On mapping values in AI governance

2 December 2021, Algorithmic Law and Society symposium, HEC Paris, Paris

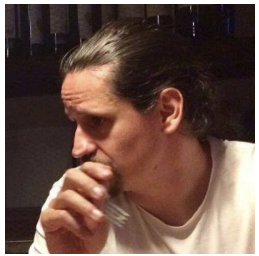
Geoff Gordon, Asser Institute, the Hague
Bernhard Rieder, Media Studies, University of Amsterdam
Giovanni Sileno, Informatics Institute, University of Amsterdam

context of this work

- RPA Human(e) AI seed grant by UvA

project: **“Mapping Value(s) in AI”**

Geoff Gordon
Faculty of Law



Bernhard Rieder
Faculty of Humanities

Giovanni Sileno
Faculty of Science

focus: algorithmic decision systems

- algorithmic decisions are increasingly used in all types of human-related activities:
  - predictive systems, recommender systems, decision support-systems…

focus: algorithmic decision systems

- algorithmic decisions are increasingly used in all types of human-related activities:
  - predictive systems, recommender systems, decision support-systems…
- these are both **objects** and **instruments** of regulatory governance

general research question

- not what values AI should satisfy, but

  how values manifest in context-sensitive computational and social processes?

general research question

- not what values AI should satisfy, but

  how values manifest in context-sensitive computational and social processes?

- this paper focuses in particular on setting the theoretical groundwork underpinning and motivating a “mapping” methodology for AI governance

key points of our contribution

1. *assemblage* as method to look at techno-regulation and regulation of technology
2. material stance on law
3. connection with critical practice of AI

1. assemblage

partial performances

- what an AI system produces cannot be defined nor observed upon the AI system alone
- the performativity of law is not defined by the sources of law alone

partial performances

- what an AI system produces cannot be defined nor observed upon the AI system alone
- the performativity of law is not defined by the sources of law alone

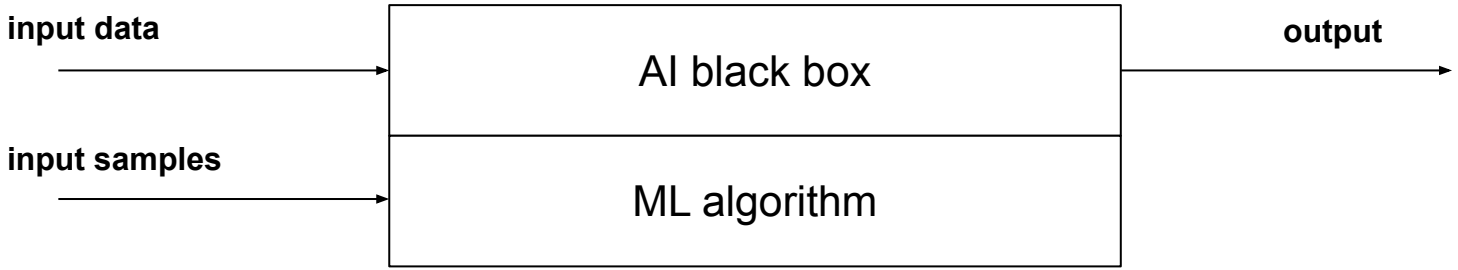
but where do people focus
their analysis the most?

partial performances

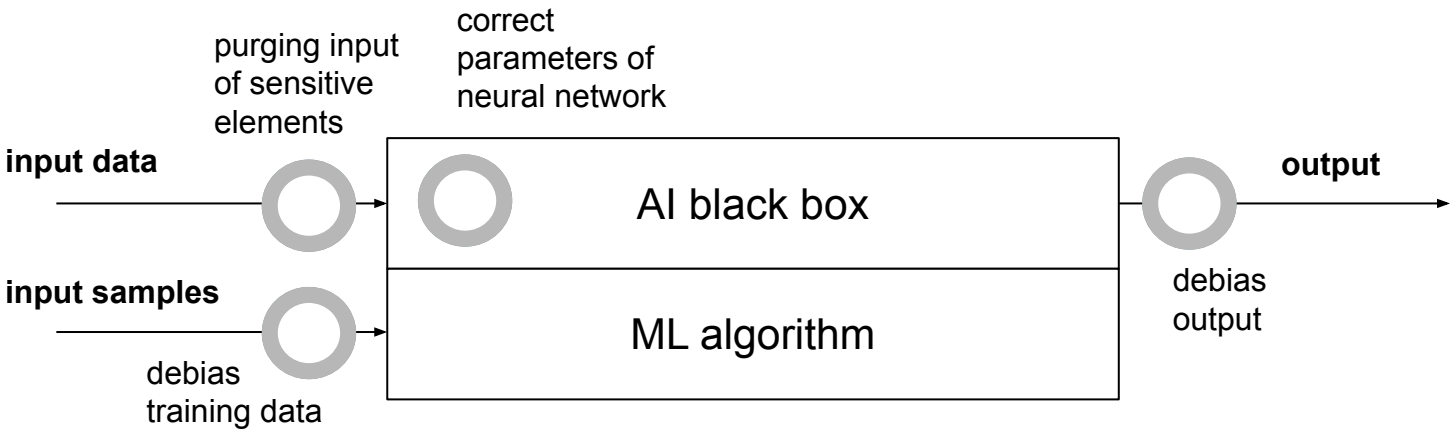
- what an AI system produces cannot be defined nor observed upon the **AI system** alone
- the performativity of law is not defined by the **sources of law** alone

but where do people focus their analysis the most?

