HRI

- Moral verification and validation challenges → derived from interaction with people. Important to ensure systems work from a human perspective
- Observability and controllability of complex useful systems → opaque and hard to maintain
- Emergent and dangerous behaviours from learning when interacting with people and the environment
- We need to understand loss and costs, to make safer systems
- Collaborating autonomy → people supervising autonomous systems. The operator can intervene to approve, the system suggests to the operators.
- Applying synthesis of behaviours → case + requirement in logic → composition and synthesis to get code. A challenge: how to include the human in the loop in this synthesis.
- In transport systems: rules for semi-controlled systems safety, mechanical safety is doable but expensive. Tomorrow → natural interactions are needed (between machines and people), contextual information is needed for each other's understanding.
- Autonomous vs. automatic systems → more difficult to define control and environment. How to ensure main users are qualified to use them when autonomous systems are meant to work with the elderly, the ill. Money will drive the verification of autonomous systems.
- Deterministic a priori verification is not applicable for systems that interact and learn in complex environments. Teaching systems what is good and bad.
- Layers for an autonomous system, and then verification of those layers (the agent or manager).
- We want systems that are useful for humans, and high assurances of safety too.
- Guidance on user interfaces → who is the user and what for? For evaluation and iterative design after testing. This loop is costly.
- System models → allow people to understand what the system is doing. Allow designing fault tolerant systems (they will go back to safe states). Allow analysing systems automatically to expose faults and problems.

Discussion

Historical progression of aircraft verification: huge literature in cognitive science on human behavior in avionic systems (human-machine interactions). This is why we have good behavior in aircraft (low accident record). It helps that aircraft are slow to respond.

Robotics are where avionics were when the Wright brothers. How is verification going to evolve for these systems? Exploiting the lessons for avionics in autonomous cars, although with a more difficult environment. Figuring out analogies between autonomous systems and other cognitive structures in engineering to be able to import. There will be lessons learned, as we don't know all

about autonomous systems. Avionics is quite conservative to change processes (e.g., they bought microprocessors that worked well, and will not change them), but this might be different for autonomous systems.

It is necessary to define what is useful, the behaviours need to be formalized, to ask questions to the machines. There is a need for trust and system understanding at the beginning. While in operation, people care that the system performs its task only, and not how or why.

Different types of humans to consider in the interactions: users as part of the decisions, pilots in aircraft that will monitor and try to understand the machines to work with them; users that just want to get the system to work for them and they do not care about what they do; people that try to take the systems to their limits to see what they can do. Users in the same system (e.g. autonomous car) can take different roles according to the circumstances. The robots might force the human into these different roles. We need to specify the role of the human and the expected behaviours.

Defining “dangerous” is difficult. Definitions are an iterative process for autonomous systems. It is important to consider all these different groups of people when designing systems, for robustness of autonomous systems.

Communication systems are important in autonomous systems that interact with people, for security and safety reasons. Important to consider communication in design for safety.

Ethical design for robots: useful but not restrictive systems. It is difficult to make systems that everybody will accept or will be able to deal with everybody (e.g., robots taking care of people with dementia, or drunk people). You might not be able to design robots that work for every type of people, but you need to substantiate these design decisions. Ethicists are needed in the conversations about specifications and modelling for designing safe systems, as some systems (e.g. autonomous cars) will need to make decisions that will endanger humans.

Is it possible to make something that will be accepted by people? Who is going to decide how autonomous systems are going to decide? Autonomous cars will be there, regardless.

Some verification is needed in the design process. But this needs to consider different categories of users and conflicts and contexts.

Also, it is possible to then put runtime monitoring to check the assumptions about the people and environment, to adjust the system. This process is challenging, starting from how to specify human behaviours in a formal way.

Liability question → when to take over a different role in the system, how to control human-robot interactions. In the security domain, people are classified according to the information they can access. These roles might need to be dynamic, but it is better to anticipate as much as it is possible. Flexible systems vs general systems.

Human needs mental model of the robot, also robot's human mental model is needed. These models are not well defined now. Models of reasoning need to be explained to the people too for accountability.

Flexible autonomy or sliding autonomy → the transition between autonomy and human overriding it needs to be studied better, e.g. in medical robots. Safety barriers are currently in place in medical

robots, to prevent the human to make mistakes. Problems ensue when the human needs to violate these barriers as part of the medical procedure, using the robot. In the transition, it is needed to communicate clearly who is doing what, which is difficult at high speed (different from the autopilot). Simplifying the interfaces might be risky.