SMU | LIBRARIES

Develop Your Own Digital Humanities Research Institute

SMU LIBRARIES

Introductions & Overview

Presenters

- Rafia Mirza is currently a Humanities Research Librarian at SMU. She is the coordinator of the Digital Humanities Research Institute @ SMU.
- Dr. Eric Godat is a data scientist and team lead for the Office of Information Technology's Research and Data Science Support Services.. He and his team provide consultative services and workshops to researchers on campus
- Jonathan McMichael is SMU Libraries' Undergraduate Success Librarian. He used his background in education and curricular design to adapt the pedagogical approach for DHRI @ SMU.
- Joanna RusselBliss edited and reviewed the DHRI modules on GitHub in the weeks leading up to the conference and facilitated online discussions throughout the virtual conference. She did this while working as a Research and User Experience Intern at Southern Methodist University.

Outline: DHRI@SMU

- DHRI @SMU is a research institute based on a foundational approach to teaching DH skills so that students can continue to learn and practice as a community.

- This presentation demonstrates the process of adapting DHRI to run locally, specifically onboarding faculty, students and librarians to engage in DH.

Background

- History of DHRI at SMU
  - https://github.com/SouthernMethodistUniversity/previous
    - https://bit.ly/tladhri

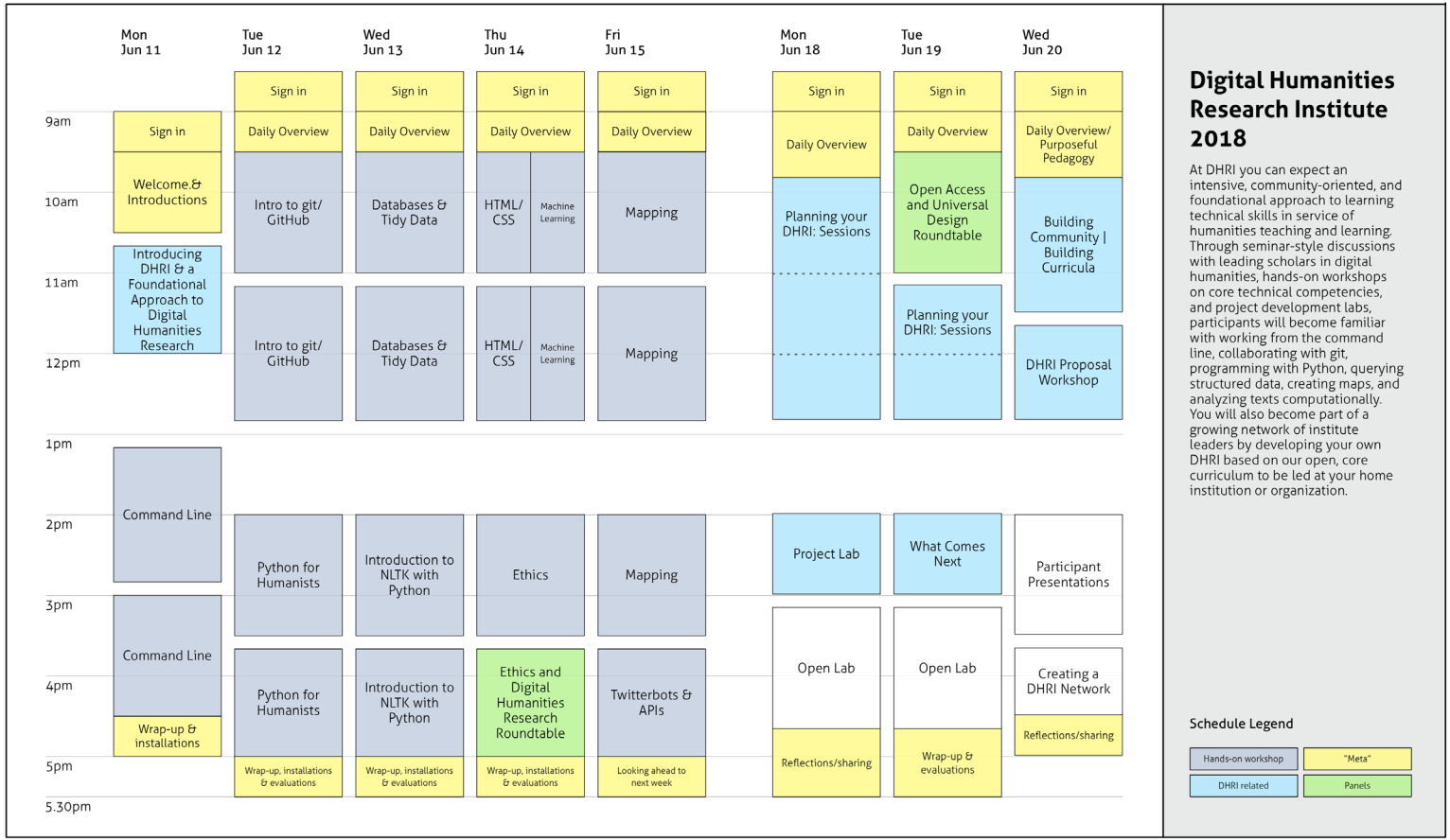
Content

DIGITAL HUMANITIES RESEARCH INSTITUTE AT SMU

Workshops build on each other such that successive workshops use skills developed in earlier ones.

