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

# Ethics on the Internet - 2
# Ethics Canvas

What is the Internet Doing to Me

**Delaram Golpayegani**
golpayes@tcd.ie

Thanks to Prof. Dave Lewis

Reminder

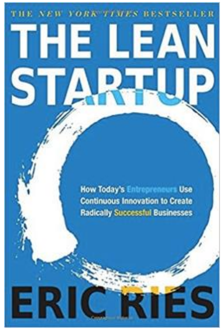
# Why Should Digital Tech Innovators be Concerned with Ethics?

- 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?**

  1. It is good for the image of your business (instrumental goal)
  2. It actually improves the service you provide! (substantive goal)
  3. It is the *good* thing to do, it contributes to your idea of a better society and being a good person (normative goal)
  4. **Law** requires it.

# Data Hungry Innovation - “Silicon Valley” Methods

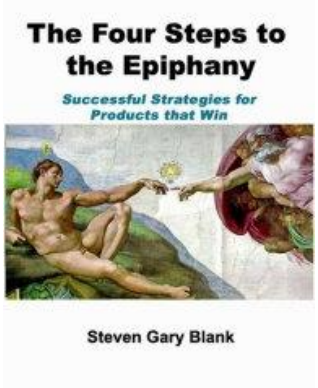


THE NEW YORK TIMES BESTSELLER

THE LEAN STARTUP

How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses

ERIC RIES

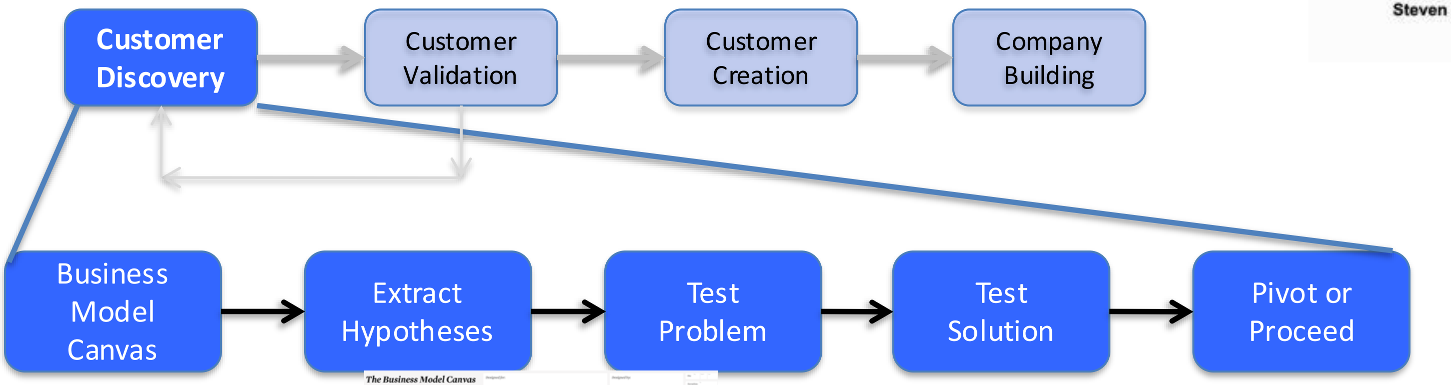


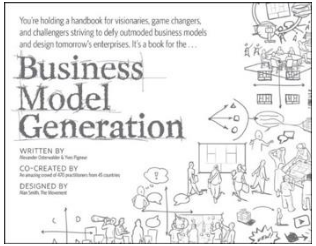
The Four Steps to the Epiphany

Successful Strategies for Products that Win

Steven Gary Blank

## The Customer Development Process





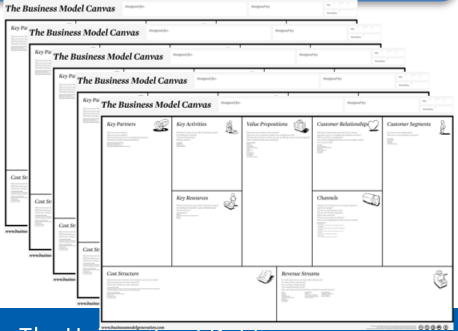
You're holding a handbook for visionaries, game changers, and challengers striving to defy outdated business models and design tomorrow's enterprises. It's a book for the ...

Business Model Generation

WRITTEN BY
Alexander Osterwalder & Yves Pigneur

CO-CREATED BY
An amazing crowd of 470 practitioners from 45 countries

DESIGNED BY
Alan Smith, The Movement



How to make ethics part of the process?

Apply ethically-focused approaches when designing, developing, and deploying and using AI

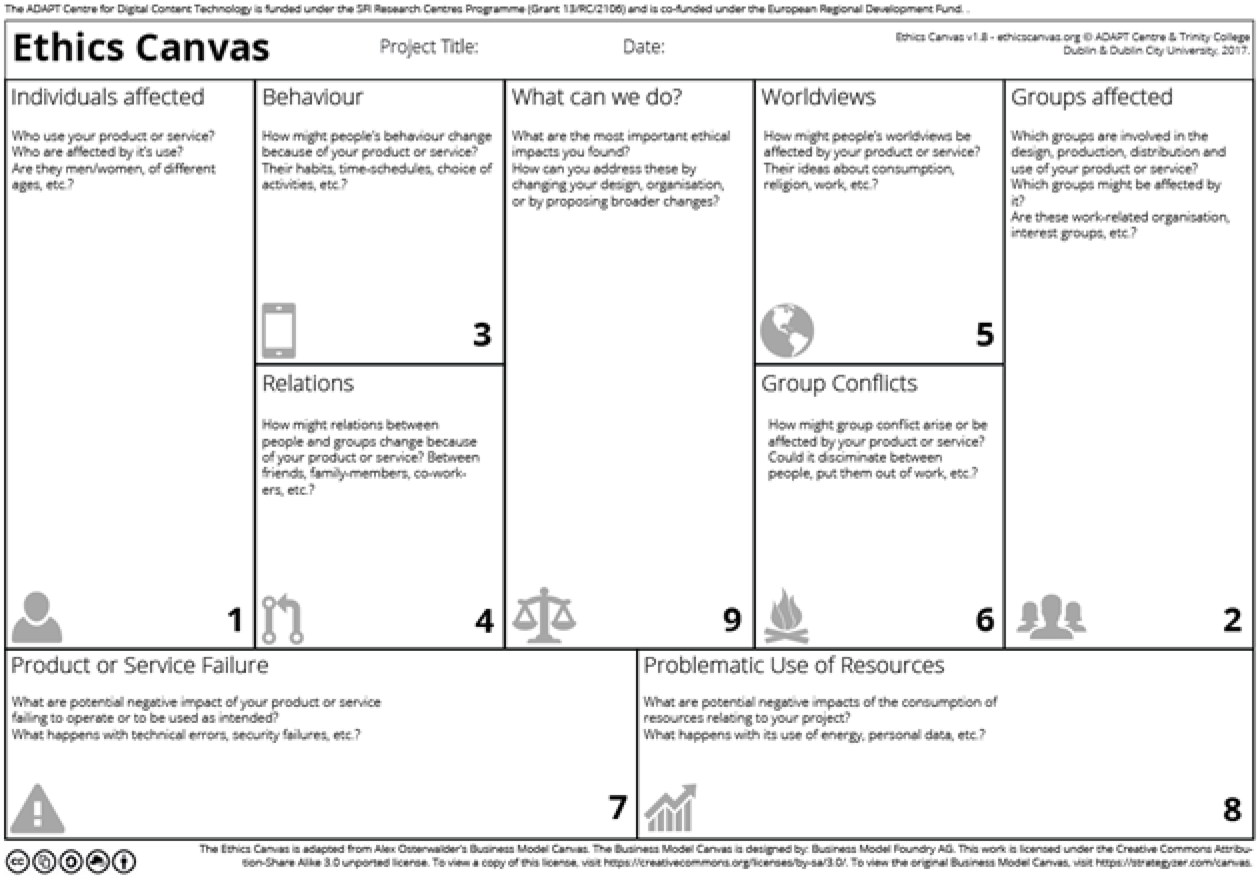
# Practicing Ethics in Responsible R&I

- 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

# The Ethics Canvas

- **Inspired by Business Model Canvas (BMC)**
- **A light-weight approach to identify, evaluate, and address ethical impacts**
- **Accessible to a wide range of stakeholders**
    - Does not need thorough background knowledge of ethical theories



The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund. .

