AI and Automated Driving: An Ethically Aligned Design

Raja Chatila

Institut des Systèmes Intelligents et de Robotique (ISIR)

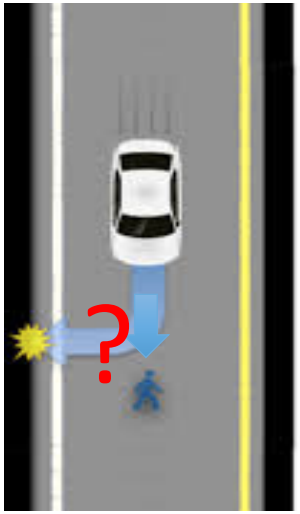
University Pierre & Marie Curie, Sorbonne Universités, Paris

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

Automated Driving: Increased Safety?

- Complex and evolving environment
- Mixed traffic
- Unexpected events
- Incomplete, noisy perception
- Non deterministic actions (non fully predictable action outcomes)
- Limited situation awareness
- Real-time decisions

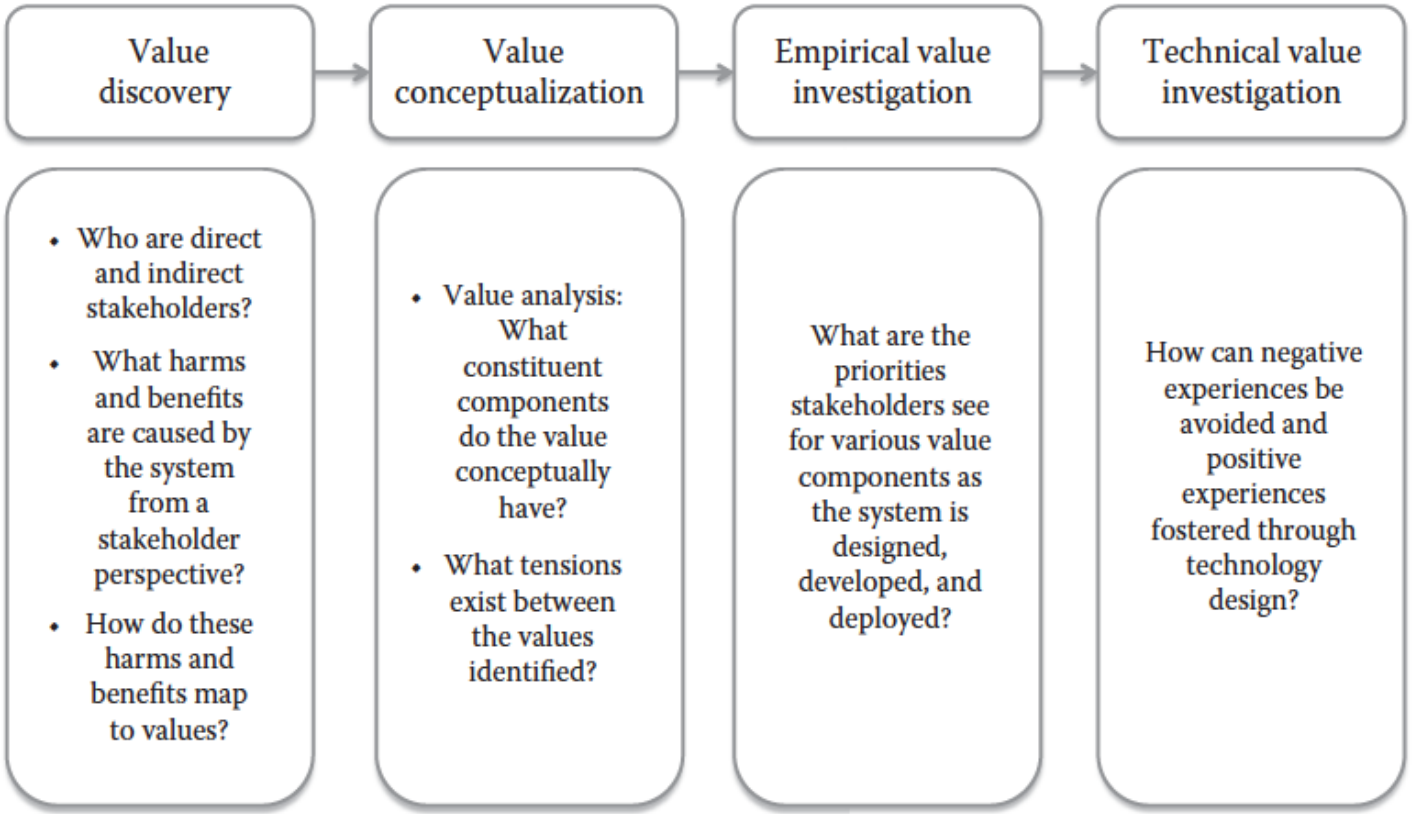
Automated Driving: Normal situations and moral dilemmas



