Winter School - IMS²

Ethics in Smart Manufacturing

Yinling LIU

Asso. Prof. at the University of Lorraine

November 2025

https://liuyinling.github.io/

Curriculum Vitae

Master

09/2013 - 04/2016

NUAA, China

Master : The analysis of risks for safety-critical systems

AADL (Architecture Analysis Design Language)

[LIU 2015, 2016, 2017a]

Ph.D.

10/2016 - 12/2019

DISP, INSA de Lyon

Thesis : Conception and Verification of Aircraft Maintenance Information Systems

Agent-Based Modeling & Simulation (ABMS); Model checking; Fuzzy-Rough Nearest Neighbor classification et prediction

[LIU 2017b, 2018, 2019a, 2019b, 2021]

Postdoc 1

01/2020 - 09/2020

DISP, INSA de Lyon

Project : Simulation of patient flows in Emergency Department in Saint-Etienne

Discrete-event simulation; ABMS

[LIU 2023]

Postdoc 2

10/2020 - 08/2022

SM@RT, IRIT, UT3

Project : Modeling and verification of requirements based on states and modes

Domain-specific languages; Model checking

[LIU 2022]

Associate Professor

Team in Epinal

1. How can we model humans in cyber-physical systems while taking into account their physiological, psychological, and social aspects?
2. How can we integrate the ethical dimension into the system (behavior, evaluation)?
3. How can we implement human-machine interactions from the engineering perspective?
4. How can we create digital twins for the human-centered cyber-physical system?

Self-driving car -1

What should the self-driving car do?

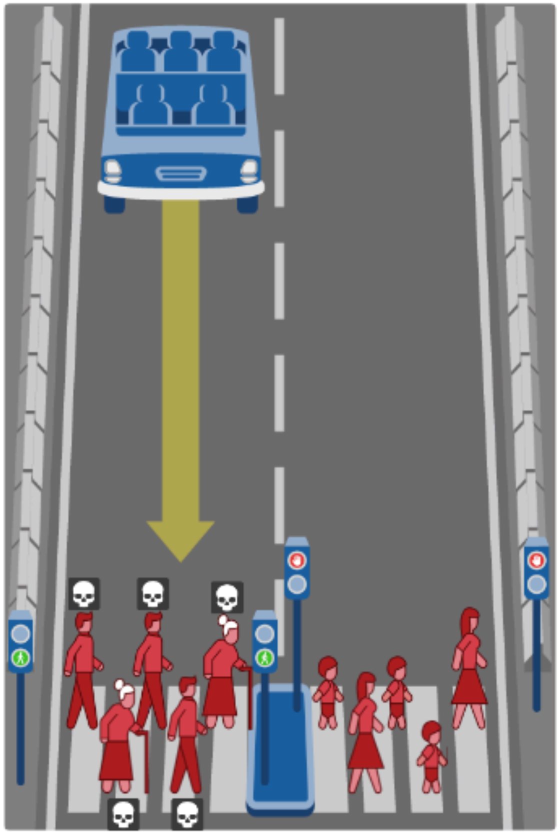
1 / 13

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in ...

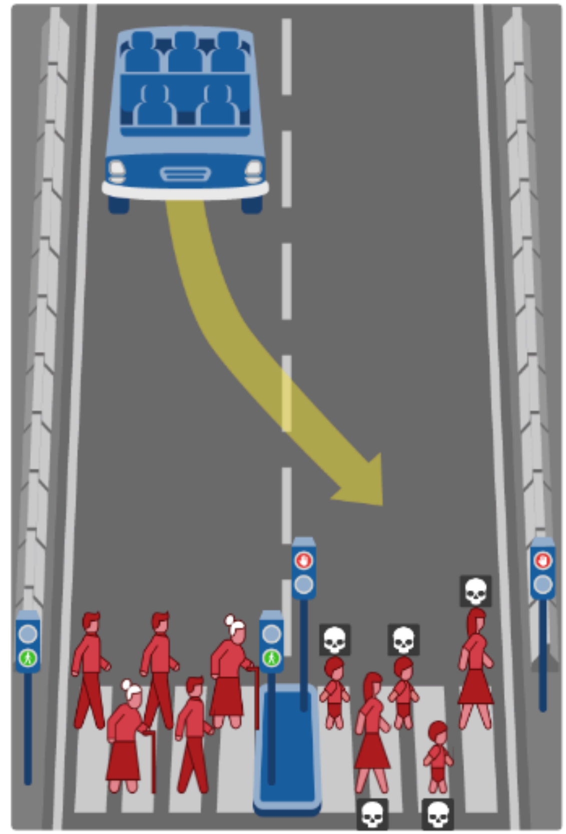
Dead:

- 3 men
- 2 elderly women

Note that the affected pedestrians are abiding by the law by crossing on the green signal.



Hide Description



Hide Description

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...

Dead:

- 3 boys
- 2 women

Note that the affected pedestrians are flouting the law by crossing on the red signal.

Game : https://www.moralmachine.net/

Self-driving car -2

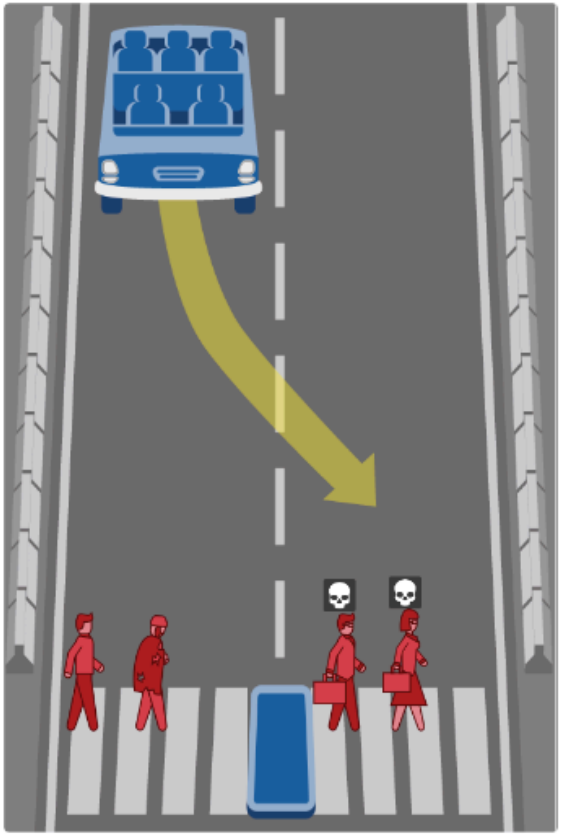
What should the self-driving car do?

2 / 13

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in …

Dead:

- 1 male executive
- 1 female executive

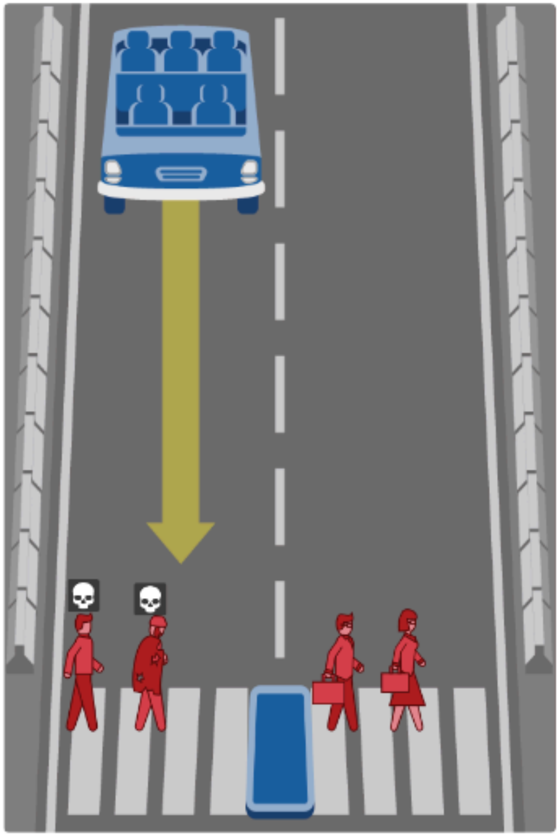


Hide Description

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in …

Dead:

