Type of Contribution:
Research Paper
Review Paper
Case Study

INTRO: JURNAL INFORMATIKA DAN TEKNIK ELEKTRO
DOI: 10.51747/intro.v4i2.426
ISSN 3025-602X

This article contributes to:
SUSTAINABLE DEVELOPMENT GOALS
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

Design and Implementation of a Raspberry Pi–Based Smart Mirror with Multimodal AI and IoT for Sustainable Outfit Decision Support

Apryadi Dwi Putra Tangalayuk[1], **Alvin Yuga Pramana**[1], **Franklin Jaya**[1], **Deny Wahyudi Asaloei**[1], **Citra Suardi**[1*], **Kasmir Syariati**[1]

[1] Department of Informatics (Makassar City Campus), Ciputra University Makassar, 90224, **Indonesia**

*citra.suardi@ciputra.ac.id

Abstract

The implementation of User-Centered Design (UCD) in mobile application development has a significant impact on user satisfaction. Based on a review of 30 scholarly articles, UCD has been shown to enhance satisfaction through a deep understanding of user needs, iterative design based on user feedback, and improvements in both functionality and intuitive user interface design. The evaluation methods commonly used to measure usability within the UCD approach include usability testing, heuristic evaluation, surveys and questionnaires, A/B testing, application usage data analysis, and cognitive walkthroughs. The findings indicate that applications developed using UCD principles tend to achieve higher adoption rates, more positive user reviews, and better overall usability. Therefore, the application of UCD is strongly recommended as an effective strategy for achieving optimal user satisfaction in mobile application development.

Keywords: Smart mirror, AIoT, Multimodal, Fast Fashion, Machine Learning

Article Info
Submitted: 2025-08-01
Revised: 2025-10-29
Accepted: 2025-12-25

Publisher
Universitas Panca Marga

1. Introduction

The development of technology has been very rapid, resulting in many innovations in various aspects of human life. Advances in communication technology allow humans to easily and quickly access important or latest information through devices such as smartphones, tablets, computers, and smart TVs [1]. As time goes by, technology develops, so a concept called IoT (Internet of Things) emerges [2]. The Internet of Things (IoT) is a concept where objects or systems can communicate with each other, collect and transfer data through the internet network without requiring direct human interaction. This technology allows various devices to connect and interact independently in various fields, including fashion [3]. IoT is one example of technological development that many people recognize. The rapid growth of information technology today has had a significant impact on various industrial sectors, including the fashion industry [4].

In the fashion world, there is a concept called fast fashion. Fast fashion is a business model that creates the impression that consumers can always look fashionable on a limited budget [5]. This concept helps people decide on their outfit every day, within a limited budget. Looking good is often one of the most important factors in building self-confidence. Along with the increasing need to look good in various situations, from work to social events, there is a common problem where people feel that the outfit they are wearing doesn't match or support their appearance.

In addition to the advantages of the fast fashion development industry that have been described previously. This industry also has negative impacts on life, especially in relation to the SDGs. The Sustainable Development Goals (SDGs), a new development framework that aims to tackle extreme poverty and address global challenges such as environmental degradation and climate change, has 17 goals and 169 targets [6]. The development of the fast fashion industry has a negative impact on SDG goal 12, responsible consumption and production. This trend encourages overconsumption behavior, where individuals continue to buy the latest fashion products without making use of the items within their lifespan [7], [8].

Previous studies on smart fashion systems have generally focused on either outfit recommendation or virtual try-on features using computer vision and image-based classification. Some smart mirror implementations have also been introduced, mainly emphasizing the display of multimedia information, outfit matching suggestions, or e-commerce integration. However, most of these systems rely solely on visual-based models, without incorporating additional modalities such as textual descriptions, voice interaction, or sustainability-related considerations in fashion decision-making.

Based on the description of the problems above, researchers are innovating by integrating the concept of fashion and AIoT to build a smart mirror system. The smart mirror concept is one of the technological innovations based on the Internet of Things (IoT) that functions as a mirror with the additional ability to display multimedia data, such as text, images, and videos [9]. Smart mirrors are designed to function not only as conventional mirrors but also as interactive devices that can present important real-time information, including outfit-related information.