To see the schedule for this year, click here.

Intro	details	repo
Data in the Humanities	details	repo
Coding	details	repo
Python	details	repo
Tools	details	repo
Access	details	repo
Project Planning	details	repo



2018: DHRI @ CUNY

SMU LIBRARIES

Digital Humanities Research Institute Schedule

Sponsored by SMU Libraries

http://www.smu.edu/dhri

	8/19/2019	8/20/2019	8/21/2019	8/22/2019
TIME	**FLE 323**	**FLE 3234**	**FLE 3232**	**FLE 3233**
9:00-9:30	Coffee/ set up	Coffee/ set up	Coffee/ set up	Coffee/ set up
9:30-10:00	Welcome (Library Dean & Academic Technology Executive Director)	daily overview	daily overview	daily overview
10:00-10:30	daily overview	git/github	data in the humanites	project planning
10:45-11:15	DH & DHRI sessions			
11:15-11:45				project lab
11:45-1:00	Lunch (Umphrey Lee)	Lunch	Lunch	Lunch
1:10-1:45	DH & DHRI sessions	r		
2:00-2:15	Command line			
2:15-2:30			geospatial data	
2:30-3:00				
3:00-3:30				
3:30-3:45	Git/gihub			
3:45-4:15				
4:15-4:45				
4:45-5:00	Wrap up/ installations	Wrap up/ installations	Wrap up/ installations	Wrap up/ installations

Digital Humanities Research Institute Schedule	Sponsored by SMU Libraries in collaboration with OIT. All sessions online through Zoom			
http://www.smu.edu/dhri				
Day/Times (all times CST)	**Monday 8/10**	**Tuesday 8/11**	Wednesday 8/12	Thursday 8/13
9:00 AM	check-in	check-in	check-in	check-in
9:15 AM	Welcome	**Python**	**Data Roundtable**	**Project Planning**
9:30 AM	Intro: Zoom & Slack			
9:45 AM	**DH**			
10:00 AM			**Data**	
10:15 AM				
10:30 AM				
10:45 AM	break	break	break	break
11:00 AM	**DH**	**Python**	**Tools**	**Project Lab**
11:15 AM				
11:30 AM				
11:45 AM				
12:00 PM	lunch	lunch	lunch	lunch
12:15 PM				
12:30 PM				
12:45 PM				
1:00 PM	**Coding**	**Python**	**Access**	**Project lab**
1:15 PM				break into small groups
1:30 PM				
1:45 PM	**Data**			
2:00 PM				
2:15 PM				
2:30 PM				
2:45 PM	break	break	break	break
3:00 PM	Installing Docker	Office Hours	Office Hours	Office Hours
3:15 PM	Office Hours			
3:30 PM				
3:45 PM				
4:00 PM				
4:15 PM				
4:30 PM				
4:45 PM				
5:00 PM	Wrap -up	Wrap -up	Wrap -up	Wrap -up
End of day	Daily exit surveys	Daily exit surveys	Daily exit surveys	Daily exit surveys
	Asynchronous Communication: will use Slack			

2020: DHRI @ SMU

SMU | LIBRARIES

DHRI@SMU August 2021

Mon	8/3 Tues	Wed	8/5 Thur	Fri
Asynchronous readings: Intro to DH & humanities data	Synchronous discussion 9:00-12:00 Intro to DH Intro to Data Afternoon Office/Lab Hours:1:00-3:00 pm DH & Data	Asynchronous readings: coding, Python	Synchronous discussion 9:00-12:00 Intro to coding Afternoon Office/Lab Hours: 1:00-3:00 pm Python	Asynchronous readings: Tools
Mon Asynchronous readings: Scoping projects, Tools *All times are Dallas Times CDT*	8/10 Tues Synchronous discussion 9:00-12:00 Intro to tools Afternoon Office/Lab Hours:1:00-3:00 pm python & tools	Wed Asynchronous readings: PMP, Access, licensing, copyright	8/12 Thur 9:00-12:00 Synchronous discussion Intro to access & PMP Afternoon Office/Lab Hours: 1:00-3:00 pm PMP, licensing, copyright etc.	Fri Asynchronous chat: any topic

