, microstate analysis, and deep learning-based decoding, providing a benchmark for computational approaches in cognitive neuroscience. A 1D 10-layer convolutional neural network (1D-CNN) was employed to classify the EEG signals. The dataset underwent extensive artifact removal using Artifact Subspace Reconstruction (ASR) and Independent Component Analysis (ICA) to enhance signal quality.

Aslan et al. [17] introduced the LieWaves dataset, a novel EEG-based dataset for lie detection, and employed deep learning techniques, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and a hybrid CNN-LSTM model, to classify deception-related neural patterns. LSTM with Discrete Wavelet Transform (DWT) feature extraction achieved the best classification accuracy of 99.88%. In a previous study [18], the same researchers experimented with a hybrid deep learning model on the publicly available Bag-Of-Lies dataset [22] DWT performed the feature extraction and decomposed the signal into subsignals, specifically D1, D2, D3, D4 and A4 subsignals, that were classified by the LSTM+NCP (Neural Circuit Policies) model.

Decision-making is also another higher cognitive process examined through EEG signals and ML undertaken by researchers. Aldayel et al. [19] built a choice prediction system using EEG signals from the SEED dataset [23]. Their research contributed to the field of neuromarketing, which is the application of neuroscience to identify consumer preferences and develop marketing strategies. Different ML and Deep Learning (DL) classifiers were built to predict individual choices. A proposed LSTM model achieved the best accuracy rate of 96%. In another neuromarketing study [20], bidirectional LSTM had the best performance in predicting consumer choices.

Suhail et al. [21] investigated neural signatures of learning and working memory tasks by combining EEG local activation with functional connectivity patterns. While local activations are measured through event-related potentials (ERPs) or EEG power spectral analysis and indicate which brain regions are active during a task, functional connectivity describes the coordination between different regions and how they interact with each other [24]. The fusion of these EEG features together classified by an ML SVM model performed the best at the Resting vs Focused Reading task with an average accuracy of 97.85%. The model held promising potential for automatic

Table 1. Studies Applying ML Classification on EEG Data

Application	Experiment	Preprocessing	Features Extraction	ML Model
Deception Detection[16]	Two player card game	ICA, ASR	Discrete Wavelet Transform (DWT)	1D-CNN
Deception Detection [17]	Visual stimuli	ICA, ASR, ATAR	DWT and fast Fourier transform (FFT)	CNN, CNN-LSTM, LSTM.
Deception Detection[18]	Visual stimuli	-	DWT and Min-Max Scaler	Hybrid LSTM+NCP.
Decision Making [19]	15 Chinese film clips	Manual artifact removal, Bandpass filtering	PSD and DE frequency-based analysis	CNN, LSTM, CNN-LSTM.
Decision Making [20]	Visual stimuli	-	PSD Spectral analysis	Bidirectional-LSTM.
Learning and Working Memory [21]	Resting, Reading, and Mental Arithmetic Operations.	Manual artifact removal, Bandpass filtering	DWT, Min-Max Normalization	Support Vector Machine (SVM), Artificial Neural Networks (ANN).

Funded by USM RUTeam grant R502-KR-RUT001-0000000472-K134.

detection of mental states when performing various cognitive tasks.

While prior studies have applied machine learning techniques to EEG data, these works primarily focused on tasks involving simple binary responses to visual stimuli or consumer preferences. Unlike these existing studies, our research is the first to examine bribery-related decision-making through a realistically framed and morally complex EEG-based task, using a custom-designed 2AFC bribery game. We also differentiate our approach by combining ERP and functional connectivity features, tailored to capture both spatial and temporal EEG dynamics.

III. Proposed Methodology

A. Experiment Design

The experimental design and data collection will be conducted in collaboration with researchers in criminology and neuroscience from the School of Health Sciences, USM to ensure domain-specific relevance and methodological rigor. The collaborators will be involved in designing realistic bribery scenarios and overseeing EEG data acquisition protocols. The entire proposed ML pipeline is displayed in Fig. 2.

