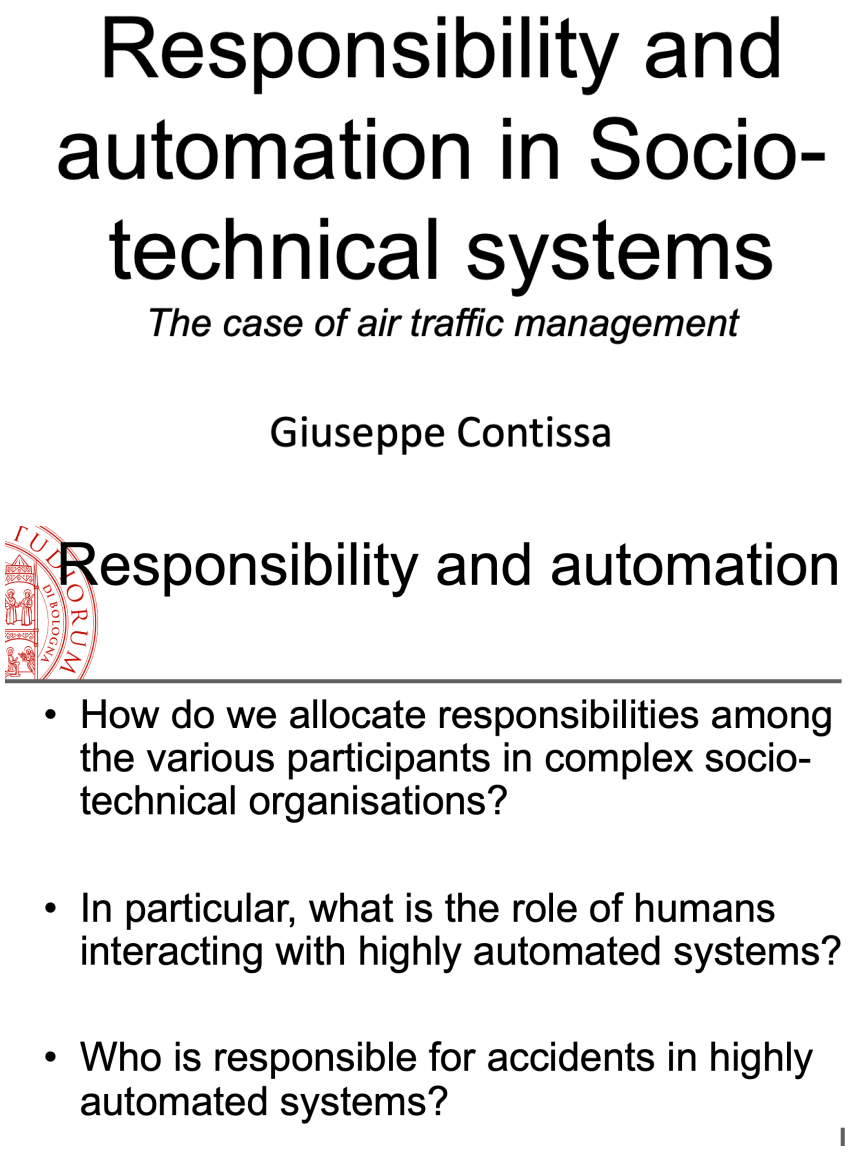
Lesson 3_Responsibility and automation in Socio-Technical systems



These are the questions behind this presentation.

We are interested in the topic of responsibility because we, as a society, can make choices, also in the design of technologies.

“responsibility”

As captain of the ship, X was **responsible** for the safety of his passengers and crew. But on his last voyage he got drunk every night and was **responsible** for the loss of the ship with all aboard.

It was rumoured that he was insane, but the doctors considered that he was **responsible** for his actions. Through out the voyage he behaved quite **irresponsibly**, and various incidents in his career showed that he was not a **responsible** person.

He always maintained that the exceptional winter storms were **responsible** for the loss of the ship, but in the legal proceedings brought against him he was found criminally **responsible** for his negligent conduct, and in separate civil proceedings he was held legally **responsible** for the loss of life and property.

He is still alive and he is morally **responsible** for the deaths of many women and children.

(Hart,H.L.A.,Punishment and Responsibility: Essays in the Philosophy of Law, 1970)

3

Hart is a famous person that with this story tries to summarize the only the possible sense in which we can use the term **responsibility**.

Here we have various ways to understand the idea of responsibility, but they have different sense.

Some of them are **active sense**: when talking about of being a responsible person we think of a person that in carrying out the task associated to his role is behaving in a responsible way.

The we have also some **passive idea** of what responsibility is: when we talk about of blaming someone for being morally responsible or legally responsible (i.e. being subject of the legal consequences of his/her action) or the think of responsibility as accountability in the sense of to be the person that is asked to give an explanation or is asked to explain what happens (this is also linked to the etymology of the term responsibility)

Different senses of “responsibility”

- **Task-responsibility.** An agent x is task-responsible for an outcome O, when x, given his role or task, has the duty to ensure that O is achieved.
- **Aretaic-responsibility.** An agent x is an aretaically-responsible agent of a certain type, if x devotes the required care to the task for which he is task-responsible.
- **Causal-responsibility.** An entity or event x is causally responsible for a harmful event H, if x has caused H. For instance a hurricane can be causally responsible for the delay of an airplane, as a controller can be causally responsible for an accident.
- **Accountability-responsibility.** An agent x is accountable for a harmful event H, if, under given x’s position, x may be requested to explain the happening of H, and may be possibly (if his explanation is inadequate to exclude blame/liability) be subject to the moral-socio-legal consequences related to H.