- 1 man
- 1 homeless person



Hide Description

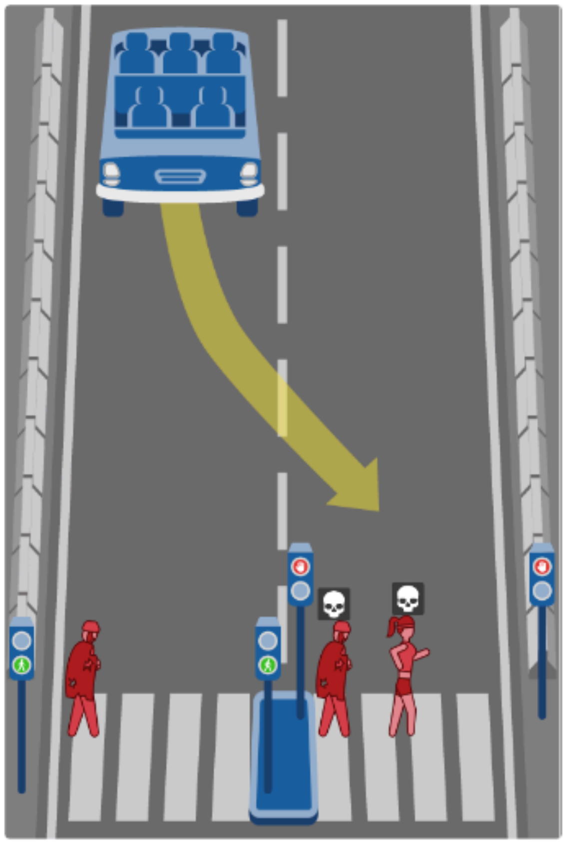
Game : https://www.moralmachine.net/

Self-driving car -3

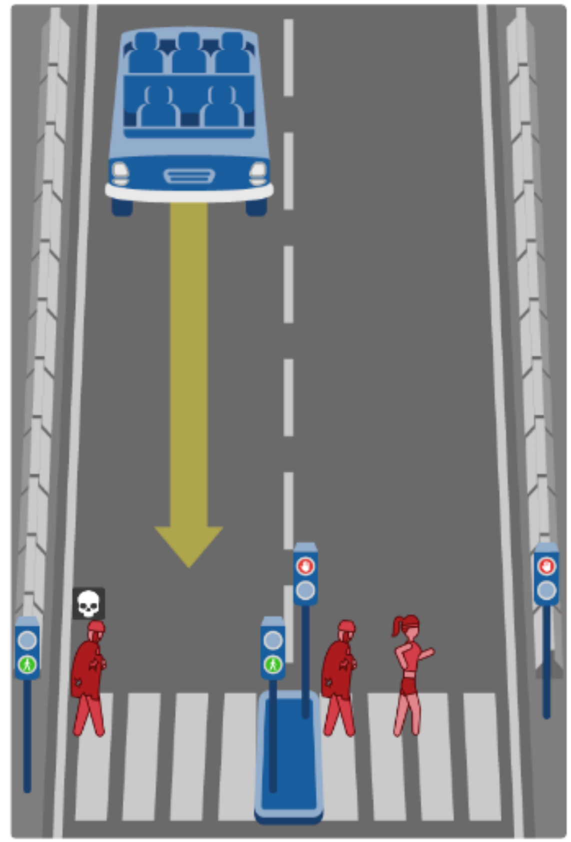
What should the self-driving car do?

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...
Dead:
- 1 homeless person
- 1 female athlete

Note that the affected pedestrians are flouting the law by crossing on the red signal.



Hide Description



Hide Description

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in ...
Dead:
- 1 homeless person

Note that the affected pedestrians are abiding by the law by crossing on the green signal.

It is difficult to judge which one is less unethical.

Industrial examples

An intelligent production monitoring system will never disclose the performances of operators.

An intelligent AGV will always stop in front of operators to avoid harms.

An intelligent production monitoring system will never over-solicitate a supervisor and will never ask him to react to a perturbation without ensuring his correct awareness of the situation.

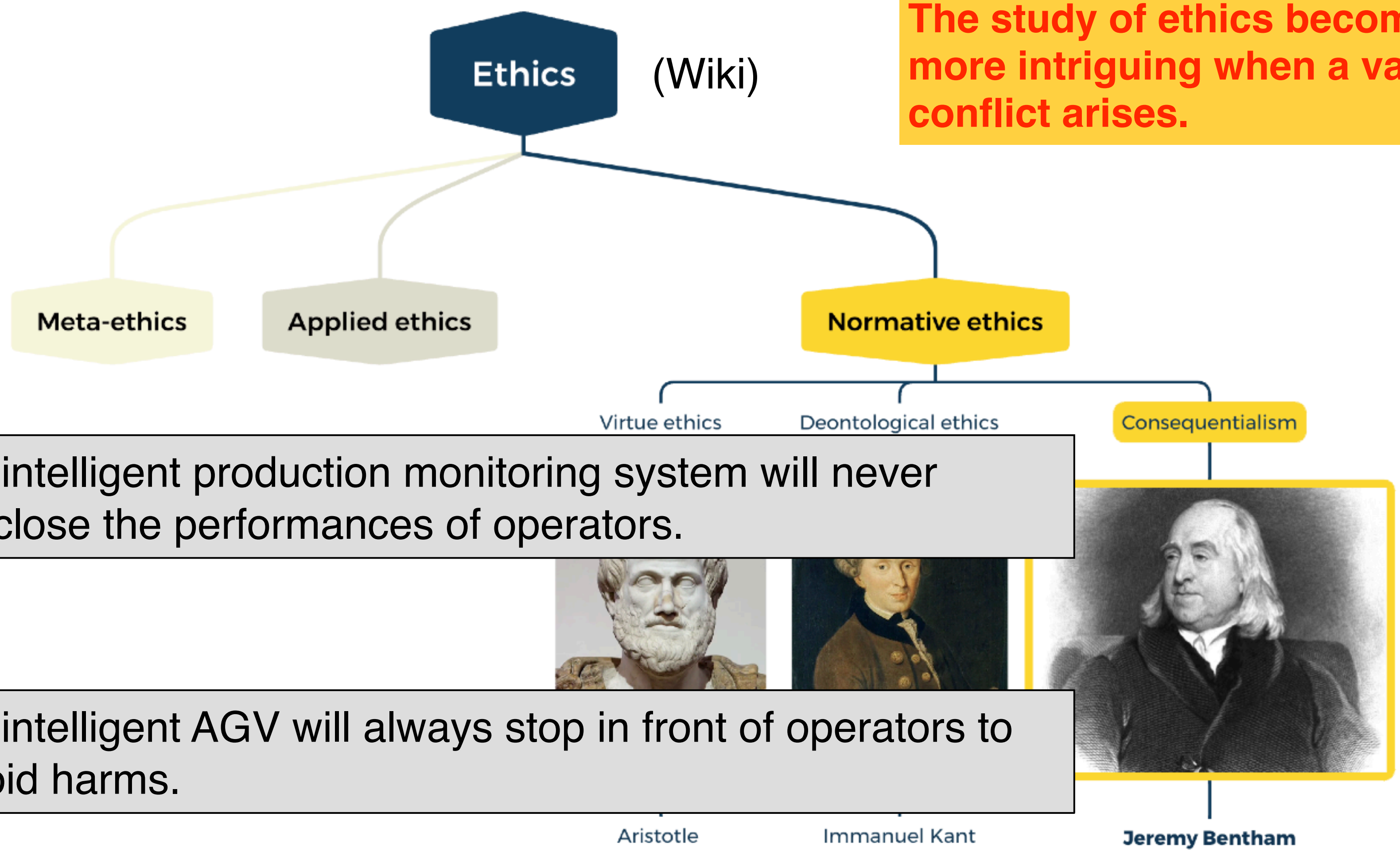
It may reflect ethical principles such as:

- Respect for human autonomy (the human remains in meaningful control).
- Non-maleficence (avoid harm caused by stress, confusion, or poor decisions).
- Transparency and explainability (the system ensures the human’s understanding).
- Accountability (humans must be able to make responsible, informed actions).

Trentesaux, D., & Caillaud, E. (2020). Ethical stakes of Industry 4.0. *IFAC-PapersOnLine*, *53*(2), 17002-17007.

Typology of Ethics

The study of ethics becomes more intriguing when a value conflict arises.



An intelligent production monitoring system will never disclose the performances of operators.

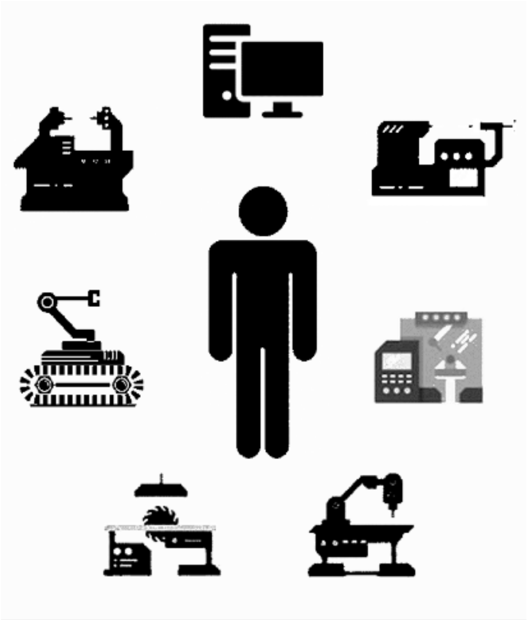
An intelligent AGV will always stop in front of operators to avoid harms.

What kind of person should I be?

What rules must I follow?

What outcome will my action bring?

Major challenge



CPS-based smart manufacturing

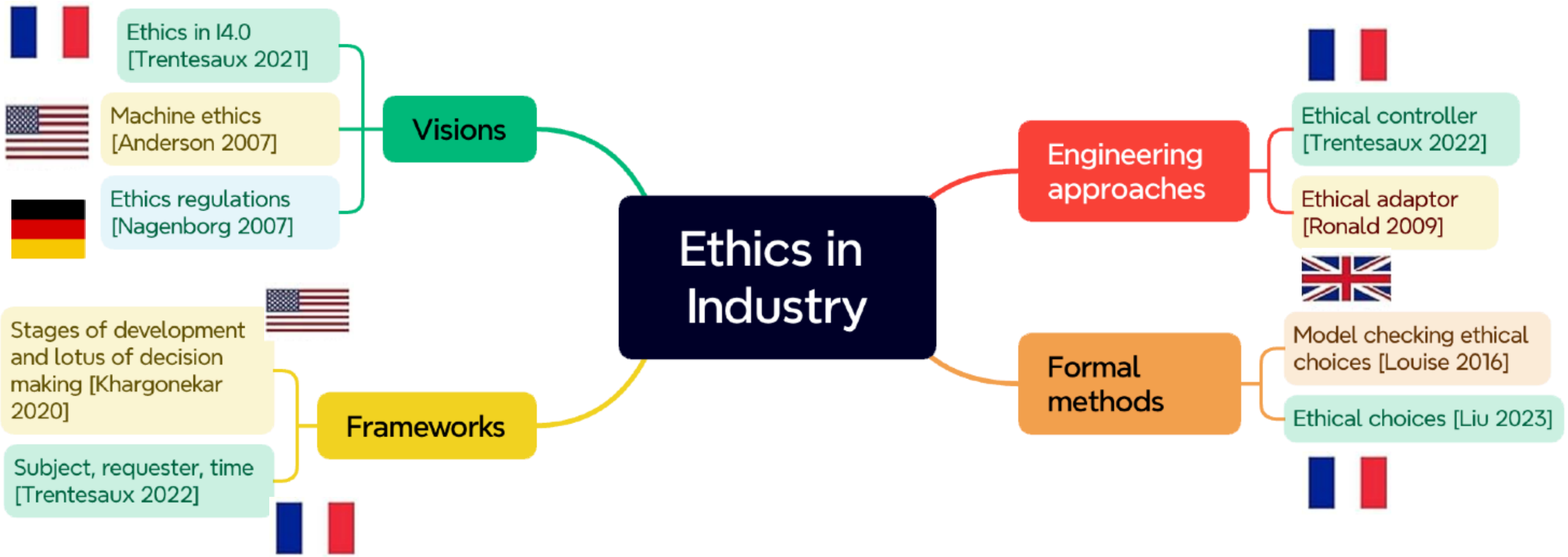
How can we enhance human trust and acceptance of decisions made by intelligent entities?

