

CompuLaw

Computable Law

Giovanni Sartor

Compulaw

- The project addresses the regulation of computations (processes and systems) through an innovative integrated legal and technological framework: it provides epistemic, technical and normative guidance for the sound development of computable law and law compliant computations.
- PI: Giovanni Sartor
- Two institutions involved:
 1. The European University (law, economics, and regulation theory)
 2. The University of Bologna
 - a. CIRSFD - Centre for Legal Informatics (legal informatics and legal logic)
 - b. Department of Computer Science and Engineering (computing and AI)

External experts will also be involved

Outline

- Introduction: the infosphere; opportunities and risks of AI
- Computable Law and explicit law-by-design
- Examples: autonomus cars & e-commerce
- Articulation of the project

Algorithmic society

- Information and communication technologies have become increasingly powerful, connected and pervasive, affecting all aspects of individual and social life.
- The **algorithmic society** (Balkin, 2016) “organized around social and economic decision making by algorithms robots, and artificial intelligence agents”

A new social ecology

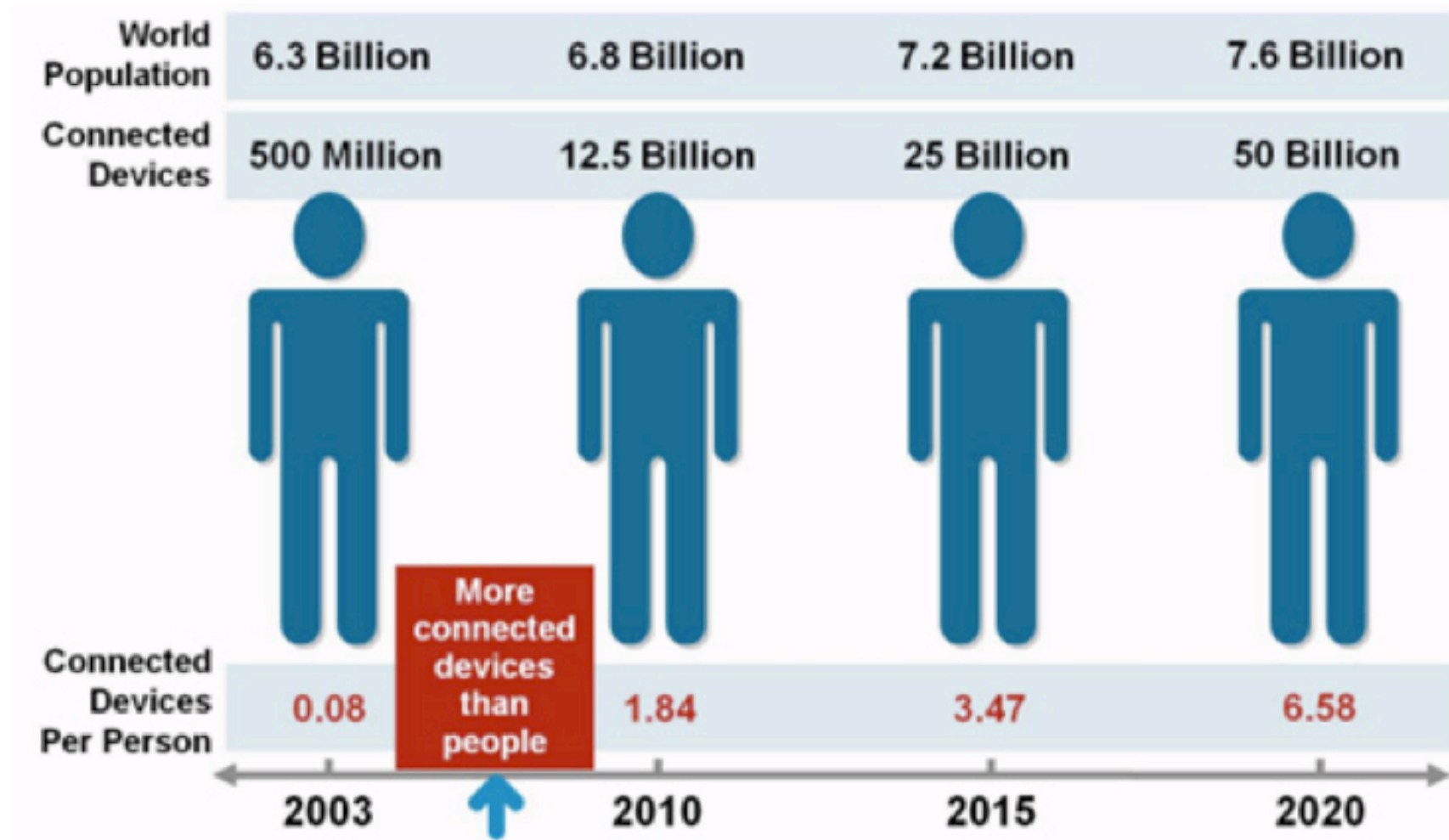
- Computations handle the storage, organisation and retrieval of data, govern the functioning of all kinds of machines, manage workflows, support individual and social activities, constitute virtual environment, envelope physical objects with digital interfaces, augment reality with content and virtual objects, animate virtual and physical automata
- **a new social ecology**, whose very existence and persistence depends upon the automated processing of information

