Using AI/ML Techniques in Systematic Literature Reviews

Emily Bolger (she/her), PhD Candidate
Computing Education Research Lab
Department of Computational Mathematics, Science, and Engineering
Michigan State University

In collaboration with: Danny Caballero, Dawson Kinsman, Anthony Kosinski, Rachel Renbarger, Ying Wang, Taylor Boyd, Noah Finkelstein, Andrea Beach, Charles Henderson, Scott Simkins, Matt Ison

Hi! I'm Emily (she/her)!

PhD Candidate; Computing Education Research Lab
Department of Computational Mathematics, Science, and Engineering

What I Do:

Use Data Science techniques (mainly, network science & NLP)
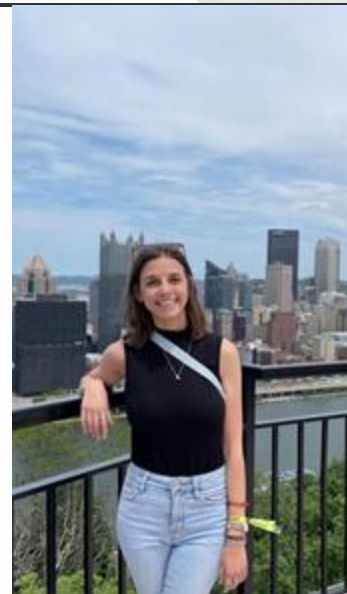to study education related data

Some other things I do:

Curriculum development for CMSE courses (ethics + AI/DS)
Graduate Women in Science

Would love to chat about:

Any AI + Education overlap, including environmental impacts
Also, books and board games!

https://egbolger.github.io/

bolgerem@msu.edu

CERL@MSU
COMPUTING EDUCATION RESEARCH LAB

What are Systematic Literature Reviews (SLRs)?
(and why do we need them?)

More than just a summary of a corpus of articles

Comprehensive, reproducible, methodological analyses
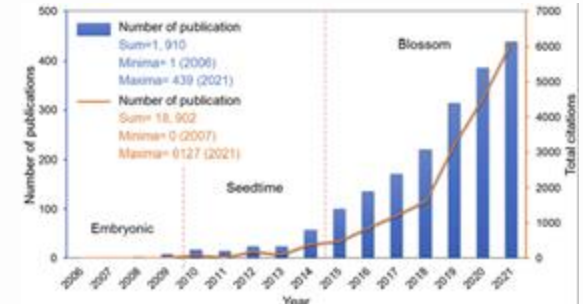
Provide comprehensive review of (sub)field

Identify research gaps in a field

Essential for early career/new to field researchers

Variety of types (e.g. meta-analyses, mapping reviews, mixed methods)

Where does ML/AI come in?

The rapid growth and expansion of published research has led to concerns that the number of articles in exclusively human-led systematic reviews may soon "threaten to exceed human cognitive processing capabilities" (p. 209, Wagner et al., 2022). (Bolger et al., 2025, under review)

Cai, Z., Zhu, J., & Tian, S. (2023). Research Progress of STEM Education Based on Visual Bibliometric Analysis. Sage Open, 13(3).

The Road So Far…

New Tools Being Developed Every Day For:

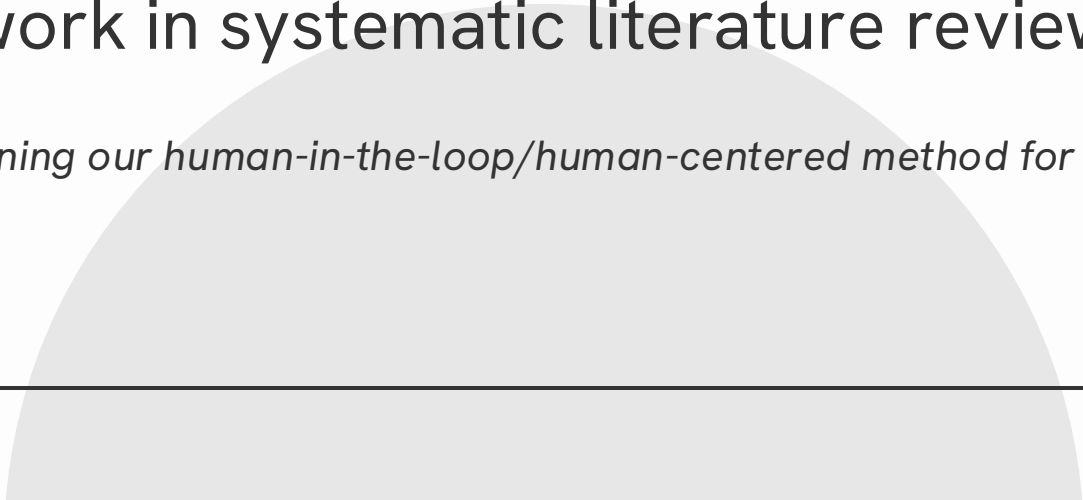|		
Article Identification	Title & Abstract Screening	Article Screening