Philosophical vs. Computational

Ethics

How can we integrate ethics into the decision-making processes of intelligent entities?

State of the art



Ethics in Industry

Issue 1: Industry 4.0 is a source of complexity

- AI, the environment, stakeholders, humans

Issue 2: Industry 4.0 is a source of risks for humans involved

- Safety issues
- Loss of skills and the replacement of humans

Ethics-related domains

- Systems engineering and safety studies
- Human-machine systems
- Robotics, embedded systems and cyber-physical systems
- Computer science and AI

They realized the obvious lack of scientific, technical, operational and mature contributions in the field of ethics when designing or imagining future industrial systems.

Trentesaux, D., & Caillaud, E. (2020). Ethical stakes of Industry 4.0. *IFAC-PapersOnLine*, *53*(2), 17002-17007.

Machine Ethics

Ethical-impact agents

Privacy, property, and power, etc.

Implicit ethical agents

Internal functions implicitly promote ethical behavior

Explicit ethical agents

Represent ethical categories, operate based on these

Full ethical agents

Make explicit ethical judgements and generally competent to reasonably justify them

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE intelligent systems*, *21*(4), 18-21.

Vision

Machine ethics is concerned with ensuring that the behavior of machines towards human users, and perhaps other machines as well, is ethically acceptable. [Anderson 2007]

Why machine ethics?

- Preventing harm from autonomous machines
- Building trust in AI and robotics
- Advancing ethical theories

How to build an explicit ethical agent?

- Choose an ethical framework
- Select a concrete domain and dilemma type
- Define case profiles and a representation for duty satisfaction/violation
- Learning a decision principle that resolves duty conflicts
- Implement the decision procedure and embed it in a system
- Evaluate against expert judgement and iterate

Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI magazine*, *28*(4), 15-15.

Ethical Regulation

- Responsibility and “autonomous robots”
  - Machine safety
  - Responsibility for complex machines
  - Roboethics and machine ethics (ethical norms)

- Machines as a replacement for humans

- Tele-presence (effects on other countries)

Nagenborg, M., Capurro, R., Weber, J., & Pingel, C. (2020). Ethical regulations on robotics in Europe. In *Machine ethics and robot ethics* (pp. 473-490). Routledge.

Ethics in CPHS

- **Stage of development** (how mature the technology is)
  - Basic research (control, communication, algo.)
  - Early stage domains (brain-machine interfaces, autonomous greenhouses)
  - Developing domains (unmanned civilian aircraft, smart and connected vehicles)
  - Mature domains (aerospace, automotive, manufacturing)

- **Locus of decision making** (who is making decisions)
  - Individuals (engineers, researchers, policymakers)
  - Organisations (companies, universities, nonprofits)
  - Governments (regulators, lawmakers)

Khargonekar, P. P., & Sampath, M. (2020). A framework for ethics in cyber-physical-human systems. *IFAC-PapersOnLine*, *53*(2), 17008-17015.

Ethics in Industrial CPS

- **Subject** (who is ethically relevant — e.g. humans, environment, organizations)
- **Requester** (stakeholders who ask for or demand functionalities or constraints)
- **Time** (stage of development)

Most of the frameworks are still at their first steps of development and need improvements to gain maturity and applicability.

Trentesaux, D., Caillaud, E., & Rault, R. (2021, November). A framework fostering the consideration of ethics during the design of industrial cyber-physical systems. In *International Workshop on Service Orientation in Holonic and Multi-Agent Manufacturing* (pp. 349-362). Cham: Springer International Publishing.

Engineering approaches

The Emergence of Ethics Engineering

Motivations

Defintion

High-level architecture

No detail about implementation is given.

Leitão, P., & Karnouskos, S. (2022, May). The emergence of ethics engineering in industrial cyber-physical systems. In *2022 IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS)* (pp. 1-6). IEEE.

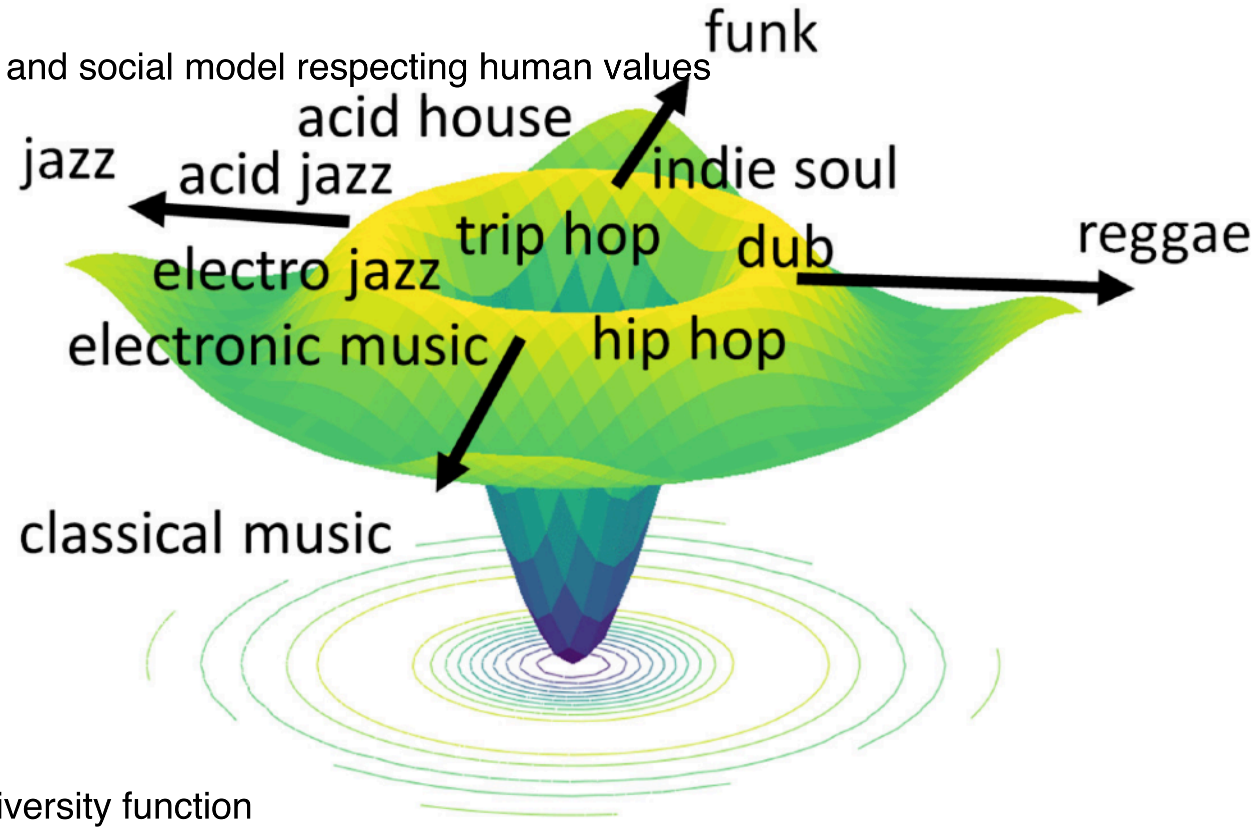
Ethics-by-Design method

Step 1: Innova

Economic and social model respecting human values

Step 2: Ethica

Respect for:
Consumers (Listeners)
Creators of Cultural Goods (Artists)
Media Services Provider
Indie Labels



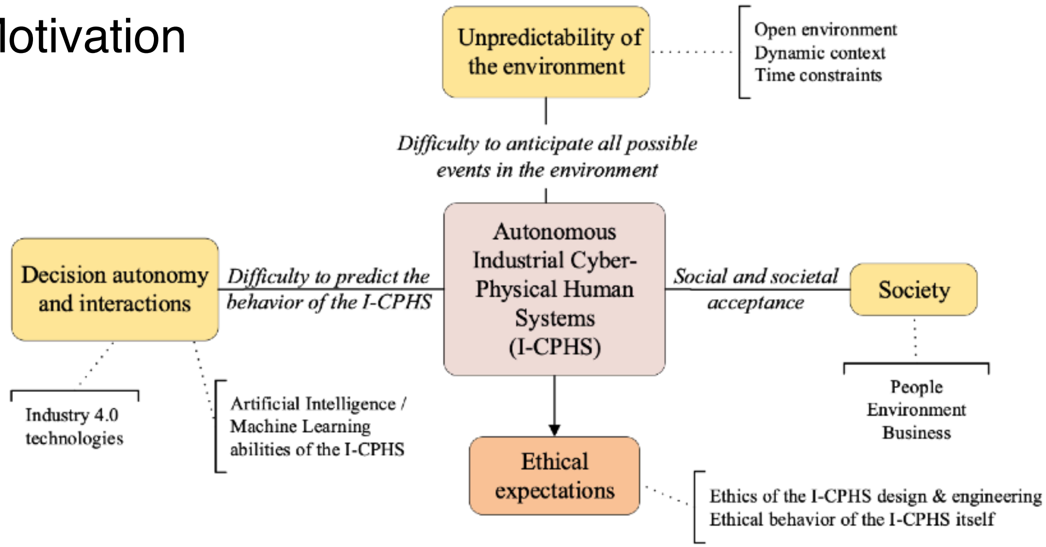
Step 3: Value

Optimal diversity function

Muhlenbach, F. (2020, October). A methodology for ethics-by-design AI systems: dealing with human value conflicts. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1310-1315). IEEE.

