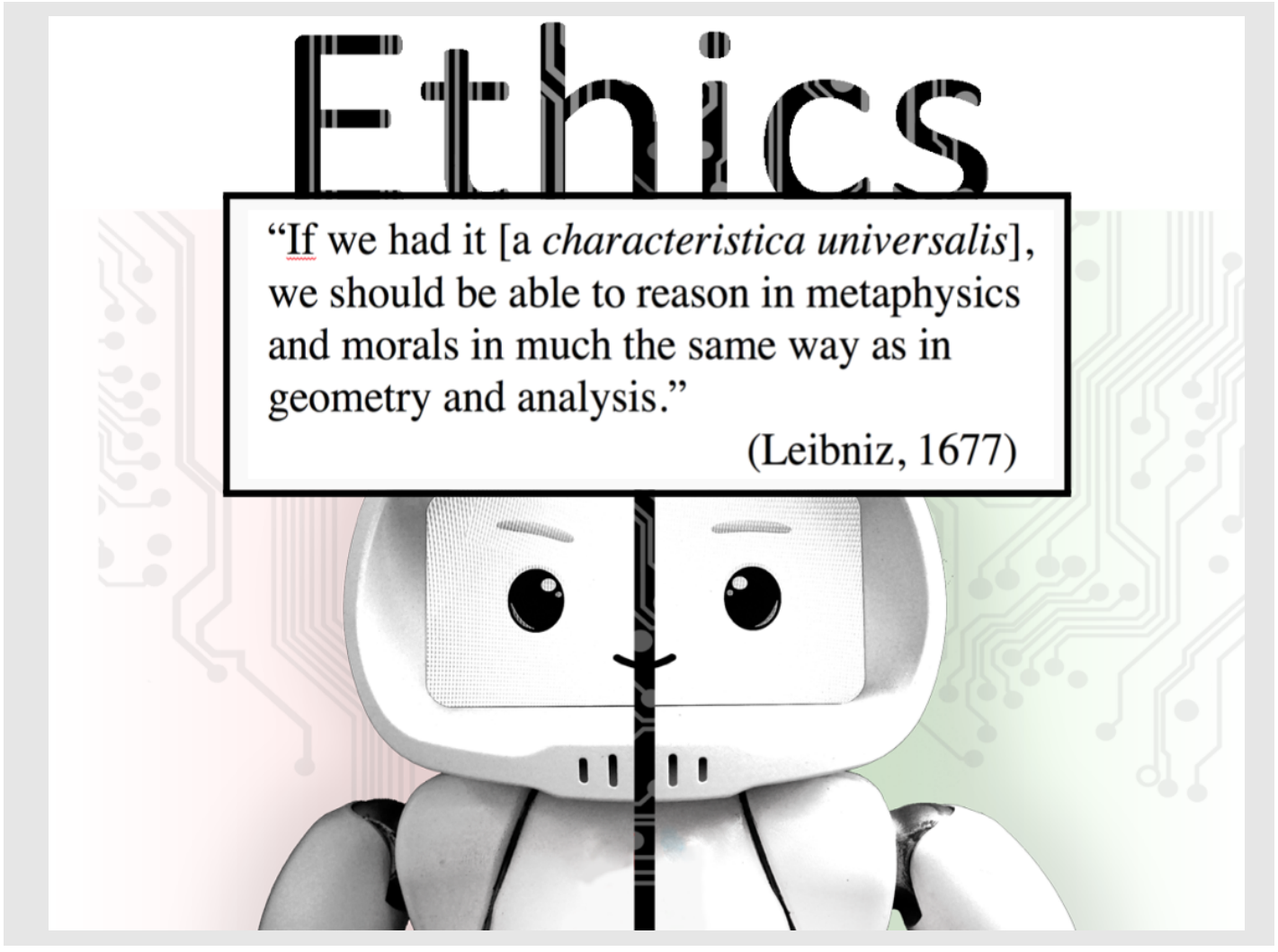
Artificial Intelligence—Position Statement: Ban on LAWS

Christoph Benzmüller
Freie Universität Berlin | University of Luxemburg

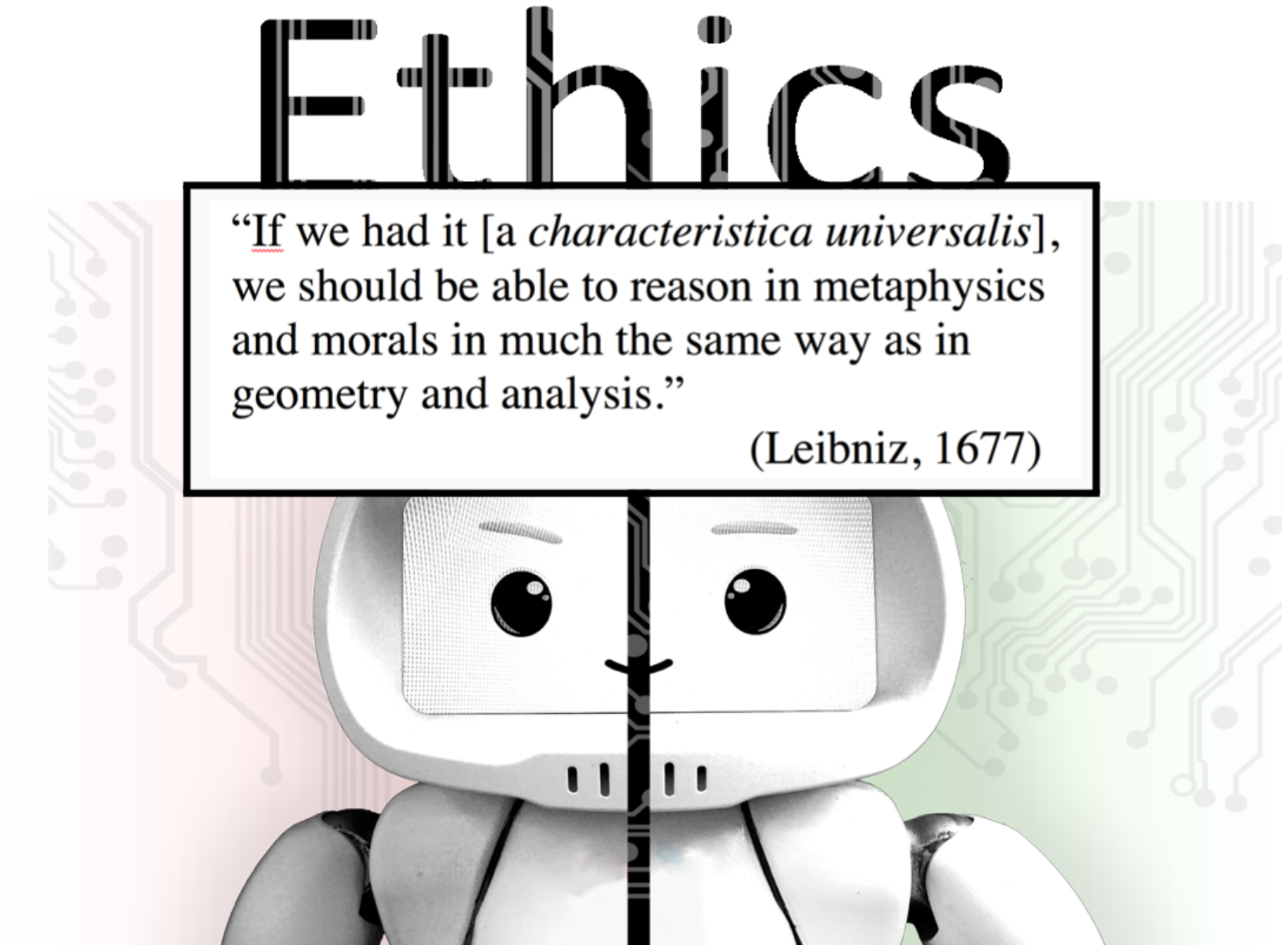


KI 2019

Why I support a Ban on LAWS

Who am I? What is my Research?

