NATIONAL LEADERSHIP GRANTS FOR LIBRARIES

Sample Application LG-259749-OLS-26
Project Category: Planning

University of Florida

Amount awarded by IMLS:	$163,016
Amount of cost share:	$0

The University of Florida will pilot a human-guided, Large Language Model-assisted bulk-work environment. The project will focus on building features to assign subject terms for text-based content. The project will add new access points to approximately 20,000 newspaper issues published in Pensacola and St. Augustine. The project will deliver a functional bulk work environment powered by NavigatorAI and HiPerGator and a website documenting the major technical decisions, as well as tested prompts and corresponding results. The project seeks to address critical needs for improving the accessibility of our digital collections at scale. This includes features such as generating transcripts for audio and video, creating alt-text for images, and providing content summaries for all file types.

Attached are the following components excerpted from the original application.

- Narrative
- Schedule of completion

When preparing an application for the next deadline, be sure to follow the instructions in the most recent Notice of Funding Opportunity for the grant program to which you are applying.

Accelerating History:

Piloting Multiple Large Language Models to Generate Subject Keywords forHistoric Florida Newspapers

Introduction

The University of Florida (UF) seeks **$163,016** over 24 months (September 1, 2026, to August 31, 2028) to pilot a human-guided, Large Language Model (LLM)-assisted bulk-work environment. Assigning subject terms to digitized collections has traditionally been a human-aided process, which is slow and can yield minimal or broad subject terms. Accurate subject metadata is vital for user discovery and retrieval. The proposed workflow framework leverages NavigatorAI, a UF service, which enables users to command multiple LLMs simultaneously for the same task, while using another UF service, HiPerGator, the fastest supercomputer owned and operated by a US university, to generate high-quality descriptive information for digital collection items at high velocity (seconds per item). In this pilot phase, we focus on building features to assign subject terms for text-based content. In celebration of the 250th anniversary of American Independence, we plan to add new access points to approximately 20,000 newspaper issues published in Pensacola and St. Augustine, the two Florida cities with close historical ties to the American Independence era. Our approach significantly scales up the process of assigning subject terms, addressing the inherent deficiencies of conventional, item-by-item practices, which are time-consuming and inconsistent in quality. Using a bulk-processing method, we can enhance the quality and consistency of subject terms for thousands of digital items within hours. The deliverables at the end of this pilot project include the following:

- A functional bulk work environment powered by NavigatorAI and HiPerGator
- Tens of thousands of new access points to about 20,000 newspaper issues hosted in the UF Digital Collections (UFDC)
- A website documenting the major technical decisions, as well as tested prompts and corresponding results
- Conference presentations
- Peer-reviewed Open-Access Journal articles

This project aligns with the agency-level goal of the Institute of Museum and Library Services (IMLS), specifically, Goal 3 Advance Collections Stewardship and Access, Objective 3.2, Promote access to museum and library collections. As well, it supports the goals of the National Leadership Grants for Libraries (NLG-L) Program, meeting Objective 1.3, to provide broad access to and preservation of information and collections through libraries and archives.

Project Justification

The Need and the Opportunity

A Large Language Model (LLM) is a type of artificial intelligence (AI) technology designed to understand, generate, and work with human language. It is trained on very large amounts of text data most gathered from the Web to learn patterns in language—such as grammar, meaning, context, and relationships between words. Users command a LLM via prompts. Prompts are like the written instructions given to an assistant for conducting specific tasks. Effective prompt construction is crucial for maximizing the alignment between a LLM's output and user expectations.

While their performance is not without limits, LLMs provide opportunities to accelerate various processes that are currently manual in nature. Our workflow ensures high-quality outcomes by leveraging multiple LLMs and integrating a human-in-the-loop review. This multi-model approach allows us to identify specific performance limitations by comparing LLM outputs. For this pilot, LLMs solve the primary challenges of our manual process: high time costs and inconsistent quality. By automating this workflow, we can now generate high-quality subject terms for thousands of digital items in a fraction of the time.

"Subjects" or “subject terms” refer to words and phrases that encapsulate the key concepts or main themes of the materials and are contained within an item's metadata. Subjects play a crucial role in organizing digital resources, making it easier, or in some cases possible, for users to discover the target materials among millions of resources. In digital collections, as essential components of the browsing structure, subjects guide users to explore and discover the content. Many digital library systems rely on search indexes that prioritize certain metadata fields, such as title, author, and subjects, when they return results to users. By enriching an item's subject terms, users can find and retrieve more relevant results. Additionally, subjects are often displayed as hyperlinked terms, prominently positioned for users to click to explore the content that shares the same subjects. Though important, subjects in digital collections are often either missing or disorganized, or insufficient to support content discovery.
The current inefficiencies are inherent to the nature of subject indexing. This two-part process requires a practitioner to first analyze content to identify core themes and then map those themes to a standardized authority file. Because this work is intellectually intensive, even a trained professional requires several minutes per item, making it difficult to scale. The results, largely determined by the experience and knowledge of the professionals, could be subjective (Alves Silva Portella 2024), then inconsistent in quality across collections over time.

