

CHAPTER I

INTRODUCTION

1.1 Background

Social media is a highly popular digital platform that has become an integral part of modern society. According to Cahyono (2016), it is a platform where people create personal profiles and connect with others to communicate. It allows two-way interaction as long as there is internet access. Every day, millions of images are shared on platforms like WhatsApp, Instagram, and X (formerly Twitter). As social media use grows, images play an increasingly important role in communication. Captions help explain these images and are also used to grab attention and boost engagement, especially for users who want to increase their presence and influence online.

In an effort to improve efficiency in daily life, humans consistently seek and develop various methods to simplify their activities (Mashlahaha & Arifin, 2023). One major innovation helping this is Artificial Intelligence, which is advancing rapidly. AI technologies, such as Generative AI, enhance productivity and creativity by creating various forms of creative content from generating ideas and automating design to manipulating images. As a result, AI plays an important role in boosting efficiency and innovation, especially in the creative and digital industries (Wijaya et al., 2024).

In the creative industry, there exists a psychological phenomenon known as *creative block*, which refers to a condition in which individuals struggle to generate new ideas or inspiration. This state often leads to a diminished ability to produce work that meets expectations or creative goals. *Creative block* can affect anyone

engaged in creative activities, including artists, writers, photographers, and digital content creators. It may be triggered by various factors such as a lack of understanding of the subject matter, mental fatigue, or excessive work-related pressure. Individuals experiencing this condition often report a loss of motivation and difficulty in formulating fresh ideas, which can negatively impact both productivity and the quality of their work. If left unaddressed, *creative block* may even become a significant barrier to career advancement (Nugroho et al., 2019).

This condition is also reflected in the results of a survey I conducted involving ten respondents from diverse backgrounds, including university students, content creators, photographers, and others. 90% of the participants stated that they often add captions when uploading photos to social media. However, many of them also admitted to frequently experiencing confusion when composing captions. Some even mentioned that this confusion leads them to delay uploading their content.

During periods of *creative block*, many creators and photographers still wish to upload content to social media, but often face difficulties in composing appropriate captions. This condition can disrupt the consistency of audience engagement and content publication rhythm. Questionnaire results show that 90% of respondents have experienced confusion when writing captions, which they managed by seeking inspiration from other users' posts, using AI-based automatic caption generators, or uploading content without captions. The duration of a creative block can vary significantly among individuals some are able to overcome it within a few hours, while others may require days or even months (GoodTherapy, 2019). During this period, many content creators, especially those with a large audience, tend to avoid prolonged absence from social media platforms. This

highlights the need for tools that can help them stay active and creative, even during creative block.

Previous research has demonstrated the feasibility of implementing image captioning technology for mobile applications using cloud-based architectures. Taufiqur & Hatta (2022) developing an image captioning application for smartphones using CNN and Transformer architectures with the Flickr8K dataset and achieving a BLEU score of 0.31. The studies adopted server-side processing to overcome mobile device computational limitations.

Based on the aforementioned background, this study focuses on the development of an Android application named "Descripix." The name *Descripix* is derived from the combination of two English words: "*describe*," meaning to provide a description, and "*pix*" or "*pict*," which are informal abbreviations of the word "*pictures*." The Descripix application is designed to automatically generate image descriptions by leveraging the LLaMA Vision model.

To improve the quality of the generated captions, this research also incorporates image metadata such as the author, date of capture, location, camera device, and other technical parameters as additional input in the caption generation process. By utilizing this information, the model is expected to produce descriptions that are more contextually relevant and meaningful. To allow users to save and view the history of generated captions, the system is supported by a backend server. This approach adopts a client-server architecture, where most of the data processing takes place on the backend, ensuring that the Android application remains lightweight and responsive for users.

1.2 Problem Identification

Based on the background, there are several problem have been identified, namely:

1. Social media users frequently experience difficulties in composing appropriate captions for the images they upload, especially during periods of creative block.
2. Creative block reduces productivity and consistency in sharing content on social media, this may negatively impact audience engagement and the credibility of content creators in digital spaces.
3. Currently, there is not many mobile applications specifically designed to assist users in generating image captions automatically, based on both visual context and image metadata such as time of capture, location, and camera device.
4. There is a lack of real-world implementation of the LLaMA Vision model for image captioning tasks in mobile applications.
5. There is a need for a lightweight and responsive Android application, enabling users to generate image captions efficiently and effectively.

1.3 Problem Scopes

The limitations of the problem limits applied in this study in order to avoid misunderstanding and to be efficient in its implementation are as follows.

- 1 The model used is the Llama Vision model taken from the Groq.com platform.
- 2 The application only works on devices with Android operating system version 15 and above.

- 3 Images that are processed for captioning are limited to general image types, without specialization to certain categories.
- 4 The application will rely on the REST API for communication with the server, so application performance can be affected by internet connection.
- 5 Application usability testing is only carried out on a small scale, that is on several users as test respondents.
- 6 This share image and caption feature can only be used if the destination application grants permission

1.4 Research Problems Statements

The following is the formulation of the problem that is the focus in this research:

1. How to design, integrate Llama Model into Descripix android application?
2. What are the implementation results of the Llama Model integration design into the Descripix application?
3. What are the results of usability testing of the Descripix application to assess ease of use and user satisfaction?

1.5 Research Objectives

Based on the problem formulation described, the objectives of this study are:

1. Design and integrate the Llama Model into the Descripix Android application.
2. Implement the Llama Model integration design into the Descripix Android application according to the architecture that has been designed.

3. Obtain the results of usability testing of the Descripix application to assess the level of ease of use and user satisfaction.

1.6 Research Result Benefits

The expected benefits of this research include.

- 1 Gain practical implementation of LLaMA Vision model for image captioning in Android applications.
- 2 Improve Understanding of LLaMA Vision for Vision-to-Text Tasks
- 3 Introduce Practical Use of In-Context Learning in Mobile AI Integration.

This project showcases implementation of In-Context Learning, serving as an example of how LLMs can be used without fine tuning the model.

- 4 Provide references for further research related to the implementation of the LLaMA Vision model for image /captioning in Android applications.

