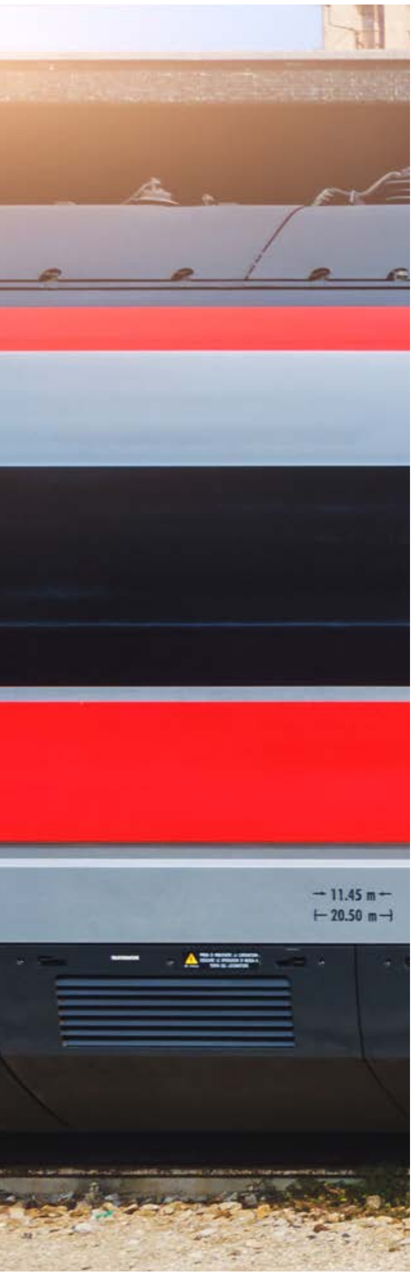
UIC SECURITY PLATFORM

Artificial intelligence-related applications for railway security

July 2025

INTERNATIONAL UNION
OF RAILWAYS

CONTENTS



EXECUTIVE SUMMARY

In an increasingly strong security context, railway companies worldwide are exploring the growing potential of Artificial Intelligence. In the past years, those technologies demonstrated that they can represent a key factor and a major asset to increase efficiency by tackling many kind of issues, even in the security domain. Especially, the emergence of Machine and Deep Learning, followed by the Large Language Model solutions, further accelerates this deep transformation in the positioning of technologies within the chains of command of any companies.

In this regard, applications of AI systems may support security experts within railway companies at each step of the security cycle in many different activities towards the protection of assets and persons, potentially including real time surveillance and reporting, access control, early detection of weapons, violence and other potentially dangerous behaviours, incident response and post-event analysis.

All the above-listed use cases are highly critical, implying personal data processing and ethics considerations, and many railway companies are now willing to start testing or implementing some of those, facing sometime difficulties related to the legal framework that is also facing the regulation of AI technologies challenge.

Thus, this document aims to provide an overview about AI technologies' main aspects (technical, ethical, regulatory, etc…), a picture of current applications of AI related technologies to the railway security domain and an illustration of AI solution deployment and usage.

This illustration is provided by SNCF, in France, that managed to use AI vision analytics tools applied on its CCTV system during the Paris Olympic Games 2024.

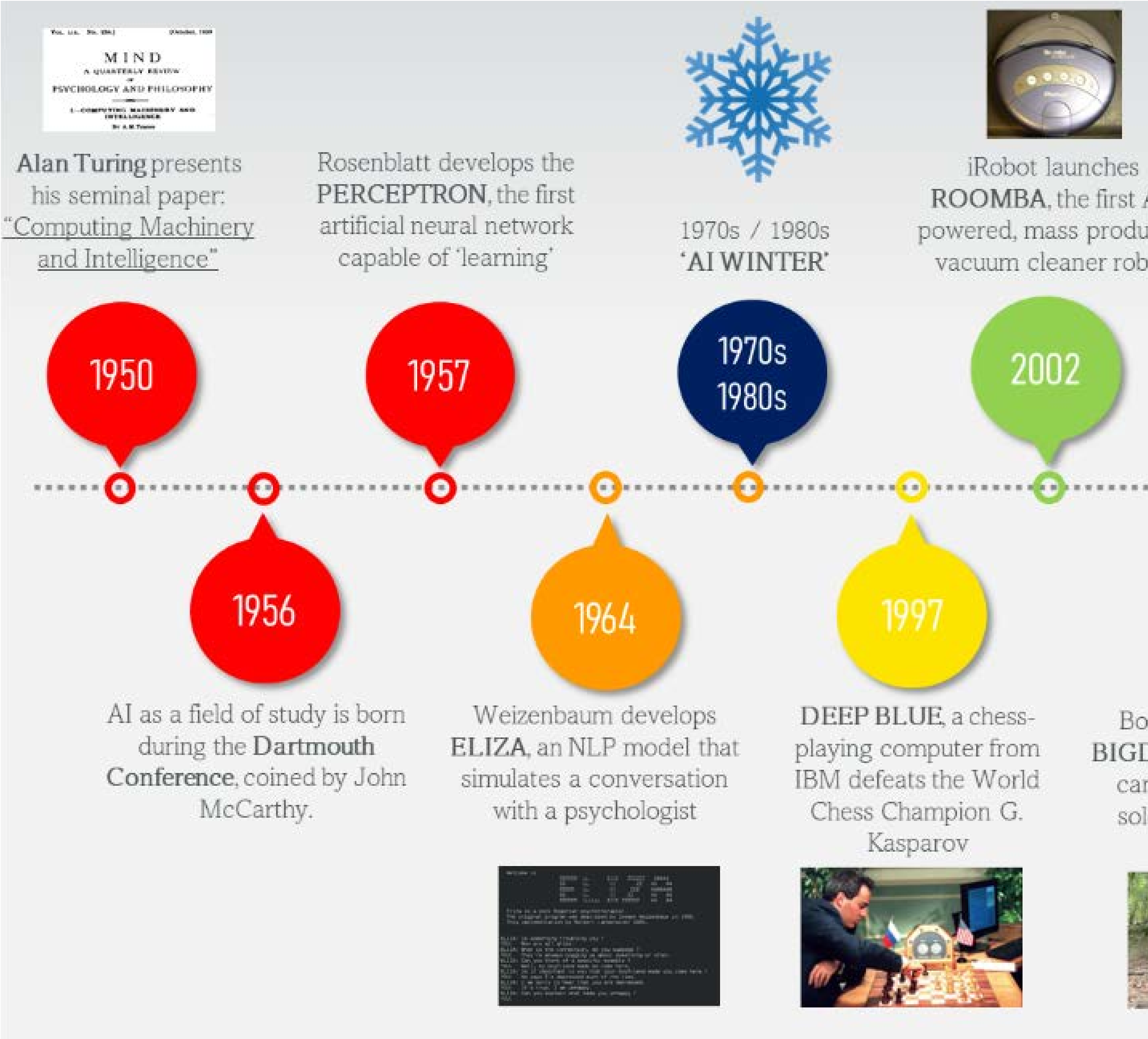
SNCF works to capitalize on its CCTV systems (80,000 cameras deployed on its sites) by developing and testing video analytics solutions since 2017 and the major events of 2023 and 2024 were an opportunity to supervise operational implementations that could prefigure a sustainable framework.

In 2023, a specific law established a framework for implementing operational experiments with video analytics algorithms and, SNCF used it to measure the operational added value of these new technologies. 4 relevant use cases to railway context have been tested: detection of abandoned objects, detection of a person in a prohibited or sensitive area, detection of abnormal density of people and detection of crowd movement. The results are very interesting and demonstrate the potential of AI technologies. On the technical side, the intrusion and overcrowding reports worked well with a very good positive alerts rate. The reporting of abandoned objects remains, as expected, more complicated and the detection of crowd movements was complex to analyse due to the low number of reports. On the operational side, the results are very positive, and the operating process implemented has demonstrated the added value of the tested technology.

With 8 video operators trained in the use of this new tool (including specific trainings about data protection and the ethical aspect), SNCF has been able to fully integrate AI into its CCTV operation. The teams involved showed great enthusiasm throughout the experiments and believes that the use of video analytics algorithms has a real interest in carrying out their daily missions.

1.

INTRODUCTION: WHAT IS ARTIFICIAL INTELLIGENCE?


MIND
A QUARTERLY REVIEW
PSYCHOLOGY AND PHILOSOPHY
I.—COMPUTING MACHINERY AND INTELLIGENCE
1950
1956
1957
1964
1970s
1980s
1997
2002

1950 – Alan Turing presents his seminal paper: "Computing Machinery and Intelligence"

1956 – AI as a field of study is born during the Dartmouth Conference, coined by John McCarthy.

1957 – Rosenblatt develops the PERCEPTRON, the first artificial neural network capable of 'learning'

1964 – Weizenbaum develops ELIZA, an NLP model that simulates a conversation with a psychologist

1970s / 1980s 'AI WINTER'

1997 – DEEP BLUE, a chess-playing computer from IBM defeats the World Chess Champion G. Kasparov

2002 – iRobot launches ROOMBA, the first A powered, mass produ vacuum cleaner rob

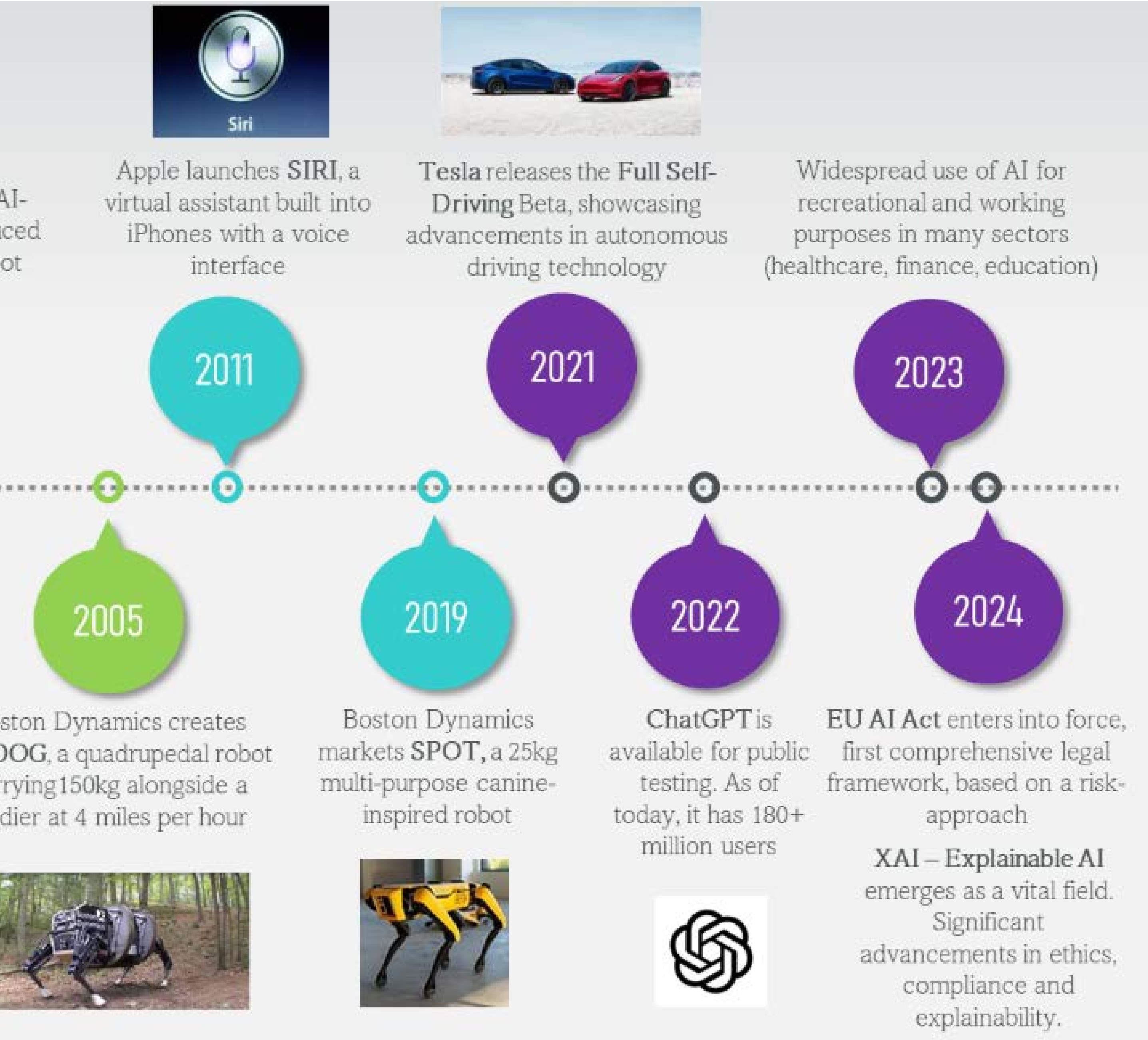
Bo BIGI car sol

Figure 1 - Artificial I

1.1. A short history of Artificial Intelligence

Early Foundations

The concept of **Artificial Intelligence (AI)** dates back to ancient myths and legends. However, the formal study of AI began in the mid-20th century. One of the first theorizations of a form of artificial intelligence owned by machines dates back to 1950, when the computer scientist and philosopher Alan Turing introduced it in his revolutionary paper [“Computing machinery and intelligence”](). In this paper, he introduced the Turing Test, a criterion for determining whether a machine can exhibit human-like intelligence.



AI-
nced
ot

2005

ston Dynamics creates **OOG**, a quadrupedal robot rying 150kg alongside a dier at 4 miles per hour

2011

Apple launches **SIRI**, a virtual assistant built into iPhones with a voice interface

2019

Boston Dynamics markets **SPOT**, a 25kg multi-purpose canine-inspired robot

2021

Tesla releases the **Full Self-Driving** Beta, showcasing advancements in autonomous driving technology

2022

ChatGPT is available for public testing. As of today, it has 180+ million users

2023

Widespread use of AI for recreational and working purposes in many sectors (healthcare, finance, education)

2024

EU AI Act enters into force, first comprehensive legal framework, based on a risk-approach

XAI – Explainable AI emerges as a vital field. Significant advancements in ethics, compliance and explainability.

ntelligence timeline

The term “Artificial Intelligence” was coined in 1956 during the Dartmouth Conference, organized by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. This conference marked the official birth of AI as a distinct field of study. Researchers were optimistic about creating machines that could perform tasks requiring human intelligence, such as playing chess, proving mathematical theorems, and understanding natural language.

The Early Years: 1950s-1970s

The initial years of AI research were characterized by enthusiasm and ambitious goals. The first model of neural network, an algorithm capable of learning and making simple decisions, was developed as early as 1957, representing the foundation ground for future developments in the field of machine learning.

Early AI programs, such as the Logic Theorist and the General Problem Solver, demonstrated the potential of machines to solve complex problems.

However, progress was slower than anticipated, leading to periods of reduced funding and interest, known as “AI winters.”

Despite these setbacks, significant advancements were made. In 1966, Joseph Weizenbaum developed ELIZA[1], an early natural language processing program that simulated a conversation with a psychotherapist. In1972, the first expert system, DENDRAL, was created to assist chemists in identifying organic molecules.

The rise of Machine Learning: 1980s-1990s

The 1980s saw a resurgence of interest in AI, driven by the development of machine learning techniques. Researchers began to focus on algorithms that could learn from data and improve over time. The introduction of backpropagation, a method for training neural networks, was a significant breakthrough.

During this period, expert systems gained popularity in various industries, providing decision support in fields such as medicine, finance, and manufacturing. The first robotic, AI-powered cars were also experimented in controlled environments (MIT Technology Review, 2016), notably with the NavLab1 Project by Carnegie Mellon University. However, the limitations of rule-based systems became apparent, leading to a shift towards more flexible and adaptive approaches.

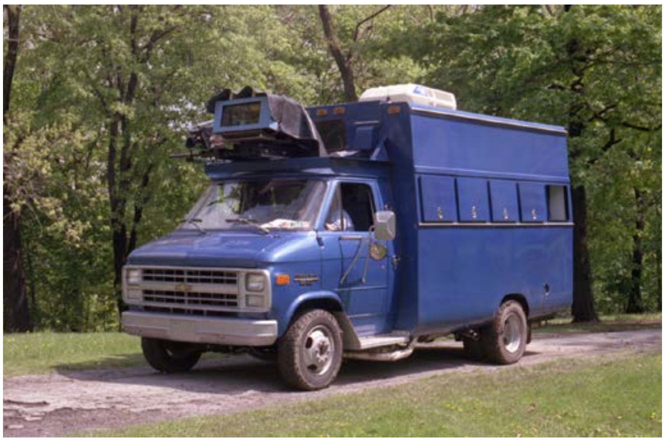


Figure 2 - NavLab1 Project by Carnegie Mellon University

[1] This simple natural language processing program gives to the users the impression to be talking to someone and being understood. The most famous variation of the original code was inspired by Rogerian psychotherapy practices. It is freely available on the web at the following link: https://web.njit.edu/~ronkowit/eliza.html

Early 2000s: The foundations of modern AI

The 21st century has witnessed unprecedented advancements in AI, fuelled by the availability of vast amounts of data and increased computational power. The early 2000s marked a period of foundational development in AI. During this time, AI research focused on improving machine learning algorithms and expanding computational power. Key advancements included **Support Vector Machines (SVMs)**, that became popular for classification tasks, offering robust performance in various applications. Early 2000s saw improvements in **Natural Language Processing (NLP)**, enabling better understanding and generation of human language. The rise of the internet led to an explosion of data, providing higher data availability, representing in turn a rich resource for training AI models.

Mid-2000s to Early 2010s: The rise of Machine Learning

The mid-2000s to early 2010s witnessed a surge in machine learning research and applications. Notable developments included:

- **Deep Learning**: The introduction of deep learning, particularly convolutional neural networks (CNNs), revolutionized image and speech recognition. Geoffrey Hinton's work on deep belief networks was particularly influential.
- **Big Data**: The proliferation of big data technologies allowed for the processing and analysis of vast datasets, enhancing AI capabilities.
- **AI in Industry**: Companies like Google, Amazon, and IBM began integrating AI into their products and services, leading to innovations like recommendation systems and voice assistants.

Mid-2010s: AI becomes mainstream

By the mid-2010s, AI had become a mainstream technology, impacting various industrial sectors:

- In 2011 Apple launched the first version of its **virtual assistant** Siri, followed in 2013 by Amazon's Alexa and Google Assistant in 2015. These devices have soon become ubiquitous, providing users with voice-activated control over their devices.

Figure 3 - AI-powered home assistants

- AlphaGo: In 2016, DeepMind's AlphaGo defeated world champion Go player Lee Sedol, showcasing the power of reinforcement learning and neural networks.
- **Self-Driving Cars**: Companies like Tesla, Waymo, and Uber made significant strides in autonomous vehicle technology, leveraging AI for navigation and decision-making.
- **Healthcare**: AI began to play a crucial role in healthcare, with applications in medical imaging, drug discovery, and personalized medicine.

Late 2010s to Present: AI in Everyday Life

The late 2010s to the present have seen AI become an integral part of everyday life.

In **2018**, OpenAI released the GPT-1 Large Language Model, expanding the capabilities of AI by employing a faster, semi-supervised model that laid the foundations for enabling the creation of human-like text. The model was followed by more refined versions GPT-2 in 2019 and GPT-3 in 2020.

In **2020,** Waymo launched its fully autonomous taxi service, Waymo One, in the Phoenix area, marking a significant milestone in self-driving car technology**.** In **2021**, Tesla released a 'Full Self-Driving Beta' version of their system, showcasing advancements in autonomous driving technology.

In November **2022** OpenAI introduced ChatGPT, an AI chatbot built on the GPT-3.5 large language model, advancing conversational AI capabilities. A rapid surge in the use of generative AI technologies brought the service to acquire 1 million users in just five days, and over 180 million users in less than two years.

In **2023**, among the many LLM developed, Meta introduced its own language model LLaMA AI, OpenAI launched its multimodal LLM GPT-4 and Google followed with its PaLM-2.

During the same year, the number of companies adopting AI technologies at least for one function has dramatically increased – according to the McKinsey Global Survey on AI (2024) – both for analytical AI (from 33% to 65% of interviewed companies) and for generative AI (from 55% to 72% of interviewed companies).

In **2024**, the main sectoral trends include an increased interest towards **Multimodal AI**: AI models that can process and generate text, images, and videos are becoming more prevalent. Following the success of text-to-image models, text-to-video models are gaining traction. At the same time, there is a growth of **Customized AI Models** which are tailored to specific needs (allowing users to fine-tune such models with little or no coding skills). Advances in AI are also driving a rapid growth in the field of **Robotics,** leading to more general-purpose robots capable of performing a wider range of tasks, achieved by training robots with large, versatile models rather than task-specific ones.

As AI's influence is growing fast, so are the concerns about **AI ethics, safety, bias, and regulation**. Governments and organizations worldwide have started to develop regulatory frameworks to ensure responsible AI use.

Disclaimer

Did you notice something 'different' in the chapter above?

It was completely elaborated by generative AI (Microsoft Copilot) and subsequently reviewed by the document editor.

1.2. How does Artificial Intelligence work?

Despite the common misconception that “AI thinks like humans”, AI systems are designed – and able, at best – to mimic the human cognitive functions in order to perform some tasks that usually require human intelligence. This is why it is commonly referred to as ‘weak’ or ‘narrow’ AI, in contrast to ‘strong’ AI that will supposedly be replicating in a much closer way (which is unlikely to happen in the foreseeable future) the human capacity to understand, learn and apply knowledge.

AI systems are often complex and encompass multiple AI algorithms into a broader framework which is usually intended to solve complex problems. These systems may display, at different degrees, four main capabilities: perception, reasoning / decision making, learning, and actuation (EU Commission, 2018).

While an **AI algorithm** is a set of formulas that shape the process of learning from data and making decisions, an **AI model** represents the patterns learnt and the knowledge extracted from the data by the algorithm. Once trained, the model will be able to make predictions or decisions based on previously unknown data.

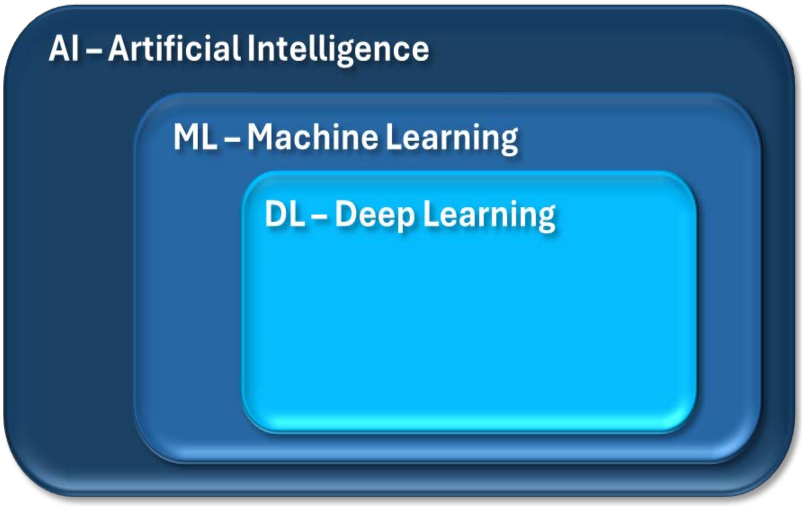


Figure 4 - Artificial Intelligence, Machine Learning and Deep Learning

Machine Learning (ML) is a branch of AI that aims at improving the way in which systems learn, make predictions and/or take decisions on new data, by using algorithms which have been/are trained on data sets. There are different types of ML. Among them (IBM, 2023):

- **Supervised Learning**: the algorithm learns from labelled data[2], leveraging known outcomes to train models. This kind of model can be very effective for some applications (e.g., image/object/ speech recognition, predictive analytics). A main drawback of this approach is that it requires extensive expert knowledge about the data and is therefore time-consuming.
- **Unsupervised Learning:** the algorithm is fed with raw/unlabelled data, and works to identify patterns, structures and relationships among them without any specific instructions. It may be especially useful when subject matter experts are unsure of common properties within a data set.
- **Semi-supervised learning:** the algorithm is fed with input data being only partially labelled.