- Professor at FU Berlin & Visiting Scholar at U of Luxembourg
- Interest: AI, Philosophy, Mathematics, Computer Science, NL
- **Symbolic AI**: knowledge representation, automated reasoning
- Developed a **universal (meta-)logical reasoning** approach
- Leading automated higher-order theorem prover in the world
- Applications: Metaphysics, Mathematics, **Machine Ethics**


Ethics
"If we had it [a *characteristica universalis*], we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis."
(Leibniz, 1677)

Preservation of Life—A Moral Value of Machines?

Pilot über die Boeing 737 Max

SPIEGEL+

"Eine Automatisierung will nicht überleben. Wir schon"

Uwe Harter ist seit 26 Jahren Pilot von Passagierflugzeugen. Er steuert A320-Jets - das Pendant von Airbus zur Boeing 737. Ein Gespräch über Notfälle im Cockpit und die Schulung der Crew. *Von Claus Hecking* mehr...

737 Max: FBI schließt sich offenbar Ermittlungen wegen Zulassung an

Abstürze der Boeing 737 Max: Welche Rolle spielten die Piloten?

- Can intelligent systems have an own ethics? **—I doubt it!—**
- Can "our" ethical principles be reliably implemented in intelligent systems **—Eventually, but we are not there yet!—**
- Does a *"human in the loop"* help? **—Obviously not!—**

Preservation of Life—A Moral Value of Machines?

Pilot über die Boeing 737 Max

SPIEGEL+

"Eine Automatisierung will nicht überleben. Wir schon"

Uwe Harter ist seit 26 Jahren Pilot von Passagierflugzeugen. Er steuert A320-Jets - das Pendant von Airbus zur Boeing 737. Ein Gespräch über Notfälle im Cockpit und die Schulung der Crew. *Von Claus Hecking* mehr...

737 Max: FBI schließt sich offenbar Ermittlungen wegen Zulassung an

Abstürze der Boeing 737 Max: Welche Rolle spielten die Piloten?

- Can intelligent systems have an own ethics? —**I doubt it!**—
- Can "our" ethical principles be reliably implemented in intelligent systems —**Eventually, but we are not there yet!**—
- Does a *"human in the loop"* help? —**Obviously not!**—

Preservation of Life—A Moral Value of Machines?

Pilot über die Boeing 737 Max

SPIEGEL+

"Eine Automatisierung will nicht überleben. Wir schon"

Uwe Harter ist seit 26 Jahren Pilot von Passagierflugzeugen. Er steuert A320-Jets - das Pendant von Airbus zur Boeing 737. Ein Gespräch über Notfälle im Cockpit und die Schulung der Crew. *Von Claus Hecking* mehr...

737 Max: FBI schließt sich offenbar Ermittlungen wegen Zulassung an

Abstürze der Boeing 737 Max: Welche Rolle spielten die Piloten?

- Can intelligent systems have an own ethics? —**I doubt it!**—
- Can "our" ethical principles be reliably implemented in intelligent systems —**Eventually, but we are not there yet!**—
- Does a *"human in the loop"* help? —**Obviously not!**—

A: Why I support a ban on LAWS

LAWS is unethical

- human dignity vs.
  algorithmic termination decision
- European position:
  - human-centric AI
  - AI should have an ethical purpose
  - ensure reliability and robustness

A: Why I support a ban on LAWS

Ethically Intelligent Systems

- research is lacking behind
- education is lacking behind

LAWS is unethical

- human dignity vs.
  algorithmic termination decision
- European position:
  - human-centric AI
  - AI should have an ethical purpose
  - ensure reliability and robustness

A: Why I support a ban on LAWS

Ethically Intelligent Systems

- research is lacking behind
- education is lacking behind

LAWS is unethical

- human dignity vs.
  algorithmic termination decision
- European position:
  - human-centric AI
  - AI should have an ethical purpose
  - ensure reliability and robustness

LAWS: Risks&Costs $>$ Opportunities

LAWS: exemplary AI area where Risks&Costs by far outweigh the opportunities

A: Why I support a ban on LAWS

Ethically Intelligent Systems

- research is lacking behind
- education is lacking behind

LAWS is unethical

- human dignity vs. algorithmic termination decision
- European position:
  - human-centric AI
  - AI should have an ethical purpose
  - ensure reliability and robustness

LAWS: Risks&Costs > Opportunities

LAWS: exemplary AI area where Risks&Costs by far outweigh the opportunities

"Develop, Deploy, then Regulate"

- not suitable for LAWS
- also not suitable for several other AI application areas

A: Why I support a ban on LAWS

Ethically Intelligent Systems

- research is lacking behind
- education is lacking behind

LAWS is unethical

- human dignity vs. algorithmic termination decision
- European position:
  - human-centric AI
  - AI should have an ethical purpose
  - ensure reliability and robustness

LAWS: Risks&Costs > Opportunities

LAWS: exemplary AI area where Risks&Costs by far outweigh the opportunities

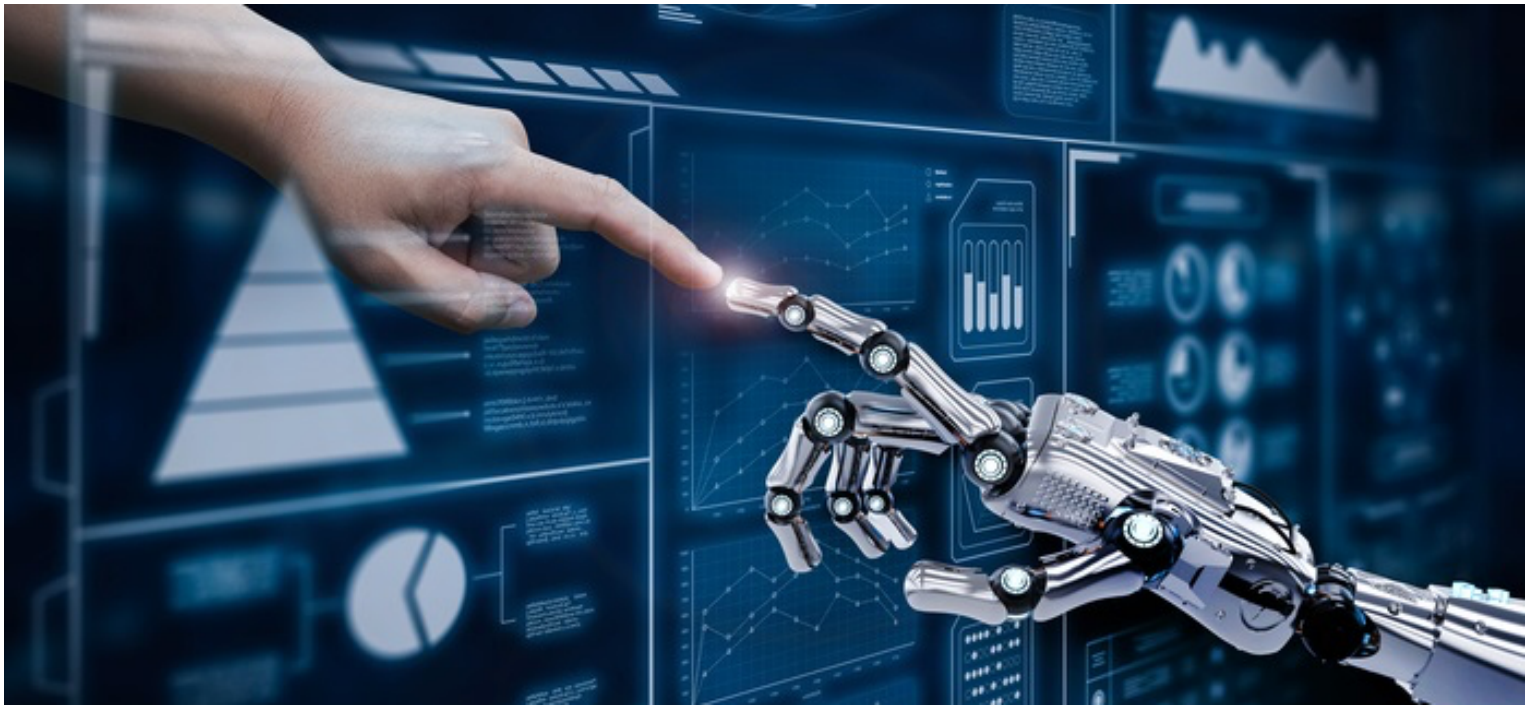
"Develop, Deploy, then Regulate"

- not suitable for LAWS
- also not suitable for several other AI application areas

Human in the Loop? Does not help much either, or does it?

