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

Data Ethics in Technology Innovation

Presented by Gaye Stephens

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

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**.

- Ethics concerns 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)

Virtue Ethics concerns cultivating ethical behaviours in individuals.

Deontological ethics concerns rules, legislation, codes of practice.

Utilitarian ethics concerns making decisions based on what will benefit the majority

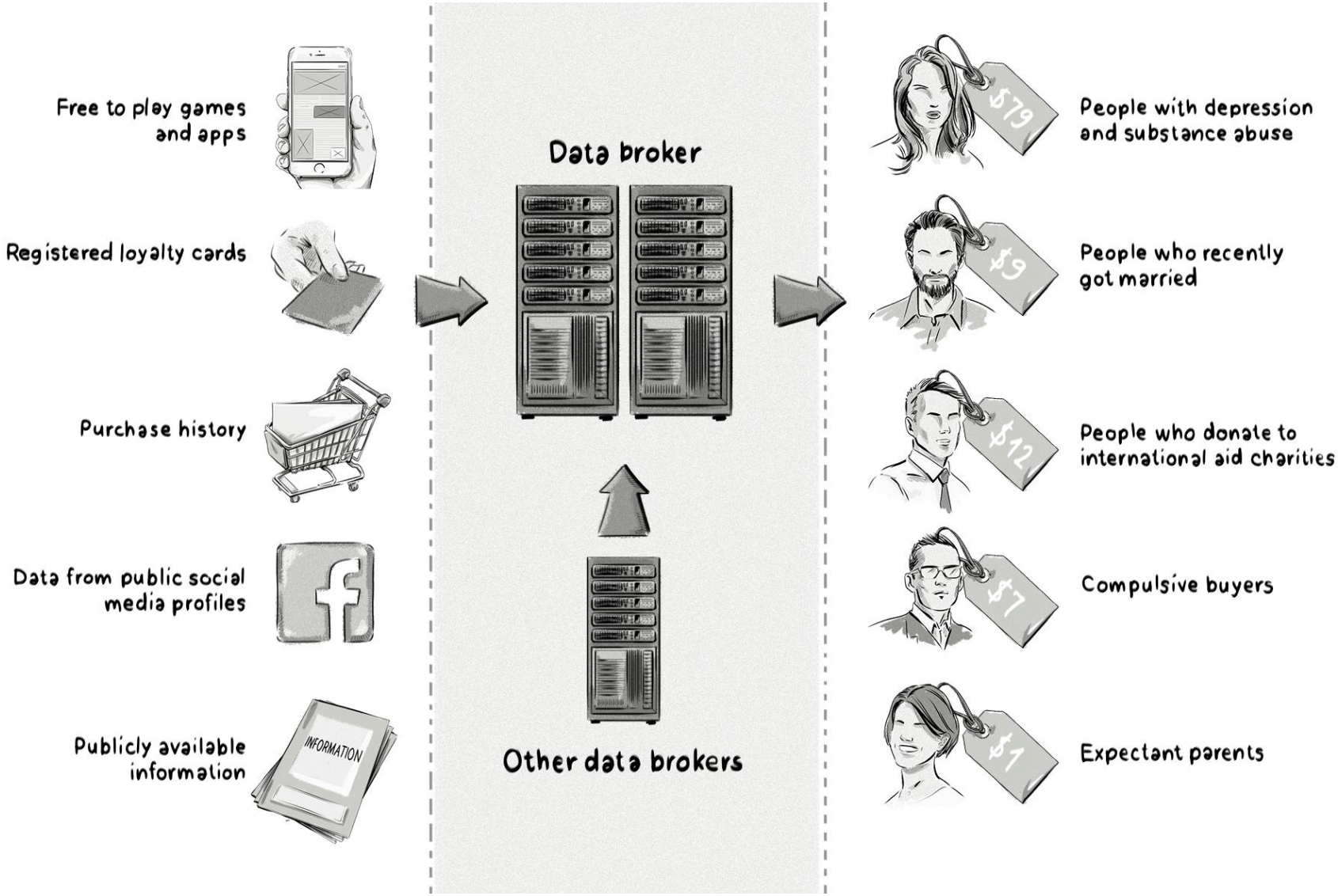
Dominant Views in Technology Ethics

- 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"

Datafication

- Collection and Processing of Data is massive source of value for digital platforms
- Digital engagement drives the Attention Economy
    - Monetizing eye-hours for advertising
- Network Effects tend to market dominance
    - Metclalf's Law: value of a network with n nodes is n^2
- Power of AI-driven digital engagement grows with the volume (and quality) of its training data
- Value beyond ads drives data sharing and data linking
    - Google+Nest+Fitbit
    - Facebook+Whatsapp
    - Apple+itunes+iwatch+ipay
    - Amazon+alexa+primetv

Data Brokerage Industry: Profiling and Targetting



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….

