

Autonomia nei Sistemi ed Etica delle Decisioni: La Prospettiva dell'Utente



Paola Inverardi

Software Engineering and Architecture Group

Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica,
University of L'Aquila, Italy

2001-2021: Ventennale Informatica Applicata, Urbino 13-09-2021.

The scope of self

- Software systems are increasingly **autonomous** in making decisions (on behalf of potential users or pro-actively).
- The power of **self** goes beyond the ability of substituting human agents in supplying (contextual) information that the system may use to make decisions while continuously running.
- Depending on the nature, **property**, and use of this information, an autonomous system may impact **moral rights** of the users, be they *single citizens, groups, or the society as a whole*



- It exceeds the system boundaries invading **user prerogatives**

Privacy and Ethics

Privacy

- It emerged with the large scale availability of automatically processable personal data
- Philosophical, regulatory and technical approaches
- It is an **ethical** dimension

Philosophical Perspective

- privacy as related to **personal** information on which we want to exercise **direct** and **unconditional** control concerning its **diffusion** and **disclosure**

W. Prosser. Privacy - California Law Review 1960.

B. Roessler. Xprivacy as a human right. Proceedings of the Aristotelian Society, 117(2):187–206, 2017

Regulatory Perspective

- General Data protection Regulation (GDPR) (May 2018)
- Art.1
 - Regulation lays down rules relating to the *protection of natural persons* with regard to the **processing** of personal data and rules relating to the **free movement** of personal data.
 - This Regulation protects **fundamental rights** and **freedoms of natural persons** and in particular their **right** to the protection of personal data.
 - ...
- Art. 2
 - This Regulation applies to the **processing** of personal data **wholly** or **partly** by **automated means** and to the processing other than by automated means of personal data which form part of **a filing system** or are intended to form part of a filing system.
 - ...

Technical Perspective

- **Privacy by design** provides high-level guidelines in the form of principles for designing privacy-preserving systems
- Privacy preferences have been historically implemented by means of **permission systems** that comprise both specification of access policies and their enforcement
- **User involvement**: users *nudged* towards better solutions. Soft Paternalism principles
 - A. Acquisti, et. al. Nudges for privacy and security: Understanding and assisting users' choices online. ACM Comput. Surv. , 50(3):44:1–44:41, Aug. 2017.
- **Privacy persona** characterizing groups of users by privacy preferences
- **Privacy assistant** human or virtual, S. Ovide. How to make data privacy real, New York Times January 19 2021

Ethics

- *“Advances in **AI, robotics** and so-called ‘**autonomous**’ technologies have ushered in a range of increasingly urgent and complex moral questions. **Current efforts** to find answers to the ethical, societal and legal challenges that they pose and to orient them for the **common good** represent a **patchwork of disparate initiatives**. This underlines the need for a collective, wide-ranging and inclusive process of reflection and dialogue, a dialogue that focuses on **the values** around which we want to organise society and on the role that technologies should play in it. “*
 - European Group on Ethics in Science and New Technologies. statement on artificial intelligence, robotics and ‘autonomous’ systems.
https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf, 2018

The autonomous car case

- Ethical problems: The trolley problem
 - The trolley problem <http://www.trolleydilemma.com>
 - MIT Moral machine <http://moralmachine.mit.edu>
 - E. Awad et others, The Moral Machine experiment, Nature volume 563, pages59–64 (2018), October 2018
- Philosophical debate
 - Mandatory ethics vs Ethical Knob
 - J. Gogoll and J. F. Müller. Autonomous cars: In favor of a mandatory ethics setting. Science and Engineering Ethics, 23(3):681–700, Jun 2017.
 - G. Contissa, F. Lagioia, and G. Sartor. The ethical knob: ethically-customisable automated vehicles and the law. Artificial Intelligence and Law, 25(3):365–378, 2017

THE TROLLEY DILEMMA

and how it relates to ethical communication

About The Trolley Dilemma

The "Trolley Dilemma" (or the "Trolley Problem") consists of a series of hypothetical scenarios developed by British philosopher Philippa Foot in 1967. Each scenario presents an extreme environment that tests the subject's ethical prowess. In 1985, American philosopher Judith Jarvis Thomson scrutinized and expanded on Foot's ideas in *The Yale Law Journal*.

Below you will find one of the Trolley Dilemma scenarios as stated by Thomson, followed by a multiple choice question. Each answer describes a unique reaction to the dilemma, and correlates with one of the five ethical paradigms of Utilitarianism, Deontology, Divine Command Theory, Ethical Relativism, and Virtue Ethics. Clicking on an answer will send you to a page that describes the corresponding paradigm and offers insight into its meaning in relation to ethical communication.

Scenario: Trolley Driver

"Suppose you are the driver of a trolley. The trolley rounds a bend, and there come into view ahead five track workmen, who have been repairing the track. The track goes through a bit of a valley at that point, and the sides are steep, so you must stop the trolley if you are to avoid running the five men down. You step on the brakes, but alas they don't work. Now you suddenly see a spur of track leading off to the right. You can turn the trolley onto it, and thus save the five men on the straight track ahead. Unfortunately, ...there is one track workman on that spur of track. He can no more get off the track in time than the five can, so you will kill him if you turn the trolley onto him" (Thomson 1985, 1395).