DHRI at SMU (August 2022)

Sponsored by: SMU Libraries

	Monday	Tuesday	Wednesday	Thursday	Friday
Time	8-Aug	9-Aug	10-Aug	11-Aug	12-Aug
	In person: FLE 323	Online	In person: FLE 323	Online	In person: FLE 323
9:00-10:30	**Introductions & Welcome**	*Introduction to Tools & Coding*	**Introduction to Python**	*Introduction to Access & Project Planning*	**One Day Project**
10:30-10:45	*Break*		*Break*		*Break*
10:45-11:45	**What is Digital Humanities?**				
11:45-1:15	*Lunch*		*Lunch*		*Lunch*
1:15-2:15	**Data in the Digital Humanities**		**Hands on: Coding**		**Project Discussion**
2:15-2:30	*Break*		*Break*		*Break*
2:30-3:30					
3:30-3:45	*Break*		*Break*		*Break*
3:45-5:00					

Collaboration

Sponsors

Thanks to the generous support of SMU Libraries and SMU OIT, this program is free to all participants.

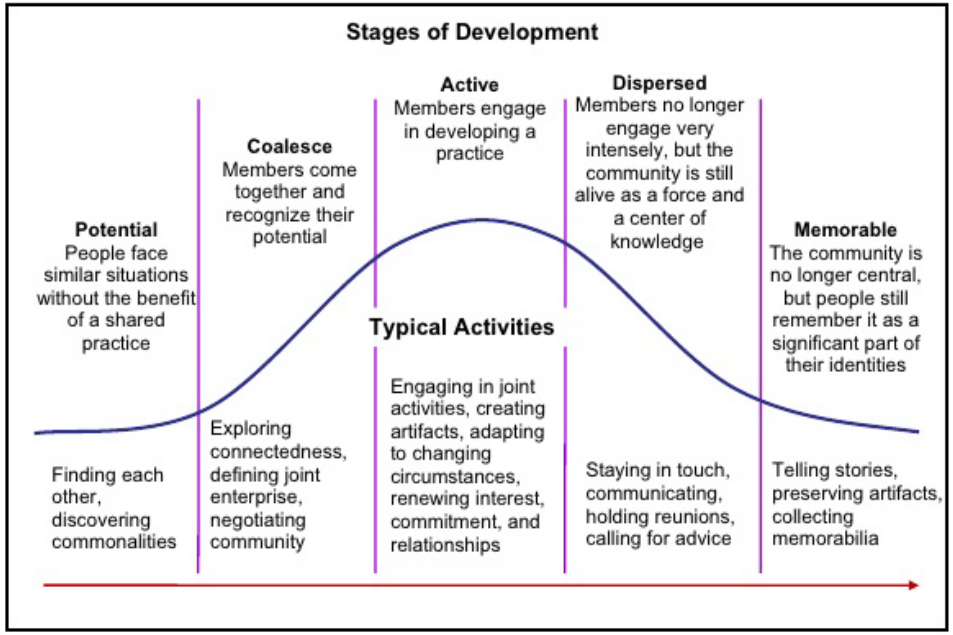
SMU LIBRARIES

Pedagogy

Pedagogy

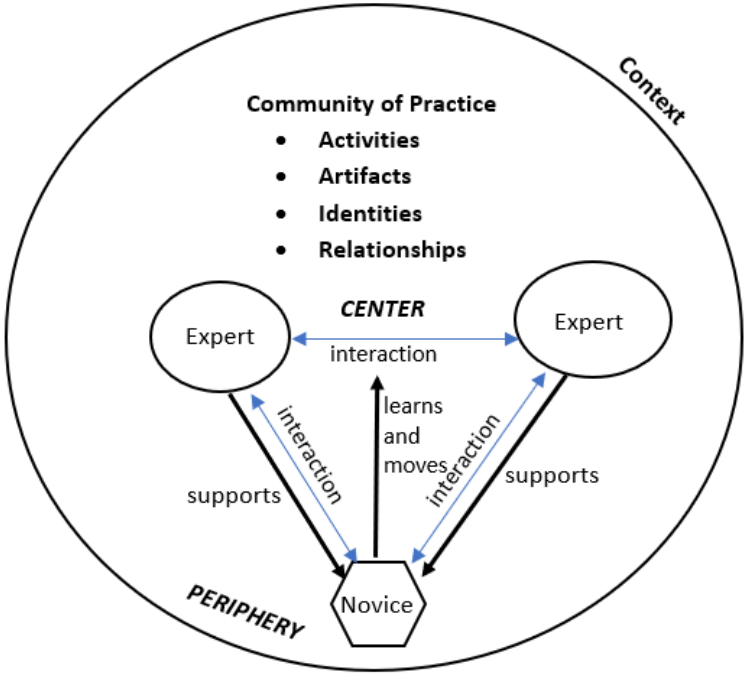
- "Although an instrumental approach satisfies a researcher's most immediate need, a foundational approach takes into consideration the long-term impact that learning core skills will have for the future professional and research needs of the scholar."
  - DHRI: Notes toward our pedagogical approach by Lisa Rhody