LLMs can address this rooted deficiency of the current approach. Using LLMs to generate terms or phrases as labels for documents, a process known as annotation, is a well-established practice in Computer Science and interdisciplinary fields such as Computational Social Science, where extensive datasets require categorization before analysis. LLMs are increasingly viewed as a robust alternative to human annotators due to their ability to accelerate the process and reduce costs (Wang et al. 2021). While out-of-the-box LLMs like ChatGPT can perform this task effectively, research emphasizes the critical need for human validation to ensure accuracy (Pangakis, Wolken and Fasching 2023). Along this line, we position the human review step at the core of the workflow.

Offering the solution to enhance the quality of assigned subjects at a scale, this project demonstrates the work to improve the discoverability within a vast digital collection, specifically for this project, the Florida Newspaper Digital Library (https://newspapers.uflib.ufl.edu/). Discoverability refers to how easily a digital object can be located within a large collection on a platform, primarily by utilizing its descriptive information, such as subjects, as access points. Discussions about methods and workflows to address the causes of poor discoverability have been ongoing since the earliest days of these platforms. While vocabulary control—the practice of using a standardized, limited set of terms for indexing and searching documents in a specific system (Haider 2017)—is recognized as crucial for improving discoverability on digital platforms, its implementation remains challenging. The substantial volume of data, coupled with the considerable human labor and time required for the process, poses significant hurdles to initiating and sustaining effective vocabulary control practices. Take the Florida Newspaper Digital Library, a sub-collection of UFDC for an example. The Florida Newspaper Digital Library contains over 1,100 newspaper titles and more than 450,000 issues, spanning history from the 1820s to the present. It has been used heavily, viewed over 21 million since 2006. However, the current limitations on access—restricted to geographic area, title, or publication date range—prevent users from searching, browsing and grouping by topical subject. This limited discoverability creates a barrier, obscuring the rich historical content held within the collection. It is understandable that a large digital collection lacks subject headings. Assigning subjects is both time-consuming and intellectually demanding, and in this specific case, the sheer

volume—over two century's worth of content—has made it impossible for trained professionals to manually review every single issue.

We need better methods. Our tests show that LLMs can assign acceptable topical subjects for a single item in seconds. Consequently, establishing a bulk-work environment that yields consistent and high-quality results within a reasonable timeframe would enable us to process vast quantities of materials. To mitigate the inherent performance limitations and potential unreliability of a single LLM, we employ a multi-LLM approach. NavigatorAI provides all major LLMs that include claude, gpt, gemini, llama and more (*Overview)*. Our principle is straightforward: drawing on results from multiple LLMs, much like seeking advice from diverse experts, enhances reliability through comparison. Crucially, all results generated by the LLMs are subject to final review by human experts.

Preliminary tests have demonstrated the approach's ability to provide human experts with quick (in minutes) and valuable insight. We conducted a test using the abstracts of 20 agriculture-related publications, asking three LLMs (gpt-5, claude-3.7-sonnet, and gemini-2.0-flash) to select three subjects focusing on specific aspects of agriculture from a pre-select list of terms (see the results analysis below). We replicated this test using the abstracts of 8 publications on Women's studies with a pre-select list of terms about specific aspects of Women's studies too. The complete results of both tests are available in the supporting documents. Both tests show that LLM insights offer human reviewers a valuable initial

