# Section 4

Ethical Issues:
The 'man' in the middle issue

sinead.barton@mu..ie

What are ethical issues?

- Remember the difference between liable and culpable?

- If something is ethical, it is morally acceptable.

- Being ethical means doing the right thing even if there is no law to stop you.

- For example: If there was no law that said there should be a safety driver in an autonomous car, would it be ethical to allow the car to be tested on a busy road?
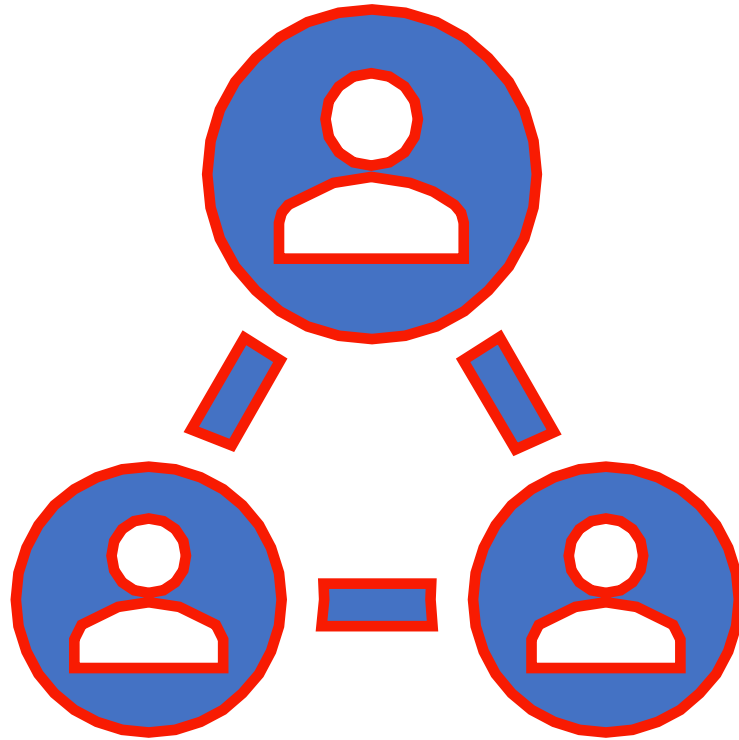
Person in the Loop / Man in the Middle Problem

- **Autonomous Robots** do not need or take any external human input in their decision making

- **Supervised/Semi-Automonous Robots** may recommend but are overseen by a human who has the final decision

Elaine Herzberg

- From our previous set of notes

- The death of Elaine Herzberg delayed the testing of autonomous vehicles in the USA

- Humans did not trust the cars, even with the safety driver

- There are things that humans do not like robots having complete control over

- In some situations a decision needs to include moral or social thinking

- It is difficult to program morality/social concepts into a robot
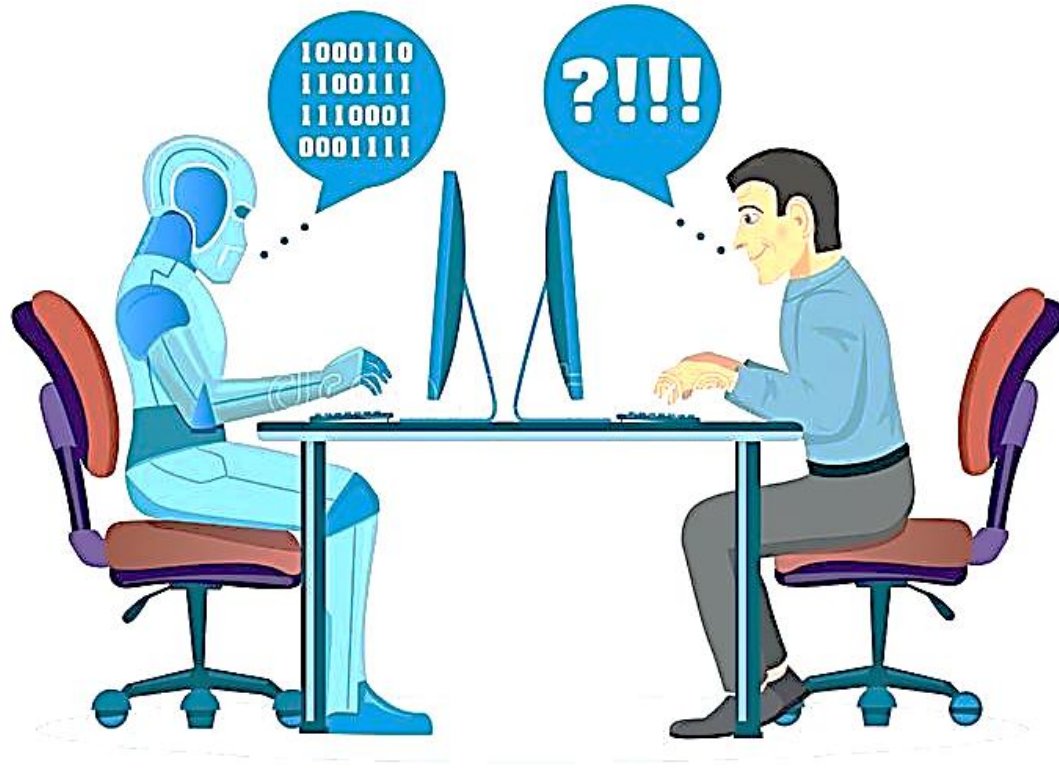
Situations where the 'man in the middle' can be problematic

