THESIS PROJECT
TITLE PROPOSAL
(CS0025 – Software Engineering 1)

Group Name: MicroMarc	Program: **Bachelor of Science in Computer Science with specialization in Software Engineering**	
Member's Name: 1. Olata, Marc 2. Coma, Alpha Romer 3. Sioson, Kristoffer Ian 4. Ong, Job Isaac	Term: [✔] 1st [ ] 2nd [ ] 3rd	Academic Year: 2024 - 2025
Proposed Project Title 1: **Augmenting Multimodal Deep Learning with Attention Layers to Recognize Addictive Short Form Content**		

1.0 Area of Investigation:

The internet has become a vast repository of information and entertainment, leading to the rise of highly engaging, short-form, and repetitive media content colloquially termed "brain rot." This content, characterized by its addictive nature, low cognitive demand, and rapid consumption cycles, has been associated with cognitive overload and reduced attention spans, particularly in younger demographics. The phenomenon has been likened to traditional media's effect on cognitive functions but amplified by the immediacy and ubiquity of the internet.

This research investigates the intersection of artificial intelligence, cognitive psychology, and digital content analysis, focusing on developing a **multimodal deep learning framework** augmented with **attention mechanisms** to detect and analyze addictive short-form video content. It integrates **visual, audio, and textual data** to identify patterns and characteristics that foster addictive behaviors among users.

By leveraging advanced AI techniques like **attention layers**, the study aspires to contribute to a deeper understanding of how short-form content impacts cognitive processes, particularly attention spans, emotional regulation, and behavioral patterns. It situates itself at the confluence of computational sciences, behavioral studies, and public health, offering interdisciplinary insights to mitigate negative effects while promoting the creation and consumption of cognitively enriching content.

Despite its widespread consumption, the long-term psychological and neurological effects of such content remain underexplored, particularly in the context of video platforms like TikTok, YouTube Shorts, and Instagram Reels. This research aims to bridge that gap by analyzing the cognitive and emotional impact of brain rot content through automated video analysis and psychological validation.

2.0 Background and Rationale of the Project/System:

The explosion of short-form video content, epitomized by platforms like TikTok, Instagram Reels, and YouTube Shorts, has revolutionized digital consumption patterns. These videos are designed for rapid engagement and mass appeal, often employing algorithmic reinforcement to

captivate viewers. However, studies highlight growing concerns over their psychological and societal impact, leading to the modern phenomena known as **“Brain Rot”**—a colloquial term for cognitive decline linked to excessive short-form media exposure.

The COVID-19 pandemic amplified these challenges as prolonged screen time became the norm, particularly among younger audiences. This era has left residual effects, including **diminished attention spans**, **content addiction**, and **reduced engagement with cognitively enriching activities**. Despite these concerns, research into the mechanisms that make such content addictive and its impact on mental health remains nascent.

Rationale:

This research is motivated by the need to:

1. **Understand the Science of Engagement**: Investigate the specific features—visual, auditory, and textual—that contribute to the addictiveness of short-form content.
2. **Leverage AI for Better Insights**: Develop an advanced **multimodal deep learning model** with attention mechanisms to analyze and predict content addictiveness.
3. **Promote Digital Well-Being**: Collaborate with mental health experts to mitigate the adverse cognitive effects of addictive digital consumption and support the creation of **cognitively enriching content**.
4. **Support Regulatory and Platform Policies**: Provide actionable insights for social media platforms to recommend healthier content while aligning with policy frameworks for digital well-being.

This study is interdisciplinary, combining **computer science**, **psychology**, and **media studies** to address a pressing societal issue. By creating a **curated dataset** of addictive short-form content and analyzing its multimodal features, the project aims to not only classify content but also propose strategies for balanced consumption, benefiting multiple stakeholders.

Contribution to Knowledge:

- **Novel Dataset**: A unique labeled dataset of addictive short-form videos, curated in collaboration with psychological experts in the Philippines.
- **AI Innovation**: Integration of **attention layers** for efficient multimodal learning.
- **Social Impact**: Insights into reducing screen-induced cognitive challenges and improving mental health outcomes.

Significance:

The study aligns with growing societal interest in the implications of digital media consumption. With **"Brain Rot"** recognized as Oxford’s Word of the Year 2024, this research addresses a timely and impactful problem, fostering discussions on digital well-being and ethical technology use.

3.0 Statement of Objectives:

The primary objective of this study is to develop an advanced deep learning system capable of identifying and analyzing addictive characteristics in short-form digital content. This system aims to leverage multimodal data—visual, auditory, and textual features—and incorporate attention mechanisms to improve the interpretability and accuracy of predictions. The following specific objectives guide the research:

Key Objectives

1. **Develop a Multimodal Deep Learning Model**
- Create and implement a **multimodal AI model** integrating visual, auditory, and textual data.
- Use **attention layers** to prioritize the most relevant features across modalities.
- Evaluate the model’s effectiveness in recognizing patterns indicative of addictiveness in short-form content.
2. **Identify Key Features of Addictive Content**

- Investigate what makes short-form content particularly engaging or habit-forming.
- Study elements such as **visual appeal (e.g., colors, rapid transitions)**, **audio cues (e.g., catchy beats, sound effects)**, and **textual hooks (e.g., captions, titles)**.

