AI and ML and Data! Oh My!

Supporting Teachers' and Learners' Work by Considering the Human Sides of Data Science

Joshua M. Rosenberg, Ph.D.
Department of Theory and Practice in Teacher Education
University of Tennessee, Knoxville

Presentation at the LEAD Graduate School and Research Network Retreat 2022-04-27
Slides: https://tiny.utk.edu/lead-retreat

A Story

How I was *planning* to start this talk

A story

machine learning methods to address the constraints each has. Finally, I conclude with cautiously optimistic ways for people with expertise in education, mathematics and statistics, and computers and computation to collaborate to advance data science in K-12 settings.

What should I title a talk with the above abstract?

The Role of Data Science in K-12 Teaching and Learning

That is actually a pretty good title for a talk with the above abstract, but could you create a title that's a bit more interesting or even funny?

How to Use Data Science in K-12 Teaching and Learning (Without Losing Your Mind)

How I was *planning* to start this talk

A story

That's closer to what I had in mind. Could you suggest any alternate titles that are maybe even more funny or thought-provoking?

Data Science in K-12: A Recipe for Success (Or Disaster)

The Benefits (and Pitfalls) of Using Data Science in K-12 Teaching and Learning

How I was *planning* to start this talk

A story

These are definitely on the right track, but they're not quite there. Any last suggestions for me?

The Promise and Peril of Data Science in K-12 Education

Pretty good, once more. How about one with a tad bit more humor, please? Last attempt.

Data Science in K-12: To Infinity and Beyond!

That was missing *something* I wanted to say

A story

- Before coming, I was a bit anxious!
- A recent event made me a bit more confident:
  - At a bakery, I ordered *Dinkel Pur* in German
  - Having family with me helped
  - But, I have much more to learn - I am humble

A little goes a long way

A story

- A key punchline from this talk is that *having some familiarity* with AI, ML, and data science can go a long way
- Drawing on our *other knowledge, capabilities, and values* is also needed

A brief aside

A story

- Many of you are already involved with these:
  - Katharina Fleig: *NLP-based learner assessment for feedback generation*
  - Lisa Minich: *The Devolopment of a Teacher Dashboard for the FeedBook System*
  - Plus working groups on *Artificial Intelligence in Education* (Xiaobin Chen) and *analyzing log file data* (Ines Pieronczyk)

But first

A story

- On a scale of 1-10 (please use the digits on your hands), **how confident are you in your knowledge and capabilities regarding**:
  - Mathematics?
  - Science?
  - Statistics?
  - Programming?
  - Education:
    - Educational research?
    - Educational systems?

Defining some terms

A story

- By using the term, **Artificial Intelligence (AI)**, I refer to simulating human intelligence through the use of computers
- By, **Machine Learning (ML)**, I am referring to a subset of AI focused on how computers acquire new information/knowledge
- By **Data Science**, I am describing analytic approaches that involve statistics and computers/computation

Where is this work taking place?

A story

- The **Learning Analytics** (LA) community focuses primarily but not only on data from learning management systems, especially in higher education settings (Wise, 2019)
- The **Educational Data Mining** (EDM) community focuses primarily but not only on data from educational technology tools, often through a highly technical lens (Baker and Inventado, 2016)

Scholarship often cuts across these fast-developing fields (e.g., Baker & Inventado, 2016; Baker & Siemens, 2012; Fischer et al. 2020)

Where is this work taking place?

A story

- The community exploring *data science in education* has two foci:
  1. **Educational Data Science** (EDS), focusing on using data science tools to support research (McFarland et al., 2021)
     - ". . . it is an umbrella for a fleet of new computational techniques being used to identify new forms of data, measures, descriptives, predictions, and experiments in education." (McFarland et al. (2021)
  2. **Data Science Education** (DSE), exploring data science as a context for learning (Wilkerson & Polman, 2019)

