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

What is the Internet Doing to Me:

Ethics in Practice: Ethic 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.

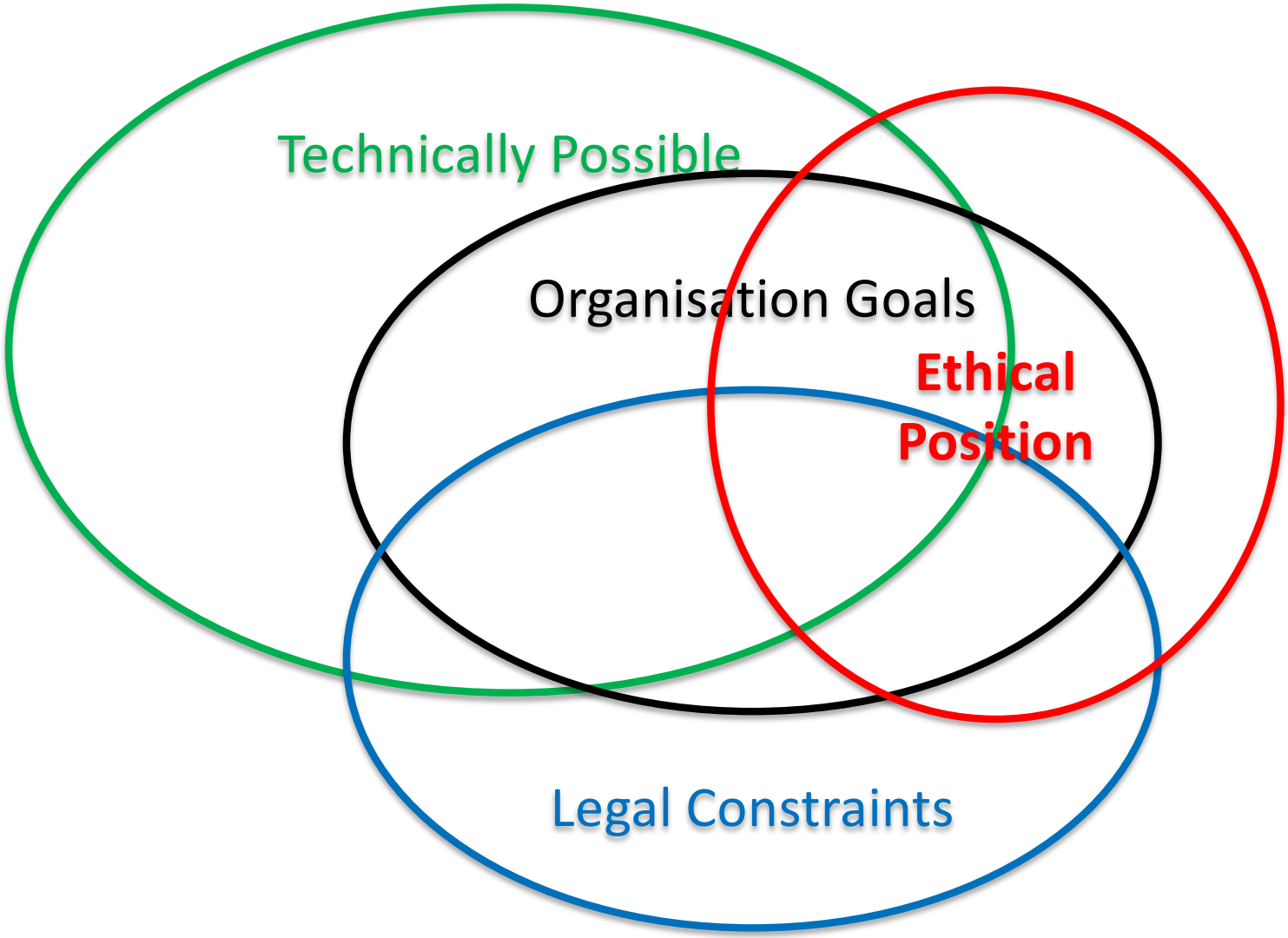
Students should not edit or modify the recording in any way, nor disseminate it for use outside of a context of teaching and learning at Trinity.

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/.