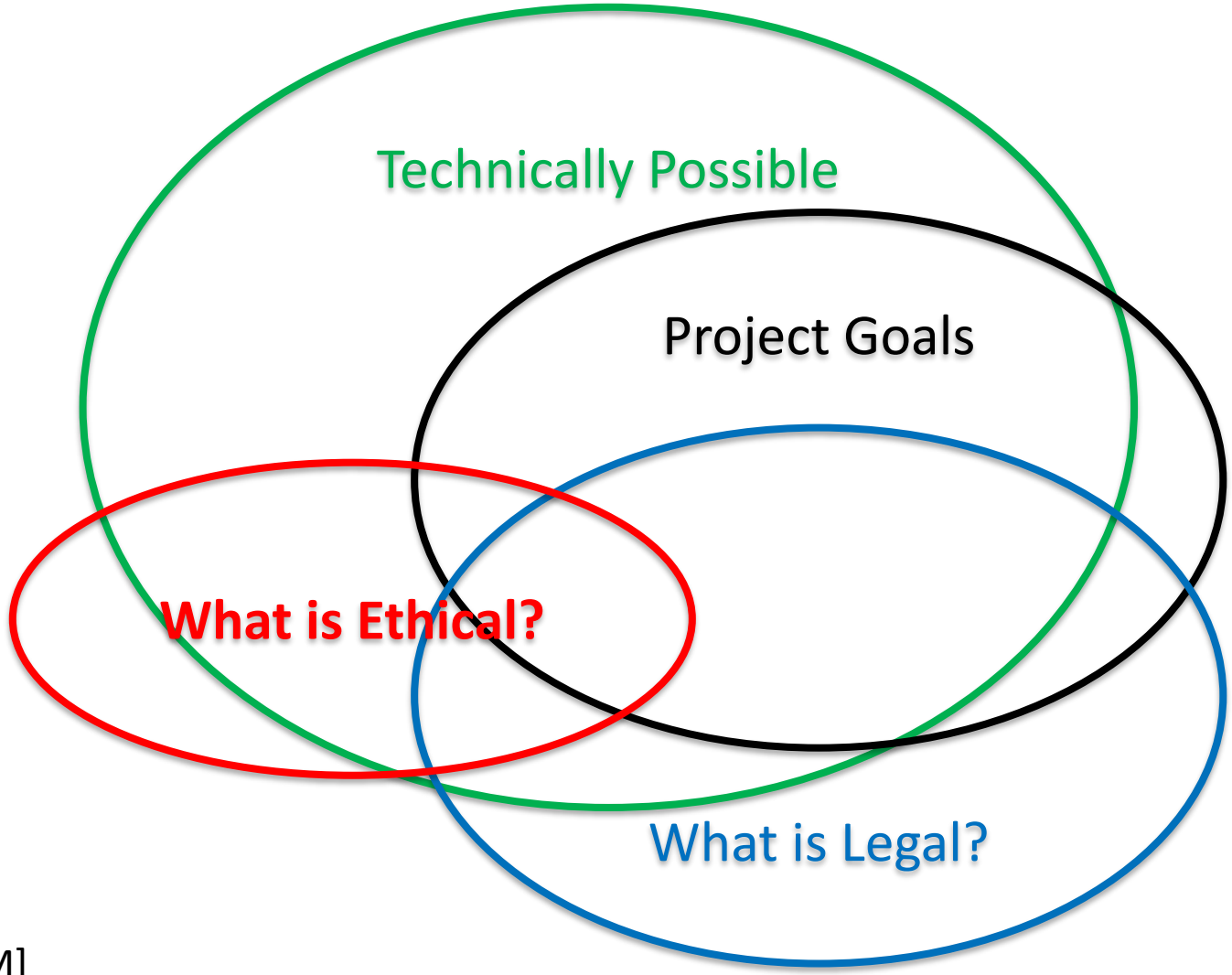
https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

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

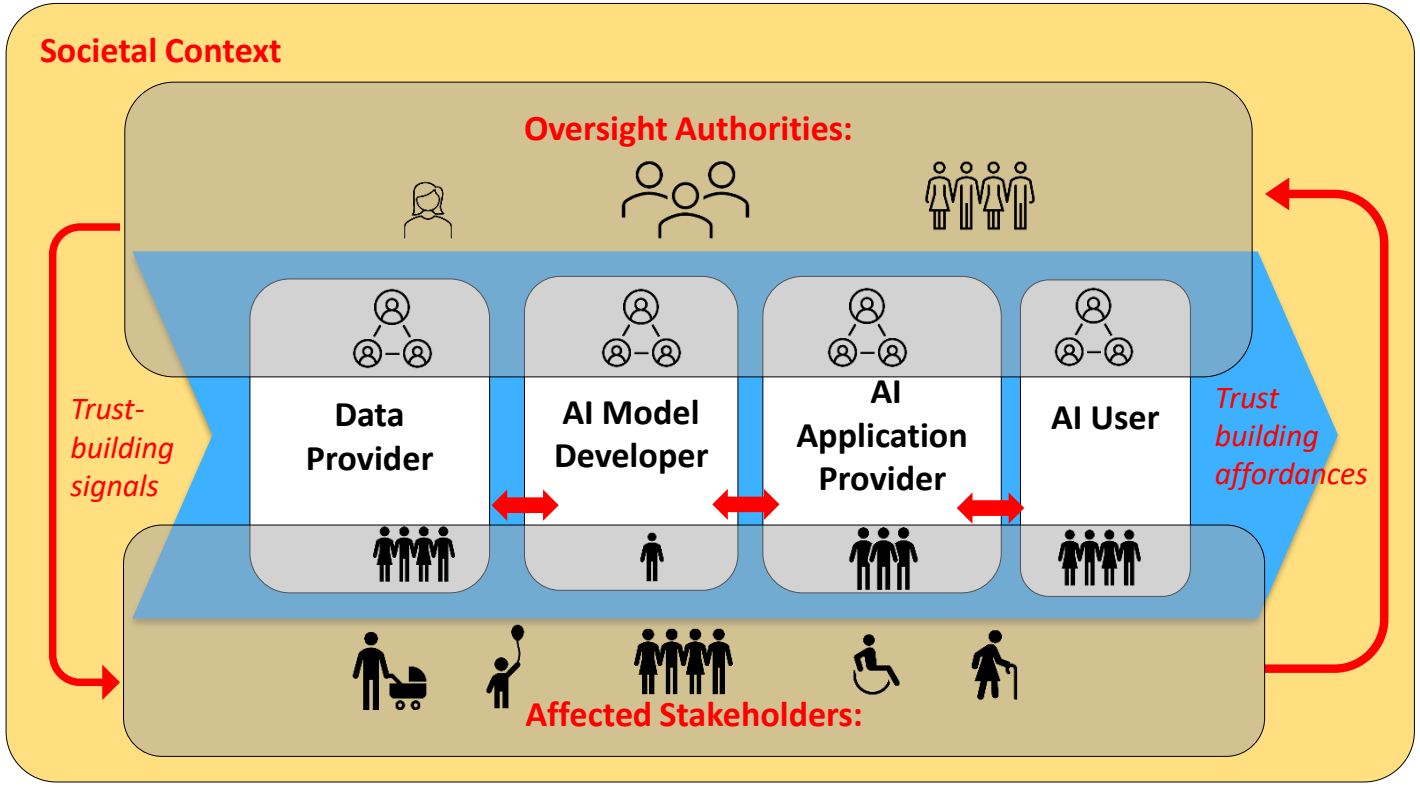
- Evidence to US Congress: https://www.youtube.com/watch?v=WQMuxNiYoz4
- Agenda: https://humanetech.com/wp-content/uploads/2019/06/Technology-is-Downgrading-Humanity-Let%E2%80%99s-Reverse-That-Trend-Now-1.pdf

Ethics in a Technology Development Project



[IBM]

Trustworthy AI and Data Governance: Co Regulation of Digital Technology



GDPR Regulates Processing of Personal Data

- EU Regulation in place since 2018
- 'Personal Data' is any information concerning or relating to an **living person who is either identified or identifiable** (i.e. the 'data subject').
- Applies whenever an individual can be identified, **directly or indirectly**, in particular by reference to an identifier or to factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that individual.
- Certain types of **sensitive personal data** are subject to additional protection:
    - Personal data revealing racial or ethnic origin.
    - Political opinions.
    - Religious or philosophical beliefs.
    - Trade union membership.
    - Genetic data and biometric data processed for the purpose of uniquely identifying a natural person.
    - Data concerning health.
    - Data concerning a natural person's sex life or sexual orientation.