Community of Practice



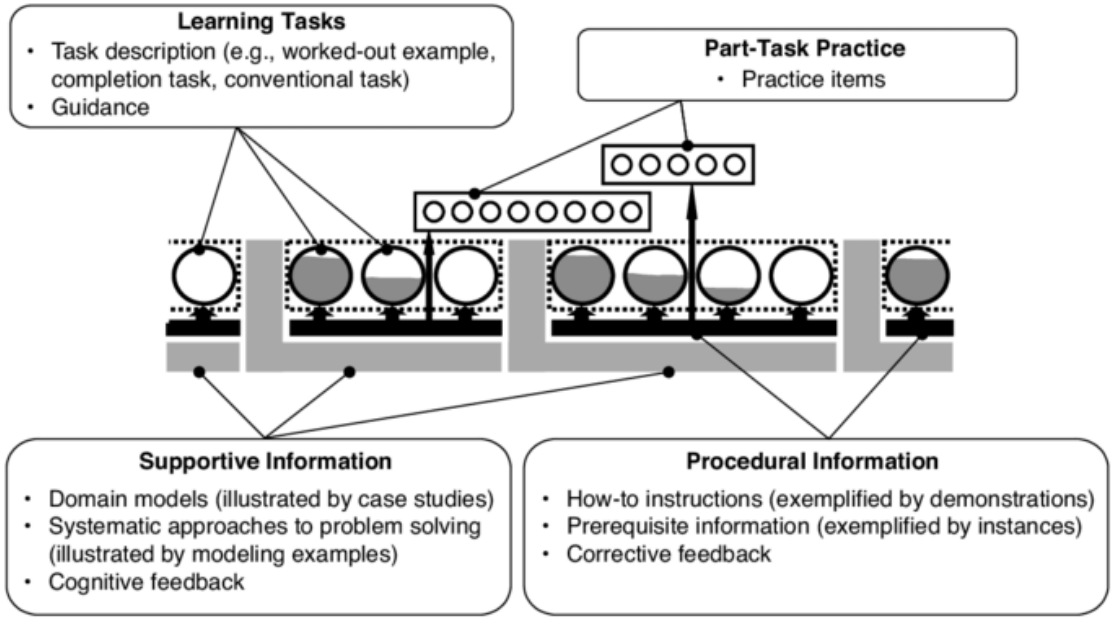
Etienne Wenger, Communities of practice learning as a social system, Systems Thinker, June 1998.

Situated Learning Model



Herrera, Sandra P.Mina. "Situated Learning Theory." *Theoretical Models for Teaching and Research*. Edited by Egbert & Roe, 2020. https://opentext.wsu.edu/theoreticalmodelsforteachingandresearch/chapter/situated-learning-theory/

4 Component Instructional Design



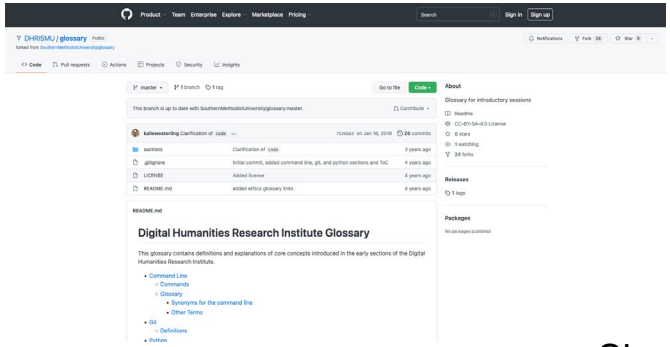
Van Merrienboer, Jeroen J. G. & Kester, Liesbeth. (2008). WholeTask Models in Education. Handbook of Research on Educational Communications and Technology.

