

DATA SCIENCE (DATA)

DATA 100 Critical Questions in Data Science (4 Credits)

An exploration of significant questions and concepts, beliefs and creative visions, theories and discoveries in Data Science for first-year students. Although topics vary, the courses also seek to improve students' communication skills beyond the written word.

College Curriculum: C100

DATA 101 Reasoning with Generative AI (3 Credits)

Reasoning with Generative AI examines how modern generative systems can be used not merely to produce text or images, but to support disciplined reasoning grounded in evidence. The course begins with the premise that useful conclusions still depend on careful observation, structured information, and transparent inference. Students learn how to move from raw inputs - textbooks, news articles, images, data - to reasoned outputs by combining traditional analytic thinking with the capabilities of contemporary generative AI tools. Throughout the semester, students practice translating observations into prompts, testing assumptions, validating model outputs, and iterating toward defensible conclusions. By the end, students understand how data is represented, how generative models process that data, and how to use these tools responsibly to augment their own reasoning.

DATA 150 First-Year Seminar (4 Credits)

An exploration of a specific topic in Data Science. A grade of C- or better fulfills the COLL 150 requirement. Although topics vary, the courses emphasize academic writing skills, reading and analysis of texts, and discussion.

College Curriculum: C150

DATA 201 Introduction to Data Science & AI (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. Data in the Age of AI prepares students to design credible analyses in a world where information is often incomplete, algorithmically mediated, or partially generated. The course introduces major AI approaches, explains why they emerged, and clarifies what kinds of questions each can answer. Students learn to judge quality, select appropriate strategies, and recognize when generative tools can extend an analysis - and when they introduce new risks or distortions. The emphasis is on building knowledge that remain transparent and defensible even when AI systems shape the data itself. All students gain the ability to navigate an environment where sound conclusions require not just reasoning with AI, but understanding how different AI modeling frameworks behave, fail, and complement one another.

Prerequisite(s): CSCI 140 or CSCI 141 or DATA 141 or DATA 101 or AAI 101

Additional fees apply

Domain: NQR

DATA 202 AI & Data Ethics (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course provides an introduction to critical, ethical, and moral issues surrounding data, AI, and society. It blends social and historical perspectives on data and algorithmic systems with ethics, policy, and case examples—from text analytics and generative models to autonomous decision-making and self-driving cars—to help students develop a workable understanding of current ethical and moral issues in data science and artificial intelligence. The course examines the ethics and morality of studying human subjects, documenting workflows, designing and deploying AI systems, and communicating results. Students debate issues surrounding privacy, surveillance, discrimination, transparency, responsibility, and trust throughout the intertwined data and AI lifecycle—from collection and creation to model training and evaluation to storage, analysis, application, and the sharing of data and model outputs.

Additional fees apply

College Curriculum: C200

Domain: CSI

DATA 209 Applied Linear Algebra & Calculus (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course introduces core mathematical foundations of artificial intelligence and computing. Topics include linear algebra, matrix decompositions, calculus, and differential equations with direct application to machine learning, optimization, and simulation. Students will explore how mathematical constructs such as dot products, eigenvectors, gradients, and transformations underpin AI systems. Concepts are taught with a strong focus on real-world implementation, using computation and visualization to build intuition over symbolic manipulation.

Prerequisite(s): DATA 101 or CSCI 140 or CSCI 141

DATA 210 Research Design and Statistics (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course introduces students to the principles of empirical inquiry, statistical modeling, and computational experimentation in data-rich and AI-supported environments. Leveraging modern tools such as Python and generative AI, students will explore how to design studies, quantify uncertainty, and perform model-driven inference using techniques foundational to machine learning, reinforcement learning, and causal reasoning. Topics include research design, simulation-based inference, likelihood-based estimation, confidence intervals, model comparison (Akaike Information Criterion [AIC] and Bayesian Information Criterion [BIC]), Bayesian updating, and the use of upper confidence bounds for decision-making. While formal mathematical prerequisites are not required, students will be expected to engage in rigorous reasoning with the support of AI-assisted coding and explanation tools. The course emphasizes reproducible workflows and theory-grounded applications.

