جامعة الفيصل
Alfaisal University

كلية الهندسة والحوسبة المتقدمة
College of Engineering and Advanced Computing

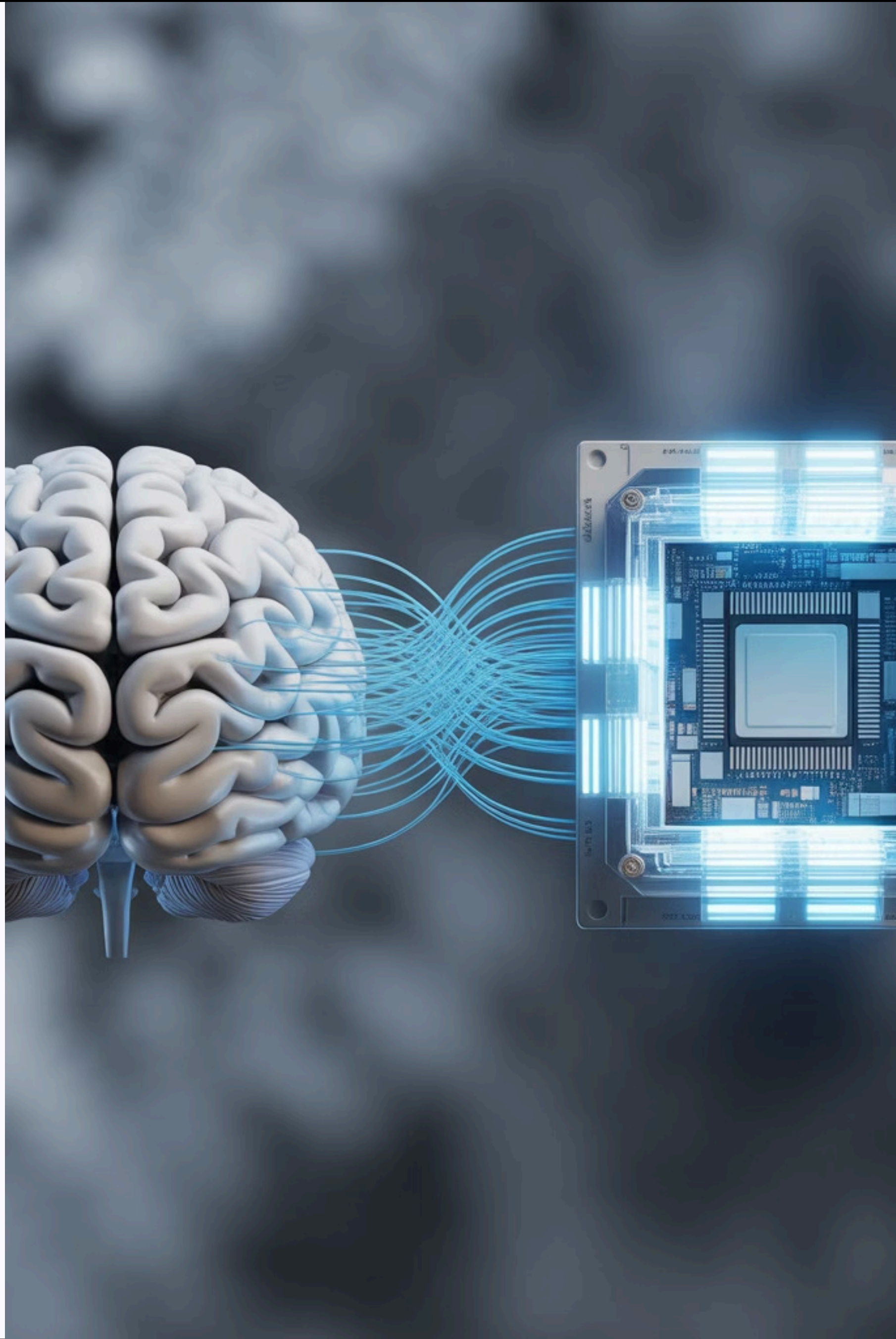
Introduction to Artificial Intelligence (SE 444)

Lecture 1: Introduction to AI

Prof. Anis Koubaa

akoubaa@alfaisal.edu

This lecture introduces the foundational concepts of Artificial Intelligence, exploring its history, key definitions, and diverse applications. We will delve into the philosophical questions surrounding intelligence, differentiate between various AI approaches, and understand the scope of AI research within the context of SE 444. Gain insights into the journey of AI from early theories to modern advancements and its impact on technology and society.



Course Outline

What is AI?

Defining **artificial intelligence** and its core **objectives**

Different approaches to AI:

Thinking vs. Acting,
Human-like vs. Rational

History of AI

Key milestones from the inception of AI to modern breakthroughs

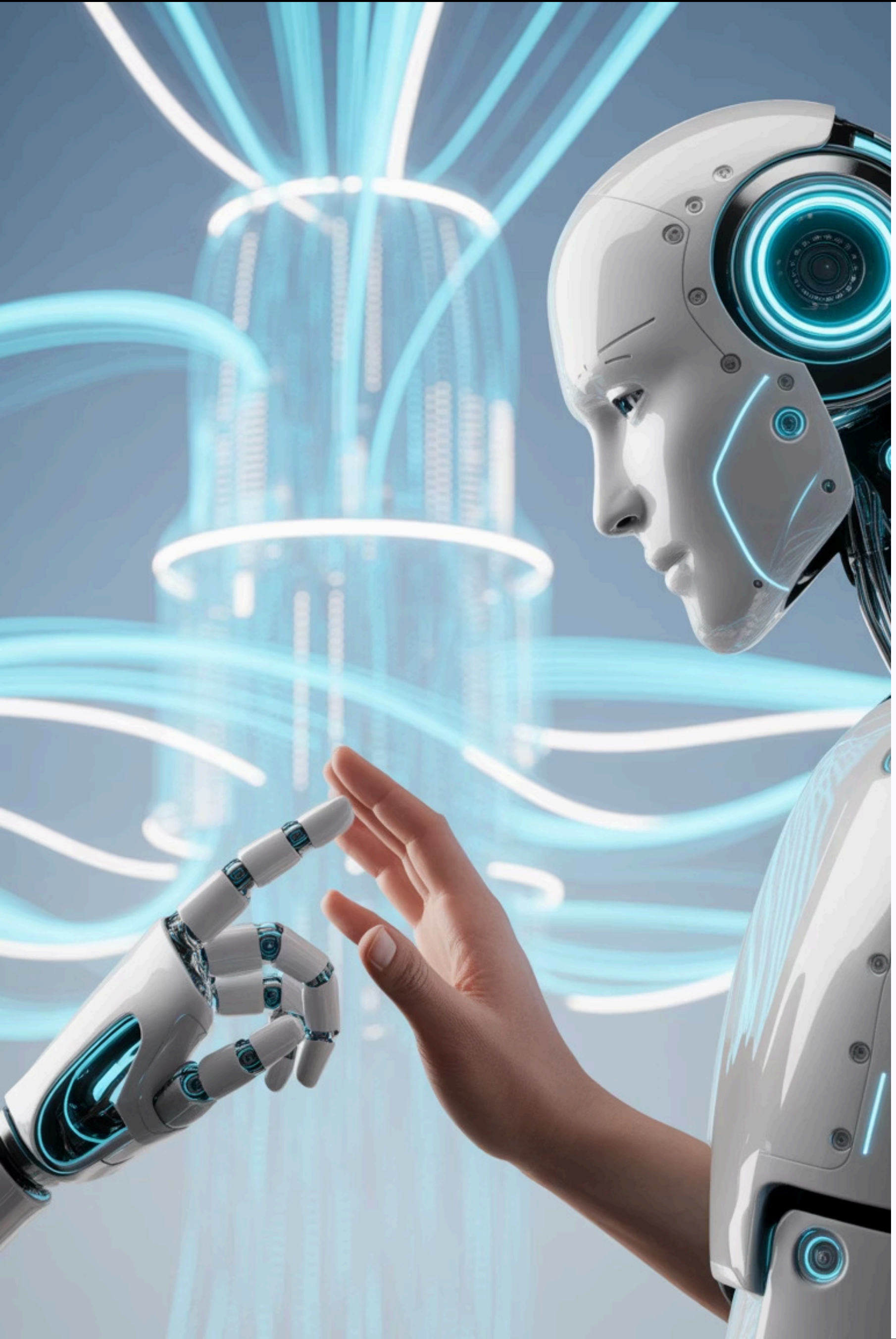
AI winters and resurgences throughout technological development

AI Applications

Real-world implementations across various domains

Current state-of-the-art AI systems and their capabilities

This introduction will provide you with a foundational understanding of AI concepts that will be built upon throughout the course. You'll gain insight into both the theoretical underpinnings and practical applications of artificial intelligence.



