IDS 2935:
Making Sense - Understanding the World with Data and AI
Quest 2

I. General Information

Class Meetings

- Fall 2023
- Tuesday 2 hours and Thursdays 1 hour
- [Location]TBA

Instructors

- Anthony Botelho abotelho@coe.ufl.edu
- Office Location: Norman Hall 501B
- Office Hours: Monday 2pm-4pm (or by appointment)
- Contact: a.botelho@ufl.edu

- Borui Zhang, Natural Language Processing Specialist (Academic Research Consulting and Services)
- Office: Communicore C2-203A
- Office Hours: Friday 3pm-5pm (or by appointment)
- Contact: boruizhang@ufl.edu 352-273-8434

Course Description

How do learners learn? Just as AI and data can help to make sense of how humans learn, existing theories of learning can also help us understand and improve how machines learn. This Quest 2 course examines a broad range of machine learning methods and practices and connects them with theories of social and behavioral sciences to solve real-world, often semi-structured problems. Focusing on use-cases within the domain of education, this course will examine how AI can be applied to model student learning, understand written text, and assess students' hand-written work. Students will learn how to examine data using statistical machine learning methods through the lens of social and behavioral learning theories. Students will also learn how to formulate research questions that can be addressed with the data and methods explored. This course bridges existing resources for applying data science techniques from across disciplines of supervised and semi-supervised learning as well as natural language processing and computer vision in the context of social and behavioral sciences.

Course Overview

This Quest 2 course focuses on the bidirectional relationship of Artificial Intelligence (AI) and theories of learning. Data, combined with methods of AI, can help us make sense of the learning behaviors that emerge through data generated as teachers and students interact with educational technologies. Similarly, many advancements in AI have resulted from theories, models, and biology related to how humans learn. The usage of digital learning platforms, learner management systems, and other technologies is growing across educational settings. Combining AI with the quality data that has been produced by using these tools can improve both teacher instruction as well as learner experiences. This data can help understand the processes of learning beyond other assessment measures like correctness and be used to inform the development of better technologies and instructional content. Like any system designed for human interaction, the data collected by these systems can be messy, incomplete, and generally ill-suited for AI applications in its raw form. The course will guide students through understanding the data, identifying problems, and examining a broad range of machine learning methods and practices to solve these real-world problems in a range of contexts. Each week will focus on drawing connections between introduced topics in AI and learning theories originating from social and behavioral science domains. Focusing on use-cases within the domain of education, students will gain hands-on experience with student interaction data (including clicks, learner-instructor interactions, and behavioral and assessment measures collected within a learning technology), as well as written text (e.g., discussion posts and course reviews), and learners' hand-written work. Assuming no prior programming experience, this course will examine processes, pipelines, and best practices in formulating research questions and using data to address them. These methods will be introduced and explored through several labs and assignments using authentic education data, as well as through discussion and a culminating group project.

Quest and General Education Credit

- Quest 2
- Social & Behavioral Sciences

This course accomplishes the Quest and General Education objectives of the subject areas listed above. A minimum grade of C is required for Quest and General Education credit. Courses intended to satisfy Quest and General Education requirements cannot be taken S-U.

Required Readings and Works

<u>Books:</u>

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit.

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. " O'Reilly Media, Inc.".

Schunk, D. H. (2012). *Learning theories an educational perspective*. Pearson Education, Inc.

<u>Papers:</u>

Baral, S., Botelho, A.F., Erickson, J.A., Benachamardi, P., & Heffernan, N.T. (2021, June). Improving Automated Scoring of Student Open Responses in Mathematics. In *Proceedings of the 14th International Conference on Educational Data Mining*, 130-138.

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two Decades of Artificial Intelligence in Education. *Educational Technology & Society*, *25*(1), 28-47.

Mathnet (2022). Research in Analysis of Student Work in School Mathematics. *Retrieved from https://www.etrialstestbed.org/mathnet*

Pilehvar, M. T., & Camacho-Collados, J. (2020). Embeddings in natural language processing: Theory and advances in vector representations of meaning. *Synthesis Lectures on Human Language Technologies*, *13*(4), 1-175.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, *2*(1–2), 1-135.

Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, No. 1, pp. 29-48).

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.

Materials and Supplies Fees: n/a

II. Graded Work

Description of Graded Work

Graded Activity	Percentage of Grade
Labs	28%
Assignments	15%
Midterm exam	20%

Group Project	22%
Discussion participation	10%

Labs (28%) - 7 labs (4% each)

Labs are designed to help you learn and practice learned AI concepts and skills through a series of scaffolded exercises aligned with real-world questions, challenges, and problems. Each lab will have multiple small coding exercises structured to address objective questions by adding to and modifying provided sample code snippets.

Assignments (15 %) - 3 assignments (5% each)

Assignments involve the application of learned skills and concepts using authentic datasets. Students will work independently to conduct analyses, following the topics covered in class, and report on their results, interpret and discuss the interpretation of results, and reflect on the applied methods and information gained from the analysis.

Assignment 1: Data analysis on ASSISTments Dataset
Assignment 2: Data analysis on MOOC Forum Dataset
Assignment 3: Data analysis on ASSISTments Image Dataset

Midterm (20%)

The Online midterm covers from Week 1 to Week 6: 10 multiple choices, 10 filling blanks, 5 true or false, and 2 short answer questions (no more than 300 words each).

Group Project (22%)

As a team of no fewer than five, choose one of the datasets utilized in class to analyze for a final group project (datasets from other sources may be used with approval of the instructor). With this data, students must propose a well-scoped question that can be answered using methods covered throughout the course. Each group will apply these methods to clean, format, and analyze their data, and then report on the findings. Students should provide discussion that targets how the project addresses a practical problem or otherwise explores new topic areas beyond what is covered in class. Finally, the team should reflect upon their experience, including their specific implementation, and identify potential future directions for the work and any limitations to either their approach or the data in addressing their respective research question.

- Team Contribution Statement (2%) (Week 8 due)
  A statement will be submitted by each team that specifies each team member name and a list of group project deliverables that each member commits to contribute. This statement will be drafted and signed by each team member to ensure a fair distribution of responsibilities.

(Accommodations for extraordinary circumstances resulting in missing team members during the semester will be considered by the instructors on a case-by-case basis)

- Group Presentation with/o preliminary findings (10%)
    - Eight-minute in-class presentation in-progress (4%) (topic, dataset information, methodology, expected outcomes)
    - Five-minute final recording video presentation (6%): upload to designated platform
- Final project report and code submission (10%)
    - Project report (1000 words without output) (6%)
    - Project coding (4%)
    - note: each team member will need to indicate which writing portion(s) of the final submissions (writing report and code files) that they made their individual contribution.
- Anonymous survey: Each group member will evaluate your participation, and that of your group members', based on your Team Contribution Statement.

Discussion Participation (10 %)

Each student is expected to come to the in-person class meetings and participate in class activities and group discussions. Additionally, each will be assigned to post to the class discussion board for 5 separate weeks throughout the course (students will be randomly assigned to a set of 5 weeks at the beginning of the semester). Each post should reflect on what was learned and identify how the covered skills may be used in practice. Posts should also discuss potential challenges, opportunities, and limitations associated with the data and/or methods covered in the respective week. Students will also be randomly assigned to 5 alternate weeks to reply to at least one student's post for that week. Replies should be thoughtful and meaningful and draw connections between the content and their own interests and experiences.

Grading Scale

For information on how UF assigns grade points, visit:
https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/

A	94 – 100%		C	74 – 76%
A-	90 – 93%		C-	70 – 73%
B+	87 – 89%		D+	67 – 69%
B	84 – 86%		D	64 – 66%
B-	80 – 83%		D-	60 – 63%
C+	77 – 79%		E	<60

Grading Rubric(s)

Participation Rubric

Components	10pts	6-9pts	1-5 pts	0pts
Class participation and group discussion	Participate in all class activities. Post to all the assigned 5 weeks of topics and respond to alternative 5 topics. The content of posts and responses meet what is described in the graded discussion work section.	Miss 1-2 class activities in the semester. Miss 1-3 posts or responses from the discussion board. The content of posts and responses meet what is described in the graded discussion work section.	Miss 3-4 class activities in the semester. Miss 4-7 posts or responses from the discussion board. Some of the content of posts and responses do not meet what is described in the graded discussion work section.	Miss more than 4 class activities in the semester. No relevant posts to the online discussion board. No posts or responses meet what is described in the graded discussion work section.

