



THE PROFESSOR-AI

PRACTICAL, HONEST AI FOR SENIOR LEADERS

WORKED EXAMPLE · VERSION ONE

AI Incident Response Plan: the first draft

The plan as an AI produces it in a single prompt, before anyone has tested it. Good, and not good enough. The five gaps are set out at the back.

ORGANISATION	VERSION	PAGES	STATUS
Fictional FE college	One, untested	Five	Template, not advice

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

STEP ONE OUTPUT · FIRST DRAFT

AI Incident Response Plan: the first draft

Midshire College, a worked example. Six hundred staff, about eight thousand students, Microsoft 365, PaperCut print management, and two teams piloting AI agents with access to student records.

READ THIS FIRST

This is the plan as it comes out of the first prompt, before it has been tested. It is a good draft and it is not good enough. The gaps are listed on the last page, and the stress-tested version fixes them. Use this one to see what a first pass looks like, and to compare against your own plan.

1. Purpose and scope

This plan sets out what the College does when an artificial intelligence system, our own or someone else's, causes or contributes to a security or data incident. It covers attacks carried out using AI tools, incidents caused by AI agents the College operates, and incidents disclosed to us by an AI supplier.

It sits under the College's Information Security Incident Policy and does not replace it. Where the two conflict, the Information Security Incident Policy takes precedence on matters of security, and this plan takes precedence on matters specific to AI systems.

2. What counts as an AI incident

- An AI agent operated by the College takes an action nobody authorised, including reading, changing or deleting records.
- An AI agent or AI-assisted attack is used against College systems.
- Personal data is exposed to, or through, an AI system, including staff use of unapproved AI tools.
- An AI supplier tells us that their system accessed, processed or exposed College data.
- An AI system produces an output that causes harm to a student, member of staff or third party.

3. Who does what

ROLE	HELD BY	RESPONSIBLE FOR
Incident Lead	Director of IT	Runs the response. Decides containment actions. Single point of decision until stood down.
Deputy Incident Lead	Head of Infrastructure	Takes the Incident Lead role when the Lead is unavailable.

Data Protection Officer	[NAME]	Decides whether the incident is notifiable. Drafts and submits the ICO report.
Accountable Officer	Principal and Chief Executive	Owns the response to the Corporation. Approves external communications.
Communications Lead	Director of Marketing	Staff, student, parent and press communications.
Clerk to the Corporation	[NAME]	Informs the Chair and Audit Committee. Records decisions.
System Owner	Named per system in the asset register	Provides access, logs and context for the affected system.

4. The first hour

TIME	ACTION, AND WHO TAKES IT
0 to 15 minutes	Whoever notices raises it with the IT Service Desk and the Incident Lead by phone, not email. The Incident Lead confirms the incident is live and starts the decision log.
15 to 30 minutes	Incident Lead isolates the affected system or account. Incident Lead notifies the DPO and the Principal. System Owner preserves logs.
30 to 60 minutes	DPO makes an initial view on whether personal data is involved and whether the 72-hour clock has started. Communications Lead prepares a holding line. Clerk informs the Chair if the incident is material.

5. Containment

The Incident Lead may disconnect a system from the network, disable an account, suspend an integration or stop an AI agent. Where an action would interrupt teaching or examinations, the Incident Lead consults the Principal before acting.

6. Reporting to the ICO

Under UK GDPR the College must report a notifiable personal data breach to the Information Commissioner's Office without undue delay and, where feasible, within 72 hours of becoming aware of it. The DPO owns this decision and the submission. Where the College decides not to report, the DPO records the reasons in the decision log.

Where the breach is likely to result in a high risk to individuals, the College must also tell those individuals without undue delay.

7. Communications

All external communication goes through the Communications Lead and is approved by the Principal. Staff are told what has happened, what they must do and what they must not say. No member of staff speaks to the press.

8. Suppliers

Where a supplier's system is involved, the Incident Lead contacts the supplier's named support contact and requests a written account of what happened, what data was affected and what the supplier has done about it.

9. After the incident

The Incident Lead produces a written review within ten working days covering what happened, what was done, what worked and what did not. The review goes to the Senior Leadership Team and to the Audit Committee at its next meeting.

WHAT THIS DRAFT MISSES

Five gaps, found by stress-testing it

Each of these came out of testing the draft above against the three real incidents from the video. They are the reason the second version exists.

1. It assumes somebody is awake

Every step starts with a person noticing and phoning someone. The fastest attack in the PaperCut campaign went from first access to full network control in five minutes, and much of that campaign ran overnight and at weekends. The draft has no standing authority for anyone to act out of hours without first assembling people.

2. It has no register of AI agents

The plan says the Incident Lead can stop an AI agent. It does not say how anybody knows which agents are running, what each one can reach, who owns it or where its off switch is. You cannot switch off something nobody has written down.

3. It treats supplier disclosure as a request, not a requirement

The draft asks a supplier for a written account after an incident. It says nothing about what suppliers must tell the College, how quickly, or what happens when a supplier discloses something months after the event. Anthropic's disclosure of a January incident came in September, and it was voluntary.

4. It never defines "becoming aware"

The 72-hour clock runs from the moment the College becomes aware. The draft does not say what awareness means, who can trigger it, or whether a supplier's late notice starts the clock. In practice that ambiguity is where organisations lose a day.

5. Nothing is written in advance

Holding statements, the ICO submission skeleton, the contact list and the decision log template are all produced during the incident, by people who are already busy. None of them exist before it starts.

THE HONEST POINT

This draft is better than what most organisations have, which is a general security incident policy with no mention of AI at all. The failure is not that the draft is bad. It is that nobody tested it before filing it.