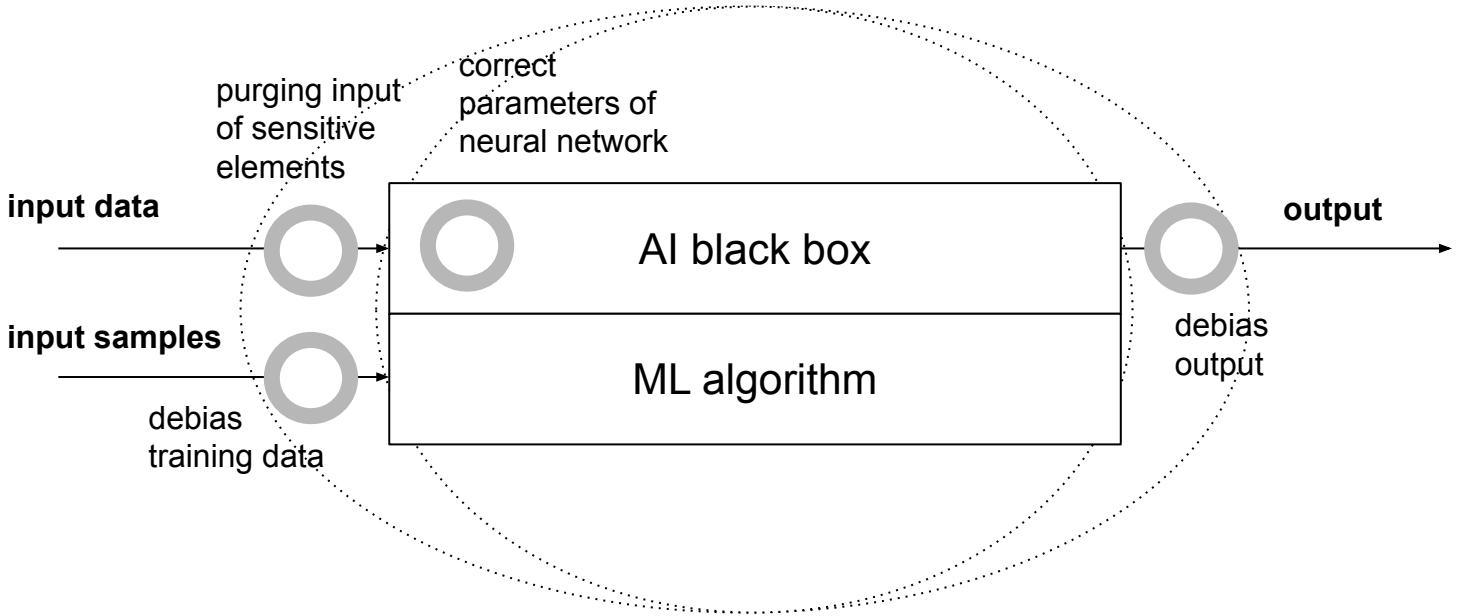
example: algorithmic fairness



example: algorithmic fairness

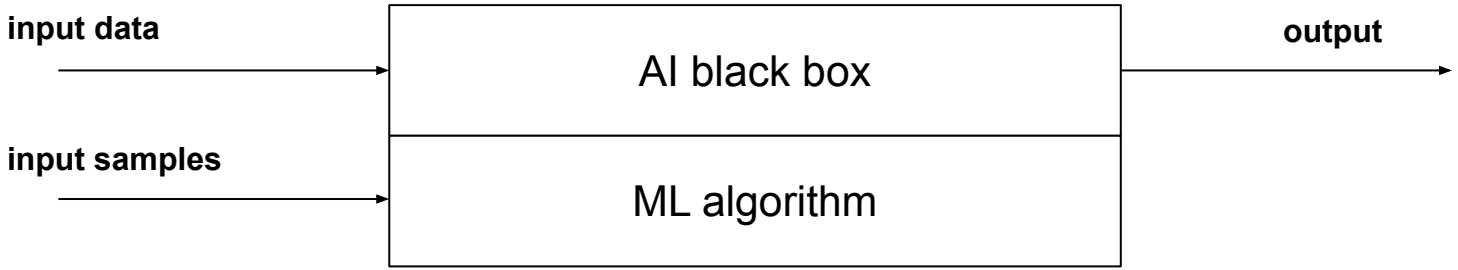


example: algorithmic fairness

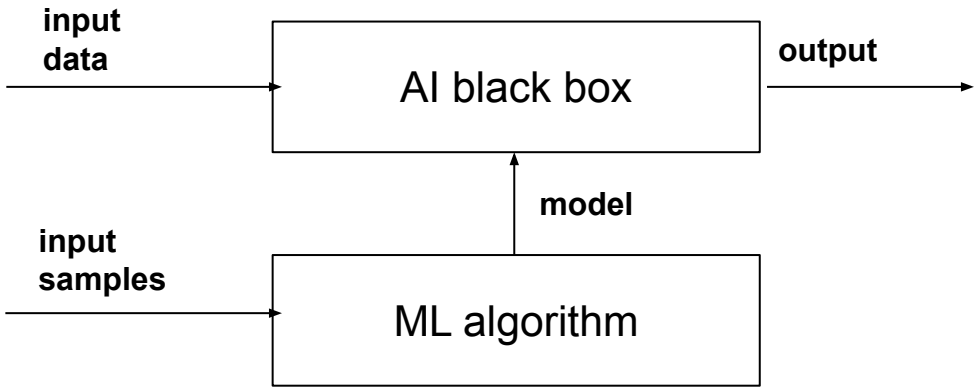