GDPR Principles for Regulatory Enforcement

GDPR enforces principles that an organisation must follow:

- Processed **lawfully, fairly and in a transparent** manner;
- Collected for specified, explicit and legitimate **purpose**s and not further processed in a manner that is incompatible with those purposes; ('purpose limitation');
- **Adequate, relevant and limited** to what is necessary in relation to the purposes for which they are processed ('data minimisation');
- **Accurate** and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate are erased or rectified without delay ('accuracy');
- Kept in a form which permits identification of data subjects for **no longer than is necessary** for the purposes for which the personal data are processed; ('storage limitation');
- Processed in a manner that ensures appropriate **security** of the personal data, ('integrity and confidentiality').

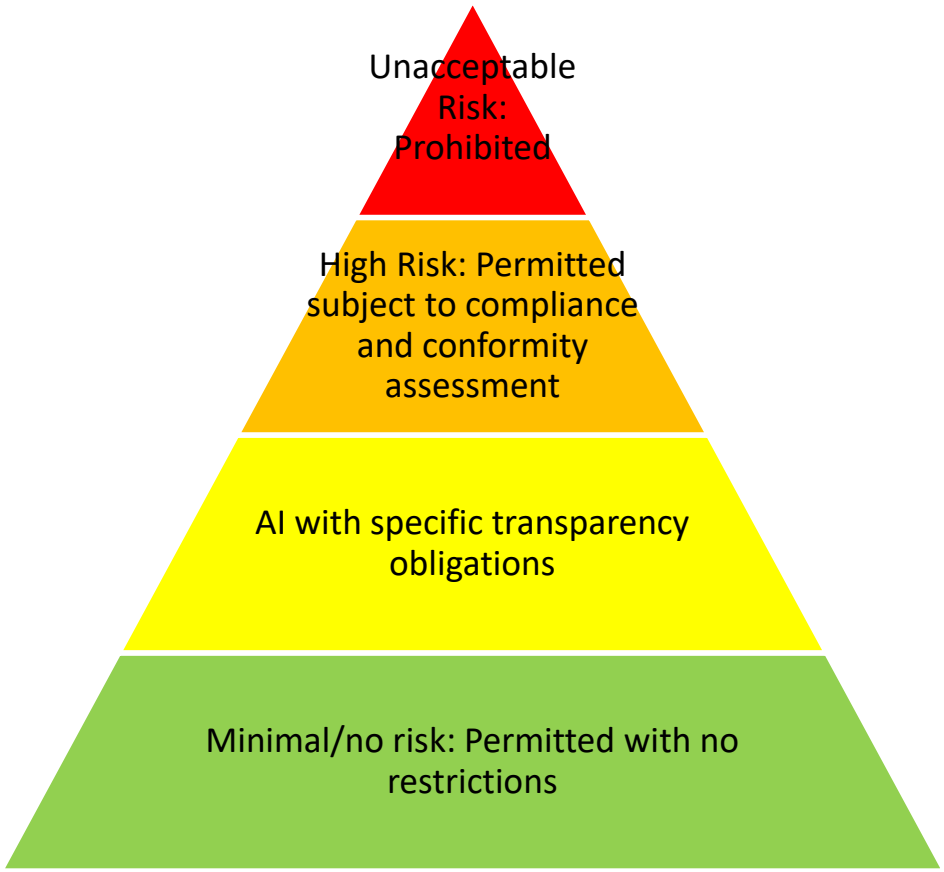
GDPR Enforcement

- GDPR enforces the following **rights for individuals** to help mitigate the risks from processing of data protection:
    - Access
    - To be Informed
    - Rectification
    - Erasure (‘to be forgotten’)
    - Portability
    - On automated decision making
    - To object to processing

- Data Subjects can submit **complaints** to Supervisory Authority if they believe a Data Controller is not honoring rights and principles.
- Data Controllers must clearly and transparently lay out what data they collect, its purpose and how to exercise right in **Privacy Policy**.
- Violators of GDPR may be fined.

13

Proposed Framework for EU AI Act



Aims to ensure AI systems are **safe** and respect **fundamental rights**

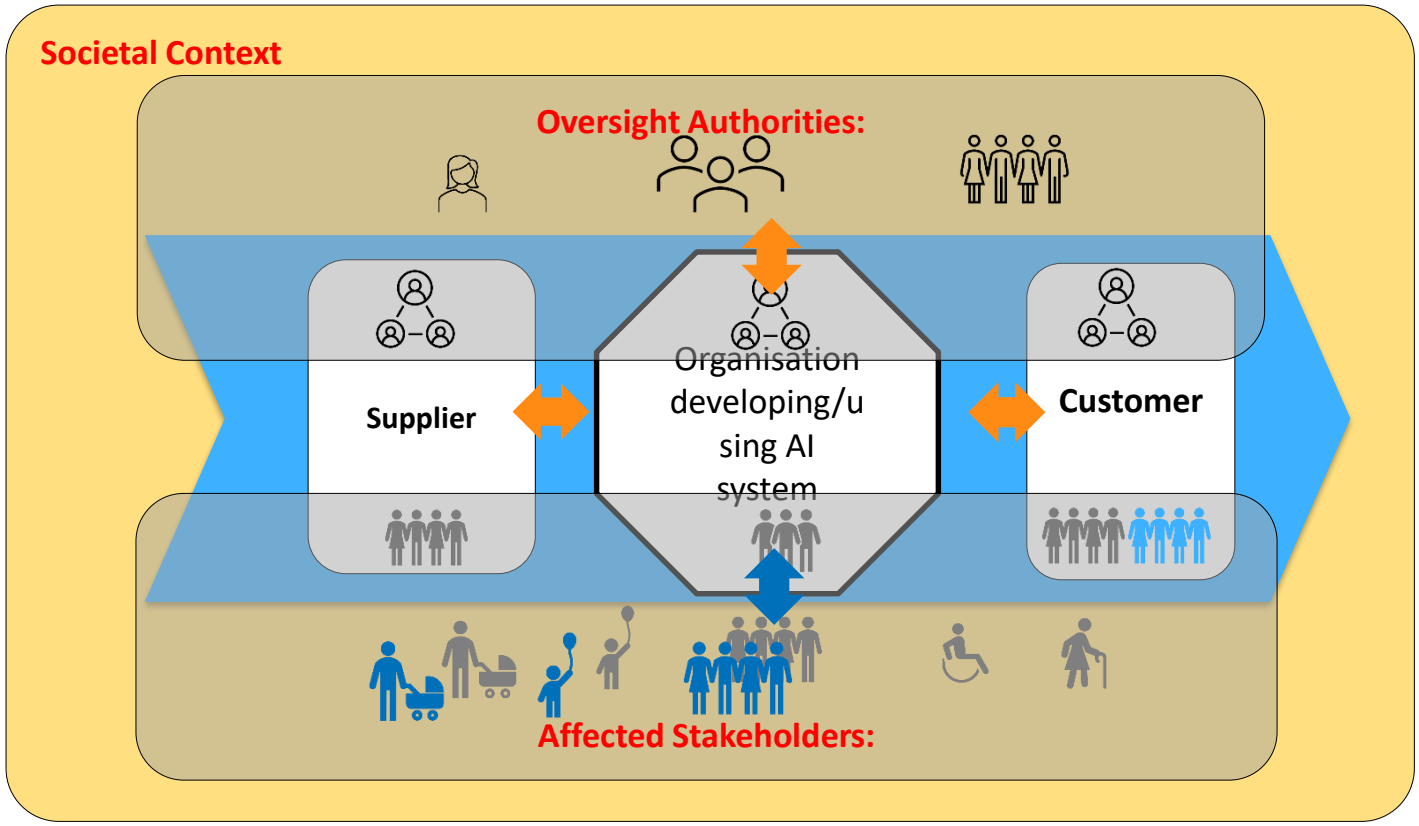
Provide legal certainty for innovation, promote public trust and support single market

