CDT-D^{2}AIR Course on
D^2 Robots and Autonomous Agents

Interacting with Humans – Some Issues and Approaches

Subramanian Ramamoorthy
School of Informatics
University of Edinburgh

1 April 2026

Roots in Early Robotics: Master-Slave Systems

Jean Vertut & Thomas B. Sheridan with 1960s master-slave manipulator

Sheridan, T. B. (2016). Human–robot interaction: status and challenges. *Human factors*, *58*(4), 525-532.

HRI: Four Principal Areas [Sheridan]

Remote interaction

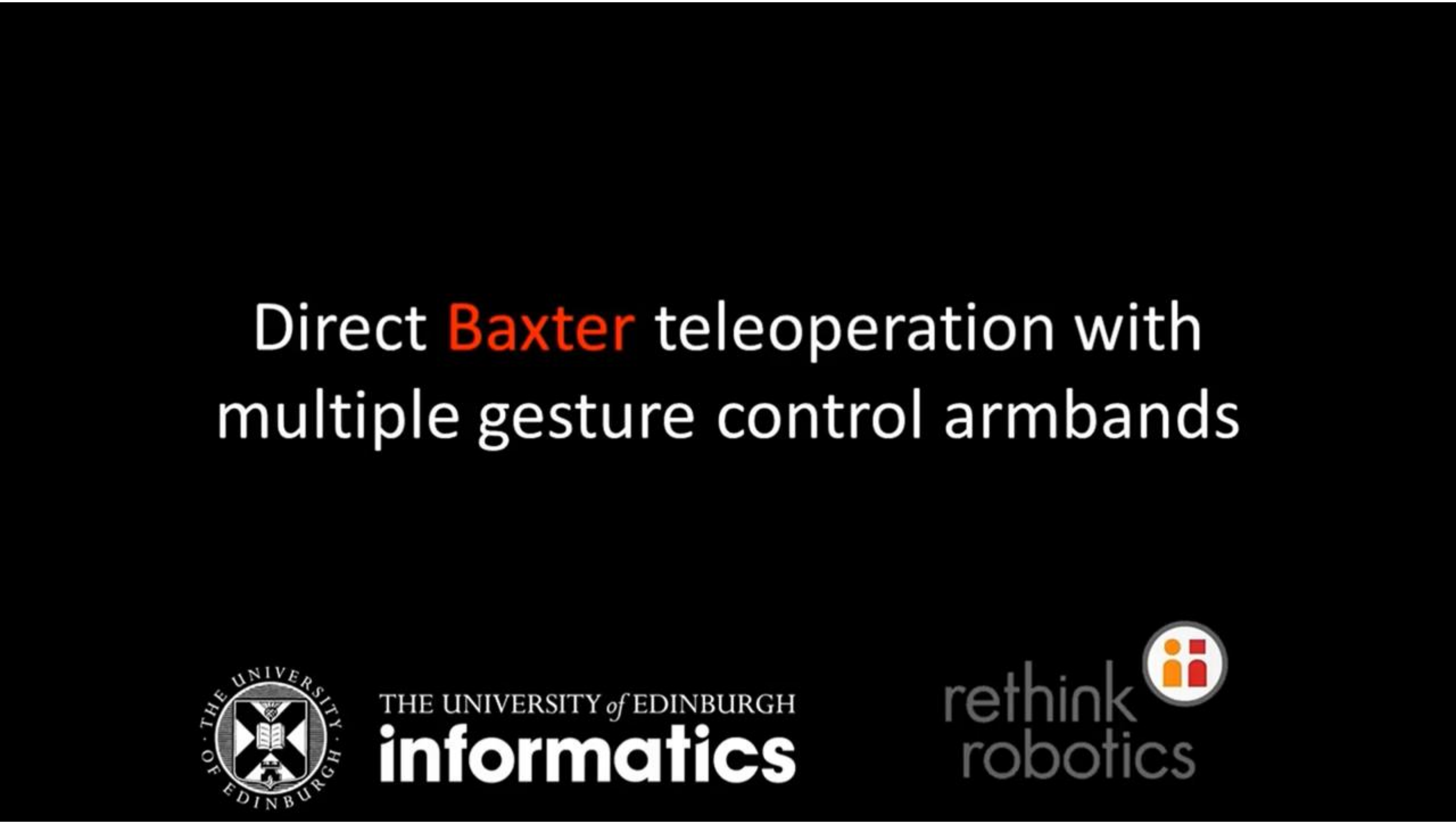
- Human supervisory control of robots performing routine tasks
  - Limited autonomy “telerobots” in manufacturing or warehouses (human is in charge at task level)
- Remote control of space, airborne, terrestrial, and undersea vehicles for nonroutine tasks in hazardous or inaccessible environments
  - “Tele-operated” robots with human in charge quite closely

Proximate Interaction

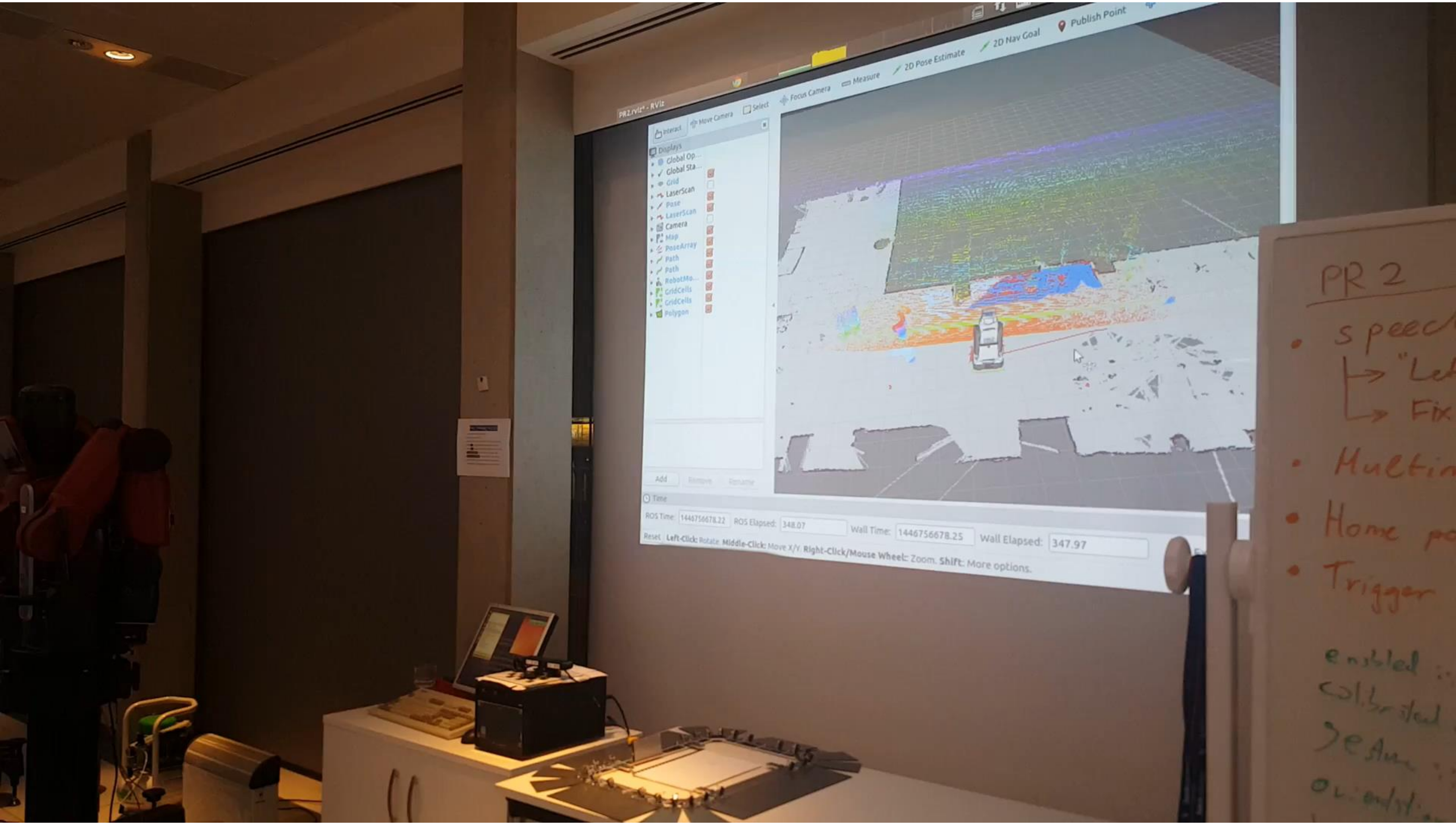
- Automated Vehicles (human is mainly a passenger)
- Human-robot social interactions (human is counter-party)