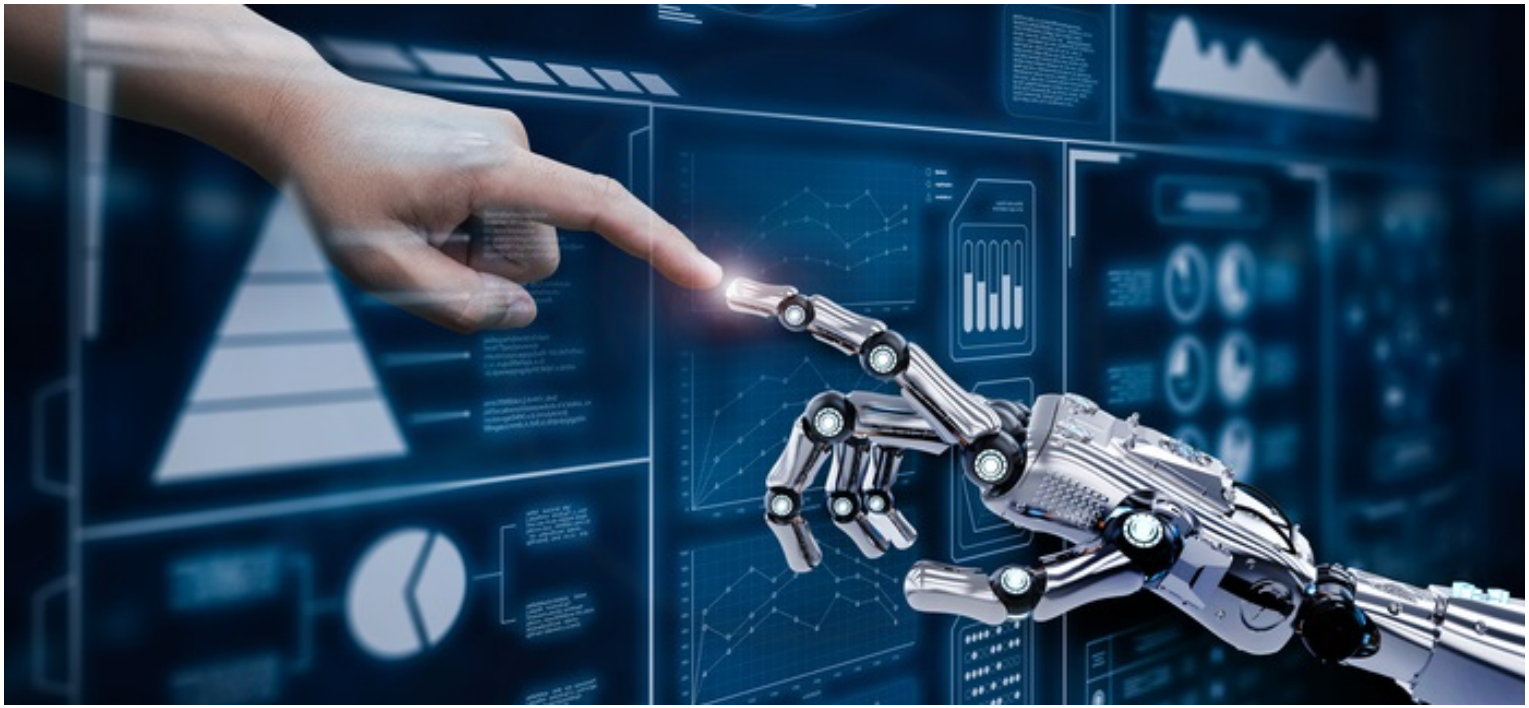
- assessment of highly complex situations
- in very short time and under extreme stress

LAWS: Inherently Unethical



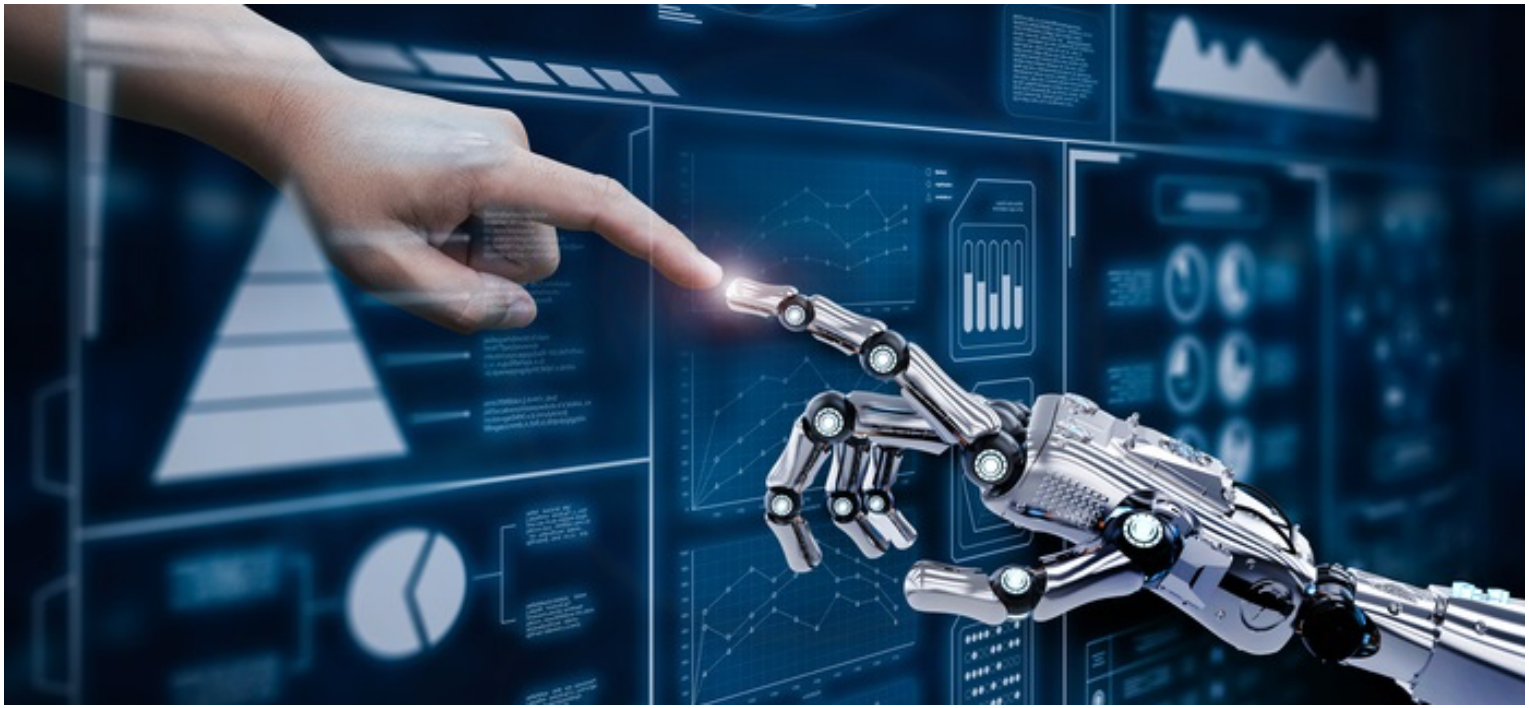
- Human Dignity versus Algorithmic Termination Decisions
- European Position: **Thrustworthy AI made in Europe**
  - AI must respect fundamental rights, applicable regulation and core principles and values, ensuring an "**ethical purpose**",
  - Technically robust and reliable implementation needed (even with good intentions, a lack of technological mastery can cause unintentional harm)

