Regulating Multiagent Systems: A Basis for Responsible AI

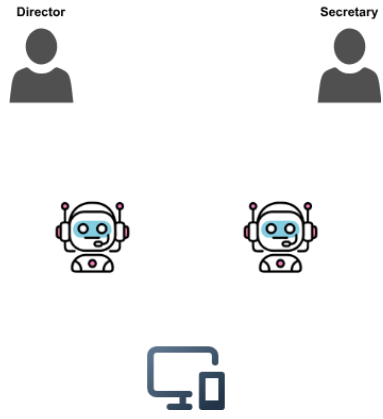
Luis Gustavo Nardin
gnardin@emse.fr

MINES Saint-Étienne

AI Agents and Human Insight – 3rd ENFIELD Webinar
November 21, 2025

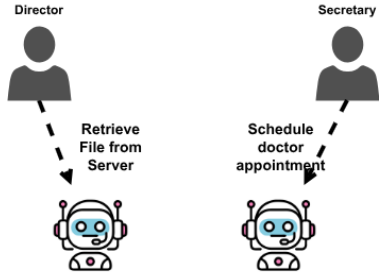
Introduction

- Human society is **increasingly becoming dependent** on IT
- Recent advances of AI promotes the **delegation** of goals to artificial agents
- Agents usually **interact** to achieve their delegated goals
- Agents interests are at best **imperfectly aligned**



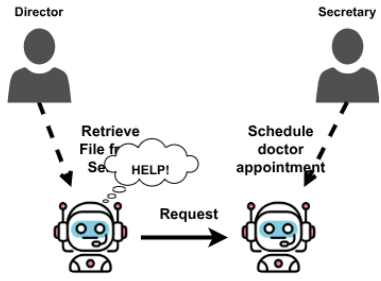
Introduction

- Human society is **increasingly becoming dependent** on IT
- Recent advances of AI promotes the **delegation** of goals to artificial agents
- Agents usually **interact** to achieve their delegated goals
- Agents interests are at best **imperfectly aligned**



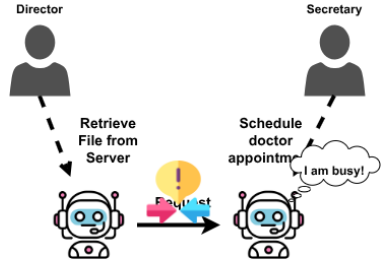
Introduction

- Human society is **increasingly becoming dependent** on IT
- Recent advances of AI promotes the **delegation** of goals to artificial agents
- Agents usually **interact** to achieve their delegated goals
- Agents interests are at best **imperfectly aligned**



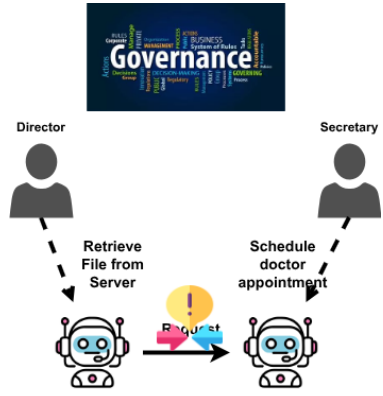
Introduction

- Human society is **increasingly becoming dependent** on IT
- Recent advances of AI promotes the **delegation** of goals to artificial agents
- Agents usually **interact** to achieve their delegated goals
- Agents interests are at best **imperfectly aligned**



Introduction

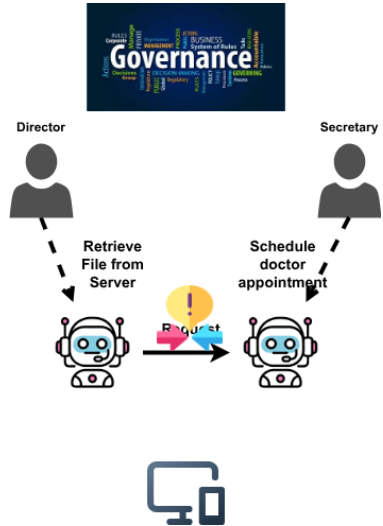
- The success of these systems relies on effective **governance**
  - how interactions are controlled
- Governance should maintain the **equilibrium** between
  - **Autonomy** of the agents
  - **Predictability** of the system



How can we govern these systems interfering the least with the autonomy of the agents?