Assignment Rubric

Components	4-5pts	2-4pts	1-2pts	0pts
Data understanding	Clear and detailed description of what you understood from the dataset, which should be well supported by a clear illustration of how you collected information from the dataset with the relevant coding methods.	A description of what you attempted to learn from the data and which methods you think will help you; generally stated about the connections between your question and methods.	A description but unclear at places about what you aim to learn from the dataset and your methods. No connection or vague connections between the question and your methods.	No relevant understanding from the data
Coding/methods	Relevant coding methods selected	State each implement step	Listed the main steps of the	No relevant coding or

	and used correctly, well-reasoned analysis steps to get the desired outputs that support the information described in the data understanding part; completed code with clear variable names labeled; runs with no errors	but lack some sufficient details; provided partial code that matches the steps in a reasonable order. Data can be run without errors, but miss some minor parts.	methods, but largely lack details; code has parts that do not match the described part or code has minor or major running errors.	methods found or used.
Result report	Show clear and straight-forward outputs that are generated by the program you implemented; a clear evaluation on how well the dataset and methods that you adopted served the goal to answer what you proposed to learn from the data. Well-stated reflections on the limitations and future directions of the problem.	Show complete output from the program with missing minor information to answer your attempted question. General evaluation on the methods, relating to your methods and output, even the output has minor parts that are missing to be difficult to evaluate.	Show the generated output of your code with errors. The evaluation misses some major point to answer the attempted question. Some general reflection on the methods you used.	No result report

Final Project Rubric

Components	Excellent	Complete	Incomplete
Team Contribution Statement (2%)	Must include all required information	NA	NA
In-class presentation (5%)	Clear demonstration of a practical topic and proposed solutions.	Not all materials and methods are stated clearly but main	Significant materials and methods are missing.

	Clearly and simply state used resources, methods, and the expected outcomes. Logically layout method steps and/or coding plan. Respect the given time limit.	topic/problem idea was appropriate and supported.	
Final recording (5%)	With the same expectation from the in-class presentation, final recording needs to be clearly articulated.	Some areas appeared hard to follow; minor errors in the slide but logically presented.	Important aspects of the presentation were missing or incomprehensible.
Final paper report (6%)	All stated deliverables are reflected completely. Clear and straight-forward outputs that are generated by the program you implemented; a clear evaluation on how well the dataset and methods that you adopted served the goal to answer what you proposed to learn from the data. Well-stated reflections on the limitations and future directions of the problem.	Show complete output from the program with missing minor information to answer your attempted question. General evaluation on the methods, relating to your methods and output, even the output has minor parts that are missing to be difficult to evaluate.	Show the generated output of your code with errors. The evaluation misses some major point to answer the attempted question. Missing aspects of reflection on the methods you used.
Final paper coding (4%)	Relevant coding methods selected and used correctly, completed code with clear variable names labeled; runs with no errors	Provided partial code that matches the steps in a reasonable order; runs without errors, but misses some minor parts.	No relevant coding or methods found or used.

Late Policy:

Labs/Assignments are released on Tuesdays classes. Student work will receive a 10% deduction if submitted within the 24 hours past due, 20% deduction within the three days past due, 30% deduction within the 7 days past due. Assignments are not accepted past the 7 days following the due date. University Accommodations rules may apply for individual adjustments.

