Comments on ETSI TS 104 158-2 v0.0.4

Securing Artificial Intelligence TC (SAI);
AI Incident Reporting
Part 2: AI Common Incident Expression (AICIE) Common Container

Heather Frase, PhD

CEO, Veraitech

February 2026

Please cite as:
Frase, H. (2025). Comments on ETSI TS 104 158-2 v0.0.4, Securing Artificial Intelligence TC (SAI); AI Incident Reporting Part 2: AI Common Incident Expression (AICIE) Common Container. February 2025. Retrieved from https://veraitechus.com/aicie-comments/

Table of Contents

1. Consistency of Measurement Within Data Fields

A recurring issue across several enumerated fields in the AICIE Common Container is what can be described as a "mixing of units" problem. In physics and engineering, mixing units means combining incompatible measures in a single calculation, such as adding meters to seconds. The result is not wrong in the sense of being a bad number; it is wrong in the sense of being meaningless. The same principle applies to enumerated data fields.

When a single field asks an incident reporter to select from a list, every item in that list should measure the same kind of thing. If some items describe what kind of harm occurred (physical, psychological), others describe where the harm occurred (critical infrastructure), and still others describe which legal framework was violated (human or fundamental rights), the field is mixing units. A reporter can select values, but the resulting data conflates distinct analytical dimensions, limiting its utility for trend analysis, cross-incident comparison, and policy development.

This problem was identified in four fields: AICIECCsystemRelationship (4.1.3), which mixed the AI system's causal role with human behavioral factors; AICIECCharmType (4.1.11), which mixed harm types with harm targets and normative frameworks; AICIECCharmConsequence (4.1.12), which mixed actual consequences with response actions and measurement categories; and to a lesser degree AICIECCpartiesAffected (4.1.14), which mixed demographic, role-based, and entity-type categories. In each case, the proposed revisions aim to ensure that every enumeration within a field answers the same question, measured in the same units.

2. AICIECCsystemRelationship (4.1.3) and AICIECCintentionality (4.1.13)

2.1 Identified Issues

2.1.1 Issues with AICIECCsystemRelationship (4.1.3)

- **Conflation of AI and human causal factors.** The field description states it captures "how the AI system(s) are related to the incident," but several enumerations ("human error," "overreliance") describe human behavior rather than the AI system’s causal role. The list mixes two different kinds of information in a single field.
- **Need and ‘unknown’ field:** Depending upon who is filing the incident, this information may not be known.
- **"Overreliance and intentional misuse" bundles opposites.** Overreliance implies misplaced trust, which is typically unintentional. Intentional misuse is deliberate abuse. These have fundamentally different implications for response, accountability, and prevention. They should be separate enumerations.
- **"Failure to act" has an ambiguous subject.** It is unclear whether the AI system failed to act (e.g., a safety system not triggering an alert) or a human failed to act on AI output. The phrasing should clarify this.
- **"Human error" does not describe the AI’s relationship.** This tells us a human made a mistake, but not what the AI system was doing. If the intent is to capture "AI functioned correctly but was used improperly," the phrasing should say that.

- **"Legal obligation omission" is a compliance framing, not a causal relationship.** This describes a regulatory status (non-compliance), not how the AI system causally relates to the incident. It is also jurisdiction dependent. *Organizational compliance posture may be better captured in AICIECCprinciplesAffected (4.1.16). This item should be dropped from AICIECCsystemRelationship.*
- **Missing categories.** The current list does not include categories for: the AI system being the target of an attack (relevant for security incidents bridging Parts 2 and 3); the AI system amplifying harm from other sources; the AI system operating as designed but the design itself was inadequate.
- **Lack of clarifying definitions.** The distinctions between enumerations are not obvious, particularly for non-expert reporters. Each enumeration should have a brief definition to support consistent use across the diverse reporting community.

2.1.2 Issues with AICIECCintentionality (4.1.13)

This field has several issues that interact with AICIECCsystemRelationship.

- **"Unintended" and "wrongful" are conflated.** These represent fundamentally different situations. "Unintended" suggests accidental harm; "wrongful" suggests deliberate misuse. Combining them in one concept obscures a critical distinction for accountability, regulatory response, and prevention.
- **Overlap with AICIECCsystemRelationship.** If AICIECCsystemRelationship includes enumerations like "overreliance" and "intentional misuse," and intentionality also captures "unintended or wrongful," the same information is partially captured in two places without clear delineation.

2.2 Recommendation

I recommend revising both AICIECCsystemRelationship (4.1.3) and AICIECCintentionality (4.1.13) as a coordinated change. The simplest to implement approach (Option A) retains a single AICIECCsystemRelationship field with revised enumerations and revises intentionality to a clean enumeration. An alternative two-field approach (Option B) is also provided. Both options address most of my issues with the two data fields.

I recommend Option B because it:

- Cleanly separates causal role from causal mechanism, enabling richer analytical queries
- Mirrors how traditional safety investigation frameworks (NTSB, ICAO) structure causal analysis
- Avoids overloading a single field with too many enumerations

However, Option A may be preferable if the committee prioritizes:

- Maintaining alignment with the OECD framework structure
- Minimizes reporter burden for the diverse stakeholder community (developers, regulators, consumers, general public).
- Capturing the essential causal story in one place, which is sufficient for trend analysis, regulatory triage, and cross-incident comparison.

