Impact of AI

Will AI threaten the survival of humanity?

1) Arguments for
   - Taking jobs away from humans
   - Misuse by authoritarian governments
   - Speed of development – not much thought about ethics and societal impact
   - Melding of human brains and AI can be dangerous
     - Security concerns
     - Malicious hackers hacking into your brain
   - Military uses
     - Autonomous drones, autonomous weapons that are allowed to kill
   - Sorcerer’s Apprentice problem
     - Unintentional killing of humans because we are in the way
     - What if AI doesn’t value human life?
   - Corporations using AI for their own purposes e.g. to make money rather than caring about societal welfare

2) Arguments against
   - AI used to help in medicine and healthcare. It has helped humans survive and thrive
   - 3 laws of robotics:
     - Can’t injure human being
     - And the other two
   - AI designers can install safeguards to prevent threats to the existence of humanity
   - Hard to build “good” AI that would be a monumental threat to humans. A lot of “narrow” AI and not much “general” AI.
• AI has helped humans be a lot more efficient
• It’s really up to us to decide what happens to AI and humanity

How can we be responsible with AI?
• Transparent development
  o Open source is generally a great idea
  o Explainable AI – can the AI system explain why it made a decision
• Green AI – AI uses a lot of power
• Establishing ethics boards
  o Detecting bias in AI decisions
• Testing and verifying AI programs
• What is the intent? Need guidelines on responsible use
  o Anticipating malicious
• Always include a manual override