III. Annotated Weekly Schedule

Week	Topics, Homework, and Assignments
Week 1 Tues/Thurs days 0824 (Thur)	● **Topic:** Overview of Machine Learning (ML) in Social and Behavioral Sciences ● **Summary:** The first lecture gives an overview of the history of the development of AI/ML and how it is used in social and behavioral sciences. In the field of education, digital learning tools are being used to improve learning and instruction. We will start with an overview of how ML is used in helping analyze teaching and learning qualities via different types of data. ● **Required Readings/Works**: o Two Decades of Artificial Intelligence in Education (Zou et al. 2022) (20 pages) o A Brief History of Natural Language Processing (NLP) - DATAVERSITY o Current relevant news or blog post (e.g., Duolingo language proficiency tests) ● **Discussion 1:** How do you define AI? What is your experience with AI or digital learning platforms? Where do you believe AI can be utilized to improve education, if at all? ● **Lab 0** (No submission required, no credit): Getting familiar with Python environments: Colabratory
Week 2 0829, 0831	● **Topic:** Data analysis with Python and Introduction to Learning Theory ● **Summary:** Getting familiar with the concepts in general data analysis; basic Python command lines for package installation, dataset importing and viewing. ● **Required Readings/Works**: o Bird Chapter 2.1 (13 pages) o Schunk Chapter 1 (Page 5-10, 6 pages) ● **Discussion 2:** Consider the theories of Behaviorism, Constructivism, and Cognitivism as it pertains to data analysis, in what ways do you believe each of these theories may be represented through data? ● **Lab 1:** Python basics (pip/import data/data overview,pandas,numpy,sklearn,nltk) ● **Lab 0 due**
Week 3 0905, 0907	● **Topics:** Behaviorist Learning Theory & Prediction Modeling ● **Summary:** Getting familiar with basic modeling pipelines while examining a behaviorist-view of data. Lecture will include data structure and correlation between features and outcomes, basic concepts of supervised learning versus

Week	Topics, Homework, and Assignments
	unsupervised learning, data splitting, and understanding basic model evaluation metrics. ● **Required Readings/Works**: o The ASSISTments Ecosystem (Heffernan et al. 2014)(28 pages) o Schunk Chapter 3 (Page 71-75, 5 pages) ● **Discussion 3:** Share an example of a potential or known application of machine learning for classification. Do you believe this model is likely to work equally well in all scenarios? What risks do you potentially see in using machine learning for classification in educational settings? ● **Lab 2** Supervised learning (Classification/regression) ● **Lab 1 due on Tuesday before class**
Week 4 0912, 0914	● **Topic:** Cognitivist Learning Theory and Featurization ● **Summary:** Real-world data in education is often unstructured or semi-structured, which needs to be featured into variables such as continuous, nominal, and ordinal variables. Lectures will discuss how these variables can be engineered to represent learning processes, making connections to cognitivist learning theory, and how different types of variables can help structure information in different ways to support modeling. ● **Required Readings/Works**: o Bird Chapter 1.2 (6 pages) o Schunk Chapter 7 (Page 278 - 281, 4 pages) ● **Assignment 1:** Identify variables (examples) and analyzing the ASSISTments dataset ● **Lab 2 due on Tuesday before class**
Week 5 0919, 0921	● **Topic:** Constructivist Learning Theory and Word-level Tokenization ● **Summary:** Text-based data is one of the common raw sourced datasets that we get in the educational setting (e.g. learning experience evaluation forms). Lecture discusses why text-based data needs to be preprocessed and the factors that determine how text should be processed: different linguistic features of languages, the format of the text (long versus short, documents versus dialogs) Various tokenizing tools (regular express, white space, pretrained tokenizers) are developed along with the NLP field. ● **Discussion 4:** Language is often used in education to construct something that demonstrates understanding (e.g. essays), which is why it is often considered from the perspective of Constructivist theory. In what ways can the meaning or usage of words change in different contexts (or different languages) and how might this affect how we interpret language data? ● **Required Readings/Works**: o P&C Chapter 3 (25 pages) o Schunk Chapter 6 (Page 230-233, 3 pages) ● **Lab 3:** Tokenization for English (examples)
Week 6	● **Topic:** Document-level Featurization and Social Learning Theory

