

Unbiased AI for Computational Poetry Analysis on Massive Digital Collections

Dr. Kahyun Choi, Principal Investigator, Assistant Professor of Information and Library Science at Indiana University Bloomington, requests \$383,955 for a three-year Laura Bush 21st Century Early Career Development project. The project develops AI-based computational poetry analysis tools that are mindful of common biases in data-driven machine learning systems. With a collaboration with HathiTrust Research Center, the project team aims at improving accessibility, diversity, and equity in the curation of massive digital collections of poetry.

Project Justification

Designing and developing machine learning (ML) algorithms that achieve the human-level understanding of literature has been an essential library and information science research topic (Cordell, 2020). Metadata derived from such algorithms can increase patrons' accessibility to massive digital collections if their quality is high enough. Also, through advanced computational tools for literature, digital humanities scholars and students can conduct analyses on large-scale collections, such as HathiTrust Digital Library's 17+ million digital items. Although AI models have achieved a deeper understanding of some text than before, such as news articles and product reviews, poetry has been rarely explored in previous research. It is because poems are very challenging even for humans to understand due to their figurative language and multiple layers of meanings. Thus, making machines understand poetry is a challenging task that tries to catch up with high-level human intelligence, pushing the limit of AI. However, applying ML methods to library research requires awareness and management of ethical concerns, such as biases, which can be caused by various reasons, including imbalanced data collection and annotation. Without putting in extra effort, AI would discriminate against underrepresented populations based on gender, race, socio-economic background, age, and more. For instance, facial recognition systems are known to be more accurate for men than women and for the light-skinned than dark-skinned.¹

This project builds on the PI's prior research on recognizing emotion and themes from song lyrics for music digital libraries. Based on her experiences in using advanced deep learning methods and her knowledge in the promotion of community-driven massive data collections, she will further explore ethical and socially responsible AI models. As an "early career development" project, it will provide opportunities for the PI to establish her long-term research agenda, secure computing and institutional resources, tighten collaboration with HathiTrust Research Center (HTRC) as well as librarians and digital humanities scholars.

Project Work Plan

This project focuses on the following research questions: (1) How to train an AI model into understanding poetry? (1-1) What kind of metadata can be drawn from the trained AI by having it analyze poems? (1-2) In such a computational analysis, how useful if some auxiliary information is available, such as word-level scores for affectiveness, the author's notes on the poem, and poem commentaries? (2) Can a computational poetry analysis tool improve the symbiotic relationship between patrons, libraries, and digital humanities scholars? (3) What kind of biases are there when an AI model attempts to understand poems, and how to reduce them? This project will address these questions by developing ethical and responsible AI models. We will also build reusable poetry worksets that reflect cultural, demographic, thematic diversity. The project team, including the PI and a graduate assistant, will regularly seek advice from the advisory board, composed of experts in English poetry, ethical AI, digital libraries, and digital humanities. The planned project activities are as follows:

Year 1. Develop the AI algorithms to estimate emotion and theme from poems. The project team will build advanced natural language processing models using deep learning that achieve a high-level understanding of poems, such as their emotion and theme. The main models will be trained using data from online poem digital libraries, such as *poets.org*. However, since the poems there are not annotated by the emotion they convey, the team will explore *data augmentation* approach by leveraging similar kinds of text collections, such as song lyrics. Because song lyrics have been better annotated than poems and are available in the form of a public dataset as well (Delbouys, 2018), a model learned from this similarly difficult text can be transferred to the proposed poetry analysis tasks with a small adjustment. Also, as for emotion detection, we also believe that incorporating the well-studied word-level affective scores into the deep learning model as a "conditioning mechanism" is a sensible approach. Moreover, in the previous research, the PI discovered that music listeners' interpretations could provide additional clues when an AI model attempts to understand themes from song lyrics. Along this line, the project team will examine whether using auxiliary data, such as the author's explanation of her own

¹<https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>

poems and the readers' commentaries, can improve the models' performance. We will evaluate the validity of the proposed method by testing the learned models' prediction accuracy compared to the domain experts' opinions on a separate set of poems. **Year 2. Disseminate the initial AI models through HTRC and improve them based on feedback.** The project team will disseminate the initial models as an open-source project to the research community and general public. In addition, based on the already established collaborative relationship with HTRC and the PI's experience with the HTRC framework, the team will incorporate the models into the HTRC's interfaces as an additional analytic tool for poetry. We envision that our dissemination activities can bring awareness of computational poetry analysis among digital humanities scholars, librarians, and patrons. In addition, it will also unleash the potential of the massive digital collection of poetry available in HathiTrust Digital Library. We will also conduct interviews and surveys to collect feedback on the proposed tools and then fine-tune the models accordingly. **Year 3 Identify and suppress the biases in AI models for poetry analysis.** Among various sources of biases in AI, the team will try to identify biases created by the training collections, for example, whether the racial and ethnic minorities are well represented in the author population. We will also identify biases stemmed from different annotation schemes: there can be an emotion or theme popular only in a certain underrepresented social group captured by only a smaller number of poets. The process will be guided and advised by the advisory board. Then, we will build unbiased worksets, from which we can re-train our machine learning models. Data augmentation schemes are another planned approach to strengthen underrepresented data categories, for instance, by transferring learned knowledge from song lyrics and auxiliary data. The final models will be provided back to the research community and public as an unbiased poetry analysis tool both via HTRC and an open-source platform. Further adjustment can be achieved by the user's participatory effort and the team's potential collaboration with digital humanities scholars and librarians. The team will also publicize the findings in various conferences and journals in the area of library and information science, digital humanities, and machine learning.

Diversity Plan

This research directly supports diversity because one of the main goals is identifying and removing ML methods' biases for poetry analysis, such as gender, class, and race. To address gender imbalance among machine learning researchers², the PI will collaborate with more women students; as a faculty mentor, the PI will work with women students through the Emerging Scholars Research Experience for Undergraduate Women program offered by IU's Center of Excellence for Women & Technology and the Women in Music Information Retrieval program.

Project Results

The project will create and publicize AI-based computational poetry analysis tools as well as unbiased poetry collections through HTRC and an open-source platform. The project activities are expected to improve patron's, librarians', and digital humanities scholars' access to diverse poetry collections in digital libraries, because the enriched metadata can provide an additional interface to the massive data collection. Since our algorithms are mindful of biases in AI, the information retrieval and recommendation results are richer and more diverse. In the long term, we envision a virtuous cycle of improved diversity in readers' access to various poems, increased participation of readers in reviews and criticism, their contribution to the unbiased AI models. As we saw in the presidential inauguration ceremony, where Amanda Gorman stirred the audience with her poem, "The Hill We Climb," the project's ultimate goal is to heal people's minds by exposing them more to literature.

Budget

The proposed budget of \$383,955 includes \$141,084 for supporting a graduate assistant for 3 years; \$80,574 for 3 years of PI's summer salary and fringe benefits; \$12,000 for conference travel; \$40,000 for building Deep Learning server for the research; \$1,500 for data annotation; \$108,797 for indirect costs.

PI

Dr. Choi, is an assistant professor in the Dept. of Information and Library Science and Data Science program at Indiana University. Her research has been to build AI-based metadata enrichment systems for massive digital collections on music and song lyrics. She commits to making AI more ethical in this project. In her current IMLS grant, she develops education materials to improve accessibility to AI in underserved populations through public libraries (LG-250059-OLS-21).

²<https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/>