Infosphere (Floridi 2103)



Outside the computer

Infosphere (Floridi 2013)



A hybrid infosphere, populated not only by humans, but also by a huge and growing number of increasingly pervasive, autonomous and intelligent computational entities.

Infosphere (Floridi 2013)



Outside the computer

As the boundaries between life online and offline break down, and we become seamlessly connected to each other and surrounded by smart, responsive objects, we are all becoming integrated into an "infosphere"

Hyper-connected “onlife” and artificial agents

- **Hyper-connected “onlife”** where artefacts can change states in autonomous ways and can do so by digging into the exponentially growing wealth of data, made increasingly available, accessible and processable” (Floridi 2013)
- **Artificial agents** endowed with high-level cognitive capacities populating the human and social world:
 - industrial robots
 - service robots in houses and health care facilities
 - autonomous vehicles and unmanned aircraft in transportation
 - autonomous electronic agents in e-commerce and finance
 - autonomous weapons in the military
 - intelligent communicating devices embedded in every ambience

Hopes and worries

Scientists and technologists have expressed their **hopes and worries** over such developments:

- Intelligent technologies will help to address the “grand challenges of humanity: maintaining a healthy environment, providing resources for a growing population, [...] overcoming disease, vastly extending human longevity, eliminating poverty” (Kurzweil 2012)
- humans may lose the race against the machines, so that human work may become redundant (Brynjolfsson and McAfee 2014)
- AI may represent an “existential risk” for humanity, that non-biological superintelligence (Bostrom 2014) will possibly spell “the end of the human race” (Stephen Hawking, Elon Musk, and others)

The advent of AI

- AI entities differ from previous mechanical devices, since they engage in complex cognitive tasks, and adapt to variable contexts, possibly learning from experience.
- Endowing computational entities with high-level cognitive capacities, AI increases their ability for diversified and autonomous action, both in physical domains (robotics) and in virtual ones (digital bots, intelligent agents).

AI: Opportunities and risks

- On the one hand, **AI systems may enhance human abilities**, improve security and efficiency, and enable the universal provision of knowledge and skills.
- On the other hand, they may increase opportunities for **control, manipulation, and discrimination; disrupt social interactions; and expose humans to harm** resulting from technological failures or disregard for individual rights and social values.

