

**MUNI**  
FI

# **Trust building via Adaptive Safety in Autonomous Ecosystems**

**Dávid Halász**



# Problems with Safety

- Intentional vs. unintentional behavior
- Supervision awareness
- Misclassification of behavior
- Feedback loops (cascading failures)
- Compatibility (with each other or humans)

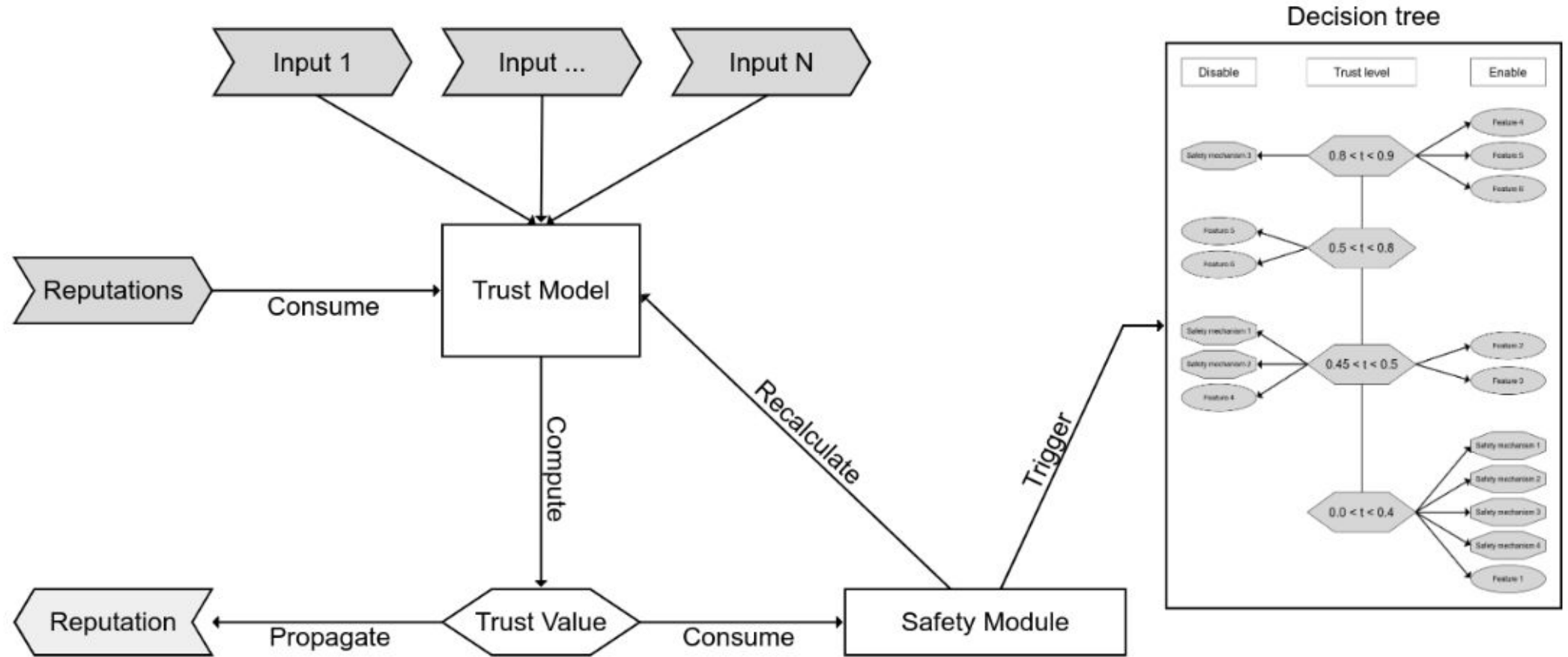