[2] Each training element/example (input) is paired with a label (output). As an example, in order to train the algorithm to recognize trains, is should be fed with images that are labelled as ‘trains’.

Deep Learning (DL) is a subset of ML and aims at better simulating the complex decision-making power of the human brain. It includes algorithms that use neural networks with many layers ("deep") to analyse complex patterns in large datasets. It allows the automation of the extraction of features from large, unstructured and unlabelled datasets, with the aim of autonomously making predictions about what the data represents.

The concept of **Generative AI** refers to DL models can also create different kinds of original content as an output (e.g., text, audio, video), and it is based on foundational models such as Large Language Models (LLMs), such as OpenAI's GPT-4 and Meta's LLaMA.

In general terms, **AI works by combining large datasets with complex AI algorithms to analyse, learn from, and make decisions based on the data**. The main phases of such a process can be described as follows:

1. **Data Collection**. AI systems gather data from various sources such as text, images, audio, and video: this data is then pre-processed to remove 'noise' and inconsistencies.
2. **Data Processing**. The collected data is fed into AI algorithms, that use statistical methods to identify patterns and relationships within the data.
3. **Learning and Training**. AI systems use ML techniques to learn from the data. Models are trained using historical data to make predictions or decisions (see above for the main different types of learning).
4. **Model Evaluation**. Once trained, the model is tested on unseen data to evaluate its performance. Metrics such as accuracy, precision, and recall are used to assess how well the model performs.
5. **Deployment**. Once the model is trained and evaluated, it is deployed to make predictions or decisions on new, unseen data (i.e., inference).
6. **Continuous Improvement**: AI systems iteratively learn and improve over time by incorporating new data and feedback. This process helps the models become progressively more accurate and efficient.

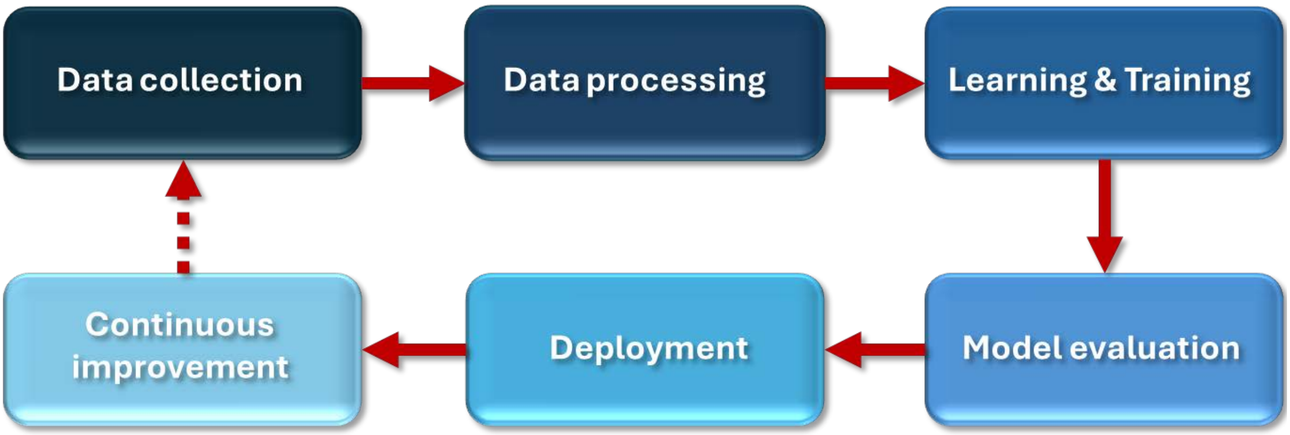


Figure 5 - Basic AI system process diagram

1.3. Ethical aspects - Biases and Algorithm transparency

AI technologies will have a positive impact on both our personal and working lives. Nonetheless, numerous legal and societal issues have also revealed the potential of these technologies to produce negative impacts. Privacy breaches, algorithmic discrimination, security and reliability issues, transparency, and other unintended consequences may lead to exacerbating already existing biases, generating discrimination or even threatening private and public security with potentially disastrous consequences.

To mitigate this risk, many states and social entities are advocating fair, ethically acceptable and sustainable development and use of this technology. Research has also developed the new field of study of AI Ethics to identify and build shared consensus over the foundational principles that should guide future AI developments and usages.

In the field of security, in particular, many concerns have been arising in the last few years, bringing important challenges to legislators and politicians around the world.

AI-powered systems may allow disproportionate **or intrusive surveillance** and an **invasion of citizen's privacy** and a limitation of civil liberties. Facial recognition and video analytics may significantly enhance security but also pose **risks to individual privacy**. The widespread use of these technologies can lead to constant monitoring, raising concerns about the right to privacy and the potential for misuse.

The decision-making processes of AI systems may be opaque due to their technical characteristics, potentially leading to a **lack of transparency** and making it difficult to understand how decisions are made. Furthermore, AI systems can perpetuate or amplify existing **biases**, potentially leading to unfair treatment of individuals.

This leads, in turn, to the **accountability issue** of clearly identifying who is responsible for the decisions taken by the system, and how to challenge them if they are incorrect or unjust. As an example, the use of AI in autonomous security systems, such as drones or robots. If an autonomous system makes a mistake or causes harm, it can be challenging to determine who may be considered responsible for it (e.g., the manufacturer, the operator, the railway company or the developer of the AI algorithm).

1.3.1. Biases

In the context of AI, the term **bias** refers to systematic errors or prejudices in the data or algorithms that can lead to inaccurate outcomes and/or to unfair treatment of individuals based on race, gender, or other characteristics. Three main classes of biases can be identified, depending on the agent that introduces/induces them into the system:

- **Bias introduced by data**. When the training data used to train AI algorithms is not representative of the real-world population, characteristics or contains historical prejudices, the AI system may learn, incorporate and ultimately perpetuate such biases and lead to unfair results/practices. As an example, facial recognition technologies have displayed, under specific circumstances, higher error rates for certain demographic groups, which can result in unfair targeting or exclusion.
- **Bias introduced by algorithms.** The AI algorithms themselves, due to a faulty design or data processing, introduce or amplify biases into the model and results.
- **Bias introduced by users.** Human users may, even unwillingly, interact with AI systems in such ways that ultimately lead to introducing bias into them, such as through feedback loops reinforcing existing patterns.

Examplary case

In January 2020, Detroit (USA) police arrested a man outside his home and subjected him to thirty hours of detention. It was the first publicly reported instance of an AI system's false face-recognition 'match' leading to a person's wrongful arrest due to a biased AI model. Following this case, police in the USA have been required to back up face recognition results with independent and reliable evidence linking a suspect to a crime before making any arrest. Staff will also be trained in face recognition technologies and their dangers.

Source: Williams v. City of Detroit | American Civil Liberties Union (aclu.org)

1.3.2. How to address ethical issues related to AI systems in the security domain?

Addressing the ethical challenges explained in the previous paragraph, in line with the aim of achieving social/public/customer acceptance of AI technologies in the railway domain, requires a balanced and comprehensive approach, including various actions:

- **Regulatory Frameworks**: Strictly adhering to the regulations that govern the use of AI in physical security to ensure it is used responsibly and ethically.
- **Ethical Guidelines**: Establishing ethical guidelines for the development and use of AI in physical security to ensure it aligns with societal values and norms.
- **Bias Mitigation**: Developing and deploying AI systems with mechanisms to detect and mitigate biases. **Inclusive design** emphasizes inclusion in the design process. The AI system should be designed with consideration for diverse groups such as gender, race, class, and culture. **Foreseeability** is about predicting the impact the AI system will have right now and over time.

Analytical techniques require meticulous assessment of the training data for sampling bias and unequal representations of groups in the training data. The source and characteristics of the dataset should be investigated and checked to ensure a fair balance has been achieved (e.g., is one gender or race represented more than the other?) (World Economic Forum, 2021).

- **Transparency and Explainability**: Ensuring that AI systems are transparent, and their decision-making processes are explainable.
- **Testing** is an important part of building or deploying a new product/service. User testing in this case refers to getting representatives from the workers (e.g., SOC operators) that will be using your AI product to test it extensively before it is deployed. Furthermore, model training should be as broad and diverse as possible, involving varying environments and contexts in order to collect new insights (WEF, 2021).
- **Public Engagement and Acceptance**: Clearly inform the public about the deployment of AI in railway security to address potential concerns. Social entities representing various stakeholders (e.g., passenger associations) may be involved in discussion rounds or exchange initiatives to build trust and consensus.
- **Valorising the Human Factor.** Labor Unions shall be involved at all stages in active and fair exchanges, clearly explaining the potential benefits brought by the adoption of AI systems into the working routine of staff, which will anyway remain at the core of processes.

By thoroughly and continuously addressing these challenges, the benefits of AI in physical security may be harnessed while minimizing potential issues and ensuring that these technologies are used in a fair, transparent, and accountable manner.

1.4. Regulatory Frameworks

Given the rapid evolution of AI technologies and their profound implications for society, adequate regulatory frameworks are crucial to timely address and mitigate significant risks including biases, discrimination, privacy violations, and unintended consequences.

Many countries around the world are currently elaborating such frameworks, establishing standards and protocols to identify, assess, and mitigate these risks, ensuring that AI technologies are developed and deployed responsibly. Regulatory frameworks ensure that these applications meet safety standards and do not compromise public welfare, thereby protecting citizens from potential harm.

Clear regulations delineate responsibilities among stakeholders, including developers, manufacturers, and users with the aim of establishing clear boundaries and eliminating grey areas of uncertainty. Furthermore, regulatory frameworks can facilitate public engagement in discussions about AI, helping to align technological advancements with societal values.

Accountability is an essential element for addressing issues of liability whenever AI systems may cause harm or make erroneous decisions, particularly in critical sectors such as transportation security. At the same time, ethical principles should be considered from the early phases of AI development, promoting fairness, transparency, and respect for human rights. By establishing rules for ethical usage of AI, frameworks help prevent misuse and foster public trust in AI technologies.

Another major achievement sought through current legislative efforts is promoting innovation by providing a clear legal landscape allowing companies to invest in AI development and/or deployment with confidence, knowing the boundaries within which they must operate to be competitive on the market.

In conclusion, as technical solutions continue to evolve rapidly, it appears clear that regulatory frameworks will play a pivotal role in shaping a future where AI technologies for security can be harnessed safely and ethically.

1.4.1. The European Artificial Intelligence Act

In order to pave the way towards future technological developments, while ensuring that AI technologies are developed and used in a lawful, ethical and transparent way all across Europe, EU institutions placed themselves at the forefront of regulatory efforts in the field of AI, developing and enacting a comprehensive Regulation – the so called "AI Act"[3] – that entered into force on August 1st, 2024.

[3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024, laying down harmonized rules on Artificial Intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

The **definition of ‘AI system’** adopted within the AI Act describes it as[4]:

“*a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments”.*

The aim of balancing the instances and opportunities brought by AI technologies with the safeguard of fundamental rights and public interests led to the adoption of a **risk-based approach** to make sure that AI systems used in the EU are safe, transparent, traceable, non-discriminatory and environmentally friendly.

The new rules establish **obligations for providers and deployers** of AI systems depending on their level of risk and contain essential requirements that all AI systems must meet to access the EU market, regardless of risk level.

To ensure proper enforcement of its provisions, the AI act establishes a **governance architecture** that includes several bodies:

- The **EU AI Office**, established within the European Commission as the centre of AI expertise across the EU. It plays a key role in implementing the AI Act - especially for general-purpose AI - fostering the development and use of trustworthy AI, and international cooperation to enforce common rules across the EU.
- A **scientific panel of independent experts** to support enforcement activities.
- The **European Artificial Intelligence Board** with Member States’ representatives to advise and assist the Commission and Member States on the consistent and effective application of the AI Act.
- An **Advisory Forum for stakeholders** to provide technical expertise to the AI board and the Commission.

It is important to note that the AI Act provisions shall not affect the competences of the Member States concerning national security, and therefore do not apply to applications used uniquely for military, defence or national security purposes[5]. Similarly, the provisions are not applicable to AI applications that are specifically developed and put into service for the sole purpose of scientific research and development.

[4] Artificial Intelligence Act, Art. 3 (‘Definitions’)

[5] Artificial Intelligence Act, Art. 2 (‘Scope’), nn.5, 6,

The following **4 levels of risk** are identified within the AI Act:

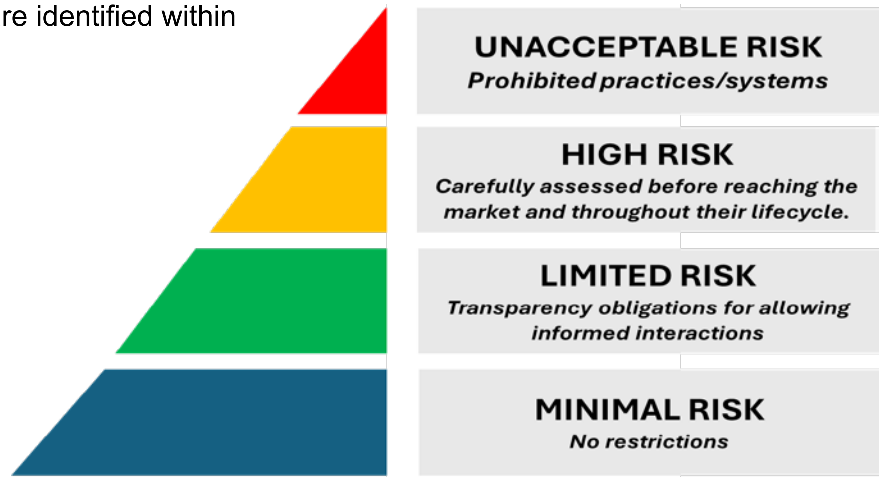

UNACCEPTABLE RISK
Prohibited practices/systems
HIGH RISK
Carefully assessed before reaching the market and throughout their lifecycle.
LIMITED RISK
Transparency obligations for allowing informed interactions
MINIMAL RISK
No restrictions

Figure 6 - The risk-based approach adopted in the EU AI Act

UNACCEPTABLE RISK **(Prohibited practices, art. 5).** This category includes applications that:

- Feature subliminal, purposefully manipulative or deceptive techniques
- Exploit vulnerabilities of persons or groups (i.e., age, disability) distorting their behaviour.
- Are aimed at classifying individuals and attributing a 'social score'
- Are aimed at assessing or predicting the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics
- Build facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage
- Infer emotions of persons at workplaces and education institutions, except for medical and safety purposes
- Deduce or infer race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation of persons to categorise individually (does not apply to labelling and filtering conducted for law enforcement on lawfully acquired datasets)
- Leverage 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, unless necessary to some specific purposes such as: identifying victims of serious crimes[6]; preventing specific, substantial and imminent threat to the life or physical safety or a genuine and present/ foreseeable threat of a terrorist attack; criminal investigations or prosecution for certain offences. In the above-mentioned use cases, authorization (or confirmation within 24 hours) by competent authorities is required, as well as notification to the relevant market surveillance authority and the national data protection authority.

HIGH-RISK **(art. 6).** A significative part of the EU AI Act focuses on the obligations for high-risk AI systems, which **may be placed on the market/deployed in the European Union**, provided that **strict requirements and constraints** are met. High-risk systems are those intended:

- as a safety component of a product, or where the AI system is itself a product, covered by the Union harmonisation legislation (European product safety legislation)

[6] Abduction, trafficking in human beings or sexual exploitation of human beings, as well as the search for missing persons

- to perform lawful usage of biometrics (remote biometric identification systems, biometric categorisation according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics; emotion recognition).
- as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity.
- for education and vocational training purposes
- for purposes related to employment, workers' management and access to self-employment.
- for regulating access to and enjoyment of essential private services and essential public services and benefits (e.g., healthcare and insurance, creditworthiness)
- for classifying and dispatching calls to emergency and first response services
- to perform lawful usages in law enforcement (e.g., assessing the risk of a natural person becoming the victim of criminal offences; to evaluate the reliability of collected evidence in the investigation and prosecution phases; for assessing the risk of a natural person offending or re-offending; to assess personality traits and characteristics or past criminal behaviour).
- to perform lawful usages in migration, asylum and border control management (e.g., to assess risks posed by a natural person who intends to enter or who has entered into the territory; for the examination of applications for asylum, visa or residence permits and for associated complaints; for detecting, recognising or identifying natural persons in the context of migration, asylum or border control management[7]).
- for the administration of justice and democratic processes

It is important to note that the systems listed above are not considered high-risk in cases where the system does not pose a significant risk of harm to the health, safety, or fundamental rights of individuals, or influence their decision making. This rule applies, specifically, if any of the conditions listed in the following are met:

- The AI system is intended to **improve the result of a previously completed human activity** or if it is intended to perform a **preparatory task** to an assessment relevant for the purposes of the use cases listed in Annex III.
- The AI system is intended to perform a **narrow procedural task**.
- The AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review.

Providers that believe their AI systems falling within the applications listed above (Annex III of AI Act) should not be regarded as 'high-risk', shall **document such an instance through an assessment** before placing the products on the market or deploying them into service.

Concerning high-risk systems, several **obligations** are stated **for AI system providers** and for the **AI system deployers**[8], as summarized in the following tables.

[7] Check of travel documents doesn't fall into this category.

[8] Artificial Intelligence Act, Art. 26.

Obligations for high-risk system providers

Several **obligations** are specified **for providers** of high-risk AI systems**:**

- **A risk management system** which should include the entire life cycle of the AI system, must be iterative and updated[9].
- **A Quality Management System** comprised of written policies, procedures, and instructions[10].
- **High-quality data** is a prerequisite for effective and ethical AI models. Therefore, several requirements for data governance (training data, validation data and testing data) for high-risk AI systems are specified. Data sets should contain accurate information, and potential bias must be identified and mitigated[11].
- **Accuracy**, **robustness**, and **cybersecurity measures**. Both organizational and technical measures for resilience purposes must be in place (e.g., backups, redundancy solutions). High-risk AI systems that continue to learn after being put into service should be designed to reduce or eliminate the risk of biased outputs that are susceptible to influence operations.
- **Technical documentation and recordkeeping** (e.g., automatic logs) **requirements** to improve transparency and deployers' knowledge about how high-risk AI systems operate and to estimate the impact of their operation before being put into service[12].
- **Permanent compliance and conformity** through continuous monitoring after the commercialization of systems to assess their performances[13].
- **Human oversight** to minimize risks to health, safety, or fundamental rights. This includes the permanent possibility, for operators, to stop the system's operation or override/disregard its output.
- **Registration.** Before being placed on the market or put into service, high-risk systems must be registered within the EU database for high-risk AI systems.
- **Serious incident reporting.**

[9] Artificial Intelligence Act, Art. 9.
[10] Artificial Intelligence Act, Art. 17.
[11] Artificial Intelligence Act, Art. 10.
[12] Artificial Intelligence Act, Art. 12.
[13] Artificial Intelligence Act, Articles 43 and 72.

Obligations for high-risk system deployers

Several **obligations** are specified **for companies deploying** high-risk AI systems, models, or services[14]. In this regard, accurate due diligence is crucial when selecting an AI system provider, as deployers may be held liable for some provider's shortcomings.

- Deployers must **use** and **monitor** the systems according to the **applicable instructions for use prepared by the provider and inform** all the relevant stakeholders (including surveillance authorities) in case of an incident, ensuring an optimal level of **cooperation**.
- **Systems shall be overseen by humans** with the possibility, for operators, to stop the system's operation or override/disregard its output.
- When the use of systems has an impact on employees' work, employees shall be informed in a transparent way about it.
- Having control over the input data, **deployers must ensure its relevance and pertinence to the purpose**. Deployers may also be obliged to carry out a DPIA (Data Protection Impact Assessment).
- Logs and records generated by the deployed systems shall be kept in accordance with any applicable regulations.

LIMITED RISK (article 50). AI systems falling under this category (e.g., chatbots, content generation tools) are considered unlikely to cause significant harm or violate fundamental rights of citizens. Nonetheless, they can still present transparency challenges, particularly in how they interact with users or generate content. For this reason:

- Providers must ensure that users are informed when they are interacting with an AI system.
- AI systems that generate synthetic content (like deepfakes or AI-generated text) must label their outputs in a machine-readable format, making it clear that the content is artificially created.
- Deployers of these systems must disclose the nature of the AI-generated content, especially in contexts where it could mislead users.

The classification of limited-risk systems will be reviewed by the European Commission every four years in order to ensure that the regulatory framework adapts to technological advancements and emerging risks.

MINIMAL RISK. AI systems in this category (e.g., AI-enabled video games, spam filters, and basic recommendation systems) are considered safe and unlikely to cause harm or violate fundamental rights. These applications, therefore, typically operate without significant ethical or safety concerns. It is encouraged, nonetheless, that developers of such systems adhere to general principles such as fairness, transparency, and non-discrimination to ensure their responsible use.

[14] Artificial Intelligence Act, Art. 26.

2. CURRENT APPLICATIONS OF AI-RELATED TECHNOLOGIES TO RAILWAY SECURITY

2.1. Why AI-related applications represent an interesting stake for railways?

After the pandemic crisis that led to a huge decrease (51% in Europe) in train ridership (ITF, 2022), it is widely estimated that railways have a vast potential to exceed the pre-covid modal share and further expand it over the next ten years. Car modal share is expected to decrease, for short/medium distance trips, by 20% to 70% depending on the region (UIC/MCKINSEY, 2022), and railways will be called upon to support this major shift, fostering sustainability goals and improving mobility (capillarity, speed, volume) and safety figures worldwide.

At the same time, AI is deeply transforming the world in many different ways, making it one of the most important and disruptive technologies of our time. This statement is firmly supported by the huge figures and increasing trends regarding research and innovation investments worldwide, both at private and public level.

Although most AI applications have not been fully implemented at scale by railway companies yet, rail transport worldwide may hugely benefit from a number of current and emergent applications of AI, as clearly identified by UIC in a study on the state of the art and perspectives of AI technologies in railways (UIC, 2021):

- **Customer assistance**.
- **Sales**.
- **Cleaning services.**
- **Predictive maintenance on infrastructure**.
- **Predictive maintenance on rolling stock.**
- **Traffic management.**

Concerning the domain of **physical security of railways**, applications of Artificial Intelligence may have a **disruptive potential.** Given the security-critical nature of many of their assets and the importance of their corporate mission, railway companies need to protect themselves from a wide range of possible attacks and to intercept in a timely manner emerging trends within the applicable threat landscape.

In this regard, applications of AI systems may support security experts within railway companies at every step of the security cycle in many different activities towards the **protection of assets** (stations, trains, tracks and other facilities) **and persons** (personnel, passengers), potentially including:

- Real time surveillance (video surveillance, site surveillance through UAVs and other robots) and reporting.
- Access control for passengers at stations and for staff at offices/company premises
- Early detection of weapons, violence and other potentially dangerous behaviours
- Incident response and post-event analysis.
- Staff training (e.g., training in complex/dangerous environments or in conditions that are difficult to replicate in a real-life exercise)
- Real-time/adaptive management and allocation of personnel and resources.
- Data gathering and integration (including open source).