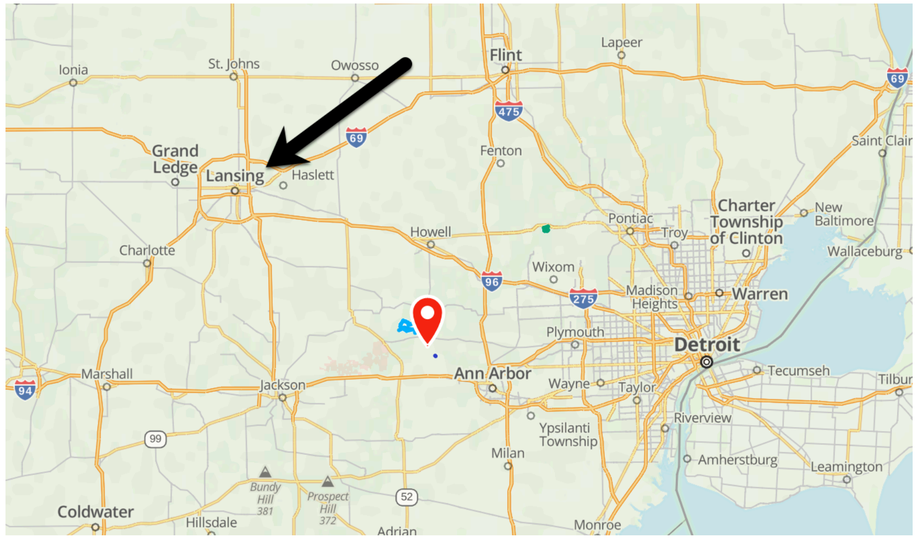
Some logistics

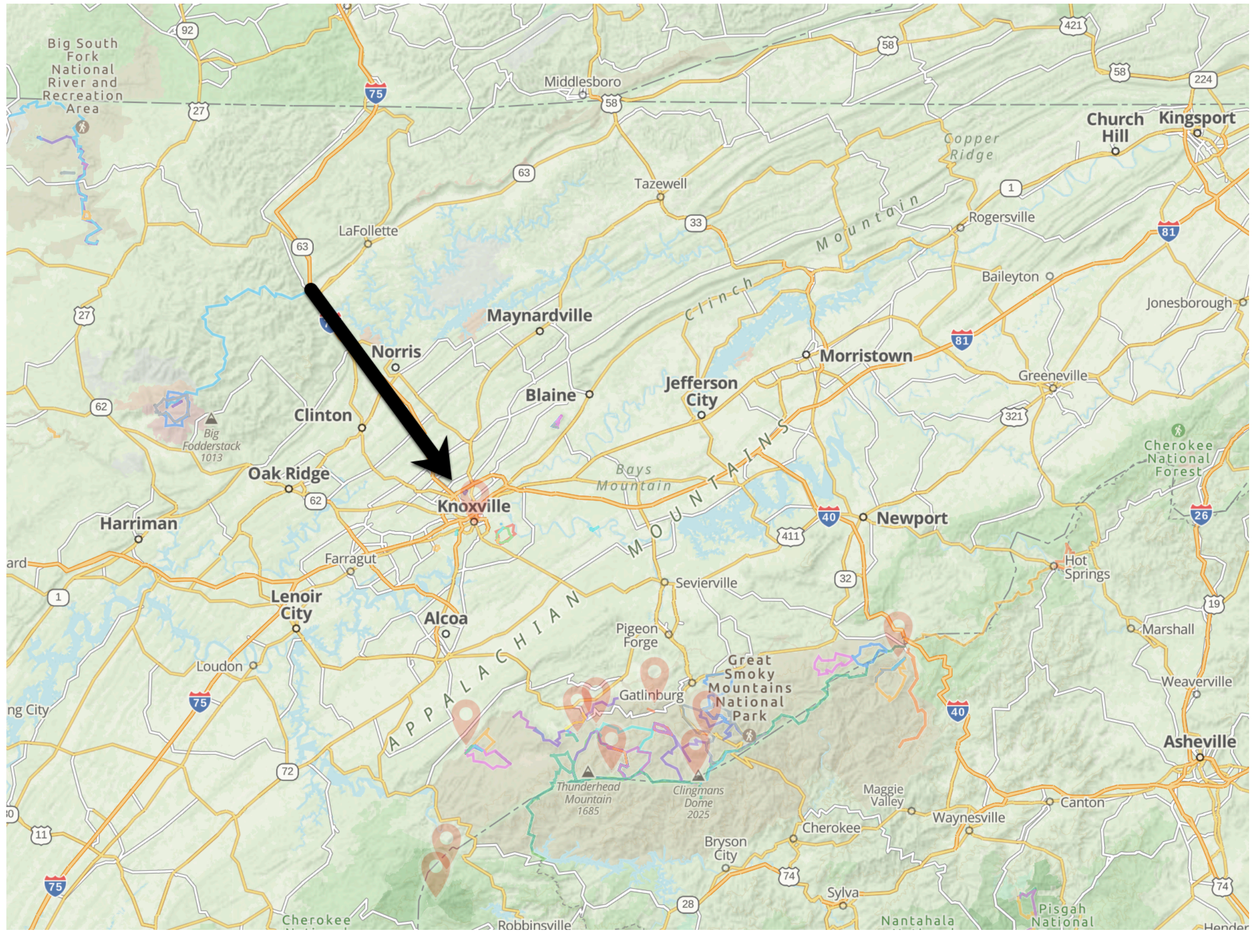
A bit about me

Some logistics

A bit about from where I come

Some logistics





A bit more about me

Some logistics

- Topically, I study teaching, learning, and educational systems
- I focus on **science education and data science in education**, but I also explore STEM disciplines
- I do so using a range of methods, but my strengths are in computation-rich approaches