- all these methods focus on **data**
- just within or nearby the *computational* ***system boundaries***

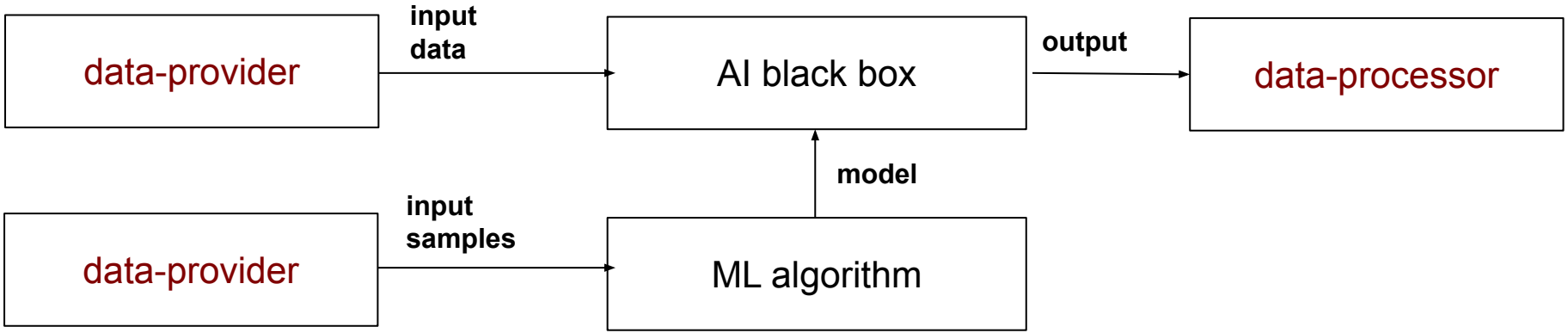
looking at the bigger picture..



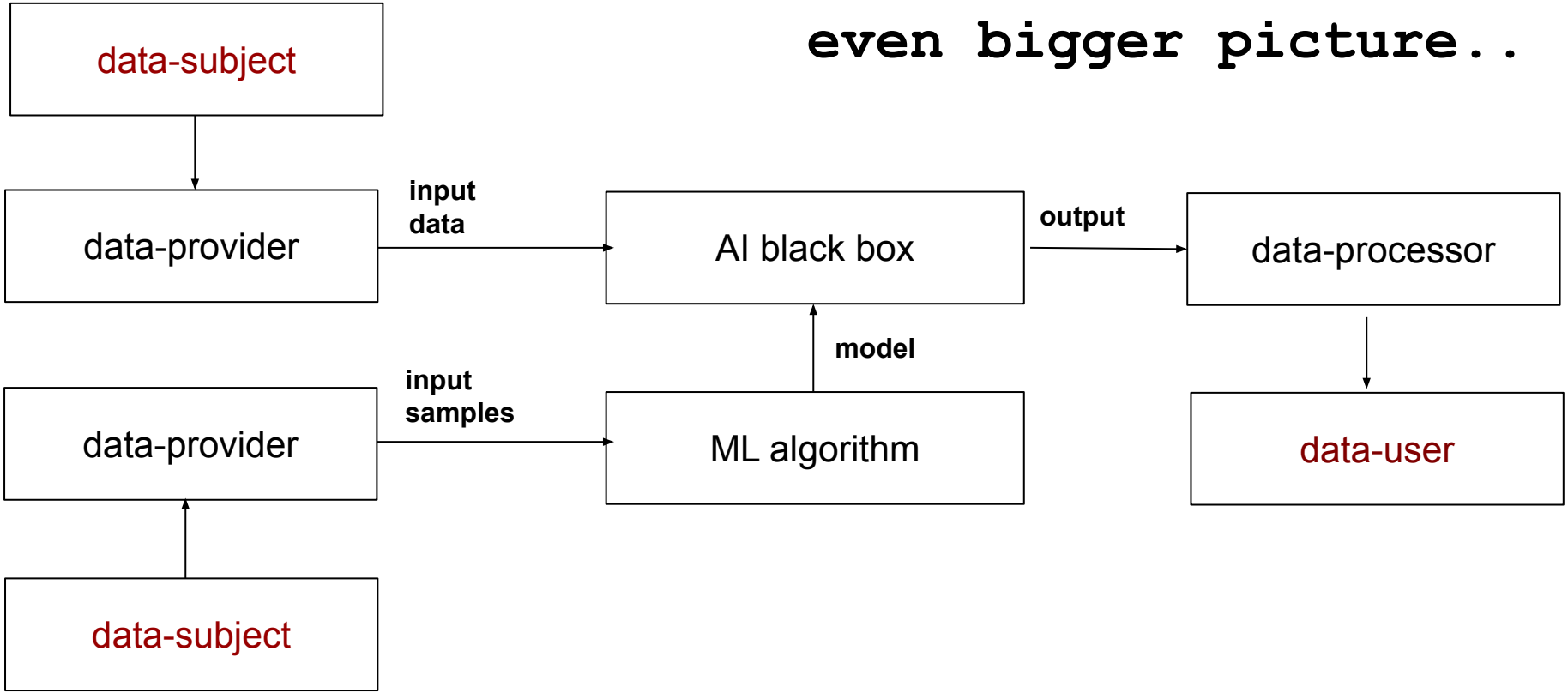
looking at the bigger picture..



