

AI and Metadata: Bridging the Gap between Digital Images and Human Understanding

Marina Morgan

Florida Southern College, 111 Lake Hollingsworth, Lakeland, FL, 33547, United States

Abstract

As digital collections continue to grow, artificial intelligence (AI) tools have been used to generate titles, descriptions, and keywords for digital collections in libraries. This poster explores whether four AI models (Claude, Gemini, Copilot, and ChatGPT 4.0), can effectively provide descriptive titles, and whether librarians and archivists can bridge the gap between digital images and human understanding, enhancing metadata quality and usability. This poster also showcases each tool's prospects for AI-driven metadata for title generation, highlighting the transformative potential of AI within digital collections.

Keywords

metadata, AI, image description, machine learning, digital collections, library, archives

1. Introduction

There are many applications of AI in libraries, from using AI tools in the classroom for research, image generation, crafting social media posts [1], to generating metadata for image descriptions [2]. In the recent past, there has been an increase in using AI in library operations, particularly “enhancing routine library operations and services” [3]. This poster will explore whether four AI models (Claude, Gemini, Copilot, and ChatGPT 4.0), can effectively describe digital images and improve the quality of existing titles in a college digital collection.

For this purpose, I focused on Child of the Sun Photographs digital collection, one of many digital collections included in the McKay Archives Digital Collections [4]. This collection was created from a selection of black-and-white and color photographs, negatives, and slides documenting the design and construction of the Frank Lloyd Wright buildings, also known as the Child of the Sun, on the Florida Southern College campus. Included in this digital collection are images of blueprints and people associated with the building construction. Additionally, the four AI tools mentioned above were compared to understand which one would produce quality metadata for titles and image descriptions for the Child of the Sun Photographs digital collection [5].

2. Methodology

Three sets of photographs were carefully selected from the McKay Archives digital collections, each set including ten images depicting different aspects of the college's history. While each image already had a title and an image description, Claude, Gemini, Copilot, and ChatGPT 4.0 were used to test how the generated titles and image descriptions compared with the existing ones without incorporating the historical context.

The first set of photographs represents different variations of the aerial view of the college, the second set features notable figures associated with the college, while the third set includes campus buildings. The images uploaded to each AI tool had the same prompt: “Please provide a title for this image”.

2.1. AI generated titles for set 1, aerial view of the college

For the aerial view of the Florida Southern College's campus set of images, Claude responded with the following titles: “Aerial View of a Town,” “Aerial View of Florida Southern College Campus,” “Aerial View of Florida Southern College Campus and Lake Hollingsworth”, or “Aerial View of Florida Southern College's Water Dome and Esplanades” clearly identifying distinctive architectural features of the Florida Southern College campus. For the same set of images, besides identifying the college, Gemini included a time period “Aerial View of Florida Southern College Campus, Early 20th Century”, “Aerial View of Florida Southern College's Bartow Campus, Mid-20th Century,” or “Aerial View of Florida Southern College Campus, Circa 1950s”. However, for one of the images [Aerial of Florida Southern College campus variation 4], the proposed title was “Aerial View of Florida Southern College's Bartow Campus, Mid-20th Century”, wrongly identifying the location of the campus. Copilot, on the other hand, with the same prompt, suggested more figurative titles, such as “Serene Academia: Aerial View of Florida Southern College,” “A Glimpse from Above: Florida Southern College in the 20th Century,” “Bird’s-Eye Beauty: Florida Southern College Campus in Black and White”, or “Aerial View of a Modernist Campus”. However, ChatGPT4o was the closest to provide a similar title to the one our digital collection has for this image: “Aerial View of Florida Southern College Campus,” “Aerial View of Florida Southern College Campus with Frank Lloyd Wright's Designed Buildings,” “Aerial View of Florida Southern College Campus by Lake Hollingsworth,” or “View of the Esplanade and Walkways at Florida Southern College”.

2.2. AI generated titles for set 2, notable figures associated with the college

The second set of photographs tested depicted persons important to the history of the college, including past presidents, deans, or Frank Lloyd Wright, the renowned architect who designed the Florida Southern College campus. For one image depicting philanthropist Annie Merner Pfeiffer and Dr. Ludd M. Spivey (President of Florida Southern College, 1925-1957) at Annie Pfeiffer Chapel dedication, using the same prompt, Claude provided the title “Enduring Affection or A Lifetime Together” to reflect “the evident bond and likely many years shared by this couple”. Gemini’s title proposal for the same image was “Text from Florida Southern College Archives mentioning Lincoln University”, while Copilot’s was “Ceremonial Handshake”. ChatGPT’s generated title however, was “Academic Ceremony at Florida Southern College”. Another image depicting Clinton Vane McClurg (College Trustee Emeritus), Frank Lloyd Wright, George Jenkins (founder of Publix Supermarkets), and Dr. Ludd M. Spivey (President of Florida Southern College, 1925-1957), was not accurately described. Titles such as “Group Portrait of “Distinguished Guests at Florida Southern College Dinner Event” (ChatGPT), “Distinguished Gentlemen at Florida Southern College Event” (Claude), “Group Portrait of Four Men in Formal Attire” (Copilot) did not accurately identify public figures in the image uploaded and processed. Moreover, Gemini provided a disclaimer: “As an AI language model, I cannot

process images directly. Therefore, I can only provide a title based on the text extracted from the image and my understanding of Florida Southern College and Lakeland, Florida”.

