TECHETHOS

FUTURE ○ TECHNOLOGY ○ ETHICS

TechEthos Policy event

Brussels & online, 14 November 2023

TECHETHOS
FUTURE ○ TECHNOLOGY ○ ETHICS

Ethics for the Green & Digital Transition

Welcome! The live will start at 10.30 CET

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

TechEthos Policy event

Brussels & online, 14 November 2023

TECHETHOS
FUTURE ○ TECHNOLOGY ○ ETHICS

Ethics for the Green & Digital Transition

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Welcome

Vivienne Parry

Event facilitator

Science Writer & Broadcaster

TechEthosEU

@TechEthosEU

Event hashtag: #EthicalTransition

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Inspired? Tag us on social media

TechEthosEU

@TechEthosEU

Event hashtag: #EthicalTransition

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Agenda - Morning

10.30-10.50	Welcome • Opening remark: **Barbara Thaler**, Member of the European Parliament & STOA Member • Introductory statement: **Mihalis Kritikos**, DG RTD
10.50-11.00	TechEthos in a nutshell: **Eva Buchinger**, TechEthos Coordinator
Ethics for the digital transformation	
11.00-11.45	Keynote: **Laura Weidinger**, DeepMind
11.45-12.15	Coffee break
12.15-13.15	Panel discussion on key ethical, social and regulatory challenges of Digital Extended Reality
13.15-14.15	Networking lunch

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Agenda - Afternoon

Ethics for the green transition	
14.15-15.00	Keynote: **Behnam Taebi**, Delft University of technology
15.00-15.15	Coffee break
15.15-16.15	Panel discussion on key ethical, social and regulatory challenges of Climate Engineering
Highlights & Outlook for the ethical governance of emerging technologies	
16.15-16.45	TechEthos in the larger context of the ALLEA Code of Conduct: **Maura Hiney**, UCD Institute for Discovery Legacies: foundation and continuation: **Eva Buchinger** (AIT), **Laurence Brooks** (University of Sheffield), **Renate Klar** (EUREC)

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Radio Tirol | Fernsehen | Landesstudio | Ganz Österr

POLITIK

Barbara Thaler wird neue WK-Präsidentin

Barbara Thaler folgt auf Christoph Walser als Präsidentin der Tiroler Wirtschaftskammer (WK). Das wurde in einer erweiterten Vorstands- und Landesleitungssitzung des Wirtschaftsbundes am Samstagvormittag beschlossen. Die EU-Abgeordnete der ÖVP ist in

MEPs

European Parliament

rbara THALER

Introductory statement

Mihalis Kritikos

Policy Analyst

Directorate-General for Research and Innovation, European Commission

TechEthosEU @TechEthosEU Event hashtag: #EthicalTransition

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

TechEthos in a nutshell

Eva Buchinger | TechEthos Coordinator

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

The ethics of new and emerging techs

"Everywhere we remain unfree and chained to technology, whether we passionately affirm or deny it. But we are delivered over to it in the worst possible way when we regard it as something neutral; (…)"

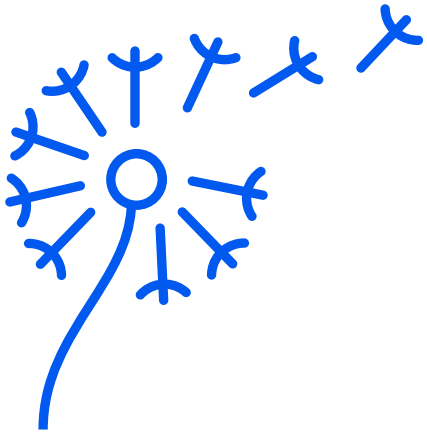
Martin Heidegger (1947/1977: 4)

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Challenge

Reconcile the needs of research and innovation and the concerns of society and reflect them in policy briefs & ethics frameworks.

- Climate engineering
- Digital extended reality
- Neurotechnologies



TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Approaches

Horizon scan

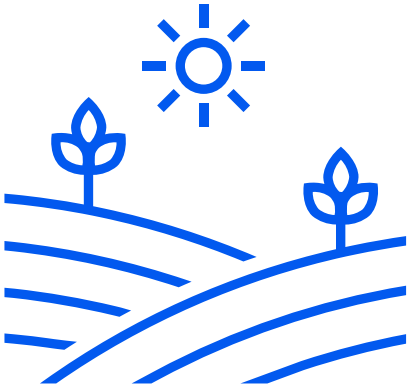
- Identify economically and ethically relevant techs

Ethics-by-design

- Make ethics an issue from the very beginning

Engagement

- 90+ experts and 300+ citizens



TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

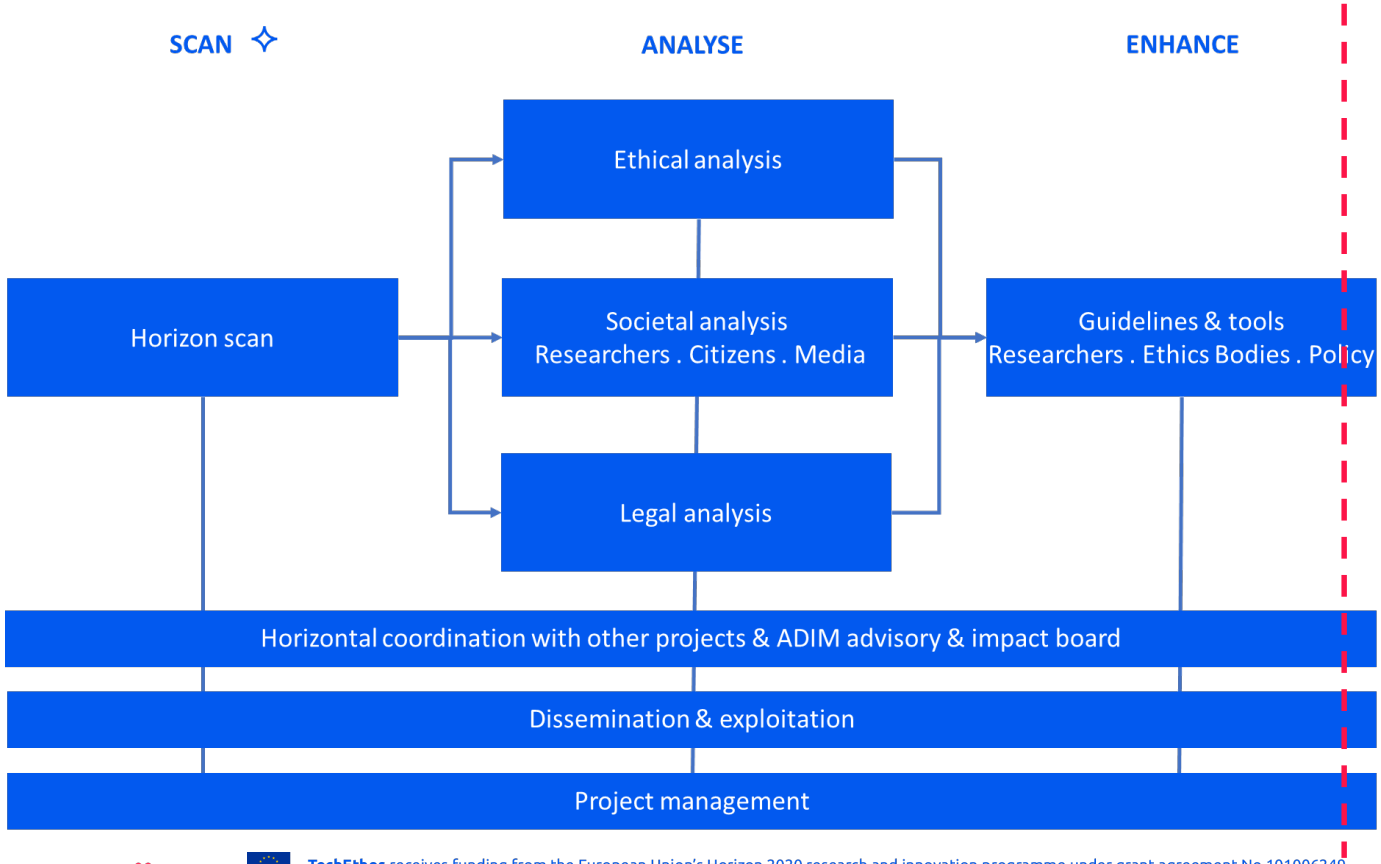
TechEthos game
Ages of Technology Impact

