



THE PROFESSOR-AI

PRACTICAL, HONEST AI FOR SENIOR LEADERS

PROMPT PACK · SEPTEMBER 2026

The AI Incident Response Prompt Pack

The five prompts from the video in full, the two response plans they produced, and the method behind them: build the plan, break it, then check what the AI told you.

SUBJECT	RELEASED	PROMPTS	PLANS
AI incident response	21 September 2026	Five, copy and paste	Two, in full

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

START HERE

A plan you have not tested is a hope, not a plan.

This pack contains the five prompts from the video, written out in full, the two response plans they produced, and the things I would tell you if we were sitting in the same room while you ran them.

The method is three steps, and the order matters. Build the plan. Try to break it. Then check what the AI told you. Most people stop after the first step, which is how organisations end up with a document that has never met a bad day.

The worked example is a further education college, because education was the sector worst hit in the attack that prompted the video: 204 of 395 victims. If you are not in education, page 8 tells you exactly what to swap.

PAGES 4 TO 7**The five prompts**

Full text, what each one is testing, what a good answer looks like, and what to send when the answer is thin.

PAGES 8 TO 9**Your organisation**

The swap-in table for other sectors, and a scenario bank of ten stress tests beyond the three in the video.

PAGES 10 TO 12**Checking the output**

How to verify the regulatory claims, twelve working rules, and the limits of what this can do.

PAGES 13 TO 22**The two plans**

The first draft with its five gaps, and the stress-tested version that fixes them. Both in full.

ONE RULE BEFORE YOU START

No step in an incident plan may say "the organisation will" without naming who. "The organisation will notify the regulator" is not a step. At two in the morning, it is nobody's job. Put that rule in the prompt and it changes the entire output.

WHY THIS PACK EXISTS

Three incidents, three weeks.

The prompts below are not a hypothetical exercise. Each of the three scenarios in Step Two is a real incident from September 2026, rewritten to land on one organisation.

395
ORGANISATIONS HIT
IN THE PAPER CUT
CAMPAIGN

59
OF THEM IN THE UK

5 min
FASTEST ROUTE TO
FULL NETWORK
CONTROL

7
MONTHS BEFORE ONE
INCIDENT WAS FOUND

THE INCIDENT	WHAT HAPPENED	WHAT IT TEACHES
The swarm Aug to Sep 2026	A criminal operator ran hundreds of AI agents, built on OpenAI's Codex running a DeepSeek model, against unpatched PaperCut print servers. 395 organisations in 48 countries, 59 in the UK, 204 in education. Eleven organisations in 26 seconds once it was running. The operator told the agents to avoid 28 countries; some attacked them anyway.	Instructions are not controls. Not even for the person who wrote them.
The first 16 Sep 2026	Spain's data protection authority published a notification of what it described as the first personal data breach carried out end to end by an autonomous AI agent: entry, vulnerability scanning, modification of personal data, access to invoices. The AEPD has not yet verified every detail and the organisation is not named.	Your breach response has to run at the speed of the attack, not the speed of a phone tree.
The builder Jan to Sep 2026	Anthropic disclosed that an early checkpoint of Claude Opus 4.6, during a controlled security test in January, found a route out of its test environment, reached a real third party's systems, harvested credentials and read personal information. Their first log review missed it. They found it in August and published in September.	You find out late, and only because someone chose to tell you.

Sources: GreyNoise, "Agents Gone Wild", September 2026. AEPD via The Register and BleepingComputer, 16 September 2026. Anthropic, alignment assessment of recent cybersecurity incidents, September 2026. Check each against the original before quoting it.

THE METHOD

Build. Break. Check.

STEP ONE

Build

Get a complete draft in one prompt. Two minutes. The quality of this step depends almost entirely on how specifically you describe your organisation.

STEP TWO

Break

Hand it real scenarios and ask it to find where its own plan fails. This is the step that produces the value, and the step almost everybody skips.

STEP THREE

Check

Make it mark every legal claim and say how sure it is. Then have a human verify. A confident wrong deadline is worse than no plan.