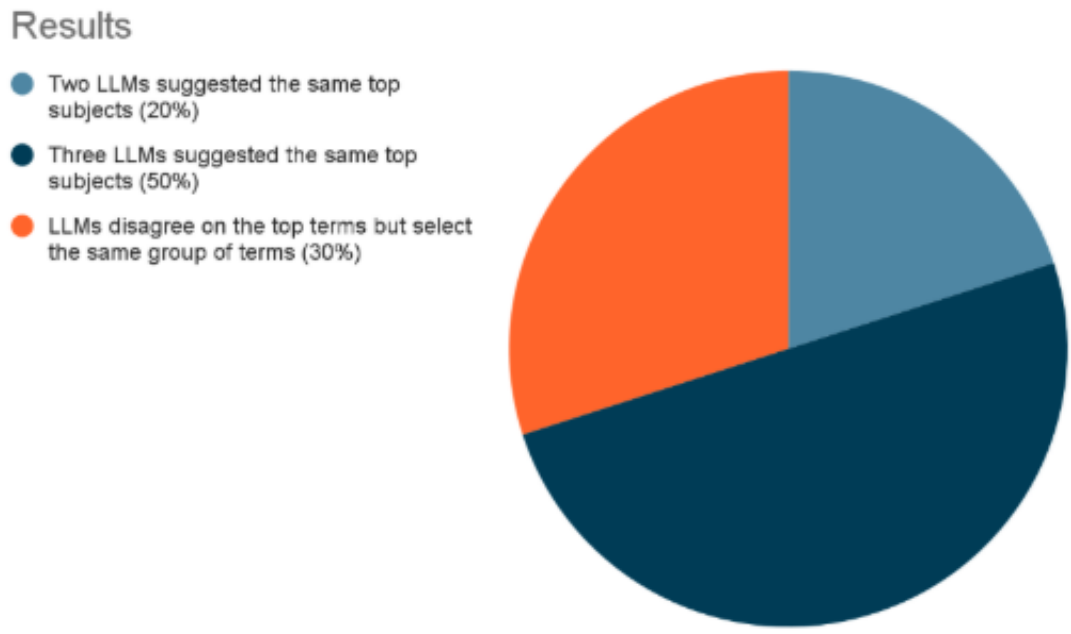


Chart 1: LLM Test Results with Agriculture Publications

step, allowing them to quickly identify the main theme and subsequently determine the most appropriate subject terms.

These new, reliably assigned terms would significantly improve content discoverability, connecting related content, and ultimately help build researchers' and learners' trust in digital collections as a major

resource. The diverse nature of the topics and the large volume of data we will successfully process demonstrate an approach that could be broadly applied in various contexts beyond research and learning.

In this pilot project, we will process about 20,000 issues of newspapers published in Pensacola and St. Augustine from the 1890s to the present day. Both of these cities are strongly linked to the American Independence Era. In 1781, the Siege of Pensacola saw Spanish forces allied with American patriots drive Britain out of West Florida, a pivotal contribution to the American Revolution. The Pensacola area also features in the later West Florida Rebellion (1810), a pro-U.S. independence movement that helped set the stage for Florida's eventual statehood. St Augustine is the oldest European-founded city in the U.S. It changed hands among Spain, Britain, and the United States around the Revolutionary period, illustrating Florida's role in the broader Atlantic conflict over sovereignty. The Florida Digital Newspaper Library hosts over 23,000 newspaper issues published in the two cities, and these issues suffer the same discoverability problems. Users need more information than just the publication area; they need a way to understand the content of each issue without manually paging through it. Our project aims to resolve this problem by adding subjects as new access points. We've selected about 20,000 issues with good quality digital files for this project. Among them are popular titles like "The St. Augustine Evening Record" and "Pensacola journal". Upon completion, users can know the thematic focus of each issue via assigned subjects.

The Innovation

To harness the power of LLMs, library service vendors like Ex Libris and JSTOR, have integrated LLMs into their products. Ex Libris has included the AI Metadata Assistant in the Metadata Editor. Using LLMs, this product provides a draft copy of a descriptive record for a single item, so library staff can review and finalize the record (The AI Assistant). JSTOR has introduced a new AI component, JSTOR Seeklight, into its Tier 3, the most comprehensive packaged services (JSTOR digital stewardship services). This service, also powered by LLMs, produces draft copies of descriptive records for multiple items with additional functions to group and then update records. As well, this service generates transcripts for Video and Audio files (Jstor seeklight). These two products target different groups of library staff. The former aims at serving catalogers who describe books one after another, and the latter, the archival staff who process similar items as a batch.

Like the other products, our project utilizes the power of LLMs for efficiency as well as maintains that human expertise for quality assurance. Unlike the others, which produce complete draft records, our focus is on a single, intellectually demanding field: subjects. This highlights the complexity of subject assignment. A complete record of library resources must encompass the title, creator information, date, genre, and subject. The information is largely factual, with the exception of the subject. For LLMs, assigning subjects based on the results of human comprehension necessitates calculations that are distinct from, and potentially more complex than, gathering factual information. By narrowing our focus to subject classification, we can isolate this specific capability and build a more robust understanding of LLM performance. This project offers a framework for using multiple LLMs in bulk to generate descriptive information about digital items, rather than targeting a specific staff demographic. Our project introduces a novel approach: we prompt and compare results from multiple LLMs, a design currently absent in known products. Furthermore, we commit to a level of technical transparency that commercial library service vendors rarely offer. We will openly share the specific LLMs, tested prompts, and all test results on the project website. This ensures that other researchers and practitioners have a solid foundation to build upon and advance this work.