Engineering approaches

Motivation



Ethical behavior

"*An autonomous industrial cyber-physical human system (I-CPHS) is said to behave ethically, if the emergent behavior of the overall system, decides and acts according to ethical expectations expressed by all the stakeholders involved or impacted by its activities.*"

Autonomous Industrial CPS

Ethical controller

- Monitor events and contexts
- Trigger ethical reasoning strategies
- Following deontology ethics and consequentialism

Very few researchers work on Ethics Engineering.

D. Trentesaux and S. Karnouskos, "Engineering ethical behaviors in autonomous industrial cyber-physical human systems" Cognition, Technology & Work, vol. 24, no. 1, pp. 113–126, 2022.

Formal verification

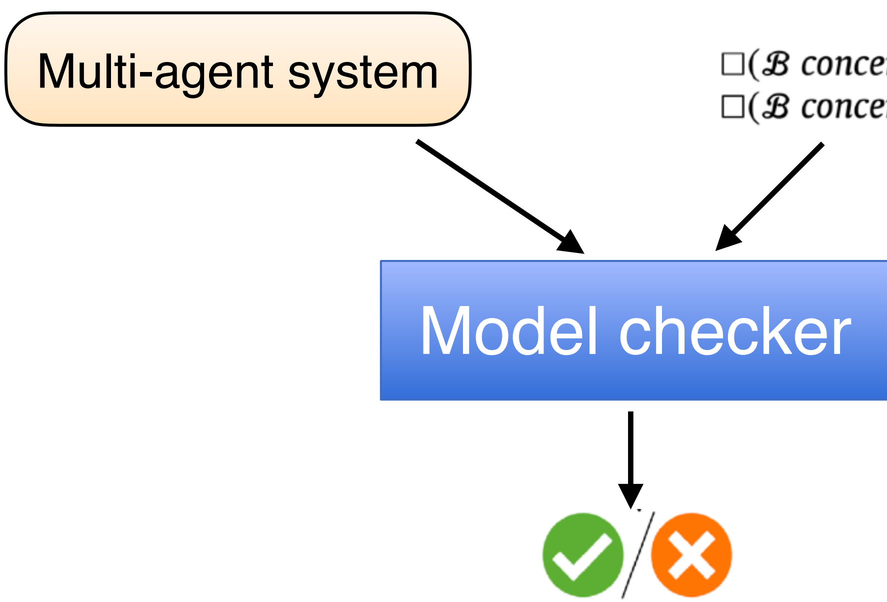
MCAPL (Model Checking Agent Programming Language framework)

GWENDOLEN

BDI (Beliefs, Desires, Intentions)

ϕ_1 = do not violate turn right rule (2);
ϕ_2 = do not stay above 500 feet rule (2);
ϕ_3 = do not collide with objects on the ground (3);
ϕ_4 = do not collide with aircraft (4).

$$\Box(\mathcal{B}\ concern(\phi_1) \rightarrow \mathcal{B}\ other\_choices\_violate([\phi_1, \phi_2, \phi_3, \phi_4]))$$
$$\Box(\mathcal{B}\ concern(\phi_2) \rightarrow \mathcal{B}\ other\_choices\_violate([\phi_2, \phi_3, \phi_4]))$$



Dennis, L., Fisher, M., Slavkovik, M., & Webster, M. (2016). Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems*, *77*, 1-14.

Formal verification

Our previous work was based on the work of [Denise 2016].

Ethical Rule Definition:

:**Ethical policy**:
E(context, ethical rule, severity)

:**Initial Beliefs**:
e(ethical rule, severity, probability)

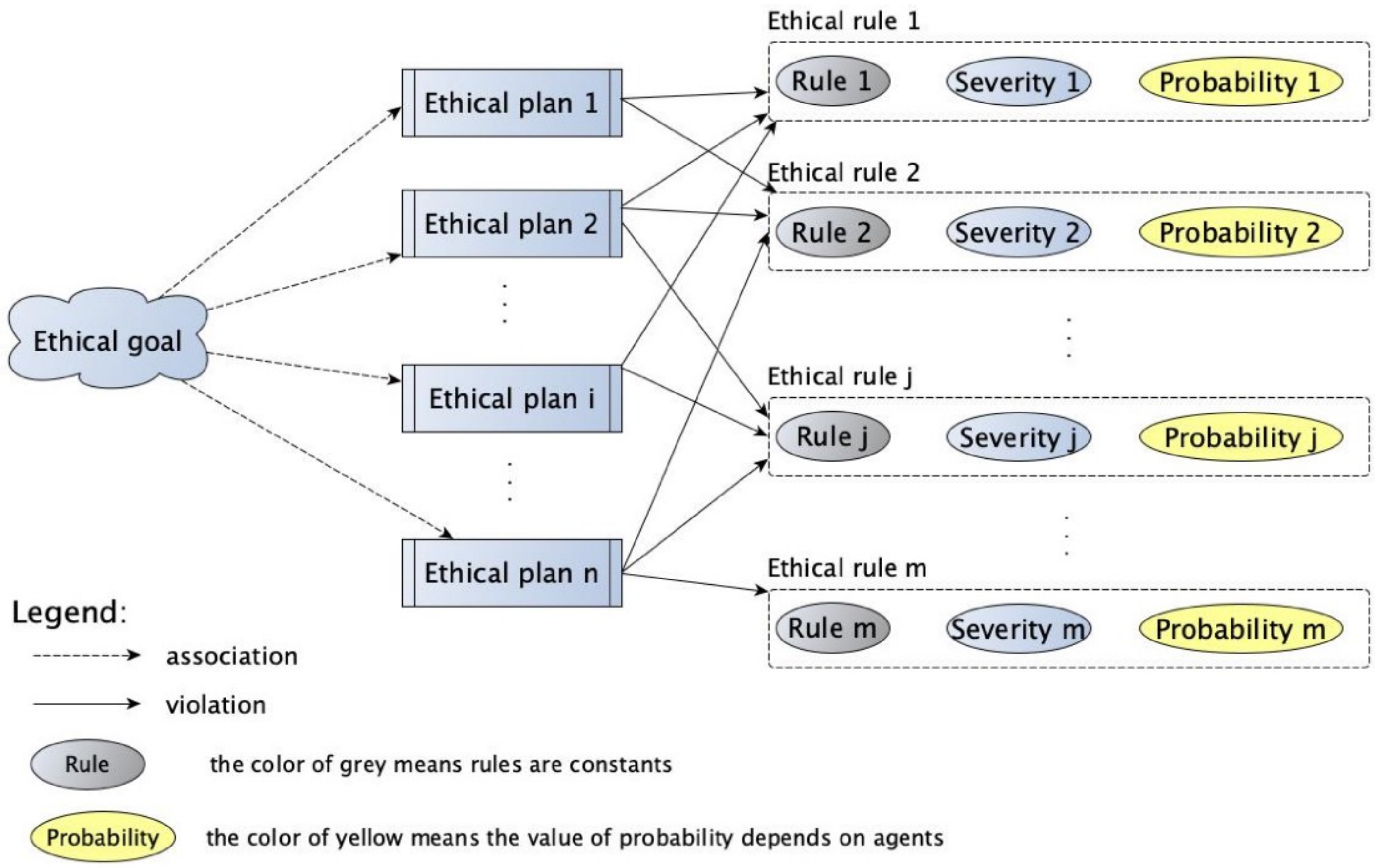
Case study:

Rules of the Air

TRACILOGIS Platform
(Intelligent Production)

Liu, Y., & El Haouzi, H. B. (2023, October). Formal verification of ethical choices in industrial cyber-physical systems. In *IEEE Conference on Systems, Man, and Cybernetics, SMC 2023*.