What is Artificial Intelligence?

A Field of Study

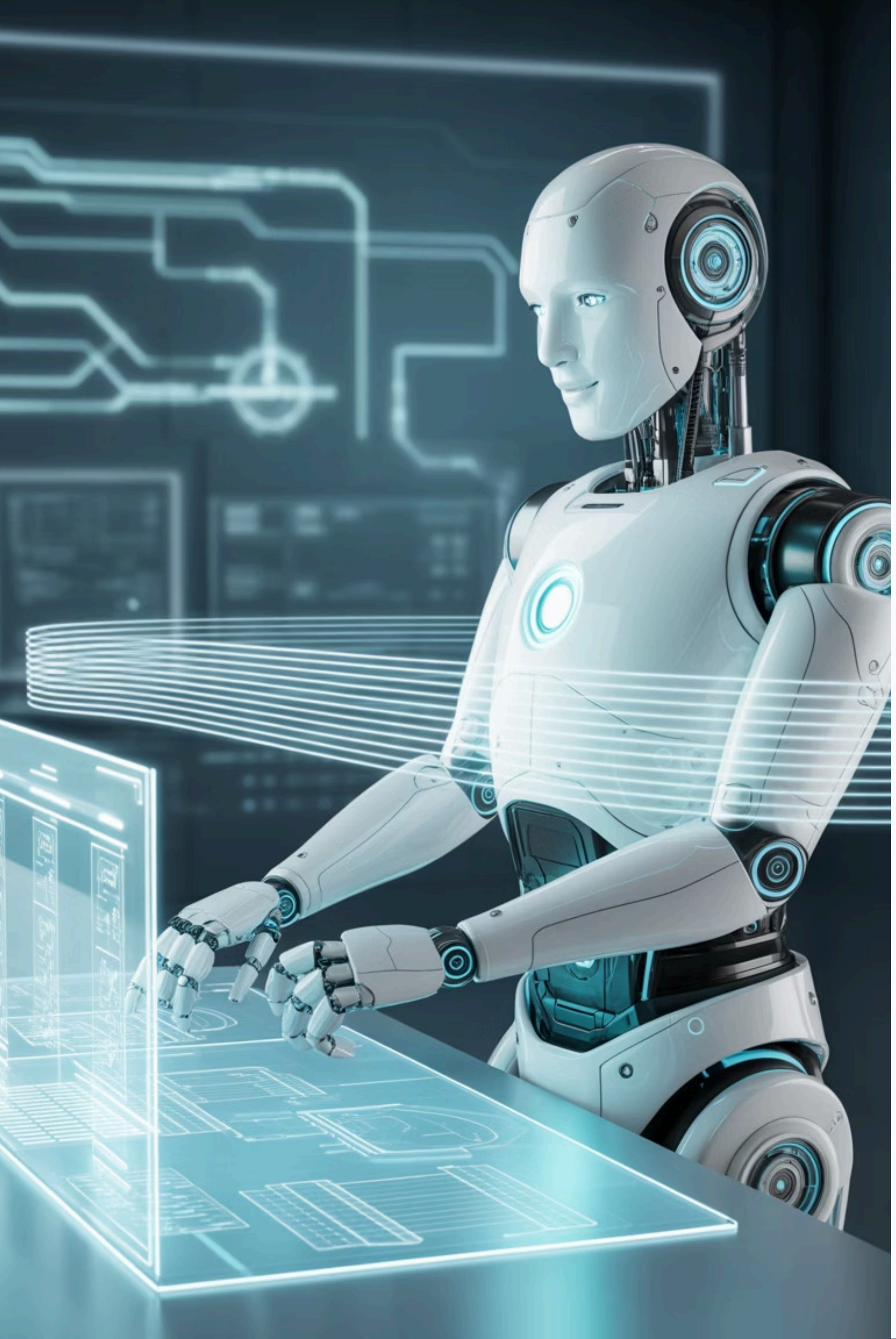
AI is the branch of computer science that aims to create systems capable of performing tasks that would normally require human intelligence.

A Collection of Technologies

AI encompasses various techniques including machine learning, neural networks, natural language processing, computer vision, and knowledge representation.

A Philosophy

AI explores fundamental questions about the nature of intelligence, consciousness, and what it means to "think" or "understand."



What is the Goal of AI?

"Have machines solve problems that are challenging for humans."

Narrow AI

Focuses on intelligent agents designed to solve a specific subproblem

- Image recognition
- Natural language processing
- Game playing
- Specialized decision-making

Artificial General Intelligence (AGI)

A hypothetical intelligent agent which can understand or learn any intellectual task that human beings or other animals can

- Human-level reasoning across domains
- Transfer learning between unrelated tasks
- Still largely theoretical

How do we achieve this?

By creating an agent that can perceive its environment, reason about it, and take actions that maximize its chances of success at some goal.

Different Views of AI

Human vs. Rational AI

This dimension addresses the fundamental question: **What kind of intelligence are we trying to emulate or surpass?**

Human-like AI

Aims to **replicate human performance**, including both **cognitive** processes and **observable** behavior. The goal is to **imitate** how humans think or behave, even if that behavior isn't always perfectly optimal.

- Mimics human thought processes.
- Replicates human behavior patterns.

Rational AI

Focuses on achieving the **best possible outcome** or making the most optimal decisions, even if the method doesn't mirror human cognition. It prioritizes logical, effective performance over human imitation.

- **Maximizes** performance and efficiency.
- Based on logical inference and **optimal** decision-making.

Thinking vs. Acting AI

This dimension distinguishes whether the focus is on the **internal thought processes** or the **external manifestation of intelligence**.

Thinking AI

Concerned with the **internal reasoning mechanisms** of the AI system. It seeks to model and understand the cognitive processes involved in problem-solving, logic, and decision-making.

- Focus on internal mental processes.
- Emphasizes reasoning and problem-solving.

Acting AI

Concentrates on the **external behavior** and observable actions of the AI system. The primary goal is to produce intelligent behavior, regardless of the internal methods used to achieve it.

- Focus on observable output and behavior.
- Emphasizes successful task completion.

🤔 How Should a Self-Driving Car Drive?

Drive Like a Human?

Should it emulate **human behavior**, including occasional *caution*, *distraction*, or even *mistakes*?

- Prioritizes "natural" interaction
- May exhibit unpredictable actions
- Relatable, but potentially less safe

Drive Rationally?

Or should it **always choose the safest**, most *efficient* action, *optimizing* every decision?

- Prioritizes optimal performance
- Always aims for maximum safety/efficiency
- Potentially less "human-like"

Think: If you're in the car, which approach do you prefer?

👉 Pause here and let students share quick opinions.

Acting Humanly vs. Acting Rationally

Acting Humanly

This approach focuses on replicating human cognition and behavior, including nuances and potential imperfections.

- Mimics human decision-making and observable actions.
- **In a self-driving car:** may exhibit human-like traits such as *occasionally passing yellow lights* or slightly exceeding speed limits.
- The ultimate goal is to create systems that can mimic the complex, often non-optimal, ways humans think and act.

Acting Rationally

This approach prioritizes achieving the **best possible outcome** based on logical reasoning and optimization, regardless of whether it mirrors human behavior.

- Seeks to **maximize performance** and efficiency in every action.
- **In a self-driving car:** would always *adhere strictly to traffic laws*, adjust speed optimally for safety, and minimize any risk of accidents.
- The goal is to produce intelligent agents that consistently make the most logical and effective decisions to achieve their objectives.

💡 **For self-driving cars:**

Acting humanly = *"drive like us"*

Acting rationally = *"drive better than us"*

Different Views of AI

Artificial Intelligence can be approached and defined from various perspectives, each emphasizing different aspects of intelligence. Understanding these views helps in categorizing and developing AI systems.