**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

# 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

# Key Benefits of the Ethics Canvas

- 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
        - Rethink or modify the design

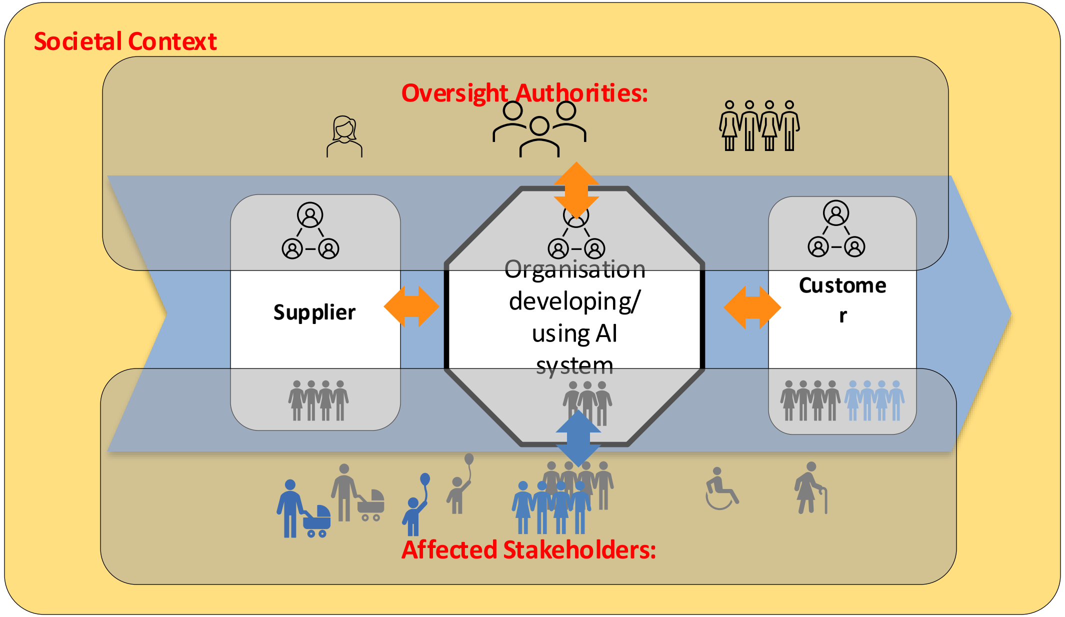
https://ethicscanvas.org

# Considerations on Ethical Impacts of Technology

- Changes in individual **behaviour**
- Relationships between **individuals**
- Relationships between **groups**
- 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)

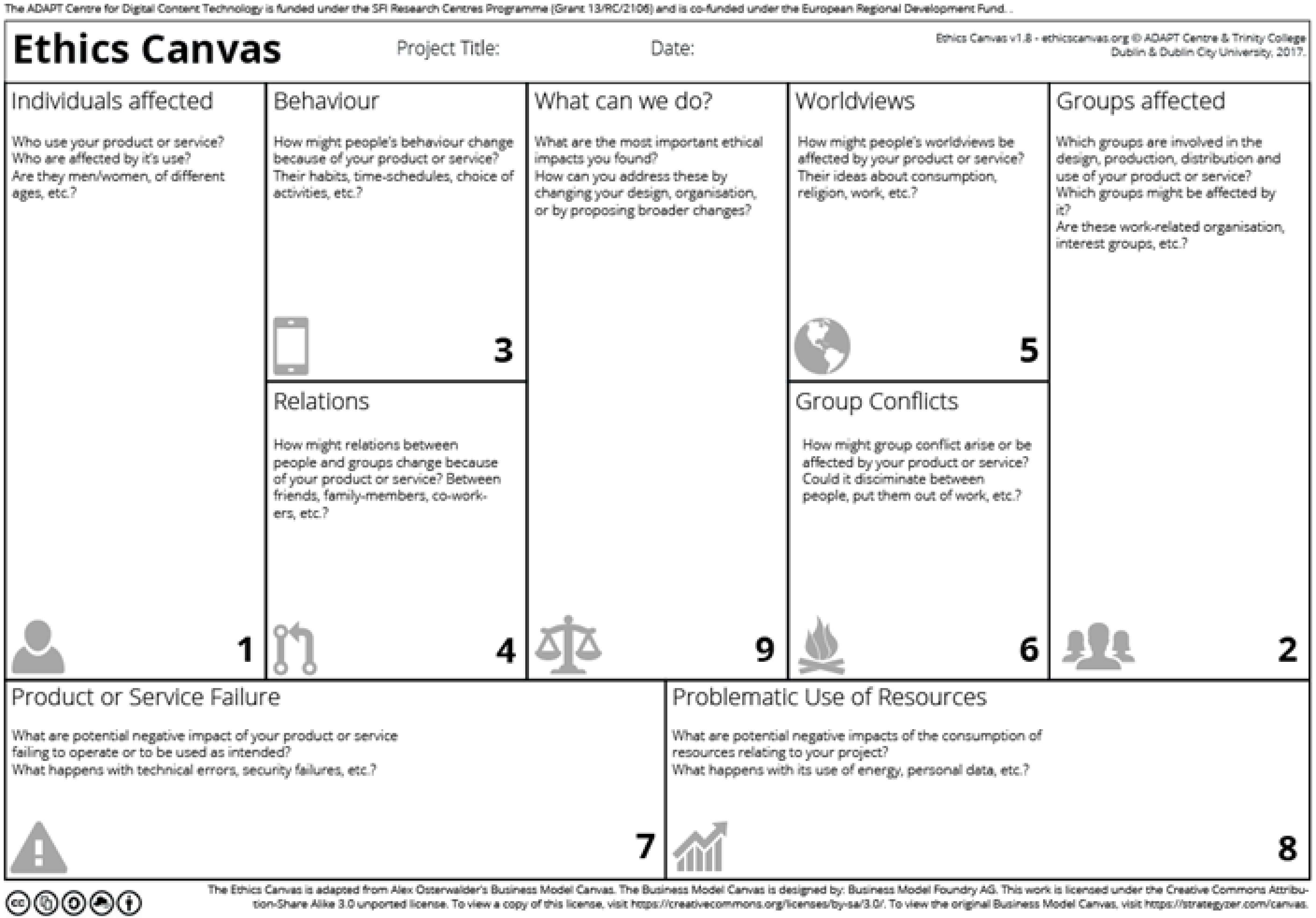
Fair Operating Procedures (suppliers, customers, regulators)

Consumer Issues (consumers)

Community Involvement and Development (local communities)

Human Rights (everyone)

Based on ISO 26000



The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund..

# 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?

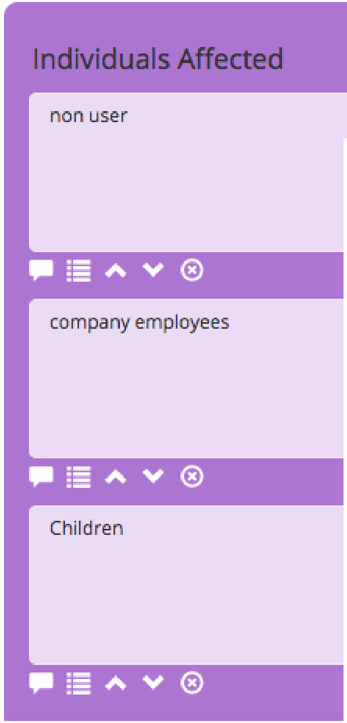
*e.g are they of different genders, of different ages, etc.?*

