Artificial Intelligence

Introduction
Chapters 1 and 2

Madhavi K. Ganapathiraju
madhavi@cs.cmu.edu

Source of Content In these Slides

This course follows Russel & Norvig's AI A Modern Approach book
Technical content is presented verbatim or edited from the book
Figures as shown in the text book are provided by book creators

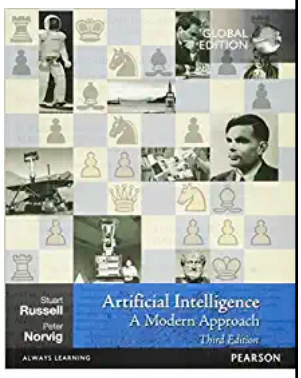
Other materials from the web are linked to their URLs in slides and URLs are also pasted at the end of the slide deck

Course Information

- **Wean Hall 2302 5:00 to 6:20**
- **Instructors:**
  - Jaime Carbonell (jgc@cs.cmu.edu)
  - Madhavi Ganapathiraju (madhavi@cs.cmu.edu)
- **Teaching Assistant:**
  - Sanket Mehta (svmehta@andrew.cmu.edu)
- **Meetings**
  - by appointment with instructors
  - During office hours with TA

Textbooks

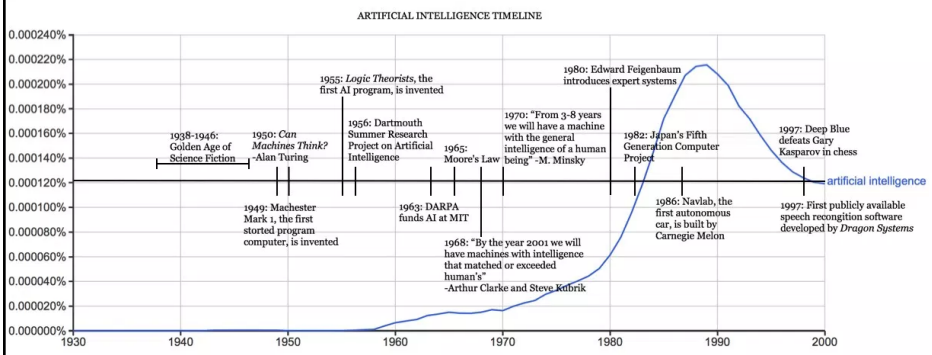
- **Stuart Russel & Peter Norvig's "AI: a modern approach"**
  - We are following 3rd edition
  - If you have 2nd edition, it should do.
- **Other books:**
  - 100 pages machine learning book for a quick intro to machine learning
  - Probabilistic graphical models by Koller and Friedman (relevant chapter numbers will be listed on course websites)



Schedule

Academic Calendar Day	Date	DoW	Lecture Number	Topic	Lecturer
Semester & Mini-1 classes begin					
	28-Aug	W	1	Introduction	Madhavi
LABOR DAY; NO CLASSES	2-Sep	M		HOLIDAY	
	4-Sep	W	2	Uninformed Search	Madhavi
Sem Course Add Deadline	9-Sep	M	3	Continue Uninformed Search + Informed Search	Madhavi
	11-Sep	W	4	Informed Search	Madhavi
	16-Sep	M	5	Constraint Satisfaction Problems	Madhavi
	18-Sep	W	6	Adverserial Search	Madhavi
	23-Sep	M	7	Pioneers in AI: Prof. Scott Fahlman	
	25-Sep	W	8	Logical Agents	Madhavi
	30-Sep	M	9	Classical Planning	Jaime
	2-Oct	W	10	Pioneers in AI: Prof. Raj Reddy	
	7-Oct	M	11	Logical Agents	Madhavi
	9-Oct	W	12	Review of Material for Midtern Exam	Madhavi/Sanket
	14-Oct	M		Midterm Exam	
	16-Oct	W	13	Ontologies	Madhavi
Mid-sem grades due	21-Oct	M	14	Proabilistic Reasoning	Madhavi
	23-Oct	W	15	Machine Learning	Madhavi
	28-Oct	M	16	Machine Learning	Madhavi
	30-Oct	W	17	Bayesian Networks	Madhavi
	4-Nov	M	18	Bayesian Networks	Madhavi
	6-Nov	W	19	Markov Decision Process	Madhavi
	11-Nov	M	20	Hidden Markov Models	Madhavi
	13-Nov	W	21	Reinforcement learning	Sanket
	18-Nov	M	22	AI Appplications in Biology, Medicine, Health	Madhavi
	20-Nov	W	23	Pioneers in AI: Prof. Jaime Carbonell	
	25-Nov	M	24	Explainable AI - Review of Papers	Students
THANKSGIVING - NO CLASSES	27-Nov	W		HOLIDAY	Thanksgiving Holiday
	2-Dec	M	25	Rule-based Systems	Jaime
LAST DAY OF CLASS (DEC 6)	4-Dec	W	26	Review of Material for Final Exam	
FINAL EXAM DAYS	9-Dec	M		NO CLASS	
FINAL EXAM DAYS	11-Dec	W		Final Exam	

History of AI



If you want to know more about these, come prepared to ask questions from invited speakers

