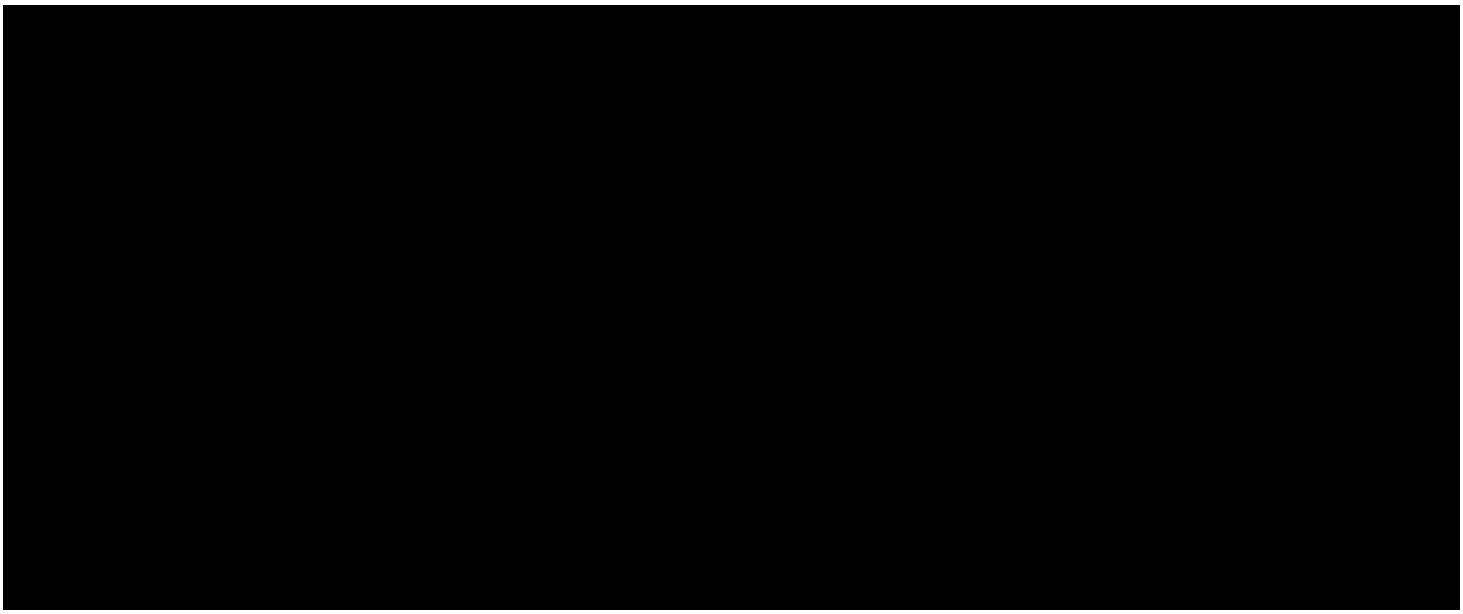
Course Syllabus

Jump to Today

DS 552/CS 552 Generative Artificial Intelligence - Syllabus



Course Overview:

Generative Artificial Intelligence (Gen-AI) is a class of machine learning models that generate new data —such as text, images, faces, voice, and artwork—that is nearly indistinguishable from data typically generated by humans. These models are trained on realistic example data sets from the real world. This course covers the underlying fundamentals of generative models and introduces the design and modeling of some of the modern generative models: Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Diffusion models, ChatGPT, and Large Language Models (LLMs) such as H2O Danube 2, Ollama, LLaMa2, BERT, and models from HuggingFace.

Several applications will be discussed, ranging from image generation for engineering or science applications to the utilization of generated data for data augmentation in AI systems. Ethical concerns related to the dangers of generative technologies—concerning issues from misinformation, bias, and data ownership—will also be reviewed.