Introduction

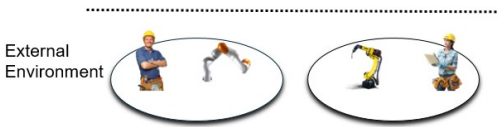
- The success of these systems relies on effective **governance**
  - how interactions are controlled
- Governance should maintain the **equilibrium** between
  - **Autonomy** of the agents
  - **Predictability** of the system



How can we govern these systems interfering the least with the autonomy of the agents?

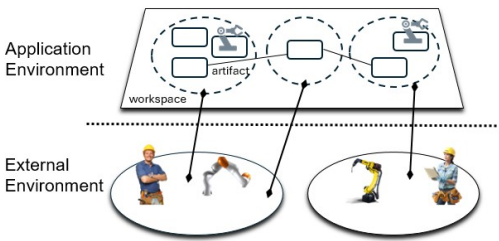
Multiagent Systems: Multiagent Oriented (MAO) Approach

(Boissier et al., 2020)



Multiagent Systems: Multiagent Oriented (MAO) Approach

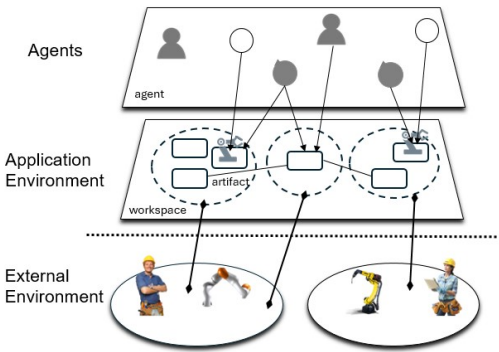
(Boissier et al., 2020)



- **Environment**: Shared medium providing the conditions for agents to exist and act
- **Agent**: Autonomous decision-makers capable of responding to events while pursuing goals and taking action to achieve them
- **Organization**: Abstractions to declare and make accessible to agents their collective structure and functioning in a shared environment

Multiagent Systems: Multiagent Oriented (MAO) Approach

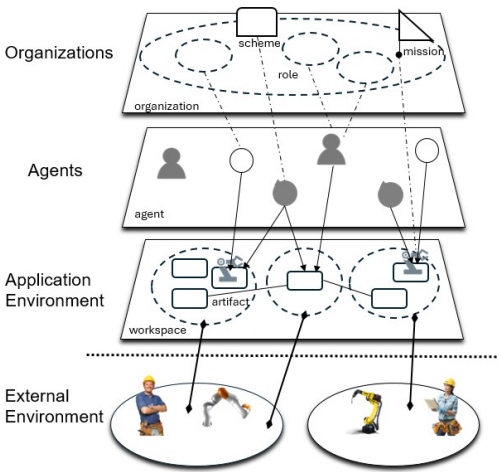
(Boissier et al., 2020)



- **Environment**: Shared medium providing the conditions for agents to exist and act
- **Agent**: Autonomous decision-makers capable of responding to events while pursuing goals and taking action to achieve them
- **Organization**: Abstractions to declare and make accessible to agents their collective structure and functioning in a shared environment

Multiagent Systems: Multiagent Oriented (MAO) Approach

(Boissier et al., 2020)

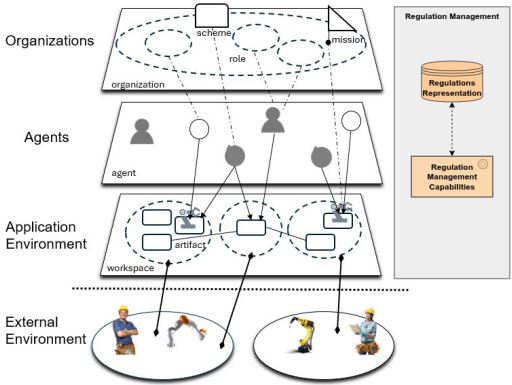


- **Environment**: Shared medium providing the conditions for agents to exist and act
- **Agent**: Autonomous decision-makers capable of responding to events while pursuing goals and taking action to achieve them
- **Organization**: Abstractions to declare and make accessible to agents their collective structure and functioning in a shared environment