Prerequisite(s): DATA 101 or CSCI 140 or CSCI 141

DATA 301 Applied Machine Learning (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course will focus on the technical application of machine learning algorithms, their nature, and discussions regarding the potential drawbacks and advantages of different classes of algorithms. Students entering into this course should have, at a minimum, a background in python and linear algebra. No single algorithm will be covered in great depth, and the course will place a focus on the code and implementation choices necessary for each class of algorithm. Topics covered will include probability, distributions, Monte-Carlo simulations, reinforcement learning, association rules, nonlinear regression, support vector machines, kernel SVM, variable/model selection, diagnostics for regression and classification, neural networks/deep learning, natural language processing, and various associated approaches. Formerly: DATA 310

Prerequisite(s): DATA 201 and (DATA 209 or MATH 211 or MATH 109) and (DATA 210 or MATH 352)

DATA 302 Databases (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. In this course students will learn about relational database design and SQL programming using Python and SQLite. Throughout the course students will demonstrate their proficiency by creating relational databases based on both real and synthetic data, and querying and updating their data using SQL. An emphasis will be placed on leveraging data validation techniques provided by SQL databases. Formerly: DATA 311

Prerequisite(s): DATA 201

DATA 303 Data Visualization (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course provides an overview of data visualization theory, helping students understand how to produce meaningful and interpretable figures from large sets of data. You will develop the capability to select between different approaches for visualization, and learn how to leverage visualizations to identify the best outcome for a given challenge. In addition, this course provides students with the technical skills to produce their own visualizations using industry-standard software, as well as by leveraging Artificial Intelligence tools.

College Curriculum: C200

Domain: NQR

DATA 305 Designing AI Agents (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. DATA 305 explores how generative AI can be used as a tool for structured, multi-step problem solving. Students will learn to design and evaluate AI-augmented workflows, develop prompt pipelines for complex tasks, and work with tools such as retrieval-augmented generation (RAG), multi-agent prompting, and function-calling APIs. Case studies will include applications in coding, data transformation, synthesis, planning, and decision-making. The course also engages with ethical challenges, including hallucination, bias, and the limits of autonomous reasoning. By the end of the course, students will be able to prototype and deploy agentic generative systems to support creative and analytical work across domains.

Prerequisite(s): DATA 201

DATA 340 Special Topics in Data Application (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. Prerequisite(s): To be determined by topic each term. Selected topics in data science that count toward the Data Applications track. The topic to be considered will be announced prior to the beginning of the semester. A student should only be able to take this course a second time if it is covering a separate topic than other times the student has enrolled. Course may count towards other Data Science tracks if approved by the instructor. A student should only be able to take this course a second time if it is covering a separate topic than other times the student has enrolled.

DATA 341 App Time Series Analysis (3 Credits)

Students will learn about and discuss relevant topics and research associated with time series analysis. The course will place a focus on the code and implementation choices necessary to perform applied time series analysis. Throughout the semester students shall replicate several time-series studies and provide replication code and analyses as part of their lab assignments. Students must apply time series skills learned throughout the course to answer their own research questions. Students shall brief progress on their projects throughout the semester as well as the final project and results during the last couple weeks of the semester. Prerequisite(s): DATA 201

DATA 390 Directed Research in Data Science (1-4 Credits)

This course is designed to permit students with a focus in Data Science to engage in directed research after completing core coursework in Data Science. Working closely with a program faculty member as an advisor, each student will conduct a substantial research project focusing on synthesis and critical analysis, to solve problems in an applied and/or academic setting, to create original material or original scholarship, and to communicate effectively with a diversity of audiences. May be repeated for credit.

DATA 405 Trustworthy AI (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course explores the ethical, legal, and technical dimensions of building trustworthy artificial intelligence systems. Topics include fairness, accountability, transparency, explainability, robustness, privacy-preserving machine learning, and algorithmic bias. Students will critically assess AI models and practices through real-world case studies and regulatory frameworks. The course prepares students to develop AI systems that are not only performant but also socially responsible and aligned with stakeholder values.

Prerequisite(s): DATA 101 or CSCI 140 or CSCI 141

Additional fees apply

DATA 415 Agentic Coordination (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course examines the design, implementation, and deployment of teams of AI agents powered by large language models and related technologies. Students will learn how modern AI agents reason, plan, use tools, retrieve information, and interact with external systems. Topics include advanced prompt engineering, tool use, retrieval-augmented generation (RAG), agent architectures, memory systems, evaluation, safety, and human-in-the-loop design. The course emphasizes building practical workflows with task-oriented agents using contemporary frameworks. By the end of the semester, students will design and deploy a functional AI agent for a real-world application.

Prerequisite(s): DATA 305

- Additional fees apply
- Additional fees apply

DATA 431 Spatial Data Discovery (3 Credits)

Most, if not all, of today's grand challenges (e.g., food, water and energy security) can be described spatially from regional to global scales and, while several individual disciplines contend to address these challenges, there is one key factor that they all have in common: the need for data. Despite our being in an age rich in data, many of the critical datasets needed for our understanding and prediction of our world are, in fact, quite limited. In this capstone course, you will get the opportunity to utilize your Python programming skills (writing scripts and creating subroutines) to connect to various types of data (e.g., GeoJSON, ASC, HDF5, and NetCDF), synthesize these data to unlock new understanding (using methods such as spatial scaling and gap-filling), create visualizations using open-source GIS software, and present to the world your own story of spatial data discovery professionally written for the web. I will take you through my own journey of data discovery, provide you with the know-how for accessing large data repositories, demonstrate methods for data harmonizing, processing, modeling and visualizing, and challenge you to think spatially.