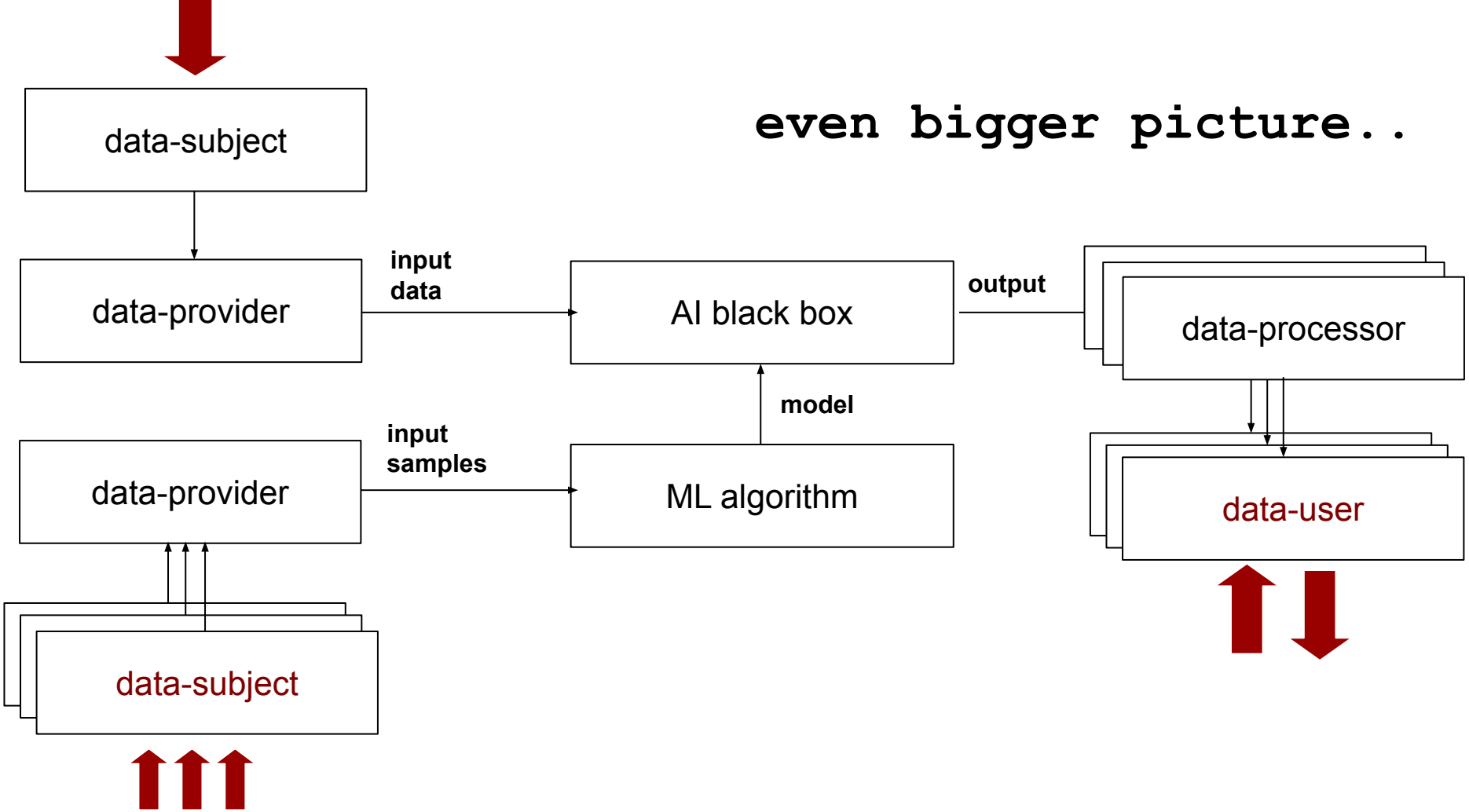
looking at the bigger picture..



