

# System Design for Increasing Adoption of AI-Assisted Image Tagging in Wikimedia Commons

Yihan Yu  
University of Washington

David W. McDonald  
University of Washington

## Abstract

In this research, we aim to investigate designs to increase the adoption and satisfaction of AI-assisted tools within commons-based peer production (CBPP) projects, with a specific focus on Wikimedia Commons. While AI-powered automation tools have long been integrated into CBPP projects for indirect tasks like content moderation, the utilization of AI for direct content generation has surged with recent advancements in generative AI algorithms. However, the impact of AI-assisted tools on human contributors and the design considerations to enhance their interaction, adoption, and satisfaction remain uncertain. This study proposes to co-design an AI-assisted image tagging tool with Wikimedia Commons contributors and users to increase adoption and satisfaction. We will perform a study of the prior WFM attempt to provide a computer-aided tagging (CAT) tool to understand the factors that led to its deactivation. We will then investigate technology designs to improve AI-assisted image tagging for structured Commons. The successful completion of this project is expected to advance the development of an AI-assisted image tagging tool on Wikimedia Commons, promoting greater adoption, usage, and satisfaction among contributors. Additionally,

the insights gained from this study can be generalized to enhance interaction and collaboration between human contributors and AI-powered automation tools in the broader Wikimedia tools ecosystem and other CBPP projects.

## Introduction

AI-powered automation tools have long been integrated into commons-based peer production (CBPP) projects for indirect tasks such as content moderation<sup>1</sup> and evaluating contribution quality<sup>2</sup>. In recent years, with the rapid advancement of generative AI algorithms, there has been a surge in attempts to utilize AI-powered tools for generating direct content in CBPP, including creating Wikipedia articles [1] and generating image annotations<sup>3</sup>. AI holds the promise of enhancing content creation efficiency and consistency, thereby addressing content gaps in areas that may have received insufficient human contributions. However, the impact of AI-assisted tools on human contributors as well as the technology designs to enhance human contributors' interaction, adoption, and satisfaction with AI-assisted tools in CBPP remain uncertain. In this proposed research, we will fill in this gap by co-designing an AI-assisted image tagging tool with contributors and users of Wikimedia Commons,

---

<sup>1</sup><https://meta.wikimedia.org/wiki/Research:Detox>

<sup>2</sup><https://www.mediawiki.org/wiki/ORES>

<sup>3</sup>[https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Computer-aided\\_tagging](https://commons.wikimedia.org/wiki/Commons:Structured_data/Computer-aided_tagging)

with the goal of increasing adoption and satisfaction.

Wikimedia Commons is a WMF project that makes multimedia resources available for free copying, usage, and modification.<sup>4</sup> However, a lack of structured, machine-readable metadata about media files has hindered its accessibility, searchability, usability, and multilingual support. Recently, WMF researchers attempted to introduce computer-aided image tagging (CAT).<sup>5</sup> Unfortunately, our prior research [2] revealed low adoption of the CAT tool by Commons contributors. Participants reported unsatisfactory usability and performance of the tool and resistance to changing their existing workflow of creating and maintaining the local category system. The CAT tool was deactivated in September 2023.

In this project, our aim is to study the previous WMF attempt to provide the CAT tool, understand the factors that led to its deactivation, and explore technology designs that enhance the adoption, usage, and satisfaction of AI-assisted tagging among Commons contributors. Our research questions are:

1. How do AI-assisted image tagging and structured data on Commons affect the work and workflows of diverse contributors and user communities in related Wikimedia projects, such as Commons, different language versions of Wikipedia, and Wikidata?
2. What are the perceptions, concerns, and preferences of Commons contributors and users regarding the quality of tags suggested by AI algorithms?

3. What usability issues and challenges do contributors encounter when using the CAT tool on Commons?
4. What technology designs can enhance the quality of suggested tags, identify appropriate tags for a Depicts statement, and improve the overall user experience with AI-assisted image tagging on Wikimedia Commons?

The successful completion of this project would advance the development of an AI-assisted image tagging tool on Wikimedia Commons, fostering greater adoption, usage, and satisfaction among contributors. This enhanced tool will provide Commons contributors with a more efficient, accurate, and user-friendly method of adding structured data to multimedia files. Improved structured data will enhance the searchability and usability of multimedia resources across WMF projects, promoting inclusivity among diverse language communities. Furthermore, the design insights from this study can be applied broadly to improve interaction and collaboration between human contributors and AI-powered tools across the Wikimedia tools ecosystem and other CBPP projects.