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

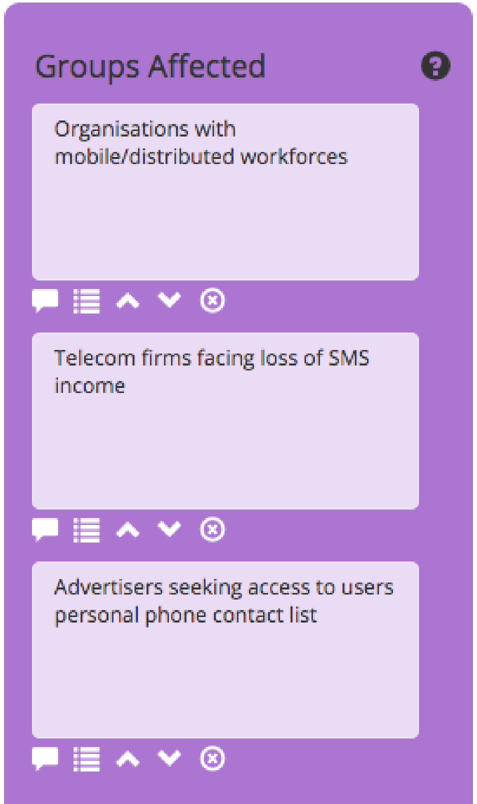


Individuals Affected

non user

company employees

Children



Groups Affected

Organisations with mobile/distributed workforces

Telecom firms facing loss of SMS income

Advertisers seeking access to users personal phone contact list

# Stage 2: Identifying Ethical Impacts

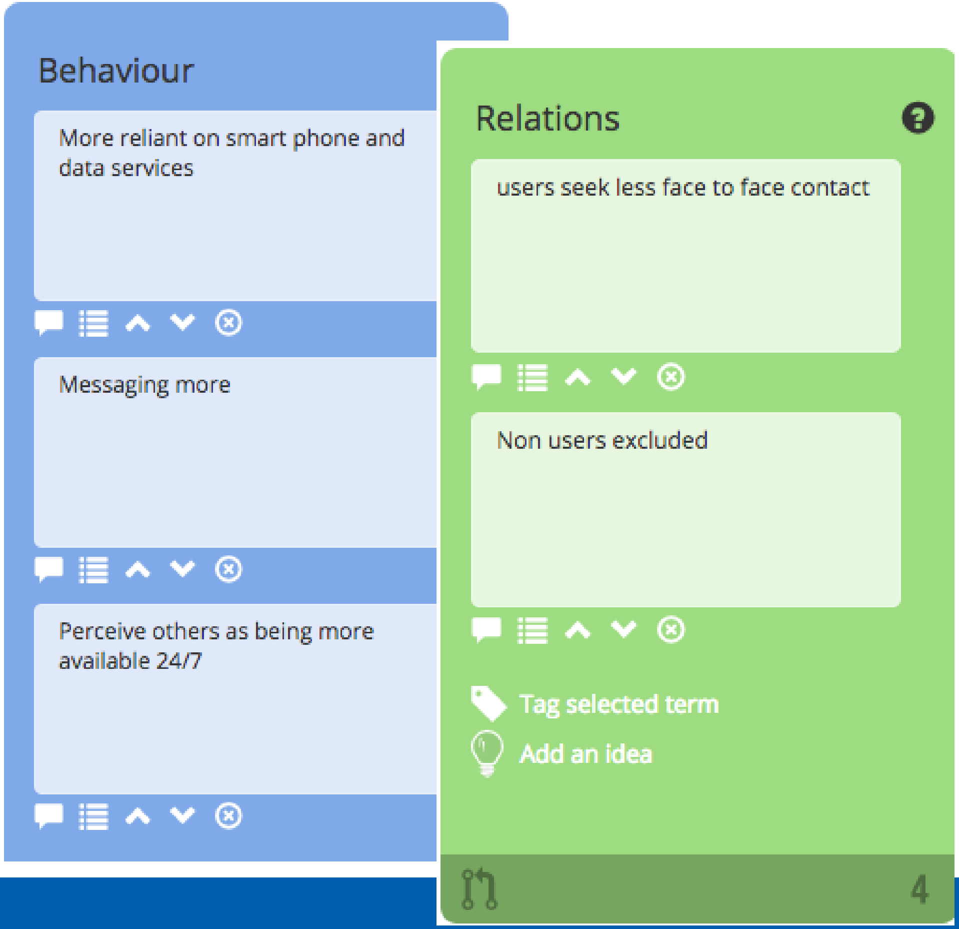
First, ‘micro’ impacts are captured by the canvas, i.e. on everyday lives of people using and living with the application

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.?*



Behaviour

More reliant on smart phone and data services

Messaging more

Perceive others as being more available 24/7

Relations

users seek less face to face contact

Non users excluded

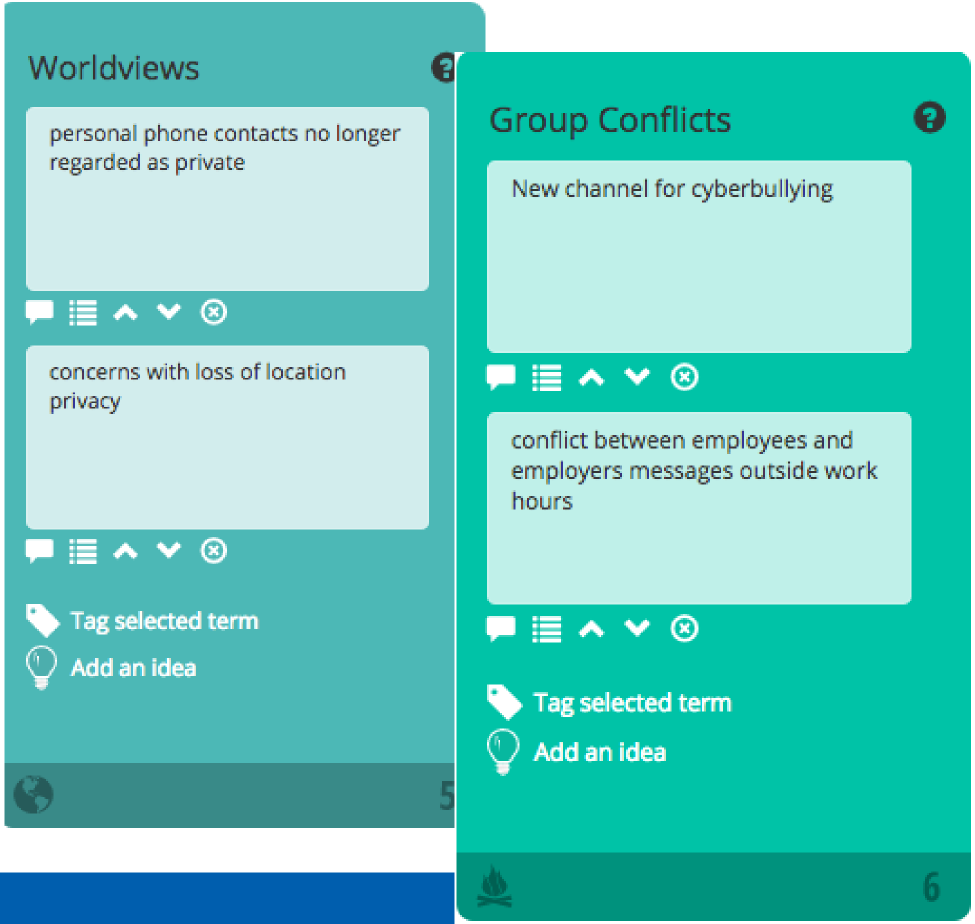
# Stage 2: Identifying Ethical Impacts

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

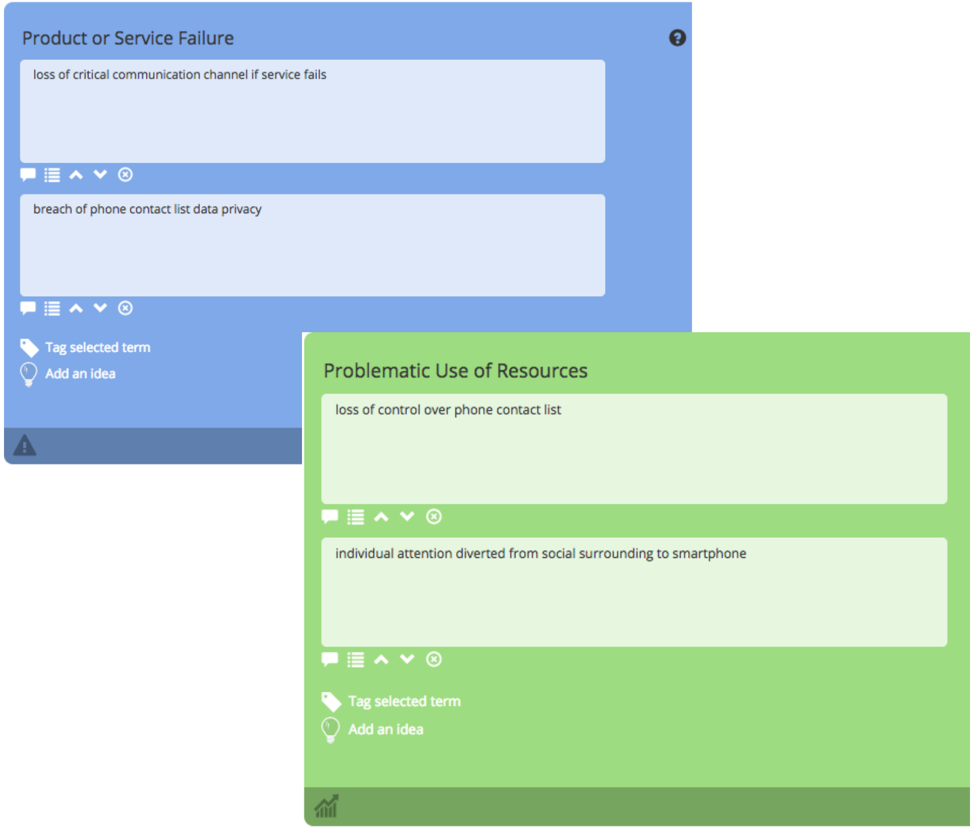
conflict between employees and employers messages outside work hours