I



V

WHAT WOULD YOU DO? (click on your answer below)

- Throw the switch in order to maximize well-being (five people surviving is greater than one).
- Throw the switch because you are a virtuous person, and saving five lives is the type of charitable and compassionate act a virtuous person performs.
- Do not throw the switch because that would be a form of killing, and killing is inherently wrong.
- Do not throw the switch because you are a Christian, and the Ten Commandments teach that killing is against the will of God.
- Do not throw the switch because you feel aiding in a person's death would be culturally inappropriate and illegal.

The harm of digital society

Citizens **moral** rights,
as well as the **social**,
economic and **political**
spheres are at danger

But ... it is **unavoidable**

We are in the Mangrove
societies, Floridi's metaphore
of the digital world



Many initiatives (the patchwork)

European bias

- Regulatory
 - GDPR, autonomous driving, AI legislation
- Scientific societies
 - USACM: Statement on algorithmic transparency and accountability
 - EUACM: When computers decide: European recommendations on machine-learned automated decision making.
- Institutional
 - European Data Protection Supervisor (EDPS)
 - Ethics Advisory Group: Towards a new Digital Ethics
 - EEC High-Level Expert Group in AI: Draft ethic Guidelines for Trustworthy AI (Apr 2019)
 - White paper on AI. European Commission 2020

The quest for an ethical approach

- EDPS in his strategy 2015-2019 sets out the goal to address the emerging challenges on data protection with an **ethical** approach.
- Ethics Advisory Group to steer a reflection on the ethical implications that the digital world emerging from the present technological trends puts forward
- In “Opinion Toward a new digital ethics” (2015) EDPS
 - identifies the fundamental **right** to privacy and the protection of personal data as **core** elements of the **new digital ethics** necessary to preserve **human dignity**.
 - calls for a big data protection **ecosystem** that shall involve developers, businesses, regulators and individuals in order to provide ‘future-oriented regulation’, ‘accountable controllers’, ‘*privacy-conscious engineering*’, and ‘**empowered individuals**’.

Ethics Guidelines for Trustworthy AI of EU High-Level Expert Group on AI 2019

- respecting the rule of law;
- being aligned with agreed ethical principles and values, including privacy, fairness, **human dignity**;
- keeping us, the **humans, in control**;
- ensuring the system's behavior is transparent to us, its decision making process is explainable; and
- being robust and safe, that is system's behavior remains trustworthy even if things go wrong.

Embedding Ethics in autonomous systems

Human at the center 1

- “the principle of **human dignity**, understood as the recognition of the inherent human state of being **worthy** of **respect**, must not be violated by ‘**autonomous**’ technologies”

European Group on Ethics in Science and New Technologies. statement on artificial intelligence, robotics and ‘au-tonomous’ systems.

https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf, 2018.

Human at the center 2

- It is more than having humans as explicit components of a system
- It is about **lifting** humans to be **actors** in the digital world by becoming autonomous systems that interact “au pair” with the rest of the digital world
 - **Empower** the user
 - From a passive to an **active** role
 - It requires an **architectural** approach

Human at the center 3

- “Accept/not accept” options do not satisfy our freedom of choice; and what about our individual preferences and moral views?
- Individuals are unprotected and powerless in their interaction with the digital world.
- In a digital society where the relationship between citizens and machines is uneven, moral values like **individuality** and **responsibility** are at risk.

Digital Ethics

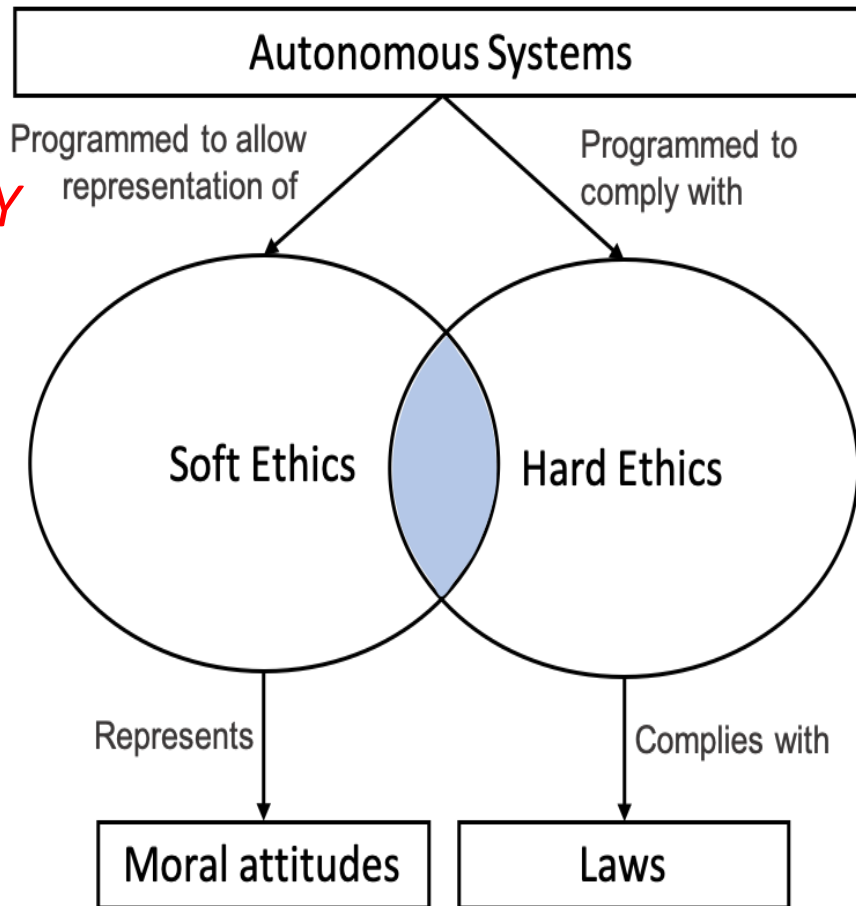
Digital ethics is the branch of ethics that aims at formulating and supporting morally good solutions through the study of moral problems relating to personal data, (AI) algorithms and corresponding practices and infrastructures.

