2023-06-16: Ethics and AI

Objective

- Introduce Artificial Intelligence
- What the Scripture and Tradition says about Knowledge
- Talk about the dangers of AI
- Talk about the appropriate and inappropriate uses of AI

Key Verse

"He who walks with integrity walks securely, But he who perverts his ways will become known.“ (Proverbs 10:9)

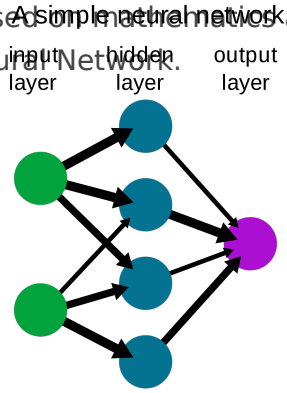
Visual Aid

https://www.dropbox.com/s/zb7om5xwkp5yx8p/Artificial%20Intelligence.pptx?dl=0

What is Artificial Intelligence

Neural Networks

- Artificial Intelligence is a branch of Computer Science that attempts to simulate human intelligence processes by machines.
- One of the most common sciences in AI is the idea of a Neural Network. And it's based on the science behind how the brain works. The brain has something called neurons that use electrical and chemical signals to send information between different areas of the brain, the spinal cord, and the entire nervous system. In 1943 a Neuroscientist and a Logician (McCulloch and Pitts) published a paper where they created a computational neural model based on mathematics and algorithms. That was the foundation of what is now called the Neural Network.



  - Neural networks work by having an input, a bunch of hidden things and an output. And it looks/sounds basic, but obviously the hard part is in the hidden things. And the hard part is in training the neural network to have the right

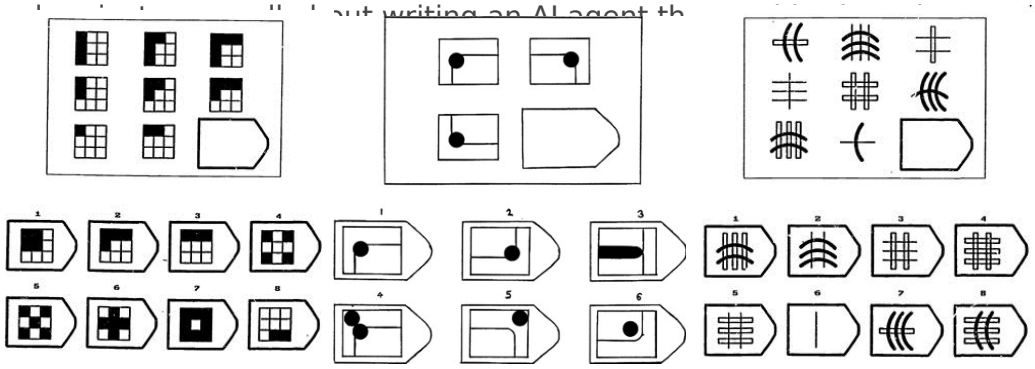
hidden things.

- https://www.youtube.com/embed/qv6UVOQ0F44

The example MarI/O helped me to better understand this concept.

- In this video, the input is the position of Mario, the location of enemies and moving objects, and the location of static obstacles
- The output is what button to press
- Basically, he wrote a program that would develop this neural network and he ran it for 24 hours, and then the neural network figured out how to beat that specific level of Mario
- So at first, it does random movements and it sees how far it gets in the level. And it gets a score for how far it got. Obviously, pushing "right" would get it a higher score. Then after it tries a bunch of things, the things with the highest scores get combined together and a little bit more randomness is introduced... and then it tries again. And it keeps trying and trying until it finally gets it.
- The algorithm that he mentions - NEAT - was developed by a professor here at UCF. I first heard about it when I took an AI class at UCF during my undergraduate studies.

Knowledge-Based AI

- In my Master's Degree, I took a class called "Knowledge-Based AI" and the assignments [illegible] writing an AI agent th[illegible]Q Test
- 

- These are called Raven's Matrices IQ Tests
- We basically just had to figure it out... and so my algorithm basically did an analysis of each of the squares given to us, and each of the choices and then it did a comparison of A-B-C and D-E-F and then G-H-Option 1, G-H-Option 2, etc. to try and make a best guess about which one was right.
- For some basic ones, it was very successful. But when you get to harder ones, it's hard enough for a human to spot the pattern, let alone a computer.
- Note that the computer is not understanding the pattern and trying to find what fits it... it is looking for patterns in different ways - by pixels. By trying to understand shapes. This one has many black pixels on the left, and this one has many black pixels on the right.
- During that class, our professor came out in the last week and said he had some news... one of our TAs this whole semester... was an AI! And we had to guess who it was. Half of the people guessed it wrong. And this was all over the news about this AI TA that no one

noticed was an AI the whole semester.
- But that AI was programmed to understand a few types of questions and answer them - usually common questions like "when is xyz assignment due" - "can I use xyz on my assignment" - "how do I contact the TAs" - etc. This is knowledge based AI

