



THE PROFESSOR-AI

PRACTICAL, HONEST AI FOR SENIOR LEADERS

WORKED EXAMPLE · VERSION TWO

AI Incident Response Plan: stress-tested

The same plan after it was tested against three real incidents: an AI agent swarm, an autonomous agent breach, and a supplier disclosing seven months late.

ORGANISATION	VERSION	PAGES	STATUS
Fictional FE college	Two, stress-tested	Seven	Template, not advice

theprofessor.info · Paul Noon · Former British diplomat, former Deputy Vice-Chancellor, AI adviser and Non-Executive Director

STEP TWO OUTPUT · STRESS-TESTED

AI Incident Response Plan: version two

The same worked example, revised after testing against three real incidents: an AI agent swarm attack, an autonomous AI agent data breach, and a supplier disclosing an incident seven months late.

ADDED Standing authority Named people can contain an incident out of hours without assembling anyone first.	ADDED AI agent register Every agent, what it can reach, who owns it, and where the off switch is.
ADDED Supplier duties What suppliers must tell us, how fast, and what we do when they are late.	ADDED Awareness defined A written test for when the 72-hour clock starts, and who starts it.

1. Purpose, scope and definitions

Scope is unchanged from version one: AI-assisted attacks on the College, incidents caused by AI systems the College operates, and incidents disclosed to us by AI suppliers.

TERM	WHAT IT MEANS HERE
Becoming aware	The College is aware from the moment any member of the Incident Team has a reasonable degree of certainty that a security incident has occurred that has led to personal data being compromised. A supplier's notification, however late, makes us aware on the day we receive it. An automated alert alone does not, but it starts the clock on investigating, and that investigation has a 24-hour limit.
AI agent	Any system that takes actions in College systems without a person approving each step, including agents inside approved products.
Containment action	Any action that stops an agent, system, account or integration from doing further harm.
Material incident	Any incident involving special category data, more than 100 data subjects, examinations, safeguarding records, finance systems, or likely press interest.

2. Detection: what tells us, and how fast

- Automated alerts. Alerts for impossible-travel logins, mass file access, new domain administrator accounts and outbound data volumes route to the on-call phone, not to a shared mailbox.

- **Agent telemetry.** Every agent in the register logs its actions to a single location that the Incident Lead can read without the agent's owner present.
- **Third-party notice.** Alerts from the supplier, the sector body, a security researcher, the police or a threat intelligence service are treated as detection, not as rumour.
- **Staff and student reports.** One phone number, published, answered out of hours.

PRE-AUTHORISED AUTOMATIC CONTAINMENT

The Corporation has approved these actions being taken automatically, without a human decision: disabling an account after confirmed credential theft, isolating a host on confirmed ransomware behaviour, and suspending an AI agent whose actions fall outside its declared permissions. Anything automatic is logged and reviewed within 24 hours. A system that only alerts a human cannot keep pace with an attack that takes five minutes.

3. Standing authority out of hours

The following people may take the following actions at any hour, without consulting anyone first, and are protected by the Corporation for doing so in good faith.

WHO	MAY DO THIS, ALONE, AT ANY TIME
On-call IT engineer Rota published weekly	Isolate a server or network segment. Disable any account. Suspend any AI agent or integration. Block an outbound destination.
Incident Lead Director of IT	All of the above, plus take a service offline entirely, including systems used for teaching.
Data Protection Officer	Start the 72-hour clock and open the ICO submission, without waiting for the full picture.
Principal and Chief Executive	Approve public communication. Authorise external incident response support and the spend that comes with it.

THE STANDING INSTRUCTION

Contain first, explain afterwards. Nobody at the College will be criticised for taking a system offline during a suspected incident and turning out to be wrong. They will be criticised for waiting for a meeting.

4. The first hour, minute by minute

TIME	ACTION	WHO
0 to 5 min	Contain. Isolate, disable or suspend. Do not investigate first.	Whoever holds standing authority and is available
5 to 15 min	Call the Incident Lead and the DPO by phone. Open the decision log. Preserve logs and snapshots before rebuilding anything.	On-call engineer

15 to 30 min	Confirm scope: which systems, which data, which agents, which accounts. Check the agent register for anything else with the same access.	Incident Lead and System Owner
30 to 45 min	Decide on notifiability and record the time of awareness. Open the ICO submission even if the answer is not yet clear.	DPO
45 to 60 min	Issue the internal holding line. Inform the Principal and the Clerk. Decide whether to call in external support.	Communications Lead, Incident Lead

5. The AI agent register

No AI agent may operate in College systems unless it appears in this register. The register is reviewed monthly by the Incident Lead and reported to the Audit Committee each term. Any agent not in the register is treated as an unauthorised system and stopped.

