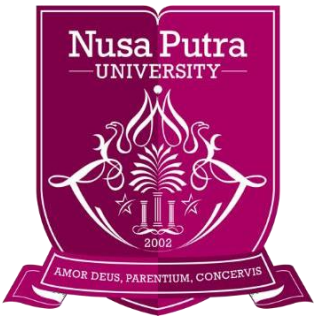
VIDEO CLASSIFICATION USING LARGE LANGUAGE MODELS APPROACH FOR FILM CENSORSHIPS IN INDONESIA



THESIS

Submitted as fulfilment of the requirements for the completion of Master of Computer Science Program

DHANY KURNIAWAN
20220130011

MASTER OF COMPUTER SCIENCE PROGRAM
SCHOOL OF COMPUTER SCIENCE
NUSA PUTRA UNIVERSITY
2024

AUTHOR STATEMENT

TITLE : VIDEO CLASSIFICATION USING LARGE LANGUAGE MODELS APPROACH FOR FILM CENSORSHIPS IN INDONESIA
NAME : Dhany Kurniawan
NIM 20220130011

"I solemnly declare and assume responsibility that this thesis is my own work, except for excerpts and summaries, each of which I have explained the source of. If at a later time there are other parties who claim that this Thesis is his work, which is accompanied by sufficient evidence, then i am willing to cancel my Master of Computer Degree along with all rights and obligations attached to the title.".

Sukabumi,

Dhany Kurniawan

APPROVAL THESIS

TITLE : VIDEO CLASSIFICATION USING LARGE LANGUAGE MODELS APPROACH FOR FILM CENSORSHIPS IN INDONESIA
NAME : Dhany Kurniawan
NIM 20220130011

This thesis has been reviewed and approved
Sukabumi, August 2024

Head of Study Program,	Supervisor,
Prof. Ir. Teddy Mantoro, PhD., SMIEEE	Prof. Ir. Media Anugerah Ayu, MSc., PhD., SMIEEE

THESIS APPROVAL

Title : VIDEO CLASSIFICATION USING LARGE LANGUAGE MODELS APPROACH FOR FILM CENSORSHIPS IN INDONESIA
Persons Name : Dhany Kurniawan
ID of Student : 20220130011

This Thesis has been tested and defended in front of the Board of Examiners in Thesis session on . In our review, this Thesis adequate in terms of quality for the purpose of awarding the Master of Computer Degree.

Sukabumi, August 2024

Supervisor 1	Examiner 1
Prof. Ir. Media Anugerah Ayu, MSc., PhD., SMIEEE	Jelita Asian, M.Sc., PhD
Supervisor 2	Examiner 2
Rahmadya Trias Handyanto, S.T., M.Kom., PhD	Umar Aditiawarman, S.T., M.Sc., PhD

FOREWORD

Praise and gratitude the author prays to Allah Subhanahu wa ta'ala, because only with His blessings and grace can the author complete this thesis. Writing this thesis is one of the requirements to achieve a Master's degree in Computers at Nusa Putra University. I, Dhany Kurniawan, known as Prince, realize that without help and guidance from various parties, from the lecture period to the preparation of this thesis, it is very difficult for the author to complete this thesis. Therefore, I would like to thank:

1. Prof. Ir. Teddy Mantoro, PhD., SMIEEE as Head of the Master of Informatics Study Program at Nusa Putra University.
2. Prof. Ir. Media Anugerah Ayu, MSc., PhD., SMIEEE as Supervisor I.
3. Rahmadya Trias Handyanto, S.T., M.Kom., PhD as Supervisor II.
4. Mr. and Mrs. Lecturer in the Computer Science Master's Program at Nusa Putra University.
5. Staff and Employees of the Master of Computer Science Program at Nusa Putra University.
6. Both parents and older brother who have prayed for, helped and supported morally and materially.
7. My beloved wife who has prayed for, helped and supported morally.
8. Hari Ambari S.Hum, M.Kom as a generous and kind friend and master who has helped and supported morally and materially.
9. Ismail Fahmi, PhD as Founder of Drone Emprit.
10. The Film Censorship Board of the Republic of Indonesia.
11. Colleagues in the Computer Science Masters Program at Nusa Putra University.
12. All parties who have helped in completing this thesis.

For further improvement, suggestions and constructive criticism will be gladly accepted. Finally, only to Allah Subhanahu wa ta'ala the author submits everything, hopefully it can be useful especially for writers in general for all of us.

