

DataRes Research

Retrieval-Augmented Generation:
BruinPlanner[©]

By: Derek Shi, Henry Liu, Alexis Adzich, Yulia Anashkina



BruinPlanner

01

**What even IS
RAG?**

02

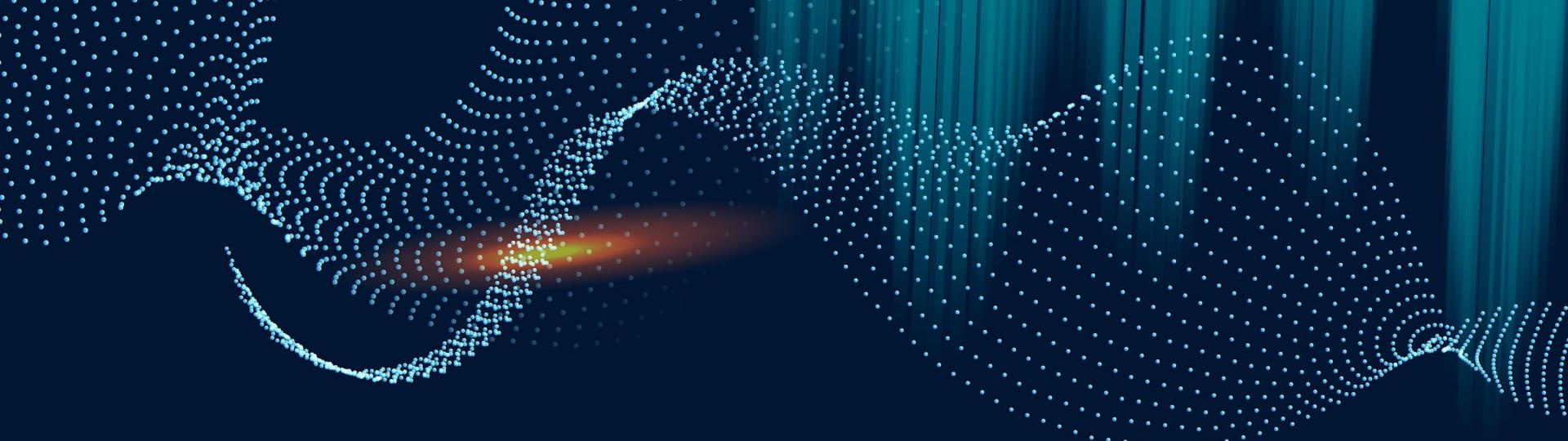
Our Project

03

Demo



BruinPlanner



What is RAG?

aka Retrieval Augmented Generation

01



BruinPlanner

RAG

Retrieval-Augmented Generation combines two key processes:

1. Retrieval:

Fetching relevant information from a database, knowledge base, or document set.



2. Generation:

Using language model (like GPT) to generate a response, incorporating the retrieved information to make the output more accurate and contextually relevant.



BruinPlanner

General Project Overview

USER INPUT

"I like coding and machine learning"

SIMILARITY SEARCH

DATA THEORY B.S.

PRI-MAJOR (1) COURSES
Students can declare the pri-major at any time while in good academic standing.

Quarter	Grade	Quarter	Grade
MATH 104P		One course from the following: STATS 103S 105	
MATH 105P			
MATH 107P		STATS 207	
MATH 108P		STATS 207	
MATH 112P		PC 104P	
MATH 154P			

All preparatory courses must be completed with a minimum overall **3.0** grade-point average and a grade of "C" or better in each course. No more than one-half of the total number of hours of these two courses may be taken before the major is declared.

THE MAJOR (16 COURSES)
Students must pass all courses before reaching 100 units (not including AP units). Other restrictions apply please see the class and catalog for more information.