In this case, the smart mirror can help users to get an objective assessment of the outfit they are wearing, as well as outfit recommendations that are more suitable for their style, occasion, or even weather conditions. The smart mirror designed by the researcher will be integrated with a multimodal model. Multimodal models are ML (machine learning) models that are capable of processing information from multiple modalities, including images, videos, and text. Multimodal systems integrate different modes of input to produce output [10]. In this research, multimodal is used by combining text and image datasets to become AI training data.

In contrast to previous smart mirror and outfit recommendation studies, this research offers several noteworthy differences. The system integrates a multimodal model that combines textual and visual inputs to assess outfits, allowing for a richer and more contextual evaluation process. It also incorporates voice-based interaction to provide real-time feedback, enabling users to communicate with the device naturally without relying on manual input. Furthermore, the proposed smart mirror aligns with sustainability objectives, particularly SDG 12, by encouraging responsible outfit selection and reducing unnecessary consumption. Through these contributions, the system serves not only as a technological innovation but also as a tool that supports sustainable fashion behavior.

2. Methods

The outfit analysis process on this smart mirror provides convenience to users by giving voice commands. The camera on the device will capture images of clothes, then the image data will be processed by the Gemini AI algorithm. After that, the smart mirror will provide analysis results in the form of assessments and recommendations delivered by voice or GTTS (Google Text To Speech), so users can easily understand the advice given without reading the text. The attractive visual appearance of the smart mirror is designed to attract users' attention and enhance their experience of using the device. With this feature, the experience of interacting with the smart mirror becomes more practical, intuitive, and efficient, and increases the convenience of choosing daily outfits.

2.1 Research Workflow

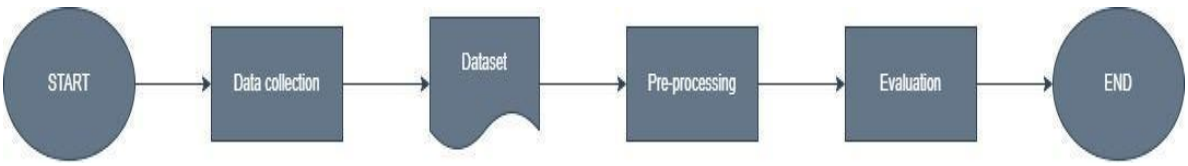

Figure 1. Research Workflow Diagram

The smart mirror development process begins with the search for datasets and relevant references to build a foundation for the outfit rating system. This step serves as the initial direction of the research, ensuring that the system's development starts with a clear technical and conceptual basis regarding how outfits should be evaluated. Building on this foundation, the next stage focused on collecting multimodal data in the form of outfit images and corresponding text descriptions. Image data was sourced from the rateyourstyle.com platform, while descriptions were generated through a generative prompting process using an AI model to align with the visual content. A total of 50 multimodal datasets consisting of image–text pairs were collected, and this dataset was then prepared as the training material for the multimodal model used in this study.

The image and text data collected during the data collection stage were then processed before being used in model training. This pre-processing stage was important to ensure that the data used was suitable for the multimodal model's requirements. For the image data, a series of adjustments were made, including normalization, resizing, and scaling, to ensure that the format and dimensions were compatible with the system's input requirements, while also maintaining consistency between data points. Meanwhile, text data was processed through sentence simplification, tokenization, and conversion to embedding representation. The embedding results from both types of data were then combined as multimodal input that could be understood and processed by the model.

During the training stage, the Gemini multimodal model was used as the core of the outfit analysis process on the smart mirror because it was able to accommodate image and text inputs in a single pipeline. This capability enabled the model to perform reasoning-based inferences so that it could assess subjective outfits while providing descriptive recommendations. The multimodal model is also more efficient than using two separate models for vision and text, and supports generative mechanisms that are suitable for the system's needs in generating outfit assessments and suggestions. During the training process, an early stopping mechanism was applied as a strategy to prevent overfitting, given the relatively small size of the dataset and the high risk of a decline in generalization capabilities. Early stopping halts training when performance on the validation data no longer improves, allowing the model to achieve better stability and convergence. This approach makes the system more adaptive and suitable for the characteristics of the dataset and research needs.

