

Knight Foundation School of Computing and Information Sciences

Spring 2024 Senior Design Project

Quantifying Copyright Infringement in Generative AI

Students: Gabriel Arrano , Gabriel Igliozi, Uriel Juarez, Alejandro Adorno

Mentor: Wenbin Zhang, Jocelyn Dzuong

Instructor/Faculty: Masoud Sadja, Florida International University

PROBLEM

The capacity to produce and create inventive results, generative artificial intelligence (AI) has revolutionized a wide range of industries, including music, writing, design, and the arts. However, the increasing popularity and use of generative AI presents a significant difficulty in regulating and evaluating copyright ownership of the results produced by AI. The complexity of AI-generated works is beyond the scope of traditional copyright rules and frameworks, creating uncertainty and legal problems around ownership, usage rights, and attribution.

SYSTEM DESIGN

The quantification program was designed using Google Colab which provides a ready-made setup without additional configuration. Due to Colab's Jupyter style notebook interface, Python was used as the primary programming language. The program connects to Bing's Image Creator API and Google's Reverse Image Searcher for image creation and reverse lookup respectively. The collected images are compared using similarity measures defined in PyPi's image similarity measures package. For additional comparison OpenCV is used. Optional module includes image comparison using OpenAI's CLIP neural network model as an additional verification step.

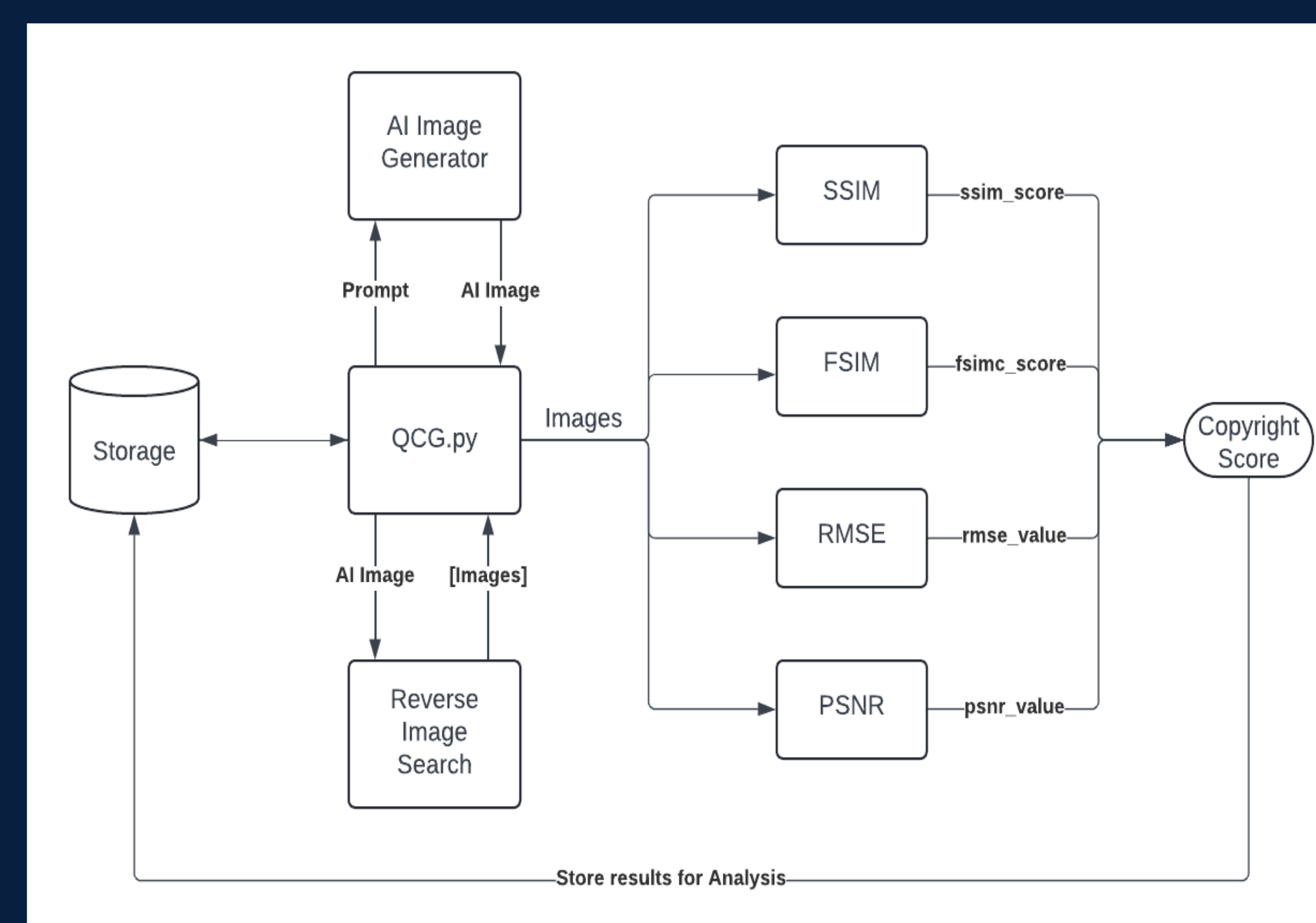
REFERENCES

1. Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2024, February 19). *Can AI-Generated Text be Reliably Detected?* arXiv. Retrieved April 15, 2024, from <https://arxiv.org/pdf/2303.11156.pdf>
2. Kanchana Kariyawasam, Artificial intelligence and challenges for copyright law, International Journal of Law and Information Technology, Volume 28, Issue 4, Winter 2020, Pages 279–296, <https://doi.org/10.1093/ijlit/eaad023>
3. Quang, Jenny. "Does Training AI Violate Copyright Law?" Berkeley Technology Law Journal, vol. 36, no. 4, 2021, p. 1407. LexisNexis. Accessed Friday January 2024.

REQUIREMENTS

The quantification program uses Google Colab in both its frontend and backend implementation. Colab allows for writing and executing Python in a web browser so a connection is required. The program also connects to Bing's Image Creator API for image creation so an API key is also needed. Lastly, a storage drive can be mounted for retention of images and results.

IMPLEMENTATION



SUMMARY

The focal point of this research is the quantification of copyright infringement specifically concerning image-generated content produced by AI. The study proposes four fundamental metrics for evaluating copyright infringement: SSIM, FSIM, RMSE, PSNR. Additionally, the research introduces a method for identifying points of interest within images, facilitating the matching of comparable features between AI-generated and copyrighted images to reinforce similarity scores.