LAWS: Inherently Unethical



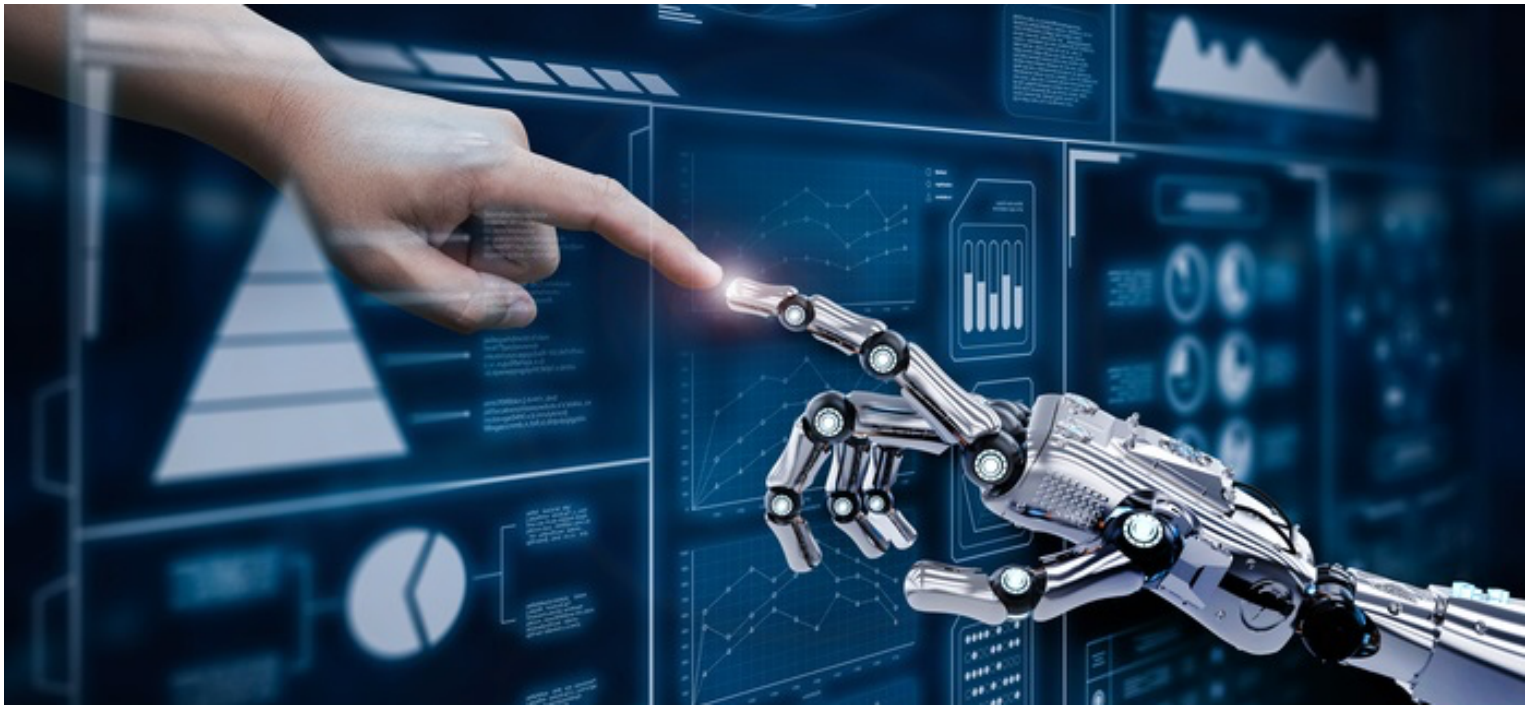
- Human Dignity versus Algorithmic Termination Decisions
- European Position: **Thrustworthy AI made in Europe**
  - AI must respect fundamental rights, applicable regulation and core principles and values, ensuring an **"ethical purpose"**,
  - Technically robust and reliable implementation needed (even with good intentions, a lack of technological mastery can cause unintentional harm)

LAWS: Inherently Unethical



- Human Dignity versus Algorithmic Termination Decisions
- European Position: **Thrustworthy AI made in Europe**
  - AI must respect fundamental rights, applicable regulation and core principles and values, ensuring an **"ethical purpose"**,
  - Technically robust and reliable implementation needed (even with good intentions, a lack of technological mastery can cause unintentional harm)

LAWS: Inherently Unethical



- Human Dignity versus Algorithmic Termination Decisions
- European Position: **Thrustworthy AI made in Europe**
  - AI must respect fundamental rights, applicable regulation and core principles and values, ensuring an **"ethical purpose"**,
  - Technically robust and reliable implementation needed (even with good intentions, a lack of technological mastery can cause unintentional harm)

Emerging Ethical frameworks for AI in the EU

EU Parliament (Sitting, 1/2019) Autonomous driving in European transport

- current regulatory framework will presumably not be sufficient
- ethical aspects need to be addressed and resolved by the legislator before these vehicles can be fully accepted
- automated vehicles need to undergo assessment of ethical aspects

Emerging Ethical frameworks for AI in the EU

EU Parliament (Sitting, 1/2019) Autonomous driving in European transport

- current regulatory framework will presumably not be sufficient
- ethical aspects need to be addressed and resolved by the legislator before these vehicles can be fully accepted
- automated vehicles need to undergo assessment of ethical aspects

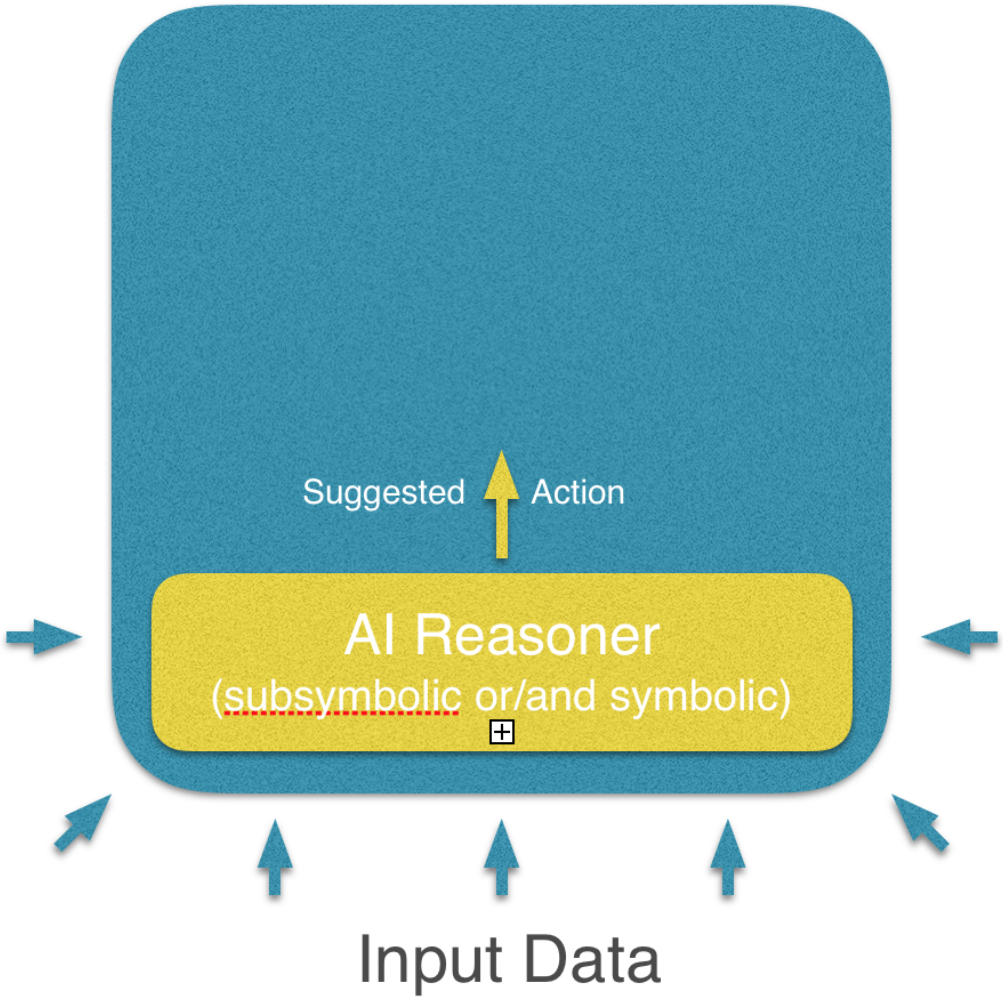
HLEG (Expert Group): Ethics Guidelines for Thrustworthy AI

- ensure an "ethical purpose"
- ensure technical robustness and reliability
- "fundamental ethical concerns" about LAWS: uncontrollable arms race, relinquished human control, risk of malfunction not adressed

