Intelligence and Decision Support

Report of the Discussions of Breakout Session

Authors: Christoph A. Thieme, Marilia A. Ramos

Autonomous systems and Artificial Intelligence

Autonomous systems are cyber physical systems. They may include Artificial Intelligence (AI), e.g., in decision making or other autonomous processes. However, although an overlap exists, an autonomous system is not a subset of AI, nor vice versa. Reasoning is an important aspect of autonomous systems and AI. Reasoning refers to the ability to explain actions and decisions. For humans, many actions are learned intuitively and do not result from reasoning. Still, the reasons for these actions can be described in retrospect, even though it was based on intuition.

AI needs to be interpreted in a simplified manner than what is currently expected in the public opinion from AI. AI is a collection of mathematical methods, helpful for solving tasks associated with intelligence. AI methods try to find regularities in sets of data. An operational perspective may help to better clarify the concept: AI is mainly comprised of some form of learning, some degree of reasoning, interaction with an environment. It should have an ability to explain how or why the AI made its internal decisions.

The definition that "AI is always the thing that humans can do and machines cannot do" is not suitable and in itself unachievable. However, AI has advantages over humans, for example, analyzing quickly large sets of data. Therefore, the human standard might not be ideal. AI may be used to learn the operational parameters for an autonomous system, identify weights for risk factors, or detect abnormalities.

The focus of AI should lie on the autonomous system, meaning that AI methods comprise tools that may help to realize autonomous systems. Autonomous systems are more than AI, since they comprise hardware and other software. By excluding the physical systems from an autonomous system and reducing it to AI, the extended Turing test is not achievable, i.e., one is unable to detect different behavior of AI and humans.

Adaptive autonomy is an often-used term in the context of autonomy and AI. However, adaptive autonomy is an ambiguous term. It may refer to a system that uses a learning (AI) system, to a system that changes the degree of autonomy during an operation, to software updates that adapt the system when needed, or

to the behavior that occurs adapting to a situation. Commonly, self-adaptive systems change their behavior based on experience.

Risk in control and decision making

For intelligent autonomous systems it is required for risk to be considered during the early design phases. Decisions of an autonomous systems need to be based on an implementation of risk considerations that are defined clearly mathematically and operationally. Risk is often defined in probabilistic terms, in a pseudo mathematical equation: risk equals probability or frequency times consequences. Successful implementations of (quantitative) risk assessments (QRA) in applied projects on autonomous systems should be developed to highlight the advantages of QRA.

AI may be seen as a factor contributing to risk. However, it is generally the complexity of the system from which risks emerge. Hence, it is important to understand the system, not only the AI. Systems might become so complex that nobody can see the full picture. Therefore, it may be impossible to understand the associated risks and failures that may occur. Methods and approaches are needed to manage and assess complex systems.

There is no essential difference between decision-making and optimization based on parameters. Autonomous systems and AI algorithms make continues choices between a spectrum of operational parameters. Making only discrete choices is not a property of an autonomous system, besides decision making; parameters are optimized to achieve the most efficient execution under the given circumstances. This resembles the behavior of, e.g., human drivers that follow a set of rules and optimizes constantly the vehicle speed and heading, and their own behavior to avoid accidents and penalties.

Therefore, risk is a cost and a constraint for operation of an autonomous system. Different types of risks emerge for autonomous systems that are not possible to be covered by only one measure to measure the risk level of operation. Identification of relevant risk measures and risk factors is one of the main tasks during development. These need to be implemented in the control system and the decision module of the system.

During the development of an autonomous system, a baseline performance needs to be defined, as reference for acceptable performance and risk. The baseline performance should not be lower than the performance by a human operator. One challenge that arises when approaching performance requirements is that the evaluation of the human performance is difficult. The performance acceptance criteria may be vague. Perceived risk versus real risk is

also relevant with respect to this evaluation. For example, car crashes occur frequently, while an autonomous vehicle accident is paid much more attention and is perceived as more severe by the public.