Sukabumi, August 2024

TABLE CONTENT

FIGURE CONTENT

LIST TABLE

LIST APPENDICES

ABSTRACT

The advancement of technology has made the dissemination of videos increasingly massive and widespread, making it very easy for anyone to access them at this time. Not only the dissemination of videos, but even the creation of videos is becoming easier and can be done by anyone. With the development of the times, now the media for video broadcasting is becoming more and more numerous and diverse with various innovations from each platform provider. Because of this, it is also related to the reach of viewing access. Viewing access that was originally public through media such as television (TV) and cinemas, is now rapidly evolving into media convergence that brings integrated viewing closer to the public. Therefore, an effort is needed to utilize technological advancements to help or alleviate the work of an institution that directly faces the impact of media convergence, one of which is the utilization of artificial intelligence (AI) for film censorship. In film censorship, there are many aspects or factors that influence whether a content or scene is deemed suitable for censorship, some of which are pornography and violence factors, of course, based on the rules established by Law Number 33 of 2009 concerning Film and Government Regulation Number 18 of 2014 concerning the Film Censorship Institution (LSF). In the film industry, the use of AI technology has experienced significant development. One important aspect in film production and distribution is the censorship process to ensure that the film complies with age classification guidelines, ethical norms, and applicable legal regulations. The film censorship process, which involves identifying and removing inappropriate content, has long been a time-consuming task involving significant human resources. With the development of Large Language Models (LLMs) that can be used to analyze video, images, sound, and text, there is potential to automate and improve the efficiency of the film censorship process while ensuring that the censorship quality is maintained.

CHAPTER I
INTRODUCTION

1.1. Background

The viewing access that was originally public through media such as television (TV) and cinemas, is now rapidly evolving into media convergence that brings integrated viewing closer to the public. Therefore, an effort is needed to utilize technological advancements to help or alleviate the work of an institution directly facing the impact of this media convergence, one of which is the utilization of Artificial Intelligence (AI) for film censorship.

Artificial Intelligence (AI) is a field developed through the combination of many subjects. In simple terms, AI involves giving machines human-like intelligence, simulating human thinking to assist people in problem-solving and to realize more sophisticated applications such as computer-assisted production and human life. Artificial intelligence, a branch of computer science, is considered one of the three most advanced technologies (genetic engineering, nanotechnology, and artificial intelligence) in the 21st century. AI modifies or adapts computers to mimic human actions (including predictions or robot control), thus making them more accurate. AI has experienced rapid development in the last 30 years and has been widely used in various academic fields in many countries. For computers to act or resemble humans, they must be equipped with knowledge and have the ability to reason. The application of AI is diverse, and the goals of AI systems can be divided into four categories:

1. System that can think like a human (Bellman, 1978)
2. System that can think rationally (Winston, 1992)
3. System that can act like a human (Rich and Knight, 1991)
4. System that can react rationally (Nilsson, 1998)

1.2. Problem Statement

The large number of films that require censorship involves significant manpower. Traditionally, this process entails playing films on a monitor/television/widescreen from data stored on CDs (Compact Discs), DVDs (Digital Video Discs/Digital Versatile Discs), flash drives, or hard drives, with each film being watched and evaluated manually by human censors. Therefore, there is a need for technological innovation to address this issue. To resolve this challenge, a technological solution is required that can alleviate and assist in the film censorship process.

1.3. Research Question

- How can technology alleviate and assist in the film censorship process?
- What testing methods can be used to facilitate and support the film censorship process?
- What achievements/results will be provided?

1.4. Research Purpose

- Conducting research on the use of AI technology that can alleviate and assist in the film censorship process.

- Testing the capabilities of Large Language Models (LLMs) to analyze video content in the film censorship process.
- Providing results in the form of recommendations for determining the age classification of a censored film.

1.5. Significance of The Study

The focus of this research is on how the use of AI can alleviate and assist in the film censorship process. This is achieved by utilizing the capabilities of Large Language Models (LLMs) to analyze video content in the film censorship process, thereby providing recommendations for determining the age classification of a censored film.

1.6. Scope Research

The scope of the research focuses on the utilization of LLMs, specifically Video-LLaVA, using Prompt Engineering.

Nusa Putra
UNIVERSITY
2002
AMOR DEUS, PARENTIUM, CONCERVIS

CHAPTER V
CONCLUSION

1. 1. Conclusion

Based on the implementation and testing conducted, the following conclusions can be drawn:

1. Testing Video-LLaVA Without Prompt Engineering: The output from Video-LLaVA testing without using prompt engineering is not optimal and does not meet the expected results.
2. Testing Video-LLaVA Using Single Age Classification Prompt Engineering: Testing with Single Age Classification Prompt Engineering produces optimal results, aligning with the expected output for each age classification or film censorship requirement.
3. Testing Video-LLaVA Using Simultaneous All Age Classifications Prompt Engineering: Testing with Simultaneous All Age Classifications Prompt Engineering successfully sorts age classifications based on the content/scenes of the film in a single process. This approach is significantly effective and efficient, yielding optimal results that meet the expected age classification outputs for film censorship. When a film is processed using this method, it provides an age classification output that accurately reflects the film's classification according to applicable standards or regulations.
4. Implications for Future Use: The results suggest that Video-LLaVA technology based on Large Language Models (LLM), when used with Prompt Engineering, could serve as an alternative solution for employing Artificial Intelligence. This approach is expected to assist and streamline the film censorship process, making the task more manageable and efficient for The Film Censorship Board of the Republic of Indonesia.