Regardless of the chosen option, explanations of the list items need to be included somewhere in the file

2.3. Recommended Language: Option A (Single Field with Revised Intentionality)

2.3.1 Revised 4.1.3 AIClECCsystemRelationship

This required value describes how the AI system(s) were involved in the incident, including as causal agent, contributing factor, or target of malicious action, using one or more of the following non-case-sensitive enumerations, including an unconstrained free-form option.

Enumeration	Description
primary cause	The AI system's behavior was the principal cause of the incident. The incident would not have occurred without the AI system's action or output.
contributing cause	The AI system's behavior was one of multiple causes of the incident. The AI system played a causal role, but other factors were also necessary for the incident to occur.
amplifying factor	The AI system scaled, accelerated, or extended harm that originated from other sources. The AI system did not initiate the harm but made it worse or more widespread.
safeguard failure	The AI system failed to detect, prevent, alert, or mitigate harm that it was designed or expected to address. This includes AI systems deployed for safety, monitoring, or protective functions that did not perform as intended.
design or specification inadequacy	The AI system operated as designed, but the design, training, specification, or validated operating conditions were inadequate for the deployment context. The system was not malfunctioning; the harm was inherent in how the system was built or specified.
correct operation with improper use	The AI system functioned correctly, but was used improperly by humans. This includes overreliance on AI outputs, use outside intended scope, misapplication of AI recommendations, or unintentional human errors in conjunction with the AI system.
tool for deliberate harm	The AI system was deliberately used by a person or organization to cause harm. The harmful outcome was the intended result of the user's actions.
target of attack	The AI system was compromised, manipulated, or attacked, resulting in the incident. This includes adversarial inputs, data poisoning, model extraction, prompt injection, or other attacks on the AI system. *If selected, reporters should also complete the security container specified in Part 3.*
incidental	The AI system was present or involved in the context of the incident but was not causally relevant to the harm. The incident would have occurred regardless of the AI system's presence or behavior.
unknown	The causal relationship between the AI system and the incident cannot be determined from available information.
… (free-form content)	Other causal relationships not captured by the above enumerations. Reporters should provide a brief description.

2.3.2 Revised 4.1.13 AIClECCintentionality

The current hybrid format (checkbox flag + free-form text describing "how an AI incident was linked in an unintended or wrongful way") should be replaced with a clean enumeration. This resolves the conflation of "unintended" and "wrongful," eliminates overlap with the revised systemRelationship, and produces structured data suitable for analysis.

Recommended language:

This optional value describes the intentionality associated with the AI incident using one of the following non-case-sensitive enumerations, including an unconstrained free-form option.

Enumeration	Description
unintentional	The harm was an unintended consequence. No party deliberately sought the harmful outcome.
intentional	The harm was a deliberately sought outcome by one or more parties.
mixed	The incident involved both intentional and unintentional elements. For example, deliberate misuse that also produced unintended secondary harms.
unknown	Intentionality cannot be determined from available information.
… (free-form content)	Other characterization of intentionality not captured by the above enumerations.

2.3.3 How the Two Fields Work Together

With these revisions, each field has a distinct purpose:

- **AIClECCsystemRelationship** answers: "What was the AI system's causal role in the incident?"
- **AIClECCintentionality** answers: "Was the harmful outcome deliberate or accidental?"

These are complementary, not overlapping. For example, an incident where an AI system was used as a "tool for deliberate harm" would have intentionality of "intentional." An incident involving "correct operation with improper use" could be "unintentional" (overreliance) or "intentional" (misuse). An incident involving "design or specification inadequacy" would typically be "unintentional." The combination of the two fields provides a richer picture than either alone.

2.4 Alternative: Option B (Two Causal Fields with Revised Intentionality)

This option has greater analytical precision but a modest increase in reporter burden. In this option, the systemRelationship field is split into two fields: one for the AI system's causal role and one for the causal mechanism. This approach separates "what was the AI's causal relationship" from "through what mechanism," mirroring how traditional safety investigation frameworks structure their analysis. *The revised intentionality (Section 2.3.2) applies to both options.*

2.4.1 Proposed 4.1.3 AIClECCaiCausalRole

This required value describes the causal relationship between the AI system and the incident using one or more of the following non-case-sensitive enumerations, including an unconstrained free-form option.

Enumeration	Description
primary cause	The AI system's behavior was the principal cause of the incident. The incident would not have occurred without the AI system's action or output.
contributing cause	The AI system's behavior was one of multiple causes of the incident. The AI system played a causal role, but other factors were also necessary for the incident to occur.
amplifying factor	The AI system scaled, accelerated, or extended harm from other sources. It did not initiate the harm but made it worse or more widespread
safeguard failure	The AI system failed to detect, prevent, alert, or mitigate harm that it was designed or expected to address.
target of attack	The AI system was compromised, manipulated, or attacked, resulting in the incident. If selected, reporters should also complete the security container specified in Part 3.
incidental	The AI system was present but was not causally relevant to the harm.
unknown	The causal relationship cannot be determined from available information.
… (free-form content)	Other causal role not captured by the above enumerations.