4 Reasons you need still need to watch for human errors when working with robots

https://blog.robotiq.com/4-reasons-you-still-need-to-watch-for-human-error-when-working-with-robotics

If a procedure changes a human should be trained in the new procedure

When a human is not familiar with a new work environment e.g. one that has included robots, they can become nervous

When a human is nervous, they can make mistakes

It is important to manage the integration between human and machine so that the 'man in the middle' is not affected

1. Humans may not feel competent around robots

2. People could get too close to moving robots

IT IS VERY DIFFICULT TO PLAN FOR EVERY SITUATION

- Robots typically have fail-safes in case a human could be injured by being close to them.

- However, industrial robots that are not expected to interact with humans might not have these fail-safes

WANDA HOLBROOK

- In 2015, she was maintaining an assembly line robot when the robotic arm malfunctioned.

- The safety systems failed, the robot moved to a restricted area of the warehouse, and Wanda was struck and killed.

3. Robots require effective programming

- A robot is only ever as good as the person who does the
  - Design
  - Manufacture
  - Programming
- If the person does not do these things properly it can cause problems in the future
- A robot can be used for multiple purposes.
  - If it is using AI the robot can identify the job
  - Otherwise a human must tell the robot to do the job
- Mistakes can cause injury/death and can cost the company a lot of money by interrupting the workflow

4. Incorrect maintenance could affect performance

- If the person who maintains the robot does not do their job correctly, the robot could become dangerous

- If a car does not get the correct type of brakes, it could injure someone

- This affects both hardware and software!

Case Study

The Uberlingen crash, 2002

AIRPLANES ARE LARGELY AUTONOMOUS

- An airplane will operate on autopilot for approximately 90% of a standard flight

- In some airports, an 'Autoland' system is used during fog

ATC (HUMAN)

- Air Traffic Control

- Directs airplanes from the ground

TCAS (AUTOMATED)

- Traffic Collision Avoidance System

- Fitted to commercial airplanes to prevent mid-air collisions, independently of ATC

What happened?

- In 2002, a DHL cargo plane and a commercial passenger flight from Moscow collided in mid air

- There were no survivors

- All four pilots and co-pilots were very experienced

- Both airplanes were under the control of Zurich ATC

- Both airplanes had TCAS installed

How did it happen?

ATC

- Despite safety procedures there was only a single person in charge of the ATC tower (2 workstations)

- Radar data was delayed

- Overworked and without reliable data the ATC did not realise until too late that the planes were too close together.

TCAS

- The TCAS system worked perfectly

- However, the ATC gave instructions to a single airplane that directly conflicted with TCAS

- By the time the pilots were able to visually confirm each others position, it was too late

Conclusion:

Robots are smart but they still need human intervention. It is important to train employees and to test their competency regularly, otherwise the robots will not effective.
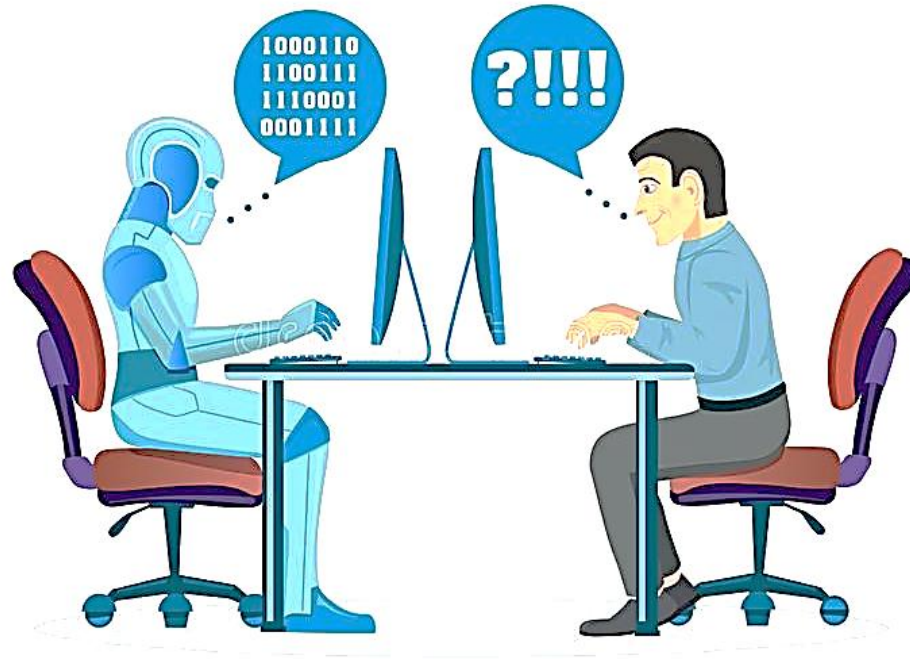
Situations where the 'man in the middle' is desirable
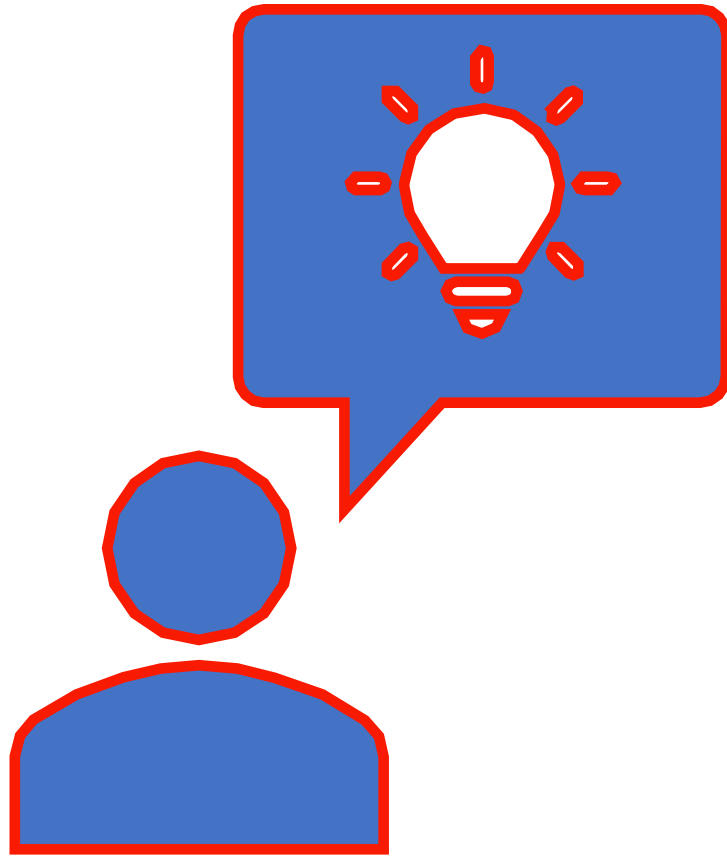
Person in the Loop

- A belief that critical decisions should always be made by a human, not by the machine

- Remember, a human is more likely to forgive another human for a mistake than to forgive a robot

Robots are useful in the medical field

Robots are:

1. Hygienic
2. Patient i.e. they do not get angry / frustrated / tired
3. Useful in remote situations
4. Able to lift large weights
5. Efficient
6. Able to do micro and macro tasks

We will talk more about this in later lectures..

However…

Would you like a machine to be diagnosing you and prescribing you your drugs – say for a life-threatening illness?

Bias in medicine

- Bias in medicine already exists

- Historically, it was male subjects that doctors used to study medicine

- It was assumed that physical symptoms would be the same in male and female patients

- However, this is not the case. For example heart attacks. A male who is having a heart attack will commonly experience severe chest pain. Women are more likely to simply feel uncomfortable or tired.

- In 2013, the top cause of death for women in the USA was heart attacks due to them being mis-diagnosed because the women did not have the typical male symptoms.

AI learns from historical data

- Remember the hiring algorithm in section 4?

- Remember we have discussed in this section that a robot is only as good as the person who programmed it?

- Unless, the AI is trained with completely un-biased and statistically representative data there is a risk that it will misdiagnose

- Can we ever be absolutely sure that the data is completely unbiased and statistically representative?