This proposal is not the director group's first effort to add machine-assisted steps into the process of assigning subjects. From 2016 to 2021, the director and the co-directors explored Machine Aided

Indexing (MAI), using a vendor-supplied tool to partially automate the subject indexing process (Digby and Dinsmore, 2018). During this project, over 76,000 English text files received 427,849 topical and geographic subjects. Many of these subjects, previously unavailable, provided new access points to the content. The MAI process has demonstrated the potential for automation, showing that thousands of items could have subjects assigned within an hour. However, the findings also underscored that the quality of machine-generated topical labels often requires human review to ensure accuracy and relevance (Ma et al., 2023). Building on learning from MAI, this project will develop an efficient review process that integrates human and machine evaluation, supplying the review outcomes to inform and direct testing efforts.

The Impact

As a novel LLM use case, this project benefits practitioners and researchers in settings that struggle to organize extensive data that include but are not limited to libraries, museums, archives, data centers, government agencies, healthcare systems, legal firms, and corporations. It demonstrates the framework and details of working with multiple LLMs, showing how human knowledge can be passed into automation processes and how the work can be scaled up without sacrificing the quality of the results. As a project to improve the discoverability within a vast digital collection in an academic library, the project benefits digital collection users including students and researchers from all backgrounds. As this project adds subjects to newspapers, the results of the work will be especially helpful to studies and projects related to American History and Culture. As an institution-built tool, its success is projected to yield considerable long-term financial savings by eliminating the need to allocate funds for comparable commercial AI services and products.

Project Work Plan

The Project

During the funded period, we will build, test, and deploy the proposed bulk work environment, and disseminate our learning via a website, present at conferences, and via peer-reviewed journal articles.

Bulk Environment Development

The bulk process assisted by LLMs will be implemented through a four-step, sequential bulk environment. These four steps can function as a framework for organizing all types of tasks that can use LLMs to scale up the work.

As detailed in Diagram 1, the four steps include

1. Group Media Files: pull out media files that include, but are not limited to JP2000, PDF, Text, and XML files from the digital collection backend folder structure;
2. Pre-Process Media Files: set up Optical Character Recognition (OCR) function on HiperGator, send OCR results to the digital collection backend, and also pass them for LLM processing;
3. LLM Tasks: bring in OCR results from Step 2, send commands/prompts via UI to multiple LLMs, and deliver results in CSV files;
4. Deliver results: write the approved results to the digital collection files and database, so the newly added data can be included in the indexing system and then used in the digital collections.

Human validation will occur between steps 3 and 4, approving satisfactory results that can be used in the digital collections and singling out items with unsatisfactory results for further analysis and more rounds of work.

For all steps, we include the User Interface (UI) design and development along with the function development, so non-programmers with expertise knowledge on subject assignment can lead and run all steps. UI refers to the visual elements and interactive components that enable a user to interact with the backend functions without being familiar with programming languages.