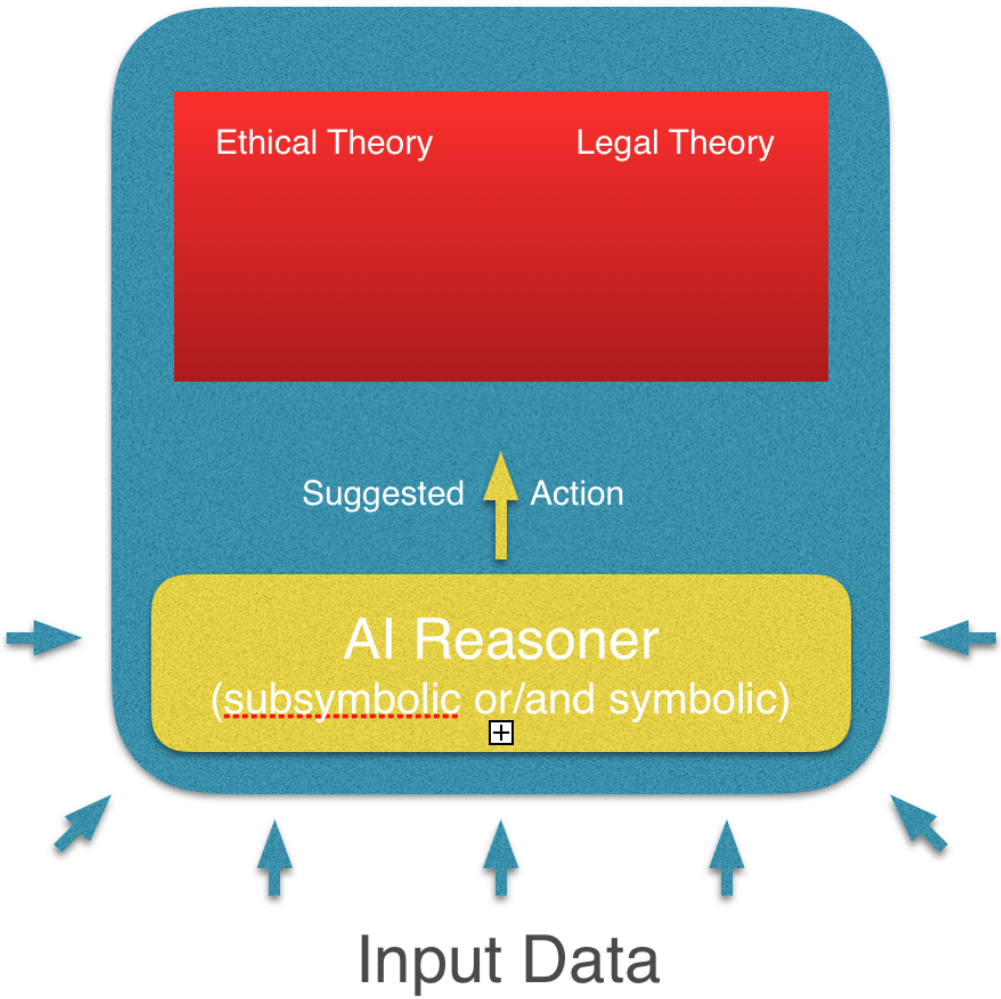
AI4People (Global Public Forum): Ethical Framework for "Good AI Society"

- define ethical framework
- recommendations how to implement such a framework

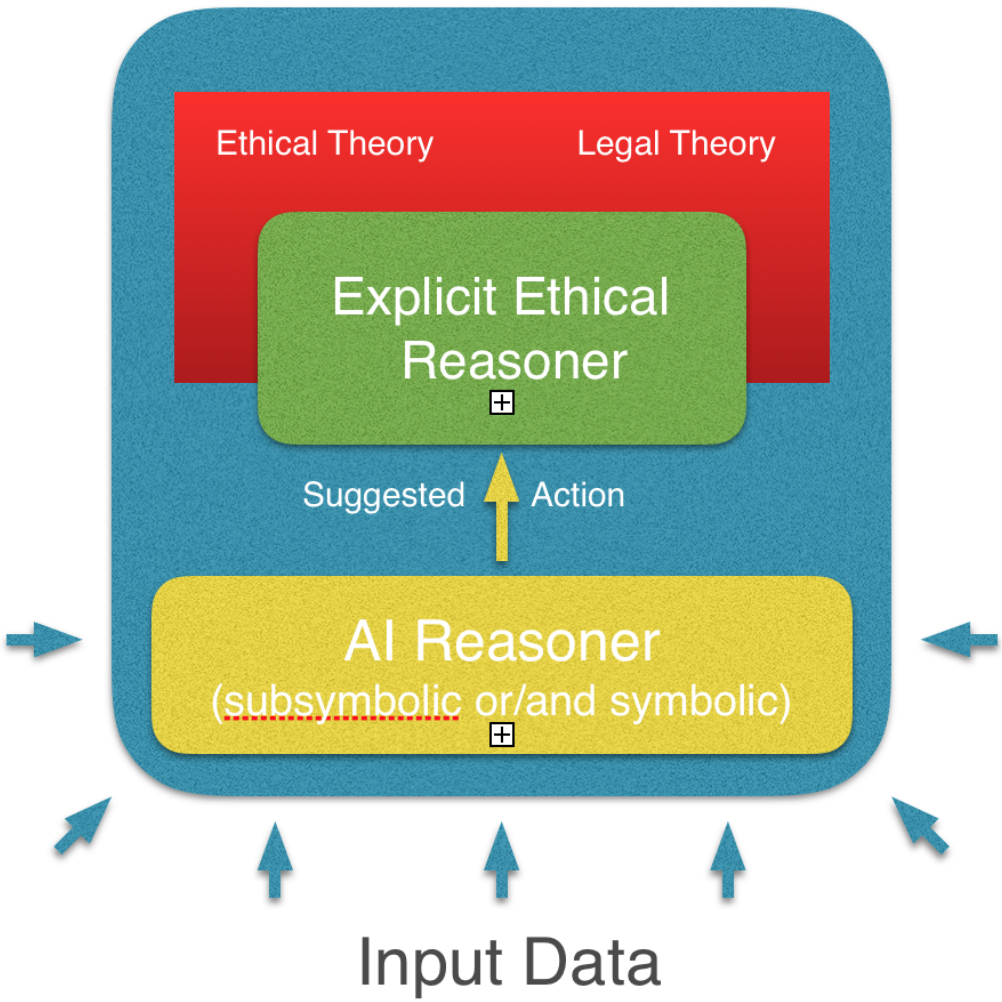
The Need for Independent Ethical Governors



