Leave the Judgment to Me
Musings on Artificial Intelligence and Subject Cataloging

By Laura E. Daniels

Subject analysis is a critical part of creating the metadata underlying library catalogs that help our users find, identify, select, obtain, and explore resources based on the topics they address. In the Program for Cooperative Cataloging (PCC) Catalogers Learning Workshop (CLW) [training modules focused on Library of Congress Subject Headings](#) (LCSH), subject analysis is described as "the process of examining a resource and figuring out what the resource is, and what it is about" (Module 1.3, slide 6). This conceptual process is a necessary predicate to subject assignment, or the "translation of the aboutness into specific notation or terms" (Module 1.4, slide 2). Multiple approaches are outlined for the process of determining this aboutness; most catalogers use a combination of approaches, and the best predominant approach may differ based on the resource being described. The complexity of this assessment—which must precede any attempts to assign terms from a controlled vocabulary—provides multiple opportunities for AI to misinterpret, mischaracterize, and introduce bias or error. This potential for error, coupled with demonstrated inconsistency in selecting appropriate terms from the LCSH vocabulary itself, leads me to have significant concerns about using generative AI to perform subject analysis and assignment.

My consideration of this topic was spurred by the experience of cataloging what I found to be a particularly challenging text in terms of LC subject assignment. The book, *[Possessed landscapes: experiments in conservation and sovereignty in Southeast Myanmar](#)*, came across my desk in May 2025 as part of the eCIP program (pre-publication cataloging performed on behalf of the Library of Congress). The publisher's summary of this title reads "anthropologist Tomas Cole grounds the Salween Peace Park's creation in the context of Indigenous concepts of ownership and autonomy and culminates in Cole's argument that the Salween Peace Park is a form of liberation conservation, in which the demand to create a protected area is deeply wedded to the demand for self-determination." As I explored the text, I jotted down the following terms: *Salween Peace Park, Indigenous sovereignty, Land conservation, Indigenous cosmologies, Indigenous ways of knowing.* I looked up information about the park itself, considering that I might need to create an authority record for it. The Salween Peace Park was established in 2019 and is managed by Indigenous Karen communities. According to the Doh Gabar website, the park "includes 168 Kaws, 34 community forests, 8 reserved forests, and 3 wildlife sanctuaries." A Kaw is defined by the website as "a territory with its own system that integrates sustainable livelihoods, nature protection, and democratic governance."

While this book described the creation of the park, it was not *about* the park. That is, someone who was looking for information about the park itself would not find much of relevance in this text. Similarly, while the text is grounded in Indigenous ways of knowing, it is not *about* "Indigenous knowledge systems" which is a "use for" reference (also sometimes called a variant) for the LCSH term "Ethnoscience." This term was somewhat applicable yet unsatisfying, as it is effectively marginalizing with the presumption of non-Indigenous as an unspoken norm while lumping together all Indigenous peoples and communities. I found no better way, however, to represent the focus on Indigenous concepts in contrast with Western assumptions about land ownership and management.

Laura E. Daniels is Director of Cataloging and Metadata Services, Cornell University Library.

Having considered and discarded one potential term and identified some appropriate secondary but not primary subjects, I tried to focus on the sovereignty and conservation aspects of the text. "Sovereignty" is an established LCSH term. Its broader terms are "International law" and "Political science," neither of which gave me confidence that I was looking in the right general area. One related term which superficially looks potentially relevant, "Self-determination, National" is distinctly tied to nationalism, which is antithetical to the approach to land management described in this book. There is friction between the Western concept of nations and the sovereignty this text refers to, which is, if anything, in opposition to the very idea of nationalism. As the author states in the introduction, the text "treat[s] notions and practices of living together with humans and more-than-humans alike… as situated and radically alternative regimes of ownership and sovereignty." In retrospect, I should have given more consideration to the term "Autonomy," which is a narrower term to "Sovereignty."

