



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

# What is the Internet Doing to Me: Ethics on the Internet – AI and Data Governance

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/](https://www.tcd.ie/info_compliance/data-protection/).**

© Trinity College Dublin 2020

# 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)



# Technology Ethics in Context

## Technology Ethics

TRL 1 > TRL 2 > TRL 3 > TRL 4 > TRL 5 > TRL 6 > TRL 7 > TRL 8 > TRL 9

Basic Research   Technology Concept   Applied Research   Laboratory Prototype   Real-world Prototype   Real-world Prototype System   Real-world Product Demo   First Commercial Product   Full-scale Commercial Product

Academia

Industry

## Research Ethics

Technology Ethics concerns how technology impact society – impact from basic research to market deployment

Research Ethics specifically addresses how research can be conducted without harming human participants

GDPR

Data protection applies at all stages of technology R&D

Ethics for publication, e.g.  
Neurips

AI Regulation

AI and Data bring a new initiatives on ICT R&D ethics

## Dominant Views in Technology Ethics

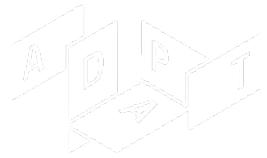
- The neutrality thesis: technologies are *instruments* that we can use to attain our own goals.
  - “People kill people”
- The determinism thesis: technologies *dictate* everything we do, they determine who we are.
  - “Guns kill people”
- The co-shaping thesis: technologies and humans together “construct”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

# Technology Impact: Example



# Software Malfeasance: Example





# Unintended Impacts - Example: Gender in Google Translate

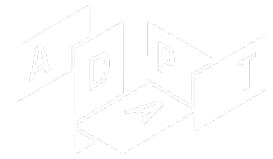
- Some languages, like Turkish, don't have gender specific pronouns
- Google translate has to guess the gender when translating in English
- Statements allocating gender to role reveal gender bias
- What is the source of this?
- Is it a problem?

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

Sample Google Translate output:

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

# Power of Big Data: Example: Cambridge Analytica



- Academic research into Psychographics (U. Cambridge) revealed the link between psychological 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....

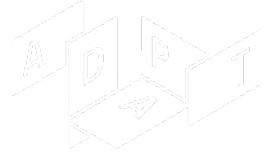
<https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-facebook-nix-bannon-trump>



Facebook, Inc. Common Stock  
NASDAQ: FB - Mar 28, 6:15 AM EDT

152.22 USD **↓7.84 (4.90%)** Facebook's share price peak  
After-hours: 151.38 ↑0.55%

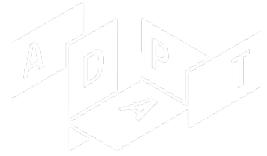




# Algorithmic Power on Behaviour & Worldview

- “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%2099s-Reverse-That-Trend-Now-1.pdf>





# Big Data and AI on the Internet

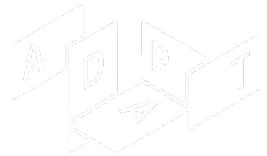
**Big Data** are extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

- Examples: location traces; social media posts/likes/comments; digital content in the form of text, audio, video; geospatial data; sensor data

**Artificial Intelligence (AI)** is a family of computational techniques that aim to mimic human capabilities such as learning and problem solving

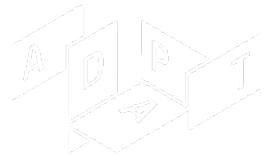
**Machine Learning** is an increasingly successful form of AI using mathematical models trained on Big Data rather than explicit coding instructions

- Example applications: media recommender systems, speech recognition, face recognition, natural language processing, machine translation, search, predictive data analytics



# AI & Data is Mainstreaming Technology Ethics

- Companies harvest and utilise personal data on a **massive scale**
- Growing concerns about the **collection, linking, use and leakage** of personal data from **mobile devices, bio-sensors, cameras, GPS trackers and social media.**
- Machine Learning deliver new levels of **insights and predictions** about an individual's behaviour and also feeds increasingly **personalised AI-driven interactive digital experiences - Ads to Alexa**
- Individuals and groups **struggle to understand** the impact of personal information processing
- Companies, especially SMEs, often lack the knowledge and expertise needed to address these **complex legal and ethical issues.**
- .

