



THE UNIVERSITY
of EDINBURGH

goals and behaviours can be assured to align with human values throughout their operation. AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity. People should have the right to access, manage and control the data they generate, given AI systems' power to analyse and utilise that data. The application of AI to personal data must not unreasonably curtail people's real or perceived liberty. AI technologies should benefit and empower as many people as possible. The economic prosperity created by AI should be shared broadly, to benefit all of humanity. Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives. The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends. An arms race in lethal autonomous weapons should be avoided. Longer-term Issues. Unless there is consensus on international treaties regarding upper limits on future AI capabilities. Advanced AI could represent a profound change in the history of civilisation and should be planned for and managed with commensurate care and resources. Risks posed by AI systems, especially those that are recursive, must be subject to planning and mitigation efforts commensurate with their expected impact. AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures. Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization. Rigour, honesty and integrity. Act with skill and care in all scientific work, and assist their development in others. Take steps to prevent corrupt practices and professional misconduct. Develop public interest. Be alert to the ways in which research derives from and affects the work of other people, and respect the intellectual property of others. Respect. Respect for life, the law and the public good. Ensure that your work is lawful and justified. Minimise and justify any adverse effect your work may have on people, animals and the natural environment. Responsibility. Responsible communication: listening and informing. Seek to discuss the issues that science raises for society. Listen to the aspirations and concerns of others. Do not knowingly mislead, or allow others to be misled, about scientific matters. Present and review scientific evidence, theory or interpretation honestly and accurately. Ensure they do not infringe on internationally recognised human rights. Prioritise metrics of well-being in their design and use. Ensure that their designers and operators are responsible and

Critical Codes Roundtable

A Public Interest Technology event on Ethical Codes of Conduct as a mode of governance for emerging technologies

Friday 18th November 2022

High School Yards, University of Edinburgh

Robert Smith, Matjaz Vidmar, Filippo Cuttica, Sophie Stone

<https://rri.ed.ac.uk/criticalcodes/>

PITCASES

Welcome!

We are delighted to welcome you to the University of Edinburgh's roundtable event on Codes of Conduct as a mode of governing emerging technologies! We are excited to have assembled around 25 scholars and practitioners with diverse experiences of studying or working with codes of conduct and ethical standards to participate in our discussions.

In an age of rapid technological development, those in power often reach for codes of conduct to guide the practices of technology developers. Codes such as the [IEEE's Ethically Aligned Design](#), the [Asilomar AI Principles](#) or the [Barcelona Declaration](#) aim to capture moral norms in succinct, widely applicable formats. Codes of conduct, and cognate ideas such as principles, standards and guidelines are now a well-established tool in the repertoire of 'emerging tech ethics'.

The goal of our roundtable is to draw together key critiques of ethical codes of conduct with interests from Public Interest Technology (PIT), Responsible Research and Innovation (RRI) and Science and Technology Studies (STS). We want to look beyond the content of bullet-pointed manifestos to think about what it takes to use these codes to transform cultures and practices. Further, is it possible to reframe ethical standards as an ongoing democratic activity?

We are one of four host institutions who have been asked by the Public Interest Technology University Network ([PIT-UN](#)) to convene a roundtable event as part of their broader Technologist Code of Ethics project. The aim of this project is to combine cross-sector expertise and experience to better understand ethical codes of conduct as a mode of governance.

Throughout the session, we will discuss four key areas of interest:

- What insights and lessons learned come from working with and studying codes?
- What relationships do codes have with broader social and institutional contexts?
- Who should be involved in elaborating codes, who decides, and what processes are needed to support inclusivity and participation?
- What, if anything, needs to change for them to function as effective forms of governance?

We have invited each of you because you bring a set of experiences related to this topic and are keen to hear your thoughts. We hope to maximise the shared knowledge in the room on the day, **so please do come prepared to think and talk about your experiences, lessons, successes, discomforts, and failures.** To that end there is a **small piece of prep work on the next page.**

We look forward to hearing your thoughts!

Rob, Mat, Filippo, and Sophie

Important Information

A small amount of preparatory work

One of our group's strengths is its diversity of experience and expertise. To make good use of that, and to guide our conversations on the day, we are asking for a very small amount of preparatory work ahead of the session.

Before Friday, **please can you identify an example of a code of conduct, set of guidelines or ethical standard that represents a 'lesson learned' about working with and engaging with codes.**

The example you think of could be a code you consider exemplary, terrible, or anywhere in between. These examples will provide a focal point for some of the structured discussion activities during the event, so please do try to contribute if you can.