REFERENCE

Alayrac, J. B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., ... & Simonyan, K. (2022). Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, *35*, 23716-23736.

Anil, R., Dai, A. M., Firat, O., Johnson, M., Lepikhin, D., Passos, A., ... & Wu, Y. (2023). Palm 2 technical report. *arXiv preprint arXiv:2305.10403*.

Awadalla, A., Gao, I., Gardner, J., Hessel, J., Hanafy, Y., Zhu, W., ... & Schmidt, L. (2023). Openflamingo: An open-source framework for training large autoregressive vision-language models. arXiv preprint arXiv:2308.01390.

Bi, B., Li, C., Wu, C., Yan, M., Wang, W., Huang, S., ... & Si, L. (2020). Palm: Pre-training an autoencoding&autoregressive language model for context-conditioned generation. *arXiv preprint arXiv:2004.07159*.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., ... & Fiedel, N. (2023). Palm: Scaling language modeling with pathways. Journal of Machine Learning Research, 24(240), 1-113.

Cornsweet, T. (2012). *Visual perception*. Academic press.

Dai, W., Li, J., Li, D., Tiong, A. M. H., Zhao, J., Wang, W., ... & Hoi, S. C. (2023). Instructblip: Towards general-purpose vision-language models with instruction tuning. ArXiv abs/2305.06500 (2023).

Driess, D., Xia, F., Sajjadi, M. S., Lynch, C., Chowdhery, A., Ichter, B., ... & Florence, P. (2023). Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*.

Gong, T., Lyu, C., Zhang, S., Wang, Y., Zheng, M., Zhao, Q., ... & Chen, K. (2023). Multimodal-gpt: A vision and language model for dialogue with humans. *arXiv preprint arXiv:2305.04790*.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve?. *science*, *298*(5598), 1569-1579.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., ... & Sifre, L. (2022). Training compute-optimal large language models. arXiv preprint arXiv:2203.15556.

Huang, S., Dong, L., Wang, W., Hao, Y., Singhal, S., Ma, S., ... & Wei, F. (2023). Language Is Not All You Need: Aligning Perception with Language Models (arXiv: 2302.14045). arXiv.

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality?. *Frontiers in psychology*, *10*, 481030.

Indonesia. 2014. Peraturan Pemerintah Republik Indonesia PP No.18 Tahun 2014 tentang Perfilman.

Lembaga Sensor Film. (2024). LAPORAN KINERJA SEKRETARIAT LEMBAGA SENSOR FILM TAHUN 2023.

Li, J., Li, D., Savarese, S., & Hoi, S. C. BLIP-2: bootstrapping language-image pre-training with frozen image encoders and large language models. CoRR abs/2301.12597 (2023). 10.48550. *arXiv preprint arXiv.2301.12597*.

Lin, B., Zhu, B., Ye, Y., Ning, M., Jin, P., & Yuan, L. (2023). Video-llava: Learning united visual representation by alignment before projection. *arXiv preprint arXiv:2311.10122*.

Liu, H., Li, C., Wu, Q., & Lee, Y. J. Visual Instruction Tuning. CoRR, abs/2304.08485, 2023. doi: 10.48550. *arXiv preprint arXiv.2304.08485*.

NUSANTARA, B. Kecerdasan Buatan, Kini dan Akan Datang.

OpenAI, R. (2023). Gpt-4 technical report. arxiv 2303.08774. *View in Article*, *2*(5).

Pinker, S. (2003). *The language instinct: How the mind creates language*. Penguin uK.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Tsimpoukelli, M., Menick, J. L., Cabi, S., Eslami, S. M., Vinyals, O., & Hill, F. (2021). Multimodal few-shot learning with frozen language models. *Advances in Neural Information Processing Systems*, *34*, 200-212.

Turing, A. M. (2021). Computing machinery and intelligence (1950).

Workshop, B., Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., ... & Bari, M. S. (2022). Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*.

Ye, Q., Xu, H., Xu, G., Ye, J., Yan, M., Zhou, Y., ... & Zhou, J. (2023). mplug-owl: Modularization empowers large language models with multimodality. *arXiv preprint arXiv:2304.14178*.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*.

Zhao, X., Ni, Y., & Jia, H. (2017, October). Modified object detection method based on YOLO. In *CCF Chinese Conference on Computer Vision* (pp. 233-244). Singapore: Springer Singapore.

Zhu, D., Chen, J., Shen, X., Li, X., & Elhoseiny, M. (2023). Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*.