Additionally, the course explores cutting-edge developments such as Retrieval-Augmented Generation (RAG), which combines generative models with information retrieval systems to produce contextually accurate and highly relevant responses. The course will include evaluation frameworks, including RAG evaluation (RAGA) and other metrics, to assess the performance and effectiveness of these systems.

Students will gain practical experience using tools such as Streamlit for deploying and sharing generative AI models as interactive web applications, allowing them to showcase their work in real-world

contexts.

Recommended Background:

Core artificial intelligence classes, such as machine learning and deep learning, or equivalent background is highly recommended.

Weekly Breakdown

Module 1: Introduction to Generative AI

- **Overview:** Introduction to the principles and significance of generative models in AI. Review of basic probability and statistics as they apply to generative AI.
- **Key Topics:** Generative models overview, applications in various fields, basics of probability, and statistics review.

Module 2: Generative vs. Discriminative Models

- **Overview:** Distinguishing between generative and discriminative models, understanding their respective roles and applications.
- **Key Topics:** Definitions, differences, and examples of generative and discriminative models.

Module 3: Variational Autoencoders (VAEs)

- **Overview:** Deep dive into VAEs, their architecture, and applications. Hands-on implementation using Python.
- **Key Topics:** VAE architecture, latent variable models, applications, coding a VAE.

Module 4: Generative Adversarial Networks (GANs)

- **Overview:** Exploration of GANs, including their structure, training challenges, and use cases.
- **Key Topics:** GAN architecture, adversarial training, applications, coding a GAN.

Module 5: Advanced GANs and Conditional GANs (cGANs)

- **Overview:** Study of advanced GAN techniques and conditional GANs, focusing on specific applications such as image-to-image translation.
- **Key Topics:** cGANs, advanced GAN techniques, image generation, practical implementations.

Module 6: Autoregressive Models and Flow-based Models

- **Overview:** Examination of autoregressive and flow-based models, their applications, and how they differ from other generative models.
- **Key Topics:** PixelRNN, PixelCNN, RealNVP, Glow, practical coding exercises.

Module 7: Diffusion Models

- **Overview:** Understanding diffusion models and their applications in generative AI, with a focus on recent advancements.

- **Key Topics:** Theory of diffusion models, practical applications, coding a diffusion model.

Module 8: Large Language Models (LLMs) in Generative AI

- **Overview:** Introduction to LLMs like GPT, BERT, H2O Danube 2, Ollama, and LLaMa2, with practical examples of their applications.
- **Key Topics:** LLM architecture, transformer models, practical applications using HuggingFace.

Module 9: Retrieval-Augmented Generation (RAG) and Ethical Considerations

- **Overview:** Exploration of Retrieval-Augmented Generation (RAG), which combines generative models with information retrieval systems to produce more accurate and contextually relevant outputs. Additionally, this module will cover the ethical implications of using generative AI.
- **Key Topics:** RAG architecture, integration with generative models, key applications (e.g., customer support, healthcare, legal research, content creation), RAG evaluation frameworks (RAGA) including precision, recall, relevance, and coherence, and other evaluation metrics (e.g., FID, IS). Ethical considerations such as bias, misinformation, and data ownership will also be discussed.

Module 10: Capstone Project – Deploying Generative AI with Streamlit

- **Overview:** In this capstone module, students will apply the concepts learned throughout the course to a final project. The project will involve creating and deploying a generative AI model using Streamlit, an open-source app framework that enables the sharing of machine learning models as interactive web applications.
- **Key Topics:**
  - Overview of Streamlit and its integration with Python for machine learning applications.
  - Hands-on creation of a generative AI model (e.g., VAE, GAN, LLM, or a RAG system).
  - Deploying the model using Streamlit, creating an interactive user interface.
  - Presenting the final project, discussing challenges, solutions, and potential improvements.
- **Deliverables:**
  - A fully deployed generative AI model using Streamlit.
  - A project report detailing the model's architecture, the deployment process, and the application's user interface.
  - A presentation of the project to the class, highlighting the key features and outcomes.