D2.2 Ethical Data Guidelines Report

Security of Critical Infrastructure by Multi-Modal Dynamic Sensing and AI

22 10 2025

ITEA Project No: 22006

sintra-ai.eu

ITEA 4

DOCUMENT VERSIONS

Version no	Date	Authors	Changes
1.0	15.07.2025	Aylin Yorulmaz (KoçSistem)	Initial Draft, Document Structure
	24.07.2025	Aylin Yorulmaz (KoçSistem)	TOC, general content drafted around SINTRA innovation, use cases and solution concept
	04.08.2025	Sultan Kırkan (ARD)	Update ARD's use case (Anomaly detection at F&B areas)
	21.08.2025	Dilara Karabulut (Koçtaş)	Update Koçtaş's use case (Anomaly detection in retail)
	22.08.2025	NL consortium	Port use case added
	22.08.2025	BE consortium	Construction use case added
	25.08.2025	Aylin Yorulmaz (KoçSistem)	Introduction section wrap up according to the whole document
2.0	31.08.2025.2025	Finland consortium	Port use case added

CONTENTS

1 ACRONYMS

Acronym	Expanded
AI	A set of technologies that enable machines to perform tasks that typically require human intelligence, such as perception, reasoning, and learning.
GDPR	The EU regulation governs the processing of personal data and guarantees data protection and privacy for individuals within the European Union.
DPIA	A formal process required under GDPR to evaluate the risks to individuals' rights and freedoms when processing personal data, especially using new technologies.
HAR	A subset of computer vision and AI systems that analyze visual and/or sensor data to classify or detect human activities and behaviors.
SHAP	A method for interpreting machine learning models by assigning importance values to each input feature that contributed to a prediction.
LIME	A tool that explains the predictions of machine learning classifiers by approximating them locally with interpretable models.
Pseudonymization	A data management and de-identification process where personal data is replaced by artificial identifiers or pseudonyms, making it harder to identify individuals directly.
Anonymization	The irreversible process of removing personally identifiable information from data sets, making individuals unidentifiable.
Federated Learning	A distributed machine learning approach that allows AI models to be trained across multiple decentralized devices or servers holding local data, without needing to transfer the data itself.
Differential Privacy	A mathematical framework that adds random noise to data to protect individual privacy while still allowing for aggregate data analysis.
OpenDP	A suite of open-source tools and libraries developed to help organizations implement differential privacy techniques in their systems.
Data Controller	The entity that determines the purposes and means of processing personal data and is legally responsible for its handling.
Data Processor	An entity that processes personal data on behalf of a data controller, typically bound by contractual obligations and subject to GDPR.
Data Subject	An individual whose personal data is collected and processed. Under GDPR, data subjects are entitled to specific rights over their data.
Edge Computing	A computing paradigm that processes data near the source of data generation (e.g., on local devices), minimizing latency and enhancing data privacy.
Five Safes	A framework for data access and sharing that includes Safe People, Safe Projects, Safe Settings, Safe Data, and Safe Outputs to ensure secure use of sensitive data.

Menlo Report	A foundational document outlining ethical principles for ICT research involving human subjects, expanding the Belmont principles to include respect for law and public interest.
Explainability	The ability of an AI system to provide understandable insights into how it arrives at specific decisions or outputs.
Surveillance Ethics	A field of applied ethics concerned with the appropriate use of monitoring technologies, balancing public safety with privacy and civil liberties.
Accountability	A principle requiring identifiable and responsible oversight of AI systems, data handling, and decision-making processes.
Data Collection Register	A formal document recording parameters of data acquisition—such as sensor types, resolution, sampling frequency, retention limits, and lawful basis for processing—in line with GDPR's accountability requirements.
Message Broker Layer	A middleware component that facilitates secure ingestion and routing of data streams, wrapping each packet in metadata for traceability and access control.
Business Definition Layer	SINTRA's analytic tier where AI-driven models execute tasks like anomaly detection or human activity recognition, maintaining links to original (anonymized) data for auditing.
Geo-fenced Cloud Infrastructure	A cloud environment restricted to specific geographic boundaries to ensure compliance with data localization and sovereignty laws.
Joint Controller Agreement (JCA)	A legally binding arrangement under GDPR Article 26, clarifying roles and responsibilities between two or more data controllers sharing decisions on processing activities.
Five Safes Framework	A governance methodology ensuring safe data sharing by vetting the individuals, projects, data types, environments, and outputs involved in any access or exchange scenario.
Transfer Impact Assessment (TIA)	A GDPR-mandated assessment that evaluates the risks of international data transfers, especially to jurisdictions without EU adequacy decisions.
Federated Analytics (FA)	A distributed data analysis approach that computes aggregate insights across decentralized datasets without centralizing personal data.
Privacy-Enhancing Technologies (PETs)	A suite of technologies—such as federated learning, homomorphic encryption, and secure multiparty computation—that enable data processing while minimizing personal data exposure.
Differential Privacy (DP)	A mathematical technique that injects noise into datasets or query results to prevent identification of individual records, even in aggregated data.
Homomorphic Encryption (HE)	An encryption method that allows computations to be performed on ciphertexts, producing encrypted results that match operations on plaintexts—thus enabling data processing without decryption.
Secure Multi-Party Computation (SMPC)	A cryptographic technique that allows multiple parties to compute a function over their inputs while keeping those inputs private.
OpenDP	An open-source differential privacy toolkit developed by Harvard and Microsoft, offering robust libraries for building DP-compliant data analysis workflows.

SmartNoise	A privacy-preserving query and data synthesis platform built on OpenDP, designed to enable compliant sharing of aggregate statistics or synthetic data.
Diffprivlib	IBM's open-source Python library implementing differential privacy for common data science workflows, supporting compliance and reproducibility.
TensorFlow Privacy	A library extension for TensorFlow that adds differential privacy mechanisms to machine learning model training.
PySyft	An open-source framework enabling privacy-preserving, federated learning and secure data science workflows using techniques such as SMPC and differential privacy.
XAIR (Explainable AI for Augmented Reality)	A framework that integrates AI reasoning into augmented reality interfaces to visualize explanations and AI decisions in real-world environments.
Adversarial Training	A method for improving model robustness by training with intentionally perturbed (adversarial) inputs to make the AI less sensitive to malicious manipulation.
Distribution Shift Detection	Techniques to identify when input data deviates significantly from the data the model was trained on, signaling possible reliability issues.
Human-in-the-Loop (HITL)	A design principle that integrates human oversight into the AI decision-making process to validate, correct, or override algorithmic outputs.
Modality Partitioning	A privacy strategy that restricts each sensor modality (e.g., audio, RFID, video) to specific analytic roles to reduce inference risks and data leakage.
Bias Audit	A formal process to evaluate and document whether AI models or fused systems demonstrate unfair treatment of specific groups across different stages.
Responsible AI Question Bank	A structured set of questions and criteria developed to guide organizations in assessing ethical, legal, and operational aspects of AI systems.
WitnessAI	An AI governance platform designed to provide visibility, policy enforcement, audit logging, and compliance automation for enterprise-grade AI systems.
SHAP/LIME Coverage Rate	A metric indicating the proportion of AI alerts accompanied by explanations generated using SHAP or LIME, used to monitor explainability in real-time systems.
PIMS (Privacy Information Management System)	A framework, aligned with standards like ISO/IEC 27701, for managing personal data responsibly and systematically within an organization.
Immutable Audit Logging	A secure logging method that prevents the alteration of records post-creation, is often implemented using blockchain-inspired or append-only data structures.
Bistable Fallback Verification	A safety mechanism in AI systems where decisions default to a safe or conservative state unless explicitly validated by multiple confidence signals or human input.

2 INTRODUCTION

The D2.2 Ethical Data Guidelines Report outlines the ethical, legal, and technical principles that guide the handling of data within the SINTRA project, ensuring compliance with privacy regulations such as the GDPR while supporting innovative AI-driven safety and security applications. This report provides a comprehensive framework for ethical data governance across all SINTRA use cases—ranging from airport surveillance to port monitoring—emphasizing privacy-by-design, transparency, accountability, and the responsible use of AI and multimodal sensor data. It also addresses challenges related to data ownership, anonymization, trustworthiness of AI systems, and the integration of privacy-preserving technologies to uphold human rights and societal values in high-stakes surveillance contexts.

2.1 Purpose of the Document

This document is a comprehensive report on the SINTRA platform, with a focus on ethical, privacy, and governance considerations for AI and multi-modal sensing systems. It explains the platform's architecture, functional layers, data flow, and processing model, as well as use-case driven design considerations for practical deployment in airports, ports, retail, construction and railway sites.

It defines the scope and objectives of the report, outlines the methodology and sources, and clarifies what the document covers and how it should be used. A significant portion is dedicated to ethical and legal concerns, discussing principles such as transparency, fairness, and accountability, alongside GDPR, data protection, privacy-by-design, and surveillance-specific considerations.

The document details data governance and security measures, including the data lifecycle, ownership, access control, cross-border governance, and privacy-preserving techniques like anonymization, pseudonymization, differential privacy, homomorphic encryption, and federated learning. It also addresses trustworthy and explainable AI, human-in-the-loop design, robustness, fairness, and adversarial safety, highlighting challenges, mitigation strategies, and ethical risks specific to multi-modal sensor fusion.

Finally, it establishes evaluation, monitoring, and compliance frameworks through periodic assessments, ethical compliance metrics, and auditing mechanisms. Recommendations are provided for developers, integrators, end-users, operators, policymakers, and regulators, offering guidance on best practices for the ethical, legal, and safe use of the SINTRA platform.

In essence, this document serves as a technical and ethical reference guide for deploying the SINTRA platform responsibly.

2.2 Scope and Objectives

The scope of this document is to provide a comprehensive framework for the ethical, legal, and operational deployment of the SINTRA platform. It covers the platform's architecture, functional layers, and data flow, along with use-case driven design considerations for environments such as airports, ports, retail, construction and railway sites.

The document addresses ethical, legal, and privacy considerations, including key principles like transparency, fairness, and accountability. It discusses GDPR compliance, data protection, surveillance-specific ethics, privacy-by-design, and privacy-preserving techniques such as anonymization, differential privacy, homomorphic encryption, and federated learning.

Data governance and management are also included, covering the data lifecycle, roles and responsibilities, data ownership, access control, and cross-border or cross-sectoral governance. It explains governance in the context of multi-modal sensor fusion, AI, and UAV surveillance.

Additionally, the document focuses on AI and multi-modal sensing, highlighting explainable AI, fairness, human-in-the-loop considerations, adversarial robustness, and ethical risks in sensor fusion. It provides frameworks for evaluation, monitoring, and compliance, including auditing, traceability, metrics for ethical compliance, and periodic assessments.

Finally, actionable recommendations and guidelines are offered for developers, integrators, end-users, operators, policymakers, and regulators, ensuring that the SINTRA platform can be deployed in a technically effective, legally compliant, ethically responsible, and operationally safe manner across multiple use cases.

3 SINTRA PLATFORM OVERVIEW

The SINTRA platform serves as the central technological backbone enabling advanced safety and security solutions across diverse operational environments, including airports, ports, construction and railway sites. Designed to support real-time situational awareness, multi-modal sensor fusion, AI-based analytics, and ethical data handling, the SINTRA platform integrates heterogeneous technologies within a scalable, modular, and privacy-preserving framework. This section details the architecture, data flow mechanisms, and how the platform's design is tailored to specific use case needs.

3.1 General Architecture and Functional Layers

The SINTRA platform architecture is structured into four core functional layers, designed to manage the end-to-end flow of data—from high-speed ingestion at the edge to aggregated insights for decision-makers via visual interfaces. These layers are:

1. Message Broker Layer: This is the foundational communication layer that facilitates real-time ingestion and transmission of sensor and system data. It supports multiple data transmission protocols to ensure compatibility with various sensor types, including CCTV, 3D cameras, lidar, acoustic sensors, and RFID. This layer is critical for enabling seamless integration of partner solutions and maintaining system responsiveness, especially under high data loads.

2. Business Definition Layer: This layer defines the logic and operational rules driving the platform's intelligence. It integrates the outputs of individual AI modules, handles sensor fusion, applies anomaly detection models, and enforces data governance policies such as access control and anonymization. Each partner in the SINTRA consortium contributes unique functional components to this layer, which supports both standalone and collaborative solutions.

3. Aggregation and API Layer: This layer aggregates processed data and facilitates access via open-source APIs. It acts as a bridge between the analytics engine and visualization tools, ensuring that processed and fused sensor data is made available securely and efficiently. It also supports real-time operations by interacting with distributed databases for fast data retrieval.

4. Visualization Layer: The user-facing layer provides dashboards and AR-powered interfaces for real-time monitoring and situational awareness. It enables operators to visualize spatial sensor data, anomaly detections, and system status across facilities. Augmented reality (AR) extensions, such as Hololens 2, enhance explainability and usability, especially in complex environments like airports and ports.

This layered approach provides modularity, enabling each layer to evolve independently while maintaining overall platform coherence. It also facilitates ethical data handling by incorporating

privacy-preserving measures at every layer—from edge-based anonymization to secure data access controls in the visualization layer.

3.2 Data Flow and Processing Model

The SINTRA platform's data processing pipeline is designed to manage high-volume, high-velocity, and high-variety data streams, while enforcing strict data protection and ethical guidelines. The core stages of the data flow model are:

Data Ingestion: Sensor data is collected in real-time from heterogeneous sources, including but not limited to CCTV, thermal cameras, 3D sensors, acoustic sensors, gyroscopes, and environmental monitors. These sources may be stationary (e.g., installed cameras in terminals or ports) or mobile (e.g., drones or vehicle-mounted sensors).

Pre-processing at the Edge: Where feasible, data is anonymized or pseudonymized at the edge using techniques such as noise injection, irreversible transformations, and real-time filtering. Federated learning and homomorphic encryption may be employed to enable privacy-preserving model training without transmitting raw data to central servers.

Sensor Fusion and AI Processing: The Business Definition Layer applies advanced AI models, including reinforcement learning, deep learning, and zero-shot learning, to detect anomalies, recognize actions, and derive context-aware insights. Multi-modal sensor fusion ensures a holistic understanding of the environment, combining video, audio, and contextual metadata (e.g., GIS, weather, user behavior).

Aggregation and Enrichment: Fused data streams are aggregated in a distributed database, where contextual enrichment (e.g., location tagging, temporal analysis, operational state) is applied. This supports temporal correlation and enables prediction and pattern recognition across longer time horizons.

Data Access and Governance: Processed data is made available to authorized users through secured APIs and user-specific access control settings. The platform enforces data ownership principles, ensuring that stakeholders retain control over their data and that sharing policies reflect regulatory and ethical constraints.

Visualization and Decision Support: Finally, data is rendered via user dashboards or immersive AR interfaces. These interfaces are customizable based on user roles (e.g., security officer, operator, analyst) and provide explainable AI outputs for trustworthiness and accountability. For instance, AR overlays may highlight anomaly sources, show object trajectories, or visualize AI confidence levels.