Translated into 6 languages

Played in Austria, Czech Republic, Romania, Serbia, Spain, and Sweden

Engaging 300+ citizens (incl. vulnerable groups)

2023



TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Achievements

https://www.techethos.eu/resources/

- SCAN: Ethical impact driven horizon scanning (D1.1, D1.2)
- ANALYSE: Identification of ethical dilemmas & values & principles (D2.1, D2.2) | Exploration of social awareness & attitudes (D3.1, D3.3) | International, EU and national legal analysis (D4.1, D4.2)
- ENHANCE: Suggestions to enhance legal frameworks (D5.2) | Suggestions for ethics framework enhancement (D5.3) | Criteria for ethical review by RECs in emerging technologies (D5.4) | Recommendation for EU law (D6.2) | Complementing the ALLEA CoC for research integrity (D5.5)
- TOOLS: TechEthos anticipatory ethics matrix TEAeM (D5.1) | TechEthos game: Ages of technology impact (D3.2) | Social readiness tool (D5.6)

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Policy briefs

https://www.techethos.eu/resources/deliverables-policy-briefs/

Policy Brief | 06 November 2023

XR and General Purpose AI: from values and principles to norms and standards

This policy brief presents TechEthos' recommendations for policy makers to ensure ethical governance, international collaboration, and public engagement in the field of eXtended Reality and Natural Language Processing (NLP).

Policy brief | 30 October 2023

Key messages for the ethical governance of Solar Radiation Modification (SRM) research

This policy brief presents TechEthos' recommendations for policy makers to ensure ethical governance, international collaboration, and public engagement in SRM research.

Policy brief | 30 October 2023

Key messages for the ethical governance of neurotechnologies

This policy brief presents recommendations for policy makers for the preparation of legislative or policy initiatives related to neurotechnologies.

Policy brief | 30 October 2023

Key messages for the ethical governance of Carbon Dioxide Removal (CDR)

This policy brief delves into the regulatory challenges within EU laws and policies surrounding CDR.

Policy brief | 28 February 2023

Enhancing EU legal frameworks for neurotechnologies

This policy brief presents TechEthos' recommendations for policy makers to protect and uphold ethical, legal and fundamental rights considerations in the development and deployment of neurotechnologies.

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

TECHETHOS
FUTURE ○ TECHNOLOGY ○ ETHICS

TECHETHOS
FUTURE ○ TECHNOLOGY ○ ETHICS

er grant agreement No 101006249.

TECHETHOS

FUTURE ○ TECHNOLOGY ○ ETHICS

Thanks to the wonderful TechEthos team

Thanks to the marvellous TechEthos ADIM Board

Thanks to the inspiring TechEthos cluster-network

Thanks to the wise guidance of the TechEthos POs

TECHETHOS

FUTURE ○ TECHNOLOGY ○ ETHICS

www.techethos.eu

Ethics for the digital transformation

Keynote by **Laura Weidinger**

Senior Research Scientist
Google DeepMind

TechEthosEU @TechEthosEU Event hashtag: #EthicalTransition

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Sociotechnical Safety Evaluation of generative AI systems

Laura Weidinger, Maribeth Rauh, Nahema Marchal, Arianna Manzini, Lisa Anne Hendricks, Juan Mateos-Garcia, Stevie Bergman, Jackie Kay, Conor Griffin, Ben Bariach, Iason Gabriel, Verena Rieser, William Isaac

TechEthos, 14 November 2023

Sociotechnical Safety

Taxonomy of Harms from multimodal generative AI

Representation Harms

E.g. Stereotypes, Exclusion

Information & Safety Harms

E.g. Dangerous capabilities, PII leak

Misinformation

E.g. Persuasion, Erosion of trust

Malicious Use

E.g. Deepfakes, Cyber attacks

Human Autonomy & Integrity Harms

E.g. Overreliance, Manipulation

Socioeconomic & Environmental Harm

E.g. Automation harm, Environmental harm

Adapted from:
Weidinger et al. (2021) “Ethical and Social Risks of Harm from Language Models” and
Shevlane et. al (2023) “Model evaluation for extreme risks”

