

# Collection development and LLMs: A practical use study

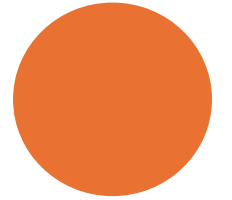
Jason A. Reuscher

Research and Instruction Librarian

Penn State Libraries

# About me:

- Research and Instruction Librarian at Penn State University Libraries
- Interested in AI's use in both scholarly research but also in practical librarianship
- Member of Penn State's Center for Socially Responsible AI (CSRAI)





## About Penn State (Libraries):

- Libraries are divided into two administrative groups under separate Associate Deans:
  - University Park (Main Campus)
  - Commonwealth Campus Libraries (everywhere else)
- By 2027, Penn State's Board of Trustees plans to close 7 Commonwealth Campuses (and their corresponding libraries)



# Penn State collections:

- Because of the planned closures, we must relocate resources throughout the Libraries:
- This includes an estimated:
  - 3,500 HathiTrust last copies
  - 40,000 Non-HathiTrust last copies
  - 140,000 multiple copies

# How do we determine what to keep?

- Shelf space is limited
  - Many libraries indicated they had little space to house closing collections
- University Libraries Annex system can only take so many items
- Last copy status important but not the only criterion



# A brief note on the relocation and/or deselection process

Oversight/advisory committee formed comprised of librarians and library staff to handle:

- Logistical issues of moving the closing library collections
- Collections analysis using catalog records and book metadata in Excel to aid subject specialists and campus librarians
- Responsible for creating a projected timeline for collections transfers and/or deselection and communication of this timeline (campuses will close in spring of 2027)



# Low-hanging fruit (media titles)

- Inventoried and tabulated in Excel worksheets
- Separate from the Book process
- Compared to books, datasets were small, manageable chunks

dvds - from YK for J... • Saved to this PC

Search

Page Layout Formulas Data Review View Automate Help Acrobat

General Alignment Number Styles

Conditional Formatting Format as Table Cell Styles

Cells Editing Sensitivity Add-ins Copilot

TX791.W64 2002 DVD

	B	C	D	E	F	
	Catalog Title	Year	Created Date	Last Activity Date	Total Charges	Item B
	Chronique d'un été (Paris 1960)	2013	4/3/2013	7/24/2013	5	000075
	The Amish : not to be modern /	2014	3/9/2016	3/22/2016	0	000078
	America and the holocaust : deceit and indifference /	2014	12/11/2014	7/26/2017	1	000077
	Rosie the riveter	2001	2/3/2009	2/6/2015	2	000065
	Hell and back again	2011	4/17/2012	6/4/2012	3	000072
	Phantom India	2007	4/17/2009	1/1/1900	0	000065
	Eyes on the prize America's civil rights years /	2010	3/8/2016	3/24/2016	0	000078
	Eyes on the prize America's civil rights years /	2010	3/21/2016	3/24/2016	0	000078
	Eyes on the prize America's civil rights years /	2010	3/21/2016	3/24/2016	0	000078
	John Lewis : get in the way /	2017	9/1/2017	9/26/2017	0	000079
	Mystery of Easter Island /	2013	2/7/2017	2/22/2017	1	000079
	The ancient Olympics athletes, games & heroes, a lecture /	1996	2/1/2010	2/26/2010	0	000067
	Ancient Chinese sports a window on Chinese history and cu	2007	2/1/2010	4/28/2010	1	000067
	LaLee's kin the legacy of cotton /	2010	4/15/2014	6/11/2014	0	000077
	The mine wars /	2016	7/29/2016	8/17/2016	0	000078
	PCN tours.	1997	10/11/2007	1/9/2008	1	000062
	PCN tours.	1998	11/12/2007	1/20/2017	5	000062
	PCN tours.	1996	10/11/2007	11/28/2007	2	000062
	PCN tours.	1997	10/11/2007	12/12/2012	4	000062

Display Settings

# Piloting large language models (LLMs) in DVD analysis



**Copilot**



- Perplexity
  - Grounded (less prone to hallucination)
- Claude
  - Not grounded, but provides more nuanced answers based on its training
- Copilot
  - PSU institutional option (paid version, possibly better results), also grounded

# Some additional factors on piloting LLMs

Privacy/Competitive  
Intelligence  
Concerns

Environmental  
Concerns

Time Constraints



## Crafting a film prompt:

**Pretend you are an expert film critic, and you are trying to decide which film titles from the attached list to add to an academic library collection.**

**Include film reception and review information for each title. Rank them by title and their reception by audiences as well as their impact on society but also sub rank them by their item legacy charges (from most charges to least charges).**