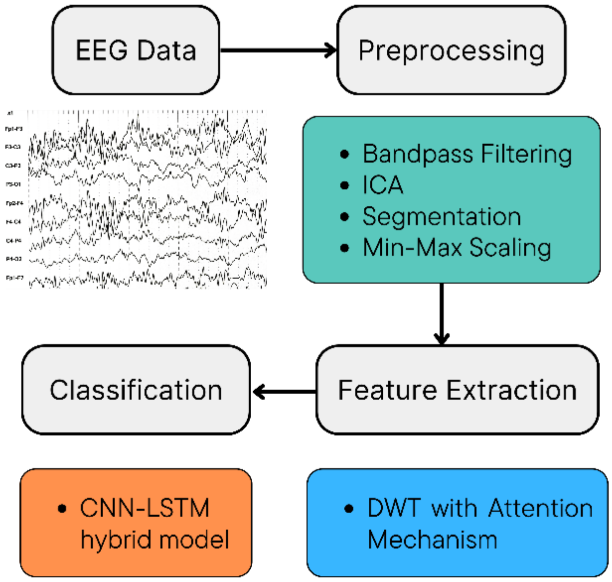

Fig. 2. Bribery Tendency Classification Pipeline Using EEG Data.

1) Participants

The sample size will be 184 participants, with 92 government officials and 92 laypersons as two distinct group. While they will be grouped together when building the ML model, this provides the opportunity to examine the differences of reactions to bribery stimuli between the two groups. The participants will be consenting Malaysian adults free of neurological disorders.

2) Task Structure: Bribery Decision Game

Participants will engage in a computerized decision-making task designed as a two-alternative forced-choice (2AFC) paradigm, adapted to capture neural and behavioural responses to bribery scenarios. This 2AFC setup aligns with the structure used by Pärnamets et al. [25], ensuring controlled comparisons and consistency across participants.

In each trial, participants take the role of a government official and are presented with hypothetical bribe offers, which include details such as the amount, risk level, and moral justification. They must respond to the prompt: *"Accept or Refuse?"*. To introduce moral complexity and cognitive conflict, the task manipulates contextual framing across three dimensions: risk-reward trade-offs, moral framing (e.g., aiding a struggling family vs. personal gain), and social pressure (e.g., bribery as organizational norm) as displayed in Fig 1.

Each trial lasts approximately 30–45 seconds, with participants completing 100 trials across multiple sessions. Research by Pardo-Vazquez et al. [26] also showed that EEG signals linked to performance monitoring are easier to detect and interpret when participants must choose between two clear options. In this bribery task, the forced choice between "Accept" or "Refuse" allows us to track how the brain responds to different moral, social, and risk-based factors during each decision.

Fig. 3. Example of a Bribery Task.

3) Data Prepocessing

Bandpass filtering (0.5–45 Hz) will be applied to isolate relevant frequency bands and Independent Component Analysis (ICA) will perform artifact removal. Artifact removal is vital so the signals are preprocessed from noise and are not corrupted by the motor movements of the participants. Afterwards, the data will be segmented into time-locked epochs, each associated with a task event. Finally, the data will be normalized using Min-Max scaling.

4) Feature Extraction

Two types of EEG features will be extracted for input into the model: event-related potentials (ERPs) features and functional connectivity features. Wei et al. [27] asserted that functional connectivity patterns elicited by deception tasks differ from those associated with honest behavior, and that research should go beyond identifying brain regions activated during lying, which can be identified by ERPs and also examine the interactions between different brain areas or their functional connectivity. Adaptive Discrete Wavelet Transform (DWT) [28] with a trainable wavelet basis function will be applied and will use attention

Funded by USM RUTeam grant R502-KR-RUT001-0000000472-K134.

mechanisms to dynamically select discriminative frequency sub-bands (delta, theta, alpha, beta).

5) Model Architecture

A Hybrid CNN-LSTM architecture that combines CNN for spatial feature extraction (identifying brain regions activated during decision-making) and LSTM for capturing temporal dynamics in EEG signals, which are crucial for understanding cognitive processes over time, can efficiently handle both ERP and functional connectivity features. This approach has been successful in similar deception detection studies [17] and provides a flexible solution for classifying decision-making behaviors related to bribery.

IV. LIMITATIONS

- The neural mechanisms identified in this study will be derived from controlled experimental conditions and may not fully capture the complexity of real-world bribery scenarios. While the task design incorporates realistic elements such as varying risk levels, moral framing, and social pressure, it cannot replicate the emotional urgency or external influences that often accompany actual bribery solicitations. In everyday contexts, factors such as fear, peer influence, or institutional norms can deeply shape ethical decision-making.
- While the interdisciplinary collaboration with criminology and neuroscience experts has enhanced the realism and validity of the experimental design, the artificial setting of a lab-based task still limits environmental validity. Bribery is a highly contextual and socially sensitive behaviour, and responses in a monitored environment may differ from those in real-world situations.
- Ethical approval has been obtained for this study, and great care is taken to ensure the dignity, autonomy, and confidentiality of all participants. Ethical considerations also require that findings not be used to label individuals as corrupt or immoral based solely on their responses in an experimental setting. The goal is to understand cognitive processes and not to assign moral judgment.