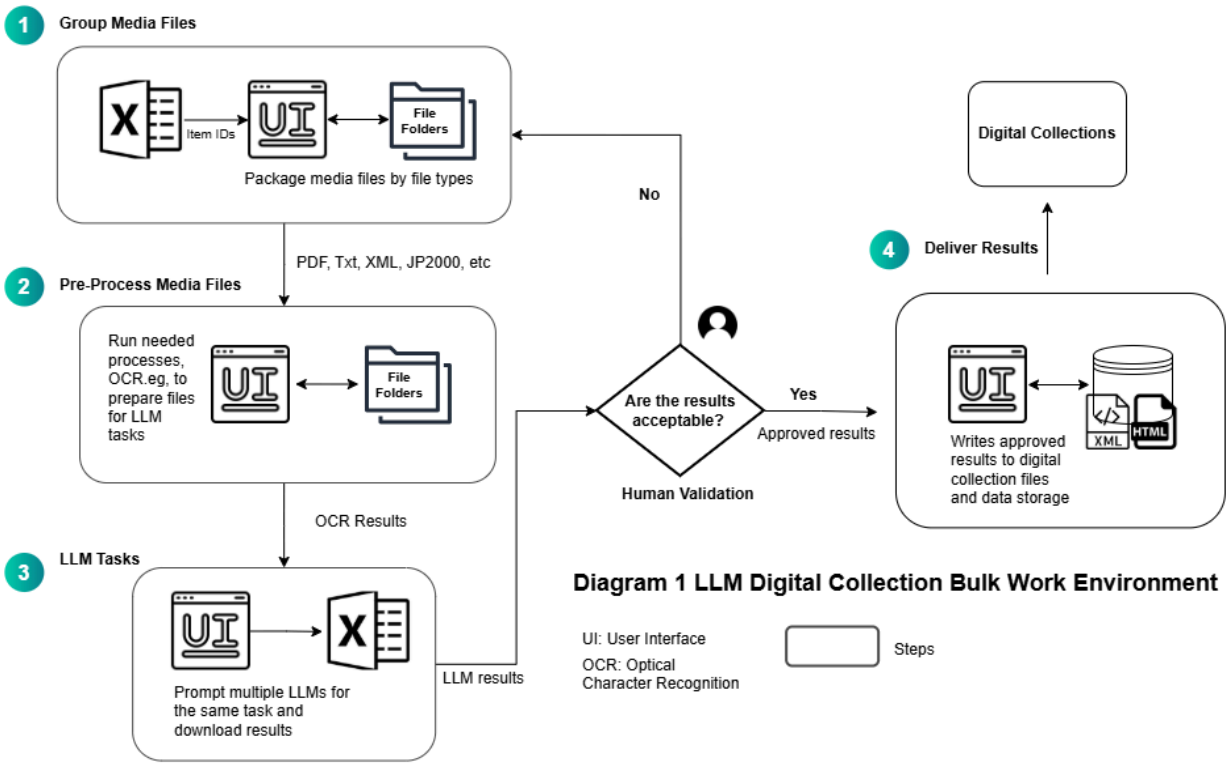


Diagram 1 LLM Digital Collection Bulk Work Environment

We anticipate that steps 2 and 3 outlined above can benefit all institutions, while the setup of step 1 and the final step 4 could vary by institution. Step 2 provides a fast environment to OCR scanned materials. OCRed materials can support in-text search and accessibility functions, e.g., enable screen readers and other assistive technology to read content aloud, aiding users with visual impairments. Step 3 uses multiple LLMs to suggest subjects for hundreds of digital items at once and delivers the results in Excel format. Many digital collection platform tools can accept data from Excel. Other institutions can easily adapt the LLMs to meet the requirements of their own platforms.

To serve institutions across the nation, we will explore the technical possibilities, the financial plan and the policy. To help us understand the challenges on this end, we will form an advisory group composed of researchers and practitioners from other institutions to provide us guidance and feedback as we march along. As well, we will disseminate our learning via a website and conference presentations.

Learning Dissemination

We will establish a project website while proposing and attending conferences where the use of LLMs in digital collections would be discussed.We will also identify open-access journals for publishing our work.

The project website will be built using a sustainable website tool identified via comparing a few products. Github.com and Omeka S are under consideration. Following platform tool selection, the next steps involve planning the site structure and choosing a template. We will establish content standards, specifically focusing on metadata to optimize for search engines. Page construction will then proceed as the project develops.

This website introduces the project, captures project milestones and documents the project's technical details as the project progresses.

The conferences we will participate in are ai4lam, code4Lib and DLF Forum. ai4lam gathers professionals in libraries, archives and museums together to organize, share and elevate their knowledge about the use of artificial intelligence (AI). The Code4Lib conference has a broader scope. It is a community interested in knowing all cases of automation applied in library settings. The focus of DLF Forum lies in all practices and theories related to digital libraries. Applying LLMs to scale up the digital collection tasks is surely included in the focus of the above conferences.

We consider a number of peer-reviewed, open-access journals, including *Educational Innovations and Emerging Technologies*, *International Journal of Digital Content Management*, *Knowledge Management & E-Learning: An International Journal*, *Library Resources & Technical Services*, *Technical Services Quarterly*, and *Catalogue and Index*, among many others. Once the article focus is determined, we will then refine our selection of target journals.

Our Yearly Goals

To successfully carry out the above activities, we have set goals for each year.

The goals of the first year are:

- build up effective communication channels among team members and establish routine communication with consultants and advisory group;
- establish the project website, and the set up of the website update routines: document the project decisions and processes, and then add documentation to the website;
- draft reports, journal articles, conference presentation proposals;
- develop and test major features.

And the goals of the second year are:

- finish and refine all features;
- start and finish adding new subjects/access points to newspaper items;
- finalize and promote the project website;
- publish journal articles and attend conference presentations.

The People

The key stakeholders for this project are diverse, including:

- **Practitioners** who will utilize this bulk environment for their daily tasks.
- **Library Administrators** responsible for library finance and personnel, as they will assess improvements in efficiency to guide resource allocation.
- **Researchers** interested in exploring the latest real-world applications of LLMs.