Thinking Humanly (Cognitive Science)

Aims to build **AI that thinks like humans**, focusing on models of **human cognition** and internal thought processes, often involving psychological experimentation.

Acting Humanly (Turing Test)

Focuses on creating **AI that acts indistinguishably from a human**, judged by external behavior rather than internal mechanisms, famously encapsulated by the **Turing Test**.

Thinking Rationally (Laws of Thought)

Develops AI systems based on formal logic, aiming to build systems that can reason correctly and deduce conclusions from given premises, regardless of human-like behavior.

Acting Rationally (Rational Agent)

Designs AI agents that act to achieve the best possible outcome or expected outcome given their perceptions, regardless of whether their thought process is human-like or logical.

Human vs. Rational AI

AI development largely follows two distinct paradigms, each with different goals and methodologies for defining and achieving intelligence.

Human-Centered AI

Aims to replicate human cognitive abilities and behavior.

- **Thinking Humanly:** Focuses on modeling human thought processes (cognitive science, neuroscience).
- **Acting Humanly:** Seeks to create AI that acts indistinguishably from a human (Turing Test, chatbots).

Goal: To build systems that emulate human intelligence, often through learning and adaptation.

Limitation: Human cognition is inherently biased, limited, and can be irrational.

Rational AI

Focuses on optimal decision-making to achieve specific goals.

- **Thinking Rationally:** Uses formal logic and reasoning to deduce conclusions (deductive systems).
- **Acting Rationally:** Designs agents to act to maximize expected outcomes, regardless of human-likeness.

Goal: To achieve the best possible outcome given available information and constraints.

Advantage: Not restricted by human weaknesses, often leading to performance that surpasses human capabilities.

Key Point: Human-like ≠ Rational

Rational AI, focused on optimal performance, is the dominant approach in modern AI applications like search, planning, and decision-making.

1. Thinking Like a Human

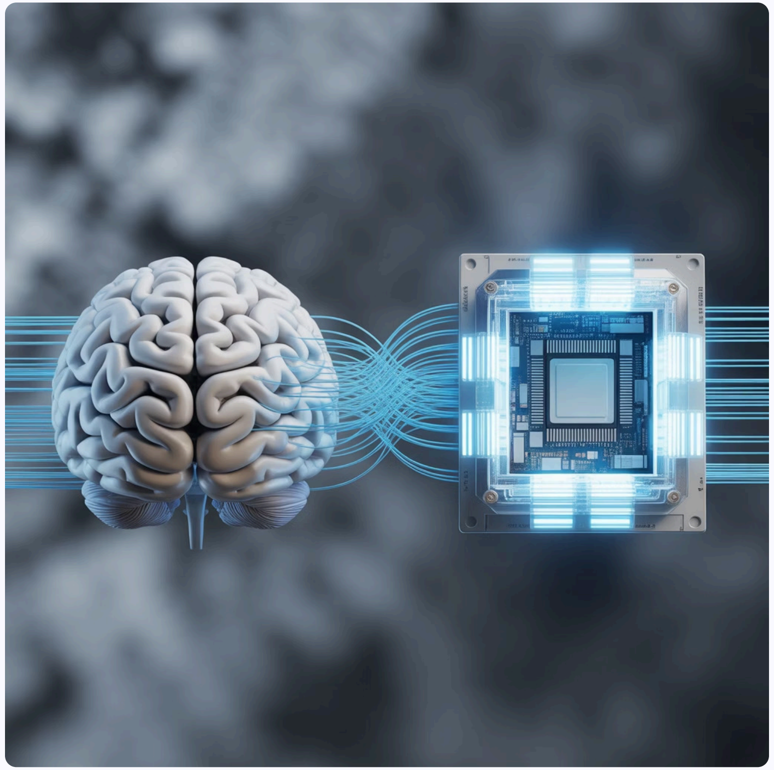
The Cognitive Science Approach

This perspective aims to create AI systems that mimic human thought processes and cognitive structures.

Key aspects include:

- Understanding how humans **process information**
- Modeling human **decision-making** and **problem-solving**
- Implementing **memory structures** inspired by human cognition
- Using **psychological experiments** to inform AI design

Note: The brain does not work like **artificial neural networks** from ML! **Biological neural networks** are far more complex, with different activation mechanisms, connectivity patterns, and learning processes.



Practical Examples of Human-Like AI

Here are concrete examples of AI systems attempting to emulate human cognitive functions:

Expert Systems

These **systems replicate human experts' reasoning** by using programmed knowledge.

For example, a medical diagnosis system would mimic a physician's rules to identify diseases.

Cognitive Architectures

Frameworks like ACT-R or SOAR are built on psychological theories to **simulate human memory**, **perception**, and **problem-solving**, offering insights into human learning.

Human-Like Errors

Some AI models are intentionally designed to **exhibit biases or errors similar to those in human cognition**, such as misclassifying objects in line with human optical illusions.

Natural Language Processing (NLP)

Modern NLP systems strive to understand, interpret, and generate **human language**, replicating abilities like understanding context, sarcasm, and emotional tone.

2. Acting Like a Human

Alan Turing (1950)

"Computing machinery and intelligence"

The Turing Test

A test designed to determine ***if a machine can exhibit intelligent behavior indistinguishable from a human***.

Required Capabilities for Passing:

- **Natural language processing** - To communicate effectively in human language
- **Knowledge representation** - To store information
- **Automated reasoning** - To use stored knowledge to answer questions and draw conclusions
- **Machine learning** - To adapt to new circumstances and detect patterns

Turing predicted that by 2000, machines would fool 30% of human judges for five minutes. ChatGPT in 2023 dramatically exceeded this benchmark!

The Turing Test Explained

- Proposed by **Alan Turing (1950)**.
- Idea: If you talk to a computer (by typing) and **you cannot tell whether it's a human or a machine**, then the machine has passed the test.

> "Can a machine pretend to be human?"

Acting Humanly

- **Acting Humanly** means: *the machine should act in a way that looks human, not necessarily the best or most logical way.*
- The **Turing Test** is exactly about this: if the AI **fools you into thinking it's human**, then it's acting humanly.

Simple Classroom Example

Imagine chatting with two people in WhatsApp:

- One is your friend.
- One is a chatbot.
  If you can't tell which is the chatbot → it **passed the Turing Test** → that's **Acting Humanly**.

How Turing Developed the Test

In his seminal 1950 paper, **"Computing Machinery and Intelligence,"** Alan Turing addressed the profound question: "Can machines think?" Instead of a direct philosophical answer, he proposed a practical experiment to test a machine's ability to exhibit intelligent behavior.

The Foundational Question

Turing's paper replaced the abstract question "Can machines think?" with a more empirical and testable one.

Introducing "The Imitation Game"

He reimagined a party game to create a concrete framework for evaluating machine intelligence, bypassing philosophical debates.

The Imitation Game Setup

Original Party Game

A human interrogator, isolated, communicates via text with two hidden individuals: a man and a woman. The interrogator's goal is to determine which person is the man and which is the woman.

Turing's Scientific Twist

Turing proposed replacing one of the hidden humans with a machine. The interrogator's new challenge: distinguish the human from the machine based solely on their text-based responses.

Turing himself did not build a chatbot, as computers in his time were rudimentary. His contribution was primarily a **thought experiment** and a **conceptual framework**.

3. Thinking Rationally

Thinking rationally focuses on the idealized or "right" way of thinking, primarily through logical reasoning.

Logic-Based Approach to AI

Dating back to Aristotle (385 BC), this approach involves:

- Describing problems in formal logical notation
- Applying general deduction procedures to solve them
- Example: "Socrates is a man; all men are mortal; therefore, Socrates is mortal."

Limitations of Pure Logic

- Describing real-world problems in logical notation is extremely challenging
- Computational complexity becomes prohibitive for complex problems
- Rational behavior in uncertain environments cannot be defined by simple rules
- Human knowledge is rarely definitive enough for strict logical deduction

While logic provides a foundation for AI reasoning, practical systems require additional techniques to handle uncertainty and complexity.

Thinking Rationally: Rule-Based AI

This approach emphasizes AI systems that reason step-by-step to arrive at correct conclusions, mirroring idealized human logic.

Classical AI Foundation

From the 1950s-1980s, rational thinking in AI was primarily realized through **rule-based systems** and formal logic.