Ethical and legal issues

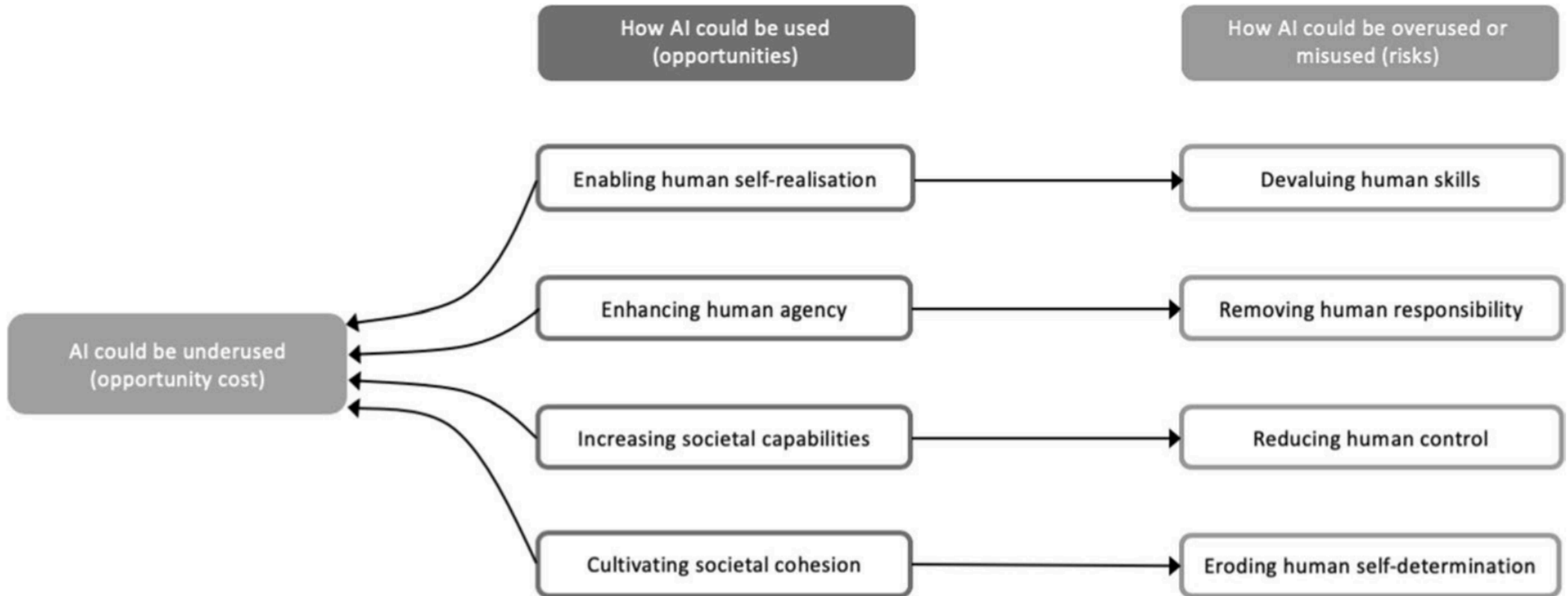
- **concrete ethical and legal issues** of AI have already emerged (civil liability, insurance, data protection, safety, contracts and crimes, ETC.)
- more and more intelligent systems leave the controlled and limited environments of laboratories to share with humans the same physical and virtual spaces (roads, skies, stock trading and other markets, etc.).
- Machine learning techniques, in the context of big data has led to computer systems that can learn how to handle individuals in efficient and cost-effective ways, though not always in the best interest of those individuals (O'Neill 2017).

Policy documents

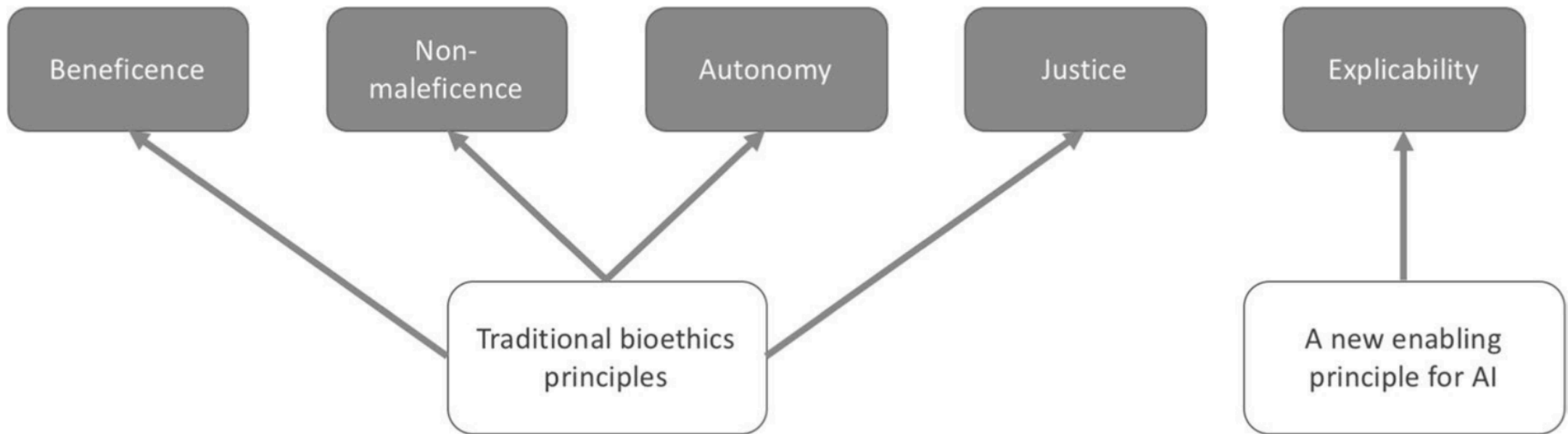
Society is becoming aware of the potential benefits and dangers of AI:

- several policy documents concerning specific application domains, (big data, unmanned aircraft, autonomous cars, service robots, autonomous weapons, etc.)
- Report by the Presidency of US (The White House 2016a, 2016b)
- 2017 Report and Resolution on Civil Laws Rules on Robotics by the European Parliament (EP 2017a, 2017b).
- Open letter (Future of Life 2016) signed by leading scientists and technologists arguing for “expanded research aimed at ensuring that increasingly capable AI systems are robust and beneficial”
- And many others..

AI4People—An Ethical Framework for a Good AI Society (Floridi et al. 2018)



AI4People—An Ethical Framework for a Good AI Society (Floridi et al. 2018)



The need for compliance

- The infosphere may become an unfriendly and dehumanising environment, where individuals are subject to persistent surveillance and to determinations over which they have no control, being outsmarted, cheated, and possibly harmed by faulty, illegal or unethical computations
- To avoid such risks, computational entities shall be designed **to comply with legal norms**, avoid causing harm, respect rights and justified expectations, transparently explain their behaviour to the individuals concerned

Compulaw: starting assumptions

- The scale and ubiquity of computations make it **impossible for humans to effectively monitor computations and anticipate** their illegal or unethical behavior
- the speed of computational entities, and the sheer number of them and of their interactions makes **real-time monitoring impossible**
- the complexity and adaptivity of their increasingly autonomous behaviour and the diversity of their deployment contexts make it **impossible to specify hard constraints at the design stage**

Computable law

- To effectively govern the infosphere, **the law must become computable**, i.e. it must also be an internal component of computational processes rather than an external constraint over them
- legal norms, values and principles must be mapped onto, and partially translated into (a) **computable representations** of legal information and reasoning, that (b) are **processed by computational entities**.

Law-by-design

- **Implicit law-by-design:** systems designed in such a way as to make illegal behaviour more difficult (*e.g. legal provisions, technical standards, guidelines*)
- **Explicit law-by-design:** systems that represent law explicitly and operate effectively on the basis of this knowledge (*computable law*)

Explicit compliance-by-design

- Compliance by design in various domains, such as data protection (Cavoukian 2013), but no general methodology that also includes explicit compliance-by-design.
- The deployment of autonomous and potentially dangerous technologies generates risks, conflicts of interests and coordination issues that must be addressed:
 - a) according to the law and
 - b) by those systems themselves at run-time, without direct human intervention
- To meet these two requirements, it is necessary that the involved systems directly apply legal norms and principles, justify their behaviour accordingly, and ask for human assistance when needed and feasible.

Artificial Legal Agents (ALAs)

- To ensure legality in the infosphere, it is necessary to develop **Artificial Legal Agents (ALAs)**, entrusted with applying the law to their own and others' behaviour, in multi-agent contexts.
- ALAs should be capable and motivated to comply with legal norms and principles that apply to their activities (driving cars, making contracts, etc.).

Humans and computations

- Computable law as a **human enterprise**, over which society should maintain full ownership and control (Mindell 2015).
- Thus, also the **socio-technical conditions** for the correct design, deployment and monitoring law compliant computations shall be addressed
- ALAs' operation depends on the “**joint cognitive**” **engagement** (Hollander and Wu 2005) of designers and deployers, whose choices determine the goals, the scope, and the limitations of the ALAs' autonomous action.