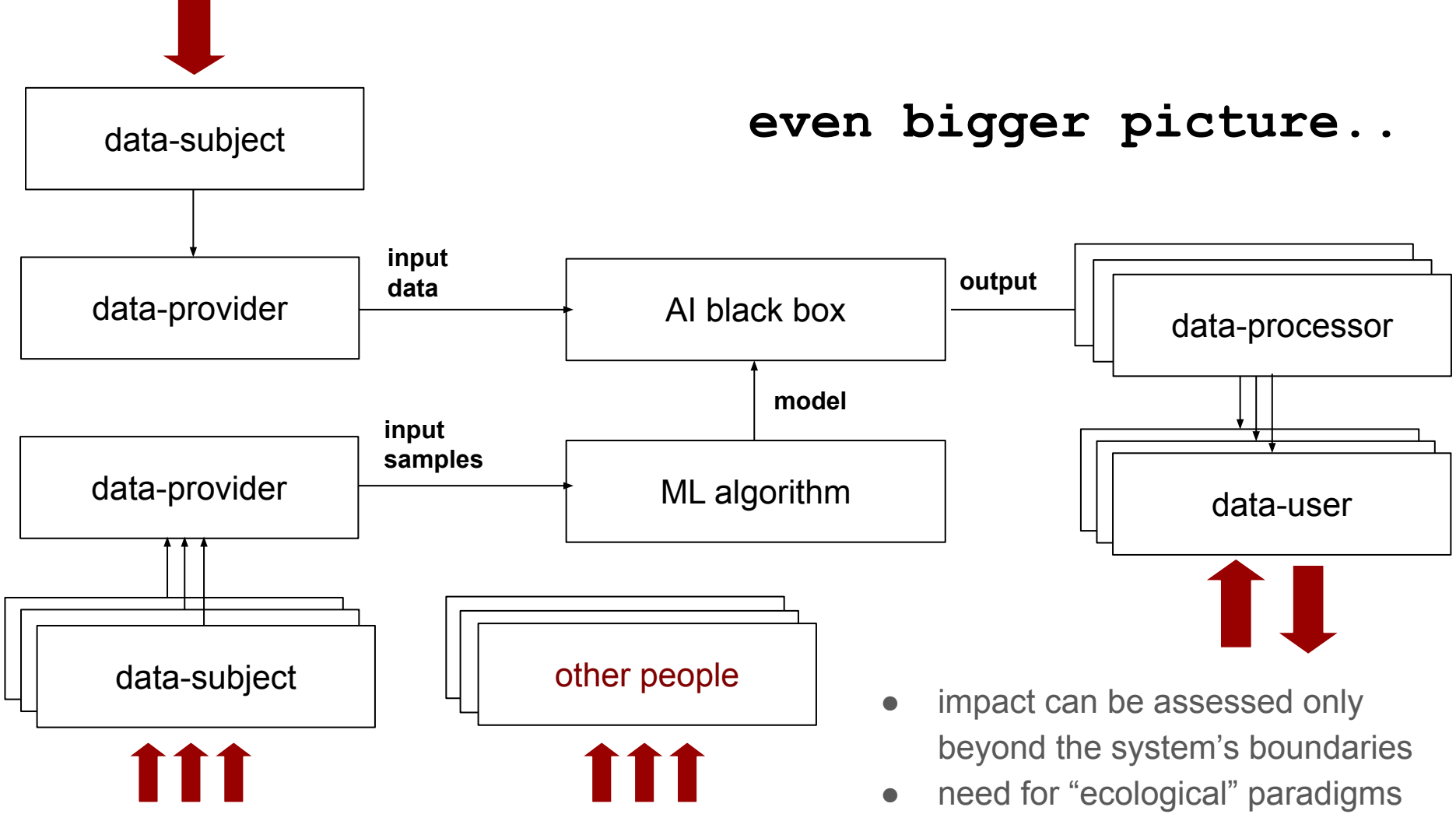
even bigger picture..



even bigger picture..



even bigger picture..