Other notes:

- **Task responsibility:** In the context of **complex systems** one important way to talk about responsibility is the **executional task:** we say that a person is responsible for the execution of a particular task, and is associated also to a particular role associated to the position that the person has inside the organization. For example we usually associate to pilots a certain task that should be carried out to make sure that the fligth will be conducted in a safe and efficient way.
- **Aretaic-responsibility** in the context of virtues behaviour.
- **Causal responsibility:** in the example of the story is the fact that the commander consider the responsibility for the ship disaster related to the storm, not to himself.
- **Accountability responsibility:** when we mean that someone has the role to explain what happened.

All these definitions of responsibility can overlap in some cases.

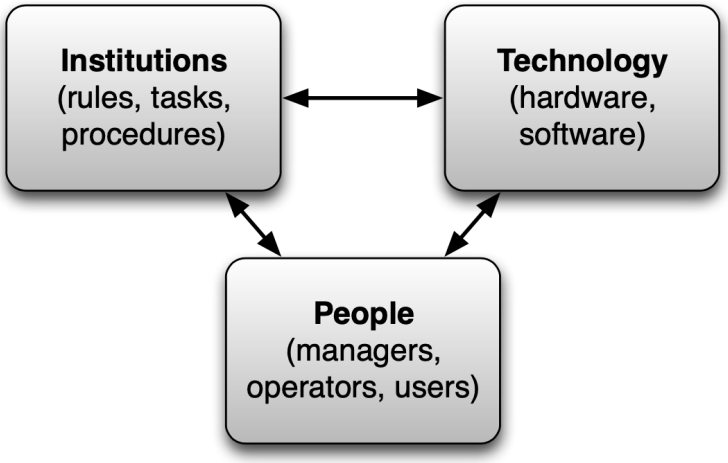
Different senses of “responsibility”

- **Blameworthiness-responsibility.** x is blameworthy for a damage H, when x caused (determined) H, and x’s action causing H represent a fault, namely the culpable violation of a standard of behaviour
- **Capacity-responsibility.** An agent x is capacity-responsible or capable if x satisfies the mental conditions which are required for liability
- **Liability-responsibility (liability)**. An agent x is liable for a harmful event H, if, given x’s connection to H, x is to be subject to the sanction (punishment or obligation to repair) connected to H.

-

- **Blameworthiness-responsibility**: moral responsibility.
- **Capacity responsibility**: a person that is adult and not child, or a person that is drunk and not not-drunk, has different responsibility in the meaning of the capacity to act
- **Liability**: in the sense of 'legally responsible'. From a legal pov the responsibility is usually associated to the obligation to compensate damages or to be subject to a punishment. Disciplinary-liability is the case in which is not the law to express the obligation but for example internal rules are the limitations.

Socio-technical systems: basic structure



To describe these complex systems that are subject of institutions, technology and people we can describe them as socio-technical systems: understand system as a combination of 3 main components.

- People working in the system
- Technology (hw or sw used inside the system)
- Institutions (all the set of rules that describes how the technology should be developed, used and how people should behave in order to fulfill the tasks, the internal rules and procedures comprising also the manual that shows how the technology should be used.)

Understand the interplay of these 3 things concerns the understanding of the role of responsibility.

Socio-technical systems: examples

Examples of socio-technical systems

- air traffic
- public administration
- healthcare systems

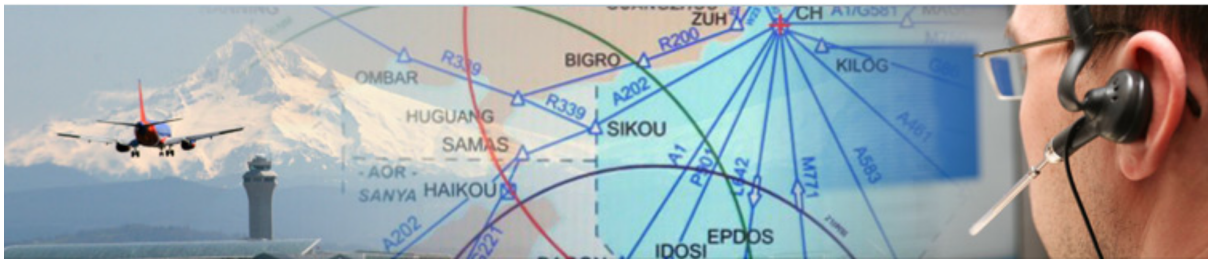
—> Systems in which we have a lot of technology and in most of the cases are highly regulated.

Responsibility is partly given by the law, but humans design the law, and also partly given by the design of the technology.

In a system in which we have a good treatment of the responsibility it means that we are in front of a situation in which we are maximizing the efficiency while reducing the costs.

The future of ATM

- In the time horizon of SESAR, that is over the next 30 years, a new generation of air traffic management systems will be developed.
- Such systems will be highly automated. They will make choices and engage in actions with some level of human supervision, or even without any such supervision.



