DATA SCIENCE (DATA)

DATA 601 Computational Foundations of Data Science (3 Credits)
This course introduces core concepts in optimization and advanced programming. Students will explore techniques such as gradient descent, convex and nonconvex optimization, and constraint-based modeling. Advanced programming topics may include software engineering best practices, debugging large-scale projects, and using libraries for scientific computation and deployment of large-scale AI models. This course deepens students' ability to implement efficient, scalable, and interpretable data science solutions.

DATA 602 Mathematical Foundations of Data Science (3 Credits)
No data scientist can work without a solid grasp of conditional probability and Bayesian reasoning. Bayes' theorem allows us to update our beliefs based on the occurrence of new events, steering the inference towards the truth and assessing uncertainty in predictions. This course offers an in-depth examination of Bayesian reasoning in data science, allowing you to grasp the fundamental components of this approach through practical examples spanning various domains. Throughout the course, you will acquire hands-on experience with concrete computational implementations, enabling students to bridge the gap between theoretical concepts and their programming applications, thereby solidifying their understanding of the material.

DATA 603 Data Ethics and Trustworthy AI (3 Credits)
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

DATA 610 Neural Networks and Deep Learning (3 Credits)
This course covers deep learning and neural networks, offering a comprehensive exploration of both foundational concepts and cutting-edge advancements in the field. Designed for students with a solid background in machine learning, this course aims to equip students with the advanced skills needed to develop, understand, explain, analyze, and implement neural network models that can tackle complex real-world problems. Throughout the course, participants will engage with a wide array of topics, including but not limited to, the architecture of deep neural networks, convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), explainability techniques, and generative adversarial networks (GANs). Special emphasis will be placed on understanding the theoretical underpinnings of these models, their optimization techniques, and how they can be applied to diverse areas such as image and speech recognition, natural language processing, and autonomous systems. Cross-listed with DATA 442.

DATA 611 Hardware and Software Systems for AI and Big Data (3 Credits)
This course introduces students to the 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 develop robust systems that integrate data ingestion, preprocessing, model training, and reporting. The course emphasizes operational efficiency, reproducibility, and the transition from research to real-world AI deployment.

DATA 612 Generative AI (3 Credits)
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 provide you with a deep background in the mathematics and computational techniques on which generative approaches are predicated, including but not limited to Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Transformer models. Throughout the course, students will: 1. Study the functional frameworks which underpin generative models: You will understand, and be able to derive as may be appropriate, equations inter-related with generative approaches (drawn from Bayesian studies and theory regarding diffusion models, variational autoencoders, and transformers, among others). 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. 4. Implement your own class of Generative AI: Building on existing models, you will implement a new architecture for an application area using Generative AI. 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.

DATA 620 Network Analysis (3 Credits)
Networks are everywhere in our lives: networks of friends on social media, the Web, networks of neurons in our brains, etc. It's amazing that such a simple representation - dots and lines - can capture a variety of relationships, whether simple or complex. 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. Cross-listed with DATA 445.

DATA 623 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, etc. Students will engage in research projects that require the development and application of novel algorithms for analyzing and interpreting web data, emphasizing the creation of new knowledge in web dynamics and information dissemination. Additionally, the course will include a critical examination of current scholarly articles and participation in discussions that explore the ethical, privacy, and security challenges associated with web technologies and data analytics. Cross-listed with DATA 449 Pre-requisite: A strong background in python is required.

DATA 631 Reinforcement Learning (3 Credits)
This course introduces the fundamentals of reinforcement learning (RL) and its applications in various domains. The students will be able to (1) understand the theoretical foundations of RL problems, (2) know how to formalize a problem as a RL problem, (3) understand a spectrum of existing RL algorithms such as Q-learning and policy gradient, and (4) how to implement a RL algorithm to the target problem of interest. 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 and apply the RL algorithms to solve example real-world problems. Cross-listed with DATA 448.

DATA 645 Graph Learning (3 Credits)
This course investigates modern machine learning methods for graph-structured data. Topics include node and edge classification, graph convolutional networks (GCNs), attention mechanisms on graphs, and unsupervised graph representation learning. Students will work with real datasets involving social networks, biological pathways, or transportation systems. The course emphasizes designing graph learning pipelines that capture both local and global structure, with applications in recommendation systems, fraud detection, and knowledge graph completion.

DATA 670 Photogrammetric and Spatial Data Analysis (3 Credits)
This course aims to provide students with a comprehensive understanding of the core concepts, principles, and practices in photogrammetric and spatial data analysis. The course consists of two primary components. The first component will focus on theoretical readings in GIS, Remote Sensing, and Spatial Statistics, where students will explore the conceptual foundations of these fields and analyze their intersections. The second component will provide hands-on training in implementing spatial data analytics in High-Performance Computing (HPC) environments. Students will gain practical experience in developing, testing, and deploying large-scale spatial analysis applications in parallel computing architectures. By the end of the course, students will have a strong foundation in the key theoretical and practical aspects of photogrammetry and spatial data analysis.