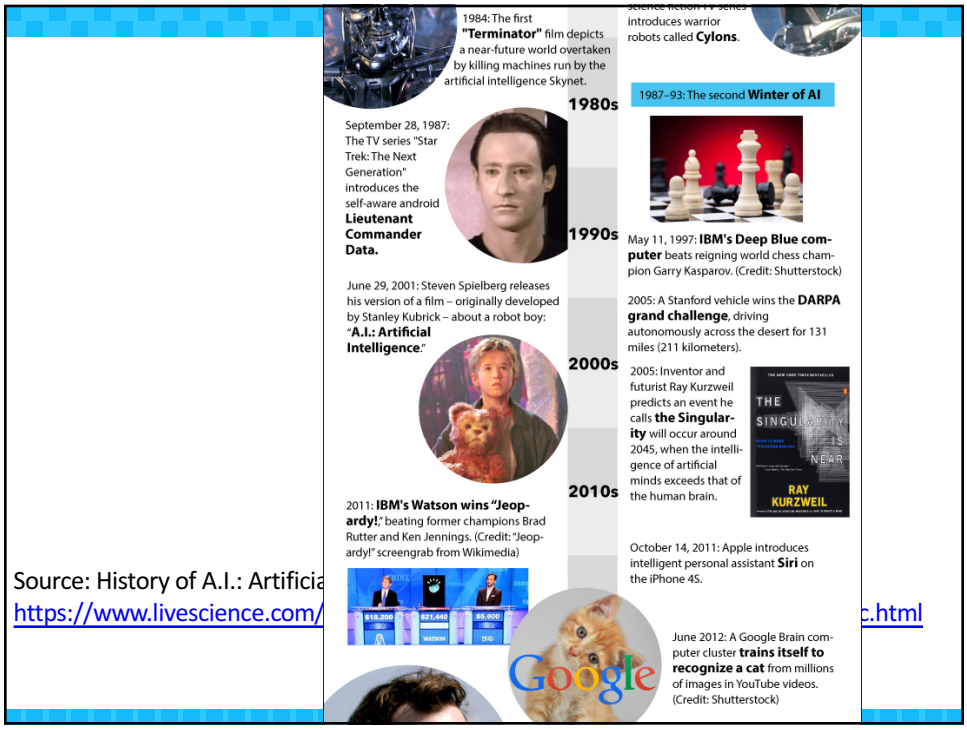
who created these histories: **Prof Raj Reddy**, **Prof. Fahlman**, **Prof. Carbonell**, (and other to-be-confirmed speakers) who are going to give talks in these lectures

Source: The History of Artificial Intelligence

http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

Also read, if interested in history of AI:

https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf



Source: History of A.I.: Artificia
https://www.livescience.com/ c.html

EXAMPLES OF

AI IN EVERYDAY LIFE

Take a few mins to..

- Think of an example area that involves technology but has no benefit from AI

Applications

AI in everyday life

- Clinical applications

Technical summary: Future of clinical development is on the verge of a major transformation due to convergence of large new digital data sources, computing power to identify clinically meaningful patterns in the data using efficient artificial intelligence and machine-learning algorithms, and regulators embracing this change through new collaborations. This perspective summarizes insights, recent developments, and

Applications

AI in everyday life

- **Clinical applications:**

Applications

AI in everyday life

- **Clinical applications**
- **AI enabled Electronic Medical Record Systems in Intensive Care Units**
  - Patients' risk of dying reduced by ½
  - Identified potentially deadly conditions that doctors missed
  - Patient recovery sped up: discharged 3 days early
- **AI-based Detection of Sepsis (Infection) reduced relative mortality by 58%**
- **Able to detect heart attack 5 mins before in 70% cases**

https://www-nbcnews-com.cdn.ampproject.org/c/s/www.nbcnews.com/mach/amp/ncna1032861

Applications

AI in everyday life

- Drug Discovery and Molecular Biology

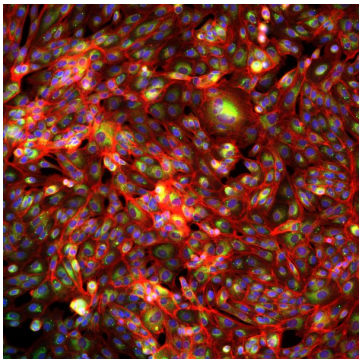
AI Drug Hunters Could Give Big Pharma a Run for Its Money

Alphabet's DeepMind cracked a problem that long vexed biologists, heating up a technological arms race in health care

Applications

AI in everyday life

- Drug Discovery
  - 400,000 - 500,000 miniature experiments per week
  - Robots apply thousands of drugs to various types of diseased cells
  - Generate 5 to 10 million cellular images
  - Machine-learning algorithms scan the images, searching for drugs that disrupt disease without harming healthy cells



https://www.bloomberg.com/news/features/2019-07-15/google-ai-could-challenge-big-pharma-in-drug-discovery

Applications

AI in everyday life

- **Transport:**

  This US city put an algorithm in charge of its school bus routes and saved $5 million

25 million US children travel on the iconic buses every day. Image: REUTERS/Brian Snyder

  - algorithm to identify the most efficient and cost-effective routes for 650 buses
  - Maps + uber/lyft + scheduling
  - Reduced costs to public transport and reduced carbon emissions

Applications

AI in everyday life

- **Personal wellbeing:**

  Can AI help prevent physician burnout?