Practitioners are core members of the project team as well as testers of the built functions. We will report our progress to library administrators via established check-in meetings and all staff meetings. As well, we invite UF researchers to be our consultants so they can provide their expertise as we progress. Experts

external to UF have been invited to join our advisory group. Their guidance and feedback will be invaluable as we prepare to expand the project's impact beyond the university. In short, the project team consists of core members and supporting staff, advised by consultants. Additionally, an advisory group has been formed to provide guidance throughout the work in preparation of the future phases.

The Project Team

The core members are formed by the director, two co-directors and the development sub-group of a Developer Contractor (5 hours/week) and a UF graduate student (10 hours/week).

Xiaoli Ma, project **director**, serves as the Metadata Librarian and the Head of the Metadata Unit at the George A. Smathers' Libraries. She will manage the project and take the lead on all lifecycle phases of the project from initiation, planning, execution, monitoring & control, to closing. Tasks include, but are not limited to, establishing the team communication channels, initiating and maintaining a smooth communication with stakeholders, testing functions, drafting prompt structure, tracking and reporting project progress, providing metadata related specifications etc. Ma's expertise lies in Metadata Creation with LLMs, Metadata Quality and Knowledge Management. Her published case studies illustrate effective methods for improving metadata quality at scale through the application of emerging technology.

Todd Digby, **co-director**, is the Chair of Library Technology Services at the George A. Smathers Libraries. He will head the development sub-group composed of a Developer Contractor and a UF Graduate Assistant and work closely with the programming advisor from the UFIT Research Computing. He will form the development group and keep their work on track. Mr. Digby's research focuses on key areas of library science, particularly how technology is utilized in specific contexts. He will draft journal articles and attend conferences to disseminate the learning.

Laura Perry, **co-director**, is the Head of the Digital Services at the George A. Smathers Libraries. Ms. Perry will head the tester group formed by a selected group of five members from the Digital Services and the Metadata Unit and organize all test tasks. As well, she will provide the technical specifications and workflow requirements to the development sub-group. Laura has been managing the digitization process, and ingest operations for the University of Florida Digital Collections (UFDC) for over 10 years, leading the efforts of adding 1 million pages of scans to UFDC annually. The 20,000 newspaper issues that the project will work on is part of the content that Perry's team made available online.

The director and co-directors have a strong history of successful collaboration, consistently achieving ambitious objectives. From 2016 to 2021, Todd Digby, Xiaoli Ma, and Laura Perry collaborated on the local implementation of MAI, a process to partially automate the bulk assignment of subjects to digital items. That project's objective partially aligns with this proposal, specifically in its goal of scaling up subject assignment by integrating automation. For that project, Todd Digby ensured the vision was communicated effectively and allocated the necessary resources to the implementation team. Xiaoli Ma and Laura Perry worked with a UF developer to localize the vendor solution, focusing on building functions, developing the UI, and establishing local workflows. Digby, Ma and Perry all presented work related to this project in national conferences.

The Developer Contractor's title will be **the AI /LLM Technology Support Specialist**. This part-time position works with the Graduate Student worker to help assist in the hands-on integration tasks across the data pipeline: exporting candidate content (text, PDFs, images with OCR), preparing prompts and batches for LLM processing, validating returned subject, and transforming outputs for ingestion. **The UF Graduate Assistant** will design and implement an innovative workflow utilizing UF's NaviGator AI and multiple LLMs to generate subject metadata for digital library collections.

Supporting Staff includes **the Programming Advisor** (a paid service provided by the UFIT Research Computing) and **Function Testers**. The advisor position guides the development sub-group to set up the bulk environment on HiPerGator and integrate NavigatorAI functions with the planned features. Testers provide users' perspective to the development sub-group as they work on developing each function area. We will recruit five current library staff to conduct this work.

Consultants

We have invited our UF colleagues to consult on different aspects of the project, leveraging their expertise areas. The core group will meet bimonthly to demo new functions and report progress, seeking feedback and advice as the project progresses. This group includes:

Erin Gallagher is the Chair of Acquisitions & Discovery Services at the George A. Smathers Libraries. She is responsible for the stewardship of the Libraries' $15 million materials budget. In her capacity, she keeps current on the vendor's latest products and services, including those that incorporate AI tools.

Daniel Maxwell, an experienced AI trainer at UFIT Research Computing, previously served for years as an Informatics Consultant at the George A. Smathers Libraries. This background gives him an understanding of library practices and expertise in using the two UF services, Navigator AI and HiPerGather.