Week	Topics, Homework, and Assignments
0926, 0928	● **Summary:** In practice, models can often be improved by incorporating features of groups within data (e.g. students in classrooms). In NLP, this often includes the use of document-level data to examine words and phrases that often appear together. Two widely-used measures of Co-Occurrence and TF-IDF are explored in this lesson as examples of how we can use higher-level information to build useful features for analysis. Lecture introduces these concepts while making connections to social learning theories and discourse. ● **Required Readings/Works**: o TF-IDF (Ramos 2003) (4 pages) o Schunk Chapter 4 (Page 119-123, 4 pages) ● **Discussion 5:** Processing documents is different from processing sentences/phrases. If you had enough resources, what kinds of AI applications would you like to build using the techniques of processing documents? Do you foresee potential challenges of your application? ● **Lab 4**: document similarity and topic modeling exercises ● **Lab 3 due on Tuesday before class**
Week 7 1003, 1005	● **Topic:** Sentiment Analysis and Cognitive-Affective-Social Learning Theory ● **Summary:** Building upon the types of features and representations discussed in previous weeks, this lesson focuses on a practical application of these features. This lesson focuses on Sentiment Analysis to measure the tone and mood implied within text, with connections to Cognitive-Affective-Social Learning Theory. Examining the usage of positive and negative words, we will examine applications of detecting emotion in text. ● **Required Readings/Works**: o P&L Chapter 1-3 (10 pages) o Turney 2002 (8 pages) ● **Discussion 6:** Sentiment analysis seems to be useful for detecting users' attitude towards learning. Do you see any problems with this analysis and how do you suggest solving them? ● **Lab 5** Sentiment analysis exercises ● **Assignment 1 due**
Week 8 1010, 1012	● Tuesday Midterm Review ● Thursday Midterm ● **Lab 4 due on Tuesday before class**
Week 9 1017, 1019	● **Topic:** Syntactically motivated data structures of text ● **Summary:** Essays, forum posts, and open response data is among the most common, and often most challenging to utilize, in education settings. This lesson will examine common NLP methods involving Part of Speech tagging, sentence parsing, and the usage of treebank corpora to analyze this type of data. ● **Required Readings/Works**: ○ Bird Chapter 5.1-5.3 (10 pages) ○ Bird Chapter 8.1-8.5 (18 pages)

Week	Topics, Homework, and Assignments
	● **Discussion 7:** Linguistic grammar structures play an important role in sentence parsing. Do you see any potential learning applications that will have a good use from sentence parsing? ● **Assignment 2:** Data analysis of MOOC forum dataset (include data cleaning, featurized variables, training methods) ● **Team Contribution Statement due on Tuesday before class** ● **Lab 5 due on Thursday before class**
Week 10 1024, 1026	● **Topic:** Deep Learning ● **Summary:** Many of the most recent advancements in machine learning have involved applications of deep learning methods. Drawing inspiration from biology and neuroscience, deep learning (DL) is a method of ML that attempts to emulate how neurons work in the human brain. Developments of libraries, frameworks, and tools have made it easier to build and apply these types of models without extensive prior knowledge requirements. This lesson introduces the basic structure of feed-forward deep learning models for supervised learning tasks. Using the widely-used MNIST dataset, focus will be devoted to examples of applying deep learning models, interpreting output, and understanding hyperparameters and model training. ● **Required Readings/Works**: o Géron2019 Chapter 10 (p278-p290) ● **Discussion 8:** Deep learning brings more possibilities to solve complex problems. It is fascinating to see the tasks that DL models have accomplished. Based on what you learned about DL, are there any complex problems (or new inventions) in the educational domain that you think DL might be great tools to solve? ● **Lab 6** : Simple Neural Network exercise
Week 11 1031, 1102	● **Topic:** Visual Competency in Learning and Image Representation ● **Summary:** Building upon the previous lesson, this week examines how deep learning models can be utilized to represent image data and contrast this with how humans interpret visual information. Drawing upon concepts of featurization introduced earlier in the course, this lesson will examine how features can be learned from groups of pixel data forming edges, shapes, and other composite features. The concepts behind convolutional neural networks will be introduced, expanding upon the previous lesson on deep learning models. ● **Required Readings/Works**: o Keras CNN for MNIST Worked Example o Schunk Chapter 4 (Pages 131- 133) ● **Assignment 3:** Analysis of the ASSISTments Image dataset (for image representation and text identification) ● **Lab 6 due**
Week 12 1107, 1109	● **Topic:** Word Representations with Deep Learning ● **Summary:** Returning to a focus on NLP, this week examines how deep learning can be applied to language data. This lesson focuses on how the word- and