2.3. AI generated titles for set 3, campus buildings

The last set tested contained images depicting various buildings, either already built or under construction. One image titled in the digital collection “Construction of the E. T. Roux Library: variation 2” had the following titles provided: “Construction Site at Florida Southern College” (Claude) and “I can’t definitively say what’s happening in the image, but here are some possible titles and descriptions: Construction Workers on a Courthouse, Building a Courthouse in Florida, or Early 20th Century Construction” (Gemini). However, Copilot and ChatGPT gave similar titles, wrongly attributed to a different building on campus: “Building the Future: The Annie Pfeiffer Chapel Under Construction” and “Construction of the Annie Pfeiffer Chapel at Florida Southern College”. Another image depicting the exterior of the Benjamin Fine Administration Building was provided the following titles: “Walkway Leading to the Water Dome at Florida Southern College” (ChatGPT), “Frank Lloyd Wright Building on Florida Southern College Campus” (Claude), “Florida Southern College Campus Building, Mid-20th Century” (Gemini), or “Florida Southern College Campus Grounds” (Copilot). Other titles provided for this set by Claude, Gemini, Copilot, and ChatGPT 4.0, aim to be general because the AI tools do not have access to detailed contextual information about each specific building or scene depicted. Their approach is to capture the main subject matter, such as a notable building or area on the Florida Southern College campus, without over-specifying elements that might not be immediately clear from the image alone.

3. AI tool comparison

The three sets of images carefully selected for this comparison had strong ties with the history of Florida Southern College. The first set of images depicted an aerial view of the college, the second pictured notable figures associated with the college, and the third was various buildings in campus. The model with the most comprehensive title was ChatGPT 4.0, even though on two occasions it relied on the previous prompt to generate a subsequent response. On the other hand, the model with the least accurate title description was Claude.

It is evident that without access to contextual or historical information, AI tools primarily rely on visual analysis for image or title descriptions. AI models are trained on large datasets of images [6] which enable them to generate descriptions based on visual features. However, without additional data sources that provide historical context, such as the dedication of a chapel, or the construction of a new building on campus, Claude, Gemini, Copilot, and ChatGPT 4.0 lacked the ability to understand and convey the background, significance, and narrative associated with the image. As a result, the content generated by Claude failed to capture the deeper historical or cultural context that gives the image its true meaning.

4. Conclusion

AI-generated titles from an image might not contain the history behind the image because AI models primarily rely on visual analysis and pattern recognition without access to contextual or historical information. AI models such as Claude, Gemini, Copilot, and ChatGPT 4.0 are

trained on large datasets of images and captions, allowing them to generate descriptions based on visual features such as actions and scenes depicted in the image. However, without additional data sources that provide historical context, these tools lacked for the most part the ability to understand the images' significance. As a result, the generated title only partially accurately described the image. Overall, the accuracy and relevance of the titles generated are limited by the AI's training data and its ability to interpret visual cues without additional context.

In conclusion, librarians and archivists play a fundamental role in bridging the gap between digital images and human understanding. While AI models excel at generating preliminary descriptions based on visual features, they lack the ability to provide the historical and contextual depth that human expertise brings. By combining AI's efficiency in analyzing large volumes of images with the specialized knowledge of librarians and archivists, we can enhance both the quality and usability of metadata. In this way, AI serves as a complementary tool, providing a foundation that librarians and archivists can refine and enrich, ensuring that digital resources are not only accurately described but also contextually meaningful.

References

- [1] E. Udell, The World of AI: How libraries are integrating and navigating this powerful technology, *American Libraries* 55 (2024), 22.
- [2] I. Reiche, The viability of using an open source locally hosted AI for creating metadata in digital image collections, *Code4Lib Journal* 56 (2023)
- [3] A. Cox, The Ethics of AI for Information Professionals: Eight Scenarios, *Journal of the Australian Library & Information Association* 71 (2022) 201–214. doi:10.1080/24750158.2022.2084885P.
- [4] McKay Archives Digital Collections, 2024. URL: <https://archives.flsouthern.edu>
- [5] Child of the Sun Photographs (1937-1957, 2006-07), 2024. URL: <https://archives.flsouthern.edu/digital/collection/flwp>
- [6] M. Breeding, AI: Potential Benefits and Concerns for Libraries, *Computers in Libraries*, 43(2023), 17–19.