**Date:** 6/15/2024 - 12/15/2024

## Related work

Wikimedia Commons is understudied and underappreciated. Our previous study [2, 3, 4] aimed to understand Commons, its contributors, and its relation to other WMF projects. This study received coverage in the research newsletter,<sup>6</sup> and the comments from Wikipedians who discussed it expressed appreciation for our efforts to investigate some of the less explored aspects of Commons.

---

<sup>4</sup><https://commons.wikimedia.org/wiki/Commons:Welcome>  
<sup>5</sup>[https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Computer-aided\\_tagging](https://commons.wikimedia.org/wiki/Commons:Structured_data/Computer-aided_tagging)

<sup>6</sup><https://meta.wikimedia.org/wiki/Research:Newsletter/2023/August>

Despite its significance for WMF projects and internet users, Commons faces challenges in content search and reuse, especially for non-English users, due to insufficient structured metadata. The Structured Commons project<sup>7</sup> aimed to address this but encountered low adoption due to unsatisfactory computer-aided tagging tools and resistance to changing the existing category system. Our study identifies a gap in understanding how to develop AI-assisted tagging interactions accepted by Commons contributors. To bridge this, our project engages the Commons community in a co-design process for contributor-algorithm collaboration.

## Methods

This study will comprise two stages. In the initial stage, we will conduct qualitative interviews with contributors and users of Wikimedia Commons to gather insights into their experiences with the deactivated WMF CAT tool. The outcomes of this first stage will inform the second stage of the study, where we will codesign solutions with contributors and users of Commons to enhance the adoption and satisfaction of AI-assisted image tagging tools.

### Stage One - Learning from a Prior AI Tagging Project

In 2020, the Structured Data on Commons team introduced a computer-aided tagging tool known as the CAT tool<sup>8</sup>. This standalone MediaWiki extension, opt-in for registered auto-confirmed contributors, utilized the Google Vision API<sup>9</sup> to suggest tags for images uploaded to Commons. These tags were then translated into Wikidata IDs through freebase-wikidata mappings<sup>10</sup>. Upon uploading, contributors

received notifications about new suggested tags, prompting them to review and confirm or ignore individual tags. Confirmed tags were subsequently added to the Depicts statements<sup>11</sup> for the image files.

The release of the CAT tool sparked discussions [5, 6, 7] primarily initiated by experienced Commons contributors regarding its utility. Our own preliminary analysis indicates that while some contributors found the experiment and output interesting, significant issues surfaced during the trial phase, necessitating thorough assessment and analysis. These challenges include the unclear goal, scope, granularity, and ontology of Depicts statements/structured data, as well as uncertainties about how depicts function on Commons. Regardless of the efforts to involve Commons members, the resulting tool suffered from usability issues such as confusing terminology (tags vs. depicts), repetitive notifications (inability to skip an image without adding tags), and the inability to edit tags, add addendums, or mark certain tags as rejected or inapplicable. Furthermore, the quality of suggested tags needs improvement, as many offered tags (e.g., 'stock photography', 'monochrome') are generic keywords that some Commons contributors maintain are unsuitable for a Depicts statement. Enhancing the tool's efficacy might involve gathering additional context about the nature of the uploaded image through image description/classification.

Following these discussions, the CAT tool was deactivated in September 2023 and officially declared a failed project following a community evaluation.

In the initial phase of our study, we will first locate the discussions where the CAT tool was

---

<sup>7</sup>[https://commons.wikimedia.org/wiki/Commons:Structured\\_data](https://commons.wikimedia.org/wiki/Commons:Structured_data)

<sup>8</sup>[https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Computer-aided\\_tagging](https://commons.wikimedia.org/wiki/Commons:Structured_data/Computer-aided_tagging)

<sup>9</sup><https://cloud.google.com/vision/docs>

<sup>10</sup><https://developers.google.com/freebase/#freebase-wikidata-mappings>

<sup>11</sup><https://commons.wikimedia.org/wiki/Commons:Depicts>

mentioned and analyze those discussions and posts to understand contributors' concerns and what they desired from the tool. Subsequently, we will broaden the dialogue beyond experienced Commons contributors by interviewing individuals with varying levels of experience, as well as those engaged in other Wikimedia projects supported by Commons, such as different language versions of Wikipedia and Wikidata. Through these interviews, we aim to delve deeper into the issues surrounding the deactivated CAT tool. The key areas of discussion include:

1. Scope and Goal of Depicts Statements/Structured Data on Commons: We will guide participants in discussing how these elements influence and impact the current work and workflows of diverse contributors and user communities, as well as the collaborations between these communities.
2. Quality of Tags Suggested by the CAT Tool: During each interview, we will present 3-5 example images and ask the interviewee to suggest Depicts statements. Following this, we will present and discuss tags suggested by the CAT tool with the interviewee to comprehend participants' perceptions, concerns, and preferences regarding AI-assisted tagging. This process aims to identify cognitive distinctions, areas of agreement on AI-generated tags, and ways to enhance the suggestions.
3. Usability of the CAT Tool: During each interview, we will showcase a live demo of the CAT tool, allowing participants to ask questions and provide feedback on the tool's usability.

