AI109: Enter the World of Artificial Intelligence

Lecture 1

AI109: Enter the World of Artificial Intelligence

- Introductory course on artificial intelligence for non-majors.
- Focus on core ideas, not advanced mathematics or programming.
- Explores technical foundations and societal implications of AI.

What Is This Course About?

- What artificial intelligence is—and what it is not.
- How AI differs from traditional software systems.
- Why AI raises ethical, legal, and philosophical questions.

Course Goals

- Develop a coherent mental model of what artificial intelligence is, how it differs from traditional software, and where its current capabilities and limits lie.
- Analyze classical AI approaches (search, logic, decision theory) and contrast them with modern data-driven methods.
- Gain conceptual fluency with machine learning, reinforcement learning, and large language models without assuming advanced mathematics.
- Evaluate how AI systems are trained, deployed, and aligned with human values and social constraints.
- Build literacy in AI-assisted programming as both a productivity tool and a source of new risks.

Learning Outcomes

- Explain what artificial intelligence is, how it differs from traditional software, and how its definitions have evolved across historical and technical contexts.
- Describe and analyze core computational ideas underlying AI systems, including programs, algorithms, representation, search, and decision-making.
- Compare and evaluate classical symbolic AI approaches and modern learning-based methods, identifying their strengths, limitations, and appropriate use cases.
- Explain how machine learning systems—including reinforcement learning and large language models—are trained, how they generalize, and where their failures arise.
- Apply AI-assisted programming tools to construct and modify simple programs while critically assessing their reliability, bias, and limitations.
- Critically assess the societal, ethical, and human implications of AI in domains such as robotics, warfare, labor, and human cognition, grounding arguments in technical understanding.

Course Structure

- Lectures and in-class discussion
- Regular quizzes and small assignments
- Three major projects plus midterm and final exams

Administrative Details

- Instructor: Richard Kelley, Ph.D.
- Meetings: Wed & Fri, 11:10–12:25, Hannan 108
- Office hours: Wednesday 10am, Pangborn 324

Communication & Course Resources

- Announcements via Brightspace and CUA email.
- Questions should be posted on Brightspace when appropriate.
- Readings and materials posted on course website.

Attendance & Technology

- Attendance required for exams; expected for lectures.
- Laptops permitted for note-taking.
- Student recording of lectures is not permitted.

Assignments & Late Policy

- Regular small assignments and major projects.
- Late work accepted with 10% penalty per day.
- Make-up exams require advance arrangement and documentation.

Use of AI in This Course

- AI use permitted or prohibited on a per-assignment basis.
- All AI use must be documented by the student.
- Students evaluated on *how* they use AI, not just results.

Grading Breakdown

- Quizzes: 20%
- Assignments: 5%
- Projects: 30%
- Exams
  - Midterm: 20%
  - Final Exam: 25%

Exams

- Midterm exam in Week 8
- Final exam on scheduled university exam date
- Exams emphasize conceptual understanding, but may include some technical detail.

Academic Integrity

- All work submitted must be your own
- Plagiarism and cheating are not tolerated
- You are responsible for properly crediting sources

University Policies & Accommodations

- University policies apply to grading and conduct – see syllabus
- Students with disabilities should contact DSS
- Reasonable accommodations will be provided when approved

Overview of Course Topics

What This Course Covers

- This course provides a comprehensive introduction to artificial intelligence, designed for students without a computational background. You will develop an understanding of how AI systems work, from classical approaches to modern machine learning.

Technical Foundations

- Programs, algorithms, and search
- Knowledge representation and reasoning
- Machine learning fundamentals
- Large language models

Broader Implications

- AI alignment and safety
- Robotics and industrial applications
- AI in warfare and ethics
- AI and human nature

Course Schedule

Week	Topic
1	Overview of AI
2	What is a Computer Program?
3	AI-Assisted Programming
4	Search Algorithms
5	Knowledge Representation
6	Optimal Decision Making
7	Reasoning and Theorem Proving
8	**Midterm Exam**

Week	Topic
9	*Spring Break*
10	Machine Learning
11	Large Language Models
12	Language Model Alignment
13	Computer Vision
14	Robots and Industrial AI
15	Robotics and Warfare
16	AI and Human Nature

Week 1 - Overview of AI

- Introduces the fundamental question: **What is artificial intelligence?**
- Explains how AI differs from traditional software systems
- Traces the historical development of artificial intelligence
- Examines the current capabilities and limitations of modern AI

- **Goal:** Develop a clear mental model of what AI is and is not in order to critically evaluate AI capabilities, risks, and societal claims without succumbing to hype or fear.

Week 2 - What is a Computer Program?

