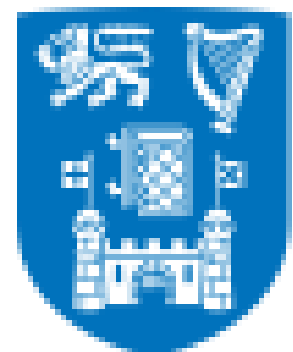


Human Factors & Ethics Canvas (Responsible AI & Addressing The Triple Bottom Line)

Dr Joan Cahill, CIHS, School of Psychology, TCD.

September 2022.



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

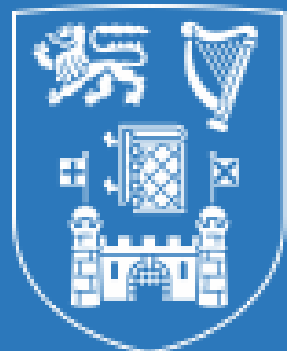


PEOPLE
PROCESS
PERFORMANCE

Centre for Innovative Human Systems

Introduction

- **Dr Joan Cahill**
- Centre for Innovative Human Systems (CIHS), School of Psychology, Trinity College Dublin
- Human Factors, Ethics, Behaviour Science
- Technology-based supports and interventions in Healthcare, Aviation, Road Transport & Financial Services



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

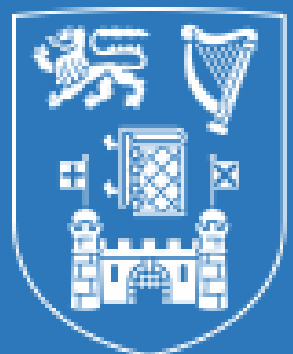
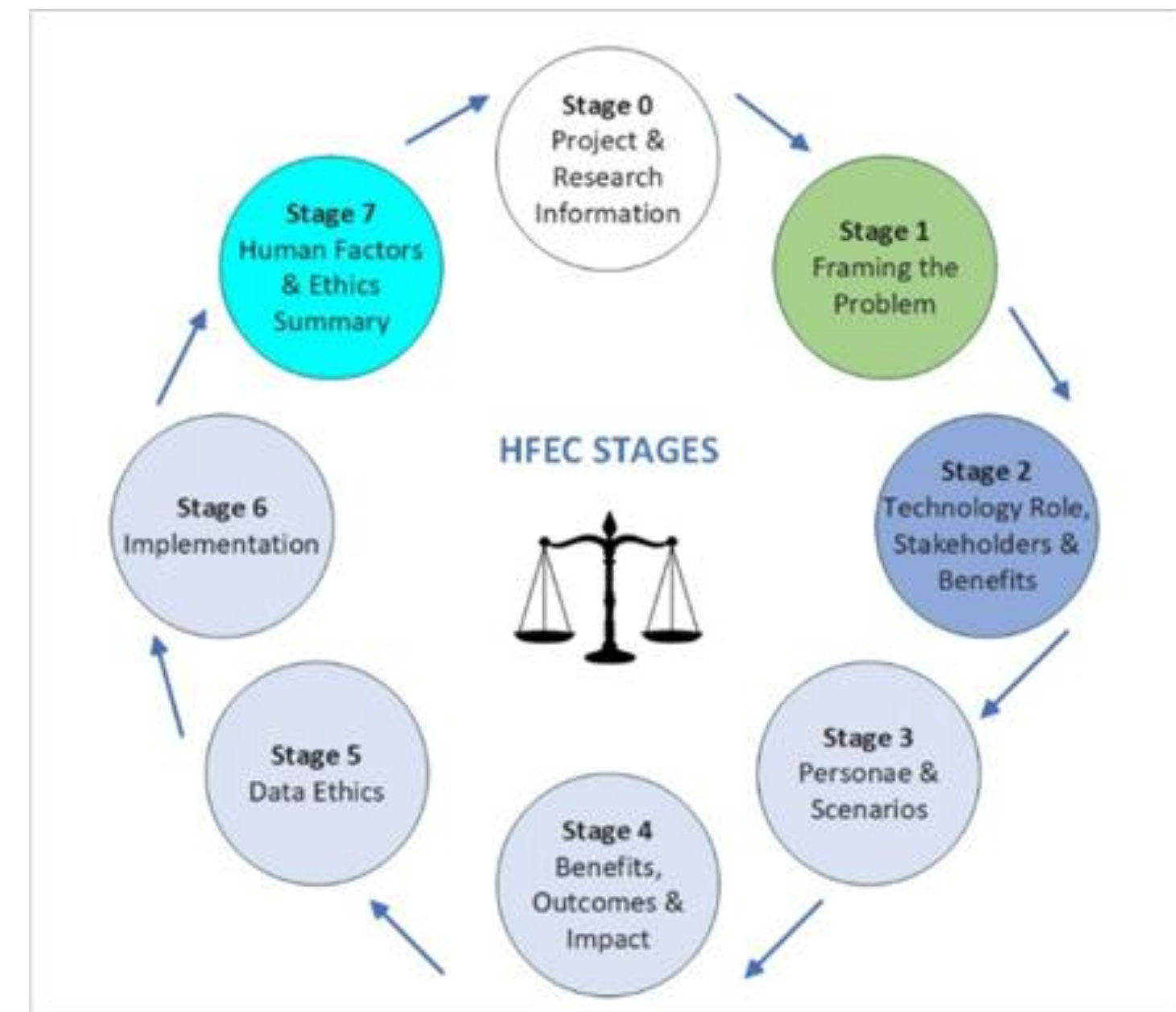


PEOPLE
PROCESS
PERFORMANCE

Topic

- Methodology to Address Advancement of Responsible Technology
- Human Factors & Ethics Canvas (HFEC)
- Ethics of technological innovation and its impacts at different levels

Impact | Uncertainty | Triple Bottom Line | Human Value | Wellbeing



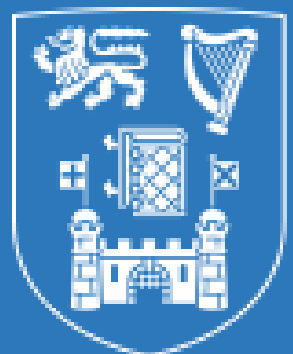
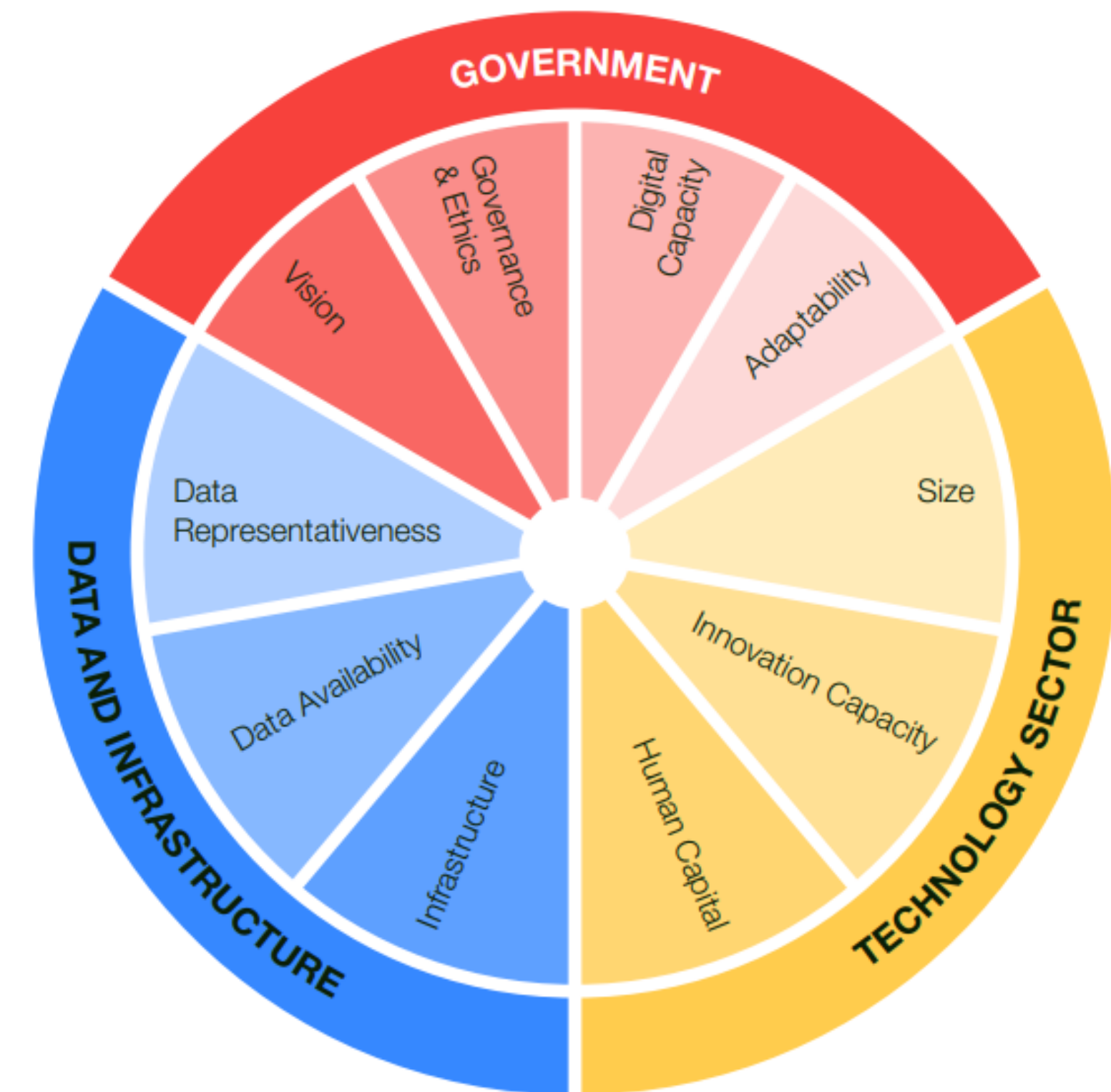
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



