Planning for an AI experiment

Fill out this worksheet to articulate success criteria, measures, risks, and benefits for an AI Use Case.

Name: LC Labs Junior Fellows

Directorate/Division: Digital Strategy/LC Labs

Date: June 25, 2024

AI Use Case – Phase II Risk Analysis

1. How important are the following considerations for measuring success?	NO importance	LOW importance	MEDIUM importance	HIGH importance	**2.** How will you monitor or measure the success for considerations of MEDIUM and HIGH importance?
a. Progress on organizational goals (short or long term)				✓	As part of the Computing Cultural Heritage in the Cloud (CCHC) project, the LC Labs' Junior Fellows are expected to create a data visualization or otherwise inter-active project that utilizes the Library's digital collections, experiments with emerging technologies, and engages marginalized communities and users otherwise isolated from the Library. The Junior Fellows will be experimenting with using machine learning to manipulate large amounts of data. The end goal is that by being able to manipulate and use large amounts of data, we can create a tool that invites members of the community to not only engage with the Library's collections but to create using these collections. Currently, we plan to do usability testing with groups at the Library that focus on education outreach and programming. Plans are also in place to build in the ability to give feedback in the interface and perhaps track site usage/visitation.

b. Accuracy of data output			✓		Due to the limited timeframe of this project (2 months), we are not able to train our own dataset. As such, we rely on the Microsoft Common Objects in Context (MS COCO) dataset in order to detect object and subjects in images. However, the dataset is limited and sometimes mislabeling occurs. Similar to how we plan to track how the use of machine learning impacts the Labs' goals, we will implement a feedback form into the UI so that users can "mark" mislabeled objects. As a pre-emptive measure, we ran multiple tests to ensure that people are at the very least consistently labeled as people. We will further ensure this by creating a manifest file that records each subject/object extracted from the data and the labels the models assigned to them, for easier review. For other objects to be extracted, we plan to: • Have an informational page that notes that mislabeling can occur due to reliance on a pre-curated dataset. • Create a python dictionary that relates like-objects to a larger synonym. Ex: ○ "Objects": {"Animal": ["dog", "cat", "squirrel"]}, {"Vehicle": ["car", "truck"]}}
c. End user feedback or impact				✓	As stated in boxes "1.a" and "1.b," we plan to incorporate feedback forms into the UI as well as have an initial testing with the Library's educational outreach division. There will be a feedback form specific to the functionality of the site (i.e., accessibility concerns) and a separate feedback form for the machine learning aspects of the site (i.e., labeling).

d. Time saved or manual burden for staff lifted				✓	We will be working with digital objects from a wide range of collections. Object detection and segmentation at such a scale would not make this tool as possible in such a short amount of time. Machine learning also makes the process iterable and scalable.
e. Lifecycle costs	✓				The MS COCO dataset is free of charge and the host site for the tool (GitHub Pages) is also of no cost. Any cost incurred will be a team member's time used to possibly update the dataset or code.
f. Unbiased results or balanced impacts across data types, classes, or demographics				✓	[Please refer to boxes "1.a" through "1.c" on how we plan to work with/mitigate bias.] As we are using a pre-curated and established machine learning data set, we cannot ensure that the results (subject/ object detection and labeling) will be unbiased. We have performed multiple tests with different Library collections of varying quantity, image quality, and subjects of different skin tones. All subjects (of varying skin tones and image quality) were captured effectively and consistently. We will also track these labels in a manifest file that can be easily and quickly referenced. Human subjects were of a higher priority with catching mislabeling as harm to animals and inanimate objects are of limited consequence. See box "1.a" to understand how we plan to work through this. Please see the "evaluations" sheet of our jfp-collections spreadsheet for reports that contain more information.

g. Feasibility for (next phase of) implementation		✓			The feasibility of the next phase of implementation is currently of low priority and it will be up to the members of the LC Labs' team to determine next steps. Currently, the work we have done should be generally applicable to more scaled projects and the data collected and code written should ensure that UI implementation is forthcoming.
h. Developing or verifying performance baselines		✓			[Please see boxes "1.a" through "1.b" and "1.f" for more information.] We opted for a quantity over quality approach due to timeframe and the nature of the tool. However, we have ensured that at the very least, human subjects (of varying skin tones) are captured and labeled correctly as "Person."
i. Adherence to quality standards over time	✓				While there could be changes to the accessibility and functionality of the model, as well as the collections available through API access, this is of low concern for the reasons listed in box "1.h."

3. Are there other considerations for measuring success not listed above? If so, what are they?

Given the limited timeline the Junior Fellows have, implementation of a UI may not be possible. In this event, the Junior Fellows will provide and updated, high-fidelity mock-up of the UI. Any code written or UI developed will then remain as a proof of concept. This limits some of the risks/concerns previously discussed.