Applications

AI in everyday life

- Crime prevention

Science & Research

Minority Report-style crime prevention with artificial intelligence is fast becoming reality

- Improvements in detection technology have all but solved credit card fraud and spam email
- Security agencies are deploying AI to detect abnormal behaviour that may suggest malicious intent

Chua Kong Ho
Published: 6:00am, 2 Apr, 2019

https://www.scmp.com/tech/science-research/article/3004167/minority-report-style-crime-prevention-artificial

Applications

AI in everyday life

- Transportation

The 25 Ways AI Can Revolutionize Transportation: From Driverless Trains to Smart Tracks

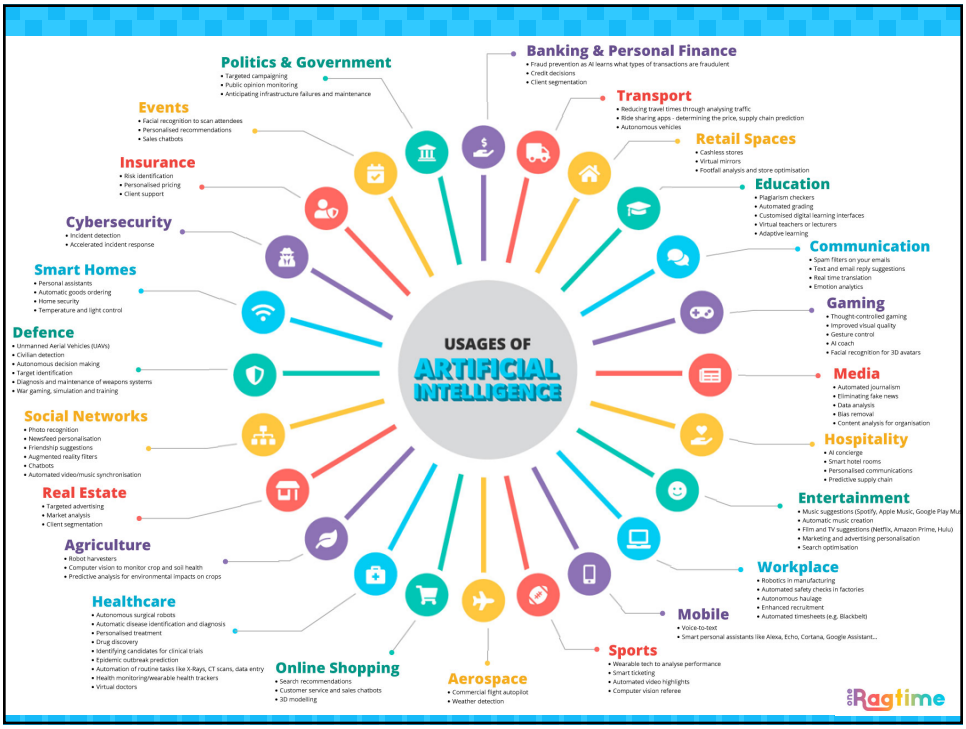
The transportation industry is in a state of flux thanks to ever-evolving technologies. Here's what to expect from AI in transportation.

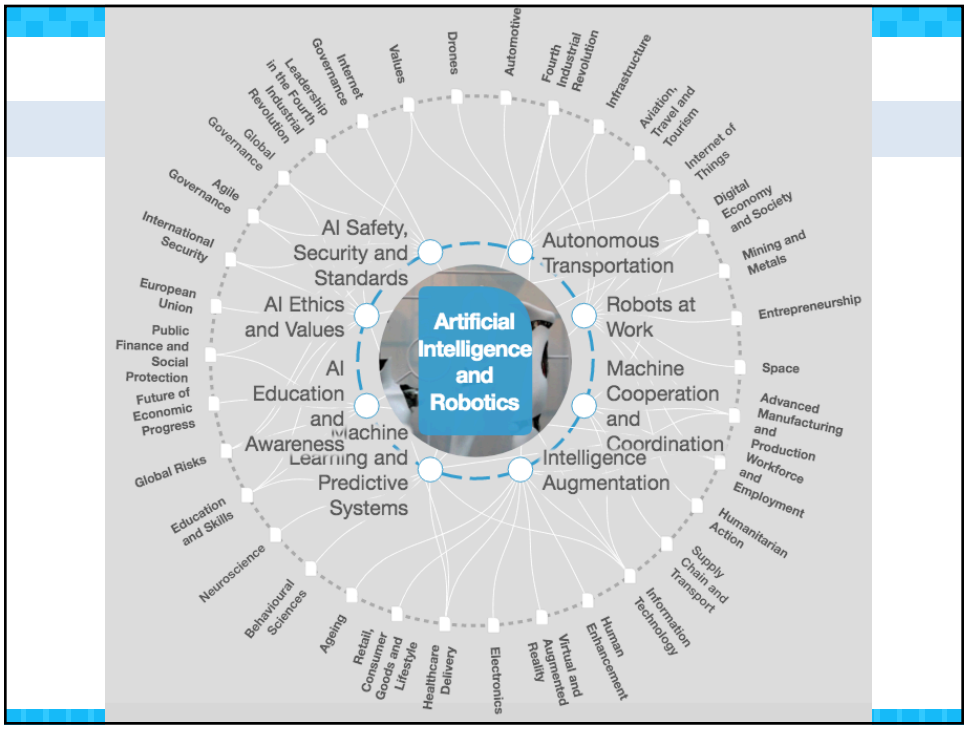
By Saoirse Kerrigan
April 22nd, 2018

Applications

AI in everyday life

- Transportation
  - Autonomous Cars, trucks, buses, trains, local carts, package delivery
  - Eliminating traffic jams, decentralizing traffic flow
  - Smart ticketing, baggage tracking
  - Smart number plates (alert in case of accident, pay fines, authenticate entry)
  - Face recognition for public and individual security, speeding up the process and personalization
  - Flight delay predictions and further arrangements
  - Taxi (Uber, Lyft, etc) and Ambulance matching & routing