Hard ethics is defined and enforced by digital legislation. Legislation is necessary but insufficient, since it does not cover everything, nor should it.

Soft ethics is the space of moral decisions that is left to the actors of the digital world, e.g., companies and citizens. It deals what ought and ought not to be done over and above the existing regulation, without trying to by-pass or change the hard ethics

- L. Floridi. Soft ethics and the governance of the digital. *Philosophy & Technology*, 31(1):1–8, Mar 2018.

FLEXIBILITY



Challenges

- It is a multidisciplinary effort across different disciplines and inside computer science
- Philosophers, sociologists, psychologists, jurists, software engineers shall work together
- Increase awareness in users and system producers

A motivating example - 1

A parking lot in a big mall;

- two autonomous connected vehicles A and B, with one passenger each, are competing for the same parking lot. Passenger in A has health problems.
- A and B are **rented** vehicles, they are **multi-user** and have a default decision algorithm (*ethic*). That is, the cars will look for the free parking lot that is closer to the point of interest, in case of contention the closest car gets in. A and B are approaching the parking lot. B is closer, therefore it will take the parking lot.
- Seems fair enough ... however ...

A motivating example - 2

- Suppose that by communicating with A, passenger in B receives the information that the passenger in A has health problems. Should passenger B follow her ethics (a virtue ethic) she would decide to leave the parking lot to A.

This use case shows many things:

- **personal privacy** is strictly connected to **ethics**: by disclosing a personal information like this, the passenger in A follows a **utilitarian** view which is related to the expectation that surrounding drivers might have a **virtue** personal ethic
- Individuals have **different** ethics also depending on the context, indeed neither a person nor a society apply moral categories separately, rather everyday morality is in constant flux among norms, utilitarian assessment of consequences, and evaluation of virtues
- A decision policy that seemed fair (to whom?) does not correspond to the personal ethic

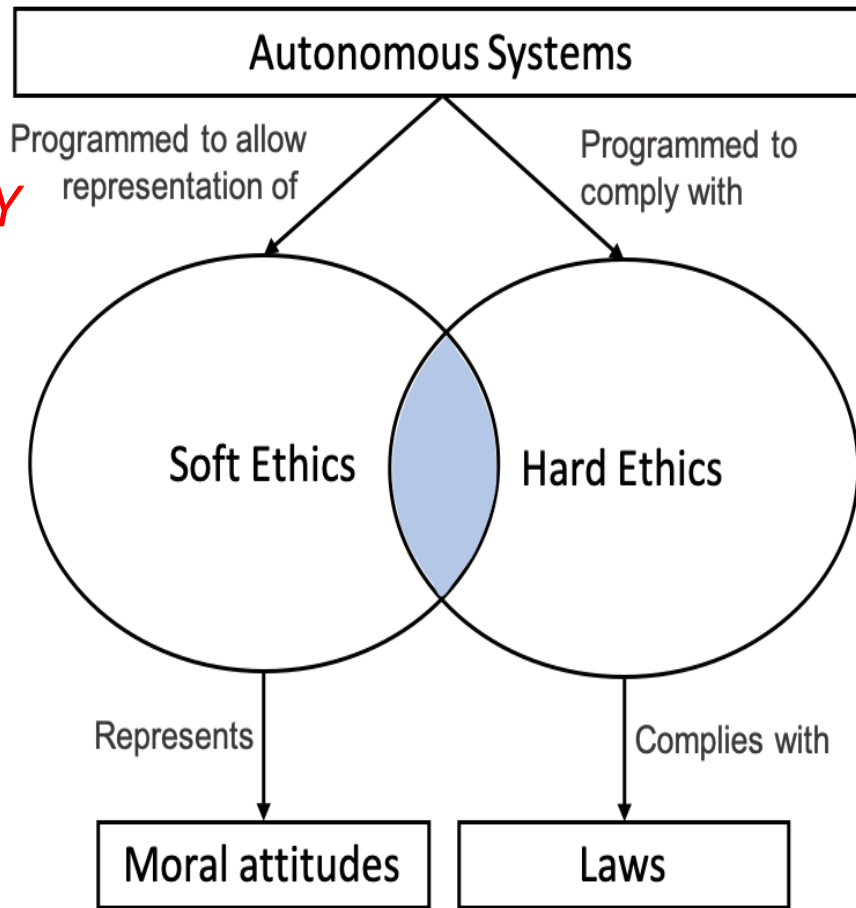
What do we learn from the example - 1

- Putting human at the center requires to have a certain level of **customization/enforcing** of the decisions of the autonomous systems
- We **postulate** that this level corresponds to the **soft** ethics of each individual
- **Soft** ethics shall live on top and be **consistent** with **hard** ethics
- Individuals use **different** ethics depending on the context also regarding their personal data

What do we learn from the example - 2

- **soft** ethics is associated to **individuals** and **hard** ethics to **systems**, i.e. autonomous cars
- The two need to combine (**moral agreement**) when an individual and a system **interact**
- Focus is on interactions of independent systems at the **architectural** level
- This puts **architectural requirements** on the autonomous systems

FLEXIBILITY



Soft ethics modeling

- **Dispositions** are those properties that individuate the causal behaviour of the entities that possess them. They dispose towards their manifestations, which occur when some conditions are met. The need of an individual to make a moral choice can be seen as the stimulus that makes her dispositions (e.g. courage, generosity) manifest.
- **Specification patterns** Specification patterns identify representative solutions to recurrent class of problems. A pattern is formulated both in structured English and in a (temporal) logic.