Quarter	Grade	Quarter	Grade
MATH 103A		STATS 105A	
MATH 108A		STATS 105C	
MATH 156		STATS 105D	
		STATS 105E	
		STATS 105F	
		STATS 105G	
		STATS 105H	

Qtr 1: Quarter sequence of Probability and Statistics
A two-quarter sequence of Probability and Statistics

Math 105E	Math 105F
One statistics elective from STATS 100C, 108, 102C, CSD 199	

Qtr 2: Intro to Probability and Mathematical Statistics
A two-quarter sequence of Probability and Statistics

STATS 100A	STATS 100B
	Two additional electives from the mathematics and/or statistics selection listed below:
1	2

One capstone course: MATH 154 and STATS 104B (to be taken in the final year).

*Students must pass this course with a "C" or better. It is strongly recommended that students take MATH 154 as one of their first upper division courses before taking MATH 156.
**Students must pass this course with a "C" or better. This course must be the least 40 units and taken for a letter grade, unless the class is a supplement to a larger course (e.g. a required lab).
***Capstone projects are available to students who have completed the major requirements.



-0.32643065,
-0.12308089,
-0.2873811 ,
-0.99628943,
-0.2503798 ,
0.24311952,
0.5662387 ,
0.17282294,
-0.1109335 ,
0.15209009 ,
0 47017908 ,
-0.19270805 ,



-0.32643065,
-0.12308089,
-0.2873811 ,
-0.99628943,
-0.2503798 ,
0.24311952,
0.5662387 ,
0.17282294,
-0.1109335 ,
0.15209009 ,
0 47017908 ,
-0.19270805 ,



OUTPUT: most similar vectors

RETRIEVAL

courses + course descriptions

► Math 33AH: Linear Algebra and Applications (Honors)

► Math 33B: Differential Equations

General Course Outline

Course Description

(4) Lecture, three hours; discussion, one hour. Requisite: course 31B with a grade of C- or better. Highly recommended: course 33A. First-order, linear differential equations; second-order, linear differential equations with constant coefficients; Power series solutions; linear systems. P/NP or letter grading.

Course Information:

In addition, two hour exams should be given. These exams are usually given in the fourth and eighth week; the exact time they are scheduled is up to the instructor. 24 of the 26 lectures are specified.

The course Math 33B has evolved over the years. At one time it was a course in infinite series, including power series solutions of differential equations. In the Fall of 1997 the infinite series course was renumbered as Math 31C, in hopes that students would take it earlier, but by the Fall of 1998 the course was back at the end of the calculus sequence with its original label Math 33B.

In 2004, the courses Math 33A and 33B were reorganized. The differential equations portion of Math 33A was moved to Math 33B, so that Math 33A is now a course in linear algebra and Math 33B is now a course in differential equations. The topics currently treated in Math 33B are as follows:

Introduction to first order differential equations

- second order linear differential equations with constant coefficients
- power series solutions of second order differential equations
- linear systems of differential equations

Textbook(s)

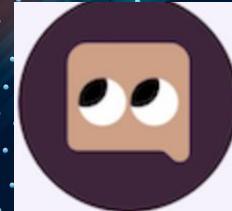
Polking, *Differential Equations*, 2nd Ed., Prentice Hall.



major course planner

BruinPlanner

How RAG is being used:



1. E-Commerce Product Recommendations

Platforms like *Amazon* or *Shopify AI* retrieve product details, reviews, and specifications and generate personalized suggestions.

2. Education and Research Tools

Elicit and *Khanmigo* (Khan Academy's AI) retrieve specific knowledge or research papers and provide relevant, AI-enhanced learning or research guidance.

3. Legal AI Assistants

Casetext's CoCounsel or *Harvey* retrieve case law, precedents, and legal statutes and generate summaries or arguments tailored to specific legal needs.