3.3 Use Case-driven Design Considerations

The SINTRA platform is not a one-size-fits-all solution; it is tailored to meet the specific operational, technical, and ethical needs of a diverse set of use cases. The requirements and constraints from each use case have directly informed platform features, such as sensor compatibility, AI model robustness, and privacy controls. Below are key design considerations derived from the use cases:

Airport Use Case

The airport scenario involves both open (halls, corridors) and semi-closed (shops, cafes) areas, each requiring different sensing strategies. Real-time anomaly detection from over 3,000 CCTV cameras is enhanced through event-driven AI processing, where only flagged frames are analyzed, reducing computational load and preserving privacy. The HAR (Human Action Recognition) system leverages 3D depth cameras and zero-shot learning to identify unusual behaviors. Privacy is reinforced via anonymization at the source and minimal image retention policies.

Augmented reality tools assist security teams in interpreting dense sensor data and understanding AI decisions. This is particularly crucial for explaining flagged behaviors or anomalies in high-traffic areas, ensuring that human operators remain in control.

Port Use Case

The port use case demands large-scale, real-time analysis of environmental, acoustic, and video data across wide geographic areas. It incorporates UAV-based surveillance, where mobile sensor platforms autonomously monitor events and relay data to the central system. Data fusion from mesh networks, smart locks, multispectral sensors, and AIS feeds requires advanced processing to detect illicit activities like smuggling or intrusion.

Special attention is paid to cross-border data governance, as ports involve international operations. Differential privacy, homomorphic encryption, and joint behavior analysis support compliance with GDPR while enabling robust threat detection.

Construction Site Use Case

In construction environments, platform flexibility is critical. The SINTRA platform must handle a mix of proprietary systems and support proactive decision-making under resource constraints. AI modules detect anomalies such as unauthorized access, unsafe behaviors, or equipment misuse. Edge-based processing and secure 5G networks are prioritized to reduce latency and support remote site operation.

Privacy risks from integrated video, audio, and positional tracking are addressed through decentralized data control, real-time anonymization, and ethical oversight mechanisms embedded into the platform's data lifecycle management.

4 ETHICAL PRINCIPLES AND LEGAL FRAMEWORKS

This section establishes the ethical and legal foundations underpinning SINTRA's data practices. Drawing from the deliverable's task definitions, SINTRA use cases, and technology innovation inputs, it elaborates:

- **Key Ethical Principles** guiding all platform operations
- Applicable **GDPR & Data Protection Regulations**
- Ethical concerns specific to **surveillance**
- Ethical challenges in **AI & multi-modal sensing**

4.1 Key Ethical Principles (Transparency, Accountability, Fairness, etc.)

Effective ethical data management is rooted in a core set of principles designed to protect individual rights, foster trust, and prevent harms:

- **Privacy**: Collect and process only what's necessary. Respect individuals' control over personal information[1].
- **Consent & Purpose Limitation**: Ensure explicit, informed consent, and restrict data use to defined purposes. SINTRA's flow architecture enforces this through documented data pipelines[2].
- **Transparency**: Maintain clarity about how, why, and by whom data is collected or analyzed. Publish algorithms, anonymization methods, and AI decision logic[3].
- **Accountability**: Assign clear responsibility for data management, including mechanisms for audits and redress. SINTRA logs every data access and model inference[4].
- **Fairness & Non-Discrimination**: Design algorithms to avoid bias—train on diverse datasets and test for disparate impacts. SINTRA deploys bias-detection modules, especially in security-critical zones[5].
- **Safety & Robustness**: Ensure AI is resilient to errors, adversarial inputs, and system failures. Real-time anomaly detection requires robust fail-safes.

[1] https://www.wired.com/story/data-ai-ethics-hippocratic-oath-cathy-o-neil-weapons-of-math-destruction/

[2] https://gdprlocal.com/gdpr-machine-learning/

[3] https://assets.publishing.service.gov.uk/media/5f74a4958fa8f5188dad0e99/Data_Ethics_Framework_2020.pdf

[4] https://transcend.io/blog/ai-ethics

[5] https://www.informationgovernanceservices.com/articles/key-data-ethics-principles/

- **Beneficence & Justice**: Maximize societal benefits, minimize harm—following the Menlo Report's addition of *public interest* to traditional bioethical principles[6]. These principles form the ethical compass of the SINTRA platform. Guided by international standards such as the UK Data Ethics Framework and frameworks like Menlo, SINTRA operationalizes ethical oversight throughout data lifecycles.

These principles form the ethical compass of the SINTRA platform. Guided by international standards such as the UK Data Ethics Framework and frameworks like Menlo, SINTRA operationalizes ethical oversight throughout data lifecycles.

4.2 GDPR and Data Protection Regulations

Operating within the EU, SINTRA must fully comply with the **General Data Protection Regulation (GDPR)**. Its key principles, and how SINTRA implements them, are:

GDPR Core Principles [7]

- **Lawfulness, Fairness, Transparency**: Each processing step is legally based (e.g., consent or legitimate interest) and transparent to users.
- **Purpose Limitation & Data Minimization**: Data is collected only for specific project goals without fresh consent or DPIA.
- **Accuracy & Storage Limitation**: Data correction mechanisms and deletion policies are enforced; recording durations in CCTV use are strictly limited.
- **Integrity, Confidentiality, & Accountability**: SINTRA uses encryption, secure APIs, audit logs, and appoints data protection officers to continually verify GDPR compliance.

Data Subject Rights

The platform fully supports:

- **Access & Rectification**: Individuals can request their data, correct errors, or withdraw at any time.
- **Erasure ("Right to be Forgotten")**: Autonomously applied when data is no longer needed or consent is withdrawn.

[6] https://studyonline.unsw.edu.au/blog/data-ethics-overview

[7] https://gdpr-info.eu/art-5-gdpr/#:~:text=5%20GDPR%20Principles%20relating%20to,lawfulness%2C%20fairness%20and%20transparency')%3B

- **Automated Decision Restrictions**: SINTRA avoids fully automated processes; human oversight is mandatory, especially in surveillance actions.

Controller & Processor Obligations

SINTRA consortium partners identify roles of controller/processor depending on the use case, ensuring legal compliance, DPIAs for high-risk sensing systems, and technical measures like differential privacy, pseudonymization, and federated learning.

Emerging AI Regulation

The EU **AI Act** complements GDPR by requiring risk-based design, transparency, human oversight, and prohibits manipulative AI. SINTRA aligns with both, preparing for future certification and compliance.

4.3 Surveillance-specific Ethical Considerations

The ethical and legal obligations within the SINTRA project extend well beyond regulatory compliance. At the heart of its design is a commitment to earning public trust, particularly as the platform operates in high-sensitivity domains like airports, ports, construction and railway sites. SINTRA incorporates AI-powered surveillance and multi-modal data fusion, and it ensures that these technologies serve societal benefits without compromising individual rights and freedoms.

A core part of SINTRA's ethical framework is grounded in fundamental principles such as transparency, accountability, fairness, privacy, and the imperative to do no harm. Transparency is maintained through both system-level and algorithmic clarity. Individuals are informed about the purpose and scope of monitoring—such as via signage in monitored areas—and the system keeps logs of why specific behaviors were flagged. For example, in the Port of Moerdijk, anomalies in vessel behavior are detected through AI, and the underlying logic (e.g., deviation in route or timing anomalies) is documented and visible for audit and review. This level of openness is crucial not only for trust but also for facilitating operator understanding and decision-making.

Accountability in SINTRA is clearly defined. Each partner organization—such as TAV, KoçSistem, or Keenfinity—is assigned roles as data controllers or processors depending on the context. A designated Data Protection Officer oversees adherence to GDPR requirements. In practice, this means that if, for instance, an AI system wrongly flags a port worker due to an emergency behavior (like rushing through a restricted area), the system is designed to log the event, trigger a human review, and use that feedback to improve future detection models. This closed-loop ensures learning and accountability in real time.

Fairness and non-discrimination are critical, especially since AI systems can inadvertently reinforce social biases if not carefully managed. SINTRA mitigates this through diverse and balanced training

datasets, pre-deployment bias audits, and extensive testing of models like Human Action Recognition (HAR) under different demographic and behavioral scenarios. A typical case is in airport terminals, where behaviors such as group prayer or prolonged standing may otherwise be misclassified as loitering or suspicious activity. SINTRA's ethical audits ensure such false positives are minimized.

Privacy is treated as a default setting rather than an afterthought. Only data strictly necessary for achieving a safety or operational purpose is collected. In many cases, identifiable details are blurred or anonymized until a valid security trigger occurs. For example, in airport retail zones, Koçtaş employs AI to detect fraudulent activity at cash registers. Routine monitoring blurs facial images and stores no identity data unless a verified incident requires a high-resolution capture for investigation.

Another foundational ethical tenet is beneficence, or the obligation to avoid harm. SINTRA avoids over-surveillance and psychological discomfort by using clear, non-intimidating notices and limiting sensor coverage to necessary zones.

From a legal standpoint, SINTRA is fully compliant with the General Data Protection Regulation (GDPR) and positions itself to meet future requirements under the forthcoming EU AI Act. Key GDPR principles such as data minimization are embedded in the platform's architecture. For instance, thermal sensors in Food & Beverage (F&B) areas are used solely to monitor crowding and temperature for safety—not to identify individuals. The platform's dashboard also allows individuals to request access to their data, correct inaccuracies, or request deletion, supporting full exercise of data subject rights. Every data transaction is logged with an audit trail to ensure transparency and redress mechanisms.

Data Protection Impact Assessments (DPIAs) are carried out for high-risk deployments, such as CCTV systems in children's play areas. These assessments evaluate potential harm and define mitigations like limiting monitoring hours or disabling sensitive features (e.g., facial recognition). This aligns with Articles 35 and 36 of the GDPR, which require such assessments for technologies likely to affect fundamental rights.

Ethical considerations unique to surveillance environments are also addressed. SINTRA applies the principle of proportionality to ensure that monitoring is not excessive. For example, in construction zones, only critical areas such as crane operation zones are monitored, while rest areas are excluded unless a specific event occurs. The platform also emphasizes edge-based processing to keep data local, reducing the need for central transmission and lowering the risk of exposure. A practical application of this is in drone surveillance scenarios, where video feeds are analyzed locally for threats like smoke or crowding, and only anonymized metadata is transmitted to central servers.

Human oversight remains a cornerstone of ethical surveillance. No AI system within SINTRA operates fully autonomously. Alerts are always reviewed by a human operator before any action is taken. For example, in retail scenarios, AI may detect patterns consistent with theft, but a trained supervisor reviews the flagged behavior before triggering a security protocol. This maintains human judgment and prevents over-dependence on algorithmic outputs.

The ethical implications of AI and multi-modal sensing are manifold. While combining data types like audio, video, thermal, and RFID improves accuracy, it also increases the risk of invasive surveillance. SINTRA addresses this through data purpose partitioning—ensuring each sensor is used only within its intended scope—and limits identity linkage across modalities unless absolutely necessary for a validated threat response.

To support accountability and user understanding, SINTRA incorporates explainable AI tools such as LIME and SHAP. These tools help operators interpret why a model flagged a particular behavior. For instance, if a fight is detected in an airport corridor, the system can explain that it was based on erratic movement, elevated noise levels, and the time of day, aiding a more informed human decision.

To prevent biased decisions, SINTRA's AI models undergo rigorous representational testing, including simulation in various settings (e.g., urban vs. rural, different lighting and noise conditions). An independent ethics board validates the fairness and inclusiveness of datasets used to train these models.

Environmental and societal impacts are also taken into account. SINTRA minimizes energy usage by using asynchronous and event-driven processing models, especially when dealing with thousands of camera feeds. Socially, the system is designed to avoid creating a sense of constant surveillance that could alter natural human behavior. In retail use cases, mmWave radar sensors are preferred over cameras for occupancy monitoring, as they do not capture any identifiable imagery, thus preserving both privacy and dignity.

In summary, SINTRA integrates ethical and legal principles at every layer of its design and operation. These principles are not isolated rules but interconnected elements that shape the system's architecture, governance model, and daily use. By embedding ethics into its foundation, SINTRA not only ensures compliance with current laws but also sets a precedent for trustworthy AI and responsible surveillance in critical infrastructure environments.

4.4 Ethical Implications of AI and Multi-Modal Sensing

The integration of Artificial Intelligence (AI) and multi-modal sensing technologies within SINTRA introduces transformative capabilities for security and situational awareness. However, these capabilities also bring significant ethical considerations that must be addressed throughout system

design, deployment, and operation. AI systems, particularly those used in surveillance and public safety, can influence human behavior, shape decision-making processes, and affect fundamental rights. Ethical data handling in such contexts is not just a regulatory obligation but a moral imperative. SINTRA's approach to managing these implications focuses on six critical dimensions: algorithmic transparency, fairness and bias mitigation, privacy-preserving AI, safety and resilience, human oversight, and societal and environmental responsibility.

Algorithmic Transparency and Explainability

One of the main ethical challenges of using AI in complex environments is the "black-box" nature of many machine learning models. Without insight into how a model reaches its conclusions—such as flagging a person's behavior as suspicious—it becomes difficult to trust or challenge the outcome. SINTRA addresses this through the integration of explainability tools like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations). These frameworks provide human-readable explanations of AI decisions, allowing security personnel to see which features (e.g., movement speed, direction changes, sound levels) contributed to a model's output. For instance, if an AI system detects a potential confrontation in an airport terminal, explainability modules can show that the conclusion was based on a combination of erratic movement and elevated audio patterns. This transparency is vital not only for internal accountability but also for public trust, particularly in contexts where decisions may have legal or social consequences. Moreover, transparency supports compliance with GDPR's requirement for "meaningful information about the logic involved" in automated processing.

Bias, Fairness, and Representativeness

Bias in AI systems is a well-documented risk, particularly when training data does not reflect the diversity of the population it serves. In the context of SINTRA, where systems analyze human behavior, movement, and environmental patterns, ensuring fairness is paramount. All SINTRA algorithms undergo demographic auditing to ensure that they perform equitably across variables such as gender, age, ethnicity, and physical ability. For example, Human Action Recognition (HAR) models are tested with varied data to ensure that behaviors like rapid movement or close-proximity interactions are not misinterpreted due to culturally normative differences or physical characteristics. The platform also includes periodic fairness audits and bias detection layers to identify and address any unintended discriminatory effects over time. In practice, this might involve testing whether older adults, children, or people using mobility aids are more likely to be flagged as anomalous in a crowded space. By ensuring representativeness in training and validation, SINTRA reduces the risk of systemic bias and supports ethical, inclusive design.