(response pattern with scope between Q and R): **between** “entering-parking-lot” **and** “exiting-parking-lot”, **when** a “weak-health-status” holds, **then** “alert-surrounding-vehicles”;

(privacy tradeoff) “weak-health-status” **is shared** only **between** “entering-parking-lot” **and** “exiting-parking-lot”

Software Architectures

- SA serve many purposes - see [TMD2009]
 - My favorite view
 - glue/connectivity *nature* that allows subsystems/components to interact, correctly
 - Define the system structure in terms of components/subsystems, their interactions in terms of functional and non functional behavior, either local and global

[TMD2009] [Richard N. Taylor](#), [Nenad Medvidovic](#), [Eric Dashofy](#), Software Architecture: Foundations, Theory, and Practice

Structuring interactions: protocols and connectors/mediators

- SA defines structure/components and **interactions**
 - **Interactions** are the *observable* actions at the interface level
 - **Interactions** are performed by following *protocols*, i.e., given ordering in the way interface operations need to be executed
 - **Connectors** are architectural elements that define how components' protocols match together
 - **Mediator**: connectors that allows the communication among compatible protocols by mediating their differences

The space of decisions

- The autonomous system takes decisions that results in *actions*
- Depending on the context, actions have an ethical implication (machine ethics):
 - Push the brake in presence of the red traffic light
 - Push the brake to avoid running on people crossing the street

Actions are finite and depend on the domain, contexts are potentially infinite but in practice made discrete (given the domain)

How and when to decide?

2

C. Benzmüller et al. / Artificial Intelligence 287 (2020) 103348

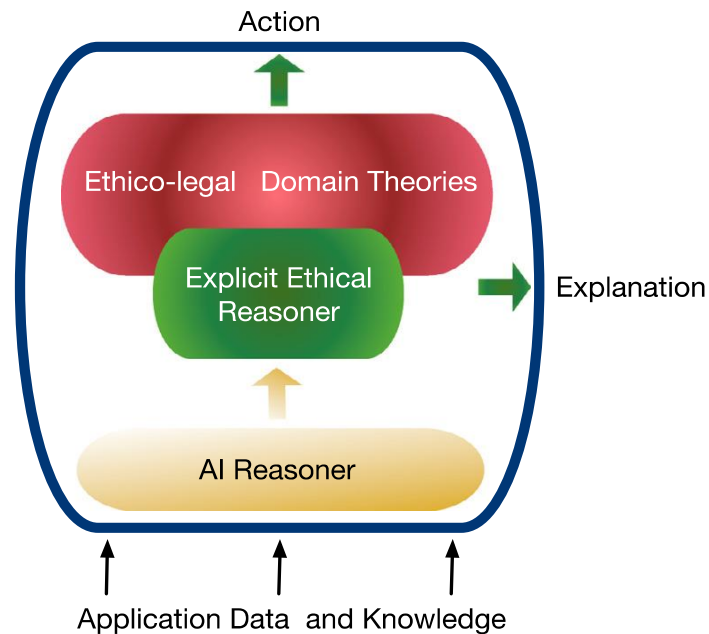


Fig. 1. Explicit ethical reasoner for intelligent autonomous systems.

Designing normative theories for ethical and legal reasoning: LogiKEy framework, methodology, and tool support ☆