# 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

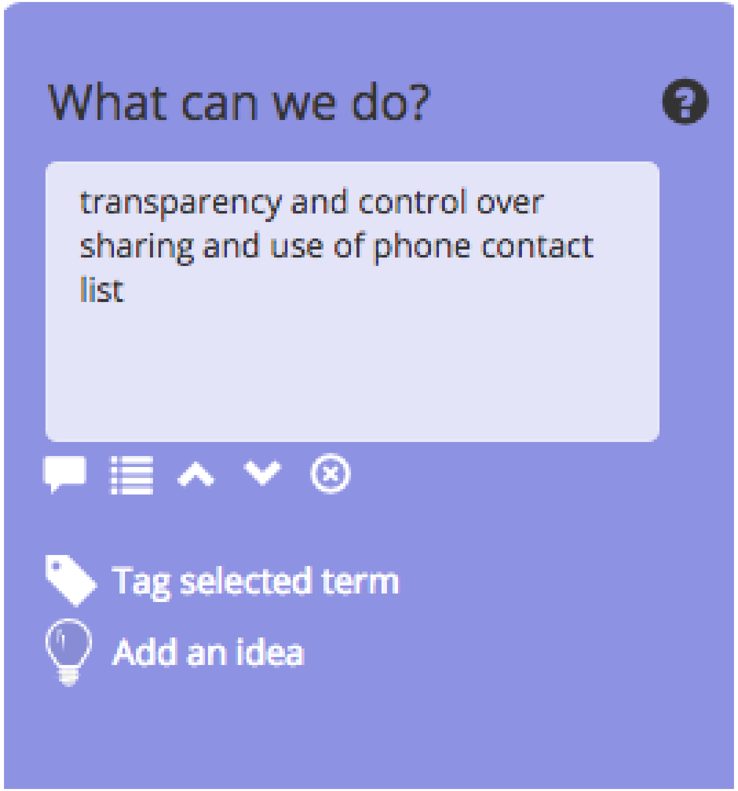
loss of control over phone contact list

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 <u>pivoting</u> your design, organisation, or by proposing broader changes?



What can we do?

transparency and control over sharing and use of phone contact list

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:
-

# Technology Impact: Example

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, with more waste

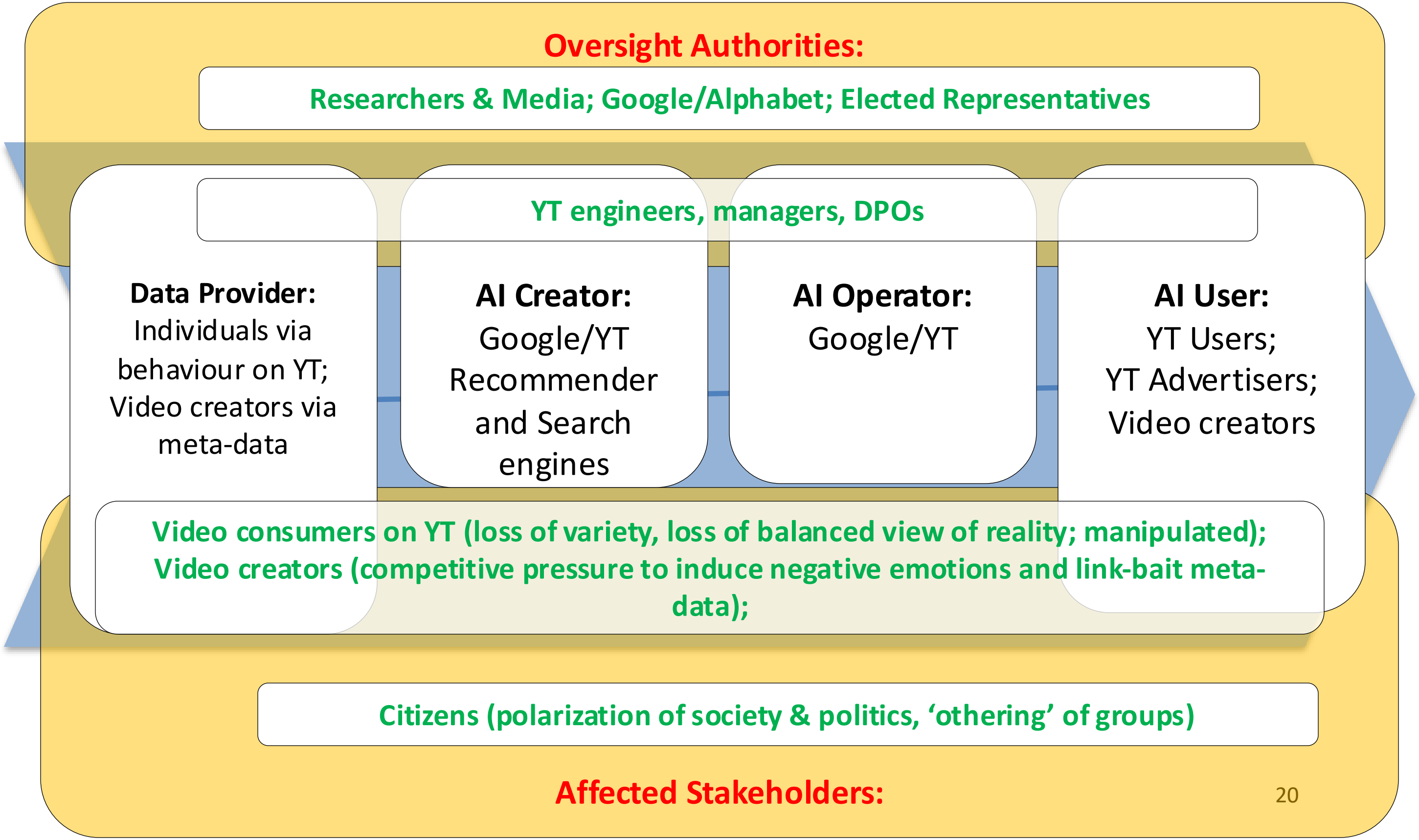
Reminder

# 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

# 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:**

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 screening of 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
- 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

Ethics Canvas
Individuals Affected
Changes in Behaviour
Social Conflicts
Group Interests
Organisations and Groups Affected
Changes in Relations
Resolving ethical impacts
Public Sphere
Products and Services provided
Impact of product or service failure
Impact of use of resources

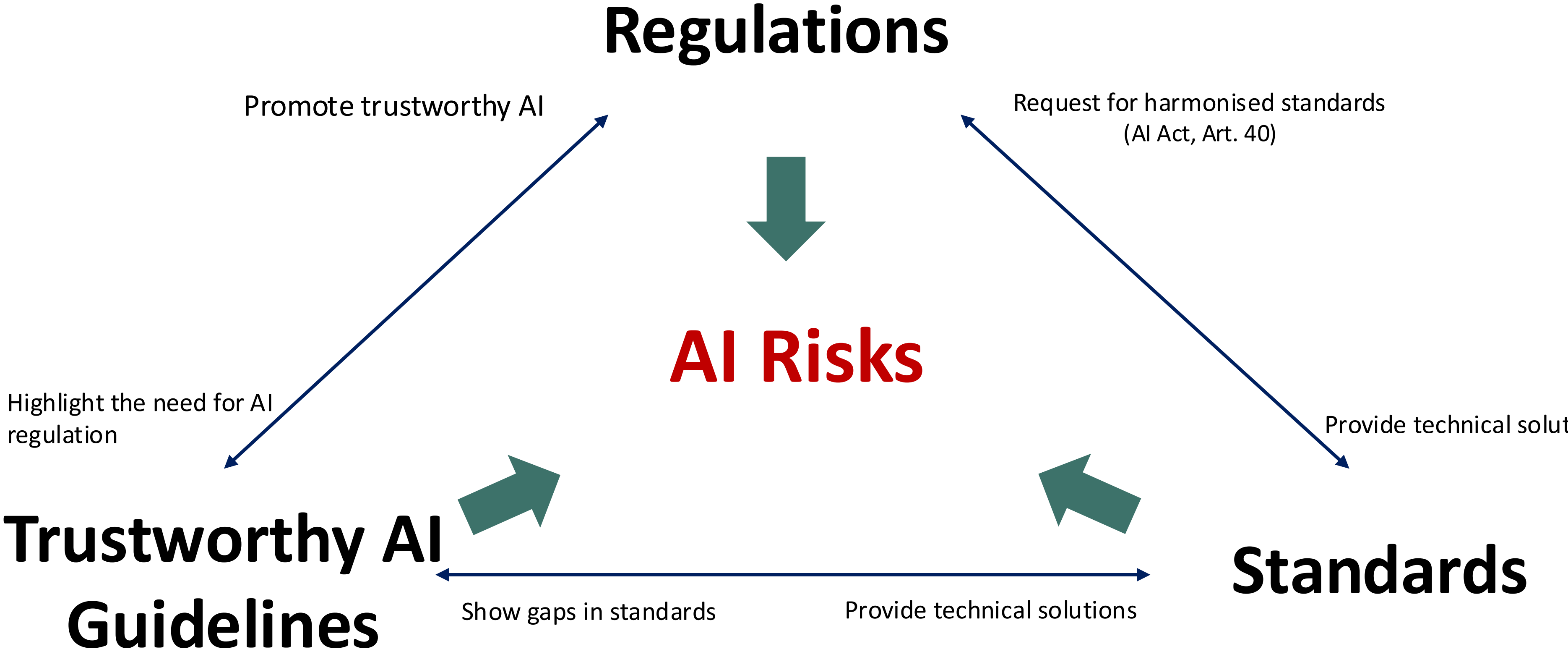
User Manual available at: https://www.ethicscanvas.org/download/handbook.pdf