V. CONCLUSION

This study presents a proposed pioneering approach to understanding the neural mechanisms behind bribery-related decision-making through EEG data analysis. By applying applying a hybrid CNN-LSTM deep learning architecture to EEG signals, this research aims to uncover specific neural mechanisms associated with bribery decision making, and possibly uncover patterns that may be indicative of cognitive conflict, risk evaluation, and moral reasoning during bribery dilemmas. The goal is to contribute to a deeper understanding of the cognitive processes involved in corruption. The study is strengthened by close collaboration with experts in criminology and neuroscience, who will guide the experimental design and data collection to align with real-world bribery dynamics.

Although there are limitations in terms of real-world applicability and experimental constraints, the findings could serve as a valuable foundation for future research in behavioral neuroscience, ethics, and corruption studies. The integration of neuroimaging with behavioural ethics and decision-making also holds promise for practical applications, such as training, behavioural interventions, or institutional reforms aimed at reducing corruption. Ultimately, this research contributes to a growing interdisciplinary effort to decode the psychological and neural basis of unethical decision-making in complex social environments. As of now, the project is actively underway, with pilot survey data collection recently conducted and pilot EEG data collection with the bribery task scheduled for the next research phase.

VI. ACKNOWLEDGEMENT

We thank Universiti Sains Malaysia for the funding of this project from RUTeam grant R502-KR-RUT001-0000000472-K134.

REFERENCES

[1] "Malaysia halves ex-PM Najib Razak's jail term in 1MDB corruption scandal." Accessed: Jan. 05, 2025. [Online]. Available: https://www.aljazeera.com/news/2024/2/2/malaysia-reduces-sentence-of-former-pm-najib-razak

[2] M. Gorsira, A. Denkers, and W. Huisman, "Both Sides of the Coin: Motives for Corruption Among Public Officials and Business Employees," *Journal of Business Ethics*, vol. 151, no. 1, pp. 179–194, Aug. 2018, doi: 10.1007/s10551-016-3219-2.

[3] "Ex-Malaysian PM Najib Razak sentenced to 12 years in jail after first trial over 1MDB scandal." Accessed: Apr. 10, 2025. [Online]. Available: https://www.sbs.com.au/news/article/ex-malaysian-pm-najib-razak-sentenced-to-12-years-in-jail-after-first-trial-over-1mdb-scandal/gfm8cdntm

[4] K. Abbink and E. Renner, "An Experimental Bribery Game."

[5] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization," *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING*, vol. 31, pp. 710–719, 2023, doi: 10.1109/TNSRE.2022.3230250.

[6] H. Hiraishi *et al.*, "Regional and Temporal Differences in Brain Activity With Morally Good or Bad Judgments in Men: A Magnetoencephalography Study," *Front Neurosci*, vol. 15, 2021.

[7] R. Sharma and H. K. Meena, "Emerging Trends in EEG Signal Processing: A Systematic Review," Apr. 09, 2024, *Springer*. doi: 10.1007/s42979-024-02773-w.

[8] J. Abrahantes, J. SERROYEN, H. Geys, G. Molenberghs, and P. Drinkenburg, "Statistical methods for EEG data," Apr. 2007.

[9] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *J Neural Eng*, vol. 16, no. 3, Jun. 2019, doi: 10.1088/1741-2552/ab0ab5.

[10] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput Biol Med*, vol. 100, pp. 270–278, Sep. 2018, doi: 10.1016/j.compbiomed.2017.09.017.

[11] N. Abe, "How the brain shapes deception: An integrated review of the literature," Oct. 2011. doi: 10.1177/1073858410393359.

[12] J. Gao *et al.*, "Effective Connectivity in Cortical Networks During Deception: A Lie Detection Study Based on EEG," *IEEE J Biomed Health Inform*, vol. 26, no. 8, pp. 3755–3766, Aug. 2022, doi: 10.1109/JBHI.2022.3172994.

[13] K. Kamble and J. Sengupta, “A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals,” *Multimed Tools Appl*, vol. 82, no. 18, pp. 27269–27304, Jul. 2023, doi: 10.1007/s11042-023-14489-9.

