Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Ethics in Technology Innovation

Dave Lewis, dave.lewis@scss.tcd.ie
Thanks to: Wessel Reijers, Arturo Calvo, Killian Levacher

Student Online Teaching Advice Notice

The materials and content presented within this session are intended solely for use in a context of teaching and learning at Trinity.

Any session recorded for subsequent review is made available solely for the purpose of enhancing student learning.

Students should not edit or modify the recording in any way, nor disseminate it for use outside of a context of teaching and learning at Trinity.

Please be mindful of your physical environment and conscious of what may be captured by the device camera and microphone during videoconferencing calls.

Recorded materials will be handled in compliance with Trinity's statutory duties under the Universities Act, 1997 and in accordance with the University's policies and procedures.

Further information on data protection and best practice when using videoconferencing software is available at https://www.tcd.ie/info_compliance/data-protection/.

Why Should Tech Innovators be Concerned with Ethics?

- Because new technologies have a **profound impact** on the way **we live**, on the **relationships we have**, on the **societal & political processes we engage in**.

- For tech innovators?
  - First: because it is good for the image of your business (instrumental goal)
  - Second: because it actually improves the service you provide! (substantive goal)
  - Third: because it is the *good* thing to do, it contributes to your idea of a better society and being a good person (normative goal)

Dominant Views in Technology Ethics

- The neutrality thesis: technologies are *instruments* that we can use to attain our own goals.
  - "People kill people"
- The determinism thesis: technologies *dictate* everything we do, they determine who we are.
  - "Guns kill people"
- The co-shaping thesis: technologies and humans together "construct" our social world.
  - "Gun-men kill people"

Technology Impact: Example

Software Malfeasance: Example

Unintended Impacts - Example: Gender in Google Translate

- Some languages, like Turkish, don't has gender specific pronouns
- Google translate has to guess the gender when translating in English
- Statements allocating gender to role reveal gender bias
- What is the source of this?
- Is it a problem?

Sample Google Translate output:

he is a soldier
she's a teacher
he is a doctor
she is a nurse

https://qz.com/1141122/google-translates-gender-bias-pairs-he-with-hardworking-and-she-with-lazy-and-other-examples/

Power of Big Data: Example: Cambridge Analytica

- Academic research into Psychographics (U. Cambridge) revealed the link between philological profiles and Facebook profiles
- Correlated major psychological types to elements in the social graph: Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism
- Cambridge Analytica applied psychographics to help target political ads in 2016 US elections….

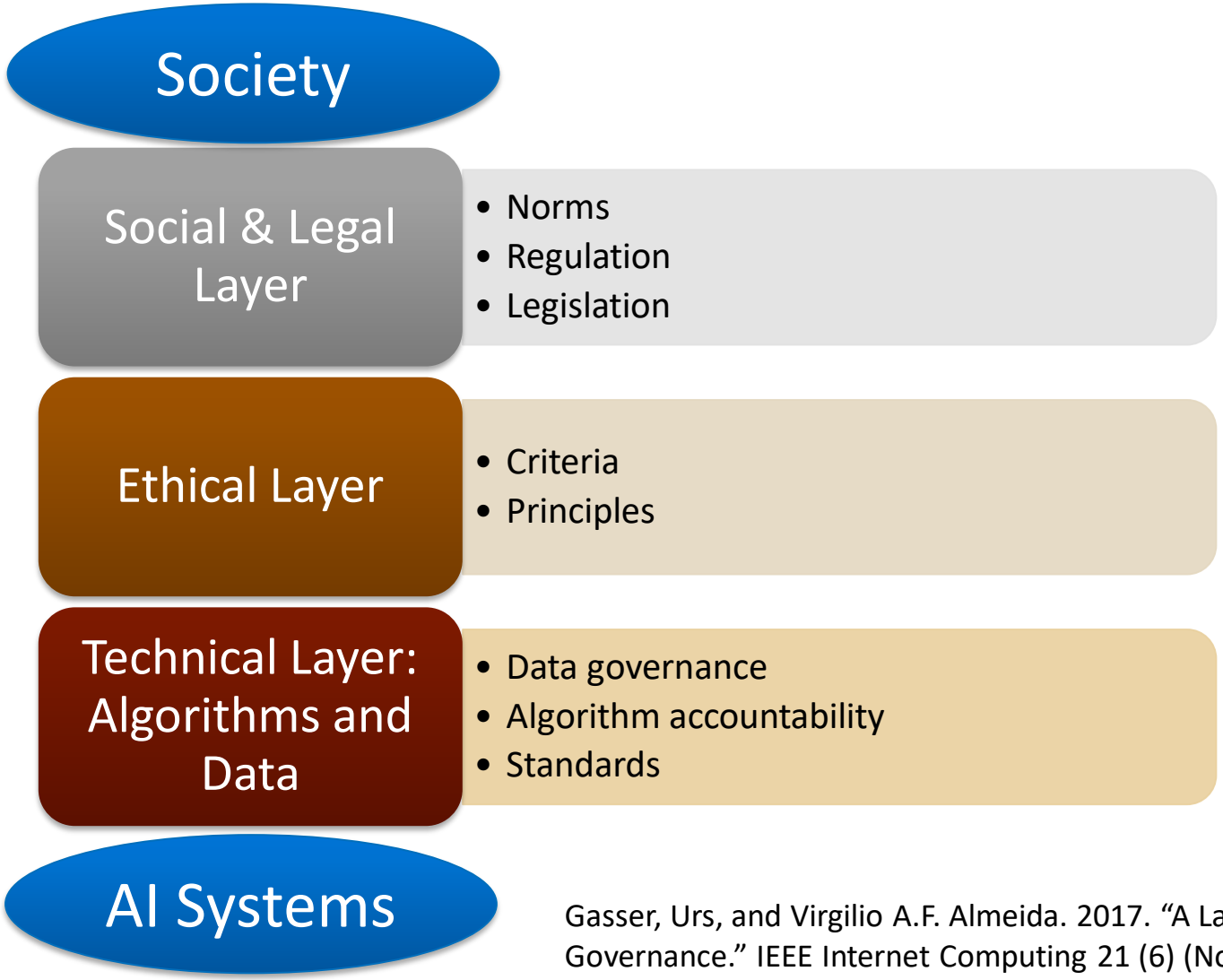
https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