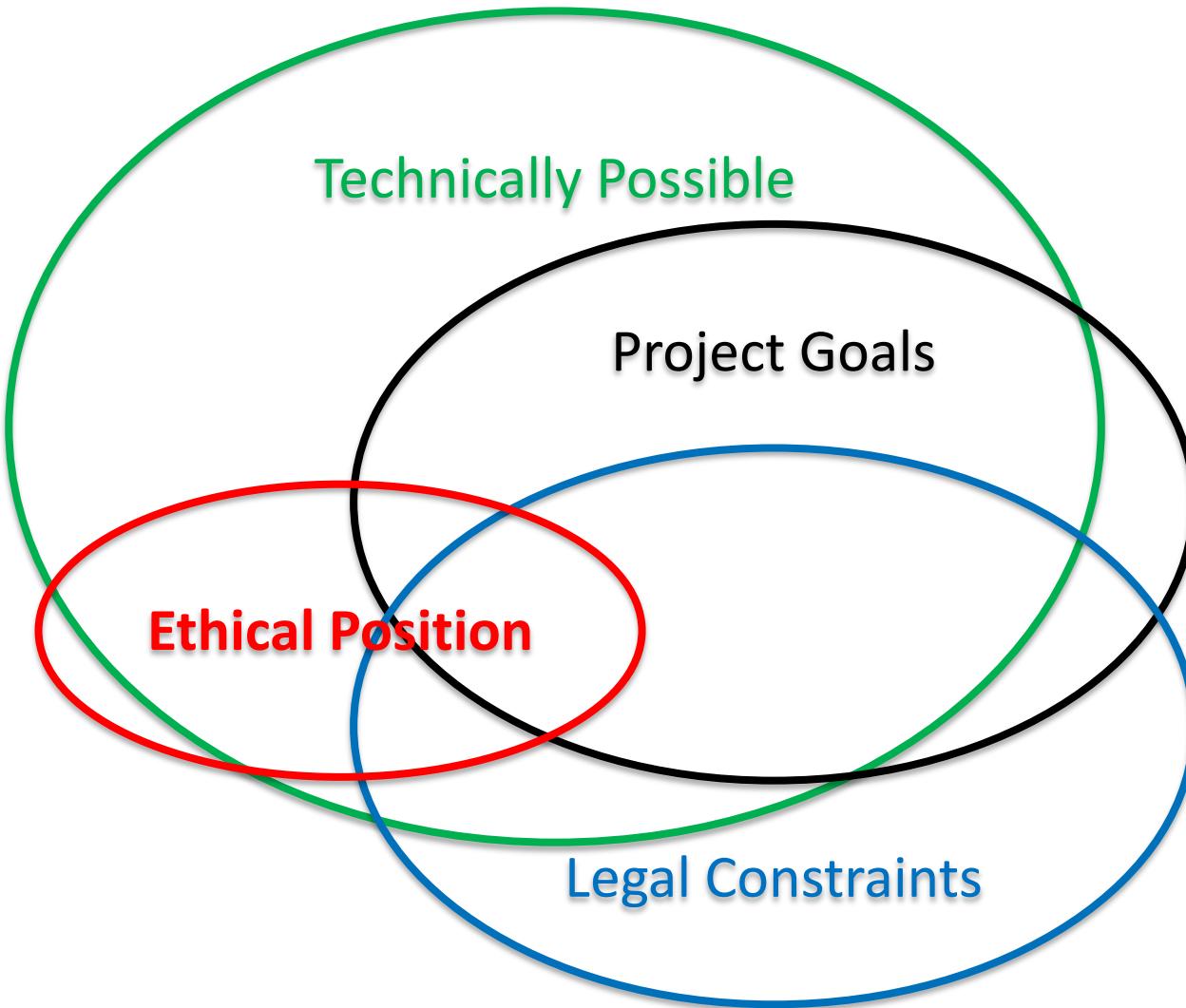


# Example of Risks: Algorithmic selection of digital content in the Internet

- Manipulation of individuals or groups,
- Diminishing variety that creates biased views and distortion of reality,
- Constraints on communication and freedom of expression,
- Threats to privacy and data protection rights,
- Social discrimination,
- Violation of intellectual property rights,
- Impact on the human brain and cognitive capacity and
- Algorithmic power over human behavior and development.

Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. *Handbook on the Economics of the Internet*, (October 2014), pp 395–425. Retrieved from <https://doi.org/10.4337/9780857939852.00028>