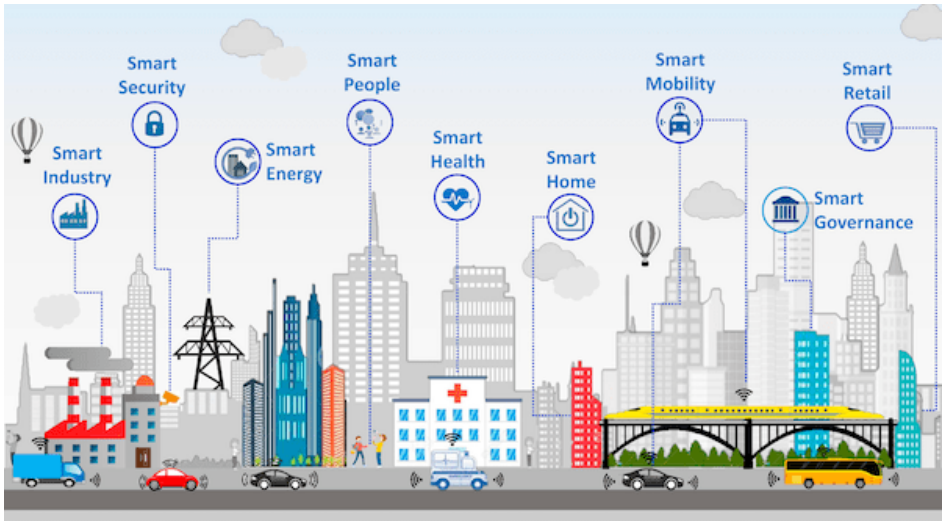
Examples of AI

AI systems are pervasive in modern world

- **Playing Chess**
- **Proving Mathematical Theorems**
- **Writing Poetry (Recognizing Sarcasm?)**
- **Driving a car**
- **Diagnosis of disease**
- **Predicting future events (not by astrology!!)**

- ***Minority Report* (2002) got it right:**
  https://www.youtube.com/watch?v=D0kovJJbcTA
  - Gesture-Based Computing
  - Driverless Cars
  - Personalized Ads
  - Voice-controlled Homes
  - Holographic Projection
  - Facial and Optical Recognition
  - Hovercraft
  - To a small extent, even PreCrime though not with same technology

Smart Cities



Is *intelligent* same as *smart*, w.r.t technology?

EXAMPLES OF

AI RESEARCH

Human Intelligence

For 1000's of years we have tried to understand

- How we think
- How our brain perceives, understands predicts and manipulates a world that is so big and complicated

Artificial Intelligence

- AI goes further, and attempts to build other intelligent systems
- AI name coined in 1956 (?)
  - Prof Fahlman's PhD thesis probably first one in AI (1973)

Pioneers in AI talk by Prof Fahlman is on September 23rd!

Readings

1988 AAAI Presidential Address

Foundations and Grand Challenges of Artificial Intelligence

Raj Reddy

Pioneers in AI talk by Prof Reddy is on October 2nd!

Last November, I got a call from Bob Simpson and Allen Sears, program managers for Artificial Intelligence at DARPA, which has been the major sponsor of AI research for the past twenty-five years. It was a call for help. "We are being asked some tough questions by the front office," they said. "What are the major accomplishments of the field? How can we measure progress? How can we tell whether you are succeeding or failing? What breakthroughs might be possible over the next decade? How much money will it take? What impact will it have? How can you effect technology transfer of promising results rapidly to industry?" They needed the answers in a hurry.

with the potential for highest payoff. As the size of investment in AI rises above the noise level, we can no longer expect people to fund us on blind faith. We are entering an era of accountability. Rather than being concerned, I think we should view this as a challenge and lay out our vision for the future.

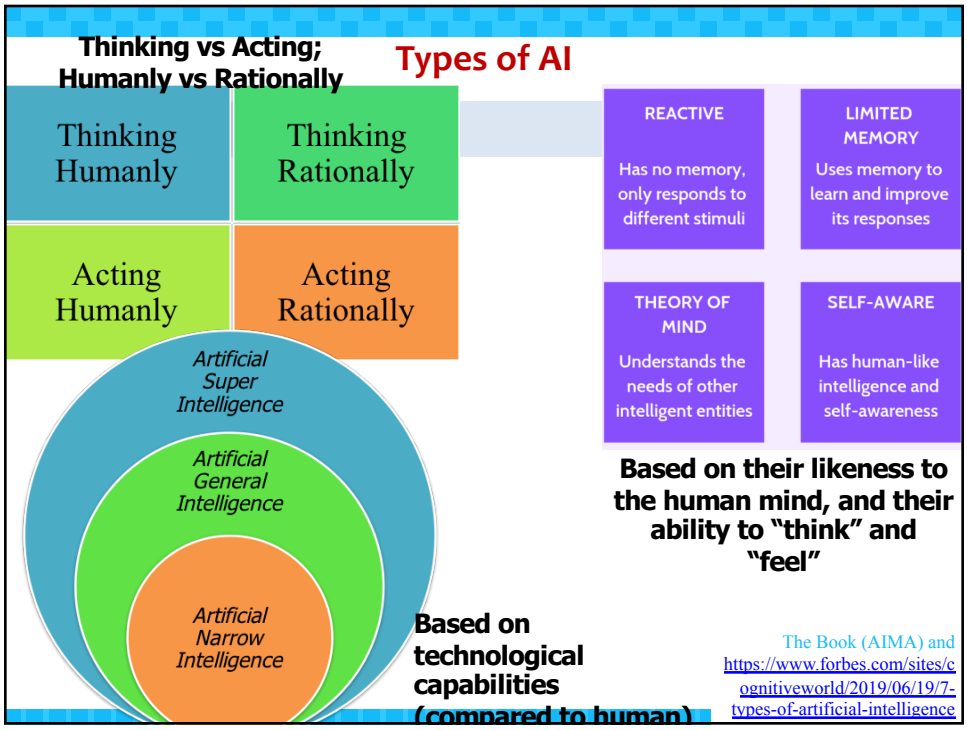
My first reaction to Simpson's request was, "Hey, we have nothing to worry about." Our track record is terrific. We have created almost a billion dollar-a-year enterprise around the globe. Previously unsolvable problems are being solved. People who have never used a computer before are solving problems with one. Vision and robotics are beginning to transform manufacturing. Planning and schedul-