[14] F. Liu *et al.*, “Emotion Recognition From Few-Channel EEG Signals by Integrating Deep Feature Aggregation and Transfer Learning,” *IEEE Trans Affect Comput*, vol. 15, no. 3, pp. 1315–1330, Jul. 2024, doi: 10.1109/TAFFC.2023.3336531.

[15] Z. Huang, K. Jiang, J. Li, W. Zhu, H. Zheng, and Y. Wang, “Discriminability of single-trial EEG during decision-making of cooperation or aggression: a study based on machine learning,” *Med Biol Eng Comput*, vol. 60, no. 8, pp. 2217–2227, Aug. 2022, doi: 10.1007/s11517-022-02557-5.

[16] Y. Chen, S. Fazli, and C. Wallraven, “An EEG Dataset of Neural Signatures in a Competitive Two-Player Game Encouraging Deceptive Behavior,” *Sci Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1038/s41597-024-03234-y.

[17] M. Aslan, M. Baykara, and T. B. Alakus, “LieWaves: dataset for lie detection based on EEG signals and wavelets,” *Med Biol Eng Comput*, vol. 62, no. 5, pp. 1571–1588, May 2024, doi: 10.1007/s11517-024-03021-2.

[18] M. Aslan, M. Baykara, and T. B. Alakuş, “LSTMNCP: lie detection from EEG signals with novel hybrid deep learning method,” *Multimed Tools Appl*, vol. 83, no. 11, pp. 31655–31671, Mar. 2024, doi: 10.1007/s11042-023-16847-z.

[19] M. Aldayel, A. Kharrat, and A. Al-Nafjan, “Predicting Choices Driven by Emotional Stimuli Using EEG-Based Analysis and Deep Learning,” *Applied Sciences (Switzerland)*, vol. 13, no. 14, Jul. 2023, doi: 10.3390/app13148469.

[20] H. Göker, “Multi-channel EEG-based classification of consumer preferences using multitaper spectral analysis and deep learning model,” *Multimed Tools Appl*, vol. 83, no. 14, pp. 40753–40771, Apr. 2024, doi: 10.1007/s11042-023-17114-x.

[21] T. A. Suhail, K. P. Indiradevi, E. M. Suhara, S. A. Poovathinal, and A. Ayyappan, “Distinguishing cognitive states using electroencephalography local activation and functional connectivity patterns,” *Biomed Signal Process Control*, vol. 77, Aug. 2022, doi: 10.1016/j.bspc.2022.103742.

[22] V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh, and M. Vatsa, “Bag-of-Lies: A Multimodal Dataset for Deception Detection,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 83–90. doi: 10.1109/CVPRW.2019.00016.

[23] W.-B. Jiang, X.-H. Liu, W.-L. Zheng, and B.-L. Lu, “SEED-VII: A Multimodal Dataset of Six Basic Emotions with Continuous Labels for Emotion Recognition,” *IEEE Trans Affect Comput*, pp. 1–16, 2024, doi: 10.1109/TAFFC.2024.3485057.

[24] N. Frolov, M. S. Kabir, V. Maksimenko, and A. Hramov, “Machine learning evaluates changes in functional connectivity under a prolonged cognitive load,” *Chaos*, vol. 31, no. 10, Oct. 2021, doi: 10.1063/5.0070493.

[25] P. Pärnamets, P. Johansson, L. Hall, C. Balkenius, M. J. Spivey, and D. C. Richardson, “Biasing moral decisions by exploiting the dynamics of eye gaze,” *Proc Natl Acad Sci U S A*, vol. 112, no. 13, pp. 4170–4175, Mar. 2015, doi: 10.1073/pnas.1415250112.

[26] J. L. Pardo-Vazquez, I. Padrón, J. Fernández-Rey, and C. Acuña, “EEG activity represents the correctness of perceptual decisions trial-by-trial,” *Front Behav Neurosci*, vol. 8, no. MAR, Mar. 2014, doi: 10.3389/fnbeh.2014.00105.

[27] S. Wei *et al.*, “Analysis of Weight-Directed Functional Brain Networks in the Deception State Based on EEG Signal,” *IEEE J Biomed Health Inform*, vol. 27, no. 10, pp. 4736–4747, Oct. 2023, doi: 10.1109/JBHI.2023.3295892.

[28] A. Narmada and M. K. Shukla, “A novel adaptive artifacts wavelet Denoising for EEG artifacts removal using deep learning with Meta-heuristic approach,” *Multimed Tools Appl*, vol. 82, no. 26, pp. 40403–40441, Nov. 2023, doi: 10.1007/s11042-023-14949-2.