Foresight in Development

Lots of people identify why technology can be great.

We need to also focus on:

- What are the foreseeable risks and limitations of this system?
- What are the foreseeable harms this system can cause?

Questions to Guide Foresight:

Work through **People * Contexts**

People

- Who are the *intended* users?
- Who are the *foreseeable* users?
  - Include unintended and malicious users
- Who are the people that may be *affected*?
- Work through *identity characteristics*
  - Age, ability status, religion…
  - With identity, there's a high risk of causing harm

Contexts

- What are the contexts in which this might be used?
- How might this be misused?
  - Used incorrectly?
  - Abused/Used maliciously?

Chart to help guide Foresight in AI

		People			
		Users		Those affected	
		Intended	Unintended Both malicious actors & people un-accounted for in development	Intended	Unintended Both people in training data & people the technology is used on
Use Contexts	Intended	🟩		🟩	
	Unintended Both harmful contexts & those unmodeled in development		🟥		🟥
	Out of scope		🟥		🟥

Unintended: Results unpredictable
Out of scope: Won’t work

Foresight in AI Chart: Text-to-Image example

Some (of many) example people and use contexts

		People			
		Users		**Those affected**	
		Intended *Artists*	**Unintended** *Malicious ex-partner*	**Intended** *Art appreciators*	**Unintended** *Unconsented artists whose art is in training data*
Use Contexts	**Intended** *Socially acceptable creative art*	🙂		🙂	
	Unintended *Mis/disinformation (e.g., Deep Fakes)*		**Revenge porn**		**Creative content that should be paid for, extrapolated in training, is plagiarised**
	Out of scope *Medical images*				