Privacy-Preserving AI

Given that AI in SINTRA operates on data from cameras, microphones, RFID readers, thermal sensors, and other modalities, the risk of re-identification and privacy intrusion is significant. To

counter this, SINTRA embeds privacy-preserving techniques directly into the AI model lifecycle. Federated learning allows model training to occur across distributed nodes (such as edge devices) without centralizing sensitive personal data. Instead of collecting data at a single server, learning happens locally, and only anonymized model updates are shared—greatly reducing the risk of data breaches. SINTRA also leverages differential privacy techniques and libraries like OpenDP to add mathematical noise to datasets, ensuring that individual contributions cannot be reverse-engineered or linked back to a specific person. These measures are particularly important in high-density environments like airports, where mass surveillance could otherwise become overly intrusive. Through these methods, SINTRA maintains a balance between operational effectiveness and individual privacy rights.

Safety and Adversarial Resilience

Another crucial ethical consideration is the system's resilience against errors and attacks. In security-sensitive environments, a false negative—failing to detect a real threat—or a false positive—unjustly flagging innocent behavior—can have serious consequences. SINTRA therefore stress-tests its AI models under a variety of conditions, including network latency, signal interference, and environmental noise (e.g., rain, wind, crowds). Special attention is given to adversarial attacks where bad actors might try to confuse AI systems using spoofed inputs or misleading signals. For instance, in drone-based port surveillance, attackers might attempt to hide heat signatures or jam radio signals. SINTRA's defense mechanisms include sensor redundancy, anomaly detection at the edge, and real-time feedback loops that identify and compensate for data inconsistencies. These efforts contribute to platform robustness and help ensure that AI outputs remain reliable even under unpredictable real-world conditions.

Human-in-the-Loop Oversight

SINTRA recognizes that AI must never be the final decision-maker in scenarios that impact human rights or safety. All AI outputs—whether they involve anomaly alerts, behavioral analysis, or threat predictions—are subject to human review before action is taken. This "human-in-the-loop" approach ensures that contextual judgment, empathy, and common sense remain part of the decision-making process. Operators are trained in AI ethics, data privacy, and responsible system use to ensure that they understand both the power and the limitations of the technology. This training includes recognizing algorithmic errors, questioning outputs, and knowing when to override automated recommendations. For example, if an AI system flags a group of people praying in an airport as a suspicious congregation, the human reviewer has both the authority and the responsibility to override that conclusion based on contextual understanding. This structure ensures that automation enhances human performance without undermining human dignity or discretion.

Societal and Environmental Consequences

Finally, SINTRA acknowledges that AI and surveillance systems, even when ethically designed, can produce broader societal and environmental effects. One risk is the normalization of surveillance, where people alter their behavior because they know they are being watched—potentially reducing their freedom of expression or movement. SINTRA addresses this by limiting persistent surveillance, using non-invasive sensors when possible (such as mmWave radar), and maintaining transparency with the public through clear communication and opt-out mechanisms where applicable. Another concern is environmental impact. AI models, especially those involving video analytics, can consume significant energy. SINTRA mitigates this through energy-efficient architecture, such as asynchronous image processing and task-triggered analytics, where sensors are only activated when certain thresholds are met (e.g., crowd density or noise level). These approaches reduce computing load, power usage, and carbon footprint. Ultimately, SINTRA's design philosophy embraces a broader view of ethics—one that accounts not just for legal compliance, but also for social well-being and environmental stewardship.

5 DATA GOVERNANCE AND OWNERSHIP

SINTRA's ethical framework is underpinned by rigorous data governance and clear ownership models. This is crucial given the complexity of handling sensitive, multi-modal data across diverse environments—airports, ports, construction and railway sites. The following subsections map out how SINTRA manages data throughout its lifecycle, delegates responsibility, controls access and sharing, and navigates cross-border and cross-sector regulatory landscapes.

5.1 Data Lifecycle in SINTRA

SINTRA's data-governance model is anchored in a six-stage lifecycle that applies uniformly across all use cases (airport, port, construction and railway site). Each stage embeds privacy-by-design controls and follows widely accepted data-lifecycle best-practice guidance.

1. Collection – Multi-modal data are generated continuously by CCTV and depth cameras, thermal imagers, acoustic arrays, RFID readers, environmental sensors and autonomous drones deployed on-site. For every sensor stream, project partners agree in advance on the minimum viable resolution, sampling rate and collection window needed to satisfy the concrete safety objective—e.g., 5 fps for corridor-wide loitering detection at airports, or 1 Hz acoustic snapshots for port-side explosion monitoring. These parameters are documented in the Data Collection Register together with the lawful basis (consent, legitimate interest, or vital interest) and an explicit maximum retention clock that starts the moment data leaves the device. Limiting collection at source fulfils GDPR's data-minimization principle and echoes industry advice to "ingest wisely—quality over quantity."

2. Ingestion & Registration – Raw packets traverse secure channels into the Message-Broker Layer, where every item is wrapped in a metadata envelope containing timestamp, GPS/zone identifier, sensor ID, data-type flag and an initial anonymity score. The broker writes this envelope to a tamper-evident catalogue that captures lineage, ownership and license terms, enabling downstream traceability and eventual right-to-erasure fulfilment. At this stage automated syntax and schema checks reject malformed or non-conforming payloads, preventing dirty data from polluting analytic models.

3. Pre-processing – Before data leave the edge, cascaded privacy filters apply blurring, hashing, or one-way transforms; personally, identifying frames are dropped unless a legal trigger exists (e.g., confirmed incident). Sensor-side pseudonymization removes MAC addresses and device serial numbers, while federated-learning routines keep raw imagery local—only model weight updates go upstream. This architecture curtails the flow of identifiable data into shared infrastructure and aligns with modern recommendations to embed confidentiality at the "earliest feasible point" of the pipeline.

4. Storage & Retention – Ingested data land in encrypted, geo-scoped clusters—hot stores for the first 24–48 hours (airport CCTV) and cold object stores for longer-term evidentiary needs (e.g., cargo-theft investigations in ports). Access paths are restricted through role-based policies and all read/write events are sealed in an immutable audit log. Retention schedules, derived from the use-case DPIAs and national regulations, trigger automatic purging or rotation; no dataset lives longer than justified. Such disciplined rotation is singled out by GDPR advisers as a hallmark of compliant lifecycle management.

5. Processing & Analysis – The Business-Definition Layer fuses stream and runs AI workloads—anomaly detection, HAR, multi-sensor correlation—inside containerized sandboxes. Each analytic result stores a back-pointer to its anonymized source files so investigators can reconstruct the evidence chain without re-exposing raw identities. Versioned model registries and data-provenance tags allow SINTRA to roll back or re-train algorithms if bias or drift is later discovered, ensuring integrity across iterative AI releases.

6. Archival or Disposal – When the legally defined retention horizon expires or a deletion request is validated, data sets follow a two-step exit path. Material needed for contractual; insurance or criminal-proceeding obligations is exported to an encrypted archive under strict legal-hold; everything else is shredded with cryptographic wipe routines and the deletion event is logged for audit. This final stage fulfils the GDPR storage-limitation principle and closes the lifecycle loop by guaranteeing that no orphaned copies persist in backups or replicas.

Together, these six stages operationalize confidentiality, integrity and availability—recognized as the primary goals of modern Data-Lifecycle Management—while translating SINTRA's ethical commitments into day-to-day technical practice.

5.2 Roles and Responsibilities

Effective data governance in SINTRA requires clearly defined roles and responsibilities to ensure legal compliance, maintain operational integrity, and uphold public trust. This structured division of duties ensures that every action within the data lifecycle—from collection to deletion—is attributable, verifiable, and aligned with ethical and regulatory standards. These assignments reflect both the **General Data Protection Regulation (GDPR)** and the **ISO/IEC 27701** standard, which extends ISO/IEC 27001 to support Privacy Information Management Systems (PIMS).

Data Controllers

Data controllers are the organizations that determine the "purposes and means" of data processing. In the SINTRA context, these are typically the **owners and operators of the environments being monitored**, such as:

- **Airport operators** (e.g., TAV Airports),

- **Port authorities** (e.g., Port of Moerdijk),
- **Construction or railway companies**.

Controllers are ultimately responsible for compliance with GDPR and carry the legal accountability for ensuring data is processed lawfully, fairly, and transparently. They are responsible for obtaining a lawful basis for processing (e.g., consent, legitimate interest), responding to data subject access requests, and defining retention periods.

For example, a port authority using SINTRA to monitor cargo areas must determine whether thermal imaging data is retained for 24 or 48 hours, whether it can be shared with customs, and how it is accessed during emergencies. The controller must also justify the necessity and proportionality of surveillance via a Data Protection Impact Assessment (DPIA).

Data Processors

Data processors are service providers that process data on behalf of the controller and only under their instructions. Within SINTRA, this includes:

- **Technology integrators** who deploy and manage sensor networks,
- **Platform providers** like KoçSistem or other backend maintainers,
- **AI model developers and cloud operators**, who run analytics workflows but do not determine surveillance purposes.

Processors are responsible for implementing technical and organizational measures to protect the data—such as encryption, access control, anomaly detection, and role-based access. Importantly, they may not use or repurpose data for any activity beyond what the controller has explicitly defined. SINTRA contracts ensure this through binding data processing agreements (DPAs) which include audit clauses, breach-notification timelines, and confidentiality terms.

Joint Controllers

In many SINTRA scenarios, multiple parties jointly determine both the purpose and means of data processing—creating joint controllership under Article 26 of the GDPR. This applies, for example, when:

- A city authority and a construction firm collaborate on mobile surveillance,
- An airport operator and a border security agency share CCTV feeds,
- Or a drone surveillance platform is jointly operated by port security and local law enforcement.

In such arrangements, a Joint Controller Agreement (JCA) is mandatory. These JCAs specify how responsibilities are divided, for example, which party handles data subject requests, who manages access rights, and how legal compliance is monitored. They also define fallback procedures in case of disputes or breaches.

Joint controllers share accountability, but individuals can exercise their rights (e.g., access, leisure) against either party. For that reason, transparency between controllers and clear documentation of roles are essential.

Data Protection Officer (DPO)

A dedicated Data Protection Officer (DPO) plays a central role in ensuring GDPR compliance across the SINTRA project. The DPO is an independent expert tasked with:

- Monitoring internal compliance,
- Advising on and overseeing DPIAs,
- Training staff and raising awareness,
- Serving as a contact point for supervisory authorities, and
- Advising on the balancing of data subject rights with SINTRA's legitimate interests in public safety.

Given the high-risk nature of SINTRA's operations, especially involving biometric and behavioral data, the appointment of a DPO is mandatory under Article 37 of the GDPR. The DPO operates independently from the controllers and processors and reports to the highest management level within the project consortium.

In practice, the DPO might oversee a DPIA for a new AI-driven behavior detection model deployed in an airport retail zone, ensuring that risks to privacy are assessed and mitigated before deployment. The DPO would also ensure appropriate transparency measures—like signage or public notices—are in place.

Alignment with International Standards

These role definitions and responsibilities are consistent with ISO/IEC 27701, which extends the ISO/IEC 27001 information security management standard to address privacy. Under ISO/IEC 27701:

- Controllers are required to implement Privacy Information Management Systems (PIMS),
- Processors must demonstrate technical controls (e.g., access policies, incident response),
- Joint controllership must be documented with a focus on legal interoperability, and

- All parties must maintain auditable records of processing activities.

By aligning with ISO/IEC 27701, SINTRA reinforces its commitment to structured accountability, cross-functional clarity, and international interoperability, especially important in a project involving cross-border deployments and multiple stakeholders.

Establishing and clearly documenting these roles is not merely a legal requirement but a foundational element of ethical governance. Each SINTRA partner, whether public or private, has specific obligations and must collaborate seamlessly to ensure that data is handled with integrity, transparency, and respect for individual rights. These defined responsibilities also form the backbone of SINTRA's incident response, audit readiness, and trust-building with both regulators and the public.

5.3 Data Sharing and Access Control

In a distributed, multi-partner platform like SINTRA—where data flows between public authorities, private operators, and AI-driven analytic systems—ensuring ethical, lawful, and controlled data access is non-negotiable. SINTRA adopts a layered access control model that combines technical enforcement mechanisms, dynamic policy evaluation, and ethical oversight to preserve confidentiality, integrity, and purpose limitation throughout the data-sharing lifecycle.

Role-Based Access Control (RBAC)

At the core of SINTRA's sharing model is Role-Based Access Control (RBAC), which restricts data access based on predefined user roles and their corresponding duties within the system. Access policies are codified in machine-readable form—known as policy-as-code—and enforced automatically through infrastructure-level orchestration. Each user, whether a security analyst, airport IT staff member, customs liaison, or system integrator, is assigned a role with scoped permissions. For example, an AI engineer might only access de-identified datasets for model retraining, while an incident response officer may access time-limited, high-resolution footage associated with a verified anomaly.

RBAC ensures that:

- No user sees more than necessary (principle of least privilege),
- Privileges are easily auditable and revocable,
- Access can be aligned with external certifications or job clearance levels.

This rigid segmentation is essential for compliance with GDPR's data minimization and integrity and confidentiality principles, and it reflects NIST's Zero Trust Architecture guidelines that emphasize contextual, need-based authorization.

Fine-Grained Access Controls (FGAC)

To complement RBAC, SINTRA implements **Fine-Grained Access Controls (FGAC)**a more dynamic mechanism that considers multiple real-time attributes such as user location, time of access, system load, security incident level, and even behavioral cues.

For instance, access to drone surveillance feeds at the port might be restricted to:

- Specific hours (e.g., during docking),
- A geofenced control room,
- Only when two-factor authentication is verified.

AI-enabled monitoring tracks usage patterns and revokes or escalates access if anomalies are detected, for example, if a user logs in from an unusual IP address, accesses an atypical volume of data, or attempts to bypass standard query tools. This intelligent adaptation protects against internal misuse, credential compromise, or procedural drift. As highlighted in enterprise AI data governance literature, such adaptive models are increasingly critical for high-risk deployments where static access rules can't keep up with evolving threats or user behavior patterns.

The "Five Safes" Framework

To evaluate and control data sharing beyond SINTRA's internal stakeholders—especially when responding to third-party requests or cross-institutional research projects—SINTRA uses the **Five Safes Framework**, a model originally developed by the UK Office for National Statistics. It consists of:

1. **Safe People** – Only qualified and trained individuals may access sensitive data.
2. **Safe Projects** – Data is shared only for legitimate and ethically approved purposes.
3. **Safe Data** – Data is de-identified or aggregated to minimize re-identification risk.
4. **Safe Settings** – Data access takes place in controlled environments, such as secure cloud sandboxes or dedicated access terminals.
5. **Safe Outputs** – Results of data use (e.g., reports, visualizations) are checked to ensure they do not disclose sensitive or identifiable information.

In practice, this means that if an academic partner or public agency requests access to SINTRA data—for example, to study crowd behavior in airports—every stage of the access workflow must pass the Five Safes evaluation. This ensures that public interest projects benefit from SINTRA's capabilities without undermining privacy or ethical standards.