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

# Ethics on the Internet - 2
# Trustworthy AI Value Chain

## What is the Internet Doing to Me

# Dealing with AI Risks



Regulations

Promote trustworthy AI

Request for harmonised standards
(AI Act, Art. 40)

AI Risks

Highlight the need for AI regulation

Provide technical solut

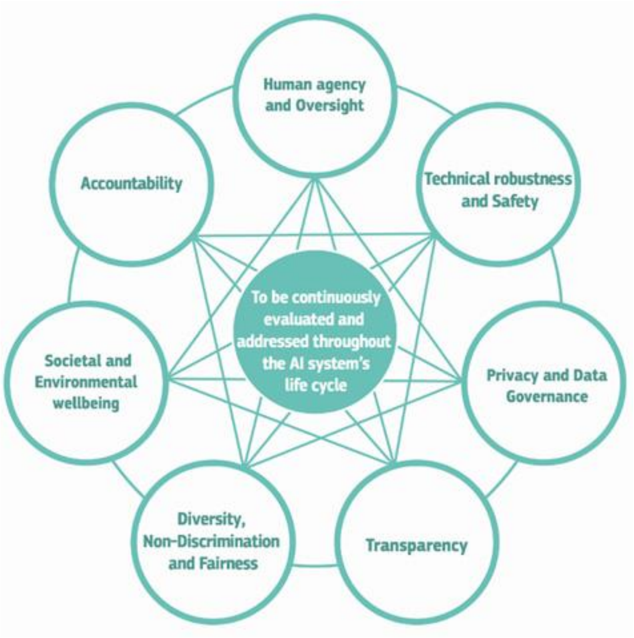
Trustworthy AI Guidelines

Standards

Show gaps in standards

Provide technical solutions

# EU Trustworthy AI Guidelines


Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

European Commission

The European Commission's

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

AI

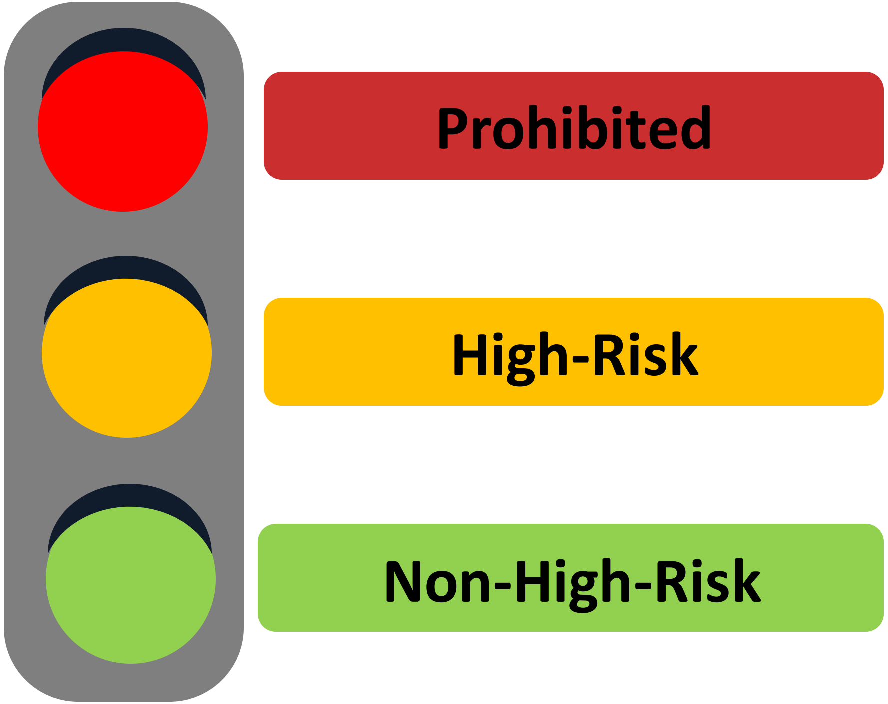
DRAFT
ETHICS GUIDELINES
FOR TRUSTWORTHY AI

Working Document for stakeholders' consultation

Brussels, 18 December 2018

https://data.europa.eu/doi/10.2759/346720

# The EU AI Act



Prohibited

High-Risk

Non-High-Risk

Promotes human-centric & trustworthy AI

**Protects against harmful effects of AI on**

- **Health**
- **Safety**
- **Fundamental Rights**

# Can the AI Act deliver Ethical AI? ……
# Not without Standards

- Existing regulation is referenced that has well established risk and quality models for health and safety
- No direct guidance on how to **protect fundamental rights** – Act references 'harmonized standards'
- Harmonized standards are international standards approved through consensus of National Standards Bodies, e.g. National Standards Authority of Ireland and approved by European Commission

standardisation is arguably where the real rulemaking in the AI Act will occur

Demystifying the Draft EU Artificial Intelligence Act, M.Veale, F.Z.Borgesius Computer Law Review International (2021), 22(4) 97-112, **https://doi.org/10.48550/arXiv.2107.03721**

# AI Standardisation

- ISO/IEC JTC1 is the global consensus forming body for ICT standards
  - Subcommittee 42 established in 2017 to develop AI standards

- CEN/CENELEC is the consensus forming body for standards in Europe
  - Joint Technical Committee 21 on AI established in 2021

- National Standards Authority of Ireland (NSAI)

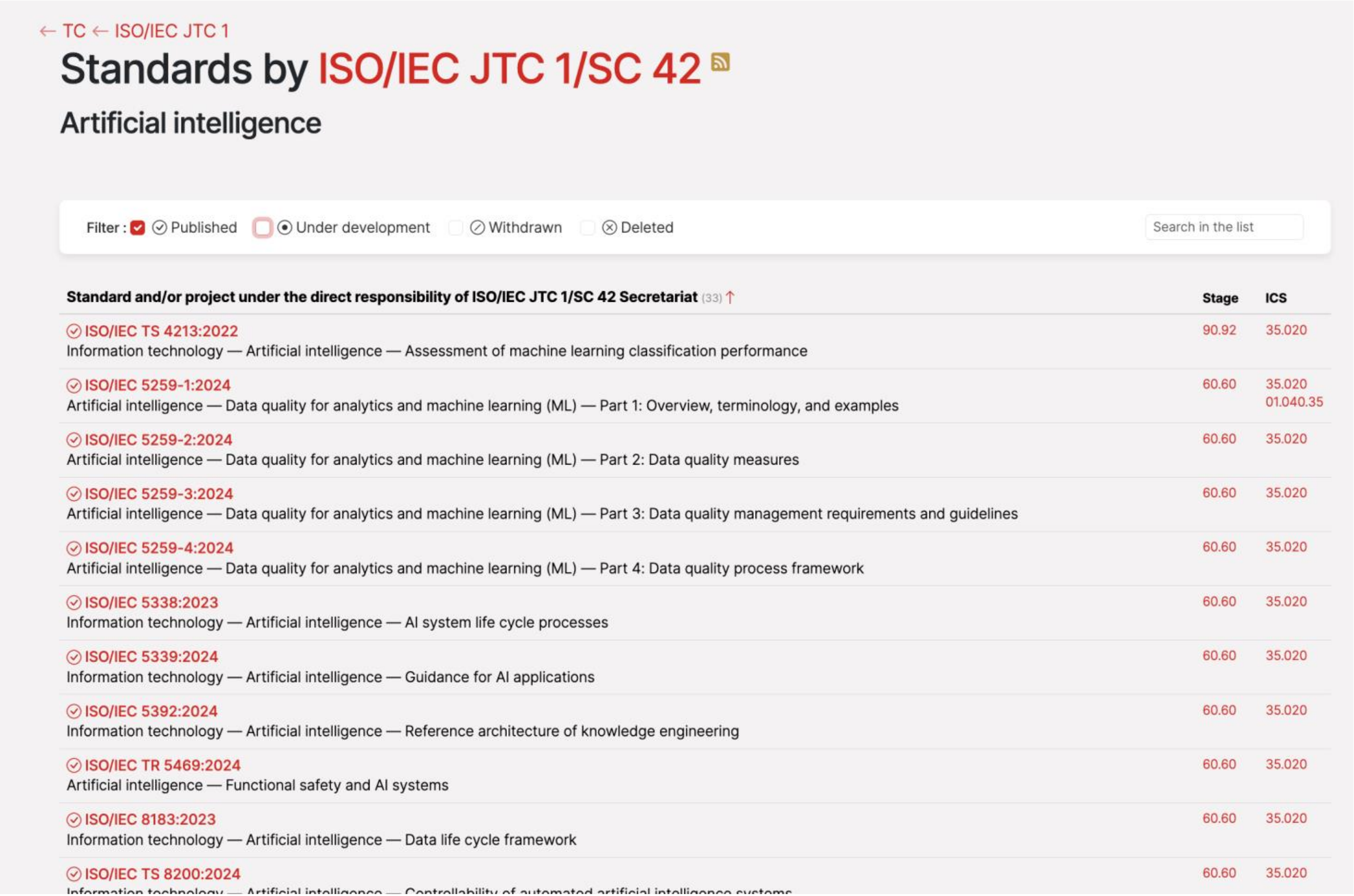
# Can International Standard Guide Ethical AI?

