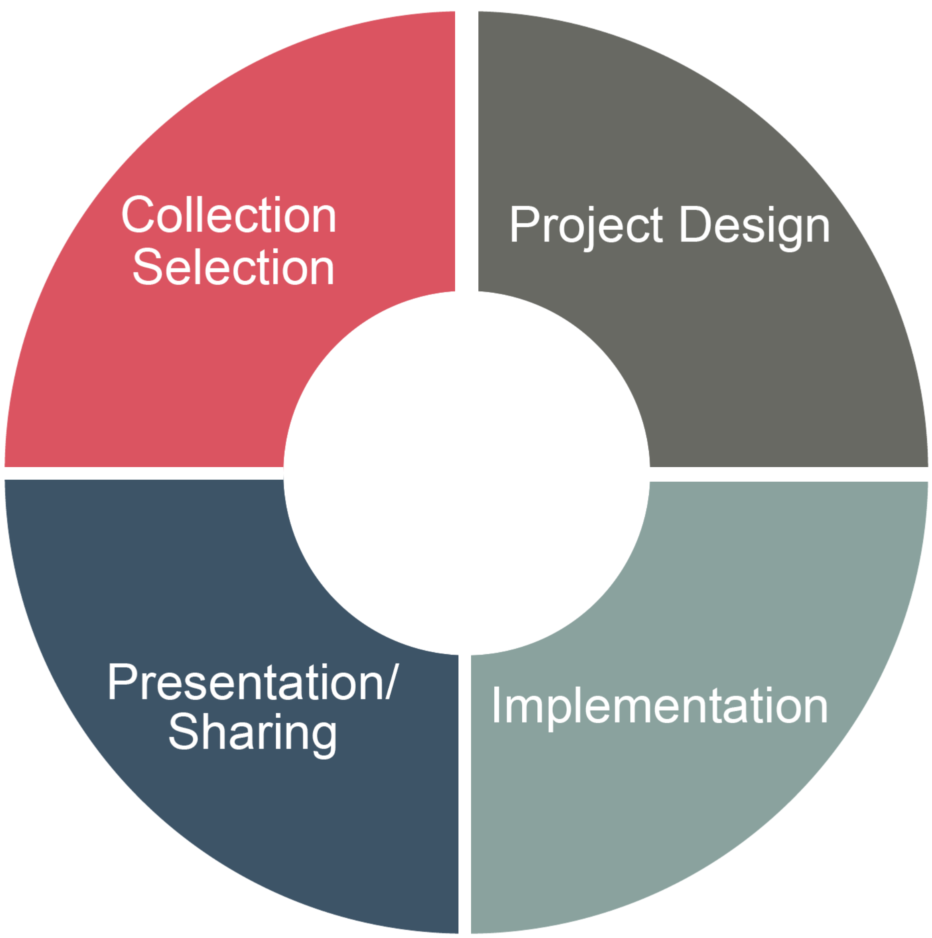
A Humans-in-the-Loop Framework

The unique challenges of machine learning as a technical solution require libraries to consider resources and staffing throughout the project lifecycle, beyond just the design and initial implementation. As machines learn from human training, outputs may improve but may also change or evolve in unexpected ways to negatively impact engagement, ethics, or usefulness. Slight variations in collection content or digital quality may be imperceptible to humans, but present new challenges to trained algorithms, so constant iteration is necessary in monitoring accuracy and bias of results and adjusting models or parameters or even machine learning approaches to course-correct.

In project or software development, user-centered design offers tools and principles to engage humans in the design and success of a project, also with a focus on human feedback and iteration, or humans in the loop. We offer a framework for development of human-in-the-loop projects in cultural heritage using the lens of iterative, user-centered design to help guide not just algorithm design but all phases of a project lifecycle, from collection selection to data integration and discovery.

The following framework outlines four stages of human-in-the-loop project development -- **collection selection, project design, implementation, and presentation/sharing** -- addressing the challenges and goals of each stage in relation to engagement, ethics, and usefulness, the humans to be involved in each phase, and tools for incorporating feedback into the identification and mitigation of risk or potential harm to humans.



Collection Selection

Goals

Engaging

- Selecting collection content that is interesting to crowdsourcing users
- Crowdsourcing volunteers feel a personal connection, a sense of ownership to the selected content

Ethical

- Exposing data from the selected collection respects the privacy of collection subjects or creators
- Potential risks to users and collection creators/subjects can be identified and mitigated

Useful

- Replicable collection selection processes for HITL projects are modeled
- Data generated from the selected collection are useful to library users
- Data generated from selected collection improves discoverability of that content
- Collection content is free of permissions restrictions to enable broad use.

Objectives

- Select collection(s) for project design
- Identify data outputs

Challenges

How do we…

- Engage a broad diversity of perspectives in selection without bogging down the process?
- Select a collection large yet homogeneous enough to benefit from a ML approach?
- Select a collection that can attract and sustain the interest of crowdsourcing users?
- Find ML methods advanced enough to generate the desired data from the collection?