Some references: Edwards et al., 2024; Fitkov-Norris & Kocheva, 2023; Marshall & Wallace, 2019; Scaccia & Scott, 202; Cierco Jimenez et al., 2022; O'Mara et al., 2015; Popoff et al., 2020; Van Dinter et al., 2021

The Road So Far...

New Tools Being Developed Every Day For:

Article Identification	Title & Abstract Screening	Article Screening

Some references: Edwards et al., 2024; Fitkov-Norris & Kocheva, 2023; Marshall & Wallace, 2019; Scaccia & Scott, 202; Cierco Jimenez et al., 2022; O'Mara et al., 2015;(Popoff et al., 2020; Van Dinter et al., 2021

However

"significant gaps … in the areas of data extraction, monitoring, quality assessment, and data synthesis." (Sundaram, 2023)

ML experts may lack an understanding of the review content while the content experts who conduct qualitative coding may lack an understanding of ML methodology.
(Chen et al., 2018)

How, and to what degree, can machine learning complement and support the work in systematic literature reviews?

Defining our human-in-the-loop/human-centered method for AI + SLRs

Our SLR

- Review of published literature on change initiatives in undergraduate STEM education

- Extension of 2011 paper ——---->

- Quality, Quantity, Content of papers has changed!

I. Disseminating: CURRICULUM & PEDAGOGY

Change Process: Tell/Teach individuals about new teaching conceptions and/or practices and encourage their use.

Examples: dissemination/training (SER, FDR), focused conceptual change (FDR)

II. Developing: REFLECTIVE TEACHERS

Change Process: Encourage/Support individuals to develop new teaching conceptions and/or practices.

Examples: reflective practice (FDR), curriculum development (SER), action research (FDR, SER)

III. Enacting: POLICY

Change Process: Prescribe new environmental features that Require/Encourage new teaching conceptions and/or practices.

Examples: policy change (HER), strategic planning (HER)

IV. Developing: SHARED VISION

Change Process: Empower/Support stakeholders to collectively develop new environmental features that encourage new teaching conceptions and/or practices.

Examples: institutional transformation (HER), learning organizations (HER)

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of research in science teaching, 48*(8), 952-984.

Project Research Questions

RQ1

What themes and categories emerge from the research on change strategies published between 2010 and 2023?

