Overview of Al

Dr. Thomas R. loerger CSCE 420 – Fall 2024

read: Ch. 1

1

What is Al?

- popular view
 - based on movies, sci-fi...
 - Hal from 2001; Data from Star Trek NG; Spielberg's AI; Ex Machina...
- landmark AI systems
 - Deep Blue (chess), AlphaGo
 - IBM Watson (general question-answering)
 - Siri, Alexa, ChatGPT
- real industrial AI applications
 - robotics, autonomous vehicles
 - mail sorting, OCR; face recognition; speech2text, translation...
 - medical diagnosis, chest radiography, protein structure prediction (AlphaFold), drug design...
 - fraud detection, legal precedent finding,
 - stock trading/portfolio risk management/quant funds/economic forecasting...

Intelligence

- What is Intelligence?
 - difficult to define
 - the thing the makes humans unique. what about animals, dogs, dolphins...?
- some properties of intelligent behavior
 - doing more than just reacting to the environment
 - acting without being programmed (autonomous decision-making)
 - having volition/motivation/intent
 - does this require free-will? consciousness? self-awareness?
 - learning from experience, adaptiveness
- What do "IQ" tests measure?
 - there are many dimensions to intelligence, not just factual knowledge or reasoning ability

Aspects of Intelligence

- acting *like a human* vs acting *ideally* what's the target?
 - rationality doing the "right" thing or "optimal" thing
 - implies there are goals/objectives
- thinking vs. doing distinction
 - are outcomes all that matter, or is how we "think" about it relevant?

 Thinking ideally (rationally) did we weigh all the pros and cons? did we make the right inference? did we do it for the right reasons? 	 Acting ideally (rationally) if we can quantify goals and costs, we can define optimal actions usually NP-hard
 Thinking humanly can we emulate decisions a human would make? practical choices in face of incomplete info. humans make mistakes, are biased by emotions 	 Acting humanly play chess like a human (grandmaster) think about "good" driving are humans always the best measure?

Perspectives on Al

- People view the *goals* of AI from different perspectives:
 - Philosophical Perspective
 - What is intelligence? Is it even possible for a machine to be intelligent?
 - Psychological Perspective
 - Al should be about creating programs that emulate human cognition.
 - Engineering Perspective
 - Forget the theoretical questions this is practical!
 - How do we build more sophisticated systems for solving more complex problems? (like autonomous driving)
 - AI is all about advanced algorithms, complexity reduction, and optimization

- attempts to define "correct thought" have a long history
 - Aristotle, Plato...Descarte, Kant, Carnap...Russell, Wittgenstein, Frege...
 - innate vs. empirical knowledge
- syllogisms laws of thought, or patterns of correct thought
 - Categorical syllogisms: All A are B; x is A; therefore x is B. Socrates is a man; men are mortal; therefore, Socrates is mortal.
 - Conditional syllogisms:

Only people with good grades get into college; Kate is in college; Kate must have gotten good grades.

• Disjunctive syllogisms:

All living things are plants or animals; Snoopy is not a plant; therefore, Snoopy is an animal.

• examples of a fallacy:

* Insects crawl; Charlotte crawls; therefore, Charlotte is an insect.

[incorrect - she is a spider, which is an arachnid, not an insect]

* Whenever it rains, the ground gets wet. The ground is wet. Therefore, it must have rained. [incorrect - there could have been other reasons. This is an example of <u>Abductive reasoning</u>]

- can a computer ever do more than it was programmed to do?
- consciousness
 - is consciousness necessary?
 - what is it? introspection? self-awareness?
- can you have an "isolated" intelligence, or must it be connected to sensors?
 - grounded understanding mind/body problem (Descarte)

- Can "intelligence" be a *computable* function? (posed by Bertrand Russell)
 - think of mapping from "situations" to "actions"
 - (remember computable functions are expressible as finite combinations of simpler functions using operators like addition, comparison, composition, recursion...)
 - the answer is not clear, but it brings out issues of role of "free will", etc
- Does intelligence rely on quantum effects? (Roger Penrose)
 - some things are physically unknowable, like joint position and momentum of an electron
 - could the (human) brain be exploiting uncertainty in states of water molecules?

