

# Preface

This collection of essays is the unexpected culmination of a 2018–2020 grant from the Institute of Museum and Library Services to the Hesburgh Libraries at the University of Notre Dame.<sup>1</sup> The plan called for a survey and a series of workshops hosted across the country to explore, originally, “the national need for library based topic modeling tools in support of cross-disciplinary discovery systems.” As the project developed, however, it became apparent that the scope of research should expand beyond topic modeling and that the scope of output might expand beyond a white paper. The end of the 2010s, we found, was swelling with library-centered investigations of broader machine learning applications across the disciplines, and our workshops demonstrated such a compelling mixture of perspectives on this development that we felt an edited collection of essays from our participants would be an essential witness to the moment in history. With remaining grant funds, we hosted one last workshop at Notre Dame to kick start writing.

The resulting essays cover a wide ground. Some present a practical, “how-to” approach to the machine learning process for those who wish to explore it at their own institutions. Others present individual projects, examining not just technical components or research findings, but also the social, financial, and political factors involved in working across departments (and in some cases, across the town/gown divide). Others still take a larger panoramic view of the ethics and opportunities of integrating machine learning with cross-disciplinary higher education, veering between optimistic and wary viewpoints.

The multi-disciplinarity of the essayists and the diversity of their research give each chapter a *sui generis* flavor, though several shared concerns thread through the collection. Most significantly, the authors suggest that while the technical aspects of machine learning are a challenge, especially when working with collaborators from different backgrounds, many of their key concerns are actually about the ethical and social dimensions of the work. In this sense, the collection is very much of the moment. Two large projects on machine learning, cross-disciplinarity, and libraries ran concurrently with our grant — Cordell 2020 and Padilla 2019, which were commissioned by major players in the field, the Library of Congress and OCLC, respectively — and both took pains to foreground the wider potential effects of machine learning. As Ryan Cordell puts it, “current cultural attention to ML may make it seem necessary for libraries to implement ML quickly. However, it is more important for libraries to implement ML through their existing commitments to responsibility and care” (1).

The voices represented here exhibit a thorough commitment to Cordell’s call for responsibility and care, and they are only a subset of the larger chorus that sounded at the workshops. We editors therefore encourage readers interested in this bigger picture to examine the meta-themes

---

<sup>1</sup>LG-72-18-0221-18: “Investigating the National Need for Library Based Topic Modeling Discovery Systems.” See <https://www.ims.gov/grants/awarded/lg-72-18-0221-18>.

and detailed information that emerged in the course of the workshops and the original survey through the grant's final report.<sup>2</sup> All of these pieces together capture a fascinating snapshot of an interdisciplinary field in motion.

We should note that the working methods of the collection's editorial team were an attempt to extend the grant's spirit of collaboration. Through several stages of development, content editors Don Brower, Mark Dehmlow, Eric Morgan, Alex Papson, and John Wang reviewed assigned essays and provided commentary before notifying general editor Daniel Johnson for prose editing, who in turn shared the updated manuscripts with the authors so the cycle could begin again. The submissions, written variously in Microsoft Word or Google Docs format, were ushered through these stages of life in team Google Drive folders and tracked by spreadsheet before eventual conversion by Don Brower into a series of TeX files, provisioned in a version controlled Github repository, for more fine-tuned final editing. Like working with diverse teams in the pursuit of machine learning, editing essays together in this fashion, for publication by the Hesburgh Libraries, was a novel way of collaborating, and we editors thought candor about this book-making process might prove insightful to readers.

Attending to the social dimensions of the work ourselves, we must note that this collection would not have been possible without the generous support of many people and organizations. We would like to thank the IMLS for providing essential funding support for the grant and the Hesburgh Libraries' Edward H. Arnold University Librarian, Diane Parr Walker, for her organizational support. Thank you to the members of the Notre Dame IMLS grant team who, at its various stages, provided critical support in managing logistics, conducting research, facilitating workshops, and analyzing results. These individuals include John Wang (grant project director), Don Brower, Mark Dehmlow, Nastia Guimaraes, Melissa Harden, Helen Hockx-Yu, Daniel Johnson, Christina Leblang, Rebecca Leneway, Laurie McGowan, Eric Lease Morgan, and Alex Papson. The University of Notre Dame Office of General Counsel provided key publication advice, and the University of Notre Dame Office of Research provided critical support in administering the grant. Again, many thanks.

We would also like to thank the co-signatories of the IMLS Grant Application for supporting the project's goals: Mark Graves (then Visiting Research Assistant Professor, Center for Theology, Science, and Human Flourishing, University of Notre Dame), Pamela Graham (Director of Global Studies and Director of the Center for Human Rights Documentation and Research, Columbia University Libraries), and Ed Fox (Professor of Computer Science and Director of the Digital Library Research Laboratory, Virginia Polytechnic Institute and State University). And of course, thanks to the 95 participants in our 2019 IMLS Grant Workshops (too many to enumerate here) and to the essay authors for sharing their expertise and perspectives in growing our collective knowledge of machine learning and its use in research, scholarship, and cultural heritage organizations. Your active engagement continues to shape the field, and we look forward to your next achievements.

## References

Cordell, Ryan. 2020. "Machine Learning + Libraries: A Report on the State of the Field." Commissioned by LC Labs, Library of Congress. <https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf>.

---

<sup>2</sup>See <https://doi.org/10.7274/r0-320z-kn58>

Padilla, Thomas. 2019. "Responsible Operations: Data Science, Machine Learning, and AI in Libraries." Dublin, Ohio: OCLC Research. <https://www.oclc.org/research/publications/2019/oclcresearch-responsible-operations-data-science-machine-learning-ai.html>.