Human Content Coding
ML/AI Codes

RQ2

How has the use of change strategies in the published literature advanced from 1995 – 2010 to 2010 to 2023?
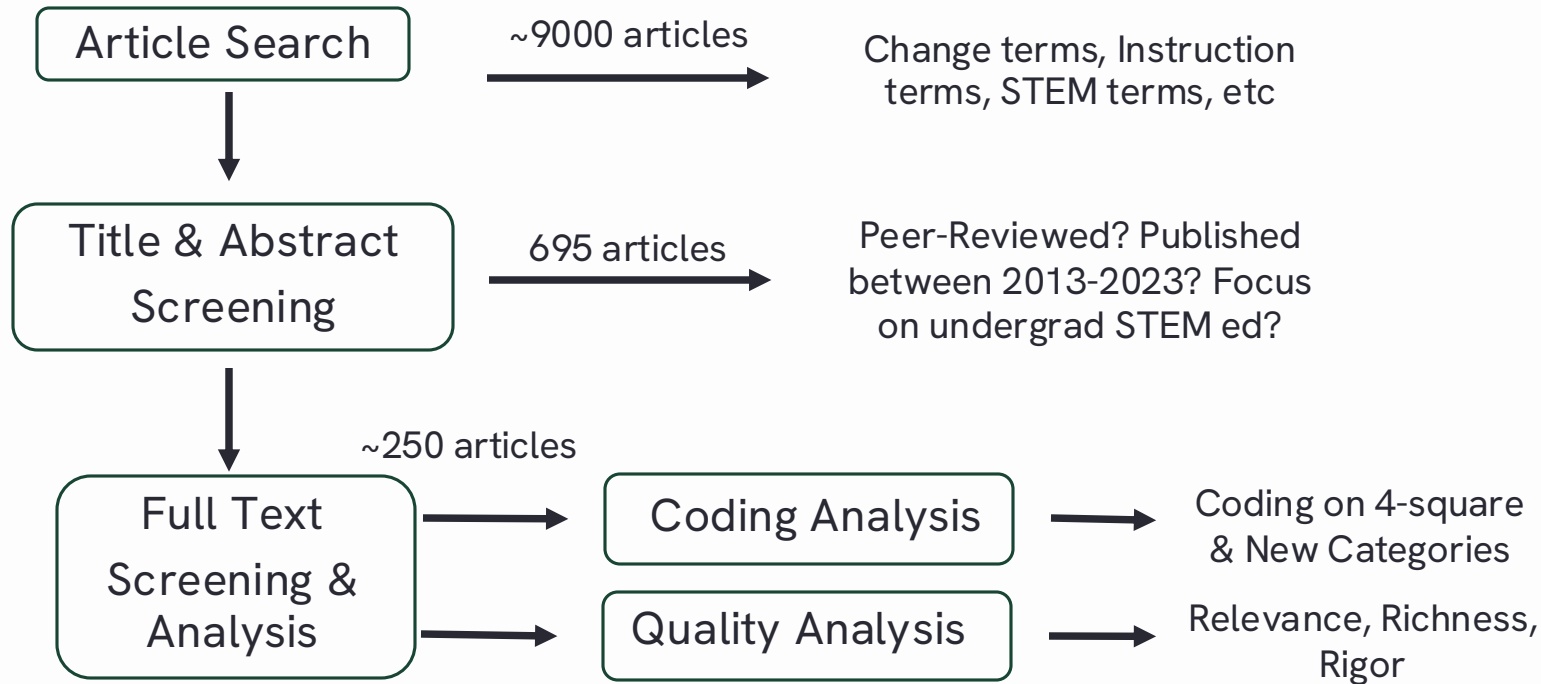
Human Quality Coding

RQ3

What are the networks/relationships between the articles, authors, and communities?

Network Science
Human Codes

What does the last decade of change literature tell us about how the field has developed and where it should go in the next decade?

Human Content & Quality Coding

Article Search

~9000 articles → Change terms, Instruction terms, STEM terms, etc

Title & Abstract Screening

695 articles → Peer-Reviewed? Published between 2013-2023? Focus on undergrad STEM ed?

~250 articles

Full Text Screening & Analysis

Coding Analysis → Coding on 4-square & New Categories
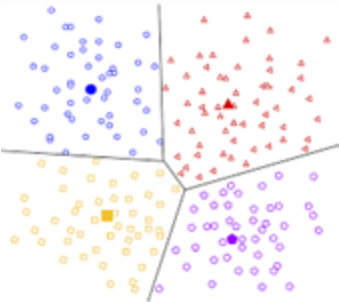
Quality Analysis → Relevance, Richness, Rigor

ML Coding - Computational Grounded Theory

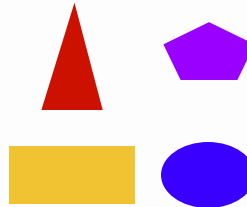Combines computational developments with resources with expert knowledge
<u>to identify themes</u>