AI and Molecular Biology are the areas that (apparently) most scientists say they wish they were working on!

I will present this applications-oriented talk on November 18th

Objectives of AI Research or

AI TYPES



Types of AI – 2

REACTIVE
Has no memory, only responds to different stimuli

Reactive Machines

- Oldest forms of AI systems
- Extremely limited capability (but have several applications)
- They emulate the human mind's ability to respond to different kinds of stimuli
- Do not have memory-based functionality
- Experience doesn't make them better
- Popular example: IBM's Deep Blue, a machine that beat chess Grandmaster Garry Kasparov in 1997

Types of AI - 2

Limited Memory

LIMITED MEMORY
Uses memory to learn and improve its responses

- Reactive machines + learn from experience
- Most applications today
  - Machine learning from labeled data
- Almost all present-day AI applications, from chatbots and virtual assistants to self-driving vehicles are all driven by limited memory AI

Types of AI - 2

Theory of Mind

THEORY OF MIND
Understands the needs of other intelligent entities

Types of AI - 2

Theory of Mind

THEORY OF MIND

Understands the needs of other intelligent entities

- Able to better understand the entities it is interacting with by discerning their needs, emotions, beliefs, and thought processes
- An area of interest for leading AI researchers
- Currently concepts or a work in progress
  - Artificial emotional intelligence is a budding industry
- Requires development in other branches of AI as well:
  - truly understand human needs
  - perceive humans as individuals whose minds can be shaped by multiple factors
  - essentially "understanding" humans

Types of AI - 2

Self-Aware

SELF-AWARE

Has human-like intelligence and self-awareness

- AI that has evolved to be so akin to the human brain that it has developed self-awareness.
- **C**urrently exists only hypothetically
- Ultimate objective of all AI research?
- Able to understand and evoke emotions in those it interacts with, but also have emotions, needs, beliefs, and potentially desires of its own
- And this is the type of AI that doomsayers of the technology are wary of.
- Although the development of self-aware can potentially boost our progress as a civilization by leaps and bounds, it can also potentially lead to catastrophe
  - Just like humans can do good and do harm to society, so can AI also (but likely with much higher impact due to its capabilities

Types of AI - 3

Technology Related

- Artificial Narrow Intelligence (ANI)
- Artificial General Intelligence (AGI)
- Artificial Superintelligence (ASI)

Types of AI - 3

Artificial Narrow Intelligence

- Existing AI, including the most complicated and capable AI
- Systems that can perform a specific task autonomously using human-like capabilities
- Can do nothing more than what they are programmed for
- Have a very limited or narrow range of competencies
- All the reactive and limited memory AI
- Even the most complex AI that uses machine learning and deep learning to teach itself falls under ANI

Types of AI - 3

Artificial General Intelligence

- These AI systems are just as capable as humans by replicating our multi-functional capabilities
- Ability to learn, perceive, understand, and function completely like a human being
- Able to independently build multiple competencies and form connections and generalizations across domains
- Can massively cut down on time needed for training

Types of AI - 3

Artificial Super Intelligence

- Human Capabilities + Faster Computation and Total Recall
- Will mark the pinnacle of AI research
- Exceedingly better at everything they do because of overwhelmingly greater memory, faster data processing and analysis, and decision-making capabilities
- Will lead to technological *Singularity*, the point at which machines intelligence and humans would merge
- Exponential growth in technology, miniaturization capabilities, mechanics, materials, genetics, biological systems, algorithms, computational power
- And while the potential of having such powerful machines at our disposal seems appealing, these machines may also threaten our existence or at the very least, our way of life

Technological Singularity

Popularized by Ray Kurzweil

- ".. in the very near future that technological advancement will be so fast,
- that we wont be able to keep-up,
- unless we augment ourselves with the technology we are creating
- Improving our physiological selves with advancements
  - Biotechnology
  - Nanotechnology
  - Artificial intelligence
- we'll become a human-machine civilization, and …"
- From https://transcendentman.com

Types of AI

Reasoning vs Behavior; Humanly vs Rationally

Thinking Humanly	Thinking Rationally
Acting Humanly	Acting Rationally

Thinking Humanly | Thinking Rationally
Acting Humanly | Acting Rationally

Acting Humanly

The Turing Test Approach

- **Proposed in 1950**
- **Designed to provide a satisfactory operational definition of intelligence**
- **A computer passes the test if a human interrogator, after posing some questions, cannot tell whether the responses come from a person or from a computer**
  - Questions and Responses are in written form
  - Total Turing Test if computer also has computer vision to and robotics to manipulate objects and move about

What abilities would the computer need?