In the last years we had a huge increasing of air traffic —> to manage correctly everything we need to introduce an higher level of automation, substitute many procedures that are currently carried by humans with technologies (and in the future with AI possible) and also integrate all systems that we have in Europe; because usually the control of the air space is considered as part of the sovranity of each single state and still in Europe we have still not a single air space. Each single air space, despite some few international rules, has its own specific rules (concerning for example how each organization is in charge of the safety or how each organization organized the space and so on)

Automation and the future ATM scenario

- New generation of ATM systems to increase capacity, safety, efficiency and sustainability
- Higher levels of automation

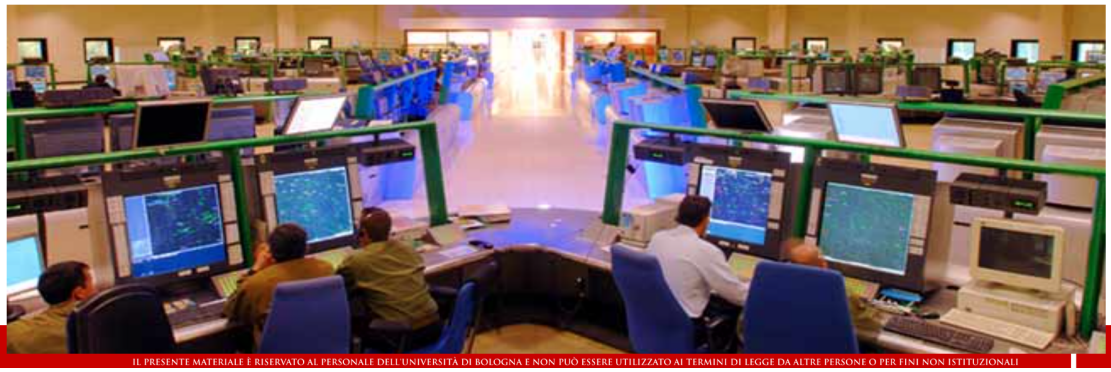
In the future every single aspect of the management of the traffic will be integrated. We are talking about technologies that will be deployed on airports, on each single aircraft.

We will manage aspects as navigation, collision avoidance, communication between aircraft and other operators on the ground and so on —> high level of automation.

What are the implication of such introduction of automation?

Implications of automation

- Delegation of task from operators to technology
- Humans as controllers and supervisors
- Hybrid agency (symbiosis/coagency → joint cognitive systems)
- Machine intelligence and autonomy (= independence + cognitive skills)
- The challenge of complexity (technological, “many hands”)



- The more we introduce automation, the more we will assist to a delegation of task from operators to technology. We have to specify that not always introducing automation means move the responsibility of a task from a human being to a technology substituting it. Introducing automation means a radical change that completely changes the way in which we deal with a particular aspect. The way in which we solve a problem is **completely changed** (not only one task).
- We still have humans most of the time that, with the introduction of automation, instead on being executors, are transformed in controllers and supervisors executing the task.
- In the future we expect not only automated systems, but systems in which we have Artificial Intelligence and in some way cognitive skills —> cognitive tasks that have a rule in the decision making processes
- At the same time is true that we are introducing the technology because we want to make the system more efficient (and in the aviation make the system more efficient means maintain the system at least as it is today) but at the same time the introduction of automation increases the complexity of the system. On one

side the we introduce the automation to increase safety, but on the other side the introduction of automation increases the technological complexity of a system and we are also making procedures more complex and also many actors will be involved in the future for the fulfillment of the task. So in the future **there is the possibility to have so many actors to be impossible to find the responsible of an accident —> problem of 'many hands'**