Christoph Benzmüller b,a,*, Xavier Parenta, Leendert van derTorre a,c

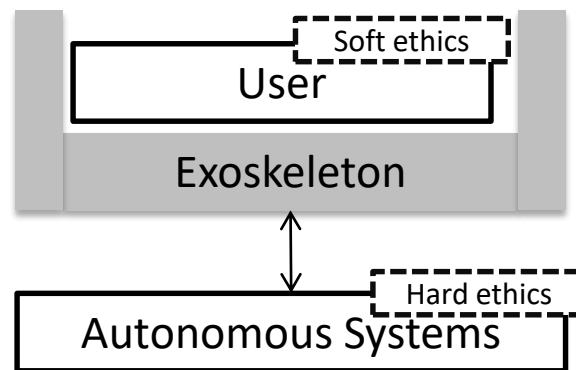


The Exosoul Project

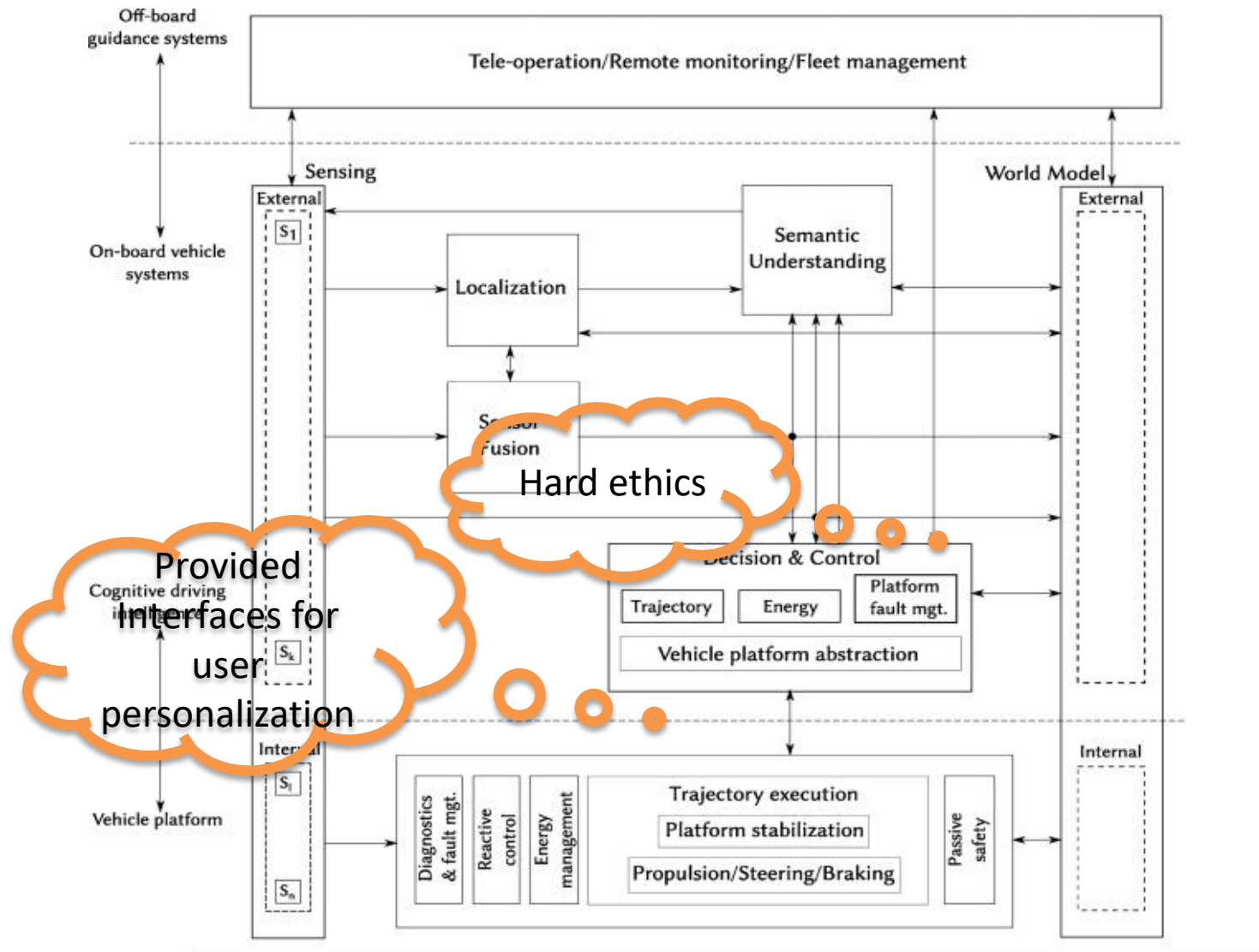
- Empowering the user with a software exoskeleton that mediates the interactions with the digital world according to her (soft-) ethics preferences.