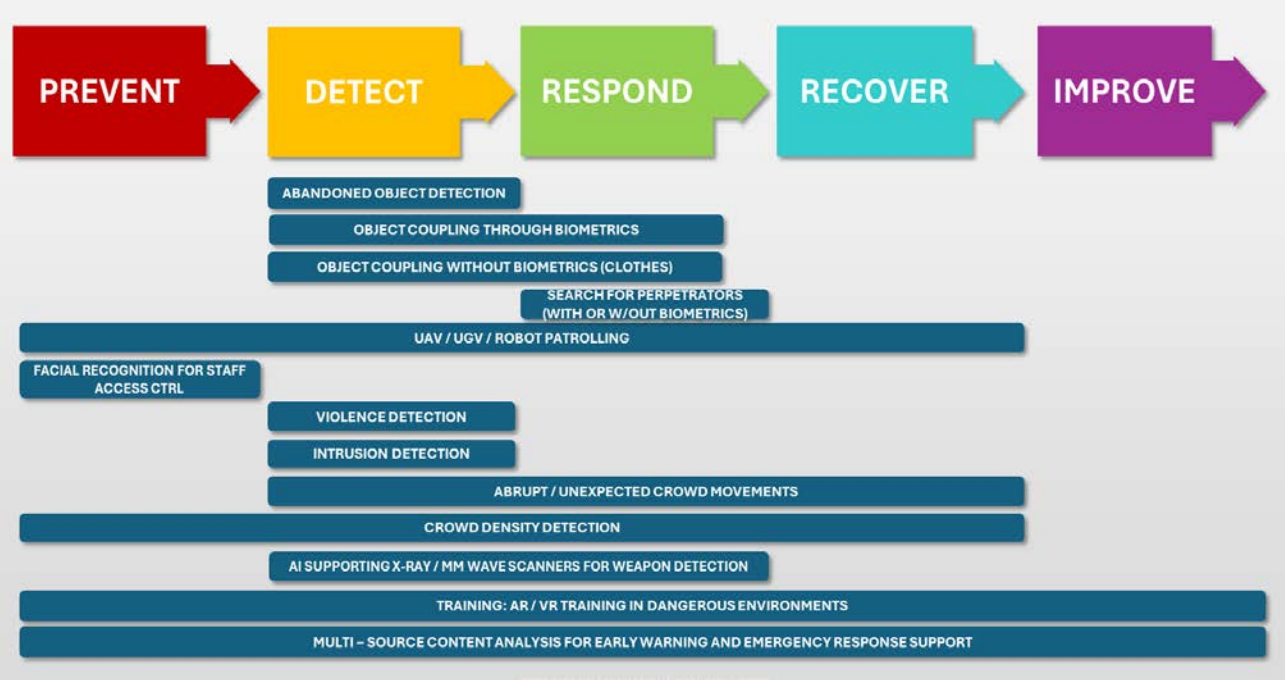
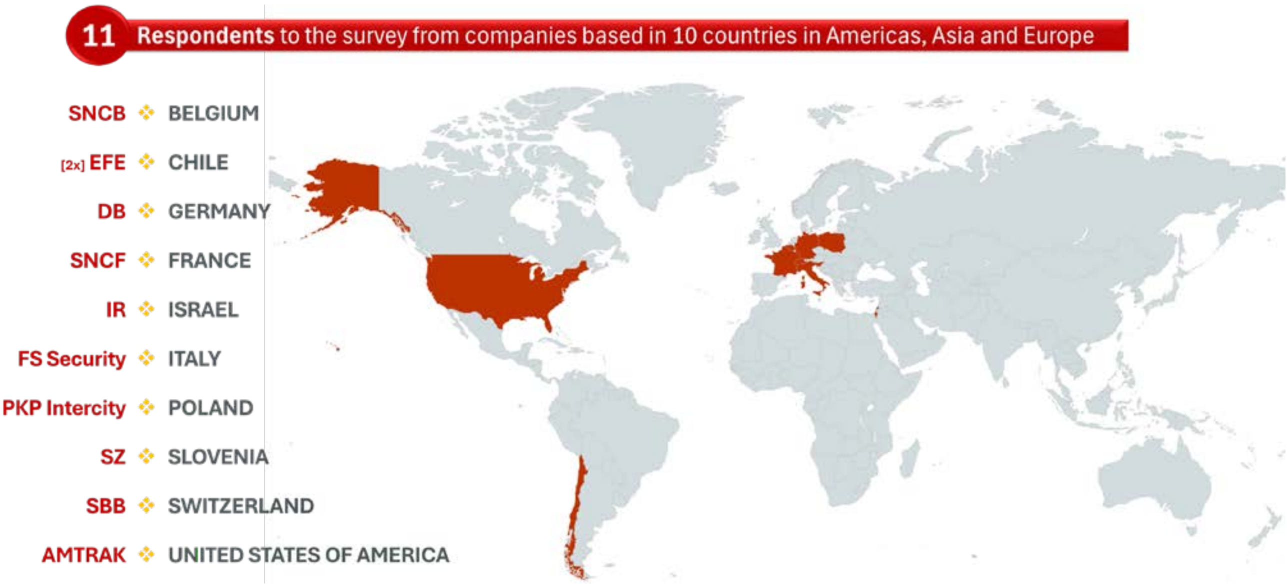

PREVENT
DETECT
RESPOND
RECOVER
IMPROVE
ABANDONED OBJECT DETECTION
OBJECT COUPLING THROUGH BIOMETRICS
OBJECT COUPLING WITHOUT BIOMETRICS (CLOTHES)
SEARCH FOR PERPETRATORS (WITH OR W/OUT BIOMETRICS)
UAV / UGV / ROBOT PATROLLING
FACIAL RECOGNITION FOR STAFF ACCESS CTRL
VIOLENCE DETECTION
INTRUSION DETECTION
ABRUPT / UNEXPECTED CROWD MOVEMENTS
CROWD DENSITY DETECTION
AI SUPPORTING X-RAY / MM WAVE SCANNERS FOR WEAPON DETECTION
TRAINING: AR / VR TRAINING IN DANGEROUS ENVIRONMENTS
MULTI – SOURCE CONTENT ANALYSIS FOR EARLY WARNING AND EMERGENCY RESPONSE SUPPORT

Figure 7 - Examples of potential use cases for AI systems in the (rail) security cycle

2.2. UIC Survey on the use of AI in the railway security domain

2.2.1. Introduction: methodology, respondents, questions and use cases



11 Respondents to the survey from companies based in 10 countries in Americas, Asia and Europe

SNCB	BELGIUM
[2x] EFE	CHILE
DB	GERMANY
SNCF	FRANCE
IR	ISRAEL
FS Security	ITALY
PKP Intercity	POLAND
SZ	SLOVENIA
SBB	SWITZERLAND
AMTRAK	UNITED STATES OF AMERICA

Figure 8 - Respondents to the UIC NTWG Survey on AI technologies in the railway security domain (2024).

The questionnaire on the topic “AI-related technologies in the railway security domain” was conceived within the UIC Working Group on New Technologies and validated in January 2024 by the members. It was then sent out in February 2024 through the UIC Network of Quick Responders, gathering **11 answers by railway security experts from 10 companies in Europe, Asia and Americas**.

In March 2024 the answers were analysed by the UIC Security Department, and this analysis was then presented during the “Workshop on AI-related technologies in the railway security domain”, organized by the New Technologies Working Group and held online on April 25th, 2024.

In this chapter, we receive the results of the questionnaire.

The survey enquired about the **13 potential sectoral use cases** of AI listed in the following:

1. **Face recognition for passenger access control.**
2. **Face recognition for staff access control at offices/facilities.**
3. **Suspicious behaviour detection.**
4. **Violence/aggression/harassment detection.**
5. **Intrusion detection.**
6. **Crowd detection/analysis.**
7. **Unattended item detection.**
8. **Tracking owner of unattended item through biometric data.**
9. **Tracking owner of unattended item through non-biometric data (clothes, accessories).**
10. **AI supporting X-Ray or Millimetre-Wave scanners for weapon detection.**
11. **UAVs/UGVs for autonomous patrolling.**
12. **Multi-source content monitoring/analysis for early warning and emergency response support.**
13. **Generative AI (e.g., reports, texts, images).**

The survey featured **10 questions**, designed to balance open and closed formats and arranged by topic:

- **Question 1: Legal framework** surrounding AI-related technologies, seeking input on their current permissibility in the rail security context.
- **Question 2: Regulatory needs** and expectations about each of the envisaged technologies/use cases.
- **Question 3: Technological maturity levels** of AI technologies for rail security, both at present and in the next years.
- **Question 4: Current deployment stage** (scouting for technologies/solutions, testing, operational deployment) of each of the envisaged technologies/use cases of AI.
- **Question 5: Operational use cases** (both at present and in the foreseeable future) for each of the technologies.
- **Question 6: Developers/Suppliers** of the AI technologies being scouted/tested/deployed.
- **Question 7: Railway companies' in-house AI development capabilities**
- **Question 8: Potential impacts of AI-related security technologies on the workforce.**
- **Question 9: Sectoral maturity of railways compared to other sectors** relating to the adoption of AI-technologies for security.
- **Question 10: Future challenges** related to the adoption of AI technologies for railway security.

2.2.2. Regulatory frameworks

The question addressed the legal framework surrounding AI-related use cases/technologies, seeking respondent's input on their current permissibility in various contexts:

"Can you indicate whether security personnel in your company are currently legally entitled to use the AI-related technologies listed in the following?"

Number of respondents: **11**

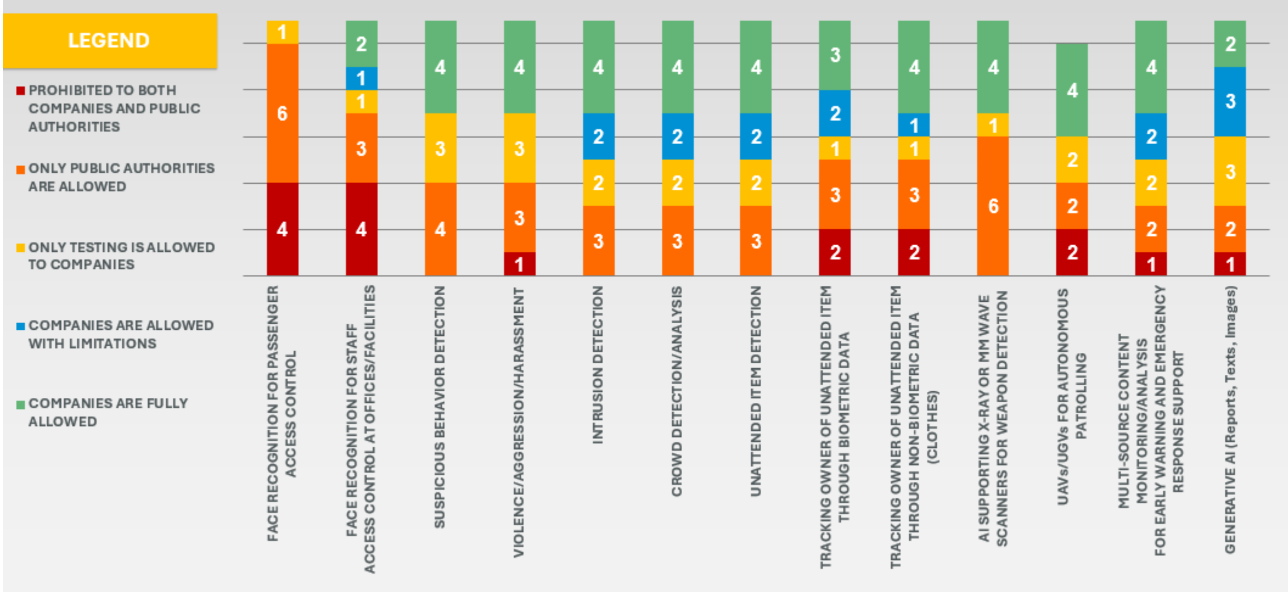


Figure 9 - UIC NTWG Survey on AI technologies in the railway security domain (2024) - Regulatory framework for AI applications

As a general remark, it should be noted that the landscape may evolve rapidly, since legislative processes are currently ongoing in many countries (as an example, the EU AI Act was not yet entered into force at the time of the survey).

National legal frameworks may vary significantly one from another but, in general terms, **non-EU Countries appear to currently be subject to less constraints** governing the usage of AI applications in the railway security domain.

Currently, the **most widely authorised/permitted** AI applications (at least for testing purposes) are represented by **intrusion detection**, **unattended object detection**, **crowd detection** and **multi-source content monitoring analysis**.

Face recognition for passenger access control is the only AI application among the investigated ones that, currently, is not operationally deployed in any of the railway companies.

The results seem to indicate that national frameworks, as of today, don't account for any difference between substantially different applications of AI, such as **Tracking owners of unattended items *by biometric data*** and ***by non-biometric data*** *(e.g., clothing)*.

2.2.3. Desired regulatory level

The question focused on the regulatory level – ranging from 'not regulated' to fully regulated through binding legislation – that expert judge suitable or desirable for each of the investigated use cases:

"According to your professional experience, do you estimate that security AI-related technologies need to be regulated to improve railway security while keeping a good balance with individual rights (e.g., privacy)? At which level?"

Number of respondents: **11**

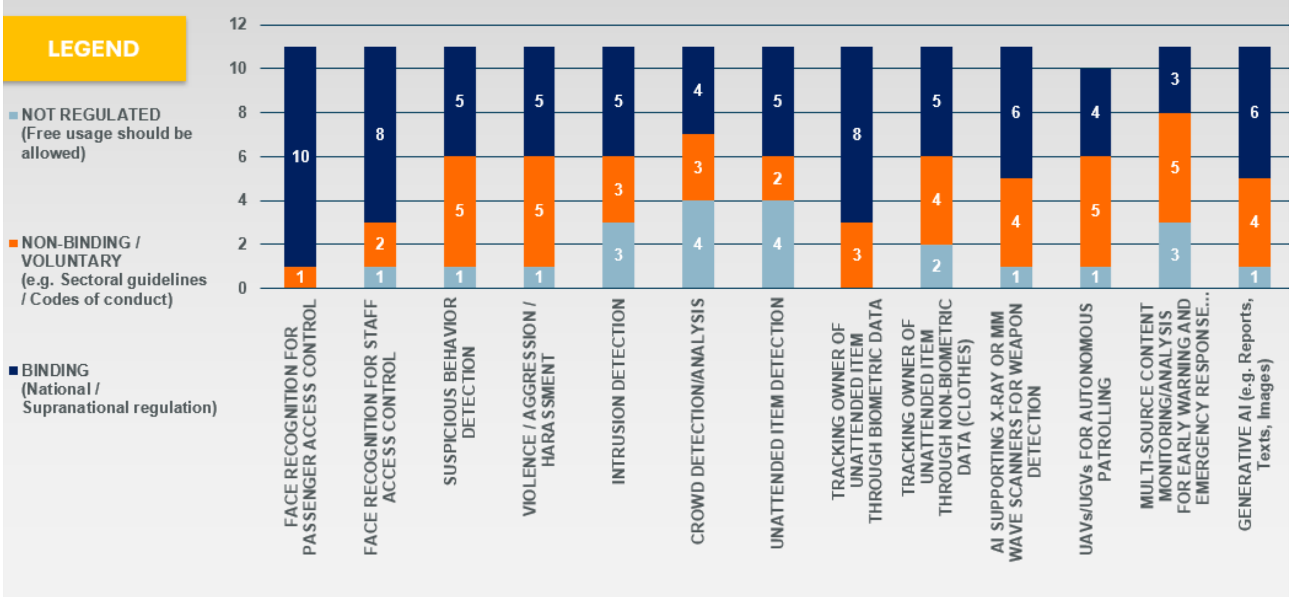


Figure 10 - UIC NTWG Survey on AI technologies in the railway security domain (2024) - Desired regulatory level for AI applications

The chart indicates that **AI applications involving human biometric data are largely regarded by experts as needing a 'strong' regulatory framework**. Binding regulation is considered as mostly desirable or needed for Face recognition for passenger access control (89.9%), Face recognition for staff access control to facilities/offices (72.7%) and Tracking of unattended item's owner through biometric data (72.7%).

On the contrary, **AI-powered technologies not involving the analysis of individuals** (e.g., Unattended item detection, Crowd detection, Multi-source content monitoring/analysis for early warning and emergency response support) **are judged to need** - according to the majority of respondents – either **a non-binding regulatory framework or to be unregulated**.

2.2.4. Current and estimated technological maturity levels

The question focused on the maturity level of investigated technologies, asking experts to estimate it both at present and in the foreseeable future:

"How high do you estimate the current maturity level of the technologies listed in the following? Do you estimate their maturity level to change significantly in the next 2-6 years?"

Number of respondents: **11**

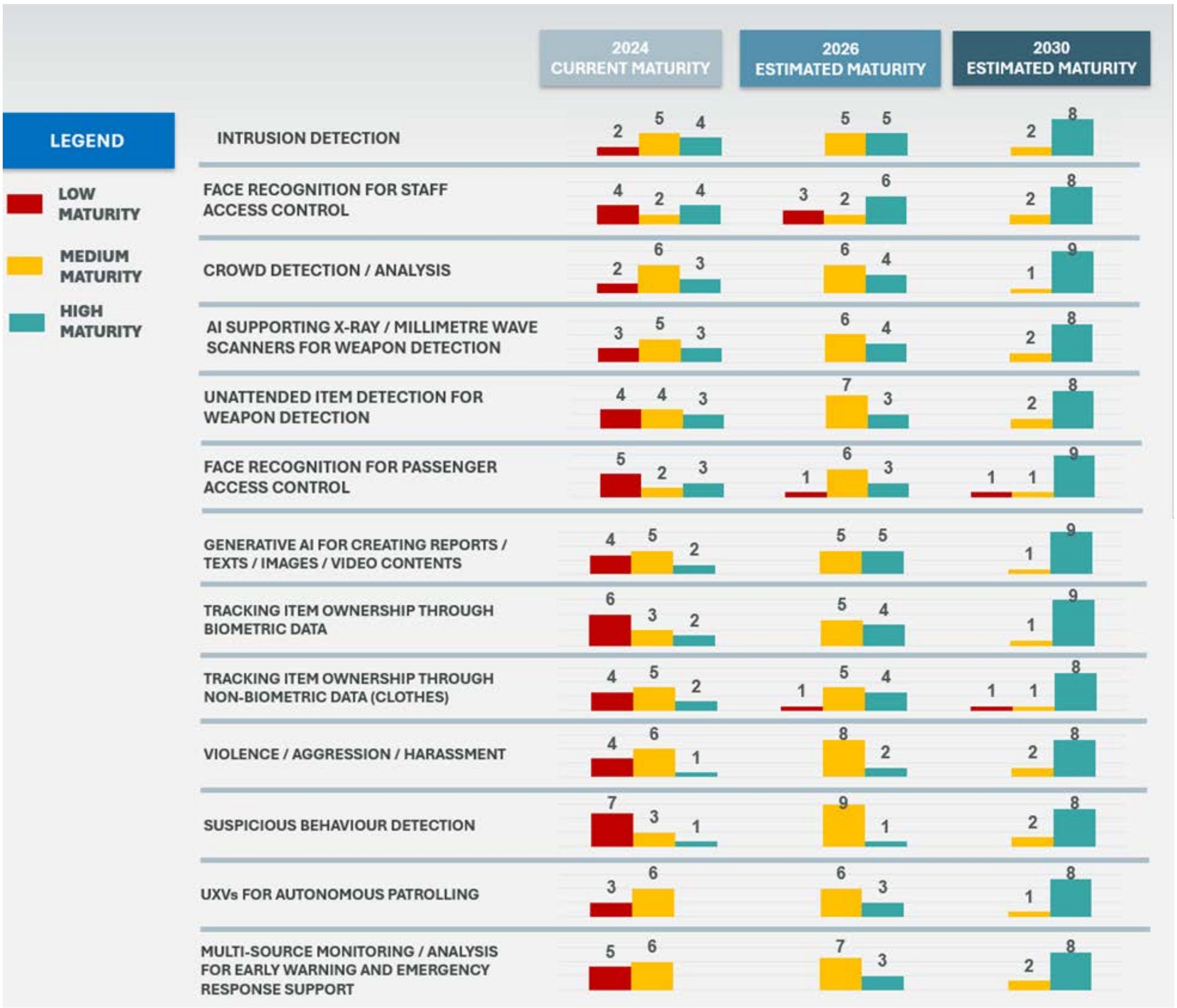


Figure 11 - Current and estimated technological maturity levels

In general terms, **most of the AI applications investigated are considered to currently display a low to medium level of technological maturity**.

Among them, **face recognition for staff access control** to premises/offices is currently estimated by the same number of respondents (4) at a low maturity level and at a high maturity level**.** This radical difference in the estimation could be attributed to the fact that some respondents may have referred to one-to-one reconnaissance technologies (generally judged to be more mature), while some others may have referred to one-to-many technologies.

Face recognition for passengers' access control will reach, according to most respondents (82%) a high level of maturity by 2030. According to one respondent (9%), though, this technology (presumably, referring to one-to-many reconnaissance technologies) will still display a low maturity level by then.

AI applications employing biometrics [Face recognition for passenger access control and Tracking owner of unattended item *through biometric data*] are mostly regarded as currently standing on a low technological maturity level. These results may also be affected by the impossibility to test such technologies in several countries due to restrictions imposed by legal frameworks.

Suspicious behaviour detection technologies are also largely judged to be still not mature (63.5% respondents). This may be mostly attributed to the difficulty of defining 'suspicious' behaviours deriving from a pattern of single actions that may be, *per se*, not suspicious nor illegal. Another challenging factor is represented by the significant risk of inadvertently incorporating biases in the algorithm training phase, ultimately leading to potential discrimination of individuals showing (even involuntarily) significative diversity from the norm in their behaviour.

Even if **Tracking owner of unattended item through *non-biometric data*** (e.g., clothes) will reach – according to 82% respondents – a high level of maturity by 2030, one expert judged it to be more challenging to achieve than using biometric data (estimating this application to still display a low maturity level in 2030).

In conclusion, the **technological landscape is deemed to be evolving very fast**: most applications will reach, according to most respondents, at least a medium maturity level within the next two years (by 2026).

By 2030, according to most respondents, **none of the investigated AI applications will stand on a low technological maturity level**. For all investigated use cases, a high level of maturity has been estimated by the majority of respondents.

2.2.5. Current deployment stage of AI technologies

The question aimed at assessing the deployment stages and challenges associated with AI technologies in the railway security domain:

"Are you/the Security function within your company currently using / testing / considering the following AI-powered technologies?"