# Ethics in a Technology Development Project



# Group Assignment – Ethics assessment

**30% of module mark**

**5000 word report by 29<sup>th</sup> Nov**

**Randomly assigned to groups of 5**

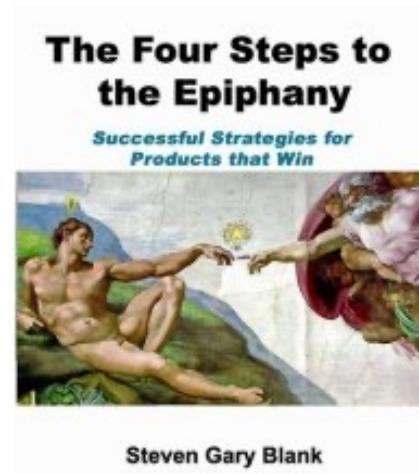
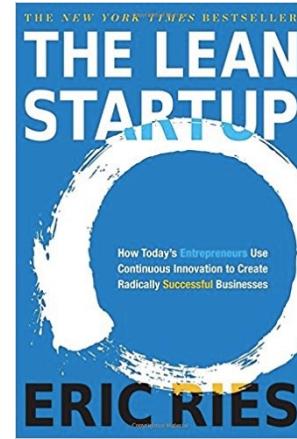
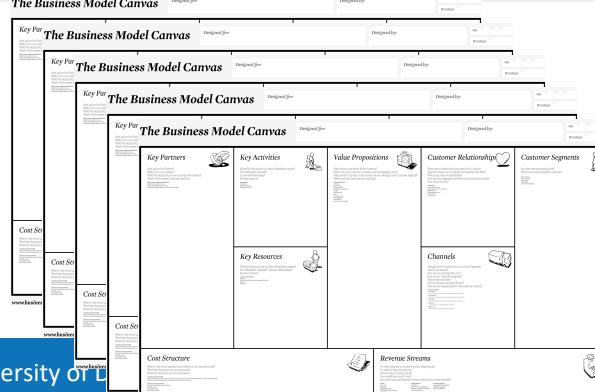
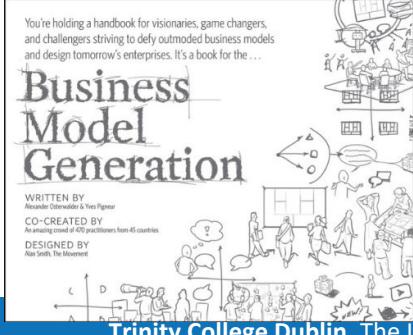
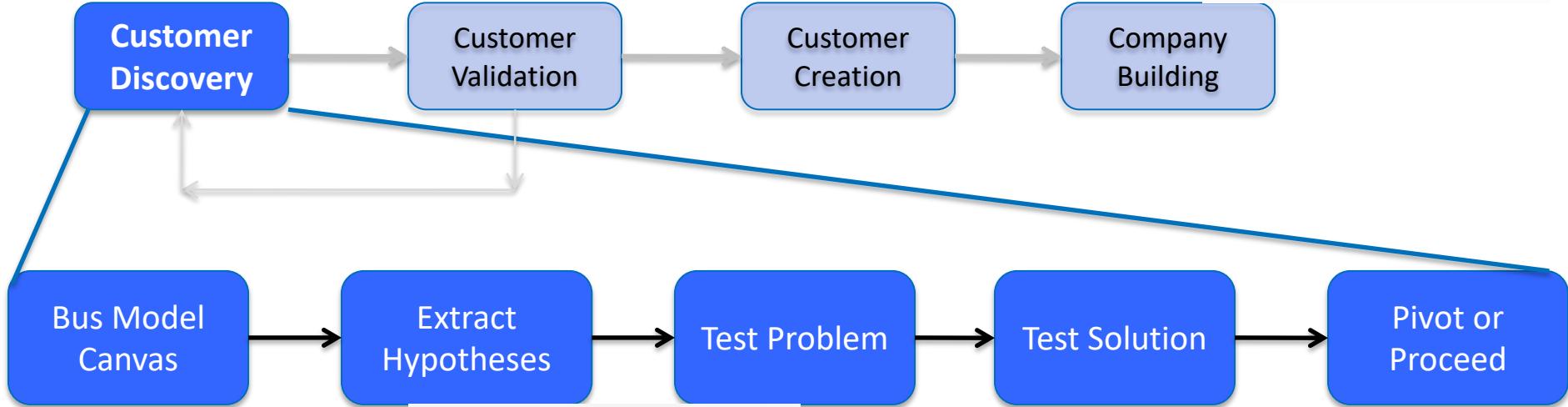
- Select and describe an internet application that makes use of data**
- Discuss ethical issues using ethic canvas catagories (next lecture)**
- Discuss where GDPR may help**
- Summarise how you broke down the work in your group**

