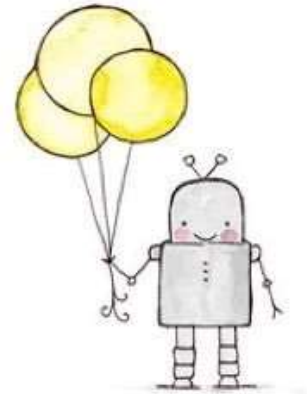


# Bridging the technical-policy gap for trustworthy AI

*Dr Jess Whittlestone*  
*Postdoctoral Research Associate, University of Cambridge*

**CFI** LEVERHULME CENTRE FOR THE  
**FUTURE OF INTELLIGENCE**





Why do we need AI policy?

## What can policy do to ensure trustworthy AI?

- Shape the way the technology is developed and used
- Shape the environment to adapt to the changes new technology brings

for example:

- GDPR shapes both how algorithms are developed and used to ensure they are transparent and accountable
- The UK AI Sector Deal is increasing investment in AI R&D which pushes forwards innovation
- Policies around technical education, reskilling and digital infrastructure aim to create an environment that can respond positively to advances in AI



How might AI  
policy go  
wrong?

# Different failure modes

- Over-regulation, stifling innovation and preventing benefits
  - Particularly likely to be a concern if those responsible for regulation aren't also responsible for realising the benefits of AI - points towards sector-specific regulation?
- Failing to manage and react to challenges as they arise
  - Failing to ensure the technology itself is sufficiently transparent, accountable, and aligned with human values to be trustworthy
  - Failing to adapt our environment to the changes AI brings - e.g. failure to adapt the job market/retrain leading to massive disruption
- Failing to anticipate new challenges
  - We may successfully navigate the challenges we're aware of right now, but new ones will inevitably arise
  - If the rate of progress continues to increase, we need to be able to think ahead even more than we have been able to so far

# Anticipating future challenges

- Need to ask: “where are AI capabilities today, where might they be at various points in the future, and how might these capabilities impact different parts of society?”
- This suggests policy needs to draw much more on technical expertise
- Gap between policy and technical work/communities in AI:
  - Very few senior policymakers have technical expertise
  - Most people developing AI aren't engaged with policy/ethics issues
  - The two groups don't always find it easy to communicate with one another



How do we bridge this gap?

## We need to:

- Improve communication between policy and technical experts
  - What are the main barriers? How much do people on each side understand about what the other does?
  - What kinds of “translator” roles might be most helpful here, and how do we create them?
- Provide opportunities for technical experts to engage more in policymaking, and with ethical/societal issues more broadly
  - Opportunities to participate on advisory boards, and to feed into policy decisions e.g the design of new centres
  - What kind of ethical/policy training should AI researchers receive, and how?
- Think about what kinds of technical understanding policymakers need, and how to provide that
  - General training is valuable, but may need something more targeted too