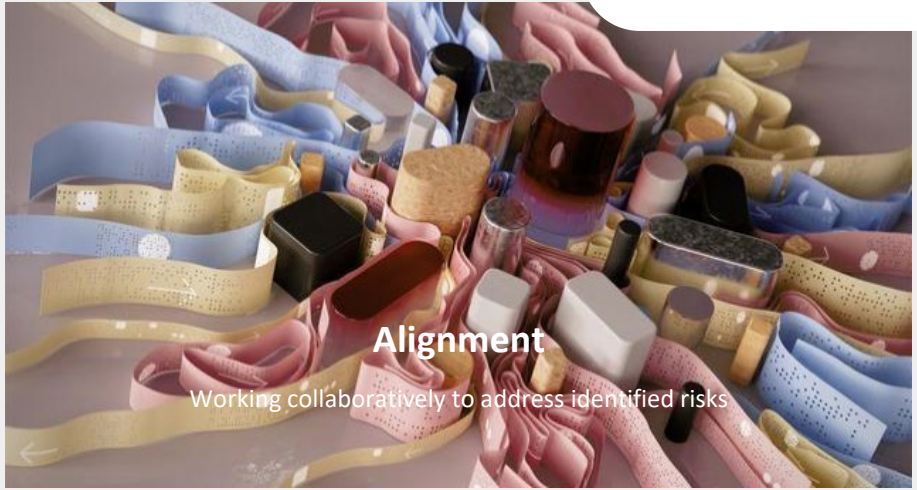
Sociotechnical Safety

Foresight

Anticipate the ethical and social risks of emerging technology

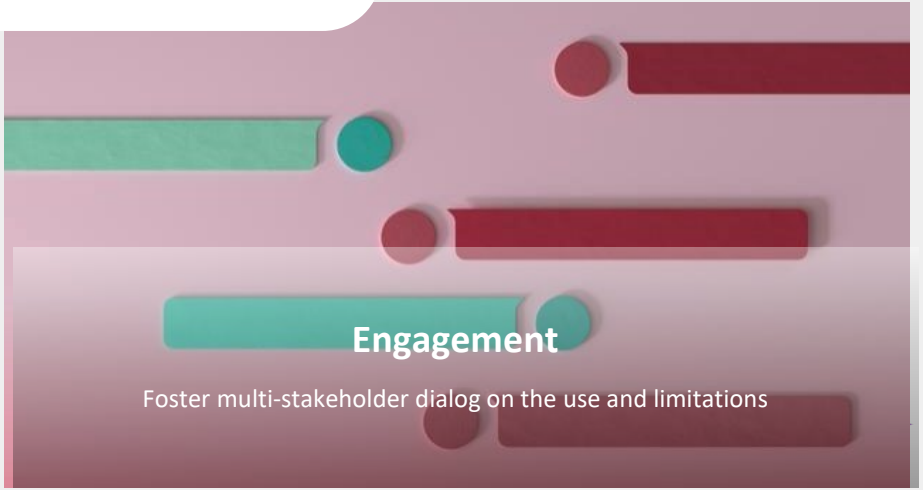
Evaluation

Translate risks into rigorous methods of assessment



Alignment

Working collaboratively to address identified risks



Engagement

Foster multi-stakeholder dialog on the use and limitations

Sociotechnical Safety

Foresight

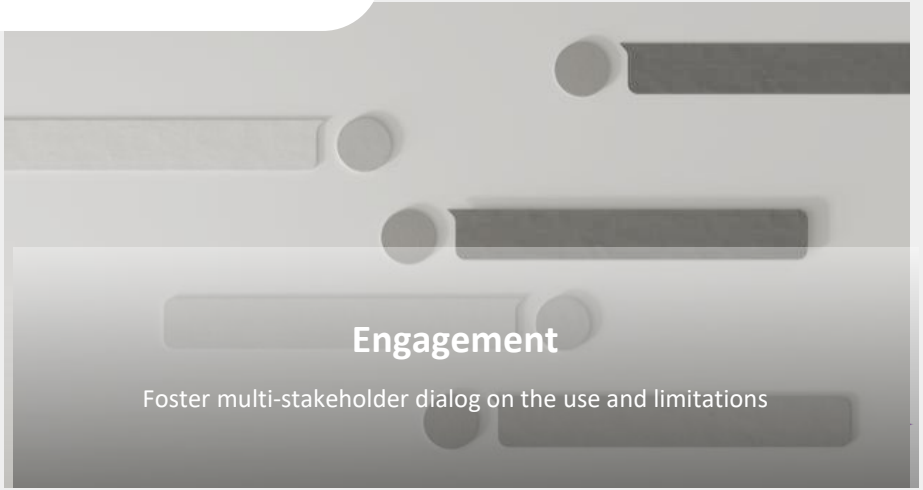
Anticipate the ethical and social risks of emerging technology

Evaluation

Translate risks into rigorous methods of assessment

Alignment

Working collaboratively to address identified risks



Engagement

Foster multi-stakeholder dialog on the use and limitations

Sociotechnical Safety

Foresight

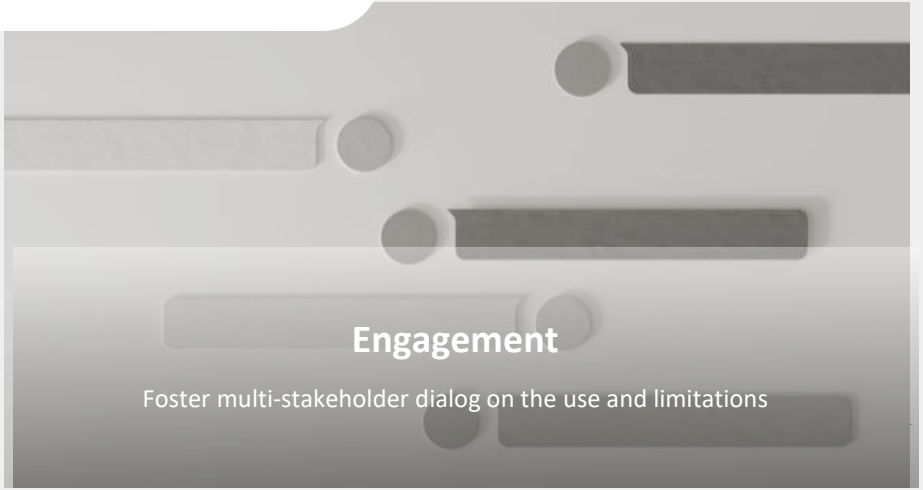
Anticipate the ethical and social risks of emerging technology

Evaluation

Translate risks into rigorous methods of assessment

Alignment

Working collaboratively to address identified risks

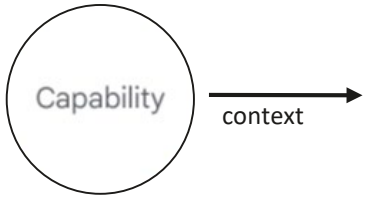


Engagement

Foster multi-stakeholder dialog on the use and limitations

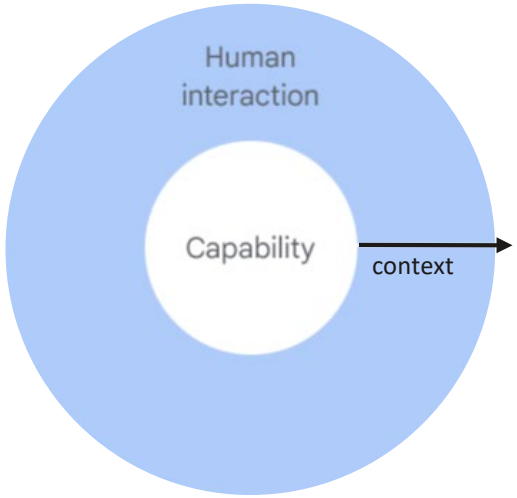
Evaluating sociotechnical safety of AI systems

- **Capability:** Assessing the full range of behaviors that a model could plausibly express during deployment.