- impact can be assessed only beyond the system's boundaries
- need for "ecological" paradigms

in doing so, we changed the framing

from totality

eg. monolithic computational module, individual

- components defined by ***relations of interiority***

to assemblage

eg. socio-technical distributed system, social context

- components defined by ***relations of exteriority***

socio-technical assemblage

- law is not defined outside the assemblage, but within it
- similarly, AI is not defined outside the assemblage

➡ law and AI reside within the same assemblage

socio-technical assemblage: concept usage

consider a lawyer trying to attack a recommendation produced by AI (e.g. a parole decision, a credit rating, inclusion on a black list, stop at borders, ...).

possible options:

- focusing on the AI output alone

socio-technical assemblage: concept usage

consider a lawyer trying to attack a recommendation produced by AI (e.g. a parole decision, a credit rating, inclusion on a black list, stop at borders, ...).

possible options:

- focusing on the AI output alone
- focusing on the code and training of the AI

socio-technical assemblage: concept usage

consider a lawyer trying to attack a recommendation produced by AI (e.g. a parole decision, a credit rating, inclusion on a black list, stop at borders, ...).

possible options:

- focusing on the AI output alone
- focusing on the code and training of the AI
- **tracing the series of interactions** in which *technical operations* and *normative judgments* are **translated** *back and forth* and from point to point

2. materiality of law

against bifurcation

- in the **code-is-law** tradition, law is materialized in code

- code must be two things at once:
  - a functioning code and,
  - a representative of law.

against bifurcation

- in the **code-is-law** tradition, law is materialized in code

- code must be two things at once:
  - a functioning code and,
  - a representative of law.

this bifurcation opens space for **misrecognition**:

- the policy-maker may ‘not get’ the code;
- the code may ‘get the law wrong’.

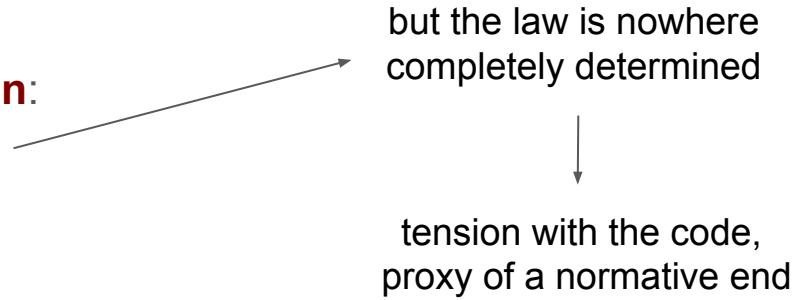
against bifurcation

- in the **code-is-law** tradition, law is materialized in code

- code must be two things at once:
  - a functioning code and,
  - a representative of law.

this bifurcation opens space for **misrecognition**:

- the policy-maker may ‘not get’ the code;
- the code may ‘get the law wrong’.



example: the SyRI case

- SyRI was developed by the Dutch Ministry of Social Affairs and Employment since 2014, and designed for end use by a variety of national agencies (e.g., the tax authority, the authority responsible for employment benefits, etc.) and municipalities.

  typical application: producing risk warnings signaling potential frauds in individual applications for social services.

SyRI in court (2020)

- the SyRI technology was recently found contrary to Article 8 of the European Convention of Human Rights, which broadly protects the right to respect for private and family life, home and correspondence.

- the final judgment of the court
  - **did not** definitively **determine competing legal interests**, due e.g. to privacy and social services administration
  - **centred on the failure** of the government to offer any **meaningful explanation** of the technology, and even less its limits.

SyRI in court (2020)

- the SyRI technology was recently found contrary to Article 8 of the European Convention of Human Rights, which broadly protects the right to respect for private and family life, home and correspondence.

- the final judgment of the court
  - **did not** definitively **determine competing legal interests**, due e.g. to privacy and social services administration
  - **centred on the failure** of the government to offer any **meaningful explanation** of the technology, and even less its limits.

(adequate) transparency is a crucial requirement, but we plausibly need **other modes of knowing and assessing the interplay of the law and technology**

SyRI in court (2020)

- the SyRI technology was recently found contrary to Article 8 of the European Convention of Human Rights, which broadly protects the right to respect for private and family life, home and correspondence.

problematic for the indeterminate nature of law

- the final judgment of the court
  - **did not** definitively **determine competing legal interests**, due e.g. to privacy and social services administration
  - **centred on the failure** of the government to offer any **meaningful explanation** of the technology, and even less its limits.

