Case Study: State-industry collaboration in the State Archives of Hawai'i to design and implement an AI technology for indexing and describing video records

Richard Arias-Hernandez[1]

Educational applications: This case study can be used in introductions of AI uses for archives, and more specifically for non-textual records. It illustrates the use of AI to enrich the description and metadata of video-records at the item level in order to improve on their access and discoverability. It motivates, exemplifies, and prompts discussions of AI for automatic speech recognition, machine translation of transcripts, automated time-coding of scenes, object/face recognition, and multi-modal summarization of video records. It also covers design and development of GPT bots to support reference in archives using only data from records as well as practical considerations for archives, such as constrained budgets, partnerships with industry, and the need for low-code/no-code solutions.

Educational topics: AI for non-textual records (video), types of AI/ML for video-archives, AI for access and archival reference, management and collaboration in AI projects for archives[2].

About: This case study is part of a series of learning materials developed by InterPARES Trust AI[3] researchers and educators to train archival professionals and students to effectively leverage artificial intelligence in their archival work. The final draft was completed on August 13th, 2024. It has a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International BY-NC-SA 4.0 license, which requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only. If others modify or adapt the material, they must license the modified material under identical terms.[4]

Since January 2022, Adam Jansen, State Archivist of Hawai'i, has been leading an InterPARES project titled "Increasing Access to Photos, Videos and Social Media records through AI-generated Descriptive Metadata" (InterPARES, 2022). According to the abstract:

"The study explores how AI and Machine Learning can be used to increase discoverability and improve access to non-textual records, such as digital photographs, films and videos, and social media posts. The study explores how AI can be used to analyze a digital photograph and add

[1] Associate Professor of Teaching, School of Information, University of British Columbia. richard.arias@ubc.ca
[2] Educational applications map to a Body of Knowledge proposed by InterPARES researchers for AI/ML for the archival professionals. https://docs.google.com/document/d/1UsjkkkGeSJrgCDJGASCAy5q0uo_ZkQpzi_Ch8XUcqYw/edit?usp=sharing
[3] This case study is an outcome of InterPARES Trust AI, an international research partnership led by Drs. Luciana Duranti and Muhammad Abdul-Mageed, University of British Columbia, and funded by the Social Sciences and Humanities Research Council of Canada (SSHRC). https://interparestrustai.org/

searchable descriptive elements that may be missing from the record creator. The study will determine to what extent AI can improve user access to the content of films and videos - and may also help determine how AI can improve archival description of films and videos - through: automatic speech recognition; machine translation of film dialogue; automated time-coding of scenes; multimodal video summarization. The study explores how AI and Machine Learning can be used to increase discoverability of Social Media posts through content and semantic analysis of text, images, and videos." (InterPARES, 2022). Jansen's study benefits from millions of digital born-records and digitized records now accessible from the Hawai'i State Digital Archives website (Hawai'i State Archives, 2024).

Since 2023, a sub-series of digital video records has become the target of experimentation in Jansen's project: video recordings of State of Hawai'i legislative hearings by the State Senate (Hawai'i State Senate, 2024). The videos are currently published on the Senate's YouTube channel (Hawai'i State Senate, 2024) and they are downloaded by the State Archives for arrangement, description, storage, access, and preservation purposes. They can be accessed by anyone interested in Hawaiian legislative research. There are thousands of bills introduced every year into the legislature and a short time for the Senate to conduct hearings and organize documents. This results in documentation of the different bills being placed in multiple containers, documents and repositories, with the oral testimonies being only video-recorded and placed on YouTube (rather than in a State-controlled system). This creates a pain point for researchers trying conducting research on a specific topic. Even if multimodality electronic records (documents, videos) are rearranged by bill and made accessible from the same website, legislative researchers find themselves rather limited in their keyword searches to the scant metadata which makes records only identifiable by the title of a specific bill, the identifier of the specific committee, and a very short description of the records explaining what the bill entails. To improve on the access and discoverability of these records, especially the video records of the bill hearings, Jansen struck a partnership with Microsoft Government Services and local consultants.

The partnership aimed to use the video recordings of the bill hearings as a test bed for several of the existing AI/ML technologies currently being developed by Microsoft in their suite of Azure AI services in the cloud, namely their Video Indexer and their OpenAI Service.

Microsoft Azure AI Video Indexer "is a cloud and edge video analytics service that uses AI to extract actionable insights from stored videos. Enhance ad insertion, digital asset management, and media libraries by analyzing audio and video content—no machine learning expertise necessary." (Microsoft, 2024). The video indexer uses multiple AI/ML technologies to analyze video and audio content, identifying objects and people (i.e., object recognition and tracking, face recognition models), scenes (i.e., scene detection), spoken words (i.e., speech-to-text-transcription; language identification; language translation; keyword extraction), and emotions (i.e., emotion/sentiment detection). Video indexer takes the output of running all these models to automatically extract metadata, such as: timestamps, transcript, speaker identification, sentiment, and topic detection. This metadata then enriches the description of this particular digital object, gets indexed, and enhances the possibilities for access and

discoverability of these records by searching by keywords related to topics being discussed in the bill hearings, names of individuals present and/or speaking in the video, words used in the hearings, and even the sentiment expressed by speakers.

