Value-Driven Adaptive Norms

Future Tech Week 2019

Barcelona, 26 September 2019

Nardine Osman

IIIA-CSIC

WENET
INTERNET OF US

The Internet of Us

WeNet in a Nutshell

WeNet – The Internet of Us

EC funded FET Proactive Project

Start Date: **1st January 2019**
Duration: **48 Months**
Total budget: **6.5 M€**
Coordinator: **Uni. of Trento (F.Giunchiglia)**

WeNet Main Goal

The main goal of WeNet is developing an online platform which will empower machine mediated diversity-aware people interactions.

WeNet Consortium

WeNet Consortium

dept. of informatics (coordinator)

WeNet Consortium

machine learning

WeNet Consortium

incentives

WeNet Consortium

norms / interaction models

WeNet Consortium

sociology

WeNet Consortium

ethics

WeNet Consortium

interaction design

WeNet Consortium

IT company

WeNet Consortium

impact creation

WeNet Consortium

use case partners

WeNet Consortium

WENET
INTERNET OF US

Norms and Normative Systems

Norms, an overview

Norms are what govern / regulate behaviour

- Regulative Norms
- Legislative Norms
- Social Norms
- Values and Principles

The Study of Norms

Sociology

- influence behaviour
- norm emergence
- motivation for norm adherence

Philosophy

- deontic logic
- axiomatic perspective
- semantic interpretations
- paradoxes

Legal Studies

- institutional norms (regulate interactions)
- norms as deontic rules
- norm enforcement
- conflict resolution

Neuroscience

- social influence, persuasion
- social conformity

Economics

- influence market behaviour

Norms in AI & Multiagent Systems

- formalising legal statements
- normative reasoning
- argumentation

AI & Law

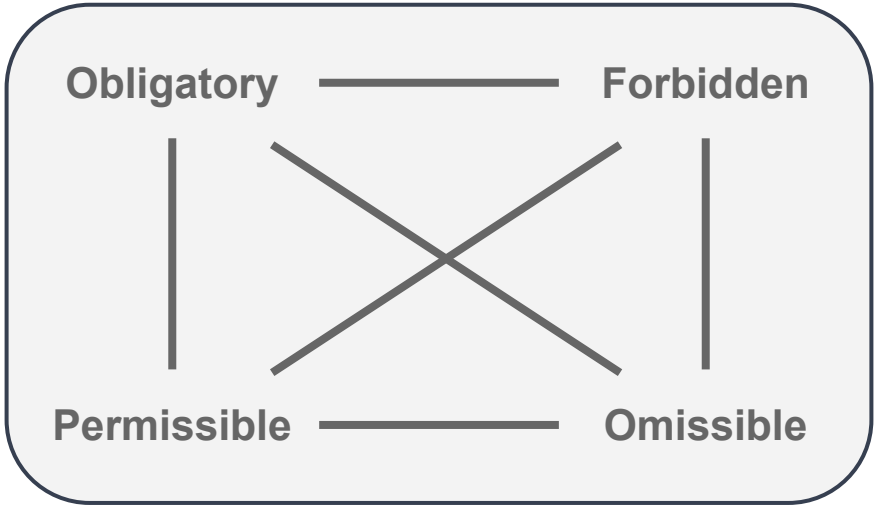
Norms in AI & Multiagent Systems

Norm Representation

Mostly based on **Deontic Logic.**

Norm Representation

Mostly based on **Deontic Logic.**



Norm Representation

Mostly based on **Deontic Logic.**

- **If-Then rules**
- **Deontic Logic**:
  - Conditional Deontic Logic with Deadlines
- **Event Calculus**
- **Expectations & Constraints**
  - Social Integrity Constraints
- **Commitments**
  - Object Constraint Language
- **Temporal Logic**
  - Hybrid Metric Interval Temporal Logic
  - Normative temporal logic (NTL)

Norm Representation

Mostly based on **Deontic Logic.**

- **If-Then rules**
- **Deontic Logic**:
  - Conditional Deontic Logic with Deadlines
- **Event Calculus**
- **Expectations & Constraints**
  - Social Integrity Constraints
- **Commitments**
  - Object Constraint Language
- **Temporal Logic**
  - Hybrid Metric Interval Temporal Logic
  - Normative temporal logic (NTL)

If the auctioneer has announced the current price and no buyer has said 'mine!', then the auctioneer can say 'next!'.

example from the SIMPLE language

Norm Representation

Mostly based on **Deontic Logic.**

- **If-Then rules**
- **Deontic Logic**:
  - Conditional Deontic Logic with Deadlines
- **Event Calculus**
- **Expectations & Constraints**
  - Social Integrity Constraints
- **Commitments**
  - Object Constraint Language
- **Temporal Logic**
  - Hybrid Metric Interval Temporal Logic
  - Normative temporal logic (NTL)

```
PERMITTED(
        (user DO appoint(regular user))
          IF
          (access level(user, register,
                'full control')))

OBLIGED(
        (buyer DO bid(product,price))
          BEFORE
          (buyer DO exit(auction house)))
```

Norm Representation

Mostly based on **Deontic Logic.**

- **If-Then rules**
- **Deontic Logic**:
  - Conditional Deontic Logic with Deadlines
- **Event Calculus**
- **Expectations & Constraints**
  - Social Integrity Constraints
- **Commitments**
  - Object Constraint Language
- **Temporal Logic**
  - Hybrid Metric Interval Temporal Logic
  - Normative temporal logic (NTL)

```
initiates(assign floor(C, S),
      status = granted(S, TO), T) ←
            role of(C, chair),
            role of(S, subject),
            holdsAt(status = free, T),
            (TO := T + 5)
```

Norm Representation

Mostly based on **Deontic Logic.**

- **If-Then rules**
- **Deontic Logic**:
  - Conditional Deontic Logic with Deadlines
- **Event Calculus**
- **Expectations & Constraints**
  - Social Integrity Constraints
- **Commitments**
  - Object Constraint Language
- **Temporal Logic**
  - Hybrid Metric Interval Temporal Logic
  - Normative temporal logic (NTL)

$$SIC = \{H(tell(X, Y, start)) \rightarrow E(pass(Y))\}$$

Norm Representation

Mostly based on **Deontic Logic.**

- **If-Then rules**
- **Deontic Logic**:
  - Conditional Deontic Logic with Deadlines
- **Event Calculus**
- **Expectations & Constraints**
  - Social Integrity Constraints
- **Commitments**
  - Object Constraint Language
- **Temporal Logic**
  - Hybrid Metric Interval Temporal Logic
  - Normative temporal logic (NTL)

```
makePendingComm(
        agent,
        a.organization,
        (openAuction(a.id),[now,now+δ],›))
```

Norm Representation

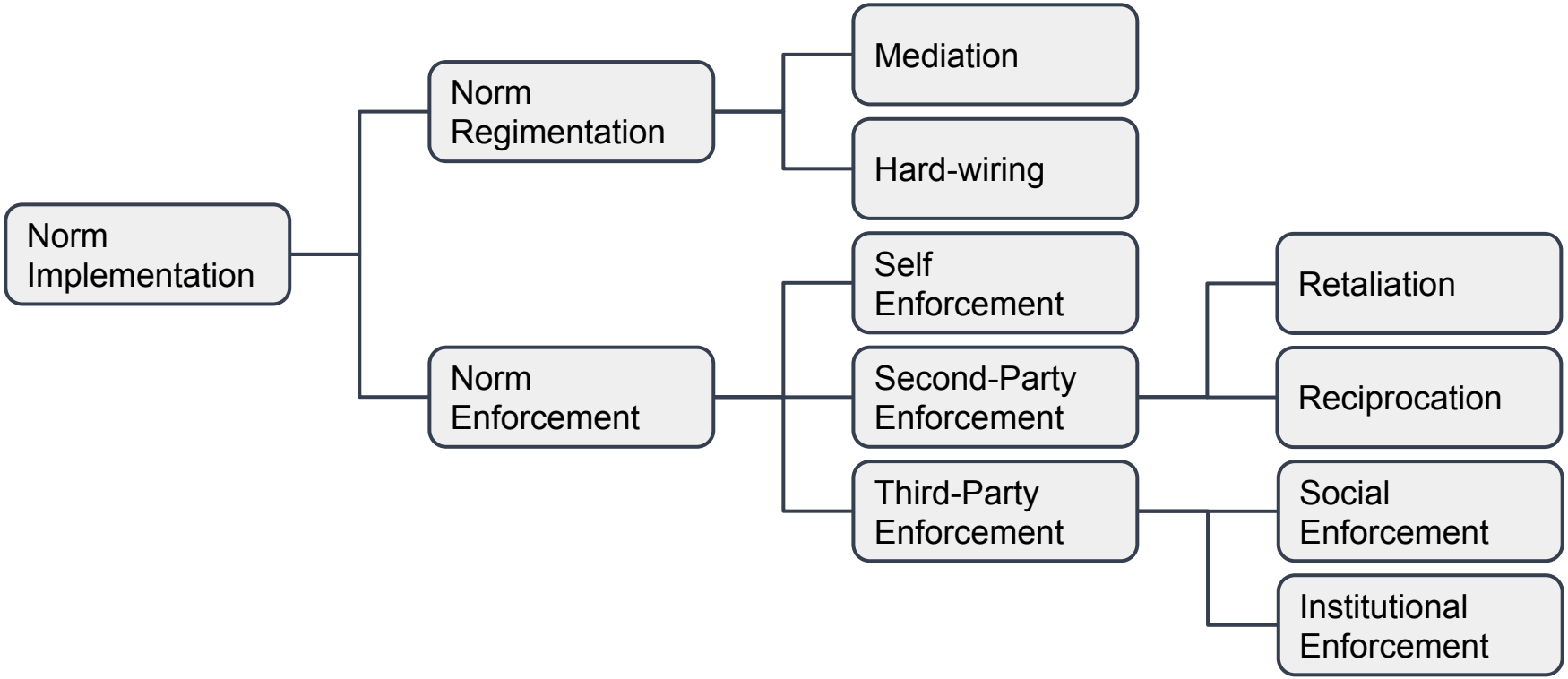
Mostly based on **Deontic Logic.**

- **If-Then rules**
- **Deontic Logic**:
  - Conditional Deontic Logic with Deadlines
- **Event Calculus**
- **Expectations & Constraints**
  - Social Integrity Constraints
- **Commitments**
  - Object Constraint Language
- **Temporal Logic**
  - Hybrid Metric Interval Temporal Logic
  - Normative temporal logic (NTL)

$$O\eta \bigcirc((wWaiting \wedge \neg wGreen) \rightarrow \neg P\eta \, \square \, wTunnel)$$

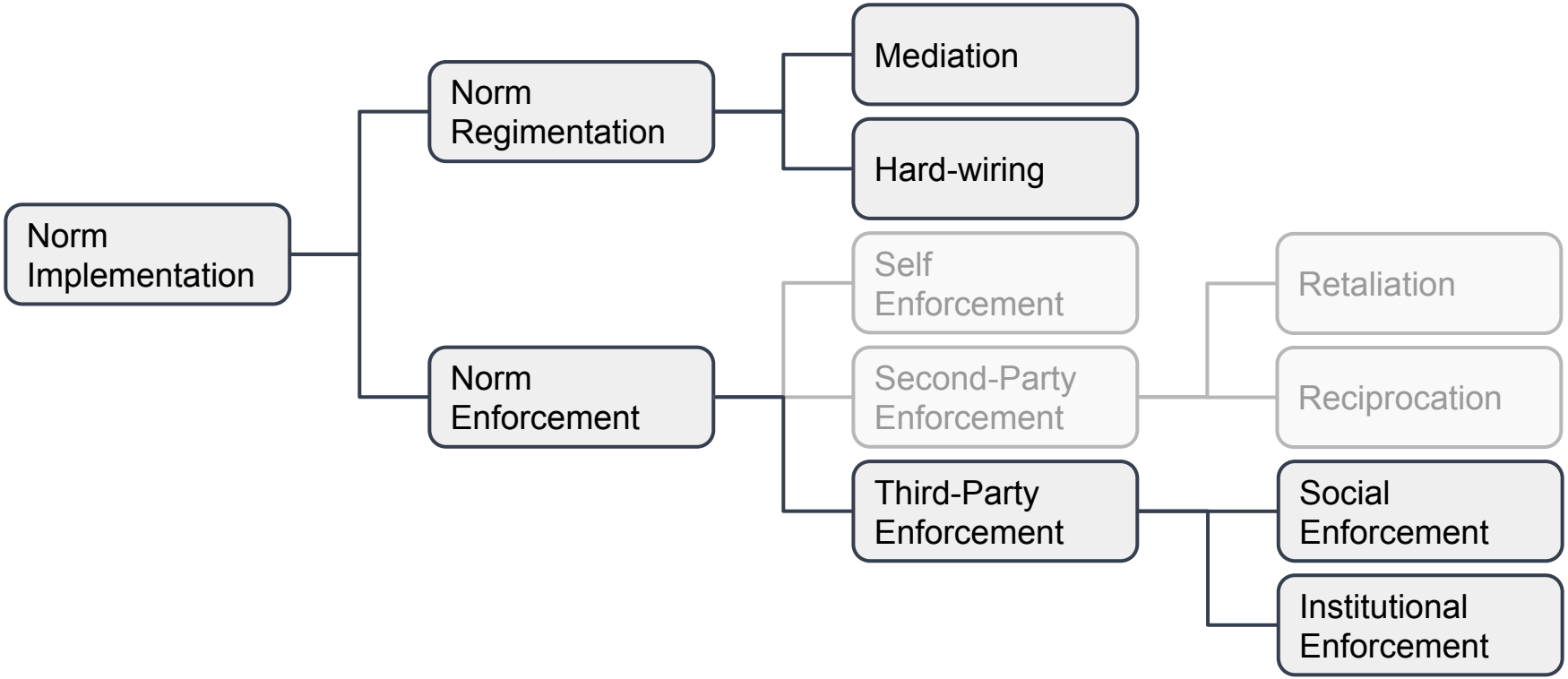
$$A\alpha \equiv O\eta\alpha$$
$$E\alpha \equiv P\eta\alpha$$

Norm Implementation



Natalia Criado Pacheco, Using Norms To Control Open Multi-Agent Systems. PhD Thesis, UPV, 2012.