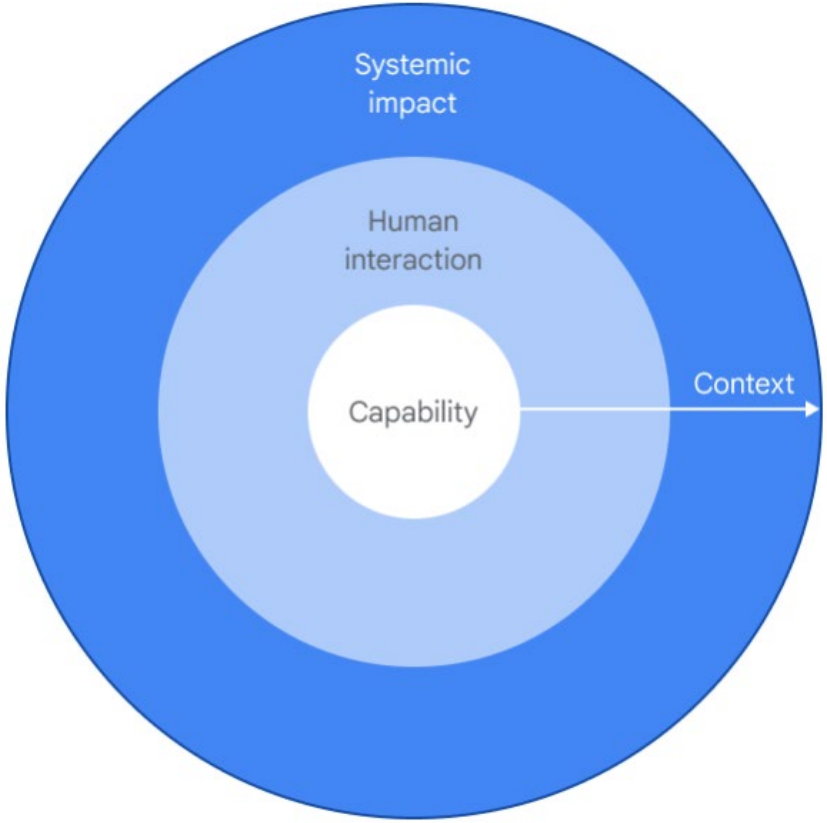
Evaluating sociotechnical safety of AI systems

- **Capability:** Assessing the full range of behaviors that a model could plausibly express during deployment.
- **Human interaction:** Assessing whether an AI model and associated elements (e.g. interface) behave as intended for a specified application.

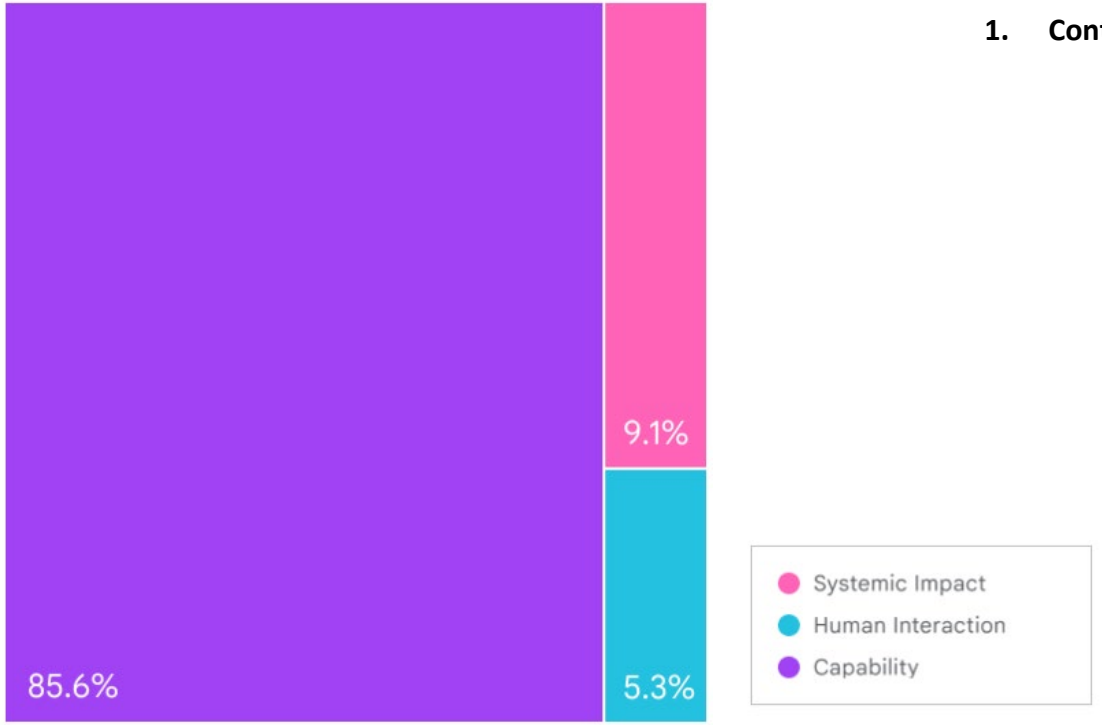


Evaluating sociotechnical safety of AI systems

- **Capability:** Assessing the full range of behaviors that a model could plausibly express during deployment.
- **Human interaction:** Assessing whether an AI model and associated elements (e.g. interface) behave as intended for a specified application.
- **Systemic impact:** Assessment of the anticipated or realized downstream effects of specific broader adoption and deployment of AI models and applications.