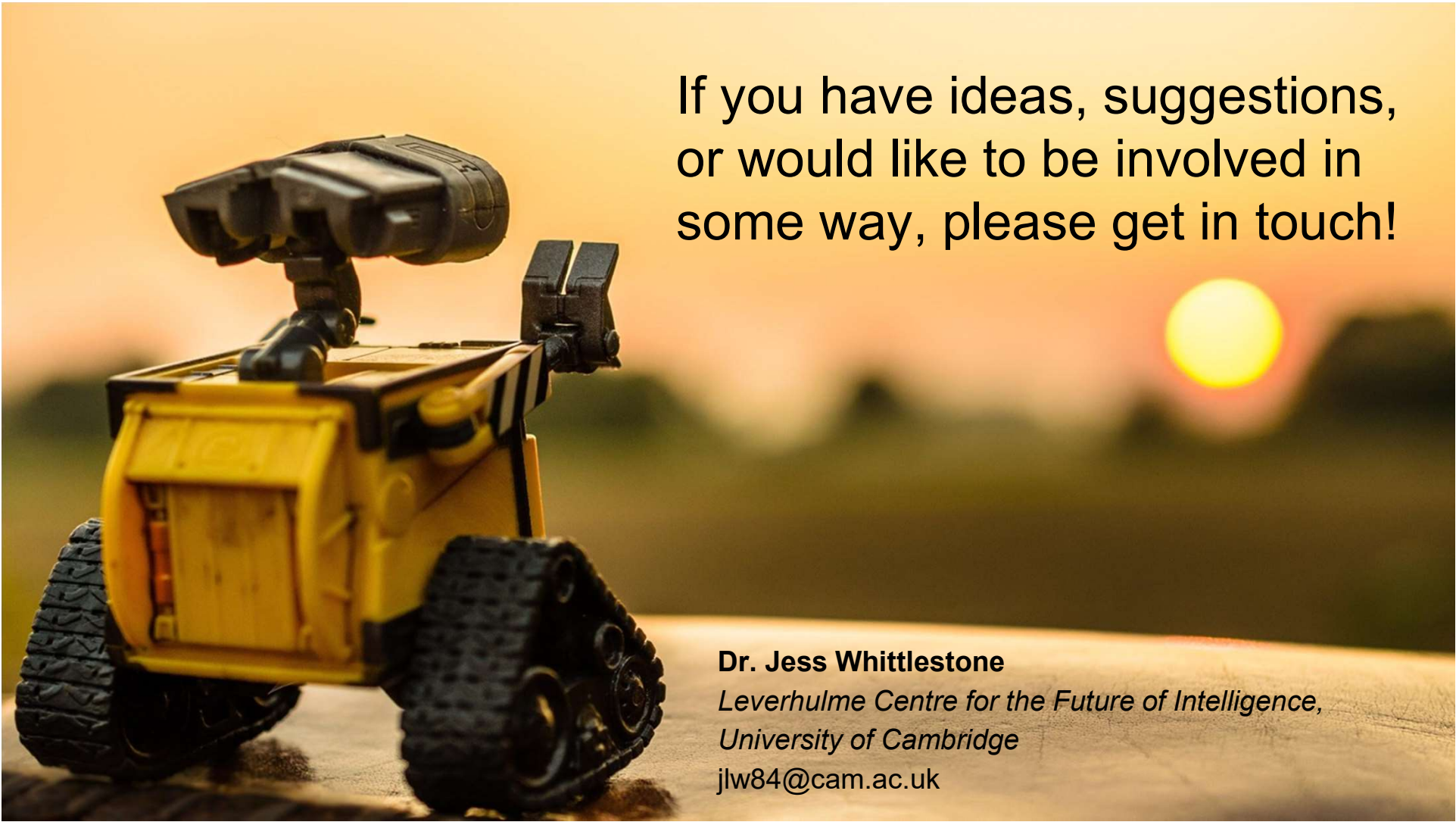


## Some things already being done

- Technical/digital training for policymakers
- “Digital coaches” for MPs
- Data ethics frameworks within government - i.e. helping people without technical expertise to ask the right questions
- Existing structure of scientific advisors within government
- Some “ethics” training for AI researchers
- Ethics and policy teams within AI companies - DeepMind, OpenAI

# This project

- Over the next year, we'll develop **concrete policy proposals** we think would help bridge the policy-technical gap for trustworthy AI, and work to build support for those proposals with policymakers (starting with the UK)
- We'll do this by talking to people working on both sides, learning from what's already been done, and drawing on lessons from past cases of science and technology policy (e.g. biotech, climate change)
- Current partners:
  - Dr. Laura James, Trustworthy Technologies Project at the University of Cambridge
  - Dr. Marc Warner, CEO of ASI Data Science
  - Dr. Tanya Filer, Bennett Institute for Public Policy



If you have ideas, suggestions,  
or would like to be involved in  
some way, please get in touch!

**Dr. Jess Whittlestone**

*Leverhulme Centre for the Future of Intelligence,  
University of Cambridge*

*jlw84@cam.ac.uk*