**Possible applications to consider could include:**

- Wearable fitness trackers, e.g. Fitbit especially after being acquired by Google**
- Smart speakers, e.g. Amazon Alexa**
- Networked children's toys with voice interactivity, e.g. Hello Barbie**
- Content recommendation, e.g. YouTube**
- Online teaching platform with automated supports/suggestions**
- Content autocomplete of personal email apps**
- Driving performance tracker apps used by insurance companies**
- Home sensors, e.g. amazon Ring or NEST (especially now acquired by Google)**
- Genealogy applications that use DNA samples**
- Dating apps**
- Marketplace/broker apps, e.g. AirBnB, Uber**
- Personalised web search**
- Messenger apps**
- Video conferencing, e.g. Zoom**

# Data Hungry Innovation - "Silicon Valley" Methods

## The Customer Development Process



How to make ethics part of the process?

# 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
- Collaborative brainstorming tool with two aims:
  - Help innovators identify, discuss and articulate possible ethical impacts
  - Bring about *pivots* in the design



<https://ethicscanvas.org>

# Considerations on Ethical Impacts of Technology

- Changes in individual behaviour
- Relationships between individuals
- Relationships between collective actors
- Relationships *between* individuals and collectives
- Impact in the public sphere, on worldviews
- Impact of technology failure
- Impacts on the environment and production processes

# Ethics Canvas

Project Title:

Date:

Ethics Canvas v1.8 - ethicscanvas.org © ADAPT Centre &amp; Trinity College Dublin &amp; Dublin City University, 2017.

Individuals affected  Who use your product or service? Who are affected by its use? Are they men/women, of different ages, etc.?	Behaviour  How might people's behaviour change because of your product or service? Their habits, time-schedules, choice of activities, etc.?   <b>3</b>	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?	Worldviews  How might people's worldviews be affected by your product or service? Their ideas about consumption, religion, work, etc.?   <b>5</b>	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.?   <b>2</b>
Relations  How might relations between people and groups change because of your product or service? Between friends, family-members, co-workers, etc.?	 <b>1</b>  <b>4</b> 	 <b>9</b>	Group Conflicts  How might group conflict arise or be affected by your product or service? Could it discriminate between people, put them out of work, etc.?	
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.?		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.?		 <b>7</b>  <b>8</b>

The Ethics Canvas is adapted from Alex Osterwalder's Business Model Canvas. The Business Model Canvas is designed by: Business Model Foundry AG. This work is licensed under the Creative Commons Attribution-Share Alike 3.0 unported license. To view a copy of this license, visit <https://creativecommons.org/licenses/by-sa/3.0/>. To view the original Business Model Canvas, visit <https://strategyzer.com/canvas>.



