



ETHICS AND AI

JAMES CALDWELL

CS 131 SECTION #17468

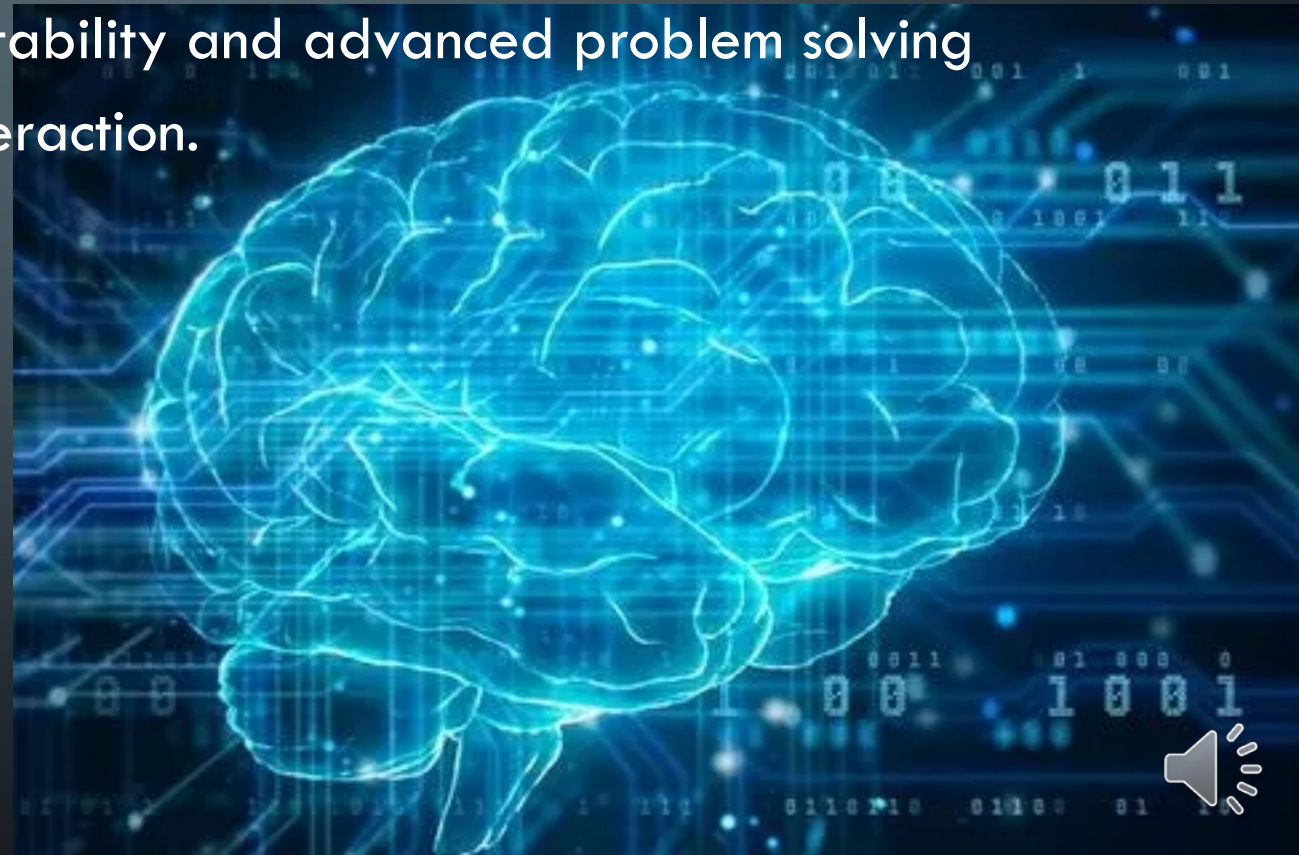
WEST LOS ANGELES COLLEGE

FALL 2022



DEFINING ARTIFICIAL INTELLIGENCE

- AI is the attempt to get computers to achieve true cognition by developing machine learning that achieves adaptability and advanced problem solving without direct and constant human interaction.