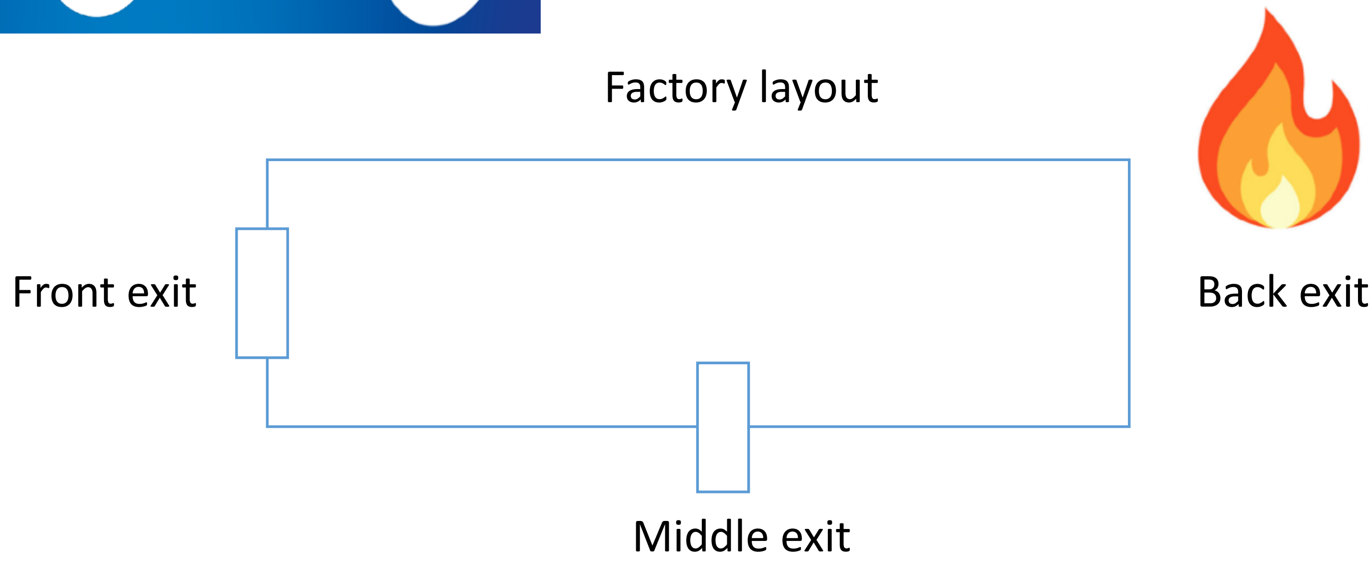
Formal verification



$$p_i = \{\forall j \in X, j \neq i, \sum_{k=1}^{j} S_k * P_k \geq \sum_{l=1}^{i} S_l * P_l\}$$

Formal verification

Factory layout

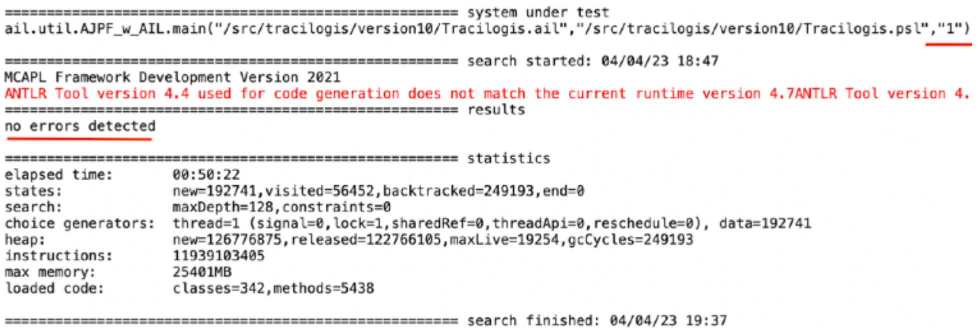


$$1: (B(iss, ethical\_rules\_1) -> <> (D(iss, delete(fire\_emergency(1)))))$$

$$2: (B(iss, ethical\_rules\_2) -> <> (D(iss, delete(fire\_emergency(4)))))$$

$$3: [](B(iss, fire) -> \sim B(prod, start))$$

R1: e(avoidWorkersInjured,3,1)
R2: e(avoidWorkersDied,10,1)



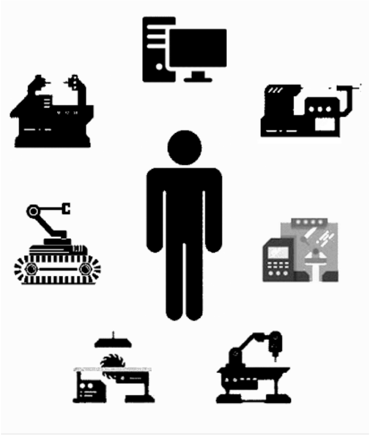
```
===================================================== system under test
ail.util.AJPF_w_AIL.main("/src/tracilogis/version10/Tracilogis.ail","/src/tracilogis/version10/Tracilogis.psl","1")

===================================================== search started: 04/04/23 18:47
MCAPL Framework Development Version 2021
ANTLR Tool version 4.4 used for code generation does not match the current runtime version 4.7ANTLR Tool version 4.
===================================================== results
no errors detected

===================================================== statistics
elapsed time:        00:50:22
states:              new=192741,visited=56452,backtracked=249193,end=0
search:              maxDepth=128,constraints=0
choice generators:   thread=1 (signal=0,lock=1,sharedRef=0,threadApi=0,reschedule=0), data=192741
heap:                new=126776875,released=122766105,maxLive=19254,gcCycles=249193
instructions:        11939103405
max memory:          25401MB
loaded code:         classes=342,methods=5438

===================================================== search finished: 04/04/23 19:37
```

Simulation results

Context

The complexity of interactions between intelligent entities and humans, made possible by the technological advances of Industry 4.0, will lead to decision-making situations requiring the use of simulation/emulation models in symbiosis with production and logistics control systems

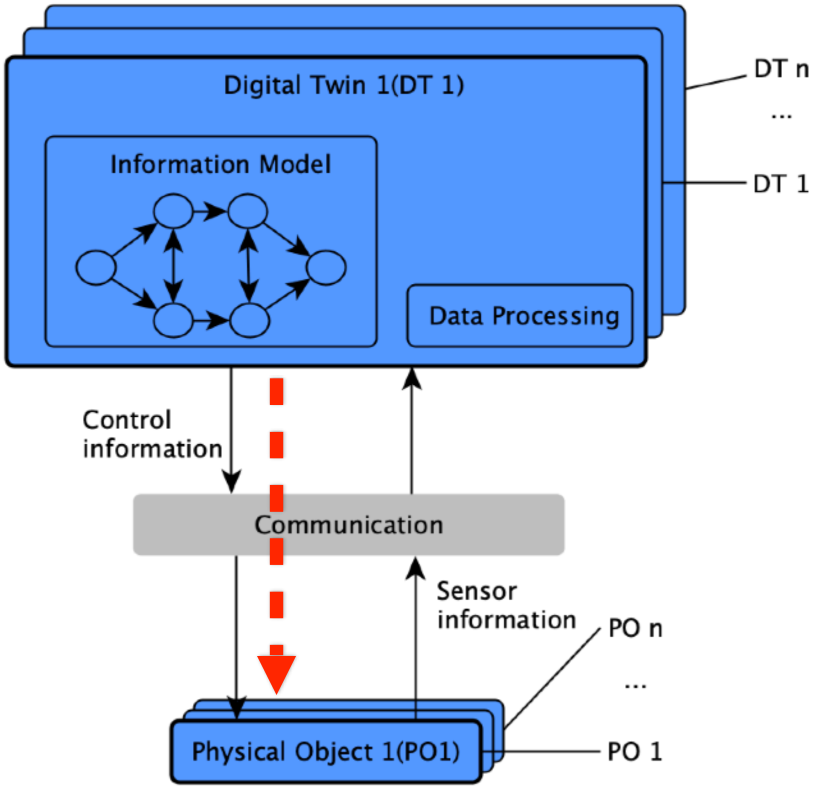
⇨ Digital Twin



The integration of humans, from design to use, is driven by uncertainties in data and their interpretation, in usage situations, in the consequences of decisions, and by the vision of a world that must remain a social system.

⇨ Human centered

Motivation



Surveillance, Simulation, Optimisation, etc.

Very few DTs have included the control from DTs to physical objects. [Lu 2020]

The users don't trust the decisions made by DTs.

Models

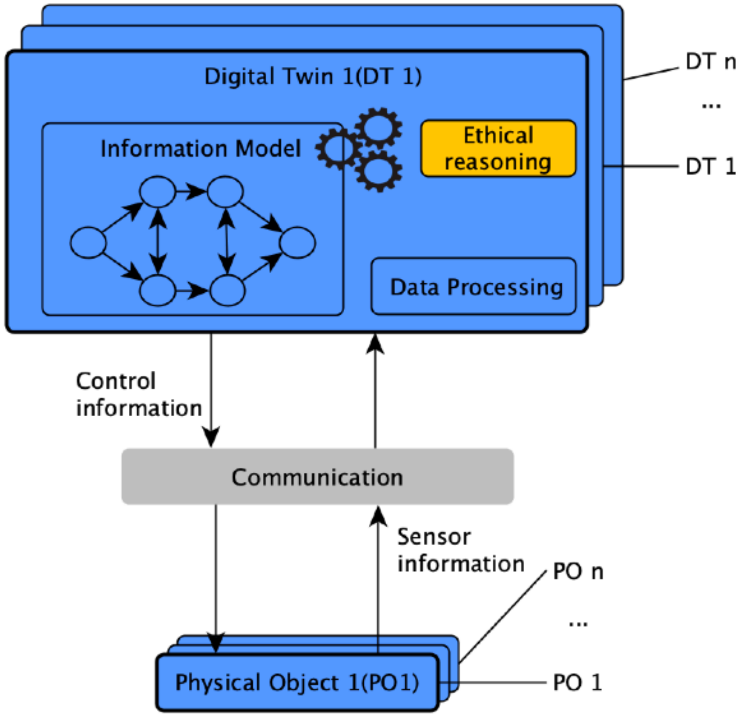
Ethical aspect

Adopting DT's decisions requires trust in the model, trust in the data, and trust in the algorithms used to update the model based on the data. [Wright 2020].