2.4.2 Proposed 4.1.3a AICIECCcausalMechanism

This optional value describes the mechanism(s) by which the AI system contributed to the incident using one or more of the following non-case-sensitive enumerations, including an unconstrained free-form option. ***This field should be completed when AICIECCaiCausalRole indicates a causal relationship other than "incidental" or "unknown."***

Enumeration	Description
unexpected system behavior	The AI system behaved unexpectedly relative to its design. This includes malfunctions, software defects, performance degradation, or responses to conditions outside its training distribution
design or specification inadequacy	The AI system operated as designed, but the design, training, specification, or validated operating conditions were inadequate for deployment. The system was not malfunctioning; harm was inherent to how it was built.
deployment context mismatch	The AI system was used outside its intended, tested, or validated operating conditions. This includes environments, user populations, or use cases differing from those for which it was designed.
improper use of correctly functioning system	The AI system functioned correctly but was used improperly. This includes overreliance on outputs, misapplication of recommendations, inadequate human judgment, or unintentional user errors.
deliberate harmful use	The AI system was deliberately used by a person or organization to cause harm. The harmful outcome was the intended result of the user's actions.
system compromise	The AI system was attacked, manipulated, or had its integrity compromised (e.g., adversarial inputs, data poisoning, or unauthorized access). *If selected, reporters should also complete the security container specified in Part 3.*
unknown	The mechanism cannot be determined from available information.
… (free-form content)	Other mechanism not captured by the above enumerations. Reporters should provide a brief description.

2.4.3 How the Three Fields Work Together (Option B)

Under Option B, the three fields each have a distinct purpose:

- **AIClECCaiCausalRole** answers: "What was the AI system's causal role in the incident?"
- **AIClECCcausalMechanism** answers: "Through what mechanism did the AI system contribute to the incident?"
- **AIClECCintentionality** answers: "Was the harmful outcome deliberate or accidental?"

This enables richer analytical queries. For example: "Show all incidents where AI was the primary cause due to design inadequacy" or "Of all safeguard failures, which were due to unexpected system behavior vs. deployment context mismatch."

2.6 Mapping from Current to Proposed Enumerations

For reference, the following table shows how each current enumeration in AIClECCsystemRelationship maps to the proposed revision.

Current Enumeration	Option A Mapping	Rationale
direct cause	primary cause	Renamed for clarity; "primary" better conveys degree of causation
contributing factor	contributing cause	Minor rename for consistency with "primary cause"
failure to act	safeguard failure	Clarifies that the AI system (not a human) failed to act, and specifies the protective function context
overreliance and intentional misuse	Split into: correct operation with improper use + tool for deliberate harm	Unbundles opposites; overreliance maps to "correct operation with improper use," intentional misuse maps to "tool for deliberate harm"
human error	correct operation with improper use	Reframed to describe the AI's role (functioning correctly) rather than the human's error
legal obligation omission	Dropped	Compliance framing, not a causal relationship; organizational compliance may be captured in AIClECCprinciplesAffected (4.1.16)
(not present)	amplifying factor	New: captures AI systems that scale or accelerate harm from other sources
(not present)	design or specification inadequacy	New: captures AI systems that operated as designed where the design was inadequate
(not present)	target of attack	New: captures security incidents; creates handoff to Part 3
(not present)	incidental	New: captures cases where AI was present but not causally relevant
(not present)	unknown	New: essential for diverse reporter base where causal role may not be determinable

3. AICIECCharmType (4.1.11)

3.1 Identified Issues

3.1.1 Mixing of Units

The enumeration list mixes three fundamentally different kinds of categories in a single field. This is analogous to a mixing of units problem: the values are not measuring the same thing.

Harm types (nature of the damage): physical, psychological, reputational, economic/property, environmental. These describe what kind of harm occurred.

Targets or contexts (where/to whom): critical infrastructure, public interest. These describe where the harm occurred or who it affected, not what kind of harm it was. An attack on critical infrastructure could cause physical harm, economic harm, environmental harm, or all three. "Critical infrastructure" is not a harm type; it is a target.

Normative/legal frameworks (which rights were violated): human or fundamental rights. This is a legal classification of the harm's significance, not a description of the harm's nature. Physical harm, psychological harm, and discrimination can all constitute human rights violations. This category crosscuts the harm types rather than sitting alongside them.

3.1.2 Duplication with Other Fields

"Critical infrastructure" is already captured in AICIECCcriticalInfrastructure (4.1.19), which provides a detailed enumeration of specific infrastructure sectors (energy, healthcare, transportation, etc.). Including "critical infrastructure" as a harm type is redundant with this dedicated field and conflates a deployment context with a type of harm.

"Human or fundamental rights" overlaps with AICIECChumanRightsImpact (4.1.15), which already provides a dedicated checkbox and free-text field for adverse impacts on human rights. However, retaining it in AICIECCharmType with guidance to complete 4.1.15 provides a useful cross-reference and ensures the rights dimension is not overlooked.