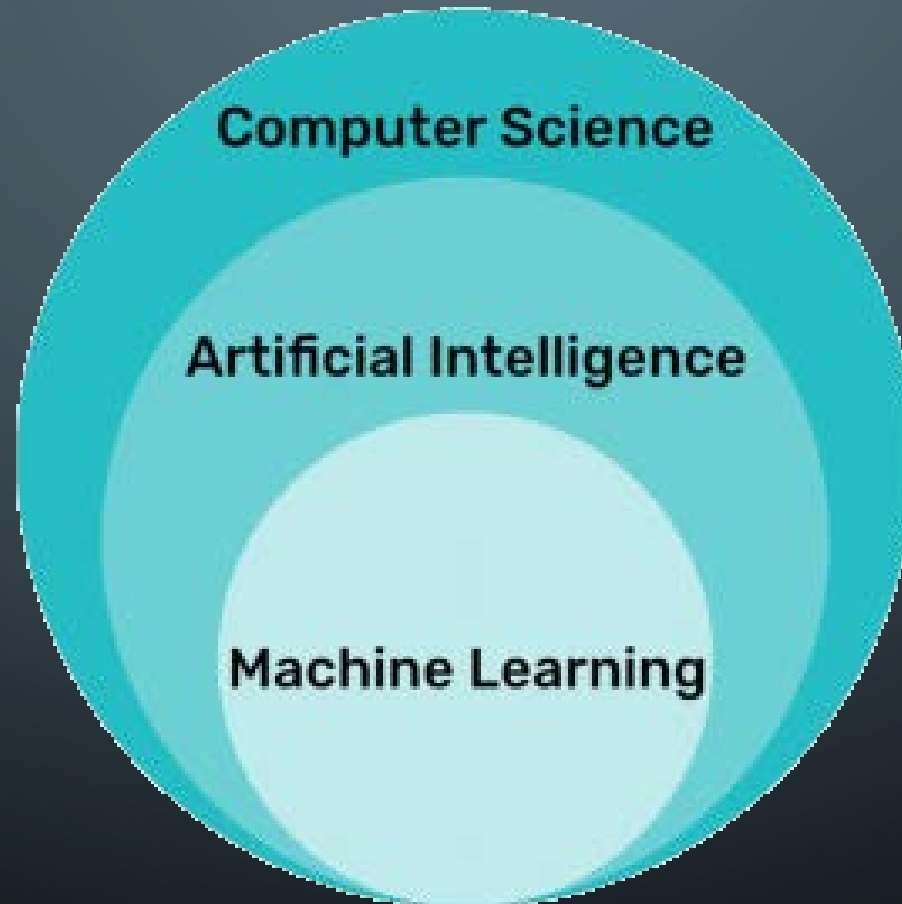


BREAKING IT DOWN

- Cognition - the mental action or process of acquiring knowledge and understanding through thought, experience, and the senses.
- Machine learning - the capability of a machine to imitate intelligent human behavior
- Adaptability - the quality of being able to adjust to new conditions



THE DEVELOPMENT OF AI IS CONSIDERED BY MANY TO BE THE LEADING EDGE OF TODAY'S COMPUTER SCIENCE



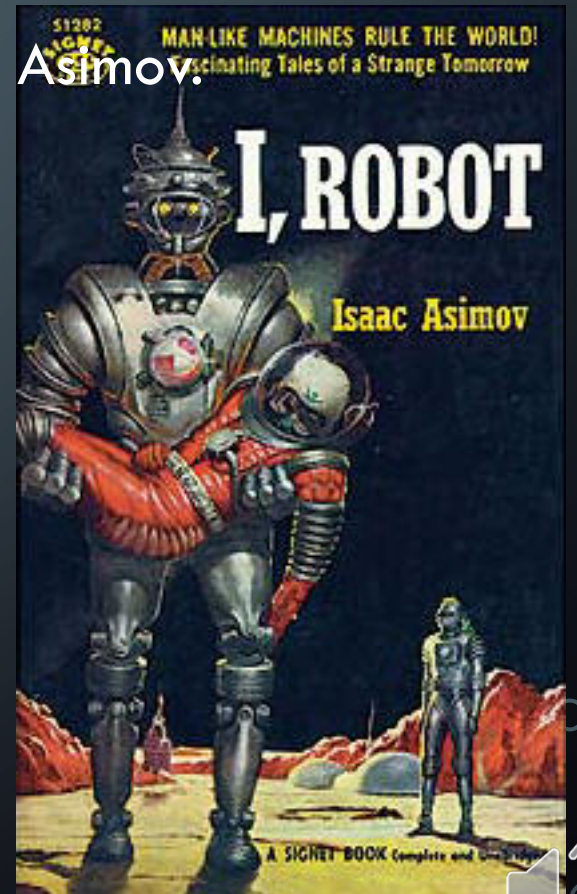
CAUTIONS

- Although many are excited to explore this field, there are also many who feel concerned about the possible disruptive potential of AI.
- This has been heightened by many fictional stories from books, and later movies, of advanced AI causing apocalyptic or dystopian futures.



ISAAC ASIMOV

- One of the most popular authors to explore this topic was Isaac Asimov.
- He wrote many “hard science” fiction books during his lifetime
- He created the famous “Three Laws of Robotics.”