Algorithmic Power on Behaviour & Worldview

- “Race to the Bottom … of the Brain Stem” Tristian Harris
- 70% of YouTube views are based on algorithmic recommendations
- Business model maximises video views to maximise ad views
- Outrage/fear/anger the most reliable reactions that drive us to keep watching
- -> Recommender algorithm inevitably drive us to content that builds outrage to keep us watching

- Evidence to US Congress: https://www.youtube.com/watch?v=WQMuxNiYoz4
- Agenda: https://humanetech.com/wp-content/uploads/2019/06/Technology-is-Downgrading-Humanity-Let%E2%80%99s-Reverse-That-Trend-Now-1.pdf

AI Governance: Layered Model

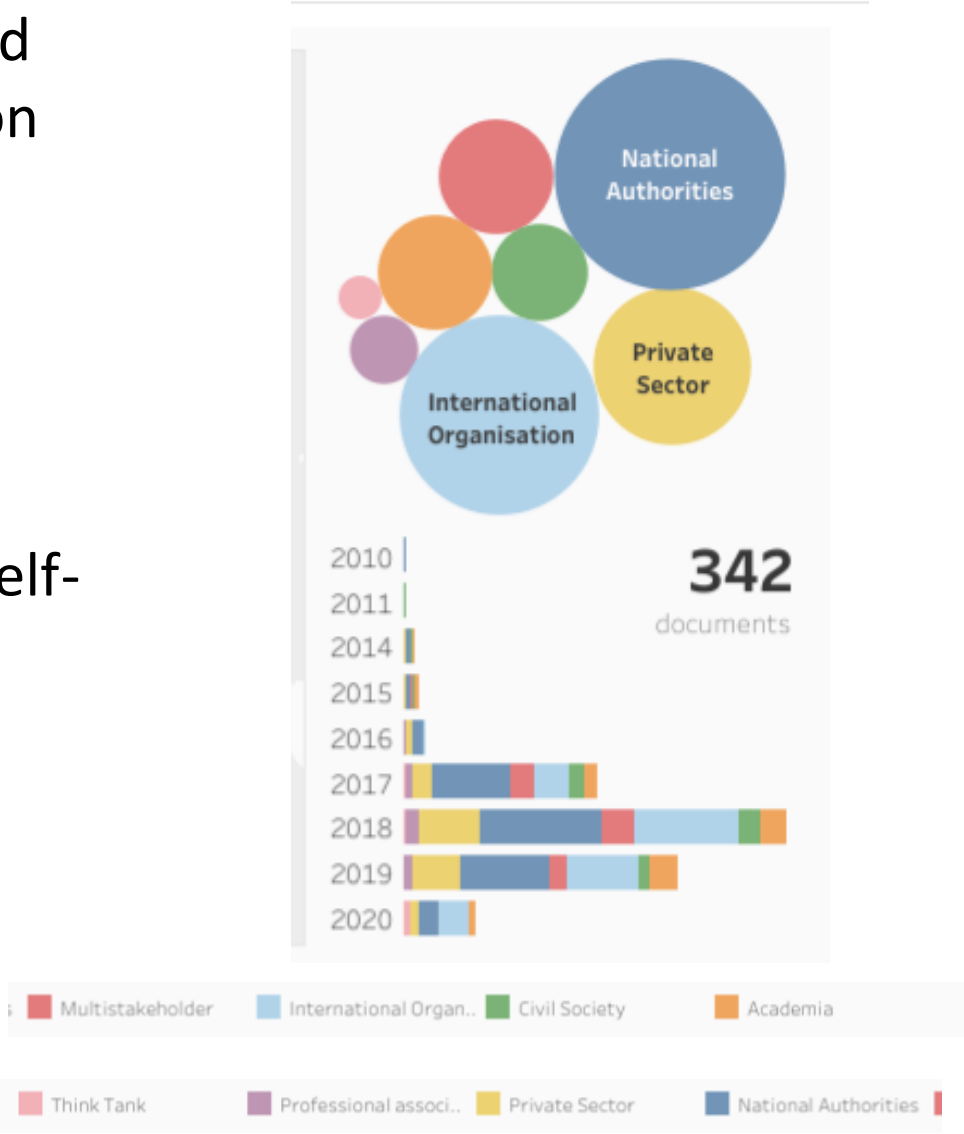


Gasser, Urs, and Virgilio A.F. Almeida. 2017. "A Layered Model for AI Governance." IEEE Internet Computing 21 (6) (November): 58–62.

AI Ethics – where next?

https://www.coe.int/en/web/artificial-intelligence/national-initiatives

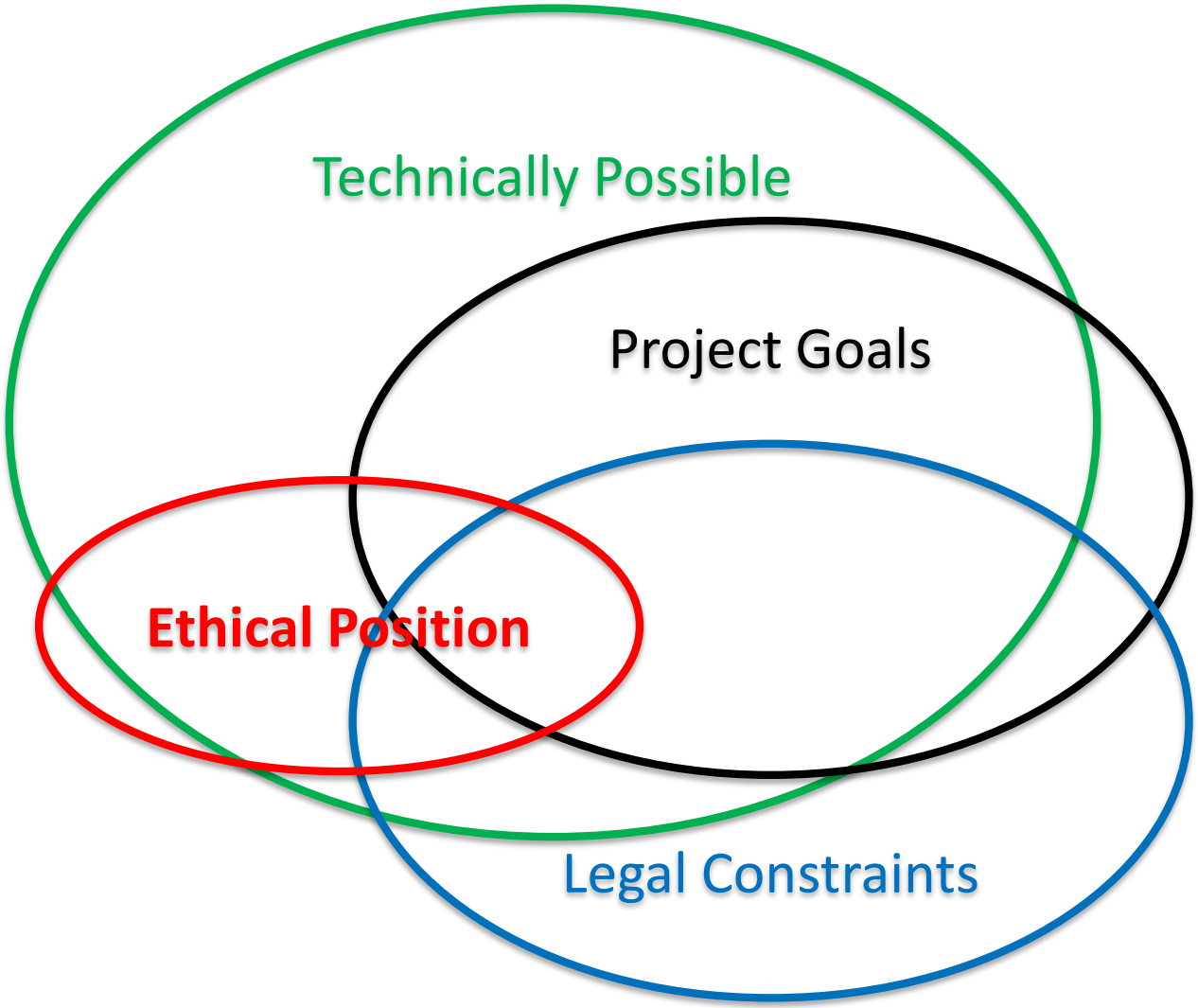
- Many countries, companies and transnational bodies working on principles or guidelines
- Some taking a wait-and-see approach BUT EU leading with recent white paper
- AI industry has failed to form self-regulation – companies trying their own, e.g. FaceBook ‘Supreme Court’
- So far not much guidance for industry practitioners