How to integrate ethical considerations into the model of DT?

Lu, Y., et al (2020). Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues. RCIM 61, 101837.
Wright, L., & Davidson, S. (2020). How to tell the difference between a model and a digital twin. Adv. M&S in Engineering Sciences, 7, 1-13.

Digital Ethical Twin (DET)



This project aims at developing digital ethical twins to enable DTs to instinctively conduct ethical reasoning about their actions so that humans can trust DTs more.

Ethical reasoning is the process for determining whether the actions of DTs are morally acceptable regarding ethical principles.

For example, an autonomous robot would choose a route with fewer workers but a longer distance to transport materials in a production line to avoid harming workers.

DET vs. Cognitive Digital Twin

DET

Concept: it integrates **ethical principles** into decision-making processes for the intention of maximizing the good and minimizing the bad.

Focus: Alignment of ethics — making sure the twin respects and replicates the same moral standards as the original (e.g., fairness, human well-being, respecting rights).

Examples:
- Testing how different ethical systems handle dilemmas (like trolley problems).
- Governance & compliance: ensuring AI behaves in a way that matches human legal or cultural norms.

Cognitive Digital Twin

Concept: A digital representation, augmentation and intelligent companion of its physical twin as a whole, including its subsystems and across all of its life cycles and evolution phases. [Adl 2016]

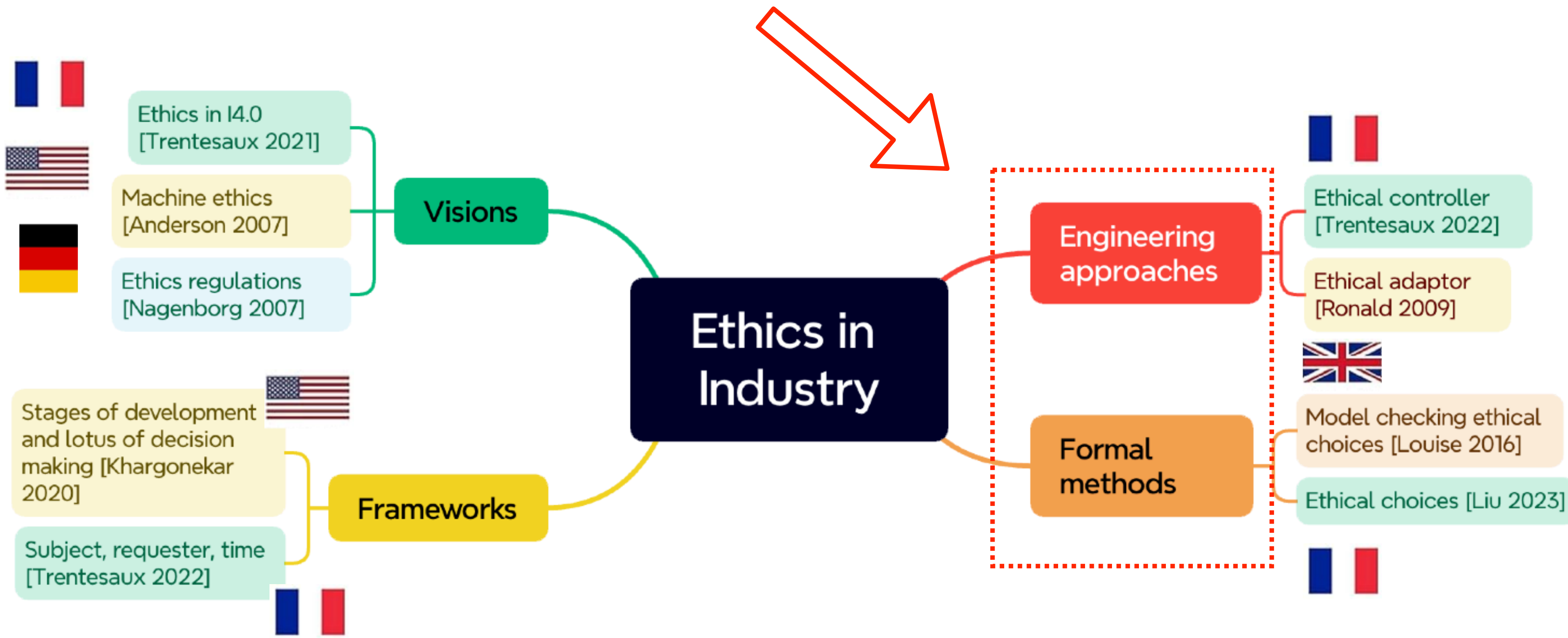
Focus: Alignment of **intelligence and cognition**, not values. The goal is to capture how someone *thinks*, not necessarily how they *ought* to act.

Example:
- Personalization (AI assistant that reasons like you).
- Simulation of decision-making in complex environments (business strategy, healthcare)

Adl, Ahmed El. 2016. The Cognitive Digital Twins: Vision, Architecture Framework and Categories. Technical Report. https://www.slideshare.net/slideshow/embed_code/key/JB6 0Xqcn.

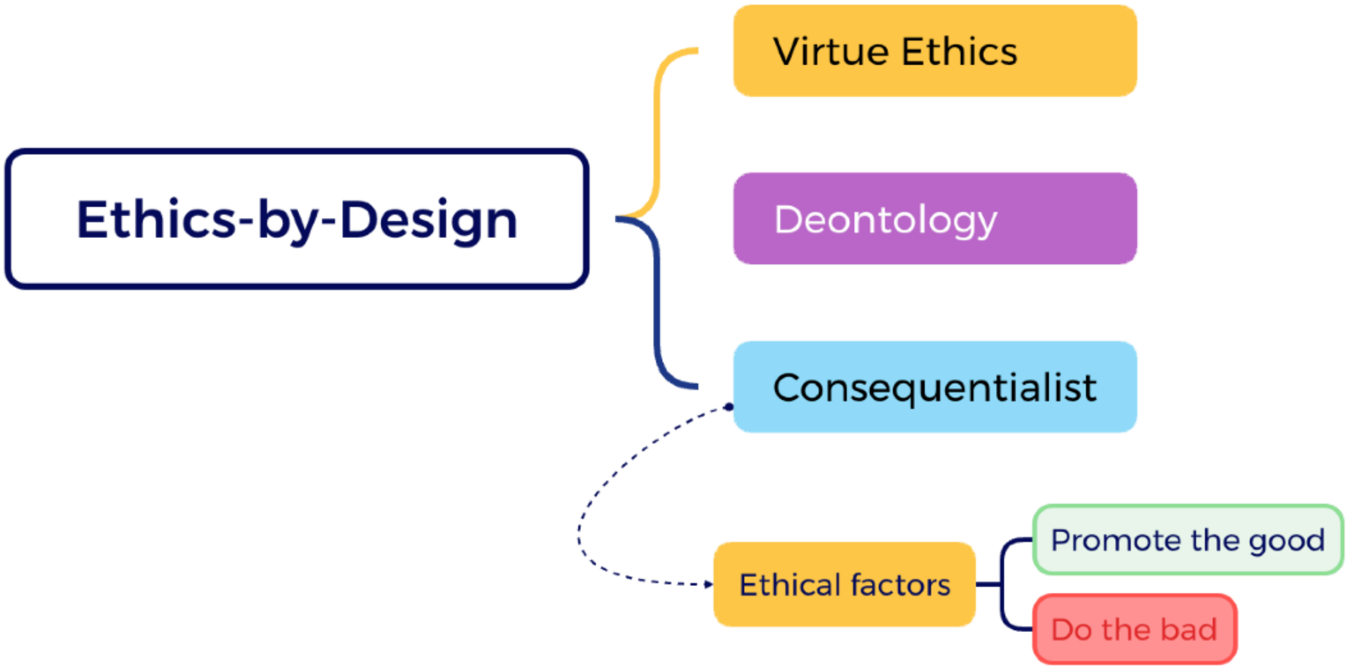
Position

DET : ethical reasoning + guarantee of ethical properties + decentralization



Arkin, Ronald C., and Patrick Ulam. "An ethical adaptor: Behavioral modification derived from moral emotions." *2009 IEEE international symposium on computational intelligence in robotics and automation-(CIRA)*. IEEE, 2009.
Louis. D., M. Fisher, M. Slavkovik, and M. Webster, "Formal verification of ethical choices in autonomous systems," Robotics and Autonomous Systems, vol. 77, pp. 1–14, 2016

Integration of ethics



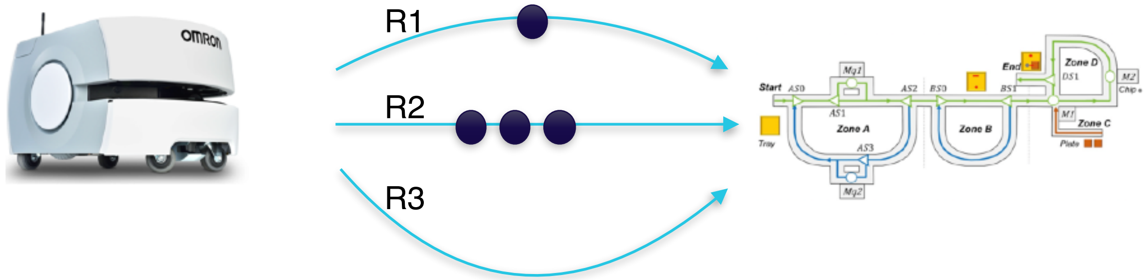
Ethical risk refers to the risk that the actions or decisions of agents lead to the violation of the following consequentialist principles [Ogien 2009]:

- Ideally, an agent performs an action if and only if that action promotes good;
- At a minimum, an agent performs an action if and only if that action entails causing the least possible harm.