Risk reduction across the system architecture

Several definitions, views and hierarchies exist for control systems. The system architecture depends on system purpose, size and complexity. A general guideline is that low levels of control are reactive, while higher levels of control are more proactive. Higher levels may include a proactive planning layer and a supervisory layer for fault detection. Using the term "executive layer" may not be useful, since everything is executed.

In each of these three suggested layers, different risk measures may be used. This depends on the layer purpose and the anticipated level of autonomy. Several risk measures may apply, e.g., probability of failure and probability of collision, mission failure, system failure. Using one measure across all levels is not sufficient. Current systems do not include explicit risk models in their control structure.

One challenge for the supervisory layer is the identification of the fault source when a fault is detected. Subsystem integration between components and systems may be inadequate and detected faults may be propagated from the real source. A clear structure and hierarchy are needed to filter faults and identify their sources.

The risk that emerges during an operation may be reduced prior to operation and during operation (post-deployment). However, risk reduction should be mainly achieved during the design and pre-deployment phases. Risk reduction may be achieved in all architectural levels.

In the post-deployment phases, the risk level and the system condition need to be monitored. The autonomous system should detect critical and pre-critical situations. It should use pre-critical situations to avoid critical situations in all equivalent systems. For identification of such situations, the autonomous system may use statistics or other machine learning (ML) approaches.

Development of safe autonomous systems

Industry practice shows that risk assessments and modeling are necessary processes in the development of autonomous systems. Using risk-informed decisions enables better design decisions. Through integration of the risk information, it is possible to identify opportunities for monitoring and prognosis of failure development. This in consequence will reduce the need for unnecessary maintenance. ML algorithms may be one tool to monitor the system. There are

"best practices" in the software industry for testing, validation and verification. However, there is not a general recipe for future developments. Simple, "if-then-else" structures, can be proven to work correctly and reliable. For ML and AI learning techniques, these methods are not available yet.

Risk models need to reflect the assumptions made in the system design. During design, these assumptions need to be identified and it must be consequently assessed how the level of risk may be affected by these assumptions. So-called legacy systems, systems that build on former generations, build on certain inherited assumptions. However, it is often undocumented why the assumptions were made. In a few cases, it is assessed if the assumptions are still reasonable.

Currently, airline pilots report anomalies based on previous experience and training. A system should be required to self-report data that can be used for further development and improvement. Near misses are a significant learning source for autonomous systems and AI. They provide more insights than just the accidents themselves.

Self-adaptive systems must be designed to detect if the adaptive behavior is performing worse than the previous learned behavior. Mechanisms need to be in place to return to proven and safe behavior in such cases. The most safety-critical parts of the system should not build on adaptive methods. A predictable and verifiable behavior is required. It is necessary to define what changes need to be verified from the outside and which can be done based on learning.

A hierarchal structure regarding the regulation of autonomous systems is required, analogous to the regulatory framework for current human operators. Since autonomous systems will become a reality relatively soon, the regulations need to be put in place. However, there needs to be room for future improvement and adaptation. The autonomous systems will not appear abruptly, and systems will change incremental. Certification for drones, for example, has requirements in place, to be commercially viable already. For consumer drones such rules are, for example, the inability to fly into no-fly zones, etc. Newly emerging companies, working on autonomous systems may be less conservative than the established companies. Hence, regulations are needed to create a common baseline.

Risk awareness in autonomous systems

Decision making

Improved intelligence and online decision-making capabilities are needed in autonomous systems. Existing control theoretic approaches are not explicitly connected to risk assessment and modelling. Some control strategies use

methods that deal with constraints and unwanted states. This leads to robust control but tends to be conservative. A clear risk definition is needed for that purpose. Risk consideration should also include events that are not known. Simple control strategies and models cannot include such considerations. There exist few control strategies for handling extreme cases with low probabilities.

