

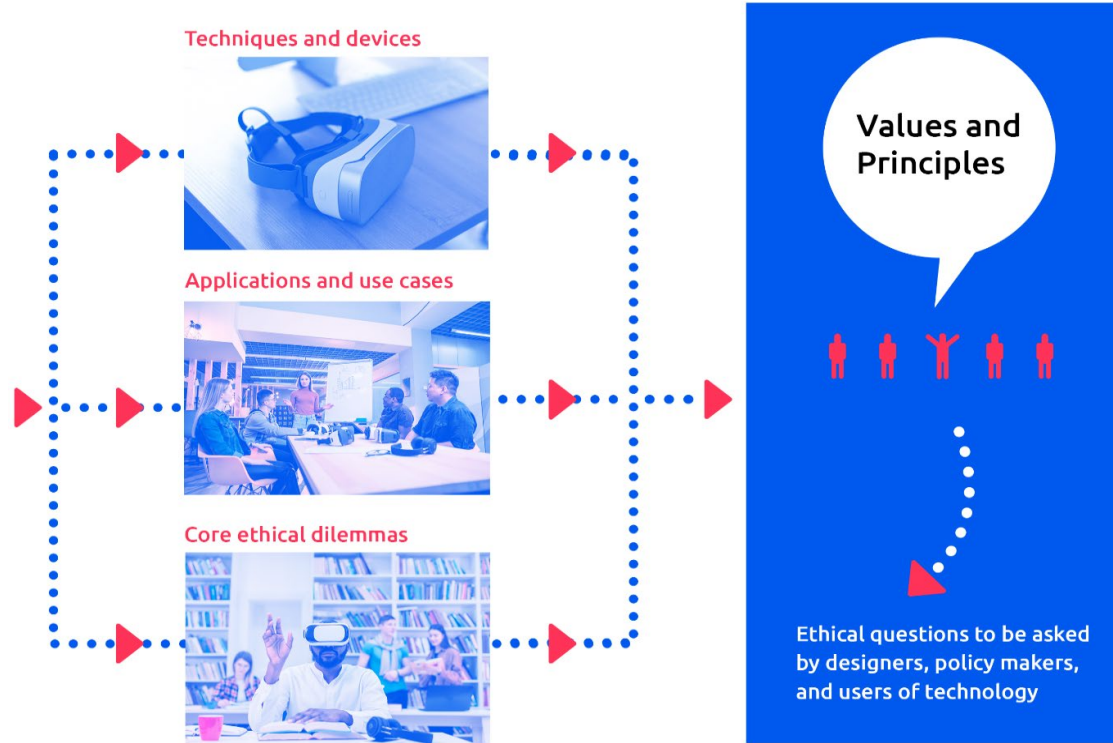


XR and AI ethics in TechEthos, iRECS, and MultiRATE

Alexei Grinbaum (CEA)



TechEthos methodology



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

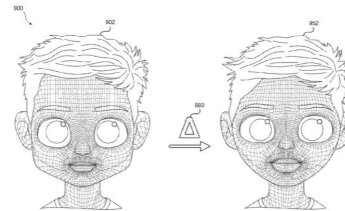
Ethical analysis: Extended Reality

❑ Technologies of extended reality (XR)

Virtual reality
Augmented and extended reality
Avatars and the metaverse
Digital twins
Affective computing in XR

❑ Applications and use cases of XR

Training and education
Health
Remote work
Romantic relationships
Social networking
Gaming



❑ Core ethical dilemmas in XR

Is there a preference for material reality?
The Equivalence Principle and its forms

❑ Values and principles in XR

Transparency
Dignity
Privacy
Non-manipulation
Responsibility
Risk reduction
Dual use and misuse
Power
Labour








Extended Reality and Natural Language Processing







eXtended Reality (XR) I

TECHETHOS
FUTURE • TECHNOLOGY • ETHICS

-  **Transparency** ♦ Should there be limits for immersion?
-  **Dignity** ♦ Can avatars simulate the presence of individuals, including the dead?
-  **Privacy** ♦ How to address privacy concerns raised by XR?
-  **Non-manipulation** ♦ Can nudging be controlled in XR?
-  **Responsibility** ♦ Should real-world sanctions be issued for virtual misconduct?

