Artificial Intelligence: Philosophy, Emotion, Modern AI

CS311
David Kauchak
Spring 2013

Adapted from notes from:
Sara Owsley Sood

Final exam

Monday, May 11, 2-5pm
- Prof. Wu will be administering

Comprehensive
- *except* Java

Notes:
- 20 pages (single sided)

Some of the big questions of AI philosophy…

What is AI, really?
- What does an intelligent system look like?
- Does an AI need—and can it have—emotions, consciousness, empathy, love?

Can we ever achieve AI, even in principle?

How will we know if we've done it?

If we can do it, should we?

Weak & Strong AI

Weak AI hypothesis:
- Machines can be created that will act as if they are intelligent (simulate thinking)
- What AI can do is to develop useful, powerful applications

Strong AI hypothesis:
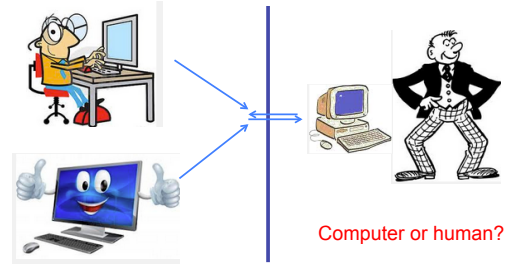- Machines can be created that will actually think (rather than simulate thinking)

Weak AI: Turing Test

Alan Turing, 1950:

(http://www.loebner.net/Prizef/TuringArticle.html)

"I propose to consider the question, 'Can machines think?'"

Weak AI: Turing Test

Computer or human?

Weak AI: Turing Test

Turing predicted: by 2000, a computer could be programmed to pass the Turing Test (for a 5 minute test) 30% of the time

Turing Test - Eliza

An early success at fooling some people: Eliza, simulating a Rogerian therapist

….Eliza; the importance of picking the correct application area

Sample interaction:
Patient: Men are all alike.
ELIZA: In what way?
Patient: They are always bugging about something or the other
ELIZA: Can you think of a specific example?
Patient: Well, my boyfriend made me come here.
ELIZA: Your boyfriend made you come here?
Patient: He says I'm depressed much of the time.
ELIZA: I am sorry you are depressed.

Jabberwacky

Uses real user conversation to learn responses.

http://www.jabberwacky.com/

CyberLover

That our computers have seduced us has long been a truism. Now, thanks to the ever-inventive internet mafia, it is becoming a literal truth. Russian cyber-crooks have reportedly unleashed a software robot, or bot, that poses as a would-be paramour in sex chatrooms. It entices randy gentlemen to reveal personal information, such as their address or birthday, or even to submit photographs of themselves. The information can then be used to break into bank accounts or carry out other forms of fraud.

It was probably inevitable. As one of Tony Soprano's sidekicks observed in a classic episode of the TV series, the two most resilient sectors of the economy are organised crime and "certain aspects of showbusiness". The aspects, that is, known as the world's oldest profession - now mixing it with the world's newest technologies.

CyberLover, as the dirty-mouthed bot is called, is quite a sophisticated piece of software. It can take on a number of different guises depending on the proclivities of its target, according to security experts at the software company PC Tools. It can play the role of a romantic lover, for instance, or masquerade as a sexual predator.

http://www.guardian.co.uk/technology/2007/dec/13/internet.crime

Can we ever achieve AI?

Can we ever achieve AI?

Argument of disability: "hey, there are lots of things that a computer can't do!"

"Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humor, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make someone fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behavior as man, do something really new."

Responses?

Some successes

What are some human-oriented tasks that computers can do better than people?

- Play chess, checkers and other game
- Inspect parts on assembly lines
- Check the speeling of text
- Steer cars and helicopters
- Diagnose diseases
- Do hundreds of other tasks as well as or better than humans
  - Computers have made small but significant discoveries in astronomy, math, chemistry, mineralogy, biology, computer science, and other fields

"No, I don't want to play chess. I just want you to reheat the lasagna."

or…

Argument of informality:

- "what people do is too complex to capture"
- Because computers can do no more than follow a set of rules, they cannot generate behavior as intelligent as that of humans

Responses?

or…

Argument of informality

More of a problem with "classic" AI
- reasoning
- knowledge representation

Some progress in incorporating background knowledge

Learning algorithms: increasing ability to operate autonomously (unsupervised learning), learn new features, prune feature spaces

Strong AI

Can machines really think?