Secure APIs and Consent Management

SINTRA enables authorized external systems—such as emergency services, customs enforcement, or traffic control—to interact with relevant data modules via Secure APIs. These interfaces are protected by mutual authentication, rate limiting, and encryption, and they enforce purpose limitation by tying every data request to a consent policy or legal basis.

Consent, when applicable (e.g., in retail analytics), is obtained via user-friendly interfaces and dashboards. All consent interactions—grants, withdrawals, scope limits—are recorded in a Consent Management Platform (CMP). In other scenarios where consent is not feasible, SINTRA documents the alternative legal justification (e.g., vital interest, public task) and logs it as metadata alongside the data access event.

This structure ensures SINTRA is aligned with GDPR Articles 6 and 7, which govern lawful basis for processing and consent requirements, as well as Article 25, which mandates data protection by design and by default.

Immutable Audit Trails

Finally, every data access, transformation, or sharing action is captured in a secure, immutable audit log. These logs are:

- Timestamped,
- Signed with cryptographic hashes,
- Stored in tamper-evident ledgers or blockchain-inspired append-only logs.

Audit trails enable real-time oversight, retrospective forensic analysis, and proactive risk detection. They also form the foundation for accountability reporting to data subjects, supervisory authorities (e.g., EU DPAs), and internal ethics boards.

By making every action traceable, SINTRA not only deters unauthorized activity but also ensures that data subjects can exercise their right to know who accessed their data, when, and for what purpose.

SINTRA's approach to data sharing and access control exemplifies how security, ethics, and utility can coexist in complex AI-driven systems. Through a blend of RBAC, FGAC, Five Safes evaluation, secure interfaces, and traceability-by-default, the platform guarantees that sensitive data is accessible only by the right people, at the right time, for the right reasons. This tightly governed structure not only ensures legal compliance but fosters the trust required for widespread acceptance of intelligent surveillance technologies.

5.4 Cross-border and Cross-sectoral Data Governance

SINTRA operates at the intersection of public security, AI innovation, and multi-modal data processing—domains that inherently span institutional, geographic, and regulatory boundaries. Given the involvement of partners across multiple EU countries (and potentially beyond), as well as the integration of data from transportation hubs, law enforcement, retail, and construction sectors, SINTRA faces complex legal and technical challenges around data movement, interoperability, and sovereignty.

To address these challenges, SINTRA adheres to a multi-layered data governance framework that supports lawful cross-border flows, cross-sector collaboration, and local control without compromising ethical standards or operational agility.

Cross-Border Data Transfers under GDPR

When personal data flows beyond national borders—particularly from the European Economic Area (EEA) to third countries, SINTRA must comply with Chapter V of the GDPR, which governs international data transfers. These transfers are permitted only if the recipient country ensures an adequate level of data protection or if other safeguard mechanisms are implemented.

Key legal tools used include:

- **EU Adequacy Decisions** – Simplify transfers to jurisdictions officially recognized as providing equivalent protection (e.g., Japan, Switzerland).
- **Standard Contractual Clauses (SCCs)** – Legally binding templates adopted by the European Commission for data-sharing between EEA and non-EEA entities. SINTRA uses SCCs for cloud vendors or AI model developers operating internationally.
- **Binding Corporate Rules (BCRs)** – Applied by multinational consortium members (e.g., Keenfinity) to enable secure intra-group data transfers.
- **Transfer Impact Assessments (TIAs)** – Assess risks posed by foreign surveillance laws and ensure that supplemental safeguards (e.g., encryption, access restrictions) are applied when transferring data to countries lacking adequacy decisions.

These instruments ensure that sensitive data—such as airport surveillance footage or behavioral analytics—can be securely processed, even when algorithms are trained or hosted across borders. This safeguards individual rights while enabling technological collaboration and scalability.

Cross-Sectoral Compliance: Public-Private Interoperability

SINTRA is uniquely positioned at the junction of multiple verticals: public safety, transportation, retail, logistics, and health. Each of these domains is governed by its own set of compliance

frameworks, risk tolerances, and ethical norms. To facilitate lawful and ethical inter-sectoral data exchange, SINTRA aligns its governance model with emerging EU policy tools such as:

- **The EU Data Governance Act (DGA)** – Promotes secure data-sharing across sectors, allowing entities to share public sector, health, and industrial data through regulated intermediaries known as "data altruism organizations" or "data cooperatives."
- **European Health Data Space (EHDS)** – Provides rules for securely sharing and analyzing health-related data (e.g., biometric indicators, thermal imagery) across healthcare and security domains.

In practice, this means that SINTRA partners—such as airports coordinating with customs and emergency medical services—can jointly access incident data under clear, legally compliant structures. By participating in these regulated frameworks, SINTRA ensures that collaboration does not erode privacy or ethical accountability.

Data Sovereignty and Local Control

An essential principle in modern data ethics is data sovereignty—ensuring that data remains under the control of the jurisdiction or organization that owns or originates it. SINTRA maintains sovereignty by implementing the following technical and operational safeguards:

- **Geo-fenced Cloud Infrastructure** – Data are stored and processed within predefined geographic boundaries (e.g., within the EU), preventing unauthorized replication or backup to third-country servers.
- **Replication Controls** – System-level rules prevent automated or accidental duplication of sensitive datasets across regions. For example, drone imagery captured in a Dutch port cannot be mirrored in a U.S.-based data center without an explicit legal justification and audit trail.
- **Data Localization Policies** – Particularly in sensitive zones (e.g., children's play areas, customs inspection zones), SINTRA restricts data flow to local edge devices and prohibits upload to central systems unless a critical threshold is crossed.

By embedding these controls, SINTRA adheres to national and EU-level data localization laws, reinforcing public trust and reducing legal exposure.

Federated Learning and Analytics: Decentralized AI

To further respect jurisdictional boundaries while enabling scalable AI training and analytics, SINTRA employs Federated Learning (FL) and Federated Analytics (FA) methodologies. These approaches allow models to be trained across distributed nodes (e.g., edge devices at airports, servers in national facilities) without centralizing raw data.

Instead of pooling sensitive datasets in a single location, each node trains a local model and shares encrypted model updates (e.g., weight vectors, gradient changes) with a central aggregator. The aggregator then refines the global model without ever seeing identifiable personal data.

This decentralized architecture offers multiple advantages:

- **Legal Compliance** – Supports data localization requirements and avoids triggering cross-border transfer clauses.
- **Privacy Preservation** – Reduces exposure of raw personal data, mitigating re-identification risks.
- **Operational Efficiency** – Enables real-time learning and adaptation in highly distributed environments like ports.

SINTRA's use of FL/FA aligns with emerging guidance from organizations such as the OECD, EDPB, and AI Now Institute, which recommend distributed AI methods for high-risk, high-sensitivity domains.

Multi-Level Interoperability: Horizontal and Vertical

SINTRA's governance framework also addresses the two main axes of data collaboration:

- **Horizontal Integration** – Sharing data across sectors (e.g., airport + customs + health) while maintaining sector-specific compliance and ethical constraints.
- **Vertical Collaboration** – Coordinating between different layers of government and oversight (e.g., local security forces, national data protection authorities, EU-level ethics boards).

Interoperability mechanisms include standardized metadata formats, consent traceability, modular policy enforcement engines, and shared vocabularies to enable seamless—but controlled—data exchange.

Cross-border and cross-sectoral data governance in SINTRA is not just a technical challenge, it is a legal, ethical, and strategic imperative. By combining GDPR-compliant transfer mechanisms, alignment with EU data governance policies, strong data sovereignty controls, and decentralized AI methods, SINTRA ensures that its platform remains interoperable without compromising on individual rights or national regulatory integrity. This positions SINTRA as a leading example of ethically aligned data governance in the age of intelligent, multi-jurisdictional surveillance.

6 PRIVACY-BY-DESIGN AND TECHNICAL SAFEGUARDS

SINTRA's privacy-by-design approach embeds strong technical and organizational safeguards across its architecture and AI-driven processes. This ensures that even as the platform delivers powerful analytics in sensitive contexts—airports, ports, construction and railway sites—privacy remains a default setting, not an afterthought. Below are key strategies and methods used to protect personal data from collection to deletion.

6.1 Privacy-Preserving Techniques

SINTRA employs a layered suite of privacy-enhancing technologies (PETs) to shield data throughout its lifecycle. These include but are not limited to:

- **Federated learning**: Distributes model training across edge devices, keeping raw data local and sharing only encrypted model updates. This maintains data sovereignty and supports several GDPR-compliant transfer scenarios.
- **Differential privacy**: Adds controlled noise to aggregated outputs to prevent leakage of individual-level information.
- **Homomorphic encryption**: Enables computation over encrypted data, allowing models to process data without revealing it.
- **Secure Multi-Party Computation (SMPC)**: Enables collaborative analysis among multiple partners without exposing raw data.

These PETs are combined strategically to form hybrid privacy frameworks—for instance, applying differential privacy atop federated learning or homomorphic encryption for secure updates—achieving both high utility and strong privacy guarantees.

6.2 Anonymization, Pseudonymization, and Differential Privacy

Anonymization and **pseudonymization** are applied to data at the edge and ingestion stages:

- **Anonymization**: Irreversible transformations such as blurring, generalization, or noise insertion render personally identifiable data unrecognizable.
- **Pseudonymization**: Assigns randomized identifiers to entities, decoupling data from real identities while preserving referential consistency for analytics.

These practices align with GDPR mandates, enabling SINTRA to classify data for less restrictive processing when anonymization is verifiable.

Differential privacy provides an additional, mathematically provable layer of protection. SINTRA's analytics Modules—especially for aggregated dashboards—use noise injection techniques or frameworks like OpenDP and SmartNoise to mask exact individual contributions. This is essential in contexts such as crowd monitoring or behavioral pattern analysis, where aggregated insights are valuable but privacy must be ensured.

6.3 Homomorphic Encryption and Federated Learning

Homomorphic Encryption (HE) allows encrypted data to be processed by AI models without being decrypted. SINTRA leverages both partial and fully homomorphic schemes to secure high-sensitivity operations, particularly during federated model updates. For example, in port drone analytics, gradient updates are encrypted before they leave local nodes, ensuring no individual data is exposed—even in transit or at rest.

Federated Learning (FL) complements this by keeping raw data local to each domain (e.g., airport zones, ports). Only encrypted model updates are aggregated. This dual approach supports legal compliance—minimizing personal data movement across borders—and technical robustness, as physics and behavioral patterns can be learned collaboratively without exposing raw sensor data.

When combined with HE, SINTRA achieves highly secure, privacy-conscious distributed training, beneficial in mixed-jurisdiction deployments.

6.4 Use of OpenDP, Diffprivlib and Other Tools

SINTRA harnesses open-source privacy toolkits to standardize and streamline privacy measures:

- **OpenDP**: A modular differential privacy library developed in Rust (with Python/R bindings), vetted for performance and safety.
- **SmartNoise**: Built on OpenDP, it enables DP-compliant queries and synthetic dataset generation for structured analytics.
- **IBM Diffprivlib, TensorFlow Privacy, and PySyft**: Offer DP and federated learning extensions compatible with SINTRA's modular architecture.

By integrating these tools, SINTRA increases transparency, ensures compliance, and benefits from the latest academic research without reinventing the wheel.

7 USE CASE-SPECIFIC ETHICAL DATA HANDLING

7.1 Airport Use Case

Applying Ethical Data Handling Guidelines to Airport Security Operations

The implementation of advanced CCTV architecture, augmented with AI capabilities, requires a careful equilibrium between strengthening security measures and safeguarding individual rights at airports. It is crucial to stay aligned with privacy-by-design, transparency, accountability, and the responsible utilization of AI and multi-modal sensor data and consistent to actual security incidents and operational contexts existing in an airport environment. Through an examination of both frequent and infrequent occurrences, our aim is to demonstrate SINTRA's dedication to developing systems that are not only effective but also ethically sound, compliant with regulations, and deserving of trust.

1. Privacy-by-Design in Action: Safeguarding Individual Rights from Inception

The guidelines emphasize that privacy considerations must be embedded into the very architecture and design of the SINTRA system, rather than being an afterthought. This means implementing privacy-preserving techniques directly into the AI model lifecycle, from data ingestion and edge processing to storage and analysis. The goal is to minimize the collection of personally identifiable information (PII) and ensure that when PII is necessary, it is handled with the utmost care and strict controls.

Use Case 1.1: Locating Lost Children/Vulnerable Individuals

Scenario: A common and distressing occurrence in busy airports is a child or a vulnerable adult (e.g., elderly person with dementia) becoming separated from their guardian. Rapid identification and reunion are critical, but so is protecting their privacy and that of others.

D2.2 Application: SINTRA's CCTV system would initially operate with a high degree of **pseudonymization** by default. All individuals captured by the cameras would be represented by temporary, non-identifiable identifiers (e.g., a bounding box with a unique, session-based ID) rather than their raw facial features or other PII. This means that general security operators viewing the live feed would see movement patterns and crowd dynamics, but not readily identifiable faces.

When a "lost child" report is received, an authorized security operator, with specific credentials and a defined purpose, would initiate a targeted search. The system would then allow temporary, controlled **de-anonymization** of relevant footage for a limited time and specific area to locate the child. This process would be meticulously logged, detailing who accessed the data, when, for what purpose, and for how long. Once the child is located and the incident resolved, the system would

revert to its default pseudonymized state for that footage, or purge it according to defined data retention policies. This ensures that identifiable data is only revealed when absolutely necessary, for a vital interest, and under stringent oversight, embodying the "privacy as the default" principle.

Use Case 1.2: Optimized Queue Management and Crowd Flow Analysis

Scenario: Airports constantly strive to optimize passenger flow through security checkpoints, immigration, and boarding gates to reduce wait times and enhance efficiency. This requires understanding crowd dynamics without tracking individuals.

D2.2 Application: For **queue management**, SINTRA's AI would analyze **anonymized aggregate data** from CCTV and other sensors (e.g., depth sensors, , mmWave radar, LiDAR). The system would focus on metrics like crowd density, flow rates, and average wait times, using techniques such as object counting and trajectory analysis of anonymized shapes. No individual passenger's journey or identity would be tracked or stored. The AI's output would be statistical insights (e.g., "Queue A is 80% capacity, average wait 15 minutes") presented on a dashboard for operational staff to reallocate resources or open new lanes. This demonstrates **data minimization** and **purpose limitation**, ensuring that data collected is strictly relevant to the operational goal and does not infringe on individual privacy.

In addition to CCTV and depth sensors, we have the potential of mmWave radar (e.g., Texas Instruments IWR1843BOOST) for monitoring passenger flow in airport environments. Radar point clouds enable the detection and localization of individuals in 3D space without capturing any biometric or personally identifiable information, ensuring full anonymity. Unlike cameras, radar is resilient to lighting changes, occlusion, and adverse weather, which makes it highly suitable for continuous operation in dynamic airport environments. Early-stage work within SINTRA has demonstrated that radar-based human detection, combined with clustering and trajectory estimation, can provide reliable crowd density and flow metrics. These insights can complement camera-based analytics while enhancing privacy-by-design, as radar inherently avoids recording visual identity features.

