**Incident Classification Assignment**

**(High School 12–18, Spain)**

**Module 8C**

1. **Purpose & Scope**

This framework classifies AI-related incidents affecting **1º ESO - 2º Bachillerato (ages 12–18)** and sets the **response level, escalation path, and legal checks** (GDPR / EU AI Act where relevant). It applies to all AI and AI-enabled platforms used for teaching, assessment, wellbeing, and administration.

1. **Severity Scoring Rubric**

Score each criterion, then sum the points to get a **provisional level** (map below). If any **red-flag trigger** is present, **escalate one level**.

| **Criterion** | **1 pt** | **2 pts** | **3 pts** | **4 pts** |
| --- | --- | --- | --- | --- |
| **Scope** (how many affected) | 1 person | <1 class | 1 class/ group | Multiple classes / whole school |
| **Sensitivity** (data/content) | Non-personal | Basic personal (name, class) | Academic/behavioural | Special category/biometric/sexual content |
| **Impact on Individuals** | Inconvenience | Minor distress | Educational harm (grades/opportunity) | Significant harm / safeguarding risk |
| **Operational Disruption** | None | Brief (≤1 lesson) | Day-level | Multi-day / exams blocked |
| **Legal/Regulatory Risk** | None | Low (unclear) | Likely GDPR duty | GDPR/AI Act duty very likely |

**Red-flag triggers (auto-escalate by +1 level):**

* Child safeguarding risk (grooming, sexualized content, doxxing).
* **Biometric** processing without valid basis/DPIA.
* **Systemic unfairness** affecting grades/admissions/support decisions.
* Evidence of **external compromise** (hack) or widespread data exfiltration.

**Level Mapping (after scoring & any red-flag bump)**

* **Level 1 (Minor):** 5-7 pts
* **Level 2 (Moderate):** 8-10 pts
* **Level 3 (Serious):** 11-13 pts
* **Level 4 (Critical/Crisis):** 14-20 pts **or any red-flag with serious harm**

1. **Response Playbook by Level**

| **Level** | **Typical Actions** | **Notifications (internal/external)** | **Examples** |
| --- | --- | --- | --- |
| **L1 -Minor** | Fix/patch, log in incident register, vendor ticket | Inform tool owner; no external | Timetabling AI double-books room |
| **L2 -Moderate** | Contain (pause feature), root-cause, teacher comms, parent notice if prudent | Notify SLT; DPO review for GDPR threshold | Proctoring false flags with **no** grade impact |
| **L3 -Serious** | Immediate containment/pause, targeted parent notices, offer remedies, formal RCA | **DPO assesses & notifies DPA within 72h if GDPR breach likely;** inform provider; brief board | Private feedback exposed to peers; grading bias impacting marks |
| **L4 - Critical** | All-hands response; suspend system; safeguarding support; legal counsel | **Police/child-protection as needed; DPA (GDPR); provider/MSA (AI Act serious incident, via provider); board/media comms** | Deepfake harassment; biometric misuse; major breach |

1. **Fast Decision Tree**
2. **Anyone at risk of harm or safeguarding?**  
   → Yes ⇒ **Level 4**. Call safeguarding lead / police; contain immediately.
3. **Personal data exposed/ accessed unlawfully?**  
   → Yes ⇒ DPO risk test. If **risk likely**, notify **DPA ≤72h** (GDPR). Level 3–4.
4. **High-risk AI involved (grading/admissions/placement) with unfair outcomes?**  
   → Likely **Level 3**; inform provider; consider AI Act “serious incident” (provider → MSA ≤15 days).
5. **Operational outage only, no harm?**  
   → Level 1–2; fix, log, inform affected staff.

**5) Worked Classifications (Spanish secondary context)**

**A) 11th-grade (1º Bachillerato) English - software bug exposed private peer-feedback comments**

* Scope: one class (3) • Sensitivity: academic feedback (3) • Impact: embarrassment + grade influence possible (3) • Ops: brief (2) • Legal risk: likely GDPR (3) → **14 pts ⇒ Level 4?**  
  No safeguarding, so **no red-flag**; total **14** maps to **L4**, but if impact is **not** significant harm, downgrade to **L3** with strong justification.  
  **Classify:** **Level 3 (Serious)** → **Contain** (disable feature), **DPO risk test & notify DPA within 72h if risk likely**, notify affected families, vendor fix, teacher oversight.

**B) AI proctoring flags 5 students incorrectly; no grade penalties issued**

* Scope: <1 class (2) • Sensitivity: behavioural inference (3) • Impact: minor distress (2) • Ops: one exam session (2) • Legal risk: low-mod (2) → **11 pts ⇒ Level 3**  
  Mitigating context (quick reversal, no records kept) can justify **Level 2 (Moderate)** with **manual review mandated**.  
  **Classify:** **Level 2–3 (document rationale).** Pause proctoring pending RCA; teacher review required.

**C) Learning-path recommender assigns easier content to girls vs. boys with same history**

* Scope: multi-class (4) • Sensitivity: educational profiling (3) • Impact: educational harm (3) • Ops: ongoing (3) • Legal: fundamental-rights risk (3) → **16 pts ⇒ Level 4** (systemic unfairness red-flag).  
  **Classify:** **Level 4 (Critical)** → Suspend feature, notify leadership & families impacted, engage vendor for urgent fix/retraining; consider AI Act serious incident route (via provider).

**D) Scheduling AI double-books a lab; teacher catches it**

* Scope: one room (1) • Sensitivity: none (1) • Impact: inconvenience (1) • Ops: brief (1) • Legal: none (1) → **5 pts ⇒ Level 1**.  
  **Classify:** **Level 1 (Minor)** → Log & patch.

**6) Escalation & Ownership**

* **Incident Coordinator:** (Name, Role)
* **DPO / GDPR Lead:** (Name, Contact]
* **Safeguarding Lead:** (Name, Contact)
* **AI Tool Owner (per system):** (Name, Dept)
* **Principal/Head:** (Name]

**Escalation timers:**

* **L2:** notify SLT within **4 hours**.
* **L3:** notify SLT/DPO within **2 hours**; draft parent comms same day.
* **L4:** immediate call-out (≤1 hour); safeguarding/legal activated.

**7) One-Page Incident Log**

* **Date/Time discovered:**
* **Reporter:**
* **System / Vendor / Version:**
* **Summary:**
* **Preliminary classification (score + level):**
* **Containment taken:**
* **Data involved (types, volume):**
* **GDPR risk test result:** Notify DPA? **Y/N** (why)
* **AI Act serious incident suspected?** **Y/N** (provider informed date/time)
* **Stakeholders notified (who/when):**
* **Root-cause analysis owner:**
* **Corrective actions & deadlines:**
* **Close-out date & final classification:**
* **Lessons learned / policy updates:**

**8) Review & Maintenance**

* **Review this matrix each term** or after any Level 3/4 incident.
* Record **all** Level 2–4 incidents in the central register; sample-audit Level 1 monthly.