PEOPLE
PROCESS
PERFORMANCE

AI Readiness Index 2021

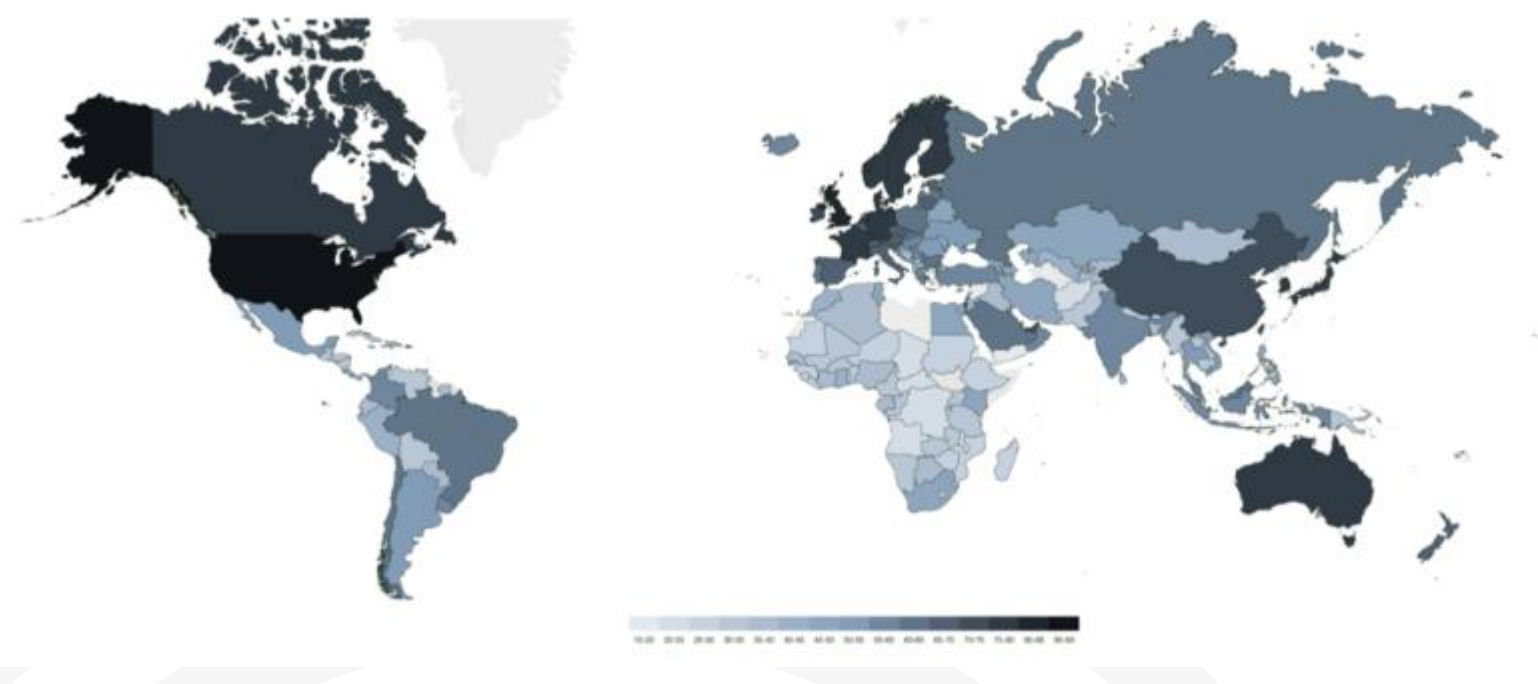
- Government Artificial Intelligence Readiness Index 2021, Oxford Insights
- Ireland ranks 17 globally (160 countries) – index score of 82.59
- Regional rank 10 (20 countries in Western Europe)
- The inclusion of ethical frameworks in a nation's AI development increases its score, as it is an important factor in promoting the responsible use of AI



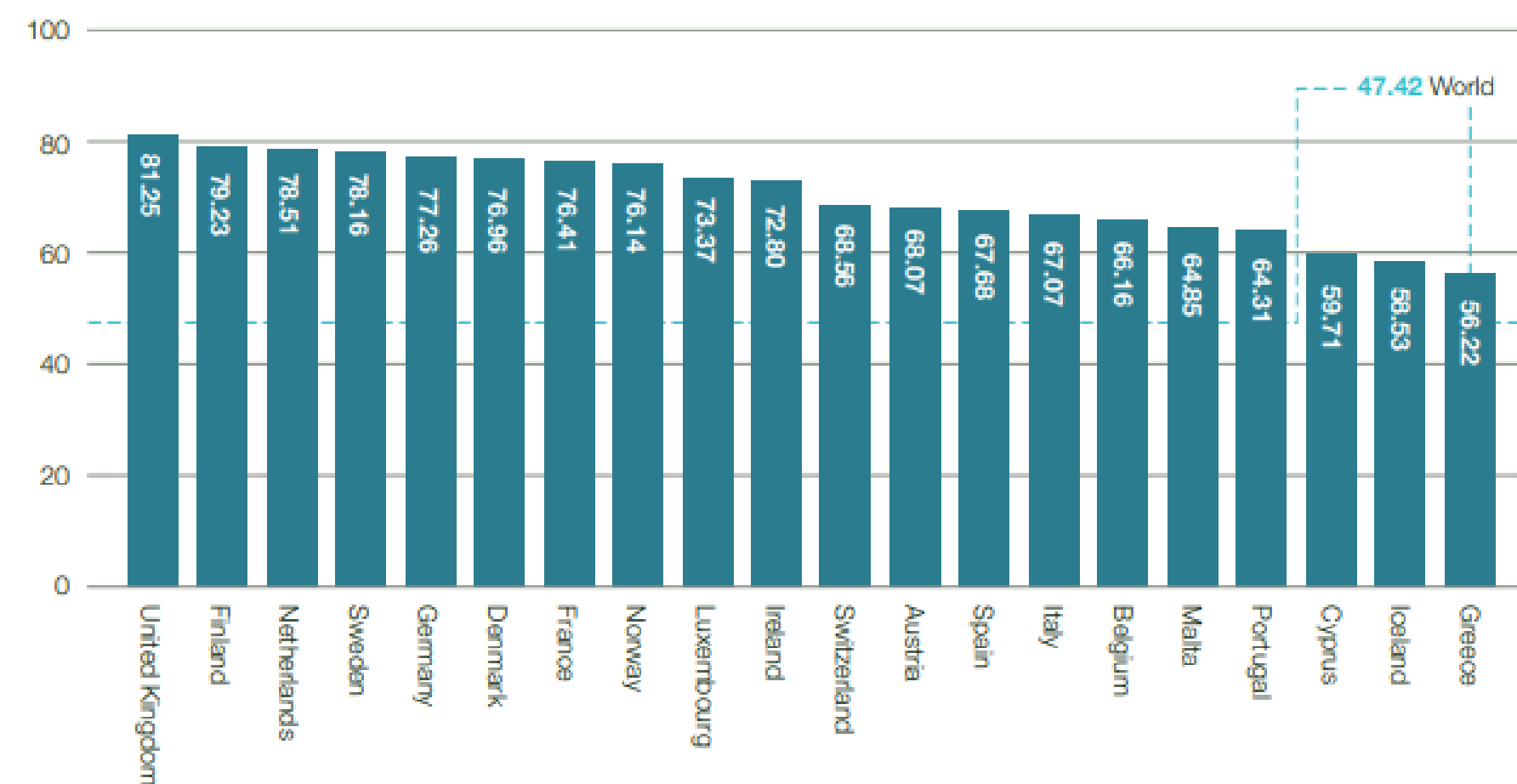
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