A **Risk-based** approach to regulating AI

Large penalties:

https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

Bridging fundamental rights, ethics and management systems



- 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: An Unmediated Reflective Approach to Innovation Ethics

- **Ethic Canvas** is a tool for identifying, evaluating and resolving ethical impacts during R&I stages
- 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**

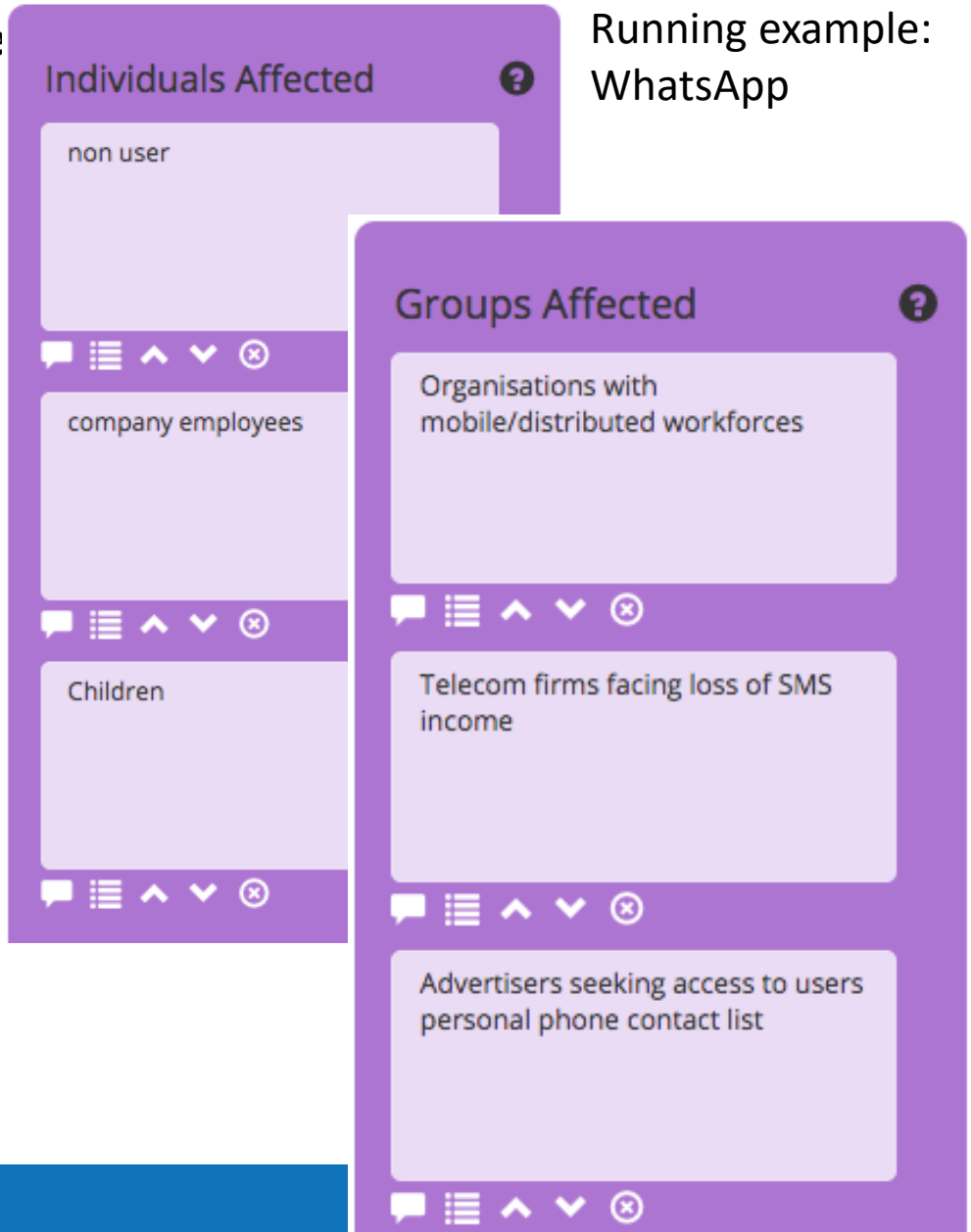
Individuals: Who use your product or service? Who are affected by it's use?

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 organisation, interest groups, etc.?