# Perplexity result? Satisfactory (hand-out 1)

Menu Pretend you are an exp... x + Create

All tools Edit Convert E-Sign Find text, tools, or help Q Share

## Ranked Film Titles with Reception, Societal Impact, and Legacy Charge Data

Title	Legacy Charges	Reception & Societal Impact
<b>A Star is Born (2006)</b>	0	Well-received remake (several versions exist, including one acclaimed 2018 film); explores fame, addiction, and tragedy.
<b>A Touch of Spice (2009)</b>	3	Celebrated Greek-Turkish film; praised for cultural depth, nostalgia, and culinary metaphors.
<b>August Wilson: The Ground on Which I Stand (2015)</b>	2	Documentary on Wilson's legacy in theater; highly praised for highlighting African American cultural contributions.
<b>Barbershop 2: Back in Business (2004)</b>	2	Received positive reviews for humor and social commentary about African American community and business.
<b>Being Mortal (2015)</b>	2	Documentary adaptation of Atul Gawande's influential work on end-of-life care; critically acclaimed for compassion and insight.
<b>Blackkklansman (2018)</b>	4	Spike Lee's Oscar-winning film; acclaimed for timely racial themes, humor, and blending fact with satire.

# Claude result? Also, satisfactory (hand-out 1)

The screenshot shows a web browser window with several tabs open. The active tab is titled "Curating films for academic library". The address bar shows the URL "https://claude.ai/chat/b35d8968-8414-4587-8f51-1fe76c76cf73". The browser's address bar and tabs are visible at the top. Below the browser window, the Claude AI chat interface is shown. The chat title is "Curating films for academic library collection". The chat content includes a request for a list of fifty essential films and a response titled "Fifty Essential Film Titles for Academic Library Collection". The response lists three films: "Bicycle Thieves", "Wild Strawberries", and "The 400 Blows". A "Reply..." input field is visible at the bottom of the chat, and a "Share" button is in the top right corner of the chat area. A footer note at the bottom of the page reads "Claude is AI and can make mistakes. Please double-check cited sources."

Inbox - Reuscher, Jason A - Outloc X | Identify individual authors that are X | Assume the role of an expert film X | Curating films for academic library X

https://claude.ai/chat/b35d8968-8414-4587-8f51-1fe76c76cf73

Research Readines... Campus: Penn Stat... Penn State - myRe... Campus Calendar |... Recipe research Dungeons and Dra... ULIBS P&T Commi... Other favorites

Curating films for academic library collection

list of fifty essential films from your collection. Let me proceed systematically with selections that represent key movements, directors, and innovations in cinema history.

### Fifty Essential Film Titles for Academic Library Collection

1. **Bicycle Thieves (Ladri di biciclette)** - Vittorio De Sica's 1948 masterpiece exemplifies Italian neorealism [Wikipedia](#), using non-professional actors and location shooting to portray post-war poverty in Rome. Ranked first in Sight & Sound's 1952 critics' poll and remains a fixture on greatest films lists [Wikipedia](#), the film's humanistic approach to social issues influenced generations of filmmakers worldwide and established new standards for cinematic realism.
2. **Wild Strawberries (Smultronstället)** - Ingmar Bergman's profound meditation on memory, regret, and mortality follows an aging professor's journey to receive an honorary degree. The film's innovative dream sequences and psychological depth demonstrate Bergman's mastery of cinematic introspection, making it essential for understanding art cinema's exploration of existential themes.
3. **The 400 Blows (Les quatre cents coups)** - Though not explicitly listed, if "Six contes moraux" represents François Truffaut's work, this would be his seminal French New Wave film about adolescent alienation. Truffaut's semi-autobiographical approach and revolutionary handheld camera work helped define the nouvelle vague movement that

Reply...

Sonnet 4.5

Claude is AI and can make mistakes. Please double-check cited sources.

# Copilot response: Satisfactory (hand out 1)

The screenshot shows a web browser window with two tabs: "Inbox - Reuscher, Jason A - Outlook" and "Pretend you are an expert film critic". The address bar shows a URL from Microsoft's cloud chat service. The browser's bookmark bar includes "Research Readines...", "Campus: Penn Stat...", "Penn State - myRe...", "Campus Calendar |...", "Recipe research", "Dungeons and Dra...", "ULIBS P&T Commi...", and "Other favorites".

The chat interface has a header with the prompt "Pretend you are an expert film critic, and you are" and a toolbar with "Auto", a share icon, a dropdown arrow, a shield icon, and a menu icon. On the left, there is a sidebar with icons for chat, search, list, share, and print.

