Part 0 Deliverable (Due 8/20)

Work through the following document as a team. Follow the instructions listed for each section. Be sure to provide sufficient detail and provide links and references to articles & documentation where applicable.

Be sure to collaborate with your team members to figure out each one of these sections.

Group Name: The NeuroVectors

Group Members: *Gennadii Ershov, Darylisha Williams, Lubna Asmi, Jessica Lawrence*

Selected Project: The Lecture Navigator

Background Research

In bulleted format, what are some relevant articles or publications that frame your research question? ***Please include at least 5 relevant links.***

- ***Artificial intelligence in education: Addressing ethical challenges in K-12 settings*** https://pmc.ncbi.nlm.nih.gov/articles/PMC8455229/#:~:text=Despite%20the%20benefits%20of%20AI,professional%20development%20%5B2%2C%2058%5D
- **AI Considerations for Teaching and Learning -**https://teaching.resources.osu.edu/teaching-topics/ai-considerations-teaching-learning
- **Artificial Intelligence in education: A systematic literature review** https://www.sciencedirect.com/science/article/pii/S0957417424010339#s0130
- **Navigated learning: An approach for differentiated classroom instruction built on learning science and data science foundations** *https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbe2.169*
- **Using artificial intelligence in academic writing and research: an essential productivity tool** https://www.sciencedirect.com/science/article/pii/S2666990024000120

Write a summary of how this research relates to your project below these bullet points

The research provided will inform and support our project. We are seeking to build an AI tool that will allow students, instructors, and TAs to query questions regarding course material and receive summarized answers with citations and slides for reference. Our project aims to adapt the concept of navigating learning to empower students to self-direct their studies. This will allow students to have more personalized learning pathways and accessibility to resources in learning any class material. Our AI tool is not a replacement for learning; it only serves as a reference and study aid. Additionally, for instructors and TAs, the integration of the tool will help in the reduction of time needed to answer common student questions and review the common

queries being given to inform their teaching. This will allow for the material to be streamlined and easier to monitor student engagement with the class material. We are taking into account the ethical usage of AI education by focusing on actual slide references, though we do note that AI can fabricate or hallucinate information. In summary, our lecture navigator seeks to create a student-focused AI tool that will help with comprehension and access to course material.

Problem Definition

What is the EXACT proble*m you are attempting to solve using AI? What would AI add to this workflow that couldn't be solved using other traditional methods?*

Our AI project seeks to provide students, instructors and TA's the ability to be able to query questions regarding the previously taught course material that provides the citations and slide numbers to our users. Our tool allows for our users to quickly access the course material without having to go through each and every slide trying to find the right material. It acts as a jumping off point to find further information about the question being asked. The AI will give the users the opportunity to quickly find information from past material without having to search through every PDF or class notes. This will be helpful for people that do not have the time to search through all of the class material.

Many learning navigators on the market are geared towards instructors and provide opportunities to take a different direction in instructing students. With our project, we focus on secondary/higher education students. With the lecture navigator, students will be able to ask questions and be provided a detailed summary for the question, as well as a citation for where the answer was found in the provided PDF. We will be replacing the instructor-centered focus and having a more student-friendly operation.

Users

Who are your primary users (e.g., students, instructors/TAs, program staff), and what jobs are they trying to get done today? What pains, constraints (time, devices, access), and contexts (in-class, homework, office hours) shape their behavior, and what would "good" look like from their perspective?

The primary users of this project would be the students and instructors/TAs. They will use AI to get answers with citations to lecture slides or the class syllabus. As students, there is a need for a tool that is easy to use and navigate to access class materials, such as lecture slides. For example, there could be a writing assignment where they would not have enough time to find credible sources from the lecture material. The lecture navigator can be a helpful tool in providing context for their questions about the coursework.

For the instructors/TAs, this tool can provide insight into the topics that students are frequently asked or needing help from the class material. This information can be useful

to help identify areas that would need further review in the class. Additionally, this can be used as a way to review any inaccuracies in the class material.

Value

What is the value of solving this problem?
Save time on finding proper citations or answers from class material.
Place to form study groups in one central location.
Track student progress/ cited sources

Limitations

What limitations and risks do you anticipate in your project? Are these limitations related to resources?