- Establishes the need to understand computation before studying AI
- Introduces core concepts: programs, algorithms, and data structures
- Explains how computers execute instructions
- Examines determinism in traditional software
- Contrasts conventional programs with AI systems

- **Goal:** Build computational literacy as a foundation for AI literacy, enabling clear understanding of how AI systems differ from traditional software without resorting to magical thinking or dismissiveness.

Week 3 - AI-Assisted Programming

- Introduces modern AI tools that can write, debug, and explain code
- Provides hands-on experience with AI programming assistants
- Teaches how to construct and modify simple programs using AI tools
- Develops skills for evaluating the reliability, correctness, and limitations of AI-generated code

- **Goal**: Gain an introductory understanding of AI-assisted programming, including how these tools can support human work and why their limitations and risks matter.

Week 4 - Search Algorithms

- Introduces the idea of AI problems as **search** through a space of possibilities
- Presents fundamental search strategies: **depth-first**, **breadth-first**, and **heuristic-guided search**
- Explains how search algorithms underpin game-playing AI, route planning, and puzzle solving

- **Goal:** Gain a basic understanding of how search algorithms are used in AI systems.

Week 5 - Knowledge Representation

- Introduces the need for knowledge representation in AI systems
- Explores how facts, relationships, and concepts can be encoded for computation
- Introduces the distinction between **symbolic** and **connectionist** approaches to AI

- **Goal:** Develop an introductory understanding of knowledge representation and why encoding human knowledge for machines is a central and challenging problem in AI.

Week 6 - Optimal Decision Making

- Introduces decision-making under uncertainty in AI systems
- Presents core ideas from **decision theory** and **expected utility**
- Explains how probability and utility are combined to guide **rational** choices

- **Goal:** Gain an introductory understanding of how decision theory guides AI choices under uncertainty and why these decisions may differ from human intuition

Week 7 - Reasoning and Theorem Proving

- Introduces logical reasoning as an early approach to artificial intelligence
- Covers **propositional logic**, **predicate logic**, and **inference rules**
- Explores **automated theorem proving**
- Examines how AI systems derive new conclusions from existing knowledge through formal reasoning

- **Goal:** Gain an introductory understanding of logic-based AI, its role in early research, and why symbolic reasoning still matters alongside modern learning-based methods

Week 10 - Machine Learning

- Introduces machine learning as learning from data rather than explicit programming
- Covers **supervised** and **unsupervised learning**
- Introduces **neural networks**
- Explores how models learn patterns, generalize to new examples, and where they fail
- **Goal:** Develop a basic understanding of machine learning as the dominant approach in modern AI, including how systems learn from data and how this process can fail

Week 11 - Large Language Models

- Introduces **large language models** and their impact on language-based AI
- Explains how models are trained on large-scale text data
- Examines how LLMs generate human-like text and exhibit **emergent** capabilities
- Explores key limitations and common failure modes

- **Goal:** Gain an introductory understanding of large language models, clarifying both their impressive capabilities and their limitations beneath the conversational interface

Week 12 - Language Model Alignment

- Introduces the alignment problem: ensuring AI systems reflect human values
- Explores **reinforcement learning from human feedback (RLHF)**
- Introduces **constitutional AI** and related alignment approaches
- Examines how these methods aim to produce helpful, harmless, and honest AI systems

- **Goal:** Develop an introductory understanding of AI alignment, why it matters for safe and responsible AI, and the challenges that remain unresolved

Week 13 - Computer Vision

- Introduces **computer vision** as enabling machines to interpret visual data
- Covers **image recognition**, **object detection**, and **scene understanding**
- Examines how **convolutional neural networks** process images
- Explores applications from facial recognition to medical imaging

- **Goal:** Gain an introductory understanding of computer vision, its major applications, and the technical and societal limits of how machines “see.”

Week 14 - Robots and Industrial AI

- Introduces robotics as the intersection of AI and the physical world
- Explores how AI supports **perception**, **planning**, and **manipulation**
- Examines **industrial automation** and warehouse robotics
- Discusses challenges of deploying AI in **unstructured** real-world environments

- **Goal:** Develop an introductory understanding of robotics as AI with physical agency, including its societal implications and the challenges of operating in the real world

Week 15 - Robotics and Warfare

- Examines the use of AI in military systems, including autonomous **drones** and targeting technologies
- Explores ethical debates surrounding autonomous and semi-autonomous weapons
- Introduces international efforts to regulate **lethal autonomous weapons systems** (LAWS)

- **Goal:** Gain an introductory understanding of military AI and the ethical questions it raises about autonomy, warfare, and societal governance

Week 16 - AI and Human Nature

- Steps back to examine the deepest questions AI raises about human intelligence and consciousness
- Explores how AI reshapes work, creativity, and human relationships
- Examines philosophical perspectives on mind, meaning, and what it means to be human in an age of intelligent machines,=