Microsoft Azure OpenAI Service is a cloud-based platform that provides access to language and image AI models. It takes advantage of Microsoft's partnership with OpenAI to use technologies such as GPT-4o and DALL-E for Generative AI applications, such as Copilot (Microsoft, 2024b).. This service was used by Jansen's team with the intention to create a "Copilot-like" service that would be trained on the data and metadata generated by the Video Indexer. Once video records of the bill hearings got processed by the Video Indexer and their description and metadata got enriched, then the outcome data was used to train a GPT model: the "Hawaii Legislative GPT model." The model then grounds a type of archival reference service in which legislative researchers can directly ask questions in natural language (text or speech) and in multiple languages. Only questions related to the enriched data and metadata from the bill hearings recorded in the videos would produce answers and citations to the original records. Any other type of questions will not be addressed or responded to by this GPT avoiding hallucinations and answers not based on the description and metadata of the actual records in the archives. One caveat of the model at the time this case study was written is that it does not support the Hawaiian Indigenous language Olelo Hawai'i, which is one of the official languages of the State of Hawai'i.

In February 2024, a demonstration of these services was provided by State Archivist Adam Jansen and his Microsoft partners during an InterPARES Trust AI symposium (Jansen & Cruz, 2024). At that moment over 4,000 videos were available to go through the process, but only about a dozen or less had been processed for the test bed and experimentation. Jansen et.al. reported that the result of their experiment showed promise to enhance access and discoverability to these video-records as originally intended. They also reported that some evaluation and refinement was necessary for the models to make it work for the State archives. One example provided by Jansen et.al. was the incorrect spelling of names of individuals in the output of the speech-to-text transcriptions due to phonetic spelling of the names by the Model that had not yet been sufficiently trained in local names/terms/phraseology. Manual inspection of the produced transcripts and correction of these misspelled names was necessary to match the correct names used by name authorities in the archival metadata. Another initial goal of the project that it was reported by the team as achieved was their initial goal code of "low-code/no-code AI services," meaning that archival professionals could use these services without the need to code or minimally having to code. Finally, the team reported that their approach ended up being "low-cost" for the archives, estimating USD$7.5/hour of video processing once the 'low code' workflow, AI Model training, and user interface had been developed (Jansen and Cruz, 2024).

Potential Discussion Questions:

1. Explain how AI can be used to enhance access to video-records. What are specific AI technologies that can be used for this purpose?
2. What are concerns related to using GPT bots to support archival reference and access to fond/records? Would you consider this project to properly address such concerns?
3. What are the potential ethical concerns related to using facial recognition technology in video records, and how should these concerns be addressed?
4. What are some of the advantages and disadvantages of this government-industry partnership to use AI in state archives?
5. What aspects of AI-powered video indexing as exemplified in this case, require close human supervision and revision?
6. Discuss the implications of using an AI technology, in this particular case, that does not support the native language Olelo Hawai'i, one of the official languages of the State of Hawai'i. How can this issue be addressed?

References

Hawaiʻi State Archives, Department of Accounting and General Services. (2024). Digital Archives of Hawaiʻi. *Welcome to the Digital Archives of Hawaiʻi*. https://digitalarchives.hawaii.gov/ Accessed on: July 23rd, 2024.

Hawai'i State Senate. (2024). YouTube Channel: Hawai'i State Senate. https://www.youtube.com/@HawaiiSenate Accessed on: July 23rd, 2024.

InterPARES. (2022). Research Studies. *Increasing Access to Photos, Videos and Social Media records through AI-generated Descriptive Metadata*. https://interparestrustai.org/trust/about_research/studies Accessed on: July 23rd, 2024.

Jansen, A. and Cruz, M. (2024). iTrust AI Public Symposium. *Using AI To Interrogate Archives - Enhancing Access to Video Using Lox Code/No Code AI Services*. Honolulu, Hawaiʻi. February 23, 2024. https://www.youtube.com/watch?v=hJjifghI6AQ&t=7074s&ab_channel=Hawai%CA%BBiStateArchives Accessed on: August 2, 2024.

Microsoft. (2024a). *Azure AI Video Indexer*. https://azure.microsoft.com/en-us/products/ai-video-indexer Accessed on August 2, 2024.

Microsoft. (2024b). *Azure OpenAI Service*. https://azure.microsoft.com/en-ca/products/ai-services/openai-service Accessed on August 2, 2024.