Typical Learning Tasks

- Readings/Lecture
- Guided asynchronous discussions
- Worked examples
- Individual or group synchronous activities
- Learner presentation/discussion
- Future individual learning opportunities

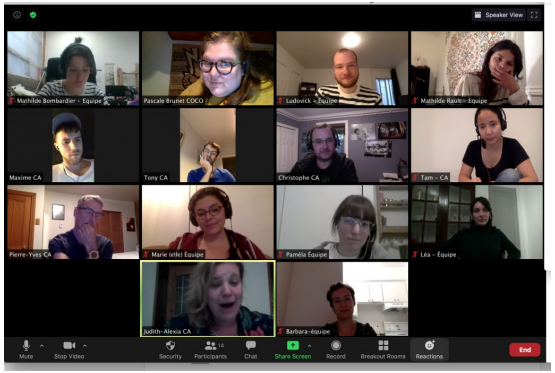
Building a Learning Ecosystem

Github Repository



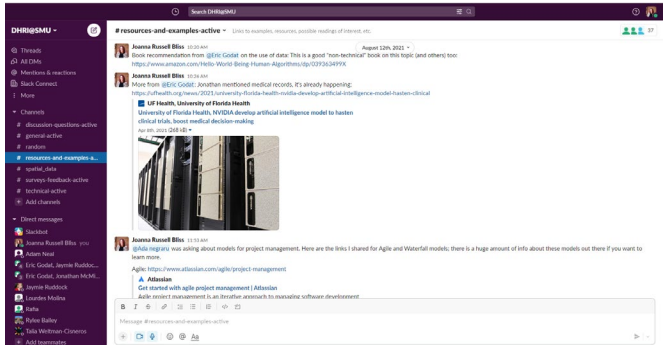
Learning Text/ Lecture Slides

Zoom or Classroom session



Archive/Back Channel

Slack (or other communication platform)



Asynchronous Learning Activities and ongoing discussion

Discussion, Coordination and Feedback

SMU LIBRARIES

Synchronous content

How did they do *this* specific project?

- ***We will be discussing this page in our synchronous meeting. You do not need to engage with this page before the synchronous session.***
- What does a project site tell you about *how* it was done?
- Activity handout

From Github Repository

Asynchronous content

Summary: What to do before this synchronous session

✅ Read the following before this session.

- We'd like each of you to go to The Pudding and read one of their data journalism pieces.
- In their archive, you can choose by topic or type of chart, based on your interests.
- Once you have read an article, put the link in Slack, as well as some of your reflections. (One of us will post an example response in Slack that you can use as a template.)

✅ Discussion questions

- *On Slack*
- Post your article and reflections, and also using the *reply in thread* option on Slack, engage with other participants posts.

Their data

- Where did their data come from?
- Is the data public?
- How big was the data? Did they use all of it?

Their analysis

- What is their thesis statement?
- What was their hypothesis?
- What choices did they make as the analysis developed?
- Were there caveats or ramifications to their choices?
- What were their conclusions?

From Github Repository

Search DHRI@SMU

DHRI@SMU

- Threads
- All DMs
- Mentions & reactions
- Slack Connect
- More

Channels

- # discussion-questions-active
- # general-active
- # random
- # resources-and-examples-a...
- # spatial_data
- # surveys-feedback-active
- # technical-active
- Add channels

Direct messages

- Slackbot
- Joanna Russell Bliss you
- Adam Neal
- Eric Godat, Jaymie Ruddoc...
- Eric Godat, Jonathan McMi...
- Jaymie Ruddock
- Lourdes Molina
- Rafia
- Rylee Bailey
- Talia Weltman-Cisneros
- Add teammates

resources-and-examples-active

Links to examples, resources, possible readings of interest, etc.

37

August 12th, 2021

Joanna Russell Bliss 10:20 AM
Book recommendation from @Eric Godat on the use of data: This is a good "non-technical" book on this topic (and others) too:
https://www.amazon.com/Hello-World-Being-Human-Algorithms/dp/039363499X

Joanna Russell Bliss 10:26 AM
More from @Eric Godat: Jonathan mentioned medical records, it's already happening:
https://ufhealth.org/news/2021/university-florida-health-nvidia-develop-artificial-intelligence-model-hasten-clinical

> **UF Health, University of Florida Health**
> University of Florida Health, NVIDIA develop artificial intelligence model to hasten clinical trials, boost medical decision-making
> Apr 8th, 2021 (268 kB)

Joanna Russell Bliss 11:53 AM
@Ada negraru was asking about models for project management. Here are the links I shared for Agile and Waterfall models; there is a huge amount of info about these models out there if you want to learn more.

Agile: https://www.atlassian.com/agile/project-management

> **Atlassian**
> Get started with agile project management | Atlassian
> Agile project management is an iterative approach to managing software development

Message #resources-and-examples-active

SMU LIBRARIES

Search DHRI@SMU

DHRI@SMU

- Threads
- All DMs
- Mentions & reactions
- Slack Connect
- More

Channels

- # discussion-questions-active
- # general-active
- # random
- # resources-and-examples-a...
- # spatial_data
- # surveys-feedback-active
- # technical-active
- Add channels

Direct messages

- Slackbot
- Joanna Russell Bliss you
- Adam Neal
- Eric Godat, Jaymie Ruddoc...
- Eric Godat, Jonathan McMi...
- Jaymie Ruddock
- Lourdes Molina
- Rafia
- Rylee Bailey
- Talia Weltman-Cisneros
- Add teammates

general-active

For general chat, introductions, etc.