Ethics in a Technology Development Project



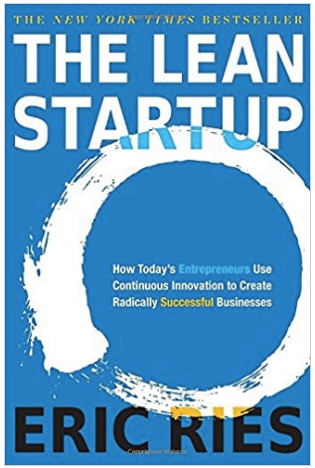
[IBM]

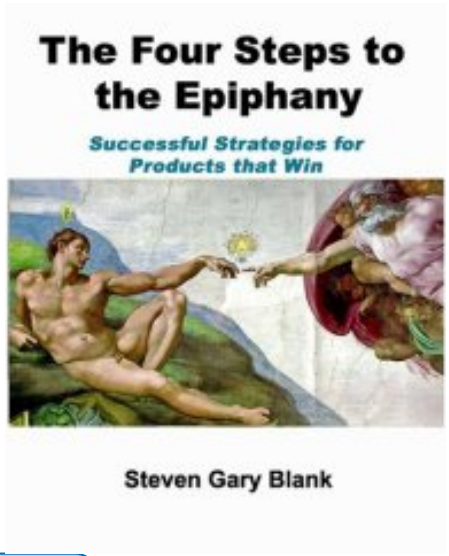
Practicing Ethics in Responsible Research & Innovation (R&I)

- 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

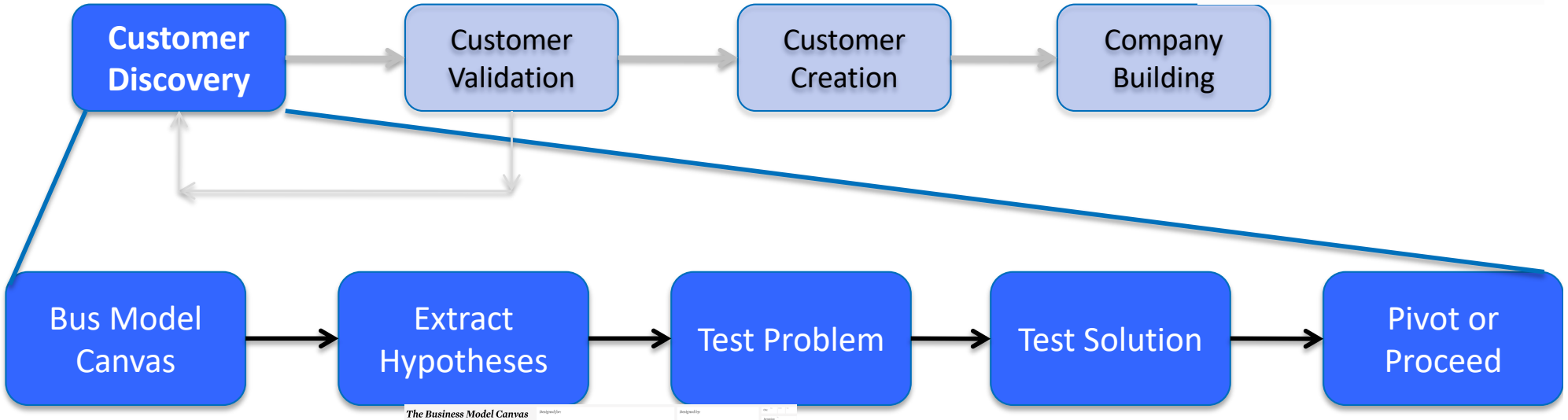
- ***Disclosure* level** neglected in current methodologies
- Need to:
  - Keep pace with **volume and speed** of innovation
  - **Accessible** to non-ethicist
    - R&I teams have an important perspective
    - R&I teams position to implement pivot to mitigate negative impact
  - Enabling a **collaborative**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

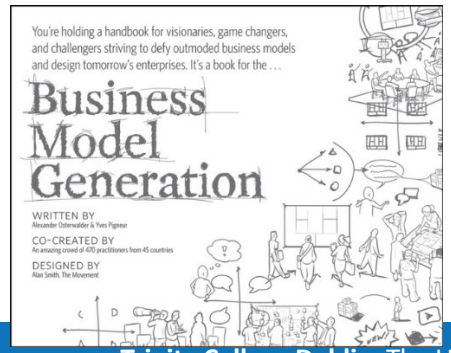
Data Hungry Innovation - "Silicon Valley" Methods