3. **Demonstrate the Value of Attention Mechanisms**
- Validate the hypothesis that attention layers enhance multimodal learning by enabling the model to weigh critical features differently across modalities.
- Assess performance improvements compared to models without attention mechanisms.

4. **Collaborate with Experts in the Field of Psychology**
- Engage with mental health professionals to define "addictiveness" using cognitive and emotional benchmarks.
- Co-create a **labeled dataset** of addictive and non-addictive short-form content for model training and validation.

5. **Evaluate and Refine the System**
- Conduct rigorous testing to measure the system's accuracy, robustness, and generalizability.
- Analyze potential biases introduced during manual content annotation.

6. **Contribute to Digital Well-Being**
- Provide insights into mitigating the effects of "Brain Rot" by developing systems that identify and potentially flag or moderate cognitively harmful content.
- Offer recommendations for social media platforms to promote **cognitively enriching and balanced media consumption**.

Long-Term Vision

The outcomes of this research aim to empower stakeholders—such as mental health professionals, content creators, and regulatory bodies—with tools to identify and mitigate the spread of addictive content, fostering a healthier digital media ecosystem.

4.0 Scope and Limitations of the Study:

Scope

This research focuses on developing and evaluating a multimodal deep learning model augmented with attention mechanisms to detect and analyze addictive short-form video content. It addresses a contemporary digital phenomenon with significant social, psychological, and technological implications.

1. **Content Scope**
- The study examines content from popular short-form video platforms such as **TikTok**, **Instagram Reels**, and **YouTube Shorts**.
- Multimodal features include **visual (e.g., colors, motion patterns)**, **auditory (e.g., soundtracks, sound effects)**, and **textual elements (e.g., captions)**.
2. **Technology Scope**
- Utilizes **multimodal fusion techniques**, such as early, late, and intermediate fusion, to integrate features from diverse data types.
- Incorporates **attention mechanisms** to enhance the system's ability to prioritize significant data patterns and improve prediction accuracy
3. **Interdisciplinary Collaboration**
- Collaborates with psychologists and mental health experts to define "addictiveness" and establish labeling criteria for the dataset.
- Includes feedback from social scientists to ensure a holistic understanding of the cognitive and behavioral implications.
4. **Dataset Creation**
- Involves the creation of a curated, annotated dataset of addictive and non-addictive short-form content.
- Manual annotation by a diverse group of raters ensures the inclusion of various perspectives in labeling.

Limitations

1. **Platform Specificity**
- The research focuses on content from a limited number of platforms, which may reduce generalizability to other forms of digital media, such as long-form content or traditional media.
2. **Subjectivity in Annotations**
- The manual annotation of content introduces subjective biases, as perceptions of addictiveness may vary between annotators based on cultural, psychological, and personal factors.
3. **Data Collection Constraints**
- Access to platform data is subject to privacy policies, copyright restrictions, and API limitations, potentially limiting the diversity of the dataset.
4. **Technical Challenges**
- Complexities in aligning and fusing multimodal data may impact system performance.
- Attention mechanisms, while powerful, may introduce computational overhead, affecting scalability.
5. **Ethical Considerations**
- Defining and labeling content as "addictive" could raise ethical concerns, such as stigmatization or misinterpretation by content creators and users.
6. **Temporal Relevance**
- As trends in short-form content are highly dynamic, the relevance of features identified during the study may diminish over time, requiring periodic updates to the model.

Primary Beneficiaries

1. **Psychologists and Mental Health Researchers:** The system will provide insights into the cognitive and emotional impact of short-form video consumption, aiding in developing interventions to mitigate negative psychological effects. Research by Su et al. (2024) on the impact of short video consumption on attention functions underscores the need for such tools in psychological research.

2. **Educational Institutions:** Educators and policymakers can use the system to identify and discourage the consumption of brain rot content among students, promoting healthier media consumption habits. Liu and Choi (2024) highlight the role of cognitive load in educational settings and the adverse effects of low-value content on learning engagement (Liu & Choi, 2024).

3. **Social Media Platforms:** Social media companies can leverage the system to develop algorithms that recommend more cognitively enriching content, enhancing user well-being and long-term engagement. Nguyen et al. (2024) demonstrate the potential for AI-based content recommendation systems to influence user behavior positively (Nguyen et al., 2024).

Secondary Beneficiaries

1. **Parents and Guardians:** The system will empower parents with insights into the types of content their children consume, enabling them to make informed decisions about screen time and media exposure.

2. **Content Creators and Influencers:** By understanding the characteristics of brain rot content, creators can pivot toward producing more meaningful and engaging content that promotes cognitive well-being.

3. **Regulatory Bodies and Policymakers:** The findings from this research can inform regulations and policies aimed at reducing the prevalence of cognitively harmful content on social media platforms, promoting digital well-being across society.