Number of respondents: **10**

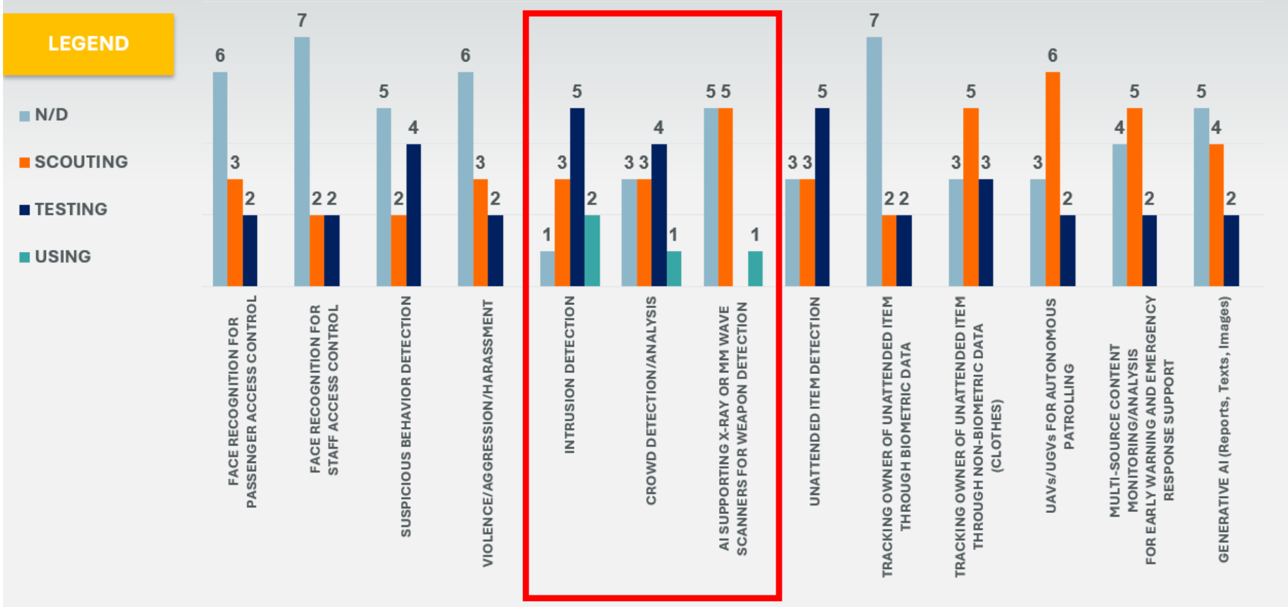


Figure 12 - Current deployment stage of AI technologies for railway security

At present, only three security-related AI technologies among the investigated ones, have been currently operationally deployed by railway companies: **Intrusion Detection** is used within 2 companies, while **Crowd Detection/Analysis** and **AI supporting X-Ray/Mm wave scanners for weapon detection** are currently employed by one company.

In general terms, two among the above-mentioned applications turned out to be considered, in absolute terms, the most exploitable according to sectoral experts. While 18% (2 companies) are currently using AI-based **Intrusion Detection**, 45.5% are testing it while further 27.5% are scouting for such technology. Results were similar for **Crowd Detection/analysis**, with one company (9%) using it, 36.5% testing and 27.5% scouting for such AI application.

On the contrary, the currently less 'appealing' technologies (in terms of exploitability) are **Face Recognition for staff access control** and **Tracking owner of unattended item through *non-biometric data*** (63.5% of experts declare that they are not using, testing or even scouting for these technologies), followed by **Face Recognition for passengers** and **Violence/Aggression/Harassment detection** (18% are testing and 27.5% are scouting these technologies).

Finally, it is worth noting that – even if not in used in any company – **Unattended Item Detection** shows a considerable level of interest among rail security practitioners, being currently tested by 45.5% of companies.

2.2.6. Current and foreseen use cases

The question aimed at inquiring about current and potential/anticipated use cases for AI-related security technologies:

"For each of the technologies that you are scouting for/testing/using, which are the current and/or expected use cases?"

Number of respondents: **11**

AI-powered **Intrusion detection systems** are currently regarded as a means to support security personnel in discovering trespassers gaining access to restricted areas at stations and other security-sensitive premises, in order to be able to alert Security Operation Centres / Authorities on time. It is also seen as a potentially effective tool to support efforts to tackle graffiti and vandalism issues.

Other use cases were also suggested, including monitoring of the railway infrastructure/line for illicit accesses to tracks and sensitive spots such as tunnels, or the possibility to set custom perimeters to be monitored for a specific purpose/time. Full integration into existing security management systems was described as a desired feature, as well as the possibility of generating automated reports to support the security analysis and risk assessment functions.

Crowd detection/analysis technologies were considered useful by experts for measuring density to mitigate (or possibly to anticipate) overcrowding situations (e.g., on platforms and at station entrances) and monitoring flows of people (with the possibility of receiving alerts for unexpected/sudden crowd movements), notably during big events.

The automated **detection of unattended items** is regarded as one of the most interesting features of AI systems, along with the possibility for it to be integrated/retrofitted to existing CCTV systems at stations or on-board trains. A complementary function of this application is represented by the possibility of **tracking the owner of the luggage** that has been left unattended. This may be achieved through two different technical systems, involving different impact levels on privacy. The first one relies on the acquisition and exploitation of **biometric data** of individuals (e.g., facial recognition), while the second one utilizes the **metadata of outfit characteristics** (e.g., colour/shape of clothes) to track individuals moving across railway premises.

Face recognition technologies, whose future development and adoption will be strictly tied to evolutions of regulatory frameworks, have been judged by experts as potentially useful for access control purposes in general, as well as for specific use cases such as enhancing access to platforms through fast lanes. That said, such technologies may be employed to ensure smooth and secure access to sensitive facilities (e.g., traffic control centres, cash management centres) for authorized staff, eliminating the need for badges or keys or even adding a supplementary layer of security.

AI systems for **suspicious behaviour detection** may be employed for spotting statistical abnormalities or unexpected behavioural patterns (e.g., loitering) and anticipate potentially harmful events, such as suicide attempts, by triggering timely intervention by operational personnel or authorities.

Detecting ongoing acts of violence or harassment (e.g., persons fighting) was envisaged as a future use case by experts, being potentially useful both at stations and on-board trains.

AI applications for **detecting weapons** concealed inside luggage or on human bodies, coupled with existing X-ray or millimetre-wave scanners can be trained to effectively spot cold/blade weapons, firearms, explosive materials. Advanced technologies such as pass-through portals would allow security checks with limited or no disruption to passenger flows. It was suggested, in addition, that these technologies may be employed at railway premises such as offices to carry out checks on company staff.

As AI-powered robotics are now one of the most promising fields of research and experimentation, different kinds of robots and **unmanned vehicles (UXVs)** may be employed for patrolling and monitoring purposes. Depending on the kind of vehicle (e.g., UAV, UGV, URV[15]) and on the array of sensors installed on-board, these vehicles may autonomously (or semi-autonomously, depending on technology and applicable regulations) support security functions such as track/station/perimeter monitoring and surveillance tasks in different conditions (day/night, adverse weather, dangerous or unsafe environments), sending out alerts to operation centres in case of events and/or taking out countermeasures directly (e.g., playing audio messages as a deterrent or even carrying/moving/delivering objects).

2.2.7. Technology suppliers and in-house development capabilities

The question aimed at analysing how AI security technologies are currently sourced by railway companies, whether purchased off-the-shelf, custom developed, or internally developed.

[15] UAV: Unmanned Aerial Vehicles (such as quadcopters, hexacopters, etc.); UGV: Unmanned Ground Vehicles equipped with wheels or similar systems for moving onto ground surfaces; URV: Unmanned Rail Vehicles, equipped with wheels for moving on the railway tracks.

"Who develops the AI-related technologies that you are currently scouting / testing / using?"

Number of respondents: **11**

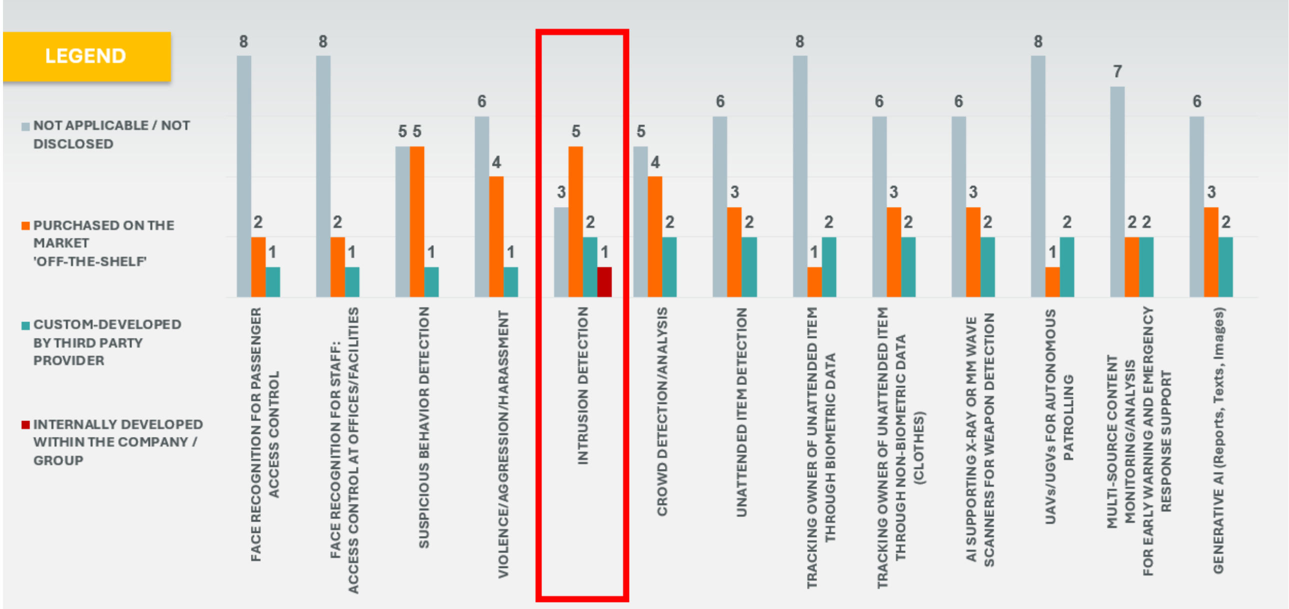


Figure 13 - Technology suppliers

Gathered data showed that, currently, companies are mostly scouting for, testing or adopting technologies purchased on the market (i.e., 'off-the-shelf'). Less frequently, they are purchasing solutions custom-developed upon their needs by external providers. Only in one case, a railway company has developed in-house a technical solution based on AI for intrusion detection.

Complimentarily to the latter question, a further enquiry was sent to experts regarding ongoing or future plans for railway companies to build up internal developing skills for AI-related applications.

"Is your company planning to achieve in-house developing skills for AI-related applications, especially in the security domain?"

Number of respondents: **7**

While just one company has already set up some capabilities to develop AI-related applications for security internally, most respondents judge that it is still too early to plan to achieve in-house development capabilities (it may depend, mostly, on future technical evolutions and regulatory frameworks).

Since developing such systems may require extensive expertise and technical equipment, some respondents estimated that a combination of solutions purchased off-the-shelf and custom ones developed on-demand by external providers will probably represent the most viable option.

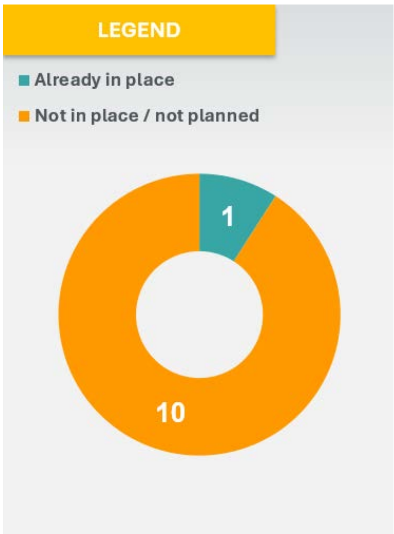


Figure 14 - In-house technology development capabilities

2.2.8. Impacts on the workforce

The question addressed the potential impacts of AI technologies on the railway security workforce, considering implications for current and new tasks, along with stress levels:

"Do you estimate that the employment of AI-related technologies in the railway security domain will have an impact on your workforce management / availability?"

Number of respondents: **11**

Answers in this regard were heterogeneous (likely because they may also vary depending on the different AI use cases), with an almost equal number of respondents judging that adopting AI technologies for security will:

- Determine the need for additional staff (at least, at an early stage) to take full advantage of its possible applications.
- Be neutral to the number of resources, helping those already employed to increase their efficiency and decrease stress levels.
- Free several human resources and make them available for other tasks.

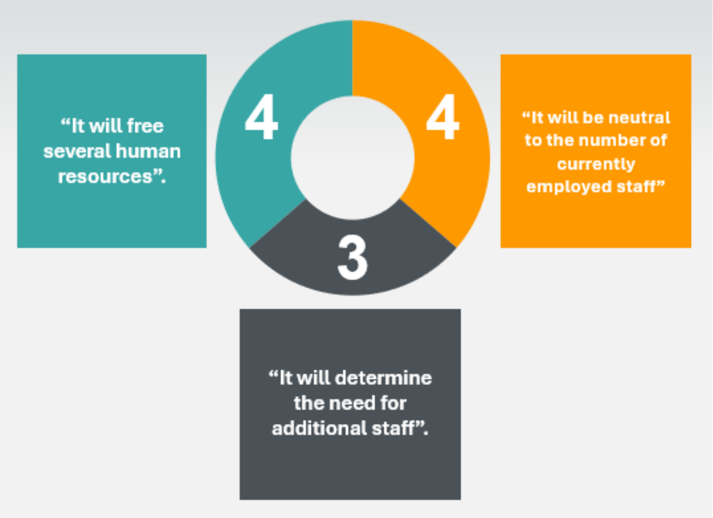


Figure 15 - Potential impacts of AI-related applications on railway companies' workforce

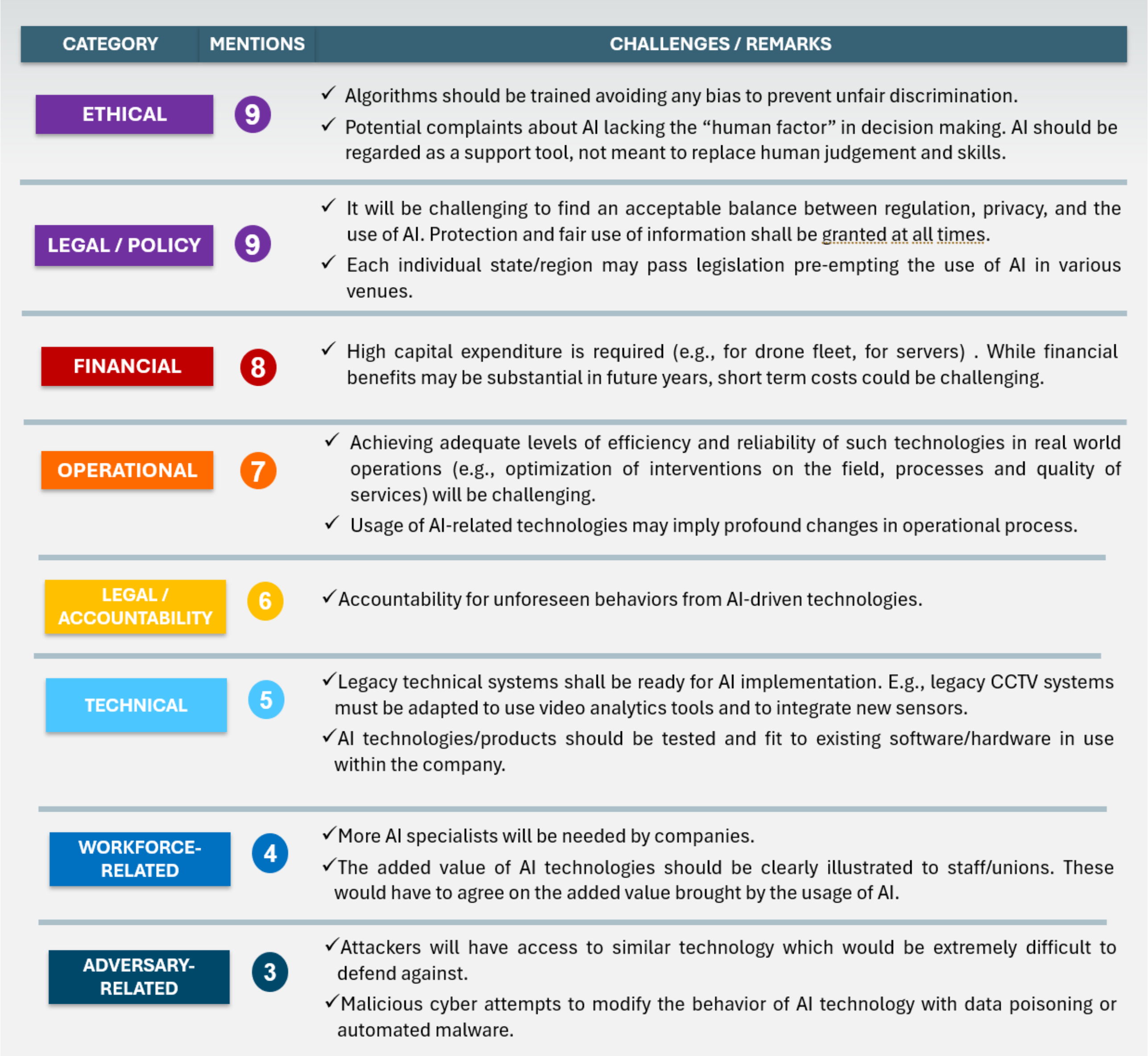
One expert argued that, even if the employment of AI technologies may enhance the quality of operations in many circumstances and tasks (e.g., incident management, detection, response, communication, reporting), **it will however unlikely result in a reduction of staff**. Companies will need qualified trainers to instruct staff to make correct and fair use of AI technologies. Furthermore, AI technology deployment is likely to occur gradually in phases, meaning the entire network will not be able to benefit from these technologies right from the start. Moreover, relevant phenomena and priorities are likely to evolve over time, necessitating continuous adaptation.

2.2.9. Foreseen challenges

The question solicited experts' insights into anticipated challenges in the steady evolution of AI technologies within railway security:

"According to your professional experience, what will be the main challenges that the railway security domain will face, relating to AI technologies?"

Number of respondents: **11**



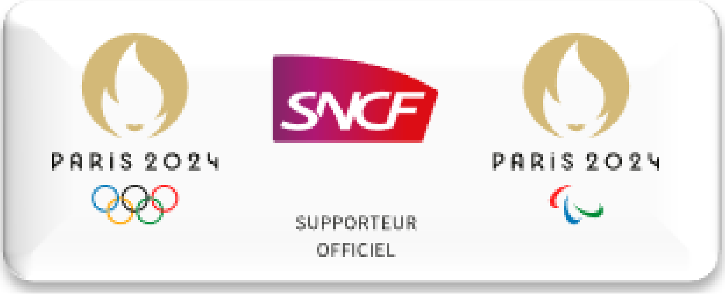
CATEGORY	MENTIONS	CHALLENGES / REMARKS
ETHICAL	9	✓ Algorithms should be trained avoiding any bias to prevent unfair discrimination. ✓ Potential complaints about AI lacking the "human factor" in decision making. AI should be regarded as a support tool, not meant to replace human judgement and skills.
LEGAL / POLICY	9	✓ It will be challenging to find an acceptable balance between regulation, privacy, and the use of AI. Protection and fair use of information shall be granted at all times. ✓ Each individual state/region may pass legislation pre-empting the use of AI in various venues.
FINANCIAL	8	✓ High capital expenditure is required (e.g., for drone fleet, for servers) . While financial benefits may be substantial in future years, short term costs could be challenging.
OPERATIONAL	7	✓ Achieving adequate levels of efficiency and reliability of such technologies in real world operations (e.g., optimization of interventions on the field, processes and quality of services) will be challenging. ✓ Usage of AI-related technologies may imply profound changes in operational process.
LEGAL / ACCOUNTABILITY	6	✓Accountability for unforeseen behaviors from AI-driven technologies.
TECHNICAL	5	✓Legacy technical systems shall be ready for AI implementation. E.g., legacy CCTV systems must be adapted to use video analytics tools and to integrate new sensors. ✓AI technologies/products should be tested and fit to existing software/hardware in use within the company.
WORKFORCE-RELATED	4	✓More AI specialists will be needed by companies. ✓The added value of AI technologies should be clearly illustrated to staff/unions. These would have to agree on the added value brought by the usage of AI.
ADVERSARY-RELATED	3	✓Attackers will have access to similar technology which would be extremely difficult to defend against. ✓Malicious cyber attempts to modify the behavior of AI technology with data poisoning or automated malware.

Figure 16 - Foreseen challenges towards the operational deployment of AI-related applications for railway security

3. CASE STUDY: SNCF’S OPERATIONAL EXPERIMENT DURING PARIS 2024 OLYMPIC AND PARALYMPIC GAMES

3.1. Introduction

While ‘normal life’ for Parisians and the usual summer tourism was still to unfold regularly, 15 million additional tourists were expected to visit Paris during the 2024 Olympic and Paralympic Games, held respectively from July 26th to August 11th and August 28th to September 8th.



Managing transportation during the Olympic Games was therefore comparable to dealing with a seamless rush-hour situation for the whole event duration (e.g., mass transportation to Stade de France – capable of hosting 80,000 persons – had to be ensured three times per day on average, for multiple consecutive days). In this context, ensuring safety and security to visitors and citizens was paramount.

These huge numbers emphasize the importance of such a once-in-a-lifetime event for French railways (SNCF) and its employees. **Carefully planning all the security aspects relating to the Olympic Games was crucial,** which started within SNCF 5/6 years before the event, and technology was identified as one of the main drivers in this regard.

In this context, SNCF had an important opportunity to carry out operational tests on Artificial Intelligence applied to CCTV image analysis during the Games, while only technical tests had been carried out in the past.

3.1.1. The French regulatory framework for testing AI technologies

A **special regulatory framework** that allowed SNCF to run tests on AI technologies was enacted by the French Government in May 2023 (the so-called JOP24 Law[16]), for a duration of two years until March 31st, 2025. The scope of the framework limited the testing activities to sports, recreational activities and high-risk events as defined by the French Government, that maintained general control on the experiment also by selecting the software solutions to be tested through a qualification process and issuing specific prefectural authorizations for each event where tests were to take place.

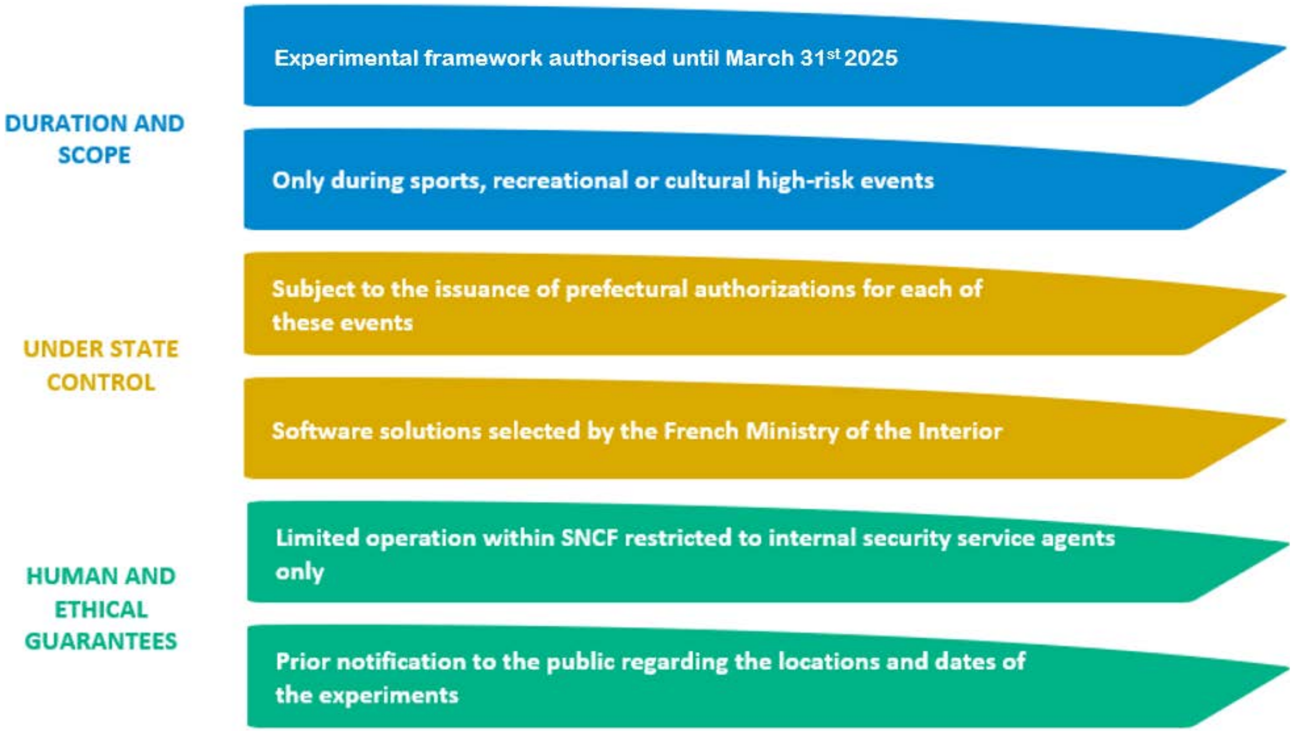

DURATION AND SCOPE
Experimental framework authorised until March 31st 2025
Only during sports, recreational or cultural high-risk events
UNDER STATE CONTROL
Subject to the issuance of prefectural authorizations for each of these events
Software solutions selected by the French Ministry of the Interior
HUMAN AND ETHICAL GUARANTEES
Limited operation within SNCF restricted to internal security service agents only
Prior notification to the public regarding the locations and dates of the experiments

Figure 17 - The overall regulatory framework for SNCF testing activities on AI.