Sheridan, T. B. (2016). Human–robot interaction: status and challenges. *Human factors*, *58*(4), 525

Example 1: Teleoperation



Example 2: Human-Robot Handover



Attributes of HRI Problems [According to Goodrich and Schultz]

Interaction, the process of working together to accomplish a goal, emerges from the confluence of the following factors:

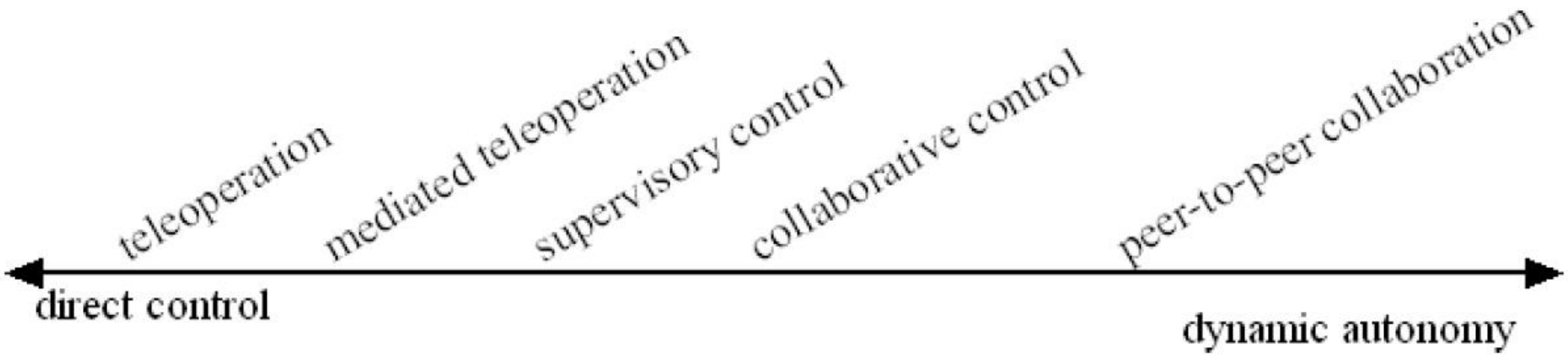
- Level and behaviour of autonomy
- Nature of information exchange
- Structure of the team
- Adaptation, learning, and training of people and the robot
- `Shape' of the task

Goodrich, M. A., & Schultz, A. C. (2008). Human–robot interaction: a survey. *Foundations & Trends® in Human–Computer Interaction*, *1*(3), 203-275.

Levels of Autonomy

1. Computer offers no assistance; human does it all.
2. Computer offers a complete set of action alternatives.
3. Computer narrows the selection down to a few choices.
4. Computer suggests a single action.
5. Computer executes that action if human approves.
6. Computer allows the human limited time to veto before automatic execution.
7. Computer executes automatically then necessarily informs human.
8. Computer informs human after automatic execution only if asked.
9. Computer informs human after automatic execution only if it decides to.
10. Computer decides everything and acts autonomously, ignoring the human.

Levels of Autonomy



This axis does not correlate 1-to-1 with difficulty.
For instance, full autonomy may be harder than peer to peer collaboration.

Problems in the middle

Ironies of automation:

When working in the supervisory setting, automation may expand rather than eliminate problems with a human operator

As automation gets better, long-term knowledge and skill may deteriorate. When the human is finally called upon, their awareness and skill may be found wanting.

L. Bainbridge, Ironies of automation, *Automatica*, 1983

Information Exchange: Many Options

- Visual displays, typically presented as graphical user interfaces or augmented reality interfaces
- Gestures, including hand and facial movements and by movement-based signalling of intent
- Speech and natural language, which include both auditory speech and text-based responses, and which frequently emphasize dialog and mixed-initiative interaction
- Non-speech audio, frequently used in alerting, and physical interaction and haptics, frequently used remotely in augmented reality or in teleoperation to invoke a sense of presence especially in telemanipulation tasks and also frequently used proximately to promote emotional, social, and assistive exchanges

Teams: Many Questions

- How many remote robots can a single human manage?
  - Depends on level of autonomy, task and means of communication
  - In practice, e.g., in rescue, we need at least two operators on scene
  - With sophisticated autonomy, perhaps one human operator will do
- Organisation of teams: who has the authority to make certain decisions: robot, interface software, or human?
- Who has the authority to issue instructions or commands to the robot and at what level: strategic, tactical, or operational?
- How are conflicts resolved, especially when robots are placed in peer-like relationships with multiple humans?
- How are roles defined and supported: is the robot a peer, an assistant or a slave?

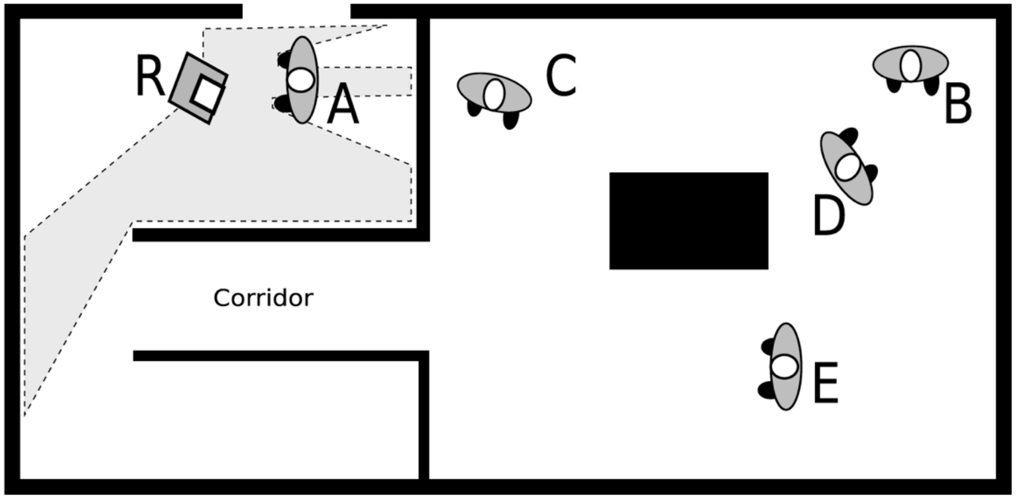
Adaptation, Learning and Training

- Tradeoff between robots not needing much operator training (e.g., in schools or nursing homes) vs. need to ensure safety through extensive training (e.g., in hazardous environments)
- Training humans to use robots (typical for remote applications): using the interface, interpreting video, controlling the robot, coordinating with other members of the team, and staying safe while operating the robot in a hostile environment.
- Training in applications involving proximate robots: to produce learning or behavioural responses with humans.
  - Therapeutic and social robots designed to change, educate, or train people, especially in long-term interactions

Problem Domains for HRI [Goodrich + Schultz]

Application area	Remote/ Proximate	Role	Example
Search and rescue	Remote	Human is supervisor or operator	Remotely operated search robots
	Proximate	Human and robot are peers	Robot supports unstable structures
Assistive robotics	Proximate	Human and robot are peers, or robot is tool	Assistance for the blind, and therapy for the elderly
	Proximate	Robot is mentor	Social interaction for autistic children
Military and police	Remote	Human is supervisor	Reconnaissance, de-mining
	Remote or Proximate	Human and robot are peers	Patrol support
	Remote	Human is information consumer	Commander using reconnaissance information
Edutainment	Proximate	Robot is mentor	Robotic classroom assistant
		Robot is mentor	Robotic museum tour guide
		Robot is peer	Social companion
Space	Remote	Human is supervisor or operator	Remote science and exploration
	Proximate	Human and robot are peers	Robotic astronaut assistant
Home and industry	Proximate	Human and robot are peers	Robotic companion
	Proximate	Human is supervisor	Robotic vacuum
	Remote	Human is supervisor	Robot construction