Running example: WhatsApp

Stage 2: Identifying Ethical Impacts

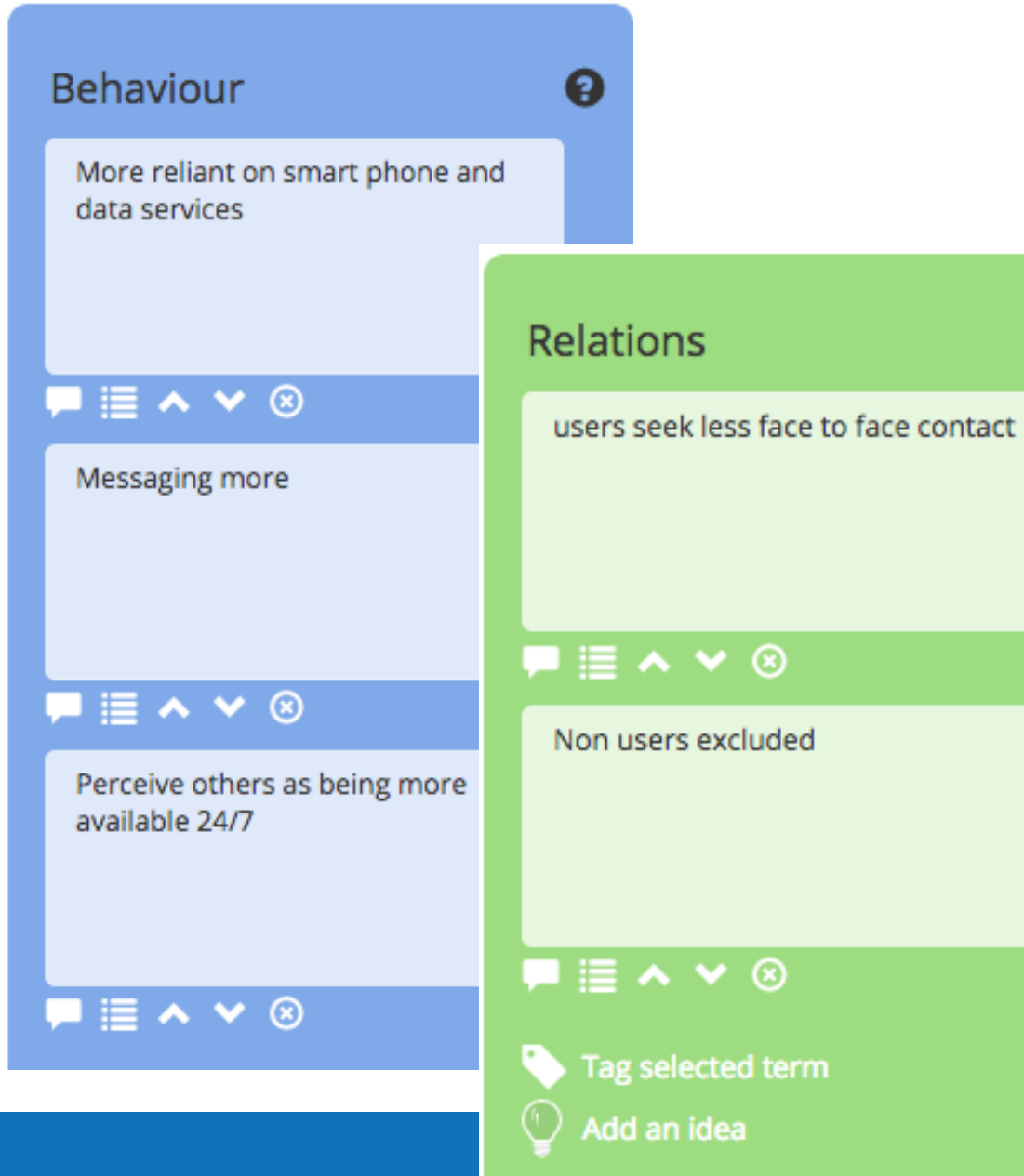
First, **'micro' impacts** are captured by the canvas, i.e. on everyday lives of people using and living with the application

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.?