Thinking Humanly | Thinking Rationally
Acting Humanly | Acting Rationally

Turing Test

Computer Needs Abilities in:

- **Natural Language Processing**
  - To communicate
- **Knowledge Representation**
  - To store what it knows or learns
- **Automated Reasoning**
  - To use stored information to draw conclusions and answer question correctly
- **Machine Learning**
  - To adapt to new circumstances and to detect patterns and extrapolate patterns

Turing Test remains relevant event today

Pioneers in AI talk by Prof Carbonell is on Nov. 20th!

Thinking Humanly | Thinking Rationally
Acting Humanly | Acting Rationally

Total Turing Test

Extends Ability for Physical Interaction

- **Computer Vision**
  - To perceive objects
- **Robotics**
  - To manipulate objects and move about

.. These 6 fields compose most of AI

Thinking Humanly | Thinking Rationally
Acting Humanly | Acting Rationally

However..

- **Imitating humans is not the most valuable aspect of AI**
- **For example, aeroplanes have become so advanced despite not really trying to imitate pigeons so perfectly that other pigeons can't distinguish them**

Thinking Humanly | Thinking Rationally | Acting Humanly | Acting Rationally

Thinking Humanly

Cognitive Modeling Approach

- Understand the workings of human mind

Methods for this include:

- Introspection (watch one's own thoughts)
- Psychology experiments
- Brain imaging

Allen Newell and Herbert Simon (1961) developed General Problem Solver: Not only match input-output pairs to human's (Acting Humanly) but also to trace reasoning

Thinking Humanly | Thinking Rationally | Acting Humanly | Acting Rationally

Thinking Humanly

Cognitive Modeling Approach

What is it?

A cylindrical can that is dented.

Why not a complicated shape can to begin with?

SATURN

My nephew, a toddler, who was just beginning to learn alphabet, read it fluently as S V T U R N

Does that show the brain thinks of rotation as a more obvious explanation than deletion?

Thinking Humanly	Thinking Rationally
Acting Humanly	Acting Rationally

Thinking Humanly

Cognitive Modeling Approach

- Cognitive Science brings computer models from AI and experimental methods from Psychology to construct precise and testable theories of human mind.
- Cognitive Science and AI continue to fertilize each other and grow.
- E.g. Computer vision and speech recognition incorporate neuropsychological evidence in computational models

What is Cognitive Science?

- **Study of Intelligent Systems**
  - Eg: The Human Mind

Slide credit: Dr. Monika Krishan

What is Intelligence?

- **Complex Construct**
- **Ability to survive ?**
  - Cockroaches have existed for millions of years, in their present form
  - More intelligent than humans?
- **Ability to respond "rationally" to the environment ?**
  - Humans are notorious for acting "irrationally"
  - Exhibit biased reasoning
  - Eg: Confirmation Bias
  - Tendency to selectively seek out evidence that supports a pre-existing belief, ignoring or devaluing contradictory evidence

Slide credit: Dr. Monika Krishan

Working Definition of Intelligent Behaviour

- **Ability to**
  - Respond to questions coherently
  - Solve problems of varying difficulty
  - Set goals for themselves and pursue appropriate courses of action towards achieving these goals
  - Create something novel or useful
- **Assumption**
  - Set of common processes that enable the above behaviors
- **Cognitive Science**
  - Study of these processes

Slide credit: Dr. Monika Krishan

Domains of Study within Cognitive Science

- **Memory**
  - What are the limits of memory?
  - Where is memory located in the brain?
  - How does the nature of the stimulus affect its encoding?
  - eg: Does the structure of the stimulus affect how easily it is
  - is recalled?
- **Vision**
  - How does the visual system derive a 3D representation of the world from a 2D image on the retina?
  - How does the process of visual recognition occur?
  - How are visual and non-visual features integrated ?

Slide credit: Dr. Monika Krishan

Domains of Study within Cognitive Science

- **Language**
  - Learning & Acquisition
  - Is language necessary for intelligent thought?
- **Reasoning & Decision Making**
  - Do humans have a common logic of reasoning ?
  - How do they act under conditions of uncertainty in the environment?
- **Cognitive Development**
  - How does the infant's intelligent system develop?
  - How does it differ from that of an adult?
  - Is it just a difference of "better" or "more"?
  - Or is it a "different" system – architecture, processes?

Slide credit: Dr. Monika Krishan

Cognitive Science: Multidisciplinary Field of Study

- **Informed by:**
  - Philosophy
  - Psychology
  - Linguistics
  - Computer Science
  - Neuroscience
  - Anthropology
  - Economics

Slide credit: Dr. Monika Krishan

The Stroop Task

Name the color with which each of these words have been written, aloud

Example: yellow

Answer: red

blue white red yellow pink red orange green orange
blue red yellow pink black red white yellow orange
pink red blue green purple blue white green orange
blue red orange yellow green pink black red white
black yellow white green yellow pink red green blue
black purple blue green

Slide credit: Dr. Monika Krishan

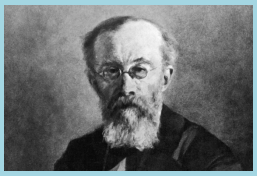
The Stroop Task

- **Note the difficulty in ignoring the color indicated by the word**
- **Reflects *automaticity* of reading**
- **Need to *inhibit* or supress textual information to carry out the task**
- **Language and visual systems can compete for attention**
- **Inhibition is a key aspect of Attention, allowing us to selectively attend to stimuli**
- **Prevents one from being overwhelmed by the environment**