Norm Implementation



Natalia Criado Pacheco, Using Norms To Control Open Multi-Agent Systems. PhD Thesis, UPV, 2012.

Norm Reasoning

- **Norm diagnosis.** Check and verify properties of norms.
- **Conflict resolution.** Check for inconsistencies.
- **Norm compliance.** Assess consequences of obeying norms.

Norm Creation

- **Top-Down Approaches**
  - **Offline design**
  - **Online norm synthesis**
    - driven by conflict detection

- **Bottom-Up Approaches**
  - **Norm Emergence**: usually focuses on internalisation of norms

Norm Creation

- **Top-Down Approaches**
  - **Offline design**
  - **Online norm synthesis**
    - driven by conflict detection
    - learning mechanisms (though limited to learning parameters, punishment, ...)
- **Bottom-Up Approaches**
  - **Norm Emergence**: usually focuses on internalisation of norms

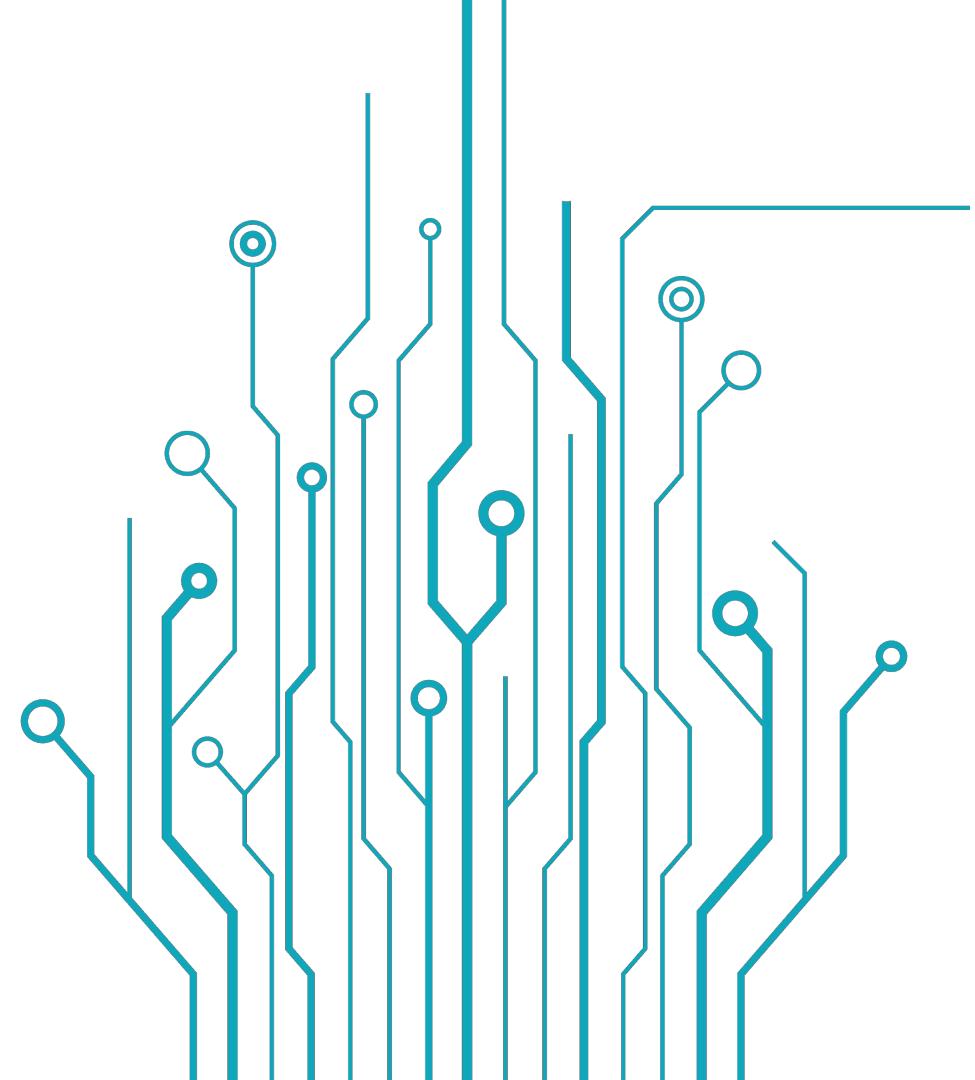
Norm Creation

- **Top-Down Approaches**
  - **Offline design**
  - **Online norm synthesis**
    - driven by conflict detection
    - learning mechanisms (though limited to learning parameters, punishment, ...)
- **Bottom-Up Approaches**
  - **Norm Emergence**: usually focuses on internalisation of norms

Norm emergence triggers top-down norm creation.

Norm Creation

- **Top-Down Approaches**
  - **Offline design**
  - **Online norm synthesis**
    - driven by conflict detection
    - learning mechanisms (though limited to learning parameters, punishment, ...)
- **Bottom-Up Approaches**
  - **Norm Emergence**: usually focuses on internalisation of norms
  - **Norm Agreement**

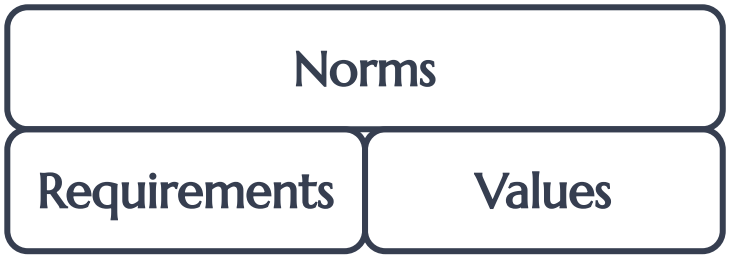


WENET
INTERNET OF US

Value-Driven Adaptive Norms

Evolution of Norms

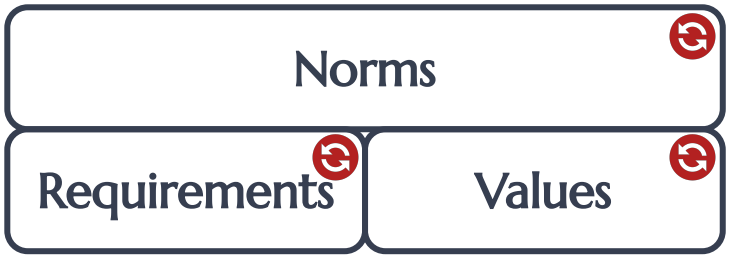
Norms are based on our ***requirements*** and ***values***.



Evolution of Norms

Norms are based on our ***requirements*** and ***values***.

And they should **adapt** to our **evolving** requirements and values!

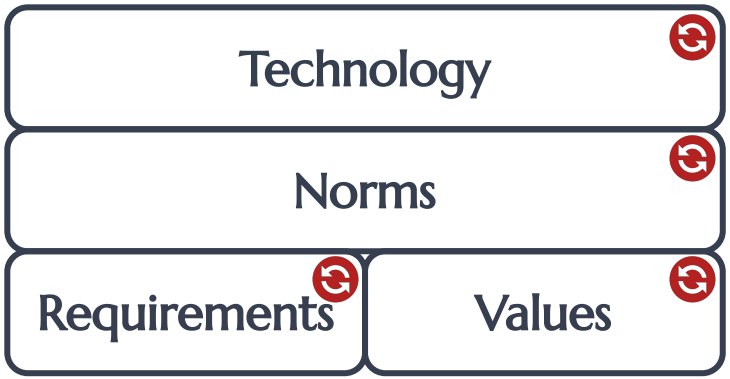


Evolution of Norms

Norms are based on our ***requirements*** and ***values***.

And they should **adapt** to our **evolving** requirements and values!

As such, our technologies should be mediated by **adaptive** norms.

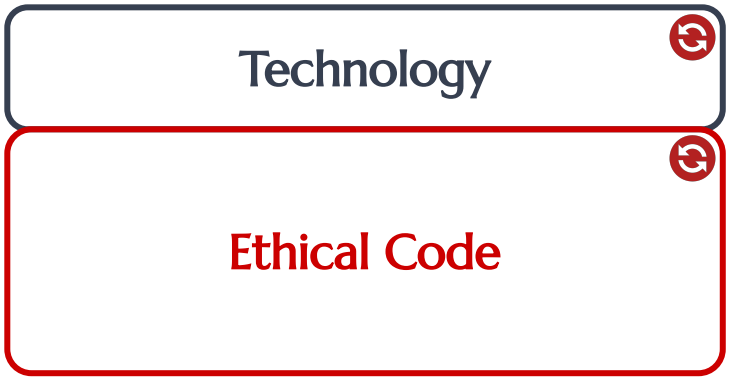


Evolution of Norms

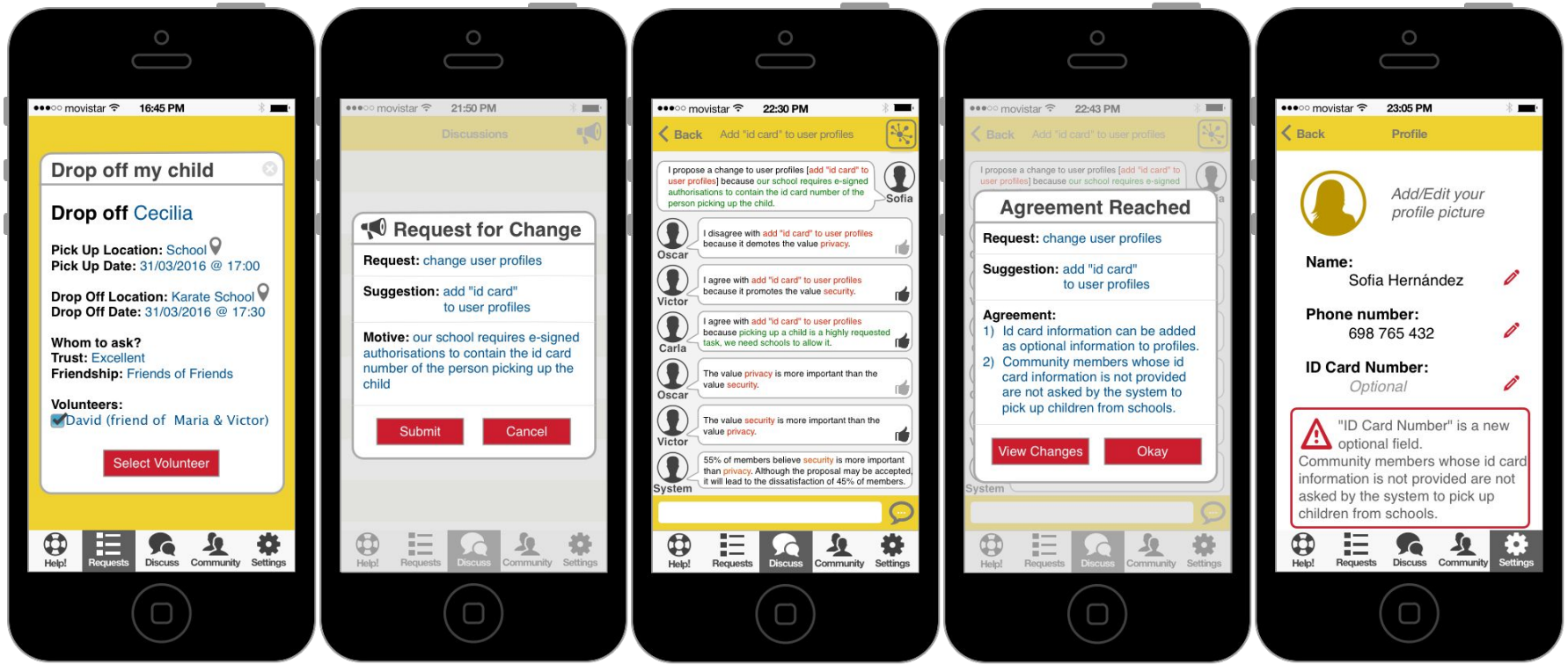
Norms are based on our ***requirements*** and ***values***.

And they should **adapt** to our **evolving** requirements and values!

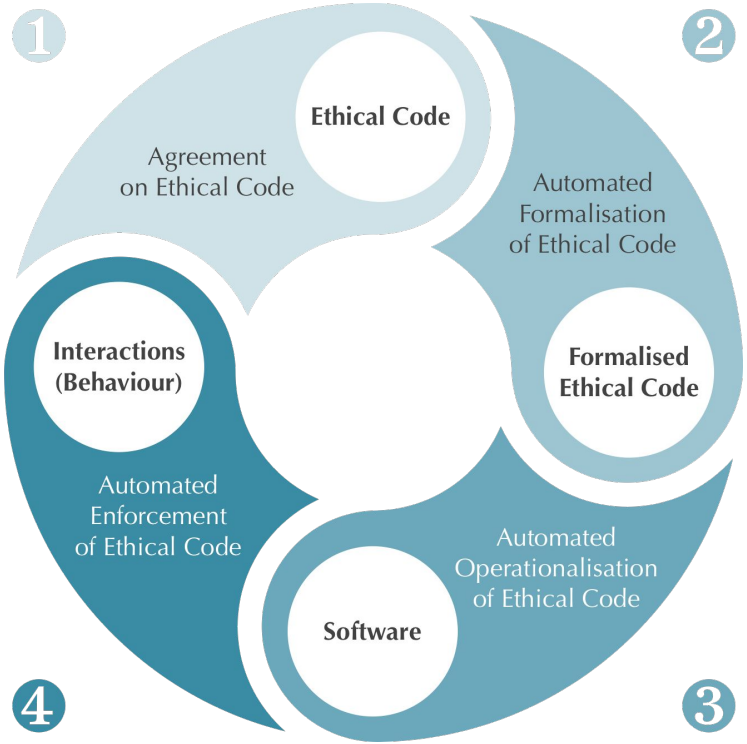
As such, our technologies should be mediated by **adaptive** norms.