1. Pattern Detection

Unsupervised
Clustering of Topics

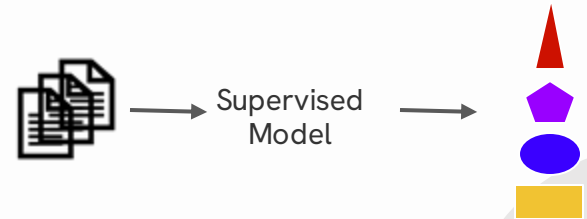2. Pattern Refinement
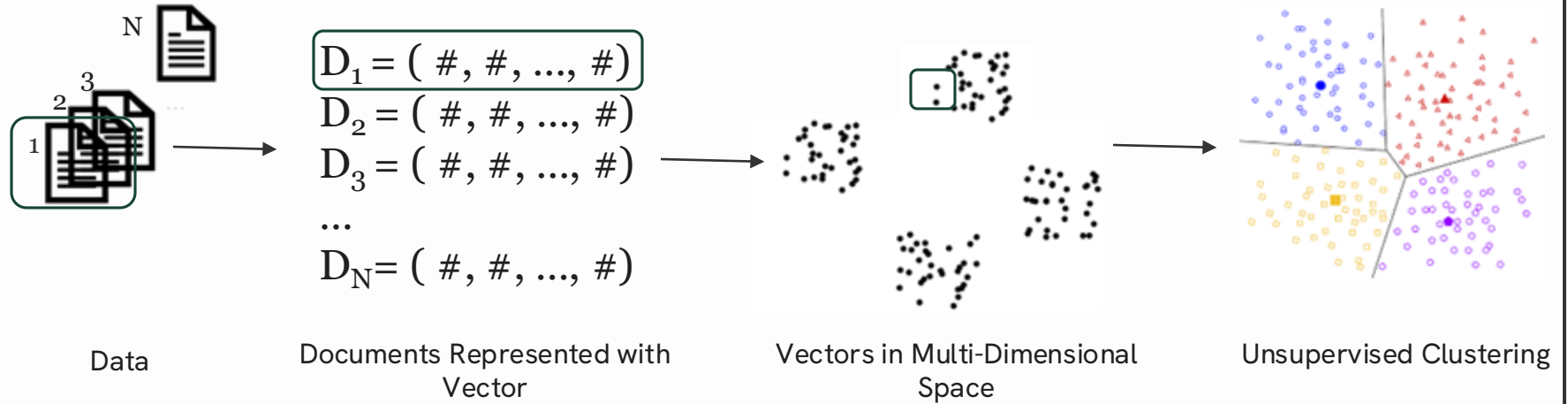
3. Pattern Confirmation

Classifying Documents to
Topics with Supervised Model

Supervised
Model

Nelson, Laura K. "Computational grounded theory: A methodological framework."
Sociological Methods & Research 49.1 (2020): 3-42.

CGT Step 1 in Steps

1. Pattern Detection
Unsupervised Clustering of Topics

2. Pattern Refinement

3. Pattern Confirmation
Classifying Documents to Topics with Supervised Model

Supervised Model

$D_1 = (\#, \#, ..., \#)$

$D_2 = (\#, \#, ..., \#)$

$D_3 = (\#, \#, ..., \#)$

...

$D_N = (\#, \#, ..., \#)$

N

3

2

1

Data

Documents Represented with Vector

Vectors in Multi-Dimensional Space

Unsupervised Clustering

Tschisgale, Paul, Peter Wulff, and Marcus Kubsch. "Integrating artificial intelligence-based methods into qualitative research in physics education research: A case for computational grounded theory." *Physical Review Physics Education Research* 19.2 (2023): 020123.
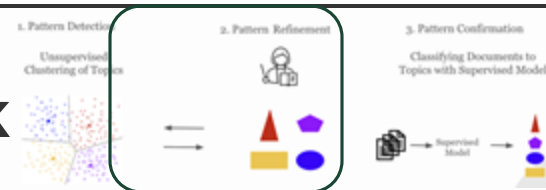
Our Human-in-the-Loop Framework
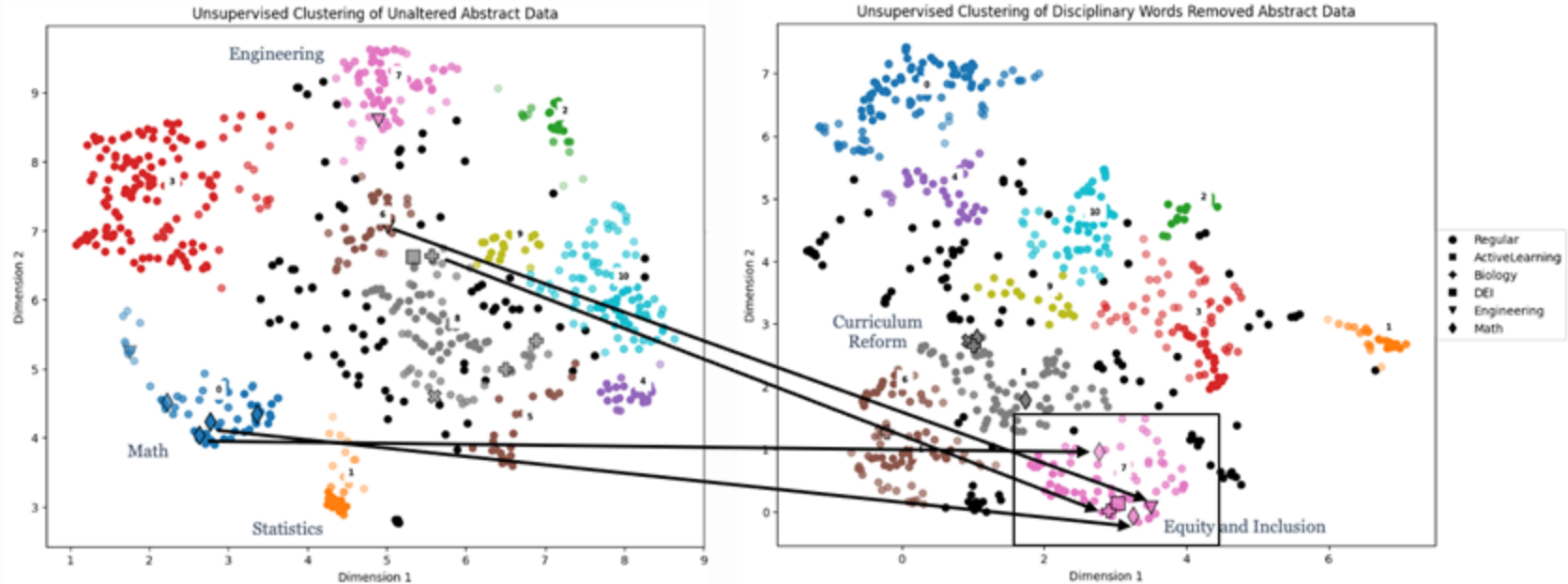
(well, a subset of it)