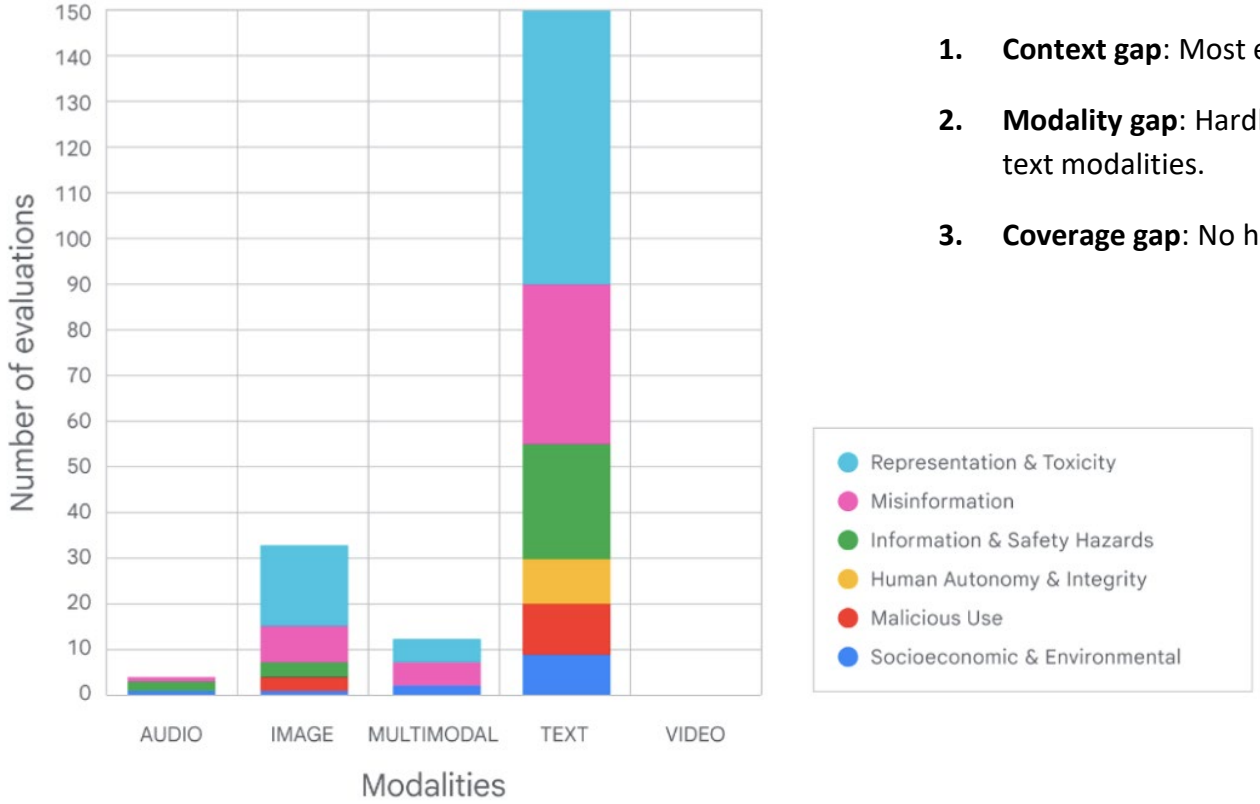


State of safety evaluations for generative AI today



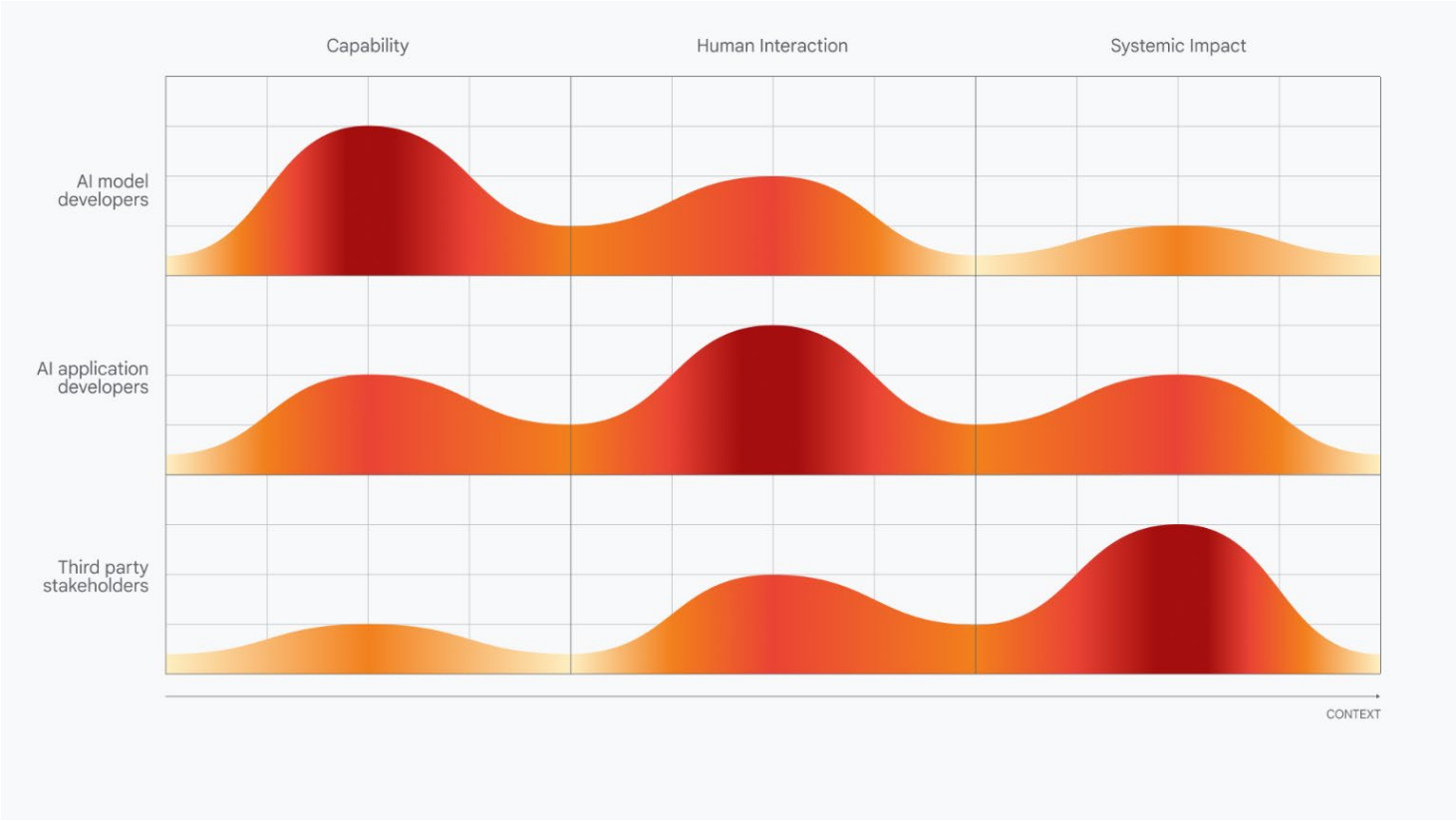
1. **Context gap**: Most evaluations are model-centric.

State of safety evaluations for generative AI today



1. **Context gap**: Most evaluations are model-centric.
2. **Modality gap**: Hardly any safety evaluations exist for non-text modalities.
3. **Coverage gap**: No harm area is covered across modalities.

Roles & responsibilities



Steps forward

1. **Thriving ecosystem for sociotechnical evaluation**
    - Sharing evaluations & where possible making them simple to run
    - Quality assurance through validating tests
    - Clarify roles & responsibilities across evaluation layers

1. **Reporting on progress**
    - Track evaluation gaps & state of the field
    - Report evaluation results

1. **Build safety evaluations**
    - Interactive & systemic impact evaluations
    - Prioritise evaluations for urgent harm areas

Thank you!