Motivating Example: uHelp app



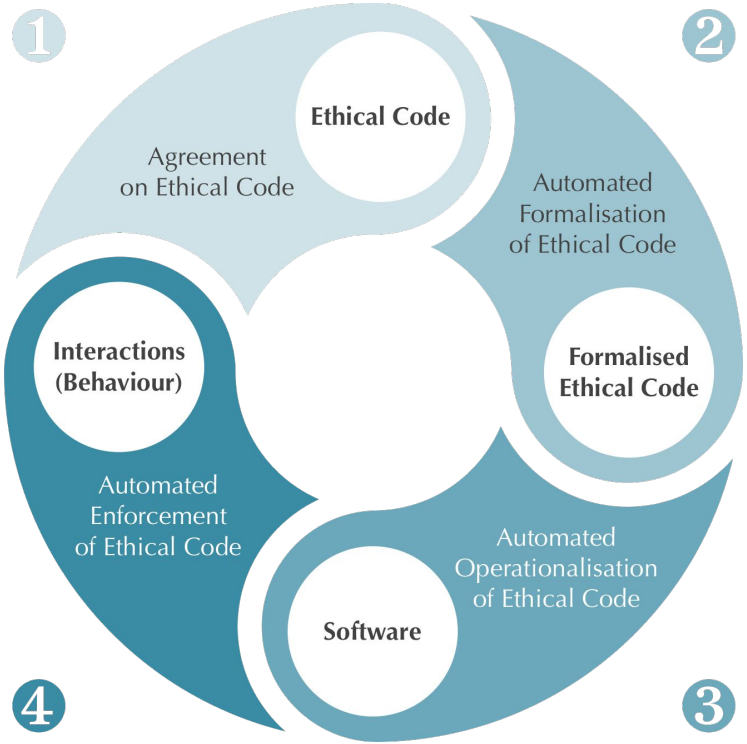
Evolution of Technology



Evolution of Technology

Agreement Technologies

Argumentation
Negotiation
Trust & Reputation
Computational Social Choice



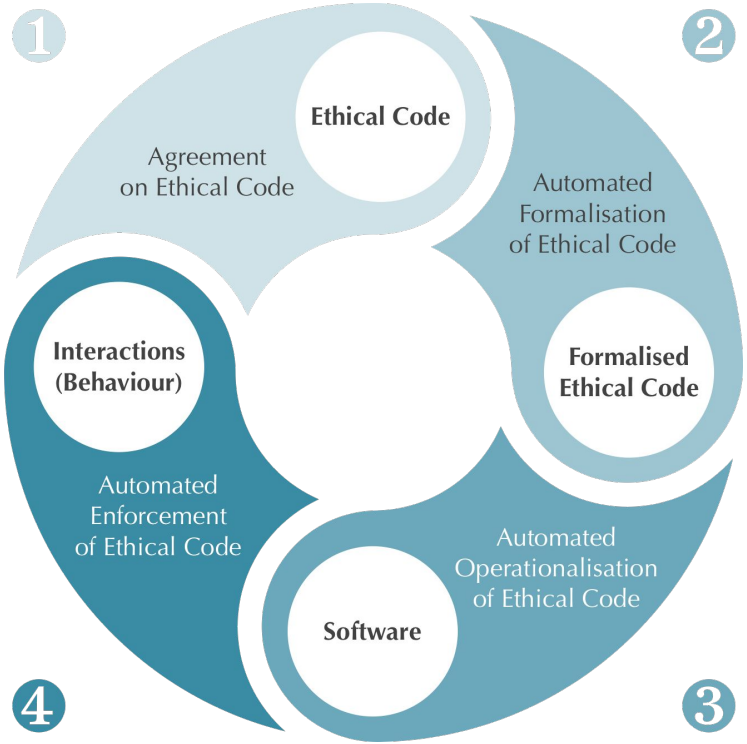
Evolution of Technology

Agreement Technologies

Learning

Learn when to change norms
Learn the best norms
Learn norm consequences

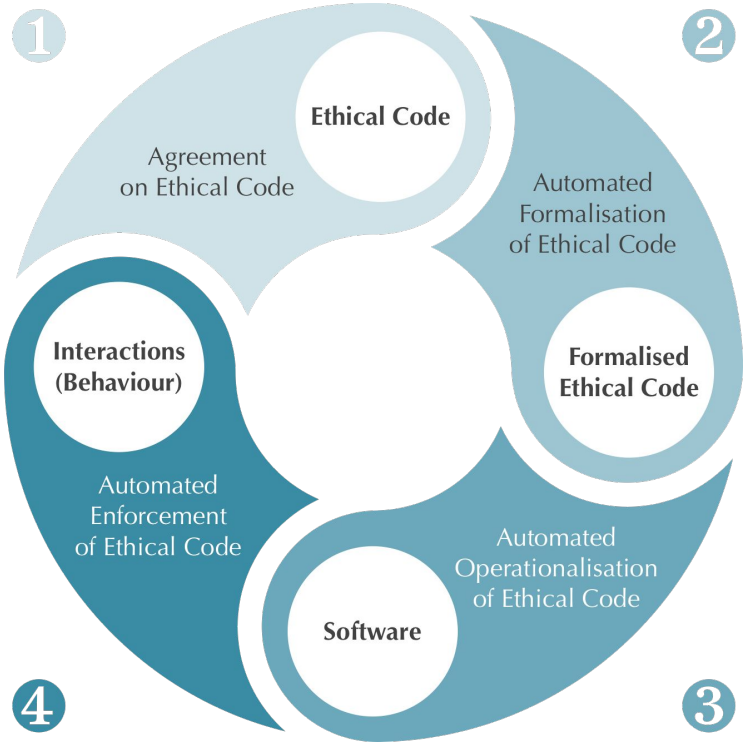
ML / CBR / simulations / sentiment analysis / analogical reasoning / coherence theory / norm synthesis



Evolution of Technology

Agreement Technologies

Learning



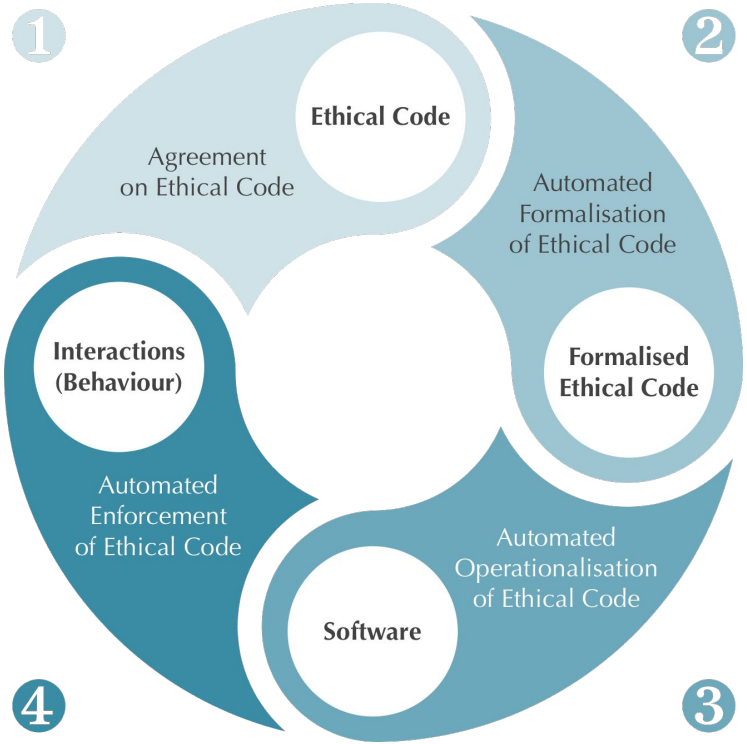
Logic for Norms

First Order Logic
Modal Logic
Deontic Logic
...

Evolution of Technology

Agreement Technologies

Learning



Logic for Norms

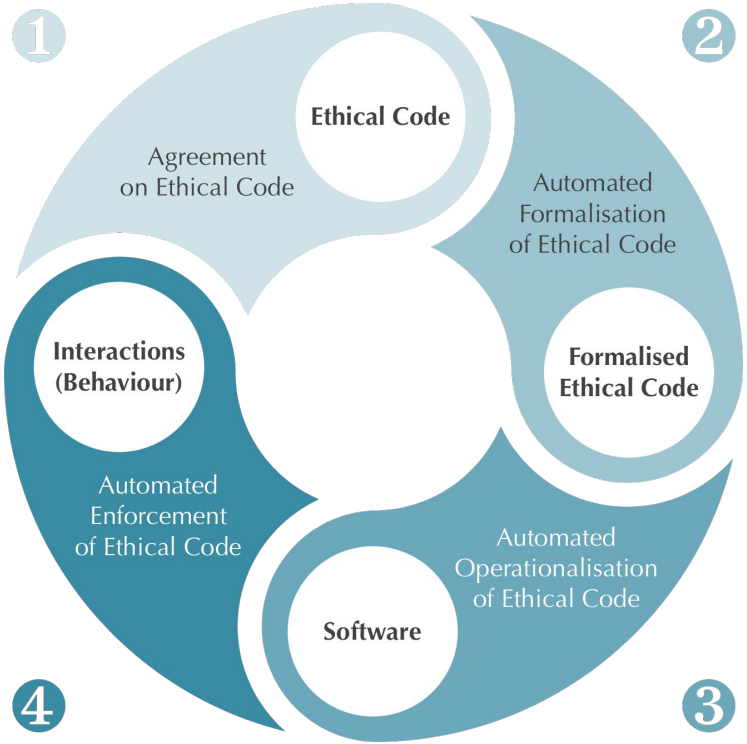
Natural Language Processing

Recognising norms
Extracting modalities & their parameters

Evolution of Technology

Agreement Technologies

Learning



Logic for Norms

Natural Language Processing

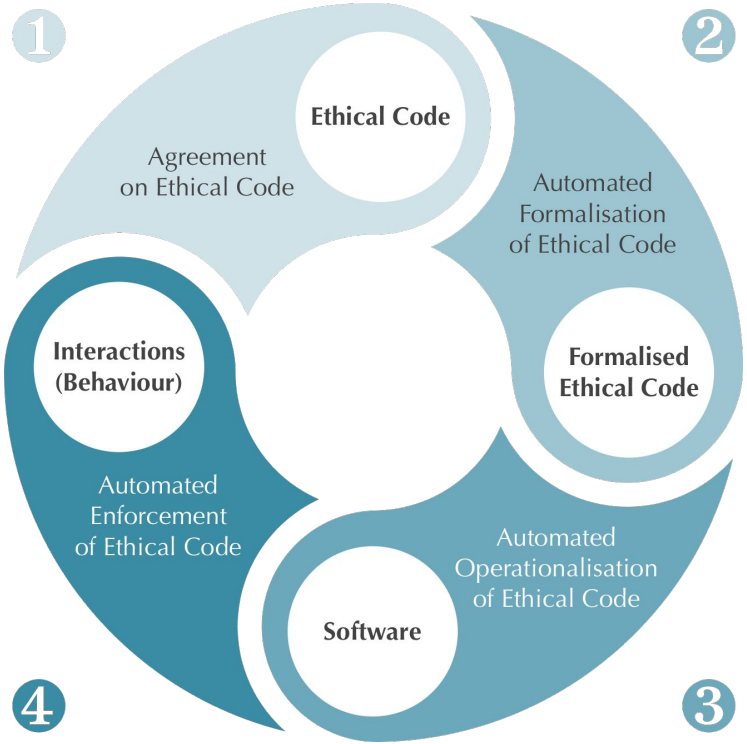
Normative Systems

Processes
If-Then statements
Constraints
...