# 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 its 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 organisation, interest groups, etc.?*



Running example:  
WhatsApp

## Stage 2: Identifying Ethical Impacts

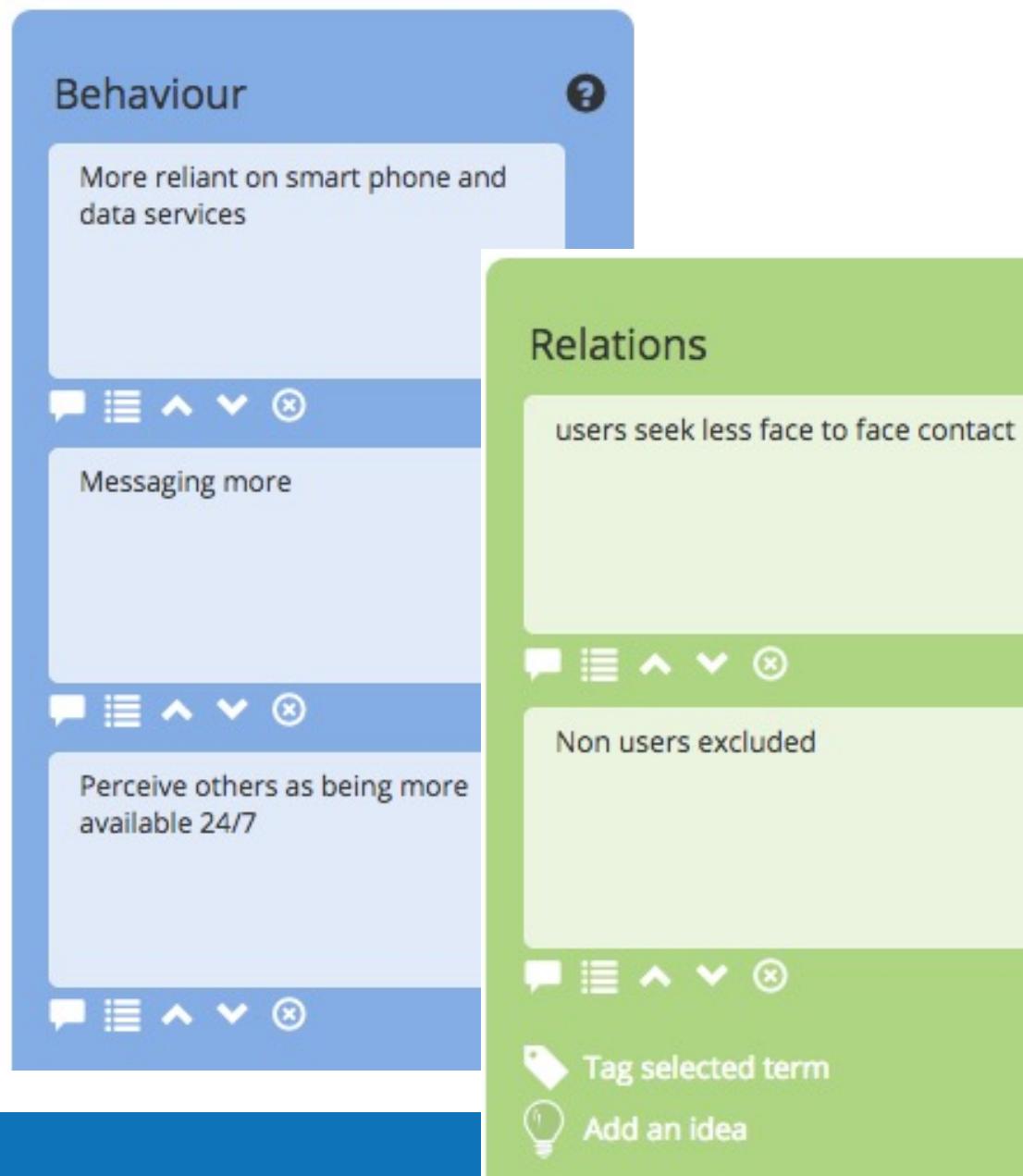
First, ‘**micro**’ impacts are captured by the canvas, i.e. on everyday lives of people using and living with the application