problematic for the complexity of technology

(adequate) transparency is a crucial requirement, but we plausibly need **other modes of knowing and assessing the interplay of the law and technology**

methodological standpoint

- rather than focusing attention on the ideal norms or values applicable to AI generally, we look at the *assemblage* and the *interactions* that constitute its *viability* and *operation*

methodological standpoint

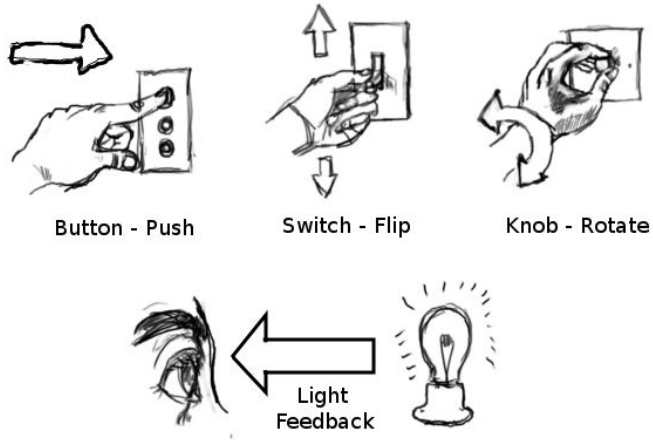
- rather than focusing attention on the ideal norms or values applicable to AI generally, we look at the *assemblage* and the *interactions* that constitute its *viability* and *operation*

- as a communicative practice that is materially situated, law coordinates **horizons of material expectations** among networked participants (including expectations of and among objects and things).

  these coordinating horizons function as ***material affordances.***

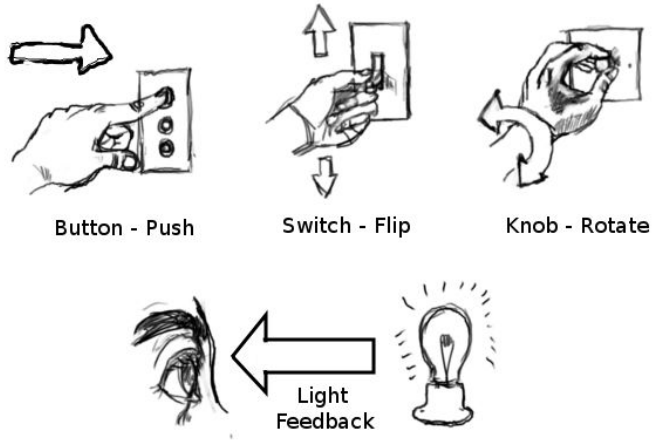
affordances

- affordance is opportunity of action: a behaviour of the agent that an environment (object) can “afford”.



affordances

- affordance is opportunity of action: a behaviour of the agent that an environment (object) can “afford”.

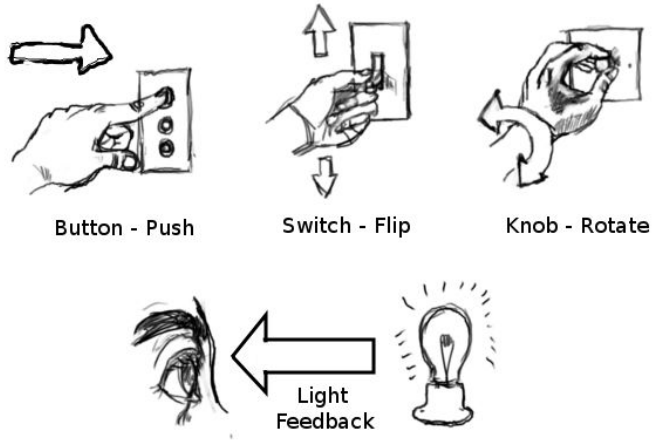


institutional affordances

the medieval port of Genoa, flourishing with the introduction of insurances, contract options and other mechanisms of risk management

affordances

- affordance is opportunity of action: a behaviour of the agent that an environment (object) can “afford”.



institutional affordances

the “reality” of law derives precisely from legal affordances defined and operating with other parts of the assemblage!

3. critical practice of AI

connection with critical practice of AI

- our methodology in line with the call for a critical practice of AI called for by Philip Agre.
- most AI research and development is centred on a single question: *does a proposed alternative solution work better?*
- As Agre makes clear, this begs a prior question, namely: ***what does it work for?***

central issue

the legal command is indeterminate

the computational command is complex

central issue

most people work on trying to disambiguate

the legal command is indeterminate

the computational command is complex

most people work on trying to make it understandable

central issue

most people work on trying to disambiguate

the legal command is indeterminate

the computational command is complex

most people work on trying to make it understandable

(suppose that) for all the reasons argued above

we renounce to approach these commands directly.

knowing by “encircling”

- encircling is a research technique recently proposed in *security studies* (De Goede, Bosma), developed to deal with problems of **secrecy**.

knowing by "encircling"

- encircling is a research technique recently proposed in *security studies* (De Goede, Bosma), developed to deal with problems of **secrecy**.

- the method 'is less focused on **uncovering the kernel of the secret**, than it is on analysing the **mundane lifeworlds of security practices and practitioners** that are powerfully **structured** through **codes and rites** of secrecy.'