THE THREE LAWS OF ROBOTICS

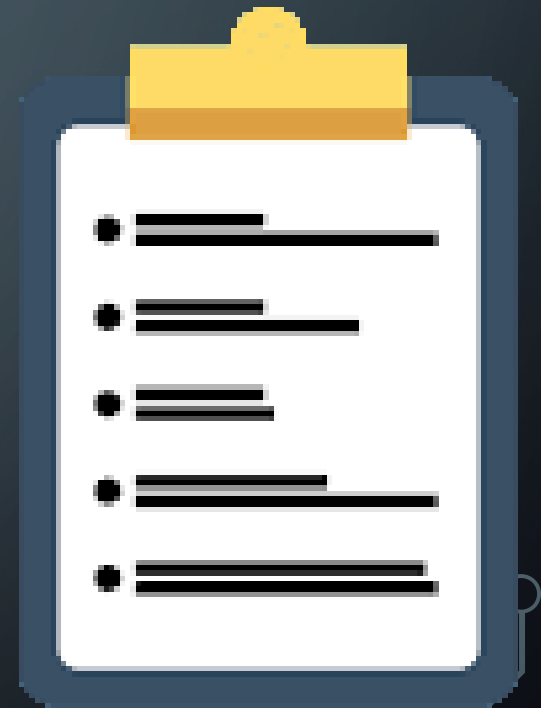


- First Law - A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- Second Law - A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- Third Law - A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.



MODIFICATIONS AND ALTERATIONS TO THE LAWS

- Asimov was ahead of his time and his Laws are often used as the basis of programming ethics into computers today.
- This does not mean that they are perfect or without issue.
- Asimov, himself, has been noted to use at least 29 variations of his own laws within his works.
- This has inspired computer scientists to think beyond the basics and charter new territory in the field.



SOME ISSUES WITH IDENTITY

- Lyuben Dilov: "A robot must establish its identity as a robot in all cases." Dilov gives reasons for the fourth safeguard in this way: "The last Law has put an end to the expensive aberrations of designers to give psychorobots as humanlike a form as possible. And to the resulting misunderstandings..."



- Nikola Kesarovski: "A robot must know it is a robot."