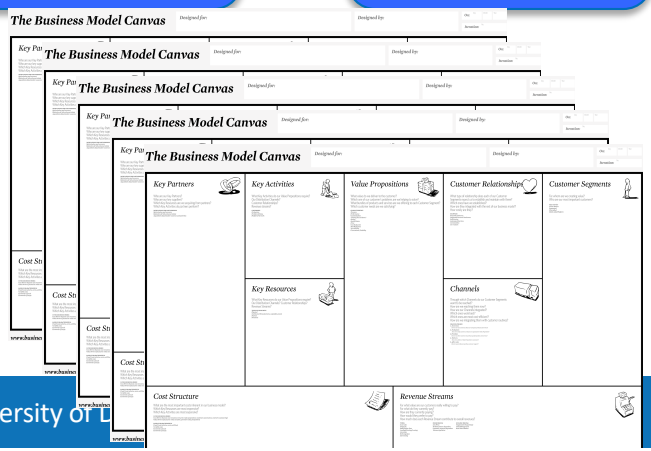




The Customer Development Process







How to make ethics part of the process?

Ethics Canvas: Lightweight approach

- **Ethic Canvas** is a **tool** 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
- Impact in the **public sphere, conflicts,** our **worldviews**
- Impact of technology **failure**
- Impacts on the **environment and other shared resources**

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 [blocks 1&2]

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

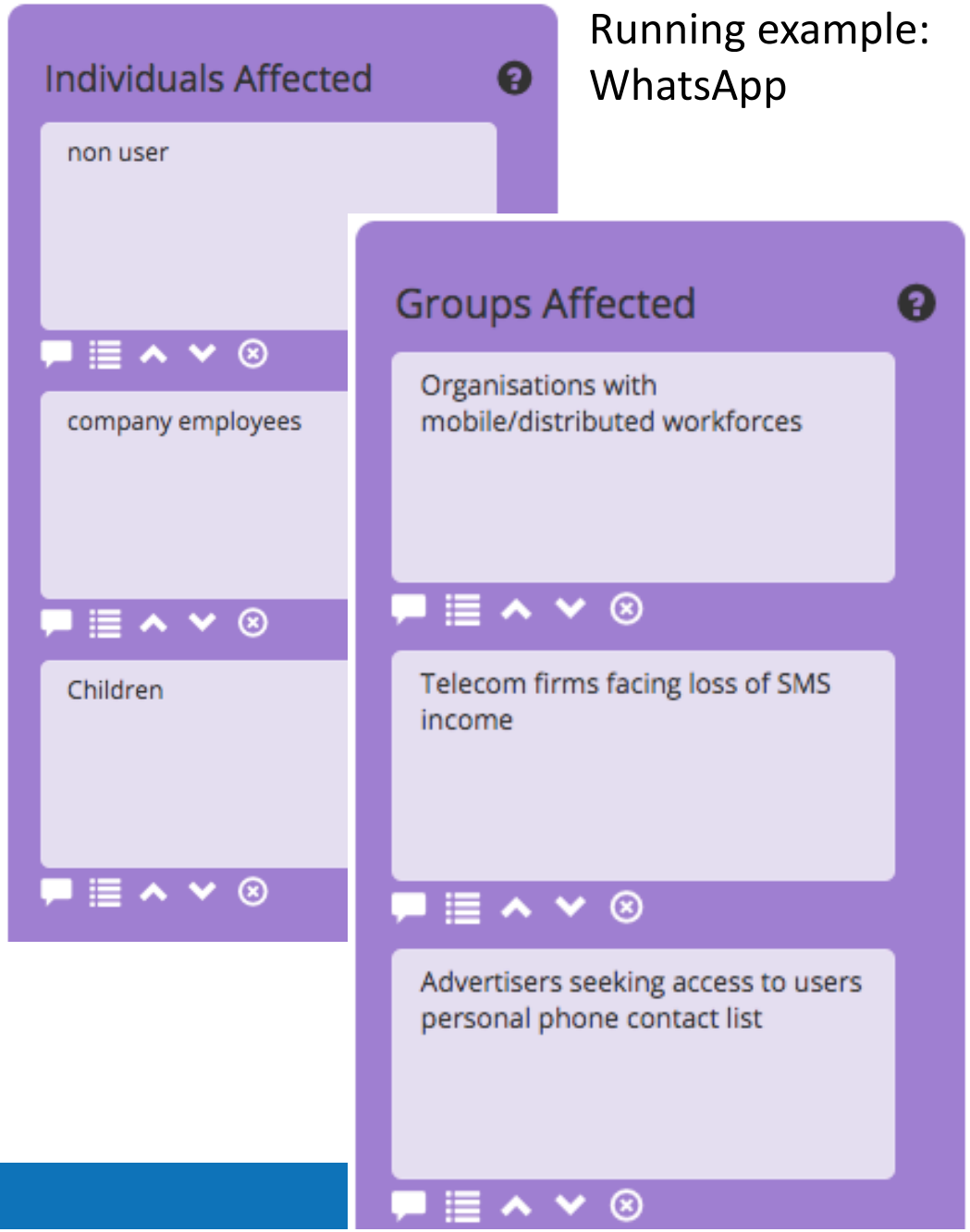
Individuals: Who use your product or service? Who are affected by it's use?