AI: Challenges to Tech Regulation

- **Pacing:** AI tech and applications develop faster that societies ability to regulate it
- **Securitisation**: International competition as AI perceived as a strategic economic/military resource
- **Innovation:** Perceived impediment to AI-based innovation and its economic and social benefits
- **Asymmetry:** Power of AI concentrated in a few digital platforms that benefit from massive network effects

Ethics in a Technology Development Project



[IBM]

Practicing Ethics in Responsible R&I

- Levels of practising ethics on responsible R&I (Brey, 2000):
  - ***Disclosure***: exploration and identification of ethical impacts
  - ***Theoretical***: frameworks to evaluate the impacts
  - ***Application***: moral deliberation to overcome negative impacts

- ***Disclosure* level** neglected in current methodologies
- Need to:
  - Keep pace with **volume and speed** of innovation
  - **Accessible** to non-ethicist
    - R&I teams have an important perspective
    - R&I teams position to implement pivot to mitigate negative impact
  - Enabling a **collaborative** process

Ethics Canvas: Lightweight approach

- **Ethic Canvas** is a **methodology** for identifying, evaluating and resolving **ethical impacts** during **R&I** stages:
  - Formation of knowledge and concepts
  - Design of the technology
  - Prototyping and testing
  - Integration of R&I outcomes into society

- **Foster ethically informed technology design** by engaging R&I teams with the ethical impacts
- Transform affordances of popular *Business Model Canvas* into an ***Ethics Canvas***
- **Collaborative brainstorming tool** with two aims:
  - Help teams identify, discuss and articulate possible ethical impacts
  - Bring about *pivots* in the design

Unmediated Reflective Approach

- We can use the Ethics Canvas as tool for **capture and reflection of ethical implications** on R&I settings
- Promotes a **reflective**, **unmediated**, **easy-to-use** and **self-service** approach to the analysis of ethical issues by researches / developers
- Reflective tool for "Value sensitive design":
  - What kind of values do we want to inscribe in our application? (our vision of the Good Life)
  - How can we operationalise these values?
  - How can we "design" technologies and their applications accordingly?

Considerations on Ethical Impacts of Technology

- Changes in individual **behaviour**
- Relationships between **individuals**
- Relationships between **collective actors** who represent groups e.g. companies, unions, professional bodies, charities, elected bodies
- **Relationships** *between* individuals and collectives
- Impact in the **public sphere, conflicts,** our **worldviews**
- Impact of technology **failure**
- Impacts on the **environment**

Ethics Canvas

Project Title:

Date:

Individuals affected

Who use your product or service?
Who are affected by it's use?
Are they men/women, of different ages, etc.?

1

Behaviour

How might people's behaviour change because of your product or service?
Their habits, time-schedules, choice of activities, etc.?

3

Relations

How might relations between people and groups change because of your product or service? Between friends, family-members, co-workers, etc.?

4

What can we do?

What are the most important ethical impacts you found?
How can you address these by changing your design, organisation, or by proposing broader changes?

9

Worldviews

How might people's worldviews be affected by your product or service?
Their ideas about consumption, religion, work, etc.?

5

Group Conflicts

How might group conflict arise or be affected by your product or service?
Could it disciminate between people, put them out of work, etc.?