Human-aware Robot Navigation



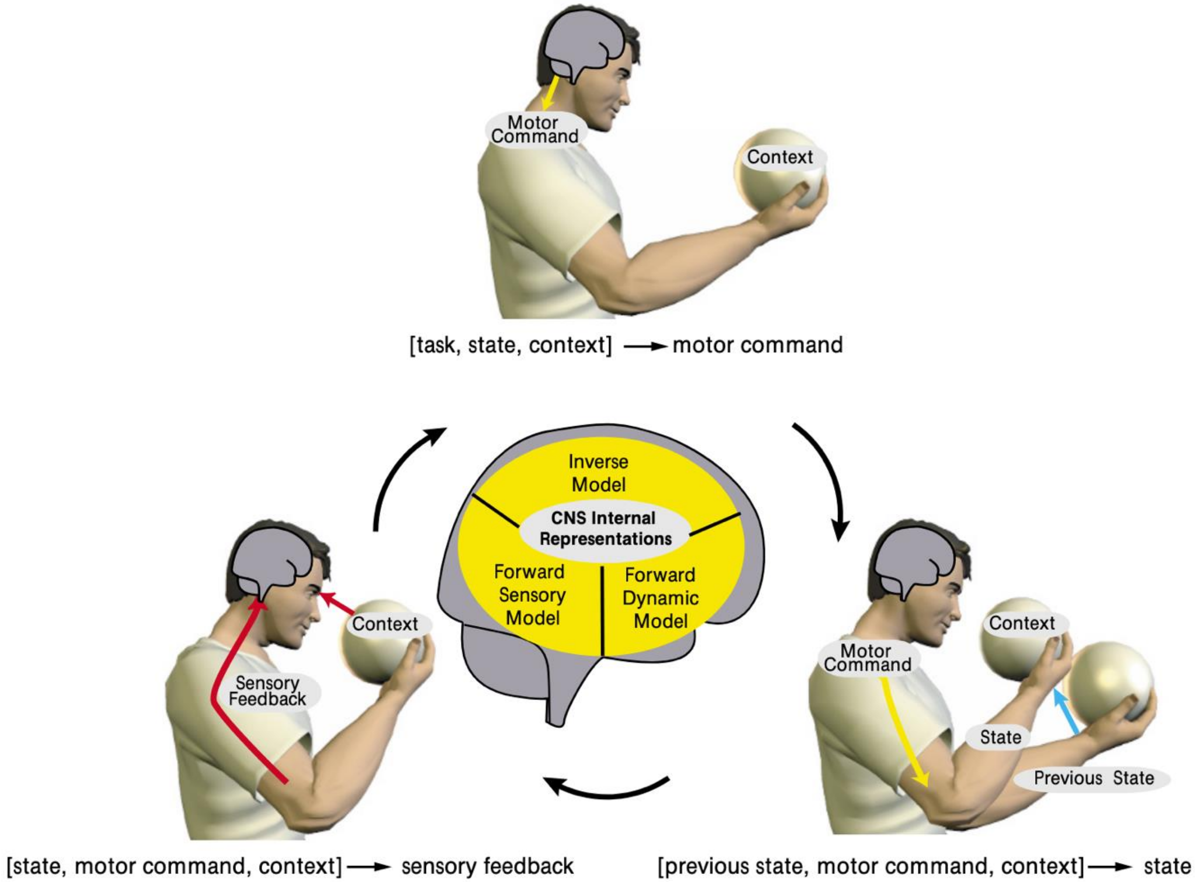
Example scenario: The task of robot R is to guide person A to person B without causing discomfort to any human present.
Dashed area shows example of robot laser range finder coverage.
Challenges: passing through the corridor in formation with person A, while avoiding incoming person E, and approaching person B without disturbing person D unnecessarily

T. Kruse et al., Human-aware robot navigation, Robotics and Autonomous Systems 2013.

Human-aware Navigation Concerns

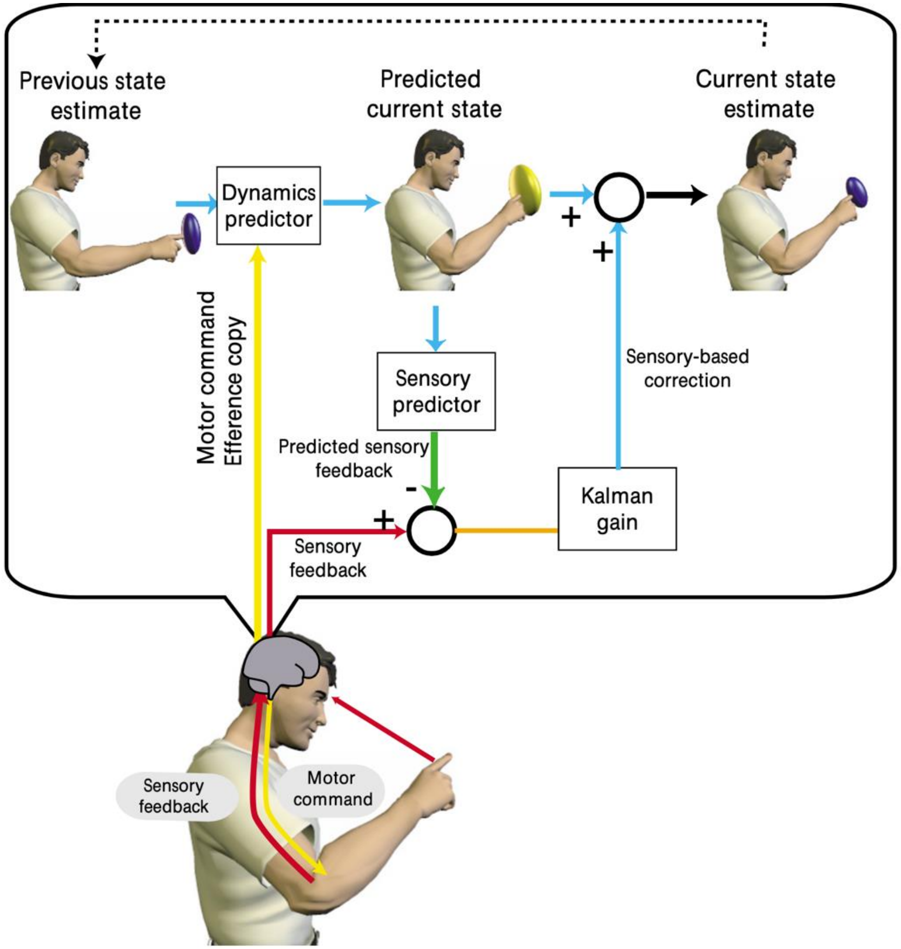
1. Comfort: absence of annoyance and stress for humans in interaction with robots
2. Naturalness: similarity between robots and humans in low-level behaviour patterns
3. Sociability: adherence to explicit high-level cultural conventions.

How to Model?



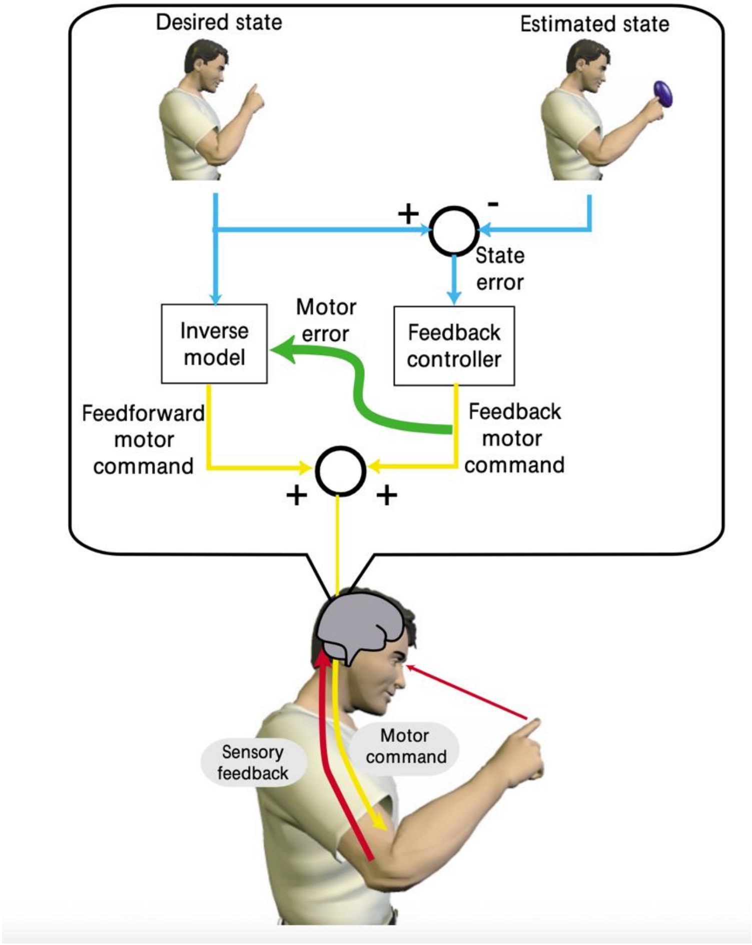
[Wolpert+Gahramani, *Nature Neuroscience* 2000]