PEOPLE
PROCESS
PERFORMANCE

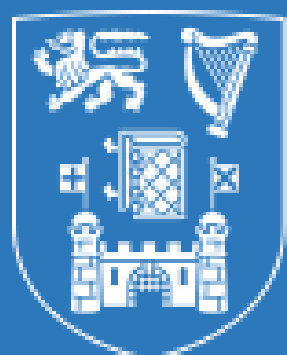


Western Europe



Global Ranking

Global Position	Country	Overall Score	Government	Technology Sector	Data and Infrastructure
1	United States of America	88.16	88.46	83.31	92.71
2	Singapore	82.46	94.88	66.69	85.80
3	United Kingdom	81.25	85.69	67.26	90.81
4	Finland	79.23	88.45	63.85	85.40
5	Netherlands	78.51	80.42	66.17	88.92
6	Sweden	78.16	80.76	67.37	86.36
7	Canada	77.73	84.36	63.75	85.08
8	Germany	77.26	78.04	67.68	86.07
9	Denmark	76.96	83.50	63.24	84.14
10	Republic of Korea	76.55	85.27	58.49	85.89
11	France	76.41	82.10	60.61	86.53
12	Japan	76.18	81.90	59.31	87.32
13	Norway	76.14	84.24	59.25	84.91
14	Australia	75.41	83.79	57.07	85.37
15	China	74.42	83.79	61.33	78.15
16	Luxembourg	73.37	82.67	50.66	86.80
17	Ireland	72.80	74.70	61.11	82.59
18	Taiwan	71.98	77.59	59.42	78.92



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



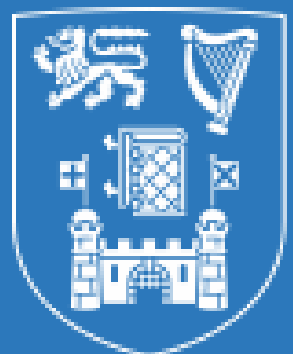
PEOPLE
PROCESS
PERFORMANCE

AI - Here for Good (2021)



People-centred approach
Ethical approach
AI used for societal good and sustainability

Manage pressures/conflicts between business demand, regulation, societal needs and values?
Manage uncertainty?
Make informed decisions and deliver benefits across the TBL?



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



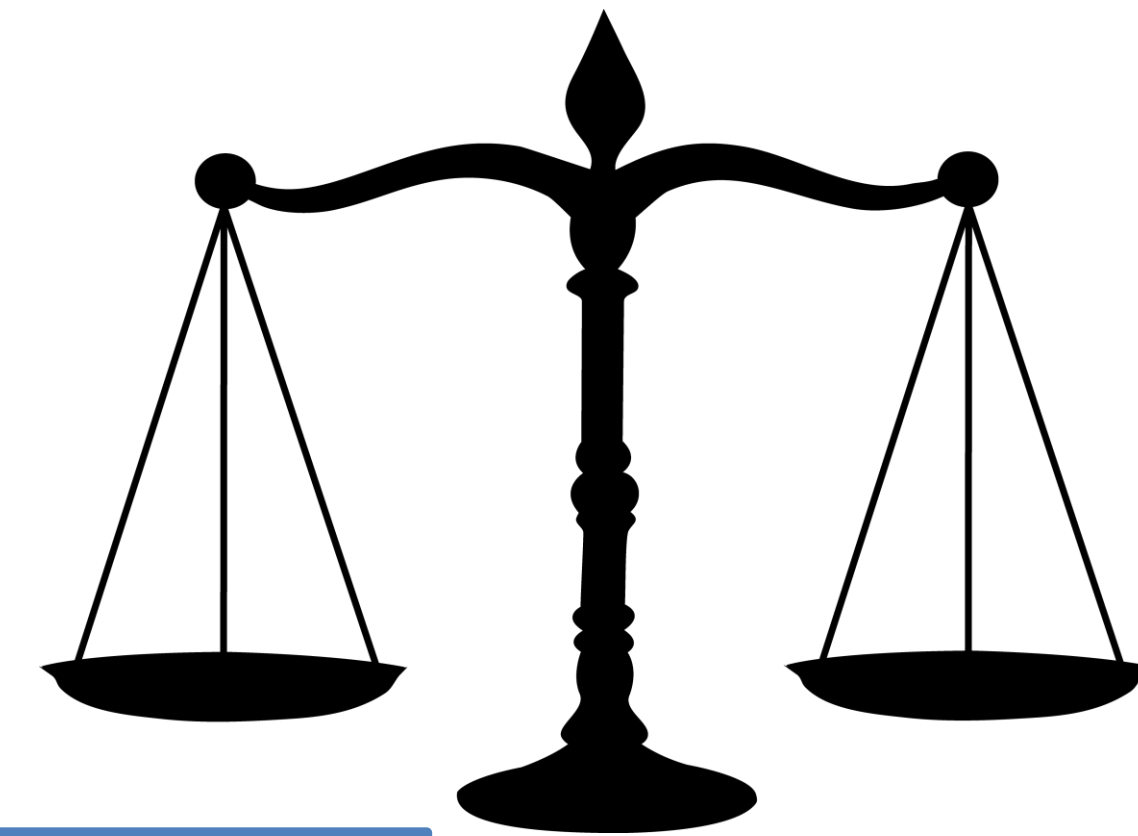
PEOPLE
PROCESS
PERFORMANCE

Human Factors & Ethics Canvas (HFEC)

- The objective of the ‘human factors and ethics’ canvas, is to create an **evidence map in relation to the specification of an ethically responsible technology solution** that properly addresses relevant human, social and ethical issues.
- Ontological design
- Ethics by design practices
- Value-based system design methodology
- Integrating ethical value requirements (EVR) and priorities into system design (Ethically Aligned Design, IEEE)
- Ethical risk review
- Impact assessment

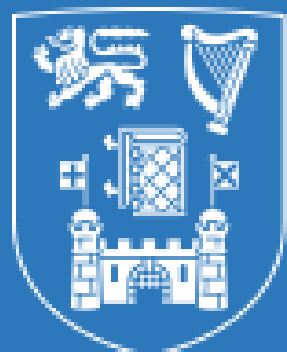
Ethics

Human Factors



Design

Impact Assessment



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



PEOPLE
PROCESS
PERFORMANCE

Development, Application & Validation

HEALTHCARE

- Assisted Living (Oneview Healthcare)
- Oncology Pathways (Oneview Healthcare, Oxford & NHS)

ROAD TRANSPORT

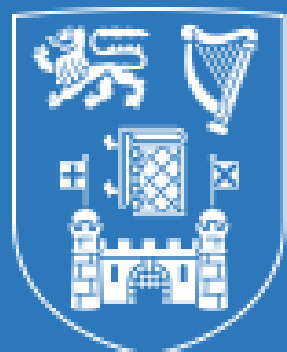
- Assisted Driving (Mazda and SFI)

AVIATION

- Lived Experience & Wellbeing Project (Enterprise Ireland)
- Robotics/sensors: Smart Hangar (SFI)

FINANCIAL SERVICES

- Intelligent Work (Zarion & Enterprise Ireland)



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

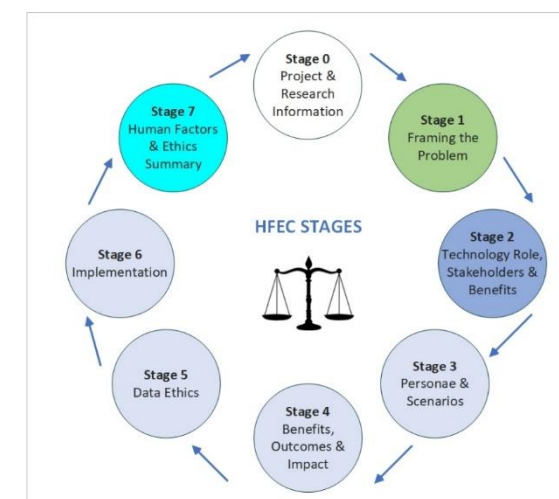


PEOPLE
PROCESS
PERFORMANCE

Understanding HFEC/Methodology & Problem Space

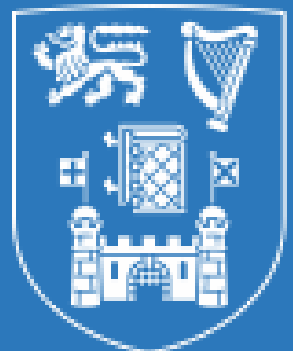
Problem/Challenge

**What it is?
Target/Outcome**



Why Care?