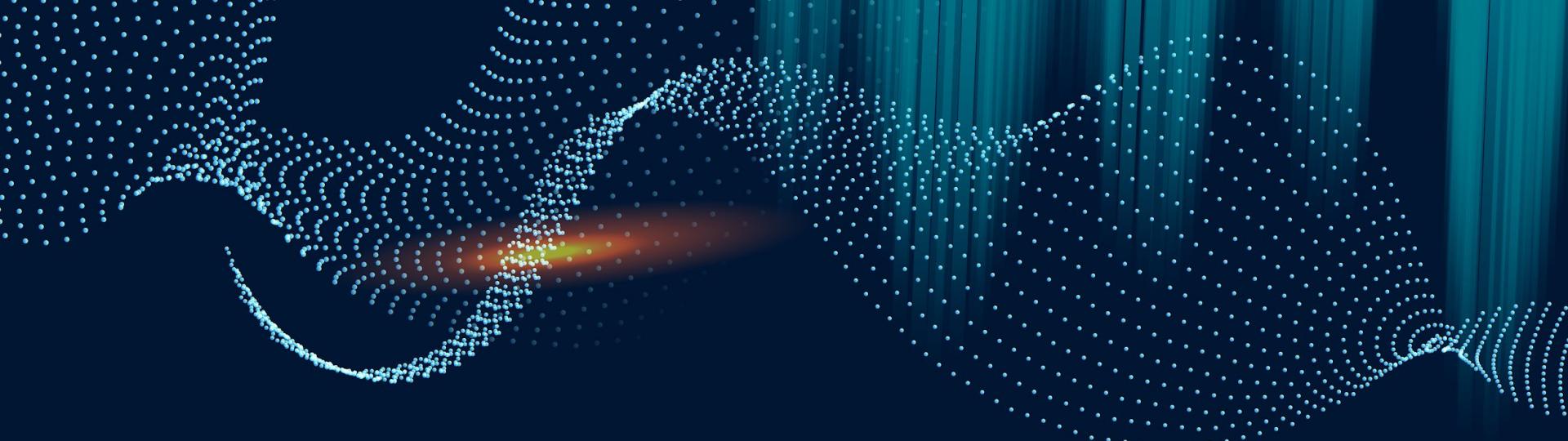
4. Healthcare Applications

MedPaLM or *Infermedica AI* use RAG to fetch clinical guidelines, patient data, and medical literature to generate insights or recommendations for practitioners.

Harvey OpenAI



BruinPlanner



02

What problem did we tackle?



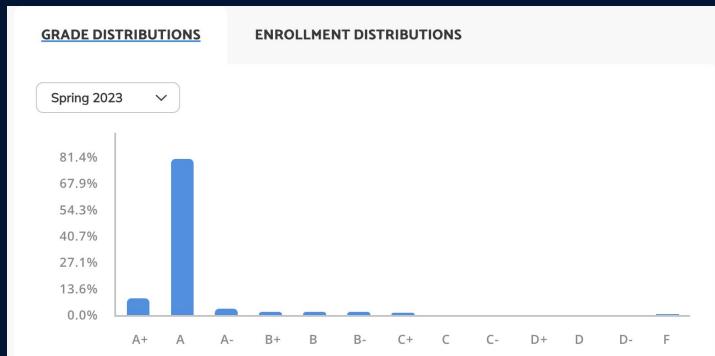
BruinPlanner

Problem Statement

Have you ever been in this position??



A screenshot of the UCLA Catalog website. The header reads "UCLA Catalog General Catalog 2024-25". Below the header, there are navigation links for College and Schools, Departments, Subject Areas, About UCLA, Undergraduate Study, and Graduate Study. The main content area is titled "Advanced Search" and includes a search bar and a message stating "You're now viewing the 2024-25 Catalog". Below the search bar are tabs for All, Majors, Minors, Courses (which is selected), and Information. At the bottom, there are filters for College/School and a note that 10000 Results found.