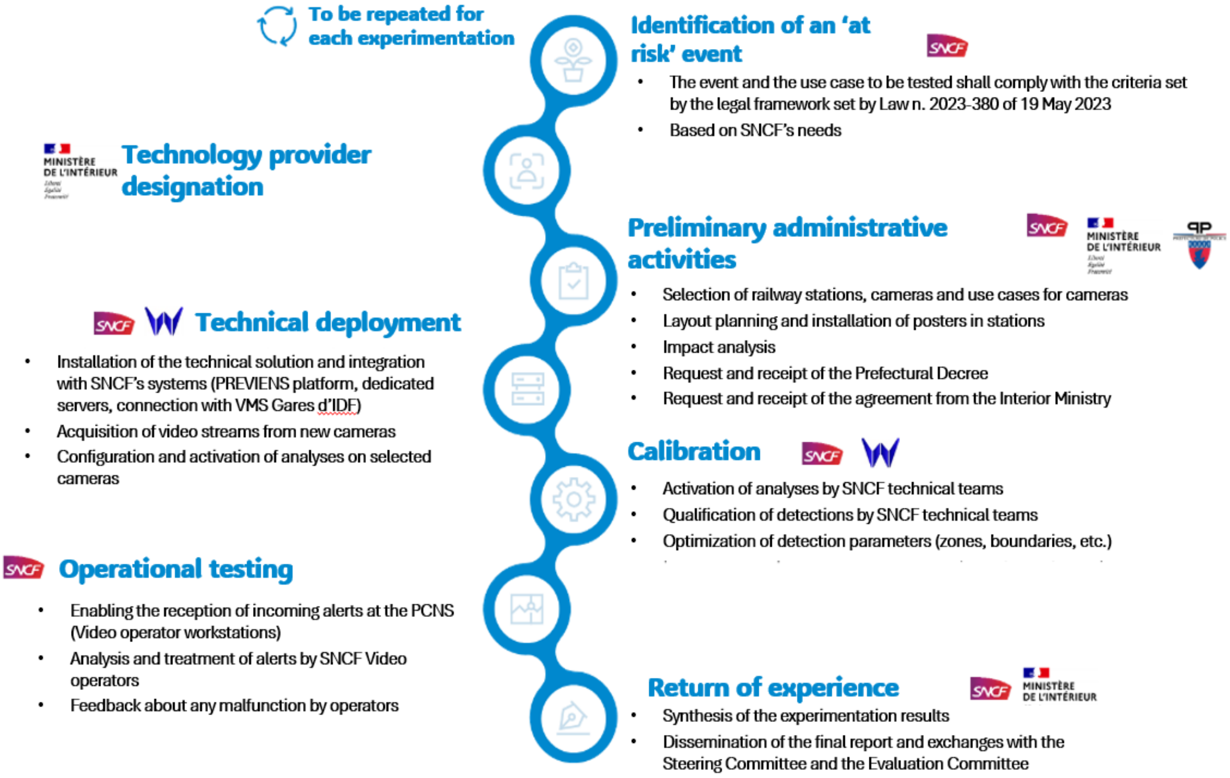
In accordance with Article 10 of the JOP2024 Law, the implementation of augmented video experiments had to be carried out within a restricted framework and to stick to the following key principles (see Picture 13):

- The selection of technological solutions to be tested was made by the Ministry of the Interior within the framework of a national public procurement process conducted from August to December 2023 and made available to SNCF through a dedicated agreement.

The **technical solution selected for testing was Cityvision by WINTICS**.

- Prior issuance of a prefectural authorization had been needed for each implementation of the experiments, whose validity period could not exceed one month.
- Compliance in processing of personal data had to be ensured also through the completion of a Data Protection Impact Assessment (DPIA), to be submitted to the **National Commission for Information Technology and Liberties** (**CNIL**).

[16] LOI n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions. Available online at : LOI n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions (1) - Légifrance



To be repeated for each experimentation

Identification of an 'at risk' event

- The event and the use case to be tested shall comply with the criteria set by the legal framework set by Law n. 2023-380 of 19 May 2023
- Based on SNCF's needs

Technology provider designation

Preliminary administrative activities

- Selection of railway stations, cameras and use cases for cameras
- Layout planning and installation of posters in stations
- Impact analysis
- Request and receipt of the Prefectural Decree
- Request and receipt of the agreement from the Interior Ministry

Technical deployment

- Installation of the technical solution and integration with SNCF's systems (PREVIENS platform, dedicated servers, connection with VMS Gares d'IDF)
- Acquisition of video streams from new cameras
- Configuration and activation of analyses on selected cameras

Calibration

- Activation of analyses by SNCF technical teams
- Qualification of detections by SNCF technical teams
- Optimization of detection parameters (zones, boundaries, etc.)

Operational testing

- Enabling the reception of incoming alerts at the PCNS (Video operator workstations)
- Analysis and treatment of alerts by SNCF Video operators
- Feedback about any malfunction by operators

Return of experience

- Synthesis of the experimentation results
- Dissemination of the final report and exchanges with the Steering Committee and the Evaluation Committee

Figure 18 - Implementation process to be followed for each experiment.

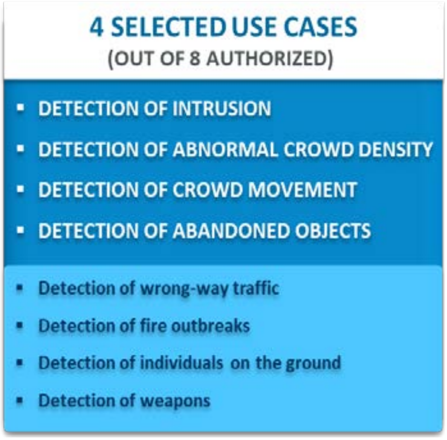
Ethical and human guarantees were in place to ensure respect for privacy and data treatment. The public had to be duly informed that an experiment involving analytics and personal data was being carried out, both online through the SNCF website and at stations.

It is also important to note that all the outputs of tested analytics (e.g., metadata, bounding boxes), excluding the alerts, had to be deleted after analysis to prioritize ethical considerations.

3.2. The testing plan

3.2.1. SNCF's overall approach to AI technologies testing

Eight use cases were allowed for testing within the legal framework, including the detection of: intrusions, abnormal crowds, crowd movements, abandoned objects, wrong way traffic, fire outbreaks, individuals on the ground, and weapons; however, **SNCF chose to execute only the first four** due to various considerations (non-applicability at stations, likelihood of getting a high rate of false alerts, relatively low technological maturity or added value brought by the solution).


4 SELECTED USE CASES
(OUT OF 8 AUTHORIZED)

- DETECTION OF INTRUSION
- DETECTION OF ABNORMAL CROWD DENSITY
- DETECTION OF CROWD MOVEMENT
- DETECTION OF ABANDONED OBJECTS

- Detection of wrong-way traffic
- Detection of fire outbreaks
- Detection of individuals on the ground
- Detection of weapons

Figure 19 - Selected use cases for testing

SNCF's overall experimental approach and methodology towards implementing AI technologies can be segmented into five steps:

STEP 1

Clearly **understanding the user needs** (also through exchanges with operators within C&C), therefore defining and prioritizing relevant use cases.

STEP 2

The second step was **gaining clear knowledge about which personal data were to be involved** in the acquisition and analysis activities and duly performing all the related activities (DPIA, security risk analysis, contractual commitments, and information to involved parties/individuals) to ensure compliance with regulations and ethics (e.g., avoid introducing biases in the training phase).

STEP 3

The third step was represented by **setting up technical experiment** (even on multiple solutions) to **assess TRLs** (technological readiness level) **and ensuring the proper functioning of the algorithm** (e.g., measuring the false alert rate) before handing it over for field tests.

STEP 4

The fourth step was a **'real-world' pilot phase involving video operators at C&C Centres** for extensive and careful performance evaluation in real life scenarios. Checking and adjusting operational processes to the new technologies was inevitably considered as an iterative activity.

STEP 5

The fifth step would correspond to the **full deployment of the solution**. Challenges in this regard may be represented by scaling up the platform, servers and deploying the solution to the entire infrastructure, estimating the ROI (Return on Investment), setting up partnerships and calls for tenders.

3.2.2. Communication towards the public

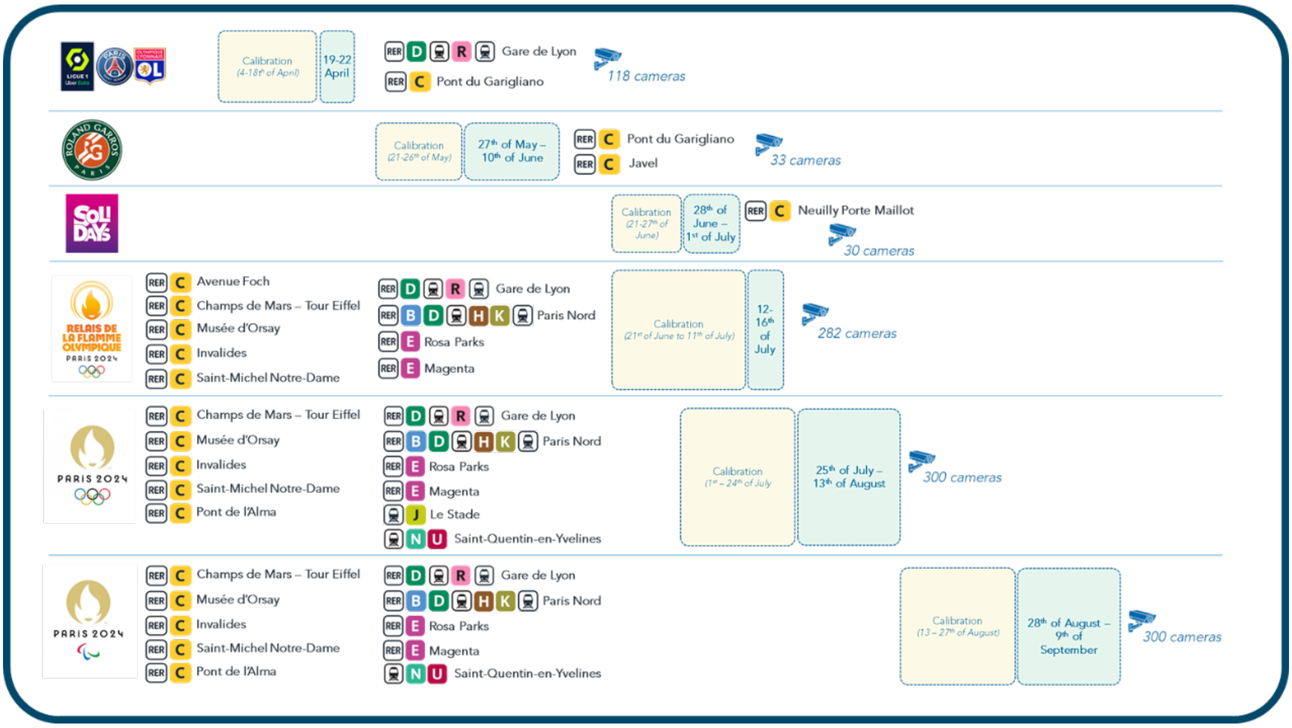
Clear and comprehensive communication with the public was crucial.

In compliance with GDPR, and to ensure proper notification to transport users circulating in the designated experimental areas, over 150 specific information posters were placed at all the locations concerned for the entire duration of the experiment.

Both content and design of such information panels were developed in collaboration with French authorities to effectively inform individuals about personal data usage and ongoing testing activities.

Additionally, a detailed informational notice, accessible via QR code, was made available on a dedicated SNCF webpage (accessible at: [Videosurveillance analysis experiment | SNCF Group](https://www.groupe-sncf.com/en/information/video-emergency-call/videosurveillance-analysis-experiment)[17]). The webpage explains the procedures for implementing the experiment and allows individuals to exercise their rights.

3.2.3. The Paris 2024 testing schedule



Calibration (4-18th of April)
19-22 April
RER D R Gare de Lyon
RER C Pont du Garigliano
118 cameras
Calibration (21-26th of May)
27th of May – 10th of June
RER C Pont du Garigliano
RER C Javel
33 cameras
Calibration (21-27th of June)
28th of June – 1st of July
RER C Neuilly Porte Maillot
30 cameras
RELAIS DE LA FLAMME OLYMPIQUE PARIS 2024
RER C Avenue Foch
RER C Champs de Mars – Tour Eiffel
RER C Musée d'Orsay
RER C Invalides
RER C Saint-Michel Notre-Dame
RER D R Gare de Lyon
RER B D H K Paris Nord
RER E Rosa Parks
RER E Magenta
Calibration (21st of June to 11th of July)
12-16th of July
282 cameras
PARIS 2024
RER C Champs de Mars – Tour Eiffel
RER C Musée d'Orsay
RER C Invalides
RER C Saint-Michel Notre-Dame
RER C Pont de l'Alma
RER D R Gare de Lyon
RER B D H K Paris Nord
RER E Rosa Parks
RER E Magenta
J Le Stade
N U Saint-Quentin-en-Yvelines
Calibration (1st – 24th of July
25th of July – 13th of August
300 cameras
PARIS 2024
RER C Champs de Mars – Tour Eiffel
RER C Musée d'Orsay
RER C Invalides
RER C Saint-Michel Notre-Dame
RER C Pont de l'Alma
RER D R Gare de Lyon
RER B D H K Paris Nord
RER E Rosa Parks
RER E Magenta
N U Saint-Quentin-en-Yvelines
Calibration (13 – 27th of August)
28th of August – 9th of September
300 cameras

Figure 21 - SNCF's 2024 Games testing schedule

The experiment schedule was dense, starting in April 2024, so as to gain as much experience as possible through several other events before the Olympic and Paralympic Games.

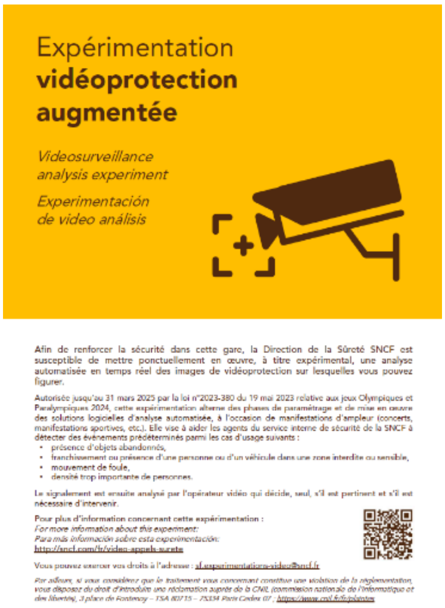
Each testing phase was preceded by a calibration phase for setting up and optimizing the camera/system parameters.

[17] Accessible online at: https://www.groupe-sncf.com/en/information/video-emergency-call/videosurveillance-analysis-experiment

The **first experiment** was carried out during a **national football game** on the days from 19th to the 22nd of April 2024. It involved 118 cameras installed at 3 stations (including the major hub Gare de Lyon).

Two **smaller-scale tests** were conducted shortly after: the first took place during the **Roland Garros tournament** (May 27th to June 10th) and involved the usage of 33 cameras deployed at 2 stations of the RER C line. The second one was carried out during the **SoliDays event** (June 28th to July 1st), with 30 cameras active at a different station on the same line. Those tests were very important because of their duration, especially for Roland Garros which took place over more than 2 weeks. They enabled the Video operators to really handle the analytics solution and to experience the management of the alerts procedure.

The **first test on a greater scale**, **related to 2024 Olympics**, was carried out from July 12th to 16th for the **arrival of the Olympic flame in Paris**. It involved 9 stations, including two major hubs (Gare de Lyon and Gare du Nord), and a total of 282 cameras.



Expérimentation
vidéoprotection
augmentée

Videosurveillance analysis experiment

Experimentación de video análisis

Afin de renforcer la sécurité dans cette gare, la Direction de la Sûreté SNCF est susceptible de mettre ponctuellement en œuvre, à titre expérimental, une analyse automatisée en temps réel des images de vidéoprotection sur lesquelles vous pouvez figurer.

Autorisée jusqu'au 31 mars 2025 par la loi n°2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques 2024, cette expérimentation alterne des phases de paramétrage et de mise en œuvre des solutions logicielles d'analyse automatisée, à l'occasion de manifestations d'ampleur (concerts, manifestations sportives, etc.). Elle vise à aider les agents du service interne de sécurité de la SNCF à détecter des événements prédéterminés parmi les cas d'usage suivants :

- présence d'objets abandonnés,
- franchissement ou présence d'une personne ou d'un véhicule dans une zone interdite ou sensible,
- mouvement de foule,
- densité trop importante de personnes.

Le signalement est ensuite analysé par l'opérateur vidéo qui décide, seul, s'il est pertinent et s'il est nécessaire d'intervenir.

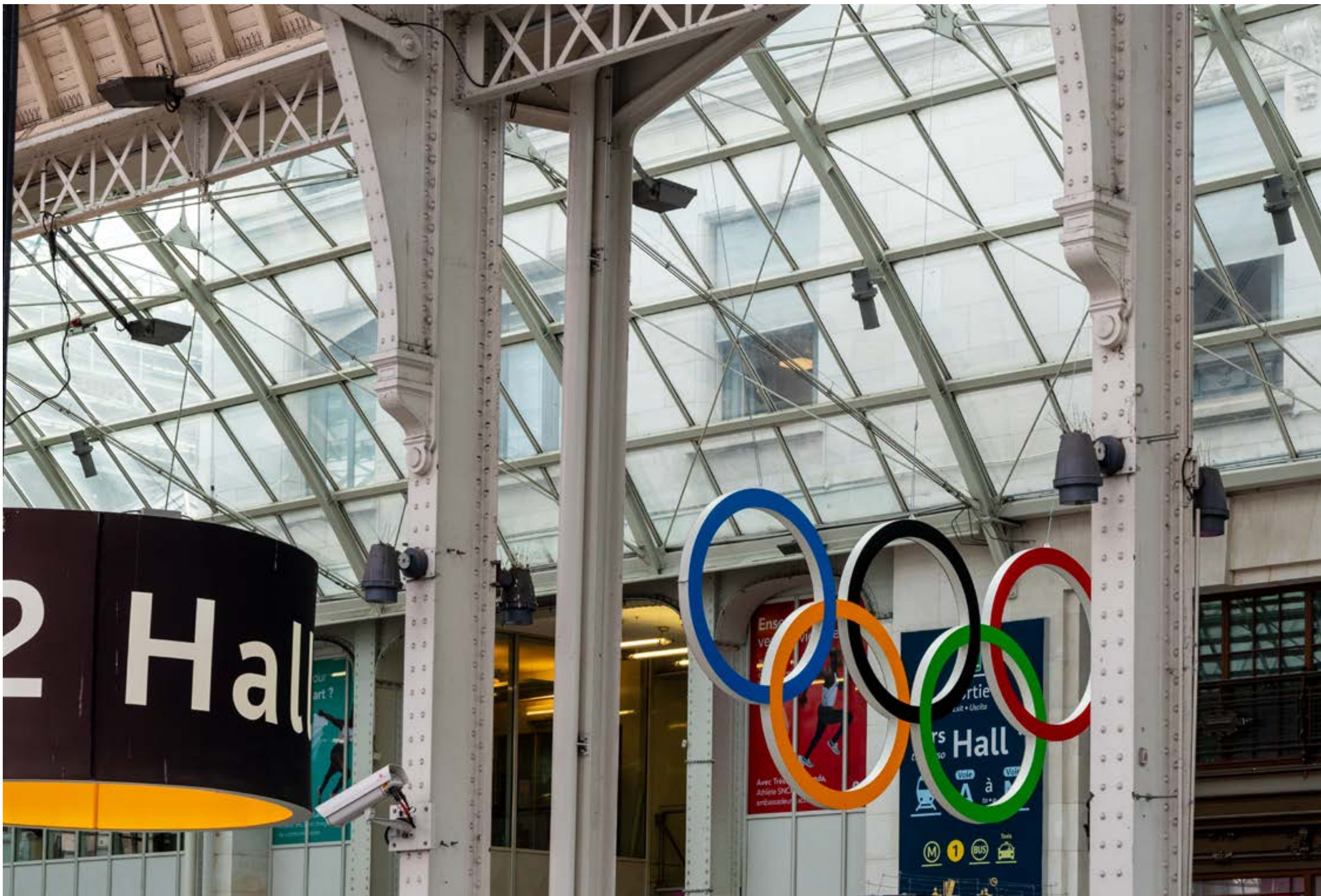
Pour plus d'information concernant cette expérimentation :
For more information about this experiment:
Para más información sobre esta experimentación:
http://sncf.com/fr/video-appels-surete

Vous pouvez exercer vos droits à l'adresse : sf.experimentations-video@sncf.fr

Par ailleurs, si vous considérez que le traitement vous concernant constitue une violation de la réglementation, vous disposez du droit d'introduire une réclamation auprès de la CNIL (commission nationale de l'informatique et des libertés), 3 place de Fontenoy – TSA 80715 – 75334 Paris Cedex 07 ; https://www.cnil.fr/fr/plaintes

Figure 20 - SNCF communication campaign on AI testing

The most extensive experiments were finally made during the **Olympic Games**, from July 25th to August 13th, as well as during the **Paralympic Games** (28th of August to 9th September), as described in the following paragraphs.



3.3. Technology

3.3.1. The overall approach

SNCF has an extensive CCTV system, including 80,000+ cameras. In the **Paris area only, around 11,000 cameras** are installed **at 400 stations**. The network is centralized, and all the footage is stored in a secure data centre (with a maximum retention period of 30 days).

This system is coordinated by a **Video Management System (VMS)** by GENETEC allowing the visualization of images in real-time and delayed, as well as the extraction of images subject to judicial requisition on operator stations. Furthermore, all these cameras, operator stations, and VMS servers are interconnected via a multi-service network initially deployed for the needs of the video surveillance system, relying on 'multicast' technology.

In preparation for major events in 2023 and 2024, several preliminary actions have been implemented by the Security Directorate of the SNCF Group, SNCF Gares & Connexions, and SNCF Transilien to optimize the video surveillance service.

Firstly, the Security Directorate conducted performance diagnostics of the video surveillance systems at 42 stations concerned by the 2024 Olympic and Paralympic Games. Based on the recommendations made, SNCF Gares & Connexions and SNCF Transilien, with the support of Île-de-France Mobilités, implemented optimizations and replacements of existing cameras, as well as the deployment of nearly 500 new cameras.

