



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

What is the Internet Doing to Me: Ethics on the Internet - 1 Technology Ethics, Data and AI

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/.

© Trinity College Dublin 2020

Why Should Digital Tech Innovators be Concerned with Ethics?

- Because new digital 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)



- Fourth:
Because the
Law requires it

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

GDPR

Ethics for publication, e.g.
Neurips

AI Regulation

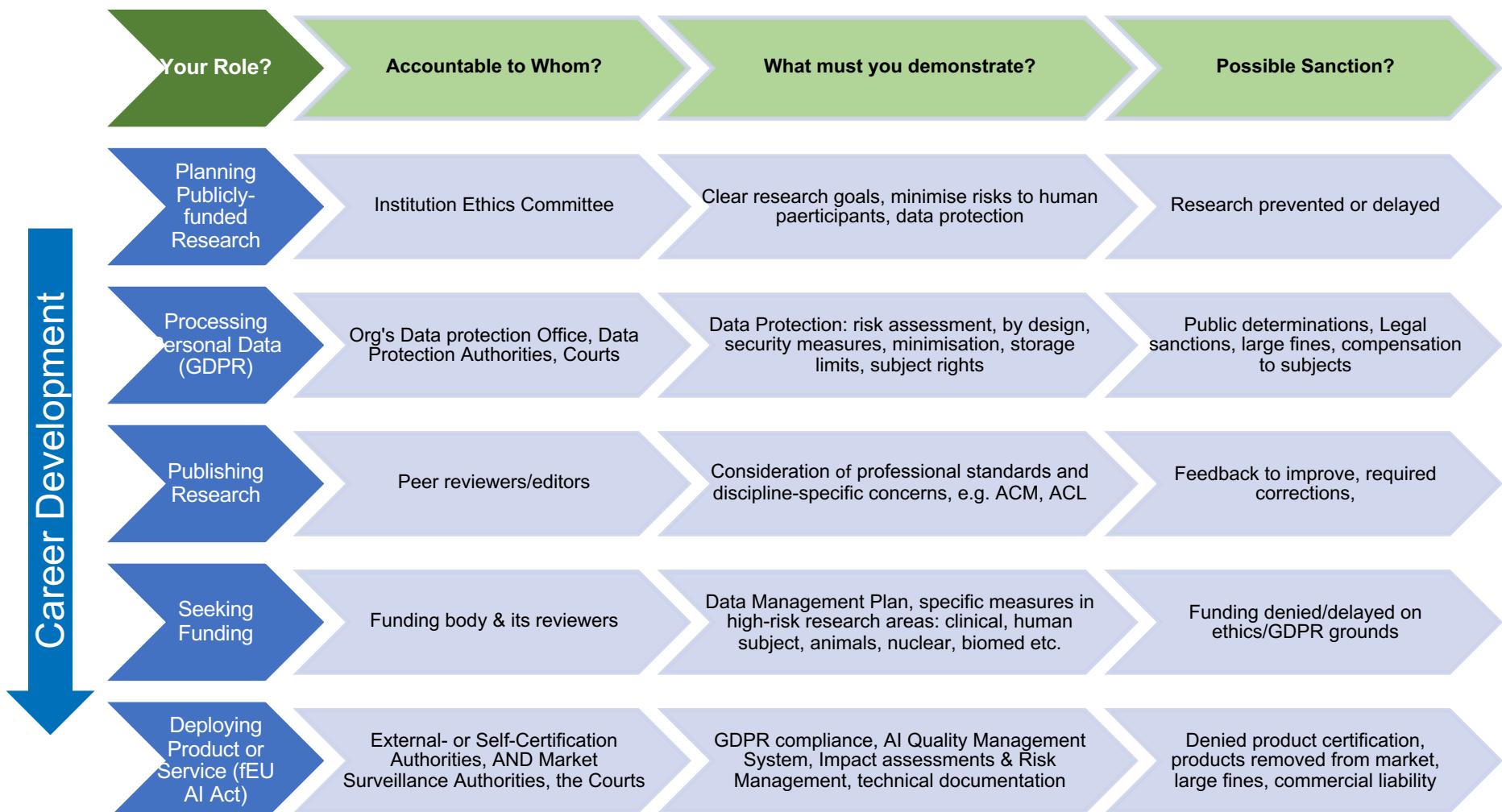
Technology Ethics concerns how technology impacts society – risks can emerge (and be mitigated) from basic research to market deployment