Regulated Multiagent Systems

- **Regulations** can be used in MAS to guide agents towards the overall system objectives, while maintaining agent autonomy
- **Regulation Management**
  - Regulations Representation
  - Regulation Management Capabilities

Regulated Multiagent Systems



- **Regulations** can be used in MAS to guide agents towards the overall system objectives, while maintaining agent autonomy
- **Regulation Management**
  - Regulations Representation
  - Regulation Management Capabilities

Regulations Representation

- **Constitutive Norms**
  - Institutional interpretations of environmental facts
  - Link *brute facts* to *institutional facts*

$$id : x \text{ count-as } y \text{ [when } t\text{] [while } m\text{]}$$

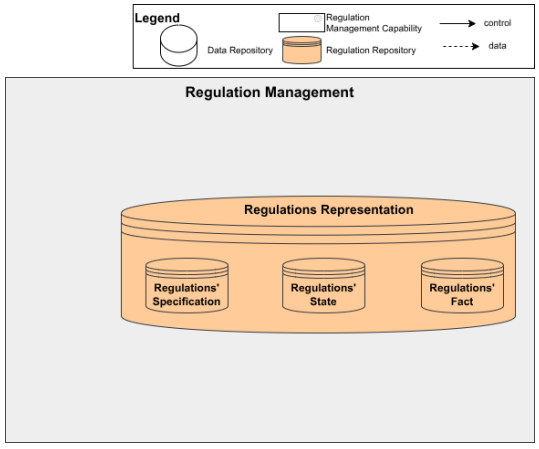
- **Regulative Norms**
  - Prescribe constraints on behavior or state of affairs that are expected
  - Qualify a behavior or state of affairs as obligations, prohibitions and permissions

$$\text{norm } id : \delta \rightarrow \psi \text{ if } \phi : sr_i(args)$$

- **Sanction rules**
  - Consequences of compliance with or violation of regulative norms

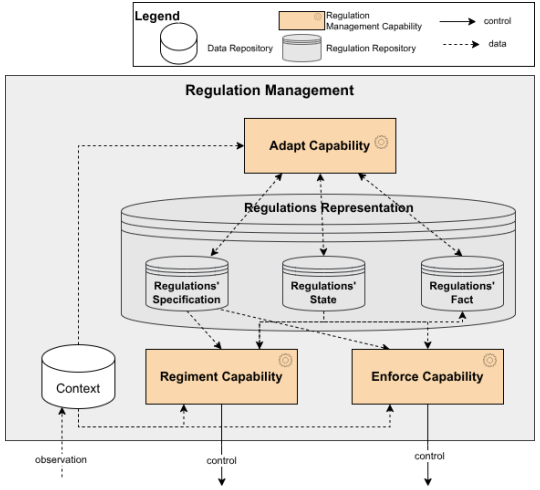
$$\text{sanction-rule } sr_i(args) : \rho \rightarrow \text{sanction}(\alpha, \gamma)$$

Regulations Representation



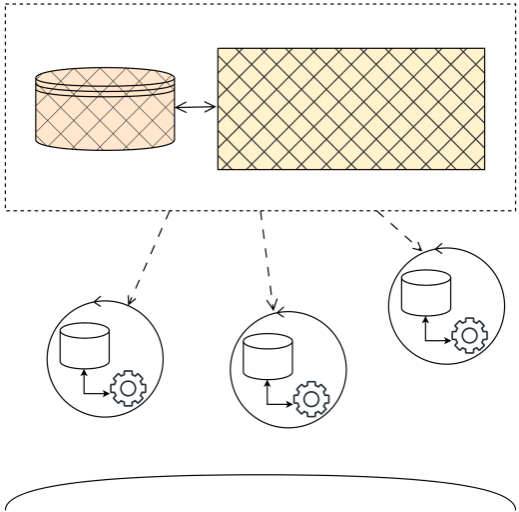
Yan, E., Nardin, L.G., Boissier, O. & Sichman, J. (2025). A Regulation Adaptation Model for Multi-Agent Systems. In 28th European Conference on Artificial Intelligence (ECAI 2025), Bologna, Italy