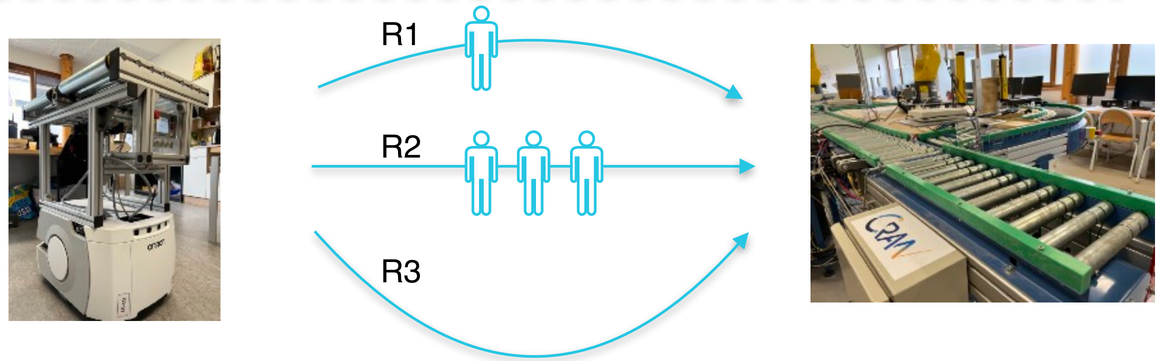
[OGIEN 2009] OGIEN, Ruwen et TAPPOLET, Christine. Les concepts de l'éthique. Book in 2009.

Ethical reasoning

Virtual world



Physical world



How does the robot make the decision?

Ethical factors
- Promote the good
  - Productivity
  - Capability of utilisation
  - ...
- Do the bad
  - Collision
  - Mental charge
  - ...

Ethics by Design (EbD)

Principles: These are the ethics principles/ethical values detailed in Part 1. An AI system is considered unethical if it violates these principles/values.

Ethical Requirements: These are the conditions that must be met for the AI system to achieve its goals ethically. These may be instantiated in many ways: through functionality, in data structures, in the process by which the system is constructed,with organisational safeguards.

Ethics by Design Guidelines: These are concerned with the processes for creating the system. In many cases guidelines are specific tasks which must be completed at specific points in the development process. The guidelines are either implementations of ethics requirements, or broader guidelines for different stages of developments that help ensure proper implementation of requirements.

AI Methodologies: There is a variety of methodologies used in AI and robotics projects (AGILE, CRISP-DM, V-Method etc). These are, at least partially, distinguished by the manner in which the development process is organized. Each methodology offers its own steps and sequence. Ethics by Design maps its guidelines onto the steps in each individual methodology.

Tools & Methods: specific tools and processes within the development process. For example, data-sheets for datasets can be employed to assess the ethical characteristics of data.

European Commission 2021 Ethics By Design and Ethics of Use Approaches for Artificial Intelligence.

Ethical principles

1. **Respect for Human Agency**: human beings must be respected to make and carry out their own decisions.
2. **Privacy and Data governance**: people have the right to privacy and data protection.
3. **Fairness**: people should be given equal rights and opportunities and shouldn't be advantaged or disadvantaged undeservedly.
4. **Individual, Social and Environmental Well-being**: AI systems should contribute to, and not harm, individual, social and environmental wellbeing.
5. **Transparency**: The purpose, inputs, and operations of AI programs should be knowable and understandable to its stakeholders.
6. **Accountability and Oversight**: Humans should be able to understand, supervise and control the design and operation of AI-based systems and the actors involved in their development or operation should take responsibility for the way that these applications function and for the resulting consequences.

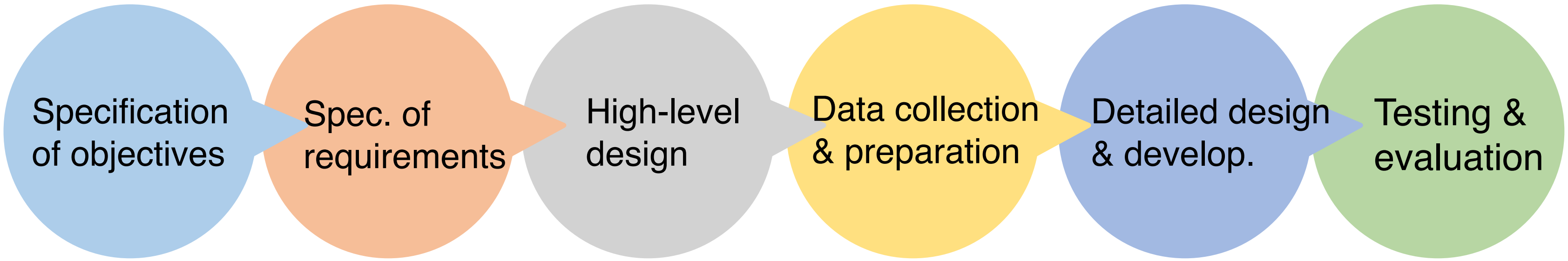
European Commission 2021 Ethics By Design and Ethics of Use Approaches for Artificial Intelligence.

Ethical requirements

1. **Respect for Human Agency**: AI systems must not negatively affect human autonomy, freedom or dignity.
2. **Privacy and Data governance**: AI systems must not violate the right to privacy and to personal data protection. They MUST use data which is necessary, non-biased, representative and accurate.
3. **Fairness**: AI systems must be developed with an inclusive fair, and non-discriminatory agenda.
4. **Individual, Social and Environmental Well-being**: Steps must be taken to ensure that AI systems do not cause individual, social or environmental harm, rely on harmful technologies, influence others to act in ways which cause harm or lend themselves to function creeps.
5. **Transparency**: AI systems should be as transparent as possible to their stakeholders and to their end-users.
6. **Accountability and Oversight**: Human oversight and accountability are required to ensure conformance to these principles and address non-compliance.

European Commission 2021 Ethics By Design and Ethics of Use Approaches for Artificial Intelligence.

Generic model for development



1. **Specification of objective**: The determination of what the system is for and what it should be capable of doing.
2. **Specification of requirements**: Requirements for building the system, including initial determination of required resources, together with an initial risk requirement and cost-benefit analysis, resulting in a design plan.
3. **High-level design**: conceptual model and architecture model.
4. **Data collection and preparation**: Collection, verification, cleaning and integration of data
5. **Detailed design and development**: implementation.
6. **Testing and evaluation**: system testing and evaluation.

European Commission 2021 Ethics By Design and Ethics of Use Approaches for Artificial Intelligence.

EbD example 1

A financial app uses AI to suggest investment options

Ethical principle

Respect for Human Agency: human beings must be respected to make and carry out their own decisions.

Ethical requirement

Respect for Human Agency: AI systems must not negatively affect human autonomy, freedom or dignity.

Generic model

Specification of objective: The system should provide clear explanations of each recommendation and let the user decide whether to act.
…
…

European Commission 2021 Ethics By Design and Ethics of Use Approaches for Artificial Intelligence.

EbD example 2

Care robots for lifting patients and their impact on the moral elements of care.

Phase of care	**Description**	**Value**
Caring about	Recognizing that someone has a need that requires attention.	**Attentiveness** - noticing and understanding the needs of others
Taking care of	Accepting responsibility for responding to those needs and deciding how to meet them.	**Responsibility** - committing to act on those needs
Care giving	Actually providing the care or taking the action to meet the need.	**Competence** - giving care skillfully and effectively
Care receiving	Observing how the care is received and whether the need was met; allowing for feedback.	**Responsiveness** - being open to the care receiver's reactions and adjusting accordingly

Van Wynsberghe, A. (2020). Designing robots for care: Care centered value-sensitive design. In Machine ethics and robot ethics (pp. 185-211). Routledge.

EbD example 2

Care robots for lifting patients and their impact on the moral elements of care.

Ethical principle

Individual, Social and Environmental Well-being: AI systems should contribute to, and not harm, individual, social and environmental wellbeing.