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 [blocks 3&4]

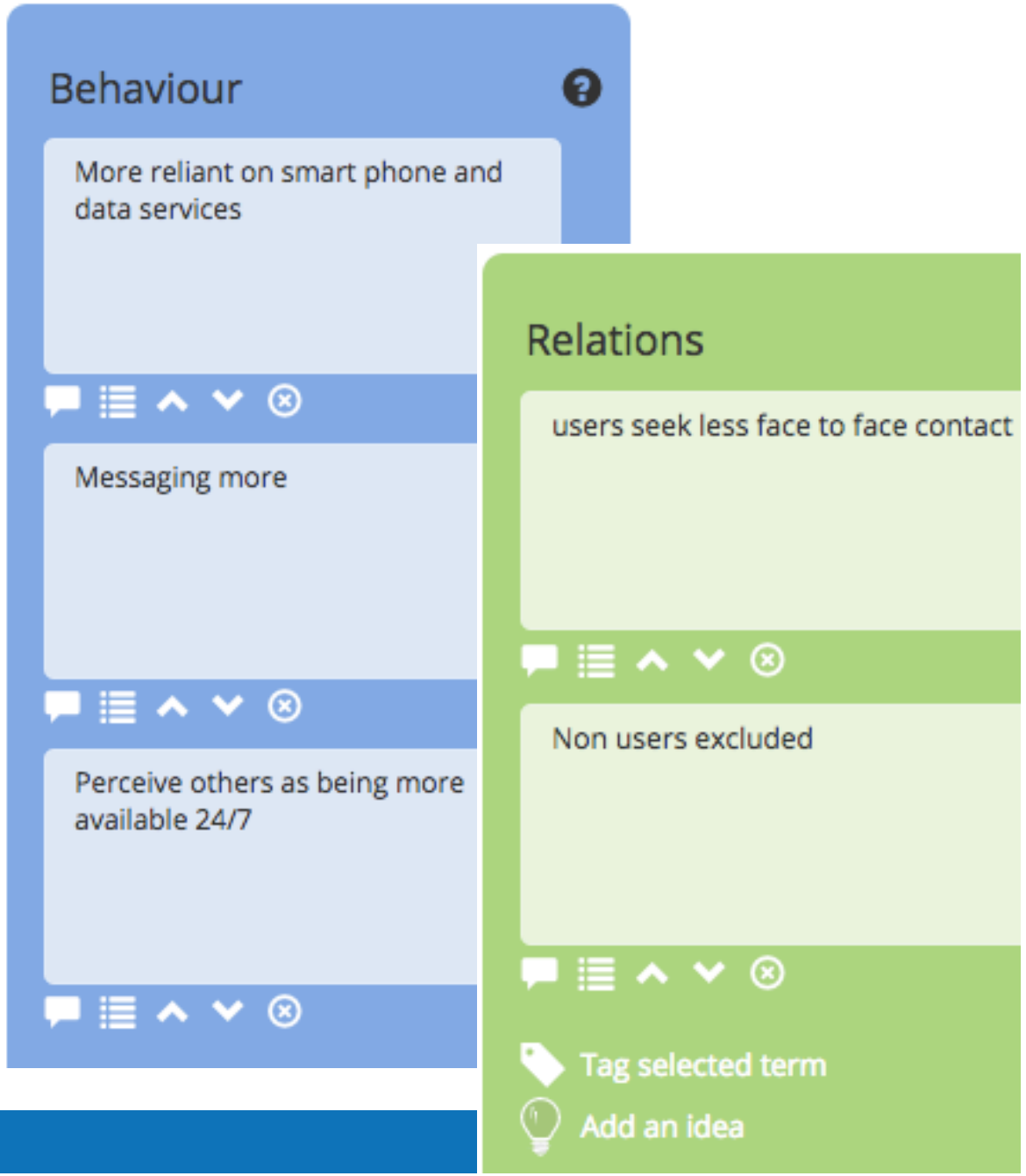
First, **'micro' impacts** are captured by the canvas, i.e. on everyday lives of people using and living with the application