System Components

This included using **propositional logic** (true/false statements), **first-order logic** (rules with variables), and **expert systems** with knowledge bases.

How It Works

Rule-based systems mimic rational thought: input facts, apply logical rules, and derive a precise conclusion.

For example, in a medical expert system:

> Rule: If (fever AND cough) → then (diagnose: flu)
>
> The AI "thinks rationally" by deducing the diagnosis directly from established rules.

Key Limitation: Real-World Uncertainty

While effective in clearly defined, structured domains (like mathematics or logical puzzles), these systems struggle with the inherent uncertainty and complexity of the real world, unlike modern ML's probabilistic approaches.

From Rules to Learning

Classical AI: Rule-Based

This approach focuses on explicit rules and logical inference, often following an "If X then Y" structure.

- Human-designed rules and symbolic reasoning.
- **Example:** "If fever AND cough → flu."
- AI deduces conclusions based on pre-defined logic.

Modern AI: Machine Learning

Instead of hand-coded rules, ML and Deep Learning learn patterns and make decisions directly from data.

- No explicit rules written by humans.
- **Example:** Trained on thousands of medical records, learns probabilities like "If fever + cough → 80% chance flu."
- Model builds its own "rules" from data.

Rule-Based AI Analogy

A teacher gives you the rules, and you just apply them.

Machine Learning Analogy

You look at many examples, discover the rules yourself, and then apply them.

While both aim for rationality, ML/DL represent a newer, data-driven path to achieve optimal decisions.

Contrasting AI Approaches: Rule-Based vs. Machine Learning

While both traditional rule-based AI and modern Machine Learning (ML) aim for intelligent behavior, their underlying mechanisms and ideal applications differ significantly.

	Thinking Rationally (Rule-Based AI)	Machine Learning & Deep Learning
Strengths	Explainable, precise in known domains	Scalable, handles uncertainty and complexity
Weaknesses	Brittle in real-world variability; hard to scale rules	"Black box" (less interpretable); data-hungry
Use Cases	Chess engines (e.g., early Deep Blue logic), legal reasoning	Image recognition, self-driving cars, predictive diagnostics
Rationality Type	Deductive (logical proofs)	Inductive/Statistical (probabilistic optimization)

Understanding these distinctions is crucial for selecting the appropriate AI paradigm for different challenges.

4. Acting Rationally

A rational agent acts to achieve the best expected outcome

Goal-Oriented

Goals are application-dependent and expressed in terms of the **utility of outcomes** - the measure of "goodness" of different results.

Maximizing Expected Utility

Being rational means acting to maximize your expected utility. "Expectation" accounts for different possible outcomes with associated probabilities.

Bounded Rationality

In practice, utility optimization is subject to the agent's knowledge and computational constraints - what's known as **bounded rationality**.

The rational agent approach doesn't require perfect reasoning or complete knowledge - just making the best decision possible with available information and resources.

4. Acting Rationally

Advantages of Expected Utility Maximization

Generality

An optimization framework that goes beyond explicit rules, allowing broader application across diverse domains.

Practicality

Adaptable to real-world problems with inherent uncertainty and complexity, making it highly applicable.

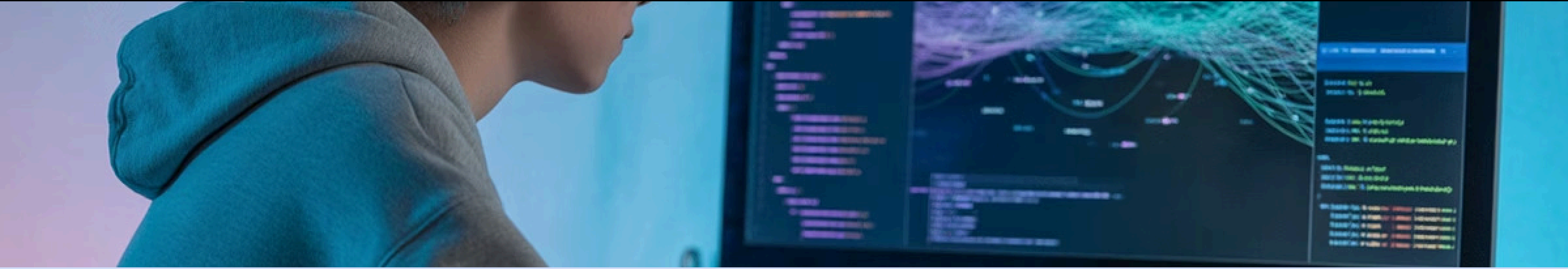
Scientific Approach

Amenable to rigorous scientific and engineering methodologies, including simulation and experimentation.

Outcome Focus

Primarily concerns the decisions and actions made, not the intricate cognitive process behind them.

This formulation provides a flexible framework for developing AI systems that can operate effectively in complex, uncertain environments while remaining computationally feasible.



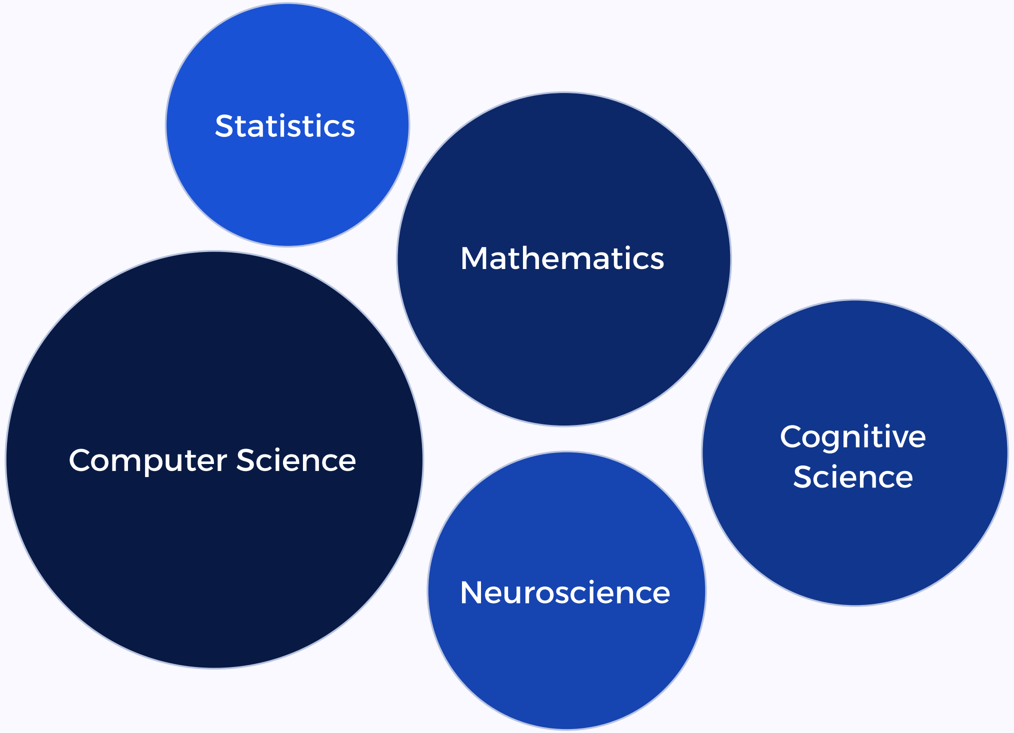
What type of AI do we cover in this course?

Create a narrow AI agent that think and acts rationally

That is, use machines to solve a specific hard problem that traditionally would have been thought to require human intelligence.

This course will teach you the fundamental principles and techniques to design intelligent agents that can perceive their environment, reason about it, and take actions to achieve specific goals.

Fields Related to AI



Artificial intelligence is inherently interdisciplinary, drawing from and contributing to numerous fields. Understanding these connections helps contextualize AI's development and applications.

Throughout this course, we'll explore how these various disciplines inform modern AI approaches and how advances in AI reciprocally impact these fields.

Components of an Intelligent Agent

> When you use **Siri**, **Alexa**, or a **self-driving car** – how do they actually ‘sense’ and ‘decide’ what to do?