A Process Model

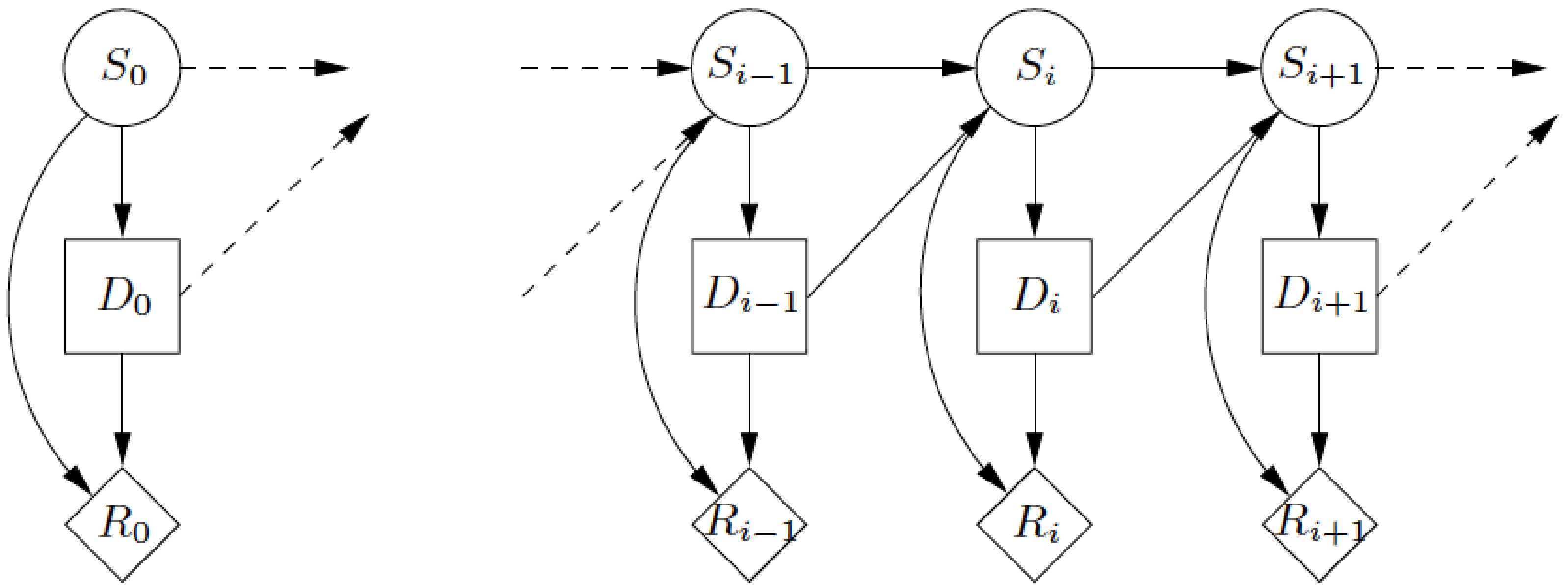


[Wolpert+Gahramani, *Nature Neuroscience* 2000]