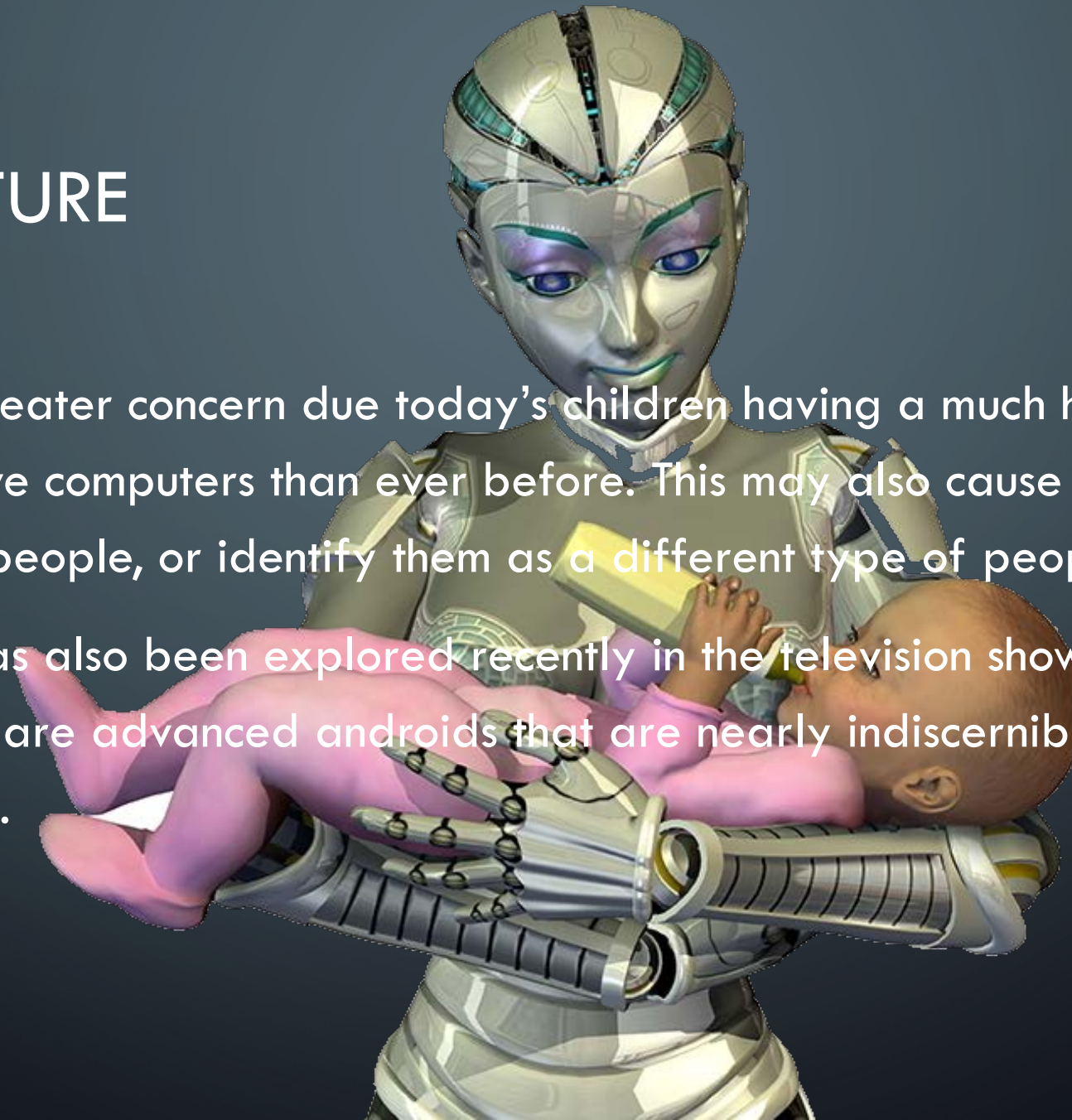
JUSTIFIED CONCERNS

- The concerns of Dilov and Kesarovski have been shown to be justified. Not just by robots being confused about their identity, but also humans!
- A very recent example is of Google Engineer Blake Lemoine, who was put on leave following his assertions that that LaMDA, an AI chatbot that his team was developing, had obtained sentience.



OUR FUTURE

- This is of greater concern due today's children having a much higher exposure to interactive computers than ever before. This may also cause them to confuse robot with people, or identify them as a different type of people.
- This idea has also been explored recently in the television show *Humans*. In the show, there are advanced androids that are nearly indiscernible from humans, at a glance.



IN MODERN MEDIA

- Since humans have programmed these androids to be eternally calm and efficient while performing their duties, human children view these beings as “better than humans.”
- There is an issue in the show of some children being identified as “synthies,” or fake synths. They pretend to be androids to the extent of pretending that they don’t have human needs such as eating or going to the restroom.



FROM ANOTHER DIRECTION

- Although interesting and thrilling to some, the development of AI separate or outside of human control is not the main concern today.
- There is a lot of focus on human-controlled AI that are quite far from the foreseen advances from “I, Robot,” but are also more advanced than people might realize
- Privacy is one of the largest issues, because a robot will not forget as long as it has space on its system and its system doesn’t malfunction.



PRIVACY

- AI is being used to obtain and process data from humans, which is then used to market products and services back to those humans.
- Advances in robotic medical may be held back due to concerns about data security. What do they do with the massive amount of data they may collect while aggregating medical files and information?
- Who owns the data at that point?



AI... ART?

- A new issue has exploded into public relevance in 2022 with the increased use of Stable Duffusion models. These are AI systems that are able to generate images based on text-based prompts. A popular example one being the NovelAI system available to subscribers.
- The issue of these systems are that they aggregate images from real artists to generate those “new” images.
- Where does the credit go? If these images are sold, who is entitled to payment?



BIAS AND SAFETY

- Another issue with machine learning is the development of bias.
- Currently, most people think of machines as perfectly neutral arbiters of data.
- This immediately becomes a mere facsimile when you figure in human control and or influence.
- They are limited by human bias or lack of fairness and ethics.



HUMAN LIMITS

- There are many instances of machines not working properly for darker skinned people due their sensors not being calibrated for that level of skin hue.
- This doesn't even require bigotry or racism on the programmer's part. It can result merely from a lack of forethought or a limited testing pool.
- However, AI developed in this capacity will therefore be limited in their ability to fairly process data.
- Since human aren't perfect, can any thing we create be so?



WARFARE

- AI is already used extensively to calculate and guide our modes of transportation.
- They are also used to quickly identify military targets.
- A newer issue is autonomous weapon development.
- There are designs for AI-controlled flight carriers. Are we signing our death warrants by entertaining these possibilities?



CONCLUSION

- This is a broad topic that I've barely touched the tip of, but the main thrust of this presentation is that the development of A.I. has always been partially tempered by fear of our possible creations.
- This seems to be changing as more people are entering the field and blazing new paths in developmental A.I.
- Although it may seem pedantic, I think the concerns of yesteryear should remain in the forefront of our psyche as we continue development of these amazing technologies. Otherwise we cannot ensure the safety of ourselves and our children.