Use Case 1.3: Restricted Area Breach Detection

Scenario: An individual attempts to enter a highly sensitive or restricted area of the airport (e.g., tarmac access points, baggage handling facilities, air traffic control towers) without authorization.

D2.2 Application: In such critical zones, SINTRA's system would prioritize **intrusion detection** over individual identification. The AI would be trained to detect anomalous movement patterns or the presence of objects in unauthorized zones. For instance, if a person crosses a virtual perimeter, the system would trigger an alert. The initial alert might involve a blurred image or a generic "intruder detected" notification. **De-anonymization** (revealing the individual's identity) would only occur *after* a human operator confirms a legitimate breach and deems it necessary for

security response. This adheres to the **proportionality** principle, where the level of data processing is proportionate to the risk and the necessity of the security incident. Furthermore, the system would employ **edge processing** to analyze movement patterns locally, minimizing the transmission of raw, identifiable video streams to central servers unless an actual security event is confirmed.

mmWave radar has also been evaluated for security-sensitive airport zones. The advantage of radar in this context lies in its ability to detect motion patterns and intrusions even in low-visibility or camera-blind areas (e.g., smoke, darkness, weather). SINTRA's radar pipeline is being extended with clustering and tracking algorithms to identify the presence and movement of humans in restricted areas.

Radar-based intrusion alerts are inherently privacy-preserving: the system only generates an abstract 3D representation (e.g., bounding boxes or trajectories) without revealing any biometric identity. If integrated with other modalities (CCTV, access control systems), radar can serve as an additional privacy-first sensing modality, supporting proportional and layered security responses as described in D2.2.

2. Transparency for Trust and Accountability: Unveiling the AI's Logic

Transparency is a cornerstone of ethical AI, fostering trust among passengers, staff, and regulators. SINTRA's guidelines mandate clarity about how data is collected, processed, and analyzed, including publishing information on algorithms, anonymization methods, and the logic behind AI decisions. This ensures that stakeholders understand the system's capabilities and limitations.

Use Case 2.1: Handling Unattended Baggage

Scenario: A passenger inadvertently leaves a suitcase or backpack unattended in a busy terminal, which could be a security risk or simply a lost item.

D2.2 Application: When SINTRA's AI flags an "unattended bag," the system's visualization layer would not only display an alert but also provide a **simple, human-readable explanation** for why it was flagged. For example, the alert might state: "Object stationary for 15 minutes in high-traffic zone, no human interaction detected within 5-meter radius." This **explainability** is a key aspect of transparency, allowing security personnel to quickly understand the AI's reasoning, assess the situation accurately, and avoid false alarms.

Furthermore, airports deploying SINTRA would be encouraged to implement **public signage** in prominent locations, informing passengers about the presence of AI-powered surveillance for security purposes, specifically mentioning detection of unattended items. This proactive communication demonstrates **operational transparency**, ensuring passengers are aware of the

monitoring and its purpose, fostering trust and compliance. All actions taken by the system (e.g., flagging, human review, resolution) would be logged in a transparent audit trail, accessible for internal review and regulatory compliance.

Use Case 2.2: Detecting Abnormal Passenger Behavior

Scenario: The AI is designed to identify unusual or potentially suspicious behaviors that might indicate a threat, such as erratic movements, sudden running, or aggressive interactions.

D2.2 Application: For **abnormal passenger behavior detection**, SINTRA's AI models would be designed with **algorithmic transparency** in mind. While the underlying neural networks can be complex, the system would provide indicators that contribute to a "suspicious" score. For instance, if an individual is flagged, the system might highlight contributing factors like "rapid change in direction," "prolonged loitering in restricted area," or "sudden increase in speed." This helps human operators understand the AI's inference rather than blindly trusting an opaque "black box."

Crucially, the **rules for flagging** certain behaviors would be clearly defined and subject to regular review by human experts and potentially an ethics board. This ensures that the AI's definition of "abnormal" aligns with ethical considerations and avoids bias. Any human override of an AI alert would also be logged, contributing to a continuous feedback loop for model refinement and demonstrating accountability in decision-making.

3. Accountability Through Clear Roles and Auditing: Ensuring Responsible Governance

Accountability is fundamental to ethical data governance, establishing clear responsibilities for data management and providing mechanisms for audits, redress, and incident response. SINTRA's guidelines mandate meticulous logging of every data access, model inference, and human intervention, creating an unalterable audit trail.

Use Case 3.1: Responding to Minor Medical Incidents (e.g., Fainting, Slips and Falls)

Scenario: A passenger experiences a sudden medical emergency, such as fainting or falling, in a public area of the airport.

D2.2 Application: When SINTRA's AI detects a "fall" or "motionless person," it triggers an alert for immediate human review. The **airport operator** (e.g., TAV Airports), designated as the **Data Controller**, holds ultimate responsibility for ensuring that appropriate medical assistance is dispatched promptly and that any associated data (e.g., time, location of fall, relevant footage) is handled in accordance with privacy regulations.

Every access to the incident's data by medical staff, security personnel, or other authorized responders is recorded in an **immutable audit log**. This log details the user, timestamp, data

accessed, and purpose, ensuring full **traceability** for retrospective analysis, internal investigations, or external regulatory audits. Should a complaint arise regarding data handling, this detailed log provides the necessary evidence for accountability and redress, aligning perfectly with GDPR's stringent accountability requirements. The **Data Protection Officer (DPO)**, a role mandated by GDPR, would oversee these processes, ensuring compliance and addressing any data subject requests.

Use Case 3.2: Managing Access Control Violations

Scenario: An unauthorized individual attempts to bypass a secure access point leading to a restricted area, such as an employee-only entrance or a baggage sorting facility.

D2.2 Application: SINTRA's system, through integrated access control sensors and AI, would detect such attempts. The principle of **clear chain of responsibility** is vital here. The security team responsible for that specific zone would be immediately alerted. Every step of the incident – from the initial detection by the AI, to the alert being sent, to the human operator's response, and any subsequent actions (e.g., dispatching security, apprehending the individual) – would be meticulously **logged**. This includes timestamps, the identity of the person attempting access, the specific access point, and the security personnel involved. This detailed logging ensures that in the event of a successful breach or a subsequent investigation, there is a complete and verifiable record, upholding **accountability** for both the system's performance and human intervention.

Use Case 3.3: Incident Response and Forensic Analysis

Scenario: Following a significant security incident (e.g., a theft, an act of vandalism, or a verified threat), security teams need to review past events to understand the sequence of actions and identify perpetrators.

D2.2 Application: For **incident response and forensic analysis**, SINTRA adheres to strict **data retention policies** as outlined in guidelines. Raw video footage and associated metadata would be securely stored for a legally permissible and operationally necessary period (e.g., 30 days, or longer if a specific incident requires it). Access to this historical data for forensic purposes would be highly restricted, requiring multiple levels of authorization and conducted within secure environments.

The system would ensure **data integrity** through cryptographic hashing and tamper-evident logging, guaranteeing that the footage and logs used in investigations have not been altered. The **audit trail** would precisely record every instance of data access for forensic review, including the specific segments of video viewed, by whom, and for what investigative purpose. This robust framework ensures that data used for critical incident response is reliable, legally admissible, and handled accountably, providing a strong foundation for post-incident analysis and legal proceedings.

4. Responsible AI and Multi-Modal Sensor Data: Ethical Deployment of Advanced Technologies

SINTRA's approach to managing ethical implications extends beyond mere compliance, focusing on the responsible development and deployment of AI. This involves ensuring algorithmic transparency, mitigating bias, preserving privacy through advanced techniques, building safety and resilience, maintaining human oversight, and considering broader societal and environmental impacts. The integration of multi-modal sensor data amplifies the need for these considerations.

Use Case 4.1: Responding to Active Shooter/Terrorist Attack

Scenario: An extremely rare but devastating event where an individual or group initiates a violent attack within the airport premises.

D2.2 Application: In this high-stakes scenario, SINTRA leverages **multi-modal sensor fusion** to provide comprehensive situational awareness. CCTV feeds are integrated with acoustic sensors (trained to detect gunshots or explosions), thermal sensors (to identify individuals in smoke or low light), and potentially ground radar for movement tracking. While AI rapidly identifies anomalies like "erratic movement," "large, sudden congregations," or "confrontation," the **Human-in-the-Loop (HITL)** principle is paramount. AI alerts are immediate and prioritized, but a human operator always reviews the situation and makes the final decision on response protocols. This prevents over-reliance on automated systems in critical, life-threatening situations and ensures ethical considerations (e.g., avoiding misidentification, ensuring proportionality of force) are maintained.

The system is also designed with **adversarial robustness** in mind, anticipating and mitigating attempts to spoof sensors or mislead the AI during an attack. This involves continuous training with diverse data, including simulated adversarial scenarios. Furthermore, any deployment of such high-impact AI features would undergo rigorous **ethical review** by an independent board, as stipulated, to ensure that the potential benefits outweigh the risks and that all safeguards are in place.

Use Case 4.2: Detecting Unauthorized Indoor Drones

Scenario: An unauthorized drone is detected flying within the airport terminal or restricted airspace, posing a significant security and safety threat (e.g., for espionage, disruption, or carrying contraband).

D2.2 Application: When an unauthorized drone is detected, SINTRA applies **modality partitioning**. Instead of solely relying on CCTV, which might inadvertently capture sensitive personal data of individuals while trying to track a drone, dedicated acoustic or radar sensors would primarily focus on the drone's presence, trajectory, and unique signatures. This approach **limits data collection to the specific threat**, reducing the scope of personal data processing. For instance, acoustic

sensors would identify the drone's distinct motor hum, and radar could track its movement without capturing any visual imagery of the surrounding environment or people. This demonstrates responsible use of AI and multi-modal sensing by ensuring each sensor is utilized only within its intended and most privacy-preserving scope. Only upon confirmed detection and threat assessment would CCTV be used in a targeted manner to identify the drone's operator, if necessary, and with strict logging.

Use Case 4.3: Ethical Congregation Detection and Management

Scenario: Large groups of people gathering, whether for planned events (e.g., protests, demonstrations) or unexpected reasons (e.g., flight delays, emergencies), can pose crowd control challenges or security risks.

D2.2 Application: SINTRA's AI for congregation detection must be developed with a strong focus on fairness and bias mitigation. The AI models would be rigorously tested on diverse datasets to ensure they do not disproportionately flag certain demographic groups (e.g., based on ethnicity, religion, or attire) as "suspicious congregations." For example, a group of passengers praying or a large family gathering should not be misidentified as a potential threat. The system would focus on behavioral patterns (e.g., sudden increase in density, aggressive interactions, blocking pathways) rather than group characteristics.

In terms of responsible management, if a large, potentially problematic congregation is detected, the AI would provide data on density and movement, but the decision to intervene or disperse the crowd would always rest with human security personnel. The system would support a graduated response, providing data for informed decisions, rather than autonomously recommending actions that could infringe on freedom of assembly or cause undue alarm.

Use Case 4.4: Perimeter Security and Intrusion Detection

Scenario: An attempt is made to breach the airport's external perimeter fence or other controlled boundaries, potentially by individuals or vehicles.

D2.2 Application: For **perimeter security**, SINTRA integrates a variety of sensors, including thermal cameras, ground radar, and fence-mounted vibration sensors, alongside CCTV. This **multi-modal approach** enhances accuracy and minimizes false positives (e.g., animals triggering alerts). The AI's primary function is **intrusion detection**, focusing on identifying unauthorized movement across boundaries. The system would be designed to filter out environmental noise and non-threatening movements.

From a **societal and environmental responsibility** perspective, the deployment of these sensors would be carefully planned to avoid unnecessary surveillance of surrounding residential areas. The system would be calibrated to focus solely on the airport's boundary, ensuring that its operational impact is confined to its intended security purpose. Alerts would be specific to

intrusion events, and visual confirmation via CCTV would only be activated upon an initial alert from a non-visual sensor, further minimizing continuous, broad-area surveillance.

5. Ethical Data Handling Guidelines for Object/Anomaly Detection in Airport Terminals

1. Purpose Limitation and Contextual Integrity

- Object detection in airport terminals must be strictly limited to safety, security, and operational efficiency goals (e.g., left luggage, perimeter intrusion, crowding).
- The detection system must avoid function creep—data collected for object detection must not be repurposed (e.g., for profiling, commercial behavior tracking, or biometric marketing) without renewed consent or legal basis.
- Each object class (e.g., luggage, weapons, unattended items) must be clearly tied to its legitimate detection purpose, and only relevant classes should be processed within sensitive zones.
- mmWave radar point clouds are used solely for presence detection, people flow analytics, and restricted-area intrusion signals—no identity inference or profiling.

2. Informed Consent and Transparency

- Passengers and staff must be informed via signage and accessible digital notices about the presence of AI-based object detection systems, including:
    - The types of objects being detected.
    - Whether images are stored, and for how long.
    - Who can access the data and under what circumstances.
- For indirect consent scenarios (e.g., public CCTV), ensure clear justification under GDPR's legitimate interest or public safety legal bases, supported by DPIAs.

3. Data Minimization and Proportionality

- Only the minimal necessary visual data must be processed:
    - Use low-resolution feeds or edge blur filters where object classification is possible without identity detection.
    - Deploy masking or cropping in areas not relevant to object detection (e.g., restrooms, food courts).

- Avoid over-surveillance by activating high-resolution or AI-enhanced analysis only in high-risk zones or on triggering events (e.g., an abandoned item or restricted entry breach).

4. Privacy-Preserving Techniques

- Apply anonymization at ingestion where feasible, especially when facial features or identity-linked behaviors may be present.
- Use privacy-enhancing technologies (PETs):
    - Federated Learning to train object detection models across different terminals without centralizing image data.
    - Differential Privacy to add noise in statistical object detection logs (e.g., frequency of unattended baggage events).
    - Homomorphic Encryption when aggregating results across terminals or agencies.
- In mmWave point cloud data, there is no visual appearance or biometric features. Data has information about position and motion of passengers.

5. Fairness and Bias Mitigation

- Ensure object detection models do not introduce bias in detection or classification:
    - Validate against multiple camera perspectives (e.g., top-down, oblique) and lighting scenarios to ensure consistent detection across terminals.
    - Test across diverse object appearances (e.g., luggage of different shapes/colors/sizes) to avoid false positives related to aesthetics or cultural norms.
- Maintain fairness especially in anomaly detection where contextual factors (e.g., cultural attire, group behavior) might falsely signal risk.
- Radar sensing is invariant to skin tone, clothing, and visual appearance; nevertheless, validate performance for mobility aids, body sizes, and carry-ons.

6. Explainability and Operator Understanding

- Use explainable AI tools like SHAP or LIME to enable operators to understand the rationale behind flagged objects or anomalies:
    - Example: “Item left unattended for 4+ minutes in a restricted zone during off-peak hours; matched suspicious size signature.”