6

Groups affected

Which groups are involved in the design, production, distribution and use of your product or service?
Which groups might be affected by it?
Are these work-related organisation, interest groups, etc.?

2

Product or Service Failure

What are potential negative impact of your product or service failing to operate or to be used as intended?
What happens with technical errors, security failures, etc.?

7

Problematic Use of Resources

What are potential negative impacts of the consumption of resources relating to your project?
What happens with its use of energy, personal data, etc.?

8

Stage 1: Identify the Relevant Stakeholders

Who might be affected by application– be **inclusive**

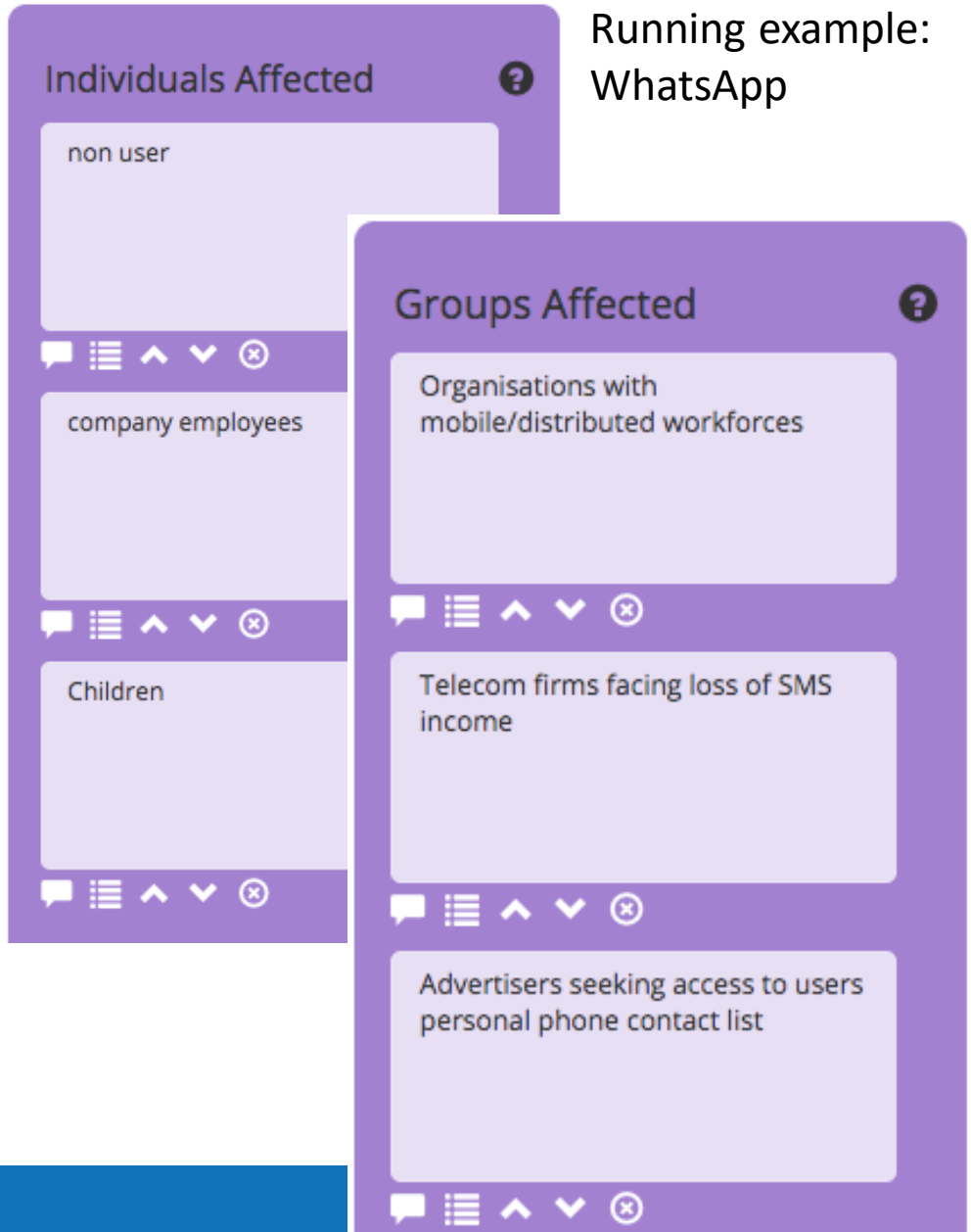
Individuals: Who use your product or service? Who are affected by it's use?

e.g are they of different genders, of different ages, etc.?