There is another gap between control theory and control practice today. Switching between discrete states is used to adapt to certain situations. There is a lack of usage of established control strategies in practice. Proactive approaches are required: Actions in case something might happen, being ready for "black swans", i.e., rare but extreme incidents. In contrast, a reactive approach would imply to act when something is happening. In any case, there is a difference between safe behavior and safe state. In certain situations, it may not be practical or safe to go into a safe state, e.g., shut down the system, or stop it. A safe alternative needs to be designed and chosen.

Model predictive control (MPC) is one control strategy that is suitable for autonomous systems. However, using only one risk measure in such a control strategy is not suitable. A vector of several risks is needed to optimally use the method; these may be probability of collision, time to collision, etc. Risk may be then a cost and a constraint in the MPC algorithm. Using risk just as an optimization criterion for minimization would lead to the system never starting, since then the risk is lowest. In addition, using the risk as a constraint enables the user to demonstrate that the system will not accept a higher risk than prescribed by a legislator. In the MPC method, this may have the disadvantage, that the system will always choose a solution close to the accepted risk limit.

Online risk models may assist in decision-making. Online risk models are models that have been developed before the mission is executed and that use data measured online to constantly update the current risk level. Necessary data measurements can be identified in risk analysis. It may be possible to sample the measures directly or it may be necessary to use risk indicators. ML may be used to tune the different risk factors and other objectives to give the behavior we want. Game theory may be useful, too. It must be taken into consideration how other entities involved might act.

An intelligent system must not only follow the rules and trust that other traffic participants do the same. An autonomous system must be able to detect or predict intentions. A good example is the maritime sector, where COLREG rules exist. However, human navigators may violate these. Initially in the aviation traffic collision avoidance system (TCAS), for example, only positions were communicated. This was not an intelligent system, since it only detected other planes, but did not coordinate maneuvers with each other. After serious

accidents, the rules of behavior had to be adapted and the system is now more prescriptive and solving traffic situations automatically.

Health monitoring

The performance of an autonomous system also affects the decision possibilities. Hence, it is necessary for an autonomous system to be aware of its health status. Parameters that define the health status are the conditions of sensors, actuators and the control system performance. In addition, mission external parameters, such as information on maintenance and available spare parts in the operation basis may affect the decision possibilities for an autonomous system.

Two different time horizons need to be taken into account with respect to health monitoring: the long time perspective gives information on degradation of components the overall system's condition that may be used for service planning, e.g., changing parts, and general maintenance. The short time perspective provides information on the system's performance degradation, its effect on the mission outcome, and the ability to handle possibly critical situations.

For a system to detect that its performance is degrading, it needs to be designed knowing the baseline performance. Risk assessments are essentially identifying what types of situations the system cannot deal with. Therefore, risk assessments aid to identify performance requirements to the design of the systems and the operational design limitations. The system is then limited to function properly in situations the designer managed to envision. Hence, the system also needs to be able to detect that it is operating outside the operational design envelope.

Input for this type of behavior needs to be supplied in manuals for sensing equipment. Similar to the commonly found curves "efficiency vs. environmental parameters", the measurement uncertainty could be described by the behavior over a combination of environmental parameters. However, a device may not be tested in all operational conditions. Then the reliability data needs to be produced by the user, e. g., NASA is producing reliability information for most of the components themselves, such as charge and discharge curves of batteries under extreme temperature conditions.

Sensors may be subject not only to physical degradation, but also to snow, fog, dust, or alike. This may inhibit the performance. In addition to monitoring the physical condition of the system, information needs to be combined and the reasons for degraded performance need to be detected. AI methods may assist to monitor the system health and detect the causes to a degraded component.

Sensor requirements

Sensors need to be reliable during an operation. The environment influences the uncertainty of measurements. Components degrade, which increases the uncertainty. An autonomous system needs to handle these facts by sensor redundancy, better sensors, etc. However, redundancy adds to costs, increases power consumption, and adds weight. One possible solution could be using the payload sensors as redundant sensors (e.g., using visual flow to validate accelerometer measurements). The ability to handle uncertain situations is also needed when facing sensor degradation.