Coffee break

We will start again at 12.15 CET

Coming next: Panel discussion on key ethical, social and regulatory challenges of Digital Extended Reality

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Panel discussion

Key ethical, social and regulatory challenges of Digital Extended Reality

- **Alexei Grinbaum**, French Alternative Energies and Atomic Energy Commission (CEA) – TechEthos partner
- **Kevin MacNish**, Sopra Steria
- **Alina Kadlubsky**, Open AR Cloud Europe
- **Ivan Yamshchikov**, CAIRO

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Final conference

14 November 2023

TECHETHOS
FUTURE ○ TECHNOLOGY ○ ETHICS

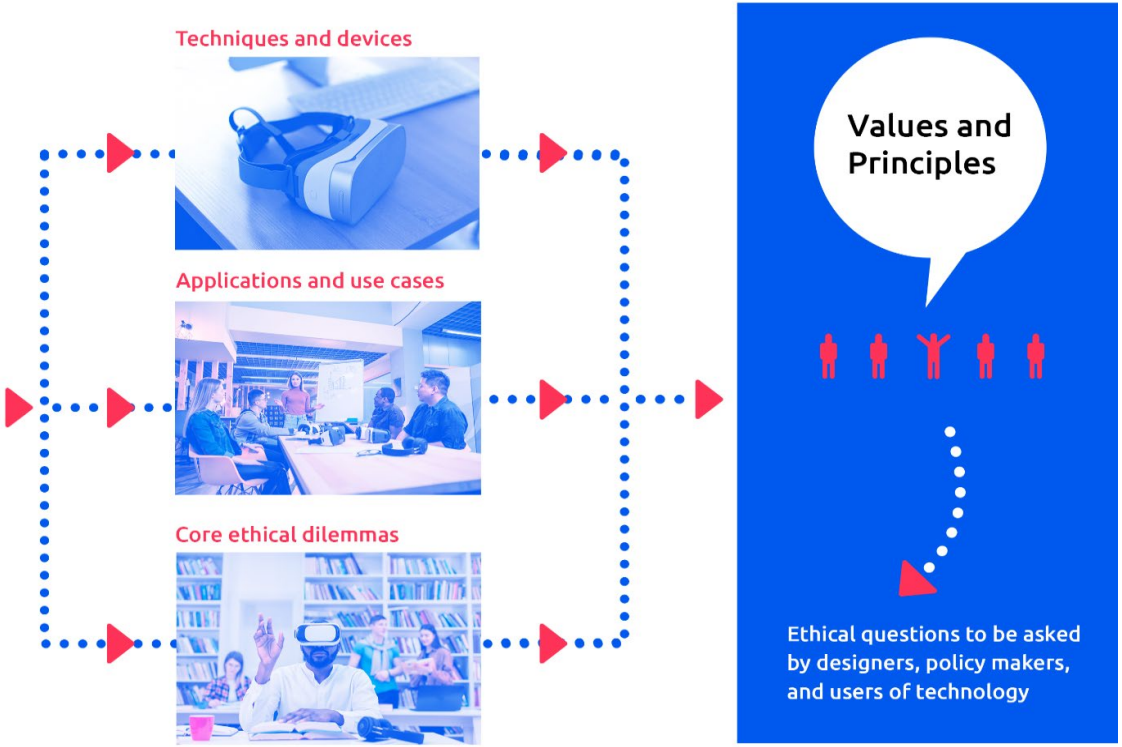
Selected messages on XR and NLP

Alexei Grinbaum (CEA)

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

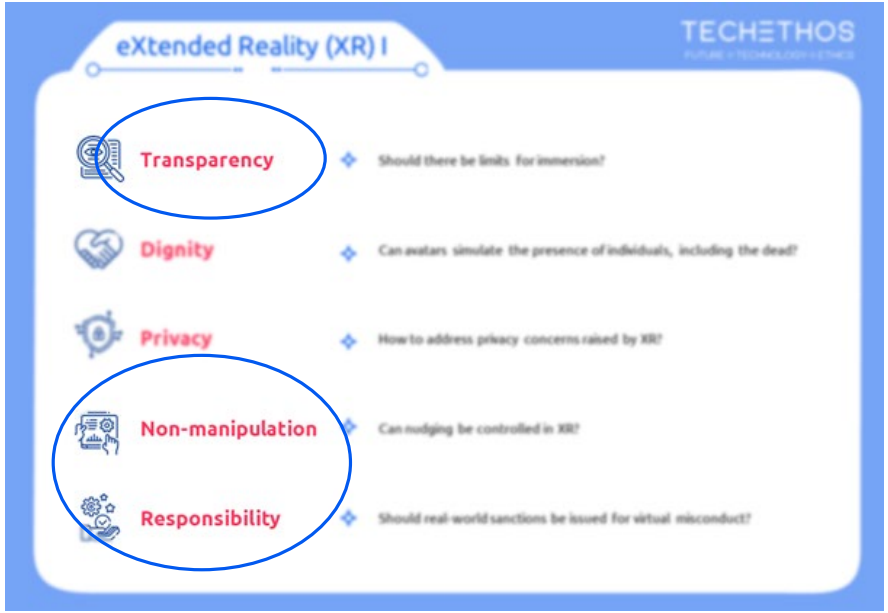
TechEthos methodology

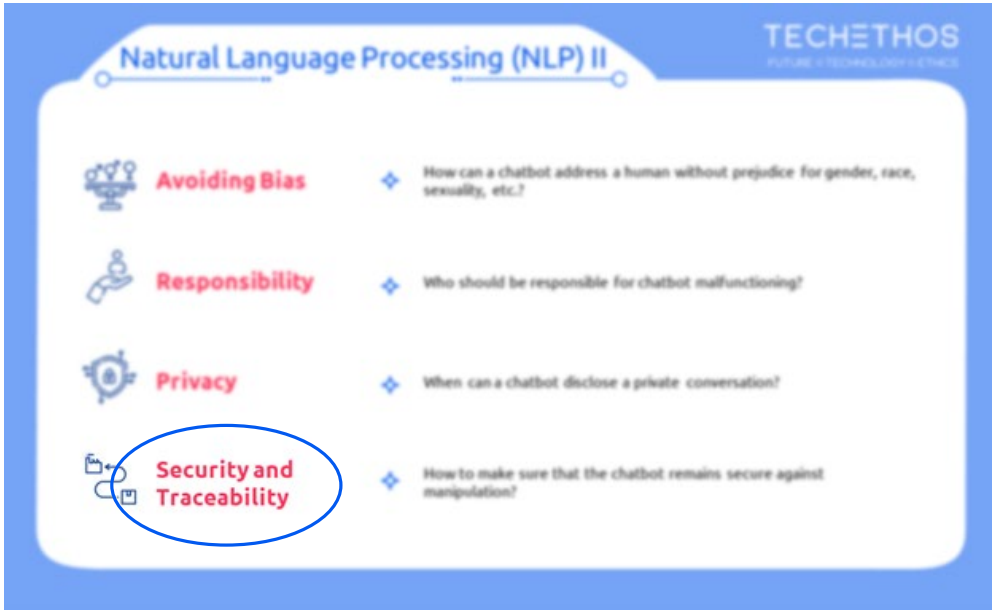
TECHETHOS
FUTURE ○ TECHNOLOGY ○ ETHICS



From speculation to reality: Enhancing anticipatory ethics for emerging technologies (ATE) in practice

Extended Reality and Natural Language Processing





TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

The Ethical Need for Watermarks

arXiv:2209.03118

Moral Equivalence in the Metaverse

NanoEthics (2022) 16: 257-270

Transparency

Maintaining status distinctions

AI-generated content: watermarks

History is important

Evaluating humans and machines separately

Avatars in XR: who's behind?