Groups: Which groups are involved in the design, production, distribution and use of your product or service?

Which groups might be affected by it?

e.g. are these work-related



Running example: WhatsApp

Stage 2: Identifying Ethical Impacts

First, **'micro' impacts** are captured by the canvas, i.e. on everyday lives of people using and living with the application

Behaviour: How might people's behaviour change because of your product or service?

e.g. habits, time-schedules, choice of activities, etc.?

Relations: How might relations between people and groups change?

e.g. between friends, family members, co-workers, etc.?

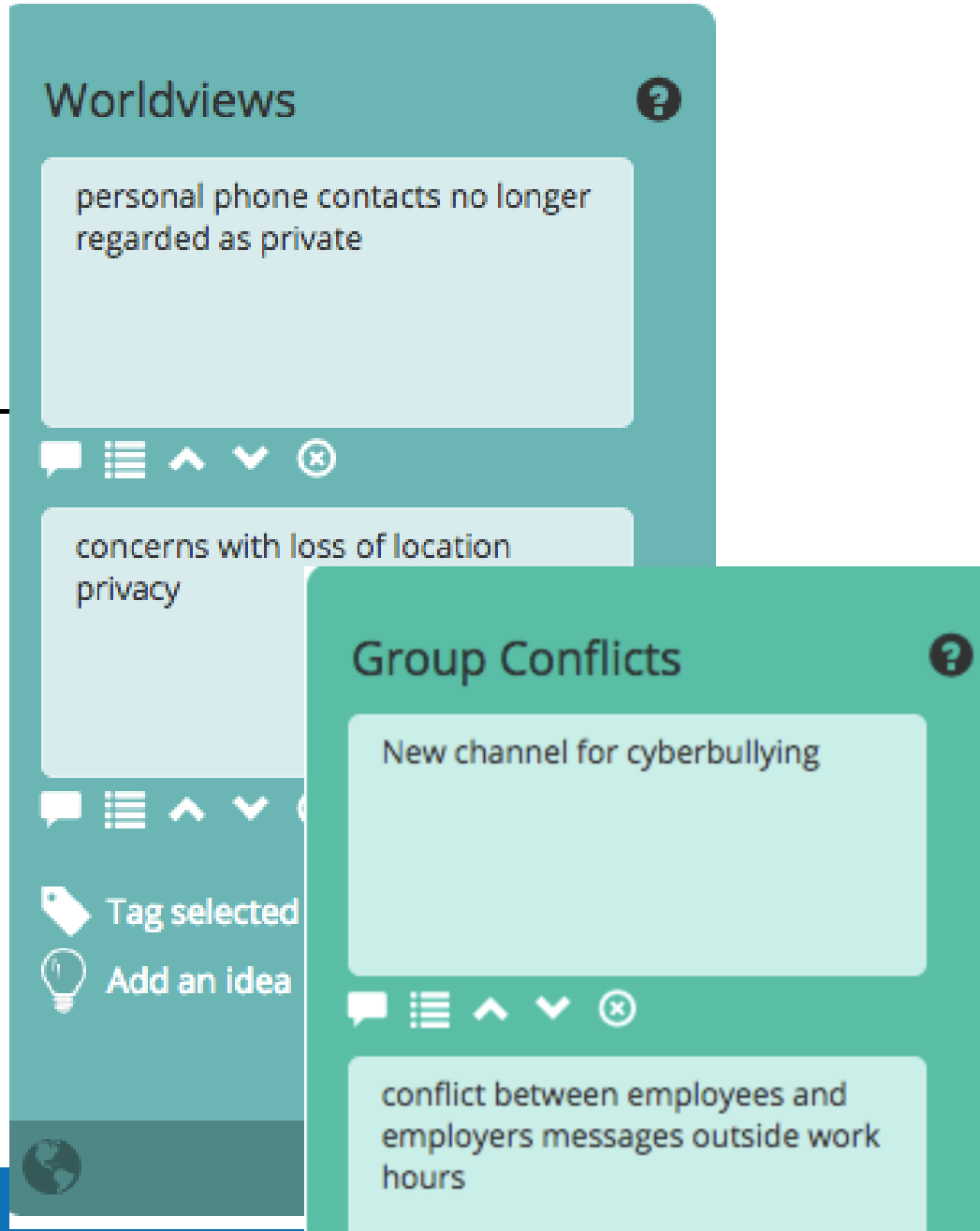
Stage 2: Identifying Ethical Impacts

Next **'macro' impacts** need to be considered.

These surpass individual's impacts - pertain *to* *collective*, social structures instead, e.g. related to political structures or cultural value-systems.