We acknowledge that parts 2 and 3 of the interview will involve interactions with the now-deactivated CAT tool, for which we do not have access. To address this challenge, we will contact the structured data team for limited access to the tool (such as through an OAuth protected LiftWing instance<sup>12</sup> to facilitate the interview process.

Alternatively, we can attempt to replicate the methodology employed by the deactivated CAT tool to generate suggested tags. This involves submitting sample images to the Google Vision API to retrieve corresponding 'labels.' Subsequently, these labels will be mapped from Freebase objects to Wikidata entities using a Freebase/Wikidata Mappings file<sup>13</sup> provided by Google. This process will result in the generation of suggested tags. Following this, we will develop an interactive demo allowing users to visualize the image and evaluate the suggested tags. This demo will closely adhere to the layouts and functionalities demonstrated in the interactive prototype of the deactivated CAT tool<sup>14</sup>. It's important to note that this entire effort is aimed at creating a version for study purposes, with the intention of utilizing the deployed, deactivated version as the primary tool, making this replication process a fallback option.

Qualitative information gathered from our stage 1 interviews will be coded using a thematic approach [8]. We will conduct open coding to identify initial themes and patterns within the data. Subsequently, we will create a preliminary coding scheme based on recurring topics, issues, or sentiments expressed by participants and iteratively refine the coding scheme as new themes emerge during the analysis. The main goals, issues, challenges, and opportunities

---

<sup>12</sup>[https://wikitech.wikimedia.org/wiki/Machine\\_Learning/LiftWing](https://wikitech.wikimedia.org/wiki/Machine_Learning/LiftWing)

<sup>13</sup><https://developers.google.com/freebase/#freebase-wikidata-mappings>

<sup>14</sup><https://wikimedia.invisionapp.com/console/share/BZTUW XU3E95>

related to structured data and the deactivated CAT tool, as revealed through participant interviews, will be summarized.

Our stage 1 findings will be posted on common community forums and related discussion pages to gather community input. These findings will also guide the design of the second stage of the study, in which we will co-design with Commons community members to address the identified issues in AI-assisted image tagging tools on Commons, aiming to better serve the goals of different communities and enhance user adoption, experience, and satisfaction.

### **Stage Two - Co-Design for Enhanced Adoption and Satisfaction**

In the second stage, we will actively involve members of the Commons community in a co-design study aimed at refining the CAT tool. Through participant interviews, we will collaboratively explore:

1. **Technology Designs:** We will investigate technology designs that utilize current AI technologies, including human-in-the-loop image annotation [9] and tag refinement methods [10] based on human-generated category information and file descriptions. The goal is to enhance the quality of suggested tags.
2. **Tag Appropriateness and Filtering:** We will explore ways to identify tags suitable for a Depicts statement and effectively filter out generic keywords or tags.
3. **Usability Enhancement Designs:** We will work on designs to improve the overall user experience and address the usability issues identified in the stage one interviews.

We will use the same iterative thematic coding process to analyze qualitative data and design artifacts from the stage two study, summarize our findings, and report them to the communities.

In both stages of the study, we will use three approaches to engage participants who do not speak English:

1. **Multilingual Interviewers:** both authors of this study speak English and/or Mandarin Chinese only. To engage participants who do not speak these two languages, we will identify, recruit, and train students from our public university who are fluent in the languages spoken by Commons contributors to conduct the interview following the approach we adopted in [11]. We will then use professional translation services to transcribe the interview scripts into English for our data analysis.
2. **Multilingual Materials:** we will make sure that all study-related materials, including demos, image tags and metadata are available in the languages spoken by the participants.
3. **Flexibility in Participation:** we will offer flexible interview schedules to accommodate contributors from various time zones and let participants decide the video conference software we use in the interview.