- When UPMC evaluated the risk of death from pneumonia of patients arriving in their emergency department, the AI model predicted that mortality dropped when patients were over 100 years of age or had a diagnosis of asthma

- While the AI model correctly analyzed the underlying data. It was incorrect to conclude that they were lower risk.

- Rather, their risk was so high that the emergency department staff gave these patients medicine before they were even registered into the electronic medical record

- Without understanding clinician assumptions and their impact on data could lead to AI-inspired protocols that harm high-risk patients

UPMC
University of Pittsburgh
Medical Centre
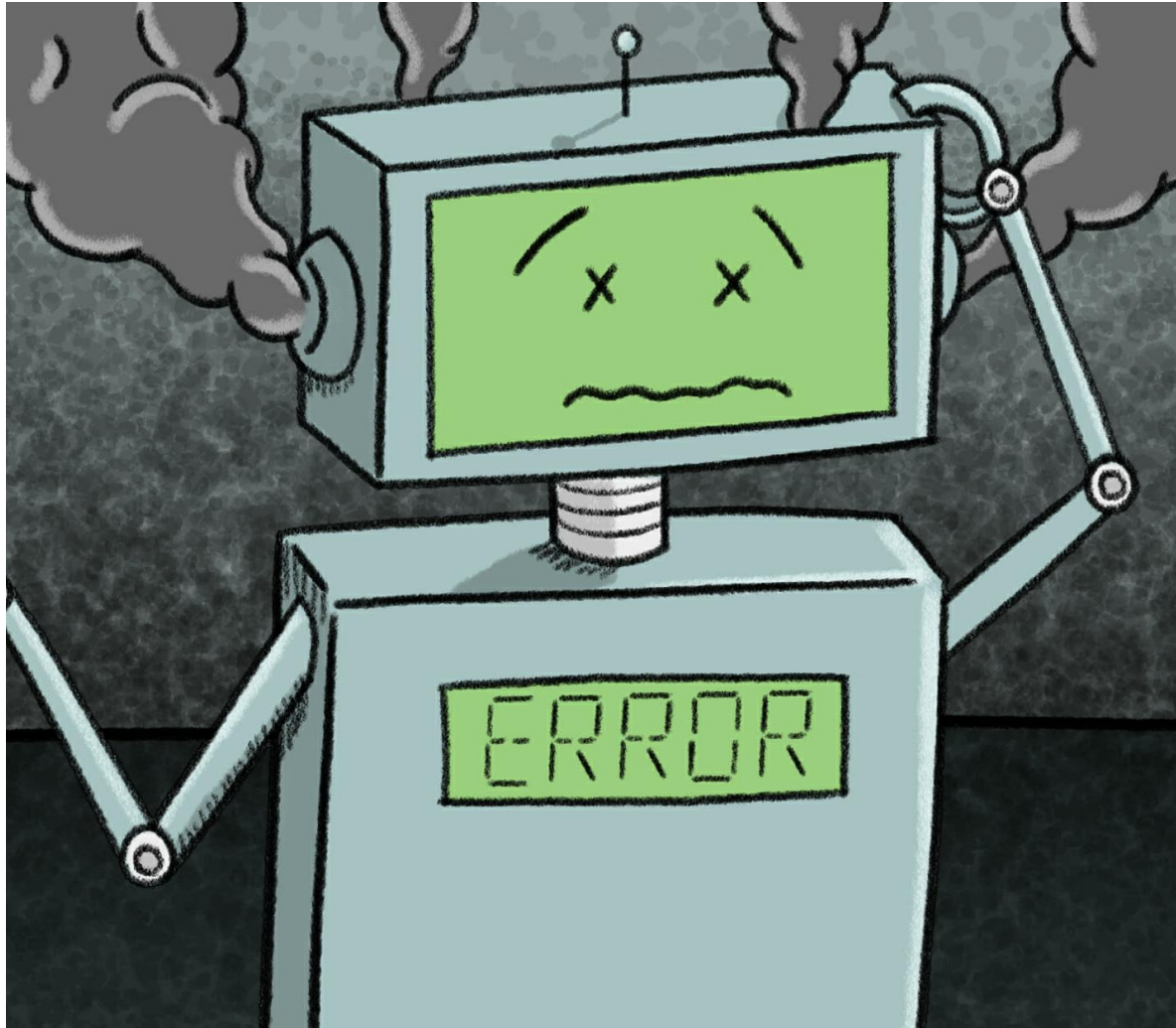
Correlation does not imply causation!

- Experienced medical staff make hundreds of informed decisions every day. It could be impossible for a robot to be programmed with all of the necessary information.

When emotions interfere / enhance...