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.?



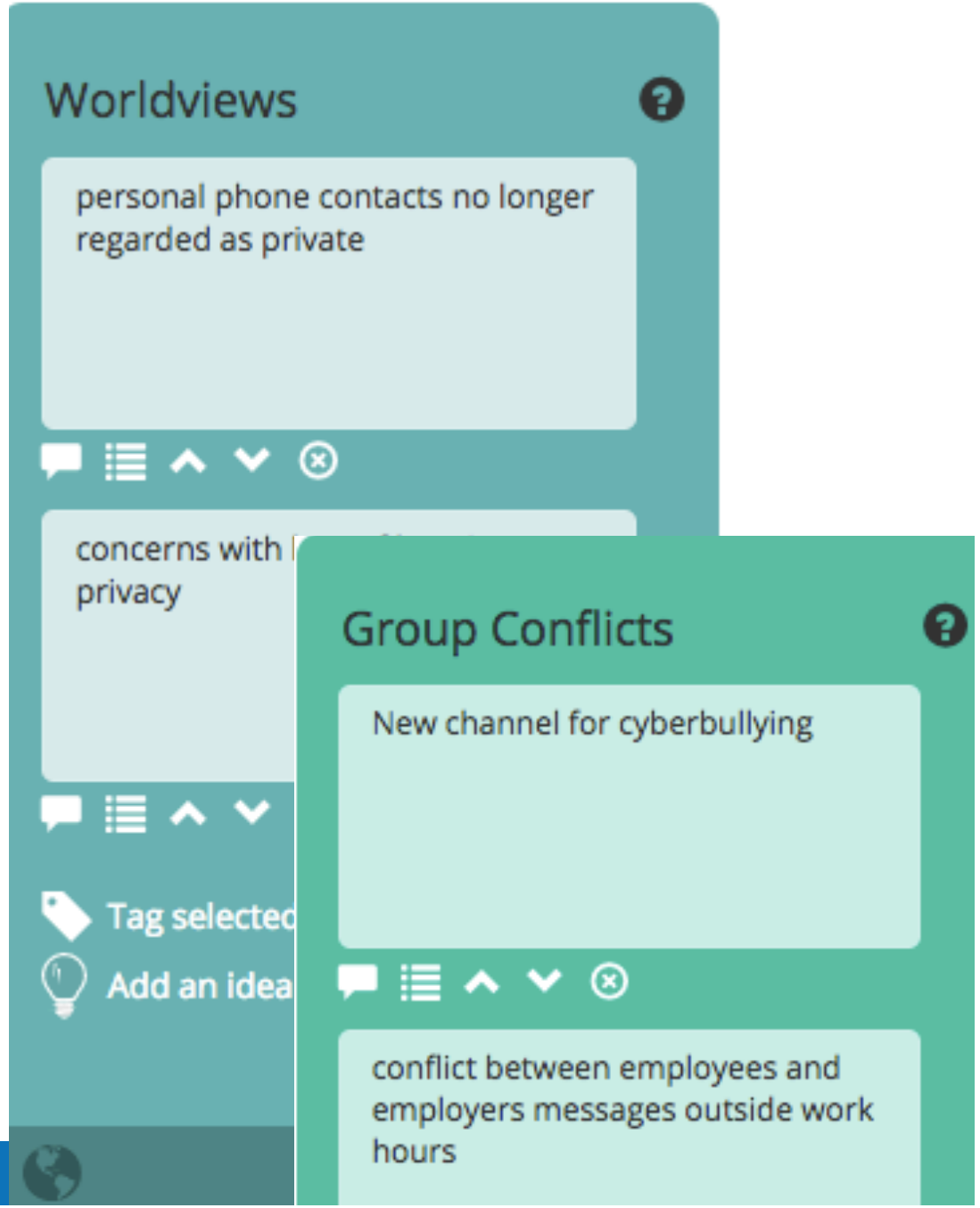
Stage 2: Identifying Ethical Impacts [blocks 5&6]