The evaluation stage was conducted to assess the system's performance from both technical and functional aspects. The technical evaluation focused on the quality of the multimodal model's output in generating consistent and relevant outfit assessments and recommendations. Functional evaluation is carried out on the smart mirror device to ensure that the integration of Raspberry Pi, camera, voice command module, and voice output through GTTS can run in real-time and responsively. The system is declared successful if it is able to provide a practical outfit assessment experience, can be understood through voice, and supports users in determining outfits efficiently.

2.2 System Workflow

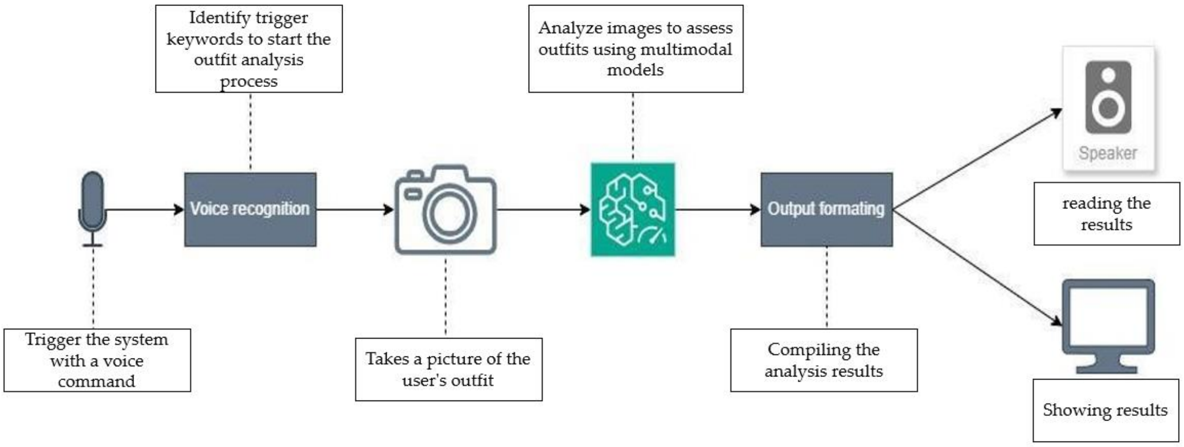

Figure 2. System Workflow Diagram

In this system, the process begins when the user gives a voice command that is recognized by the voice recognition module as a trigger for outfit analysis. Once triggered, the camera captures an image of the user's outfit and sends it to a multimodal model for analysis and assessment based on visual characteristics. The analysis results are then formatted for easy understanding before finally being delivered to the user through two feedback channels, namely in audio form via speakers and visually via a screen.

Based on the predefined search strategy, a total of 226 articles were retrieved from the Scopus database. These records were then screened to remove duplicates and inaccessible papers, resulting in 152 articles available for further review.

During the next screening phase, titles and abstracts were examined to assess relevance to the core keywords, such as User-Centered Design and mobile applications. As a result, 9 articles identified as systematic reviews or literature reviews and 65 articles that were not aligned with the research scope were excluded.

In addition, 12 articles could not be downloaded in full text and were therefore removed from the dataset. After completing the screening process, a total of 66 articles remained for the subsequent eligibility stage.

3.Results and Discussion

The following is a system design for the outfit rating feature in the system to produce sound output made based on the flowchart:

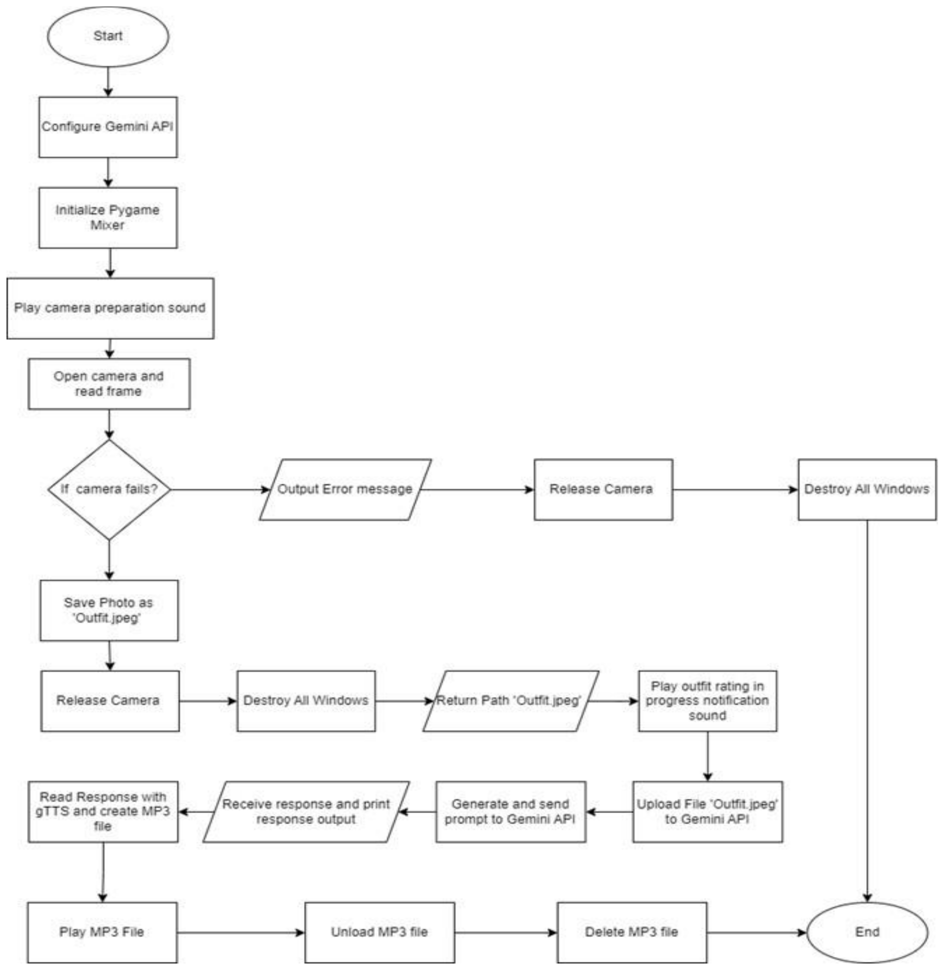

Figure 3. Outfit Rating System Flowchart

3.1 Presenting the Results

The figure above is a flowchart designed by researchers to create an AI system that performs outfit assessment that provides voice output. Starting from configuring the Gemini AI API, which processes the input given by the user in the form of sentences that are spoken directly that can trigger the smart mirror. The system will capture the input given by the user, then open the camera using OpenCV (infrastructure for computer vision applications) to detect the user's outfit. If the system fails to open the camera, an error message will appear. However, if successful, the image captured by the camera will be converted to 'Outfit.jpeg' by the system, and the camera is closed. Then, the image file obtained by the system will be detected using Gemini AI, and the system will provide output in the form of a voice saying that AI is assessing the outfit. After that, Gemini AI will perform a prompt to assess the image obtained by the system. Furthermore, when Gemini AI has finished analyzing, the resulting output will be printed as text. Then, it is read by gTTS, and the results read by gTTS are put into an MP3 file.

In the final stage, the system plays the MP3 file, which is the result of gTTS, and deletes it when the MP3 file has finished playing. The voice output produced by AI is the result of AI model training conducted by researchers. This training is done so that AI can recognize outfits properly. Thus, AI can assess the outfit according to the dataset given by the researcher. To train the AI model, researchers used 50 outfit image datasets and outfit rating text datasets for training. The image dataset is taken from the website rateyourstyle.com, while the text dataset which is a description and rating of the previous image dataset is obtained from Yeschat.AI. In this research, both datasets are labelled with an input dataset containing outfit images and an output dataset containing outfit descriptions and rating text, which can be seen in Figure 4, Sample dataset.