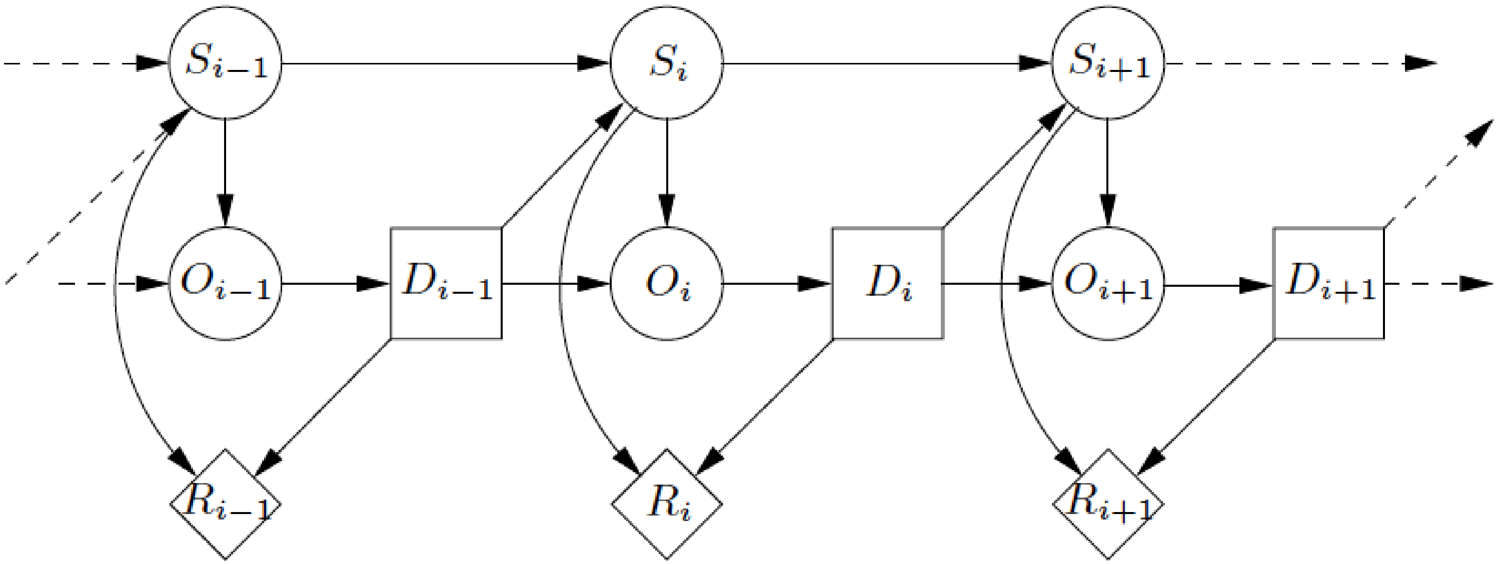
Crucial Concept: Internal Models



How to Model? Example Representation: MDP as an Influence Diagram

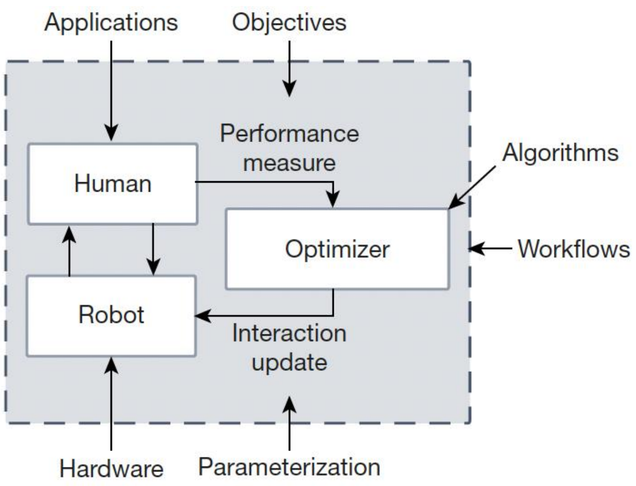


Partially Observable (PO) MDP as an Influence Diagram

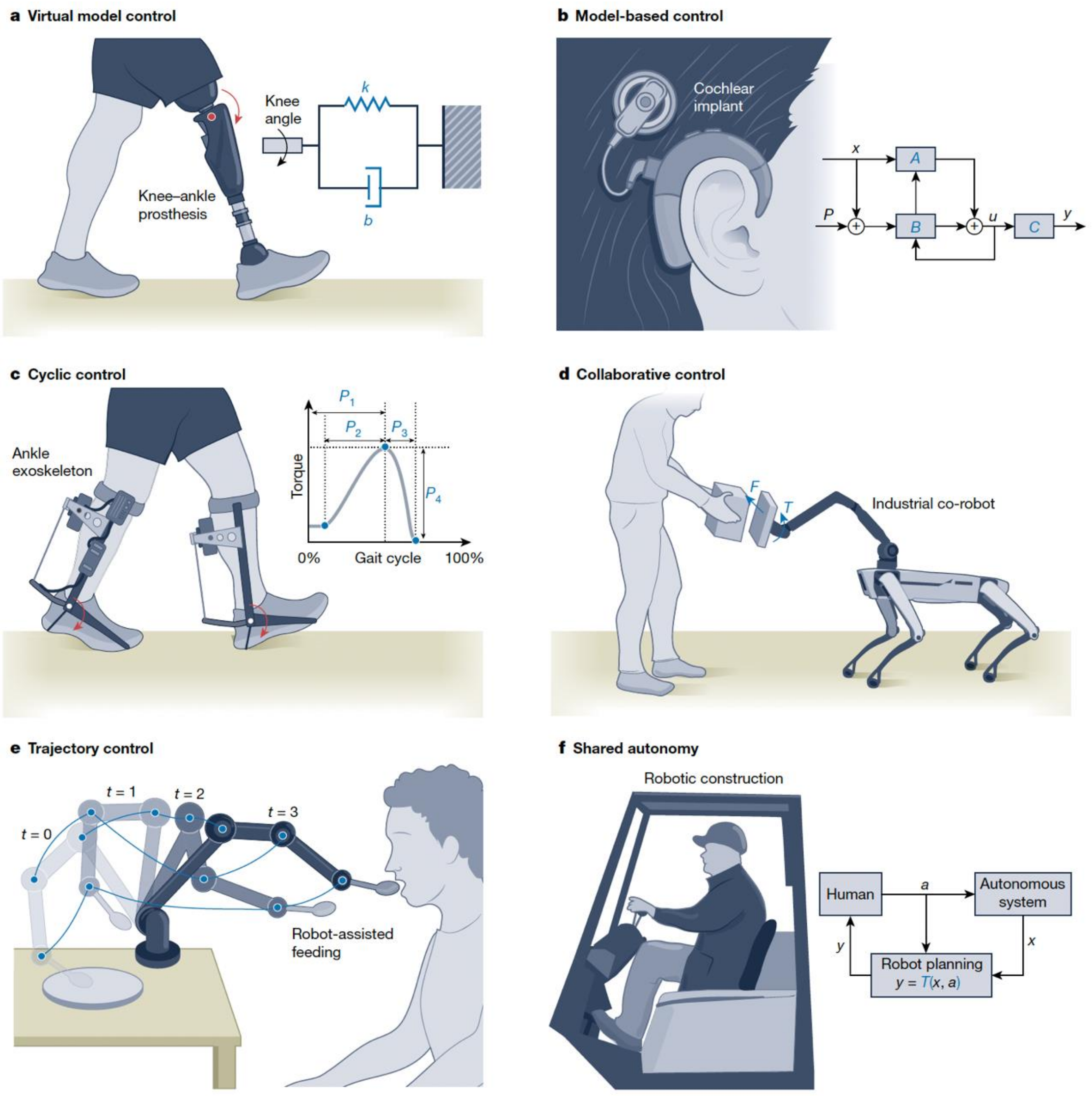


For a model structured as this one is, we may not want to proceed naively, by unrolling as a tree. Instead use, e.g., approximate inference & message passing.

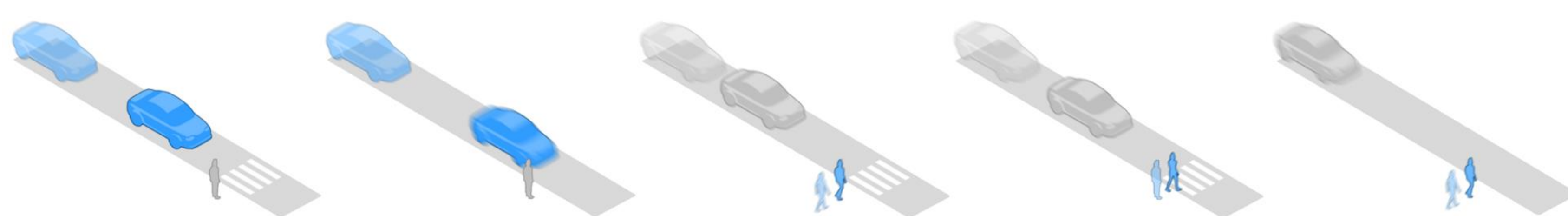
Concept: Human-in-the-loop Optimization

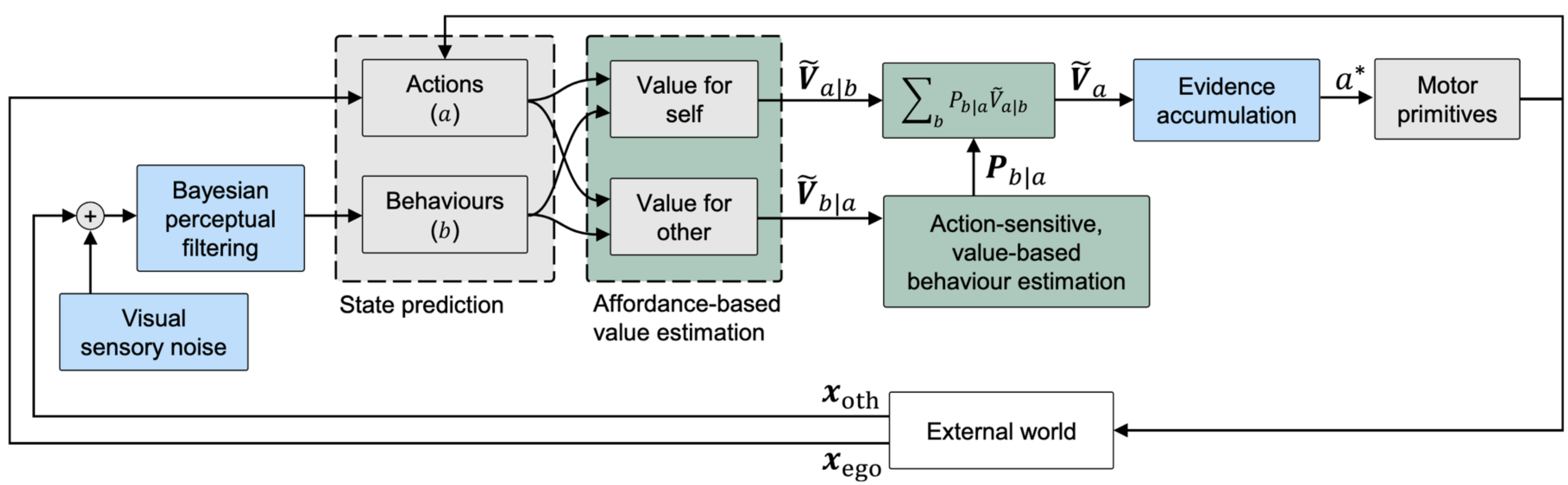


[Slade et al., *Nature* 26 Sep 2024]