What does it mean to think?

Do we have to have a brain to have a mind? to think?

"brain in a vat" experiment

Is physicality crucial for intelligence?

Matrix scenario: a brain is supported, bodiless, in a vat, and signals simulating a virtual world are fed in/out of the brain

Is being hungry the same as some rule:
   DyingFor (Me, Pizza)
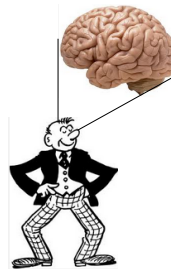
Could you tell the difference?

"brain in a vat" experiment

Moravec (robotics researcher/functionalist) is convinced that his consciousness would remain unaffected

Searle (philosopher and biological naturalist) is equally convinced his consciousness would vanish

"brain prosthesis" experiment

Technology advances where we can create an artificial neuron:

Exact same electrical/physiological responses as a real neuron.

We can copy an existing neuron.

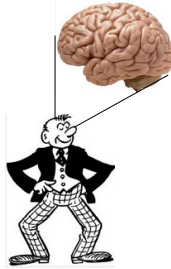## "brain prosthesis" experiment
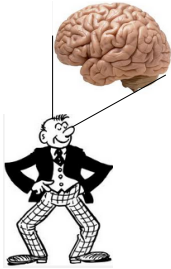
Technology advances where we can create an artificial neuron:

If I exchange one real neuron for one artificial will you notice?
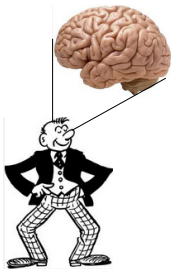
"brain prosthesis" experiment

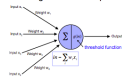Technology advances where we can create an artificial neuron:

If I exchange two real neuron for two artificial will you notice?

"brain prosthesis" experiment

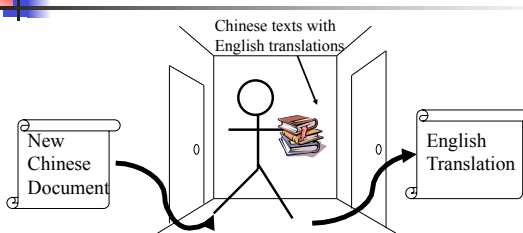Technology advances where we can create an artificial neuron:

If I continue this process, when will you notice?

http://www.smbc-comics.com/index.php?db=comics&id=1879

Welcome to the Chinese Room

Chinese texts with
English translations

New
Chinese
Document

English
Translation

You can teach yourself to translate Chinese
using *only* bilingual data (without grammar books,
dictionaries, any people to answer your questions…)

The Chinese Room

John Searle, 1980

Human who knows only English; stacks of paper with
Chinese symbols; *rule book in English*, stating which bit
of paper to give in response to a given (Chinese) input

Human who knows only Chinese on outside of room;
passes in Chinese query, receives Chinese response

Do you know Chinese?

你好

Creative

Having the ability or power to create: Human beings are
creative animals.

Productive; creating.

Characterized by originality and expressiveness;
imaginative: creative writing.

How do people write stories?

Can Computers Be Creative?

Two paintings produced by Harold Cohen's Aaron
software:

http://www.kurzweilcyberart.com/

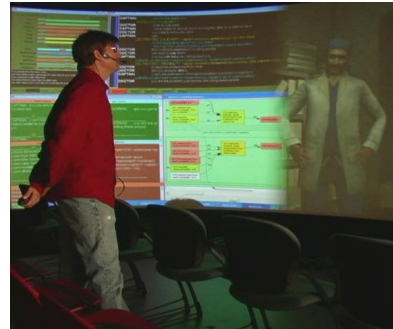http://www.kurzweilcyberart.com/aaron/aim_clip_cohen.html

Say Anything

Corpus based story telling

http://sayanything.ict.usc.edu/SayAnything/

http://people.ict.usc.edu/~gordon/publications/ICIDS09.PDF

Can Computers Understand and Express Emotion?