- Provide visual overlays or AR dashboards that show bounding boxes, alert confidence levels, and object classification rationale.

7. Data Storage, Retention, and Access Rights

- Video and detection metadata should be stored only:
    - For durations legally required (e.g., 7–30 days for most EU jurisdictions), or
    - Until resolved security incidents.
- Individuals must be able to request access, correction, or deletion of data in compliance with GDPR Articles 15–17.
- Use tiered access control with immutable audit logs for all retrieval and inference events.

8. Security and Adversarial Robustness

- Apply robust defenses against tampering or adversarial attacks:
    - Test detection models for spoofing vulnerabilities (e.g., printed objects mimicking threats).
    - Use redundant sensing (thermal + visual) to verify object presence.
    - Monitor for model drift or abnormal error spikes over time.

9. Human Oversight and Ethical Governance

- Ensure all object detection alerts pass through a human-in-the-loop mechanism before triggering law enforcement or automated response.
- Train operators on:
    - The limits of AI accuracy.
    - Situational judgment during ambiguous or low-confidence detections.
    - Ethical protocols for escalation or investigation.

10. Continuous Evaluation and DPIAs

- Run Periodic Ethical Reviews:
    - Audit detection logs for false positives/negatives and their consequences.
    - Conduct post-incident assessments if false detection led to unjust treatment (e.g., racial profiling, unnecessary searches).

- Refresh Data Protection Impact Assessments (DPIAs) with every major model update, camera relocation, or terminal configuration change.

Anomaly Detection in Airport Food & Beverage Areas

Scenario: Food and beverages areas like cafés and restaurants in airports are high-density zones where passenger health and safety risks may occur due to incidents such as coughing, fainting, falling, running, or fighting. Detecting these anomalies early is critical for timely intervention by airport safety and security teams.

D2.2 Application: In the scenario, SINTRA applies multi-modal AI anomaly detection that uses a combination of cameras, acoustic, and air quality sensors to monitor passenger behavior in food and beverage zones.

The system operates with privacy-by-design principles: visual inputs are anonymized (i.e. face blurring) using bounding boxes, and acoustic sensors detect abnormal patterns without recording speech and video. The system also includes real-time crowd detection, generating heat maps to visualize occupancy levels and detect abnormal crowding. When an anomaly is detected — such as a passenger/person collapsing or a sudden increase in loud, aggressive behavior. Only in confirmed cases does the system allow temporary logged access to de-anonymized footage for authorized responders.

Data is processed locally where possible and retained only for incident resolution under strict audit controls. Visualizations such as clustered heat maps, help call indicators, and comfort-level metrics are made available on a dashboard designed for operational use. This ensures safety is enhanced ethically, preserving the dignity and privacy of all individuals while enabling fast, effective responses to near real-time threats.

Anomaly Detection in Retail

Scenario: Retail environments such as checkout areas in DIY and home improvement stores are sensitive zones where loss or fraud may occur, including situations like passing items without scanning, concealing products, or unusual customer behavior around the point of sale. Detecting such anomalies in real-time is critical for preventing shrinkage and supporting store security operations.

D2.2 Application:

In this scenario, Koçtaş applies multi-modal AI anomaly detection that combines video analytics from in-store IP cameras with transactional data from POS systems. Camera feeds are analyzed

using human pose estimation and object detection models to identify both customer movements and product handling. These are then cross-referenced with POS log entries (e.g., product name, timestamp, price) to detect inconsistencies such as an item being bagged without being scanned.

The system operates under privacy-by-design principles: video inputs are anonymized, focusing only on human motion key points and product bounding boxes, without capturing identifiable features. POS logs are parsed locally to extract only necessary metadata, ensuring no personal information is processed. Anomalies trigger alerts to security staff, while final decision-making remains with human operators to ensure fairness and accountability.

Data Handling and Security:

- All data is processed locally on GPU-enabled edge devices (e.g., NVIDIA Jetson), minimizing external transmission.
- No customer identity or personal information is collected; only product and motion data are analyzed.
- Processes comply with GDPR and KVKK regulations, ensuring anonymization and strict audit logging.
- Access to raw data is restricted, and only flagged events may be temporarily de-anonymized for authorized security staff under strict controls.

Models and Methods:

- Human Movement Analysis: YOLOv8-Pose for detecting keypoints and identifying abnormal actions.
- Product Detection: YOLOv8 object detection re-trained for Koçtaş product categories.
- Tracking: DeepSORT (with ByteTrack as alternative) for tracking customers and items over time.
- Transaction Integration: POS logs parsed from local TXT files, enabling alignment of product scans with detected visual items.
- Anomaly Detection: Inspired by STG-NF (Normalizing Flows for Human Pose Anomaly Detection, ICCV 2023), future model development will focus on detecting suspicious postures or behaviors such as concealing or bypassing items.

Current Status and Next Steps:

- Real-time human and product detection has been successfully implemented.

- Multi-object tracking and POS log integration are functional.
- Next phase focuses on pose-based anomaly detection and full integration of all modules into a live demo.

Ethical Considerations:

The system is designed to avoid bias or discriminatory outcomes, applying equal treatment across all individuals. AI alerts are always subject to human review to prevent false positives. Data is used strictly within the project scope and is not shared with third parties, ensuring ethical handling of sensitive retail surveillance data.

7.2 Port Use Case

Introduction

The SINTRA Port Use Case focuses on developing AI-driven, multi-modal surveillance systems at ports (Port of Moerdijk, DP World, and Port of Kemi) to detect criminal activities, threats, and anomalies in real-time. The project leverages diverse data sources including video feeds, acoustic sensors, multispectral imaging, AIS, GIS, logistics timetables, social media, and IoT-based smart locks.

Given the sensitivity of the collected data—especially personally identifiable information (PII) and location-based information—the SINTRA consortium commits to embedding ethical principles and regulatory compliance throughout all data processing stages. This document provides guidelines to ensure responsible, transparent, and legally compliant data use in the project.

Research at the university level creates the foundation for AI-driven multimodal development work. GDPR guidelines and other relevant regulations must be considered throughout the entire lifecycle, from laboratory work to empirical research in the port area. Privacy in AI development work will follow service-oriented architecture principles (SOA).

Applying Ethical Data Handling Guidelines to Port Security Research Operations in Kemi

Use Case 1.1: Testing the secure data flow from devices in the Jyväskylä University laboratory environment

Scenario: D2.2 The essential thing is that we can replicate and simulate the device connections to be installed in the seaport, similarly, to ensure that the system works as intended in a cybersecure way. By preparing for the field tests in this procedure, we reduce the possibility that the entire combination of devices and software to be moved to the seaport area is vulnerable to harmful

events. The procedure belongs is part of the proof-of-concept idea. Laboratory tests ensure that the tests performed at the port are compliant with the guidelines by following GDPR requirements. AI laboratory where research work will be done is a closed space, where only a limited number of persons have access. The area (space) for research is well protected, and the laboratory is separately locked with locks. The immediate environment is protected so that potential damage cannot occur. The devices cannot be used by outsiders. The analysis takes place in a technically closed environment. Security of actions is ensured in a way that raw data is not shared with external actors. The closed community formed ensures secure management of information and data sharing, handling, and processing

Application Use Case 1.1

Before conducting the empirical test in the Seaport, researchers at the University of Jyväskylä will test the sensor and camera configurations in the laboratory before installing them in the Kemi Seaport area. Security testing (e.g, pentesting) for systems and device testing of the system combination is carried out to comply with the requirements of field research in relation to the GDPR. Everything must be in order when the devices, hardware, and software are transferred to Kemi seaport. The laboratory test ensures that sensors, surveillance cameras, and video recording systems can be connected in a way that all use cases are possible to carry out by following the privacy requirements that are listed in this document. The laboratory environment is also so called command and control center where data feeds are handled. The AI laboratory plays a central role in the research work; therefore, data to be processed must be anonymized so that certain identifiers are erased. One process is to pre-process the real-world data that is processed by the stakeholders.

Use Case 1.2: Securing the seaport area by detecting abnormal events in a multimodal way

The scenario: Abnormal phenomena may occur simultaneously in the port area. In order to identify that the scenario aims to gather relevant weak signals and data from the environment by using the cameras and sensors (emergence points where events are illustrated). CCTV/IP 3d Cameras will be monitored in the container area, loading arms, and the fences. Those will be used to recognize and gather data from the containers and observe abnormal technical/natural/human-based (cyberphysical) events in the area. Audio sensors primarily gather noise and vibration from areas near the fences. Monitoring the use of locks and parameters is one task. Abloy Smart lock systems have a map platform for the operator that demonstrates where the locks are situated and can transmit signals securely about the use of locks and detected environmental changes (e.g., temperature). Procedures in the port area are carried out while protecting the privacy of individuals in the area by complying with the GDPR guidance.

The video monitoring and recording systems will use controlled anonymization of relevant footage in a specific area, for example, to locate the sea container. This process would be meticulously logged, detailing who accessed the data, when, for what purpose, and for how long.

D2.2 Application: Access to the area is restricted, and port security guidelines are followed in all activities. The Sensoan will design and set up the secure sensor network for the fences, cranes, and unloading arms. The sensors are mainly used to monitor the mechanical movements of objects. The monitoring activities carried out by designated individuals are carefully recorded, detailing who has used the data, when, for what purpose, and for how long. This means that operators viewing the live feed would see movement, but not readily identifiable faces or other irrelevant data. Video recording must be expressed in action, and irrelevant factors must be eliminated as efficiently as possible. Details of individuals are sought to be obscured or avoided in advance. For example S-VMW video recording system from Teleste strictly follows the GDPR guidelines, also through signature automation.

Use Case 1.3: Following (tracking) the movement of the sea container and changes in the container

Scenario: In this scenario, we will follow the movement and changes of the sea container in the seaport area by using multimodality that forms by using several data sources and acoustic sensors. The installed camera detects suspicious information flow from the lifted container. By comparing information with each other, it is possible to identify differences that generate an alarm. The scenario consists of following the movement of the sea container in the seaport area by comparing information provided, such as descriptions, weight, and location, but also monitoring the location and direction of the movement. It is crucial to be sure that sea containers remain intact, and the declared information corresponds to the content. Movement is possible to track by combining different data from sensors and cameras. The Teleste video recording & Management system and the Sensoan sensor system form a combined sensor network. The Abloy smart lock system gathers information such as user data from locks and environmental changes. When the given data does not match the received combination of sensor and camera data causes an alarm in the Secapp emergency system. An authorized operator, with specific credentials and a defined purpose, would initiate the start and confirm the verification process of the deviation.

D2.2 Application: All operations are performed without recording identifiable people. Access to data is ensured only for named people with limited access rights. All actions are accurately logged and noted in the detailed log data. Surveillance solutions will be used for real-time video monitoring, audio, data transmission, and recording abnormal events and especially the movements of the containers in the area, but also changes in the interior of the containers.

Use Case 1.4: Recording trajectories of loading arms and fuel pipes to prevent damage

Scenario: The safe transfer of fuel or gas between the port and the ship is of paramount importance. Potential vessel movements may cause waves that affect the insecure gas or fuel loading process. Also, vibration of the loading arms may cause unwanted events. Therefore, it is important that ship loading by using the loading arms progresses safely and the operation is stable. The aim is to monitor the smooth and secure connection of the gas or fuel pipes to the cargo ship by using video surveillance cameras and sensors. The elements will be set up to the pipeline's unloading arm so that it is possible to analyze trajectories and other limit value changes that have been set. Pipelines of liquid and gas must be connected to the vessels without vibration. So, pipeline brackets are also a monitored target. The main aim is to monitor only technical events.

D2.2 Application: Video recording of trajectories and fuel pipes can be done in such a way that people are completely excluded, because the purpose is to find out about the factors that cause changes in the trajectory of the pipe when it is connected to the ship. The camera will be set up on the unloading arm. Similarly, when examining a loading arm, it is important to aim the camera so that the object of examination is monitored during the repetition of the trajectory. The following guidelines of privacy related to GDPR will be the same as previous use cases.

Privacy guidelines in AI Service-oriented architecture

1. Encryption: Data at rest and in transit are encrypted to prevent unauthorized access and interception.

2. Access control: Access control mechanisms are in place to restrict access to services and data based on user roles and permissions. SINTRA access protocols are based on the idea of Role-Based Access Control (RBAC).

3. Anonymization: Sensitive data are anonymized or pseudonymized to protect individual privacy while still enabling the analysis and use of aggregated data. Used data is always pseudonymized or anonymized in a way that privacy aspects are considered. Anonymization should be done as early as possible.

4. Privacy by design: Privacy considerations are built into the SOA design and development process from the outset, rather than being an afterthought.

5. Auditing and monitoring: Mechanisms are in place to monitor service interactions and usage patterns so that potential privacy violations can be identified and corrective actions can be taken. We will use suitable tools and techniques to audit and monitor data in all processes, such as logging, tracing, or reporting, to record and analyze the data flows, transactions, or operations in AI SOA.

6. Standards and Guidelines: Stakeholders will use leverage standards to create a framework for privacy-aware SOA implementations. (Applicable standards are, for example, ISO 27001, 27701, and 42001 for managing AI lifecycles)

7. Privacy Policies: Stakeholders will follow clear privacy policies, such as GDPR, that describe how personal data is collected, used, and protected.

8. Trust Management: Trust models and mechanisms are established to ensure that services can be trusted to handle data responsibly and in accordance with privacy policies. All use cases to be performed are based on a holistic review of security and safety aspects.

9. Data Minimization, Retention, and Deletion: Only necessary data is collected and exchanged between stakeholders and services to minimize the risk of privacy breaches. Data is only retained as long as is necessary for the specific research purpose for which it was collected. Data is only retained as required by regulations. Data will be deleted when it is no longer needed, taking into account the requirements of the regulations, including GDPR.

10. Security policies: Security policies will be implemented to guide the processing of sensitive data and ensure compliance with data protection regulations. The main principle, for example, at the port of Kemi is to avoid sensitive data collection.

Applying Ethical Data Handling Guidelines to Port Security Operations

Contextual principles

For example, the research team will follow up on all Sintra-related requirements as mentioned in this document, such as privacy of GDPR, national and EU regulations, and other cybersecurity guidelines.

- Object monitoring and tracking in seaport areas must be strictly limited to safety, security, and operational efficiency goals (e.g., monitoring and tracking sea containers and lifting arms in Kemi Seaport). Only designated individuals have the right to view live video, analyse, or handle data.

- All devices that collect data must avoid function creep. Data collected, for example, for video surveillance and management systems or in the servers must be used for agreed purposes; data of object monitoring and tracking, or systems functions must not be used for extended purposes without new consent, agreement in accordance with legislation.

- Objects are clearly tied to their legitimate monitoring purpose, and only relevant classes should be processed within sensitive zones.

The devices used in the research, such as video management and surveillance systems, are utilized solely for research purposes. Data is not used for identity inference or profiling. The collection, storage, use, and analysis of data are systematically followed in accordance with SINTRA information security guidelines and EU regulations.