## **Expected output**

This project will produce three outputs:

The first output will comprise research reports for the Wikimedia community. Throughout the project, we will distribute two interim reports, providing updates on our research progress. The initial interim report, scheduled for August

2024, will present findings from the first stage of the project, during which we learn from the deactivated CAT tool. Subsequently, the second interim report, anticipated in November 2024, will present the results from the second stage, focusing on co-designing AI-assisted image tagging systems for enhanced adoption and satisfaction. These interim reports will be shared with the research participants and the Wikimedia community via our Meta-Wiki research page and the village pump to solicit feedback. Upon project completion in December 2024, we will issue a final research report. This report will summarize and contextualize our study findings and offer actionable insights to aid community decisions concerning addressing contributors' and users' concerns and improving the design and implementation of AI-assisted image tagging tools on Commons. We will publish the final report and present our results from the report at Wiki Workshop 2025. Additionally, this report will serve as a first step for the research team to initiate our next project: the development, deployment, and testing of an enhanced AI-assisted image tagging tool on Commons.

The second output will be at least one paper manuscript to be submitted to leading HCI conferences such as CSCW or CHI. This paper is intended for academic researchers, practitioners in human-computer interaction, and the Wikimedia research community. Its aim is to contribute to the academic understanding of the opportunities, challenges, and implications of designing to support interactions and collaborations between human contributors and AI-powered automation tools in the Wikimedia ecosystem and more broadly, in other CBPP projects.

The third output will involve presenting our research progress and outputs at events such as Wiki Workshop 2024, Wikimania 2024, and Wiki Workshop 2025 for Wikimedia contributors, developers, and interested community members.

## Risks

AI algorithms can introduce bias, resulting in inaccurate or unfair tagging, especially in sensitive content areas like medical images [12] or diverse cultural contexts. In this study, we will employ two approaches previously utilized in the deactivated CAT tool to ensure ethical implementation of AI algorithms. Firstly, we will integrate the SafeSearch<sup>15</sup> function provided by the Google Vision API to detect explicit content, such as adult or violent content, within images. This function classifies content into five categories (adult, spoof, medical, violence, and racy) and assigns a likelihood score for each category's presence in a given image. By leveraging this capability, we can effectively identify and temporarily flag files susceptible to inaccurate or unfair tagging for manual review. Secondly, we will implement Commons' blocklist feature<sup>16</sup>, which filters out tags considered unsuitable for Commons for various reasons when suggested by the algorithm. In addition to these technical measures, we will also seek the opinions of participants in our study, particularly those with expertise in sensitive domains, on how to address bias and/or unfair tagging.

Contributors may resist workflow changes and be reluctant to adopt the new AI-assisted tagging system, particularly if they've had unsatisfactory experiences with similar tools before. To address this, we will adopt a co-design research methodology [13] in our study

---

<sup>15</sup><https://cloud.google.com/vision/docs/detecting-safe-search>

<sup>16</sup>[https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Computer-aided\\_tagging/Blocklist](https://commons.wikimedia.org/wiki/Commons:Structured_data/Computer-aided_tagging/Blocklist)

to empower the contributors by giving them a sense of ownership in the development of the new system. [14, 15] When individuals feel their input is valued and directly contributes to the final product, they are more likely to embrace and support the changes.

## Community impact plan

We will engage individuals with varying levels of experience with Commons, as well as those involved in other Wikimedia projects supported by Commons, such as different language versions of Wikipedia and Wikidata, in our interviews. Additionally, we will actively involve Wikimedia Commons contributors, particularly those engaged in image categorization, in a user-centered co-design process for the AI-assisted tagging system. Simultaneously, we will seek input from the broader Wikimedia community. Regular research updates will be provided through our project page, and we will utilize community channels and mailing lists to gather insights and address concerns. Furthermore, we will present our findings at various events such as Wiki Workshop and Wikimania, and publish our paper under open access to maximize dissemination.

## Evaluation

The success of the research project will be measured by its positive impact on improving the adoption, usage, and satisfaction of AI-assisted image tagging tools on Wikimedia Commons, as demonstrated by the Wikimedia community's discussion, acceptance, and implementation of proposed system designs. Impact in relevant academic fields will be measured by citations of our academic research papers, which will serve as a foundation for further analysis.

## Budget

We do not request a budget for conference/event travel and accommodation because we will apply to receive travel support from our departmental student travel support program.

Ph.D. student stipend: Half-Time (50% FTE)  
Monthly Rate: \$3,332 \* 6 months = \$19,992

Participant compensation: \$20 \* 50 participants  
= \$1,000

Total: \$20,992

## Response to reviewers and meta-reviewers

**Meta:** Help us understand how the project will engage Commons contributors who do not speak English.

**Authors' Response:** We have added our approaches to engage participants who do not speak English in the Methods section.