- **SC42 follow established model of identifying specific consideration (for AI) within existing standards**
    - Management System, Risk Management, Quality Management
    - Organisation and data governance
- **AI-specific standards identify types of technical metrics that can be used:**
    - Bias
    - Testing of Neural Networks

# Trustworthy AI Standards: Some Key Concepts

ISO JTC1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

- Trustworthiness: ability to meet stakeholder’s expectations in a verifiable way [JTC1 AG]
- Stakeholder: any individual, group, or organization that can affect, be affected by, or perceive itself to be affected by a decision or activity [ISO/IEC 38500:2015]
- Accountable: answerable for actions, decisions, and performance [ISO 31000:2018]
- Risk: effect of uncertainty on objectives [ISO 31000:2018]
- Control: measure that maintains and/or modifies *risk* [ISO 31000:2018]
- Bias: favouritism towards some things, people, or groups over others



← TC ← ISO/IEC JTC 1

# Standards by ISO/IEC JTC 1/SC 42

## Artificial intelligence

Filter : ☑ Published □ Under development □ Withdrawn □ Deleted     Search in the list

| Standard and/or project under the direct responsibility of ISO/IEC JTC 1/SC 42 Secretariat (33) ↑ | Stage | ICS |
|---|---|---|
| ISO/IEC TS 4213:2022<br>Information technology — Artificial intelligence — Assessment of machine learning classification performance | 90.92 | 35.020 |
| ISO/IEC 5259-1:2024<br>Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples | 60.60 | 35.020<br>01.040.35 |
| ISO/IEC 5259-2:2024<br>Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 2: Data quality measures | 60.60 | 35.020 |
| ISO/IEC 5259-3:2024<br>Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 3: Data quality management requirements and guidelines | 60.60 | 35.020 |
| ISO/IEC 5259-4:2024<br>Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 4: Data quality process framework | 60.60 | 35.020 |
| ISO/IEC 5338:2023<br>Information technology — Artificial intelligence — AI system life cycle processes | 60.60 | 35.020 |
| ISO/IEC 5339:2024<br>Information technology — Artificial intelligence — Guidance for AI applications | 60.60 | 35.020 |
| ISO/IEC 5392:2024<br>Information technology — Artificial intelligence — Reference architecture of knowledge engineering | 60.60 | 35.020 |
| ISO/IEC TR 5469:2024<br>Artificial intelligence — Functional safety and AI systems | 60.60 | 35.020 |
| ISO/IEC 8183:2023<br>Information technology — Artificial intelligence — Data life cycle framework | 60.60 | 35.020 |
| ISO/IEC TS 8200:2024<br>Information technology — Artificial intelligence — Controllability of automated artificial intelligence systems | 60.60 | 35.020 |

https://www.iso.org/committee/6794475/x/catalogue/p/1/u/0/w/0/d/0

# Can International Standard Guide Ethical AI?

- Standards should not and will not resolve consensus on disputed concepts which often frame ethical issues

- Example: should ‘fairness’ in allocating education or healthcare resources be based on:

    1. Sameness/equality?

    2. Deservedness/meritocracy?

    3. Need?

- Such societal-level disputes must be resolved through political processes, not by technical experts employed by large companies

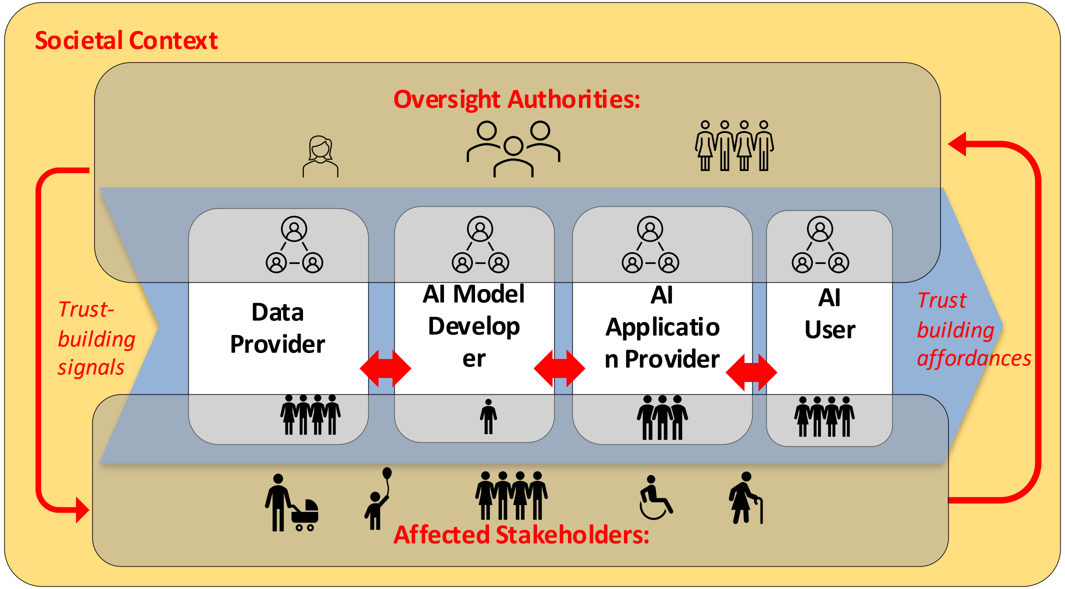
- Standards may be able. To provide ‘knobs and levers’, e.g. definition of tests for bias

- It is a societal responsibility to define acceptable levels of risk

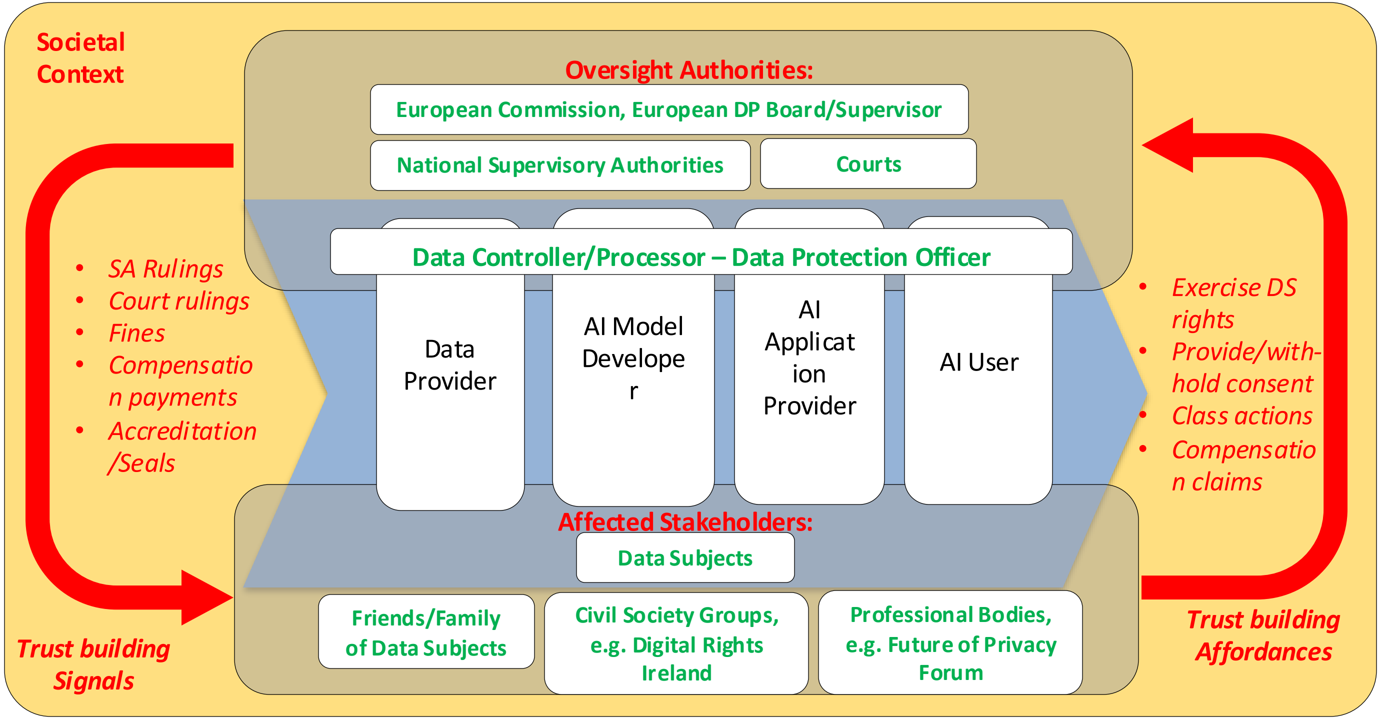
    - E.g. risk of mis-recognizing speech from those with less common accents