Humans 🤝 Machines

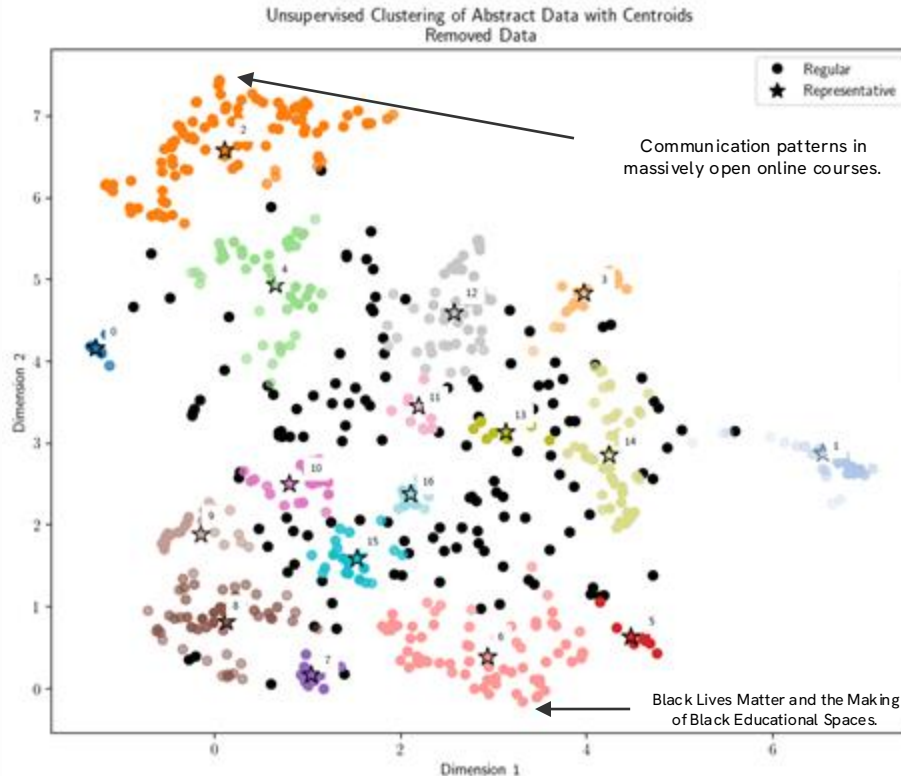| 2. Pattern refinement	2a. Creating cluster output guides	ML	Creating user-friendly walk-throughs and information guides of model results
	2b. Reviewing guides	ML + method girls	Making edits for specific team needs (e.g., time, priorities)
	2c. Reducing and reviewing clusters	Method girls	Identifying most relevant models (here, from 7 to 3) for discussion/analysis
	2d. Interviewing	ML + method girls	Consensus on most relevant models; further edits specific to team
	2e. Team reviews reduced set of clusters	Whole Team	Identifying most relevant models (from 3 to 1) for discussion/analysis
	2f. Interview in pairs	ML + Whole Team	Consensus on most relevant models and applicability to larger analysis
	2g. Synthesizing cluster labels	ML + method girls	Deciphering most relevant clusters and associated labels from interviews and Google Forms
	2h. Defining cluster labels	Whole Team	ML presented the team with identified and labeled clusters. Whole Team reaches consensus.