Thus, during major events, all 11,000 cameras in Île-de-France were able to be utilized, as is usually the case, by the users of the Île-de-France video surveillance system, notably by the agents of the SNCF internal security service.

3.3.2. The PREVIENS Experimentation Platform

Alongside this video surveillance system, the SNCF Security Directorate has deployed, since 2017, an **experimentation platform for augmented video solutions** as part of its **PREVIENS program**.

This platform, interconnected with the Île-de-France video surveillance system, allows the secure use of any of the 11,000 existing cameras to evaluate internal or external image analysis solutions under real conditions.

It relies on high-performance technical equipment (servers and CPU, RAM, GPU components) forming the foundation on which augmented video solutions under test are installed, including the Cityvision solution proposed by WINTICS as part of the experiments covered in this report.

The diagram below details the architecture of the PREVIENS platform and its interconnection with the existing video surveillance system.

As part of the experiments conducted by SNCF, **up to 8 servers** were made available to WINTICS to allow **real-time analysis of up to 300 cameras simultaneously**, on which **a total of 420 algorithmic processes** were applied.

The employed servers, each valued at approximately 15,000 euros, consist of the following high-performance components:

- **6 NVIDIA Tesla T4 graphics cards**
- **2 Intel Xeon Silver 4216 processors**
- **128 GB of RAM**
- **4 TB of SSD storage.**

It is crucial to note that – for cybersecurity reasons – no remote access to this platform is possible. Each external participant (WINTICS, resources from the Ministry of the Interior and Overseas, members of evaluation or steering committees) had to be on-site in order to access the platform, and signing their access on a register. Furthermore, given the personal nature of data accessible through the platform, each participant from WINTICS had to read and sign an individual confidentiality agreement and the PREVIENS experimentation platform usage policy before any intervention.

3.3.3. Data Treatment

In the architecture implemented by SNCF, the images from the cameras selected by the Security are not retained by the algorithmic processes, which are limited to real-time analysis and generation of reports in case of detection. This report is displayed in the Genetec software by the notification of an alert specifying the use case concerned, the camera concerned, as well as the date and time of detection. This alarm is accompanied by a “bookmark” or marker positioned at the time of detection in the video recording accessible only via the Genetec software.

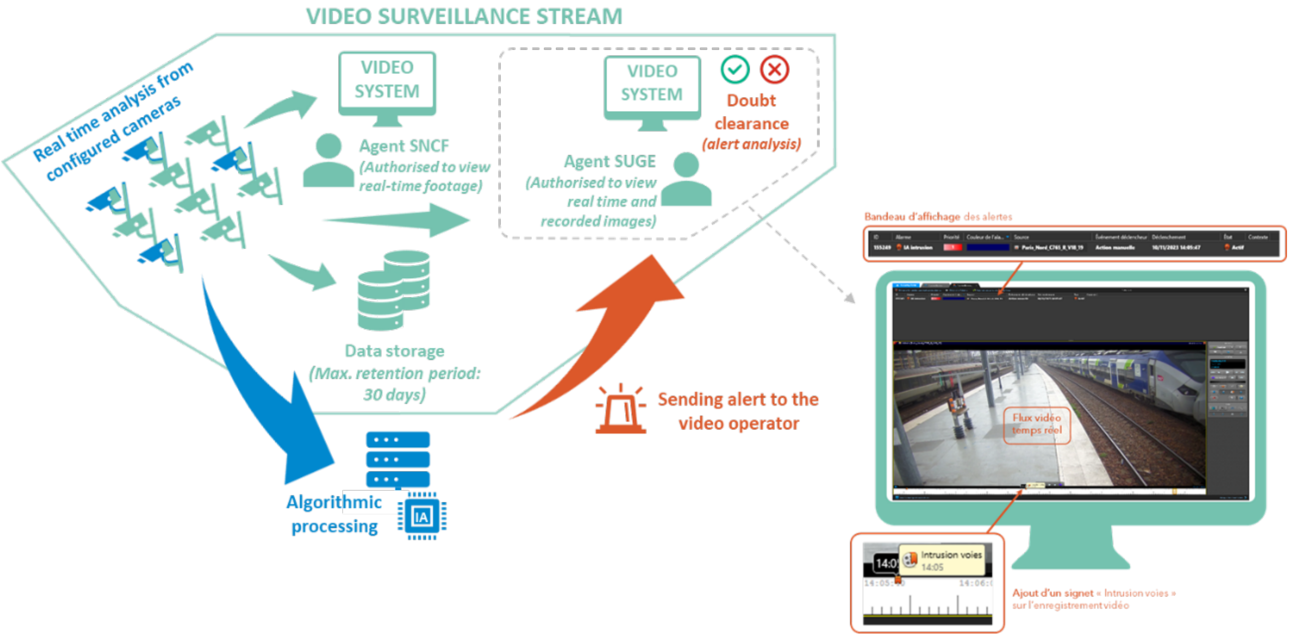


Figure 23 - Overall structure of the tested system

The diagram below illustrates the operating principle of the augmented video system in parallel with the existing video protection system.

The **processes implemented do not exploit any biometric data** (e.g., facial features, gait) and are **limited to classifying the content of the images** (people, objects, trains). Based on the configuration (detection zone, duration of presence, number of people, speed of movement, etc.) and circumstances, they generate an alert/event report.

Furthermore, no element highlighting the source of the alert/report is displayed to the video operator or retained by the Cityvision solution. Thus, the video operator must, through their own analysis, verify the relevance of the alert/report generated by the algorithmic processes in order to acknowledge the associated alarm.

Finally, activity logs ensure the traceability of all actions performed by the system or users, particularly for audit purposes.

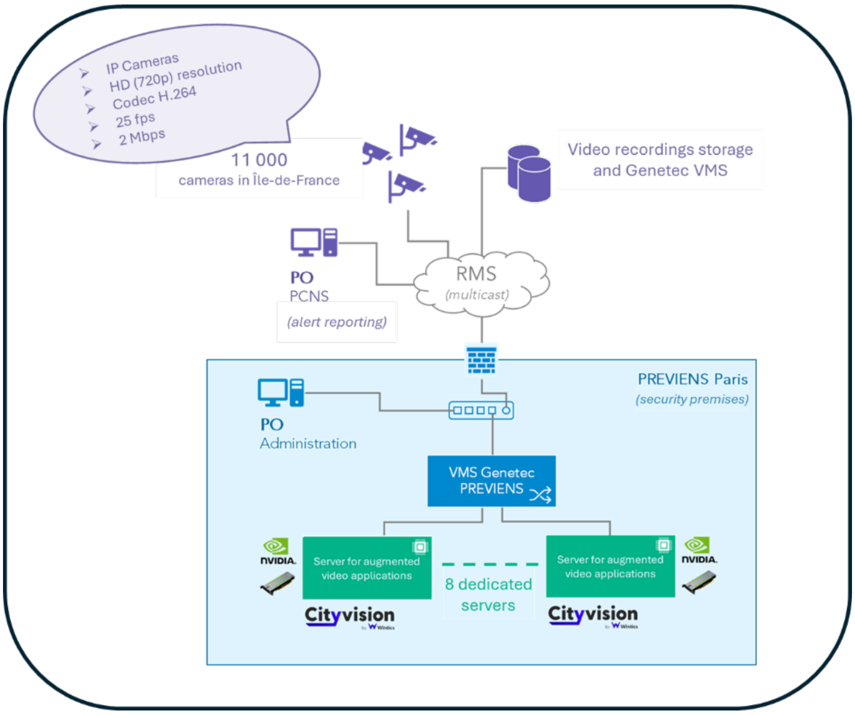


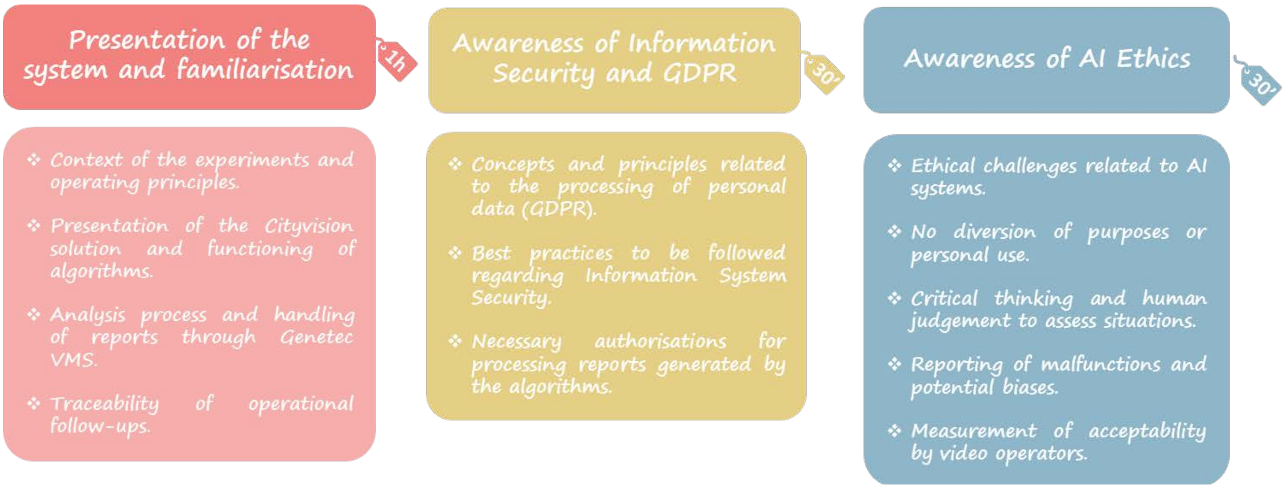
Figure 22 - The PREVIENS Experimentation Platform

3.3.4. Staff training

As established within the legal framework ruling the experiment, only **specifically appointed and trained SNCF security personnel** were entitled to test the technology.

Eleven SNCF agents in total took part in the experiment:

- **Project team**: 3 agents in charge of managing the experiments and supporting SNCF users.
- The team members received a specific training from WINTICS to acquire the necessary skills for administering the Cityvision solution (configuration, activation and deactivation of processes, etc.).
- **Video operators**: 8 operational agents (enrolled on a voluntary basis), forming the video patrol team of the National Security Command Post (PCNS).



Presentation of the system and familiarisation

1h

- Context of the experiments and operating principles.
- Presentation of the Cityvision solution and functioning of algorithms.
- Analysis process and handling of reports through Genetec VMS.
- Traceability of operational follow-ups.

Awareness of Information Security and GDPR

30'

- Concepts and principles related to the processing of personal data (GDPR).
- Best practices to be followed regarding Information System Security.
- Necessary authorisations for processing reports generated by the algorithms.

Awareness of AI Ethics

30'

- Ethical challenges related to AI systems.
- No diversion of purposes or personal use.
- Critical thinking and human judgement to assess situations.
- Reporting of malfunctions and potential biases.
- Measurement of acceptability by video operators.

Figure 24 - Training program for C&C operators testing AI at SNCF

Each of the video operators taking part in the tests had to be formally authorized to take part to the experiment. The clearance could be successfully achieved after the successful completion of **a training program consisting of three modules**: introduction to the system, analyse alerts (1 hour); awareness of personal data protection (GDPR and information security, 30 minutes); awareness about ethical aspects of AI (30 minutes). The primary focus was on protecting personal data and adhering to the EU regulatory framework on AI.

It is important to note that only the agents of the SNCF project team had direct access to the Cityvision solution in order to avoid any accidental errors or potential misuse of the implemented processes (modification of settings, deactivation of processes, etc.). The operational agents have thus only used their usual video management software (VMS Genetec), enhanced by the reception of alerts.

Finally, it is paramount to underline that **the aim of the test was not to replace video operators with technology**, **but to empower them** by providing more time to make decisions based on processed data, **keeping it as a 'human-centric' activity**.

3.3.5. Operational implementation in the SNCF's Command & Control Centre – Olympic and Paralympic Games

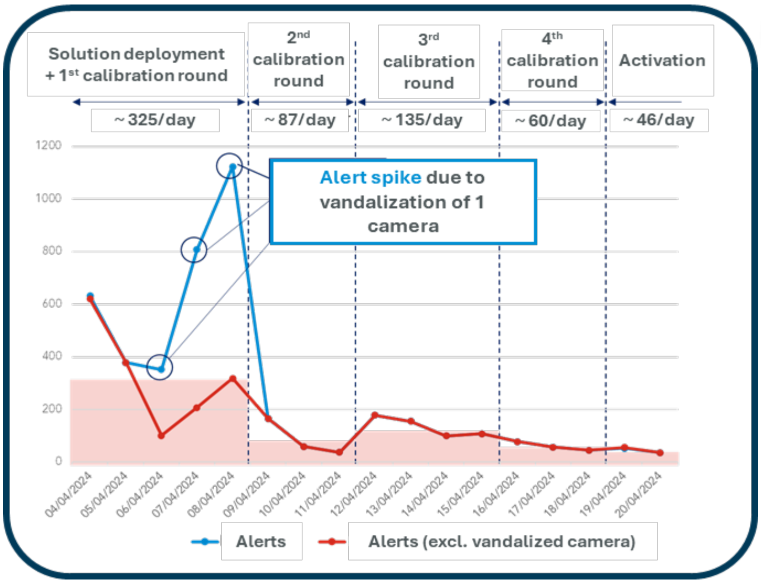
Calibration phase

Timespan: July 1st - 24th 2024 for Olympic Games; August 20th - 27th 2024 for Paralympic Games.

Pre-operational phase of **installation**, **calibration** and **definition of operational parameters** was paramount, as analytics solutions don't work properly just out of the box. Instead, these need to be re-calibrated several times to achieve algorithm optimization. In this phase, it must be ensured that all the variables (e.g., weather, lighting, attendance, ongoing construction works) are taken into due account. The **goal of this phase was to reduce the number of false alerts** and improve the quality of the results by studying the detected events and consequently adjusting timing, detection zones, and alert thresholds.

For each camera, many different parameters were set and – progressively - fine-tuned:

- **Detection zones**, allowing the delimitation of different analysis areas in the image (accessible by SNCF and WINTICS administrators).
- **Detection thresholds** for people, trains, and objects allowing the definition of minimum confidence scores regarding the classification performed by the AI (restricted to WINTICS administrators only).
- **Detection rules** for different use cases (e.g., duration of presence in an area, number of persons) that trigger reports (partially accessible by SNCF and WINTICS administrators).
- **Time slots/schedules** for activating analyses, allowing the definition of rules about times (accessible by SNCF and WINTICS administrators).
- **Minimum delay between two detections** to limit the number of reports generated for the same event on a camera (accessible by SNCF and WINTICS administrators).
- As shown in the figure below, multiple calibration rounds progressively brought down the number of alerts from 600+ to around 40 per day, eliminating false positives for the most part.



CALIBRATION OPERATIONS

- ✓ **OBJECTIVE :** Limiting the number of false alarms and ensuring optimal performance
- ✓ **Multiple calibration operations** are necessary to take different variables into account :
  - ➢ Weather
  - ➢ Attendance in the station
  - ➢ Construction works
  - ➢ New train station furniture
- ✓ These operations mainly involve studying the detected events and **modifying the parameters** to prevent false alarms.:
  - Detection zones
  - Timing
  - Detection thresholds for people, objects, and trains
  - Time windows for activating analyses

Figure 25 - Variation in the number of alerts over the calibration phase

Deployment phase

Timespan: July 25th to August 13th, 2024, for **Olympic Games**; August 28th – September 9th, 2024 for **Paralympic Games**.

During the events, **316 cameras** (300 max. simultaneously in use) were employed across – respectively – 11 and 10 stations in Paris (see above, Figure 16).

All the video patrol tables allocated to the Paris area were mobilised for the tests**,** which were conducted during the working hours of video operators (06:30 am to 02:00 am during the Olympics).

No changes were made to the standard operating procedures related to security (e.g., coordination with SUGE[18], Law Enforcement Agencies) and to traffic (e.g., coordination with transport operators, station management). In this regard, the alerts produced by the tested AI systems had been **integrated into the GENETEC production VMS, in order to minimize the operational changes while enhancing the operators' capabilities.**

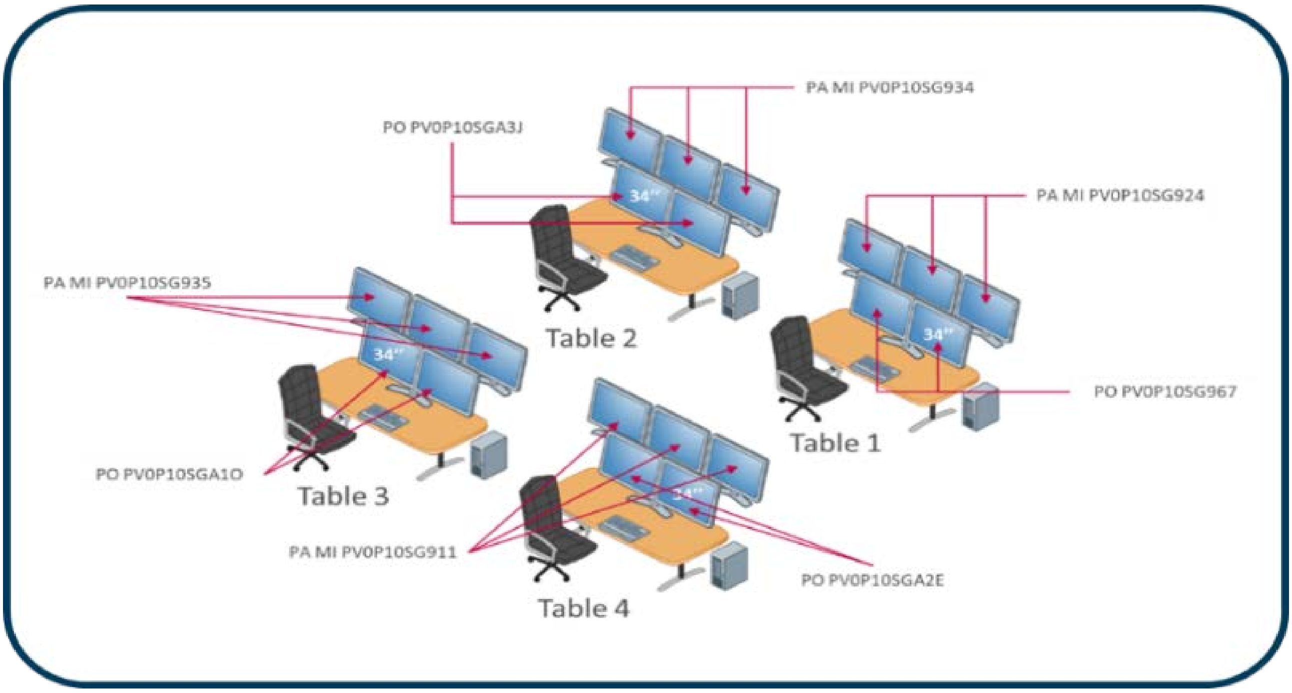


Figure 26 - Video surveillance workstation scheme

One of the major **challenges** faced was that of **effectively allocating video analytics** among the stations while coordinating with the number of agents deployed on the ground during events, in order to avoid unnecessary tests at stations that were already fully secured by field staff and authorities, thereby saving valuable calculation capabilities to be profitably employed at other assets.

[18] The acronym stands for Surveillance Générale and today designates the Railway Security function (Sûreté Ferroviaire) within the SNCF Group.

Integration within pre-existing systems and processes

To enable the implementation of the augmented video system within the National Security Command Post (PCNS) and to facilitate its adoption by the video operators while preserving ergonomics and ease of use, the SNCF project team sought a **strong integration** of the tested technologies on the video operator workstations and the existing VMS.

In the event of an incident detection, the Cityvision solution generates an alert for the SNCF Genetec VMS, allowing:

- The **creation of a bookmark/marker** on the video recording of the camera concerned at the time of detection.
- The **triggering of a real-time alert** (visual notification could also be coupled with an audio notification), materialized by a pop-up notification, specifying the type of event detected and the camera concerned.

  Once the alert is displayed, the video operator analyses the corresponding situation using the images from the camera concerned by the alert (real-time images and video recording) as well as other nearby cameras accessible through the VMS, even if no algorithmic processing is configured on them.

  The video operator can then acknowledge the alarm using three commands:

  - **Positive** – alert of operational interest.
  - **False positive** – alert without any operational interest
  - **Not processed** – alert not handled by the video operators (outside service hours)

In the event of a "positive" alert, the authorized video operator is able decide, with the support of the PCNS manager (responsible for coordinating the operational response), on the actions to be taken to address the observed situation (e.g., requesting field intervention of a SUGE team, canine team, law enforcement, the traffic operations centre, or the station manager).

All operational follow-ups are then recorded at several levels:

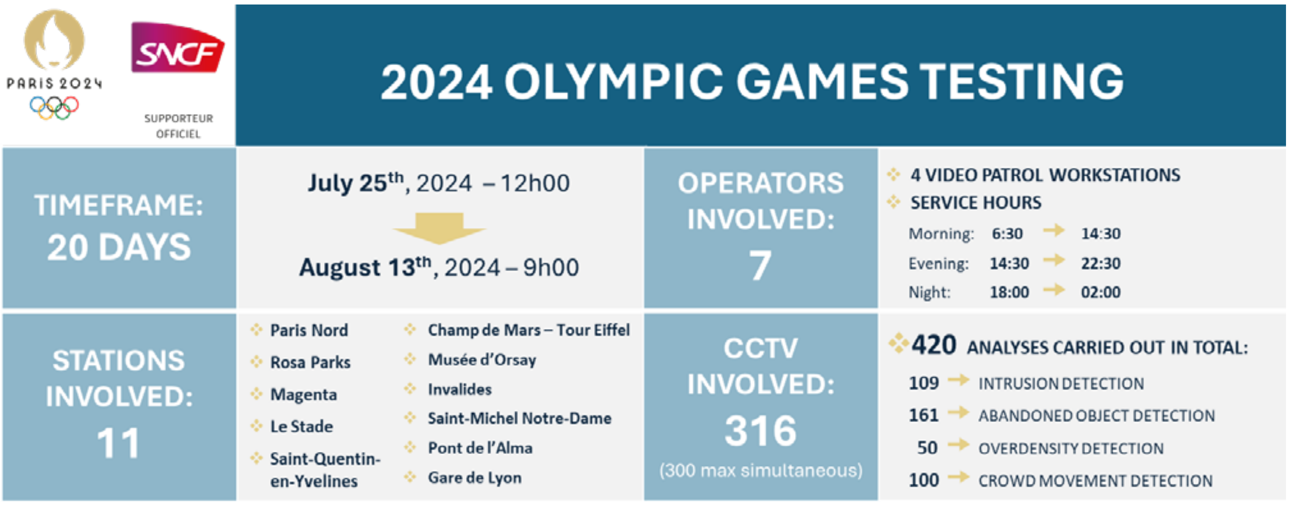
- By the video operator, in the logbook written and signed at the end of each shift.
- By the PCNS manager, in an event log/report completed in real-time and kept in the Security Hypervisor;
- By the intervening SUGE team, in the logbook written and signed at the end of each shift.