We aim at producing and delivering Exosoul software components

The Exosoul architecture



Automotive Functional architecture



Empowering the user

Soft ethics

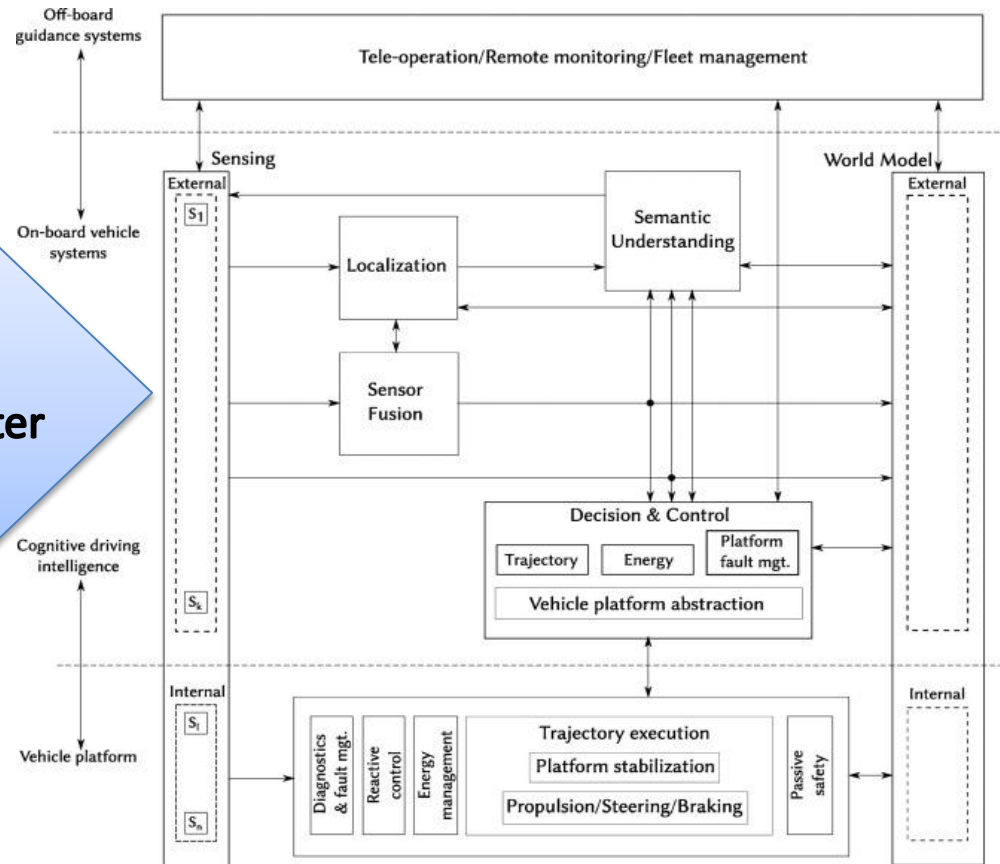


EXOSKELETON
preferences interface:
To realize Virtue ethics
behaviours

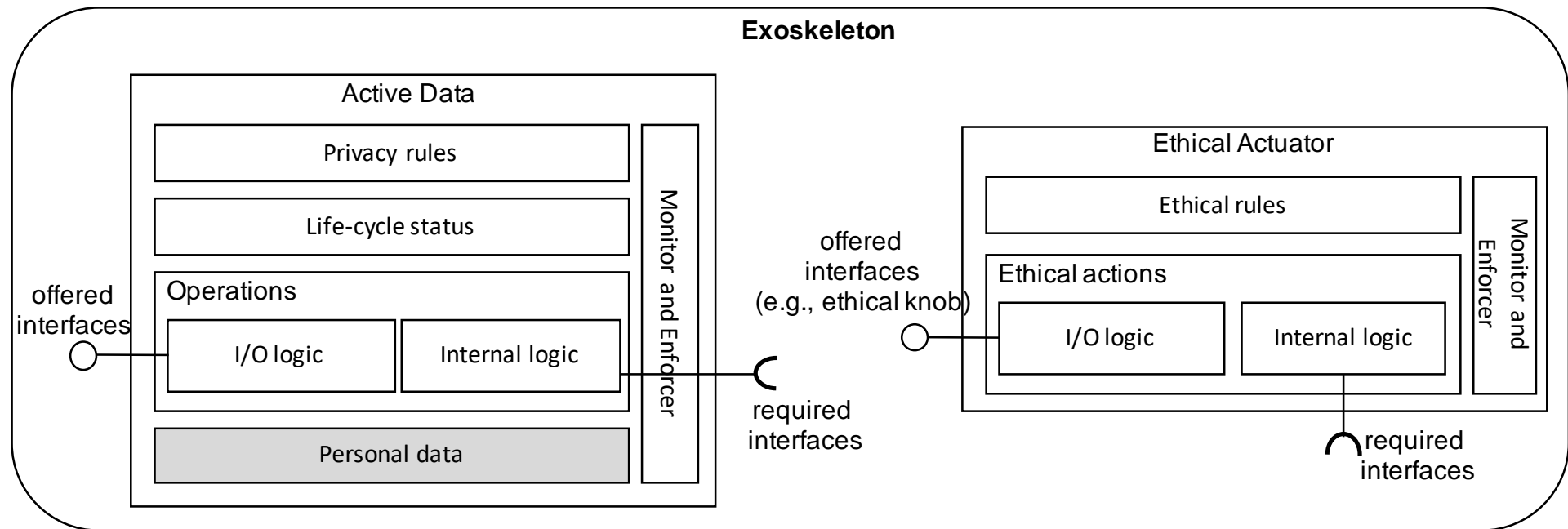
On board



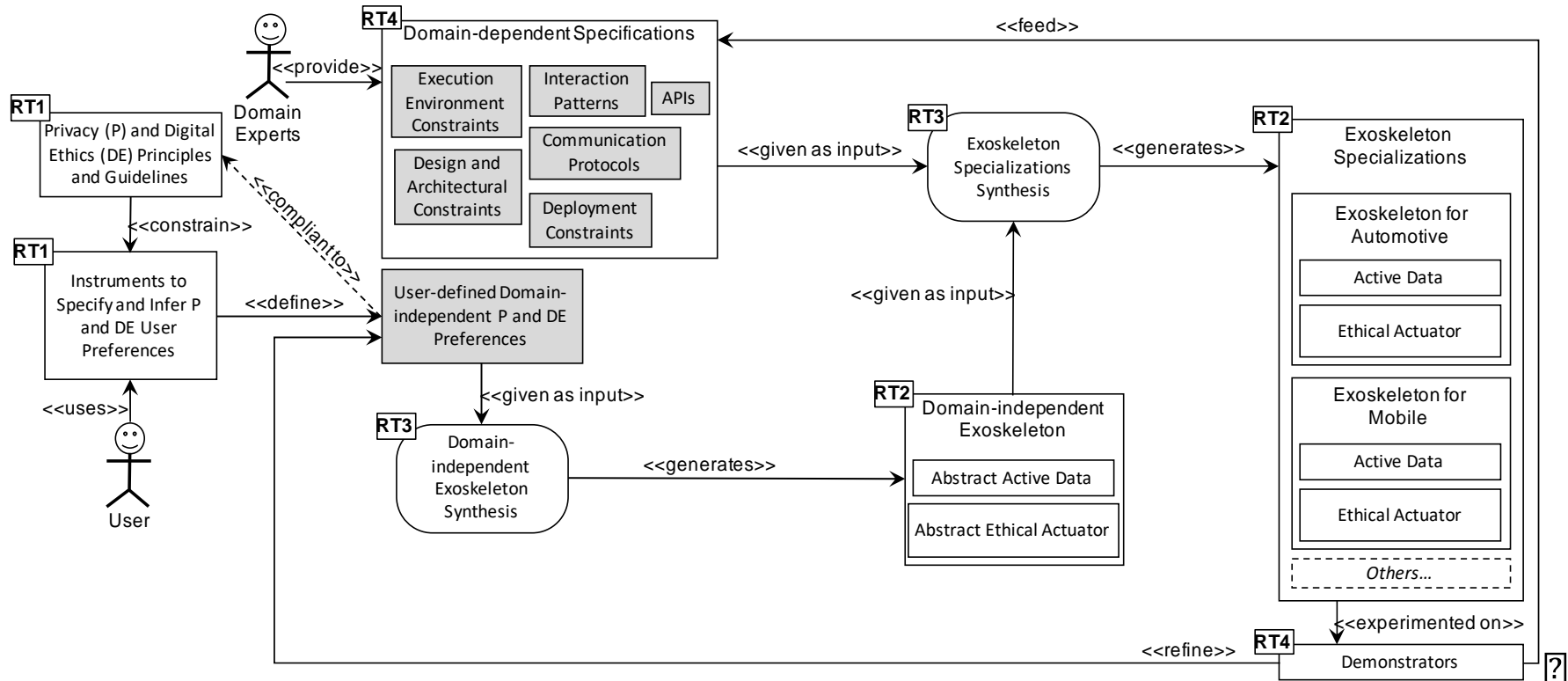
Connector/
mediator/adaptor

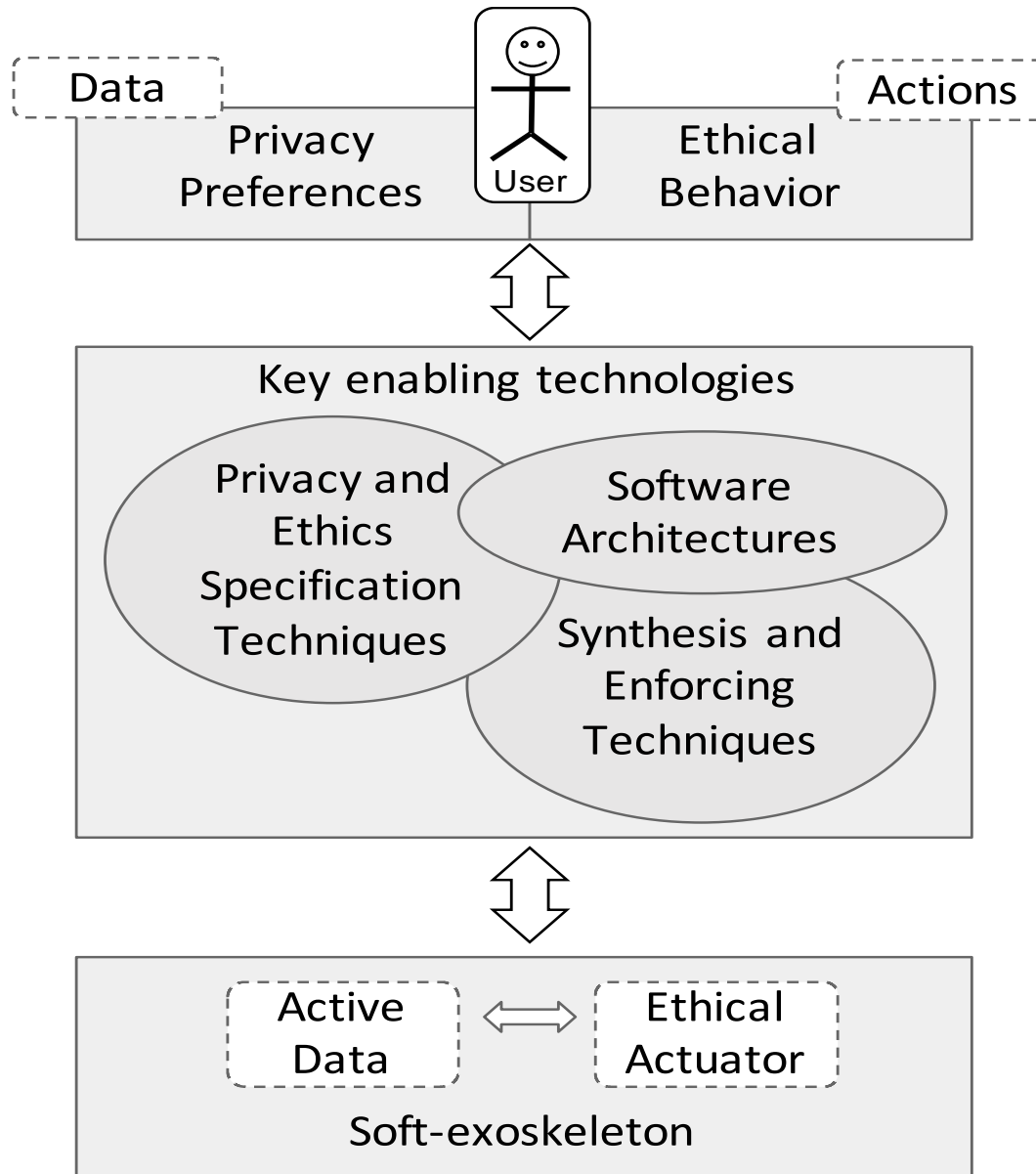


The shield architecture



The methodology





Overview

Challenges

Challenges - 1

- Express the user (soft-) ethical preferences. Top down and bottom up: ethical categories, dispositions, specification patterns, social psychology, privacy persona, etc.
- Automate the exoskeleton production: synthesis and model driven engineering
- Design the system's *self* boundaries to hard ethical decision: 2 domains (automotive, mobile)
- Define the system interface and protocol requirements to allow matching with the user's exoskeleton (protocols)
- Bridge the gap between ethical preferences and actual decision making: model 2 model transformations