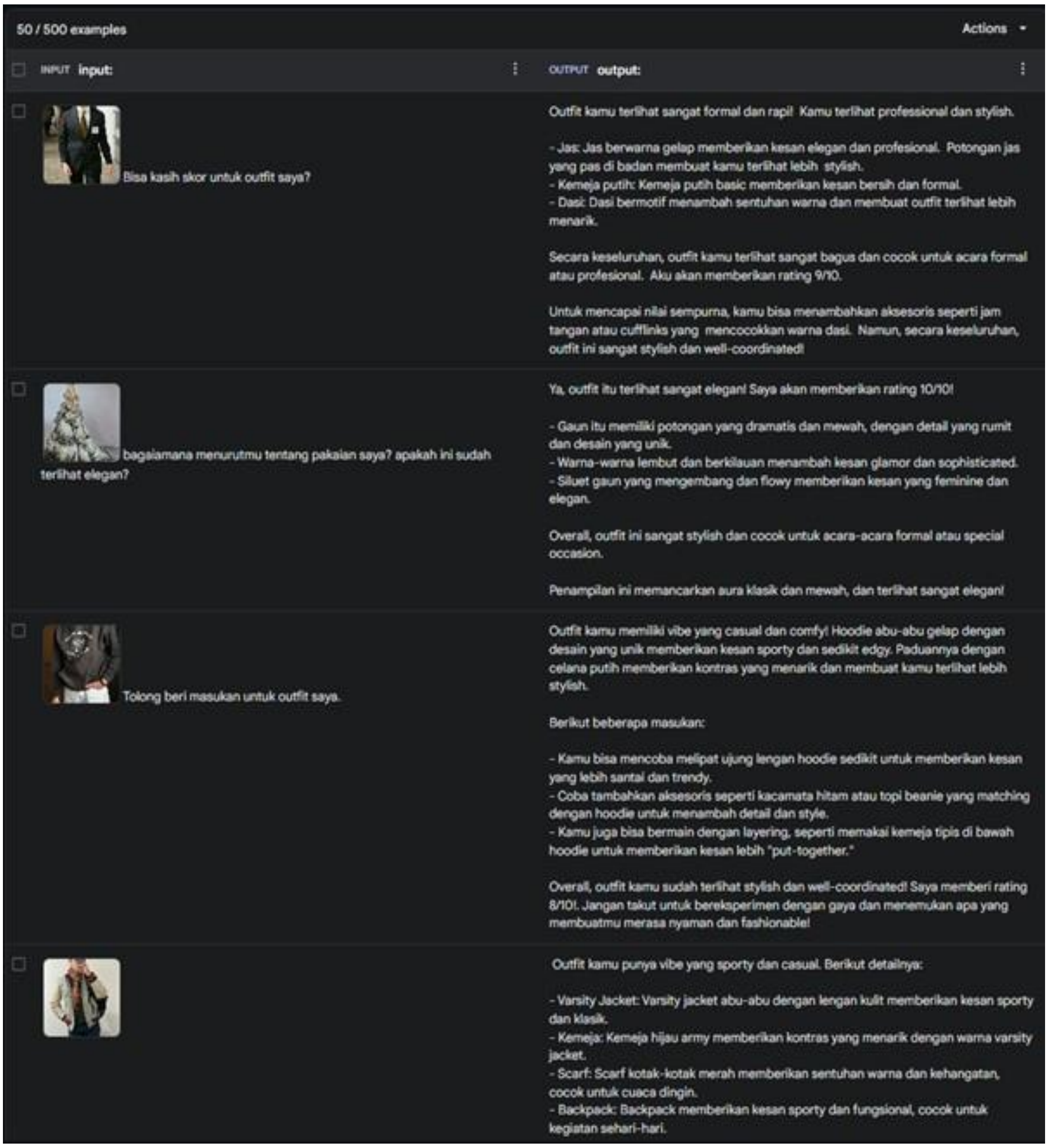

Figure 4. Sample Dataset

Every time the data training process is completed, it will produce an epoch; the entire epoch will produce a loss. Epoch is the number of iterations determined by the researcher to produce a loss graph, while loss itself is a scalar value that we try to minimize during model training. The lower the loss graph, the closer our prediction is to the actual label.