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

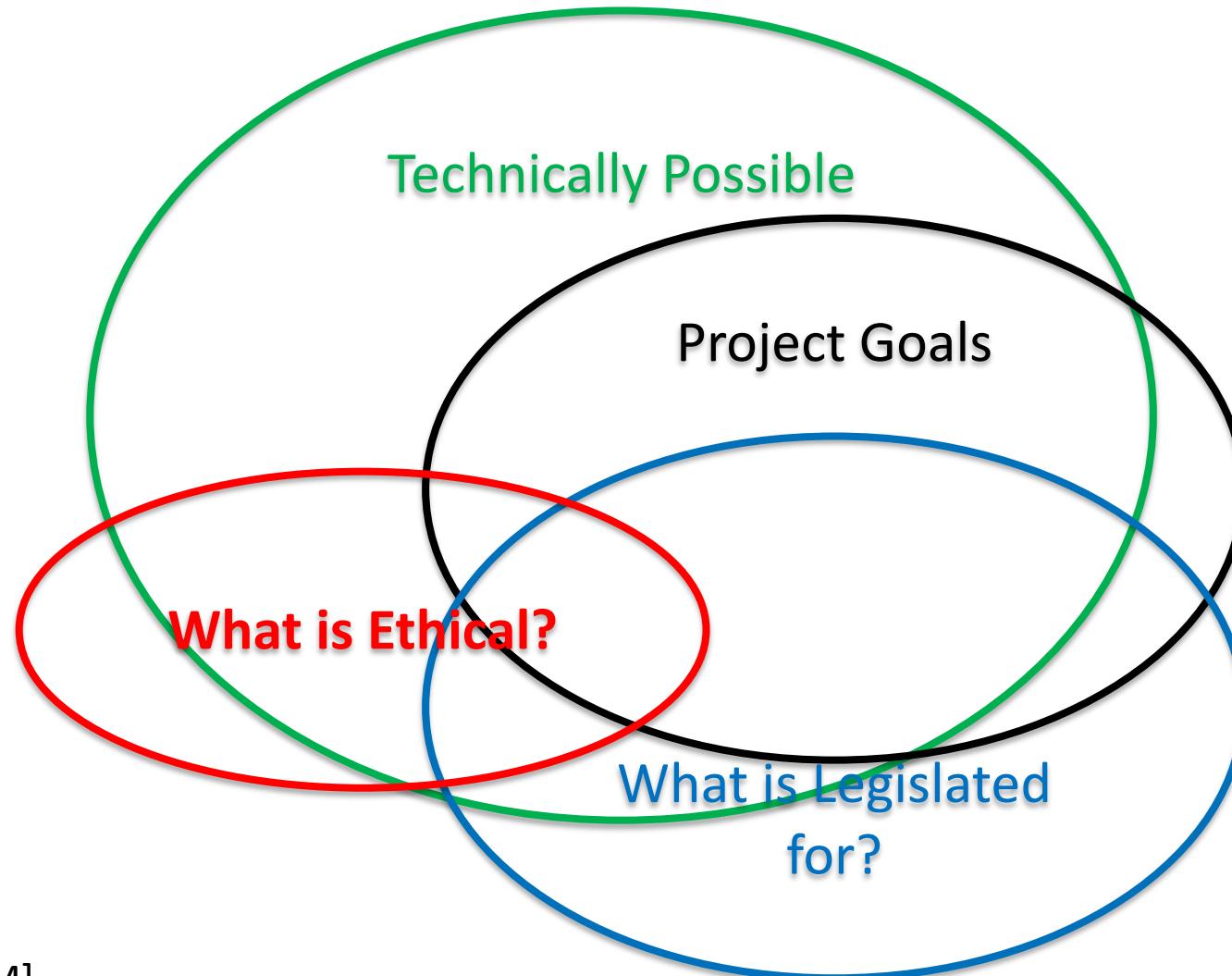
GDPR applies at all stages of technology R&D if using personal data

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

AI Research and Innovation in the EU: Who am I accountable to? What should I do? What verdicts can I be subject to?



Ethics in a Technology Development Project



[IBM]

Dominant Views in Technology Ethics

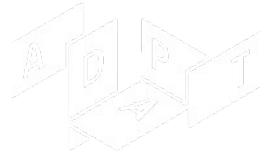
- The neutrality thesis: technologies are *instruments* that we can use to attain our own goals.
 - “People kill people”
- The determinism thesis: technologies *dictate* everything we do, they determine who we are.
 - “Guns kill people”
- The co-shaping thesis: technologies and humans together “construct”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

Technology Impact: Example



Software Malfeasance: Example





Unintended Impacts - Example: Gender in Google Translate

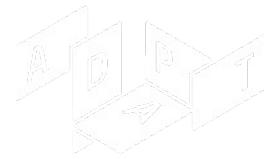
- Some languages, like Turkish, don't have gender specific pronouns
- Google translate has to guess the gender when translating in English
- Statements allocating gender to role reveal gender bias
- What is the source of this?
- Is it a problem?

<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....

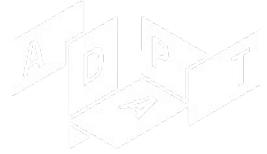
<https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-facebook-nix-bannon-trump>



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%**

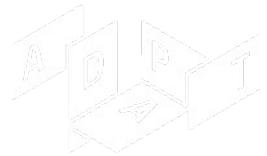




Algorithmic Power on Behaviour & Worldview

- “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's-Reverse-That-Trend-Now-1.pdf>





Big Data and AI on the Internet

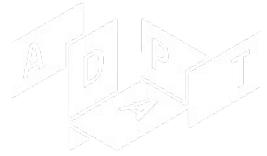
Big Data are extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

- 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 also feeds increasingly **personalised AI-driven interactive digital experiences –Digital Engagement From 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.**

2022 - Step Change in AI Capabilities

Language Models

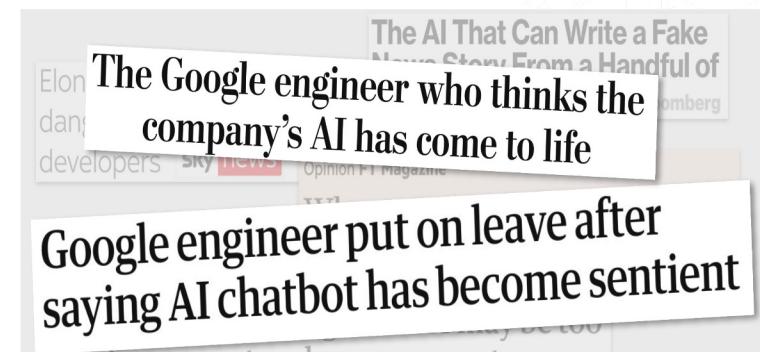
- Machine learning model that underpin Natural Language Processing tasks
- translation, question-answering, speech recognition, summarization, entity recognition etc

Large Language Models (LLM)

- Trained on vast content data sets crawled from the Web
- Surprised that LLM excel at a wide range of tasks

Foundational LLMs

- Models that can be easily adapted to new tasks
- Prompt Engineering, Reinforcement Learning from Human Feedback, Model Fine Tuning



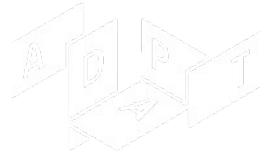
How ChatGPT Managed to Grow Faster Than TikTok or Instagram

HOME > ECONOMY

ChatGPT may be coming for our jobs. Here are the 10 roles that AI is most likely to replace.