Phenomenology in Some H-R Interactions

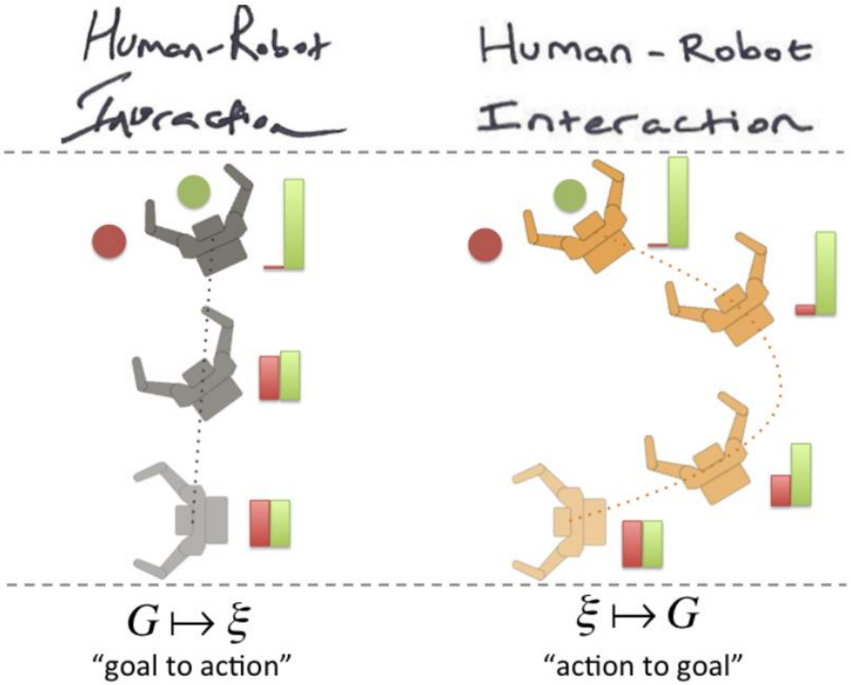


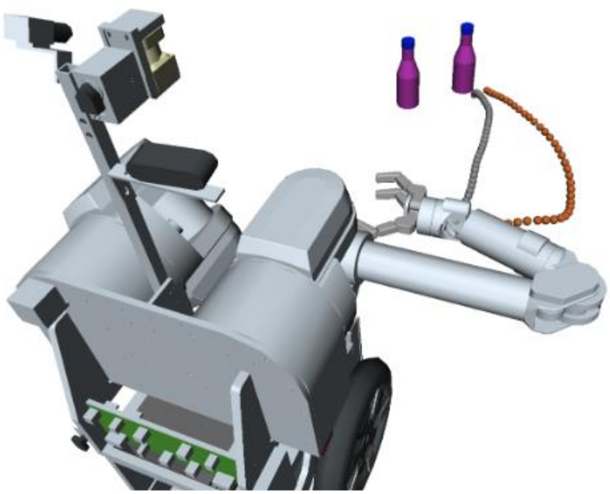


[Markkula et al., *PNAS Nexus* 2023]

Related Issues: Legibility and Predictability

Legible motion = motion that enables an observer to confidently infer the correct goal configuration G after observing only a snippet of the trajectory,





A. Dragan et al., Legibility and predictability of robot motion, HRI 2013

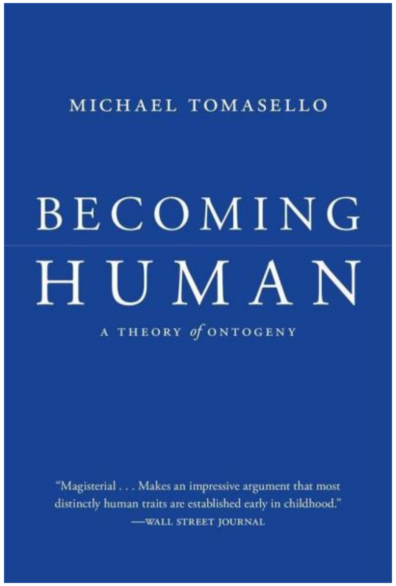
What is Even Harder to Model?

Infants are deeply and innately interactive:

https://www.youtube.com/watch?v=Z-eU5xZW7cU

How do we account for this?

Collaboration evo-develops; *teaching* via embodied communication is crucial…



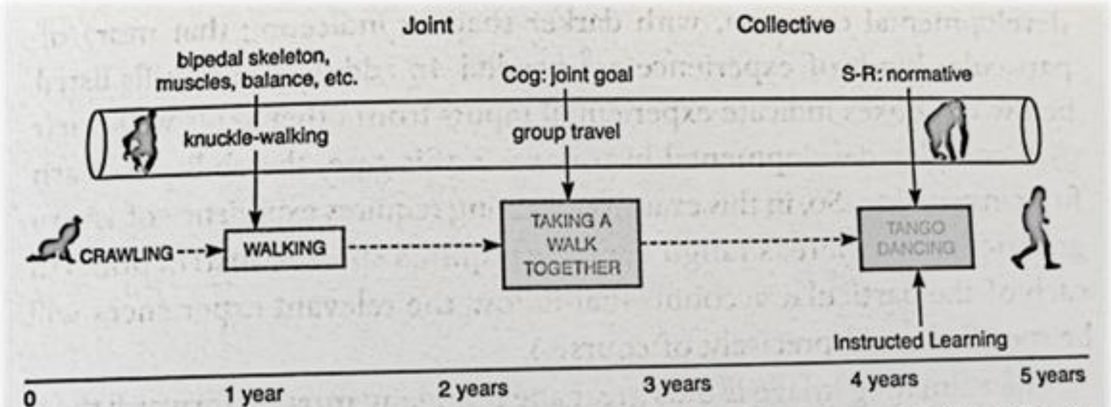
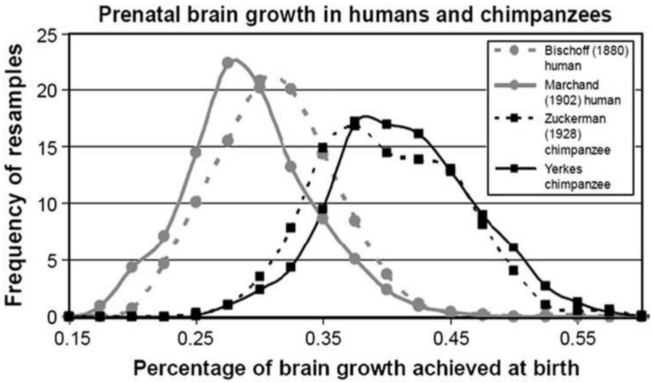
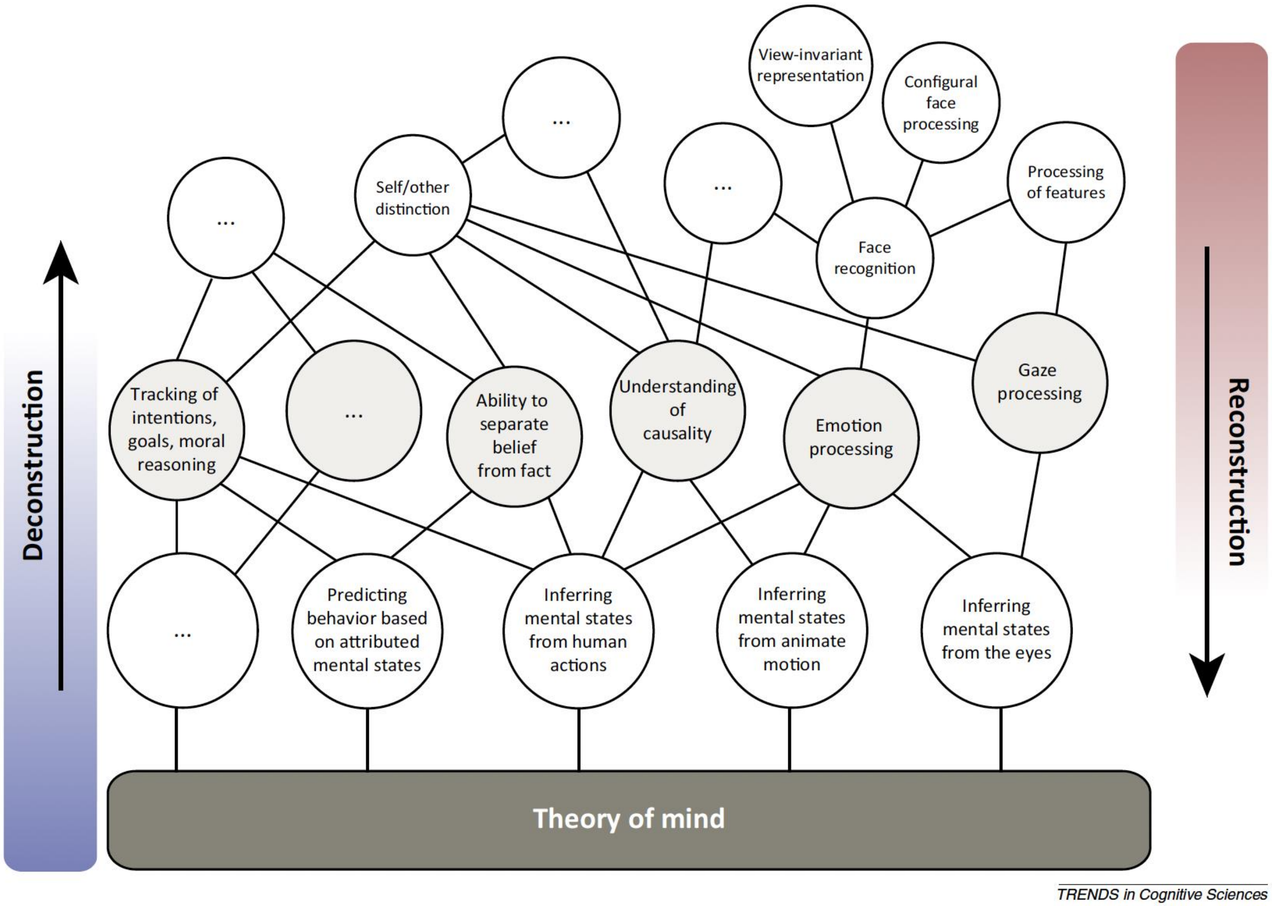


Figure 2.5 Illustration of a developmental diagram: the ontogenetic pathway for learning to dance the tango. Abbreviations: Cog = cognitive capacity; S-R = executive self-regulation.