Week	Topics, Homework, and Assignments
	document-level featurization methods introduced earlier in the course can be combined with deep learning to represent language with examples of Word2Vec and GLOVE as pivotal representation methods used in practice. ● **Required Readings/Works**: o Word2Vec (8 pages) o GLOVE ● **Lab 7:** Pre-trained GLOVE model exercise ● **Assignment 2 due on Tuesday before class**
Week 13 1114, 1116	● **Topic:** Appropriation of Knowledge and Foundation Models ● **Summary:** While the previous lesson introduced applications of deep learning to represent language, the focus of this lesson is to examine the strengths and weaknesses of state-of-the-art NLP methods in practice. As humans, we constantly appropriate knowledge from other sources when learning a new concept, and machine learning can often benefit from this same practice. We will introduce additional practical models of BERT and GPT-NEO for tasks of text understanding and text generation. ● **Required Readings/Works**: o BERT documentation o GPT-3 Examples ● **Discussion 9:** Do you think how neural language models are trained is similar to how humans learn a language too? Why or why not? ● **Assignment 3 due**
Week 14 1121, 1123 (holiday)	● **Topic:** Foundation models for Computer Vision ● **Summary:** The purpose of this lesson is to showcase state-of-the-art methods that combine supervised learning, NLP, and computer vision. We will examine foundation models for both text-to-image and image-to-text, with particular focus on how these methods can be utilized in practical settings using the assignment 3 dataset. ● **Required Readings/Works**: o DALL-E 2 Overview ● **Discussion 10:** How do you think Computer Vision can benefit digital learning? ● **Lab 7 due on Tuesday before class**
Week 15 1128, 1130	● **Topic:** Workshop week ● **Summary:** This week is devoted to helping students identify, articulate, and address practical challenges that they are facing within their group projects. Each group will be asked to share a challenge that they are facing with their project to share and discuss as a class. Students will then be given time to work together within their groups to formulate a plan to incorporate feedback.
Week 16 1205 (last class)	● **Topic:** Group progress presentation ● **Summary:** In the final class period, each group will be asked to give a short pitch describing their overall project and their progress as of the class period. These

Week	Topics, Homework, and Assignments
	presentations should focus on how the project addresses a practical problem and/or explores new topics beyond what has been covered in class. ● **Group project presentation video due on the final exam week 12/12** ● **Finished project report + coding link due on the final exam week 12/12**

IV. Student Learning Outcomes (SLOs)

At the end of this course, students will be expected to have achieved the Quest and General Education learning outcomes as follows:

Content*: Students demonstrate competence in the terminology, concepts, theories and methodologies used within the discipline(s)*.

- Identify, describe, and explain the terms and concepts related to the application of AI to study practical problems in educational contexts, as well as the ethical and methodological challenges and limitations of such methods (Quest 2, S). Assessment: Labs, assignments, exam, and the group project.
- Recognize connections between types of real-world data and appropriate problem-solving techniques involving AI modeling (Quest 2, S). Assessment: Labs, assignments, exam, and the group project.
- Apply the basic strategies of extracting useful information from massive data pools to address practical problems (Quest 2, S). Assessment: Labs, assignments, exam, and the group project.

Critical Thinking: *Students carefully and logically analyze information from multiple perspectives and develop reasoned solutions to problems within the discipline(s).*

- Analyze various types of data at scale using basic AI tools to examine students' learning outcomes and behavior, learners' attitudes to address potential learning problems, and evaluate or suggest better learning practices (Quest 2, S). Assessment: Assignments, the group project.
- Draw connections between the education-focused datasets and problems and their respective appropriate AI methods to other disciplines (Quest 2, S). Assessment: Discussion, the group project.
- Reflect on both ethical and technical perspectives of the AI methods used in the analyses of the educational datasets (Quest 2, S). Assessment: Assignments (individual level reflection), the group project and discussion board (group level reflection).