Ok fine...this'll do

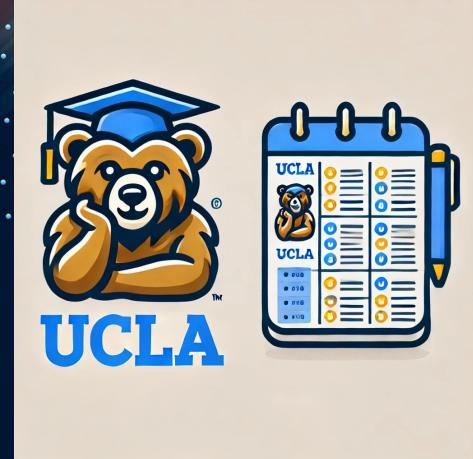


BruinPlanner

PROJECT OVERVIEW

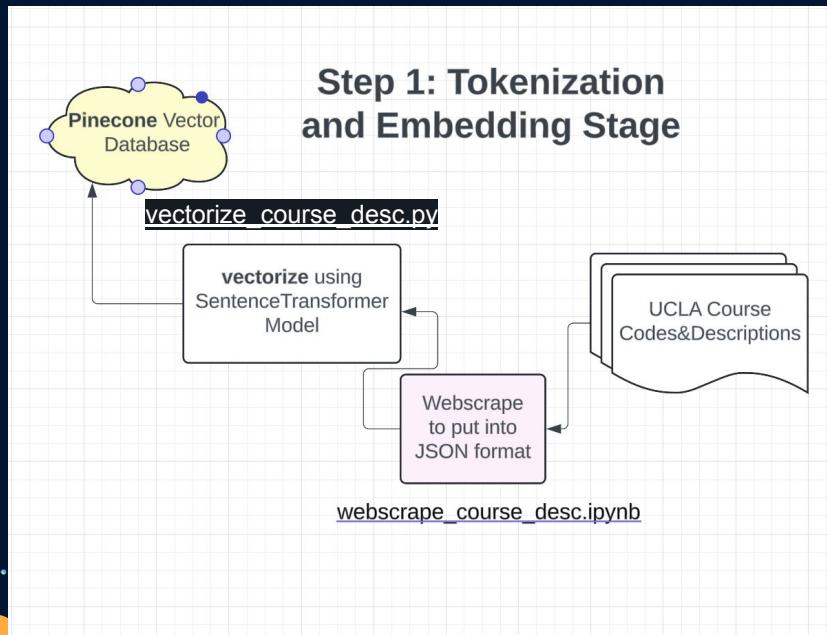
COURSE PLANNER CHATBOT- BruinPlanner

- Input your major and a sentence about your interests
- We output courses and GEs that match your interests and apply to your degree requirements
- Take classes that interest you!!!
So that you look like this in class



BruinPlanner

Creating our Database (1)