Risk & Benefit Analysis: The Agency or Department

4. What are the related goals or priorities this new or enhanced capability could support?

As part of the Computing Cultural Heritage in the Cloud (CCHC) project, the LC Labs' Junior Fellows are expected to create a data visualization or otherwise inter-active project that utilizes the Library's digital collections, experiments with emerging technologies, and engages marginalized communities and users otherwise isolated from the Library. The Junior Fellows will be experimenting with using machine learning to manipulate large amounts of data. The end goal is that by being able to manipulate and use large amounts of data, we can create a tool that invites members of the community to not only engage with the Library's collections but to create using these collections.

In our case, we are creating a collage tool using images from the Library's collections. The success of this collage tool requires that the user can extract or "cut" subjects and objects from the images. Machine learning ensures that we can do this at scale and iterate this process should the Labs team want to expand upon the tool.

5. Describe potential **benefits** to the Agency.	Rate potential benefits and risks			
[Please refer to box 4.] The scalability that machine learning allows for will allow us to create a collage tool that engages a younger audience (which does not often engage with Library collections remotely).	1 -LOW benefit	2	3	4 – HIGH Benefit ✓
6. Describe potential **risks** to the Agency and how you will mitigate the risks. A potential (future) risk to the agency would most likely surround data use and harvesting for more model training. However, this is not currently of concern as we are not training the model but using a pre-curated data set). Another potential (future) risk may be regarding a change in the licensing of the model. We do not have any plans in place should this happen, but do not expect it to be an issue.	1 - HIGH risk	2	3	4 – LOW Risk ✓

Risk & Benefit Analysis: End Users

7. How will end users (customers or citizens) interact with the new or enhanced capability?

The intended audience of the tool are teachers, who may use this tool for educational purposes, and students (late elementary to early middle school) who may use this tool for educational play in a less-guided classroom setting (after-school care, for example) or in a guided classroom setting (history lesson with a teacher).

Users will not be interacting with machine learning directly, so direct risks are limited. The assets that users will be interacting with will already be manipulated by the machine learning model and ready for use.

8. Describe potential **benefits** to end users.	Rate potential benefits and risks			
	1 -LOW benefit	2	3	4 – HIGH Benefit
Our hope is that this tool will allow users to engage in shared history and become familiar with the resources available to them via the Library.				✓

9. Describe potential **risks** to end users and how you will mitigate the risks.	1 - HIGH risk	2	3	4 – LOW Risk
Potential risks to end users may be inaccurate representations of historical figures with whom they share identities or the misrepresentation of marginalized histories. These risks would more likely than not be due to subjects and objects being displayed out of context. Further risks could involve bad actors placing marginalized subjects in compromising or unfavorable contexts. In order to mitigate this, we will be placing the image in its full context into an image browser, with an outline of what will be extracted from the image. Once the user clicks on the image, and more options will be displayed in the tool bar to get information on the image via the Library's item site. Furthermore, we have ensured that subjects and objects will be extracted "whole" from the image. There will be no functionality with which a user can extract attributes of subjects, such as eyes, mouths, etc. This is to ensure that bad actors do not use a "dissected" image of marginalized groups for nefarious acts.		✓		

Risk & Benefit Analysis: Staff

10. How will staff interact with the new or enhanced capability?

There are no foreseeable risks to staff. However, staff that work in educational outreach may see this tool as beneficial when showing citizens how the Library's collections may be of service to them.

11. Describe potential **benefits** to staff.	Rate potential benefits and risks

Potential benefits will most likely be regarding outreach efforts. Such a tool can demonstrate to teachers and community members how beneficial the Library's collections are. It will also help foster awareness of and interest in the Library in a younger demographic.	1 -LOW benefit	2	3	4 – HIGH Benefit ✓
12. Describe potential **risks** to staff and how you will mitigate the risks. There are no foreseeable risks to staff.	1 - HIGH risk	2	3	4 – LOW Risk ✓

Risk & Benefit Analysis: People depicted in the data				
10. How will people represented in the data be depicted by the new or enhanced capability? While users will be able to "cut" subjects from images to create a collage, subjects will first be displayed in context and information about the image will be readily accessible to the user.				
11. Describe potential **benefits** to people depicted in data. Most of the subjects in the collections we are working with for this tool are not living. Despite this, benefits to their memory and remaining community will be the sharing of their history.	Rate potential benefits and risks			
	1 -LOW benefit	2	3	4 – HIGH Benefit ✓

12. Describe potential **risks** to people depicted in the data and how you will mitigate the risks. [See Box 9 for more information.] Concerns for the use of machine learning in this project mostly centers how the model would interact with older images, darker skin tones, and how it may label people due to biases. Based on tests done on various collections using the model, this risk is low and mostly affects animals and inanimate objects. In order to limit the harm of using the tool to "distort" the images of marginalized communities, we ensured that only the whole image of the individual can be extracted. There are concerns about removing subjects from their initially agreed upon context, but we plan to build this context into the tool by showing the user the whole image and will an option in the tool bar to view more information about the image.	1 - HIGH risk	2 ✓	3	4 – LOW Risk