- Turing Test
 - proposed by computer scientist Alan Turing (1950) "Can Machines Think?", CACM.
 - how it works
 - 15 min conversation via terminal (blinded)
 - restricted to a topic/subject, like pets, or gardening (but not current events)
 - human assessors judge whether human or computer
 - statistical criterion pass if can fool >50% of judges indistinguishable from humans

- Turing Test (proposed by computer scientist Alan Turing)
 - advantages:
 - Integrates reasoning and understanding and communication
 - Empirical evaluation based on performance (doing): doesn't matter if it is "conscious", or how it is implemented internally
 - Comparison to the "average" intelligence of humans (by averaging against multiple decoys, avoiding individual differences in intelligence)
 - Not "subjective", in the sense that it does not depend on a single judge's opinion (it is averaged over multiple judges)
 - disadvantages:
 - non-conversational aspects of intelligence recognizing objects visually, tying shoes...
 - does computer have delays answers and make mistakes to emulate humans? does it have to appreciate humor? analogies?

- Turing Test (proposed by computer scientist Alan Turing)
 - Loebner competition annual competition, \$100,000 prize
 - chatbots
 - Eliza emulates a psychiatrist; pattern-based: "Tell me more about X..." "How did that make you feel?"
 - Eugene Goostman emulates a 13-yr old Ukranian boy (which induces people to forgive his conversational mistakes)

Eliza transcript

- ELIZA: Please tell me what's been bothering you.
- USER: I feel happy to be writing right now, to be typing this as we speak. Lately, I've had writer's block.
- ELIZA: Can you elaborate on that ?
- USER: I only really feel happy when I'm writing.
- ELIZA: How long have you been writing ?
- USER: Always, but I've been writing seriously since I was eighteen.
- ELIZA: What incident are you thinking of ?
- USER: Um.
- ELIZA: What does that suggest to you ?
- USER: I don't think it's enough, to only be happy when you write.
- ELIZA: Does that trouble you ?
- USER: ...Um, because you can't write all the time. And then there's the real world, and you still have to deal with real people.
- ELIZA: How does that make you feel ?

- How do humans think?
 - Humans are our best examples of intelligent beings
 - If we can write programs that think/act like humans, it proves we understand it.
- Example: Natural Language Understanding
 - How do humans parse and interpret sentences?
 - How do we turn sentences into meaning?
 - How can we correct for mistakes? (grammatical, spelling...)
 - How do we resolve ambiguities?
 - I saw the man on the hill with the hat.
 - I saw the man on the hill with the tree.
 - I saw the man on the hill with the telescope.
 - How do we fill-in missing/foreign words, or recognize a pun?
 - How do children learn to speak (language acquisition)?

^{7/27/2024} When/why do we commit a slip-of-the-tongue, or a faux pas?

Another example: Can *chess programs* be designed to think and choose moves "like a human"?

- Humans are great at ...
 - making complex decisions in the face of uncertainty (situation assessment; multiobjectives; deadlines, costs, tradeoffs; ethical conflicts...)
 - interpreting perceptual input (recognizing a tiger, or the sound of a tiger), even if there is interference/noise, occlusions, shadows...
 - using context to resolve ambiguities
 - reasoning about other humans' intentions
 - discovering unexpected solutions ("Eureka" effect) (creativity Douglas Hofstadter)
- But humans also have limitations:
 - humans make mistakes, and are subject to fatigue
 - limits on short-term memory; calculation speed; attention (distraction)
 - decisions can be influenced by emotions
 - biases: recency bias, expectation/confirmation bias
 - long-term memory is malleable, subjective, incomplete

7/27/2 decisions are often sub-optimal (satisficing="good enough")

- Paradigms: Behaviorism vs. Information Processing
 - behaviorism: stimulus-response
 - Info Proc: can form internal representations of the world based on sensory input, reason about it, make predictions and decisions
- Cognitive Science
 - concept representations; inter-connections (links) among concepts
 - memory (models of short-term and long-term memory)
 - activation and retrieval processes
 - classification
 - focus of attention
 - mental imagery

- Neuroscience
 - Brains have inherent parallelism
 - Cognitive processes like perception, memory, and reasoning operate in parallel
 - Distributed nature (and plasticity) makes brains fault-tolerant
- The mind compensates for *missing info*, and constantly tries multiple ways to interpret and explain the data it senses
- Psychological models of Consciousness
 - debate among modern philosophers, e.g. John Searles, Daniel Dennett
 - can it be explained as an "illusion" of perceiving/comparing streams of perceptual states?