THE UNIVERSITY OF TENNESSEE KNOXVILLE

And how I became interested in these topics

Some logistics

- *2011*: Saw high school students become proud of a graph they created
- *2014*: Learned R in a class on multi-level modeling
- *2015*: Was exposed to educational research on students' *data modeling*
- *2015-Present*: Engaged in cycles of learning, applying, and learning more around the topic of *data science in education*

Goals for today

Some logistics

1. **Articulate a thesis** about the nature of data science in education
2. **Describe some prior research** (mine and others') that speaks to this thesis
3. **Work toward a model** of how humans can use machines for research in a cooperative mode

Thesis

The potential of data science

Thesis

- " . . .**the digital age is having a profound impact on statistics and the nature of data analysis**, and these changes necessitate reevaluation of the training and *education practices* in statistics.
- "In particular, **computing is an increasingly important and necessary aspect of a statistician's work**, and needs to be incorporated more fully into statistics *training*."

Nolan and Temple Lang (2012), *The American Statistician*

See also: Bargagliotti et al. (2021), Hardin et al. (2015), Hardin et al. (2021), LaMar & Boaler (2021), Lee & Campbell (2020), Rosenberg et al. (2020), Wilkerson & Polman (2018)

Others have expressed some caution

Thesis

- Is data science education really that different from statistics education? (Rubin, 2019)
- How should we think about the factors influencing how students engage with data? (Lee et al., 2021)
- Who benefits from abundant data and computing power? (D'Ignazio & Klein, 2020)

An initial attempt at a thesis

Thesis

- Educational data science has a role in education, but **how we do this work matters** in terms of impact sustainability over time
- But, I want to use this retreat to rethink this work and progress toward a model of what this “how” can look like

Why does this matter?

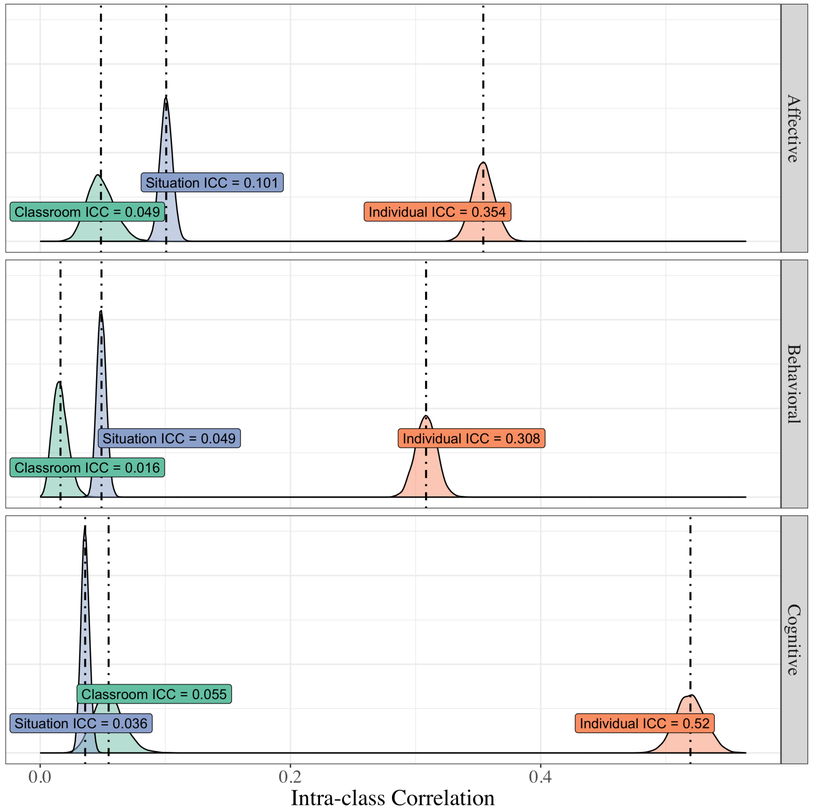
Thesis

- To step back from this thesis, **I hope working toward this model can inform the work of the LEAD Graduate School and Research Network** (and this retreat)
- I also hope to **complement the talks of other scholars**—external and members of the network