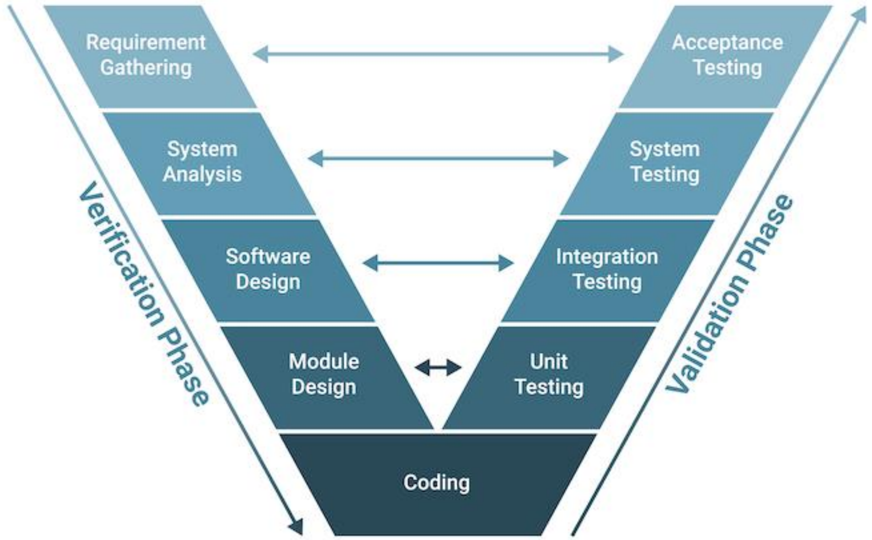
[De Silva et al., J. Human Evolution 51 (2006)]

… and builds on many inter-locking pieces

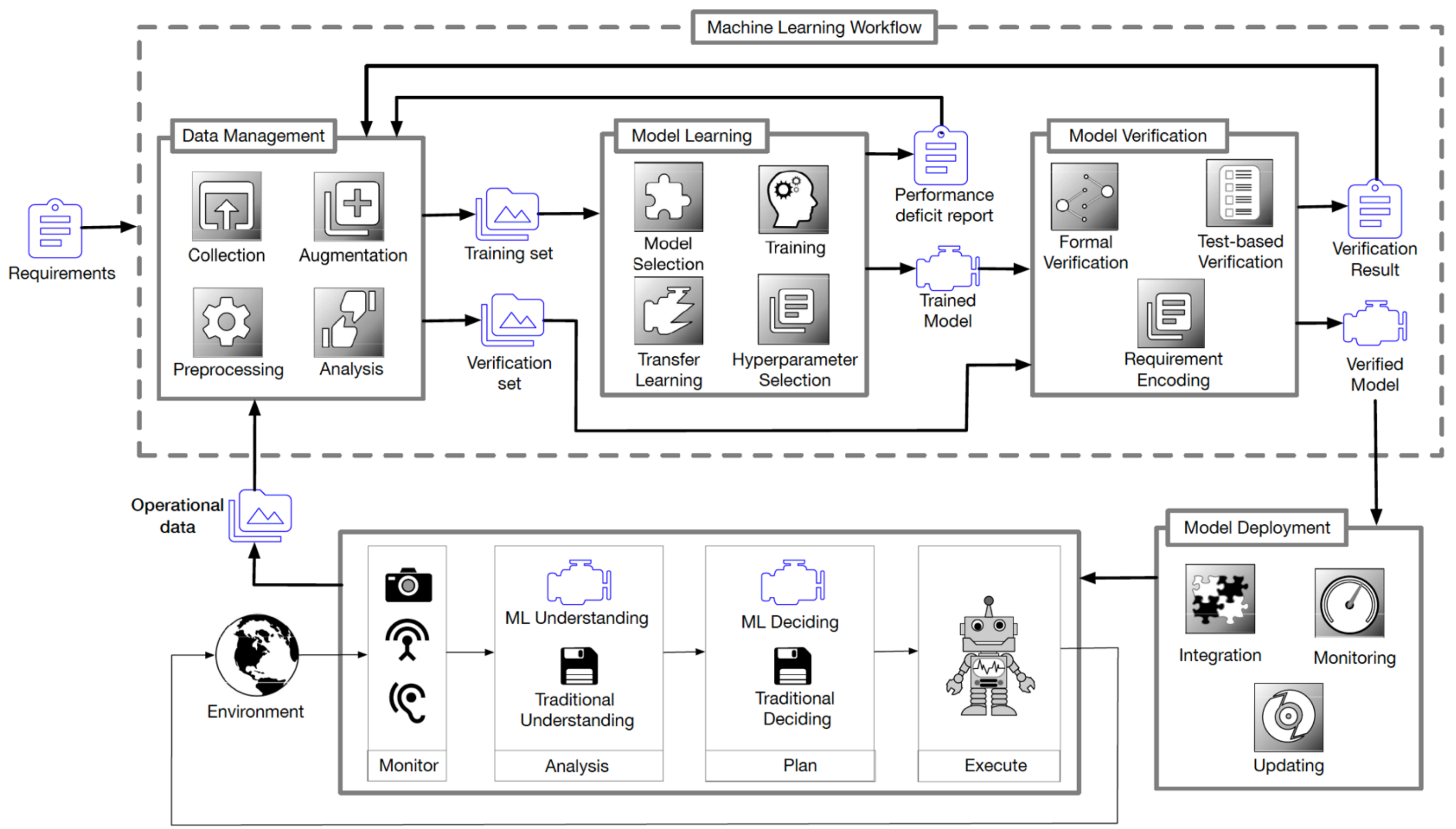


[Schaafsma et al. 2015]

Stepping Back:
Human interactions in Design Processes
- a variety of considerations



Where and when decisions are made: ML workflow



R. Ashmore, R. Calinescu, C. Paterson, Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *arXiv:1905.04223, 2019.*

Risk and Mitigation: Some issues

- These technical processes address safety and technical concerns around autonomy

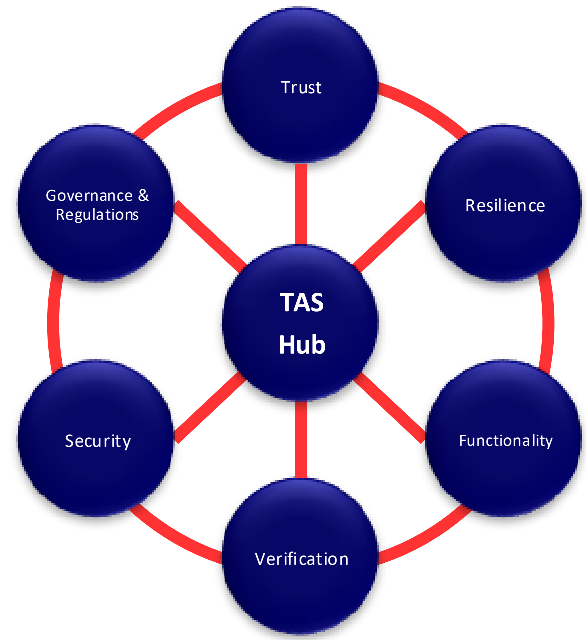
- Organisational and regulatory design also plays crucial role. Example, consider different approaches:
  - Type certification
  - Change control protocols for deployment of AI
  - Post-market surveillance and incident databases

Some Lessons from Our Recent Research: UKRI TAS Programme

The TAS Hub was funded as part of the Strategic Priorities Fund (SPF) which funds multi- and interdisciplinary research across 34 themes in response to strategic priorities and opportunities.