Rules for shared responsibility

Virtual sanctions that have no material analog

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

https://www.sfchronicle.com/projects/2021/jessica-simulation-artificial-intelligence/
https://www.washingtonpost.com/technology/2023/08/09/ai-dead-children-tiktok-videos/

“Intellectually, I know it’s not really Jessica, but your emotions are not an intellectual thing.”

“Thanks to extensive media archives of RAI, we were able to collect enough material to successfully generate a synthetic human of Maria Callas.”

Source IBC Accelerator
https://pluxbox.com/blog/creating-synthetic-humans-for-next-gen-storytelling/

The Jessica Simulation:
Love and loss in the age of A.I.

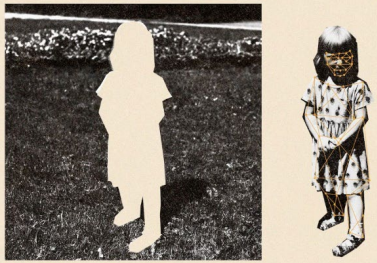
The Washington Post Sign in

TECH Help Desk Artificial Intelligence Internet C

AI is being used to give dead, missing kids a voice they didn’t ask for

By Jennifer Hassan

August 9, 2023 at 3:17 a.m. EDT



(Washington Post illustration; iStock)

Non-manipulation

Policy Brief

XR and General Purpose AI: from values and principles to norms and standards

Nudging or manipulation to the sole benefit of the manufacturer or the operator should be prohibited, while nudging to the benefit of the user should be evaluated on a case-by-case basis depending on context.

Does it depend on type of technology or degree of immersion / non-distinction?

Are there legitimate goals for manipulation?

Personalised nudging and emotional AI: new political and regulatory concerns

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Panel discussion

Key ethical, social and regulatory challenges of Digital Extended Reality

- **Alexei Grinbaum**, French Alternative Energies and Atomic Energy Commission (CEA) – TechEthos partner
- **Kevin MacNish**, Sopra Steria
- **Alina Kadlubsky**, Open AR Cloud Europe
- **Ivan Yamshchikov**, CAIRO

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Lunch is available on the ground floor at **Bambino**

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Lunch break

We will start again at 14.15 CET

Coming next: Keynote – Ethics for the Green transition

Behnam Taebi, Full Professor of Energy & Climate Ethics

Delft University of Technology

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Agenda - Afternoon

Ethics for the green transition	
14.15-15.00	Keynote: **Behnam Taebi**, Delft University of Technology
15.00-15.15	Coffee break
15.15-16.15	Panel discussion on key ethical, social and regulatory challenges of Climate Engineering
Highlights & Outlook for the ethical governance of emerging technologies	
16.15-16.45	TechEthos in the larger context of the ALLEA Code of Conduct: **Maura Hiney**, UCD Institute for Discovery Legacies: foundation and continuation: **Eva Buchinger** (AIT), **Laurence Brooks** (University of Sheffield), **Renate Klar** (EUREC)

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Ethics for the green transition

Keynote by **Behnam Taebi**

Full Professor of Energy & Climate Ethics
Delft University of Technology

TechEthosEU @TechEthosEU Event hashtag: #EthicalTransition

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Coffee break

We will start again at 15.15 CET

Coming next: Panel discussion on key ethical, social and regulatory challenges of Climate Engineering

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Panel discussion

Key ethical, social and regulatory challenges of Climate Engineering

- **Dominic Lenzi**, University of Twente – *TechEthos partner*
- **Benham Taebi**, Delft University of Technology
- **Dušan Chrenek**, Directorate-General for Climate Action, European Commission
- **Matthias Honegger**, Perspectives Climate Research