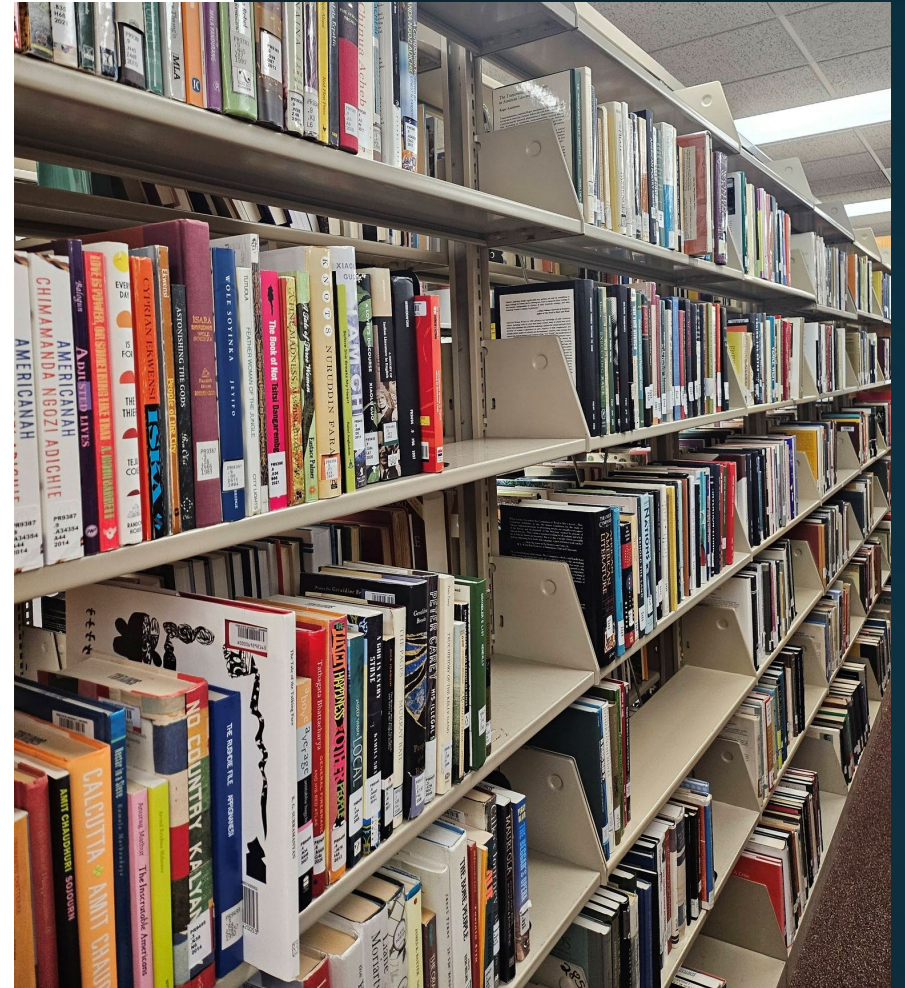
The response is structured as follows:

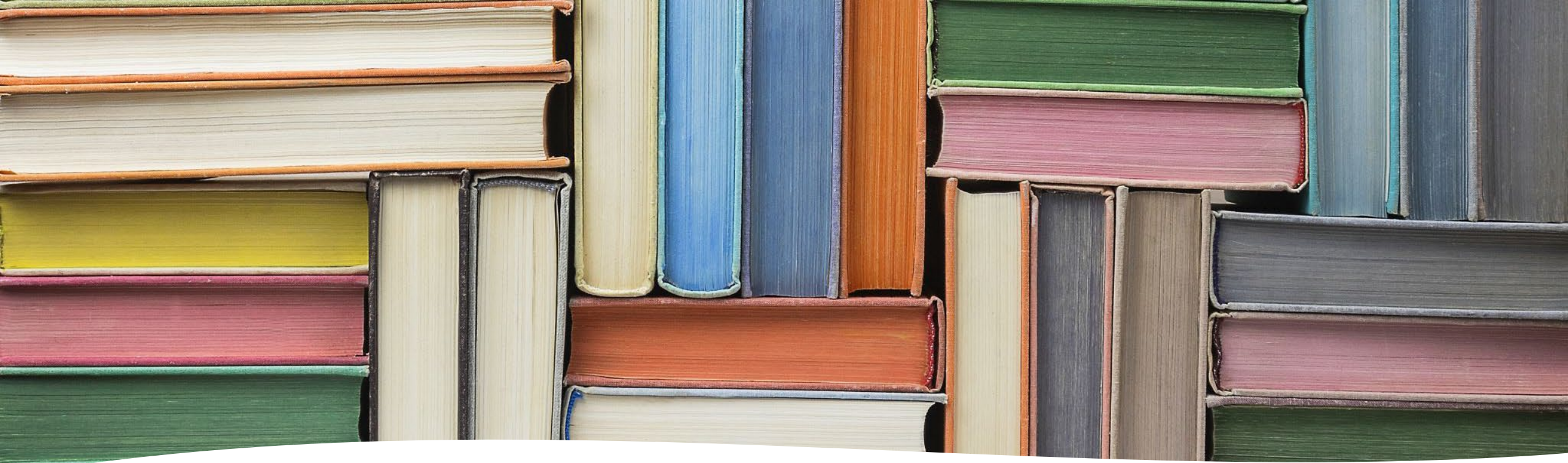
- ACCATONE (1961, Pier Paolo Pasolini)**
  - Item legacy charges: 0 (low circulation but high scholarly value)
  - Reception & Criticism**  
Highly praised by critics for its raw neorealist depiction of Rome's underclass; controversial on release due to its brutal tone and nonprofessional actors. dvds - from YK for Jas...
  - Societal Impact**  
Marked Pasolini's explosive entrance into cinema and reshaped Italian social realism, influencing global art cinema and political filmmaking. dvds - from YK for Jas...
  - Academic Justification**  
Essential for Italian cinema, Marxist cultural critique, and post-neorealism.
  - Retain
- BICYCLE THIEVES (Ladri di biciclette) (1948, Vittorio De Sica)**
  - Item legacy charges: 4

At the bottom, there is a "Message Copilot" input field with a plus icon on the left and a microphone icon on the right. A footer note states "AI-generated content may be incorrect".

# Prompting LLMs for book retention decisions

- More data to simplify for LLMs to handle
- More interest in connecting the book data with data outside of the catalog
- Can AI determine which listed authors had or have a PSU affiliation?





A new prompt is needed:

**Identify individual authors that are or were affiliated with the Pennsylvania State University (PSU) from the uploaded list of authors and their published books.**

**Once you have identified the authors, generate a numbered list of the authors names with a brief biographical summary whenever possible.**

# Results: Perplexity (free version)

Julia Spicher  
Kasdor (Yes)

Toby  
Thompson  
(Yes)

John Marsh  
(Yes)

Michael  
Bérubé (Yes)

James L. W.  
West III (Yes)

# Results: Claude (free version)

Lester P. Nichols  
(Yes)

H. or Hans Arthur  
Meyer (Yes)

Maynard Hensch  
(Maybe)

Josephine  
Kennedy Warwick  
(Unsure)

No fifth author  
listed

# Results: Copilot (paid version)

Martin E. P.  
Seligman (Univ. of  
Pennsylvania)

Nicholas Rescher  
(University of  
Pittsburgh)

John H. Harvey  
(University of  
Iowa, Texas Tech  
University)

Lynn S. Zubernis  
(West Chester  
University of  
Pennsylvania)

Ed Jacobs (West  
Virginia  
University)

# What happened?

- Perplexity provided the most recent affiliation matches (or did it?)
- Claude found historical individuals affiliated with Penn State
- Copilot found no authors with Penn State affiliation

# Are LLMs useful for collection development?

- For the DVD data, LLMs were able to look at the titles and match them to reviews regardless of whether the AI model was grounded or ungrounded. They did not rank them by check-out as I would have wished though.
- For my larger book dataset, the free LLMs seemed somewhat useful in analyzing the book data with my affiliation request. However, their selections of authors varied based on their training and grounding.
- It really depends on which model is used for the type of question answered. The best solution for this type of query is to use many models with the same prompt (for the moment).

# Concluding thoughts on LLM usage in collection development

- Think before you prompt. Iterate several drafts of a prompt to make sure you hit every point clearly and concisely before you interact with an LLM.
- If uploading data, make sure the model can handle the file size (every AI model is different, and each model handles data differently based on the user's access level).
- Consider carefully the role that you want the LLM to take, the information that you want it to consider, and the output that you wish to see.
- Expect to check the results of each interaction. LLM answers will vary across models and even within the same model when used repeatedly.

# Some further questions and thoughts on LLMs and book metadata

- Would a custom or niche LLM trained on book and publisher metadata be more useful than the existing models (either for deselection or selection purposes)? Or do grounded models suffice?
- Are LLMs merely a step along the way to another form of AI?
  - How would agentic AI compare to existing library profiles with existing book vendors?

Questions?