HCI

Cliff Nass

Example human-human situation
- Someone tries to give you help and their timing is bad,
- you try ignoring then frowning or glaring,
- an intelligent person picks up on that feedback, interprets what it means, and acts accordingly (backs off)

Analogous human computer situation
- So if a computer tries to give you help at a bad time (aka - clippy),
- you try to ignore it and then frown or glare
- An intelligent systems would receive that signal, interpret that signal, and react appropriately

Emotion

Can we build systems to
- Detect it?
- Express it?

Detection

Gestures

Facial Expressions

Speech/Text

Physiological Cues
- Blood volume pressure
- Skin Conductivity
  - Glove (other WEARABLE DEVICES!)

Detecting emotion via wearable devices

2001 - 81% accuracy in (forced decision)
- detection of 8 emotions:
  - Neutral, anger, hate, grief, platonic love, romantic love, joy, reverence
- Person dependent - trained for at least 4 weeks
- GROUND BREAKING!

Since then, lots more work has been done
- http://www.youtube.com/watch?v=ceP-vcbFxh0

Applications?

Ethics

http://www.smbc-comics.com/index.php?db=comics&id=2956#comic

What we've done

How many slides?
- ~628 slides

How many pages of notes?
- ~135 pages

How many functions/methods did we look at in class?
- 170

How many lines of code have we looked at in class?
- 1,564

How many lines have **you** written?
- 1,673 (well, that's how many the solutions have)

What we covered

Python!
- variables
- functions
- loops
- conditionals
- recursion
- higher order functions
- classes
- file I/O
- many other, sub-topics
  - lists, tuples, dictionaries, …
  - exceptions
  - turtle graphics

What we covered

Context free grammars

Neural Networks

Search
- algorithms
- problem solving
- adversarial search and game playing

DFAs/NFAs/Turing machines

Artificial Intelligence

Where we started

```
# David Kauchak
# CS150
# 9/14/11

# This program figures out the number of hot dogs
# needed for a BBQ
tim = 1
amy = 2
todd = 2 * amy
brenda = todd - 1
mark = (brenda+1)/2 + 1 # add 1 to brenda to round up

total_hotdogs = tim + amy + todd + brenda + mark
print total_hotdogs
```

Where we ended

```
class NimGame:
    """ Class to keep track of a nim game. """

    def __init__(self, starting_piles):
        """
        Construct a new game with the list starting_piles
        as the piles
        """
        self.piles = starting_piles[:]

    def get_piles(self):
        """ Returns a copy of the current piles """
        # return a copy to avoid anybody messing with the internal
        # state of the game
        return self.piles[:]

    def make_move(self, pile_number, num_to_remove):
        """
        Move num_to_remove from pile_number.

        Return True if the move was valid, False otherwise.
        """
        if pile_number < 0 or pile_number >= len(self.piles):
            return False
        elif num_to_remove < 0 or num_to_remove > self.piles[pile_number]:
            return False
        else:
            self.piles[pile_number] -= num_to_remove
            return True

    def is_over(self):
        """ Is the game over? """
        return sum(self.piles) == 0

    def __str__(self):
        return str(self.piles)
```

Where we ended

```python
def play_nim(player1, player2, game_state):
    # player1 will start
    player1_turn = True

    while not game_state.is_over():
        if player1_turn:
            print "Player 1's turn"
            print game_state
            (pile_number, num_to_remove) = player1(game_state.get_piles())
            print "Player 1:",
        else:
            print "Player 2's turn"
            print game_state
            (pile_number, num_to_remove) = player2(game_state.get_piles())
            print "Player 2:",

        print str(num_to_remove) + " from pile " + str(pile_number)
        print

        if not game_state.make_move(pile_number, num_to_remove):
            print "ILLEGAL MOVE!"

        player1_turn = not player1_turn

    # game's over, see who won
    if player1_turn:
        print "Player 2 won!"
    else:
        print "Player 1 won!"
```