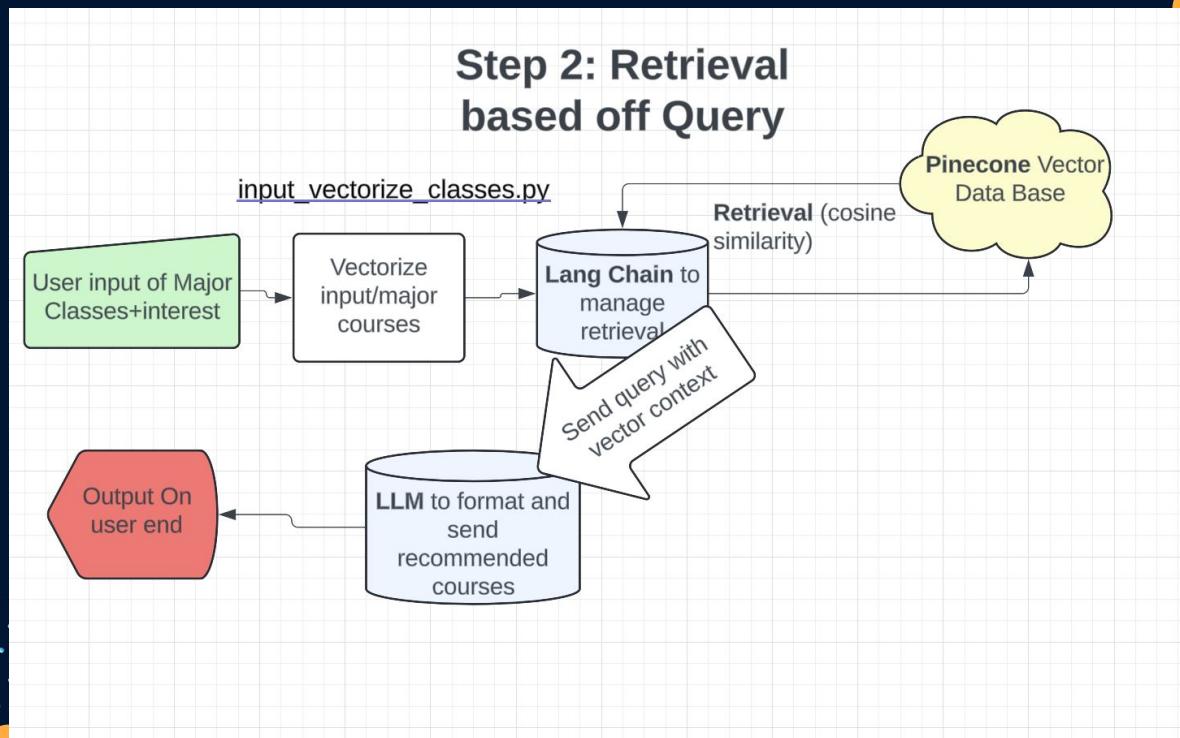


Course Desc Vector DB

	ID	VALUES	
1	STATS 188SA	-0.0136837019, -0.054542426, -0.0525...	
	SCORE	0.0339	
	METADATA		
	category:	"Statistics Elective"	
	description:	"Description: Tutorial, to be arranged. Enforced corequisite: Honors Collegium 1..."	
	sequence:	"None"	
2	MATH 31B	-0.0979411453, -0.0143987751, -0.01864...	
	SCORE	0.0222	
	METADATA		
	category:	"Required Mathematics"	
	description:	"Description: Lecture, three hours; discussion, one hour. Requisite: course 31A ..."	
	sequence:	"None"	



Retrieval Based on Input (2)



Vectorized Inputs for Similarity Matching

ID	VALUES	
1	i like hard math -0.0282162242, -0.0414193794, -0.0231516864, 0.00686302176, -0.0925763249, ...	
SCORE 0.9996	METADATA category: "user_input"	
2	ID VALUES MATH 151A -0.0280138217, 0.0891988054, -0.0487038158, -0.0508943424, -0.136488199, -0....	
SCORE 0.5603	METADATA category: "requirements_with_options"	
3	ID VALUES MATH 182 0.00745979976, 0.0380325615, -0.0577021129, -0.00715054572, -0.127301425, -0...	
SCORE 0.5543	METADATA category: "requirements_with_options"	



DEMO



BruinPlanner

A Data Theory major interested in Neural Networks....

Enter your sentence of interest: I am interested in Neural Networks

Searching for courses...

Recommended Courses:

Course ID: math-156

Similarity Score: 0.4522

Description: Lecture, three hours; discussion, one hour. Requisites: courses 115A, 164, 170A or 170E or Statistics 100A, and Computer Science 31 or Program in Computing 10A. Strongly recommended requisite: Program in Computing 16A or Statistics 21. Introductory course on mathematical models for pattern recognition and machine learning. Topics include parametric and nonparametric probability distributions, curse of dimensionality, correlation analysis and dimensionality reduction, and concepts of decision theory. Advanced machine learning and pattern recognition problems, including data classification and clustering, regression, kernel methods, artificial neural networks, hidden Markov models, and Markov random fields. Projects in MATLAB to be part of final project presented in class. P/NP or letter grading.