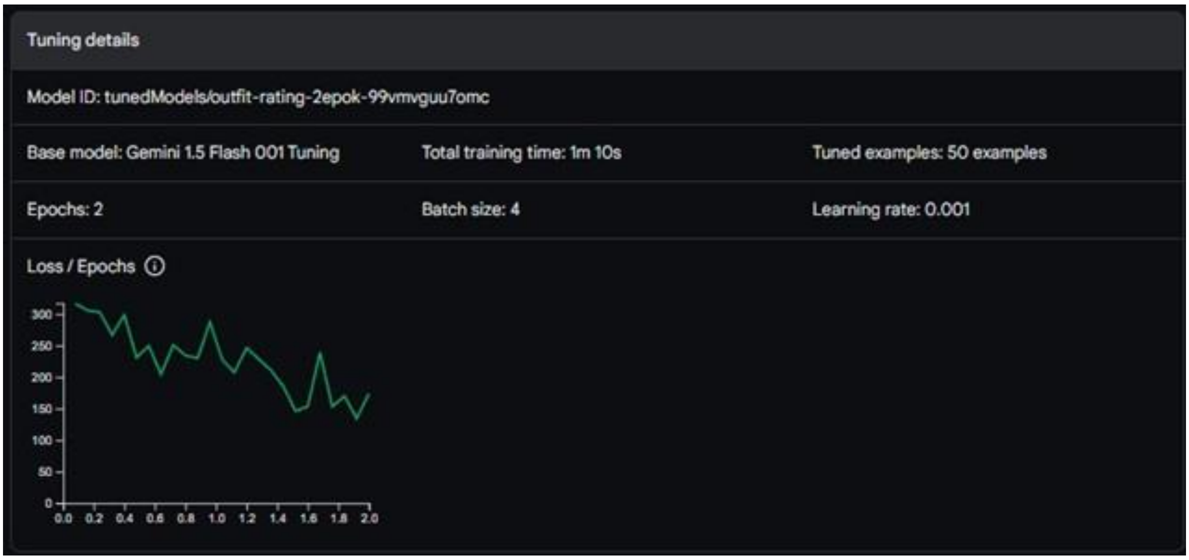

Figure 5. Graphic Loss Epochs 2

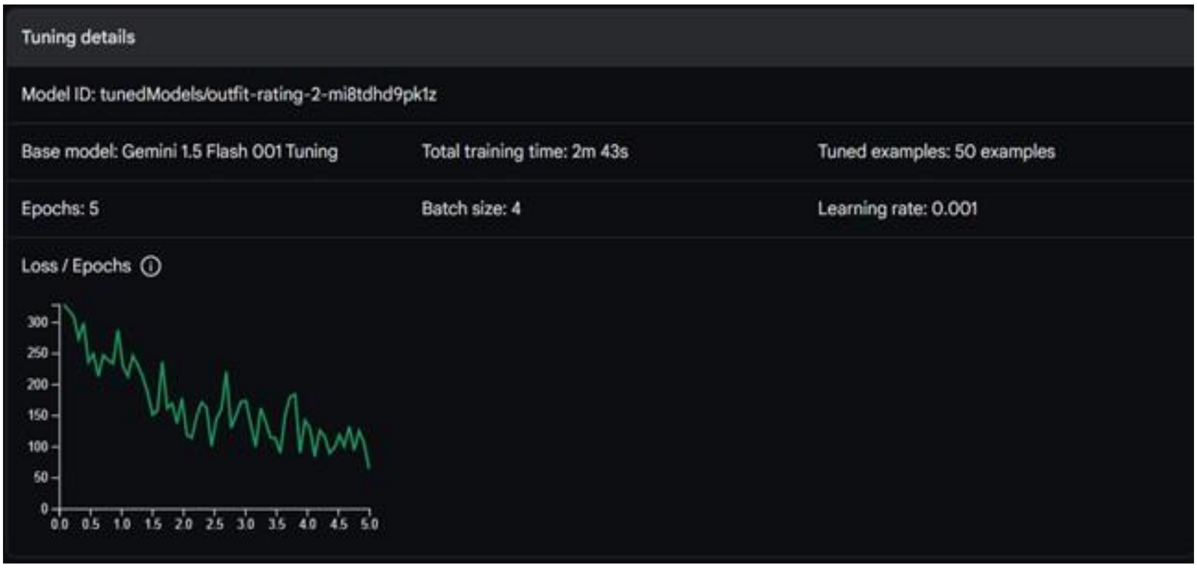

Figure 6. Graphic Loss Epochs 5

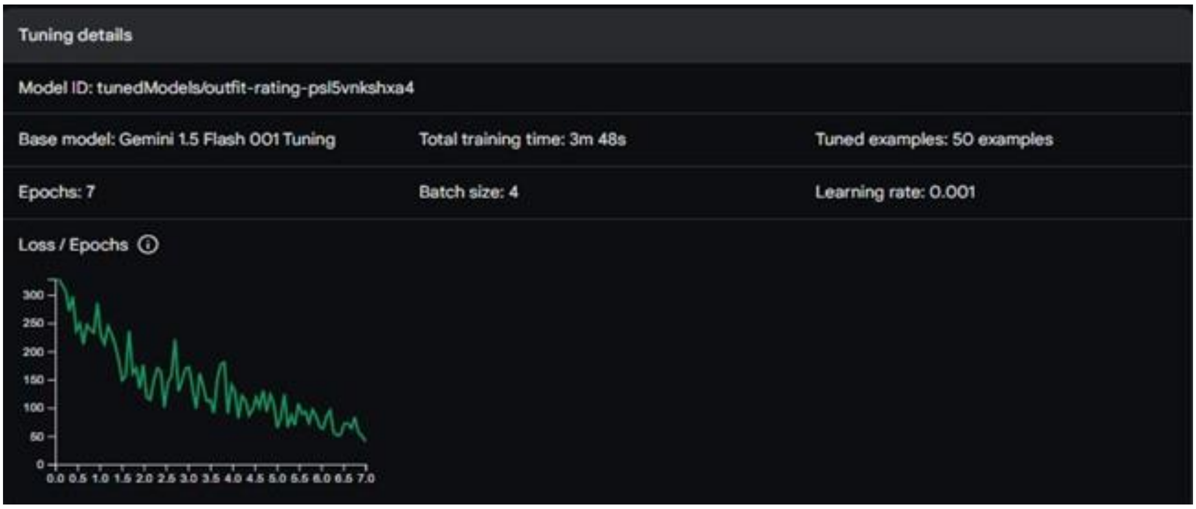

Figure 7. Graphic Loss Epochs 7

Each graph shows a loss/epochs plot, which illustrates the performance of the model during each epoch (cycle) of training. A decrease in the loss value indicates that the model is learning and getting better at predicting the output based on the given data. In all graphs, there is a decreasing trend in the loss value as the epochs increase, which indicates that the model is converging. In the model with 5 epochs, there is a more obvious decrease in the loss value compared to the 2-epoch session. With each additional epoch, the model gets more opportunities to learn from the data, resulting in better convergence and more accurate predictions.

This area of research encompasses a wide range of technological expertise, so to perform smart mirror classification goes beyond mere technical proficiency, requiring a deep understanding of user needs and the appropriate environment in which to use this technology. Classification models may suffer from limitations in interpretation and relevance due to a lack of comprehensive understanding of user needs and the associated cultural context. It is imperative for researchers to engage in collaborative efforts with experts in technology and user psychology to ensure the fidelity and convenience of the features developed.

Given the many constraints and complexities in this study, it is imperative for researchers to have a comprehensive point of view when conducting feature classification and smart mirror design. To improve accuracy and navigate ethical issues, a balance between technological rigor, sensitivity to user needs, data variation, and an adaptive development methodology is required.

This research is related to Sustainable Development Goal (SDG) 12: Responsible Consumption and Production, particularly in terms of more efficient and sustainable fashion consumption. The outfit assessment feature on the smart mirror can support users in choosing clothes based on more measurable knowledge and preferences, thereby reducing impulsive consumption behavior and the potential for excessive clothing purchases. By helping users maximize their existing outfits, this system plays a role in extending the use cycle of clothing, which in turn can reduce textile waste and the environmental impact caused by the fashion sector. This process contributes to more responsible consumption behavior, in line with SDG 12's target of promoting sustainable consumption patterns and raising awareness of the impact of resources in everyday life.