**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.?*

Worldviews

personal phone contacts no longer regarded as private

concerns with loss of location privacy

Group Conflicts

New channel for cyberbullying

conflict between employees and employers messages outside work hours

## 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, etc.?

### Product or Service Failure

loss of critical communication channel if service fails



breach of phone contact list data privacy

### Problematic Use of Resources

loss of control over phone contact list



individual attention diverted from social surrounding to smartphone

## 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?

What can we do?

transparency and control over sharing and use of phone contact list



Tag selected term



Add an idea

Individuals Affected: -	Behaviour: -	What can we do?: -	Worldviews: -	Groups affected: -
Relations: -			Group Conflicts: -	
Product or Service Failure: -		Problematic Use of Resources: -		

<b>Individuals Affected:</b> Consumer of food -	<b>Behaviour:</b> - Less time preparing meals  Easier to live singly/independently  More consumption of ready meals	<b>What can we do?:</b> Find other reasons to eat together as a family  Microwave fresh rather than processing meals  Switch to air fryer	<b>Worldviews:</b> - More individualistic outlooks - Devaluing food preparation and cooking skills	<b>Groups affected:</b> Cooked food vendors – less business  Fresh food vendors: more value in pre-processed food as convenience attractive to consumers
<b>Relations:</b> Less family interaction at meal times -			<b>Group Conflicts:</b> ?	

**Product or Service Failure:**

Only way of warming food

Microwave unit leaks

**Problematic Use of Resources:**

More processed food and packaging

<p><b>Individuals Affected:</b></p> <ul style="list-style-type: none"> <li>- Everyone accessing youtube</li> <li>Children</li> <li>Content posters</li> </ul>	<p><b>Behaviour:</b></p> <ul style="list-style-type: none"> <li>- More screen time due to recommendations</li> <li>Access to violent or disturbing content</li> <li>Access to age inappropriate content</li> <li>Open to false messages/information</li> <li>Open for harmful body images</li> </ul>	<p><b>What can we do?:</b></p> <ul style="list-style-type: none"> <li>- Green energy for data centres and networks</li> <li>- Screen time reporting and rationing</li> <li>- Better screen on inappropriate content</li> </ul>	<p><b>Worldviews:</b></p> <ul style="list-style-type: none"> <li>- Increase in belief in conspiracy theories</li> <li>- increase in extremist and polarized views</li> </ul>	<p><b>Groups affected:</b></p> <ul style="list-style-type: none"> <li>- News providers</li> <li>- Advertisers</li> <li>- Content providers</li> <li>- youTubers</li> <li>- Content moderators</li> </ul>
	<p><b>Relations:</b></p> <ul style="list-style-type: none"> <li>- Less consuming video as a group</li> <li>Less consuming same video as social contacts, less common experience to share</li> </ul>		<p><b>Group Conflicts:</b></p> <ul style="list-style-type: none"> <li>- fakenews and distortion of facts impact civic and democratic processes</li> <li>Employer harms on content moderators</li> <li>Displacement of local news sources</li> </ul>	

**Product or Service Failure:**

- Loss of advertising opportunities
- Loss of video for promoting services or providing information, e.g. how-tos

**Problematic Use of Resources:**

- Data center power consumption

# 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>