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 [block 7&8]

Aspects that *indirectly* impact our lives..

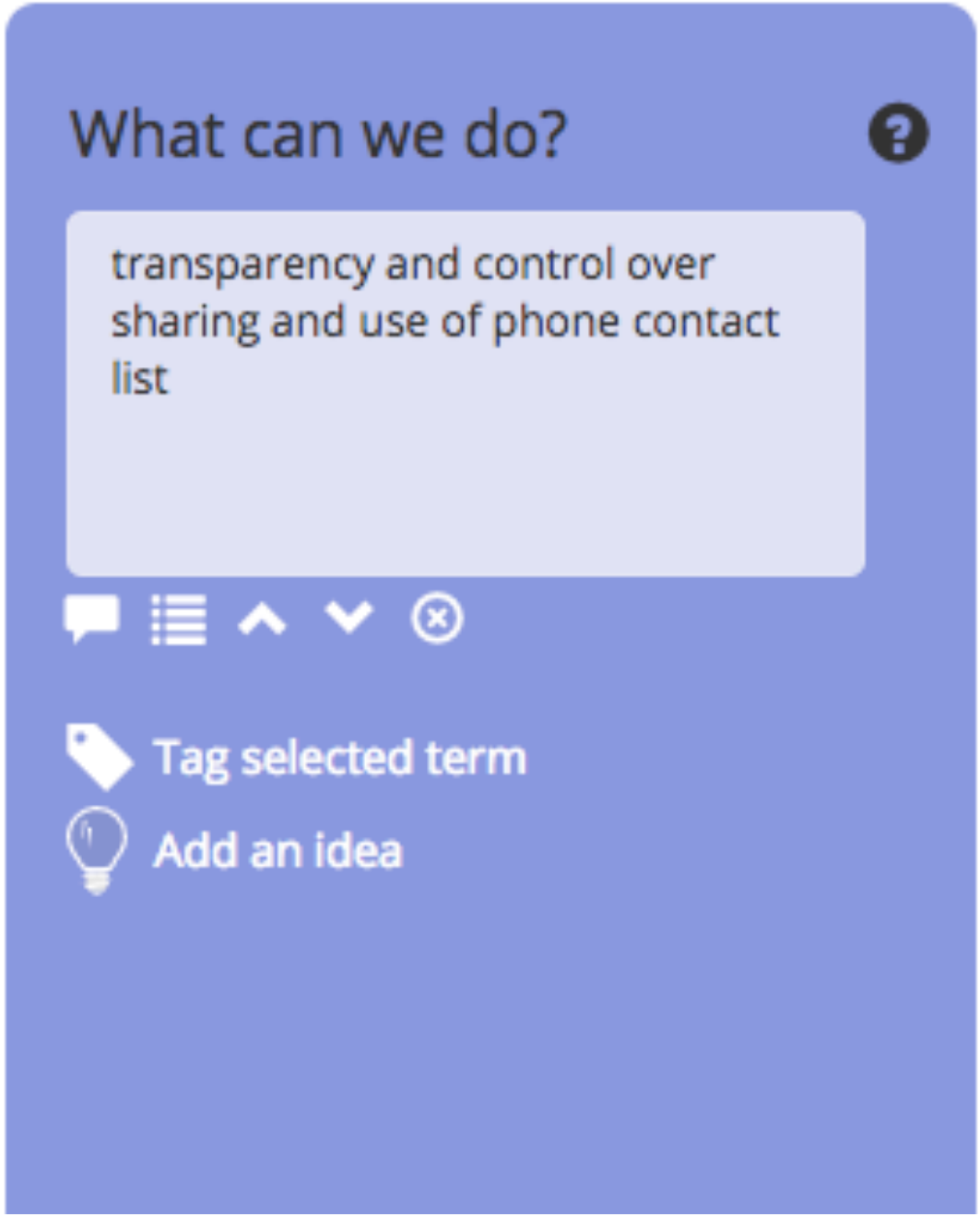
Potential negative impact of your **product or service failure**? e.g. what happens with technical errors, security failures, etc.?