Components of an Intelligent Agent

Essential Components

Environmental Interface

Sensors to perceive the environment and actuators to affect it

Knowledge Representation

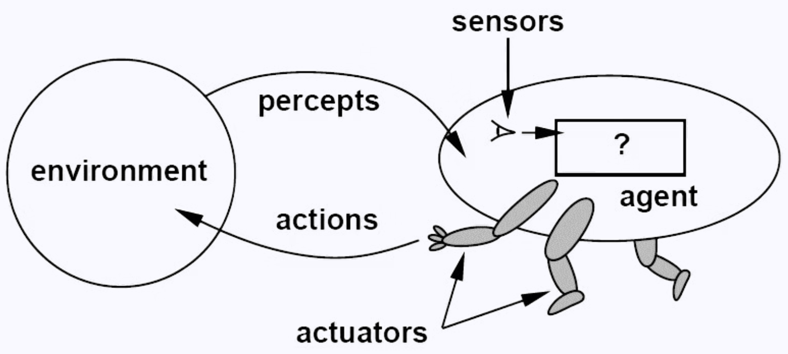
Methods to store and organize information about the world

Reasoning & Planning

Algorithms to process information and determine actions to achieve goals

Learning (Optional)

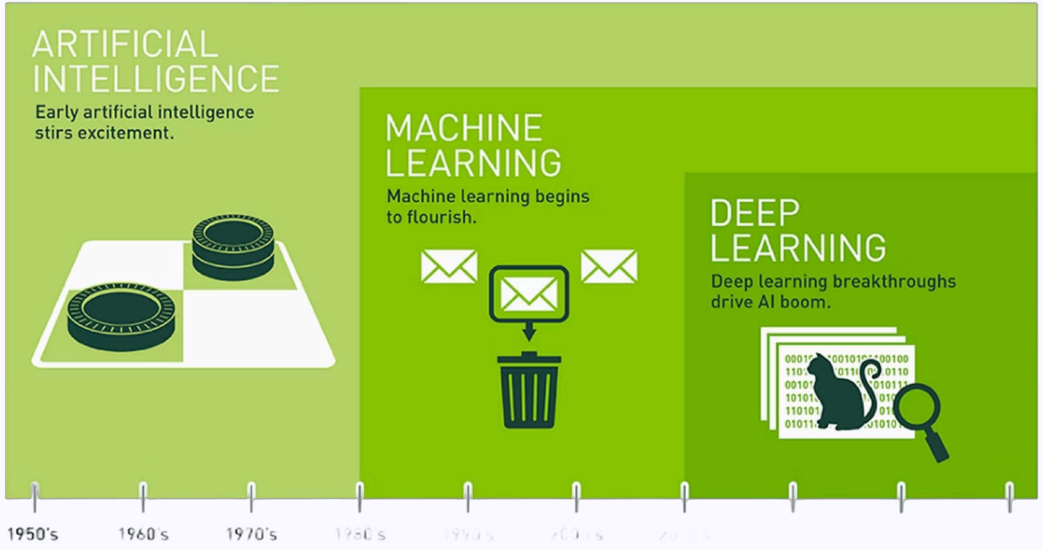
Ability to improve performance based on experience



Agent interacting with the environment

[Artificial Intelligence: A Modern Approach, Editions 1-3]

An intelligent agent requires both the ability to interact with its environment and the cognitive capabilities to process information, make decisions, and potentially learn from experience.



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

From AI to Deep Learning

1

Artificial Intelligence

The broad concept of machines being able to carry out tasks in a way that we would consider "smart"

2

Machine Learning

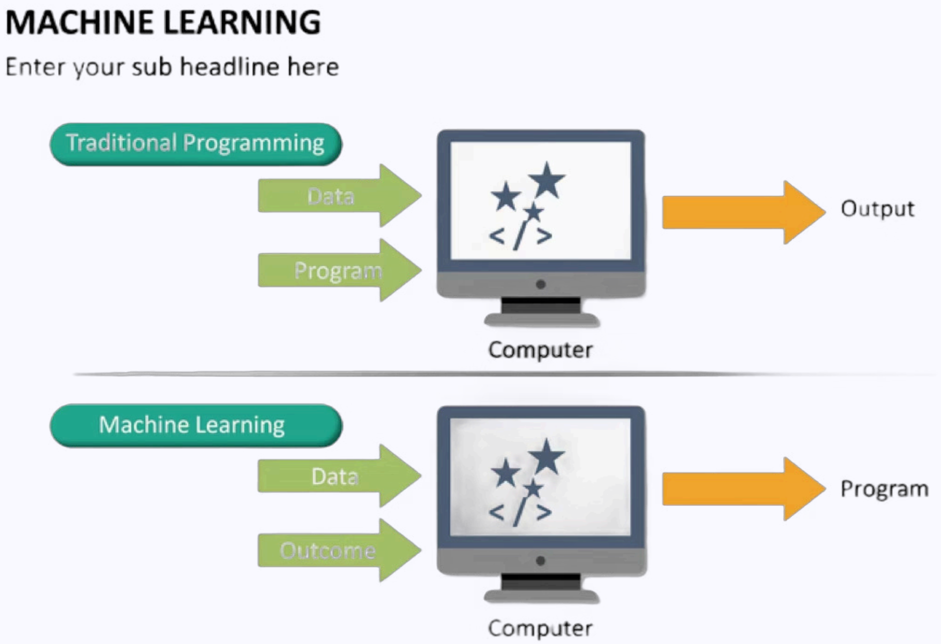
Systems that can learn from data without being explicitly programmed

3

Deep Learning

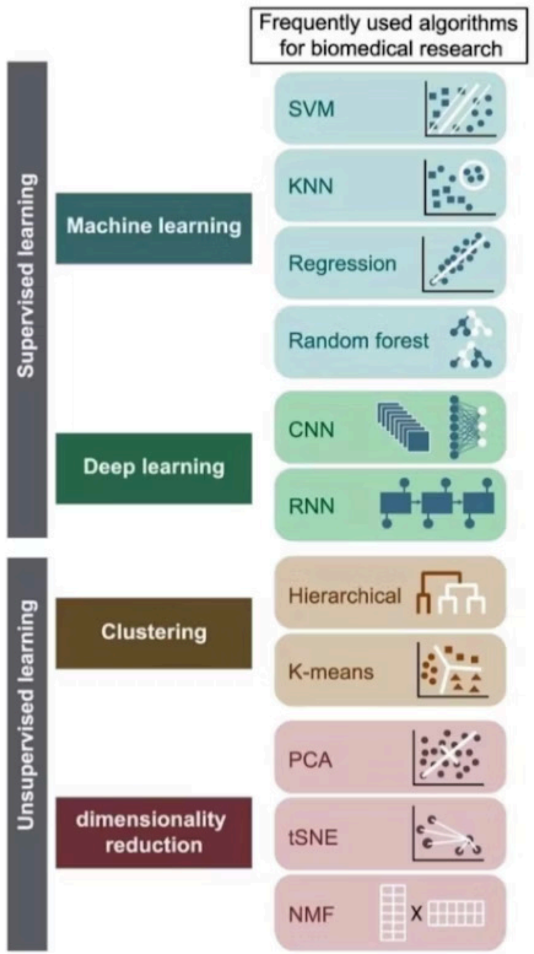
Advanced ML using neural networks with multiple layers to process complex patterns