- **Goal:** Develop an introductory perspective on how AI challenges traditional views of human intelligence and what it means to be human in an age of intelligent machines

The History of Artificial Intelligence

From Turing to Transformers: 1940s - Present

Timeline Overview

Seven decades of AI evolution

- 1940s-1950s: **Foundations** - Turing, birth of the field
- 1956-1974: **Golden Years** - Optimism, early programs
- 1974-1993: **AI Winters** - Setbacks, reduced funding
- 1993-2010: **Revival** - Machine learning rises
- 2010-2020: **Deep Learning** - Neural networks triumph
- 2020-Present: **Generative AI** - LLMs transform society
- Future: What's Next? - AGI, ethics, regulation

Part I

Foundations & Early Ideas

1940s - 1950s

Philosophical Predecessors

The dream of thinking machines predates computers

- 1673 - Gottfried Leibniz: Built the **Stepped Reckoner**, a mechanical calculator. Dreamed of a "calculus ratiocinator" - a universal logical calculator.
- 1837 - Charles Babbage: Designed the **Analytical Engine**, the first general-purpose programmable computer concept.
- 1843 - Ada Lovelace: First to recognize computers could go beyond calculation. Wrote the first algorithm for machine execution.

"The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform." — Ada Lovelace, 1843

Alan Turing: Father of AI

The mathematician who asked "Can machines think?"

- 1936: Introduced the **Turing Machine** concept - theoretical foundation for all computing
- 1950: Published "Computing Machinery and Intelligence" - proposed the Turing Test
- The **Imitation Game**: If a machine can fool a human interrogator 30% of the time, it exhibits intelligent behavior
- Turing's Vision: "We can only see a short distance ahead, but we can see plenty there that needs to be done."

The Dartmouth Conference (1956)

Where "Artificial Intelligence" was born

- John McCarthy coined the term "**Artificial Intelligence**"
- Key Attendees: Marvin Minsky, Claude Shannon, Nathaniel Rochester, Allen Newell, Herbert Simon
- The Proposal: "Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."
- This 8-week workshop established AI as a formal field of research

Part II

The Golden Years

1956 - 1974

Early AI Programs

The first demonstrations of machine reasoning

- **Logic Theorist** (1956): Newell & Simon - proved mathematical theorems, found novel proofs
- **General Problem Solver** (1959): Newell & Simon - attempted to create a universal problem-solving machine
- SAINT (1961): Slagle - solved **symbolic integration** problems at college freshman level
- These programs showed machines could perform tasks previously thought to require human intelligence

ELIZA: The First Chatbot

Joseph Weizenbaum, MIT, 1966

- Simulated a Rogerian psychotherapist using pattern matching
- Used simple rules to transform user input into responses
- The **ELIZA Effect**: Users attributed human-like understanding to the program despite its simplicity
- Weizenbaum was disturbed by how readily people formed emotional bonds with his creation

Early Neural Networks

The Perceptron and its limitations

- 1958: Frank Rosenblatt's **Perceptron** - first trainable neural network
- Could learn to classify patterns through training
- 1969: Minsky & Papert's "Perceptrons" book showed fundamental limitations
- Could not solve **XOR problem** - this criticism led to the first **AI winter** for neural networks

The Era of Optimism

Bold predictions that didn't come true

- Herbert Simon (1965): "Machines will be capable of doing any work a man can do within 20 years"
- Marvin Minsky (1967): "Within a generation the problem of creating 'artificial intelligence' will be substantially solved"
- These predictions fueled massive government and corporate investment
- The gap between promise and reality would soon become apparent

Part III

The AI Winters

1974 - 1993

The First AI Winter (1974-1980)

When reality caught up with the hype

- 1973: The **Lighthill Report** (UK) criticized AI's failure to achieve "grandiose objectives"
- **DARPA** cut AI funding dramatically in the United States
- Key Problems: **Combinatorial explosion**, limited computing power, narrow applicability
- Many AI labs closed; researchers left the field or rebranded their work

Expert Systems: A Brief Revival

Knowledge-based AI in the 1980s

- **MYCIN** (1970s): Diagnosed bacterial infections with 69% accuracy (better than some doctors)
- R1/XCON (1980s): Configured computer systems for DEC, saved $40M annually
- Japan's Fifth Generation Project (1982): $850M government investment in AI
- **Expert systems** market grew to $1 billion by 1988

The Second AI Winter (1987-1993)

The collapse of the expert systems market

- Expert systems proved brittle, expensive to maintain, and couldn't handle uncertainty
- The AI hardware market collapsed as desktop computers became more powerful
- Japan's **Fifth Generation Project** failed to meet its ambitious goals
- "AI" became a dirty word - researchers avoided the term to get funding