An autonomous system is not only about sensors, but it must be able to comprehend the meaning of the sensor measurements. Many systems can detect if the weather and climate conditions exceed the design limitation.

Data requirements for safety and reliability analysis and safety monitoring

Data and information from the sensors should be reliable and available when needed. In addition to pure measurements, sensors should give information on the sensors' uncertainty of the measurement. In this way, the uncertainty and the effect on the system may be assessed mathematically.

It is known that navigation systems are prone to both noise and design flaws, which may be undetected until their effect is experienced. However, when the system is deployed, it may be too late to correct the error. Hence, an appropriate design process needs to be chosen, to ensure that necessary data is collected in the appropriate frequency and quality.

One approach may be to use information trees, which are similar to fault trees. The challenge with such a tree structure is the interpretation of the Boolean logic. The trees can be used to identify:

- What is the information that needs to be gathered (the top event)?
- What needs to be measured based on what the needed information?
- What types of data and sensors are needed to meet the knowledge condition in the top event?
- Which data types are dependent or independent?
- What are the success metrics?
- Where are the best places to collect information?

Internal and external data uncertainty

Three types of uncertainty can be differentiated:

1. Measurement or data uncertainty
2. Model uncertainty
3. Interaction uncertainties

Measurement uncertainties are well defined and inherent to the measurement system and method. Methods for describing this uncertainty are well established, e.g., Gaussian distributions. The uncertainty can be expressed numerically. Its effect can be propagated through the system and the effect can be assessed. Sensors should give information on the certainty of the measurements.

Model uncertainty reflects the completeness and correctness underlying the models that are used in a system. Statistical distributions may not be able to capture this type of uncertainty and some parameters that are used in a model may be highly uncertain, e.g., turbulence is difficult to capture numerically. Assumptions need to be made that are imperfect. Model uncertainty may be introduced to keep the system simple. Adding many parameters, whose effect is uncertain, will not improve the model. Hence, parameters may be neglected, if the effect is highly uncertain or negligible, in order to make the system more efficient.

The third type of uncertainty is the uncertainty with respect to the interaction with other parties, humans or manually operated/autonomous systems. The behavior of others is difficult to predict. An autonomous system may be “perfect” in itself. However, other traffic participants may cause an accident.

The system is a conservative system if the estimated uncertainty exceeds the real uncertainty. Unsafe behavior is to be expected if the estimated uncertainty is lower than the real uncertainty in a given situation. Risk analysis is required to estimate the uncertainties with respect to the control environment. The analysis needs to include the operators or supervisors and the autonomous system themselves. The state of the operator needs to be reflected in the control system.

With respect to the third type of uncertainty, one challenge is to robustly detect and identify obstacles and other participants. For AI methods, such ML and deep learning, it is difficult to predict their output, due to their prediction accuracy and often in tracible behavior. Both, identification and prediction are time dependent. Small variations in timing may affect the predictions. A test approach may require a very low uncertainty level, which will correspondingly take a lot of time. Without a verifiable equation, it is difficult to quantify this

uncertainty. There are methods for handling noise, i.e., environmental disturbances, but the theory is lacking when it comes to working with probability density functions.

An approach to verification should be to test first the algorithms, e.g., through simulations. These tests, then need to be validated in the real operational environment. There is need for clear guidelines and checklists for building an autonomous system.

If an operator or supervisor is involved in an autonomous operation, it may be necessary to monitor the operator and assess the uncertainty with respect to the operator's ability to cope with a situation. Information could be extracted from the performance during the current task and projected on the execution of the next task. The visualization of uncertainty to the supervisor/ operator remains one challenge.

A core demand for AI-based systems is that they need to be able to detect if they are outside of their operating range. The system needs to detect if the uncertainty in a given situation is too high. This includes the detection of anomalies that were not included in the training data sets and the appropriate reaction to these. This can be compared to a human driver who will adapt to a new situation and identify untrained situations.