Aaron Mok and Jacob Zinkula Updated Jun 4, 2023, 3:09 PM GMT+1



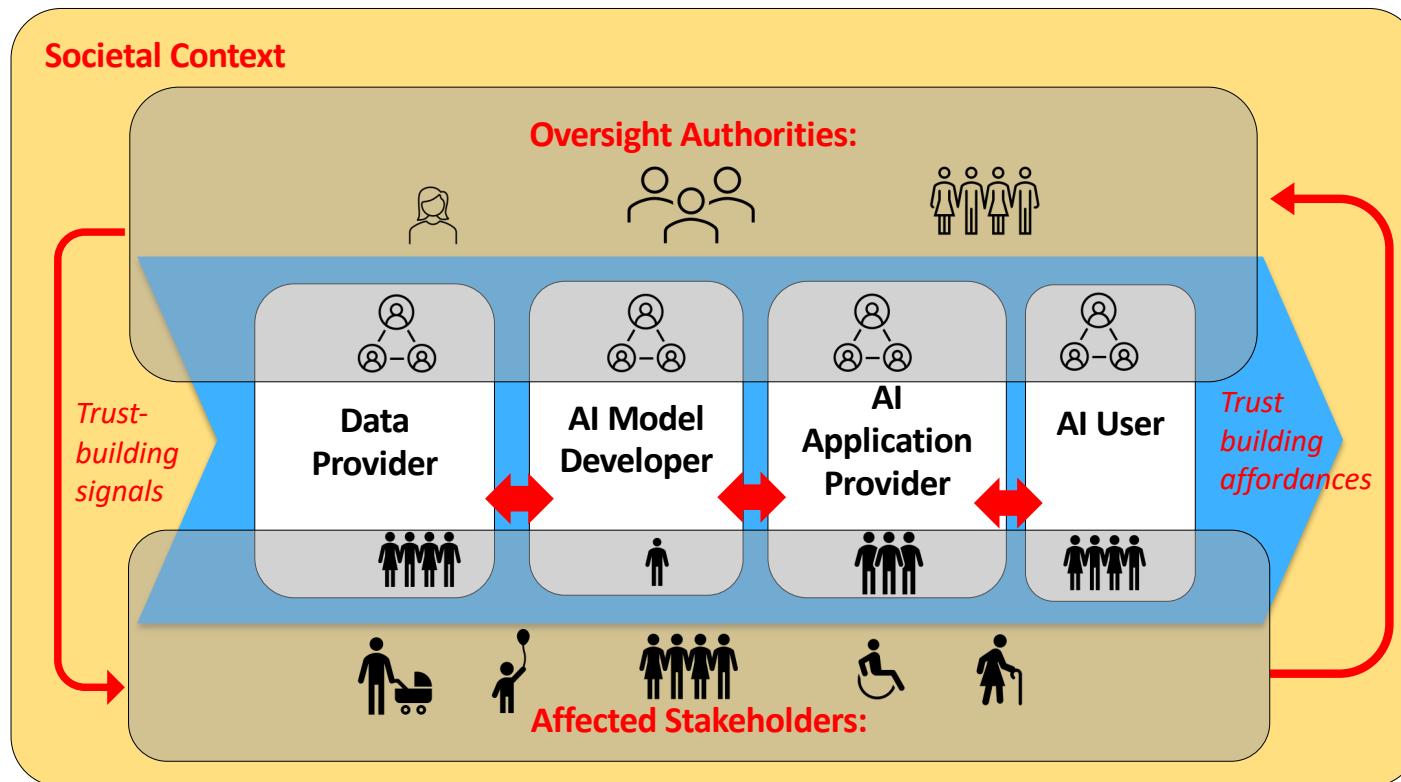


Risks: Algorithmic selection of digital content on the Internet

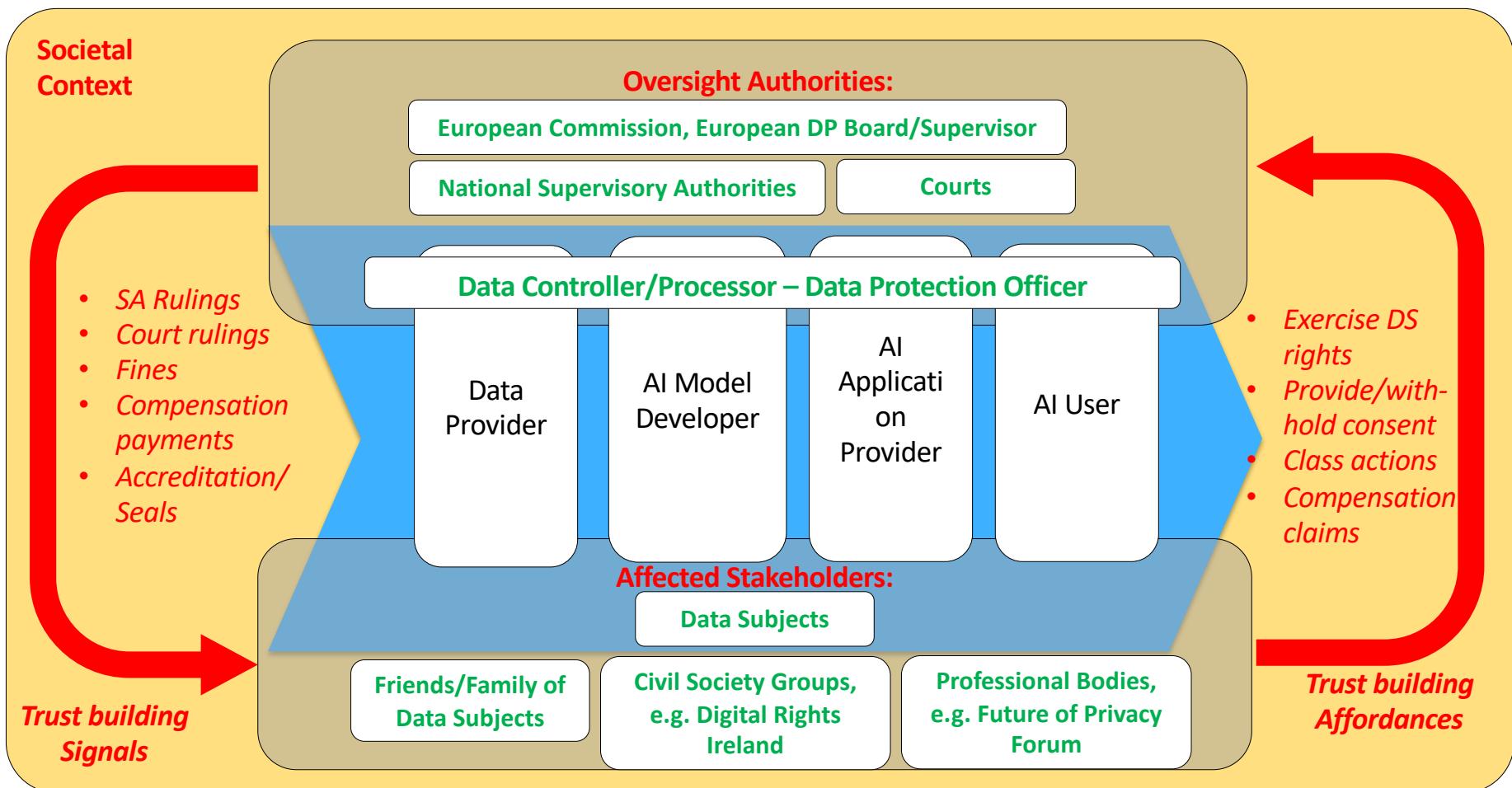
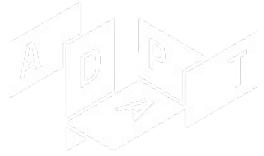
- Manipulation of individuals or groups,
- Diminishing variety that creates biased views and distortion of reality,
- Constraints on communication and freedom of expression,
- Threats to privacy and data protection rights,
- Social discrimination,
- Violation of intellectual property rights,
- Impact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

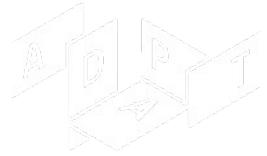
Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. *Handbook on the Economics of the Internet*, (October 2014), pp 395–425. Retrieved from <https://doi.org/10.4337/9780857939852.00028>

Trustworthy AI and Data Governance: Systems of Co Regulation of AI/Data based Digital Technology



The Scope and Role of GDPR on Trustworthy AI/Data governance





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

Example: Gender Bias in Machine Translation (MT)



Oversight Authorities:

Language/Technology Researchers (highlight bias);