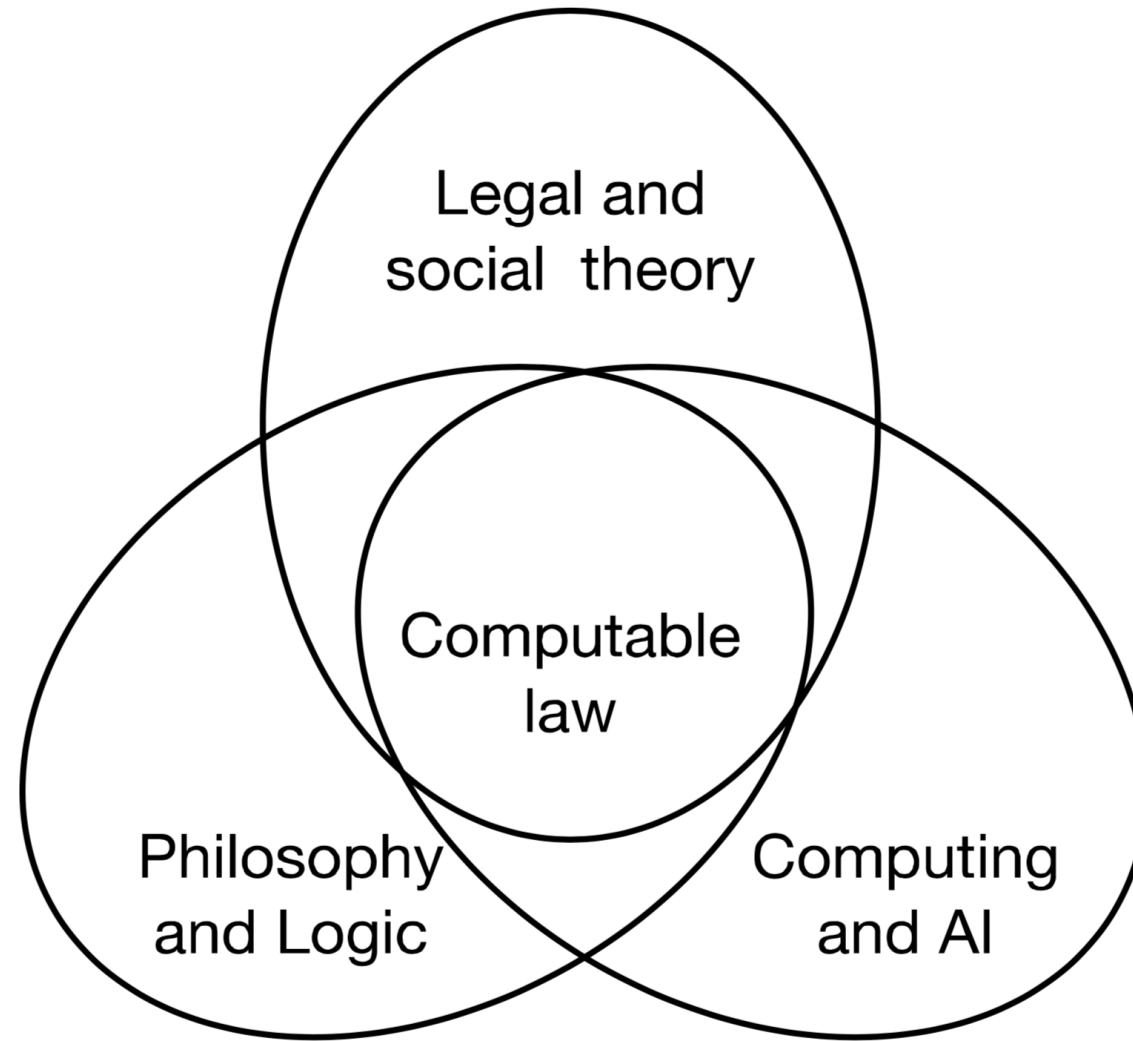
Computable Law: crucial aspects and issues

- what legal requirements in what domains and to what extent should be made computable
- what technologies and architecture should be used
- how to combine top-down compliance with predefined rules and learning from cases
- how to addresses normative conflicts
- how to apply rules according to legal values/principles
- what formal/substantive legal constraints should govern the design and deployment of ALAs

Computable law: the need for a theoretical framework

- A vast amount of research exists on computable models of legal contents and reasoning, on the normative governance of multiagent systems, on logical models of norms and argumentation, on natural language processing and machine learning as applied to the law, on information/algorithms/machine ethics, on roboethics, etc.
- However, these ideas, models and methods have had limited application to operational environments
- **A comprehensive theoretical framework, informed to both legal theory and computer science, is still missing**