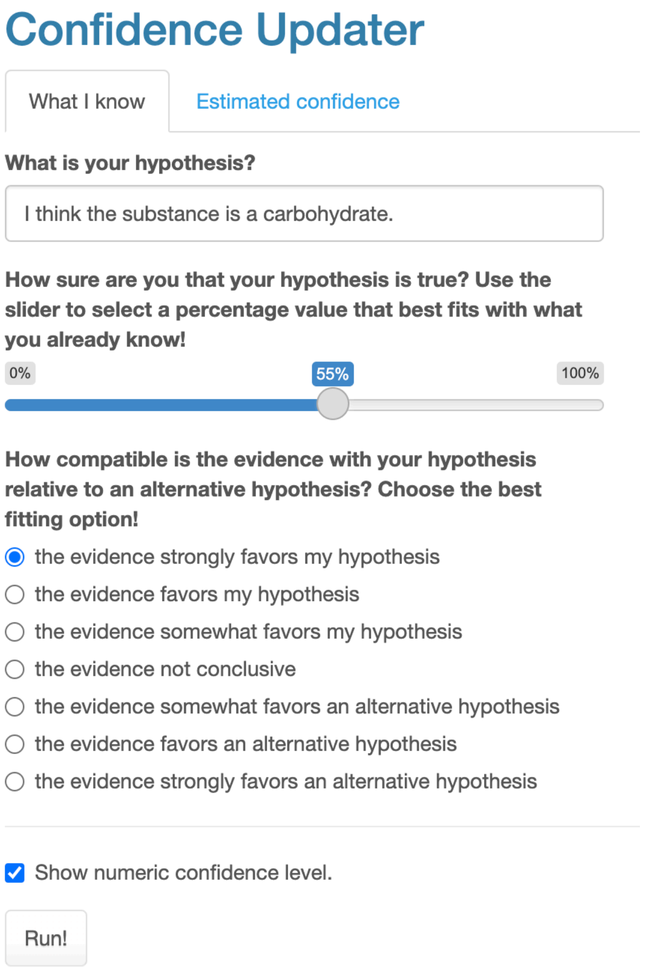
Prior Research

Supporting learners' to use Bayesian methods

Prior research



[Rosenberg et al. (2020), How does situational engagement vary between learners, situations, and classrooms?](#) (NSF-supported; see: https://github.com/picsul/short-message-survey)



[Rosenberg et al. (2022), Making Sense of Uncertainty in the Science Classroom: A Bayesian Approach](#)

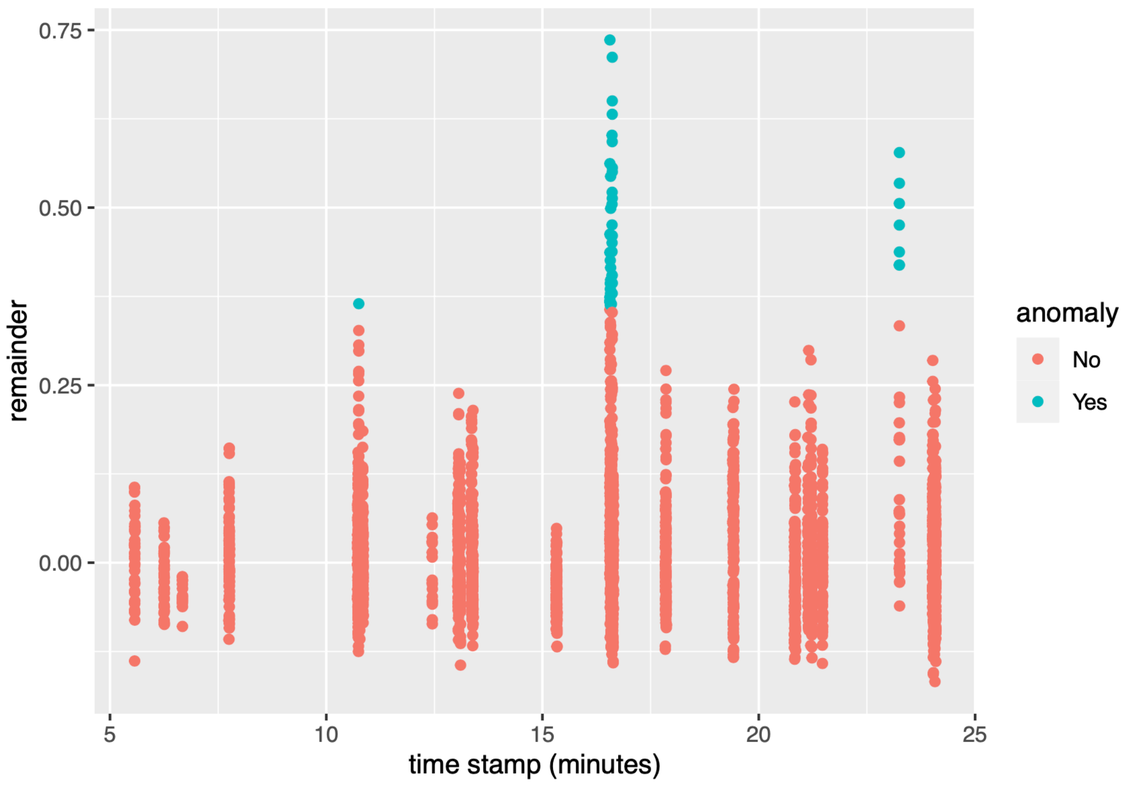
Combining qualitative and machine learning methods