Units: 4.0

Course ID: stats-c161

Similarity Score: 0.4243

Description: Lecture, three hours. Requisites: course 100B, Mathematics 33A. Introduction to pattern analysis and machine intelligence designed for advanced undergraduate and graduate students. Concurrently scheduled with course C261. P/NP or letter grading.

Units: 4.0

Course ID: math-168

Similarity Score: 0.4156

Description: Lecture, three hours; discussion, one hour. Requisites: courses 115A, 170E (or 170A or Electrical and Computer Engineering 131A or Statistics 100A). Introduction to network science (including theory, computation, and applications), which can be used to study complex systems of interacting agents. Study of networks in technology, social, information, biological, and mathematics involving basic structural features of networks, generative models of networks, network summary statistics, centrality, random graphs, clustering, and dynamical processes on networks. Introduction to advanced topics as time permits. P/NP or letter grading.

Units: 4.0

Course ID: math-182

Similarity Score: 0.3484

Description: Lecture, three hours; discussion, one hour. Requisite: course 3C or 32A, and 61. Not open for credit to students with credit for Computer Science 180. Graphs, greedy algorithms, divide and conquer algorithms, dynamic programming, network flow. Emphasis on designing efficient algorithms useful in diverse areas such as bioinformatics and allocation of resources. P/NP or letter grading.

Units: 4.0



A Data Theory major interested in Markov Chains...

Enter your sentence of interest: I am interested in Markov Chains

Searching for courses...

Recommended Courses:

Course ID: math-171

Similarity Score: 0.5962

Description: Lecture, three hours; discussion, one hour. Requisites: courses 33A, 170E (or 170A or Statistics 100A). Discrete Markov chains, continuous-time Markov chains, renewal theory. P/NP or letter grading.

Units: 4.0

Course ID: stats-102c

Similarity Score: 0.4781

Description: Lecture, three hours; discussion, one hour. Requisite: course 100B. Introduction to Markov chain Monte Carlo (MCMC) algorithms for scientific computing. Generation of random numbers from specific distribution. Rejection and importance sampling and its role in MCMC. Markov chain theory and convergence properties. Metropolis and Gibbs sampling algorithms. Extensions as simulated tempering. Theoretical understanding of methods and their implementation in concrete computational problems. P/NP or letter grading.

Units: 4.0

Course ID: math-168

Similarity Score: 0.4193

Description: Lecture, three hours; discussion, one hour. Requisites: courses 115A, 170E (or 170A or Electrical and Computer Engineering 131A or Statistics 100A). Introduction to network science (including theory, computation, and applications), which can be used to study complex systems of interacting agents. Study of networks in technology, social, information, biological, and mathematics involving basic structural features of networks, generative models of networks, network summary statistics, centrality, random graphs, clustering, and dynamical processes on networks. Introduction to advanced topics as time permits. P/NP or letter grading.

Units: 4.0

Course ID: math-42

Similarity Score: 0.3991

Description: Lecture, three hours; discussion, one hour. Requisites: courses 31A, 31B, 32A, 32B, 33A, one statistics course from Statistics 10, 12, 13, one programming course from Computer Science 31, Program in Computing 10A, Statistics 20. Introduction to data-driven mathematical modeling combining data analysis with mechanistic modeling of phenomena from various applications. Topics include model formulation, data visualization, nondimensionalization and order-of-magnitude physics, introduction to discrete and continuous dynamical systems, and introduction to discrete and continuous stochastic models. Examples drawn from many fields and practice problems from Mathematical Contest in Modeling. P/NP or letter grading

.

Units: 4.0



Some more sample questions we asked

Question: Why should i take stats 100A if I'm interested in statistics?