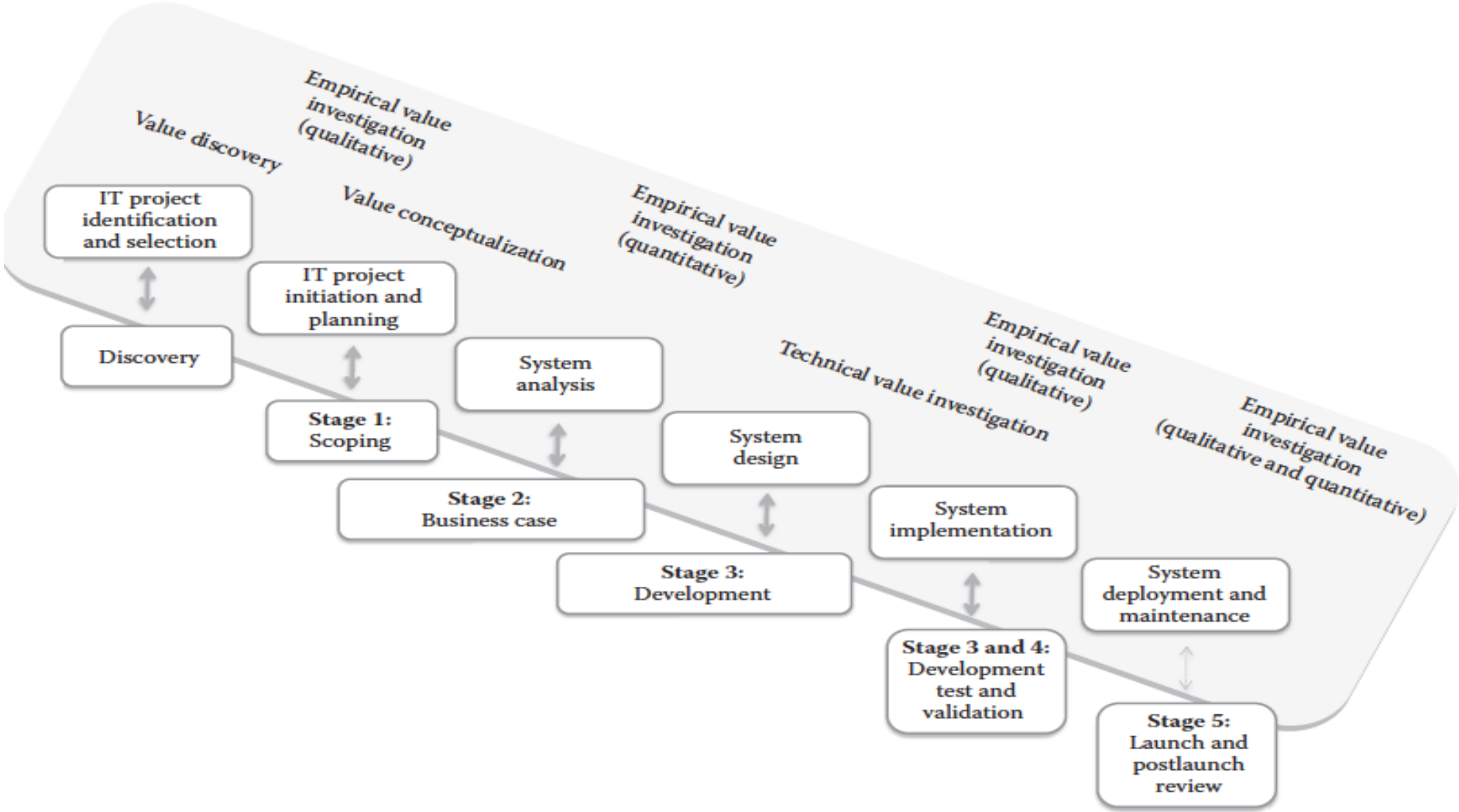
A Design Approach for an AI Moral Agent: Value by Design

- Dilution of responsibilities: Designer? Manufacturer ? Dealer? Operator? Owner? The Autonomous agent?
- Creation of a moral personhood for autonomous agents? Probably a bad idea
- Following an ethical approach provides guidelines and common grounds
- Operational rules can be implemented and explained
- Shared norms, and local adaptation
- Problems:
  - acceptability and consensus on the ethical values
  - Rule set completeness
  - Uncertainties

Value-Sensitive Design



Sarah Spiekermann, Ethical IT Innovation, CRC Press , 2016


Value discovery
Empirical value investigation (qualitative)
Value conceptualization
Empirical value investigation (quantitative)
Technical value investigation
Empirical value investigation (qualitative)
Empirical value investigation (qualitative and quantitative)
IT project identification and selection
Discovery
IT project initiation and planning
Stage 1: Scoping
System analysis
Stage 2: Business case
System design
Stage 3: Development
System implementation
Stage 3 and 4: Development test and validation
System deployment and maintenance
Stage 5: Launch and postlaunch review

Conclusions: Design Guidelines

- Pragmatic moral approaches based on learning and validation may be more appropriate
- Clearly define limits of machine capacities: decision-making, control algorithms, perception.
- Qualify uncertainties.
- Design for tracing robot decisions.
- Design for predictability of human-machine system.
- Transparent design of decisional system