Professional Bodies for Translators (advise on translation ethics);

Translation Clients (perform translation QA);

Data Provider:

Translation DBs;
Translation Clients;
Translators;
Web Content writers/
publishers

AI Creator:

MT software providers (e.g. Google MT, Iconic Translation Machines)

AI Operator:

Language Service Providers (e.g. Lionbridge, EU translation service);
Browser vendors (e.g. Google)

AI User:

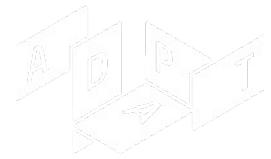
Translators/
posteditors;
Translation clients;

Reader of translated content (inaccurate content);

Groups misrepresented by translated content (experience further bias);
Writer of translated content (author's moral rights)

Affected Stakeholders:

Power of Big Data: Example: Cambridge Analytica



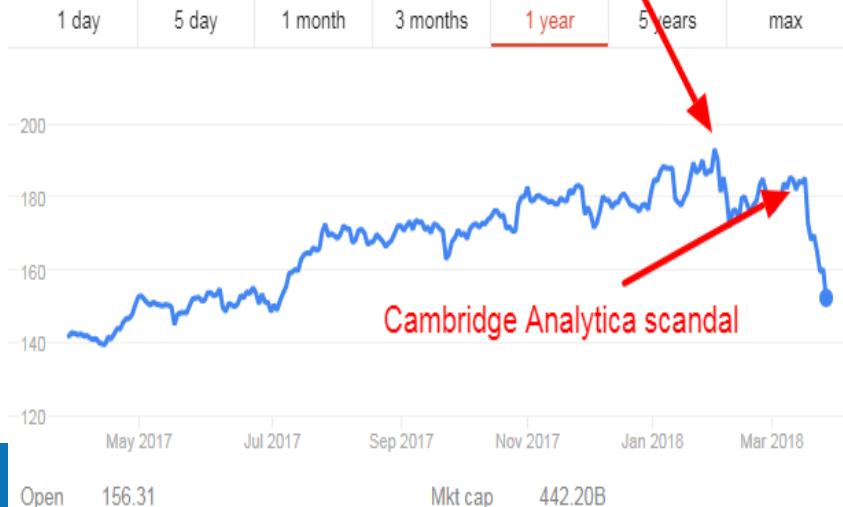
- Academic research into Psychographics (U. Cambridge) revealed the link between psychological profiles and Facebook profiles
- Correlated major psychological types to elements in the social graph: Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism
- Cambridge Analytica applied 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%**