Evolution of Technology

Agreement Technologies

Learning



Logic for Norms

Natural Language Processing

Normative Systems

Formal Verification

Model Checking / Automated Theorem Proving

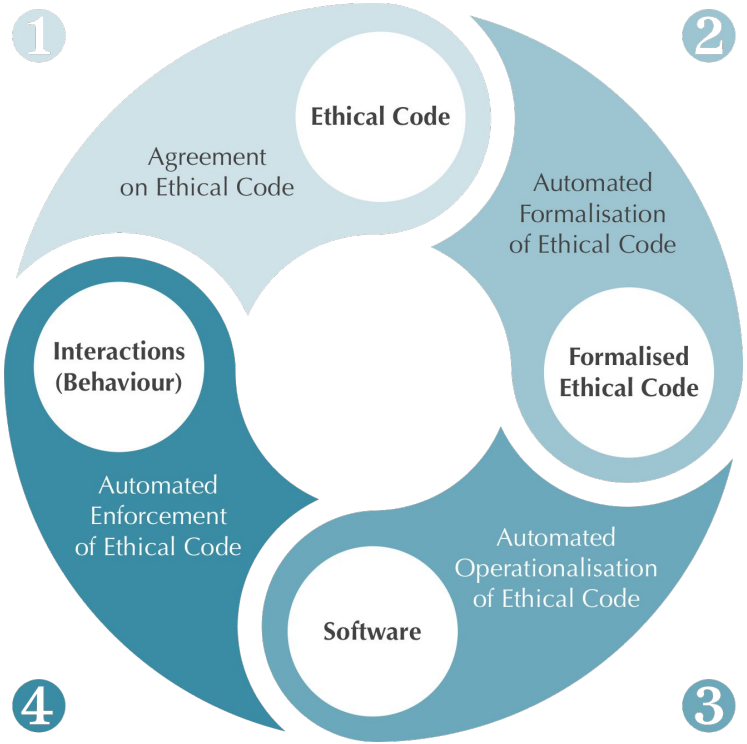
Evolution of Technology

Agreement Technologies

Learning

Norm Enforcement

Providing incentives to comply "Punish" defects



Logic for Norms

Natural Language Processing

Normative Systems

Formal Verification

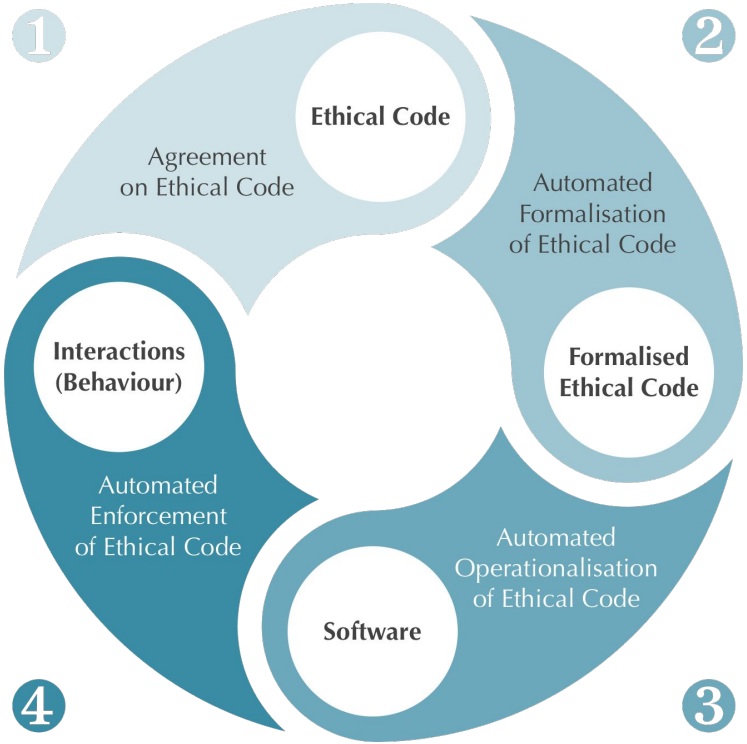
Evolution of Technology

Agreement Technologies

Learning

Norm Enforcement

Automated GUIs



Logic for Norms

Natural Language Processing

Normative Systems

Formal Verification

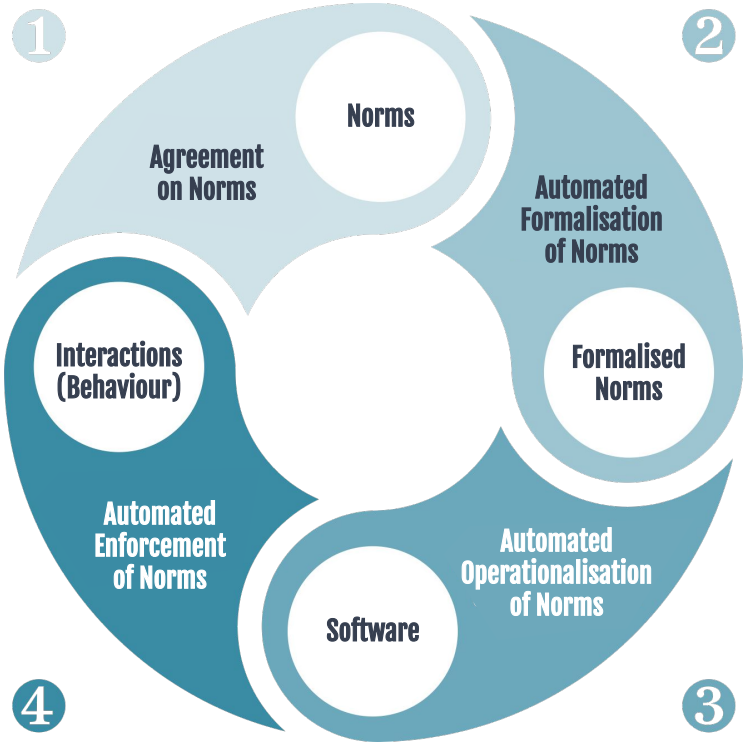
Evolution of Norms

Agreement Technologies

Learning

Norm Enforcement

Automated GUIs



Logic for Norms

Natural Language Processing

Normative Systems

Formal Verification

Research Challenges

- **Closing the cycle**
- **Introducing values**
  - Linking values to norms
  - Value driven agreements
  - Specifying values
  - Verifying values
- **Giving people control**

Research Challenges

- **Closing the cycle**
- **Introducing values**
  - Linking values to norms
  - Value driven agreements
  - Specifying values
  - Verifying values
- **Giving people control**

The Tragedy of the Commons

"Freedom in a commons brings ruin to all."

Garrett Hardin (1968)

The Tragedy of the Commons

forests & meadows in Switzerland

mountain meadows in Japan

huerta irrigation in Spain

zanjera irrigation in Philippines

fisheries in Indonesia

The commons have been effectively managed by some!

Elinor Ostrum (1990)

Governing the Commons

8 Design Principles for Common Pool Resource (CPR) Institutions.

- **Clearly defined boundaries**
- **Congruence**
- **Collective-choice arrangements**
- **Monitoring**
- **Graduated sanctions**
- **Conflict-resolution mechanisms**
- **Minimal recognition of rights to organise**
- **Nested enterprises** *(for systems that are part of larger systems)*

Governing the Commons

8 Design Principles for Common Pool Resource (CPR) Institutions.

- **Clearly defined boundaries**
- **Congruence**
- **Collective-choice arrangements**
- **Monitoring**
- **Graduated sanctions**
- **Conflict-resolution mechanisms**
- **Minimal recognition of rights to organise**
- **Nested enterprises** *(for systems that are part of larger systems)*

Governing the Commons

8 Design Principles for Common Pool Resource (CPR) Institutions.

- **Clearly defined boundaries**
- **Congruence**
- **Collective-choice arrangements**
- **Monitoring**
- **Graduated sanctions**
- **Conflict-resolution mechanisms**
- **Minimal recognition of rights to organise**
- **Nested enterprises** *(for systems that are part of larger systems)*

Governing the Commons

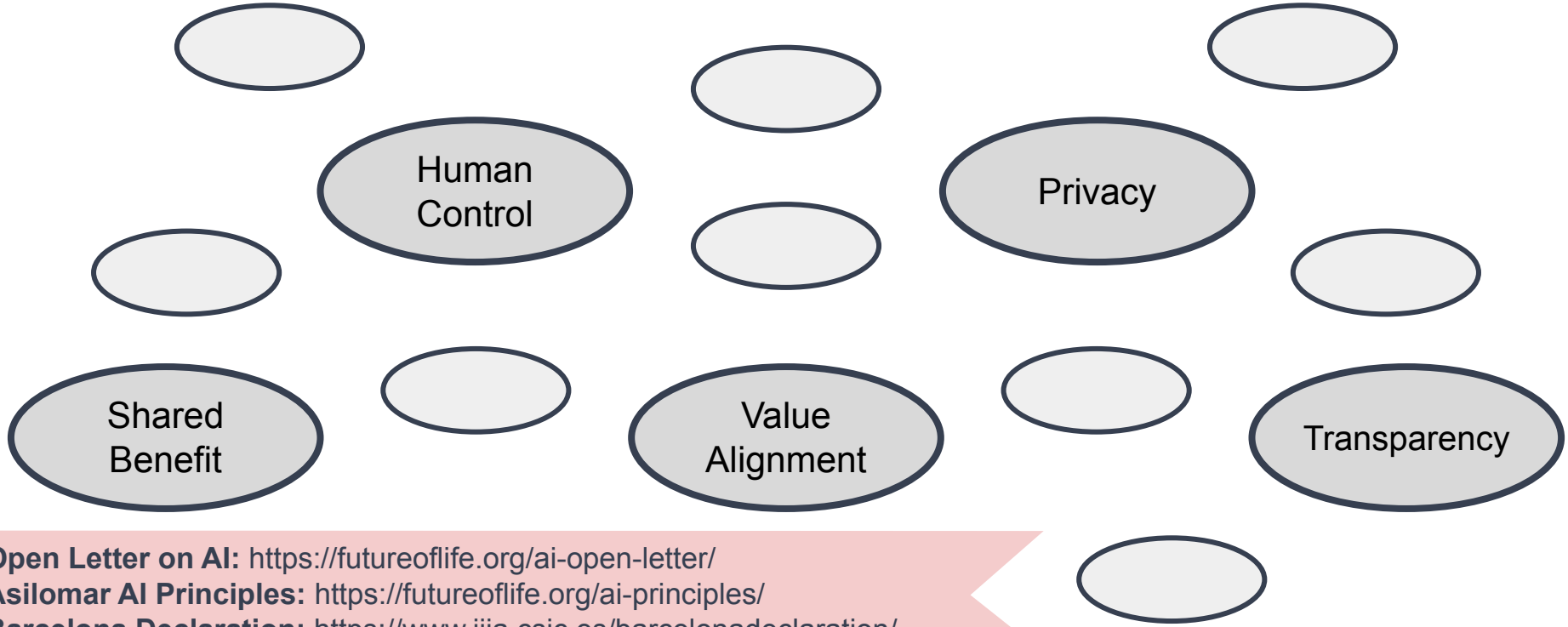
8 Design Principles for Common Pool Resource (CPR) Institutions.

- **Clearly defined boundaries**
- **Congruence**
- **Collective-choice arrangements**
- **Monitoring**
- **Graduated sanctions**
- **Conflict-resolution mechanisms**
- **Minimal recognition of rights to organise**
- **Nested enterprises** *(for systems that are part of larger systems)*