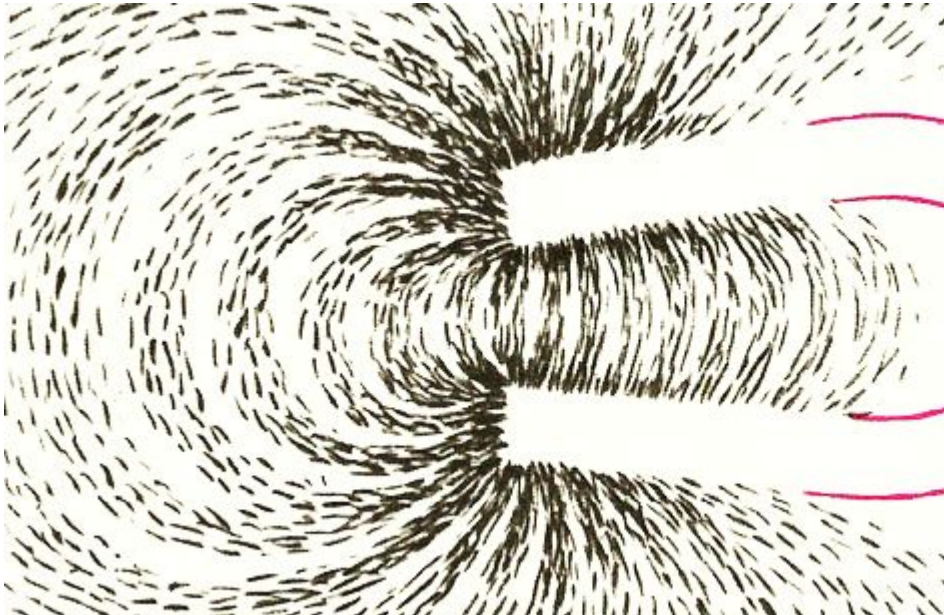
“mapping” values via encircling

- *our goal:* studying how **value attributions** denoting relative worth, merit, or importance, take form in ecologies of human and computational agents.

“mapping” values via encircling

- *our goal:* studying how **value attributions** denoting relative worth, merit, or importance, take form in ecologies of human and computational agents.

we cannot see “values”, but we
can see how people/AI deal with them

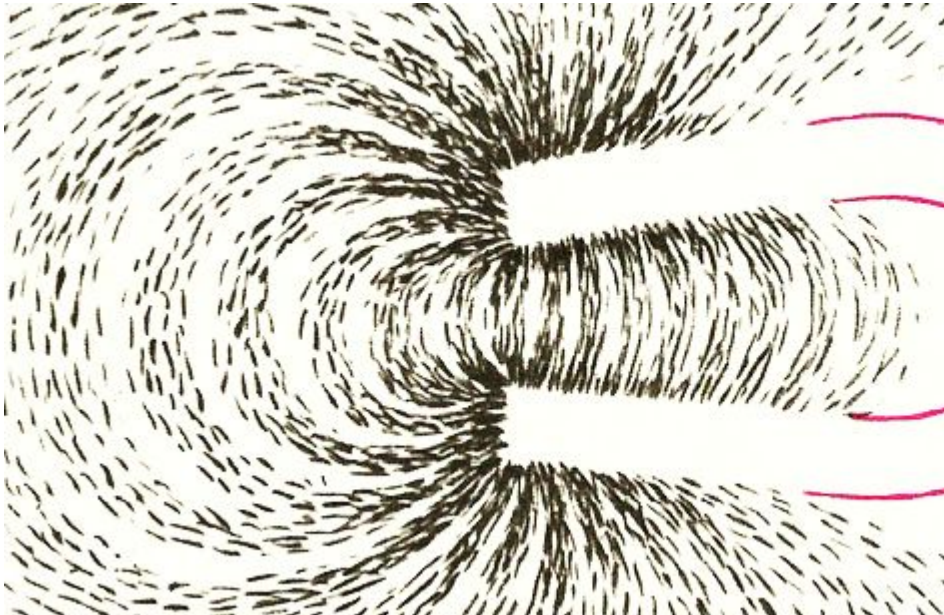


“mapping” values via encircling

- *our goal:* studying how **value attributions** denoting relative worth, merit, or importance, take form in ecologies of human and computational agents.

  we cannot see “values”, but we
  can see how people/AI deal with them

- possible axes (parallel work):
  1. **ambient technical knowledge**
  2. **local design conditions**
  3. **materialized values**



Conclusions

Wrapping up

- we are not arguing against rights-based and rule of law programs; but we elaborate on their limits,
- new research programs can be designed that go beyond such limits, so in complementation to standard programs.

Perspectives

possible uses of the proposed framework:

- **academics:** a wider array of interactions among incentives, pressures, materialities and routines for analysis
- **AI practitioners:** reflective standpoints relevant for design, development and deployment phases,
- **legal practitioners:** a wider number of sites to contest
- **regulators and judges:** fuller perspective on the normative conditions and stakes at play in any given outcome.

On mapping values in AI governance

2 December 2021, Algorithmic Law and Society symposium, HEC Paris, Paris

Geoff Gordon, Asser Institute, the Hague
Bernhard Rieder, Media Studies, University of Amsterdam
Giovanni Sileno, Informatics Institute, University of Amsterdam