Example: Cambridge Analytica

Oversight Authorities:

Media & Whistleblowers; Advertising regulators;
Data Protection Regulator; Political Campaign Rule regulators/courts; Elected
Representatives

FB and CA engineers, managers, DPOs

Data Provider:
Individuals via
Facebook Social
Graph

AI Creator:
Cambridge
Analytica
targeting engine

AI Operator:
Cambridge
Analytica using
Facebook as a
platform

AI User:
Cambridge
Analytica Clients;
FB users
receiving
personalized
targeted
messages

Targeted FB users (data used without consent, view manipulated);

Election candidates (suffer unfair competition);
Citizens in a democracy (integrity of system of government damaged)

Affected Stakeholders:

Group Assignment – Ethics assessment

- Select a high risk application area where this model could apply (from domains/purposes defined in the AI Act – next week)
- Outline the value chain for this AI application
- Discuss and document ethical issues using ethic canvas categories (next)
- Identify risks and mitigations for AI subjects and other affected stakeholders (guided by ISO 26000)
- Summarise how you broke down the work in your group

25% of module mark

Randomly assigned to groups of 5

By 9/10th Nov: Contact group

By 16th Nov: Choose model and application

By 1st Dec: Submit report



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

Ethics on the Internet – Technology Ethics Assessment – The Ethics Canvas

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.

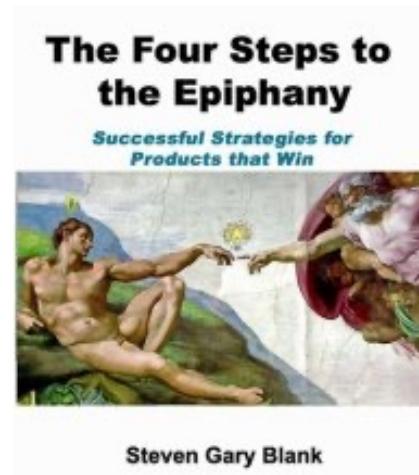
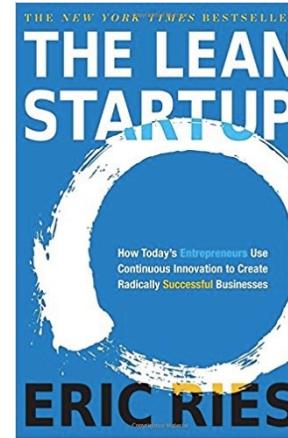
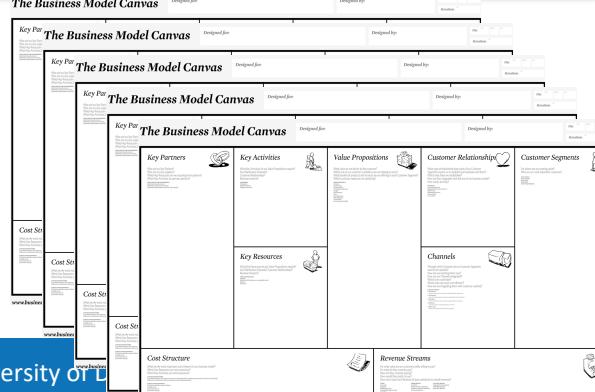
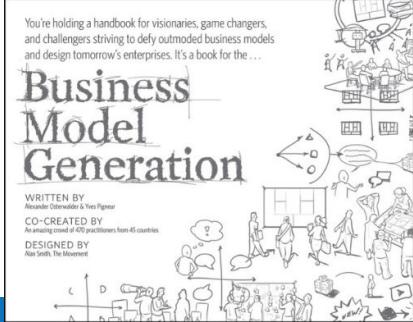
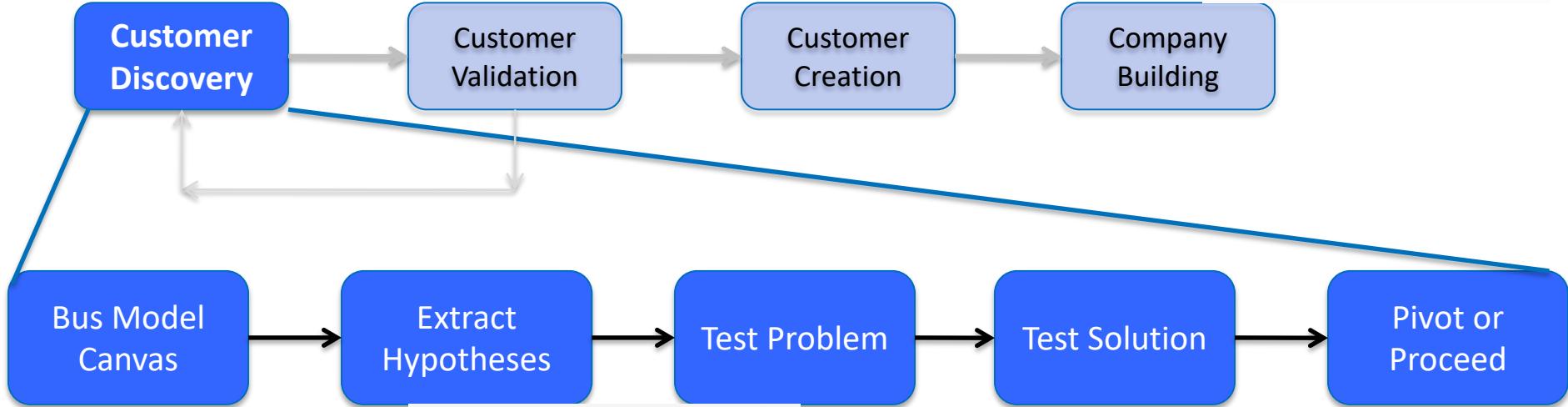
Recorded materials will be handled in compliance with Trinity's statutory duties under the Universities Act, 1997 and in accordance with the University's [policies and procedures](#).