3.4. Technical and Operational results

The results of the testing activities have been measured by SNCF through many different KPIs (technical, operational, perceptual), as described in the following:

- **Algorithmic precision**. Percentage describing the rate of 'positive' alerts (events that should trigger an alert, based on the system programming/calibration) over the total number of alerts. The remaining percentage, therefore, describes 'false positives'.
- **Processing rate.** Percentage describing the rate of alerts that were treated in real time by human operators, over the total number of alerts generated by the system. In this regard, it is important to note that the systems were kept operational 24h/day, registering events and sending alerts even at night, when human operators were not physically present at their workstations.
- **Operational interest.** Percentage describing the rate of 'operationally positive' alerts (relevant from a security perspective) over the total number of alerts processed. Measuring this data was important, especially considering that the fact that the system is working and sending alerts correctly, doesn't imply that all the recorded events/alerts are relevant to the operators from a security perspective.
- **Operational interventions.** Number of interventions on the field by operational staff that were triggered on the basis of system-generated alerts.
- **Acceptance/Perception by the public.** This information was deduced based on the number of registered interactions with the public: number of image access requests addressed to SNCF, number of information requests, number of views of the informative page about the experiment published on the company website.

3.4.1. Results of the tests conducted during the Olympic Games



PARIS 2024

SNCF SUPPORTEUR OFFICIEL

2024 OLYMPIC GAMES TESTING

TIMEFRAME: 20 DAYS

July 25th, 2024 – 12h00

August 13th, 2024 – 9h00

OPERATORS INVOLVED: 7

- 4 VIDEO PATROL WORKSTATIONS
- SERVICE HOURS

Morning:	6:30	→	14:30
Evening:	14:30	→	22:30
Night:	18:00	→	02:00

STATIONS INVOLVED: 11

- Paris Nord
- Rosa Parks
- Magenta
- Le Stade
- Saint-Quentin-en-Yvelines
- Champ de Mars – Tour Eiffel
- Musée d'Orsay
- Invalides
- Saint-Michel Notre-Dame
- Pont de l'Alma
- Gare de Lyon

CCTV INVOLVED: 316 (300 max simultaneous)

- 420 ANALYSES CARRIED OUT IN TOTAL:

109	→	INTRUSION DETECTION
161	→	ABANDONED OBJECT DETECTION
50	→	OVERDENSITY DETECTION
100	→	CROWD MOVEMENT DETECTION

Within the **11 stations involved**, **420 analytics were calibrated on 326 cameras** (maximum 300 in simultaneous use): 161 were allocated for abandoned objects detection, 109 for intrusion detection, 100 for crowd movement detection and 50 for crowd density (deployed at usually uncrowded spots).

Four video patrol operator tables received **989 alerts in total** (~50 per day), with a daily average falling below the acceptability threshold agreed internally by SNCF involving the video operators (maximum 100 alerts/day).

The **overall algorithmic precision** was registered at 63%, meaning that 687 events were correctly identified as 'positives', and alerts were sent to operators. Regarding the **operational KPIs**, the rate of **alerts processed in real-time by operators** was 61% (601 alerts). During working hours, such a percentage rises to 99%: this is mainly related to the high volume of intrusion alerts registered at night, triggered by SNCF maintenance personnel duly working in areas labelled as 'forbidden' by the systems.

Regarding the **reception and perception by the public** about the experiment, SNCF didn't receive any formal requests from individuals to access the recorded images. Similarly, no information requests were addressed to the company, while the dedicated information page on the website scored only 32 visits despite the wide visibility given to informational resources (see above, paragraph 3.2.2)

Overall, most alerts were related to the intrusion detection (714) and abandoned objects (270) use cases, with no alerts recorded for crowd movement and just a few concerning excessive crowd density.

Intrusion Detection

The intrusion detection use case scored the highest number of alerts, 714 in total.

The experimented systems achieved an **algorithmic precision rate of 80%** on the 109 CCTVs configured for this use case.

The **real-time processing rate achieved was 54%.** This is mainly due to the fact that many intrusions (mostly SNCF staff or other station workers) were detected during the night hours, when CCTV operators were not working.

The **operational interest was measured at 65%**, with 251 relevant alerts out of a total of 388. It should be noted that cases of 'intrusion' on the tracks by maintenance personnel were intentionally kept as alerts (i.e., individuals wearing orange safety vests may have been excluded during the programming/calibration phase) to be evaluated by the human operator on a case-by-case basis. Finally, it is important to note that **three system-generated alerts led to relevant operational responses, all of which were successful**. An individual attempting to illegally enter the train was detected inside the restricted Eurostar area, prompting coordinated intervention with the station manager. Two additional cases were recorded: an intrusion in a tunnel was treated by SUGE agents, and an illegal track crossing led to a police intervention.

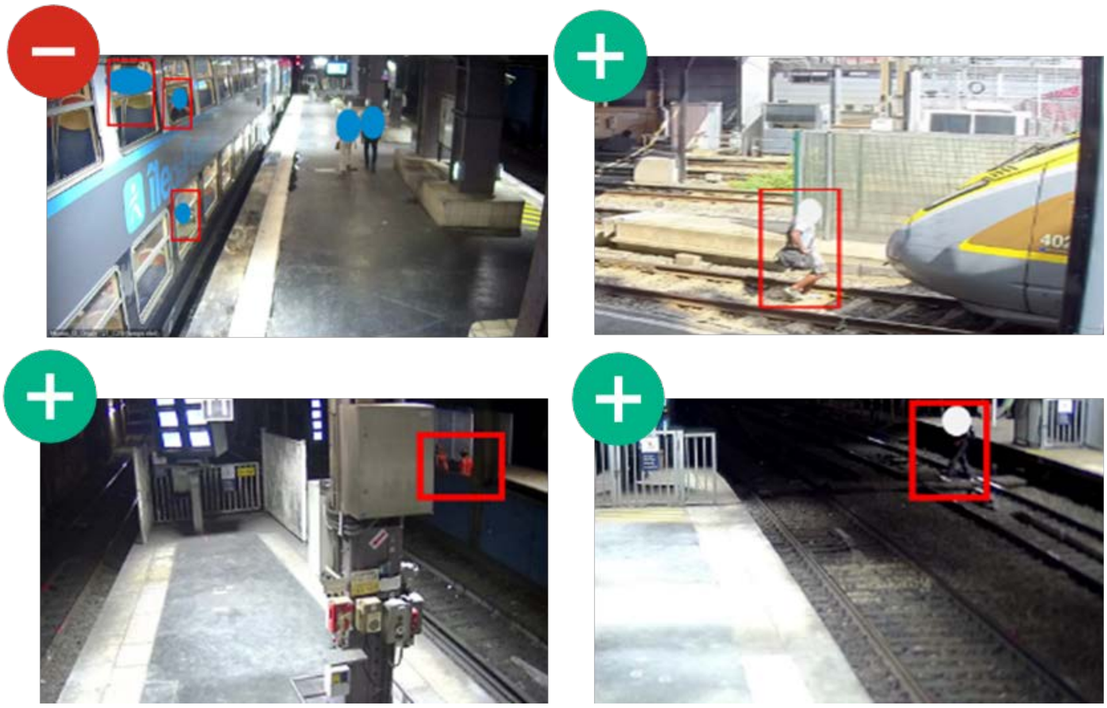


Figure 27 - Positive (green) and false positive (red) intrusion detections by the system

Challenges encountered: the false-positive detections were primarily due to the erroneous non-detection of trains on the track (in this case, passengers on trains were misidentified as individuals standing on the tracks). Other false positives were ascribed to the presence of individuals in close proximity of the detection area (even if outside of it), and to misclassification due to factors such as weather conditions or the presence of animals.

Abandoned object detection

The system scored an **algorithmic precision of 38%** generating **270 alerts**, with the most part of those (168) being false positives. A total of 208 alerts were treated in real-time by human operators, with a **processing rate of 77%**.

The relevant alerts for security were 21, showing an **operational interest rate of 10%.** Most of those positive detections were achieved in conditions of low crowd density and in situations that didn't present any complications. It is worth noting that objects in use by maintenance personnel (e.g., cleaning buckets, trash cans) were purposedly considered by SNCF as positive detections (although not operationally relevant), therefore needing the attention of a human operator to be assessed conclusively. Opaque trash bags, on the contrary, were considered as operationally relevant from a security perspective.

Challenges encountered. Most false positive (90 occurrences) were related to station furniture, mistakenly considered by the system as abandoned objects (e.g., a new SNCF information desk installed near the platform area). An additional 13% of false alerts were generated due to different environmental conditions such as light reflections, puddles, water drops on camera lenses. In some cases (7%), the item owner was not properly detected by the system (e.g., seated persons), therefore considering their luggage as unattended.

During the Olympic games testing, no operational procedures for abandoned object handling were triggered following the alerts generated by the system (on the contrary, an unattended item procedure was successfully triggered during the Paralympic Games).

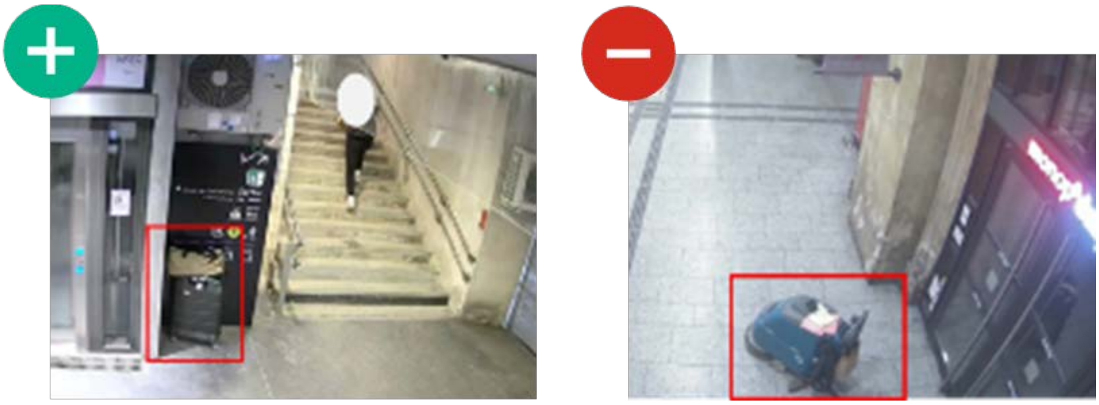


Figure 28 - A positive recognition of an abandoned object (green) and a false alert generated by a cleaning robot (red)

Crowd density detection

The system performed outstandingly relating to this use case, displaying an **algorithmic precision of 100%** by detecting **5 cases** of crowd density above the threshold set by SNCF (i.e., presence of more than X persons within a defined perimeter, with a density higher than X individuals per square meter). The low number of detections may be reconducted to a low level of actual occurrences, in turn related to the numerous SNCF agents deployed at stations to facilitate passenger flows and avoid overcrowdings.

All the detected positives were deemed as relevant (**operational interest rate**: 100%) and cases were treated in real time by operators (**processing rate**: 100%).

It should be noted that the strategy chosen by SNCF was to deploy this system to areas that aren't usually crowded (e.g., station entrance/ exit, corridors), so that any gathering there may be considered as a potential anomaly to be analysed by a human operator. All the detections were treated through active monitoring by video operators, although none of the circumstances required a physical intervention on the field (as an example, one gathering at a station entrance was due to many people seeking shelter from heavy rain).

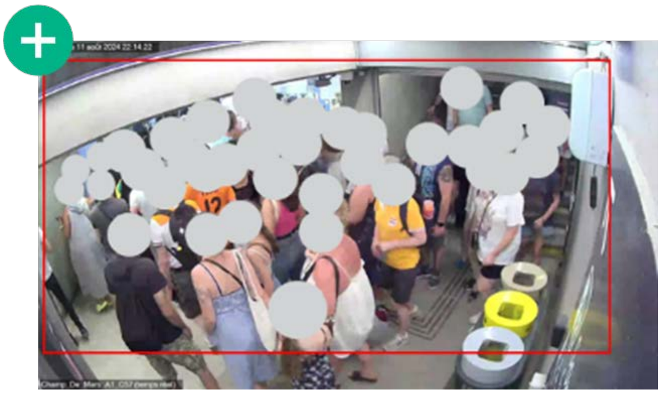
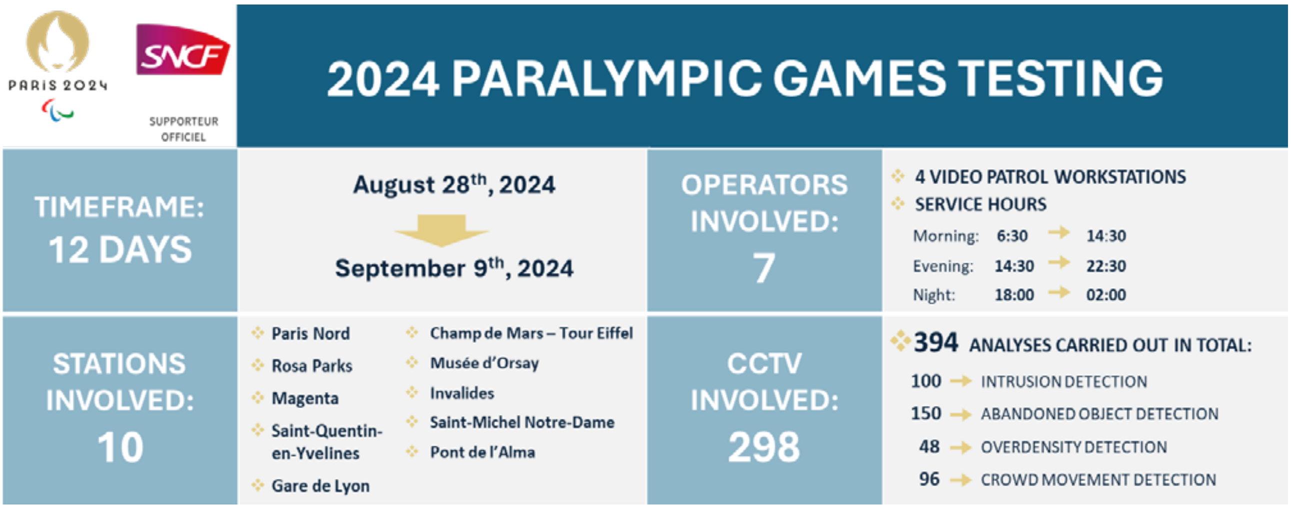


Figure 29 - Positive detection of high crowd density

Crowd movements detection

The extremely low number of occurrences during the testing phase (no alerts during the Olympic games and one detection during the Paralympic Games) make difficult to draw definitive conclusions about this use case.

3.4.2. Results of the tests conducted during the Paralympic Games



PARIS 2024

SNCF SUPPORTEUR OFFICIEL

2024 PARALYMPIC GAMES TESTING

TIMEFRAME: 12 DAYS	August 28th, 2024 September 9th, 2024	OPERATORS INVOLVED: 7	4 VIDEO PATROL WORKSTATIONS SERVICE HOURS Morning: 6:30 → 14:30 Evening: 14:30 → 22:30 Night: 18:00 → 02:00
STATIONS INVOLVED: 10	Paris Nord Rosa Parks Magenta Saint-Quentin-en-Yvelines Gare de Lyon Champ de Mars – Tour Eiffel Musée d'Orsay Invalides Saint-Michel Notre-Dame Pont de l'Alma	CCTV INVOLVED: 298	394 ANALYSES CARRIED OUT IN TOTAL: 100 → INTRUSION DETECTION 150 → ABANDONED OBJECT DETECTION 48 → OVERDENSITY DETECTION 96 → CROWD MOVEMENT DETECTION

Within the **10 stations involved**, **394 analytics were deployed and calibrated on 298 cameras**. Specifically, 150 systems were set up for abandoned objects detection, 100 for intrusion detection, 96 for crowd movement detection and 48 for crowd density. Seven human operators, working at four video patrol workstations, received **326 alerts in total** (~27 per day), which fell largely below the acceptability threshold agreed internally by SNCF involving the video operators.

The **overall algorithmic precision** was calculated at 51%, meaning that 156 events were correctly identified as 'positives', and alerts were sent to operators. Concerning **operational KPIs**, the rate of **alerts processed in real-time by operators** was 84% (96% during working hours), totalling 273 alerts.

The overall **operational interest rate** of alerts **was measured at 47%**, with **54 operationally relevant alerts** and **one successful intervention on the field** being initiated because of automated reports generated by the system.

Compared to the results achieved during the Olympic Games testing, a lower total number of alerts and the higher processing rate were achieved. This result should be mainly ascribed to the fact that **SNCF decided to disable the systems for four hours at night** (00:30 to 04:30).

Overall, the numerical outcomes varied slightly from those registered during the Olympics experiment. A relatively higher number of alerts was related to abandoned objects (251), and a lower one (56) to intrusions. Eight cases of excessive crowd density, along with a single one for (unexpected) crowd movement were also registered.

Concerning the **perception by the public** about the experiment, like in the case of Olympic Games, SNCF didn't receive any formal requests from individuals to access the recorded images. Similarly, no information requests were addressed to the company, while the dedicated information page on the website scored only 36 visits despite the wide visibility given to informational resources (see above, paragraph 3.2.2).

Intrusion Detection

The **overall number of intrusion alerts (56 events)** decreased greatly compared to the outcomes of the Olympic Games experiment, while the **real-time processing rate rose to 91%** (51 events)**.** This was mainly due to the choice by SNCF to disable analytics at night, in order to avoid the (already experienced) massive detection of its own maintenance personnel working on the tracks from 00:30 to 04:30 (see above).

The registered **algorithmic precision rate was 46%** (26 positives vs. 30 false positives)**.** Such low value, nonetheless, may be largely ascribed to the large volume of alerts (17) generated by the temporary malfunctioning of a single camera, due to the fact that it had been purposedly moved from its intended position by an unknown individual.

The **operational interest was measured at 53%**. While 27 alerts were considered as operationally relevant from a security perspective, video operators did not judge necessary (also based on real-time monitoring, situational analysis, on-site presence of staff) to initiate any intervention.

Challenges encountered: The challenges already identified during the previous testing phase were still encountered, such as the erroneous non-detection of trains on the track, the misclassification of individuals in proximity of the detection boundary or due to weather/environmental conditions). Additionally, the sudden change in the planned positioning of a camera (probably due to vandalization) made clear the importance to closely monitor the performance of the systems, along with the capacity to intervene promptly by managing the system (i.e. disabling the analytics coming from the damaged/ camera and initiating maintenance procedures).

The image above describes an interesting case of a ‘useful false positive’. The person in the red rectangle is standing on the platform, but it has been mistakenly considered by the system as being standing on the tracks. Although technically representing a false positive, the detection of this event may undoubtedly display an operational interest linked to preventing dangerous (or suicidal) behaviours and justify the initiation of a staff intervention on-site.

Abandoned object detection

The system scored an **algorithmic precision of 22%** generating **251 alerts** in total, with the most part of those (195) being false positives. A total of 205 alerts were treated in real-time by human operators, with a **processing rate of 82%**.

It is worth noting that objects in use by maintenance personnel (e.g., cleaning buckets, trash cans) were purposely considered by SNCF as positive detections (although not operationally relevant), therefore needing the attention of a human operator to be assessed conclusively. Opaque trash bags, on the contrary, were considered positive and operationally relevant from a security perspective.

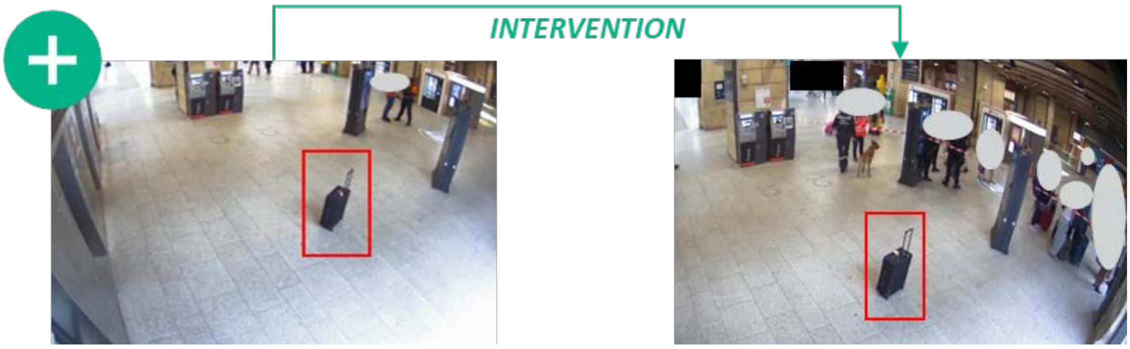


Figure 31 - Successfully initiated intervention procedure triggered by an abandoned object alert

The alerts relevant for security were 12, scoring an **operational interest rate of 6%.** Like in the case of the Olympics experiment, most of the relevant positive detections were achieved in conditions of low crowd density and in situations that didn't present any complications. Nonetheless, an unattended item procedure was initiated following the abandonment of a suitcase, leading to the mobilisation of a canine detection team and the issuance of a penalty to the owner.

Challenges encountered. Most false positives (161) were related to the unexpected presence of station furniture, mistakenly considered by the system as abandoned objects. As an example, an SNCF flag installed near the platform area generated 114 false alerts. Compared to the previous experiment, furthermore, a higher number of positives was due to missed identification of the item's owner, even when the individual was in ideal position (not seated).

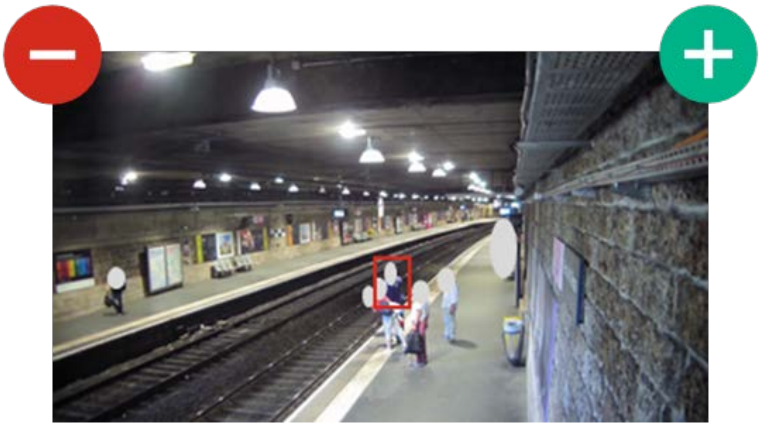


Figure 30 - A false positive detection displaying an operational interest

The rate of false alerts due to different environmental conditions (e.g., light reflections, puddles) and to missed identification of seated persons dropped to a negligible percentage of 1% each.

Crowd density detection

Similarly to the tests carried out during the Olympic Games, also during the Paralympics the system performed outstandingly relating to this use case, displaying an **algorithmic precision of 100%** by detecting **18 cases** of crowd density above the threshold set by SNCF for the specific area where deployed.

All the detected positives were deemed as relevant (**operational interest rate**: 100%) although none of the circumstances required a physical intervention on the field, and almost all the cases were treated in real time by operators (**processing rate**: 89%).

The relatively higher number of detections, compared to the previous experiment (5 events), may be largely attributable to the circumstance that, when Paralympic Games took place, most Parisians were already back in the city from their summer holidays, thereby cumulating commuting/schooling and tourism traffic.

Crowd movements detection

While the only detection was processed in real time and matched the established parameters (i.e., six persons running slow in the same direction) being therefore positive, the event was deemed as operationally irrelevant by the video operator.

3.4.3. Lessons learnt and next steps

Operational conclusions and next steps

During the Olympic and Paralympic Games, four successful interventions were initiated due to events reported by the systems under test. These were carried out on-site by SUGE Teams, Canine Units and National Police.