- Data collected from sensors and cameras is only for research use and may not be reused without new consent and a new agreement in accordance with legislation. Stakeholders cannot unilaterally deviate from agreed operational guidelines.

GDPR and AI act-related requirements steer all actions in the seaport area and in the laboratory of the University of Jyväskylä research.

Handling of Metadata

Metadata is handled in a way that personal data will not be stored. Metadata is handled in such a way that there is no personal data stored. Screenshots are taken using the camera system, but people are not identifiable. We use a system where audio data is stored in the data recorder's memory protected by a certificate, which is done automatically with a digital signature. Downloaded files are password-protected. Encryption is the default setting, which meets EU regulatory requirements, such as GDPR. All actions are detailed in the logs very precisely. Who takes actions, with which camera, time of action, time of action, e.g., viewing, etc.

Livestreaming and GDPR

Live streaming does not generate or produce any personal registers. The recorded video system does not generate personal register data (people in the video). Username information may contain personal data indirectly, making it subject to GDPR. Log data may contain personal data, and personal data, such as video recordings, is processed appropriately following GDPR.

Aligned with stakeholders' privacy policies (E.g., TELESTE):

Data storage in the S-VMW video recording system

An extended possibility to protect recordings against unauthorized use and misuse.

- Provides encryption capability for video recordings in motion.

- Data stored on the recorder is further protected against tampering by an integrated tamper-proof system.

- When authorized users download video data from S-VMX, XADES (XML Electronic Signature) – a digital file signature is automatically created and is available for later verification of file authenticity when uploaded with the video file.

- The uploaded file can also be password-protected, so that no unauthorized person can access the file.

- Encryption is enabled by default (factory setting) with plug-ins that can be used with SVMX systems in the EU. Decryption is only possible for authorized users and the operation is recorded in detail in the event log database.

Core Ethical Principles

The SINTRA platform operationalizes the following key ethical principles:

Privacy

- Only data necessary for threat detection is collected.
- Individual control over personal information is respected.
- Privacy-preserving methods, including differential privacy, pseudonymization, and homomorphic encryption, are integrated into data pipelines.

Consent and Purpose Limitation

- Explicit, informed consent is obtained where personal data is collected.
- Data is strictly used for defined objectives such as anomaly detection, crime prevention, and situational awareness.
- Data flow architectures document the purpose of each processing stage to prevent misuse.

Transparency

- Algorithms, AI decision logic, and anonymization methods are publicly documented for auditability.
- Port authorities, stakeholders, and relevant oversight bodies have access to information about how data is processed and analyzed.

Accountability

- Clear responsibilities for data management are assigned to consortium partners.

- Audit logs track every access, modification, and AI inference.
- Mechanisms for corrective action and redress are established.

Fairness and Non-Discrimination

- AI models are trained on diverse datasets to minimize bias.
- Bias-detection modules are implemented to prevent discriminatory surveillance, particularly in critical security zones.
- Regular evaluation ensures equitable treatment across all actors, including port workers, visitors, and vessels.

Safety and Robustness

- AI systems are designed to resist adversarial inputs, hardware failures, and environmental conditions (e.g., Arctic weather for Port of Kemi).
- Real-time anomaly detection incorporates fail-safes to prevent false positives or missed threats.

Beneficence and Justice

- All processing aims to maximize societal benefits (public safety, crime prevention) and minimize harm (invasion of privacy, unfair profiling).
- The Menlo Report framework guides ethical decision-making, emphasizing public interest in surveillance.

GDPR and Data Protection Compliance

SINTRA adheres to GDPR requirements and EU data protection principles:

Core GDPR Principles

- **Lawfulness, Fairness, Transparency:** Data collection and processing are legally justified (e.g., consent or legitimate interest) and clearly documented.
- **Purpose Limitation & Data Minimization:** Only data necessary for port threat detection and anomaly analysis is collected.
- **Accuracy & Storage Limitation:** Data is corrected or deleted when inaccurate or no longer needed; CCTV recordings have strict retention limits.

- **Integrity, Confidentiality, Accountability:** Data is encrypted, access-controlled, and logged; designated data protection officers ensure compliance.

Data Subject Rights

- **Access & Rectification:** Individuals can request copies of their data and correct inaccuracies.
- **Erasure:** Data is deleted when consent is withdrawn or retention purposes end.
- **Automated Decision Restrictions:** Human oversight is mandatory for all AI-driven surveillance actions, preventing fully automated high-stakes decisions.

Controller & Processor Obligations

- Roles of data controllers and processors are clearly assigned to consortium partners.
- Data Protection Impact Assessments (DPIAs) are conducted for high-risk sensors and AI systems.
- Technical safeguards include federated learning, anonymization, and secure cloud integration.

Core principles in data sharing of research in Kemi seaport

The Kemi Port environment and laboratory environment in Jyväskylä are simultaneously similar regarding data sharing.

At the stakeholder level, responsibilities include transforming data into standardized formats (GDPR taken into account), maintaining transmission clients and endpoints, and ensuring reliable system uptime. For instance, sensor devices installed on-site broadcast raw data streams to designated recipients (e.g., X, Y, Z), each of whom determines its own internal mechanisms for receiving and interpreting the data.

The data transfer protocols are standardized for interoperability and encryption. For example, video streams are transmitted using Real-Time Streaming Protocol (RTSP), secured via HTTPS or RTSPS (RTSP over TLS), and delivered in RTP packets. These streams can later be post-processed into MP4 or MKV formats using tools like FFmpeg, ensuring flexibility for archival or offline analysis.

The platform emphasizes multi-modal data capture, such as combining video, system logs, environmental metrics, and device telemetry, which enables comprehensive, context-aware analysis. This data is processed using machine learning and AI methods tailored for incident detection, anomaly recognition, and predictive analytics. Processed outputs are fed into an event alarm system provided by Secapp, which converts findings into structured alerts with relevant incident information. Alerts are distributed to pertinent decision-makers and port operators. This system ensures real-time responsiveness while preserving the independence of operational infrastructure

AI-Specific Ethical Guidelines

The SINTRA project's AI deployment follows EU AI Act principles and anticipates emerging regulations:

- **Risk-based Design:** AI models are classified by risk level, with stringent measures for high-risk use (e.g., UAV surveillance in populated port areas).
- **Human Oversight:** Operators supervise AI-driven anomaly detection, with manual intervention capability.
- **Transparency of AI Decisions:** Model logic and anomaly scoring are auditable and explainable.
- **Prevention of Manipulation:** AI is safeguarded against adversarial attacks that could compromise detection accuracy.

AI ACT regulation applies to:

- providers placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the Union, irrespective of whether those providers are established or located within the Union or in a third country;
- deployers of AI systems that have their place of establishment or are located within the Union;
- providers and deployers of AI systems that have their place of establishment or are located in a third country, where the output produced by the AI system is used in the Union;
- importers and distributors of AI systems;
- product manufacturers placing on the market or putting into service an AI system together with their product and under their own name or trademark;
- authorized representatives of providers, which are not established in the Union;

- affected persons who are located in the Union.

Multi-Modal Data Governance

Data Collection

- Sensor data (video, acoustic, multispectral, AIS, IoT) is collected only from designated port areas.
- Mobile UAV surveillance adds flexibility without compromising privacy: flight patterns are unpredictable but respect lawful monitoring constraints.

Data Storage and Processing

- Secure storage solutions and encrypted APIs ensure integrity and confidentiality.
- Data fusion across modalities occurs in edge devices or secure cloud platforms with access logging.

Data Sharing

- Public-private collaboration respects cross-organizational privacy standards.
- Aggregated or anonymized data is shared for research, while individual identities remain protected.

Monitoring and Continuous Improvement

- Regular audits validate compliance with ethical principles and GDPR.
- AI models undergo continuous testing to identify biases, false positives, and system vulnerabilities.

Ethical Considerations: Conclusion

The SINTRA Port Use Case operationalizes ethical and legal data management principles through every stage of data collection, processing, analysis, and sharing. By embedding privacy, transparency, accountability, fairness, and safety, SINTRA ensures that AI-driven surveillance contributes to societal security while safeguarding individual rights.

The principles of use of Artificial Intelligence-aided threat hunting require following the guidelines of the GDPR and the complementary AI ACT. Several factors must be taken into account when developing an AI-based system, such as risks associated with designing the mechanism, human

resources required for supervision operations, the transparency of AI-aided procedures and decisions, and methods for preventing manipulation. That means it is not enough to exploit AI in some way; it is more necessary to design a threat hunting approach for the ecosystem in a cyber-secure manner that falls under the overall security umbrella. To consider all regulatory requirements and suitable standards forms the governance roof for a successful enterprise concept.

These ethical guidelines provide a robust framework for consortium partners, port authorities, and system integrators, ensuring responsible and compliant deployment of advanced multi-modal threat detection systems.

7.3 Construction Site Use Case

Introduction

The SINTRA Construction Site Use Case focuses on monitoring industrial construction sites to ensure the **security of critical infrastructure** and **safety of workers and visitors**. The approach leverages multi-modal sensors including ANPR cameras, audio sensors, live video streaming from drones, and integrated proprietary monitoring systems.

The primary objectives are:

- Integrating heterogeneous data streams for a **common situational picture**.
- Detecting **anomalous or unsafe behaviors** in real-time.
- Providing **actionable decision support** to security personnel.

Given the sensitive nature of surveillance data and the need to respect privacy, the SINTRA consortium embeds ethical principles and GDPR compliance into all aspects of data collection, processing, analysis, and sharing.

Core Ethical Principles

Privacy

- Collect only necessary data for safety and security monitoring.
- Respect the privacy rights of workers, visitors, and contractors.
- Integrate privacy-preserving techniques such as **pseudonymization, differential privacy, and anonymization** in multi-modal data fusion pipelines.

Consent and Purpose Limitation

- Obtain explicit, informed consent where personal data is processed.
- Limit data use strictly to defined purposes: anomaly detection, threat prevention, and proactive monitoring.
- Document all data flows and processing steps to enforce purpose limitation.

Transparency

- Maintain clarity about **how, why, and by whom** data is collected or analyzed.
- Publish algorithms, AI decision logic, and anonymization techniques to relevant stakeholders.
- Ensure construction site personnel and visitors are informed about surveillance measures.

Accountability

- Assign clear responsibilities for data management among consortium partners.
- Maintain comprehensive logs for all data access, model inferences, and system alerts.
- Establish mechanisms for audit, review, and redress in case of ethical or legal violations.

Fairness and Non-Discrimination

- Design AI models to avoid bias toward specific groups of workers, visitors, or contractors.
- Train models on diverse datasets reflecting realistic site conditions.
- Evaluate algorithms for disparate impacts, particularly in security-critical zones.

Safety and Robustness

- Ensure AI models and monitoring systems are resilient to errors, adversarial attacks, and technical failures.
- Implement fail-safe mechanisms for real-time anomaly detection to prevent false alarms or missed events.

Beneficence and Justice

- Prioritize societal benefits such as workplace safety, infrastructure protection, and accident prevention.
- Minimize harm, including intrusion into personal privacy and undue surveillance.

- Apply guidance from the Menlo Report, emphasizing public interest and ethical oversight.

GDPR and Data Protection Compliance

The SINTRA C-Site Use Case operates fully under the **EU General Data Protection Regulation (GDPR)** framework:

GDPR Core Principles

- **Lawfulness, Fairness, Transparency:** Data collection and processing are legally justified (e.g., consent, legitimate interest) and documented.
- **Purpose Limitation & Data Minimization:** Only data required for anomaly detection and safety monitoring is collected.
- **Accuracy & Storage Limitation:** Data correction mechanisms and automatic deletion policies are enforced; CCTV and drone recordings have strict retention limits.
- **Integrity, Confidentiality, Accountability:** Data is encrypted, access-controlled, and monitored via audit logs; data protection officers verify compliance.

Data Subject Rights

- **Access & Rectification:** Workers and visitors can request copies of their data and correct errors.
- **Erasure ("Right to be Forgotten"):** Automatically applied when data is no longer needed or consent is withdrawn.
- **Automated Decision Restrictions:** Human oversight is mandatory for AI-driven surveillance decisions, preventing fully automated high-risk actions.

Controller & Processor Obligations

- Roles of controllers and processors are clearly assigned among consortium partners (e.g., drone operators, sensor providers, AI developers).
- High-risk sensing systems are subject to **Data Protection Impact Assessments (DPIAs)**.
- Technical safeguards include **federated learning, anonymization, and secure cloud integration**.

AI and Multi-Modal Data Governance

Risk-Based AI Design

- AI systems are classified based on potential risk to privacy and safety.
- High-risk components (e.g., drone video analysis, real-time anomaly detection) implement **enhanced oversight and validation**.

Human Oversight

- Security personnel maintain decision-making authority over AI alerts.
- AI recommendations are provided as **supportive guidance**, not automatic enforcement.

Multi-Modal Data Fusion

- Data from cameras, drones, audio sensors, and proprietary systems is fused securely and privacy-preservingly.
- Only anonymized or pseudonymized outputs are used for further analysis whenever possible.
- Access to raw identifiable data is limited to authorized personnel.

Continuous Monitoring and Evaluation

- Regular audits check for algorithmic bias, system vulnerabilities, and compliance with ethical standards.
- AI models are periodically retrained with updated and diverse datasets to maintain fairness and robustness.

Proactive and Ethical Decision Support

- AI-driven alerts are designed to provide actionable recommendations while avoiding unnecessary intrusion.
- Examples: dispatching drones to verify suspicious activities, triggering localized alerts to on-site personnel, or activating additional sensors.
- Decision support ensures interventions are proportionate, justified, and transparent, preserving both safety and privacy.

8 TRUSTWORTHY AND EXPLAINABLE AI

SINTRA's mission is to deploy AI systems that are not only technically robust but also ethically aligned and trusted by users, operators, and the public. This requires a holistic approach spanning model transparency, explainability through immersive interfaces, resilience to attacks, and meaningful human oversight.

8.1 AI Model Transparency and Fairness

At the core of trustworthy AI is transparency—a commitment to making AI models understandable, auditable, and justifiable. SINTRA ensures transparency through comprehensive model documentation detailing architecture decisions, training data provenance, performance metrics, demographic fairness tests, and known failure modes. Such detailed documentation aligns with emerging best practices in AI governance Prior to deployment, each model undergoes rigorous fairness evaluations, where test datasets reflect demographic and behavioral diversity—covering gender, age, ethnicity, mobility aids, and cultural norms. Bias-detection tools assess whether certain groups are disproportionately flagged or misclassified. When necessary, retraining or adjustment of decision thresholds ensure that fairness metrics (e.g., equal false-positive rates across groups) are met. These processes fulfill ethical standards and help comply with legal frameworks such as the EU's emerging AI Act, which mandates fairness in AI systems.