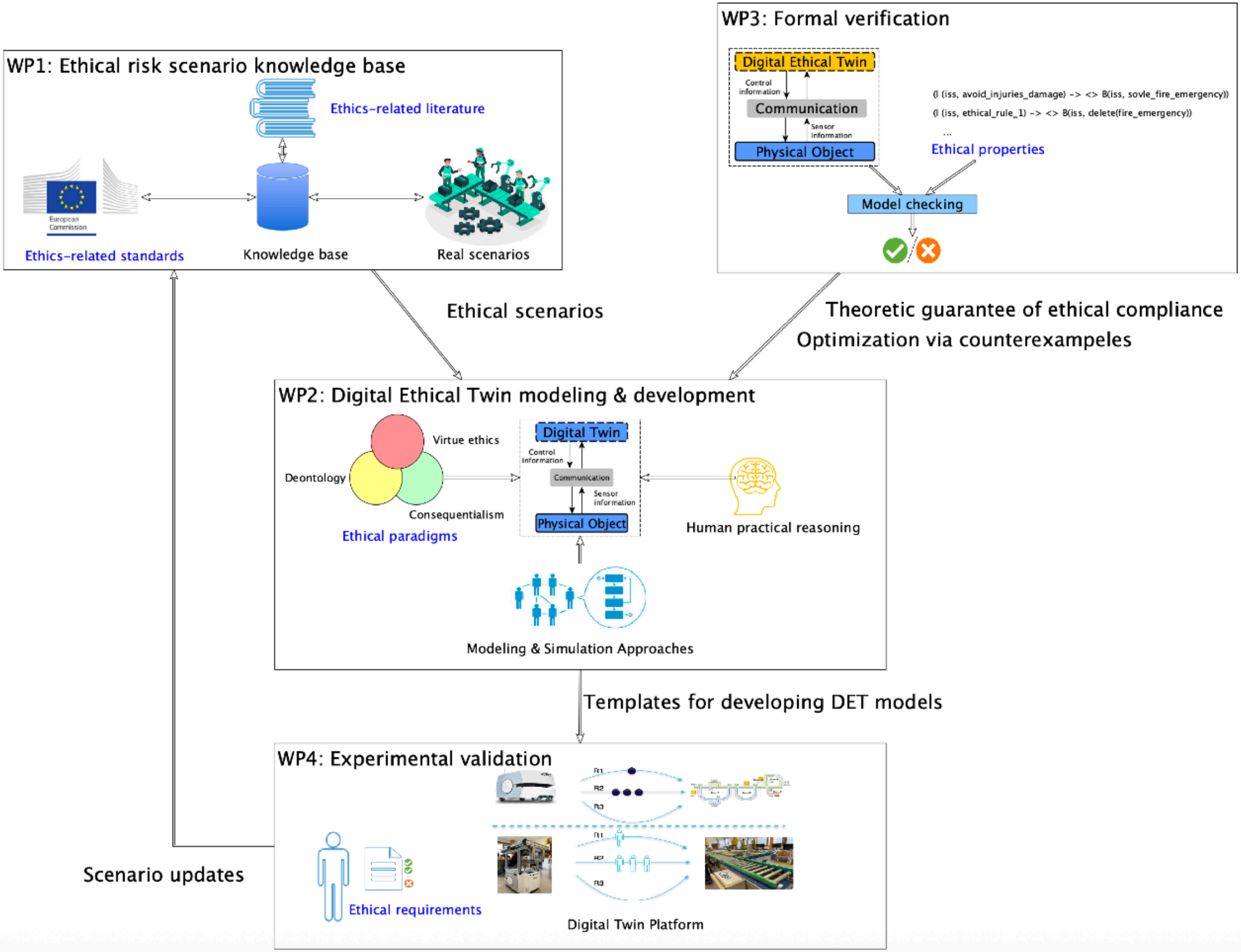
Ethical requirement

Individual, Social and Environmental Well-being: Steps must be taken to ensure that AI systems do not cause individual, social or environmental harm, rely on harmful technologies, influence others to act in ways which cause harm or lend themselves to function creeps.

Generic model

Specification of objective: Robots must be able to detect relevant cues that indicate a care receiver's needs.
…
…

DET methodology



Work organization

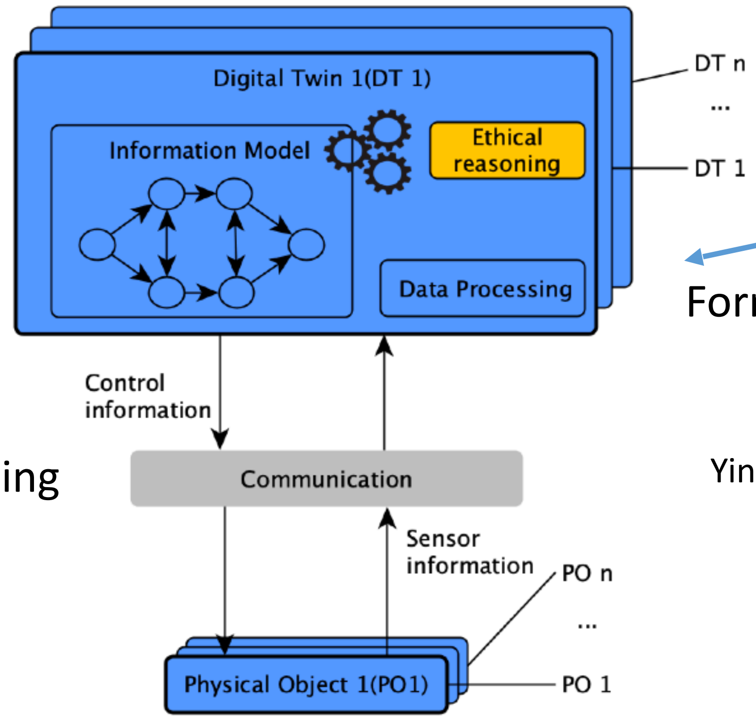
V. Govaere, INRS, Nancy Govaere, V., et al. (2019). Smart planning-approaching the characteristics of a valid, balanced transport round. In *Congress of the International Ergonomics Association*

A. L. Halftermeyer, Univ. d'Orléans Lefeuvre-Halftermeyer, et al. (2016). Typologie des risques pour une analyse éthique de l'impact des technologies du TAL. *Revue TAL: traitement automatique des langues*, *57*(2), 47-71.

1

Knowledge Base

Ethical factors



3

Model checking

Formal guarantees

Ethical reasoning

Human-thinking theory

2

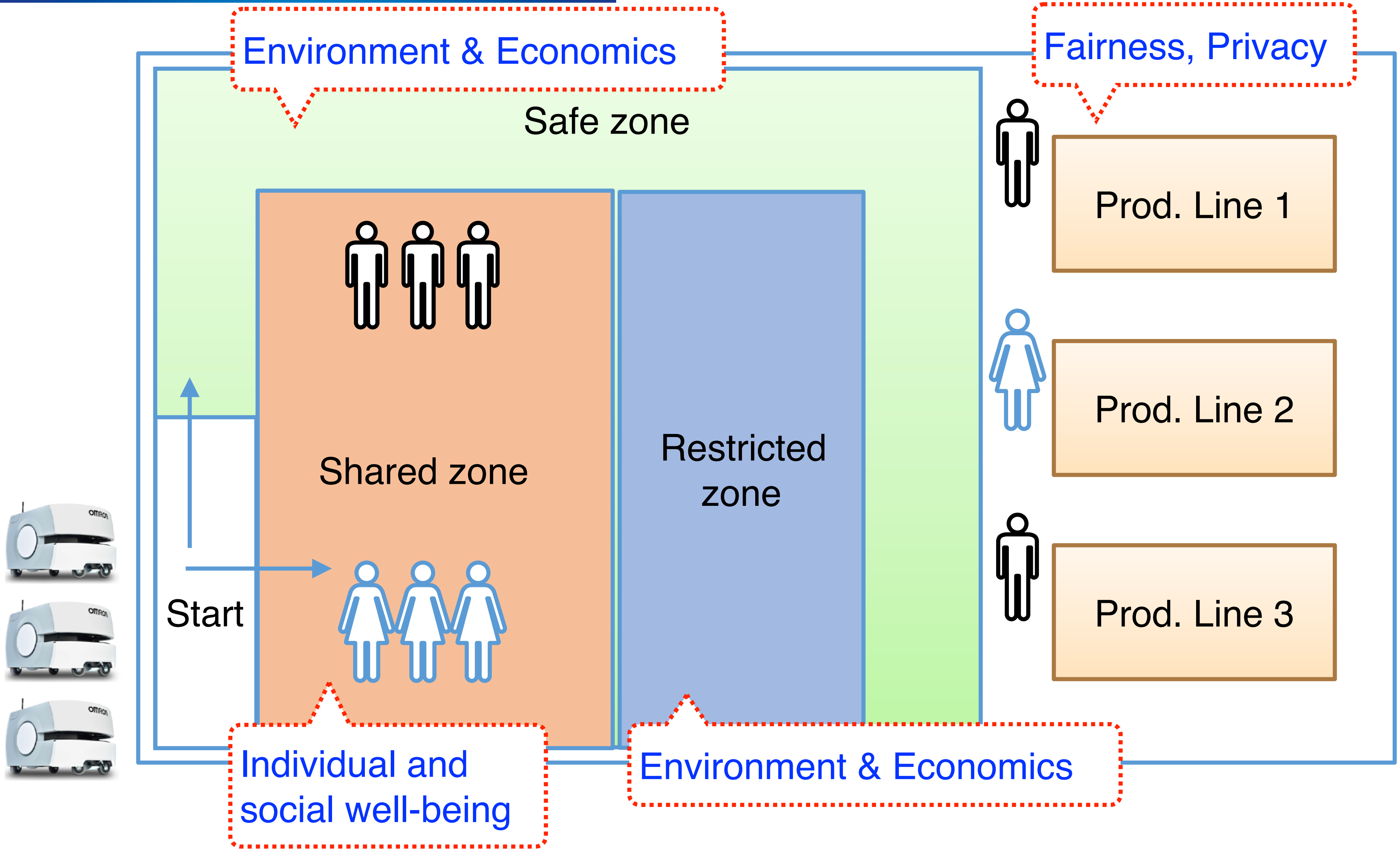
Yinling LIU, Hind Bril El Haouzi, CRAN

LIU Yinling, Hind Bril El Haouzi. Formal verification of ethical choices in industrial cyber-physical systems. IEEEConference on Systems, Man, and Cybernetics, SMC 2023, Oct 2023, Hawaii-Honolulu, United States

Hind Bril El Haouzi, CRAN El-Haouzi, H. B., & Valette, E. (2021). Human system integration as a key approach to design manufacturing control system for Industry 4.0: Challenges, barriers, and opportunities. IFAC-PapersOnLine, 54(1), 263-268.

Yinling LIU, CRAN LIU, Yinling, MOYAUX, Thierry, BOULEUX, Guillaume, et al. Hybrid simulation modelling of emergency departments for resource scheduling. Journal of Simulation, 2023, p. 1-16.

Case study



How would the robots make decisions about the route and the production line?

Thank you!

Q&A

https://liuyinling.github.io/