Daisy Zhe Wang is Professor and Director of Data Science Research Lab at UF. Her latest research includes multimodal data. Multimodal data is information collected from more than one modality or sense, such as text, images, sounds, video, or sensor readings, that describe the same event or phenomenon. Digital collections represent a valuable source of multimodal data. This presents a future opportunity to integrate and apply her multimodal data research to enhance digital collection practices and infrastructure.

Advisory Group

To prepare for the future expansion of the LLM bulk environment, we've invited scholars, researchers and practitioners across the U.S. and formed an advisory group. The following experts have agreed to be part of this project. The project team and the advisory group will meet quarterly.

Rebecca Bakker is the Digital Collections Librarian at the Florida International University. Her expertise covers AI, metadata and data analysis.

Julia Simic is the Director of Digital LIbrary Services at the University of Oregon Libraries. She spearheaded the transition of Oregon Digital, a massive digital collection shared by Oregon State University Libraries and Press and the University of Oregon Libraries, to a new digital collection platform. This involved coordinating the efforts of several library teams.

Hannah Tarver is the Supervisor of Digital Lab at the Digital Libraries Division, one of the eight divisions comprising University Libraries at the University of North Texas.Tarver has been one of most innovative researchers in Metadata Quality of academic digital libraries since 2010. Her publications have provided new methods to address critical and fundamental questions in metadata practices.

Yifan Wang is an Assistant Professor in the Department of Computer Science at the University of Hawaii at Manoa. He specializes in database architecture design for AI development and deployment.

Macia Lei Zeng is a Professor Emeritus of Information Science at Kent State University and the recipient of the 2024 ASIS&T Award of Merit. Her research interests include knowledge organization systems, metadata, and digital humanities, with over 100 research papers and six books.

To further enrich our advisory group, we intend to incorporate principal contributors from other AI initiatives that are transforming digital collection methodologies. Specifically, we have identified the following relevant projects: The *AI for Transcription in Digital Collections* initiative at the Minnesota Digital Library and *The Semantic Search for Digital Collections* project currently underway at Northwestern University.

Project Results

The deliverables at the end of this pilot project includes the following:

- A functional bulk work environment powered by NavigatorAI and HiPerGator
- Tens of thousands of new access points to about 20,000 newspaper issues
- A website documenting the major technical decisions as well as tested prompts and correspondent results
- Conference presentations
- Peer-reviewed Open-Access Journal articles

Leveraging two advanced UF services, our LLM assisted human steering bulk work environment provides the opportunity to scale up digital collection processes without compromising the quality. Specifically during the pilot phase, we address the rooted deficiency of the conventional item-by-item process by providing rapid work with consistent quality in results. Users will gain greater insight into two centuries of newspaper issues in the Florida Digital Newspaper Library through the addition of new subject headings. They can explore the historical transformation of two Florida cities, tracing the changes that occurred as the United States developed as an independent nation.

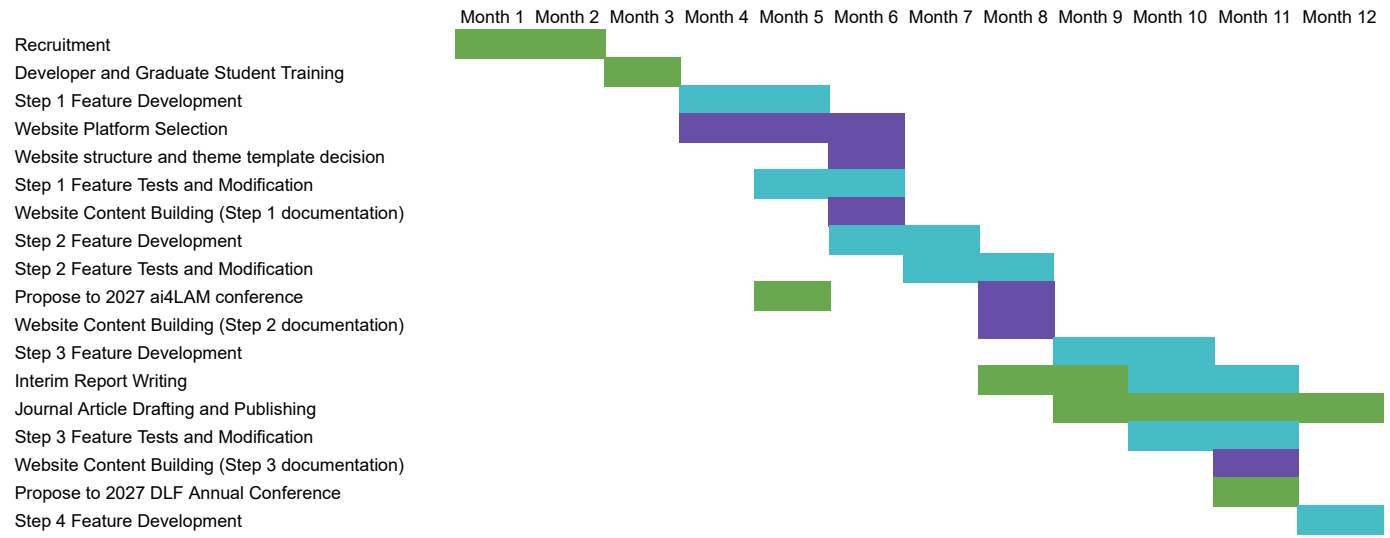
We will publicize our learnings on our website and shape the website as the central location to store and share all technical documents and our conference presentations. We will participate in three conferences: ai4lam, code4Lib and DLF Forum so we can disseminate our learnings to a wide audience and also seek institutional partners to expand the project beyond this pilot phase.