4. Conclusion

Smart Mirror is an IoT system integrated with artificial intelligence (AI) to combine IoT technology with outfit assessment features, thereby improving user effectiveness and convenience in choosing clothes. This study successfully implemented a multimodal AI model on a smart mirror using 50 image and text datasets as training data. Based on the training results in Figures 4 and 5, the more epochs given, the greater the chance for the model to learn data patterns better. In Figure 6, the training process stopped at the 7th epoch due to the early stopping mechanism, which is automatic termination when there is no significant decrease in loss value to prevent overfitting while saving time and resources. The implementation results show that the system is capable of providing quick and accurate outfit assessments and recommendations through voice output, thereby helping users determine suitable outfits in a short time. This integration also shows the potential contribution to SDG 12 through support for more responsible and sustainable fashion consumption behavior.

This research contributes scientifically to the development of a multimodal AI-based smart mirror system capable of visual-textual reasoning and generating generative recommendations in the context of outfit assessment. Future research development can focus on increasing the number and variety of datasets so that the model can recognize more styles and user preferences in a more inclusive manner. In addition, improving the algorithm using more adaptive and advanced AI models can increase the accuracy of assessments. The potential for integration with external fashion applications or platforms is also an area of exploration to provide recommendations that are more relevant to current trends, while strengthening the system's contribution to supporting more sustainable fashion consumption.

References

[1] M. Warschauer and T. Matuchniak, “New Technology and Digital Worlds: Analyzing Evidence of Equity in Access, Use, and Outcomes,” Review of Research in Education, vol. 34, no. 1, pp. 179–225, Mar. 2010, doi: 10.3102/0091732x09349791.

[2] M. N.O. Sadiku, Ed., “Internet of Things,” in Emerging Internet-Based Technologies, CRC Press, 2019, pp. 1–15. doi: 10.1201/9780429020247-1.

[3] M. M. Yusri et al., “Smart mirror for smart life,” in 2017 6th ICT International Student Project Conference (ICT-ISPC), IEEE, May 2017, pp. 1–5. doi: 10.1109/ICT-ISPC.2017.8075339.

[4] Julianto, “Analisis Digital Fashion Dalam Perspektif Teknologi, Sistem Informasi, Dan Bisnis,” Adopsi Teknologi dan Sistem Informasi (ATASI), vol. 2, no. 1, pp. 71–78, Jun. 2023, doi: 10.30872/atasi.v2i1.726.

[5] F. Caro and V. Martínez-de-Albéniz, “Fast Fashion: Business Model Overview and Research Opportunities,” 2015, pp. 237–264. doi: 10.1007/978-1-4899-7562-1_9.

[6] Goal 9: Industry, Innovation and Infrastructure. Taylor & Francis, 2019. doi: 10.4324/SDG-G009.

[7] K. Niinimäki, G. Peters, H. Dahlbo, P. Perry, T. Rissanen, and A. Gwilt, “The environmental price of fast fashion,” Nature Reviews Earth & Environment, vol. 1, no. 4, pp. 189–200, Apr. 2020, doi: 10.1038/s43017-020-0039-9.

[8] O. Naumova and M. Naumova, “SUSTAINABLE DEVELOPMENT AND GREENWASHING: AN ANALYSIS OF SUFFICIENT CONSUMPTION IN THE FASHION INDUSTRY,” Economics, Finance and Management Review, no. 1(17), pp. 16–24, Mar. 2024, doi: 10.36690/2674-5208-2024-1-16.

[9] D. Gold, D. Sollinger, and Indratmo, “SmartReflect: A modular smart mirror

application platform," in 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE, Oct. 2016, pp. 1–7. doi: 10.1109/IEMCON.2016.7746277.

[10] S. Oviatt, "Multimodal Interfaces," 2012, pp. 405–430. doi: 10.1201/b11963-22.