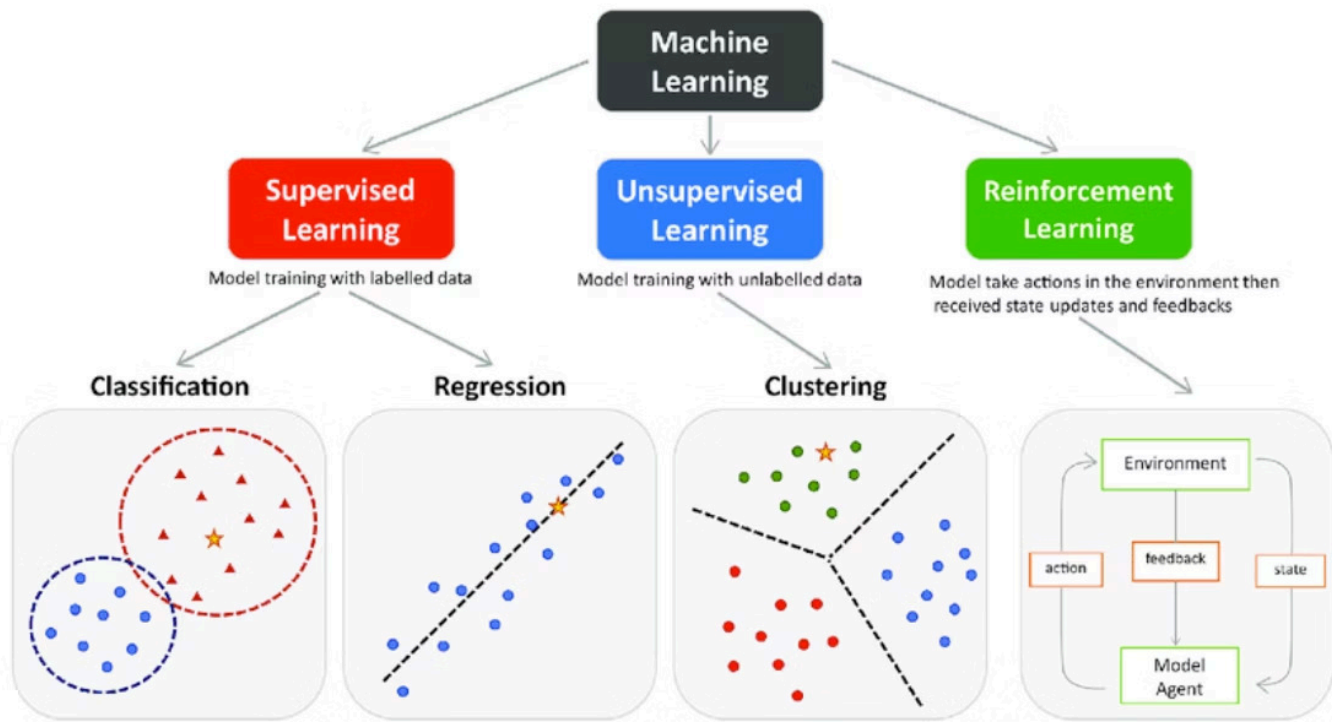
Traditional Programming vs. Machine Learning



- **Rule-based** vs **Pattern-based:** Traditional uses explicit instructions, ML discovers from data
- **Improvement Method:** Traditional requires manual updates, ML adapts automatically
- **Development Process:** Traditional needs step-by-step coding, ML builds own algorithms
- **Pattern Recognition:** Traditional handles simple rules, ML excels with complex relationships

MACHINE LEARNING
SUPERVISED vs. UNSUPERVISED





The History of AI

1 — 1940s-1950s: Birth of AI

McCulloch & Pitts create first computational model of a neuron (1943)

Turing Test proposed (1950)

Dartmouth Conference coins the term "Artificial Intelligence" (1956)

2 — 1960s-1970s: Early Optimism

Early AI programs demonstrate problem-solving abilities

Development of LISP programming language

First expert systems are developed

3 — 1974-1993: AI Winters

First AI Winter (1974-1980): Funding cut after unrealistic expectations

Brief revival with expert systems

Second AI Winter (1987-1993): Market collapse of specialized AI hardware

4 — 1990s-2010s: Revival

AI focuses on specific problems rather than general intelligence

Deep Blue defeats chess champion Garry Kasparov (1997)

Statistical approaches gain prominence

5 — 2010-2022: Deep Learning Era

Deep learning revolution begins with GPU acceleration

Transformer architecture enables large language models (2017)

Generative AI models like DALL-E and ChatGPT emerge (2022)

6 — 2023-Present: Generative AI and Agentic AI Era

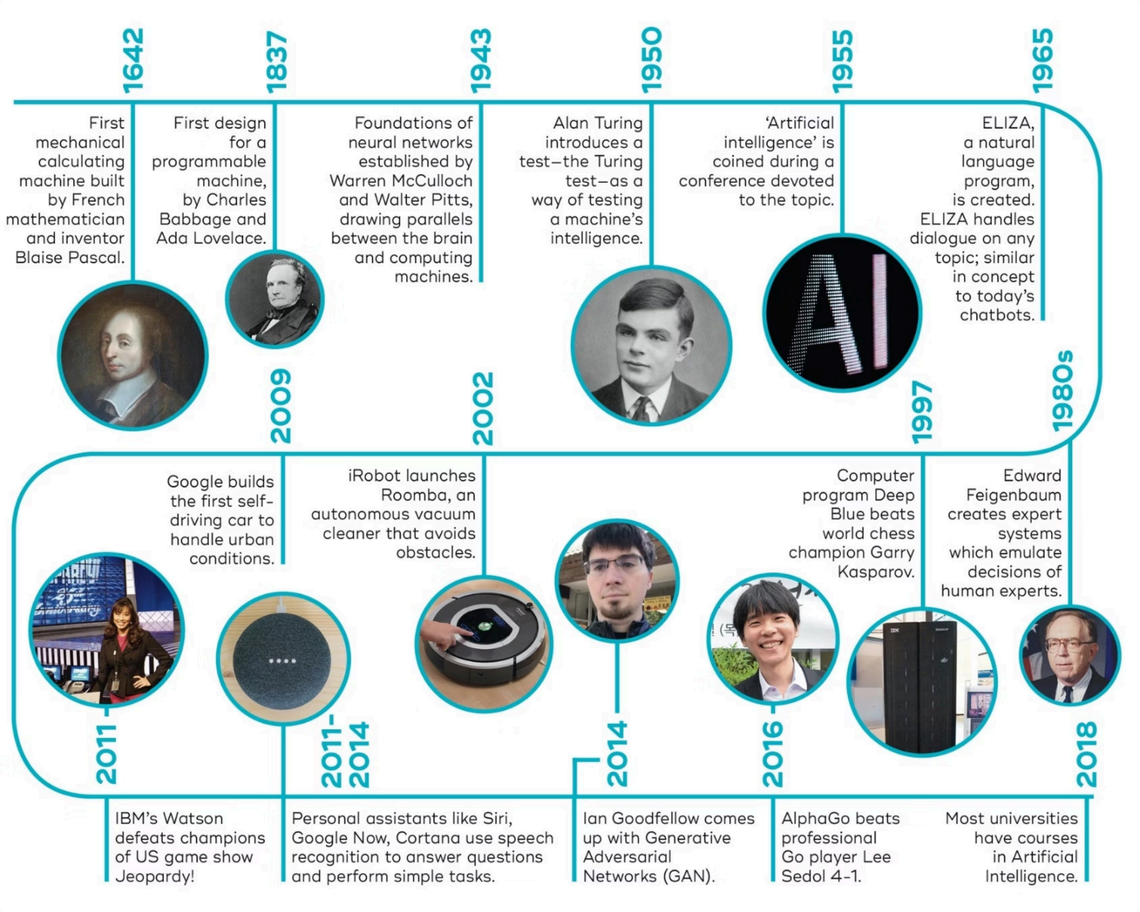
Widespread adoption and impact of Large Language Models (LLMs)

Rapid advancement in generative models (text, image, video, audio)

Emergence of agentic AI systems that can plan and execute complex tasks

The history of AI has been characterized by cycles of excitement and disappointment, with recent breakthroughs finally delivering on some of the field's earliest promises.

AI Milestones



Source: **https://qbi.uq.edu.au/brain/intelligent-machines/history-artificial-intelligence** with additions

The history of AI has been marked by cycles of hype and disappointment, but recent years have seen unprecedented practical advances.

Ethical Considerations in AI

As AI grows, addressing ethical implications is crucial for responsible development.

Fairness & Bias

AI can reinforce **societal biases** from unbalanced data, leading to discriminatory outcomes.

Transparency & Explainability

AI decisions, especially in critical applications, must be **understandable to humans.**

Privacy & Data Protection

AI's reliance on **vast personal data** raises serious concerns about privacy and security.

Safety & Accountability

Determining **responsibility for autonomous AI mistakes** (e.g., self-driving cars) is a complex challenge.

Impact on Jobs & Society

Widespread AI automation **may display jobs**, requiring societal adaptation and new economic models.

What Accounts for Recent Successes in AI?

Specialized Hardware

GPUs and purpose-built AI accelerators enable massive parallel processing.

Big Data

Unprecedented amounts of training data from the internet, sensors, and digital services.

Machine Learning Focus

Shift from rule-based systems to data-driven approaches that can discover patterns.