WENET
INTERNET OF US

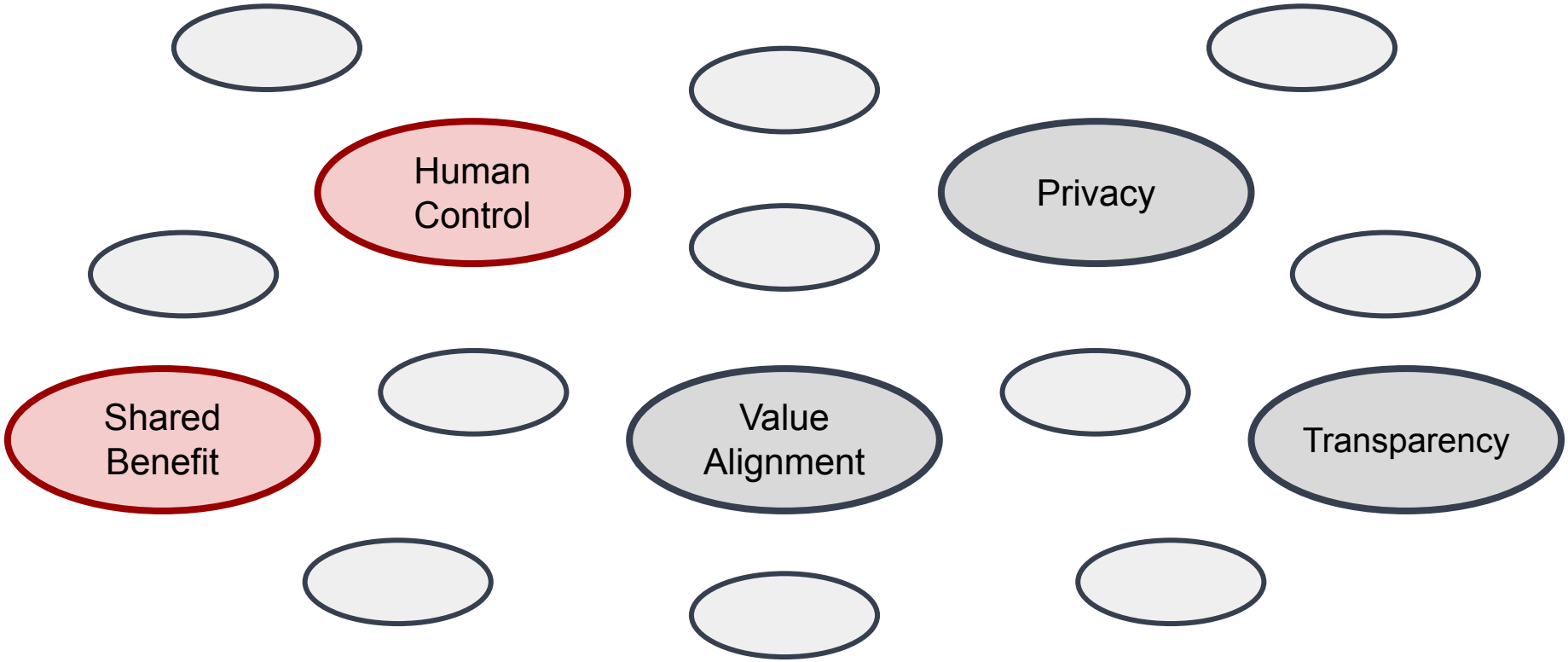
Ethical Implications

Today's Ethical Concerns

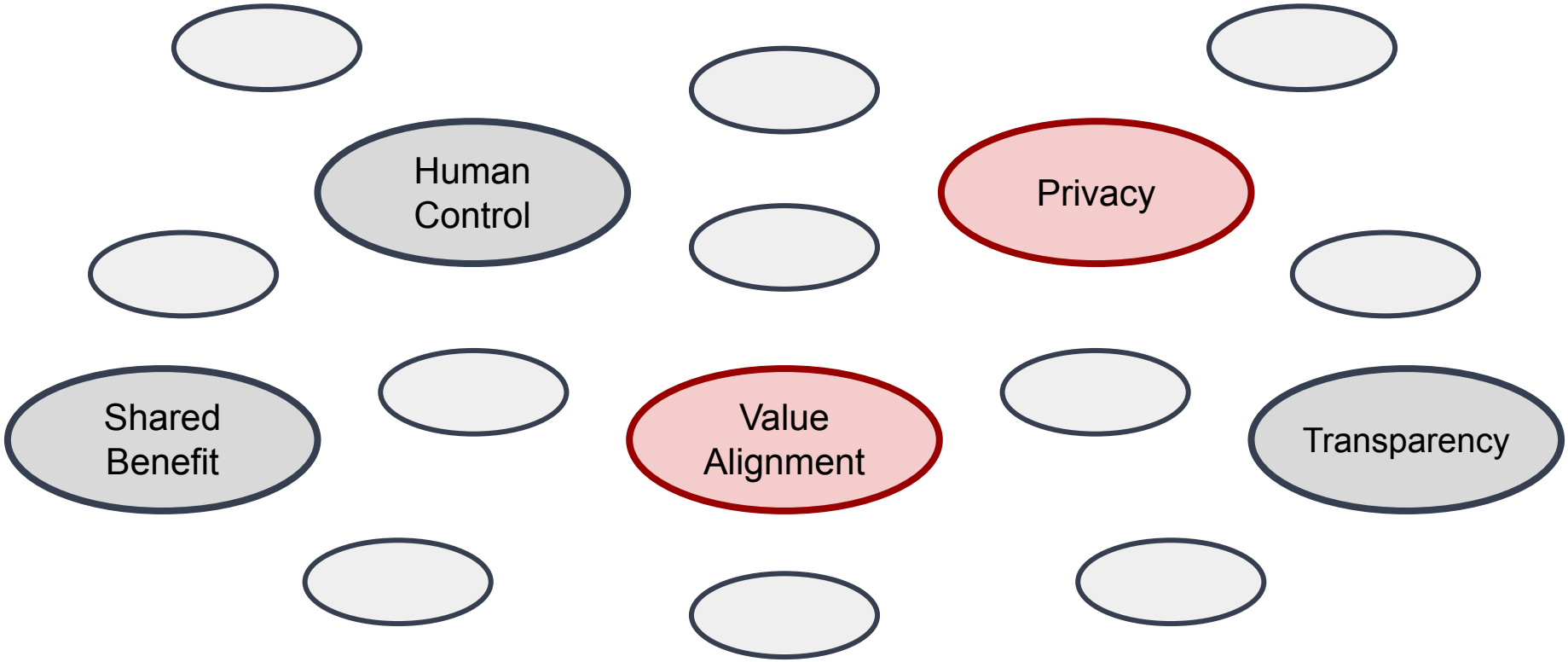


Open Letter on AI: https://futureoflife.org/ai-open-letter/
Asilomar AI Principles: https://futureoflife.org/ai-principles/
Barcelona Declaration: https://www.iiia.csic.es/barcelonadeclaration/

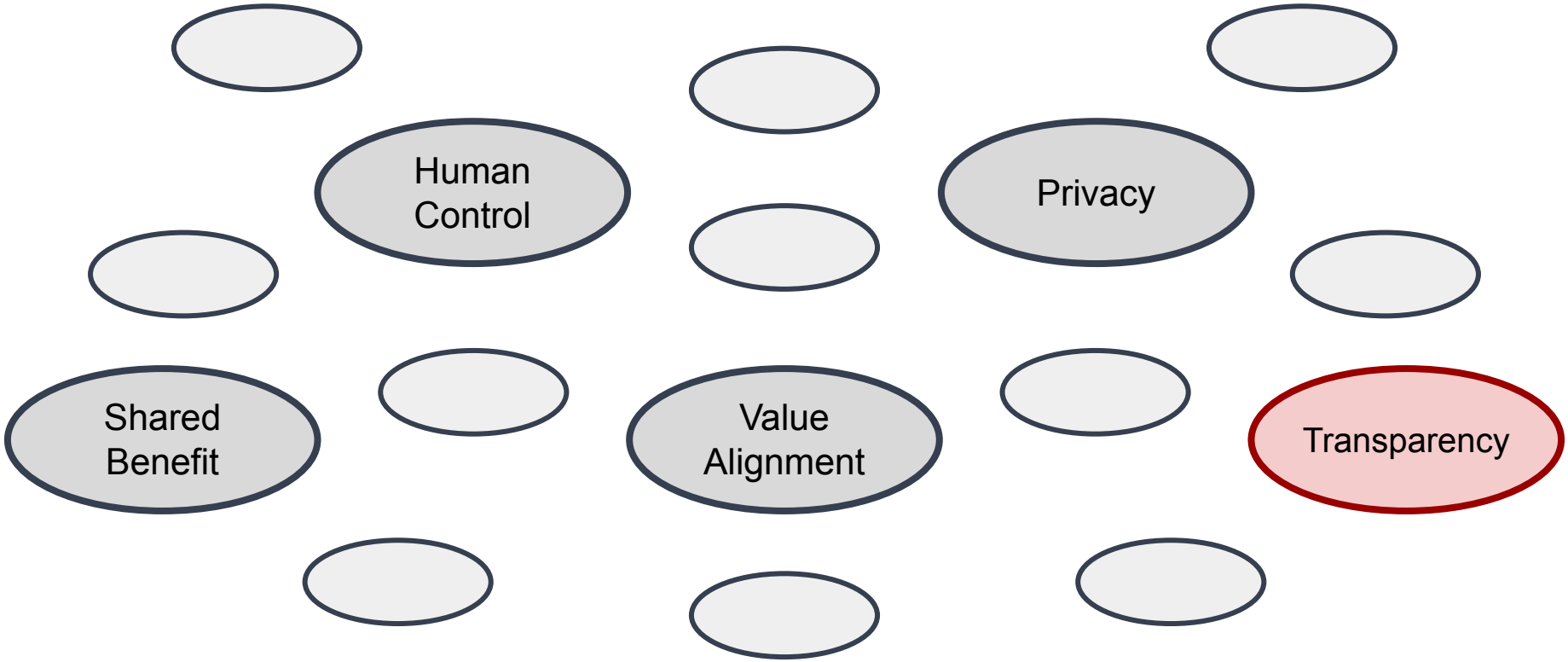
Today's Ethical Concerns



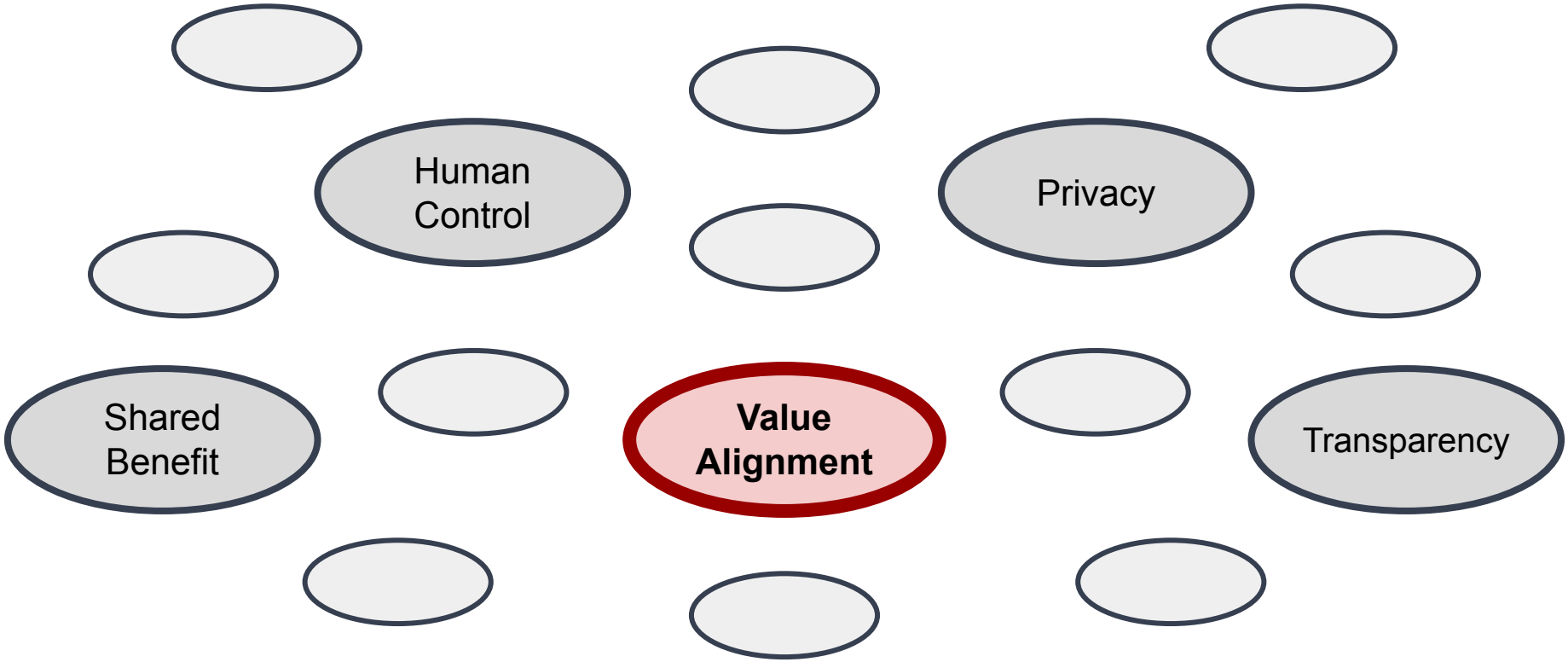
Today's Ethical Concerns



Today's Ethical Concerns



Value Alignment



Value Alignment

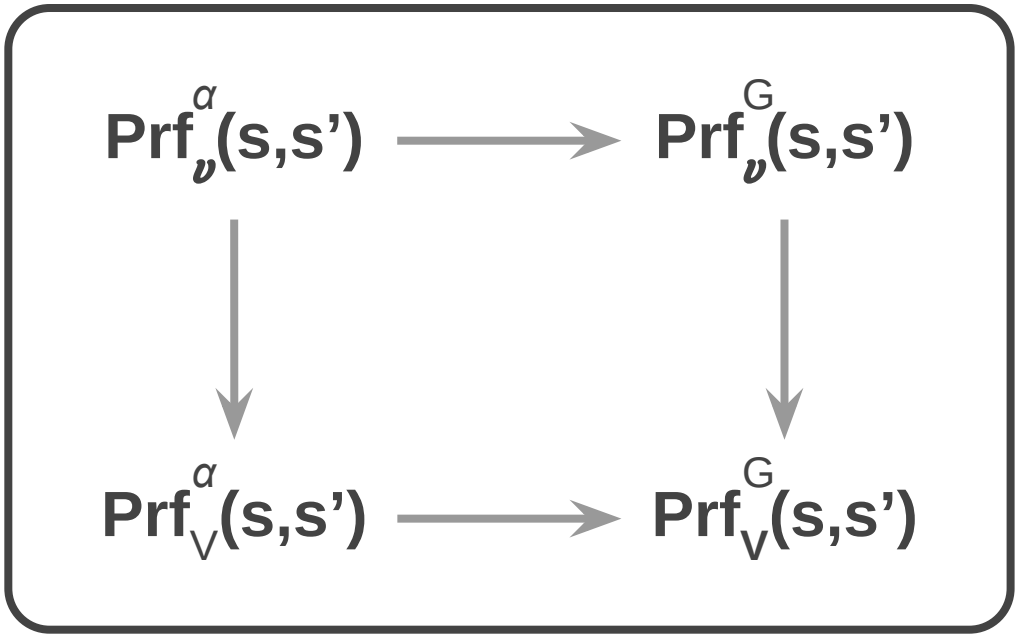
Values as Preferences.

Values are understood as preferences over behaviour, or preferences over the states of the world: $Prf_{v}^{\alpha}(s,s')$

Sierra, C., Osman, N., Noriega, P., Sabater-Mir, J., Perello-Moragues, A. (2019):" Value alignment: a formal approach", Responsible Artificial Intelligence Agents Workshop (RAIA) @ AAMAS 2019.

Value Alignment

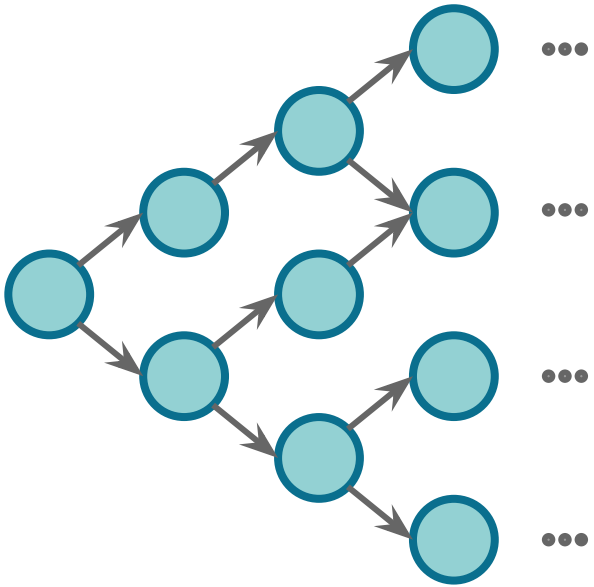
Aggregation of Value-Based Preferences.