How might people's **Worldviews** be affected by your product or service? *e.g. their ideas about consumption, religion, work, etc.?*

Social conflicts: How might **Group Conflict** arise or be affected? *e.g.discriminate between people, put them out of work, etc.?*

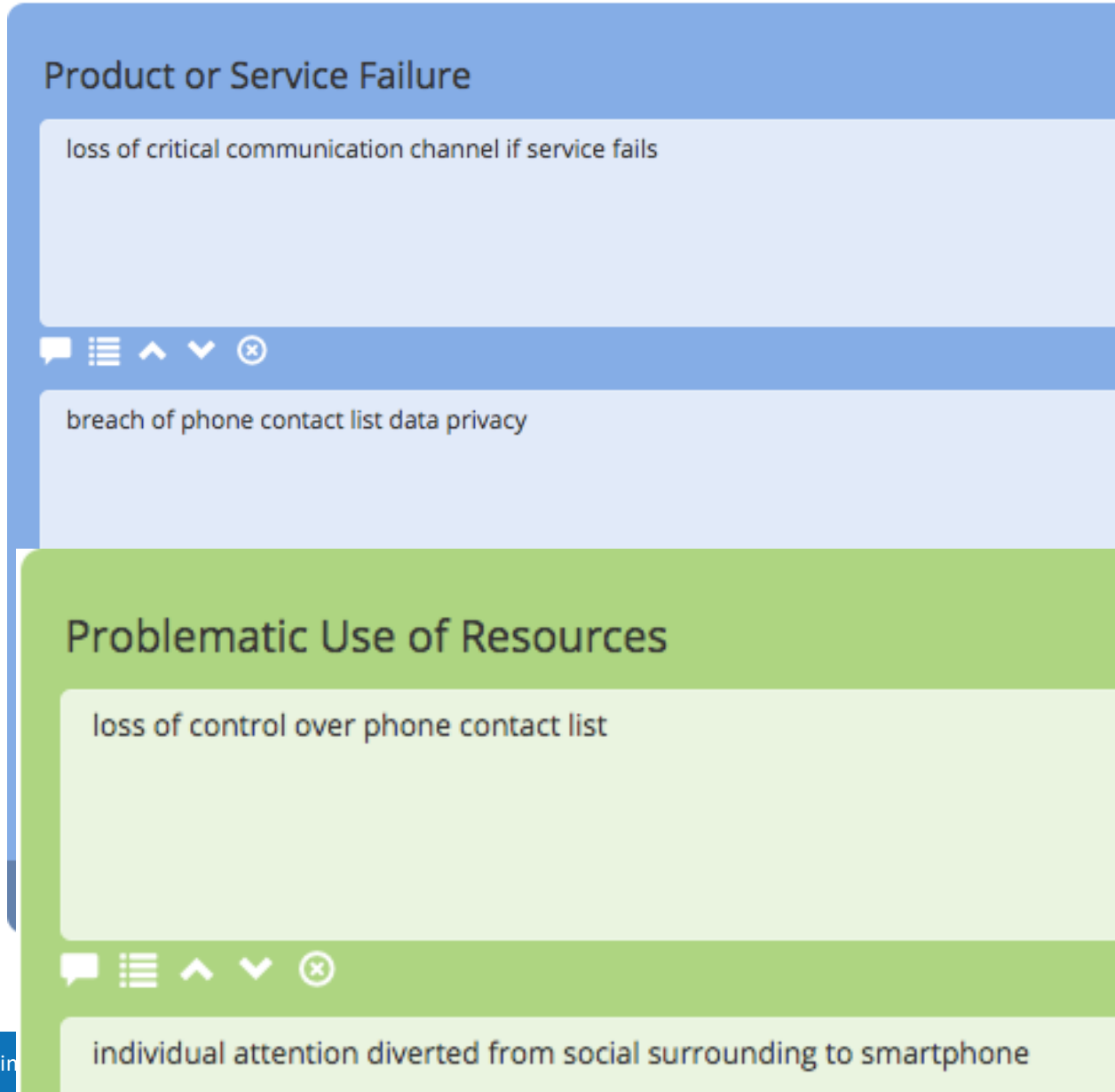


Stage 2: Identifying Ethical Impacts

Aspects that *indirectly* impact our lives..

Potential negative impact of your **product or service failure**? e.g. what happens with technical errors, security failures, etc.?