New Algorithms

Breakthroughs in deep learning architectures and optimization methods.

The convergence of these factors has created a perfect storm for AI advancement, enabling systems that outperform humans on specific tasks while becoming increasingly accessible to developers.

AI is Harder Than Originally Thought

> "It is not my aim to surprise or shock you– but ... there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until–in a visible future–the range of problems they can handle will be coextensive with the range to which human mind has been applied. **More precisely: within 10 years a computer would be chess champion, and an important new mathematical theorem would be proved by a computer**."

– Herbert Simon, 1957

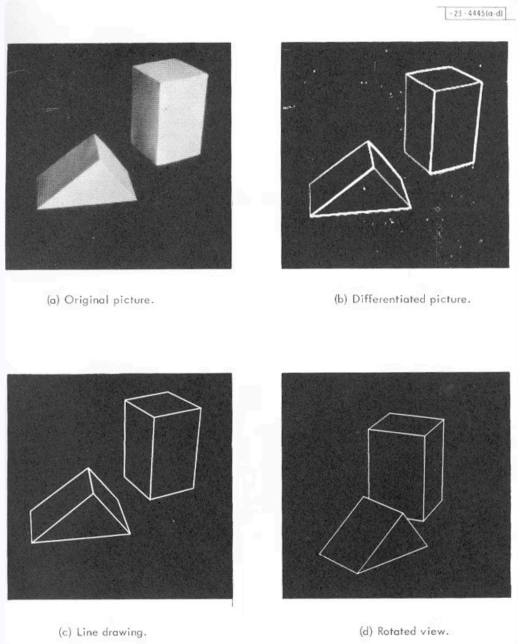
Simon's prediction came true — but 40 years later instead of 10

- Chess grandmaster defeated by Deep Blue in 1997 (40 years after prediction)
- Automated theorem proving became practical in the 1990s
- Many other "simple" human tasks proved extraordinarily difficult to automate

AI's history is marked by overly optimistic predictions. What seems simple to humans often involves complex cognitive processes, proving difficult for AI to replicate.

From Blocks World to Modern Object Recognition

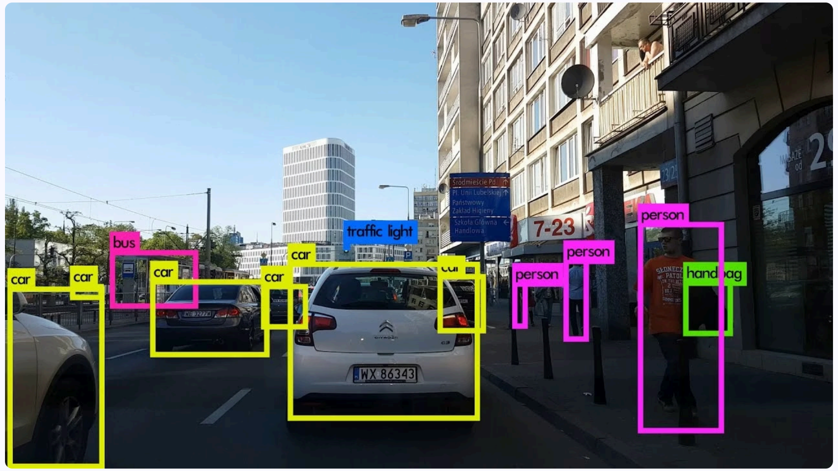
1963: Early Computer Vision



- Controlled Lighting
- Basic Geometric Shapes
- High Contrast Backgrounds
- Limited Object Types

These simplified "blocks worlds" were used to develop fundamental algorithms for edge detection and shape recognition.

Now: Real-World Computer Vision

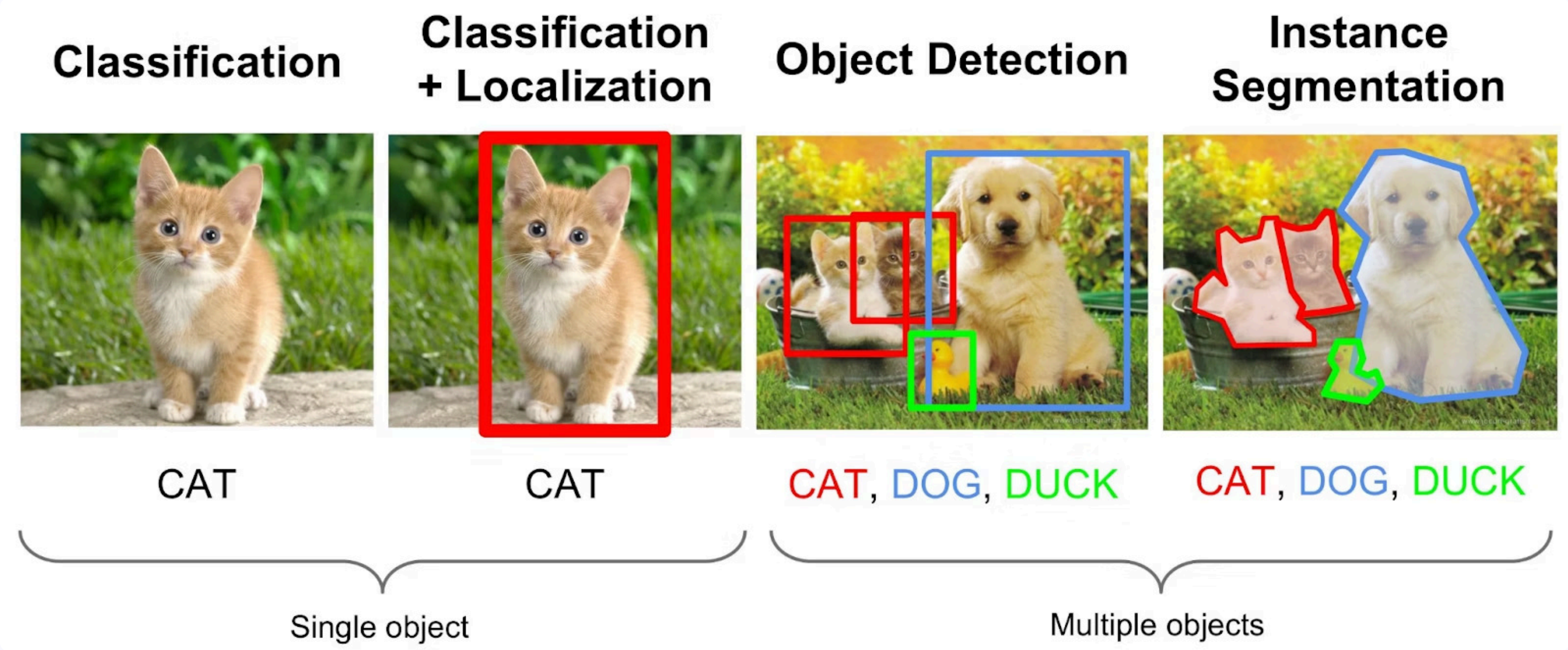


- Complex Natural Environments
- Variable Lighting Conditions
- Overlapping Objects
- Thousands of Object Categories
- Partial Occlusion
- Varying Perspectives