DATA 691 Topics in Data Science (3 Credits)
The course will cover a broad range of topics in Data Science, and the specific subject and text will be selected by the instructor. This course is intended to provide students with an in-depth understanding of the latest trends and technologies in the field of Data Science. Throughout the course, students will be required to engage in independent research and critical analysis of the selected subject matter. The course will utilize various mediums including readings, case studies, and other resources as appropriate. Students will be expected to demonstrate their ability to synthesize and apply the knowledge gained from the course in practical applications. By the end of the course, students should have gained a deep understanding of the selected subject matter in Data Science and be able to apply the knowledge gained to both real-world scenarios and their own research. Course may be repeated for credit.

DATA 692 Readings in Data Science (0-6 Credits)
Directed Readings in Data Science is designed to promote the study of subject matter beyond topics covered in regular course offerings. It is most appropriate for students who have already exhausted offerings in the area of study that interest them. Working closely with a faculty advisor, students will study, through readings and other mediums as may be appropriate, a Data Science related topic selected by the instructor. Students will be expected to illustrate their understanding of the material, as well as ability to synthesize and critically discuss key topics presented. May be repeated for credit.

DATA 696 Data Science Capstone (3 Credits)
The capstone project aims to provide students an opportunity to integrate and apply core knowledge and skill components in Data Science that were acquired during the program in a complex real-world project driven setting. The problems we will address in this course will be related to the general theme of Data Science, and will therefore include extensive use of various big data sources and machine learning tools. In this environment students will be required to define possible scenarios, identify key challenges, explore possible solutions and deliver an effective solution. Given its nature, a capstone project often goes beyond a single discipline and require the application of varied disciplines to the solution of a single large-scale problem.

DATA 700 Master's Thesis (1-12 Credits)
This course provides the structure and academic credit for thesis research. Under the supervision of a faculty advisor, students will conduct an original, independent investigation on a topic of relevance to data science or its application in a domain discipline. The project must demonstrate the student's ability to formulate a research question, apply appropriate data science methods, and contribute new insights to the field. The experience culminates in a written thesis and oral defense evaluated by a committee of faculty members.

DATA 706 Research Methods in Data Science (3 Credits)
This course develops advanced strategies for managing the full data lifecycle in support of reproducible, transparent, and responsible research. Students examine data planning, acquisition, documentation, cleaning, versioning, storage, governance, quality control, and long-term preservation, with emphasis on practices that allow research results to be replicated and extended. The course addresses responsible data dissemination, including open data practices, restricted access models, metadata standards, licensing, privacy protection, disclosure risk, and ethical obligations to research participants, communities, sponsors, and the public. Students apply these concepts through the design of data management and dissemination plans for complex research projects.

DATA 707 Scientific Communication in Data Science (3 Credits)
This course prepares students to translate original data science research into usable products, workflows, decisions, and communication outputs for scholarly, public, policy, and professional audiences. Students develop skills in research prospectus writing, grant writing, academic paper writing, validation, interpretation, stakeholder communication, implementation planning, and assessment of real-world use. The course includes substantial attention to visualization and visual analytics, with students developing interactive and static visual products that support interpretation, exploration, explanation, and decision-making.

DATA 751 Teaching and Professional Development in Data Science (3 Credits)
This course prepares doctoral students for teaching, mentoring, and professional practice in data science. Students will examine evidence-based teaching methods, classroom practice, assignment and assessment design, responsible use of generative AI in instruction, and strategies for explaining technical material to varied audiences. The course connects students to the intellectual life of the department through colloquium participation, structured reflection, and professional development activities. Students will leave the course with a teaching portfolio, sample instructional materials, a professional development plan, and a stronger foundation for future work as instructors, researchers, mentors, and data science leaders.

DATA 792 Readings in Data Science (3 Credits)
This course is designed to promote the study of subject matter beyond topics covered in regular course offerings. It is most appropriate for students who have already exhausted offerings in the area of study that interest them. Working closely with a faculty advisor, students will study, through readings and other mediums as may be appropriate, a Data Science related topic selected by the instructor. Students will be expected to illustrate their understanding of the material, as well as ability to synthesize and critically discuss key topics presented.

DATA 799 Directed Research (1-12 Credits)
Work on research that forms basis for doctoral dissertation.

DATA 800 Dissertation Research (0-12 Credits)
This course provides the structure and academic credit for doctoral dissertation research in Data Science. Under the mentorship of a dissertation advisor and committee, students will conduct sustained, original research that advances the theory, methodology, or application of data science in significant and demonstrable ways. The dissertation must clearly articulate a novel research question, apply or develop rigorous data science methods, and produce insights that contribute meaningfully to scholarly discourse and practice. The experience culminates in a formal dissertation and public defense, evaluated by a faculty committee in accordance with the standards of doctoral scholarship.