Slide credit: Dr. Monika Krishan

Thinking Humanly: Methods Used to Study Human Thinking

1. **Introspection (Wilhelm Wundt, 1879)**
   - Subjects asked to reflect on their thoughts, images and feelings in response to controlled stimuli (eg: light or sound)



   - Ended around 1920s
   - Highly Unreliable
   - Inconsistent responses to identical stimuli
   - Act of introspection can itself alter the experience
   - Mental processes are largely inaccessible to the conscious mind

eg: *Your perception of this 3 dimensional room with its objects and people at specific locations is immediate. Can you describe the mental steps/processes that lead to this percept?*

Slide credit: Dr. Monika Krishan

Thinking Humanly: Methods Used to Study Human Thinking

2. Experimental Study of Behaviour

- Human subjects
- Assigned simple specific tasks (input)
- Performance (output) recorded
- Inferences drawn about mental processes/capacities that lead to these outputs

- 2 kinds of Experimental studies

  a) **Hypothesis Testing**: Tasks designed to test specific hypotheses, typically conducted in a controlled lab setting
  - Eg: "People can hold 7-9 discrete items of information in their working memory at a time"

  b) **Observational or Exploratory Studies**: "See what might happen" approach, often in naturalistic environments
  - Eg: "Are people at pizzerias happier than those at coffee shops"

Slide credit: Dr. Monika Krishan

Thinking Humanly: Methods Used to Study Human Thinking

3. Anatomical and Physiological Techniques

- Study of the Physical Brain

  Through injuries resulting from strokes, wounds, illnesses

- eg: Clive Wearing Contracted Encephalitis in 1985
- Hippocampus Damaged
- Loss of all memory prior to illness
- Inability to form lasting new memories

Slide credit: Dr. Monika Krishan

Thinking Humanly: Methods Used to Study Human Thinking

The Case of Clive Wearing

- 20 second memory on average
- Max of 2 minutes
- However ….
- 2 past memories fully preserved
- Guess ?
- Piano skills and Wife !!!
- Implications

⇒ Different *Types of* Memory

⇒ Stored in Different Areas

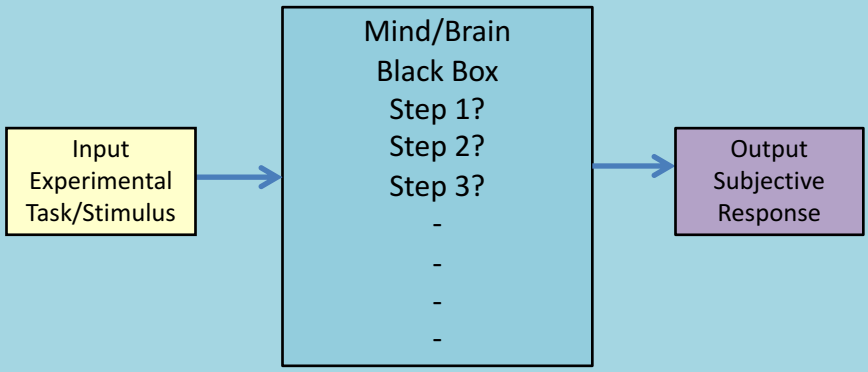
Slide credit: Dr. Monika Krishan

Methods Used to Study Human Thinking: Anatomical and Physiological Techniques…. continued

- Animal Studies
- Imaging of blood flow in the brain during a Task
- Recording of electrical activity from single neurons
- Brain studies useful
  Constrain possible models of Intelligence (human)
- Brain studies challenging to interpret
- Mapping brain activity/structure to specific mental functions not straightforward!
- eg: The brains of octo/nonagenarian nuns who lived active lives to the very end were found post-mortem to exhibit signs of advanced Alzheimer's Disease. (no symptoms!)
- Physical and functional aspects of the brain not necessarily correlated

Slide credit: Dr. Monika Krishan

Computational Modelling of Cognitive Processes



A model of the computational *process* that transforms the input into the observed output

Not merely and input-output plot!

Slide credit: Dr. Monika Krishan

Thinking Humanly | Thinking Rationally | Acting Humanly | Acting Rationally

Thinking Rationally

"Laws of Thought" Approach

- **Aristotle – codify right thinking**
- **Syllogisms: patterns for argument structures that always yield correct conclusion given premise**
- **Socrates is a man**
  - ➔ **All men are mortal**
  - ➔ **Socrates is mortal**
- **These laws of thoughts were supposed to govern operations of mind. Origins of *Logic***

Thinking Humanly	Thinking Rationally
Acting Humanly	Acting Rationally

Thinking Rationally

"Laws of Thought" Approach

- By 1965 programs existed that in principle could solve any solvable problem that was described in logical notation
- Logicist approach hopes to build on such problems to create intelligent systems

Thinking Humanly	Thinking Rationally
Acting Humanly	Acting Rationally

Acting Rationally

Rational Agent Approach

- An agent is something that acts
- All computer programs do something. But agents are expected to:
  - operate autonomously
  - Perceive their environment
  - Persist over prolonged time
  - Adapt to change
  - Pursue goals
- A rational agent is one that acts so as to achieve the best outcome (or best expected-outcome in the presence of uncertainty)

Thinking Humanly	Thinking Rationally
Acting Humanly	Acting Rationally

Acting Rationally

Rational Agent Approach

- Acting rationally also involves thinking rationally but sometimes there may not be a "best" outcome, or sometimes taking "some" action may be more appropriate that deliberating all possible options (e.g. taking hand off a hot plate)
- Skills required for Acting Humanly (knowledge representation and reasoning) also useful