1) AI hallucinations
2) Can fabricate or provide inaccurate citations
3) Risk of bias or inaccuracies
4) Currently, unable to track student progress or citations provided due to lack of resources
5) Having to decide the relevant slides to be included with the project

Reference(s):

https://teaching.resources.osu.edu/teaching-topics/ai-considerations-teaching-learning
https://www.digitalinclusion.org/resource/2025-dn-toolkit/
https://navigatorlabs.org/research-publications/

Solutions

How do you plan to address these risks? Keep in mind that narrowing the scope of your project is an acceptable answer.

We plan to use the provided lecture material from Phase 1 and Phase 2 to address any risks or limitations from our project. This will narrow down the scope of the project to reduce our limitations or complexities in our project outcome.

Additional Data

Which datasets might you want to add for this project?
We believe the lecture slide PDFs provided in Phase 1 and Phase 2 are sufficient for our project goals.

Data Privacy Considerations

What data will you ingest and store, and why is each field necessary? Who can access raw data and what practices (env vars, no commits) will you enforce to ensure that sensitive data is not exposed?

We will ingest PDF lecture slides using a Python script(?) that extracts slide text into dictionaries with lecture_name, slide_number, and slide_text. This structure supports fast search and accurate citation.
Raw data access is limited to team members. Sensitive info (e.g. API keys) will be stored in .env, which is excluded via .gitignore to prevent exposure.

Tools

What approach best fits your problem and constraints: RAG (you need verifiable course facts with citations), agents (you need multi-step tool use or routing with escalation), a fine-tuned model (you need consistent tone/format at scale and have ≥200–500 high-quality examples), or a vector-DB-only lookup (you need semantic search without generation)? Why is this choice superior to the alternatives given your data availability, privacy needs, latency/cost targets, and team skills?

RAG best fits our needs because it allows verifiable answers with slide citations. It balances accuracy, no cost, and works well with our team's skills. Other options are either too complex (agents), require more data (fine-tuning), or lack generation (vector-only).

Testing Data

What test set (at least 10–20 examples) will you create that covers easy/medium/hard, near-duplicates, noisy inputs, and out-of-scope cases and what labels (answers, citations, rubric scores) define correctness? Why do these cases represent real classroom usage and edge conditions?

We'll create test questions covering easy, medium, hard, near-duplicates, noisy inputs (typos), and out-of-scope cases. Each will have a correct answer and slide citation. These reflect real classroom use and test the system's accuracy, robustness, and edge-case handling.

Question, expected answer, slide number and name of PDF slide

1) Question : what is a neuron?
   Expected Answer : A neuron is a cell that takes input and provides output via electrical signals (action potential) once we cross a certain threshold potential. + (Slide: Introduction to Neural Networks-slide 7)
   Test type: Easy

2) Question: What are the applications of Neural Networks?
   Expected Answer: Image recognition via Convolutional Nets
   Time series prediction via Long-Short Term Memory Nets
   Embedding generation via Word2Vec
   Word prediction via Transformers
   (Slide: Introduction to Neural Networks-slide 16)
   Test type: Easy

3) Question: What is z?
   Expected Answer: $z = W1X1 + W2X2 + \ldots + WnXn$
   (Slide: Introduction to Neural Networks-slide 13)
   Test type: Hard

4) Question: What are other Options for Gradient Descent?
   Expected Answer: Instead, we utilize either stochastic or mini-batch gradient descent.Both are more cost-effective and lead to approximately the same optimum.
   (Slide: Introduction to Neural Networks-slide 51)
   Test type: Medium

5) Question: Can you show me a deep neural network diagram?
   Expected Answer:
   (Slide: Introduction to Neural Networks-slide 15)
   Test type: Hard

6) Question: What's in your mind?
   Expected Answer: Not in slides or syllabus.
   Test type: noisy

7) Question: Can you provide all the links in - Introduction to Neural Networks pdf?
   Expected Answer: ' listing all links'
   Test type: Hard

8) Question: Explain how perceptrons handle binary categories
   Expected Answer: Discuss perceptrons as binary classifiers ("on/off"), coefficients/weights (W) to an output using the dot product. Include the role of heavyside step function for activation. Mention examples like fraud/not fraud or dog/not dog.
   Test type: Hard

9) Question : what are Epochs?
   Expected Answer: Epoch count is a hyperparameter that you can control.

However we will see that in Keras we can implement a
callback which halts training as soon as we have a “good enough” model.
(Slide: Introduction to Neural Networks-slide 62)
Test type: Easy

10) Question : What is TKH?
Expected Answer: Not in slides or syllabus.
Test type: noisy

GitHub Link

https://github.com/imwaymaran/lecturer-navigator.git