**M U N I**  
**F I**

**Granularity (Trust) + Graduality (Safety)**

# Trust-based Adaptive Safety

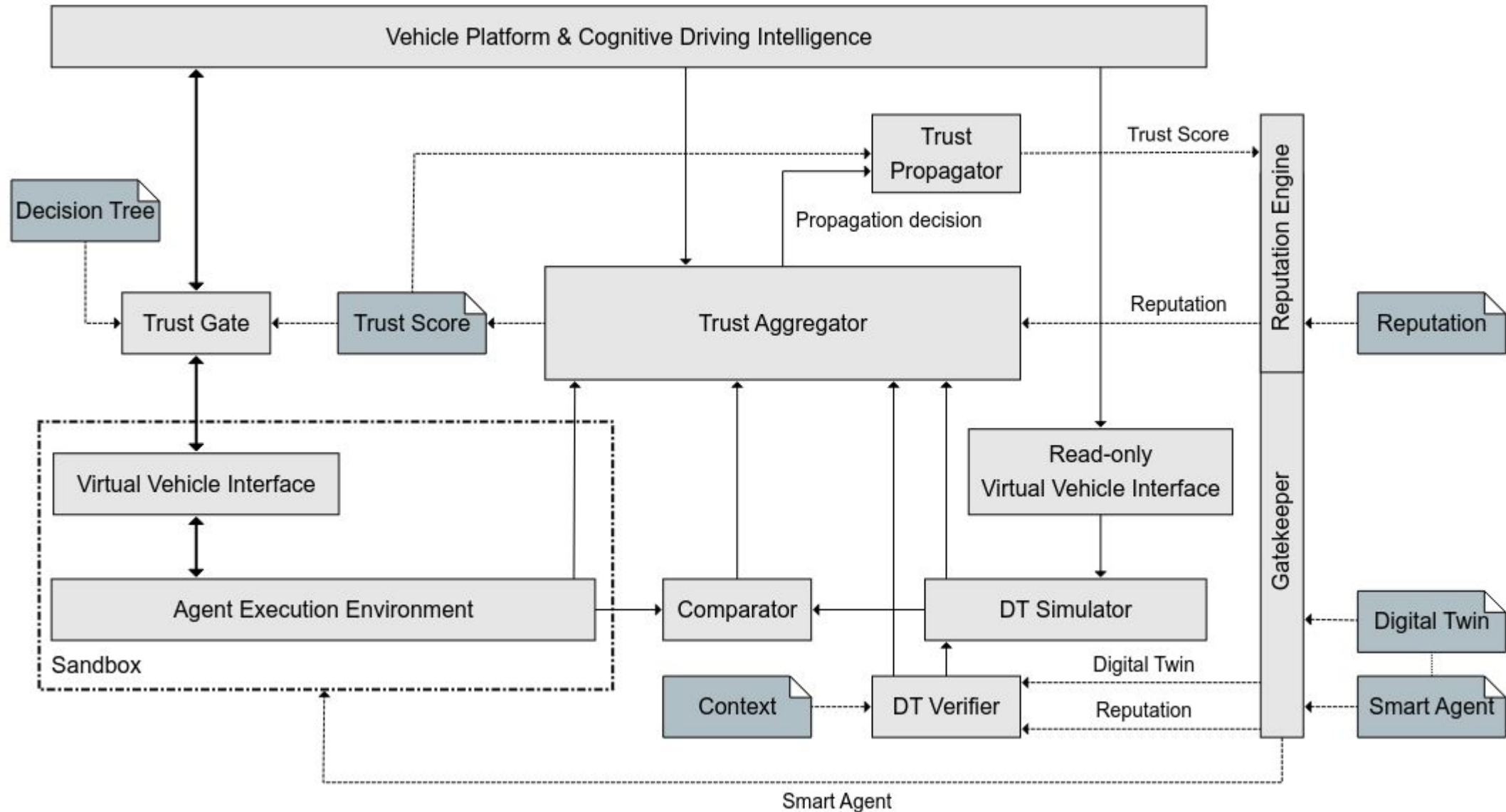
- Real-time trust assessment
  - Predictive Simulation using Digital Twin
  - Live compliance checking
  - External reputation
  - Non-binary complex trust score (%)
- Dynamic toggling of safety mechanisms
  - Adapt safety methods to the trust score
  - Consider the possible impact on other actors

