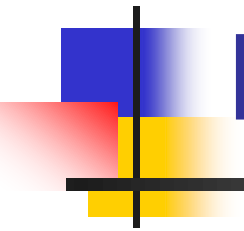
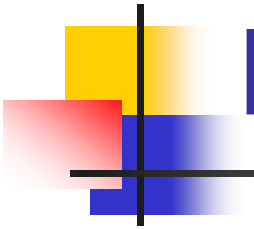
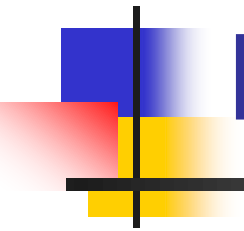


History/Overview of AI



But first:

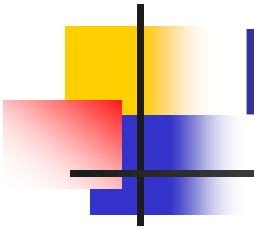
- Assignment 1 clarification: *Also* submit your commented file. No late penalties if submitted today.
- Assignment 1 bonus points – did you get them?



Definitions of AI

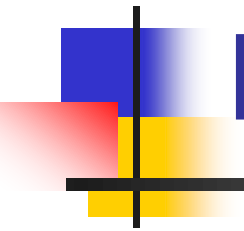
AI according to Luger

- Using computers to do symbolic reasoning
- Problems that do not respond to algorithmic solutions -> heuristic search
- Problem-solving using inexact, missing or poorly defined information -> how to represent such information?
- Capturing and reasoning over qualitative (as opposed to quantitative) features
- Semantic vs. syntactic features
- "Sufficient" answers, vs. exact, complete or optimal answers
- Domain-specific knowledge
- Meta-level knowledge, and reasoning about reasoning



More generally:

"The art of creating machines that perform functions that require intelligence when done by people." (Kurzweil, 1990)



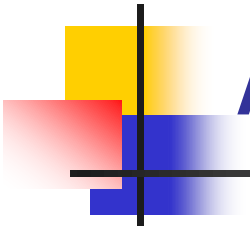
History of AI

Ada Lovelace (1815-1852)

daughter of Romantic poet Lord Byron

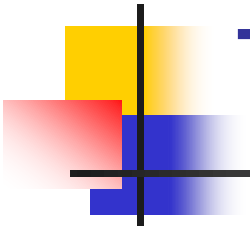
"[Babbage's] Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform."

People have *thought* about the nature of intelligence and the possibility of AI for a long time...



Alan Turing (1912-1954)

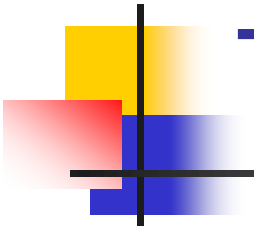
- The Turing Test
- Turing Machine (TM)
- Arguably the father of modern computing
- WW2 code breaking genius



The ‘Dark Ages’ (1943-56)

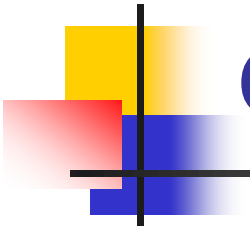
- McCulloch, Pitts: Early neural networks. Showed to be TM equivalent.
- von Neumann: many contributions to CS, including ENIAC.
- Shannon: information theory, heuristics vs. brute search
- McCarthy, Minsky: turned AI into a field...

The Rise of AI (1956-late 60s)

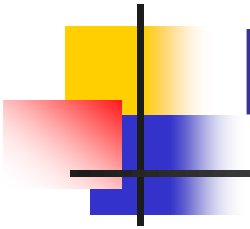


- Classical AI, including much early work on reasoning, planning, means-end analysis and representation and search.
- Many of these ideas have been absorbed into general CS.
- However, these techniques did *not* work on some 'basic' problems: natural language (e.g. translation), 'common-sense' reasoning, image processing, robot control, etc.

Unfulfilled promises (late 60s, early 70s)

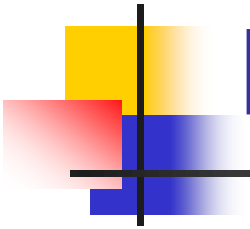


- Funding dries up!
- The difficulty of the 'real' problems sinks in. Many are seen to be 'AI-complete' – that is, they require general-purpose artificial intelligence to solve. Think about machine translation...