Automation: not all or nothing

- Not just **substitution of a human operator**
- Support to human capabilities

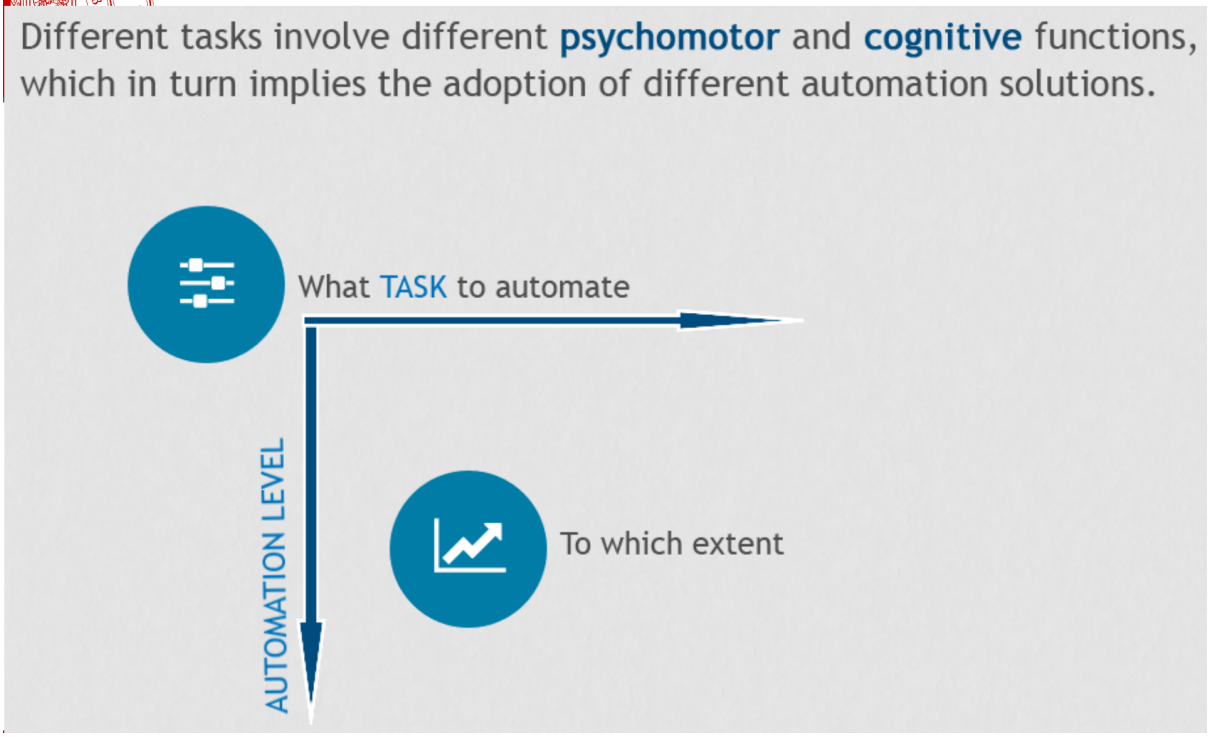
in performing tasks

- Some degree of cooperation is usually required

Automation is not just substitution of human operators, but it is introduced to support the humans in doing the task.

The problem is, always from the pov, is that some of the operations require cooperation between humans and technology: **human-machine interaction**.

Automation: not all the same



Automation is not present in the same way in all the different situations. Different tasks will involve different functions (physical or cognitive) which in turn implies the adoption of different automation solutions.

So we will have **different degrees of automation in relation to different kind of functions → this will change the way in which we will understand the relationship between the humans and the automation and, as consequence, also the way in which responsibility is allocated between all the actors.**

The level of automation taxonomy (SESAR 1)

From INFORMATION to ACTION

INCREASING AUTOMATION

A INFORMATION ACQUISITION	B INFORMATION ANALYSIS	C DECISION AND ACTION SELECTION	D ACTION IMPLEMENTATION
A0 Manual Information Acquisition	**B0** Working memory based Information Analysis	**C0** Human Decision Making	**D0** Manual Action and Control
A1 Artefact-Supported Information Acquisition	**B1** Artefact-Supported Information Analysis	**C1** Artefact-Supported Decsion Making	**D1** Artefact-Supported Action Implementation
A2 Low-Level Automation Support of Information Acquisition	**B2** Low-Level Automation Support of Information Analysis	**C2** Automated **Decsion Support**	**D2** Step-by-Step Action Support
A3 Medium-Level Automation Support of Information Acquisition	**B3** Medium-Level Automation Support of Information Analysis	**C3** Rigid Automated **Decsion Support**	**D3** Slow-Level **Support** of Action Sequence Execution
A4 High-Level Automation Support of Information Acquisition	**B4** High-Level Automation Support of Information Analysis	**C4** Low-Level Automatic **Decision Making**	**D4** High-Level **Support** of Action Sequence Execution
A5 Full Automation Support of Information Acquisition	**B5** Full Automation Support of Information Analysis	**C5** High-Level Automatic **Decision Making**	**D5** Low-Level **Automation** of Action Sequence Execution
		C6 Fulll Automatic **Decision Making**	**D6** Medium-Level **Automation** of Action Sequence Execution
			D7 High-Level **Automation** of Action Sequence Execution
			D8 Full **Automation** of Action Sequence Execution

A condensed version of the LOAT matrix

13

LOAT = Level of Automation and Taxonomy (taxonomy of automation used in aviation)

4 columns describe the 4 main cognitive functions related to the automation level (raws)

ROT / Use of video cameras in the control tower

A
INFORMATION ACQUISITION

- **A0** Manual Information Acquisition
- **A1** Artefact Supported Information Acquisition
- **A2** Low Level Automation Support of Info Acquisition
- **A3** Med. Level Automation Support of Info Acquisition
- **A4** High Level Automation Support of Info Acquisition
- **A5** Full Automation Support of Info Acquisition

The system **supports** the human in **acquiring information** on the process s/he is following. Filtering and/or highlighting of the most relevant information are **up to the human**.

14

Dynamically use the resource of traffic control : instead of having a fixed number of controllers, you have a controller that is not required to be in an airport, that can be assigned dynamically to different airports.

Activation of speed vectors by controllers

B
INFORMATION ANALYSIS

- **B0** Working-memory based Information Analysis
- **B1** Artefact Supported Information Analysis
- **B2** Low Level Automation Support of Info Analysis
- **B3** Med. Level Automation Support of Info Analysis
- **B4** High Level Automation Support of Info Analysis
- **B5** Full Automation Support of Info Analysis

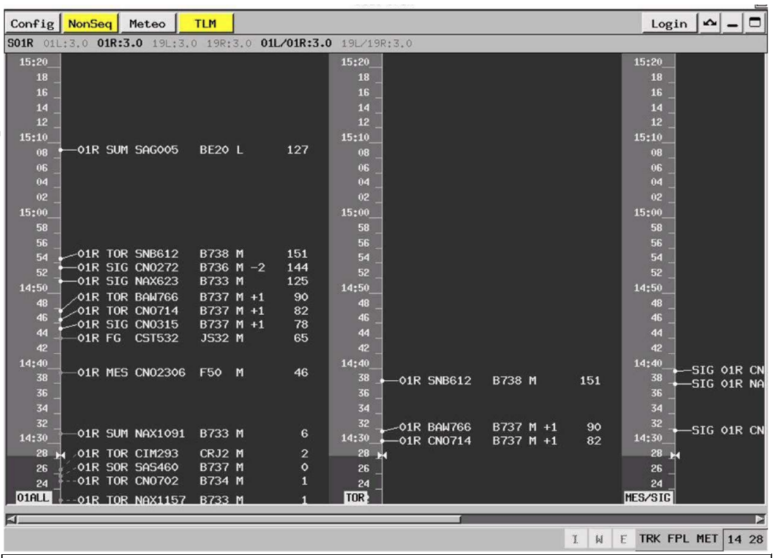
Based on user's request, the system **helps** the human in comparing, combining and analysing different information items regarding the status of the process being followed.