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 [block 9]

What are the most important ethical impacts you found?

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



Thinking about Stakeholders

- **Stakeholder types for Social Responsibility issues** [ISO 26000:2010]
    - Human Rights: everyone
    - Labour practices: workers
    - The environment: future generations
    - Fair operating practices: customers and providers
    - Consumers
    - Local Community involvement and development

Stakeholders: Human Rights

Risks	Treatments
• **Legal**, from impacts in equality, privacy, access to justice • **Reputational**, from impacts to dignity, physical and mental integrity • **Complicity** in partner violations of rights • **Conflicts between stakeholder**, e.g. investors vs consumers, suppliers vs local communities • **To civil & political rights**: e.g. fake news social media bots, deep fake video impacting elections, filter bubbles, censorship • **To economic, social, cultural rights**: education, healthcare, wellbeing • **To just and favourable work**: casualised and deskilling labour of gig and click workers	• *Due diligence*: Human rights policy • *Avoid* value chain partners that may commit violations • Establish *grievance and redress mechanism*: transparent, accessible, external scrutiny, e.g. AI explanations • *Monitor for discrimination* towards vulnerable groups in AI decision making, e.g. insurance, justice, recruiting • *Education and access* for all groups to benefit of AI • Ensure worker *freedom of association* and collective bargaining

Stakeholders: Labour Practices - Workers

Risks

- Legal: arising from **discrimination** in AI assisted recruiting
- Reputational: **worker dissatisfaction**, e.g. to intrusive monitoring
- AI automation leading to **labour displacement**
- **Deskilling** of work, e.g. translators correct machine translations
- To worker **physical and mental health**

Treatments

- *Recognition* of secure employment & decent working conditions
- Engage in *social dialo*gue with workers and affective professional and community representative
- Employee *retraining*
- *Health and safety practices*, e.g. for robot coworkers, offensive content moderators
- Protect *personal data of employees*
- Seek *assurance* of good labour practices in value chain partners

Stakeholders: The Environment – Future Generations

Risks	Treatments
• Increased **carbon emission** due to AI training and service operation • Resource usage and **pollution** from AI-driven product creation and disposal, e.g. sensors, batteries, obsolete smart phones	• Monitor and plan reduction of non sustainable energy and resources use over whole product lifecycle • Make AI services available for environmental monitoring and analysis

Stakeholders:
Fair Operating Procedures: Suppliers, Customers, Regulators

Risks	Treatments
• Use of AI in **corrupt or anti-competitive practices**, e.g. finance, investment, procurement • Use of AI to **undermine the public political process**, e.g. through deep fakes, targeted manipulation or misinformation online • Violation of **intellectual property** rights	• Ensure *transparency and other safeguards* against abuse of power or complicity, e.g. protecting whistleblowers • Promote responsible behaviour in *value chain partners*, e.g. through requesting ethical impact assessment from AI partners • Identify, respect and fairly compensate *right holders*, e.g. annotators, translators, content providers

59

Stakeholders: Consumers

Risks

- Conveying **deceptive, misleading, fraudulent or unfair** information to consumers
- Endangering **consumer health** and safety, e.g. mental health, self image
- **Unsustainable consumption**
- Misuse of **personal data**
- Denial of access to **essential services**

Treatments

- Clearly *identify promoted content* and its sponsors
- Monitor and benchmark *safety performance* and correct problems promptly
- Consumer '*nutrition labels*', e.g. performance and failure envelope, energy usage
- Clear and accessible *complaint and redress* mechanisms
- Compliance with *privacy regulations*, e.g. transparent on data held or shared and its use
- Consumer *education and awareness* raising, regardless of their capabilities or accessibility needs
- *No discrimination or censorship* in access to services and information

