Carnegie Mellon University
HeinzCollege
INFORMATION SYSTEMS • PUBLIC POLICY • MANAGEMENT

Carnegie Mellon University

Introduction to AI

Course Developer: Professor David Steier

Course Description

Driven by the combination of increased access to data, computational power, and improved sensors and algorithms, artificial intelligence (AI) technologies are entering the mainstream of technological innovation. These technologies include machine learning, computer vision, natural language processing and the emerging area of generative AI.

After an introduction of some basic AI concepts and techniques, the course illustrates both the potential and current limitations of these techniques with examples from a variety of applications. We spend some time on understanding the strengths and weaknesses of human decision-making and learning, specifically in combination with AI systems and on ethical and policy implications of new AI capabilities. Exercises will include hands-on application of basic AI techniques as well as selection of appropriate technologies for a given problem and anticipation of design implications. In a final project, groups of students will participate in the creation of an AI-based application.

This is a full-semester course, planned in general around two lecture sessions per week of 1 hour 20 minutes apiece.

Learning Objectives

Students will be able to:

1. Identify problems where artificial intelligence techniques are applicable.
2. Apply selected basic AI techniques; judge applicability of more advanced techniques.
3. Participate in the design of systems that act intelligently and learn from experience.

The second part of the course will cover societal implications and governance, including public policy. Issues we will cover include:

- Economic and labor impact: Large generative models can be powerful tools to empower humans but could also automate many tasks that were previously done by humans, potentially leading to job loss and disruption in the job market.
- Information and ecosystem impact: Large generative models can be used to help moderate polarizing conversations, but they can also generate fake information (called hallucinations) and can be used to create deepfakes, poisoning the information commons.
- Data ownership impact: AI relies on training data in the form of text, images, and videos to learn patterns to understand and generate related output. However, open questions remain on how to compensate content owners and reward the creativity of content producers in the age of AI.

Class sessions and assignments will involve hands-on AI activities, including the creation of audio and video "deepfakes" and "red teaming" of models. Discussion topics will dissect AI public policy debates across the US, EU and international realms. Additionally, we will explore techniques for responsible AI development and release at organizations.

**Note: AI tools such as ChatGPT and Microsoft Copilot were used to help generate themes to cover in this course and to find readings, generate activities to support this curriculum, and help edit this syllabus for clarity.*

The course is designed for any graduate student to experience the latest AI technology and explore its societal implications. It is ideal for students on policy, business, design, data science or technology tracks.

Course Prerequisites

This course is primarily aimed at students with technical backgrounds who wish to design and develop products and services using AI. A background in basic statistics is required for the course. Students need at least a basic knowledge of Python to complete the assignments for this course. Students who have not taken 90-812 or 95-888 or have equivalent background will be required to complete supplementary work to learn Python at the beginning of the course.

Learning Objectives

The primary reference for the course is Russell, S. & Norvig, P. *Artificial Intelligence: A Modern Approach*, Pearson, 2020. This is the fourth edition of the leading textbook in AI, generally accepted as the most comprehensive reference on the subject. It is a substantial update to the third edition, so investing in the latest edition is worthwhile, especially if you plan to do further work in AI. This text will be supplemented by on-line material as listed in the course outline below.

Homework, final projects and grading

There will be 6 homework assignments:

Assignment

1) Classification and Clustering
2) Reinforcement learning
3) Computer vision
4) Natural language
5) Fairness in AI
6) Artificial General Intelligence

Assessments	Percentage of Final Grade
Homework assignments	50%
Final project presentation and report	20%
Three quizzes	30%
Class attendance and participation	20%

Course Schedule

Week #	Topics	Readings/Assignments
1	**Introduction to AI** - Introductions - Course structure and policies - What is AI (using case studies) - History of AI - Proposing and evaluating AI applications	Russell & Norvig, Chapter 1, "Introduction" in *Artificial Intelligence: A Modern Approach*, 2020 **HW #1** – Classification and Clustering
2	**Introduction to Machine Learning** - What is machine learning? - Supervised vs. unsupervised learning - Regression -- linear, logistic, ridge - Classification – decision trees, SVM, random forests - Model evaluation - Dimensionality reduction: PCA - Clustering – k-means, hierarchical clustering ○ Recitation topic: Implementing machine learning	Chapter 19: Learning from Examples" in Russell & Norvig, *Artificial Intelligence: A Modern Approach*, 2020 C. Aggarwal, Chapter 9 "Unsupervised Learning" in *Artificial Intelligence: A Textbook*, 2021, Springer Huneycutt, J., "An introduction to clustering algorithms in Python", May 29, 2018, https://towardsdatascience.com/an-introduction-to-clustering-algorithms-in-python-123438574097
3	**Machine Learning (continued)** - Semi-supervised learning - Expectation maximization - Reinforcement learning - Monte Carlo Tree Search	Chapter "Reinforcement Learning" in *Artificial Intelligence: A Modern Approach*, 2020 (optional) van Engelen, J.E., Hoos, H.H. A survey on semi-supervised learning. Mach Learn 109, 373–440 (2020). https://link.springer.com/article/10.1007/s10994-019-05855-6 **HW #2** – Reinforcement learning