Scope & Areas of Focus



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



PEOPLE
PROCESS
PERFORMANCE

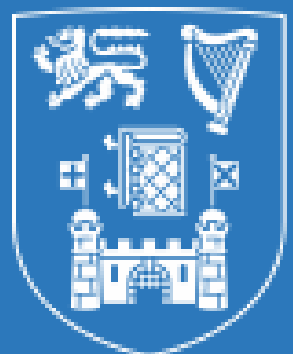
What is Ethical/Responsible AI?

Triple Bottom Line



Human activity must not compromise the long-term balance between the
(1) **economic**, (2) **environmental** and (3) **social pillars** (Elkington, 1994)

- Achieve equitable, inclusive, and ecologically sustainable AI development (European Commission, Ethics guidelines for trustworthy AI, 2018)
- Human Centric approach (EU Charter)
- Increase of human well-being as a metric of our progress (IEEE, Ethically Aligned Design)
- IEEE 7000-2021, IEEE 7010-2020

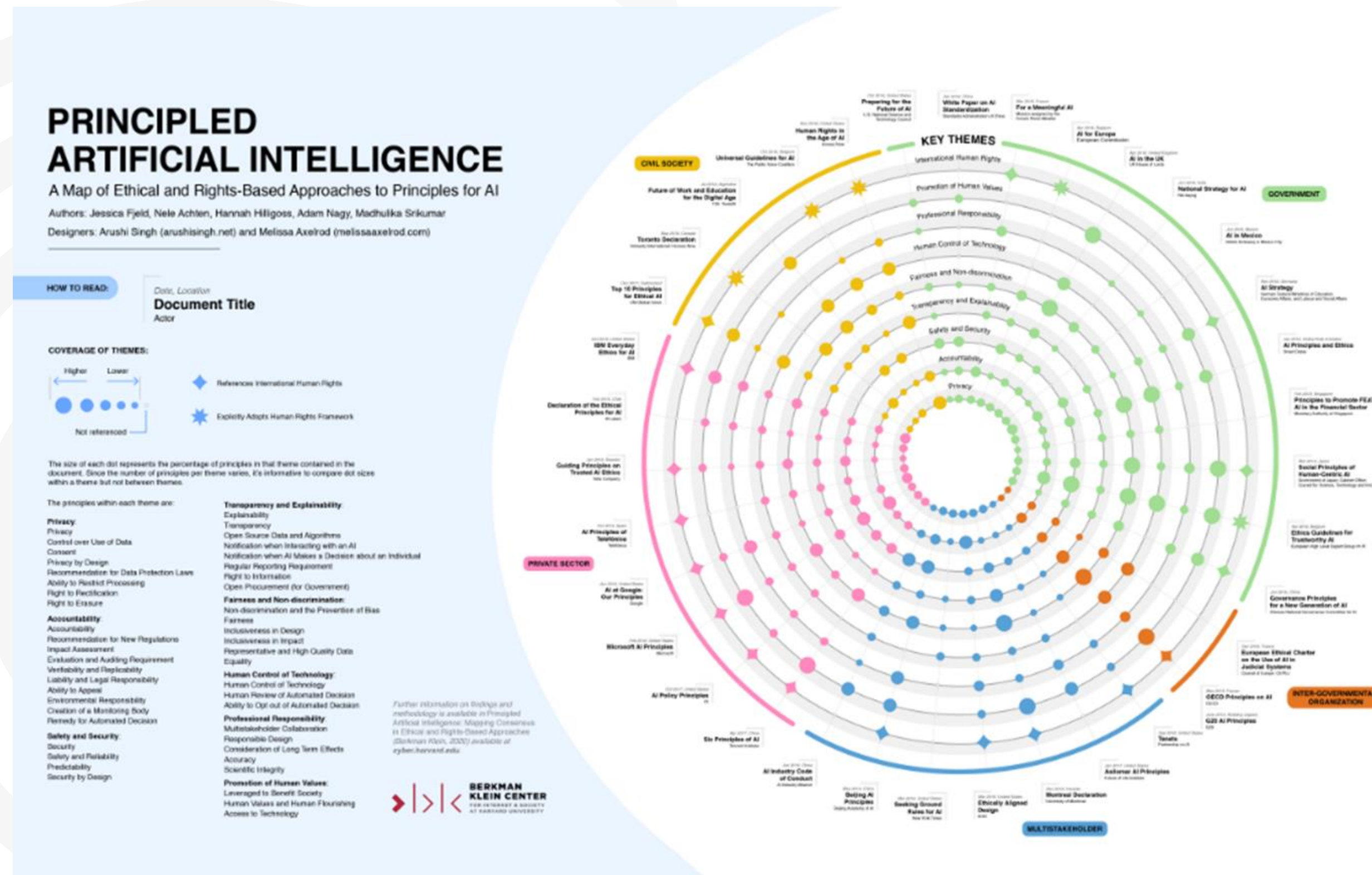


Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



PEOPLE
PROCESS
PERFORMANCE

Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI

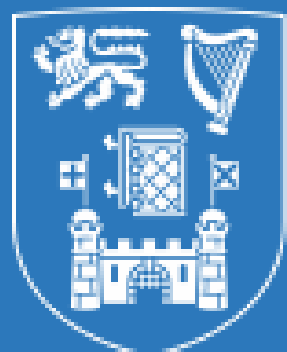


Growing consensus around **eight key thematic trends:** privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, **and promotion of human values**

- EU Ethics guidelines for trustworthy AI
- IEEE Ethically Aligned Design (EAD)
- OECD Principles on AI/G20 AI Principles
- OECD Privacy Guidelines
- UNI Global Union's top ten principles
- Asilomar AI Principles
- Beijing AI Principles
- Montreal Declaration
- Toronto Declaration

<https://cyber.harvard.edu/publication/2020/principled-ai>

(The Berkman Klein Center for Internet & Society Research Publication Series)