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Final conference

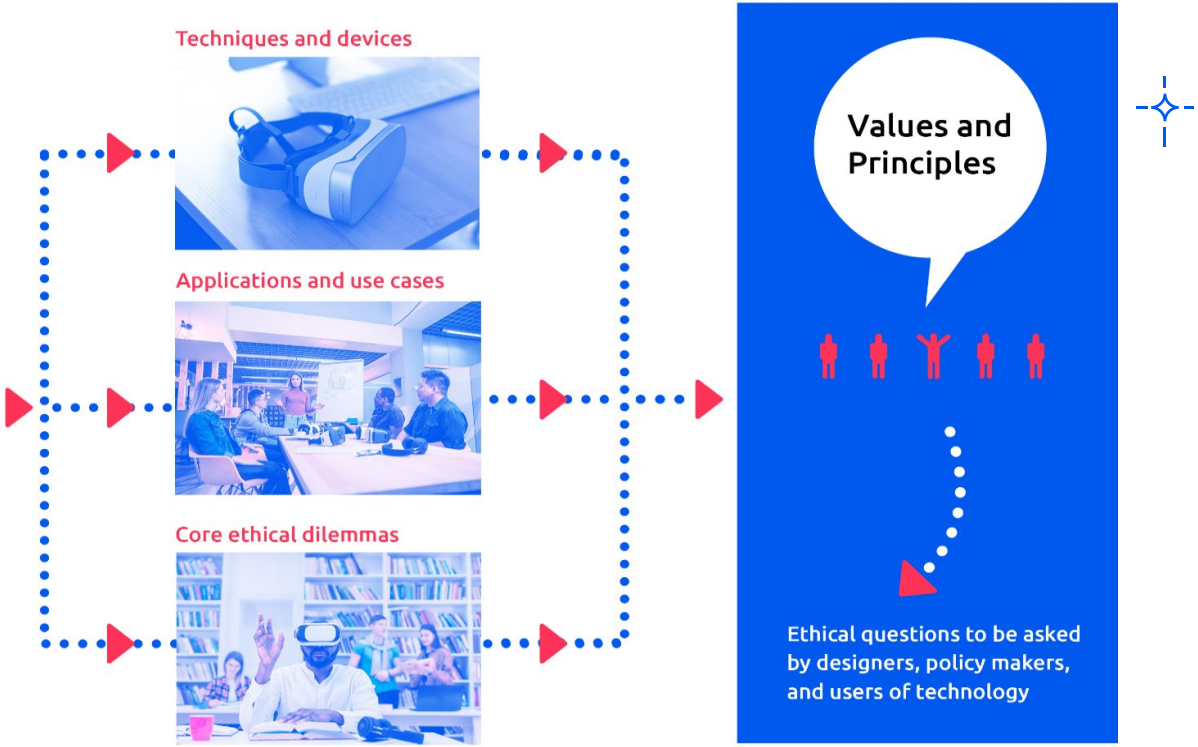
14 November 2023

TechEthos: key messages on Climate Engineering

Dr. Dominic Lenzi

University of Twente

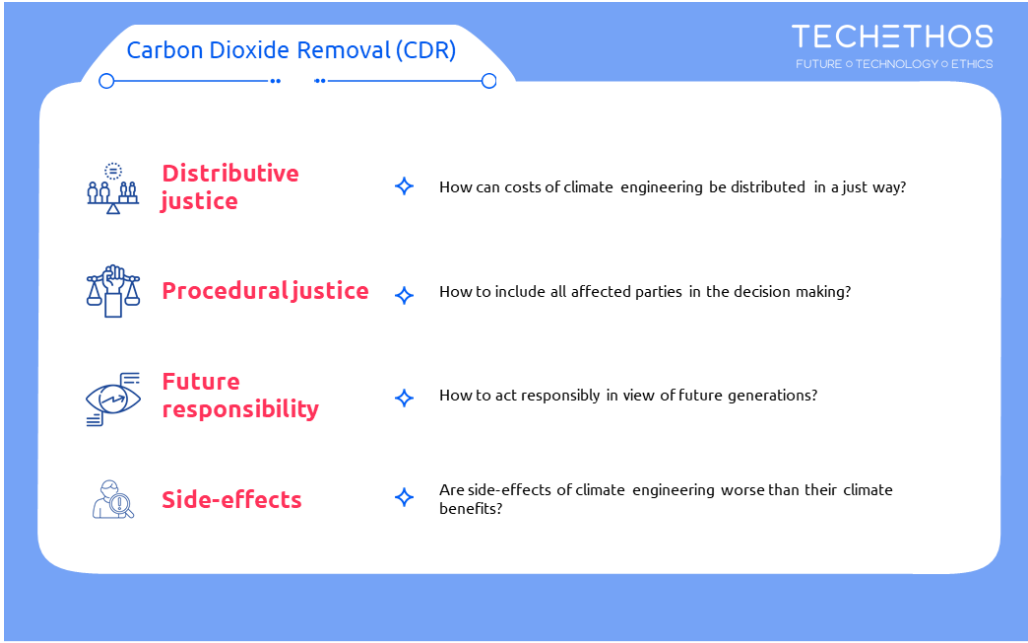
TechEthos methodology



From speculation to reality: Enhancing anticipatory ethics for emerging technologies (ATE) in practice

Carbon Dioxide Removal: ethical and governance concerns

- A ‘moral hazard’ effect, → slower emissions reduction
- Distribution of costs, incl. role of ‘carbon majors’
- Side effects of implementation, e.g. biodiversity, food security, water, human rights
- Public participation in decision-making, implementation, siting



TechEthos receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101006249.

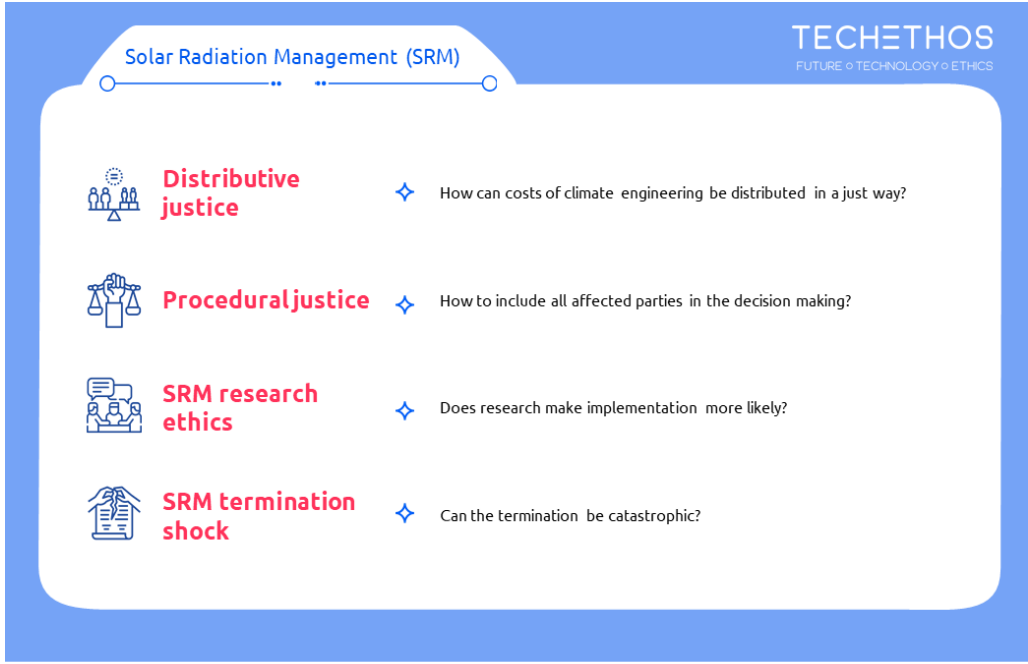
Carbon Dioxide Removal: key messages

- Clarify implications of EU principles, esp. Do No Significant Harm principle and Leave No-one Behind principle
- Clarify how CDR can be implemented in accordance with the UNFCCC's principle of Common But Differentiated Responsibilities and Respective Capabilities
- Scrutinize the role of the fossil fuel industry in CDR deployment. CBDR-RC includes the Polluter Pays Principle and Ability to Pay Principle
- Clarify how CDR can be implemented in accordance with EU's Biodiversity Strategy 2030; consider 'nature-based' forms of CDR

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Solar Radiation Modification: ethical concerns

- A ‘moral hazard’ effect, → slower emissions reduction
- Distribution of harms on most vulnerable
- Procedural justice, incl. ‘all affected principle’
- Research ethics: need for effective and legitimate governance of research



TechEthos receives func

Solar Radiation Modification: key messages

- Pursue a common definitions of SRM research, field testing, & deployment; esp. “deployment with a scientific basis” in UNCBD decision on CE
- Refine governance framework; develop a precautionary approach guided by ethical guardrails when assessing risks of SRM research programmes, against risks associated with not pursuing research
- Ensure SRM research governance is based on int. partnerships with wide representation; develop global participation & accountability mechanisms
- Include legitimacy and global justice when assessing the implications of SRM and SRM research, → protection of human rights
- Facilitate communication and knowledge-sharing of SRM research activities; limit acquisition of intellectual property

TechEthos receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101006249.

Panel discussion

Key ethical, social and regulatory challenges of Climate Engineering

- **Dominic Lenzi**, University of Twente – *TechEthos partner*
- **Benham Taebi**, Delft University of Technology
- **Dušan Chrenek**, Directorate-General for Climate Action, European Commission
- **Matthias Honegger**, Perspectives Climate Research

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Highlights & outlook for the ethical governance of emerging technologies

TechEthosEU

@TechEthosEU

Event hashtag: #EthicalTransition

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

TechEthos in the larger context of the ALLEA Code of Conduct

Maura Hiney

University College Dublin – Institute for Discovery

TechEthosEU @TechEthosEU Event hashtag: #EthicalTransition

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

Legacies: foundation and continuation

- **Eva Buchinger**, Austrian Institute of Technology
- **Laurence Brooks**, University of Sheffield
- **Renate Klar**, EUREC

TechEthosEU @TechEthosEU Event hashtag: #EthicalTransition

TechEthos receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006249.

TECHETHOS

FUTURE ○ TECHNOLOGY ○ ETHICS

Thank you

TechEthosEU

@TechEthosEU

www.techethos.eu