Area Control Centers: most of the work of traffic controller is not in the airport but in the traffic area control. Controllers make use of technology to acquire more information in a form that has been already analyzed by the computer. For example the traffic controller may ask to the computer to show vectors that can give visually an idea of the direction of each single flight.

AMAN sequence of landing aircraft

C
DECISION AND ACTION SELECTION

- **C0** Human Decision Making
- **C1** Artefact Supported Decision Making
- **C2** Automated Decision Support
- **C3** Rigid Automated Decision Support
- **C4** Low Level Automatic Decision Making
- **C5** High Level Automatic Decision Making
- **C6** Full Automatic Decision Making



https://www.eurocontrol.int/sites/default/files/article/content/documents/nm/fasti-aman-status-review-2010.pdf, page 16

The system **proposes one or more decision alternatives** to the human, leaving freedom to the human to generate alternative options. The human can **select** one of the alternatives proposed by the system **or** her/his own one.

AMAN = Assistant fo Arrival Manager → it is a technology that creates a plan for the departures and landings in an airport. The idea is that the system proposes one or more decisional alternatives to the human and leave the freedom to the human to select one of the options provided by the system or to move to a completely new strategy out of the box

Autopilot

D
ACTION
IMPLEMENTATION

- **D0** Manual Action and Control
- **D1** Artefact Supported Action Implementation
- **D2** Step by step Action Support
- **D3** Low Level Support of Action Sequence Execut.
- **D4** High Level Support of Action Sequence Execut.
- **D5** Low Level Automation of Action Sequence Exec
- **D6** Medium Level Automat. of Action Seq. Execut.
- **D7** High Level Automation of Action Seq. Execut.
- **D8** Full Automation of Action Sequence Exec

The system performs automatically a sequence of actions **after activation by the human**. The human can **monitor** all the sequence and can **interrupt** it during its execution.

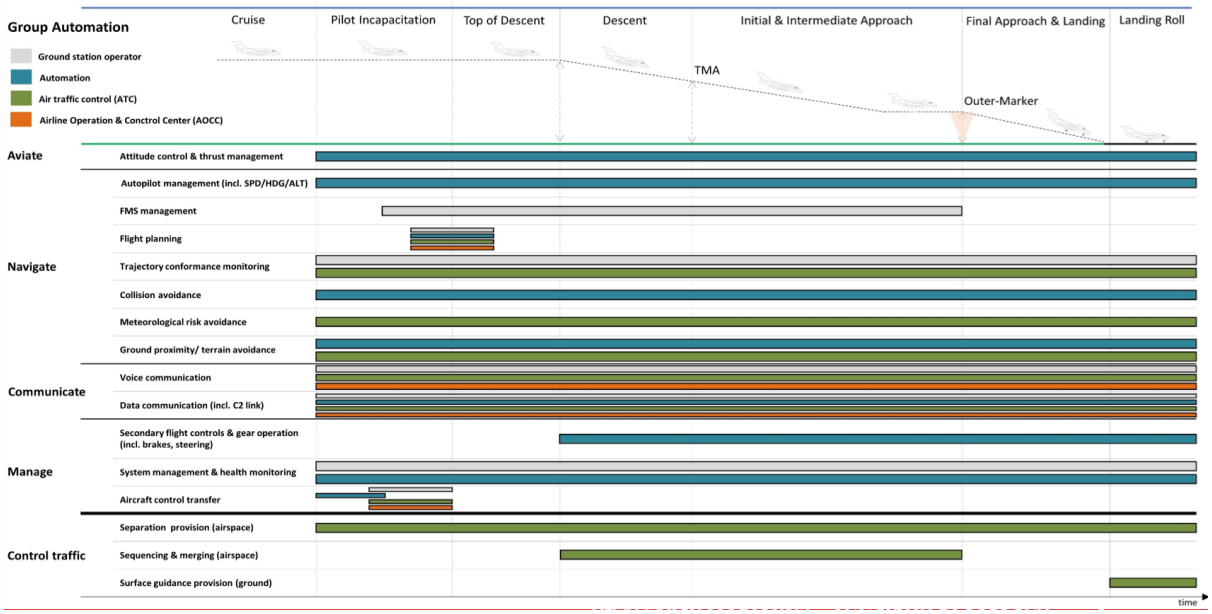
Autopilot: here we have automation in execution of the actions. The technology is driving the aircraft. To do that the technology should be activated by the human and the human has always the possibility to override the function of the autopilot → the human is supposed to monitor the functioning of the system and the interrupt the execution by the function of the autopilot.

ARGOS: technology developed by the government that in the future is supposed to replace a traffic controller. Traffic controllers have very limited resources. In the future, in order to co-op with the traffic is important to find a way to replace partially or completely a traffic controller with the technology. ARGOS is different from the autopilot because is the technology that has to activate the human when is the case and not the opposite. So we can fo other activities until ARGOS call us. It solves the problem of managing all the flights, ensuring that all the flights are sufficiently separated each other to do not incur in risky situations.