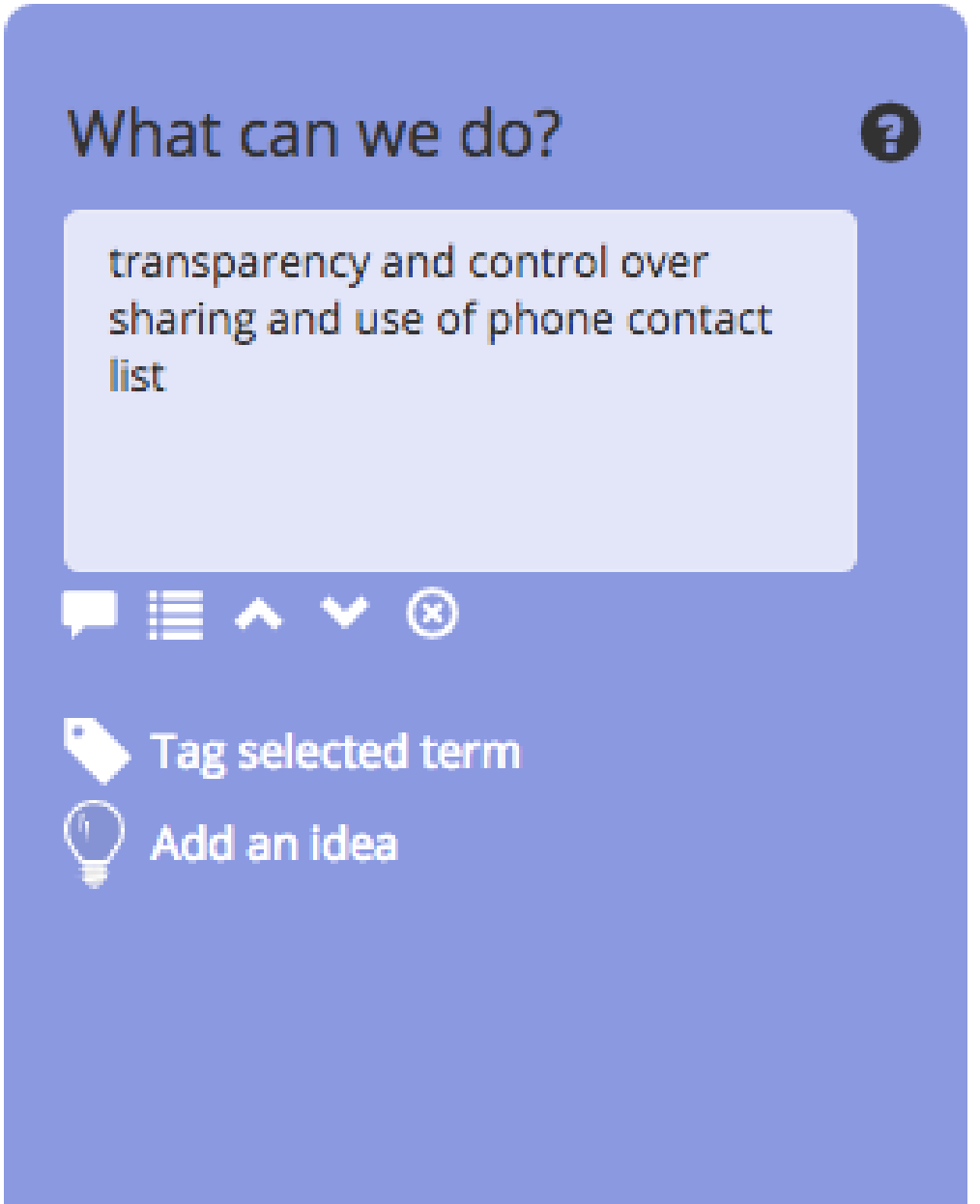
Potential negative impacts of the **consumption of resources** relating to your project? e.g. what happens with its use of energy, personal data,



Stage 3: How to Address Ethical Impacts

What are the most important ethical impacts you found?

How can you address these by pivoting your design, organisation, or by proposing broader changes?



The Ethics Canvas

- Canvas current version: **1.8**

- Web version:

https://ethicscanvas.org

- License: **Creative Commons Attribution Non-Commercial 3.0 Unported**

- User Manual available at:
- https://www.ethicscanvas.org/download/handbook.pdf

Conclusions

- As tech becomes more powerful and ubiquitous, risks of individual and societal impact and harm grows
- Tech Ethics becoming a priority for governments and companies, e.g. for AI, Big Data, Robotics, IoT etc
- Modern innovation techniques feeding AI and Big Data applications need appropriate forms of ethical consideration – agile, accessible
- Ethic Canvas is a simple tool to help innovation teams reflect on ethical issues across application design iterations

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Thank You