AGENT	OWNER	CAN REACH	CAN IT WRITE?	OFF SWITCH	REVIEWED
[Name and product]	[Named person]	[Systems and data]	[Yes or no]	[Exact steps, and who can do it at 3am]	[Date]

Every row must name a person, not a team. "IT" cannot switch anything off at three in the morning; a person can.

6. The regulatory clock

The College reports a notifiable personal data breach to the ICO without undue delay and, where feasible, within 72 hours of becoming aware, using the definition in section 1. Where the College cannot provide full information in that window, it submits what it has and completes the report in phases. Where individuals are at high risk, they are told without undue delay, in plain language, with one action they can take.

POINT IN TIME	WHAT MUST EXIST
Hour 0	Time of awareness recorded, with the name of the person who made the judgement and why.
Hour 12	Draft ICO submission open, with known facts only. Nothing invented to fill a field.
Hour 24	Decision made: notifiable, not notifiable, or still investigating with a reason recorded.
Hour 48	If notifiable, submission reviewed by the DPO and the Principal.
Hour 72	Submitted, or a documented reason why it was not feasible, plus the date it will be.

7. Suppliers: what we require, and what we do when they are late

These terms are required in every new or renewed contract for a product with AI features, and the Director of IT maintains a list of existing contracts that fall short.

- Notification to the College within 24 hours of the supplier becoming aware of any incident affecting College data or systems, including incidents during the supplier's own testing.
- Notification of any unauthorised access by the supplier's own models or staff, whether or not the supplier judges it harmful.
- A named contact and an out-of-hours number, tested annually.
- Log data made available to the College on request, in a usable form, within five working days.
- No unilateral change to the model, its permissions or its data handling without notice to the College.

WHEN A SUPPLIER DISCLOSES MONTHS LATER

Treat the day of notification as the day of awareness and run this plan from hour zero. Do not treat it as historic. Ask three questions in writing: what data was touched, when did the supplier first know, and why did the notification take this long. Record the answers, report them to the Audit Committee, and take the delay into account at renewal.

8. Communications, written in advance

Three holding statements are drafted, approved and stored outside the affected systems: one for staff, one for students and parents, one for press. Each says what is known, what the College is doing, what the reader should do, and when they will hear next. Nobody drafts from scratch during an incident.

9. Evidence

Preserve before you repair. Logs, memory captures, agent transcripts and email headers are exported to storage that the incident cannot reach, before any system is rebuilt. Agent transcripts matter as much as server logs: they are the record of what the agent decided and why.

10. Review, rehearsal and reporting

- Written review within ten working days, to the Senior Leadership Team and the Audit Committee.
- Tabletop exercise twice a year, at least one out of hours, using the scenarios in the prompt pack.
- Quarterly board report: agents in the register, incidents, near misses, supplier notifications, and any contract that still lacks the disclosure terms.
- This plan is reviewed annually, or after any material incident, whichever is sooner.

APPENDIX

Templates to fill in before you need them

A. Contact sheet, kept on paper as well as online

ROLE	NAME	MOBILE	BACKUP
Incident Lead			
On-call engineer			
Data Protection Officer			
Principal			
Clerk to the Corporation			
Cyber insurer, policy number			
External incident response			
Key AI suppliers			

B. Decision log, started in the first fifteen minutes

TIME	DECISION	REASON, AND WHAT WAS KNOWN AT THE TIME	TAKEN BY

Written at the time, not reconstructed afterwards. This is the document a regulator, an insurer and your auditors will all ask for.

C. Five questions the board should be able to answer today

1. How many AI agents operate in our systems, and who owns each one?
2. Who can switch one off at three in the morning, without calling a meeting?
3. What must our AI suppliers tell us, and how quickly, in writing?
4. When does our 72-hour clock start, and who decides?
5. When did we last rehearse any of this?

BEFORE YOU ADOPT THIS

This is a worked template for a fictional college, produced to show a method. It is not legal advice and it is not a compliance product. Your data protection officer and your legal adviser should review it, and your board should approve the standing authority in section 3 before anyone relies on it.