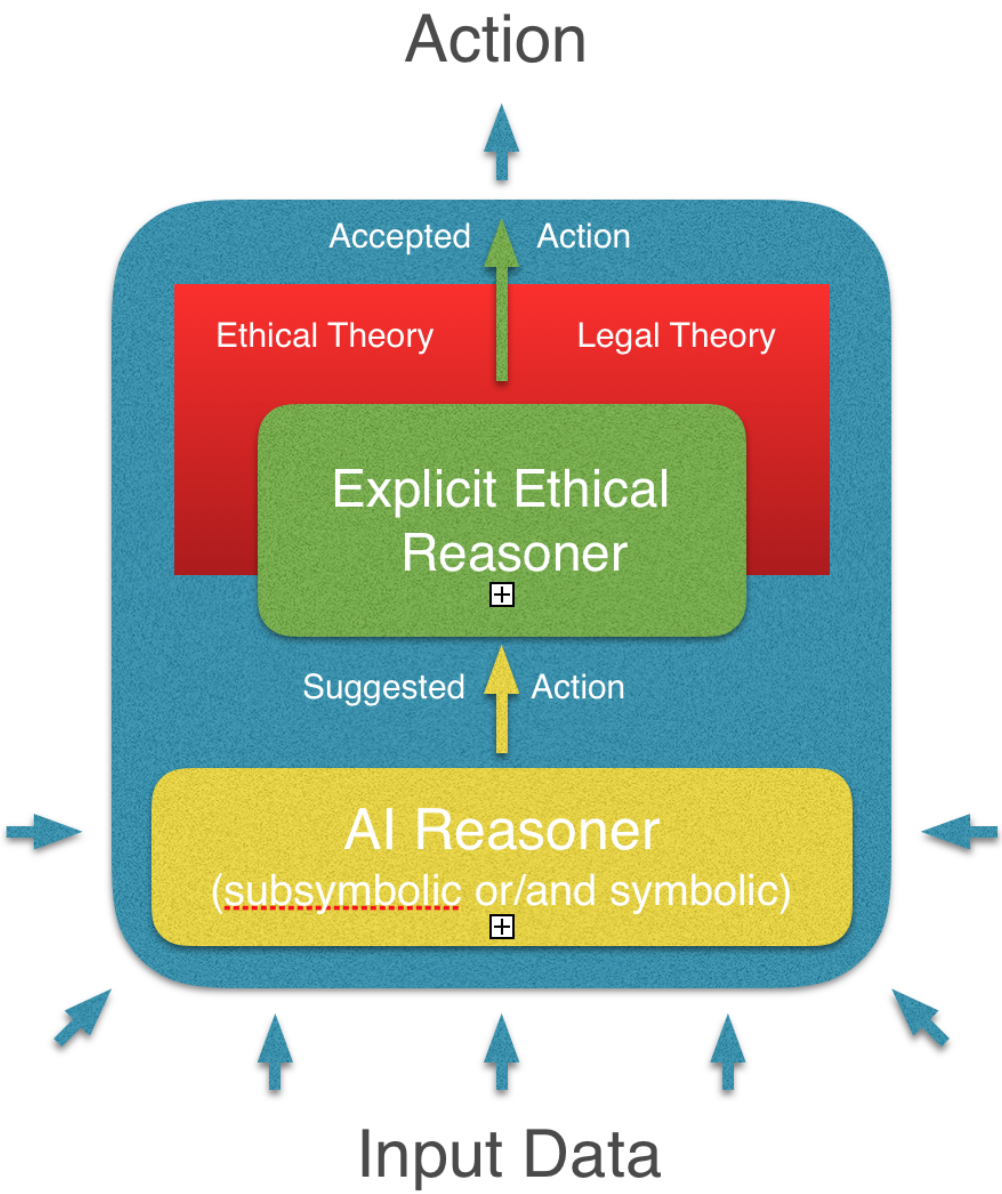
The Need for Independent Ethical Governors



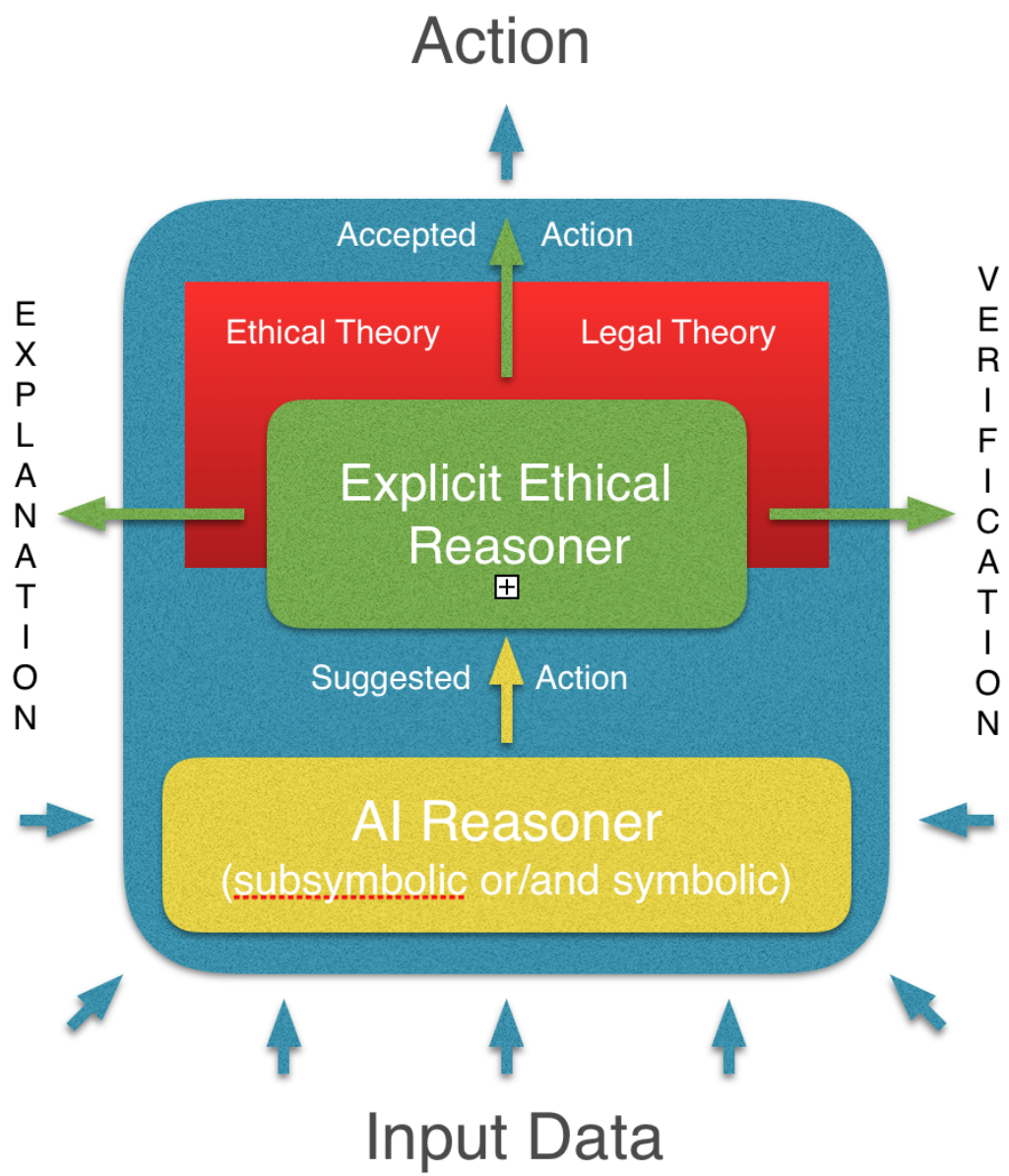
The Need for Independent Ethical Governors



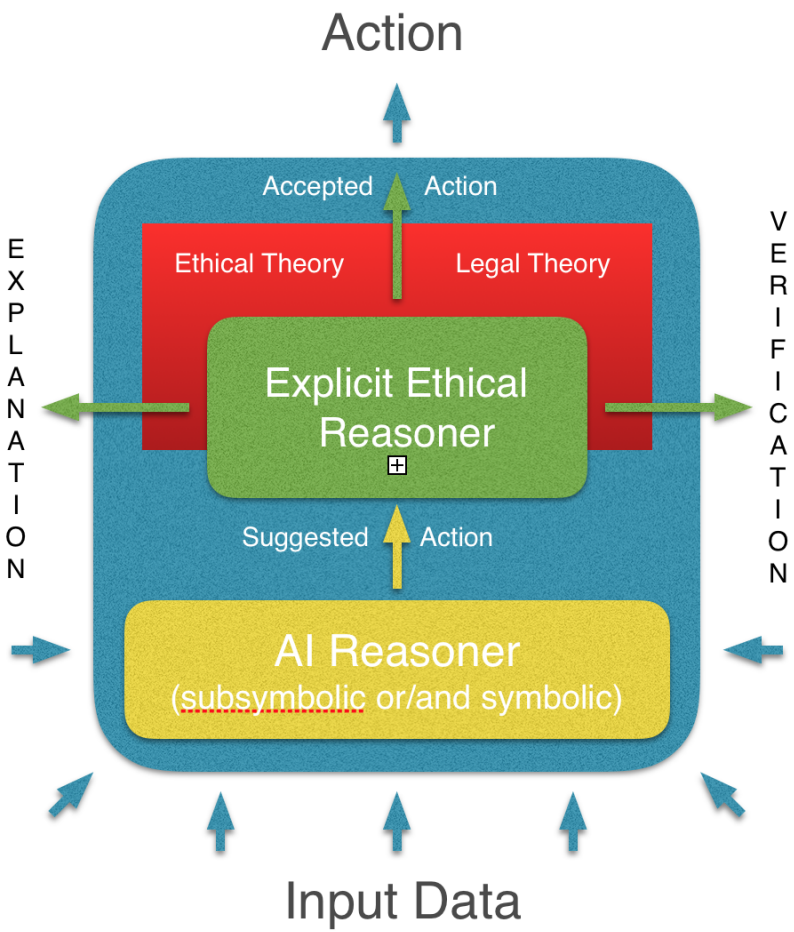
The Need for Independent Ethical Governors



The Need for Independent Ethical Governors



The Need for Independent Ethical Governors



Related Work

- Artificial Moral Agents
  - [Wallach&Allen, 2008]
- Ethical Governors
  - [ArkinEtAl., 2009, 2012]
  - [Dennis&Fisher, 2017]
- Ethical Deliberation in ART
  - [Dignum, 2017]
- Programming Machine Ethics
  - [Pereira&Saptawijaya, 2016]
- Calculi for Ethically Correct Robots
  - [Bringsjord *et al.*, 2018]
- ...