    - Same for education, ambulance dispatch, asylum?

# 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



Societal Context
Oversight Authorities:
European Commission, European DP Board/Supervisor
National Supervisory Authorities
Courts
Data Controller/Processor – Data Protection Officer
Data Provider
AI Model Develope r
AI Applicat ion Provider
AI User
Affected Stakeholders:
Data Subjects
Friends/Family of Data Subjects
Civil Society Groups, e.g. Digital Rights Ireland
Professional Bodies, e.g. Future of Privacy Forum
SA Rulings
Court rulings
Fines
Compensatio n payments
Accreditation /Seals
Trust building Signals
Exercise DS rights
Provide/with-hold consent
Class actions
Compensatio n claims
Trust building Affordances

# Example: Gender Bias in Google Translate

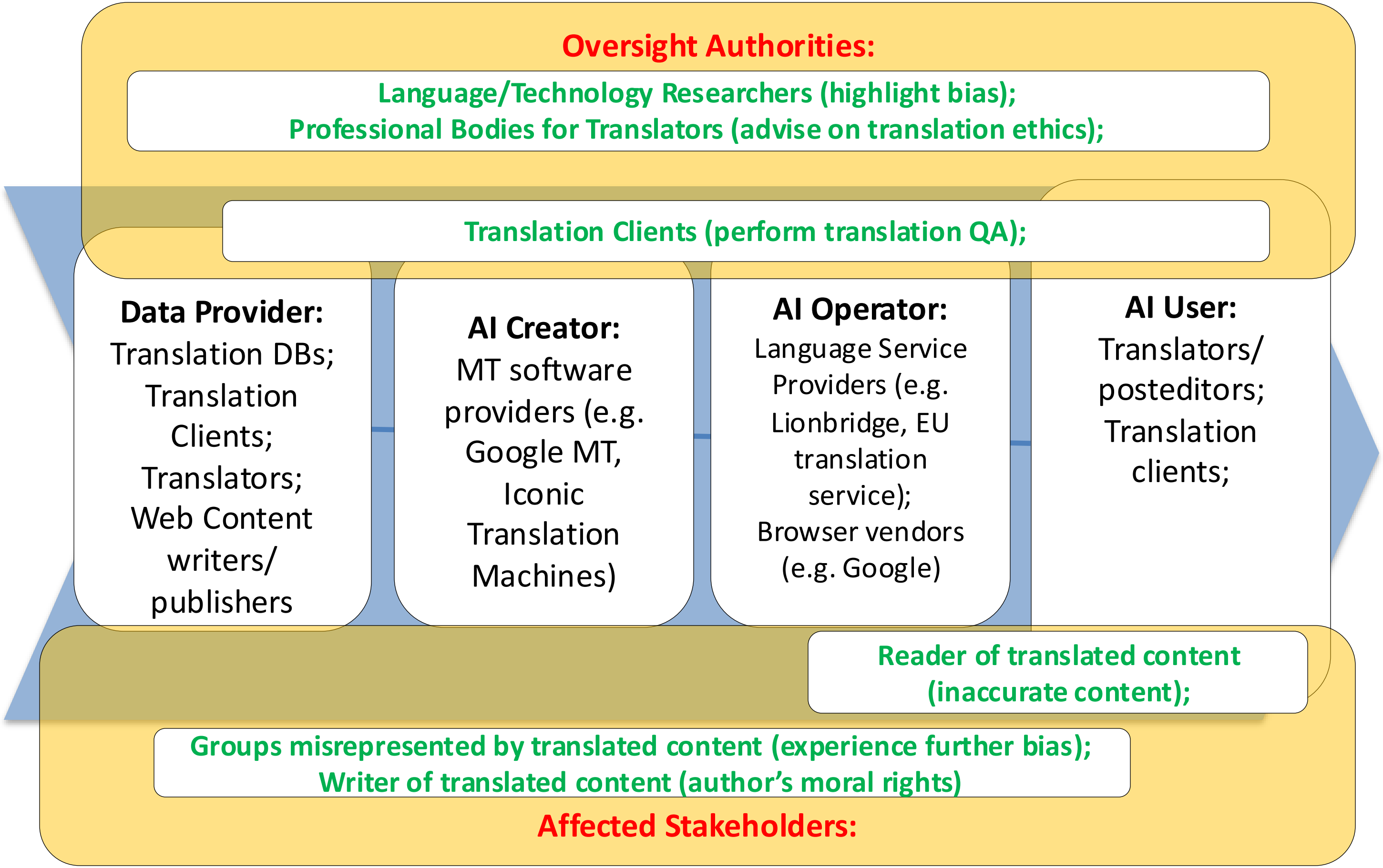
- **Some languages, like Turkish, don’t have gender specific pronouns**
- **Google translate has to guess the gender when translating in English**
- **Statements allocating gender to role reveal gender bias**