4	**Deep Learning** - Neural networks and back-propagation - Convolutional neural networks - Recurrent neural networks and LSTMs - Transfer learning - Deep reinforcement learning - Recitation: Implementing deep learning in PyTorch	Russell & Norvig, Chapter 21, "Deep Learning" in *Artificial Intelligence: A Modern Approach*, 2020 (optional) 3Blue1Brown, "But what is a neural network," 2017 (20 min video) https://www.youtube.com/watch?v=aircAruvnKk&vl=en **Quiz #1**
5	**Computer Vision** - Introduction to computer vision - Image segmentation - Edge and motion detection - Object classification - Pre-trained vision models	Russell and Norvig, Chapter 25, "Computer Vision" in *Artificial Intelligence: A Modern Approach*, 2020 PyTorch Dataloader video: https://www.youtube.com/watch?v=zN49HdDxHi8 (optional) TensorFlow, "Image Recognition", July 30, 2018, https://www.tensorflow.org/tutorials/images/image_recognition **HW #3 –** Computer vision
6	**Natural language understanding** - Intro to natural language understanding - Machine translation - Sentiment analysis - Transformers - Recitation: Natural language understanding and transformers	Russell & Norvig, "Chapter 23: Natural Language Processing" and "Chapter 24: Deep Learning for Natural Language Processing," in *Artificial Intelligence: A Modern Approach*, 2020 (optional) Devlin, J. et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", 24 May 2019, https://arxiv.org/pdf/1810.04805.pdf

7	**Introduction to Large Language Models** - Large language models for natural language processing - Open vs. closed source LLMs - Vector databases - Fine tuning - Case studies: ChatGPT	Brown, T.B. et al, "Language Models are Few Shot Learners"(GPT-3), May 28, 2020, https://arxiv.org/pdf/2005.14165.pdf OpenAI, "ChatGPT: Optimizing Language Models for Dialog," November 30, 2022, https://openai.com/blog/chatgpt/ L. Monigatti, "A Gentle introduction to Vector Databases," August 1, 2023, https://weaviate.io/blog/what-is-a-vector-database (optional) Ouyang, L. "Training language models to follow instructions with human feedback", 4 Mar 2022, https://arxiv.org/pdf/2203.02155.pdf **HW#4 –** Natural language
8	**Improving Large Language Models** - Image synthesis - Vision transformers - Multi-modal models - Retrieval-augmented generation - Synthetic Data - Small vs. large language models - AI Agents	Y. Gao, et. al. "Retrieval-Augmented Generation for Large Language Models," January 4, 2024, https://arxiv.org/abs/2312.10997 (optional) Devlin, J. et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", 24 May 2019, https://arxiv.org/pdf/1810.04805.pdf **Quiz #2**

9	**Ethics in AI** - Privacy - AI and the future of work - Algorithmic bias - AI alignment	Russell & Norvig, "Chapter 27: Philosophy, Safety and Ethics of AI" in *Artificial Intelligence: A Modern Approach*, 2020 E. Ntoutsi, et. al., "Bias in data-driven artificial intelligence systems—An introductory survey,"Wiley Online Library, 03 February 2020, https://onlinelibrary.wiley.com/doi/full/10.1002/widm.1356
10	**Robotics and Artificial General Intelligence** - Introduction to robotics - Collaboration and autonomy - Artificial General Intelligence - Prospects for Artificial Superintelligence	J. Reed, et. Al, "A Generalist Agent", May 19, 2022, https://arxiv.org/pdf/2205.06175.pdf Russell & Norvig, "Chapter 26.8 Humans and Robots" in *Artificial Intelligence: A Modern Approach*, 2020 L. Fridman, "Human-Centered Autonomous Vehicle Systems: Principles of Effective Shared Autonomy", 3 Oct 2018, https://arxiv.org/pdf/1810.01835.pdf **HW #5 –** Fairness in AI
11	**Infrastructure for AI** - Parallel and distributed computing for scalability - MLOps	https://a16z.com/2020/10/15/the-emerging-architectures-for-modern-data-infrastructure/ Google Cloud, MLOps: Continuous delivery and automation pipelines in machine learning, https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning

12	**AI Applications** - AI for Software development - AI and healthcare - AI and financial services - AI and the public sector	**HW #6 –** Artificial General Intelligence
13	**AI Governance; Futures of AI** - AI governance frameworks - AI policies and regulations - Brain computer interfaces - Other emerging developments in AI - Quantum computing and AI	Russell & Norvig, "Chapter 28: The Future of AI", in *Artificial Intelligence: A Modern Approach*, 2020 **Quiz #3**
14	**Final Project Presentations** - Final presentations	Final project presentations Final report