# Safety Assurance Framework



# Adaptive Safety for Smart Agents

- Smart Agents
  - Coordination of heterogeneous systems
  - Ensuring safety of the whole ecosystem
  - Providing additional features
- Safety & security concerns
  - Possibly malicious third-party software
  - Faulty behavior
- Trust and Digital Twin predictive simulation

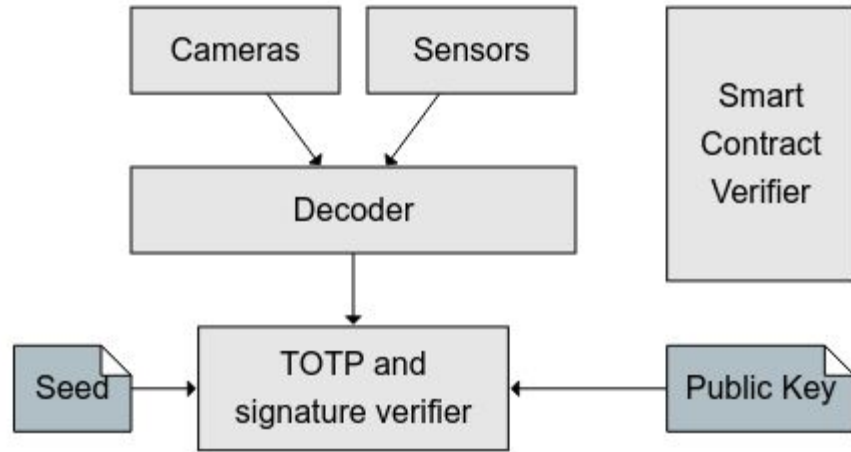




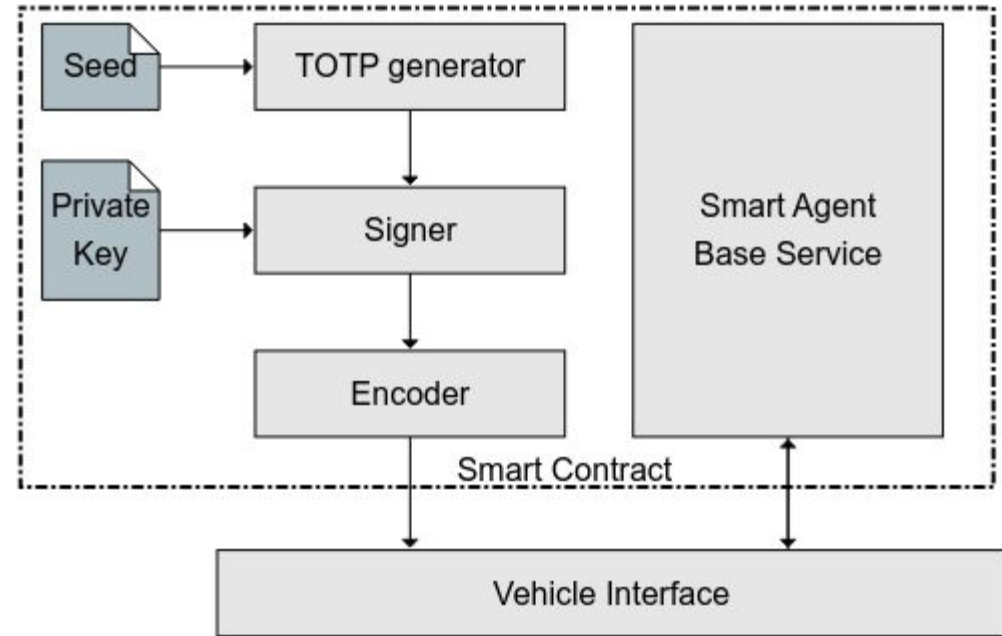
# Trusted Execution

- Mandatory running of Smart Agents
  - Executed on the right device?
  - Executed with the right privileges?
  - Executed without instruction tampering?
- Blockchain and Micro-movements