Part IV

The Revival Era

1993 - 2010

Machine Learning Emerges

A shift from rules to learning from data

- **Statistical methods** replaced **hand-coded rules**
- **Support Vector Machines** (1995): Powerful classification algorithms
- Random Forests, Boosting: Ensemble methods improved accuracy
- Key insight: Let the machine discover patterns rather than programming them explicitly

Deep Blue vs. Kasparov

May 11, 1997 - A watershed moment

- IBM's **Deep Blue** defeated world chess champion Garry Kasparov
- Evaluated 200 million positions per second using specialized hardware
- Proved machines could defeat humans at intellectual games
- Critics noted it used **brute force**, not "real" intelligence - but the public perception shifted

AI Goes Mainstream

Quiet successes in the 2000s

- **Search Engines**: Google's PageRank and machine learning for relevance
- **Spam Filters**: Bayesian classification became ubiquitous
- Recommendation Systems: Netflix Prize (2006) spurred innovation
- **Speech Recognition**: Dragon NaturallySpeaking, early Siri development

Part V

Deep Learning Revolution

2010 - 2020

The ImageNet Moment (2012)

When deep learning proved its power

- **AlexNet** won ImageNet competition with 15.3% error rate (vs. 26.2% for runner-up)
- Created by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton
- Key innovations: Deep CNNs, ReLU activation, dropout, **GPU training**
- This single result reignited the entire field and launched the deep learning era

The GPU Revolution

Hardware enables deep learning scale

- **GPUs**: Originally for graphics, perfect for parallel matrix operations
- NVIDIA **CUDA** (2007): Made GPU programming accessible
- Training time dropped from weeks to hours
- **Moore's Law** + Specialized Hardware = Exponential capability growth

AlphaGo: Beyond Human Intuition

DeepMind conquers the "impossible" game

- March 2016: **AlphaGo** defeated Lee Sedol 4-1 in Go
- Go has 10^170 possible positions - brute force impossible
- Combined deep learning with **Monte Carlo tree search**
- **Move 37** in Game 2: A creative move no human would make, praised as "beautiful"

Natural Language Processing

Teaching machines to understand language

- **Word2Vec** (2013): Words as vectors capturing semantic meaning
- Sequence-to-Sequence (2014): Enabled machine translation breakthroughs
- **Attention Mechanism** (2015): Models learn what to focus on
- **Transformers** (2017): "Attention Is All You Need" - foundation for modern LLMs

Part VI

The Modern Era

2020 - Present

The GPT Revolution

Language models that can do almost anything

- **GPT-2** (2019): OpenAI initially withheld it, fearing misuse
- **GPT-3** (2020): 175 billion parameters, few-shot learning capabilities
- **GPT-4** (2023): Multimodal, passing bar exams and medical boards
- Key insight: Scale + simple objective (predict next word) = **emergent capabilities**

ChatGPT: AI Meets the Masses

November 30, 2022 - The tipping point

- Reached 100 million users in 2 months (fastest-growing app ever)
- Made AI accessible to everyone through natural conversation
- Sparked global conversation about AI capabilities and risks
- Triggered massive investment: Microsoft ($10B in OpenAI), Google (Bard), Meta (LLaMA), Anthropic (Claude)

Multimodal AI

Beyond text: Images, audio, video

- DALL-E, Midjourney, Stable Diffusion: **Text-to-image generation**
- Whisper: Near-human **speech recognition** across 99 languages
- GPT-4V, Claude 3, Gemini: Vision + language understanding
- Sora, Runway: **Text-to-video** generation emerging

AI Applications Today

Transforming every industry

- Healthcare: Drug discovery (**AlphaFold**), medical imaging, diagnosis assistance
- Finance: Fraud detection, algorithmic trading, risk assessment
- Transportation: **Self-driving vehicles**, route optimization
- Creative: Writing assistance, art generation, music composition, code generation

Part VII

Looking Forward

The Future of AI

The Path to AGI

Artificial General Intelligence: The ultimate goal?

- **AGI**: AI that matches or exceeds human capability across all cognitive tasks
- Timeline predictions vary wildly: 5 years to "never"
- Current debate: Are LLMs a path to AGI or a dead end?
- Unknown unknowns: We may not recognize AGI when we see it

Ethical Considerations

The challenges we must address

- Bias & Fairness: AI systems can perpetuate and amplify societal biases
- Privacy: Training data concerns, surveillance capabilities
- Misinformation: Deepfakes, automated propaganda at scale
- Economic Disruption: Job displacement, wealth concentration, access inequality

Key Takeaways

What we've learned from 70+ years of AI

- Progress is non-linear: AI winters followed by explosive growth
- Simple methods + scale often beat complex approaches
- Hardware advances enable algorithmic breakthroughs
- The gap between narrow AI and AGI remains vast but shrinking
- Human-AI collaboration may be more important than human-level AI

Questions?