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

© Trinity College Dublin 2020

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



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

- 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

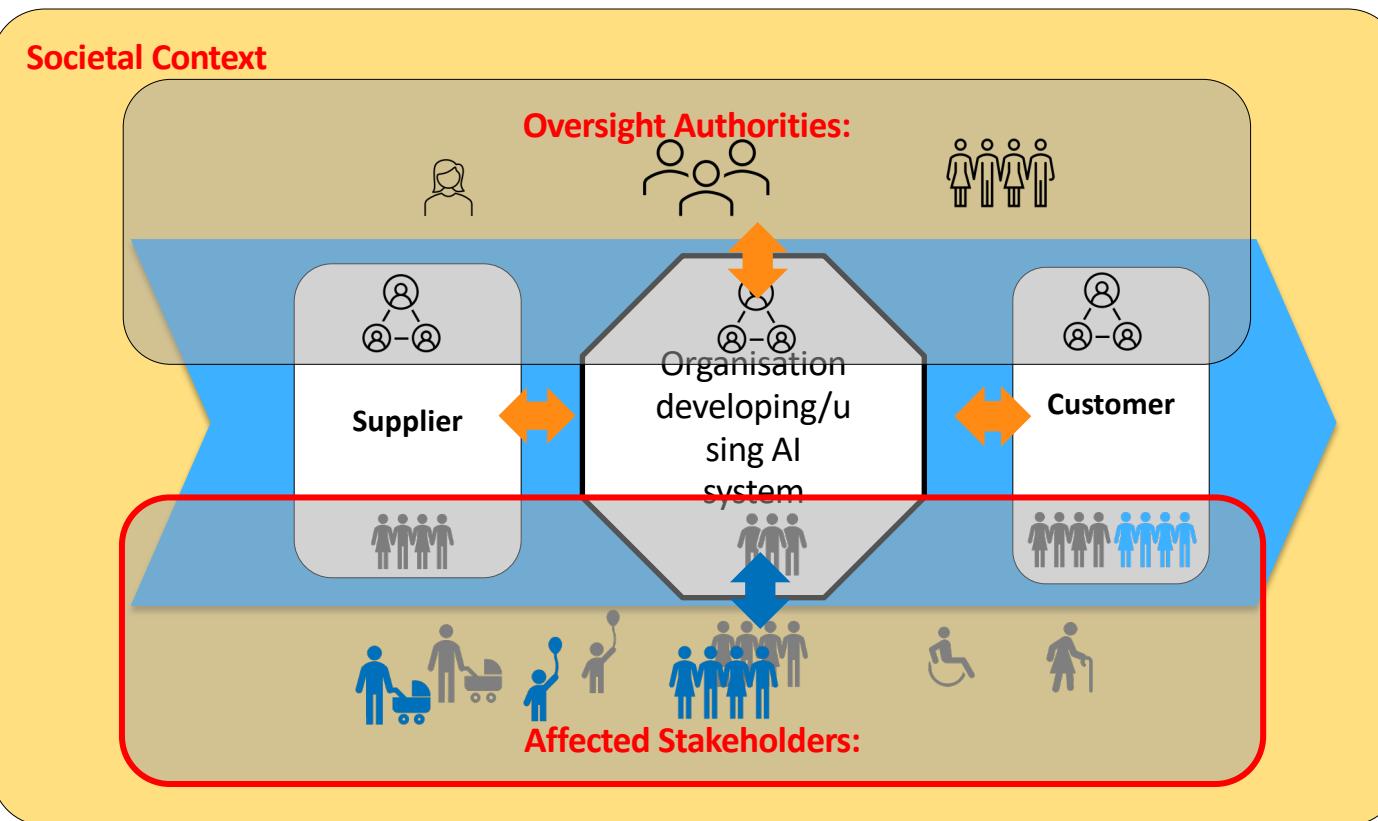


<https://ethicscanvas.org>

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

Identifying Stakeholders in AI/Data Value Chains Social Responsibility Perspective



- Labour Practices (workers)
- The Environment (future generations)
- Fair Operating Procedures (suppliers, customers, regulators)
- Consumer Issues (consumers)
- Community Involvement and Development (local communities)
- Human Rights (everyone)

Based on ISO 26000

Ethics Canvas

Project Title:

Date:

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

Individuals affected	Behaviour	What can we do?	Worldviews	Groups affected
Who use your product or service? Who are affected by its use? Are they men/women, of different ages, etc.?	How might people's behaviour change because of your product or service? Their habits, time-schedules, choice of activities, etc.?	What are the most important ethical impacts you found? How can you address these by changing your design, organisation, or by proposing broader changes?	How might people's worldviews be affected by your product or service? Their ideas about consumption, religion, work, etc.?	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.?
	 3		 5	
1	Relations How might relations between people and groups change because of your product or service? Between friends, family-members, co-workers, etc.?		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.?	
 2	 4	 7	 9	 6
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.?	
		 8		