3.1.3 Vagueness of "Public Interest"

"Public interest" is undefined and could encompass democratic processes, public trust in institutions, information ecosystem integrity, social cohesion, or any collective societal concern. Different reporters in different jurisdictions will interpret this differently, reducing consistency and analytical value. The OECD likely kept this term generic to accommodate jurisdictions without democratic governance structures, but the resulting vagueness undermines its utility for trend analysis.

3.1.4 Missing Harm Types

Several well-established harm types relevant to AI incidents are not represented in the current enumeration:

- **privacy or personal data** (privacy violations, data exposure, surveillance, unauthorized data collection). A common categorie of AI incidents has no clean mapping in the current list.
- **Discrimination or bias** (differential treatment based on protected characteristics). The most frequently reported AI-specific harm type, currently subsumed under "human or

fundamental rights," is invisible for trend analysis. This is among the most common harm types seen in AIID.

- **Autonomy or agency harm** (manipulation, deception, coercion by AI systems). Increasingly relevant with generative AI, recommender systems, and dark patterns.
- **Information ecosystem harm** (misinformation, disinformation, synthetic media, content pollution). A rapidly growing incident category that currently has no home in the enumeration.
- **Legal or regulatory harm** (AI-created legal liability, regulatory violation, interference with legal processes).
- **Civic, institutional, societal, and cultural harms** (harm to governance structures, public trust, labor markets, cultural heritage). Currently partially captured by the vague "public interest" but without sufficient specificity.

3.2. Recommendation

The proposed revision takes a targeted restructuring approach: retain the well-established OECD harm types as the foundation, remove items that belong in other fields, replace vague items with specific alternatives, and add well-established missing harm types. Every change traces back to the OECD framework, framed as clarification and extension rather than replacement.

Key design principles:

- Each enumeration should describe the nature of the harm, not the target, context, or legal classification.
- Items already captured in dedicated fields should be removed from AICIECCharmType with a cross-reference.
- "Public interest" is replaced by three specific categories (information ecosystem, civic or institutional, societal or cultural) using a distinguishing test: Does the harm affect what people can know or trust as true? Does it affect how society governs itself? Does it affect how people live, work, or express identity?
- "Human or fundamental rights" is retained as a crosscutting category with explicit guidance to also complete field 4.1.15.

3.3 Proposed Revised Language for 4.1.11 AICIECCharmType

This required value describes the types of harm associated with the AI incident using one or more of the following non-case-sensitive enumerations, including an unconstrained free-form option. Multiple selections are expected when an incident involves more than one type of harm.

Enumeration	Description
physical	Bodily injury, death, or harm to physical health or safety.
psychological	Harm to mental health, emotional wellbeing, or sense of personal safety. Includes distress, trauma, harassment, and intimidation.
reputational	Harm to the standing, credibility, or public perception of an individual or organization.
economic or property	Financial loss, property damage, loss of livelihood, or economic disadvantage.
environmental	Harm to the natural environment, including pollution, resource depletion, or ecological damage.
privacy or personal data	Harm to individuals through violation of privacy, unauthorized collection or exposure of personal data, surveillance, or misuse of personal information.
discrimination or bias	Differential treatment, exclusion, or disadvantage based on protected characteristics. Includes algorithmic bias causing unfair outcomes.
autonomy or agency	Loss or undermining of an individual's capacity for free, informed decision-making. Includes induced actions, beliefs, or preferences that the affected party would not have formed absent AI system intervention, or that were shaped without the affected party's understanding or meaningful consent. May result from manipulation, deception, coercion, dark patterns, or algorithmic behavioral influence.
information ecosystem	Harm to the integrity of public information and discourse. Includes misinformation, disinformation, synthetic media, content pollution, and erosion of the ability to distinguish authentic from artificial content. Distinguished from 'privacy or personal data,' which addresses harm to specific individuals through their personal information
civic or institutional	Harm to governance structures, civic processes, public trust in institutions, or the rule of law. Includes electoral interference, suppression of civic participation, and undermining of judicial or regulatory processes.
societal or cultural	Broad structural harm to communities, labor markets, cultural heritage, or social fabric. Includes labor market disruption, harm to creative industries, cultural homogenization, and erosion of social cohesion.
legal or regulatory	Harm to affected parties through legal mechanisms. Includes wrongful liability (e.g., AI-hallucinated citations causing sanctions), fraudulent legal filings, unauthorized legal advice relied upon to detriment, and interference with legal processes
human or fundamental rights	Violation of internationally recognized human rights or fundamental freedoms. This category crosscuts other harm types and should be selected in addition to the relevant specific harm type(s). *If selected, reporters should also complete AICIECChumanRightsImpact (4.1.15).*
… (free-form content)	Other harm types not captured by the above enumerations. Reporters should provide a brief description.

3.4 Mapping from Current to Proposed Enumerations

The following table shows how each current enumeration maps to the proposed revision, with rationale for each change.