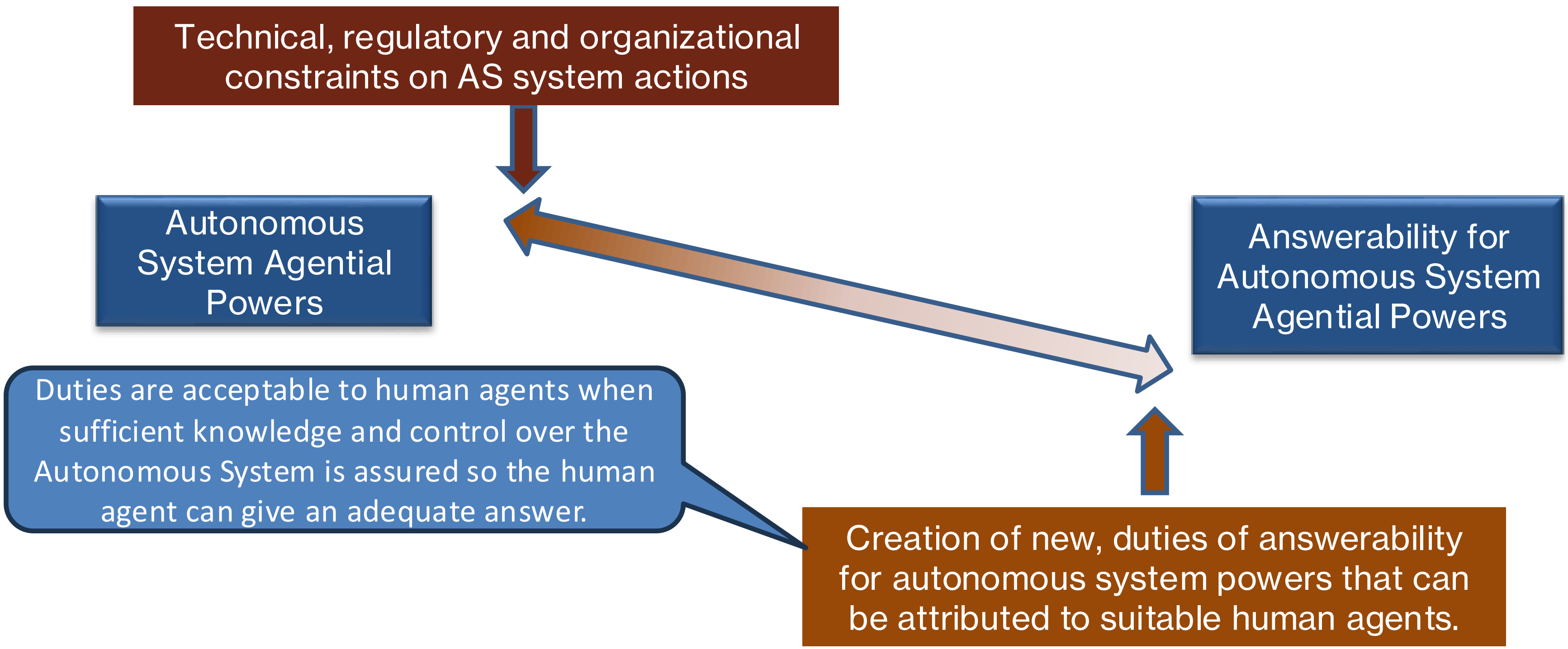
Total Funding	Universities	Industry Partners
£33m over 4 years	20+	100+
Funding	**Researchers**	**Disciplines**
Hub: £11.7m Nodes : £3m each	130+	10+

World's largest research programme in Trustworthy AI and Autonomous Systems



The following slides draw on work done within the UKRI Research Node on Trustworthy Autonomous Systems Governance and Regulation

Anticipate: Responsibility Framework: Balancing Powers and Answerability

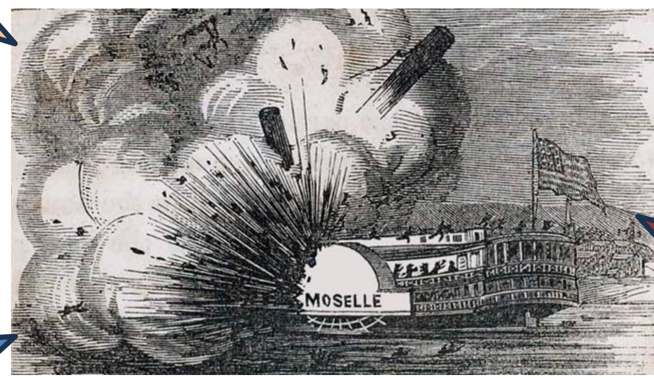


S. Vallor and B. Ganesh, Artificial intelligence and the imperative of responsibility: Reconceiving AI governance as social care, in *The Routledge Handbook of Philosophy of Responsibility*, Routledge, 2023.

Reflect: Use the historical record: Steamboat Regulation: Don't reinvent the wheel

Informational: Establish good information gathering and sharing to contribute to safe operation

Mechanical: devise generic technical mechanisms and processes that contribute to safe operation



Regulatory: Identify specific roles and duties with appropriate training and restrictions on entry to the roles

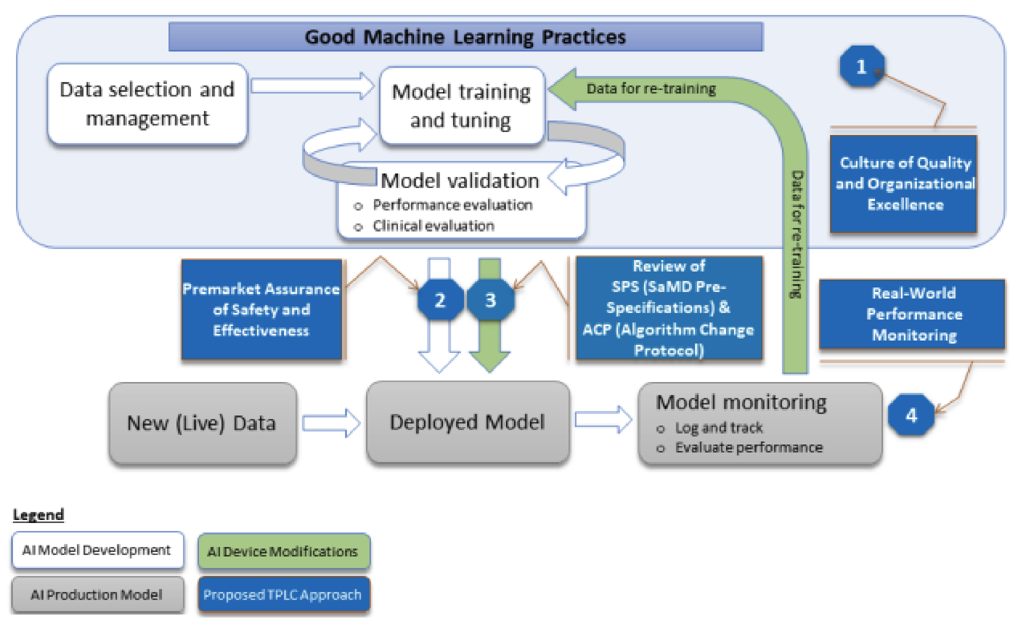
Liable: identify human agents that are identified as carrying the liability for accidents.

B. Ganesh, S. Anderson, S. Vallor, If It Ain't Broke Don't Fix It: Steamboat Accidents and their Lessons for AI Governance. *WeRobot* 2022.

Engage: Workshops with Stakeholders

- How to include stakeholders, e.g. MHRA, FDA, NHS, Vendors
- Workshop 1: Challenge of approval, evidential standards, UK/US alignment.
- Workshop 2: The EU dimension, MDR and the AI Act.
- Workshop 3: Operation, Post-market Surveillance, PCCP, ACP.

Overlay of FDA's TPLC Approach on AI/ML Workflow



Act: Tools to enable users to explore models