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**

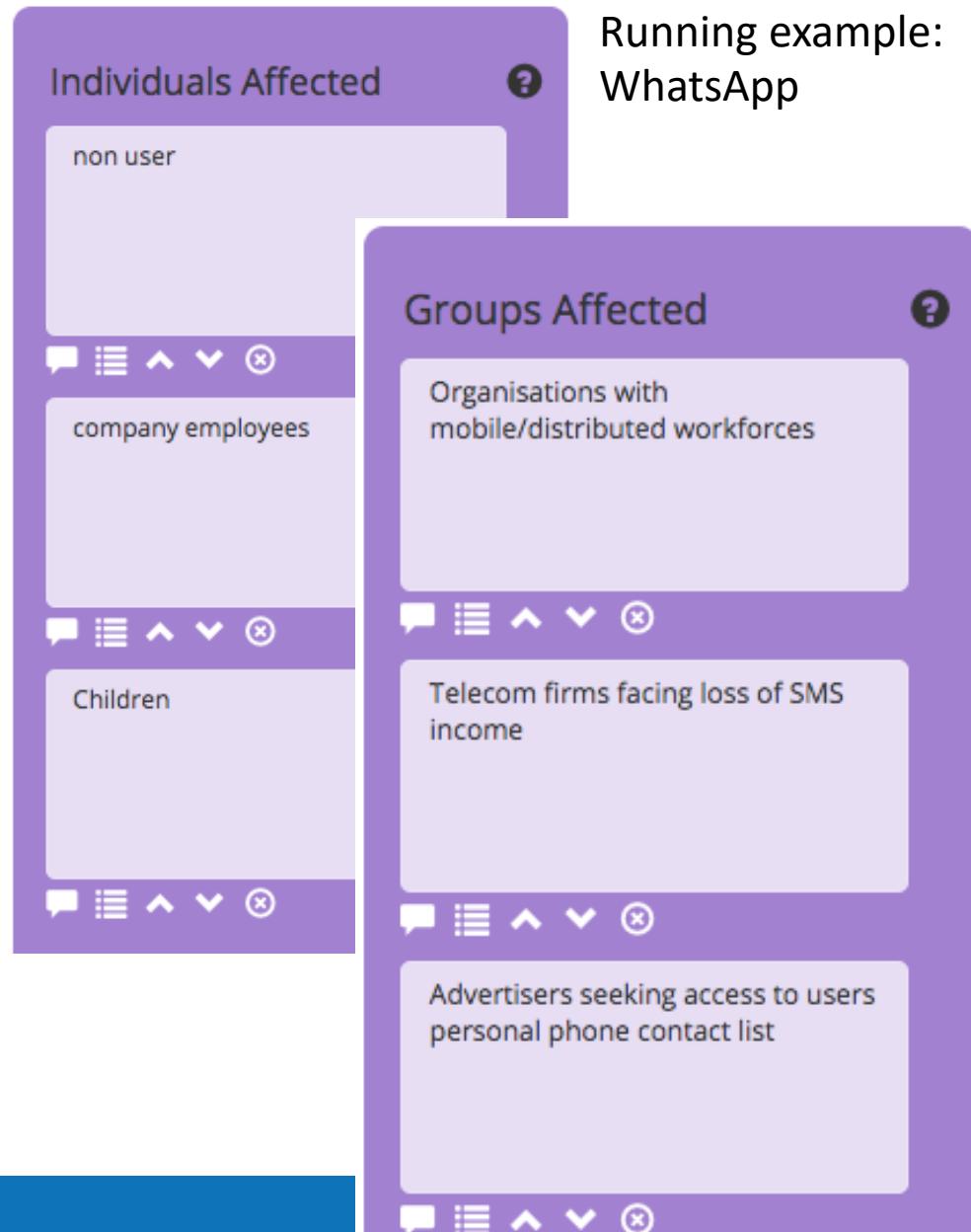
Individuals: Who use your product or service? Who are affected by its use?