A framework for computable law: the 3 main clusters



The 3 overlapping clusters

1. **Legal-social cluster:** specific domains of the law at issue in different contexts (e.g., traffic law, financial market law), aspects of ICT law (e.g., data protection, e-commerce), general legal doctrines (e.g., criminal and civil liability, contracts)
2. **Philosophical-logical cluster:** theories of normativity (of norms, values, institutions), agency (of mind, intentionality, action), language use (semantics, pragmatics, argumentation), interactions (causality, influence, decisions and games), and corresponding logical models (deontic logic, logics of action, cognition and intentionality, game theory). It also includes the normative foundations of law (fundamental rights, principles and social values)
3. **Computing-AI cluster:** relevant models of knowledge representation, automated reasoning, multi- agent interactions, and learning.

Examples: 2 case studies

1. Autonomus cars
2. E-commerce

Autonomus cars

- **autonomous cars**, a rapidly developing technology, currently evolving from partially automated driving to full automation, also including safety critical functions.
- In leading projects, self-driving cars are supposed to take the passenger to destination without any human contribution.

Autonomus cars: autonomus compliance

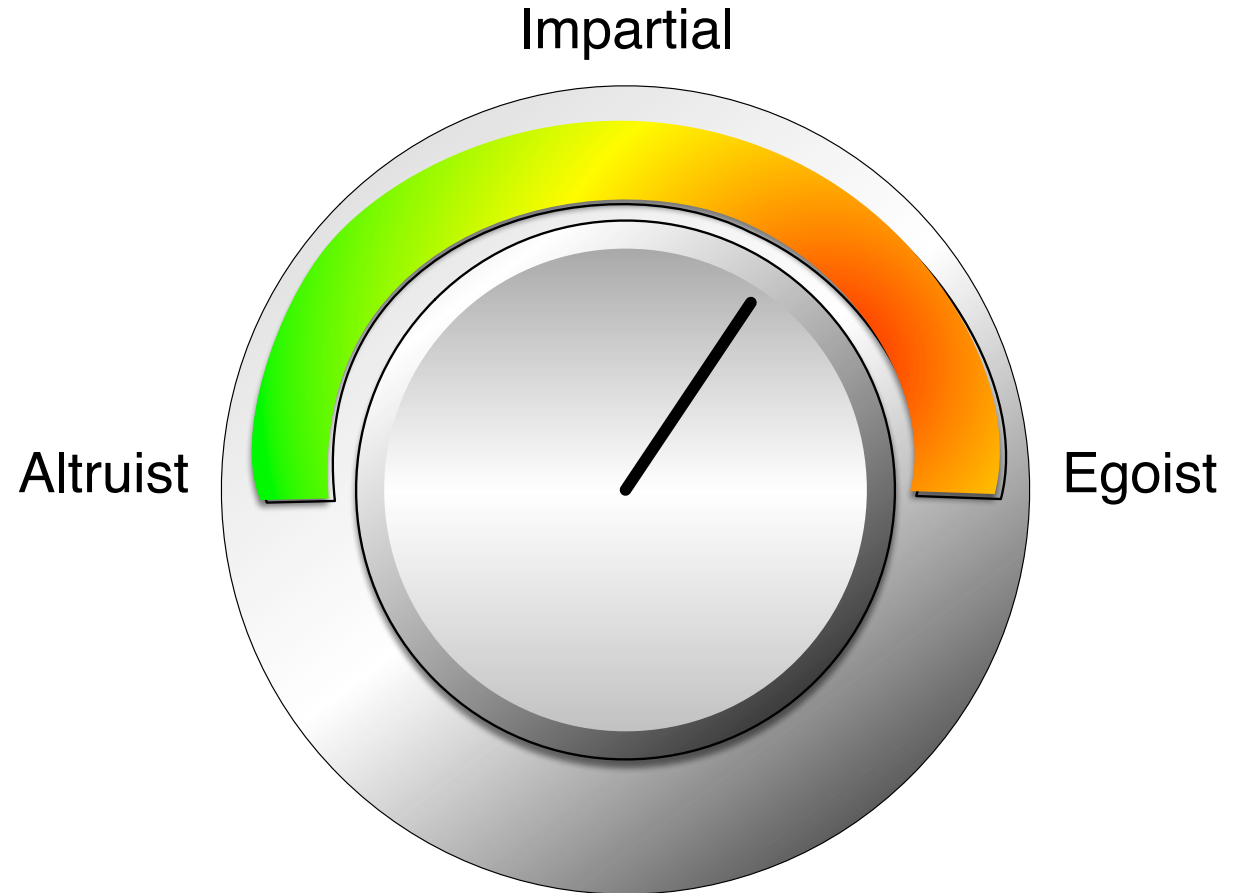
- Given the complexity of traffic regulations **only explicit law-by- design can support autonomous compliance**
- The car must self-govern its behaviour according to a computable model of the applicable laws. It should match its current situation to the relevant rules (speed limits, right of way, distances to be kept, etc.), and make consistent determination
 - rejecting outcomes that violate normative constraints
 - handling exceptions to traffic rules, including cases where rules should be violated
- It should also take into account legal principles, and address cases where **competing legally relevant interests and values** are at stake, as when avoiding an impact with third parties may put at risk the safety of passengers (Bonnefon et al. 2015).