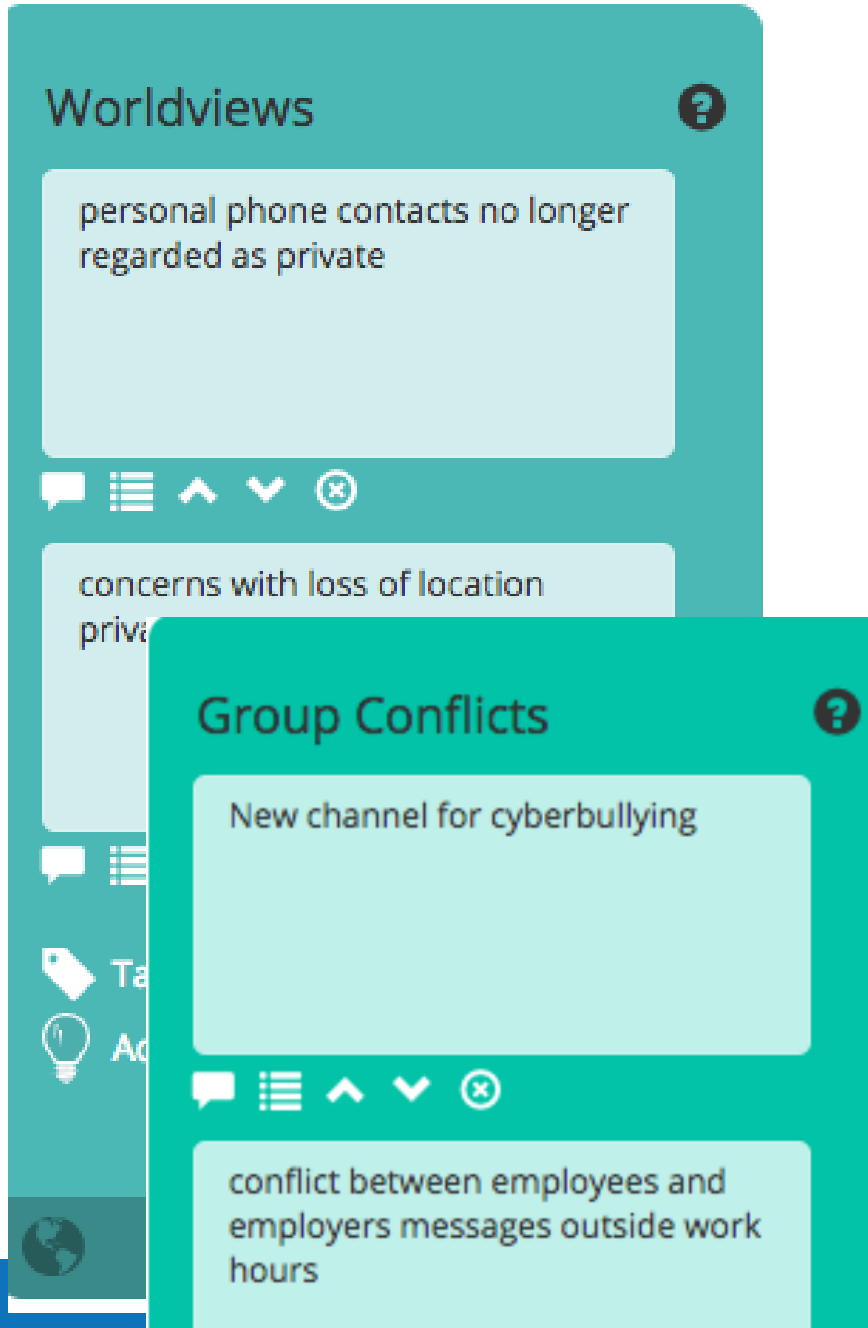
Stage 2: Identifying Ethical Impacts

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.?*

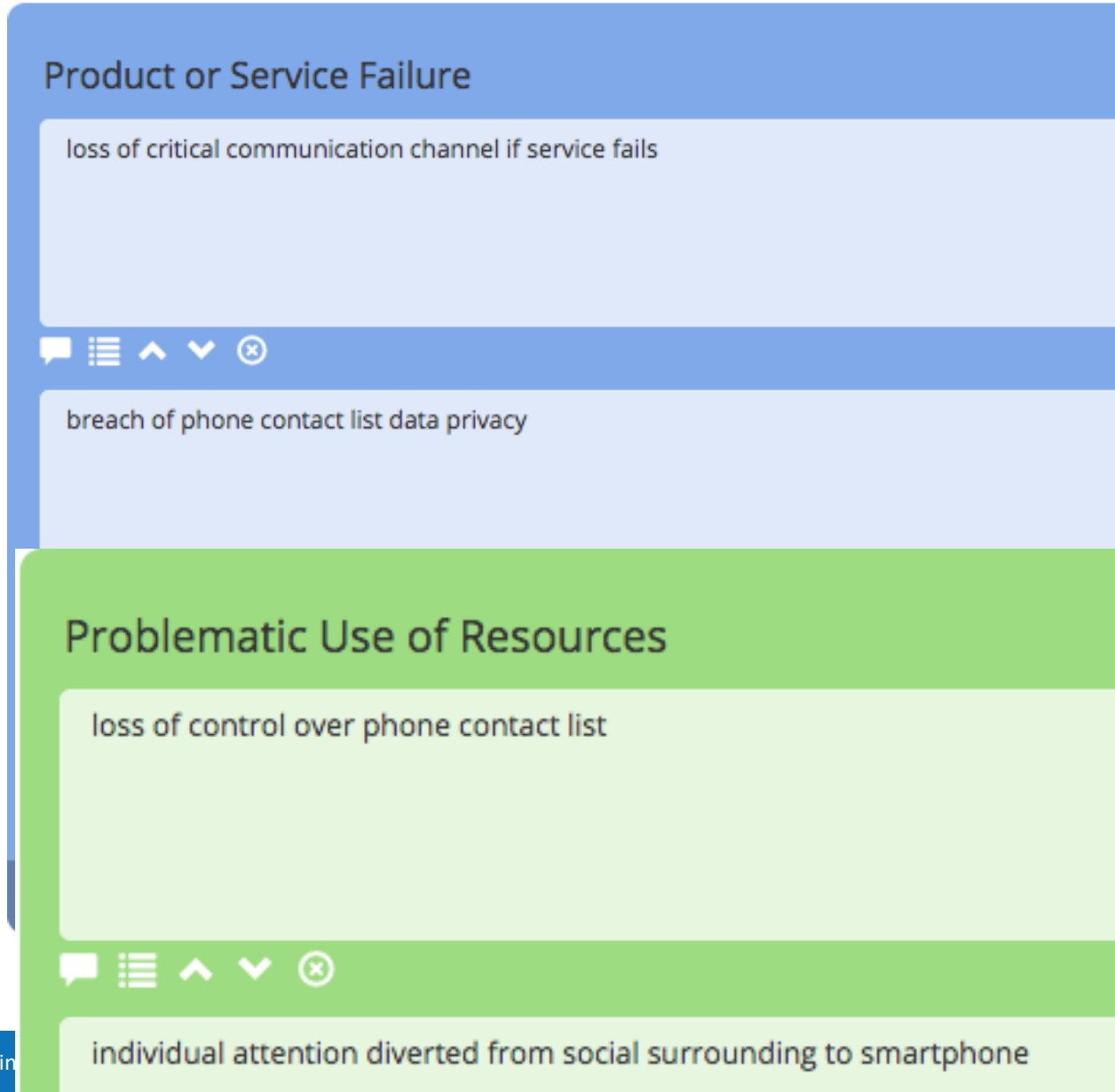


Stage 2: Identifying Ethical Impacts

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.?