Communication: *Students communicate knowledge, ideas and reasoning clearly and effectively in written and oral forms appropriate to the discipline(s).*

- Communicate data and analysis outputs, make sense of the real-world by drawing facts from abstract, massive information, propose explainable narratives for making personal and group decisions (Quest 2, S). Assessment: Assignments, group projects.
- Work in a team-setting, develop and present a potential problem/topic and the solution in the education field via a set of real-world data, with a list of deliverables that can explain how your

selected methods can solve/understand the problem (Quest 2, S). Assessment: Assignments, group projects.

Connection: *Students connect course content with meaningful critical reflection on their intellectual, personal, and professional development at UF and beyond.*

- Connect the ongoing challenges and opportunities using AI in advancing student learning, as a junior research in-progress, how will you contribute to the field and make impact to human society (Quest 2). Assessment: Assignments, discussion, group projects.

V. Quest Learning Experiences

1. Details of Experiential Learning Component

- AI interactive applications - students gain hands-on experience in the practical applications and limitations of AI, while encouraging students to develop ethical awareness in the early stages in their academic career. Some selected AI applications that students will be experiencing in the class:
    - Duolingo: language learning application uses AI to improve the quality of customized language proficiency tests.
    - DALL-E2: artwork generating application uses AI to enable natural language to directly instruct image drawing.
    - OpenAI Playground: a collection of AI applications that are powered by language models which can response any task that user selected including summarize for a 2nd grader, Q and A, parse unstructured data
- Assignments - students will work closely with authentic datasets collected in education settings to gain experience in formatting, cleaning, and analyzing to address real problems in the domain.
- Group projects - will provide a structured team setting with committed deliverables and timelines, post-project anonymous internal team grading, and a real-life conference presentation format.

2. Details of Self-Reflection Component

Self-reflections are included in discussion posts, three assignments, and the final group project, with each emphasizing different perspectives in connecting course content to real world applications within and beyond the domain of education. Students are provided with opportunities to reflect not only on the AI methods and applications, but also on the structure and usage of the authentic datasets with accompanying challenges and limitations.

VI. Required Policies

Attendance Policy

Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies that can be found at: https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx

Students Requiring Accommodation

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting https://disability.ufl.edu/students/get-started/. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

UF Evaluations Process

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at https://gatorevals.aa.ufl.edu/students/. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via https://ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at https://gatorevals.aa.ufl.edu/public-results/.

University Honesty Policy

UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code (https://www.dso.ufl.edu/sccr/process/student-conduct-honor-code/) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Counseling and Wellness Center

Contact information for the Counseling and Wellness Center: http://www.counseling.ufl.edu/ , 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.

The Writing Studio

The writing studio is committed to helping University of Florida students meet their academic and professional goals by becoming better writers. Visit the writing studio online at http://writing.ufl.edu/writing-studio/ or in 2215 Turlington Hall for one-on-one consultations and workshops.

In-Class Recordings

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A "class lecture" is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To "publish" means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

Appendix:

Explanation of Social and Behavioral Science Integration Throughout the Course

Advancements in not only methods of AI, but also the improved accessibility and usability of libraries and tools to build statistical and machine learning models, have led to increased usage across fields – including social and behavioral sciences. With even the most basic understanding of a small number of programming concepts, researchers can build and apply state-of-the art methods in their respective fields of study.

In areas of social and behavioral sciences, especially in education, various modeling and machine learning methods (commonly referred to collectively as AI or *methods* of AI – though we will not argue nomenclature) are used in two primary capacities: 1) they are used to study and test theories of learning using data – quantitatively examining concepts that would previously be limited to qualitative examination (such as emotion/affect) or examining concepts at finer granularities (i.e. studying the moment-to-moment processes of learning), but also 2) AI methods are conversely themselves examined, revised, and improved through the lens of social science theory.