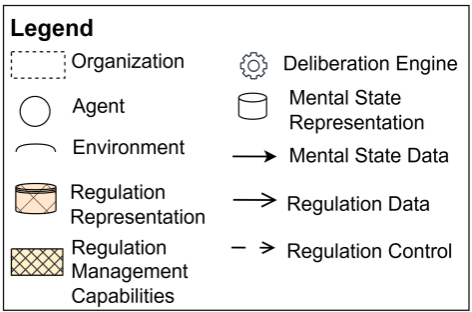
Regulation Management Capabilities



Yan, E., Nardin, L.G., Boissier, O. & Sichman, J. (2025). A Regulation Adaptation Model for Multi-Agent Systems. In 28th European Conference on Artificial Intelligence (ECAI 2025), Bologna, Italy

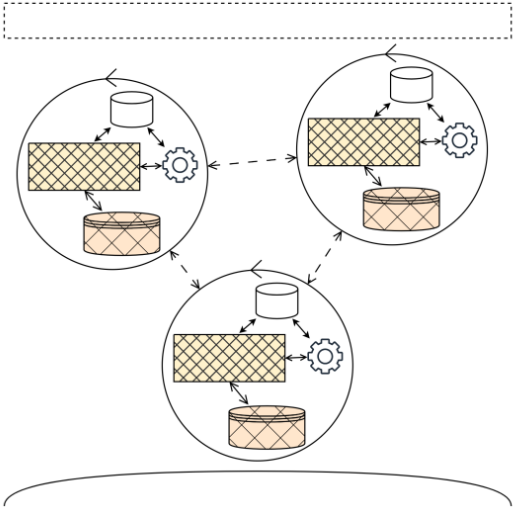
Organization-Centric Regulation Management Perspective

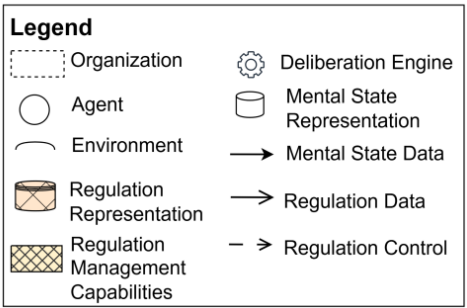




Yan, E., Nardin, L.G., Boissier, O. & Sichman, J. (2025). A Unified View on Regulation Management in Multi-Agent Systems. In Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XVIII

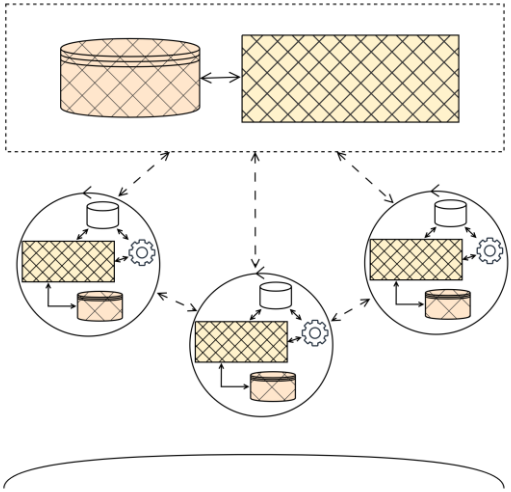
Agent-Centric Regulation Management Perspective

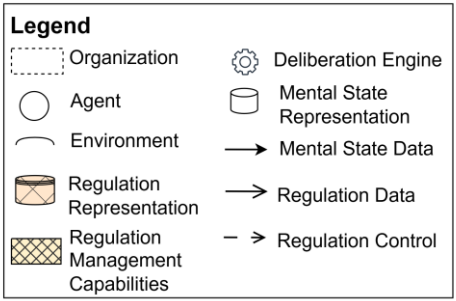




Yan, E., Nardin, L.G., Boissier, O. & Sichman, J. (2025). A Unified View on Regulation Management in Multi-Agent Systems. In Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XVIII