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



PEOPLE
PROCESS
PERFORMANCE

Scope/Areas of Focus

Data

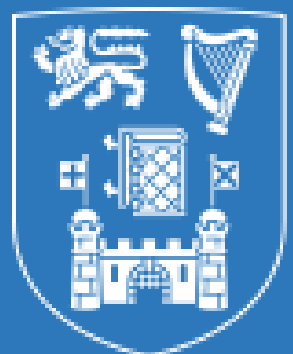
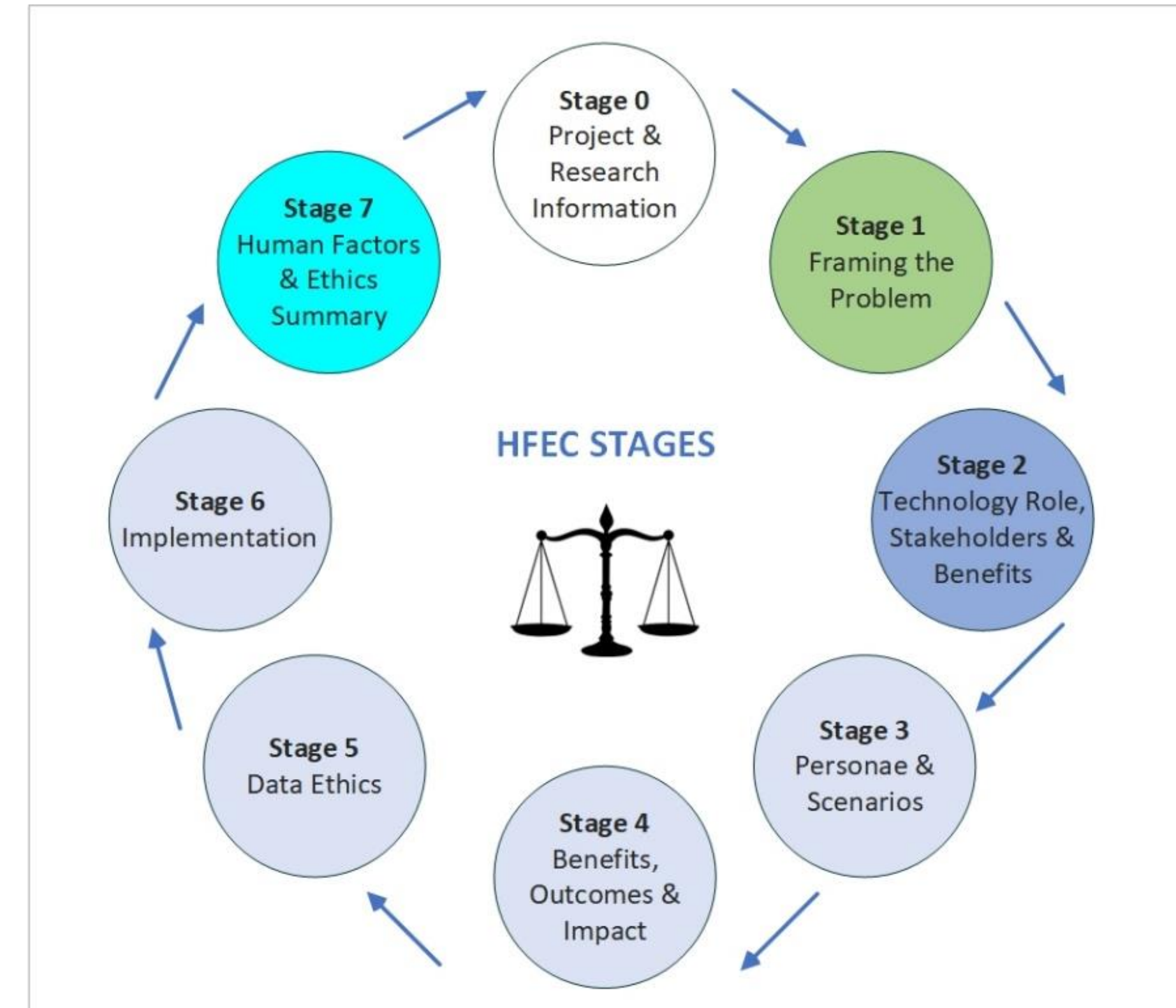
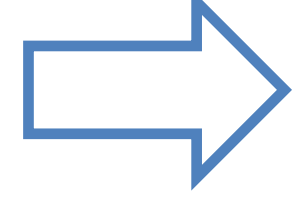
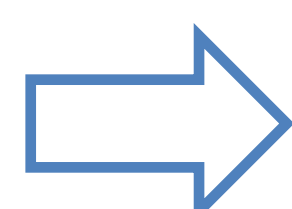
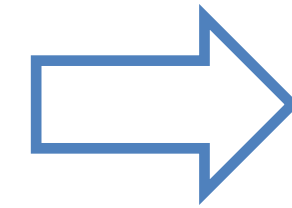
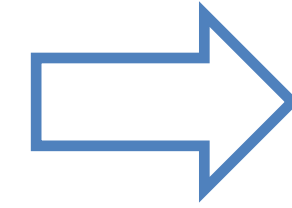
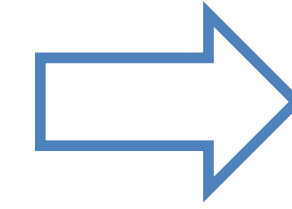
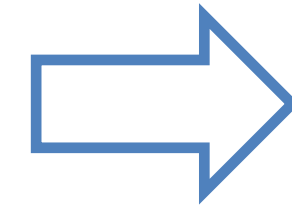
Model & Algorithm

Framing Problem & Role of Tech

Purpose, Use & Implementation

Human Role

Impact, Benefits & Consequences



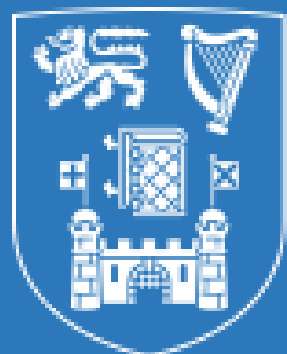
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



PEOPLE
PROCESS
PERFORMANCE

Why Care/Important

- Legal/compliance and costs of breaches .
- Business as corporate citizens within the larger social ecosystem (CSR). Are we doing enough?
- Not enough that technologies designed to be legal and safe in their usage (responsibility to consider human/societal needs and values).
- Missed opportunity – deliver positive benefits across 3BL.
- Cannot ignore uncertainty in relation to human experience and impact 3BL.
- Ethnocentrism.
- Focus on what matters.
- Impact on occupational health and safety/future of work.



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

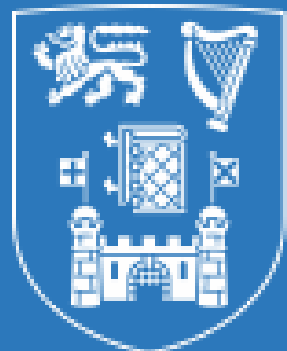


PEOPLE
PROCESS
PERFORMANCE

Ethical AI + Promotion of Human Values...

Legal | Ethical | Acceptable | Desirable

- Balances potential benefits and harms
- Respects human values and principles
- Upholds human rights (**autonomy, privacy**)
- Maintains the bonds of human solidarity among people - across generations, cultures, regions (**equitable human development**)
 - Serves and benefits both people and planet (**beneficence, wellbeing**)
 - Focuses on the right problems
 - Evidence based (need, benefits, impact)



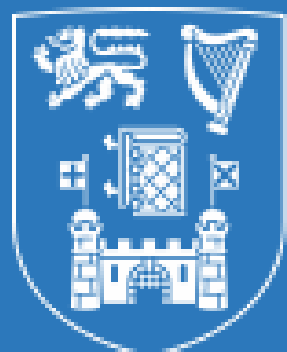
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



PEOPLE
PROCESS
PERFORMANCE

A/IS & Ethics Challenge

- **Technology availability (risk) .v. progress**
- **Understanding:** scope of responsible technology - beyond data autonomy, bias, transparency/ethical black box
- **Benefits:** new technologies have the potential to deliver positive benefits (economic, societal, environmental), if these are considered in design/development process (beyond compliance)
- **Awareness & Choice:** Design/technology teams exercise choice in relation to what is valued and advancing technology that delivers benefits (and not harm). Rationale & responsibility for these choices needs to be managed as part of the design process.
- **Uncertainty:** Limited range of empirical data to use. Technology designers must examine the ethical implications of things which may not yet exist, or things which may have impacts we cannot predict (Capurro, 2009)
- **Focus on using technology to address what matters**



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



PEOPLE
PROCESS
PERFORMANCE

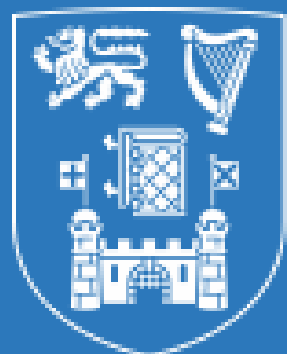
A/IS & Ethics Challenge (Uncertainty)

1. **Contribution to human wellbeing**
2. **Gap between the intended use** of a system and its use in terms of what is **implemented**
3. **Psychosocial effects** of certain technologies are not always apparent
4. **The beneficent uses of technology have a way of being co-opted for other purposes**
5. **Impact on behaviour, activities and societal values**