For example, throughout the course, we will examine various methods through the lens of Behaviorist, Cognitivist, and Constructivist learning theories (among others, with these being the most prominent). As Behaviorist learning theory perceives the human mind as a "black box" through which learning processes cannot be directly observed, we will examine each of the methods introduced after week 2 from this perspective first. This contrasts Cognitivist learning theory which views the mind from the perspective of undergoing a set of processes during learning (information is received, it is processed, compared to memory of prior experiences, and is used to revise mental models of ideas and concepts); we will compare the cognitivist view of methods with behaviorist starting in week 2, with greater emphasis being given from week 4 onward – models, too, can be thought of as a set of processes, implicitly or explicitly comparing new data to its existing "understanding" of that data that is being modeled. With this, we will also examine how theory and domain knowledge can help guide methodological choices to address common misconceptions regarding the applications and limitations of AI.

In short, concepts and theories of social and behavioral sciences become the focus of each module, and define the lens through which we examine each of the methods introduced throughout the course.

Flipped Classroom Design

The structure of the course is inspired by a flipped-classroom design. Very few existing instructional materials exist that make explicit connections between methods of AI and theories common to social and behavioral sciences (thus part of our motivation for this course). Students will be supplied with resources (i.e. readings, videos, and examples), to be considered outside the classroom, which are designed to provide students with sufficient foundational knowledge aligned to each module; class period itself, therefore, can focus on drawing connections between the methods and theories along with potential for application and associated limitations (both from the methods' and theories' perspectives, as both of these come with their own assumptions and contextual considerations).

Structure of Assignments

This is not a class designed to teach students programming. While we will examine several methods that will utilize some foundational programming concepts (i.e. the ability to look at, change some values, and run scripts of code), we acknowledge and emphasize that this class intends to support students of varying experiences and backgrounds (i.e. we make no assumptions of any prior programming knowledge or experience). Even with this, improvements in usability and accessibility of AI tools makes it easy for someone with no experience to apply even the most complex state-of-the-art methods. Consider this example code:

```python
from sentence_transformers import SentenceTransformer
from sklearn.metrics import cohen_kappa_score
from datautility import read_csv

SBERT = SentenceTransformer('bert-base-nli-mean-tokens')
data, headers = read_csv('student_text.csv')

decision_tree = DecisionTreeClassifier(max_depth=5)
decision_tree.fit(SBERT.encode(data[:, 2]), data[:, 3])

test_data, _ = read_csv('student_text_test.csv')
predictions = decision_tree.predict(data[test_data])

print(cohen_kappa_score(data[:, 3], predictions))
```

In these 10 lines of Python code (the programming language utilized in the course), we are reading in a dataset, applying a Deep Learning Neural Network-based Natural Language Processing method developed and made available by Google (called SentenceBERT – see week 13 of the syllabus), training a decision tree machine learning model (see week 3 of the syllabus), and evaluating it using a common measure utilized in social and behavioral sciences (Cohen's Kappa, often used to establish agreement between raters in field observation settings – it was appropriated from this context and has now become a common measure used in machine learning). In other words, students are not being asked to write code to implement these methods/models from scratch, but will instead be using 1-2 lines of code that simply applies the method from existing publicly-available sources. This example is meant to illustrate how these tools and methods are accessible and approachable by even those with no programming experience with the proper instruction and scaffolding; while some of the lines may look a little confusing to someone without programming experience, learning what is happening in just 10 lines of code is enough to apply even the most advanced topics covered in the course.

For each lesson and assignment requiring any programming, students will be supplied with carefully-designed scaffolding such that they are not expected to produce code themselves from scratch (unless they want to), but rather can modify starter-code to examine how changes to model parameters (i.e. changing the "max_depth" of the decision tree in the above example) alters the model performance as well as how it may alter the interpretation of its output. Beyond the example above, we will examine how the output of the model can be aligned with theories of learning to quantitatively examine human learning strategies, processes, and behaviors as alluded to throughout the syllabus.