LOGOS BOUND

Reason Binding Structures

P O L I C Y U N I T

When AI Hallucinations Become Public Decisions

The West Midlands Police Copilot Failure and What It Means for Every Public Body

February 2026

Executive Summary

On 6 November 2025, West Midlands Police banned travelling supporters from a Europa League fixture on the basis of intelligence that included a fabricated football match generated by Microsoft Copilot. The Chief Constable subsequently denied AI involvement twice before Parliament, before admitting the error. He resigned. The IOPC is investigating.[1]

This paper argues that the failure was not technological but architectural — a failure of governance, not of AI. Five structural failures are identified: no verification protocol, no provenance marking, no competence framework, no decision audit trail, and no institutional honesty. A seven-principle governance framework is proposed, together with five implementation tools (Appendices A–E).

Key findings:

- The same AI tool (Microsoft Copilot) is deployed across hundreds of UK public bodies through Microsoft 365 enterprise licences, largely without governance frameworks
- Existing statutory obligations — PACE 1984, the Equality Act 2010, the Human Rights Act 1998, UK GDPR, and administrative law — already require governance of AI outputs but are not being applied
- The Bridges v South Wales Police [2020] EWCA Civ 1058 judgment confirms that AI use by public authorities is subject to equality duties
- The real risk is not spectacular fabrication but quiet, plausible errors accumulating across thousands of decisions affecting millions of people

THE COST OF GOVERNANCE IS LOW

Provenance marking is procedural — a convention applied to existing documents.
Verification is existing professional discipline applied to a new source.
Training requires hours, not infrastructure. Competence assurance, not system rebuilds.
Audit trails require logging — the same discipline applied to every other evidence source.

This is not an AI problem. It is a governance gap.

1. The Incident

In October 2025, West Midlands Police provided intelligence to Birmingham's Safety Advisory Group ahead of a Europa League fixture between Aston Villa and Maccabi Tel Aviv on 6 November 2025. The intelligence dossier included a reference to a previous fixture between Maccabi Tel Aviv and West Ham United that had raised security concerns.[2]

The match never happened. It was fabricated by Microsoft Copilot.[3]

That fabrication was included within the intelligence material provided to a multi-agency Safety Advisory Group, which recommended banning travelling supporters from the fixture.[4] That decision affected the rights of real people on the basis of intelligence that included an event that did not exist.

The Chief Constable subsequently appeared before Parliament twice — in December 2025 and January 2026 — and on both occasions stated that AI had not been used in preparing the intelligence. On 6 January 2026, he told the Home Affairs Select Committee: *"We do not use AI."* On 12 January, he wrote to the Committee admitting that the fictitious match had arisen from Microsoft Copilot.[5]

The Home Secretary stated she had lost confidence in the Chief Constable.[6] He resigned on 16 January 2026. The Police and Crime Commissioner made a voluntary referral to the Independent Office for Police Conduct to examine the circumstances around the decision.[7] The force's own ethics panel had questioned officers about AI governance — including Copilot access and training — one month before the intelligence failure occurred.[8]

2. The Problem Is Not Hallucination

Much of the public commentary has focused on AI "hallucination" — the tendency of generative models to produce plausible but false information. This framing is understandable but insufficient. Hallucination is a known, well-documented characteristic of large language models. It is not a bug. It is a feature of how probability-based text generation works. Every major AI provider acknowledges it. Every responsible deployment framework accounts for it.

The problem in the West Midlands case was not that the AI hallucinated. The problem was that no system existed to catch it.

The failure was not technological. It was architectural. It was a failure of governance — of the structures, processes, and human oversight mechanisms that should sit between an AI output and a public decision. Responsibility for that failure sits at three levels: the individual who used the tool without verification, the institution that deployed it without governance, and the sector that adopted it without standards.

3. Five Governance Failures

3.1 No Output Verification Protocol

The fabricated match was included in an official intelligence dossier without independent verification. No officer checked whether the fixture had actually occurred. No second source was consulted. No football intelligence database — which the force maintains — was cross-referenced. The AI output was treated as fact and passed directly into a decision-making process.