Natural Language Processing (NLP) II

TECHETHOS
FUTURE • TECHNOLOGY • ETHICS

-  **Avoiding Bias** ♦ How can a chatbot address a human without prejudice for gender, race, sexuality, etc.?
-  **Responsibility** ♦ Who should be responsible for chatbot malfunctioning?
-  **Privacy** ♦ When can a chatbot disclose a private conversation?
-  **Security and Traceability** ♦ How to make sure that the chatbot remains secure against manipulation?

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



G7 Hiroshima process

International Draft Guiding Principles

for Organizations Developing Advanced AI systems

(draft for consultation)

7. Develop and deploy reliable content authentication and provenance mechanisms such as watermarking or other techniques to enable users to identify AI-generated content

Task 2.2

AI in Healthcare

Extended Reality



Recommendations

AI in Healthcare

- 1) **Adapt the composition of RECs to include IT professionals**
- 2) **Set uniform and coherent 'AI in healthcare' guidelines across EU member states**
- 3) **Develop REC methodologies beyond compliance**

Extended Reality

- 1) **Establish Digital Ethics Committees (DECs) on a par with biomedical RECs**
- 2) **Maintain distinctions between AI-generated and human-generated content**
- 3) **Limit and regulate surveillance capabilities in virtual work environments**

Digital ethics in industrial design

PROJECT MULTIRATE

AI FOR SECURITY

ETHICS READINESS LEVELS



PROJECT SOPRANO

AI FOR ROBOTICS

ETHICS-BY-DESIGN
PROCESS IN FOUR
CONCRETE USE CASES

Score ($0 \leq 4$)

ELPRL 0 – Ethical and Privacy considerations lacking.

ELPRL 1 – Identified Ethical and Privacy Issues: Ethical and privacy considerations raised by the system have been identified.

ELPRL 2 – Characterized Ethical and Privacy Interactions:

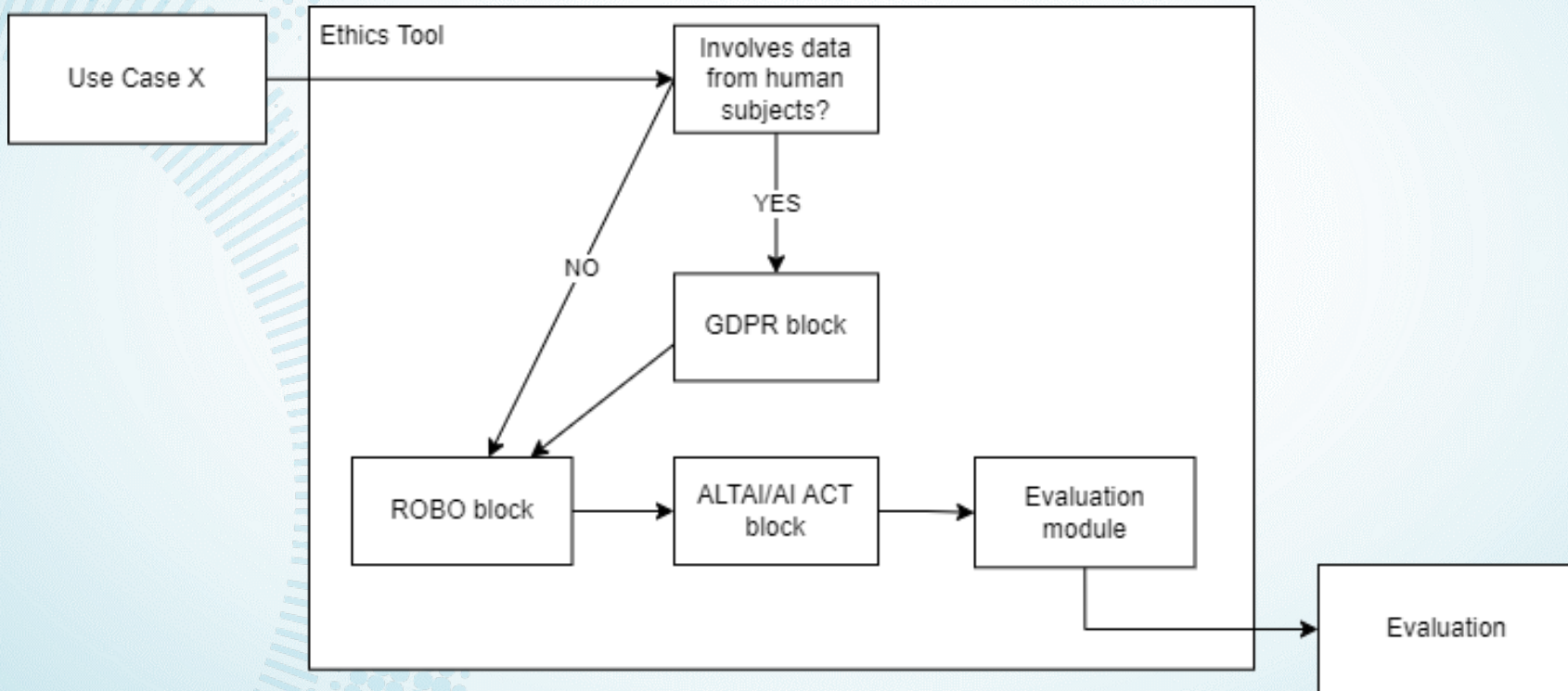
The interactions between different ethical and privacy considerations have been characterized.