Uncertainty

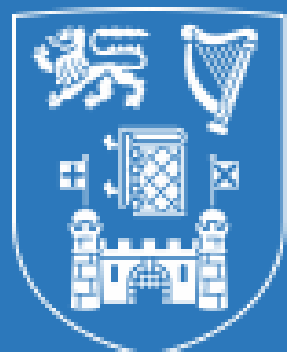
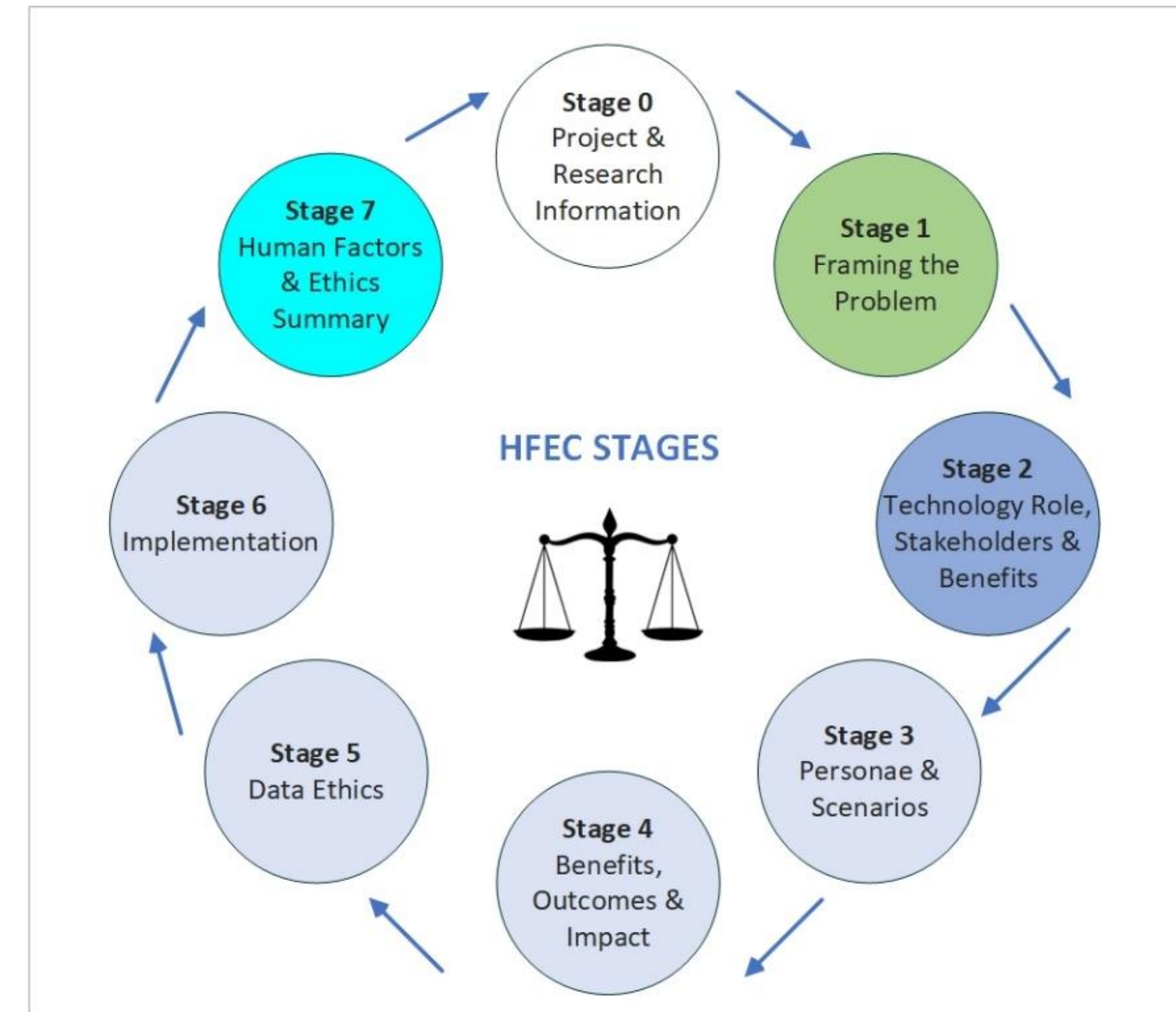


*The primary problem for the ethics of emerging technology is the problem of **uncertainty**. That is, how to deal with the uncertainty of future products, uses and consequences, and associated ethical issues that will result from an emerging technology (Sollie, 2007).*



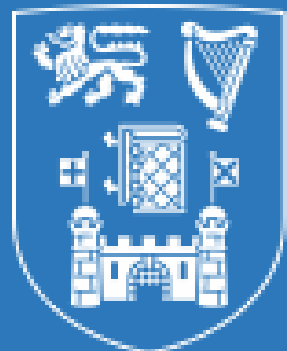
Human Factors & Ethics Canvas (HFEC)

- **Integration of ethics and HF methods**, particularly around the collection of evidence using **stakeholder evaluation methods, personae-based design, scenario-based design** approaches
- Community of practice (internal and external stakeholders)
- Address uncertainty
- **3BL underpins analysis**
- **Combines anticipatory/foresight approaches and participatory/deliberative ethics approach**
- Makes use of **ethical theories/perspectives** that are used in relation to the analysis of technology innovation in relation to the analysis of benefit versus harm (Principlism, Consequentialism & Deontology)

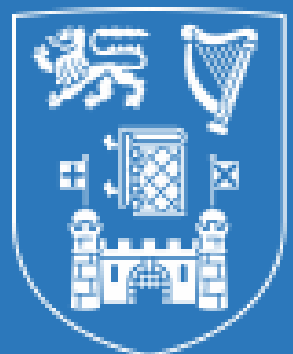
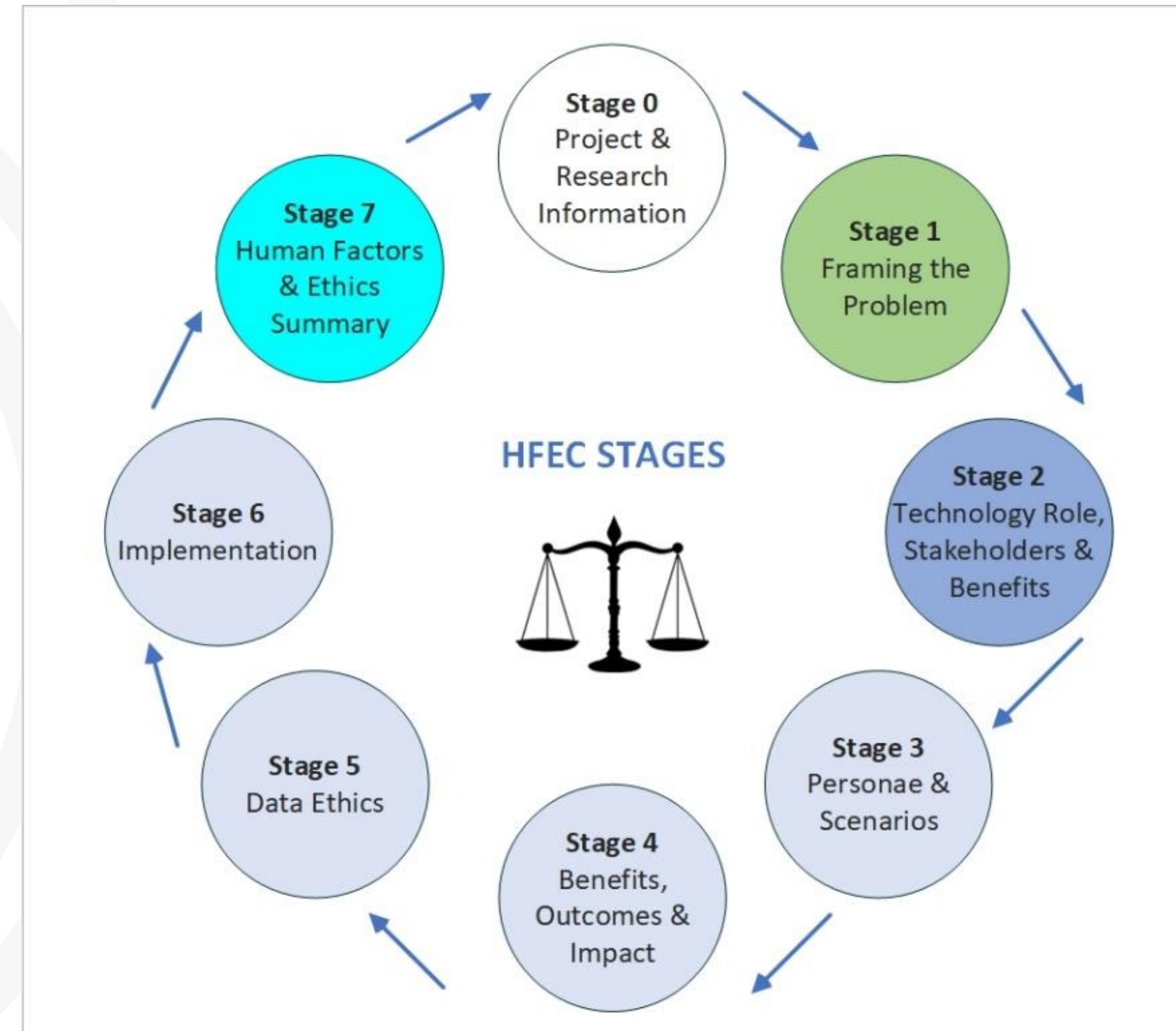