Value Alignment

Value Alignment.

One is aligned with a value if their actions move them towards preferred states.

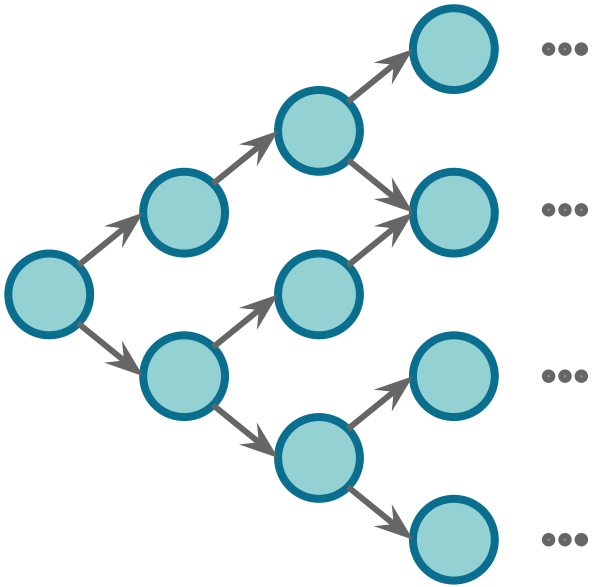


Value Alignment

Value Alignment.

One is aligned with a value if their actions move them towards preferred states.

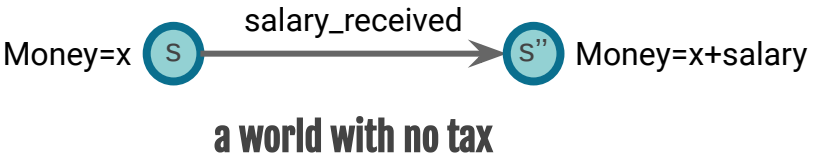
Actions move us toward/away from preferred states.

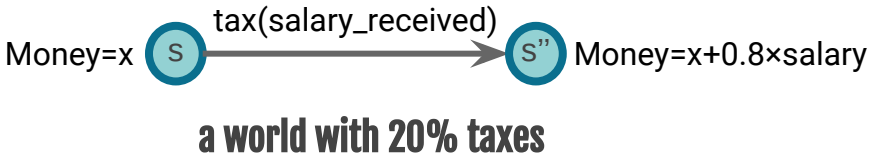


Value Alignment

Alignment of Norms with Values.

Behaviour is governed by norms.



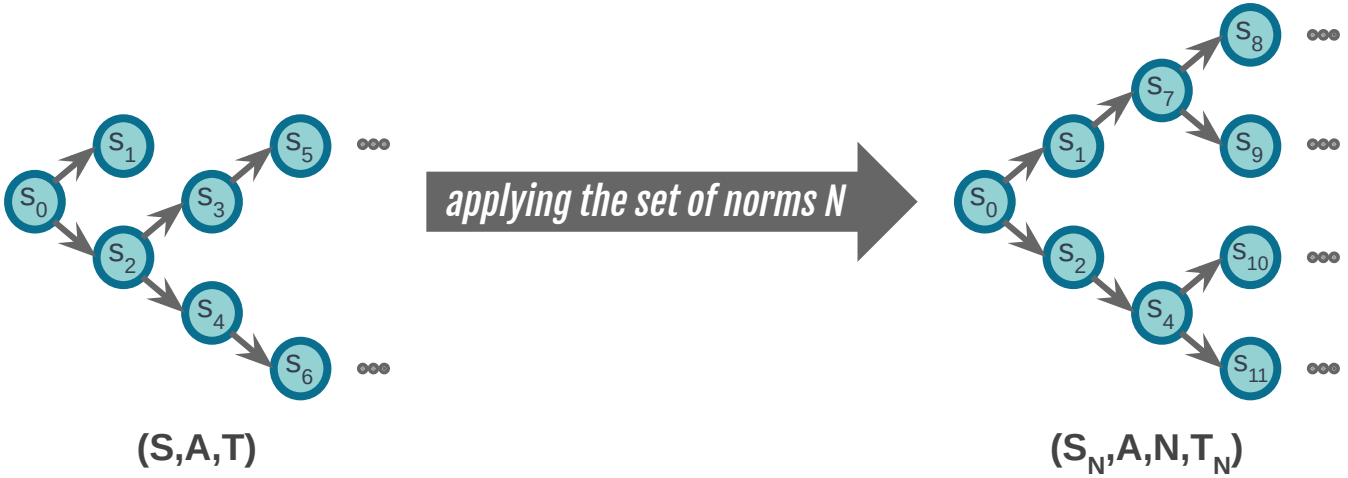


Value Alignment

Alignment of Norms with Values.

Behaviour is governed by norms.

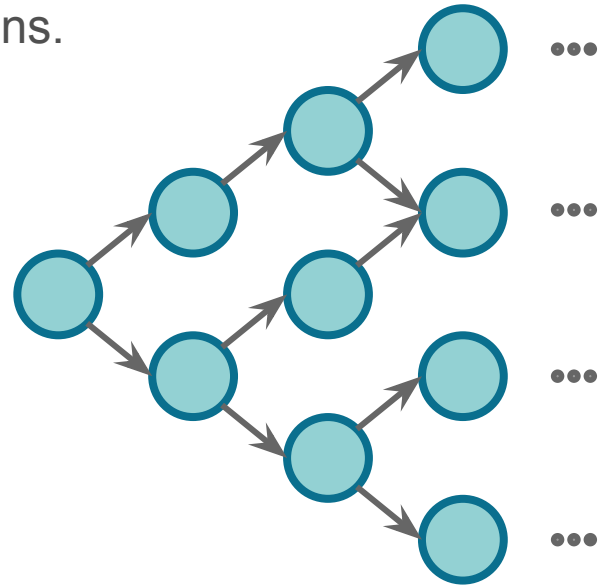
Norms change the world it is applied to.



Value Alignment

Alignment of Norms with Values.

The degree of alignment of a norm n with a value v for agent α is the accumulation of preferences along the transitions.

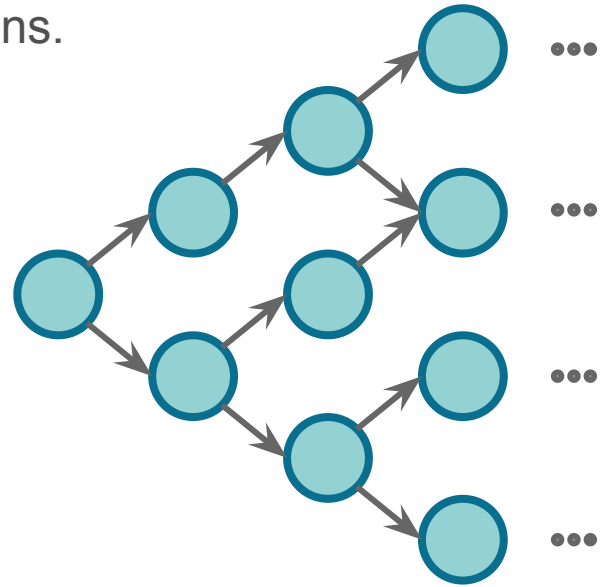


Value Alignment

Alignment of Norms with Values.

The degree of alignment of a norm n with a value v for agent α is the accumulation of preferences along the transitions.

We consider **all possible paths**.



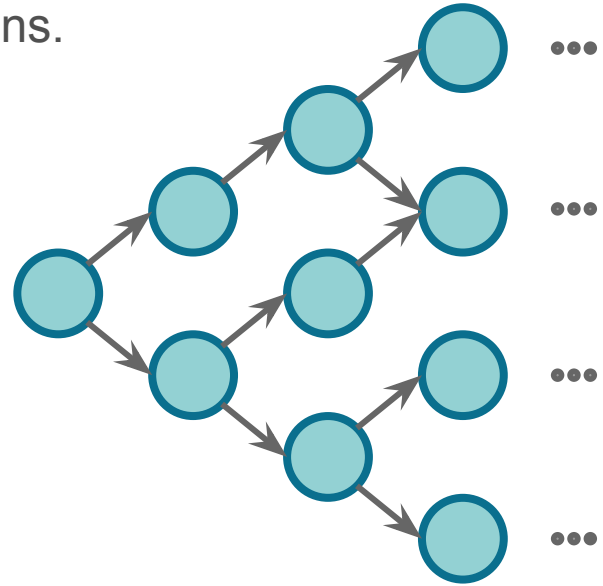
Value Alignment

Alignment of Norms with Values.

The degree of alignment of a norm n with a value v for agent α is the accumulation of preferences along the transitions.

We consider **all possible paths**,

giving **equal weight** to all paths and all transitions.



Value Alignment

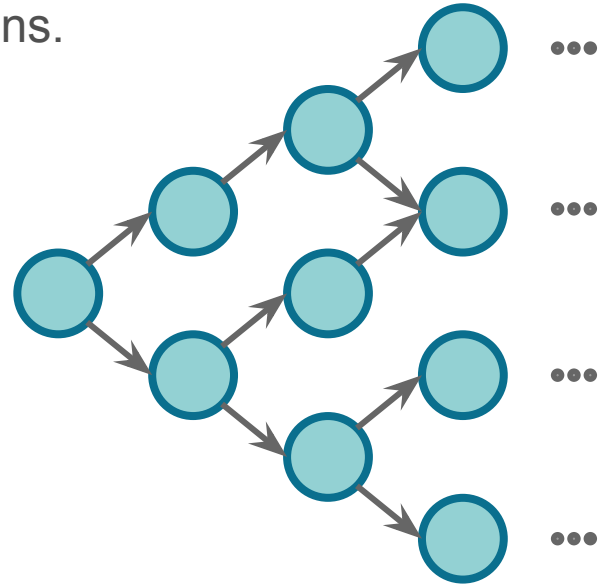
Alignment of Norms with Values.

The degree of alignment of a norm n with a value v for agent α is the accumulation of preferences along the transitions.

We consider **all possible paths**,

giving **equal weight** to all paths and all transitions.

$$\mathsf{Align}^{\alpha}_{n,v}(\mathcal{S},\mathcal{A},T) = \frac{\sum_{p\in paths}\ \sum_{d\in[1,length(p)]} \mathsf{Prf}^{\alpha}_{v}(p_I[d],p_F[d])}{\sum_{p\in paths} length(p)}$$



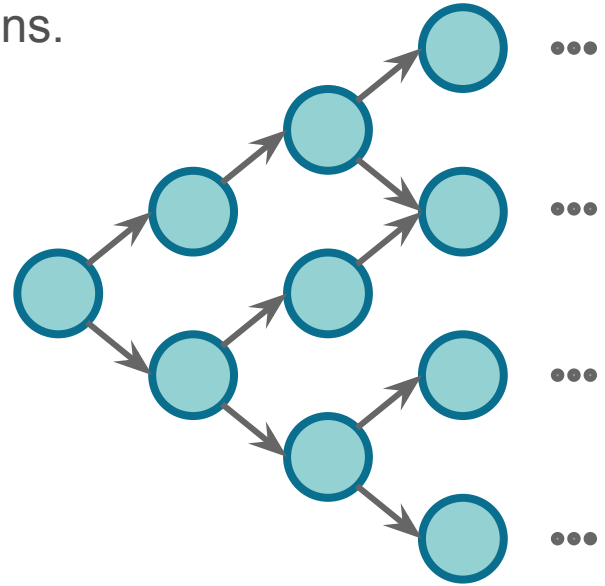
Value Alignment

Alignment of Norms with Values.

The degree of alignment of a norm n with a value v for agent α is the accumulation of preferences along the transitions.

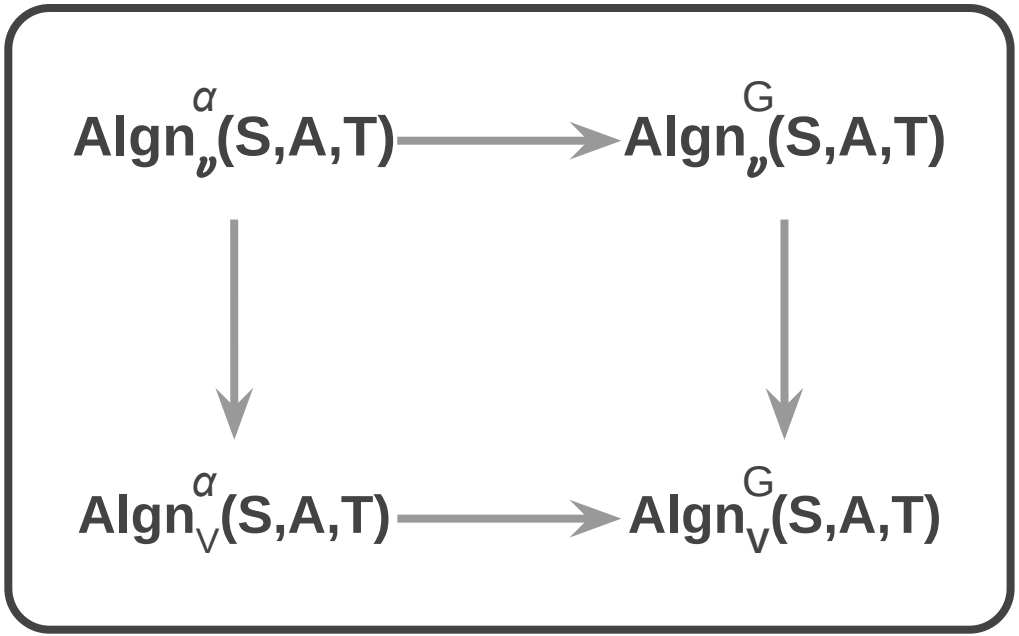
For efficiency, we follow a **Monte Carlo sampling** approach, where x is the number of sampled paths, and l the path length:

$$\mathsf{Align}^{\alpha}_{n,v}(\mathcal{S},\mathcal{A},T) = \frac{\sum_{p \in paths'} \sum_{d \in [1,l]} \mathsf{Prf}^{\alpha}_{v}(p_I[d], p_F[d])}{x * l}$$



Value Alignment

Aggregation of Alignment.