I ended up assigning the following subject terms:

Indigenous peoples--Land tenure--Burma

Protected areas--Social aspects--Burma

Protected areas--Political aspects--Burma

Land use--Social aspects--Burma

Land use--Political aspects--Burma

Ethnoscience--Social aspects--Burma

Ethnoscience--Political aspects--Burma

Conservation of natural resources--Social aspects--Burma

Conservation of natural resources--Political aspects--Burma

Considering the nuance and complexities I encountered identifying appropriate subject terms for this work, I was curious about what an AI tool might do. I used this as an opportunity to play with the relatively new [LCSH recommendation tool](#) available as a Chrome browser extension. This extension uses the title, author, table of contents and an abstract—all of which are readily available from the eCIP system—to suggest Library of Congress subject terms and verify them against LC's linked data service ([id.loc.gov](#)). I chose this tool because of its ease of use and relative robustness (that is, its use of both the table of contents and abstracts as well as title). The tool recommended the following subject terms for the text described above:

Salween Peace Park (Myanmar)

Indigenous people--Land tenure--Myanmar

Nature conservation--Myanmar

Self-determination, National--Myanmar

Political ecology--Myanmar

Myanmar--Politics and government

I am not sure what happened with the place name—Burma, not Myanmar, is the authorized form. Of more concern, however, is the suggestion for the primary subject to be the one I initially determined was not relevant. Tables of contents and publisher summaries are often inadequate for accurate

and complete subject analysis. I suspect I would not have explored all the options I did had I used the AI suggestions as a starting point, and my subject assignment would have been too superficial.

I performed a similar experiment on several other books while creating catalog records for them and have continued to compare results from the AI tool to records in the Library of Congress catalog. I acknowledge that this is a preliminary, superficial examination using a single AI tool. I am not attempting to assess or evaluate this tool, rather using it as an exemplar to interrogate the suitability of AI for subject analysis and assignment. In addition, the AI suggestions are based on minimal information, not on the full text of the resource, and some of the shortcomings I identify would likely be ameliorated by supplying more information. I often find the title, table of contents, and publisher's summary inadequate for accurately ascertaining the range and scope of a book. In fact, titles and chapter titles can be misleading; [LC Subject Headings Manual H 180](#) acknowledges that titles may be misleading, cryptic, or more general than the focus of the work itself. Currently, though, full text cannot be made available to commercial LLMs due to copyright; this is a significant issue that would be costly to address adequately.

Based on my initial experiments, I observe that AI tends to suggest terms that are both broader and narrower than I find appropriate. In the [LC Subject Headings Manual H 180,](#) we are instructed to "assign headings that are as specific as the topics they cover," and to "follow the hierarchical reference structure built into the subject authority file to find as close a match as possible between the topic of the work and the headings that exist to express that topic in the Library of Congress subject heading system." This is the sort of grappling that we do in attempting to align written records of complex and sometimes messy human thoughts with a relatively rigid, necessarily limited collection of terms. Identifying what exactly is "as specific as the topics they cover" requires awareness of and attention to context. When considering a subject with which I am not familiar, I frequently look to see what other materials have been assigned the same term or combination of terms to ensure that my application is consistent. One could quibble with my choices above (and I am sure someone will), but that very quibbling is also something machines cannot do.

Would it save the cataloger time to start with suggestions from AI? I am not convinced that it would. Each term still needs to be verified as both authorized and appropriate. In many cases, terms that a cataloger would never assign will be suggested and investigating those would be a waste of time and effort. Similarly, some very appropriate terms will not be suggested, and either the cataloger will need to spend time identifying and evaluating those, or the quality of the description will be compromised. In the end, the burden is on the cataloger to interpret the nature of the resource, to find and identify appropriate subject terms, and to propose new terms when necessary. And we already have excellent tools (at least for those who have a subscription to ClassWeb) for searching, viewing, and browsing related subject headings. Moreover, effective understanding and searching of controlled vocabularies is a critical skill for catalogers to develop and maintain; relying on AI has the potential to make humans less efficient and less effective in our evaluation of AI's results. It seems careless to expend the tremendous resources consumed by AI to make these marginally useful recommendations.

Why is it that so many of us seem to jump straight to subject analysis and assignment when considering potential applications of AI to metadata creation? Superficially, it is an evident and time-consuming component of cataloging. But as a cataloger, I would much rather AI assisted me by identifying potential duplicate records (as OCLC is using it to do) or by transcribing descriptive metadata. Technology should assist us in doing our work better, not attempt to do our work for us. The large language models underlying AI are excellent at predicting, replicating, and generating convincing imitations of human creations. But that is not what subject cataloging demands. When

we consider potential applications of AI in metadata creation, I believe we are asking the wrong questions. If AI is going to be useful at all, it will be to give us more time to devote human energy to the tasks that require judgment and understanding of context. Subject analysis and subject term assignment are not and will never be an exact science; they are more of an art, requiring significant judgment. And judgment is exactly what AI is lacking.