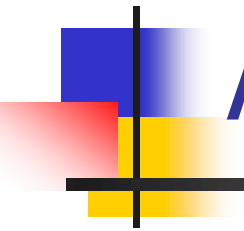
Expert Systems (70s – 80s)

- A viable set of problems is discovered: expert reasoning over a narrow domain.
- Examples: chemical analysis (DENDRAL), diagnosis of infectious blood diseases (MYCIN), mineral exploration (PROSPECTOR)
- Still in use, and useful, if the domain is chosen well.



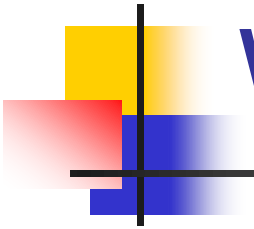
Modern AI (late 80s onward)

- Machine learning (e.g. neural networks)
- Evolutionary computation
- Fuzzy logic
- Cognitive modeling
- Natural language processing
- Agent-based systems



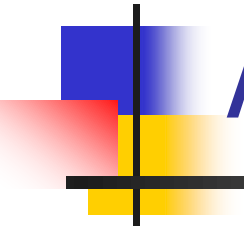
Aims of AI

What are we trying to do?



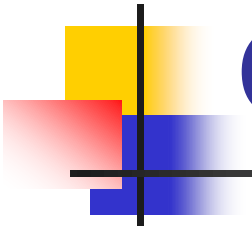
One or more of the following:

- To produce artificial humans/animals
- To build systems that exhibit human intelligence
- To learn more about human intelligence by trying to emulate it in a machine
- To build systems that behave (super-) intelligently
- To build systems which can do (boring, dangerous) tasks that (only?) humans now do
- To build useful machines using mechanisms inspired by human intelligence



AI Application Areas

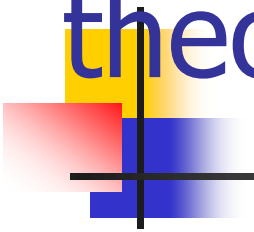
Or rather, a selection of sub-fields, important techniques, and catchphrases



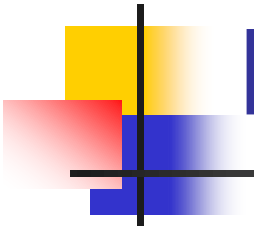
Game playing

- Require intelligence in humans
- Used to be a populist benchmark
- Constrained domain and well-defined rules are easily represented on a computer
- Exhaustive search not effective (compare chess and tic-tac-toe)
- Planning and strategy might be applicable to other problems
- Game playing programs have a market!

Automated reasoning and theorem proving

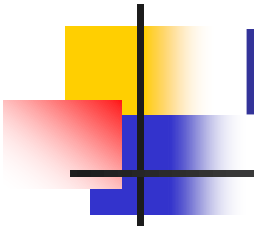


- Requires intelligence in humans
- Constrained domain and well-defined rules are easily represented on a computer
- Exhaustive search not effective
- Novel proofs and theorems would demonstrate creativity
- Useful if better than humans



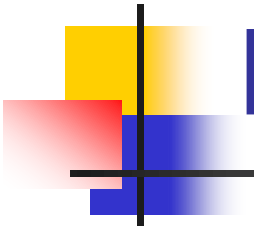
Expert systems

- Attempt to capture human expertise (e.g. medical knowledge *and reasoning*)
- Useful if successful, but problems include:
  - Lack of `deep' knowledge, or common sense
  - Lack of robustness/flexibility
  - Inability to provide deep explanations
  - Difficult to verify solution
  - Little or no learning
- We will go into ways to solve these problems in some detail.



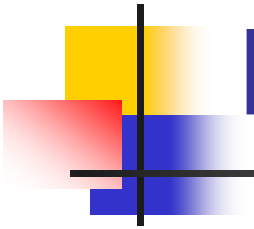
Modeling human performance

- Primarily a research tool, to refine and test models of human behavior
- Machine must not only do *what* humans do, but do it *how* humans do it
- Has revolutionized the human sciences, especially cognitive science, linguistics, psychology, physiology and brain science



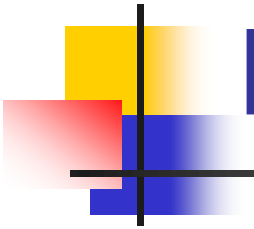
Robotics

- Autonomous (or semi-autonomous) machines which must sense, manipulate, and move through their environment
- Usually not very humanlike, except when dealing with human-engineered environments
- Useful for tasks which are too dangerous or too boring for humans
- Related fields: machine vision, autonomous agents, planning, behavior-based robotics



