

# Bad Machines Corrupt Good Morals

Alternative Title: How the MMCI 2021 class got  
rid of all their Smart Devices...

Import Pandas...



# Overview



- Paper: [Bad machines corrupt good morals | Nature Human Behaviour](#)
- Humans are generally good...but what happens when external pressure is applied by AI forces?
- How do you identify that pressure?
- If a machine does it: is it truly unethical/ethical?

# Social Roles

These are four roles that both humans and machines can influence ethical behavior:

1. Role Model
  - Whom you look for guidance/judgement. Children → Adult. What role do algorithms play here? Who influences more, the AI or the human?
2. Advisor
  - Can you make AI impartial? Is Alexa more of an assistant or an advisor now? Can machines give corruptible advice? Financial implications?
3. Partner
  - Partners in crime = two who act together for mutual benefit. Who have long depended on machines but like to deflect blame. Less emotional investment.
4. Delegate
  - Delegate unethical tasks to AI agents. Plausible deniability via blackbox. Marketing/advertising? Deep fakes?

# Breakout Activity

---

**Objective:** List potential concerns of AI in these roles, and potential solutions. (Policy/Industrial/Technical/Legal/etc)

---

Groups 1 + 2 = Role Model

---

Groups 3 + 4 = Advisor

---

Groups 5 + 6 = Partner

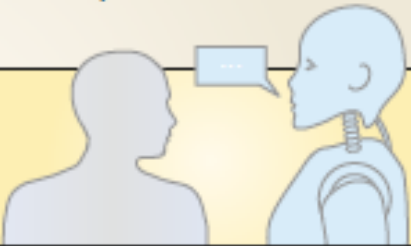

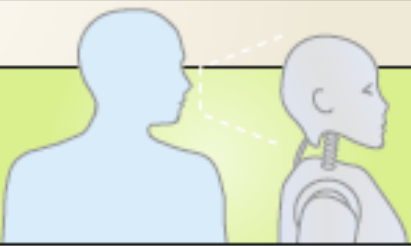
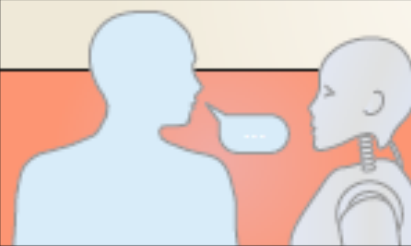
---

Groups 7 + 8 = Delegate

---

# Thoughts?

## Ethical decisions

AI as influencer			AI as enabler		
AI agents nudge humans towards unethical behaviour by displaying or recommending actions with harmful consequences			AI agents are involved in (un)ethical behaviour and allow people to pursue selfish goals while reducing guilt		
<ul style="list-style-type: none"> <li>• Conformity</li> <li>• Inertia</li> <li>• Diffusion of responsibility</li> </ul>	Customers buy harmful products on the basis of recommender systems			Students teaming up with NLG algorithms to create fake essays	<ul style="list-style-type: none"> <li>• Collaboration</li> <li>• Shared benefits</li> <li>• Shared responsibility</li> </ul>
<b>Mechanisms</b>	<b>Fears</b>	<b>Advisor</b>	<b>Partner</b>	<b>Fears</b>	<b>Mechanisms</b>
<ul style="list-style-type: none"> <li>• Observation</li> <li>• Imitation</li> <li>• Conformity to social norms</li> </ul>	Online traders imitate manipulative market strategies of trading algorithms			Outsourcing online pricing leading to algorithmic collusion	<ul style="list-style-type: none"> <li>• Displacing responsibility</li> <li>• Distancing harm</li> <li>• Anonymity</li> </ul>
<b>Mechanisms</b>	<b>Fears</b>	<b>Role model</b>	<b>Delegate</b>	<b>Fears</b>	<b>Mechanisms</b>