As long as the system is not able to find the solution or make some mistakes it is still possible to call the human to help it.

SAFELAND project

supporting flight and landing of aircraft operated by a single pilot, in case of partial or total incapacitation of the pilot.



European project.

Idea: in the future in order to reduce the cost of the pilot we will have only one pilot. What if the pilot is unable at some point for some problems to keep the control of the flight? Who is going to make sure that the pilot will land safely? Idea is to have technologies to check whether or not the pilot is unable to control the flight and in combination with additional technologies and humans on the ground be able, by a re-distribution of the tasks associated to the responsibilities, to safely land the aircraft.

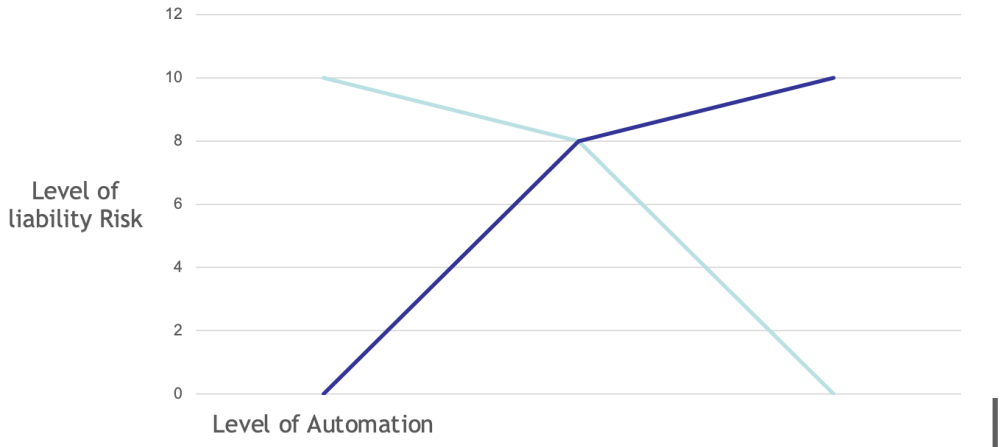
Some questions

- How automation transforms operators' roles and tasks? What impact on their responsibilities?
- Who is responsible for the behaviour of systems that humans cannot fully monitor and control?
- Who is responsible for information supplied by automated systems that the human cannot verify?

We need to discuss responsibilities in which a human has to take a decision under the suggestion of a technology. What is expected from a human operator? Which is the responsible for the action in that case?

Level of automation and responsibility

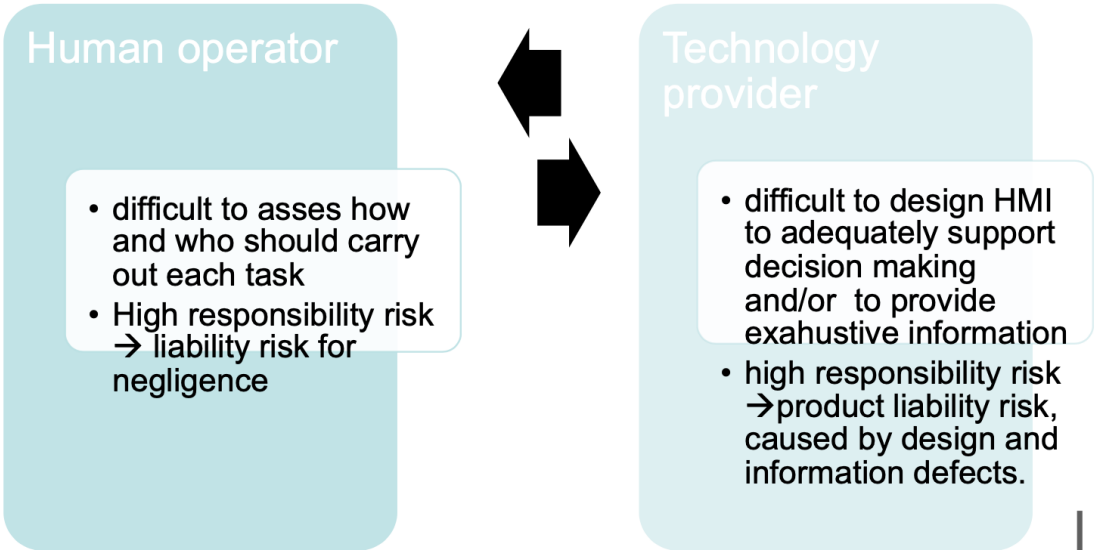
- Increasing the level of automation will proportionally increase the responsibility for the technology provider, and decrease the responsibility risks for the human operator.
- However the employment of technologies with **intermediate levels of automation** may result in a higher risk of being considered responsible, both for the technology provider and the human operator



- If there is a technology in which there is no automated task of course the responsibility, like the legal liability, is under the human responsibility
- If I substitute all humans with a technology I will have the opposite situations.
- Can we have situations in the middle in which the responsibility is 50-50? No, because situations in which we have an interplay between humans and technologies are situations in which the system is much more complicated, first of all in terms of human-machine interaction and therefore there are many things that need to be taken into account, many more tasks to be carried out by the technology and by the humans. Therefore, whenever you have a task, you have the risk to have the task responsibility; namely the fact that one of the 2 actors (the technology or the human) is not able to execute the task.