**Meta:** Strengthen the impact statement of your proposal for Stage II and consider reducing the budget request (and scaling down the scope as relevant) to increase the chances of success.

**Authors' Response:** We have strengthened the impact statement of our proposal in the Introduction and Expected Outputs section. Additionally, we have scaled down the scope of our proposal from an entire user research, co-design, development, deployment, and testing cycle to user research and co-design only. The findings from this project will inform our future research in developing, deploying, and testing an advanced AI-assisted tagging tool on Commons.

**Reviewer 1 (R1):** Research contributions could be made clearer.



**Authors' Response:** We have clarified our research contributions in the Introduction and Expected Outputs sections.

**Reviewer 1 (R1):** Improve the clarity of the research question.

**Authors' Response:** We have clarified our research questions in the Introduction section.

**Reviewer 2 (R2):** It is unclear whether the insights gathered in the first part of the research plan will be used to implement a new automated tagging system.

**Authors' Response:** We have provided a more detailed research design in the Methods section.

**Reviewer 2 (R2):** More detail is needed on the actual machine learning pipeline to implement the automated labeling.

**Authors' Response:** We have provided additional details on the machine learning pipeline in the Methods section.

**R1 & R2:** The section about risks would benefit from a more comprehensive discussion of how AI is going to be implemented in an ethical way.

**Authors' Response:** We have added discussions on how we ensure ethical implementation of AI algorithms in the Risks section.

**Reviewer 3 (R3):** It would be really useful to know what percentage of full-time equivalent (FTE) the doctoral student commits to the project outside their other obligations.

**Authors' Response:** The doctoral student is expected to work half-time (50% FTE; 20 hours per week) on the project for six months. Also, this proposed project is related to her doctoral thesis.

## References

1. Simonite, T. (2018). Using artificial intelligence to fix Wikipedia's gender gap.

2. [https://meta.wikimedia.org/wiki/Research:Studying\\_Wikimedia\\_Commons](https://meta.wikimedia.org/wiki/Research:Studying_Wikimedia_Commons)
3. Yu, Y., & McDonald, D. W. (2022). Unpacking Stitching between Wikipedia and Wikimedia Commons: Barriers to Cross-Platform Collaboration. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-35.
4. Yu, Y., & McDonald, D. W. (2023). "Why do you need 400 photographs of 400 different Lockheed Constellation?": Value Expressions by Contributors and Users of Wikimedia Commons. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2), 1-34.
5. [https://commons.wikimedia.org/wiki/Commons\\_talk:Structured\\_data/Computer-aided\\_tagging](https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Computer-aided_tagging)
6. <https://phabricator.wikimedia.org/T339902>
7. [https://commons.wikimedia.org/wiki/Commons:Village\\_pump/Archive/2020/02/#Misplaced\\_invitation\\_to\\_%22tag%22\\_images](https://commons.wikimedia.org/wiki/Commons:Village_pump/Archive/2020/02/#Misplaced_invitation_to_%22tag%22_images)
8. Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis?. *Qualitative research in psychology*, 18(3), 328-352.
9. <https://docs.aws.amazon.com/sagemaker/latest/dg/a2i-use-augmented-ai-a2i-human-review-loops.html>
10. Liu, D., Hua, X. S., Wang, M., & Zhang, H. J. (2010, October). Image retagging. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 491-500).
11. Bipat, T., Alimohammadi, N., Yu, Y., McDonald, D. W., & Zachry, M. (2021). Wikipedia Beyond the English Language Edition: How do Editors Collaborate in the Farsi and Chinese Wikipedias?. *Proceedings of the ACM on*



- Human-Computer Interaction*, 5(CSCW1), 1-39.
12. Menking, A., & McDonald, D. W. (2020). Image Wishlist: Context and Images in Commons-Based Peer Production Communities. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1-21.
  13. Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *Co-design*, 4(1), 5-18.
  14. Bødker, S., Ehn, P., Sjögren, D., & Sundblad, Y. (2000, October). Co-operative Design—perspectives on 20 years with ‘the Scandinavian IT Design Model’. In *proceedings of NordiCHI* (Vol. 2000, pp. 22-24).
  15. Trischler, J., Pervan, S. J., Kelly, S. J., & Scott, D. R. (2018). The value of codesign: The effect of customer involvement in service design teams. *Journal of Service Research*, 21(1), 75-100.