Probability

- Another class that I took in my Master's was called "AI for Robotics" and it was all about programming a basic self-driving car
- It was basically a probability class
- We had one module about figuring out the probability of lane detection and how to stay in your lane for steering
- We had one module about navigational systems
- We had one module about using sensor data to predict what was a car and what wasn't
- But the main take-away from the class, for me, was that it is all probability and probability distributions

Machine Learning

- Back in 2012, Google did something crazy... they said that their computers could now tell if a picture contains a cat or not.
- https://blog.google/technology/ai/using-large-scale-brain-simulations-for/

Uses of AI

- Natural Language Processing
  - Google Translate
  - Grammarly
- Computer Vision
  - Self-Driving Cars
  - Image Categorization
  - Surveillance Cameras
  - Facial Identification
  - Agriculture
  - Healthcare
    - Scanning MRIs/X-Rays to find tumors
    - Scanning MRIs to find commonalities between large numbers of patients that have the same diagnosis - that may not be visible to the human eye
- Personalization
  - Targeted Advertising
  - Personalized Shopping
  - Personalized Learning
- Gaming
  - Playing strategic games like chess
  - Giving a unique gaming experience every time
- Travel
  - Suggesting hotels, rentals, etc.

Is AI Good or Bad

What is Intelligence?

- Kinds of Knowledge
  - Sensuous (comes through sight, touch, smell, taste, hearing)
    - This is the most basic of knowledge - the sky is blue, the grass is green, the pizza is hot, feseekh smells bad, medicine tastes bad, airplanes are loud, etc.
  - Sourced (books, newspapers, films, mass media)
    - This is the stuff you learn in school or from reading books or from watching movies/documentaries
  - Scientific (through hypothesis, experimentation, observation - knowledge of the physical world)
    - This is something you would learn through the scientific method, setting up an experiment, observing its results and drawing conclusions
  - Social (through experiences and relationships)
    - This is emotional knowledge - understanding how people feel, relating to them
    - HH Pope Shenouda did not mention this one in this book
  - Sophia/Wisdom (through the fear of God and by knowing His Scripture and His Commandment)
    - Wisdom in Greek is Sophia
    - This is the knowledge of discernment and being able to discern what is right or wrong
    - This is the knowledge of the Scriptures and Commandments of God
  - Spiritual (God reveals to His holy people what He wants them to know
    - There is knowledge revealed by God to those who are holy
    - Like when people would approach HH Pope Kyrillos and he would know what they were going to ask
    - Or like when God spoke through His Prophets
    - Or like when God showed St John the Revelation, or St Peter the vision
    - Or when the Lord guided St Paul to go to Macedonia
  - Ten Concepts Chapter 10 - The Concept of Knowledge ~ HH Pope Shenouda III
    - Note that HH Pope Shenouda talks about all of those, but does not present them in that order, nor does he use the same language there but rather I tried to find a "S" word for all of them
- There is useful knowledge and there is harmful knowledge
  - "There is harmful knowledge like that in which our father Adam and our mother Eve fell. It led them to lose their innocence and simplicity and live in the dualism of good and evil, truth and falsity, lawful and unlawful." (Page 134)
  - Many people wept and regretted knowledge which they came upon
    - e.g. someone will say "I just want to know how alcohol tastes" - but that knowledge brought him to alcoholism
    - e.g. someone will say "I just want to try this drug once... to see how it feels" - but that knowledge brought him nothing more than addiction
    - e.g. someone will say "I just want to watch porn to see what sex is like" - but that knowledge was false and brought him addiction, depression and destroyed

    - his relationships with others and gave him a false idea of what sex is like
    - e.g. someone will say "I just want to see the video of someone being killed" - but that knowledge haunted them and gave them nightmares and to an extent, desensitized or numbed them from the act
- The fact is that all knowledge that comes to your mind will affect your sense and feelings and relationships with others; it will affect and influence your subconscious in the form of suspicions, thoughts, dreams, memories. It will extend within you and outside to a great extent and you may not be able to stop it. - HH Pope Shenouda (Ten Concepts)
- Which types of knowledge would you expect an AI like ChatGPT to have?
    - Mostly Sourced Knowledge; some sensuous knowledge (e.g. computer vision or audio processing). But that's it.
    - ChatGPT cannot know what is right or wrong, cannot make wise decisions, does not and cannot know God, cannot relate to people or understand their emotions
    - Can you compare Human Intelligence to Artificial Intelligence? It is incomparable

What is Artificial?