- Can intelligent systems have an own ethics? — **I doubt it!**
- Can "our" ethical principles be reliably implemented in intelligent systems — **Eventually, but we are not there yet!**

What arguments opposing LAWS are most likely to resonate with scientists?

Possible Arguments

Full verification of complex AI systems hardly feasible

Increasing deployment of machine learning technology

- complex and powerful, but also
- intransparent
- lack of explication
- vulnerable to malicious attacks
- ...

Reliable identification of compatants and non-compatants?

Assessment of complex situations in very short time error-prone

Human in the loop does not help much

Responsibility and liability issues may be blurred

Killer Robot technology will become widely available

What arguments opposing LAWS are most likely to resonate with scientists?

Possible Arguments

Full verification of complex AI systems hardly feasible

Increasing deployment of machine learning technology

- complex and powerful, but also
- intransparent
- lack of explication
- vulnerable to malicious attacks
- ...

Reliable identification of compatants and non-compatants?

Assessment of complex situations in very short time error-prone

Human in the loop does not help much

Responsibility and liability issues may be blurred

Killer Robot technology will become widely available

Confront scientists with the question:

Is it possible to develop provably correct operating LAWS?

How to get scientists involved?

How to get Scientists involved? Difficult, since . . .

Scientists always need to get . . .

this one next paper written (deadline tonight)

the lecture course for the next day prepared

this important funding proposal sent off (deadline next week)

these 9 conference papers reviewed (deadline in two weeks)

the invited presentation prepared for next week

the overdue project report written for the funding agency

the panel discussion prepared for tomorrow

the job application written (contract ends in 3 months)

the travel organised for the upcoming conferences

these 2 doctoral theses reviewed that will be defended soon

this big research project proposal reviewed for the funding agency

these 3 recommendation letters written

the lecture course exams from last week corrected

. . . ahh—I forgot—some scientist also have kids . . .

these research questions addressed that brought me into academia

How to get Scientists involved? Difficult, since ...

Scientists always need to get ...

this one next paper written (deadline tonight)

the lecture course for the next day prepared

this important funding proposal sent off (deadline next week)

these 9 conference papers reviewed (deadline in two weeks)

the invited presentation prepared for next week

the overdue project report written for the funding agency

the panel discussion prepared for tomorrow

the job application written (contract ends in 3 months)

the travel organised for the upcoming conferences

these 2 doctoral theses reviewed that will be defended soon

this big research project proposal reviewed for the funding agency

these 3 recommendation letters written

the lecture course exams from last week corrected

... ahh—I forgot—some scientist also have kids ...

these research questions addressed that brought me into academia

How to get Scientists involved? Difficult, since . . .

Scientists always need to get . . .

this one next paper written (deadline tonight)

the lecture course for the next day prepared

this important funding proposal sent off (deadline next week)

these 9 conference papers reviewed (deadline in two weeks)

the invited presentation prepared for next week

the overdue project report written for the funding agency

the panel discussion prepared for tomorrow

the job application written (contract ends in 3 months)

the travel organised for the upcoming conferences

these 2 doctoral theses reviewed that will be defended soon

this big research project proposal reviewed for the funding agency

these 3 recommendation letters written

the lecture course exams from last week corrected

. . . ahh—I forgot—some scientist also have kids . . .

these research questions addressed that brought me into academia

How to get Scientists involved? Difficult, since ...

Scientists always need to get ...

this one next paper written (deadline tonight)

the lecture course for the next day prepared

this important funding proposal sent off (deadline next week)

these 9 conference papers reviewed (deadline in two weeks)

the invited presentation prepared for next week

the overdue project report written for the funding agency

the panel discussion prepared for tomorrow

the job application written (contract ends in 3 months)

the travel organised for the upcoming conferences

these 2 doctoral theses reviewed that will be defended soon

this big research project proposal reviewed for the funding agency

these 3 recommendation letters written

the lecture course exams from last week corrected

... ahh—I forgot—some scientist also have kids ...

these research questions addressed that brought me into academia

How to get Scientists involved? Difficult, since ...

Scientists always need to get ...

this one next paper written (deadline tonight)

the lecture course for the next day prepared

this important funding proposal sent off (deadline next week)

these 9 conference papers reviewed (deadline in two weeks)

the invited presentation prepared for next week

the overdue project report written for the funding agency

the panel discussion prepared for tomorrow

the job application written (contract ends in 3 months)

the travel organised for the upcoming conferences

these 2 doctoral theses reviewed that will be defended soon

this big research project proposal reviewed for the funding agency

these 3 recommendation letters written

the lecture course exams from last week corrected

... ahh—I forgot—some scientist also have kids ...

these research questions addressed that brought me into academia

How to get Scientists involved? Difficult, since ...

Scientists always need to get ...

this one next paper written (deadline tonight)

the lecture course for the next day prepared

this important funding proposal sent off (deadline next week)

these 9 conference papers reviewed (deadline in two weeks)

the invited presentation prepared for next week

the overdue project report written for the funding agency

the panel discussion prepared for tomorrow

the job application written (contract ends in 3 months)

the travel organised for the upcoming conferences

these 2 doctoral theses reviewed that will be defended soon

this big research project proposal reviewed for the funding agency

these 3 recommendation letters written

the lecture course exams from last week corrected

... ahh—I forgot—some scientist also have kids ...

these research questions addressed that brought me into academia

How to get Scientists involved? Difficult, since ...

Scientists always need to get ...

this one next paper written (deadline tonight)

the lecture course for the next day prepared

this important funding proposal sent off (deadline next week)

these 9 conference papers reviewed (deadline in two weeks)

the invited presentation prepared for next week

the overdue project report written for the funding agency

the panel discussion prepared for tomorrow

the job application written (contract ends in 3 months)

the travel organised for the upcoming conferences

these 2 doctoral theses reviewed that will be defended soon

this big research project proposal reviewed for the funding agency

these 3 recommendation letters written

the lecture course exams from last week corrected

... ahh—I forgot—some scientist also have kids ...

these research questions addressed that brought me into academia

How to get Scientists involved? Difficult, since ...

Scientists always need to get ...

this one next paper written (deadline tonight)

the lecture course for the next day prepared

this important funding proposal sent off (deadline next week)

these 9 conference papers reviewed (deadline in two weeks)

the invited presentation prepared for next week

the overdue project report written for the funding agency

the panel discussion prepared for tomorrow

the job application written (contract ends in 3 months)

the travel organised for the upcoming conferences

these 2 doctoral theses reviewed that will be defended soon

this big research project proposal reviewed for the funding agency

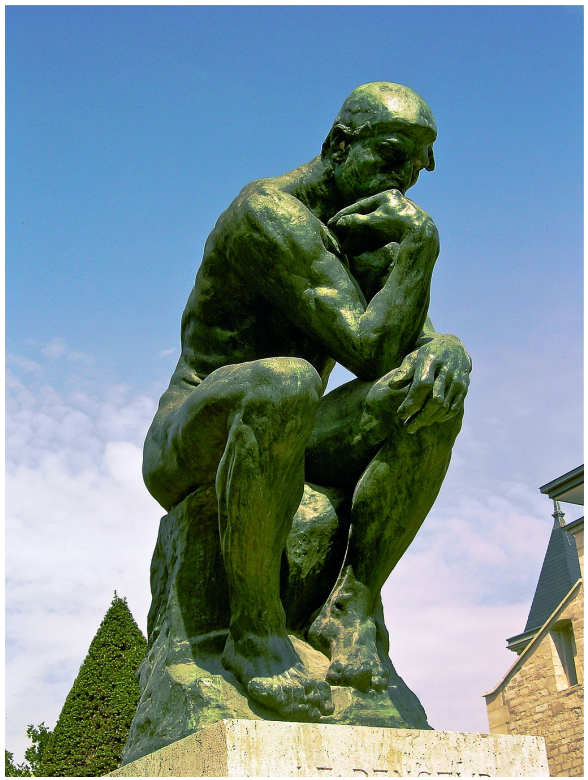
these 3 recommendation letters written

the lecture course exams from last week corrected

... ahh—I forgot—some scientist also have kids ...

these research questions addressed that brought me into academia

Theory vs. Reality



Theory

Reality

You want to get scientists actively involved? Liberate them from the rat race!