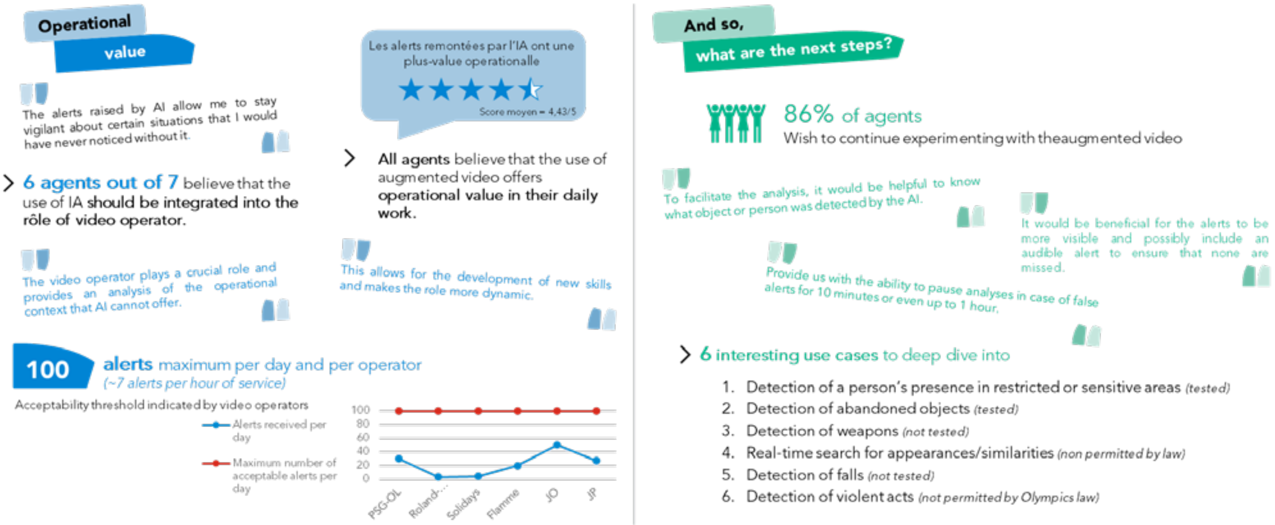
The low number of interventions triggered should be regarded as a positive and encouraging result, considering the low number of security-relevant occurrences during the 2024 Games in general, the limited scale of the experiment and the massive presence of SNCF personnel, police and other first responders directly at the interested premises, contributing to a strong deterrent effect on potential attackers and perpetrators of malevolent acts.

Overall, SNCF estimated that the tested technologies may reach their full potential, generating more detections and leading to interventions on-site, if employed in 'normal times' that don't imply heightened security, specifically designed measures and deployment of extra personnel. It has been seen for instance during the experiments conducted previous the Olympic and Paralympic Games, with 3 operational interventions triggered following a positive detection in only 2 days, and on limited perimeter during the very first experimentation.

In order to learn and retain the most from the experiments, as well as to effectively plan the next steps, **operational feedback** was sought **from the video operators** that took part in the tests: they were asked to fill out a survey and to have an individual debriefing interview about their experience.

While the number of false alerts received during the daily working/testing routine was judged as limited and acceptable (well above the threshold of maximum 100 alerts per operator/day that they had suggested) the few positive alerts were deemed to enable agents to stay particularly active and vigilant during their duties.

All the seven video-operators declared that the **use of Artificial Intelligence and analytic capabilities may bring an added value to their daily working routine** while still leaving them in command of the decision-making process, with most of them also **wishing AI technologies to be integrated into their role on a permanent basis** and therefore being **willing to continue experimenting with AI solutions**.



The alerts raised by AI allow me to stay vigilant about certain situations that I would have never noticed without it.

> **6 agents out of 7** believe that the use of IA **should be integrated into the rôle of video operator.**

> **All agents** believe that the use of augmented video offers **operational value in their daily work.**

The video operator plays a crucial role and provides an analysis of the operational context that AI cannot offer.

This allows for the development of new skills and makes the role more dynamic.

100 **alerts** maximum per day and per operator *(~7 alerts per hour of service)*

Acceptability threshold indicated by video operators

86% of agents
Wish to continue experimenting with theaugmented video

To facilitate the analysis, it would be helpful to know what object or person was detected by the AI.

It would be beneficial for the alerts to be more visible and possibly include an audible alert to ensure that none are missed.

Provide us with the ability to pause analyses in case of false alerts for 10 minutes or even up to 1 hour.

> 6 interesting use cases to deep dive into

1. Detection of a person's presence in restricted or sensitive areas *(tested)*
2. Detection of abandoned objects *(tested)*
3. Detection of weapons *(not tested)*
4. Real-time search for appearances/similarities *(non permitted by law)*
5. Detection of falls *(not tested)*
6. Detection of violent acts *(not permitted by Olympics law)*

Figure 32 - Infographic summarizing the feedback received from video operators

Six use cases were also suggested by video operators to be further investigated and (whenever possible) set up for testing in future: intrusion and abandoned objects detection (already tested), detection of weapons, detection of persons falling on the ground, real-time searching capabilities based on similarities in appearance such as clothes (this use case is not currently allowed under the French law) and detection of violent acts (not allowed under the legal framework regulating in France the current experimental phase, also referred as 'Olympics law').

Technical conclusions and future improvements

Intrusion Detection. Overall, this use case was judged as very valuable by SNCF, matching most expectations in terms of algorithmic precision and operational interest. The hardware resources required to effectively run this system, as well as its maturity level defined by the need for further development effort, were also deemed positively. The main aspects needing future improvement for this promising use case are described in the following.

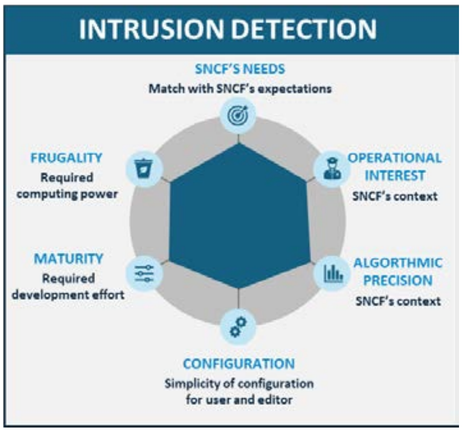


Figure 33 - Technical assessment of intrusion detection use case

A reliable **detection of trains on tracks** is needed to avoid generating false alerts (especially in case of trains standing partially out of the camera field of view). Whenever a train is not properly detected, people sitting in the coaches (if they are visible to the cameras) may be considered as standing on the tracks.

The **camera positioning and angles** should be carefully considered in order to ensure adequate visibility of tracks and restricted areas: in this regard, cameras placed parallel to the tracks have shown better detection results. Detection of humans has shown optimal results when the body is fully visible and totally inside the ‘intrusion zone’: a better **management of occlusions** is needed to enhance the detection of individuals standing behind others or in case their position is overlapping over the allowed and restricted zone.

The **identification of objects** (other than trains and humans) should be improved in order to avoid the misclassification of animals and environmental elements such as leaves, shadows, reflections, puddles. The capacity to quickly handle **changes in the environment** (e.g., construction works, weather conditions) as well as **unexpected camera movements** should be improved, also through the possibility – for operators – to temporary set on pause the analysis on one/more cameras to prevent unnecessary alert repetition.

Abandoned objects detection. Despite the difficulty, the hardware resources and the developments further required to effectively set up such a use case in operational conditions, its operational interest within SNCF’s context was judged to be paramount.

It should also be noted that a higher rate of false positive detections may be considered strategically acceptable when compared to the potential impact brought by a single real threat related to this use case, along with considerable savings in terms of time, cost and reputation.

The main areas of technical improvements for this use case are described in the following.

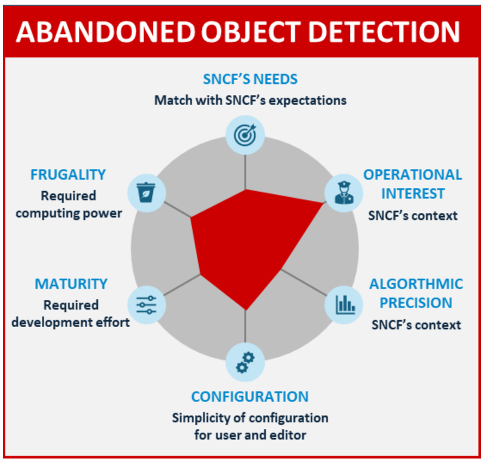


Figure 34 - Technical assessment of left object detection use case

While objects standing isolated in a plain and relatively empty view field are well recognizable, the **classification of objects** should be further improved and allow operators to select specific classes to be included/excluded from the detection (e.g., station furniture) in order to allow the system to better distinguish between luggage and suspicious items. Further desirable improvements would be the **capacity to exclude from the analysis environmental elements** such as puddles and stains, as well as to better estimate the distance between an object and a person. As mentioned above for intrusion detection, enhancements in the system’s **ability to properly recognize individuals in any condition** (standing or seating, in plain sight or partially hidden) would support a strong development of this use case, as well as refinements in the algorithm’s **capability to distinguish some peculiarities of clothing** (e.g., hoods). Finally, also in this case the **capacity to quickly handle changes or disruptions** by temporarily setting on pause the analysis on one/more cameras would allow preventing unnecessary alert repetition.

Crowd density detection. Overall, this use matched most expectations in terms of algorithmic precision, with relatively low requirements in terms of hardware resources and efforts needed for further developments.

The experimentation of this use case, though, showed that the operational interest for security purposes is relatively low, being mostly potentially relevant for other corporate areas, notably for station management.

The usability and performances of this use case may be further enhanced adopting a **different approach** than the one experimented, by first determining the standard occupancy rate (nominal density) of an area after an observation period and subsequently conducting a comparative analysis against the defined nominal density.

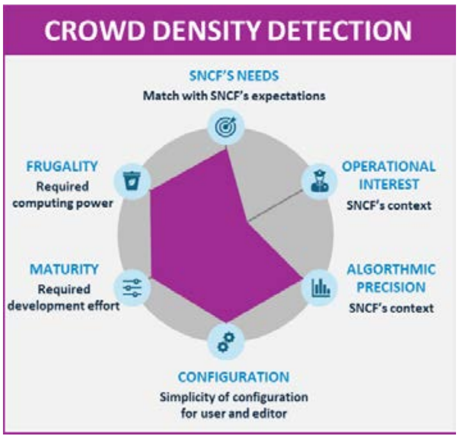


Figure 35 - Technical assessment of crowd density detection use case

Crowd movements detection. With just one detection during the testing phases, it was difficult to draw definitive conclusions about the operational potential of this use case.

This algorithm was also estimated by SNCF as particularly challenging to set up and calibrate, because threshold parameters had to be set relating to a minimum number of persons involved, along with a numerical quantification of their movement vector ('collective' approximate direction and minimum average speed).

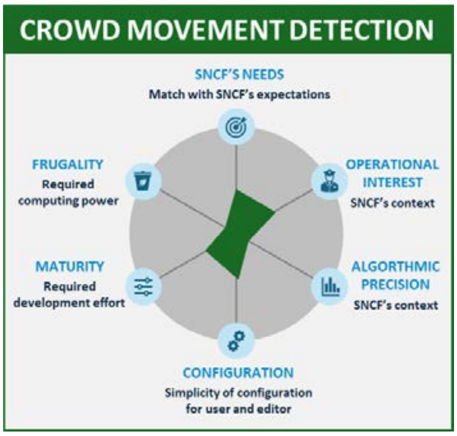


Figure 36 - Technical assessment of crowd movements detection use case

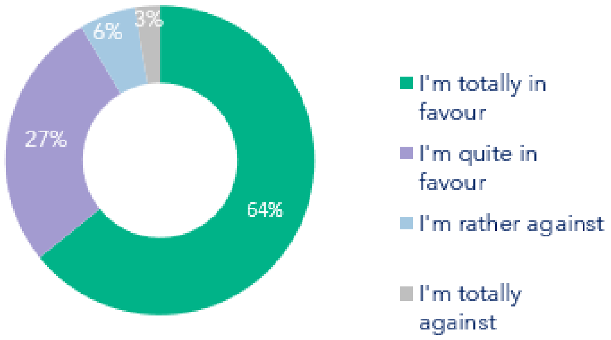
Perception and acceptance of augmented video applications by station users

As part of another experimentation activity with AI, carried out from December 21st, 2024, to January 6th 2025, SNCF wanted to collect further and more detailed feedback from station users on their perceptions about augmented video (for communication towards the public and feedback specifically about the Olympics, see above par. 3.2.2.).

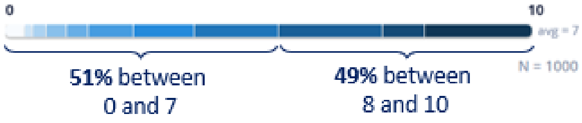
The aim of this study, in particular, was to find out how users who have walked through the stations during the testing period felt about the usage of augmented video**,** and to ask them about potential further uses of such technology.

The study involved **1000 participants** who were **recruited through social media channels** and asked to answer **16 questions**.

What is your opinion on the presence of video protection cameras in stations and on trains?



How confident are you in the usefulness of CCTV cameras in stations and trains? (from 1 to 10)?



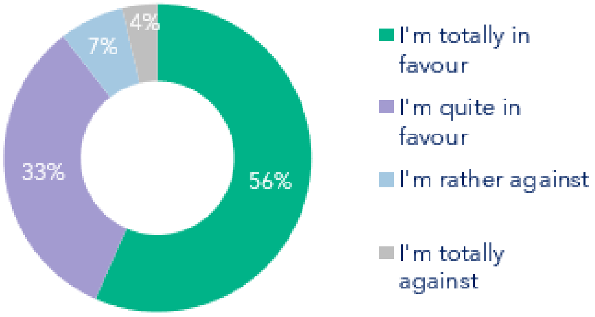
On average, respondents rate their **level of confidence in the usefulness of video protection at 7/10.**

Among the main findings, **most respondents (91%) declared to be in favour of video protection (CCTVs) in stations and trains and are quite confident in its usefulness**.

Regarding – more specifically – the usage of **augmented video tools for security** purposes, respondents are **generally in favour of using augmented video tools for security purposes (89%)**. Notably, when respondents were informed about the process for augmented video usage in place at SNCF, which puts people at the centre of decision-making, they demonstrated to be even more in favour of such use of technology (95%).

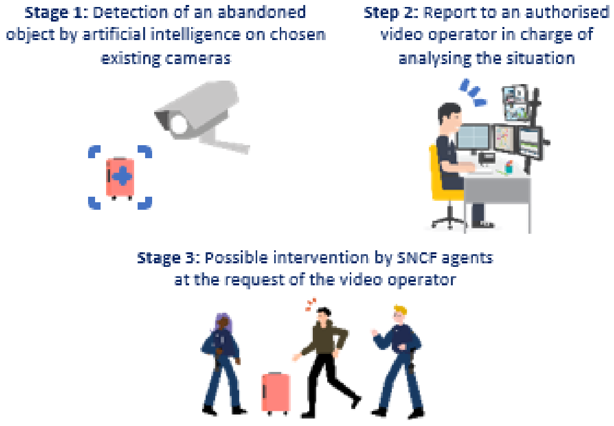
Among the preferred use cases, respondents mentioned that the **detection of weapons would be the most useful and desirable augmented video tool**, followed by the detection of violence (flights or brawls) and the fall of persons on the ground.

Are you in favour of using image analysis tools in real time to detect possible security incidents?



89% of respondents **are in favour of using augmented video tools** to detect security incidents

SNCF's augmented video usage process



Stage 1: Detection of an abandoned object by artificial intelligence on chosen existing cameras

Step 2: Report to an authorised video operator in charge of analysing the situation

Stage 3: Possible intervention by SNCF agents at the request of the video operator

95% of respondents who gave an opinion are **in line with this use**

When asked to express what augmented video means to them, respondents' opinions were divided. Spontaneously, this technology was associated with safety and security, artificial intelligence and remains a source of fear. In fact, **34% of respondents declared to have concerns** about augmented video, especially on the following areas (in order of priority):

1. **Infringement of freedom**
2. **Misuse**
3. **Mass surveillance**
4. **Errors and biases**
5. **Cybersecurity**
6. **De-humanization**
7. **Facial recognition**

Communication
Communicate on the use of augmented video tools, the objectives and the data protection measures put in place.

Legal framework
Ensure that the use of these tools complies with the laws in force and put in place strict policies to prevent abuse.

Audit
Set up regular audits, by independent structures, to ensure that the tools are used in accordance with the established rules.

Training
Train employees in the ethical and responsible use of augmented video tools.

Restricted access
Restrict access to sensitive data to authorised persons only, respecting the principle of necessity.

Cybersecurity
Using secure technologies to prevent hacking or information leaks.

These fear factors should be duly and exhaustively addressed through many parallel actions:

Finally, when asked about their opinion about the adoption of such technologies on a permanent ground in future, **a large percentage of respondents (88%) stated that they would like SNCF to continue its work on augmented video in the long term and on a daily basis.**

4.

CONCLUSIONS

Striving to constantly enhance security for users should be regarded as a core element within railway companies' mission, functional to effectively supporting the sectoral development of rail transport through current and future evolutionary processes.

While technology undoubtedly represents – together with organizational and human factors – one of the pillars of security, it should be subject to the continuous process of monitoring, scouting, testing and adoption that leads to stable and long-lasting improvement.

The steep evolution of AI technologies started in the latest years is currently stretching the boundaries of many disciplines and unfolding previously unthinkable use-cases in every industrial domain.

Investing in AI applications for security, therefore, cannot be regarded as a merely technological upgrade, but rather as a strategic way forward for railway companies. The benefits of enhanced threat detection, improved operational efficiency, and cost reduction make a compelling case for adopting AI technologies. As the railway industry continues to evolve, embracing AI will be crucial in ensuring the safety and security of passengers, assets, and infrastructure.

As railway transport may be considered less mature than other critical sectors relating to the testing and deployment of AI-powered technologies in the physical security domain, this may be mostly related to the necessity to abide strict safety standards and policies that govern the adoption of new and not fully mature technologies, frequently imposing or suggesting a cautious posture for railway companies. Furthermore, high costs for research, development and purchasing of such technologies, as well as potential uncertainty regarding the evolution of regulatory frameworks and public perception and acceptance of such technologies are also playing a major role in this regard.

This *status quo* is probably going to evolve significantly in the coming years. The maturity levels of technologies investigated are deemed to be growing fast in the next years, reaching high levels for most of them by 2030 according to the experts interviewed. National and supranational regulatory initiatives are going to stimulate the development of safe and secure AI systems while regulating their adoption by setting clear boundaries and shedding light on grey areas of uncertainty (mostly related to ethical, legal and accountability aspects) that represent, currently, major challenges towards their adoption by railway companies.

Public acceptance about the use of AI systems for rail security represents another paramount factor enabling their potential deployment in the next years. In this regard, clear information and communication strategies by companies towards users may play a major role, explaining the benefits potentially brought by such technologies to their user experience, while clearing potential doubts about ethical aspects and reaffirming the central role of humans in supervision of such systems.

Finally, as demonstrated by the SNCF experiment, actively involving personnel/operators since the preparatory steps of testing may bring many advantages in terms of efficiency and usability of AI systems, helping workers to fully understand the potential benefits brought by such technologies to their daily activities (adequate training pathways are paramount to this end), while supporting companies in adopting sustainable strategies towards their development, implementation and operational deployment.

BIBLIOGRAPHY

ACLU (January 2024), 'Williams v. City of Detroit', online, accessed on 21/10/2024 at: Williams v. City of Detroit | American Civil Liberties Union (aclu.org)

California Institute of Technology (Caltech) website (August 2024) '*Exploring AI vs. Machine Learning*', online, accessed on 10/09/2024 at: Link: Exploring AI vs. Machine Learning - Caltech

Correa et al. (2023), Patterns, Volume 4, Issue 10, 10085, '*Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance*', online, accessed on 17/10/2024 at: Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance - ScienceDirect.

European Commission – High-level expert group on artificial intelligence (December 2018), '*A definition of AI: main capabilities and scientific disciplines*', online, accessed on 11/09/2024 at: A definition of AI (europa.eu)

European Commission (February 2020b), "*On Artificial Intelligence - A European approach to excellence and trust*", online, accessed on 18/11/2024 at: commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission website (October 2024), "AI Board", online, accessed on 18/11/2024 at AI Board | Shaping Europe's digital future

IBM Company website (March 2021) '*Supervised versus unsupervised learning: What's the difference?*', online, accessed on 10/09/2024 at: https://www.ibm.com/blog/supervised-vs-unsupervised-learning

IBM Company website (March 2023) '*What is supervised learning?*', online, accessed on 10/09/2024 at: What Is Supervised Learning? | IBM

ISACA (2024), '*Understanding the EU AI Act: Requirements and Next Steps*', online, accessed on 15/11/2024 at: Understanding the EU AI Act: Requirements and Next Steps

ITF – International Transport Forum (December 2022), *'Statistics Brief – Trends in the Transport Sector'*, online, accessed on 25/09/2024 at: Modal shift to cleaner transport fails to materialise (itf-oecd.org)

McKinsey (May 2024), *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*', online, accessed on 18/10/2024 at: the-state-of-ai-in-early-2024-final.pdf (mckinsey.com)

Mirror, (27/03/2023) '*Commuters absolutely stunned to see robot dog trotting along train platform*', online, accessed on 10/09/2024 at: Commuters absolutely stunned to see robot dog trotting along train platform - Mirror Online

MIT Technology Review (08/11/2016) '*In the 1980s, the Self-Driving Van Was Born*', online, accessed on 10/09/2024 at: In the 1980s, the Self-Driving Van Was Born | MIT Technology Review

NIST (March 2022), '*There's More to AI Bias Than Biased Data, NIST Report Highlights*', online, accessed on 21/10/2024 at: There's More to AI Bias Than Biased Data, NIST Report Highlights | NIST

PresseCitron (16/06/2022), '*Métro : on a rencontré le nouveau robot-chien de la RATP*', online, accessed on 10/09/2024 at: Métro : on a rencontré le nouveau robot-chien de la RATP (presse-citron.net)

Rail Market News (02/10/2023), '*DB Cargo: Dog robot inspects wagons',* online, accessed on 10/09/2024 at: DB Cargo: Dog robot inspects wagons | Latest Railway News (railmarket.com)

SNCF Company website (April 2024), "*Video-surveillance analysis experiment*", online, accessed on 25/10/2024 at: Videosurveillance analysis experiment | SNCF Group

SNCF Company website (March 2024), "*Intelligence artificielle*", online, accessed on 18/11/2024 at: Innovation & recherche : intelligence artificielle | Groupe SNCF

SNCF Company website, "*Railway Security*", online, accessed on 18/11/2024 at: Railway Security—protecting people and property | SNCF Group

The AI Navigator website, online, accessed on 29/09/2024 at: The AI Navigator | Your Guide Through the World of AI

UIC (March 2021), '*Artificial intelligence. Case of the railway sector – State of play and perspectives*', online, available at: Artificial intelligence - Case of the railway sector - State of play and perspectives (uic.org)

UIC/MCKINSEY (August 2022), '*Boosting passenger preference for rail',* online, accessed online on 18/11/2024 at: Boosting passenger rail | McKinsey

World Economic Forum (July 2021), '*Research shows AI is often biased. Here's how to make algorithms work for all of us',* online, accessed on 21/10/2024 at: Research shows AI is often biased. Here's how to make algorithms work for all of us | World Economic Forum (weforum.org)

INTERNATIONAL UNION OF RAILWAYS
16, rue Jean Rey - 75015 Paris - France
Tel. +33 (0)1 44 49 20 20
Fax +33 (0)1 44 49 20 29
E-mail: info@uic.org

Published by: UIC Security Platform - New Technologies Working Group
Director of publication: Marie-Hélène Bonneau
Authors: Bruno De Rosa (UIC Security Department), Armand Raudin (SNCF Direction de la Sûreté)
Cover and layout: Ludovic Wattignies
Photo credit: Adobe Stock

ISBN 978-2-7461-3488-1
Copyright deposit: July 2025

www.uic.org

in X YouTube

#UICrail