Prior research

n	Classification algorithm	Percentage agreement	Quadratic Weighted Kappa
169	Naïve Bayes	.60	.47
	Support vector machine	.59	.56
	Sequential neural network	.56	.53
1885	Naïve Bayes	.66	.62
	Support vector machine	.70	.66
	Sequential neural network	.70	.65

Rosenberg & Krist (2021), Combining Machine Learning and Qualitative Methods to Elaborate Students' Ideas About the Generality of their Model-Based Explanations

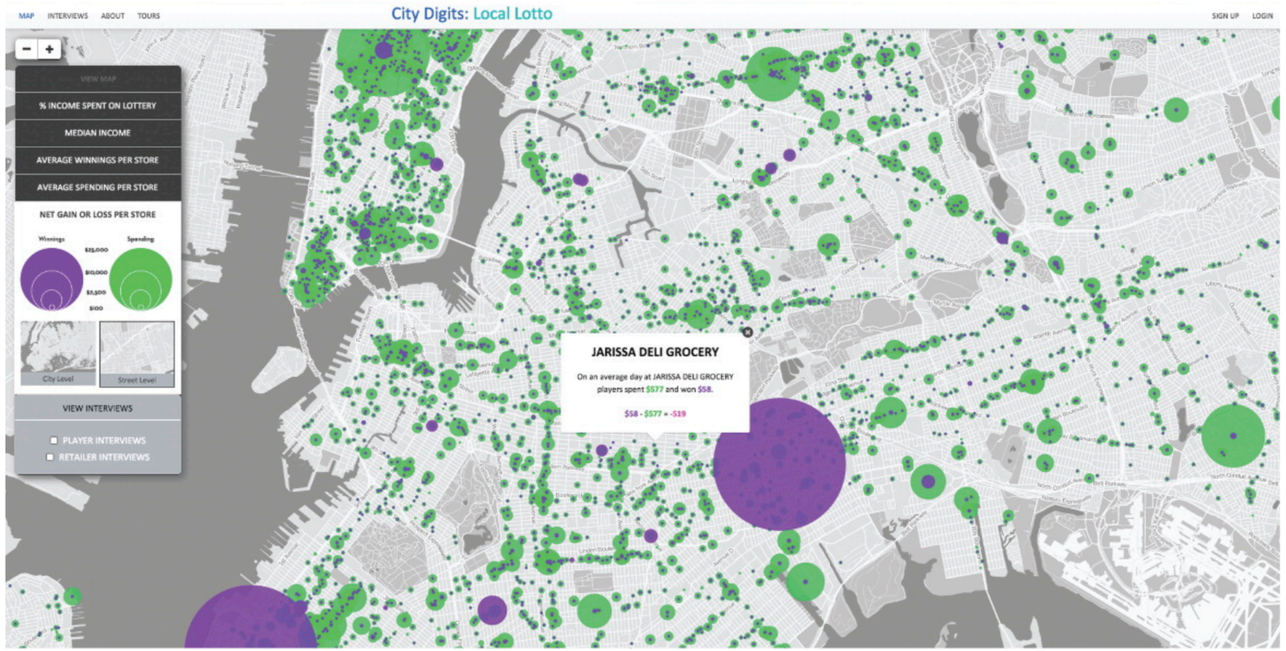
https://faast.shinyapps.io/generality-shiny/



Theory-based Computational Analysis of Classroom Audiovisual Data
(NSF-supported)

Exploring justice through data science

Prior research



[Rubel et al. (2016), Teaching Mathematics for Spatial Justice: An Investigation of the Lottery]