**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/

# 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)

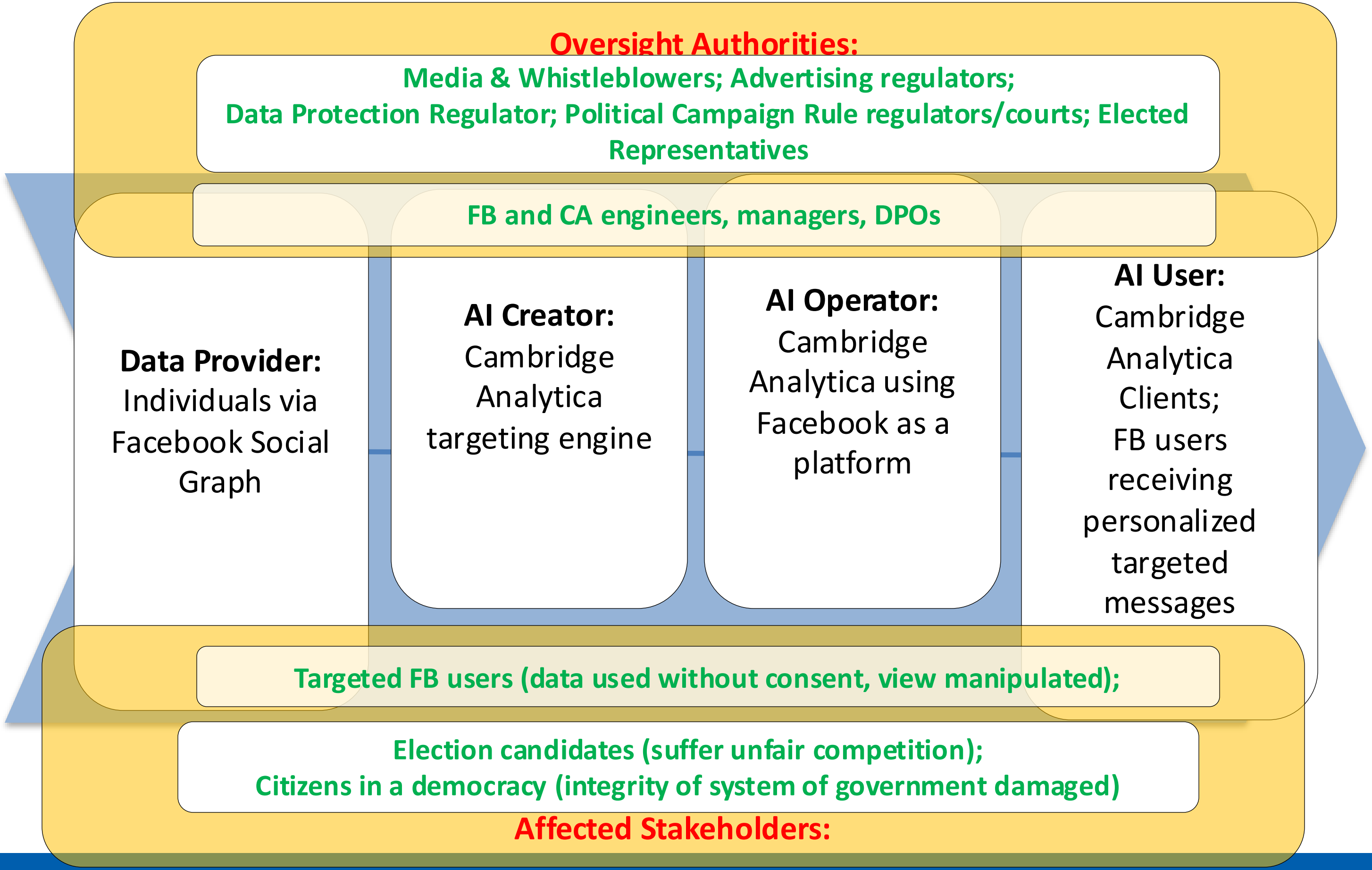
**Affected Stakeholders:**

# 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-faceook-nix-bannon-trump

# 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

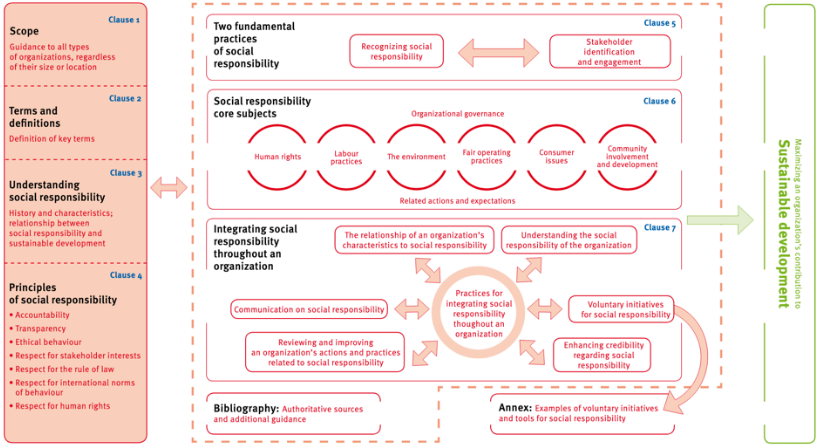
Targeted FB users (data used without consent, view manipulated);

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

**Affected Stakeholders:**

# Social Responsibility for AI

- **Ethical and Societal Issues:**
    - ISO need international consensus BUT avoids importing specific value-sets
    - Needs principles, which ones?
- **ISO already has non-ICT specific principles: ISO 26000 – Social Responsibility**
- **Stakeholder identification and engagement is key**


Clause 1
Scope
Guidance to all types of organizations, regardless of their size or location
Clause 2
Terms and definitions
Definition of key terms
Clause 3
Understanding social responsibility
History and characteristics; relationship between social responsibility and sustainable development
Clause 4
Principles of social responsibility
• Accountability
• Transparency
• Ethical behaviour
• Respect for stakeholder interests
• Respect for the rule of law
• Respect for international norms of behaviour
• Respect for human rights
Clause 5
Two fundamental practices of social responsibility
Recognizing social responsibility
Stakeholder identification and engagement
Clause 6
Social responsibility core subjects
Organizational governance
Human rights
Labour practices
The environment
Fair operating practices
Consumer issues
Community involvement and development
Related actions and expectations
Clause 7
Integrating social responsibility throughout an organization
The relationship of an organization's characteristics to social responsibility
Understanding the social responsibility of the organization
Practices for integrating social responsibility thoughout an organization
Communication on social responsibility
Voluntary initiatives for social responsibility
Reviewing and improving an organization's actions and practices related to social responsibility
Enhancing credibility regarding social responsibility
Bibliography: Authoritative sources and additional guidance
Annex: Examples of voluntary initiatives and tools for social responsibility
Sustainable development
Maximizing an organization's contribution to

ISO 26000 structure

# What is Social Responsibility?

``Responsibility of an organization for the **impacts** of its decisions and activities on society and the environment, through transparent and ethical behaviour that

– contributes to sustainable development including health and the welfare of society;

– takes into account the expectations of stakeholders;

– is in compliance with applicable law and consistent with international norms of behaviour; and

– is integrated throughout the organization and practised in its relationships

# Social Responsibility in ISO 26000

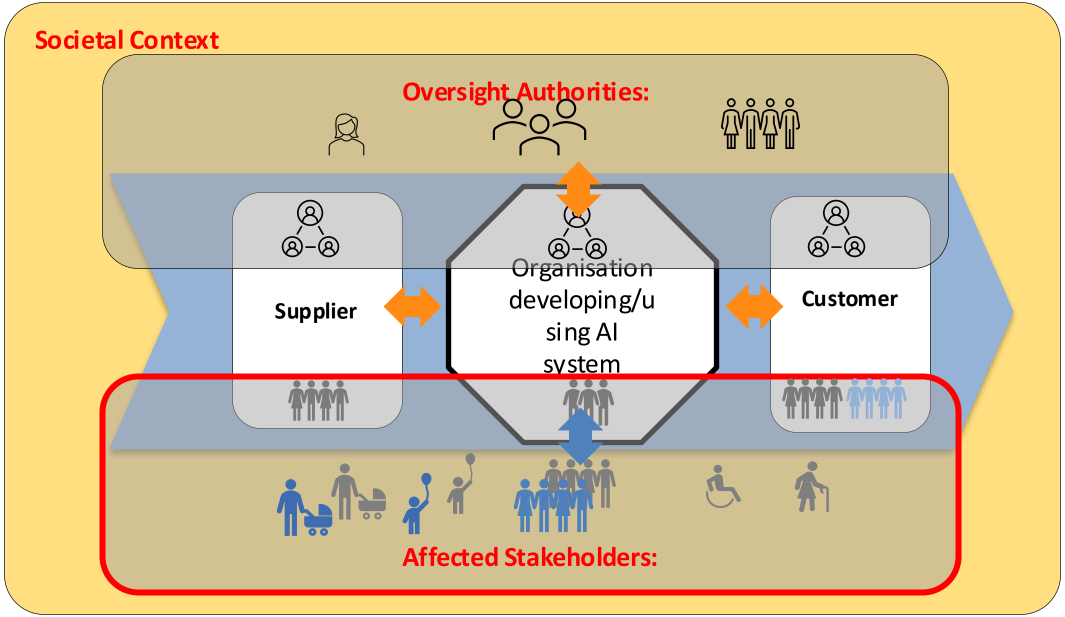
### Principles of social responsibility

- Accountability
- Transparency
- Ethical behavior
- Respect for stakeholder interests
- Respect for the rule of law
- Respect for international norms of behaviour
- Respect for human rights

### Social Responsibility Core Subjects

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

# 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

# **Human Rights** issues for Social Responsibility

## Risks

- **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

## Mitigations

- *Due diligence*: Human rights policy , Fundamental Rights Impact Assessment
- *Avoid* value chain partners that may commit violations
- Establish *grievance and redress mechanism*: transparent, accessible, external scrutiny, AI explanations
- *Monitor for discrimination* towards vulnerable groups in AI decision making, e.g. insurance, justice, recruiting
- *Education and access* for all groups to benefits of AI
- Ensure worker *freedom of association* and collective bargaining

UN human rights: https://www.un.org/en/about-us/universal-declaration-of-human-rights

EU fundamental rights: https://commission.europa.eu/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights_en

# **Labour Practice** issues for Social Responsibility

## Risks

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

## Mitigations

- *Recognition* of secure employment & decent working conditions
- Engage in *social dialo*gue with worker 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

# **Environment** issues for Social Responsibility

### Risks

- 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

### Mitigations

- Monitor and plan reduction of non-sustainable energy and resource use over whole product lifecycle
- Make AI services available for environmental monitoring and analysis

# **Fair Operating Procedure** issues for Social Responsibility

### Risks

- 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

### Mitigations

- 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

# **Consumer** Issues in Social Responsibility

### Risks

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

### Mitigations

- 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. transparency 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

# **Community Involvement and Development** issues in Social Responsibility

## Risks

- 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

## Mitigations

- *Consult* with early and widely 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

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

# Thank You