Prerequisite(s): DATA 201 and GIS 201

College Curriculum: C400

DATA 440 Special Topics (1-4 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. Selected topics in Data Science. Course is designed to meet the capstone requirement. The topics to be considered will be announced prior to the beginning of the semester. Instructors may require prior approval for registration. Students enrolled in this section will complete a project to fulfill the capstone requirement. Course may count towards other Data Science tracks if approved by the instructor.

Prerequisite(s): DATA 141 or CSCI 140 or CSCI 141

College Curriculum: C400

DATA 441 Advanced Applications of AI (3 Credits)

This course focuses on creating a compelling story of data discovery where students will develop skills to access, read and analyze disparate data sources from a variety of open-access international, governmental, and private organizational databases and will learn about the challenges associated with real data. To tackle these issues, students will use a variety of advanced analytical and machine learning methods including variable selection, nonparametric regression, functional data analysis and deep learning. Students will communicate their findings through the web by creating a data science paper in HTML format that engages a broad audience to understand the data being addressed, the research methods applied, and the results obtained.

Prerequisite(s): (DATA 301 or DATA 310) and (MATH 211 or MATH 109)

College Curriculum: C400

DATA 442 Neural Networks & Deep Learning (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course teaches the foundations of Neural Networks and Deep Learning. Students entering into this course should have, at minimum, a background in data preprocessing, cleaning, manipulation, and dimensionality reduction within python. Through an applied learning project, you will learn how to implement a machine learning project from design to implementation in the context of neural networks. Topics we will cover include the basic building blocks of neural networks, RNNs, convolutional networks and computer vision, backpropagation basics and strategies (including inductive transfer approaches), differences between technical implementations (i.e., TensorFlow, Keras, Torch), and more.

Prerequisite(s): DATA 301 or DATA 310

College Curriculum: C400

DATA 444 Agent-Based Modeling (3 Credits)

In this course, students will use openly accessible, global, near present-time, high-resolution satellite, household survey and CDR data, with machine learning and spatial statistics methodologies to construct agent-based models of human development processes. Each student will select and describe an administrative subdivision, its demographics, and its built and natural environments in order to estimate social and economic, complex and adapting, agent-based decision, movement and land use models. Students will construct modules that project demand for infrastructure (transportation, water, and electricity) and social services (health care, education, and public safety) as well as simulate an infectious disease outbreak, a natural disaster and unabated urbanization. The programming languages python, java, and R will be used in this course.

Prerequisite(s): DATA 301 or DATA 310

College Curriculum: C400

DATA 445 Network Analysis (3 Credits)

In this course, we will survey a broad range of fundamental topics in network science, relevant to students from data/computer science and engineering, informatics, business, biology, physics, statistics, social sciences, etc. For example, we will explore the properties of social networks and the key role of hubs, and how directed and weighted networks affect the spread of information and misinformation in social media. These topics are important and useful in many job sectors from marketing to technology, management to design, and from biology to the arts and humanities.

Prerequisite(s): (DATA 301 (may be taken concurrently) or DATA 310) or MATH 351

College Curriculum: C400

DATA 446 Generative AI (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course offers an in-depth exploration of Generative Artificial Intelligence (AI), a branch of AI focused on creating models that can generate new content, such as images, text, and sounds, mimicking human-like creativity. The curriculum is designed to explore the foundational theories and practical applications of generative models, including but not limited to Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Transformer models. Throughout the course, students will: 1. Understand the Core Principles of Generative AI: Grasp the theoretical underpinnings and algorithms that enable machines to generate novel, realistic content. 2. Explore Different Generative Models: Learn about various generative architectures, including GANs, VAEs, and transformers, understanding their unique characteristics and applications. 3. Apply Generative AI in Practice: Engage in hands-on projects that involve training, tuning, and deploying generative models to produce creative content across different domains, such as art, music, and natural language generation. Programming assignments and projects will be carried out in Python, leveraging popular machine learning libraries like TensorFlow and PyTorch. By the end of this course, students will be proficient in designing and implementing generative AI models, capable of producing innovative and complex outputs that reflect aspects of human creativity.

Prerequisite(s): DATA 301 or DATA 310

DATA 447 Bayesian Reasoning in Data Science (3 Credits)

No data scientist can work without a solid grasp of conditional probability and Bayesian reasoning. Bayes' theorem allows to update our beliefs based on the occurrence of new events, steering the inference towards the truth and assessing uncertainty in predictions. This course provides an introduction to Bayesian reasoning in data science and will let you appreciate the basic building blocks of this approach through real-world examples across different areas. During the course you will learn concrete computational implementations, that will help students connect what they have read and heard with what they can program, reinforcing the material.