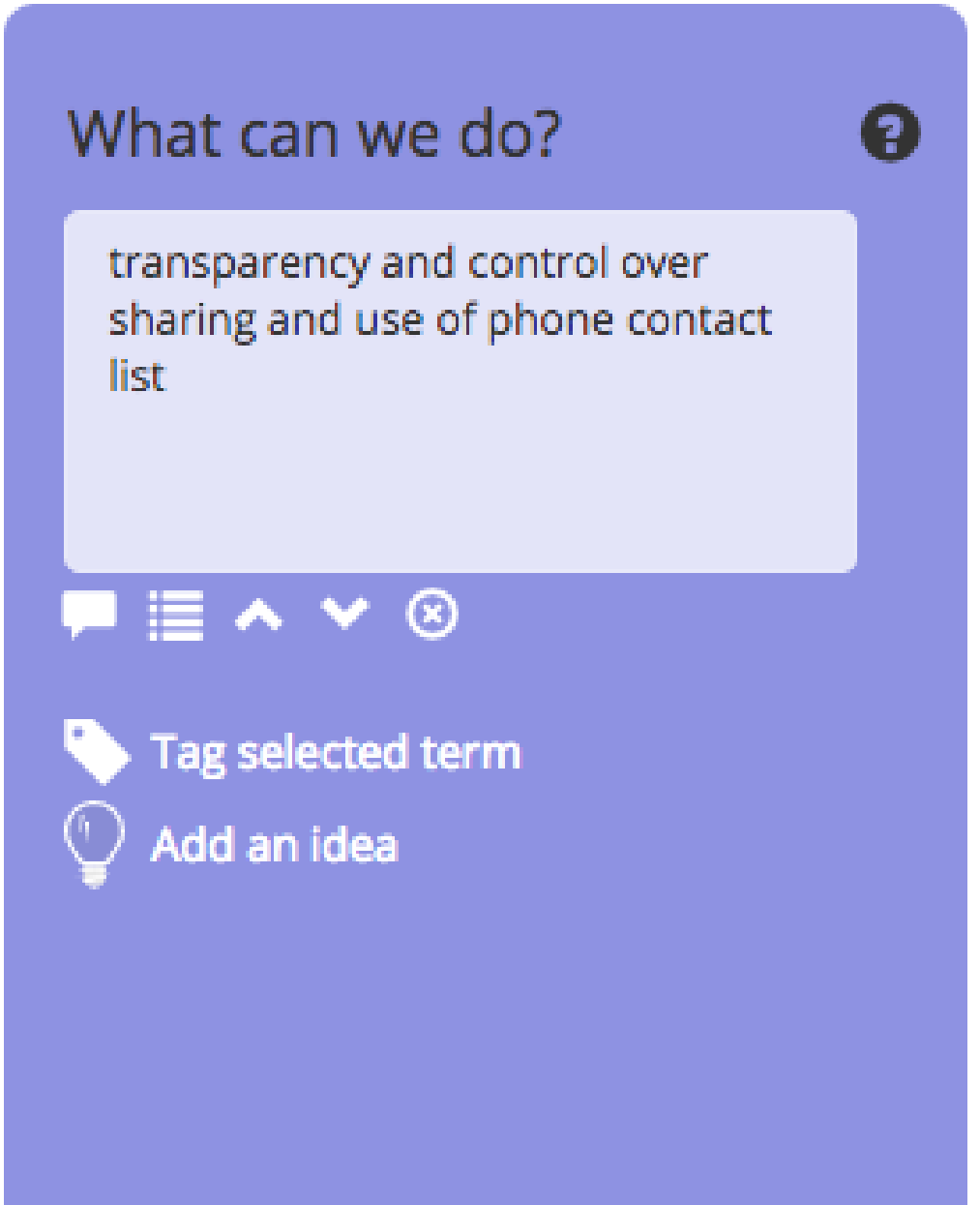
Potential negative impacts of the **consumption of resources** relating to your project? e.g. what happens with its use of energy, personal data, etc.?



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?



Ethics Canvas, Group: Title: Date:

Individuals Affected:
-

Behaviour:
-

Relations:
-

What can we do?:
-

Worldviews:
-

Group Conflicts:
-

Groups affected:
-

Product or Service Failure:
-

Problematic Use of Resources:
-

Ethics Canvas, Group: Title: Microwave Example Date:

Individuals Affected:

Consumer of food

Behaviour:

Less time preparing meals

Easier to live singly/independently

More consumption of ready meals

Relations:

Less family interaction at meal times

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

Group Conflicts:

?

Groups affected:

Cooked food vendors – less business

Fresh food vendors: more value in pre-processed food as convenience attractive to consumers

Product or Service Failure:

Only way of warming food

Microwave unit leaks

Problematic Use of Resources:

More processed food and packaging

Ethics Canvas, Group: Title: YouTube Date:

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

Relations:

Less consuming video as a group

Less consuming same video as social contacts, less common experience to share

What can we do?:

- Green energy for data centres and networks
- Screen time reporting and rationing
- Better screen on inappropriate content

Worldviews:

- Increase in belief in conspiracy theories
- increase in extremist and polarized views

Group Conflicts:

fakenews and distortion of facts impact civic and democratic processes

Employer harms on content moderators

Displacement of local news sources

Groups affected:

- News providers
- Advertisers
- Content providers
- youTubers
- Content moderators

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

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

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

Thank You

1. Innovation Plan

Prepare and submit to blackboard a **group innovation plan** which contains

A brief description of your innovation idea and how it relates to your assigned dissertation (max 1 page).

An initial business model canvas (BMC) covering at least the value proposition and customer segment components.

Supporting text to elaborate on the content included in your BMC including the value proposition, customer segments and any other BMC section you have completed.

An ethics canvas relating to your technology based business model.

A summary of the hypotheses you plan to test, how you are planning to test these hypotheses (though interviews etc)

A short summary of how you have organised your work as a group.

2. Group critique of assigned dissertation

As a group and using feedback form templates provided contribute to a **critique of the dissertation** assigned to your group. Based on lecture notes, learning from the group assignment processes and marking scheme provided complete the critique feedback form to capture your groups' views on the dissertation and justification for these views.

3. Innovation Presentation

Contribute to a **group presentation** on your business model development submitted as a 10 min max video. The audience for your video are potential investors. In your video, you should aim to be:

Clear in the messaging, including on the testing of hypotheses

Convincing and direct in the presentation of evidence gathered and the persevere/pivot decisions made.

Realistic in your assessment of the viability of your current business model.

1. Research Plan

Individually, submit an **initial research plan** related to your MSc research project and captured in a set of 9 slides providing the following information

research canvas (as per template given in class) supported by relevant references from your reference list.

motivation statement (approx. 150 words)

glossary of key terms in your canvas.

list of at least 7 references which support your research idea, method and motivation

research project Gantt chart to capture the timeline for your project.

using the template given in class, include a literature review protocol. The protocol should indicate how you will conduct your initial literature review. The evidence resulting from this literature review should provide support with placing your research in a context of your chosen research field.

an explanation of whether or not you require to get ethical approval for your research.

list technical and research skills you have or hope to develop to carry out your research project.

list 2 key pivots you made to your initial research plan based on discussion with a classmate.

2. Dissertation Critique

Individually, select a dissertation of interest to you from https://publications.scss.tcd.ie/theses/diss/dissertation-index.mcs.php#2024_MCS. Based on lecture notes and learning from the group assignment processes and marking scheme provided during class critique this dissertation. Complete the critique feedback form to capture your views on the dissertation and justification for these views.