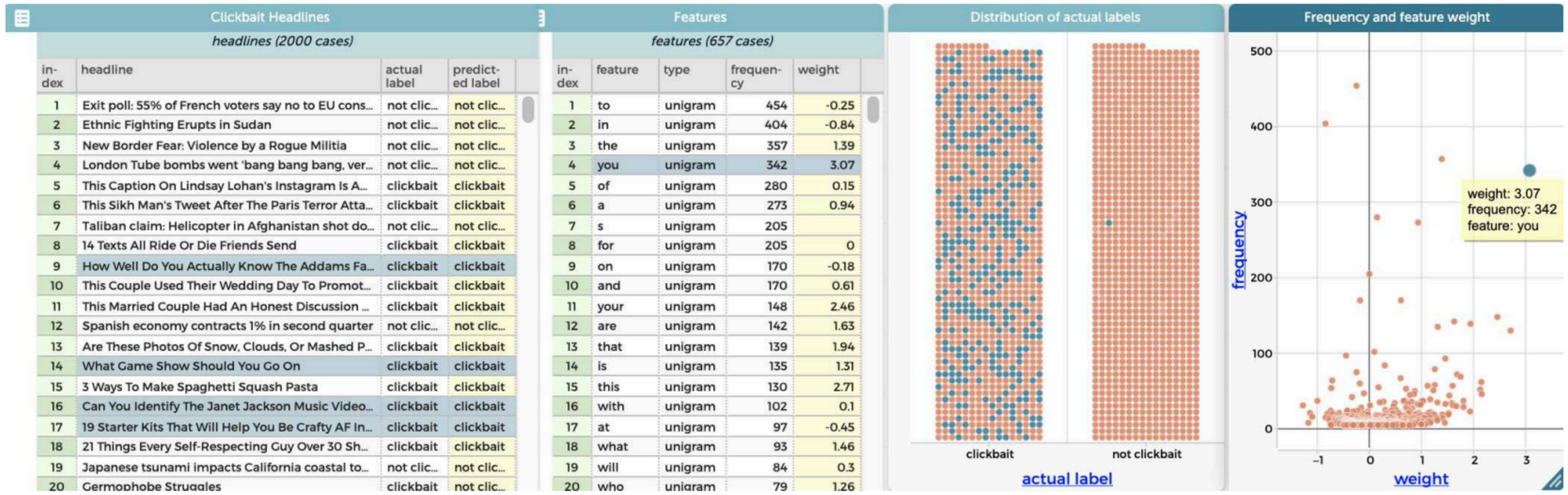
"And notice **it's ranked number one in certain areas, right**. Uh, ok and let's look—so it's orange around Inglewood and here this area between Compton and Carson."

[Philip et al. (2016), Becoming Racially Literate About Data and Data-Literate About Race: Data Visualizations in the Classroom as a Site of Racial-Ideological Micro-Contestations]

ML and AI Education

Prior research

Building on research on how students make informal predictions ([Konold, 1989, Informal Conceptions of Probability](#))



Clickbait Headlines

headlines (2000 cases)

index	headline	actual label	predicted label
1	Exit poll: 55% of French voters say no to EU cons...	not clic...	not clic...
2	Ethnic Fighting Erupts in Sudan	not clic...	not clic...
3	New Border Fear: Violence by a Rogue Militia	not clic...	not clic...
4	London Tube bombs went 'bang bang bang, ver...	not clic...	not clic...
5	This Caption On Lindsay Lohan's Instagram Is A...	clickbait	clickbait
6	This Sikh Man's Tweet After The Paris Terror Atta...	clickbait	clickbait
7	Taliban claim: Helicopter in Afghanistan shot do...	not clic...	not clic...
8	14 Texts All Ride Or Die Friends Send	clickbait	clickbait
9	How Well Do You Actually Know The Addams Fa...	clickbait	clickbait
10	This Couple Used Their Wedding Day To Promot...	clickbait	clickbait
11	This Married Couple Had An Honest Discussion ...	clickbait	clickbait
12	Spanish economy contracts 1% in second quarter	not clic...	not clic...
13	Are These Photos Of Snow, Clouds, Or Mashed P...	clickbait	clickbait
14	What Game Show Should You Go On	clickbait	clickbait
15	3 Ways To Make Spaghetti Squash Pasta	clickbait	clickbait
16	Can You Identify The Janet Jackson Music Video...	clickbait	clickbait
17	19 Starter Kits That Will Help You Be Crafty AF In...	clickbait	clickbait
18	21 Things Every Self-Respecting Guy Over 30 Sh...	clickbait	clickbait
19	Japanese tsunami impacts California coastal to...	not clic...	not clic...
20	Germophobe Struggles	clickbait	not clic...