52

August 3rd, 2021

Rafia 10:46 AM
Going back to 'helping you to learn to ask questions' better questions, look at about pages helps you to figure out how something was done, and if something is mentioned often -digitization, data cleaning, it is also helpful to ask if this is something that is done a lot/there is a professional community around, are there best practices i can look for?

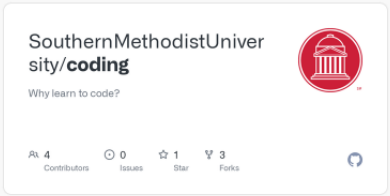
Joanna Russell Bliss 11:18 AM
Copying a question from @Joshua Dodds in Zoom for people to consider over the next couple of days: I think a true pioneer in DH, maybe proto-DH, is Joseph Weinebaum who is the famed creator of ELIZA, the world's first chat bot. The way people reacted to the algorithm scared him and he wrote a book called Computer Power and Human Reason in which he talks about the (natural and cultural) limits of computers and what they should and should not be be used for.Is anyone familiar with Weizenbaum, and is talking about what should and should not be done with digital projects a part of DH? What is the place of ethics in sets of data? Can the full human experience be understood in terms of data, or sets of data? And is DH really even trying to do that?

Rafia 12:04 PM
Today went really well, hearing everyone's comments and questions was really interesting! Looking forward to possibly seeing some of you again this afternoon, and seeing everyone this Thursday. (edited)

The instructors will be back online in 1 hour/1 pm Dallas time if anyone wants to discuss more/has additional questions. We can return to the data discussion over Slack and next T

12:06 For this Thursday, read the Coding Repo https://github.com/SouthernMethodistUniversity/coding

GitHub
GitHub - SouthernMethodistUniversity/coding: Why learn to code?
Why learn to code? . Contribute to SouthernMethodistUniversity/coding development by creating an account on GitHub. (67 kB)



For the Broussard reading mentioned https://github.com/SouthernMethodistUniversity/coding/blob/master/sections/async.md you can access it using your SMU log on, if you don't have an SMU log on you can access it here https://smu.box.com/s/28fyvmll393spisov8g0yomxi1c6rjw2

Message #general-active

SMU LIBRARIES

Search DHRI@SMU

DHRI@SMU

- Threads
- All DMs
- Mentions & reactions
- Slack Connect
- More

Channels

- # discussion-questions-active
- # general-active
- # random
- # resources-and-examples-a...
- # spatial_data
- # surveys-feedback-active
- # technical-active
- Add channels

Direct messages

- Slackbot
- Joanna Russell Bliss you
- Adam Neal
- Eric Godat, Jaymie Ruddoc...
- Eric Godat, Jonathan McMi...
- Jaymie Ruddock
- Lourdes Molina
- Rafia
- Rylee Bailey
- Talia Weltman-Cisneros
- Add teammates

discussion-questions-active

14

August 3rd, 2021

Justin Fisher 7:27 AM

Article: Help a computer win the New Yorker Cartoon Caption Contest

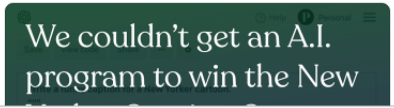
Their data: This is arguably "digital humanities", broadly construed, but is not the paradigmatic sort of processing-a-big-humanities-dataset sort of digital humanities project. The goal of their project is to use AI to produce competitive captions for the New Yorker magazine's weekly contest where (mostly human) contributors compete to produce the best caption for a given New Yorker cartoon. The AI they use is "the GPT-3 model davinci-instruct-beta, through the OpenAI API Playground" which "has read a lot of text on the internet that has described New Yorker cartoons and their captions. It's also read a lot of New Yorker articles and has some understanding of the magazine's readership and what they find interesting or funny. It has generally seen a lot of comics with associated text so presumably understands what is meant by a 'cartoon caption.' " For each week of the 7-week (and counting) trial, they provided this AI with a text description of the week's cartoon, and perhaps additional information like descriptions of past winning captions for other cartoons (their approach and various parameter settings varied from week to week), and had the AI generate one or more proposed captions.

Their Analysis: In some weeks they report a single caption as having been produced. In other weeks they report multiple captions, or report that multiple captions were produced, and either given to human raters to evaluate, or one of the produced options was chosen by their team. That approach involved quite a lot of license for human involvement and judgement calls, more than one would often expect for scientific research, but probably well within the broad umbrella of human-curated "digital humanities" undertakings.