Stakeholders:
Community Involvement and Development

Risks	Treatments
• Difficulty **identifying communities** suffering negative impact of AI use (including non-users), e.g. social networks, road users, medical patients • Negative impact on **local health, employment and wellbeing**, e.g. deskilling, child development, culture wars • **Concentration of AI's wealth and income creation** away from local communities • Perpetuating **local dependence on philanthropic** activities	• *Consult* with early and *widely* with communities, especially where vulnerable • Be *transparent* on engagement with local authorities • Promote *education and preservation* of local cultures • Support *employment creation and skills development* in impacted communities and along value chain • Direct AI to *solve local* social and environmental issues • Enhance *local scientific and technological* development and entrepreneurship • Promote *economic diversification*, support local suppliers and employment

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

References

- Principled Artificial Intelligence: Mapping Consensus in Ethical and Right-based Approaches to Principles for AI, Jessica Fjeld, Adamn Nagy, Berkman Klein Centre, Jan 15, 2020, https://cyber.harvard.edu/publication/2020/principled-ai
- AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, Floridi, L., Cowls, J., Beltrametti, M. et al. Minds & Machines (2018) 28: 689. https://doi.org/10.1007/s11023-018-9482-5
- Gasser, Urs, and Virgilio A.F. Almeida. 2017. "A Layered Model for AI Governance." IEEE Internet Computing 21 (6) (November): 58–62.
- Hagendorff, T., (2019) "The Ethics of AI Ethics – An Evaluation of Guidelines", https://arxiv.org/pdf/1903.03425.pdf
- Adamson, G., Havens, J. C., & Chatila, R. (2019). "Designing a Value-Driven Future for Ethical Autonomous and Intelligent Systems". Proceedings of the IEEE, 107(3), 518–525. https://doi.org/10.1109/JPROC.2018.2884923
- Leenders, G. (2019). "The Regulation of Artificial Intelligence — A Case Study of the Partnership on AI". Medium, April, Retrieved from: https://becominghuman.ai/the-regulation-of-artificial-intelligence-a-case-study-of-the-partnership-on-ai-c1c22526c19f
- EU High Level Expert Group on AI, (2019). "Ethics Guidelines for Trustworthy AI", April 2019, Retrieved from https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines
- "Mapping regulatory proposals for artificial intelligence in Europe", Access Now, Nov 2018. Retrieved from https://www.accessnow.org/mapping-regulatory-proposals-AI-in-EU
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Daumé, H., Crawford, K. (2018) "Data sheets for Datasets", in Proceedings of the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning, Stockholm, Sweden, 2018, available at: https://arxiv.org/abs/1803.09010
- Buchanan, B., Miller, T. (2017) "Machine Learning for Policy Makers - What it is and Why it matters" Belfer Centre, available from: https://www.belfercenter.org/sites/default/files/files/publication/MachineLearningforPolicymakers.pdf
- Joshua A. Kroll , Joanna Huey , Solon Barocas , Edward W. Felten , Joel R. Reidenberg , David G. Robinson & Harlan Yu, "Accountable Algorithms", 165 U. Pa. L. Rev. 633 (2017). Available at: https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3
- Calo, Ryan, "Artificial Intelligence Policy: A Primer and Roadmap" (August 8, 2017). Available at SSRN: https://ssrn.com/abstract=3015350 or http://dx.doi.org/10.2139/ssrn.3015350
- Future of Life Institute. (2017). "Asilomar AI Principles". Future of Life Institute. Retrieved from https://futureoflife.org/ai-principles/
- Scherer, Matthew U., "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies" (May 30, 2015). Harvard Journal of Law & Technology, Vol. 29, No. 2, Spring 2016. Available at SSRN: https://ssrn.com/abstract=2609777 or http://dx.doi.org/10.2139/ssrn.2609777
- Saurwein, Florian and Just, Natascha and Latzer, Michael, "Governance of Algorithms: Options and Limitations" (July 14, 2015). info, Vol. 17 No. 6, pp. 35-49, 2015. Available at SSRN: https://ssrn.com/abstract=2710400

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

Thank You