Features

features (657 cases)

index	feature	type	frequency	weight
1	to	unigram	454	-0.25
2	in	unigram	404	-0.84
3	the	unigram	357	1.39
4	you	unigram	342	3.07
5	of	unigram	280	0.15
6	a	unigram	273	0.94
7	s	unigram	205	
8	for	unigram	205	0
9	on	unigram	170	-0.18
10	and	unigram	170	0.61
11	your	unigram	148	2.46
12	are	unigram	142	1.63
13	that	unigram	139	1.94
14	is	unigram	135	1.31
15	this	unigram	130	2.71
16	with	unigram	102	0.1
17	at	unigram	97	-0.45
18	what	unigram	93	1.46
19	will	unigram	84	0.3
20	who	unigram	79	1.26

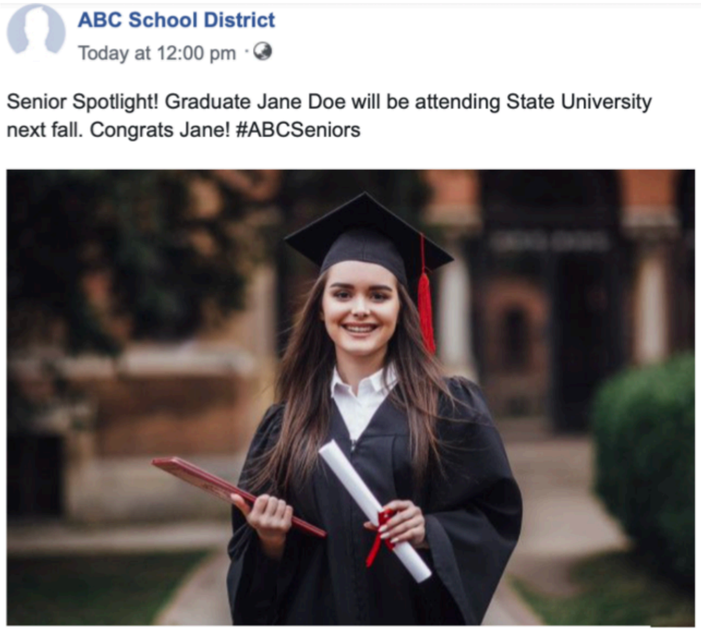
Tatar et al. (2021), Modeling Unstructured Data: Teachers as Learners and Designers of Technology-enhanced Artificial Intelligence Curriculum

The affordances of new forms of media—and drawbacks

Prior research

Most participants in a Twitter-based network were involved in substantive conversations (Rosenberg et al., 2020, Idle chatter or compelling conversation?)

Posts on Twitter about the NGSS were very positive—much mores than posts about the CCSS (Rosenberg et al., 2020, The Next Generation Science Standards on Twitter)