Once you have identified an example, please [email it to Sophie Stone](mailto:sophie.stone@ed.ac.uk) (sophie.stone@ed.ac.uk) by **Thursday 17th November, 5pm**. We will print out copies for the discussion. That's it!

Documenting conversations

We have been asked by the Public Interest Technology University Network ([PIT-UN](#)) to produce a report that summarises our discussions and responds to a series of questions they have.

To do this, we will be documenting the roundtable in two ways:

- We'll record speaker talks and plenary discussions. We'll use the recording to provide summaries of this aspect of the event. After producing the summaries, the recording will be deleted.
- For all other parts of the session, our work will be captured using a digital whiteboard called [Miro](#). For each discussion a member of the Science, Technology and Innovation Studies (STIS) department will be responsible for documenting the discussion. Unless you have been contacted directly to assist with this, you will not need to know how to use Miro, so please don't worry about learning new tools ahead of the workshop!

We will be adopting Chatham House rule, i.e. you are free to discuss the roundtable but please do not attribute positions to individuals without their prior consent.

Writing up

After the roundtable, we will take the insights from the day and generate a report to submit to the PIT-UN network in mid-December. We would like this to be a collaborative activity, are keen to draw on your knowledge, and do not want to misrepresent positions taken on the day. As part of that synthesis phase, we will circulate a draft paper to you all for your comment and contribution. To support that activity, we may also follow up on some of the ideas or experiences to have emerged from the session. We hope you will be keen to join us in an ongoing dialogue about the insights that are generated!

Programme

Further instruction and running times for each of the sessions will be shared on the day. However, please find below some short summaries that will give you an indication for the way we will be structuring the day.

10:30 – 11:00am: Meeting, greeting, and introductory statements

We hope you will be able to join us for some light refreshments. We will make some Introductory statements at 10:50am, with a view to starting the session promptly at 11am.

Teas, coffees, and biscuits will be made available.

11:00am – 3:00pm: Roundtable discussions

Morning Session: Critiquing Codes

The theme of the morning session is 'Critiquing Codes'. This session will open with two short talks or provocations on the theme from two of our speakers, Emma Frow and Nina María Frahm. We will then spend the rest of the session drawing on your experiences to reflect on the lessons learned from engaging with codes and situating them within their broader contexts. This session will comprise two parts (with a short break in between) and the Critical Codes team have prepared some exercises to help guide the discussions that will be introduced on the day.

Lunch break: Approximately halfway through the event, there will be a 45-minute lunch break.

Sandwiches, teas, and coffees will be provided.

Afternoon Session: Rethinking Codes

The theme of the afternoon session is 'Rethinking Codes'. This session will open with another two short talks or provocations on the theme from two more of our speakers, Shannon Vallor and Luke Stark. We will then spend the rest of the session drawing on your experiences to discuss themes of inclusion and participation in the development of codes of conduct, as well as focusing on ways to develop codes within their broader contexts to be more democratic and effective. Once again, the afternoon session will comprise two parts (with a short break in between) and the Team have prepared a further set of exercises to structure those discussions. We will share all the details on the day.

After the session

Whilst the formal session will end at 3pm, anyone who does not have to leave immediately is welcome to join us for informal drinks and bite to eat in Edinburgh to carry on our discussions.

Please let a member of the Critical Codes Team know if you are interested in joining us after the event!

People

Organisers

Robert Smith	University of Edinburgh
Matjaz Vidmar	University of Edinburgh
Filippo Cuttica	Independent Consultant
Sophie Stone	University of Edinburgh

Speakers

Nina María Frahm	Arhaus University
Emma Frow	Arizona State University
Luke Stark	University of Western Ontario
Shannon Vallor	University of Edinburgh

Participants

Cennydd Bowles	Independent Consultant
Martin Boyle	CodeBase
Jane Calvert	University of Edinburgh
Benedetta Cantanzariti	University of Edinburgh
Morgan Currie	University of Edinburgh
Steven Drost	CodeBase
Steven Earl	University of Edinburgh
Jess Freaner	IDEO
James Garforth	University of Edinburgh
Nina Kojima	University of Glasgow; Partisan Media
Reuben Message	University of Edinburgh
Lena Podoletz	University of Edinburgh
Vardev Sachdev	University of Edinburgh
Jamey Wetmore	Arizona State University
Xiao Yang	University of Edinburgh
Andrew Youngson	CodeBase