Current Enumeration	Proposed Mapping	Rationale
physical	physical	Retained; description added for clarity
psychological	psychological	Retained; description added for clarity
reputational	reputational	Retained; description added for clarity
economic/property	economic or property	Retained; minor rename for readability
environmental	environmental	Retained; description added for clarity
public interest	Replaced by: information ecosystem, civic or institutional, societal or cultural	Original term too vague for consistent use. Replaced with three specific categories using jurisdiction-neutral language
critical infrastructure	Dropped	Already captured in dedicated field AICIECCcriticalInfrastructure (4.1.19). Not a harm type; describes the target/context of the harm
human or fundamental rights	human or fundamental rights	Retained as crosscutting category; added guidance to also complete AICIECChumanRightsImpact (4.1.15)
(not present)	privacy or personal data	New: captures privacy violations, data exposure, and surveillance harms. One of the most common AI incident types
(not present)	discrimination or bias	New: captures algorithmic bias and differential treatment. Most frequently reported AI-specific harm, previously invisible within "human or fundamental rights"
(not present)	autonomy or agency	New: captures manipulation, deception, and coercion by AI systems. Increasingly relevant with generative AI
(not present)	legal or regulatory	New: captures AI-created legal liability and regulatory violations

The three categories replacing "public interest" can be distinguished by asking which aspect of collective life the harm affects:

Information Ecosystem	Civic or Institutional	Societal or Cultural
Affects what people can know or trust as true	**Affects how society governs itself**	**Affects how people live, work, or express identity**
Misinformation, disinformation, deepfakes, content pollution, erosion of trust in information sources	Electoral interference, suppression of civic participation, erosion of trust in institutions, and undermining of the rule of law	Labor market disruption, harm to creative industries, cultural homogenization, and erosion of social cohesion

This test helps reporters select the appropriate category. Some incidents may involve more than one of these categories (e.g., AI-generated political deepfakes could affect both the information ecosystem and civic processes), which is why the field supports multi-select.

3.5 Adoption Considerations

List length. The proposed list has 14 enumerations (including free-form) compared to the current 8. Each item is a genuinely distinct harm type.

Crosscutting nature of "human or fundamental rights." This category intentionally crosscuts the others. A discrimination incident may involve both "discrimination or bias" and "human or fundamental rights." This is by design: the specific harm type enables trend analysis, while the rights flag ensures the normative dimension is captured and triggers completion of AICIECChumanRightsImpact (4.1.15).

Relationship to OECD framework. All five original OECD harm types are retained. The changes extend the OECD framework by replacing vague categories with specific ones and adding well-established harm types from the AI safety literature. This can be characterized as clarifying and extending the OECD approach, not replacing it.

Jurisdictional neutrality. The term "civic or institutional" was chosen over "democratic" to accommodate jurisdictions without democratic governance structures while still capturing harms to governance processes, rule of law, and civic participation. This addresses a known limitation of frameworks that use democracy-specific language.

Distinction from systemRelationship. The proposed 'legal or regulatory' harm type describes harm experienced BY affected parties through legal mechanisms (e.g., wrongful liability, interference with legal processes). This differs from 'legal obligation omission' in the current AICIECCsystemRelationship, which describes organizational compliance status. The former is a type of harm suffered; the latter is an organizational attribute unrelated to causal role.

4. AICIECCharmConsequence (4.1.12) and Proposed AICIECCseriousIncidentIndicators

4.1 Identified Issues

4.1.1 Issues with AICIECCharmConsequence (4.1.12)

- **Mixing of units.** The current enumeration mixes actual consequences (what happened) with response actions (compensation) and measurement approaches (quantification). These are fundamentally different kinds of information.
- **Limited utility for severity filtering.** The current field does not enable efficient identification of high-severity incidents that may trigger regulatory notification obligations or warrant priority attention.
- **Numeric precision is frequently unavailable.** Review of incidents in the AI Incident Database found that many lack quantitative information about scale. Structured numeric fields will likely produce mostly "unknown" values.

4.1.2 Gap: No Standardized Severity Indicators

The current specification lacks a standardized way to identify whether an incident crossed commonly recognized severity thresholds (death, serious injury, critical infrastructure disruption, rights violations). This gap impedes:

- Regulatory compliance: Organizations cannot efficiently determine reporting obligations
- Priority filtering: Databases cannot surface high-severity incidents for attention
- Trend analysis: Researchers cannot track serious incidents over time
- Cross-framework interoperability: Different regulatory regimes use similar severity concepts but have no common data structure

4.2 Recommendation

I recommend two coordinated changes:

- Add a new field AICIECCseriousIncidentIndicators with binary indicators for severity thresholds commonly used across regulatory frameworks
- Revise AICIECCharmConsequence to a minimal structure capturing approximate scale, ongoing status, and narrative context

The new AICIECCseriousIncidentIndicators field uses binary (yes/no/unknown) indicators rather than numeric scales or subjective severity ratings. This design choice is informed by analysis of established incident reporting frameworks in aviation (ICAO, NTSB, NASA ASRS), medical devices (FDA MAUDE), data protection (GDPR), cybersecurity (NIS2), and autonomous vehicles (NHTSA), all of which use categorical thresholds rather than numeric severity scales. See Appendix A for detailed background.

The field specifications avoid reference to any specific regulation to support international adoption while still enabling compliance with frameworks that use similar severity concepts.