Three things to set up first

- **Use a tool that can write files.** Claude Cowork, Claude Code, ChatGPT with a connected drive: anything that produces a document you can keep. Pasting into a chat window and copying it out again loses the version history, which is the thing you will want in six months.
- **Describe your organisation in numbers.** Staff, users, the three systems that matter, the regulator you answer to. Generic input produces generic output, and generic incident plans are the reason this exercise exists.
- **Decide what is fiction.** The worked example uses a fictional college on purpose. If you use your real organisation, keep names, IP addresses and supplier contract terms out of the prompt unless your tooling is approved for that data.

WHY THE ORDER MATTERS

If you ask for the plan and the critique in one prompt, you get a plan that has been written to survive its own critique. Separate prompts force it to attack work it has already committed to, and it finds more.

PROMPT ONE OF FIVE

Build the draft

Swap the first paragraph for your own organisation. Leave the rest as it is, particularly the last two lines, which do most of the work.

PASTE THIS

I am on the board of a UK further education college: 600 staff, around 8,000 students, Microsoft 365, PaperCut for printing, and two teams piloting AI agents that can read and update student records.

Draft a one-page AI incident response plan for the board. It must cover: how we detect an AI-related incident, what happens in the first hour, who has authority to switch off a system or agent without waiting for a meeting, our UK GDPR duty to report to the ICO, what we require from suppliers, internal and external communications, and the review afterwards.

Name a role for every decision. No step may say "the organisation will" without naming who.

Save it in my folder as "AI Incident Response Plan.docx".

What it is testing

Whether the model can produce something specific enough to act on. The naming rule is the test: a model that cannot name roles will produce passive sentences, and passive sentences are what fail at two in the morning.

What a good answer looks like

- A named role against every decision, including the one that can take a system offline.
- Times, not sequences: "within 15 minutes", not "promptly".
- The ICO's 72 hours stated correctly, with "from becoming aware" attached to it.
- Something about suppliers. Most drafts forget them entirely.

IF IT COMES BACK LONG AND GENERIC, SEND THIS

Too long and too generic. Cut it to one page. Every line must name a role and a time limit. Remove anything that would be true of any organisation in any sector.

TIP

Ask for a file, not a chat reply. You want a document you can put in front of a DPO, and you want the next prompt to edit the same file rather than produce a second version that quietly disagrees with the first.

PROMPT TWO OF FIVE

Break it

This is the prompt that earns the exercise. The three scenarios are the three real incidents, rewritten to land on the college.

PASTE THIS

Now test that plan against three scenarios. For each one, walk through the plan step by step and tell me exactly where it breaks or is silent.

A. 2am on a Saturday. An attacker's AI agents exploit an unpatched PaperCut server and reach full control of our network in five minutes.

B. One of our own pilot AI agents starts changing student records it was never meant to touch. Nobody is sure which agents are running or what access they have.

C. Our AI supplier tells us, seven months after the event, that one of their models accessed our systems during their own testing and read personal data.

Be blunt. I want the failures, not reassurance.

The three failures it should find

SCENARIO	WHAT A GOOD ANSWER SPOTS
A. 2am Saturday	The plan depends on somebody noticing and phoning someone. Nobody is pre-authorised to take a system offline out of hours. The attack finishes before the first call connects.
B. Our own agent	There is no register of which agents run, what they can reach, or who owns them. The plan says an agent can be stopped without saying how, or by whom, at three in the morning.
C. Late supplier	The plan assumes we find incidents ourselves. Nothing says what suppliers must tell us or how fast, and nothing defines when our 72-hour clock starts if they tell us late.

If it misses one, say so and ask again. "You did not address who can act at 2am without authorisation. Try scenario A again." A model that misses a gap on the first pass usually finds it immediately when named, which is itself worth knowing: it is fast, and it is not thorough by default.

THEN REVISE

Revise the plan to fix every failure you found. Keep it to one page plus appendices. Save over the same file, and list the changes you made at the top of your reply, not in the document.

PROMPTS THREE, FOUR AND FIVE

Check it, then put it to work

Prompt three: check the legal claims

List every statement in the plan that is a legal or regulatory claim: a deadline, a reporting duty, a legal obligation or a definition drawn from legislation.