Value Alignment

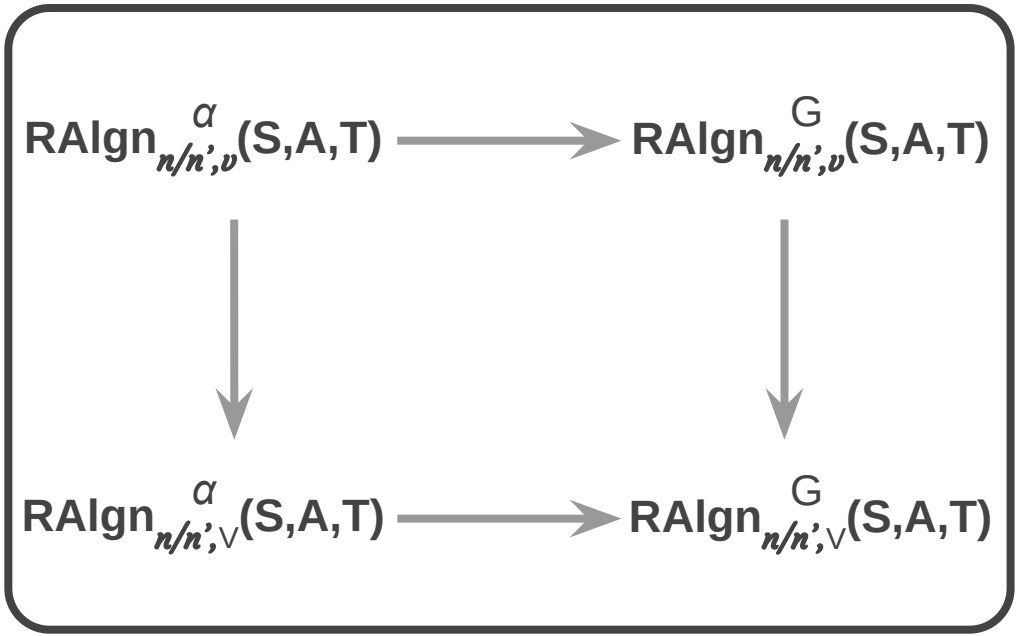
Relative Alignment.

The relative alignment of norm $n1$ w.r.t. norm $n2$:

$$\mathrm{RAlgn}^{\alpha}_{n_1/n_2,v}(\mathcal{S},\mathcal{A},T) = \mathrm{Algn}^{\alpha}_{n_1,V}(\mathcal{S},\mathcal{A},T) - \mathrm{Algn}^{\alpha}_{n_2,V}(\mathcal{S},\mathcal{A},T)$$

Value Alignment

Aggregation of Relative Alignment.



Value Alignment

Research Questions.

- How to define the aggregation functions?
- What if not all paths and transitions are equiprobable?
- Given a set of norms N, a set of values V, and a set of agents G:
  - What is the subset of norms $N^* \in N$ with optimal alignment for group G?

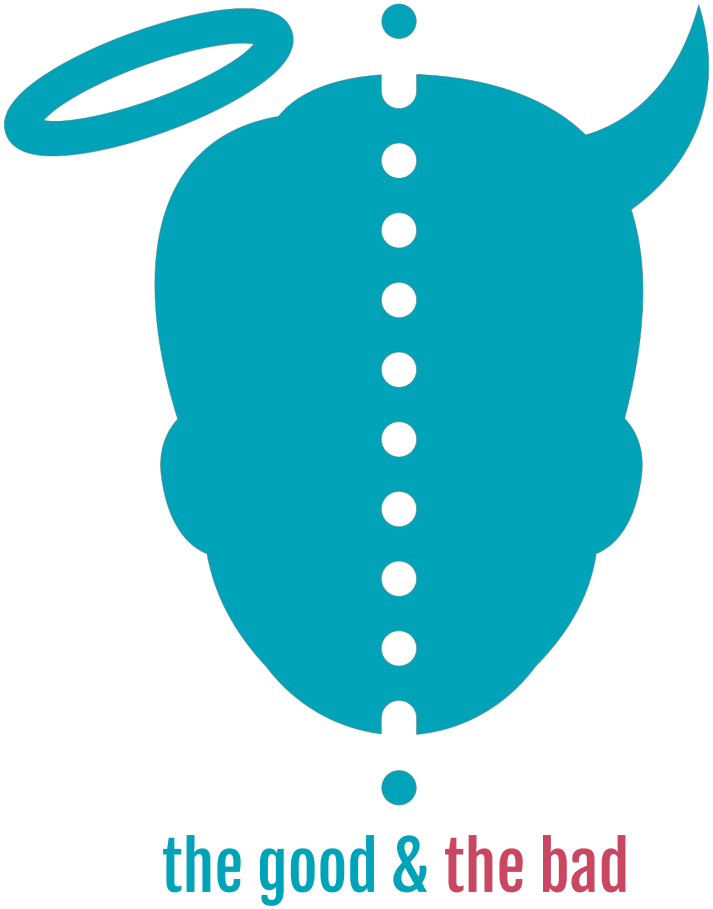
$$N^* = \arg \max_{N' \subseteq N} \mathsf{Algn}^{G}_{N',V}$$

  - What is the subset of agents $G^* \in G$ better aligned with norms N?

$$G^* = \arg \max_{G' \subseteq G} \mathsf{Algn}^{G'}_{N,V}$$

  - What is the optimal social preference aggregation function?

$$f^* = \arg \max_{f \in F} \mathsf{Algn}^{G'}_{N,V}(f\{\mathsf{Prf}^{\alpha}_{V}\}_{\alpha \in G})$$



WENET
INTERNET OF US

Ethical Concerns

Ethical Concern: Human Control

Are we sure we want to give the users full control?

Ethical Concern: Human Control

What if people make “wrong” decisions?

Ethical Concern: Human Control

What if people make “wrong” decisions?

Wrong decisions are defined as:

- **illegal**, or
- **unethical**

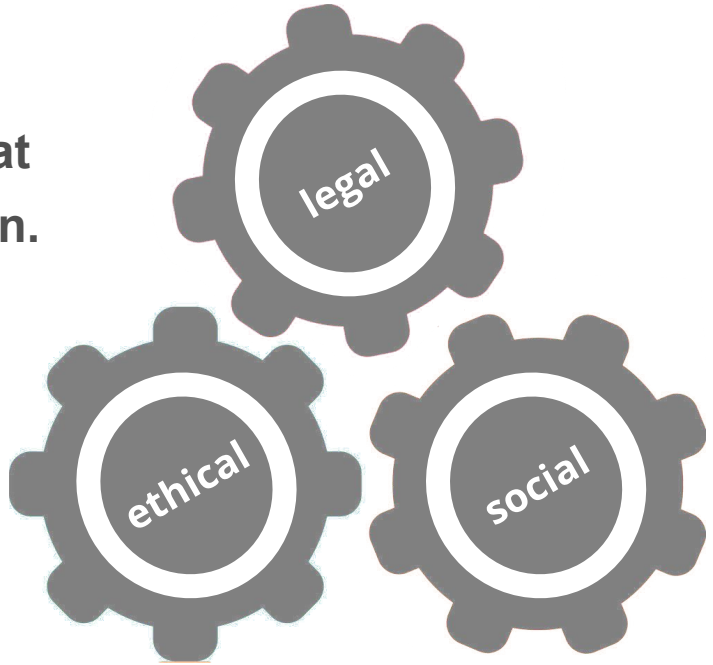
Ethical, Legal, & Social Requirements

What if people make “wrong” decisions?

Ethical, Legal, & Social Requirements

What if people make “wrong” decisions?

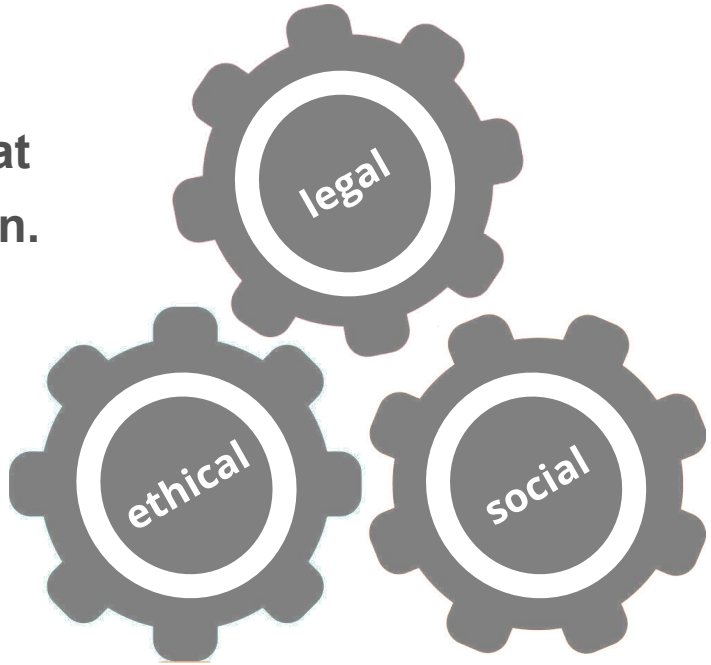
Answer. **Give people control, yet maintain that legal and ethical requirements are not broken.**



Ethical, Legal, & Social Requirements

What if people make “wrong” decisions?

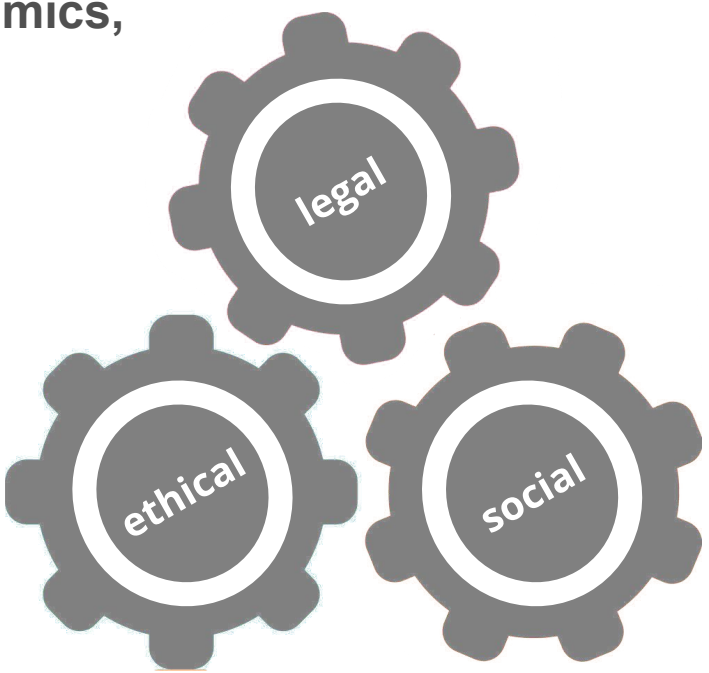
Answer. **Give people control, yet maintain that legal and ethical requirements are not broken.**



What is driving the evolution of our technologies is the evolution of our ***ethical***, ***legal***, and ***social*** requirements.

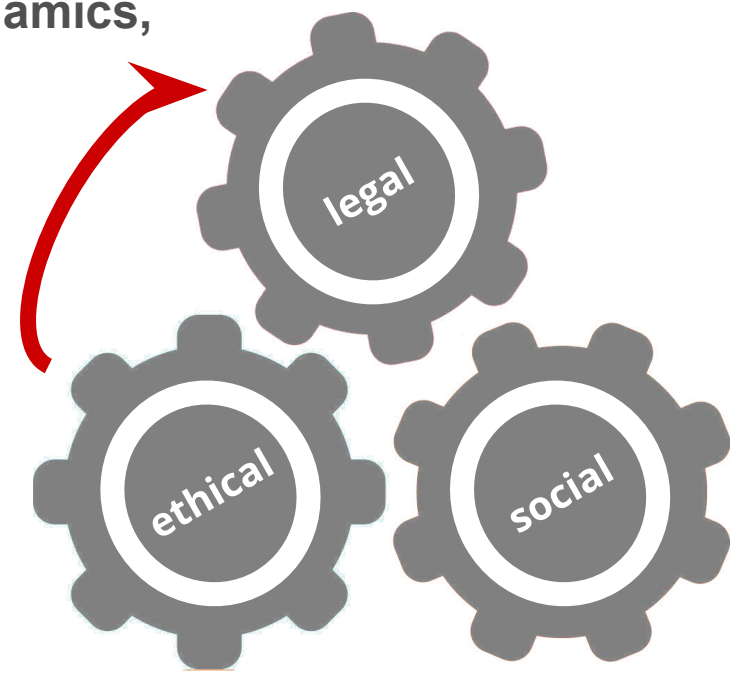
Research Challenges

1. Understand the ethical-legal-social dynamics, and how does one impact another.
2. Encode ethical and legal requirements into the system.