Hybrid-Centric Regulation Management Perspective

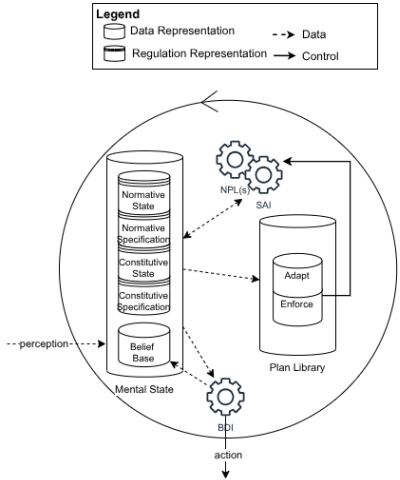




Yan, E., Nardin, L.G., Boissier, O. & Sichman, J. (2025). A Unified View on Regulation Management in Multi-Agent Systems. In Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XVIII

Normative Agent Architecture

- BDI normative agent architecture
- NPL(s) and SAI Engine
- Regulations Repository
- Plans
  - Adapt
  - Enforce



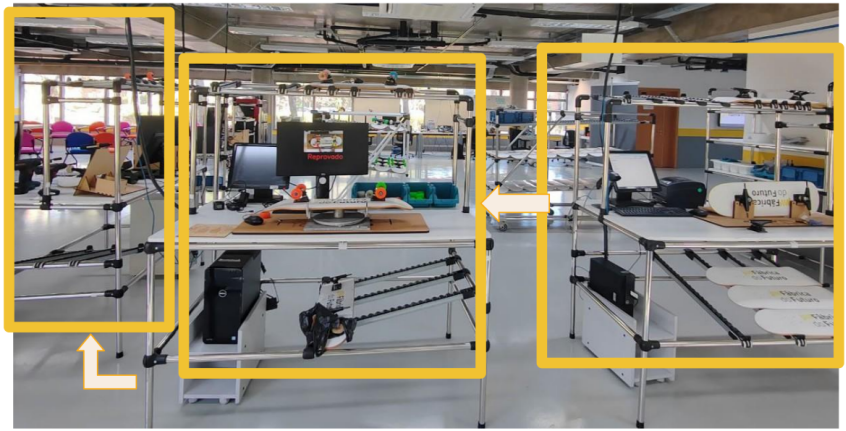
Yan, E., Nardin, L.G., Boissier, O. & Sichman, J. (2025). A Unified View on Regulation Management in Multi-Agent Systems. In Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XVIII