Challenges - 2

- It is a multidisciplinary effort across different disciplines and inside computer science
- Philosophers, sociologists, psychologists, jurists, software engineers shall work together
- Increase awareness in users and system producers

The Exosoul Team

- Paola Inverardi, Marco Autili, Davide Di Ruscio, Patrizio Pelliccione, Massimo Tivoli, Gianluca Scoccia
- Simone Gozzano, Guglielmo Tamburrini, Marco Segala, Donatella Donati
- Patrizio Migliarini

References on Ethics in AS/AI

- Association for Computing Machinery US Public Policy Council (USACM). Statement on algorithmic transparency and accountability. https://www.acm.org/binaries/content/assets/publicpolicy/2017_usacm_statement_algorithms.pdf, 2018.
- P. Inverardi The European Perspective on Responsible Computing, CACM Special Session on Europe ... 2019
- J. L. et al. When computers decide: European recommendations on machine-learned automated decision making. <http://www.acm.org/binaries/content/assets/public-policy/ie-euacm-adm-report-2018.pdf>, 2018.
- EDPS. Opinion 4/2015, towards a new digital ethics - data, dignity and technology. https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf, 2015.
- EDPS. Leading by example, The EDPS Strategy 2015-2019. https://edps.europa.eu/sites/edp/files/publication/15-07_30_strategy_2015_2019_update_en.pdf, 2015.
- European Group on Ethics in Science and New Technologies. Statement on artificial intelligence, robotics and 'autonomous' systems. https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf, 2018
- J. P. Burgess, L. Floridi, A. Pols, and J. van den Hoven. Towards a digital ethics - edps ethics advisory group. https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf, 2018.
- L. Floridi. Soft ethics and the governance of the digital. *Philosophy & Technology*, 31(1):1–8, Mar 2018.
- The DECODE project. <https://decodeproject.eu>.
- G. Contissa, F. Lagioia, and G. Sartor. The ethical knob: ethically-customisable automated vehicles and the law. *Artificial Intelligence and Law*, 25(3): 365–378, 2017.
- J. Gogoll and J. F. Müller. Autonomous cars: In favor of a mandatory ethics setting. *Science and Engineering Ethics*, 23(3): 681–700, Jun 2017.
- Ethics Commission Automated and Connected Driving, Appointed by the German Federal Minister of Transport and Digital Infrastructure, Report (extract) <https://www.bmvi.de/SharedDocs/EN/Documents/G/ethiccommission-report.pdf>, June 2017
- Corinne Cath, Sandra Wachter, Brent Mittelstadt, Luciano Floridi Editors. 'Governing artificial intelligence: ethical, legal, and technical opportunities and challenges' *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A: MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES* Royal Society, Nov. 2018

- R. Anjum, S. Noer Lie, and S. Mumford. Dispositions and Ethics. In *What Tends to Be: The Philosophy of Dispositional Modality*, Routledge, 2018.
- Liao, B., Slavkovik, M., van der Torre, L.: Building jiminy cricket: An architecture for moral agreements among stakeholders. In: *AAAI/ACM Conf. AI, Ethics, and Society*, pp. 147–153 (2019)

Our references

- P. Inverardi. The european perspective on responsible computing. *Commun. ACM*, 62(4):64–64, Mar. 2019.
- P. Inverardi. Ethics and Privacy in Autonomous Systems: A Software Exoskeleton to Empower the User. *SERENE* 2019.
- G. L. Scoccia, I. Malavolta, M. Autili, A. Di Salle, and P. Inverardi. Enhancing trustability of android applications via user-centric flexible permissions. *IEEE Transactions on Software Engineering (TSE)*, 2019.
- G. L. Scoccia, S. Ruberto, I. Malavolta, M. Autili, and P. Inverardi. An investigation into android runtime permissions from the end users perspective. In *5th IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft 2018)*, 2018.
- Marco Autili, Davide Di Ruscio, Paola Inverardi, Patrizio Pelliccione, Massimo Tivoli: A Software Exoskeleton to Protect and Support Citizen's Ethics and Privacy in the Digital World. *IEEE Access* 7: 62011-62021 (2019)
- Scoccia GL., Fiore M., Pelliccione P., Autili, M., Inverardi P., Russo A. Hey, my data are mine! Active data to empower the user, *IEEE/ACM 39th International Conference on Software Engineering (ICSE2020-NIER)* 2020
- Migliarini P., Scoccia GL., Autili M., Inverardi P. On the Elicitation of Privacy and Ethics Preferences of Mobile Users, *7th IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft 2020, Vision Track)* 2020
- Scoccia GL., Autili M., Inverardi P. A self-configuring and adaptive privacy-aware permission system for Android apps, *1st IEEE International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS 2020)* 2020.