Note on involvedMinors: Section 5.1.1 identifies the inclusion of 'children' as a demographic characteristic in AICIECCpartiesAffected as a mixing-of-units issue, since age is orthogonal to party role. The inclusion of involvedMinors in AICIECCseriousIncidentIndicators may appear inconsistent. The distinction is one of purpose: AICIECCpartiesAffected categorizes affected parties by their relationship to the AI system (user, subject, bystander), while AICIECCseriousIncidentIndicators identifies regulatory severity triggers regardless of category type. Protection of minors represents a near-universal regulatory concern that elevates reporting obligations, enforcement attention, and public interest across frameworks. Its inclusion reflects regulatory reality rather than taxonomic purity.

4.3 Proposed New Field: AICIECCseriousIncidentIndicators

This recommended field contains binary indicators for severe incident outcomes that commonly trigger regulatory notification obligations, enforcement actions, or elevated public interest. These indicators enable filtering for high-priority incidents and support compliance across multiple regulatory frameworks.

These indicators capture multiple dimensions commonly used by regulatory frameworks to identify serious incidents: outcome severity (fatality, serious harm, property/environmental damage, rights violations, critical systems disruption), affected populations of special concern (minors), and incident scope (widespread harm).

Guidance: Complete this field to identify whether an incident crossed commonly recognized severity thresholds. Select "unknown" when information is unavailable or unverified rather than leaving blank.

Enumeration	Description
involvedFatality	Indicates whether the incident resulted in the death of any person. Values: yes | no | unknown. Select "yes" if death occurred as a direct or reasonably proximate consequence of the incident, including deaths within 30 days where a causal connection is established or suspected.
involvedSeriousPhysicalHarm	Indicates whether the incident resulted in serious physical harm to any person's health or bodily integrity. Values: yes | no | unknown. Serious physical harm includes conditions that are life-threatening, require or prolong hospitalization, result in temporary or permanent impairment, or require medical intervention to prevent such outcomes.
involvedCriticalSystemsDisruption	Indicates whether the incident caused serious disruption to critical infrastructure or essential services. Values: yes | no | unknown. *If "yes," also complete AICIECCcriticalInfrastructure (4.1.19) to identify the affected sector(s).*
involvedRightsViolation	Indicates whether the incident resulted in violation of legally protected rights or freedoms, including constitutionally protected rights, human rights under international instruments, or rights protected by applicable law. Values: yes | no | unknown. *If "yes," also complete AICIECChumanRightsImpact (4.1.15).*
involvedSeriousPropertyOrEnvironmentalHarm	Indicates whether the incident caused serious harm to property or the natural environment. Values: yes | no | unknown. Assess seriousness based on economic impact, cultural significance, permanence of damage, and scale.
involvedMinors	Indicates whether the incident specifically affected or targeted children or minors (persons under the age of 18). Values: yes | no | unknown.
widespreadHarm	Indicates whether the incident caused or is capable of causing harm across a large population, multiple organizations, or multiple geographic regions. Values: yes | no | unknown. Select "yes" if the incident affected individuals across multiple jurisdictions, a large number of persons, multiple unrelated organizations, or critical services with broad public dependency.

4.4 Revised Field: AICIECCharmConsequence (4.1.12)

This optional field provides a qualitative description of incident consequences, including approximate scale, duration, and resolution status.

Enumeration	Description
approximateScale	Free text. Approximate number of individuals, entities, or systems affected, if known. Use precise counts when available (e.g., "1 person," "47 employees"). Use qualitative descriptors when precise counts are unavailable (e.g., "thousands of users," "population-scale"). Record "unknown" when scale information is unavailable.
ongoing	Enumeration: yes | no | unknown. Indicates whether incident effects are still occurring or have been resolved. Select "yes" if harmful effects continue, affected systems remain compromised, or remediation is incomplete.
consequenceNotes	Free text. Additional context about consequences not captured in other fields, such as reversibility of harm, duration, secondary effects, or remediation actions taken.

4.5 Field Relationships

The proposed fields complement existing AICIE fields without duplication:

Field	Question Answered	Relationship
AICIECCharmType (4.1.11)	What kind of harm?	Category taxonomy; orthogonal to severity
AICIECCseriousIncidentIndicators (NEW)	Did it cross severity thresholds?	Binary triggers for regulatory/priority filtering
AICIECChumanRightsImpact (4.1.15)	Which rights affected?	Specificity; complete if involvedRightsViolation=yes
AICIECCcriticalInfrastructure (4.1.19)	Which sector affected?	Specificity; complete if involvedCriticalSystemsDisruption=yes
AICIECCharmConsequence (4.1.12)	What was scale/duration/status?	Contextual details; free-text for flexibility

5. AICIECCpartiesAffected (4.1.14)

5.1 Identified Issues

5.1.1 Mixing of Units

The enumeration list mixes several different classification dimensions in a single field:

- **Role/relationship to AI system:** consumer, workers
- **Demographic characteristic:** children
- **Entity type:** business, trade union, government, civil society
- **Scope indicator:** general public

These are not measuring the same thing. A child could be a consumer, a worker, or a bystander. A worker could be in business, government, or civil society. The categories cross-cut rather than complement each other.