Fragmentation of tasks

The **fragmentation of tasks** may results in uncertainity and complexity of procedures



In situations in which humans and technologies interact you have **high fragmentation of tasks.**

If we have a technology which does not expect any intervention of the humans, we are limiting human -machine interaction, which means that if we look at the side of the manufacturing of the technology, we do not need to make sure that the technology should inform the operator in the correct way. On the other side you do not need to train the operator to make him know hoe to manage the technology.

For an ethical and legal point of view we always to ensure that a technology is under human control → take care of the design of the HMI.

On the other side we always to have in mind that considering interaction between technologies and humans make the systems more complicated and is source of responsibility risk and also from a legal point liability risk.

Level of automation and responsibility

- Increasing the level of automation will (as a very general trend) increase the responsibility for the technology provider, and decrease the responsibility risks for the human operator.

- Individual responsibility: 2 contrasting trends
  - It shall persist only when the human acted with an intention to cause harm or with recklessness (Just Culture)? OR..
  - always, as humans are the «moral crumple zone» (Elish 2018)?

There are 2 important trends in the topic of how to share responsibility and liability.

1. When in the future we will have full automated systems we do not want the humans to be the frontline of the responsibility. We want to the system to be safe → whenever in the system there is a person that knows that in case of a failure or situations in which you are near to a failure, the responsibility can be traced back to that person —> the person will have no incentives to report about the failure. If we have a situation, for example in which we have a pilot interacting with a technology and the pilot makes a mistake, which maybe partially reconducted both to a bad behaviour of the pilot but also partially to a bad design of the HMI; let's alspo say that at the endy of the day the pilot is very lucky and we do not have the accident → this is what the expert of safety call a "near accident" a "near failure". So unless it is the pilot to take the initiative to inform the rest of the system that there was that risk, you will not know anything about this risk and there this failure will disappear. The idea is that we want to find a good balance between creating the incentive to people to report their on failures and to the other side avoid that these people instead prefer to do not so (because for example if you are a pilot and you confess that you made a mistake you will be sent back to training for example). On the other side you do not want that people know that they are exonered by liability because otherwise you would have the opposite risk → people taking unnecessary risk → this is a

problem not only for the people that deliberative want to make an accident —> this is the case in which a pilot over-trust the technology → being so confident of the capability of the system to protect myself regardless of my failures that for example I will decide to put the aircraft inside a storm because i know that the technology will support me in navigate the aircraft.

So idea of this trend is that we should limit responsibility of humans.

Open issue: Decision making authority

- ***Effective* decision-making authority in socio-technical systems**
  - **Joint cognitive systems?**
  - **The model described (or prescribed) by laws, regulations, procedures:**

- **Aviation:** ICAO Annex 2, sec. 2.3.1 Responsibility of pilot-in-command (ultimate responsibility)

- Vienna Convention on **Road Traffic**, Art. 1(v) "Driver" means any **person** who drives a motor vehicle or other vehicle (but recent proposal of amendment of art 8(5) for ADS)

To what extend we can relate the decision making authority to the humans?

AI and EFFECTIVE DECISION-MAKING AUTHORITY

What about decisions to be taken jointly with AI, in conditions of limited resources – time, information, explanations? E.g.:

- **Medical diagnosis** assisted by AI (Lagioia, Contissa 2020)
- **Frontex border** controls: «12 seconds to decide»

Machine intelligence is fundamentally ***alien****, and often, the entire purpose of an AI system is to learn to do or see things in ways humans cannot[..]*

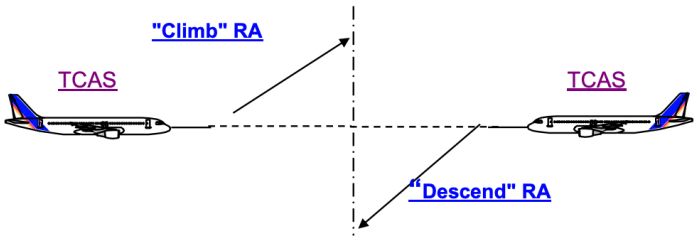
Ultimately, the ***lack of a principled basis to contradict AI predictions implies that the reasonableness of an action in individual cases must be tied to the decision to use AI as a general matter.*** *(Selbst 2019)*

Owing to the ***evidence*** *in their favor (stipulated by definition), it is more appropriate to think of* ***expert robots as above average in their ability to make decisions that will produce desirable outcomes […]***

This fact suggests that ***granting a general decision-making authority to human experts will be problematic once expert robots are properly on the scene****.*
(Millar, Kerr 2018)

Traffic Collision Avoidance System (TCAS)