# Trusted Execution



Partial Architecture of a Road-Side Unit

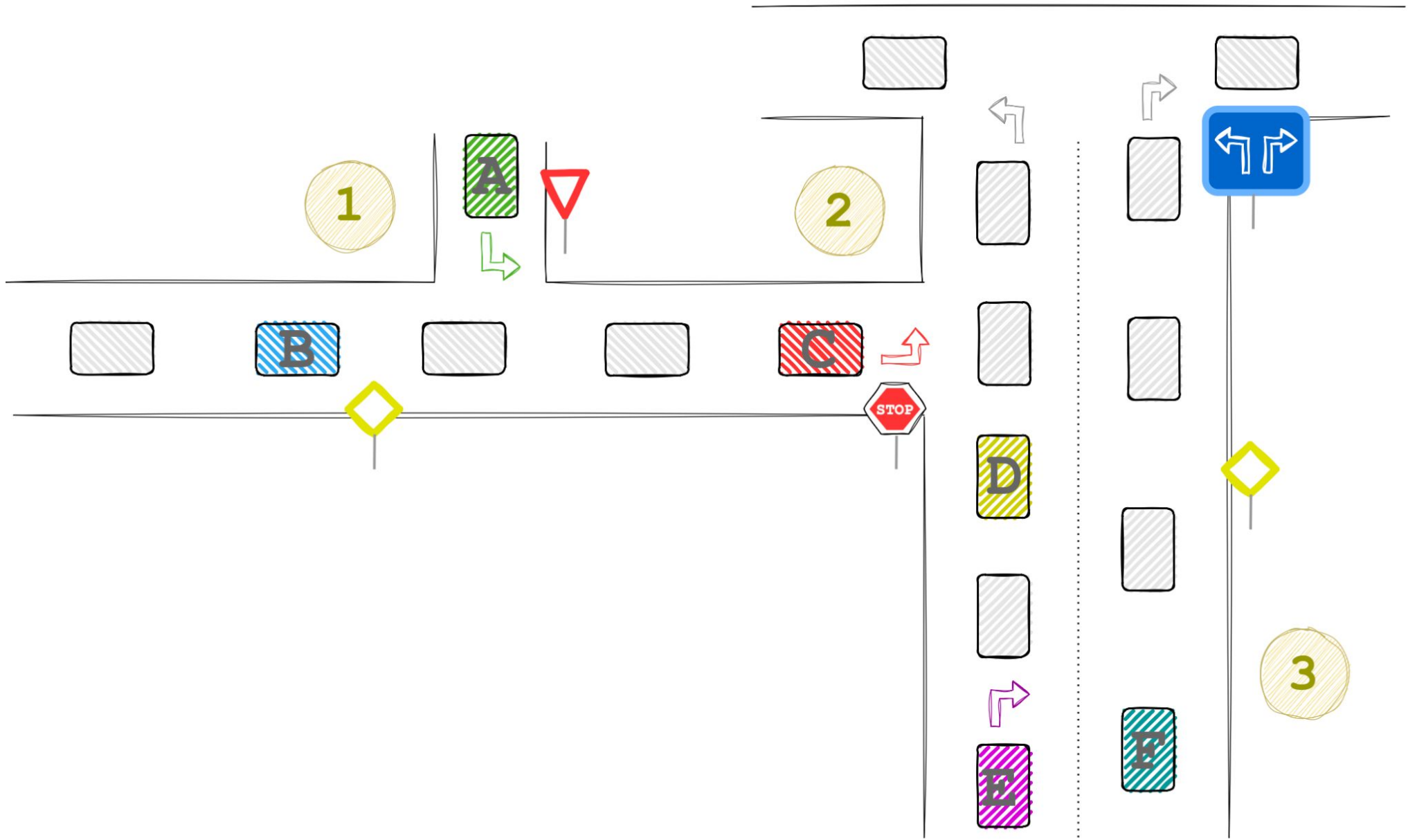


Conceptual Architecture of a Smart Agent

**MUNI**  
FI

# **Incentivizing fair Behavior**

With RNDr. Daša



# Monetization of Fairness

- Fair behavior described by Smart Contracts
  - By vehicles for individual gain
  - By the ecosystem for collective impacts
- Purchase fair behavior with Fairness Tokens
  - Token value decreases over time
- Social credit system for Autonomous Vehicles

**M U N I**  
F I

**Work with us :)**

Pleeease