5.1.2 Overlap with Other Fields

- **"Children"** is now redundant with the proposed AIClECCseriousIncidentIndicators.involvedMinors, which specifically captures whether minors were affected.
- **Environment and infrastructure** are sometimes considered "affected parties" but are already captured in AIClECCharmType (environmental) and AIClECCcriticalInfrastructure (4.1.19). The field name "parties" implies persons or organizations, but the scope is unclear.

5.1.3 Vague and Overlapping Categories

- **"General public"** overlaps with all other categories. Consumers, workers, and civil society are all members of the general public. It is unclear when to select this versus more specific options.
- **"Civil society"** is undefined and could mean NGOs, community organizations, activists, or the public generally.
- **"Consumer"** may overlap with "general public" depending on interpretation.

5.1.4 Missing Relevant Party Types

The current enumeration omits several analytically useful categories:

- **Subjects:** Persons about whom the AI system made decisions or predictions, who may not have directly used the system (e.g., job applicants screened by AI, individuals flagged by predictive policing)
- **Bystanders:** Persons indirectly affected who were neither users nor subjects of the AI system

5.2 Recommendation

I recommend revising the field to focus specifically on human parties (persons and organizations) with categories based on their role or relationship to the AI system or incident. Non-human affected entities (environment, infrastructure, property) are already captured in other fields.

Key design principles:

- Each enumeration describes the affected party's relationship to the AI system or role in the incident
- "Demographic characteristics" (age, protected classes) are not party types. Minors are captured in AICIECCseriousIncidentIndicators.involvedMinors; other demographic details may be noted in AICIECCharmType or AICIECChumanRightsImpact free-text fields.
- Organizational types are consolidated rather than enumerated individually
- Cross-references guide reporters to appropriate fields for non-human affected entities

5.3 Proposed Revised Language for 4.1.14 AICIECCpartiesAffected

This optional value describes the types of persons or organizations affected by the AI incident using one or more of the following non-case-sensitive enumerations, including an unconstrained free-form option. This field captures human parties; the affected environment is captured in AICIECCharmType, and the affected critical infrastructure in AICIECCcriticalInfrastructure (4.1.19).

Enumeration	Description
users	Persons who directly interacted with or used the AI system.
subjects	Persons who were the focus of the AI system's operations, whether through decisions, predictions, assessments, profiling, or data processing. Subjects may not have directly used or been aware of the AI system.
bystanders	Persons indirectly affected who were neither users nor subjects of the AI system. Includes persons harmed by downstream consequences or collateral effects.
workers	Persons affected in connection with their employment or labor. This category captures labor-related harms and may be selected alongside other categories when relevant (e.g., a user who is also a worker experiencing workplace harm). Includes workplace safety incidents, labor displacement, and employment-related harms.
organizations	Businesses, institutions, associations, or other organizational entities affected by the incident. Use free-form content to specify type if relevant (e.g., "small business," "nonprofit," "trade union").
government entities	Government agencies, public sector bodies, or officials affected in their governmental capacity.
... (free-form content)	Other affected party types not captured by the above enumerations. Reporters should provide a brief description.

5.4 Mapping from Current to Proposed Enumerations

The following table shows how each current enumeration maps to the proposed revision, with the rationale for each change.

Current Enumeration	Proposed Mapping	Rationale
consumer	users	"Users" is more precise; consumers are users in a commercial context
children	Dropped	Now captured in AICIECCseriousIncidentIndicators.involvedMinors. Age is a demographic characteristic, not a party type
workers	workers	Retained with a clarified description
business	organizations	Consolidated; organizational subtypes can be specified in free-form
trade union	organizations	Consolidated into a general organizational category
government	government entities	Minor rename for clarity
civil society	organizations	Consolidated; "civil society" was too vague for consistent application
general public	Dropped	Too vague; use specific categories (users, subjects, bystanders) or multiple selections
(not present)	subjects	New: captures persons affected by AI decisions or data processing who did not use the system directly
(not present)	bystanders	New: captures persons indirectly affected

5.5 Cross-Reference Guidance

When completing AICIECCpartiesAffected, reporters should also consider:

If this was affected	Complete this field
Natural environment	AICIECCharmType: "environmental"
Critical infrastructure	AICIECCcriticalInfrastructure (4.1.19)
Property	AICIECCharmType: "economic or property"
Minors specifically	AICIECCseriousIncidentIndicators.involvedMinors
Human rights	AICIECChumanRightsImpact (4.1.15)

6. Additional Observations

While reviewing the systemRelationship and intentionality fields, the following observations about the broader schema may be worth noting for future discussion. These are not part of the current proposal but may inform subsequent revisions.