Skateboard Assembly Line

Objective: Assembly custom skateboards based on customer orders

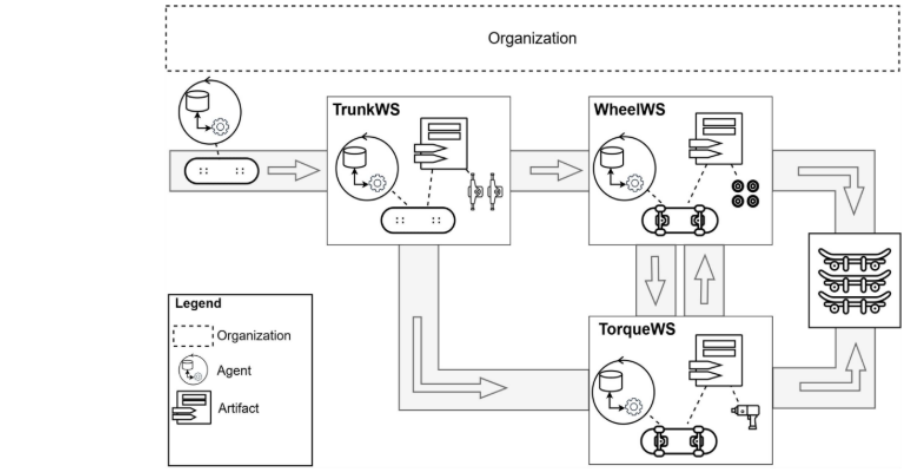
Workstations

- TrunkWS
- WheelWS
- TorqueWS



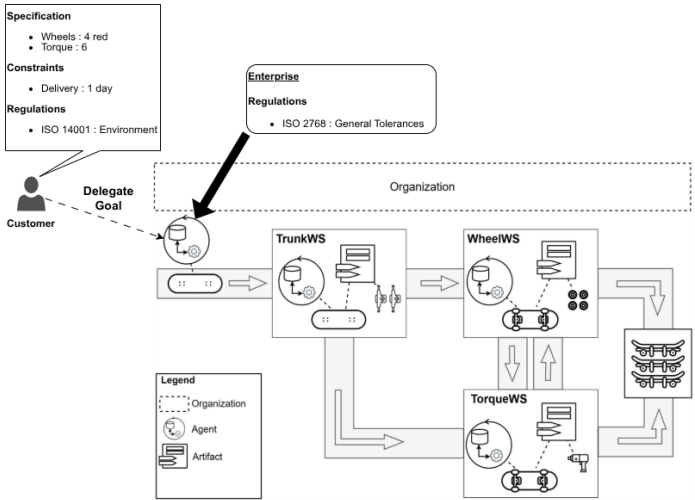
Yan, E., Nardin, L.G., Hübner, J., Boissier, O. & Sichman, J. (2025). Perspectives on Regulation Adaptation in Multi-Agent Systems: from Agent to Organization Centric and Beyond. In Anais do XIX Workshop-Escola de Sistemas de Agentes, seus Ambientes e Aplicações, Fortaleza/CE

Skateboard Assembly Line



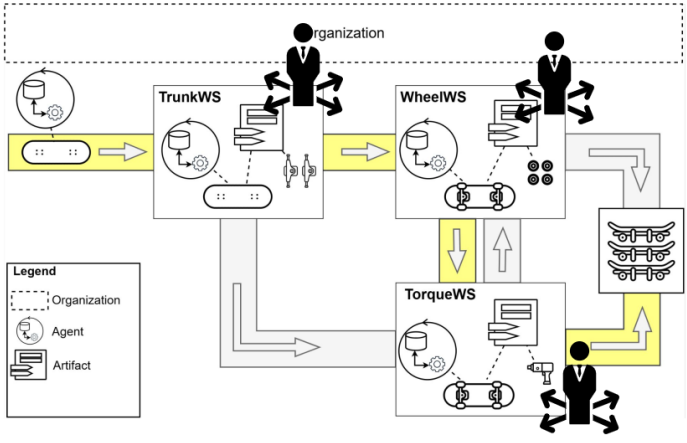
Yan, E., Nardin, L.G., Hübner, J., Boissier, O. & Sichman, J. (2025). Perspectives on Regulation Adaptation in Multi-Agent Systems: from Agent to Organization Centric and Beyond. In Anais do XIX Workshop-Escola de Sistemas de Agentes, seus Ambientes e Aplicações, Fortaleza/CE

Skateboard Assembly Line



Yan, E., Nardin, L.G., Hübner, J., Boissier, O. & Sichman, J. (2025). Perspectives on Regulation Adaptation in Multi-Agent Systems: from Agent to Organization Centric and Beyond. In Anais do XIX Workshop-Escola de Sistemas de Agentes, seus Ambientes e Aplicações, Fortaleza/CE

Skateboard Assembly Line



Yan, E., Nardin, L.G., Hübner, J., Boissier, O. & Sichman, J. (2025). Perspectives on Regulation Adaptation in Multi-Agent Systems: from Agent to Organization Centric and Beyond. In Anais do XIX Workshop-Escola de Sistemas de Agentes, seus Ambientes e Aplicações, Fortaleza/CE

Support for Responsible AI

- **Collaborative Decision-Making**
  - Facilitate balanced decisions based on explicit expected behavior
  - Establish mechanisms for detecting conflicts or violations enabling quick response
- **Transparency and Explainability**
  - Enhance the comprehension by humans of agents decision-making processes
- **Accountability**
  - Identify clearly the distributed responsibilities, enabling the creation of checks for accountability
- **Adaptability**
  - Enable easy adaptation to new or changing regulatory frameworks

Thank you!

Partially funded by ANR-FAPESP NAIMAN project (ANR-22-CE23-0018-01, FAPESP 2022/03454-1)