Bolger, E., Renbarger, R., et al.. (2025a). Integrating Machine Learning into Systematic Literature Reviews: Advancing a Human- in-the-Loop Framework. Campbell Systematic Reviews. [Under Review]

The Model is Only as Good as the Data We Give It

Human expertise is critical for model interpretation and development

CGT Final Clustering on Abstracts

Unsupervised Clustering of Abstract Data with Centroids
Removed Data

Communication patterns in
massively open online courses.

Black Lives Matter and the Making
of Black Educational Spaces.

UMAP Parameters
n_neighbors = 10
min_dist = 0.1

HDBSCAN Parameters
min_cluster_size = 10
min_samples = 5

-1	"Noise"
0	Flipped Classrooms
1	Sustainability
2	Technology Integration in Learning
3	Developing Quantitative Skills in Disciplinary Context
4	Adaptive Learning
5	Service Learning
6	Inclusive Teaching Practices
7	Institutional Change
8	Community-based Change
9	Assessing Instructor Professional Development
10	Active Learning Instructional Strategies
11	Inquiry-Based Instructional Strategies
12	Project-Based Learning
13	Instructional Strategies to Develop Critical Thinking
14	Undergraduate Research
15	Core Course Reform
16	Assessment for Course and Program Improvement

Full Text coming soon!

The Pros and Cons of CGT

Pros

- New ways to look at the data, see relationships between articles
- Summarizes information quickly
- Allow for comparison with 4-square
- Validate themes from human coding themes

Cons

- Integration <u>requires</u> human interpretation
- Time consuming and **iterative process**
- Coherence v relevance
- Expectations v capabilities of ML

The Systematic Review Pipeline

1 ──── **2** ──── **3** ──── **4**

Article Identification

- Research Rabbit
- Connected Papers
- Semantic Scholar
- any LLM

Title & Abstract Screening

- EPPI
- Rayyan
- Elicit
- Abstrackr
- LaserAI

Full Text Screening

Full Text Coding

- CGT
- Other ML tools
- Scholarly
- Zotero + ChatGPT

Do they even help????

*not even close to a comprehensive list

The AI Market is Growing Rapidly

We need a sustainable, adaptable framework for evaluating and integrating AI/ML tools into systematic literature reviews

Upcoming Work (pending NSF funding):

Adapt our existing HitL methods to include commercial AI systems

Design human-centered & AI-enabled practices to that allow researchers to deepen their understanding of chosen field

Develop shareable and adaptable framework of tools and practices for using AI in SLRs

Hi! I'm Emily (she/her)!

PhD Candidate; Computing Education Research Lab
Department of Computational Mathematics, Science, and Engineering

What I Do:

Use Data Science techniques (mainly, network science & NLP)
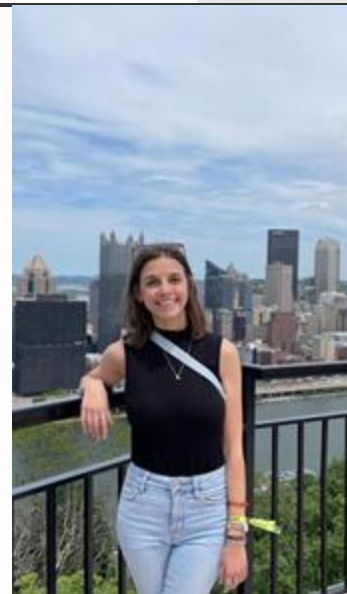to study education related data

Some other things I do:

Curriculum development for CMSE courses (ethics + AI/DS)
Graduate Women in Science

Would love to chat about:

Any AI + Education overlap, including environmental impacts
Also, books and board games!

https://egbolger.github.io/

bolgerem@msu.edu

CERL@MSU
COMPUTING EDUCATION RESEARCH LAB