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

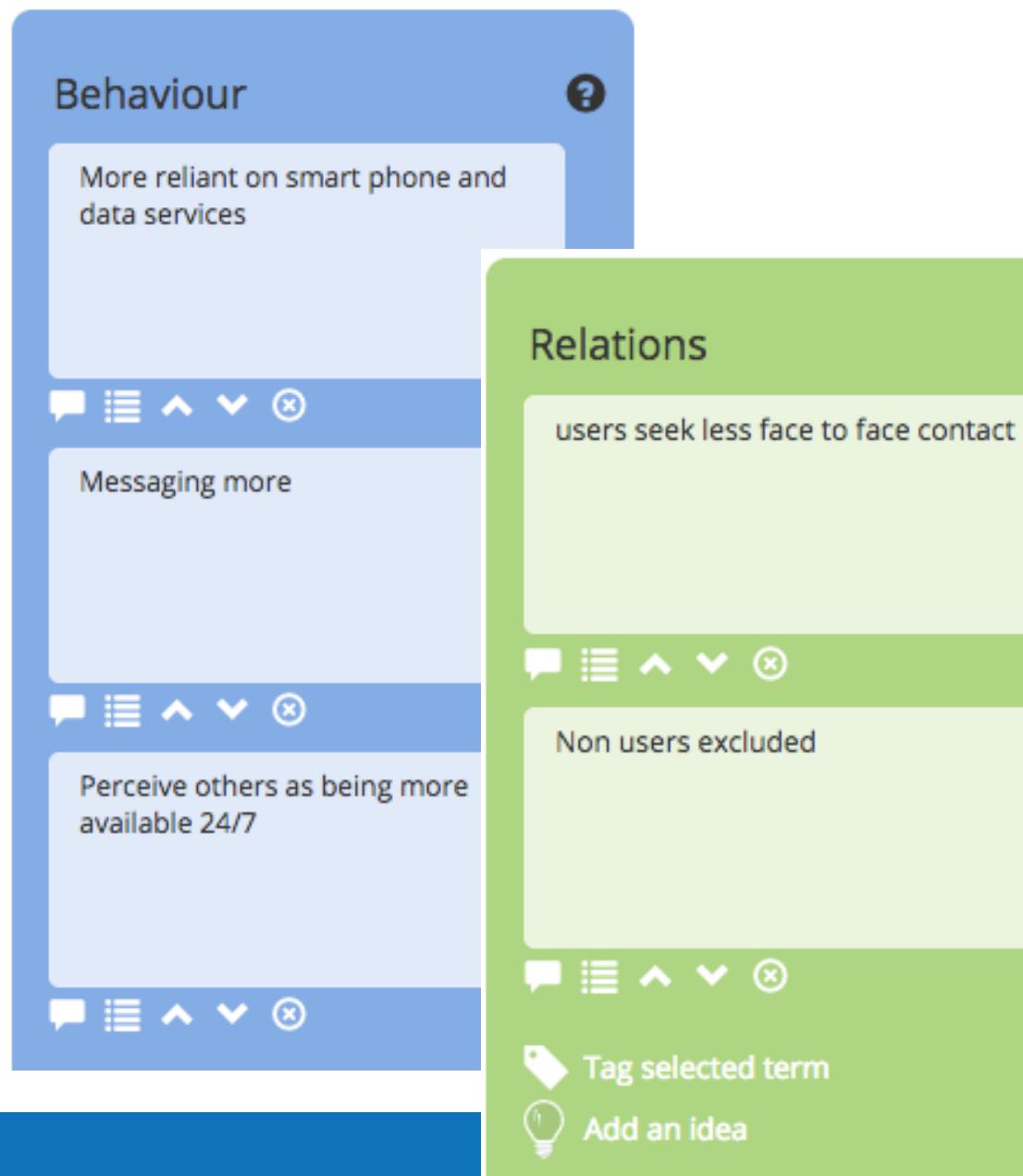
First, ‘**micro**’ impacts are captured by the canvas, i.e. on everyday lives of people using and living with the application

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

Tag selected

Add an idea

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

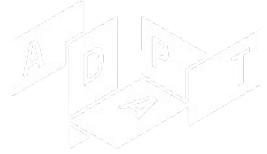
Individuals Affected: -	Behaviour: -	What can we do?: -	Worldviews: -	Groups affected: -
Relations: -			Group Conflicts: -	
Product or Service Failure: -		Problematic Use of Resources: -		

Technology Impact: Example



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

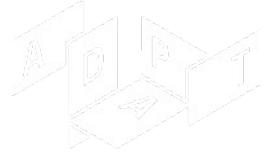
Product or Service Failure: Only way of warming food Microwave unit leaks	Problematic Use of Resources: - More processed food and packaging, with more waste
--	---



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's-Reverse-That-Trend-Now-1.pdf>





Example: YouTube (YT)

Oversight Authorities:

Researchers & Media; Google/Alphabet; Elected Representatives

YT engineers, managers, DPOs

Data Provider:
Individuals via behaviour on YT;
Video creators via meta-data

AI Creator:
Google/YT
Recommender and Search engines

AI Operator:
Google/YT

AI User:
YT Users;
YT Advertisers;
Video creators

Video consumers on YT (loss of variety, loss of balanced view of reality; manipulated);
Video creators (competitive pressure to induce negative emotions and link-bait meta-data);

Citizens (polarization of society & politics, 'othering' of groups)

Affected Stakeholders:

41

Ethics Canvas

Project Title:

Date:

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

Individuals affected	Behaviour	What can we do?	Worldviews	Groups affected
Who use your product or service? Who are affected by its use? Are they men/women, of different ages, etc.?	How might people's behaviour change because of your product or service? Their habits, time-schedules, choice of activities, etc.?	What are the most important ethical impacts you found? How can you address these by changing your design, organisation, or by proposing broader changes?	How might people's worldviews be affected by your product or service? Their ideas about consumption, religion, work, etc.?	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.?
	 3		 5	
1	Relations How might relations between people and groups change because of your product or service? Between friends, family-members, co-workers, etc.?		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.?	
 2	 4	 7	 9	 6
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	
			 8	
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 .				



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>.

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

Individuals Affected: Everyone accessing youtube Children Content posters	Behaviour: More screen time due to recommendations Access to violent or disturbing content Access to age inappropriate content Open to false messages/information Open for harmful body images	What can we do?: - Green energy for data centres and networks - Screen time reporting and rationing - Better screening of inappropriate content	Worldviews: - Increase in belief in conspiracy theories - increase in extremist and polarized views	Groups affected: - News providers - Advertisers - Content providers - YouTubers - Content moderators
	Relations: - Less consuming video as a group Less consuming same video as social contacts, less common experience to share		Group Conflicts: fakenews and distortion of facts impact civic and democratic processes Employer harms on content moderators Displacement of local news sources	

Product or Service Failure:

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

Problematic Use of Resources:

- Data center power consumption

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

<https://ethicscanvas.org>

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