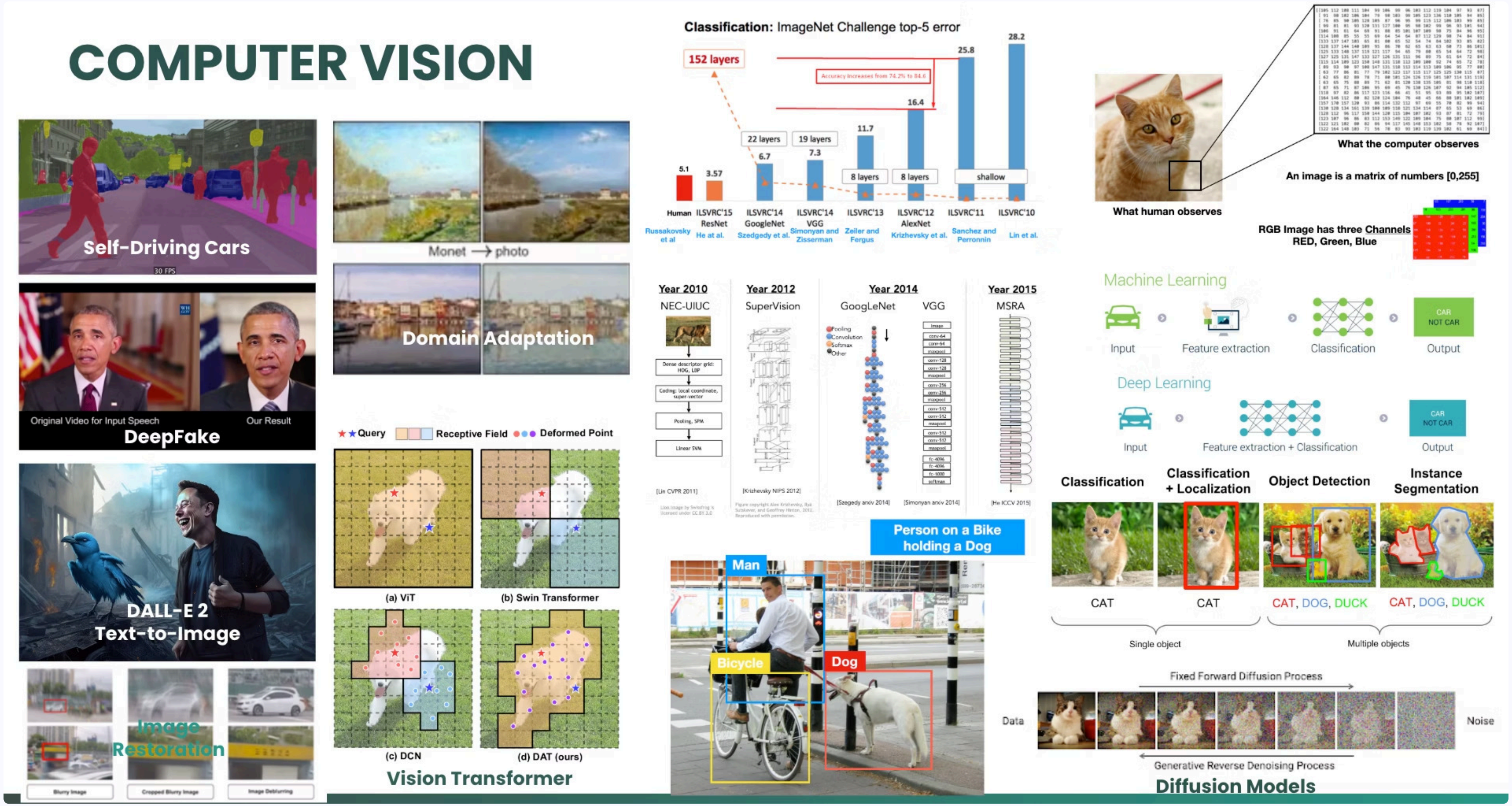
This is significantly more challenging, yet modern AI can now achieve superhuman performance in many visual recognition tasks.

The evolution from blocks world to modern computer vision illustrates both how we underestimated the complexity of "simple" human tasks and how far AI capabilities have advanced.

Computer Vision Tasks



Evolution of Computer Vision



Vision and Image Processing Applications

OCR (Optical Character Recognition)

Enables reading license plates, and handwriting recognition for mail sorting and document digitization.

Face Detection

Now standard for smartphone cameras, security systems, and social media tagging and verification.

Vehicle Safety

Powers features like lane detection, obstacle recognition, and driver monitoring in modern cars.

Visual Search

Allows users to find similar products online or identify objects and landmarks directly from images.

Image Generation

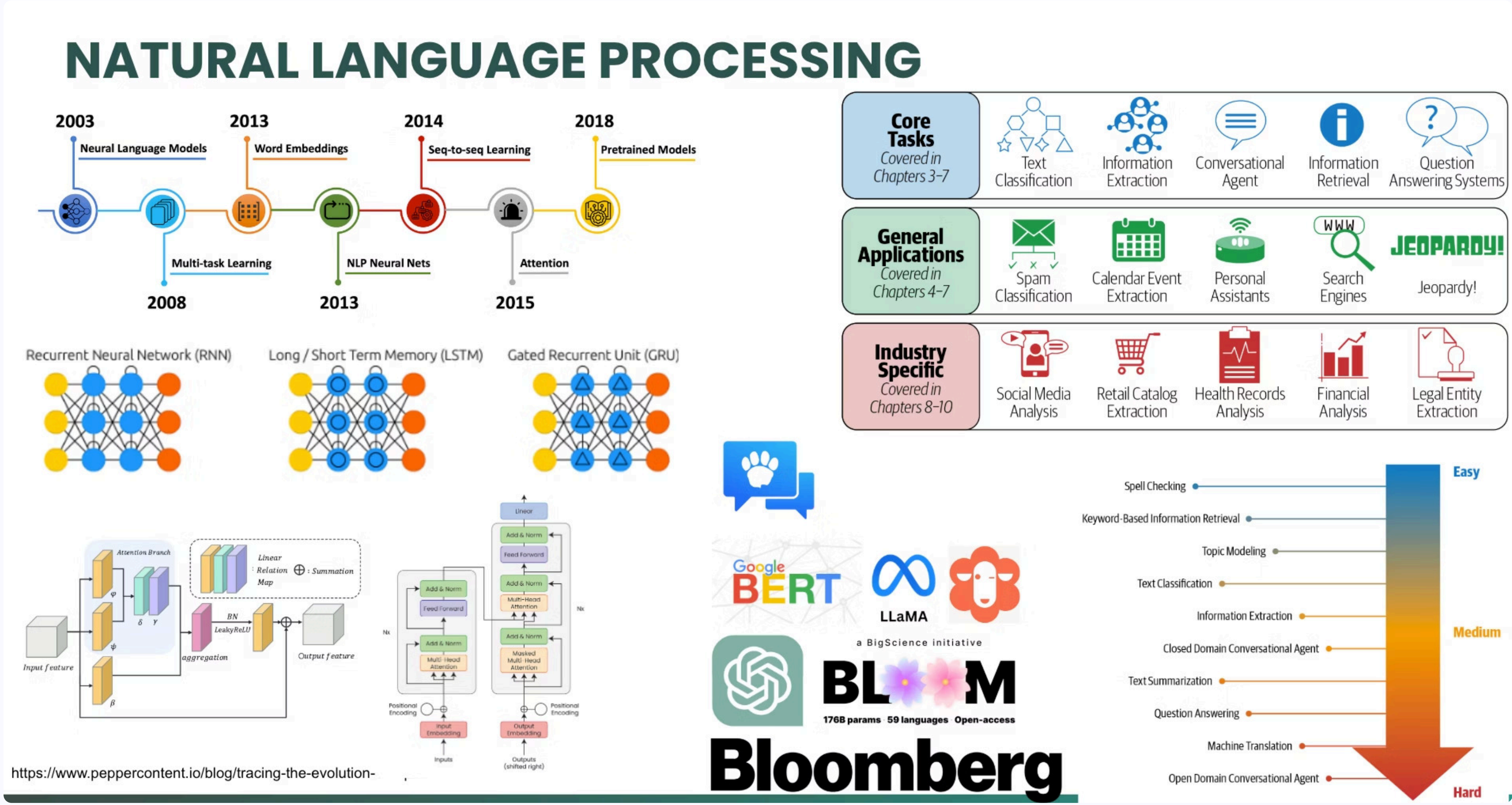
Utilizes AI to create entirely new and realistic images from simple text descriptions or other inputs.

Medical Imaging

Assists in detecting anomalies, diagnosing diseases, and guiding surgical procedures with enhanced precision.

Many of these technologies now operate at superhuman performance levels in their specific domains, transforming various industries.

Evolution of Natural Language Processing



AI Applications

Computer Vision
AI enables image understanding and detection

Natural Language
Comprehension, generation, and translation

Robotics
Autonomous control and adaptive behaviors

Healthcare
Diagnostics, treatment planning, and drug discovery

Finance
Fraud detection, risk modeling, and algorithmic trading

Gaming
Adaptive NPCs and procedural content generation

Transportation
Self-driving systems and traffic optimization

Manufacturing
Predictive maintenance and smart automation

Artificial intelligence has moved from research labs to become deeply integrated in numerous industries and everyday applications.

Throughout this course, we'll explore both the theoretical foundations and practical implementations across these diverse domains.