Autonomus cars: what is needed

- **models of proportionate normative decision-making** which also consider broadly defined legal/ethical principles and values, together with user's preferences
- **machine learning methods** should complement guidance through explicitly formulated legal norms.
- The car, exercising its **adjustable autonomy**, should determine when a hard issue should be referred to the human passenger/driver, or when, this being impossible for contextual constraints, the car should act autonomously.
- The car should be able to **explain its behaviour**, including the rejection of users' requests (e.g., to speed up), by appealing to the facts of the case and the applicable norms/principles.
- The **interaction** between the car, its user, and smart city infrastructures should be transparent
- A **liability model** should be created, through new regulations and re-interpretation of the principles of liability, to ensure compensation of the victims, and provide adequate incentives to all actors

The Ethical Knob

- Ethical Knob: a device enabling passengers to ethically customise their AVs, namely, to choose between different settings corresponding to different moral approaches or principles.
- the passenger has the task (and burden) of deciding what ethical approach should be adopted in unavoidable accident scenarios
- the AV has the task to implement the users ethical choice, based on its risk assessment.
- thus, the passenger would provide the “ethical customisation” of the AV, which would then behave accordingly.



Contissa, Lagioia, Sartor (2018)

E-commerce

- AI applications in e-commerce: **intelligent e-commerce agents**. They may act autonomously, according to the specification of their users
- Their operations should respect multiple legal requirements: **laws on contract, consumer protection, data protection, taxation, as well as principles of fairness, good faith, transparency and non-discrimination**.
- When implementing a contract, an e-commerce agent should apply both provisions in the contract and the applicable norms, **monitoring its own compliance as well as the compliance of its counterparts**.
- When required, the agent should be able to **explain the reason for its behaviour** (e.g., refusing a loan, raising an insurance price, implementing a contract in an unexpected way).

E- commerce

- Deployers of e-commerce agents may adopt a (more or less) **open delegation** (Castelfranchi and Falcone 1998), empowering the agent itself to determine how to achieve the purposes indicated, according to its planning and learning.
- Thus, the system should be able to combine its goal-directed behaviour and compliance with legal and user-generated constraints.
- Research themes /issues:
 - the **illegal or unfair/unethical behaviour of AI applications** in e-commerce
 - the extent to which, and the ways in which **AI itself**, as implemented in e-commerce agents, or in agents controlling them, **may contribute to deliver legality and fairness**.

CLAUDETTE

An Automated Detector of Potentially Unfair Clauses

Lippi et al.
(2018)

Claudette found 4 potentially unfair clauses (displayed in **bold**) out of 5 sentences.
By hovering your cursor over each unfair sentence, you can see the most likely unfairness category.

[...]

Unilateral Change unfair clause

As such, the Services may change from time to time, at our discretion.

We may stop (permanently or temporarily) providing the Services or any features within the Services to you or to users generally.

We also retain the right to create limits on use and storage at our sole discretion at any time.

We may also remove or refuse to distribute any Content on the Services, suspend or terminate users, and reclaim usernames without liability to you.

Share link

Save results

Try Again

Contact

Machine powered analysis of consumer contracts and privacy policies

<http://claudette.eui.eu/demo/>

Articulation of the project

3 Subprojects:

- **Subproject 1:** Regulatory issues and frameworks in computable law focuses on legal issues
- **Subproject 2:** Computational methods and technologies, focuses on the computational modelling of normative requirements and on the implementation of ALAs
- **Subproject 3:** A legal-computational approach for the hybrid infosphere, merges law and technology into an integrated framework.