Built upon the achievements of the pilot phase, we plan to extend LLM support to address critical needs for improving the accessibility of our digital collections at scale. This includes features such as generating transcripts for audio and video, creating alt-text for images, and providing content summaries for all file types. We will also incorporate materials in other languages.Since the core development is already complete, expanding these features requires only trivial additional work. The main focus will be on thorough testing and refinement of the prompts for various specific tasks and this work can be included into UF staff's daily tasks. Following the pilot phase, the annual maintenance cost for this environment—primarily for NavigatorAI and HiperGator—is projected to be $1,160. This amount is notably minimal, only a fraction of the subscription costs for conventional library vendor services. Given the demonstrably fast and high-quality results, the administration at the George A. Smathers Libraries is anticipated to incorporate this expense into their standard operational budget. Opening up our digital collection's bulk work environment to partner institutions represents a significant, complex endeavor. This expansion demands extensive planning and the involvement of an expanding group of stakeholders to address critical challenges such as security, legal compliance, financial overhead, and customer service.

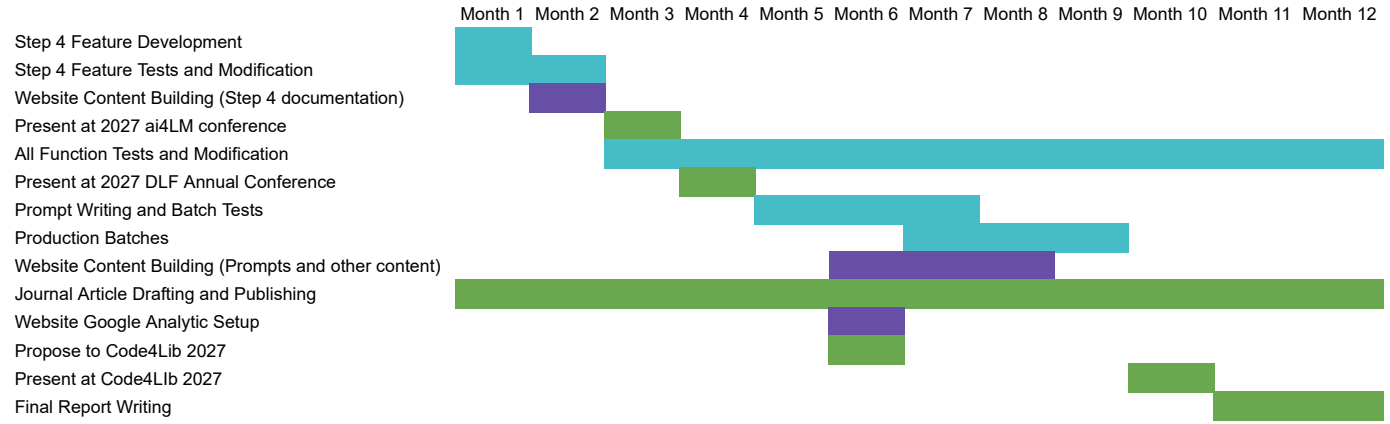
Applicant Name: Xiaoli Ma, Todd Digby, Laura Perry

Project Title: Accelerating History--Piloting Multiple Large Language Models to Generate Subject Keywords for Historic Florida Newspapers

Year One (2026 Sep. to 2027 Aug.)



Year Two (2027 Sep. to 2028 Aug.)



Meeting: The project team will meet every month, or more often if a specific task requires it. The project team and the consultant group are scheduled to meet every two months. Meetings between the project team and the advisory group will occur every quarter.