- **Severity enumerations (4.1.10).** The value "incident" appears twice in the enumeration list. Additionally, the ordering (serious hazard, hazard, incident, serious incident, incident) does not follow an obvious severity progression. Consider whether the list should be reordered or deduplicated.
- **JSON schema consistency.** Several fields in Annex A that support multi-select in the specification text are typed as "string" with a single "enum" rather than as an array. The schema should be reviewed for consistency with the multi-select intent described in the field definitions.
- **AI tasks (4.1.25).** The list omits "content generation," which appears in the OECD framework (Table B-1, row 25) and is relevant to a significant and growing category of incidents.
- **Demographic characteristics of affected parties.** The specification lacks a structured field for capturing demographic information about affected parties, such as age groups (other than minors), race/ethnicity, gender, disability status, or other protected characteristics. This information is relevant for analyzing patterns of disparate impact and would support discrimination or bias trend analysis. Currently, such details can only be captured in free-text fields (AICIECChumanRightsImpact, AICIECCadditionalInfo), limiting their utility for structured queries. The proposed AICIECCseriousIncidentIndicators.involvedMinors addresses one demographic dimension; others remain unstructured.
- **Clarifying definitions for data fields and enumeration values.** The proposed revisions in Sections 2–5 include brief descriptions for each data field and their enumeration values. This supports consistent application across the diverse reporting community. Most other data fields and their enumerated values lack definitions. Terms like "serious hazard" versus "hazard" versus "incident" versus "serious incident" lack distinguishing criteria, inviting inconsistent use. Consider adding brief definitions for all enumeration values, either inline in the field descriptions or in a consolidated glossary annex.

Appendix A: Background and Rationale for Section 4

A.1 The Regulatory Landscape

A.1.1 EU AI Act

The EU AI Act establishes mandatory incident reporting for providers of high-risk AI systems. Article 73 requires notification to market surveillance authorities of "serious incidents" as defined in Article 3(49). The definition establishes categorical thresholds rather than numeric severity scales: death, serious health harm, critical infrastructure disruption, fundamental rights violations, and serious property or environmental harm.

Reporting timelines are driven by outcome category: incidents involving widespread fundamental rights infringement or critical infrastructure disruption require notification within 2 days; incidents involving death require notification within 10 days; other serious incidents require notification within 15 days. This demonstrates that regulators need to quickly identify which threshold was crossed, not precisely how many people were affected.

A.1.2 GDPR

The General Data Protection Regulation provides instructive precedent for incident reporting under conditions of uncertainty. Article 33 requires notification within 72 hours of becoming aware of a personal data breach, and explicitly requires only "the approximate number of data subjects concerned" (emphasis added). The regulation acknowledges that precise counts are often unavailable at time of notification.

A.1.3 NIS2 Directive

The Network and Information Security Directive (NIS2) requires reporting of significant cybersecurity incidents based on both categorical triggers (operational disruption, cross-border impact) and quantitative guidance (e.g. financial loss exceeding €100,000 or 5% of turnover). NIS2 employs three-stage reporting: early warning within 24 hours, incident notification within 72 hours, and final report within one month.

A.1.4 Sectoral Frameworks

Analysis of established incident reporting frameworks in high-reliability sectors revealed consistent patterns:

- **NHTSA Standing General Order** for autonomous vehicles uses binary categorical indicators: fatality (yes/no), hospital-treated injury (yes/no), vulnerable road user struck (yes/no). The order explicitly acknowledges that "a reporting entity may not become aware of all circumstances related to the crash."
- **FDA MAUDE** (Manufacturer and User Facility Device Experience) for medical devices uses categorical patient outcomes: Death, Life-Threatening, Hospitalization, Disability, Required Intervention, Other, No Known Outcome, Unknown. The FDA acknowledges the database contains "incomplete, inaccurate, untimely, unverified, or biased data."
- **ICAO ADREP** for aviation uses clinically anchored injury categories: Fatal (death within 30 days), Serious (e.g.hospitalization exceeding 48 hours, bone fracture, severe hemorrhage, burns affecting more than 5% of body surface), Minor, None.

- **NASA ASRS** uses narrative-driven voluntary reporting with no required structured severity field. Severity emerges from expert analyst coding after submission.

A.2 Key Findings

Across all frameworks examined, several patterns emerged:

- Binary thresholds drive regulatory action, not severity scales. Regulators need to know whether an incident crossed a threshold (death occurred, critical infrastructure disrupted), not precisely how severe it was on a 1 to 10 scale.
- Definitions are anchored to observable criteria. ICAO's definition of "serious injury" as hospitalization exceeding 48 hours provides objective, clinically verifiable criteria.
- All frameworks acknowledge uncertainty. GDPR uses "approximate," NHTSA acknowledges incomplete awareness, FDA warns of data quality limitations.
- Numeric counts are rarely available. Review of AIID incidents suggests most lack quantitative scale information.
- Staged reporting recognizes evolving understanding. Multiple frameworks require updates until remediation is complete.

A.3 Design Philosophy: Jurisdictionally Neutral but Regulation Ready

The field specifications avoid explicit reference to any specific regulation for several reasons:

- International adoption: The AICIE specification should serve incident reporters globally, not only those subject to EU regulation.
- Regulatory evolution: Incident reporting requirements are emerging rapidly. Tying specifications to current regulatory text creates a maintenance burden.
- Regulatory harmonization: Multiple jurisdictions are developing similar severity concepts. A jurisdictionally neutral specification can serve as a common infrastructure.
- Research utility: Academic researchers need consistent severity indicators regardless of which legal framework applies.

Despite avoiding explicit regulatory references, the proposed fields are designed to fully support compliance with the EU AI Act and comparable frameworks. An organization completing the AICIECCseriousIncidentIndicators field will capture all information needed to determine whether an incident meets serious incident criteria and which reporting timeline applies.