Certainly! Taking **STATS 100A** is a valuable choice if you are interested in statistics for several reasons:

1. Foundational Knowledge

- **Probability Theory:** STATS 100A provides a solid foundation in probability theory, which is essential for understanding statistical methods. Probability theory forms the backbone of statistical inference, hypothesis testing, and modeling.
- **Mathematical Rigor:** The course is mathematically rigorous, helping you develop a deep understanding of the concepts and techniques used in statistics.

2. Prerequisites for Advanced Courses

- **Sequence Requisites:** STATS 100A is a prerequisite for several advanced courses, such as **STATS 100B** (Introduction to Mathematical Statistics) and **MATH 170S** (Introduction to Statistics). These courses build on the foundational knowledge gained in STATS 100A.
- **Capstone Preparation:** If you are interested in the capstone course **MATH M148** (Data Science Projects), taking STATS 100A will prepare you with the necessary statistical background to tackle real-world data science problems.

3. Career and Research Opportunities

- **Data Science and Analytics:** A strong understanding of probability and statistics is crucial for careers in data science, machine learning, and quantitative analysis. Many industries, including finance, healthcare, and technology, value candidates with a strong statistical background.
- **Academic Research:** If you are considering further academic pursuits, such as a master's or Ph.D. program in statistics, data science, or related fields, a strong foundation in probability and statistics is essential.

4. Practical Applications

- **Real-World Problems:** The skills and knowledge gained in STATS 100A can be directly applied to real-world problems. Whether you are analyzing data for a project, conducting research, or working in an industry, a solid understanding of probability and statistics will be invaluable.
- **Statistical Software:** While STATS 100A focuses on theoretical aspects, the concepts learned can be applied using statistical software, enhancing your practical skills.

5. Complementary Skills

- **Interdisciplinary Applications:** Statistics is a versatile field with applications in various disciplines, including economics, psychology, biology, and engineering. Taking STATS 100A will complement your studies in other areas and broaden

Question: I have taken stats 10 and stats100A and am interested in neural networks. What class should I take?

Given your background (having taken STATS 10 and STATS 100A) and your interest in neural networks, MATH 156 is the most suitable course for you. Here's a detailed breakdown of why MATH 156 is a good fit and what you can expect:

MATH 156: Introduction to Mathematical Models for Pattern Recognition and Machine Learning

Prerequisites:

- **Requisites:** Courses 115A, 164, 170A or 170E or Statistics 100A, and Computer Science 31 or Program in Computing 10A.
- **Strongly Recommended:** Program in Computing 16A or Statistics 21.

Key Points:

1. Content:

- **Probability Distributions:** You will learn about both parametric and nonparametric probability distributions, which are fundamental in understanding the data you work with.
- **Dimensionality:** The course covers the curse of dimensionality, correlation analysis, and dimensionality reduction techniques, all of which are crucial for handling high-dimensional data.
- **Decision Theory:** Concepts of decision theory are introduced, which will help you make informed decisions based on data.
- **Machine Learning and Pattern Recognition:** Advanced topics include data classification and clustering, regression, kernel methods, artificial neural networks, hidden Markov models, and Markov random fields.

2. Neural Networks:

- **Artificial Neural Networks (ANNs):** One of the key topics in this course is artificial neural networks, which aligns perfectly with your interest in neural networks. You will gain a solid understanding of the theory and practical applications of ANNs.
- **Kernel Methods:** These are closely related to neural networks and provide a different approach to learning from data, which can be very useful.

3. Projects:

- **MATLAB Projects:** You will work on projects using MATLAB, which is a powerful tool for implementing and experimenting with machine learning algorithms, including neural networks.

4. Preparation:

- **Requisites:** You have already taken STATS 10 and STATS 100A, which are part of the prerequisites for this course. Additionally, if you have taken or are planning to take Program in Computing 10A, you will meet the programming prerequisite.

Where we hope to go from here?