http://www.skybrary.aero/index.php/TCAS

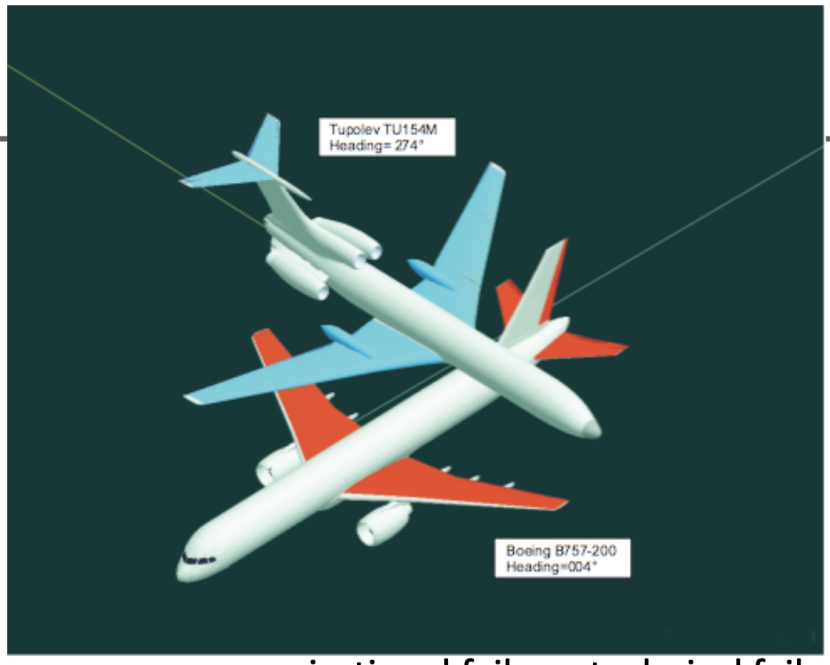


- Visual and aural advices
- 2 types of advisories: **TA (Traffic Advisory)** and **RA (Resolution Advisory)**

- RA shall be executed by the crew; The system decides the best option and informs the human
- During the execution by the pilot the system provides guidance through continuous visual and aural feedback

This technology is particular because contrarily to the assumptions, here is the technology to have the 'last word' in the sense that this is a technology that makes the last safety net before a collision. Theoretically in a perfect world this technology would not be used. We talk about situations in which we are in the 40 s before the crush. TCAS is deployed on each aircraft. Whenever 2 TCAS come in contact they generate a warning in the traffic advisor. After informing the pilot the 2 aircraft negotiate a way to avoid collision. The best way usually is that one aircraft goes down and one goes up, to increase the space. In this moment **is the technology to give commands and the humans has only to execute the orders. In those 30s the pilot disregard any other order.** Best way to avoid the accident is following the technology.

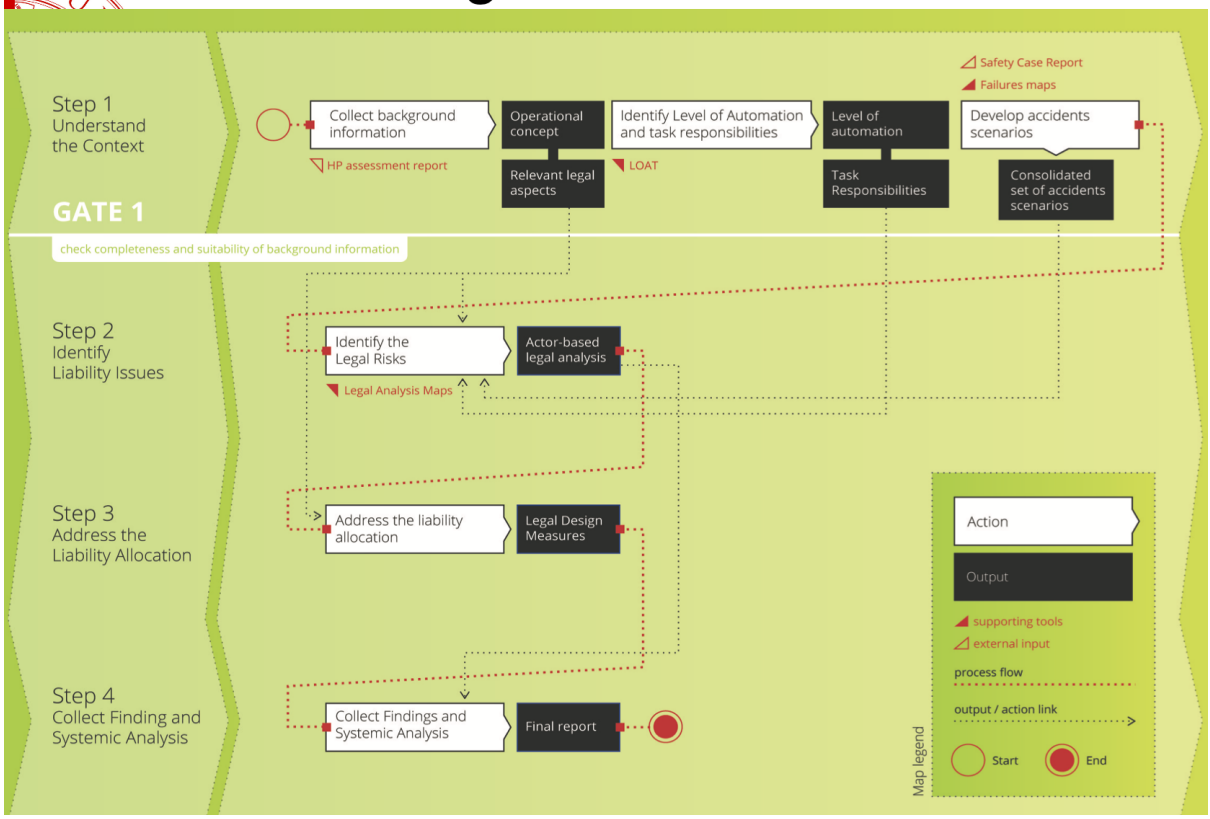
Überlingen mid-air collision (2002)



Human error, organizational failure, technical failure

https://www.youtube.com/watch?v=NlKu7BtMe8I

Legal Case



Legal Case is a technology which is developing to judge the responsibility of different cases and provide suggestions and ways to solve the problem also with intervention of the designer of the technology.