Prerequisite(s): DATA 301 or DATA 310

DATA 448 Reinforcement Learning (3 Credits)

A \$250 per credit fee will apply for students for whom initial enrollment at W&M begins in Fall 2025 or later. This course introduces the fundamentals of reinforcement learning (RL), a type of machine learning paradigm where an agent learns to make decisions by interacting with an environment. The course will cover Markov decision processes, reinforcement learning, planning, and function approximation. By the end of this course, the students will be able to (1) understand the fundamentals concepts of an RL problem, (2) know how to formalize a problem as a RL problem, and (3) learn classic RL algorithms such as Q-learning and policy gradient. There will be several hands-on projects throughout the course. Programming will be done in the Python language. By the end of the course, the students should be able to implement classical RL algorithms such as Q-learning and policy gradient methods.

Prerequisite(s): DATA 301 or DATA 310

DATA 449 Web Science (3 Credits)

Provides an overview of the World Wide Web and associated decentralized information structures, focusing mainly on the computing aspects of the Web: how it works, how it is used, and how it can be analyzed. Students will examine a number of topics including: web architecture, web characterization and analysis, web archiving, Web 2.0, social networks, (creating/detecting) social bots, collective intelligence, search engines, web mining, information diffusion on the web, Semantic Web, and more.

Prerequisite(s): DATA 301 or DATA 310

College Curriculum: C400

DATA 451 Artificial Intelligence Systems (3 Credits)

This course introduces undergraduate students to the foundational principles of automation in artificial intelligence and data science, with an emphasis on building repeatable, scalable, and production-ready workflows. Topics include orchestration tools, automated data pipelines, data quality assurance, task scheduling, and system monitoring. Students will prototype robust systems that integrate data ingestion, preprocessing, model training, and reporting. The course emphasizes the transition from theory to real-world AI deployment.

Prerequisite(s): DATA 301

DATA 460 Automation & Workflows (3 Credits)

Real-world data analyses are iterative: errors surface, new data arrive, others need to reproduce results, and many organizations depend on recurring or continuous reporting. These realities demand automation and well-documented workflows, especially in team settings, and require deliberate design choices that balance abstraction and generalization with clarity and usability. In this course, students complete projects centered on reproducibility, efficiency, and maintainability, using real or synthetic data and, when appropriate, their own datasets. They develop program prototypes and documentation for peer testing and critique, culminating in a final project that includes written documentation and a working demonstration.

Prerequisite(s): DATA 302 and (DATA 303 or DATA 301)

Additional fees apply

College Curriculum: C400

DATA 490 Independent Research in Data Science (1-4 Credits)

This course is designed to permit students with a focus in Data Science to engage in independent research. Working closely with a program faculty member as an advisor, each student will conduct a substantial research project focusing on synthesis and critical analysis, to solve problems in an applied and/or academic setting, to create original material or original scholarship, and to communicate effectively with a diversity of audiences. May be repeated for credit.

College Curriculum: C400

DATA 491 Mentored Data Science Teaching (1 Credit)

Student must have achieved an A or above in the class in which they will assist, or be otherwise approved to enroll. Approval must be given by the instructor. This course provides a mentored experience in Data Science teaching through short discussion sessions and selected readings as may be appropriate. Teaching skills will be developed by assisting in approved Data Science classes. May be repeated for credit.

DATA 495 Honors - Data Science (3 Credits)

Honors thesis research conducted by Data Science majors under the supervision of a faculty member. Students enrolled in this course sequence (DATA 495 and DATA 496) are required to write an honors thesis based on a review of the literature and the findings from their own research project. Students are expected, in most cases, to enroll in DATA 495 during the fall semester, and DATA 496 in the spring semester. For College provisions governing the Admission to Honors, see catalog section titled Honors and Special Programs.

College Curriculum: C400

DATA 496 Honors - Data Science (3 Credits)

Honors thesis research conducted by Data Science majors under the supervision of a faculty member. Students enrolled in this course sequence (DATA 495 and DATA 496) are required to write an honors thesis based on a review of the literature and the findings from their own research project. Students are expected, in most cases, to enroll in DATA 495 during the fall semester, and DATA 496 in the spring semester. For College provisions governing the Admission to Honors, see catalog section titled Honors and Special Programs.

College Curriculum: C400

DATA 498 Internship (1-6 Credits)

This course is designed to allow students to gain first-hand knowledge through practical experience in real-world settings. Students will be supervised by and will meet regularly with department faculty members during the course of their internship. Students wishing to receive academic credit for an internship program must request and obtain faculty approval prior to participation in the program. This course is designated as pass/fail only.