5.0 References:

Local Study:

Baltazar, K. R., Cajipe, R. H., Espanol, K., Charles Andrew, G., & Segundo Jr., S. P. (2023). Video content analysis through computer vision and natural language processing. [School Library].

Foreign Study:

Beskow, D. M., Kumar, S., & Carley, K. M. (2020). The evolution of political memes: Detecting and characterizing internet memes with multi-modal deep learning. *Information Processing & Management*, *57*(2), 102170.

Ceder-Thorin, F. (2024). Social media use at various levels of engagement and its emotional consequences or How I learned to stop worrying and love the scroll. https://www.diva-portal.org/smash/record.jsf?dswid=-1804

Eckert, J. (2024). TikTok poetry: The death of art or a new frontier? https://openresearch.okstate.edu/entities/publication/2333dcc2-ca0d-45ff-876d-ffc534f73a1d

Issahaku, F. Y., Liu, X., Lu, K., Fang, X., Danwana, S. B., & Asimeng, E. (2024). Multimodal deep learning model for COVID-19 detection. *Biomedical Signal Processing and Control, 91,* 105906. https://www.sciencedirect.com/science/article/abs/pii/S1746809423013393

Li, Y.-F., Wang, H., & Sun, M. (2024). ChatGPT-like large-scale foundation models for prognostics and health management: A survey and roadmaps. Reliability Engineering & System Safety, 243, 109850–109850. https://doi.org/10.1016/j.ress.2023.109850

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (pp. 689-696).

Nguyen, K. M., Nguyen, N. T., Ngo, N. T. Q., Tran, N. T. H., & Nguyen, H. T. T. (2024). Investigating consumers' purchase resistance behavior to AI-based content recommendations on short-video platforms: A study of greedy and biased recommendations. *Journal of Internet Commerce*, *23*(3), 284–327. https://doi.org/10.1080/15332861.2024.2375966

Liu, Z., & Choi, A. Y. (2024). Research on information architecture design of short-form video social platforms based on cognitive psychology. David Publisher. https://www.davidpublisher.com/Public/uploads/Contribute/65bc8953cca2a.pdf

Rahman, W., Hasan, K., Lee, S., Zadeh, A., Mao, C., Morency, L.P., Hoque, E. (2020). Integrating Multimodal Information in Large Pretrained Transformers. https://aclanthology.org/2020.acl-main.214.pdf

Sharma, D., Purushotham, S., Reddy, C. (2021). MedFuseNet: An attention-based multimodal deep learning model for visual question answering in the medical domain. https://www.researchgate.net/publication/355109958_MedFuseNet_An_attention-based_multimodal_deep_learning_model_for_visual_question_answering_in_the_medical_domain

Tabassum, N. (2024). A MULTI-HEAD ATTENTION APPROACH WITH COMPLEMENTARY MULTIMODAL FUSION FOR VEHICLE DETECTION. doi: 10.25394/PGS.25270039.v1

Yan, T., Su, C., Xue, W., Hu, Y., & Zhou, H. (2024). Mobile phone short video use negatively impacts attention functions: An EEG study. *Frontiers in Human Neuroscience*. https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2024.1383913/full

Yuan, S., Li, W., Ling, Y., Huang, X., Feng, A., Tan, S., He, N., Li, L., Li, S., Xu, A., & Lyu, J. (2023). Associations of screen-based sedentary activities with all cause dementia, Alzheimer's disease, vascular dementia: A longitudinal study based on 462,524 participants from the UK Biobank. *BMC Public Health*, *23*(1), 2141. https://doi.org/10.1186/s12889-023-17050-3

Foreign Literature:

Deep Cognition and Language Research (DeCLaRe) Lab (2022). Multimodal Deep Learning. https://github.com/declare-lab/multimodal-deep-learning

Moledina, S., & Khoja, A. (2018). Letter to the editor: Digital dementia-Is smart technology making us dumb? *Ochsner Journal*, *18*(1), 12. https://www.ochsnerjournal.org/content/18/1/12

Naness, M. (2024). Unpopular opinion: The TikTok brain rot must be stopped. The Colgate Maroon News. https://thecolgatemaroonnews.com/52356/commentary/unpopular-opinion-the-tiktok-brain-rot-must-be-stopped/

Oxford University Press (2024). 'Brain rot' named Oxford Word of the Year 2024. https://corp.oup.com/news/brain-rot-named-oxford-word-of-the-year-2024/

Potrimba, P. (2023) Multimodal Models and Computer Vision: A Deep Dive. https://blog.roboflow.com/multimodal-models/

Pulapakura, R. (2024). Multimodal Models and Fusion - A Complete Guide. https://medium.com/@raj.pulapakura/multimodal-models-and-fusion-a-complete-guide-225ca91f6861

6.0 Group Tasks

- Olata: Statement of Objectives, Scope and Limitations
- Sioson: Area of Investigation, Background/Rationale of the study
- Coma: Compilation of references, ideas, background of the study
- Ong: Target Beneficiaries