8.2 Explainability and Augmented Reality Interfaces

Explainability goes beyond policy—we embed intelligibility directly into SINTRA's interfaces, making AI decisions transparent to end-users. Operators receive structured explanations (via SHAP, LIME, or rule-based descriptors) that clarify *why* an event was flagged, for instance, showing that a person's rapid movement, trajectory change, and audio spike triggered an alert.

Critically, SINTRA augments explainability with augmented reality (AR) interfaces. Using frameworks like XAIR, SINTRA overlays contextual AI insights—highlighting suspect activities, displaying confidence levels, and visualizing heatmaps—directly on live camera feeds or physical environments. AR empowers operators to intuitively engage with AI reasoning, reducing cognitive load and improving situational awareness. For example, an AR headset in an airport control room may overlay bounding boxes and alert rationales on a live hall feed, allowing faster and more informed human intervention.

8.3 Adversarial Robustness and Safety

Trustworthy AI must be **robust**—resilient to noise, adversarial manipulation, and real-world variability. SINTRA leverages several defenses:

- **Adversarial training**, introducing perturbed samples during training, reduces responsiveness to engineered inputs

- **Anomaly and distribution-shift detection** methods identify when inputs drift outside training norms, prompting reviews or fallbacks
- **Human-in-the-loop mechanisms** intercept suspicious edge-cases or adversarial alerts, allowing expert judgment to validate or dismiss AI outputs
- **Redundant sensor inputs** (audio + video + thermal) help detect spoofing attempts, for example, ensuring a spoofed visual signal without matching thermal evidence triggers an alert rather than blind action.

Rigorous security testing—such as penetration assessments and adversarial scenario simulations conducted before live deployment. These measures align with national AI safety protocols and are essential in high-stakes contexts like drone surveillance or airport security.

8.4 Human-in-the-loop Considerations

Human oversight is a cornerstone of ethical AI in SINTRA. No automated alert leads directly to action—every decision remains subject to human validation. Analysts are trained to interpret model outputs, question recommendations, and correct mistakes. This checks-and-balances system mitigates risks of over-reliance and error accumulation.

Examples of HITL in operation include:

- Pre-deployment **model calibration** sessions where operators review system behavior against real scenarios.
- **Interactive annotation** during model uncertainty, where users label ambiguous events to refine training data, supported by active learning frameworks
- **Standard operating procedures** require secondary approvals for critical decisions—such as initiating law-enforcement response.

This structured engagement preserves **human agency** and embeds ethical accountability within SINTRA's operational fabric—ensuring that technology supports human judgment, not replaces it.

By combining transparent documentation, immersive explainable interfaces, security-conscious model design, and human oversight, SINTRA delivers a trustable AI ecosystem. This ecosystem balances cutting-edge performance with ethical safeguards, ensuring that surveillance benefits are realized responsibly enhancing public safety without compromising rights or dignity.

9 CHALLENGES AND MITIGATION STRATEGIES

Across SINTRA's multi-domain deployments—airports, ports, construction and railway sites—managing the intersection of AI, sensor diversity, and privacy creates complex technical and ethical challenges. This section explores four core risk areas and outlines SINTRA's tailored mitigation strategies.

9.1 Ethical Risks in Multimodal Sensor Fusion

Challenge: Combining audio, video, thermal, lidar, RFID, and radar data enhances situational awareness but risks aggregating sensitive personal information or creating novel inferences (e.g., identifying individuals from gait or linking behavior patterns to locations). Multimodal fusion can **amplify bias**—where biases in one modality (e.g., demographic skews in video) propagate into fused outputs

Mitigation: SINTRA employs a multi-stage fusion governance:

- **Modality partitioning**: Each sensor stream is restricted to a specific analytical task; e.g., RFID is used for access control only, not identity profiling.
- **Bias audits on fusion models**, including both intrinsic (in-model) and extrinsic (outcome) bias metrics
- **Explainable fusion pipelines**: Using attention maps and fusion transparency layers to trace which sensor contributed most to decisions.
- **Privacy filters** at each fusion stage ensure sensitive data is blurred, obfuscated, or deleted before deeper fusion.
- **Ethics Review Board checks** new fusion methods, informed by a survey on multimodal fairness

9.2 Balancing Security and Privacy

Challenge: Achieving real-time security monitoring while preserving individual privacy creates tension—especially in densely populated public settings. Over-surveillance risks chilling effects on public behavior; under-surveillance opens safety gaps

Mitigation:

- **Proportional distancing**: Only high-risk zones (e.g., customs checkpoints) receive full sensor coverage. Other areas use low-resolution or anonymized sensors.

- **Edge-first anonymization** before data transmission preserves privacy while enabling local alert generation.
- **Consent-friendly opt-outs** in retail or non-critical zones, managed via visible signage and accessible interfaces.
- **The Five Safes framework** ensures shared outputs are appropriately de-identified and purpose-bound (as detailed in Section 4.3).
- **Continuous DPIAs** update risk assessments with evolving sensor configurations.

9.3 Managing Bias in Data and Algorithms

Challenge: Bias can permeate data collection, model training, and deployment. Historical imbalances—such as Joy Buolamwini's findings on facial recognition bias—can reproduce harm in SINTRA.

Mitigation:

- **Diverse training datasets** reflect different ethnicities, ages, and mobility aids.
- **Algorithmic fairness tools**, such as IBM Fairness 360, used in preprocessing, in-processing, and post-processing stages.
- **Human-in-the-loop auditing** to review uncertain or borderline cases, using interactive bias-adjustment tools like D-BIAS
- **Scenario-based testing**, including synthetic scenarios for underrepresented groups (e.g., children, people with wheelchair), aligned with XR ethical risk studies.
- **Ongoing bias monitoring**, tracking metrics like equal false-positive rates across demographic groups.

9.4 Scalability, Performance, and Network Constraints

Challenge: SINTRA's real-time analytics demands—high-resolution video, distributed drone data, AI computation—strain networks, edge devices, and cloud infrastructure, particularly in remote or congested environments.

Mitigation:

- **Edge computing architecture** minimizes data transmission and latency; only alerts and metadata are uplinked.

- **Adaptive processing strategies**: The platform adjusts sensor rates dynamically based on congestion, utility, and threat levels.
- **Federated learning and encrypted model aggregation** reduce network, and storage demands while preserving data sovereignty (see Section 5.3).
- **Robust network design**, including mesh and fallback paths, ensures resilience in construction or railway sites.
- **Performance-driven sensor fusion**, using low-pass filters and early fusion only when necessary to manage workload
- **Continuous system stress-testing** with adversarial load injections and network failure simulations to verify reliability.

Table: Summary of Strategic Responses

Risk Area	Mitigation Strategy
Multimodal bias & privacy	Modality partitioning, bias audits, explainable fusion
Privacy vs. Security	Anonymization, consent systems, DPIAs
Systemic algorithmic bias	Fairness toolkits, human-in-loop auditing
System performance & scale	Edge processing, federated methods, load testing

10 EVALUATION, MONITORING, AND COMPLIANCE

Robust evaluation, continuous monitoring, and stringent compliance mechanisms are essential to ensure that the SINTRA platform remains trustworthy, legally sound, and ethically aligned. This section details the framework, metrics, and audit mechanisms SINTRA uses to maintain accountability and adapt to evolving standards.

10.1 Periodic Assessment Framework

SINTRA implements a recurring **Ethical & Technical Assessment Cycle**, typically executed quarterly or semi-annually, to validate ongoing compliance with GDPR, the upcoming EU AI Act, and ethical standards such as those from the Menlo Report. Each cycle includes:

- **Governance Review** – The Data Protection Officer and Ethics Board review DPIA updates, consent notices, and configurations against approved ethical policies and emerging regulations.
- **Model & System Audit** – Periodic assessments of AI models, reviewing training data, performance drift, and bias incidents using frameworks like the Responsible AI Question Bank to align procedures with EU requirements.
- **Performance & Traceability Audit** – Inspection of system logs to verify access controls, data retention events, and ethical trigger queries using tools and checklists recommended by the EDPB.
- **Risk Reassessment** – Stress-testing systems in evolving operational scenarios (e.g., dense airport crowds or cross-border traffic) to identify latent risks and recalibrate sensors, policies, and thresholds.
- **Reporting & Feedback Loop** – Findings are documented in structured reports and actioned in follow-up cycles using formal **AI Governance Platforms**, such as those described by WitnessAI, enabling measurable improvements.

10.2 Metrics for Ethical Compliance

To operationalize ethics, SINTRA tracks quantitative and qualitative metrics across critical dimensions:

Domain	Key Metrics
Fairness & Bias	Equalized false-positive rates across demographics; incidence of false alerts
Explainability	Percentage of AI alerts accompanied by clear model rationale (SHAP/LIME)
Transparency	Rate of processed access requests; audit success scores
Privacy Protection	Volume of anonymized vs. cleartext data; differential privacy noise levels
Resilience & Safety	Adversarial attack detection rate; AI failure recovery time
Governance Maturity	% of models certified through ethical review; completion rate of training courses

These KPIs are monitored via SINTRA's central dashboard and updated each cycle to reflect changes in legal standards or risk posture. The approach aligns with best practices in AI governance metrics, ensuring a balance of compliance, performance, and ethical integrity.

10.3 Auditing and Traceability Mechanisms

Effective oversight relies on strong auditing and traceability systems:

Immutable Audit Logging

Every data access, transformation, AI inference, and sharing event is recorded in tamper-evident logs. These logs include timestamps, user IDs, roles, purposes, and metadata changes, enabling robust forensic traceability. Encryption and cryptographic signing provide additional defense against tampering.

AI-Specific Auditing Procedures

Following the **EDPB AI Auditing Checklist**, SINTRA performs periodic audits including:

- Model documentation
- System maps of data flow
- Bias & adversarial audit modules
- Human oversight thresholds and bistable fallback verification

Adherence to PIMS standards (e.g., ISO 27701) ensures audit readiness and complete accountability.

Third-Party and Internal Audits

Annual external audits by certified third parties validate compliance with GDPR, AI Act provisions, and ethical standards. Internally, the DPO and Ethics Board conduct monthly reviews of access patterns, DPIA updates, and incident handling summaries. Findings from both audits feed into a continuous governance improvement plan, as advocated in ethics-based audit literature.

Compliance Automation & Dashboard

Governance platforms like WitnessAI enable SINTRA to automate compliance tracking through:

- Model registries
- Versioned policy enforcement
- Real-time KPI dashboards
- Notifications for non-conformance

This automation accelerates incident response, reduces manual overhead, and strengthens stakeholder confidence.

SINTRA's investment in structured evaluation, ethical KPI tracking, and layered audit mechanisms ensures constant alignment with legal and ethical demands. These integrated systems—from immutable logs to federated audits—make SINTRA a resilient, transparent, and future-proof platform for deploying AI in sensitive surveillance and public safety domains.

11 RECOMMENDATIONS AND GUIDELINES

To ensure the SINTRA platform's successful, ethical, and effective deployment, this section presents tailored recommendations for developers and integrators, end-users and operators, and policy makers and regulators. These guidelines are based on SINTRA's goals, use case insights, and state-of-the-art advancements—and enriched with current best practices and standards.

11.1 For Developers and Integrators

1. **Embed Privacy by Design**
   - Incorporate anonymization, pseudonymization, and data minimization techniques at sensor edge during development, as required by GDPR and privacy design frameworks.
   - Use open-source libraries like OpenDP, DiffPrivLib, and SmartNoise to enforce differential privacy in analytics.
2. **Ensure Transparency and Explainability**
   - Leverage SHAP or LIME to generate model explanations that are meaningful to users.
   - Maintain model cards and documentation outlining training data demographics, known biases, version history, and performance metrics—aligned with the "Datasheets for Datasets" principle.
3. **Implement Robust Access and Governance Controls**
   - Use policy-as-code to enforce RBAC and FGAC, including contextual checks on time, location, and user credentials.
   - Implement "Five Safes" principles to determine appropriate data access based on user type, purpose, and data sensitivity.
4. **Validate Security and Adversarial Defenses**
   - Incorporate adversarial training and continuous security testing in deployment pipelines to counter spoofing and tampering.
   - Ensure sensor fusion uses data from multiple modalities; require corroboration before sending alerts during uncertain conditions.
5. **Support Federated Learning and Encryption**
   - Build models compatible with federated learning and homomorphic encryption to maintain data sovereignty and enhance scalability.

- Partition data across administrative and geographical zones while enabling shared intelligence.

6. **Document and Support Auditing**
   - Maintain traceable data lineage and immutable logs for all processes.
   - Support DPIAs, third-party audits, and internal reviews, especially for high-risk modules (e.g., behavior analytics, facial recognition).

11.2 For End-users and Operators

1. **Commit to Human-in-the-loop Practices**
   - Always review AI-generated alerts before acting.
   - Use explainability tools to understand why alerts occur and respond with context-aware judgment and discretion.
2. **Adhere to Ethical Use Standards**
   - Respect data subject rights, including responding promptly to access or deletion requests.
   - Employ privacy-preserving defaults, for example, blurring non-essential data.
3. **Participate in Ongoing Training Programs**
   - Engage regularly in training on AI ethics, privacy, and security.
   - Practice simulation scenarios to enhance awareness of potential biases, adversarial attempts, or drifted models.
4. **Promote Transparency with Affected Individuals**
   - Display informative signage in monitored areas.
   - Be prepared to explain monitoring purposes when approached, demonstrating transparency in operation.
5. **Monitor, Report, and Escalate Incidents**
   - Track false-positive rates, data leaks, and access anomalies.
   - Follow established protocols to escalate incidents and contribute to post-event analysis.

11.3 For Policy Makers and Regulators

1. **Define Clear Norms for Ethical Data Collection**
    - Facilitate the adoption of standards for differential privacy, federated learning, and context-aware access limits.
    - Provide guidelines (e.g., sector-agnostic DPIA templates) tailored to multi-modal AI systems.
2. **Promote Compliance Frameworks and Certifications**
    - Encourage certifications akin to ISO 27701 for privacy-aware AI systems.
    - Support regulatory sandboxes for testing bias audits, adversarial safety, and privacy in real-world environments.
3. **Enable Transparent Oversight**
    - Require standardized audit trails and incident reporting for public sector AI deployments.
    - Support interoperable metadata formats for easier cross-jurisdictional audits.
4. **Ensure Public Accountability**
    - Mandate signage and public access information about surveillance systems, their purpose, and avenues for redress.
    - Integrate third-party ethics review boards or ombudsman mechanisms within high-stakes AI systems.
5. **Foster International Harmonization**
    - Aligning GDPR, AI Act, Data Governance Act, and sector-specific regulations (e.g., EHDS) to support cross-border and multi-industry deployments.
    - Collaborate globally to set unified norms for data portability, privacy, and algorithmic fairness—especially for transnational infrastructure.

These recommendations ensure that innovation remains safe, fair, and respectful across the lifecycle of AI-enabled safety systems. By following these guidelines, stakeholders at all levels—from developers to regulators—can build systems that not only meet current ethical and legal requirements but also adapt to evolving norms and public expectations.