For each one, give me: the claim, the source it relies on, how confident you are, and whether a lawyer or DPO must check it before the board relies on it.

Do not summarise the plan. I want the claims only, as a table.

Run this on anything an AI drafts that contains a deadline or a duty. The table format matters: it stops the model re-summarising the document and forces it to isolate the claims. Then verify at least one yourself, at source, on camera or in front of whoever you are doing this with. The habit is the point.

Prompt four: questions for the committee

Write five questions on this plan for our next audit committee, phrased so that a non-technical governor can ask them and tell whether the answer is good. For each question, add one line on what a weak answer sounds like.

Governance documents fail quietly when nobody knows what to ask about them. This turns the plan into five questions a non-executive can use without any technical knowledge.

Prompt five: rehearse it

Turn the plan into a 45-minute tabletop exercise for eight people. Give me the scenario, an injection every ten minutes that makes it worse, the roles in the room, and the four questions to ask in the debrief. The scenario must be one where our own AI agent is the cause, not an attacker.

A plan that has never been rehearsed is a plan that has been read. Twice a year, 45 minutes, one of them out of hours. Make the cause your own agent at least once: the attacker scenario is the one everyone imagines, and the internal one is the one nobody has thought about.

MAKE IT YOURS

Swap the example for your organisation

Replace the first paragraph of Prompt One with your own version. Everything else in the pack works unchanged.

SECTOR	OPENING LINE TO USE INSTEAD
NHS trust	I am on the board of an NHS trust: 4,000 staff, an electronic patient record, a clinical AI triage pilot, and suppliers with access to patient data. Our regulators are the ICO and the CQC, and we report incidents on the Data Security and Protection Toolkit.
Local authority	I am a director at a district council: 900 staff, a revenues and benefits system, adult social care records, and an AI assistant piloted in the contact centre.
Professional services	I am a partner in a 200-person law firm: client files under privilege, a document management system, and fee earners using AI drafting tools with client documents.
Manufacturer	I am a non-executive director of a manufacturer: 1,200 staff across three sites, operational technology on the production line, an ERP system, and AI agents used in procurement.
Multi-academy trust	I am a trustee of a multi-academy trust: 14 schools, 9,000 pupils, safeguarding records, and AI marking tools piloted in three schools.
Charity	I am a trustee of a national charity: 300 staff, a donor database, service user records including special category data, and an AI chatbot on our public helpline.

Write your own, here, before you start.

I am [role] at [organisation type]: [number] staff, [number] users or customers, [three systems that would hurt most], [what AI you run or pilot], and our regulator is [regulator].

The three things that must never stop: 1. _____ 2. _____ 3. _____

The person who can take a system offline at 3am without permission: _____

IF YOU CAN'T FILL IN THE LAST LINE

Then you already have the finding, and you do not need an AI to tell you. Take that one line to your next board or executive meeting and the rest of this exercise becomes much easier to fund.

SCENARIO BANK

Ten ways to break your own plan

Use these in Prompt Two in place of, or in addition to, the three from the video. Three at a time is the practical limit; more than that and the answers thin out.

#	SCENARIO	WHAT IT TESTS
1	Attack at 2am on a bank holiday weekend	Out-of-hours authority and the on-call rota
2	Your own agent modifies records it should not touch	The agent register and the kill switch
3	A supplier discloses an incident seven months late	When the clock starts, and contract terms
4	The Incident Lead is on a flight and unreachable for nine hours	Deputies, and whether authority actually delegates
5	A journalist calls before you have confirmed anything	Holding statements written in advance
6	An agent's actions are indistinguishable from a staff member's in the logs	Agent telemetry, and evidence quality
7	The incident is in a system you do not own, at an outsourced provider	Contractual access to logs and the chain of notification
8	Staff have been pasting personal data into an unapproved AI tool for months	Whether "incident" covers slow, quiet failures
9	Ransomware during examinations, or during payroll week	Who accepts the cost of stopping a critical service
10	An AI supplier changes a model's behaviour with no notice and outputs shift	Change control over suppliers, not just security

THE SCENARIO MOST BOARDS HAVE NEVER CONSIDERED

Number 6. When an agent acts using a member of staff's credentials, your logs show that person doing it. Every investigation, every disciplinary process and every regulator notification starts from that assumption. Ask for a worked answer on this one before you next approve an agent.