DOCTORS VS ROBOTS

- Doctors don't want to lose their job
- Doctors will ask for help if they are not sure
- Robots cannot 'think outside the box'

PATIENTS

- People in hospital can panic
- People could react badly to a robot giving them a bad diagnosis
- Robots could make people feel alone

Not Just a Medical Problem

Programmer: a 'timer' and electronic co-ordinator of signals from thermostats.

Pump: fitted at an appropriate point on the flow pipe of the primary circuit.

Motorised valves: switches direction of the primary circuit.

Cylinder thermostat: regulates the flow of primary water through the heat exchanger.

Thermostatic Radiator Valve (TRV); enables room-by-room temperature control (by closing down radiators).

Room thermostat: controls the space heating by switching off boiler when a set temperature is reached.

Typical controls on a vented central heating system

Why do you want to have a human-controllable temperature setting on your home-heating system?

Setting it to the temperature it thinks is best for you?
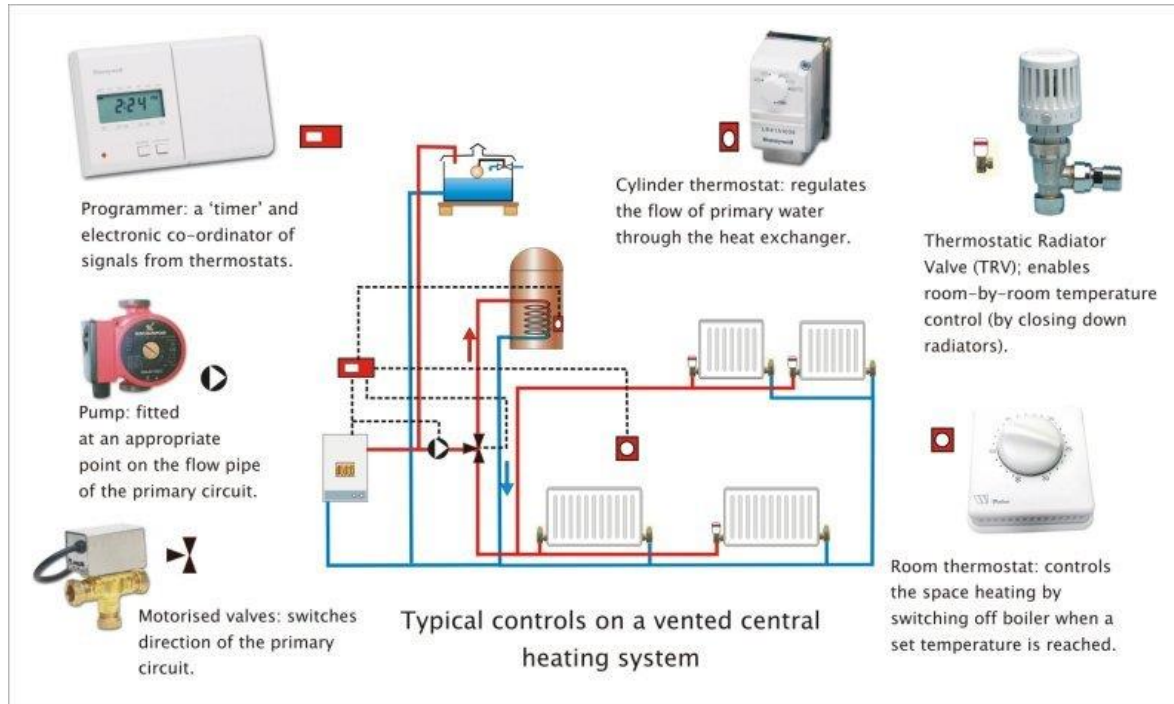
Questions

- **Ask yourself.. You code up something that takes a 1000 lines of code. The result is that you have to input a pattern of switches.**

- **One pattern gives you a reward.**

- **Any other pattern will shock you with 10 kV (non-lethal)**

- **Would you truly trust your own code... would you trust your friends? Would you trust a stranger?**

Questions

- **Is it reliable, dependable, will it ever break?**

- **Do we trust them with our lives to make the right decision?**

- **Is there a sense of a loss of control?**
  - **e.**g. does your Dad like being a passenger when you drive?

Person in the Loop / Man in the Middle

- It's about control and confidence/trust

- It's about a fear from movies that show technology in a purely negative light e.g. terminator

- It's worrying about what happens if no-one is watching – if the machine ages, breaks, gets confused. Who steps in?

Summary

Task

- **Identify 5 tasks where you want a human to be in charge – these should be normal every-day tasks.**

- **Why??**