Team & Roles (Internal Stakeholders)

- Diverse practices across R&I & commercial projects
- Multi-disciplinary teams
- Roles
 - **Human Factors & Ethics Co-Ordinator (or Ethicist, if available)**
 - Product owner
 - Designers/user experience
 - Developers
 - Data ethics
 - Human Factors
 - Privacy/DPO
 - Management/executive



Overview of Stages



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

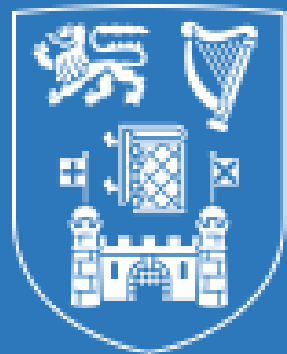


PEOPLE
PROCESS
PERFORMANCE

Stage 0: Project Information & Summary



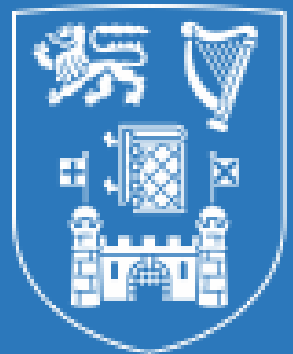
1	Date	
2	Project Name	
3	Product Owner	
4	HF & Ethics Coordinator	
5	HF & Ethics Canvas Version No.	
6	Research & Innovation Phase	
7	Summary of Research Completed & Key Sources of Information/Evidence	
8	Research Ethics Approval & Date DPIA Approval & Date	



Stage 1: Formulating the Problem & Posing the Question



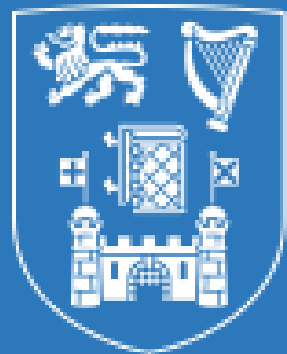
1	What is the problem that the proposed technology will address?	
2	Who is it a problem for? Key stakeholders? Who effect (directly and indirectly?)	
3	Setting & Environment?	
4	Causes of the problem?	
5	Ethical codes that apply in this setting?	
6	Ethics embedded in the problem definition?	
7	Ethics & Impact of Problem. Individual Level. Societal level. Ethics of acting/not acting?	
8	Summary of ethical issues to be addressed.	
9	Summary of relevant ethics principles and frameworks.	
10	Ethics & Key KPI	



Stage 2: Understanding Technology & Fit to Problem/Stakeholder Needs & Expected Benefits



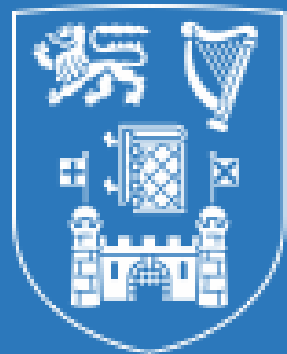
1	What is the technology? How does the proposed technology address the problem? What part of the problem does it address?	
2	What is the goal/objective? Intended purpose/function of technology?	
3	Setting & Environment?	
4	Direct users of technology? Goals? Needs? Expected Benefits?	
5	Other stakeholders impacted by technology? Goals? Needs? Expected Benefits?	
6	Design Decisions & Safeguards	



Stage 3: Stakeholders, Benefits, Consequences & Impact



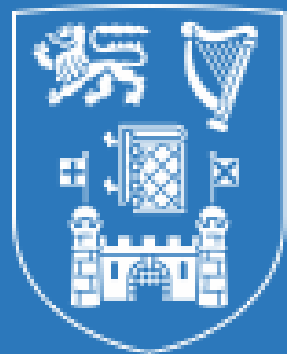
1	Overall benefits and outcomes: key stakeholders? Expected positive impacts?			
2	Expected Impact for key stakeholders (psycho-social themes). Individual level? Societal Level?	(A) Human role in the system	(B) Human Identity	(C) Lived experience, wellbeing, quality of life
		(D) Social Interaction & Relationships	(E) Activity & Behavior	(F) Attitudes & Values
3	What could go wrong? Potential failures? Potential negative impacts? Psychosocial? Environmental?			
4	Unintended consequences			
5	Unknowns			
6	Design Decisions & Safeguards			



Stage 4: Deep Dive (Personae & Scenarios)



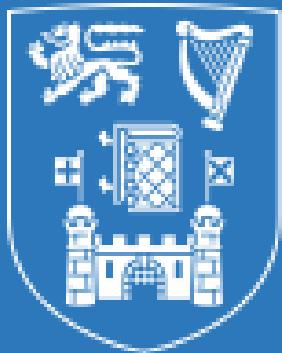
1	Example Scenario	
2	Example Personae	
3	How is it expected to work?	
4	What does success look like? Benefits for whom? Expected positive outcomes and for whom?	
5	What could go wrong? Potential failures? Potential negative impacts?	
6	Unintended consequences?	
7	Unknowns?	
8	Design Decisions & Safeguards	



Stage 5: Deep Dive (Data Ethics)



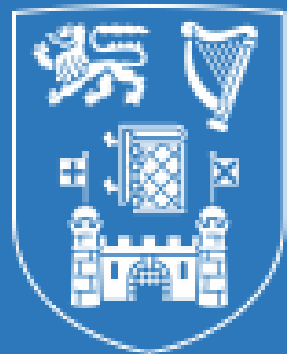
1	Ethical issues relevant to data collection? What data? Why collecting? Potential for bias in data collection?	
2	Ethical issues relevant to data, model & algorithms? Potential for bias? Potential for harm and risk? Explainability & transparency?	
3	Ethical issues relevant to data use & predictions (i.e., application of model/algorithms)?	
4	Data sharing & protections.	
5	Design Decisions & Safeguards	



Stage 6: Deep Dive (Implementation)



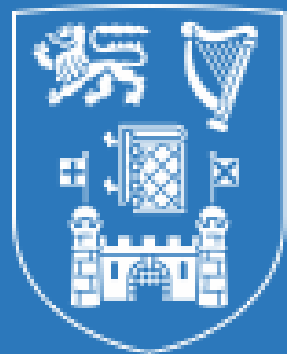
1	Implementation Approach	
2	Implementation Enablers	
3	Implementation Barriers	
4	Systems Perspective: Addressing Ethics as part of Implementation. • People. • Process. • Technology. • Culture. • Training & Education.	
5	Design Decisions & Safeguards	



Stage 7: Human Factors & Ethics Summary

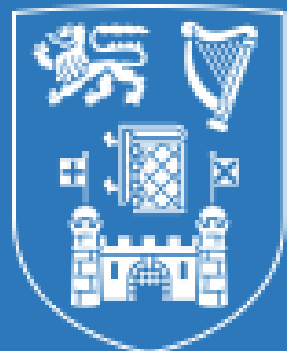


1	Key stakeholders? Who is this technology designed for?	
2	What does success look like? Success for whom?	
3	Human/Societal Vision & Technology Role/Purpose.	
4	Summary of Key Ethical Issues to be Addressed?	
5	Ethical Principles Underlying Technology Design	
6	Design Approach: Balancing Benefits & Harm. How managing ethics issues? How increasing potential positive impacts? How preventing risk/harm? How managing potential negative impacts and unintended consequences? How addressing unknowns?	
7	Data Ethics Summary.	
8	Implementation Summary	
9	Ethics & Key KPI	



Using the HFEC: Key Points

- Ethical principles need to be both articulated and then embedded in the design concept
- Personae/scenarios are useful in relation to considering and documenting the needs/perspectives of different stakeholders and adjudicating between conflicting goals/principles
- The translation of system objectives in relation to wellbeing and human benefit objectives (and associated metrics) ensures that wellbeing and human benefit are both a reference point and a design outcome.

