



# Artificial Intelligence (AI)

## Introduction



# What is AI?

- Designing Machines which can mimic the Human behaviour & Intelligence
- Human vs Rational
- Thought vs Behaviour



# Acting humanly: The Turing test approach

- Can a machine think?
- Computer passes the test if a human interrogator, after posing some written questions, cannot tell whether the written responses come from a person or from a computer.
  - Natural Language Processing
  - Knowledge Representation
  - Automated Reasoning
  - Machine Learning
  - Computer Vision
  - Robotics

# Thinking humanly: The cognitive modelling approach

- Introspection
  - Trying to catch our own thoughts as they go by.
- Psychological experiments
  - Observing a person in action.
- Brain imaging
  - Observing the brain in action.
- Breakthrough: Brain-computer interface (BMI)

# Thinking rationally: The “laws of thought” approach

- “Socrates is a man, all men are mortal, therefore Socrates is mortal” by Aristotle – Greek philosopher
- Logic
- Probability
- Rational **thought** itself is not enough for intelligent **behavior**, rather it requires rational **action**.

# Acting rationally: The rational agent approach

- **Agent** is just something that **acts**
  - Operate autonomously, perceive their environment, persist over a prolonged time period, adapt to changes and create & pursue goals.
- **Rational agent** is one that acts so as to achieve the best outcome
  - Making correct inferences (take decision after careful deliberation)
  - Reflex action (doesn't involve inference)

# Beneficial machines

- Self-driving car
- Objective: Reach destination safely
- Other factors:
  - Other errant drivers
  - Equipment failure etc.
- Staying in the garage.
- Standard model (Acting rationally) is useful guide but probably not the right model!

# Beneficial machines

- Value alignment problem- achieving agreement between our true preferences and the objective we put into the machine.
- Provably beneficial to humans!





# Foundations of AI

- Philosophy
- Mathematics
- Economics
- Neuroscience
- Psychology
- Computer Engineering
- Control theory and Cybernetics
- Linguistics

# Different AI terminology

- Artificial Narrow Intelligence (ANI)
- Generative AI
- Human level AI (HLAI)
- Artificial General Intelligence (AGI)
- Artificial Super Intelligence (ASI)

# Risks & Benefits of AI

- Benefits: Hard to find any field untouched by AI or AI not having potential to increase performance
- Risks:
- Gorilla problem
- King Midas problem

# AI Universe