Upload courses across all majors



Create a user interface



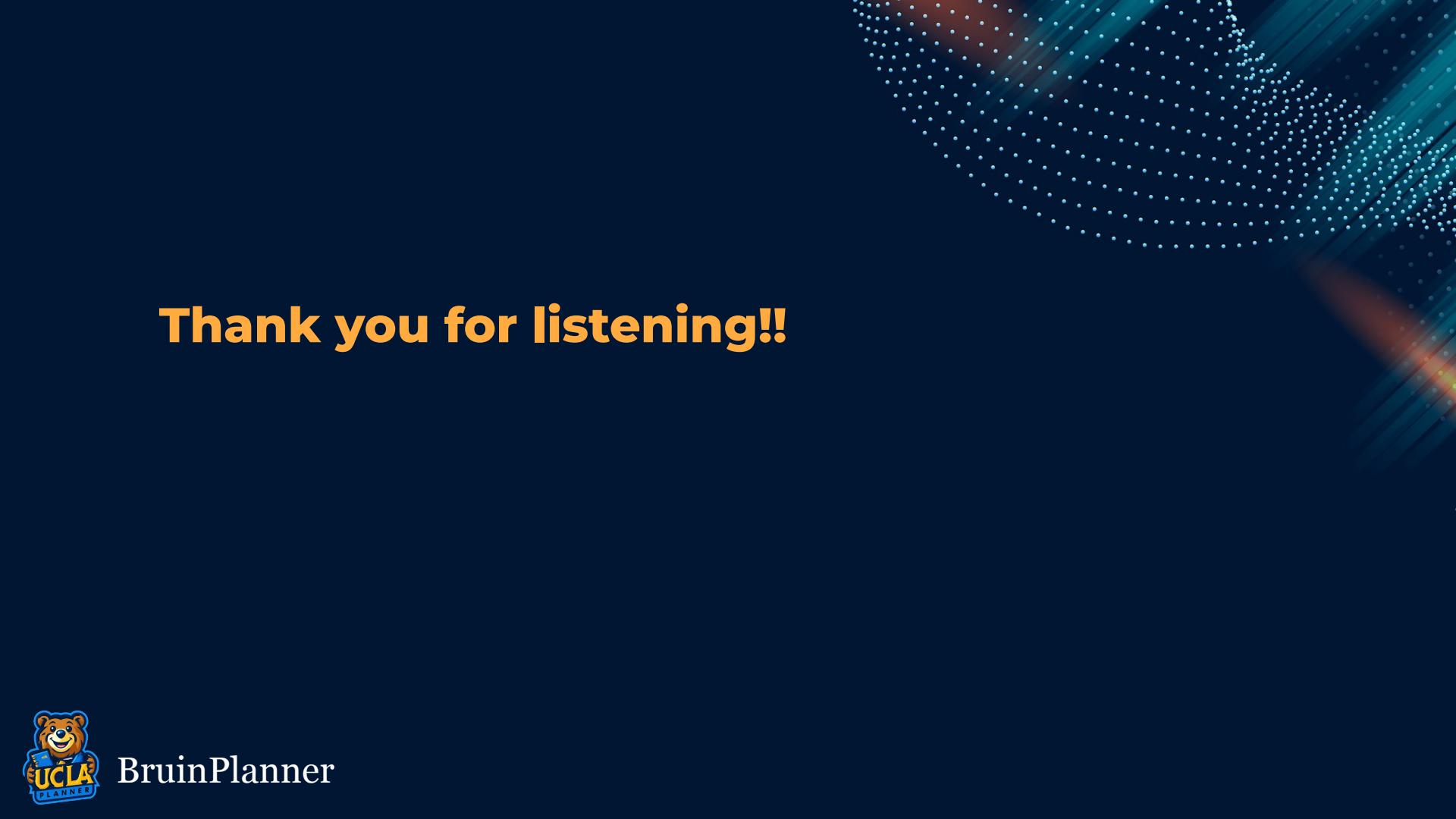
Add in GE classes



Include grade distributions from BruinWalk



BruinPlanner



Thank you for listening!!



BruinPlanner