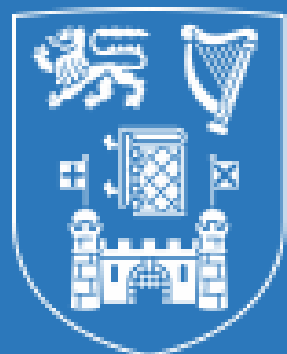


Using the HFEC: Key Points

1. Define principles underpinning system
2. Focus on stakeholders (persona and scenarios)
3. Define KPI (wellbeing etc.)
4. Define Impacts
 - Positive impacts
 - Negative impacts
 - Specific psychosocial impacts
 - Specific environmental impacts
 - Unintended consequences and unknown impacts

Design

- Embed ethical principles in system concept
- User functions & design features map to ethical principles
- Positive impacts must be supported by the system concept
- Potential negative impacts must be carefully managed in relation to the design concept and execution
- Recording design decisions and safeguards
- Iterative...



Progress, Human Benefit & TBL

Advancement, Moving Forward, Improvement

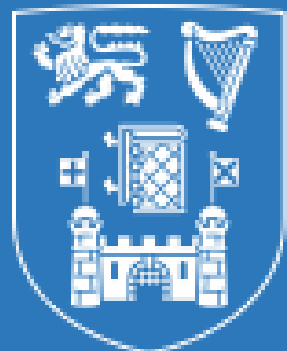
TECHNOLOGY
AVAILABILITY

LIVED EXPERIENCE &
VALUES



Address non-trivial negative and unintended consequences of new technology

Human activity must not compromise the long-term balance between the
(1) **economic**, (2) **environmental** and (3) **social pillars** (Elkington, 1994)



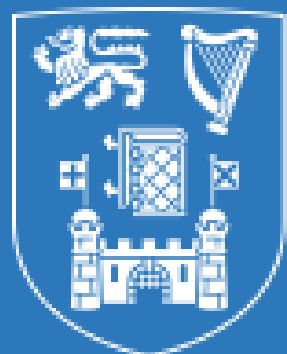
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



PEOPLE
PROCESS
PERFORMANCE

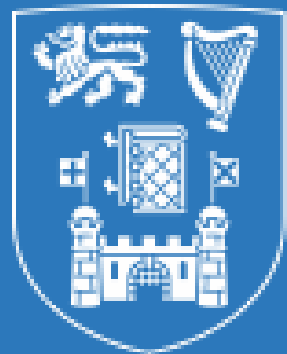
Conclusion

- Need to take actionable steps to address ethical issues
- Much convergence between the analysis of new technology from an ethics and human factors perspective
- Assessing the ethical implications of things which may not yet exist, or things which may have impacts we cannot predict, is very difficult
- This should not be barrier to posing important questions and ensuring that these questions are addressed as part of the design process
- Thinking about both potential positive, negative consequences and unintended consequences enables designers to build in protections into the design concept
- The specification of an ethics canvas as part of a broader human factors design approach ensures that ethical issues are considered.



Checklists, Standards, Certification

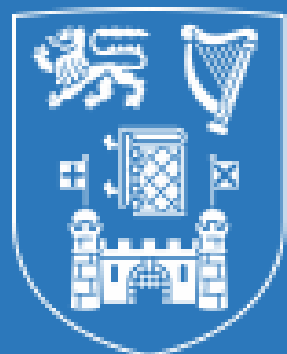
- United Nations Universal Declaration of Human Rights.
- Ethics guidelines for trustworthy AI (European Commission, 2019)
- Assessment List for Trustworthy AI (ALTAI) for self-assessment (European Commission, 2020)
- NIST 800.53 (security and privacy)
- IEEE 7000-2021, IEEE 7010-2020
- IEEE: Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)



Publications

Cahill J. (2020) Embedding Ethics in Human Factors Design and Evaluation Methodologies. In: Duffy V. (eds) Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Communication, Organization and Work. HCII 2020. Lecture Notes in Computer Science, vol 12199. pp 217-227 Springer, Cham. DOI: 10.1007/978-3-030-49907-5_15

- Cahill, J., Howard, V., Huang Y., Yeju, J, Ralph S & Dillon, A. (In Press). The Future of Work and Work Systems: Machine Learning & Artificial Intelligence for Designing Smart, Safe, Healthy & Ethical Work in Financial Services. Human-Automation Interaction: Manufacturing, Services and User Experience. Automation, Collaboration, E-Services, Vol. 10, Vincent G. Duffy et al. (Eds): Human-Automation Interaction. ISBN: 978-3-031-10779-5. <https://link.springer.com/book/9783031107825>.
- Cahill J., Howard V., Huang Y., Ye J., Ralph S., Dillon A. (2021). Addressing Human Factors and Ethics in the Design of ‘Future Work’ and Intelligent Systems for Use in Financial Services - Person Centered Operations, Intelligent Work & the Triple Bottom Line. In: Duffy V.G. (eds) Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body, Motion and Behavior. HCII 2021. Lecture Notes in Computer Science, vol 12777, page 3-13. Springer, Cham. https://doi.org/10.1007/978-3-030-77817-0_1
- Cahill, J., Cromie, S., Crowley, K, Kay, A., Gormley M, Kenny, E., Hermman, S, Doyle, C, Hever, A and Ross, R (2020). Ethical Issues in the New Digital Era: The Case of Assisting Driving. Book Chapter in "Security and Privacy From a Legal, Ethical, and Technical Perspective," ISBN 978-1-83881-206-5: <http://mts.intechopen.com/articles/show/title/ethical-issues-in-the-new-digital-era-the-case-of-assisting-driving>
- Cahill, J., Cromie, S., Crowley, K, Kay, A., Gormley M, Kenny, E., Hermman, S, Doyle, C, Hever, A and Ross, R. (2020) Advancing a ‘Human Factors & Ethics Canvas’ for New Driver Assistance Technologies Targeted at Older Adults. In: Duffy V. (eds) Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Posture, Motion and Health. HCII 2020. Lecture Notes in Computer Science, vol 12198. pp 503-520. Springer, Cham. DOI: 10.1007/978-3-030-49904-4_37
- Cahill, J. McLoughlin, S., O Neil, N. & Wetheral, S. (2020). Ambient assisted living and ethical issues pertaining to patient and care monitoring. “Elderly Care: Current Issues and Challenges”. Nova Sciences Publishers. <https://novapublishers.com/shop/elderly-care-current-issues-and-challenges/>. BISAC: MED032000. ISBN: 978-1-53618-446-4. <https://doi.org/10.52305/ATUP8209>
- Cahill, J.; Cullen, P.; Anwer, S.; Gaynor, K.; Wilson, S. (2020). The Requirements for New Tools for Use by Pilots and the Aviation Industry to Manage Risks Pertaining to Work-Related Stress (WRS) and Wellbeing, and the Ensuing Impact on Performance and Safety. Technologies (2020), 8, 40. <https://www.mdpi.com/2227-7080/8/3/40>. <https://doi.org/10.3390/technologies8030040>



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



PEOPLE
PROCESS
PERFORMANCE

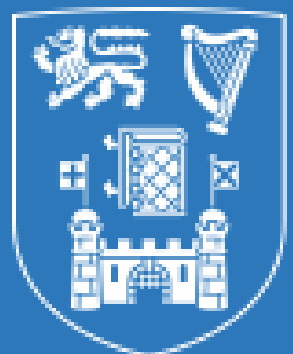
Contact

Dr Joan Cahill, CIHS, Trinity College Dublin cahilljo@tcd.ie

Linkedin: ie.linkedin.com/pub/joan-cahill/55/513/b97/

@joanccahill

<http://www.tcd.ie/cihs>



Trinity College Dublin

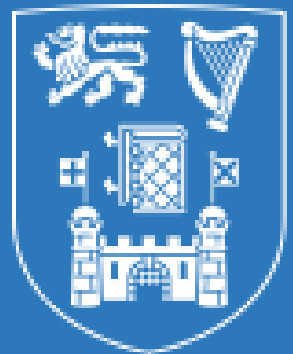
Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin



PEOPLE
PROCESS
PERFORMANCE

Thanks & Questions



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



PEOPLE
PROCESS
PERFORMANCE