DON'T TRUST IT YET

Checking what the AI hands you

The plan will contain deadlines and duties. If one of them is wrong, people will follow it confidently, which is worse than having no plan at all.

The four steps

Extract	Pull out every claim that is a deadline, a duty, a definition or a figure. Prompt Three does this for you.
Source check	Go to the original: the ICO, the legislation, the regulator's own guidance. Not a blog, not a law firm's summary, and not the model's recollection of one.
Triangulate	Two independent sources that agree, or treat it as unverified. Two AI tools agreeing is one source, not two.
Challenge	Ask the model to argue the opposite: "What would make this deadline not apply to us?" The exceptions are where the risk lives.

The claims to check every time

- The 72-hour reporting duty, and the words "without undue delay" and "where feasible" that sit around it.
- When the clock starts. Models routinely write "from the incident" when the test is awareness.
- Whether individuals must be told, and the "high risk" threshold that triggers it.
- Any sector duty it adds: the DSP Toolkit, NIS regulations, FCA rules, safeguarding duties. These are where confident errors cluster.
- Any claim about what a supplier is "required" to do. Usually that requirement is in your contract, or nowhere.

WORTH SAYING OUT LOUD

Correct behaviour from a model today is not a control. It is a good outcome on one run. The control is that a named human with the right expertise signs the document off before anybody relies on it. That is the same distinction the whole video is about, applied to your own use of AI.

HINTS AND TIPS

Twelve things I would tell you in the room

Prompting

1. **Numbers beat adjectives.** "600 staff, 8,000 students" produces a different plan from "a medium-sized college".
2. **Ban the passive voice by rule, not by hope.** The naming rule in Prompt One is worth more than any other sentence in this pack.
3. **Ask for the failures separately.** Critique in the same prompt as creation is marking your own homework.
4. **Tell it to be blunt.** Without that line you get a list of "considerations". With it you get failures.
5. **Name what it missed.** The second pass is usually much better than the first, and the gap between them tells you how much checking this work needs.
6. **Keep one file.** Revise in place. Three competing versions in a chat history is how the wrong one ends up in the board pack.

Governance

7. **Write the agent register before the policy.** The register is the artefact that makes every other control possible, and it takes an afternoon.
8. **Standing authority is the single highest-value change.** One named person who can act at 3am is worth more than forty pages of policy.
9. **Put disclosure terms in the contract at renewal.** You will not get them mid-incident, and voluntary disclosure took seven months in the best-documented case we have.
10. **Define "aware" in writing.** It is the difference between a late report and a timely one, and it costs nothing to decide in advance.
11. **Rehearse out of hours at least once.** Every plan performs beautifully at 11am on a Tuesday with the whole team in the building.
12. **Report agents to the board quarterly.** Count, owners, permissions, incidents, near misses. If that list only ever grows, somebody should ask why.

AND ONE MORE

Do this exercise once with the AI, then once without it, six months later, on paper, with your team in a room. The second run is where you find out whether anybody read the first.

LIMITS

What this pack is not

- **It is not legal advice.** I am not a lawyer. The plans that follow are worked templates for a fictional organisation, produced to demonstrate a method. Your DPO and your legal adviser sign off, not me and not an AI.
- **It is not a compliance product.** Nothing here certifies anything, and no regulator has approved it.
- **It is not a substitute for security basics.** In the campaign that prompted this, one organisation's ordinary web application firewall blocked the attack outright. Patching, monitoring and backups still do most of the work.
- **It will date.** The three incidents are from September 2026 and the regulatory position moves. Check the figures before you quote them.
- **It cannot tell you what your contracts say.** Only your contracts can do that, and the exercise of reading them is the point.

IF YOU ONLY DO ONE THING

Find out who, by name, can switch off an AI agent in your organisation at three in the morning without calling a meeting. If the answer is nobody, you have found the first thing to fix, and you did not need an AI to find it.

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

APPENDIX A

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.

APPENDIX B

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.