How to contest automated decisions: A rule-based modelling

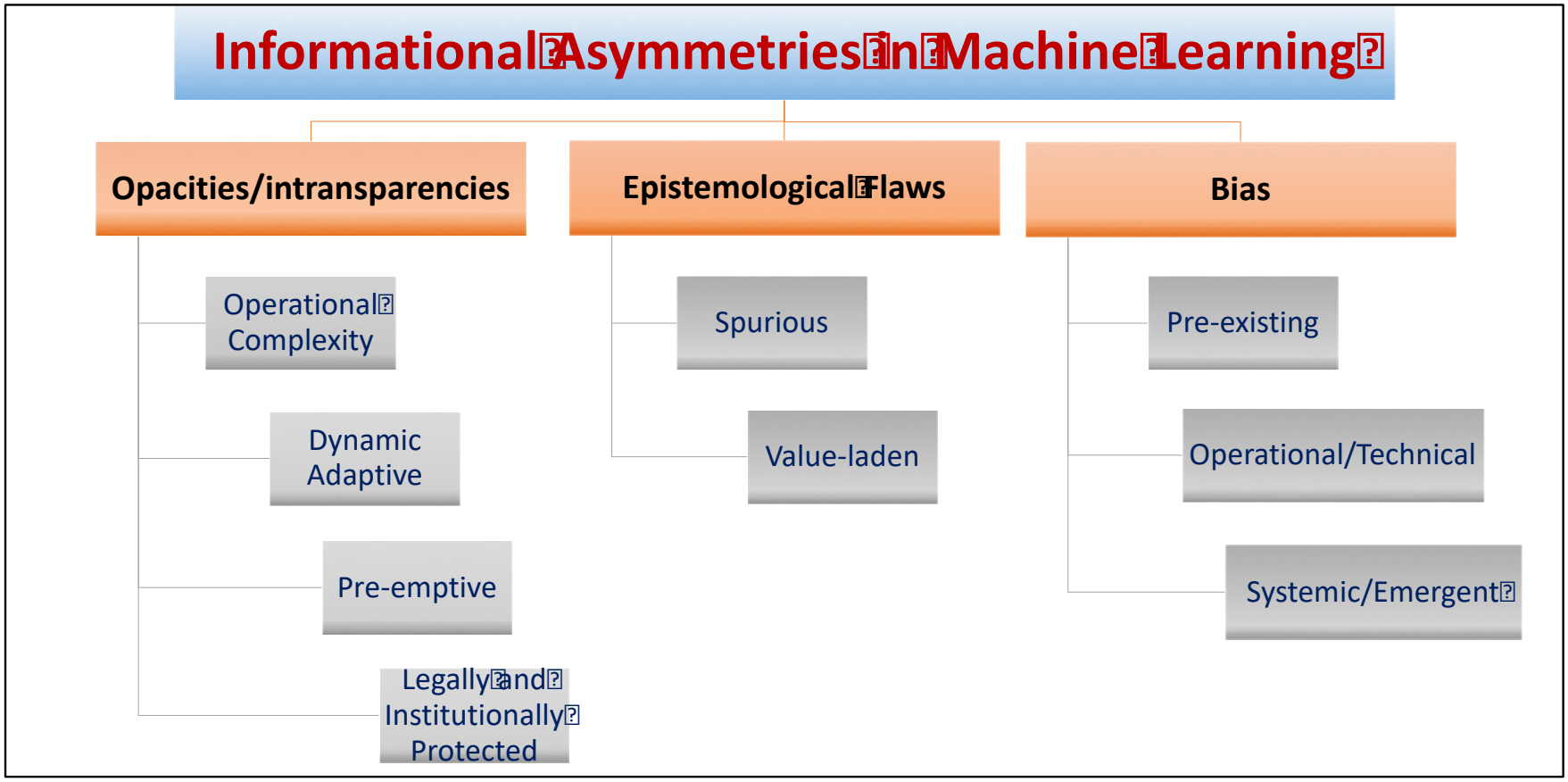
Emre Bayamlıoğlu

Which emails trigger which ads?

Which prior searches trigger which prices?

Transprency conuncdrum

- Data-driven decision-making is **unintelligible** in the sense that the recipient of the output (e.g., a classification decision), cannot construct any **concrete mapping** of how or why a particular classification has been arrived at from the given input.


Informational Asymmetries in Machine Learning
Opacities/intransparencies
Epistemological Flaws
Bias
Operational Complexity
Dynamic Adaptive
Pre-emptive
Legally and Institutionally Protected
Spurious
Value-laden
Pre-existing
Operational/Technical
Systemic/Emergent