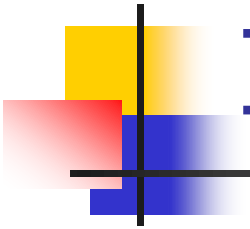
Planning and scheduling

- Now considered a separate field, but rooted in AI
- A set of tools and techniques that allow programs to solve complex, highly constrained problems
- Very useful, but not very human any more



Machine learning

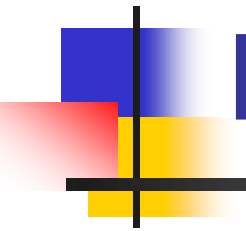
- Self-modification based on observation/feedback/successes/failures etc.
- Many different techniques, ranging from ES-related knowledge capture to neural networks
- Sometimes intended to be humanlike, sometimes not
- Representation issues, e.g. how do you know what has been learned?



Intelligent interfaces

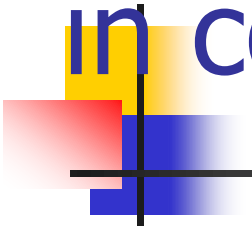
Sometimes the best way to communicate with a human is to be human-like, so it can be useful to:

- Use and understand language
- Have and recognize facial expressions
- Have and recognize gestures
- Understand what the user's beliefs and goals are, and (perhaps) have beliefs and goals too.



Evaluating AI systems

The research cycle (esp. common in cognitive modeling)



- Model
- Implement
- *Evaluate*
- Refine model
- Rinse, repeat!

Sub-Turing Tests: comparisons with humans

- Take some subset of human intelligent activity. Can a human judge tell the difference between the human and the computer performing this task?

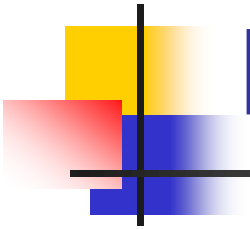
e.g. JAPE, many many other AI systems.

Spot the JAPE jokes I

- What's the difference between money and a bottom? One you spare and bank, the other you bare and spank.
- What do you give a hurt lemon? Lemon aid.
- What do you call a sour assistant? A lemon aide.
- What do you call Martian beer? An ale-ien.

Spot the JAPE jokes II

- What kind of pig can you ignore at a party? A wild bore.
- What animal runs round the the forest making the other animals yawn? A wild bore.
- What do you get when you cross jewelry and a bobcat? Cuff lynx.
- What do you get when you cross the Atlantic with the Titanic? About halfway.



Direct competition

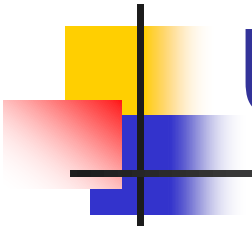
- Ideal for tasks which have a clear metric (time, distance, points etc.)
- Can the computer beat the human (or another computer)?

Game-playing computers are the classic example.

Validity of model

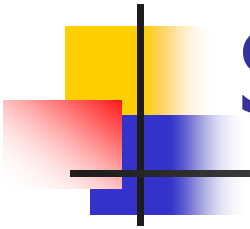
- Does the model show something interesting/relevant/new about how humans do the task? What can we learn from its success?

A problem with non-symbolic systems – what exactly do we have??



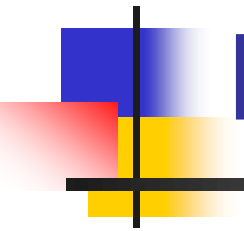
Utility

- Is it useful? Does it do the job? Does it meet the specifications?
- Software/hardware engineering style of evaluation.
- Often a sign that this particular technique or sub-field will soon no longer be called AI!

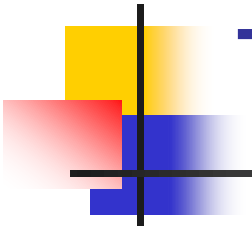


Summary

- AI is a broad field, covering many technologies, with a range of goals, but with common themes
- AI problems often have a certain mystique – and if we solve them, the mystique goes away, and they stop being problems requiring “intelligence”!