How to get scientists involved?

Open letter (IJCAI 2015)

Signed by 4502 AI/Robotics researchers and 26215 others

Too busy to get active themselves

Confront them: many will feel sorry being passive

Directly approach them

. . . like you did with me for today's meeting

. . . remind them about their signature

. . . request that they address (whenever suitable) the LAWS issue in their lecture courses, in public talks, in interviews, in their writings

. . . provide slides and teaching material to them for reuse in courses

What scientists may like

Activities that do not deplete valuable time and energy

. . . signing well written letters with clear objectives

. . . (informal) interaction with policy-makers

. . . give position statements in the media

How to get Scientists involved?

Some Further Ideas

"Hypocritical oath" for AI could eventually make sense
.3em]

"Not for military use" in AI code and in AI research papers?

Provide transparent information on:

- who is engaged in pro-LAWS project?
- who is funded by pro-LAWS institutions and industry?

Encourage anti-LAWS scientists to defend their position whenever possible

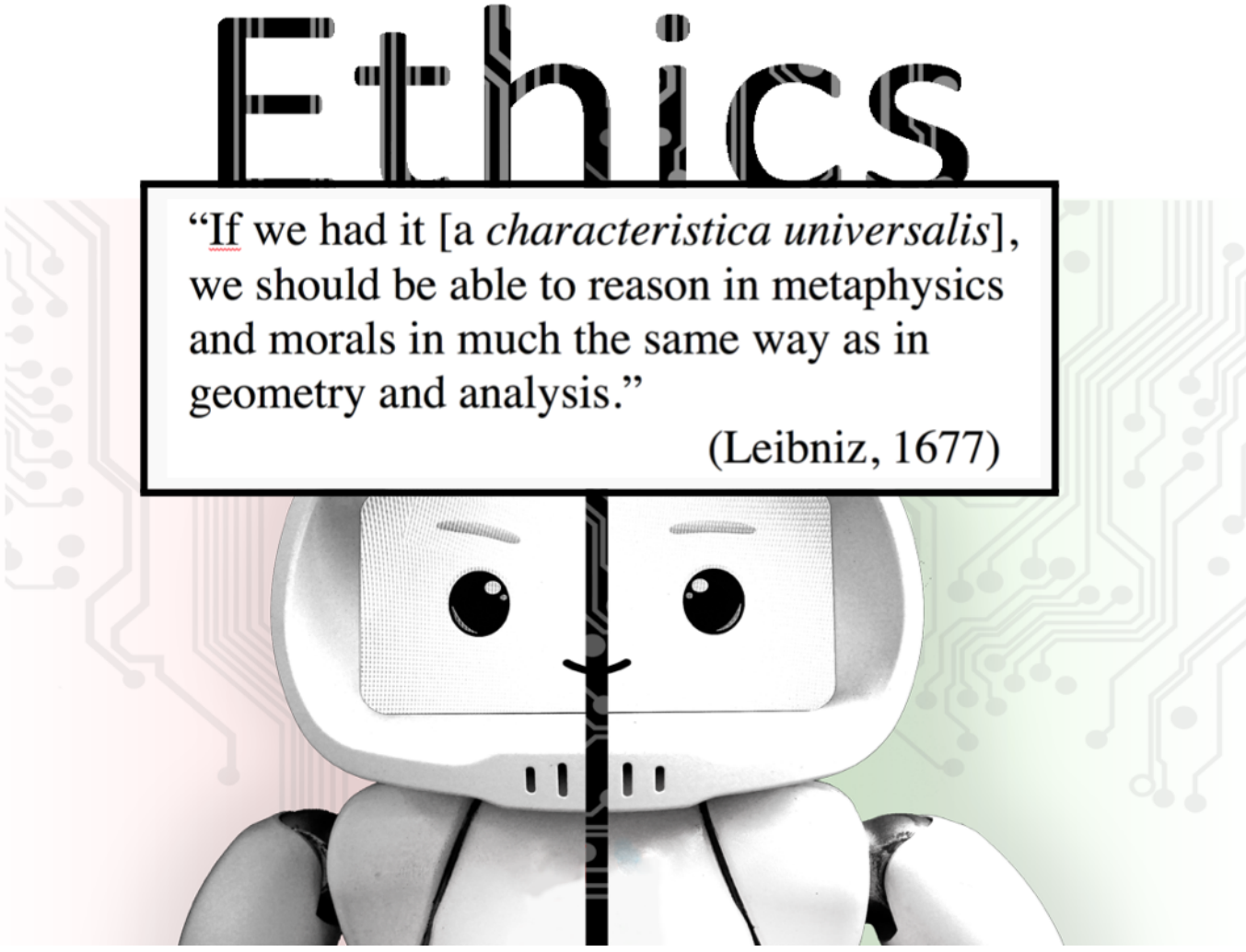
Foster anti-LAWS Research

Think about incentives:

- Funding of research projects
- Research prices
- Publication opportunities
- Membership in committees

Support and Encourage AI Scientists to become Policy-Makers

Foster Projects on Machine Ethics



- Can intelligent systems have an own ethics? **—I doubt it!—**
- Can "our" ethical principles be reliably implemented in intelligent systems **—Eventually, but we are not there yet!—**

Is LAWS compliant with European Position on Autonomous Cars?

EU Parliament—Resolution of 15. Jan 2019—Autonomous driving in European transport (2018/2089(INI))

The European Parliament,

20. Notes that the **existing liability rules**, such as . . . , **were not developed to deal with the challenges posed by the use of autonomous vehicles** and stresses that there is growing evidence that the current regulatory framework, especially as regards liability, insurance, registration and protection of personal data, **will no longer be sufficient** or adequate when faced with the new risks emerging from increasing vehicle automation, connectivity and complexity;

21. . . . calls, therefore, on the Commission to . . . introduce, if necessary, **new rules on the basis of which responsibility and liability are allocated**; calls also on the Commission to assess and monitor the possibility of introducing additional EU instruments to keep pace with developments in AI;

35. Calls on the Commission to lay down clear ethical guidelines for AI;

37. Stresses that **ethical aspects of self-driving vehicles need to be addressed and resolved by the legislator before these vehicles can be fully accepted** and made available in traffic situations; emphasises, therefore, that automated vehicles need to **undergo a prior assessment** to address these ethical aspects;

Personal experience with application for doctoral school

- Ethically Intelligent Systems

—risks&costs—

versus

—opportunities—

- Data-driven computational modelling and applications

Personal experience with application for doctoral school

- Ethically Intelligent Systems

—risks&costs—

versus

—opportunities—

- Data-driven computational modelling and applications

What do you think: Which doctoral program got funded?

Intelligence without Moral Values and Norms?

Intelligence is a Context Dependent Notion

Relative use: e.g., one animal is more intelligent than another one
More absolute use: comparison against mental capabilities of humans

My Def.: Artificial Intelligence

Science of computational technologies being developed to achieve and explain *intelligent* behaviour in machines.

My Def.: Intelligence

A collection of mental capabilities that enable an entity

1. to solve (or learn to solve) hard problems, —solve problems—
2. to successfully act in known, unknown and dynamic environments (requires perception, planning, agency, etc.), —master the unknown—
3. to reason abstractly and rationally, avoiding inconsistencies and self-contradiction, —be rational&abstract—
4. to reflect upon itself and to adjust its own reasoning with upper goals and norms, and —be self-reflective—
5. to interact socially with other entities and to align own values and norms with those of a society for a greater good. —be social—