Autonomous systems´ interaction with the operator

Autonomy in many cases shall reduce the number of operators needed to run a system safely. There are benefits to be realized with higher levels of autonomy. A system does not need to be fully autonomous to be cost effective. There are different areas or operational time intervals, where it is better to be more autonomous, e.g., for the maritime industry at open sea. In any case, there will be an interaction with humans even for fully or highly autonomous systems.

In many planned autonomous systems, the operator takes a supervisory role. The operator is used as backup to cope with situations that the system may not be able to handle. When such a situation is detected, the level of autonomy can be changed. It is critical that the communication between autonomous system and pilot is adequate. Information needs to be presented clearly and comprehensively.

The operator needs to know the state of the vehicle when receiving control. There should be a smooth transition between autonomous piloting and human piloting. It is important to identify the necessary information for the operator to carry out the necessary actions. The system needs to be designed accordingly. Recent accidents in the aviation industry show that pilots need to be

trained sufficiently in order to not fight against the autonomous systems. It needs to be defined what is part of the "autonomy" and what is the human's role.

The state of the human operator must be taken into consideration when attempting to give control. The workload for the operator may increase and it may be safer to continue autonomously, as there might not be enough time for the human to react or if the human may be unable to react. For an unmanned aircraft, it is better to use the autonomous system when taking off and landing, because of the increased stress levels and reduced perception capabilities of the operators during these tasks.

One concern is that the human operator suffers from skill degradation over time without continuous training. The operator may also suffer from a low workload and decrease of situational awareness. Similarly, one aspect that needs to be assessed in design is the confusion by sudden error messages to the operator, so called automation etiquette. The design of warning and handover messages needs to be clear. The autonomous system cannot just be stopped in the middle of the operation, e.g., this may create hazards for other participants. It is critical that the autonomous system relies on the operator when operating outside the design envelope rather than in a predefined set of situations that may be actually manageable by the autonomous system.

Industries that are currently attempting to automate their systems and products, such as, the automobile and maritime, must learn from aviation. Especially, skill degradation is widely researched in this field. Just assuming that the human is a good backup when the autonomous system reaches its operational limitations, is not viable.

Autonomous systems interaction with each other/ other systems

To improve cooperation and the predictability of the behavior of autonomous systems communication of planned actions is needed. Consequently, communication standards are necessary to be developed. In the future, an autonomous system might communicate with an infrastructure to get high-resolution maps or similar information about the area or attain feed forward information from a non-autonomous agent. It may enable people and other systems to better understand the current state than just by looking at the current behavior. Communication may also reduce time-delays, which is especially relevant for slow responding systems, such as ships.

Non-autonomous systems may benefit from using the information on the future actions and intentions of an autonomous system. For example, in the

maritime sector, information on pilots´ actions (rotating the steering wheel) could be fed forwarded to the autonomous system and communicated to other systems nearby, such that these do not need to detect that the ship is turning.

Group Participants

Adrian Arjonilla
UAS Consulting, USA

Arne Ulrik Bindingsbø
Equinor, Norway

Edmund Brekke
Department of Engineering
Cybernetics, NTNU, Norway

Ole Jakob Mengshoel
Department of Computer Science,
NTNU, Norway

Rudolf Mester
Norwegian AI Lab, NTNU, Norway

Sebastien Gros
Department of Engineering
Cybernetics, NTNU, Norway

Simon Blindheim
Department of Engineering
Cybernetics, NTNU, Norway

Sverre Rothmund
Department of Engineering
Cybernetics, NTNU, Norway

Sverre Torben
Rolls Royce Marine, Norway

Tarannom Parhizkar
Department of Marine Technology,
NTNU, Norway

Tom Mace
UAS Consulting, USA

Tor Arne Johansen
Department of Engineering
Cybernetics, NTNU, Norway