THIS BOOK/COURSE FOCUS ON GENERAL PRINCIPLES OF RATIONAL AGENTS

Introduction to

RATIONAL INTELLIGENT AGENTS

Rational Agent

- **An agent is something that acts**
- **All computer programs do something**
- **Agents are expected to**
  - Act autonomously
  - Perceive their environment
  - Persist over prolonged time
  - Adapt to change
  - Persue goals
- **A *Rational Agent* is one that acts so as to achieve the *best* outcome in the presence of uncertainty**

Rational Agent

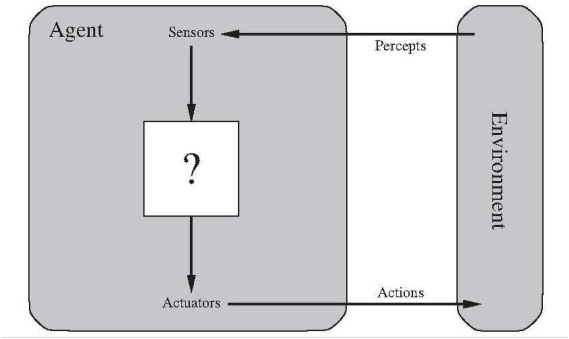
- **A *Rational Agent* is one that acts so as to achieve the *best* outcome in the presence of uncertainty**
- **It also involves *Thinking Rationally* but sometimes:**
  - There may not be a best outcome
  - Sometimes taking "some" action is better than thinking through all options (e.g. take hand OFF of a hot plate)
  - Skills required for *Acting Humanly* also help a rational agent
- **The text book AIMA, and therefore this course which is based on the book, focus on general principles of Rational Agents and on components for constructing them**

INTELLIGENT AGENTS

Intelligent Agents

Agents ⇔ Environments

- An agent perceives the environment through sensors
- Acts on the environment through actuators



Rational Agent

- Percept: sensory input at a given instant
- Percept sequence: history of inputs till date
- An agent's choice of action can depend on percept or percept sequence, but not on future sensory inputs
- An agent's behavior is determined by the agent function that maps percept sequence to action
- Agent Function is a mathematical description
- Agent Program is an implementation running on a physical system

Rational Agent

Good Behavior

- **Concept of Rationality:** based on or in accordance with reason or logic
- A rational agent "does the right thing"
- A sequence of actions by agents causes the environment to go through a sequence of states
  - If the sequence of states is "desirable", the agent performed well
- **Robot needs to cross the road**
  - Sensors: camera, microphone and radar to sense hurdles, road and traffic through
  - Actuators: accelerator, break, steering to make itself move
- **What is the *environment* that goes through a sequence of states?**

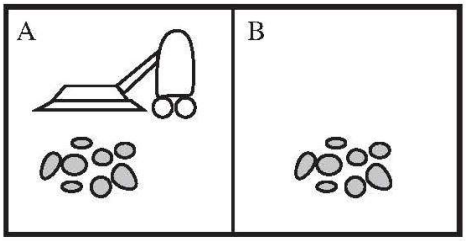
Rational Agent

Good Behavior – Concept of Rationality

- **A rational agent does the *Right Thing***
- **Actions cause changes to state of the system**
- **If actions' changes are desirable then the agent has performed well**
- **Performance measure evaluates the desirability of the sequence of environment states**

Rational Agent

Vacuum Cleaner Example



- **It can sense "clean or dirty"**
- **Can move left or right, suck and dump**
- **Goal is to clean the two locations A and B**
- **Performance Measure options:**
  - *suck max amt of dirt (problem: suck, dump, suck)*

Better to set performance measure based on desirable outcome rather than what you think an agent should do

Rational Agent

What is Rational

- **What is Rational at any given time depends on 4 aspects:**
  - Performance measure that defines what is success
  - Agent's prior knowledge of the environment
  - Actions that the agent can perform
  - Agent's percept sequence to date

Definition: For each percept sequence, a rational agent should select an action that is *expected* to maximize its performance according to the performance measure given the evidence provided by the percept sequence and its own knowledge base

URLs to content presented in slides

The History of Artificial Intelligence: http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/
https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf
History of A.I.: Artificial Intelligence (Infographic): https://www.livescience.com/47544-history-of-a-i-artificial-intelligence-infographic.html
https://www.media.mit.edu/projects/artificial-intelligence-for-drug-discovery-and-clinical-trials/overview/
https://www-nbcnews-com.cdn.ampproject.org/c/s/www.nbcnews.com/mach/amp/ncna1032861
https://www.bloomberg.com/news/features/2019-07-15/google-ai-could-challenge-big-pharma-in-drug-discovery
https://www.youtube.com/watch?v=8_l85n1OZ6U&feature=youtu.be
https://www.weforum.org/agenda/2019/08/this-us-city-put-an-algorithm-in-charge-of-its-school-bus-routes-and-saved-5-million/
https://www.medicaleconomics.com/technology/can-ai-help-prevent-physician-burnout
https://www.scmp.com/tech/science-research/article/3004167/minority-report-style-crime-prevention-artificial
https://interestingengineering.com/the-25-ways-ai-can-revolutionize-transportation-from-driverless-trains-to-smart-tracks
https://intelligence.weforum.org/topics/a1Gb0000000pTDREA2?tab=publications
https://www.youtube.com/watch?v=D0k0vJJbcTA
http://thecorememory.com/Reddy.pdf
https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence
https://transcendentman.com