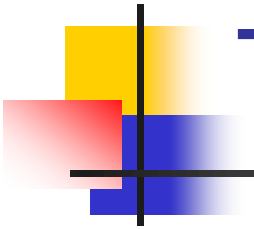


Philosophy of AI



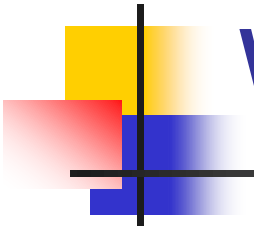
Today's topics

- The Turing Test
- The Physical Symbol System Hypothesis and Strong AI
- The Chinese Room
- Weak AI



The Turing Test

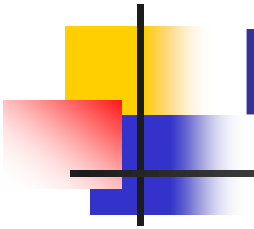
- Human (tester) communicates with a human and a machine via a typing interface
- Conversation is totally unconstrained: Any subject, any duration, any language (including slang), lying allowed, etc.
- Tester must determine which is the machine. If no better than chance, must grant that machine is intelligent (acc to Turing).



What's the point?

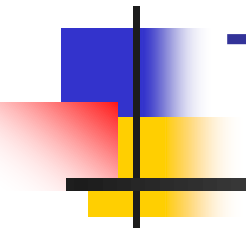
- To cut through philosophical discussions of "what is intelligence?" and "can a machine think?" with a simple test.
- A common misconception is that the Turing Test is *too easy* – in fact, very very difficult, and no program has ever passed it.

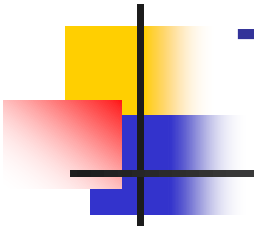
Play with a ChatBot



- Go to www.alicebot.org
- Click on "Chat with A.L.I.C.E." (or whatever bot is offered) under "Get Started".
- What are the strengths and weaknesses of this kind of chatbot? How do you think it works? Do you think this approach will lead to a program that can pass the Turing Test?
- If you want, you can build your own ALICE-style chatbot – if you do, post the link to Laulima!

Some objections to the Turing Test (there are others)





The Theological Objection

- Thinking is a function of the soul
- Machines don’t have a soul, therefore they can’t think

The argument from consciousness

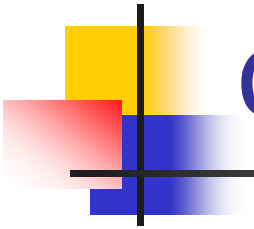
Two parts:

- Self-awareness
- *Qualia*: 'really feeling' some sensation or emotion.

Turing's response:

- Can't know about consciousness unless you *are* the thinker (even with other humans)
- If there are *any* external manifestations, they will show up in the Turing Test – and if there aren't, who cares? [paraphrased ☺]

Argument from various disabilities



- Turing lists a number: machines can't… be kind… have a sense of humor… make mistakes…fall in love…use language…be creative…
- Turing's response is more or less that these are areas for research, but that he doesn't see any particular reason why they can't do these things.

Lady Lovelace's Objection

- That is, machines can only do what they're programmed to do.

Turing responds that machines surprise him all the time…

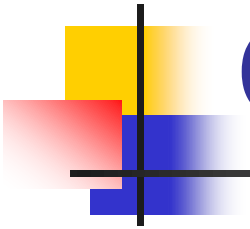
Also, what if they learn? Evolve? Change in such a way that their behavior is surprising even to their programmers?

Argument from Informality of behavior

- Humans don't strictly follow rules; computers do.

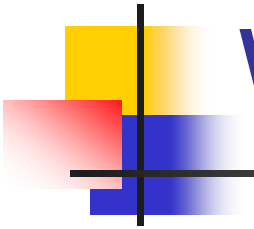
Turing responds:

- Humans *do* follow the laws of physics, at least, and probably higher level laws of behavior.
- Machines can break 'rules of conduct' as easily as humans can.



Other objections to the TT

- Ignores other kinds of intelligence (e.g. a dolphin couldn’t pass, nor a very clever Mars rover)
- Overemphasizes linguistic fluency
- Why should an intelligent computer pretend to be a human?



What do you think?

- Is the Turing Test a reasonable way to establish whether or not a machine is intelligent? Why or why not?
- What would you propose instead?

Constrained Turing Tests

- *Not* used to establish intelligence – but are used to support claims that program is 'human-like' in some specific way
- Can be limited by time (e.g. < 5 min interaction), subject matter (e.g. must talk about sports), or medium (e.g. art, music, etc.
- Often used to evaluate domain-specific AI apps.
- *Check out* The Loebner Prize… (www.loebner.org)