ELPRL 3 – Compatibility of Solutions and Ethics by Design:

The system's ethical and privacy considerations have been designed to be compatible with each other. This means that improving one aspect (e.g., system security) does not negatively impact another aspect (e.g., user accessibility).

ELPRL 4 – Control Over Ethical and Privacy Issues:

The system has sufficient working control mechanisms in place to manage its ethical and privacy considerations and ensure accountability.



All cases: Human autonomy, System security, Accessibility and usability, Impact on end-users, Auditability.

AI Act Block:

- **Human Agency and Oversight:** Evaluates how AI systems affect human autonomy and decision-making.
- **Technical Robustness and Safety:** Assesses the system's resilience to adversarial attacks and overall security.
- **Transparency:** Checks if the system's purpose, criteria, and limitations are clearly communicated to users.
- **Fairness and Non-discrimination:** Ensures the system does not disproportionately affect certain user groups.
- **Societal and Environmental Well-being:** Evaluates the system's impact on society and the environment.

AI in ROBO Block:

- **Worker Dignity**
Assesses if the design and deployment of AI in robotics respect the dignity of workers.
- **Control**
Evaluates the capacity of the user to take control from the robot and that of the machine to take control from the human under certain circumstances.
- **Imitation and Social Interaction**
Evaluates the robot's communication strategy in terms of their resemblance with or imitation of living beings.

GDPR Block:

- **Lawfulness, Fairness, and Transparency:** Ensures data processing is lawful, fair, and transparent.
- **Purpose Limitation:** Data should only be collected for specified, explicit, and legitimate purposes.
- **Data Minimization:** Only necessary data should be collected.
- **Accuracy:** Ensures the data collected is accurate and up-to-date.
- **Storage Limitation:** Checks if data is not kept longer than necessary.
- **Integrity and Confidentiality:** Assesses the measures taken to secure data during processing.
- **Accountability:** Procedures for complying with GDPR principles.

“And Tobias went out to wash his feet, and behold a monstrous fish came up to devour him. And he being afraid of him, cried out with a loud voice, saying: Lord, he cometh upon me. And the angel said to him: Take him by the gill, and draw him to thee. And when he had done so, he drew him out upon the land, and he began to pant before his feet. Then the angel said to him: Take out the entrails of the fish, and lay up his heart, and his gall, and his liver for thee: for these are necessary for useful medicines.”

Tobit 6:2-5



FILIPPINO LIPPI, NATIONAL GALLERY, WASHINGTON

Alexei Grinbaum
LES ROBOTS ET LE MAL



DESCLEE DE BROUWER

ALEXEI GRINBAUM

**PAROLE
DE MACHINES**



DIALOGUER AVEC UNE IA

humenSciences