Research Plan **indicative** rubric

	>70%	60%-69%	50%-59%	<50%
Supporting slides – Motivation statement, glossary, literature review protocol, Gantt chart, Ethics statement, skills list, reference list, key pivots (40%)	Excellent understanding demonstrated of activities involved in completion of a research project e.g. ethical approval, dissertation planning, literature review, research and technical skills development The work is comprehensive showing detail of the planned work on a Gantt chart. References are from excellent sources and are used to support motivation, domain, research approach, evaluation and technology. Clear motivation aligned to the aims of the research. Considered literature review protocol included. Evidence of consideration of the overall plan based on pivots mentioned.	The work is comprehensive but not backed up with a detailed plan. Some understanding demonstrated of activities required to complete a research project. Some activities may be missing or poorly aligned with others. References are included but may be from sub optimal sources or not of the required mix i.e. to support a mix of motivation, domain, research approach, evaluation and technology. A research motivation is included but may not be adequately aligned to the research aim. All required elements included.	The content is not comprehensive e.g. very high level objectives, generic details used on the Gantt chart. Poor understanding of activities required to complete a research project. Some activities may be missing or poorly aligned with others. More of a focus on building a piece of technology rather than answering a research question. The motivation statement may be a description of the domain rather than addressing why the student thinks it is a good piece of research to undertake. References are from poor sources, not a good breadth (i.e. covering motivation, domain, research approach, evaluation, technology) There may be scoping issues. Some required elements may be missing.	Little or no understanding of what is involved in presenting a research project plan. Little of no understanding of what is involved in completing a research project. Little or no understanding of scheduling tasks to complete a research project. Too few references included.
Research Canvas (50%)	Clear alignment between the objectives and Method/approach on the research canvas. Evaluation section is aligned with the research aim. Research aim and contribution are aligned appropriately. There is coherence between the information on the research canvas and the Gantt chart. Citations to references are included on the research canvas.	There is alignment between the objectives and Method/approach on the research canvas. There is coherence between the information on the research canvas and the Gantt chart. Research aim and contribution may not be aligned. Evaluation may not be aligned with research aim.	Some lack of coherence between objectives, approach, evaluation, aim, motivation and/or contribution. Some lack of alignment between Gantt chart and research canvas.	Incomplete research canvas Lack of coherence between objectives, approach, evaluation, aim, motivation and/or contribution. Lack of alignment between research canvas and overall research plan.
Presentation (10%)	Excellent presentation demonstrated in style, consistency and clarity of content.	Very good presentation that clearly presents a consistent research plan. May be lacking in style.	Adequate consistent presentation. Lack of clarity and style.	Poor presentation lacking consistency, style and clarity.

Innovation presentation **indicative** rubric

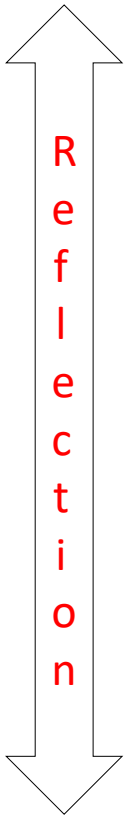
	>70%	60%-69%	50%-59%	<50%
Business model canvas and supporting text (30%)	All elements attempted, consistent (and colour coded) across different CS, no confusion between elements. Some elements beyond VP, CS, CH, CR and RS attempted. Strong overall cohesion between elements. CS comprehensively described. VP clearly aligned with CS, addressing specific, prioritised pains/gains for each CS.	Product market fit between VP and CS is clearly present. CH, CR and RS clearly address requirements of product-market fit. Element vary in their contribution to the consistency and cohesiveness of the model. CS clearly described. Reasonable alignment of VP and CS, with some expression of prioritisation.	VP, CS, CH, CR and RS have entries that are consistent. Reasonable alignment of VP and CS, though they not be prioritised. CS described.	Have not clearly identified core elements for market fit (VP, CS) or logically consistent roles for CH, CR and RS. Conceptualisation of product/service is unclear, features not clearly defined and relationship to CS are not clear CS not adequately described.
Hypothesis formation. Experiment Results and action/pivots (50%)	Clear set of hypothesis established across VP, CS, CH, CR and RS, and others. Clear individual pass/fail criteria. Hypothesis testing clearly defined. Data collected. MVP mock-up used to good effect for (and only for) customer discovery. Data collection well documented, including unanticipated insights. Clear pass/fail assessment made. Clear pivots/persevere action identified based on facts gleaned from hypothesis testing or ethics canvas considerations.	Overall testable hypothesis for PMF (VP+CS), CH, CR and RS identified. Some CS data collection undertaken. Hypothesis testing clearly related to business model. Results clearly documented, including numbers of and selection of participants. Outcome of hypothesis testing not clear in all cases. Pivot/persevere decision not fully documented	Overall questions identified but not in a testable form. Broad assessment planned, but does not seem to test a specific hypothesis. No attempt made to identify or collect data from relevant to CS discovery.	Little/No hypothesis formation attempted. No consideration of hypothesis testing to perform. Investigations bear no relate to business model.
Challenges faced (10%)	Very competent in recognising, planning for and overcoming challenges in testing hypotheses.	Some degree of planning for and attempts to overcome challenge in hypotheses testing.	Some understanding of the challenges but not really addressed.	No reflection of the challenges faces
Presentation and Conclusions (10%)	Crystal clear description of the final business model arrived at, including on the testing of hypotheses. Convincing, persuasive and direct presentation of evidence gathered. Clear, evidenced statements on the persevere/pivot decisions made. Realistic assessment of the viability of your final business model	Not all elements of the business model, or the cohesiveness between them is well presented. Some gaps in the narrative of how the evidence was collected. Reasonable assessment of the viability of your final business model.	Business model is unclear, and is not presented in a way that might convince a potential investor. The testing of hypotheses is not clear and its impact on the development of the business model is not apparent. No assessment of the viability (or otherwise) of the business model.	It is not apparent that a single business model proposal has been arrived at. It is not apparent that any testing of hypothesis contributed to business model aspects. No conclusion on business model assessment presented.

Key:
BMC - Business Model Canvas
VP - Value Proposition
CS - Customer Segment
CH - Channels
CR - Customer Relationships
RS - Revenue Stream
PMF - Product-Market Fit
MVP - Minimum Viable Product

Innovation Methods and Research

Iterative steps related to your innovation tasks on module

- Your innovation has a source(s)- Literature, dissertation, discussion, innovativeness.
- You have a theoretical framework (BMC) for structuring, discussing and presenting your innovation
- The BMC provides you with a structure to form hypothesis to guide you in gathering evidence to test the hypothesis.
- Conclusions (pivots) are made based on evidence resulting from hypothesis testing using questions, MVP, ethics canvas.
- Targeted communication of your innovation, evidence collected and pivots in the form of a plan and a final presentation t potential investors



Iterative steps related to your research plan and project on module and for your MSc research project.

- Your research question has a source(s)- Literature, dissertation, discussion, innovativeness.
- You have a discipline research question specific way for structuring, discussing and presenting your research.
- Literature, knowledge and research project description will structure and guide you in gathering evidence to answer your research question.
- Conclusions are made based on evidence resulting from data collection using research question specific methods.
- Targeted communication of your research in the form of a plan (this module), a presentation and a dissertation(MSc research project).