- (Major paradigms for intelligence and problem-solving)
- Symbolic AI (John McCarthy; Herb Simon; Alan Newell)
 - Symbol Systems Hypothesis thoughts are like discrete abstract tokens that relate to and generate other tokens (e.g. by reminding or classification or inference)
 - GPS (General Problem Solver) algebra problems, checkers, questionanswering...
 - ACT (Jon Anderson) psychology-based simulation of memory and rules (production system) that simulates human subject task performance
 - SOAR (Alan Newell) Universal Subgoaling (when stuck, humans create a new search space to figure out how to get unstuck)

- (Major paradigms for intelligence and problem-solving)
- Symbolic AI (John McCarthy; Herb Simon; Alan Newell)
 - Symbol Systems Hypothesis
 - Is this the (only) path to truly intelligent behavior?
 - strengths: tractable for computing (world idealized as discrete states; can simulate reasoning with symbolic rules)
 - weaknesses:
 - 1. uncertainty/partial matching is difficult to deal with;
 - 2. need to program a huge network of concepts and rules to do anything productive ("knowledge engineering")
 - 3. brittleness

- (Major paradigms for intelligence and problem-solving)
- Connectionism (Marvin Minsky)
 - Perceptrons, Neural Networks, Deep Networks
 - concepts are distributed as weights throughout network
 - output nodes are 'activated' collectively from connections among many neurons
 - weights are tuned through prior experience
 - Is this the (only) path to truly intelligent behavior?
 - strengths:
 - ideal for perceptual and motor-control tasks
 - all you need is training examples
 - weakness:
 - can't "explain" why decisions are made (because concepts are distributed as weights throughout network)

- what do we need to built programs that solve complex problems?
- need more sophisticated *algorithms* in our toolbox
- performance is everything; it doesn't matter how it is achieved (thinking ideally or like a human is irrelevant; consciousness is not required)
- task-focused intelligence; "wholistic" (integrated) AI is not necessary

- Core Concepts
 - Search Algorithms
 - Knowledge Representation and Inference
 - Machine Learning

- Search
 - almost everything in AI reduces to a Search problem (in some space)
 - problem solving, planning, learning, parsing, automated deduction...
 - we will study a variety of Search Algorithms
 - especially Heuristic Search (A*)

- Edward Feigenbaum: "In the knowledge is the power."
- Knowledge-based programming
 - rule bases
 - declarative programming vs. procedural programming: say "what" not "how"
- Knowledge Representation
 - formal logic(s) (propositional/Boolean, predicate/First-Order logic, epistemic logics, temporal logics, default logics, fuzzy logic...)
 - ontologies for qualitative reasoning about events, time, spatial relationships, processes...
 - challenge: how to express uncertainty? probability, confidence factors...

- Inference Algorithms
 - how can you generate deductions from a knowledge base?
- Is an action or decision a logical consequence of what I know?
- How to generate proofs?
- Is a formula satisfiable? Is there a faster way than trying all combinations of variables?
- (Automated inference has diverse applications to reasoning about mathematics, proving theorems in number/set/group theory, and computer program verification.)

- *Expertise* in high-performing humans
 - chess grandmasters; physicians; five-star chefs; auto mechanics; military generals; football coaches; air-traffic controllers...
 - typically accumulated from experience (hence, think: knowledge base)
- Expert Systems
 - example: MYCIN asks questions to diagnose blood diseases and infections
 - commonly used ES shells: PROLOG, CLIPS, JESS
 - can often generate explanations of decisions
- Major limitation of ES: *brittleness*
 - ES often draw inaccurate/silly conclusions outside scope of what they were designed for

- Machine Learning
 - who says machines can't be adaptive and learn from experience?!
 - optimizing performance for an objective function
 - many learning methods:
 - statistics (regression)
 - concept generalization, features, classification trees
 - rule revision (e.g. adding antecedents, special conditions)
 - clustering, pattern recognition
 - neural networks
 - reinforcement learning: sequences of actions to achieve goals in dynamic environments
 - to develop/build/program/implement a system, you need to train it on examples

- Complexity
 - we will see that most "AI" problems are computationally difficult
 - often NP-complete or worse
 - huge search spaces (exponential or infinite)
 - much of AI focuses on using knowledge and heuristics to reduce complexity and find an acceptable solution (often sacrificing completeness/optimality)
 - this connects AI to optimization
 - greedy algorithms; approximation algorithms
 - control theory; operations research