Humans

We involve…

- **Community managers**, who understand what tasks are engaging to volunteers;
- **Reference specialists**, who understand what data researchers are searching for and why;
- **Collection curators**, who understand the content within collections and can speak to potential risk;
- **Digital collection specialists**, who understand how to work with digitized collection objects and metadata;
- **Machine learning experts**, who understand what data generation tasks are possible to do with algorithms;
- **Program specialists**, who understand how to connect humans across organizational divisions in support of HITL projects.

Feedback Mechanisms

We learn from…

- **Cross-functional brainstorming workshops** to bring diverse collaboration to idea generation;
- **Risk/benefit analysis** to help to identify risks and mitigation strategies early in the project;
- **User stories** to show how and what collection content will be useful as data.

Project Design

Objectives

- Model collection content as structured data
- Design ML pipeline
- Design crowdsourcing tasks
- Define QC measures
- Define volunteer outreach plan

Goals

Engaging

- Defined crowdsourcing tasks are enjoyable to volunteers
- Volunteers get deep exposure to collection content
- Volunteers understand the value of their contributions to the greater good

Ethical

- Volunteers understand how their contributions will be used especially in relation to machine learning processes
- Risks to volunteers from potentially offensive content are identified and mitigated
- Potential unintended consequences of machine learning processes are identified and documented

Useful

- Data is modeled in a way that it can be shaped for various kinds of use

Challenges

How do we…

- Support a wide range of structural complexities in collection content as they surface?
- Generate enough training data to test and select initial ML processes?
- Design crowdsourcing tasks that support the ML pipeline and are also interesting to volunteers?
- Identify a "good enough" threshold for ML accuracy?
- Ensure crowdsourcing data is accurate enough to use as ground truth?

Humans

We involve…

- **Collection curators,** who understand the collection content to be modeled;
- **Reference specialists,** who understand how researchers could use collection content as data;
- **Metadata and digital collection specialists,** who understand how to model content as data;
- **Machine learning experts,** who understand ML processes for extracting data;
- **Community managers,** who understand what tasks are engaging to volunteers;
- **Library staff,** who can create training data for initial ML explorations;
- **Volunteers,** who can offer early feedback on possible crowdsourcing tasks.

Feedback Mechanisms

We learn from…

- **User interviews** with volunteers to help identify engaging types of tasks;
- **Ground truth accuracy testing** to help measure the fit of an ML process for a data generation task.

Implementation

Objectives

- Build crowdsourcing application
- Build ML pipeline
- Track data flows and outputs
- Test and refine ML processes

Goals

Engaging

- Volunteers have the opportunity to learn about the collection and related materials
- Volunteers are able to understand and track the progress of their contributions
- Volunteers are able to choose tasks and switch between different tasks

Ethical

- Data provenance and accuracy of machine learning generated data is tracked
- Use of machine learning is clear and understandable to volunteers

Useful

- Accuracy of machine learning-generated data can be improved through ground truth testing

Challenges

How do we…

- Achieve consistent results from varying collection content?
- Retrofit existing crowdsourcing platforms for new projects?
- Refine ML processes to improve accuracy as more data is generated from crowdsourcing tasks?
- Communicate ML processes and interactions to volunteers in ways that are clear and digestible?

Humans

We involve…

- **Machine learning experts,** who understand how to implement and refine ML processes;
- **Software developers,** who understand how to build crowdsourcing platforms and architect data flows;
- **Digital collection specialists,** who understand how to integrate digital collection content into crowdsourcing and ML pipelines;
- **UX designers,** who understand how to present tasks and information to users in a clear and accessible way;
- **Community managers,** who understand the wide range of volunteer needs for a crowdsourcing platform;
- **Volunteers,** who contribute data to train and test ML processes and can offer feedback on crowdsourcing tasks.

Feedback Mechanisms

We learn from…

- **User testing** with volunteers to help improve the user experience of the crowdsourcing platform
- **Ground truth accuracy testing** to help improve ML processes throughout the course of the project
- **Workflow database** to track provenance and accuracy of all tasks, processes, and data for input on platform/pipeline improvement

Presentation/ Sharing

Goals

Engaging

- Users are presented with a variety of pathways to explore the content
- Collection data is presented through interesting and pleasing interface designs and experiences

Ethical

- Provenance of data is communicated to library users
- Library users understand potential biases and incompleteness of data

Useful

- Users can discover data or collection content that meets their research needs
- Users can download and use data in tools they are familiar with

Objectives

- Provide access to structured data
- Integrate data into discovery systems

Challenges

How do we…

- Convey provenance of data without overwhelming users?
- Communicate the incomplete and dynamic nature of data generated through large-scale ML processes to users?
- Integrate large volumes of data into discovery systems without diluting search results?

Humans

We involve…

- **Collection curators,** who understand collection contents and how they should be navigated;
- **Reference librarians,** who understand researcher needs;
- **Digital collection specialists,** who understand how to connect and display digital objects and metadata;
- **UX designers,** who understand how to communicate and display data in a clear and accessible way;
- **Library users,** who can offer feedback on website usability.

Feedback Mechanisms

We learn from…

- **User persona development** to help in understanding user needs and brainstorming interface functionality;
- **Wireframing** to help imagine the potential of an interface to test with users;
- **User testing** of wireframes with library users to help understand how interface design and the data driving it meets stated and implicit research goals.