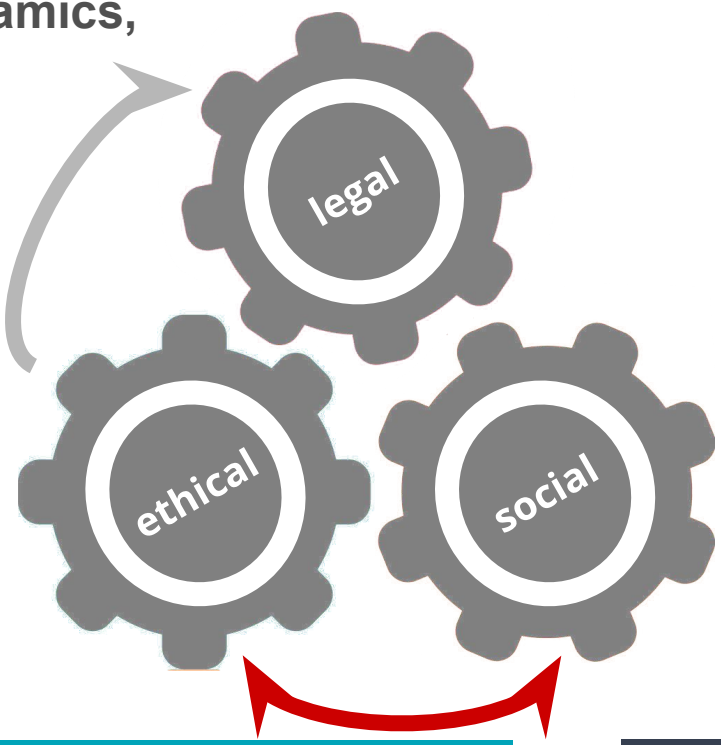
Research Challenges

1. **Understand the ethical-legal-social dynamics, and how does one impact another.**
   - From ethical to legal:
     - GDPR
     - Violating privacy when needed!



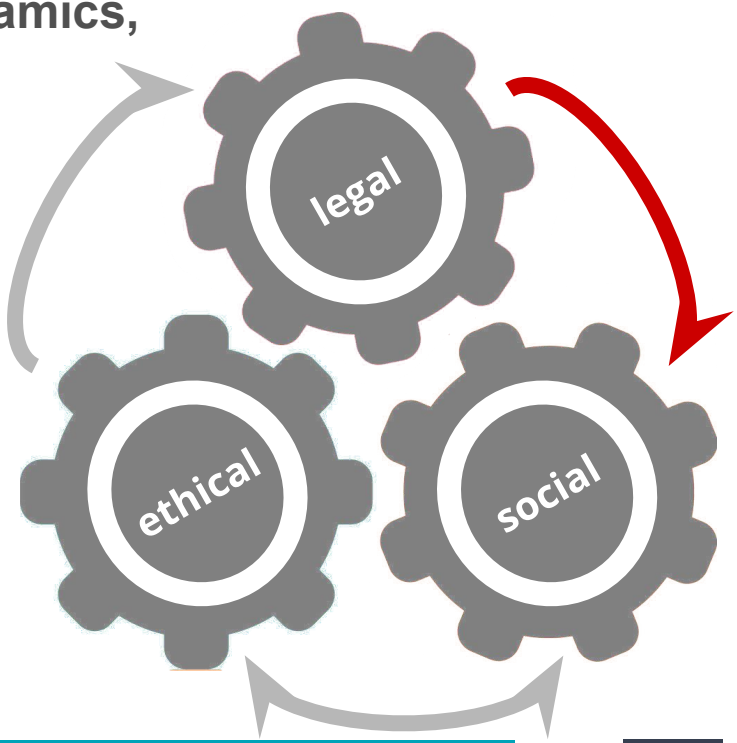
Research Challenges

1. **Understand the ethical-legal-social dynamics, and how does one impact another.**
   - From ethical to legal:
     - GDPR
     - Violating privacy when needed!
   - From social to ethical, and vice versa:
     - GDPR
     - Gender equality



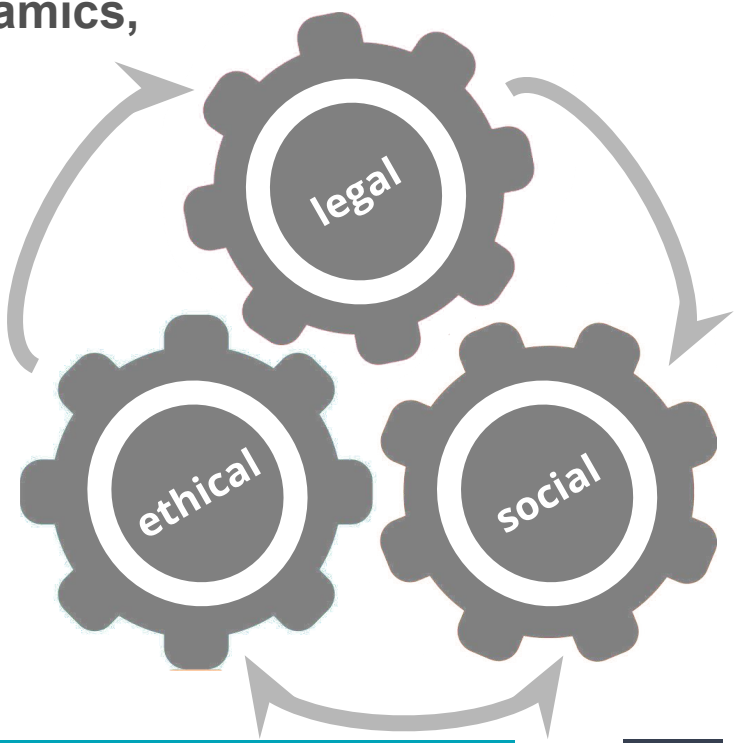
Research Challenges

1. **Understand the ethical-legal-social dynamics, and how does one impact another.**
   - From ethical to legal:
     - GDPR
     - Violating privacy when needed!
   - From social to ethical, and vice versa:
     - GDPR
     - Gender equality
   - From legal to social:
     - Speed limits
     - Smoking



Research Challenges

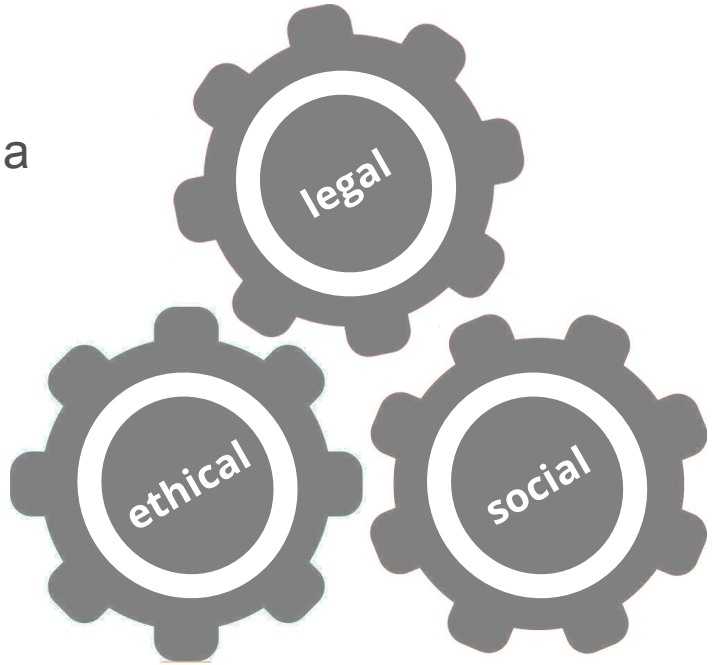
1. **Understand the ethical-legal-social dynamics, and how does one impact another.**
   - From ethical to legal:
     - GDPR
     - Violating privacy when needed!
   - From social to ethical, and vice versa:
     - GDPR
     - Gender equality
   - From legal to social:
     - Speed limits
     - Smoking



Research Challenges

2. **Encode ethical and legal requirements into the system.**

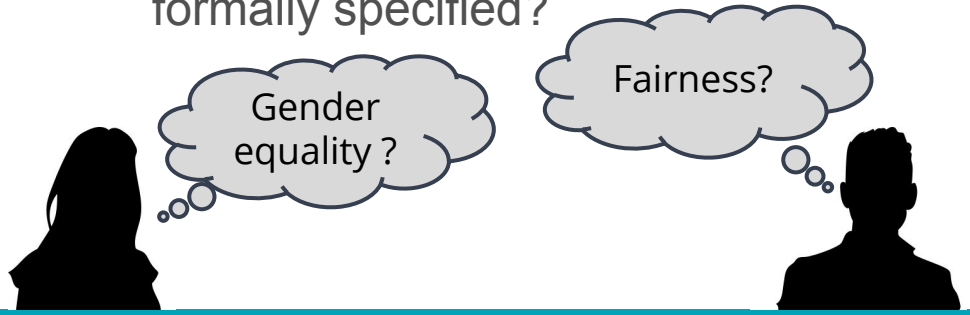
   - Legal requirements may be encoded in a way similar to social requirements.
   - Ethical requirements are more tricky.
     - Can all ethical requirements be formally specified?

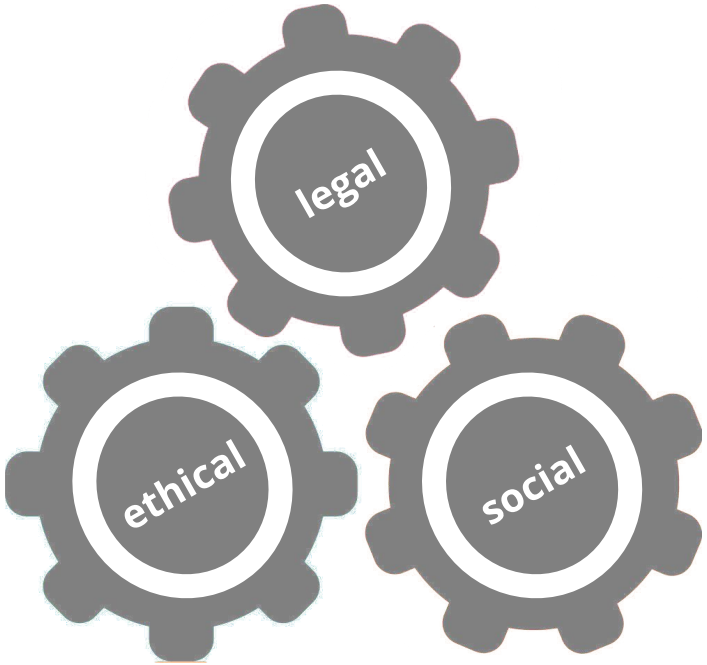


Research Challenges

2. **Encode ethical and legal requirements into the system.**

   - Legal requirements may be encoded in way similar to social requirements.
   - Ethical requirements are more tricky.
     - Can all ethical requirements be formally specified?

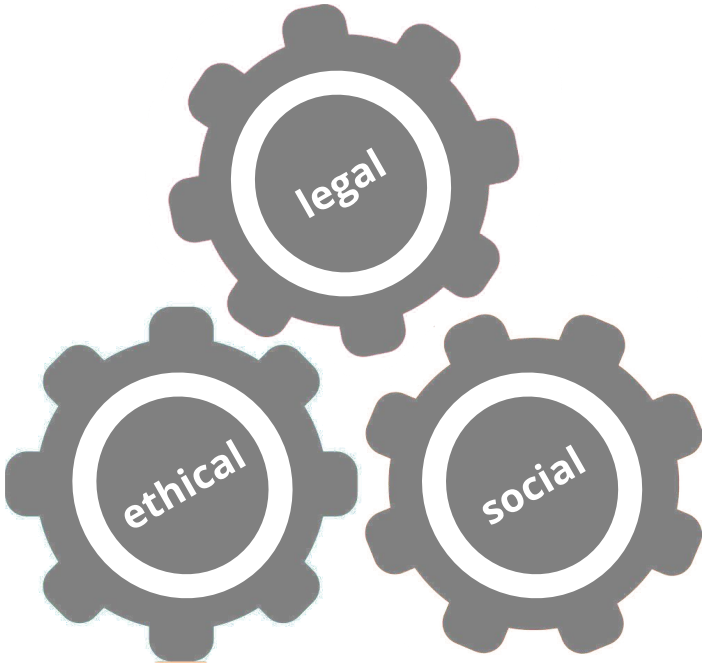




Research Challenges

2. **Encode ethical and legal requirements into the system.**

   - Legal requirements may be encoded in way similar to social requirements.
   - Ethical requirements are more tricky.
     - Can all ethical requirements be formally specified?
     - Differentiate between absolute moral principles and values.



WENET
INTERNET OF US

Take-Home Message

Take-Home Message

- Norms regulate behaviour
- Norms *should* adapt to our evolving requirements and values
- Adaptive norms require extensive research in different fields
- Ethical concerns are super interesting to address, yet tough work!
  - Understanding the value-norms relationship
  - Formalising the value alignment problem *(we've just started)*
  - Understanding the ethical-legal-social dynamics
  - Encoding ethical & legal requirements *(specification, implementation, verification)*

GET IN TOUCH

Website
www.internetofus.eu

Email
info@internetofus.eu

Twitter
@WeNetProject

WeNet project is funded by the EU's Horizon2020 programme under Grant Agreement number 823783.

WENET
INTERNET OF US

Many thanks to:

Carles Sierra, Pablo Noriega, Pilar Dellunde, Jordi Sabater-Mir, AppPhil & CIMBVAL members

GET IN TOUCH

Website
www.internetofus.eu

Email
info@internetofus.eu

Twitter
@WeNetProject

WeNet project is funded by the EU's Horizon2020 programme under Grant Agreement number 823783.

WENET
INTERNET OF US

Thank you!

Nardine Osman
nardine@iiia.csic.es