Visibility of a different type

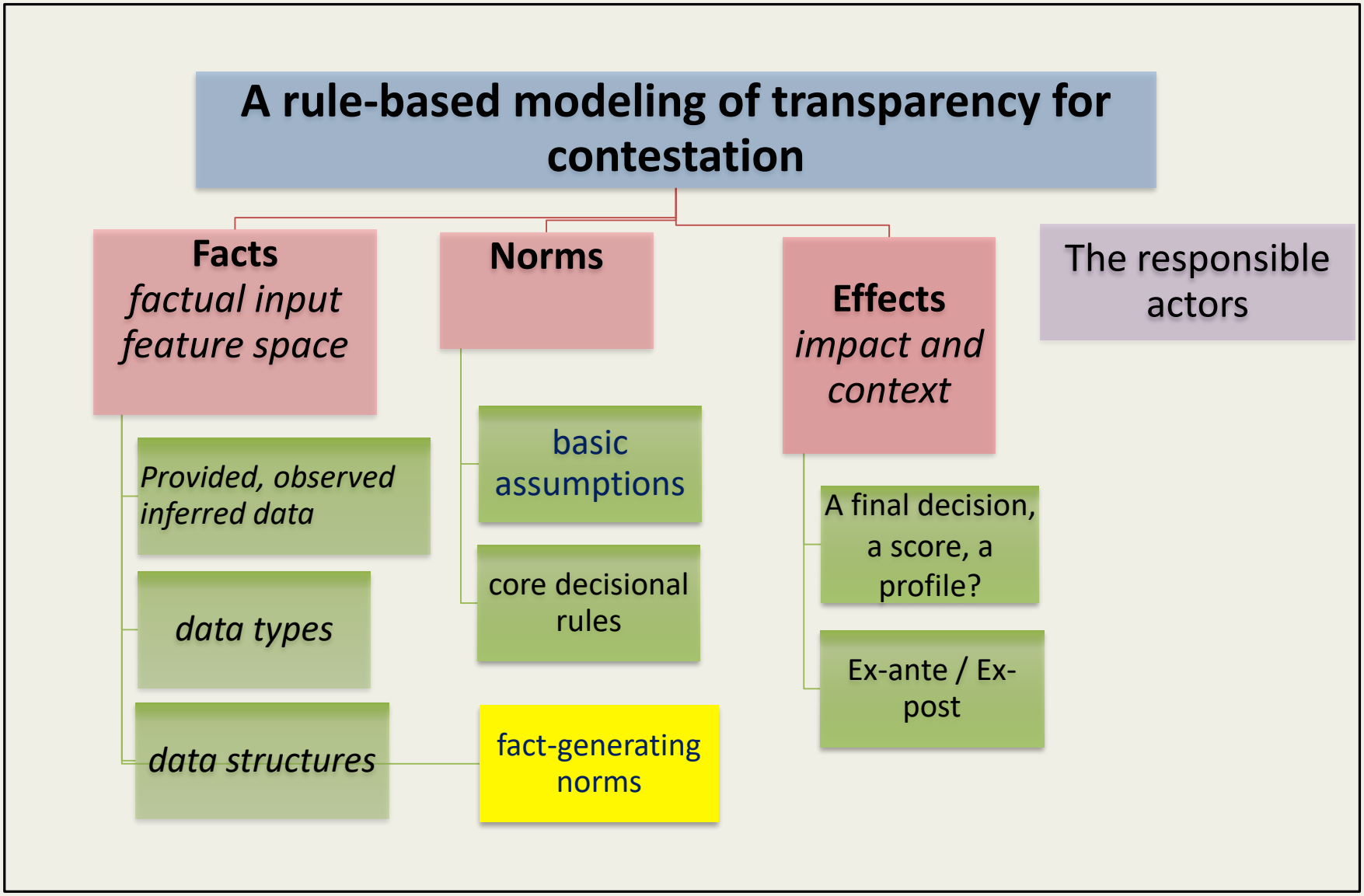
- *Actionable transparency* as an instrument to enforce rights .
  - interpretable,
  - reviewable,
  - reproducible,
  - inferable
  - engageable

Normativity: key to transparency

- *"Algorithmic decision-making necessarily embodies contestable epistemic and normative assumptions"*

  — Ruben Binns

- Conceptualisation of the outcome as a process based on ***facts, norms,*** and ***decisions/effects*** in the most abstract sense


A rule-based modeling of transparency for contestation
Facts
factual input
feature space
Norms
Effects
impact and
context
The responsible
actors
Provided, observed
inferred data
data types
data structures
basic
assumptions
core decisional
rules
fact-generating
norms
A final decision,
a score, a
profile?
Ex-ante / Ex-
post

A rule-based “explanation” of the system

- How and why a person, event, or situation is classified in certain ways, and what consequences follow from that?
- **Normative intelligibility** will mean that, given certain factual input the result could be **verified, justified** or alternatively **contested with reference to that rule**

Factual input /Feature space

- “data” is regarded **not as a tool of insight**, but simply as **informational or factual input** similar to the facts in a legal case.
- Observations and the feedback in the form of data are constructed as representations of “reality” for the system.

“synthetic method”

- Reverse engineers (dissects) the decional process
  - for a reconstruction on the basis of facts, norms and the following effects for the purposes of contestation
- A synthetic method for understanding of the “reality” by means of actual model-building

The content of the right to human intervention and contestation

- ***What to contest***
  - The Scope and the extent of the analysis
  - Accuracy of the data
  - Accuracy, appropriateness / expediency of the calculation
  - Normativity, Interpretation / assumptions (How normativity is defined, on what values/basis)

- ***Against whom***
  - H2H
  - H2M
  - M2M

Operationalisation of Transparency

i. **Physical access/level -** Conventional transparency – access, openness, visibility, notification and disclosure.
   - Failure against complexity

ii. **Algorithmic scrutiny** – Audit - Output transparency.
   - *Solution as a response to complexity*.

iii. **Algorithmic intervent**ion: Transparency by design - protection embedded.
   - Solution within complexity

Arguments from various impedements

- **Computational**
  - Complexity,
  - probabilistic reasoning
  - adaptive rule-making

Arguments from various impedements

- **Legal**
  - **Within the Article 22: individual right**
  - **IP rights**
    - IP claims hindering access or limiting disclosure
    - Use of IP protected elements in statistical investigation methods
    - Interoperability of auditing software with data processing systems
    - IP protection of audit tools (software, design features, metrics
  - **Contractual dimension/Freedom of contract**
  - **Right to knowledge/Freedom of speech**
  - **Machine integrity/Algorithmic privacy**

Arguments from various impedements

- **<u>Economic / business</u>**
  - Integrity of the system (gaming of the algorithm)
  - Feasibility
  - *Is "costs v. risks" the right paradigm?*