- AI is never certain
    - AI is either Probability or Trial and Error, or a combination of both
    - In MarI/O, it's Trial and Error... and mostly error
    - In the Raven IQ Tests, it was sometimes wrong and never completely sure
    - In Self-driving cars, although we could see clearly that those were cars, it was only 60% sure
- AI can be manipulated by input
    - Adversarial Machine Learning is a whole field of study on how to fool AI Agents
    - By introducing noise to an image, the classification of it can be completely changed
    - By introducing perturbations to a physical stop sign, the AI Agent is completely fooled
- AI can be manipulated by its owners
    - This was recently all over the news when SnapChat AI came out
    - The AI was heavily influenced by liberal ideas and would respond differently to conservative or Christian ideas

AI is a Tool

- A tool can be very useful in the right hands, and very dangerous in the wrong hands
- A saw is very good at cutting wood to make furniture... but in the wrong hands it can be used to murder
- The internet is very good at connecting people who are geographically far away... but in the wrong hands it can be used to exploit people from very far away
- AI, in the right hands with the appropriate uses, can be very helpful. But in the wrong hands, or for the wrong uses, can be very harmful.

Dangers of AI

AI Can (and Will) Lie to You

- The SnapChat AI was famously in the news for lying about knowing your location or not
- A Lawyer was disbarred because he used ChatGPT to find relevant case citations and they all ended up being fake - made up by ChatGPT. He did not verify the information he got from it. When he presented them in court, the court could not find anything he was referencing.
    - https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html

AI Can Manipulate You

- AI Virtual Assistants simulate human interaction (mimic human voice, facial expressions, body language) and can create the illusion of a personal relationship.
- Deepfake videos, images and audio are highly realistic simulations of people saying or doing things they never did
- AI is used to limit what information each person can see (e.g. on social media) and can ensure that only information that reinforces certain beliefs is prevalent, creating an echo chamber

AI Has a Fingerprint

- College students used ChatGPT to write their essays in a Philosophy class. The professor said "it was really well-written wrong" and that was a red flag. So he simply asked ChatGPT if it thought these essays were written by AI. ChatGPT said "I am 99% sure these were written by ChatGPT."
- The students failed the class
- https://www.businessinsider.com/chatgpt-essays-college-cheating-professors-caught-students-ai-plagiarism-2023-1
- AI Generated results are easy to spot (especially by another AI)

How to use AI

Inappropriate Uses for Artificial Intelligence

- **Plagiarism:**
    - Asking ChatGPT to write you an essay - even if you're planning to "edit it"
    - Asking an AI to solve math problems for your homework
    - Why is it unethical?
        - You would not be putting in the work, and getting a grade you didn't work for (i.e. stealing)
        - You would finish a paper or assignment and learn nothing (i.e. wasting resources)
        - You are misrepresenting something as your own work (i.e. lying)
        - Like the lawyer
    - Imagine that you used ChatGPT to write a poem for your loved one
        - What would they think if/when they found out it was not from you?
        - How would they feel if they had built some feelings around a false idea of you?
- **Cheating:** Requesting answers or solutions to exam questions from ChatGPT

- **Cyberbullying:** Using AI to harass, intimidate or bully others online
- **Fabricating Data or Sources:** Requesting fake data or sources or references to support your paper or project

Appropriate Uses for Artificial Intelligence

- **Educational Support**: Using ChatGPT to clarify concepts, ask questions or seek explanations on topics you're studying.
- **Writing Assistance**: Grammarly or even ChatGPT can fix your grammar or give you some writing advice on an essay you have written yourself
- **Learning New Subjects**: Ask questions about various topics
- **Finding Sources**: Asking ChatGPT for relevant sources to a certain topic, and then exploring, reading, and analyzing those sources on your own
- **Language Practice**: If you're studying a foreign language, you can engage with ChatGPT to practice having conversations, ask for translations, improve your proficiency

Contemplation

Resources

- HH Pope Shenouda
  - Words of Spiritual Benefit Volume I
  - Ten Concepts
- All other sources related to AI are linked above

Revision #8
Created 2023-06-13 17:31:13 UTC by Morcous Wahba
Updated 2023-06-17 04:17:06 UTC by Morcous Wahba