Subprojects 1 and 2 provide input to Subproject 3, which delivers the final comprehensive output of the project

Subproject 1: “Regulatory issues and frameworks.”

- **Subproject 1** addresses the **governance of computations**, focusing on intelligent systems.
- It critically considers (a) the ways in which computations affect legally relevant interests (both individual rights and social values) in different socio-technical contexts, (b) the influence chains and legal responsibilities linking computations to their deployers and designers, (c) the extent to which harm is prevented and remedied and computational justice and fairness are supported.
- The subproject will focus on the in-depth analysis of two case studies: **autonomous transportation**, and **e- commerce**.
- Connections with other domains will be considered, such autonomous weapons, where advanced technologies for legal compliance are already being developed (Arkin 2009).

Subproject 2: “Logical-Computational methods and technologies”

- **Subproject 2** focuses on **methods and technologies for the legal governance of computations**. The focus will be on **computable laws**, and on the architectures for law compliant **ALAs**. It will also address issues connected to the legal/ethical assessment of ML methods and outcomes.
- The architecture of **ALAs** will mainly be based on the BDI (Belief-Desire-Intention) models, expanded with norms, legitimacy and trust.
- Subproject 2 will provide:
 - **computable models of rules, principles, and cases**
 - **formal specifications** (normative concepts, socio-cognitive notions, social relations, rules, values and goals, defeasible argumentation, dialectical interactions, institutional structures, norm-related cognitive processes)
 - **regulations and guidelines directed to designers of computable laws and ALAs**, to support the correct specification of computable laws and their effective implementation

Subproject 3: “A techno-legal approach to computable law”

- Subproject 3 finalises a framework for the creation, and implementation of computable laws, to be complied with by ALAs having different degrees of autonomy and cognitive capacity.
- The proposed framework includes methodologies, technologies and substantive suggestions, to combine: 1) norms governing human behaviour; 2) norms specifying functional requirements of computational systems, directed at designers and deployers; 3) norms directed at computational entities.
- Special emphasis on communication and argumentation, namely, on the ways in which (intelligent) systems should explain their choices, demonstrating compliance with the applicable norms.
- Subproject 3 will deliver (1) a **methodological toolkit**; (2) a set of **substantive guidelines**.

1. Methodological toolkit

- The **methodological toolkit** will suggest how the issues identified in the Subproject 1 can be addressed, with the methods and techniques identified in Subproject 2. It will include:
 - (a) ways to translate legal requirements into computational standards, including computable legal rules and principles;
 - (b) architectures for ALAs, conditions and methods for legitimate, effective and safe delegation of compliance and enforcement to ALAs;
 - (c) methods to verify the correctness of computable norms and to simulate/anticipate social outcomes of norm-governed interactions involving ALAs;
 - (d) protocols for dialectical interactions between humans and ALAs and between ALAs;
 - (e) argumentation schemes through which ALAs can explain and demonstrate compliance with the applicable norms;
 - (f) methods to assess and explain the behaviour of machine-learning systems
 - (g) institutional arrangements to deliver computational norms to certified ALAs, involving public agencies and/or the initiatives of designers/deployers/users, supported by private standards.

2. Substantive guidelines

- The **substantive guidelines** will propose hybrid regulatory frameworks (covering selected domains, including the two cases studies) to govern the creation and implementation of computable law, to preserve and enhance the rule of law and ensure respect of fundamental rights and principles. They will address all of the following:
 - (a) obligations of users, owners, and designers/producers of ALAs;
 - (b) standards for designing ALAs;
 - (c) ways to delegate (more or less openly) tasks and responsibilities to computations;
 - (d) obligations to which intelligent computation should respond;
 - (e) normative powers of intelligent computations (conditions for the validity of contracts and legal acts accomplished by them);
 - (f) civil and criminal liabilities of designers/producers for the harm caused by computational entities acting under their choice, influence or control;
 - (g) conditions for correct and friendly human-machine interaction in the hybrid infosphere (e.g., respect for privacy and data protection, non-discrimination, and transparency).

Thank you!

Giovanni Sartor