Their visualizations: Their primary output was a visual copy of the week's cartoon with their computer generated caption underneath, made to appear much as a cartoon would in the New Yorker. Looking deeper, they also provided the full text prompt given to GPT-3, verbally describing the cartoon and asking for a caption, and in some weeks they offer a list of runner-up responses also produced by the AI.

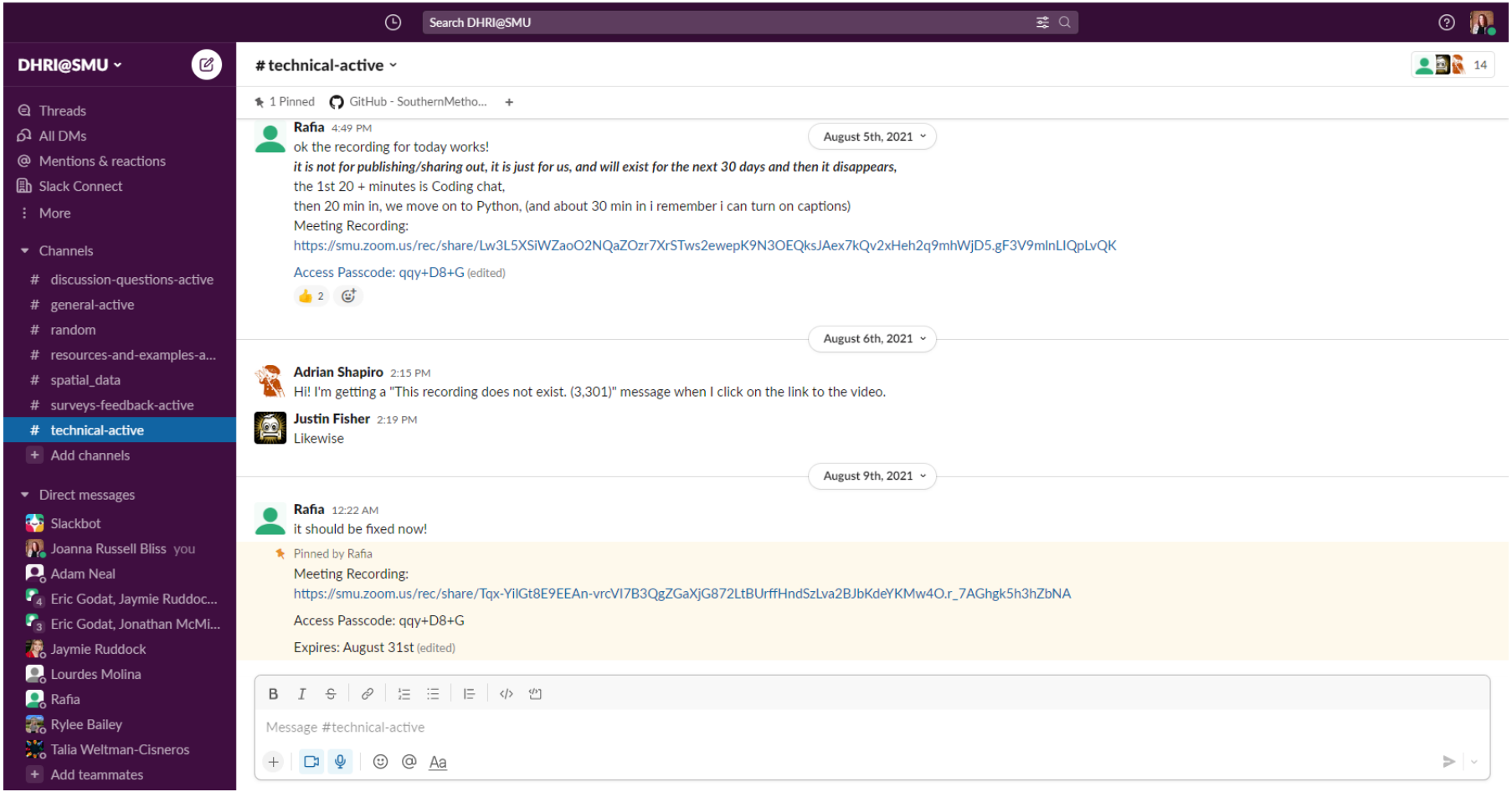
Overall: This was an interesting undertaking, somewhat akin to the classic Turing Test which tests whether computers can pass as human in text-message conversation, though since this format offered no opportunity for follow-up conversation it's difficult to ascertain just how well the AI "understood" the task or its contributions. The scientist in me would prefer to see this done with much larger data-sets (i.e., more cartoons and more runner ups for each cartoon) and much less human intervention. I suspect that the team's "curation" of results shuffled numerous not-so-good results under the rug and presented readers with a misleading picture of what all the AI actually produced. It may well be that we are more seeing the human team's sense of humor in selecting among various randomish results the AI produced, rather than the AI's own "sense of humor". So this probably isn't very ground-breaking in terms of research on AI and humor, but it may still be successful science journalism, helping to attract a broader audience and introduce them to some of the current capabilities of AI in a fun way. (edited)