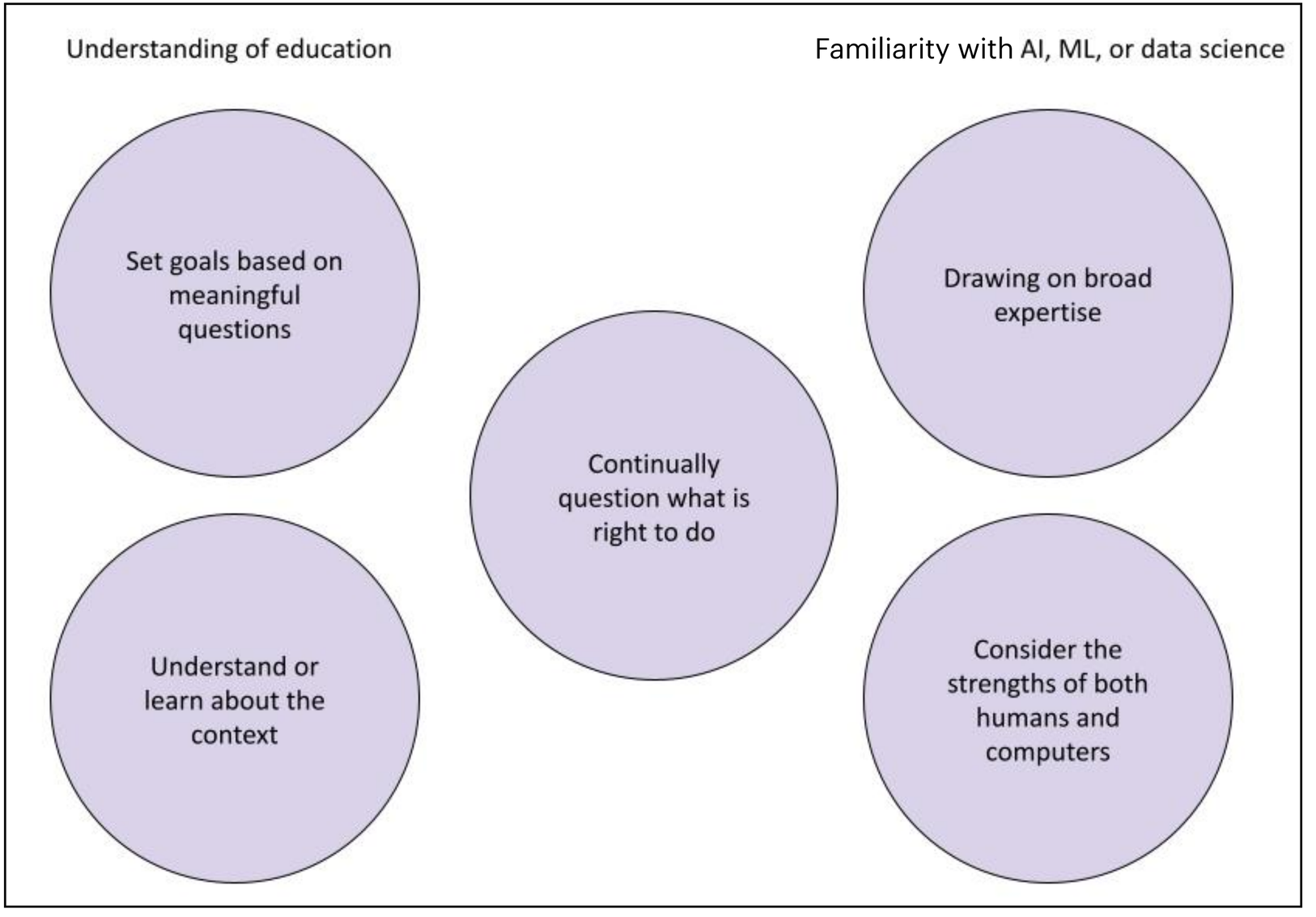
Rosenberg et al. (2022), Schools' Posts on Facebook Systematically Risk Students' Privacy

Toward a Model

Back to the thesis

Toward a model

- So, **how should research related to data science in education progress**?
- Let's play the "modeling game" with *a focus on how* (cf. Schwarz et al., 2009
  - How did the work with which I was involved work well?
  - How did others' work work well?


Understanding of education
Familiarity with AI, ML, or data science
Set goals based on meaningful questions
Drawing on broad expertise
Continually question what is right to do
Understand or learn about the context
Consider the strengths of both humans and computers

A cooperative view of working with machines

Toward a model

- There is room for humans and computers in our work, but we need to know about both to be able to answer some hard questions (Kubsch et al., 2022)
  - What are the tasks?
  - Who or what (human or computer) is doing which tasks?
  - **Who or what *should* be doing which tasks?**

AIs for humans (teachers and learners)

Toward a model

- By inquiring into the epistemic functions humans and computers serve, **we can draw attention to and not lose sight of the human side of this work**
- These may matter even more as the capabilities of AI develop (Russell, 2022)
- Key questions related to AI:
  - Do we want use AI to control?
  - Or find out what *teachers* and *learners* want?

Why familiarity can matter

Toward a model

- We can ignore, embrace, or **work out for which tasks under what conditions** we should use AI, ML, and data science (Kissinger et al., 2022)
- If we can become more familiar, we can impact their use in varied, positive ways

Please share the questions or remarks you have!

Conclusion

- **Thank you** very much to:
  - Alisa Schmid, Fitore Morina, and the the LEAD Scientific Coordination Team
  - LEAD members (special thanks to Christian Fischer)
  - LEAD Post-docs, Interns, and Students
  - Those hosting us here at Untermarchtal
  - And, the baker of the Dinkel Pur
- **Contact** information:
  - Joshua Rosenberg, Ph.D.
  - https://joshuamrosenberg.com
  - jmrosenberg@utk.edu
- **Slides**: https://tiny.utk.edu/lead-retreat

I wish to **acknowledge collaborators who shaped all of the ideas I shared today**, including those closely involved with the research I presented, including:

Conrad Borchers (UO), Patrick Beymer (UC), Macy Burchield (UTK), Mine Dogucu (UCI), Elizaeth Dyer (UTK), Christian Fischer (UT), Christina Krist (UIUC), Marcus Kubsch (IPN), Alex Lishinski (UTK), T.J. McKenna (BU), Vicky Phun (MSU), Joshua Reid (MTSU) Jennifer Schmidt (MSU), Sondra Stegenga (UU), Tayla Thomas (UTK), and E.J. Wagenmakers (UA)