> **The Pudding**
> Help a Computer Win the New Yorker Cartoon Caption Contest
> This is a weekly experiment to see if an artificial intelligence program can produce real humor. (107 kB)



Message #discussion-questions-active

SMU LIBRARIES



SMU LIBRARIES

SMU LIBRARIES

Evolution & Offshoots

Future plans

Evolution

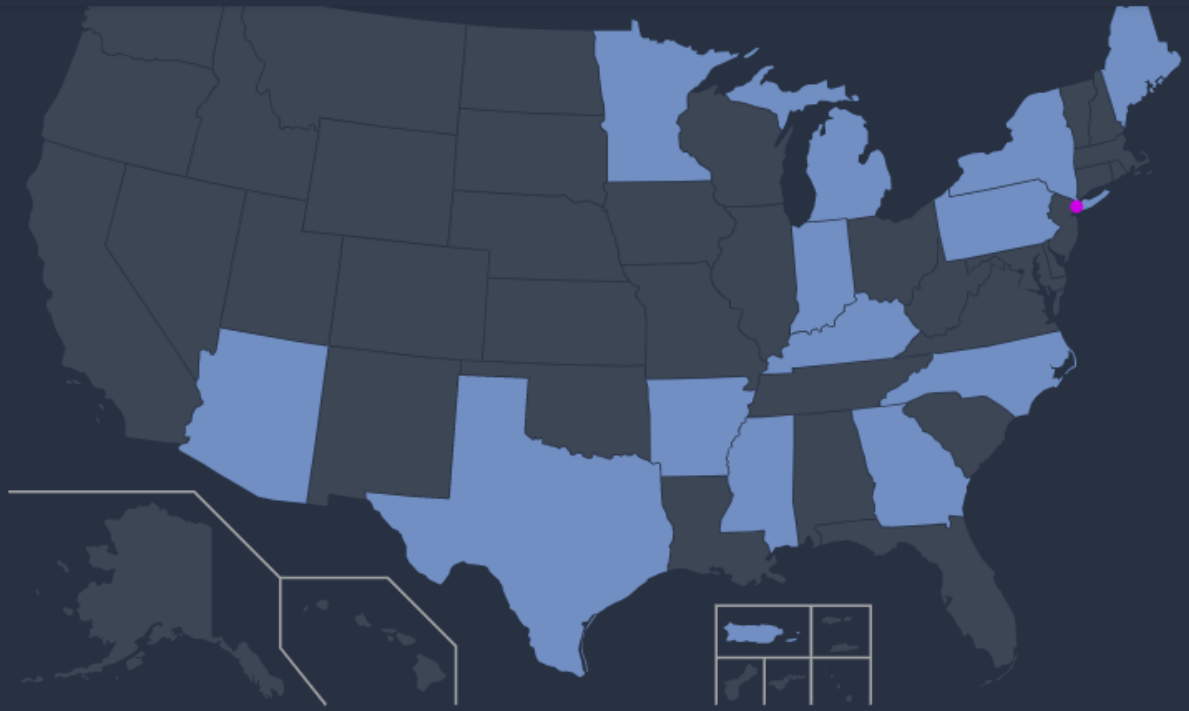
- Workshop series
  - Text and Data Mining (TDM) institute
  - Discipline-specific TDM workshops
  - Database-specific TDM workshops
    - Constellate
    - HathiTrust
    - Voyant
- Possible future iterations: Social Media, Machine Learning/ AI, GIS

SMU | LIBRARIES

Digital Humanities Research Institute

Click on a state on the map to the right to visit the institute's individual pages.

Below, you can read more about each of the individual Digital Humanities Research Institutes that have taken place already. For some of them, you can see interviews with the current Community Leaders.



SMU LIBRARIES

Questions for You to Consider

How: Funding & Technical Infrastructure

- Who is available to work on this?
  - What expertise do they have?
- When can you schedule this?
- Where can you have this?
- Why are you doing this?
  - What is the organizational context?
  - What funding is available? What technical infrastructure is available? What are your (pedagogical) goals?
  - More questions to consider in this handout

SMU LIBRARIES

Thanks!

Any questions?

https://www.smu.edu/dhri

https://bit.ly/tladhri