> **Principle:** *Every AI-generated claim that will inform a public decision must be independently verified before inclusion. Verification is not optional. It is the minimum condition for responsible use.*

3.2 No Provenance Marking

The intelligence report did not identify which elements were AI-generated and which were derived from verified sources. When the fabrication was discovered, the force could not initially determine where it had come from. The Chief Constable attributed it first to “social media scraping,” then to “a Google search,” before eventually identifying Copilot.

> **Principle:** *AI-generated content must be marked at the point of generation, not retrospectively. Every element of a public document must carry a provenance indicator so that decision-makers know which claims rest on verified evidence and which rest on AI output.*

3.3 No Competence Framework

The officer who used Copilot did so without specific training on the limitations of generative AI. The ethics panel had raised the question of training one month earlier. No training had been delivered. The force had a written policy on Copilot access but had not ensured that officers understood the fundamental characteristics of the tool — including its propensity to generate plausible fabrications.

> **Principle:** *Any public body deploying AI tools must ensure that every user understands, at minimum: what the tool can and cannot do, how it generates output, why that output may be false, and what verification steps are required before any output is used in an official context. This is not optional training. It is a precondition of deployment.*

3.4 No Decision Audit Trail

The Safety Advisory Group made its decision on the basis of a report that included AI-generated content. There is no evidence that the SAG was informed which elements of the report were AI-assisted. The decision therefore rested on a foundation that the decision-makers could not evaluate because they did not know its composition.

> **Principle:** *Any decision informed by AI-generated content must include, in its audit trail, a clear record of what AI tools were used, what outputs they produced, what verification was conducted, and what weight was given to AI-derived material versus independently verified evidence. Without this, accountability is impossible.*

3.5 No Institutional Honesty

When the fabrication was discovered, the Chief Constable denied AI involvement — twice, under oath, before Parliament. This was not a governance failure in the technical sense. It was a failure of institutional honesty that compounded every other failure. If the use of AI had been acknowledged immediately, the force could have addressed the governance gap transparently. Instead, denial converted a technological error into a crisis of public trust that cost a Chief Constable his career and triggered an IOPC investigation.

> **Principle:** *The use of AI in any public decision must be disclosable, without qualification, at any point. If an institution cannot honestly state that AI was used, AI should not have been used. Transparency is not a post-hoc remedy. It is a structural requirement.*

4. The Copilot Paradox

There is a particular irony in this case. Microsoft Copilot is currently deployed across large parts of the UK public sector through Microsoft 365 enterprise licences. It is available to officers, social workers, teachers, health professionals, and administrators across hundreds of public bodies. Many of these users have received no specific training on generative AI. Many do not understand the difference between a search engine retrieving indexed information and a language model generating probabilistic text. Many treat Copilot output as equivalent to a database query — a reasonable-looking answer to a reasonable question.

The West Midlands case is not an outlier. It is the first visible failure in a system where invisible failures are occurring every day. Across local authorities, NHS trusts, police forces, and government departments, AI-generated content is being incorporated into reports, assessments, briefings, and recommendations without provenance marking, without verification protocols, and without the decision-makers downstream knowing that AI was involved.

The question is not whether this will happen again. It is how many times it has already happened without being detected — given the ubiquity of Copilot access across the public sector and the near-total absence of governance controls.

The underlying institutional truth is straightforward: people use AI without governance because it is fast, convenient, and invisible. It reduces administrative burden. It produces professional-looking output. And nobody is checking. The incentive to use AI ungoverned is powerful precisely because the governance that should constrain it does not yet exist. Until the cost of ungoverned use exceeds the cost of compliance, the pattern will continue.

5. A Governance Framework for AI in Public Decision-Making

The following framework is proposed for any public body using AI tools in any capacity that may inform, support, or contribute to decisions affecting the rights, safety, or welfare of individuals.

For the purposes of this framework, a “public decision” is any determination by a public body that affects an individual’s rights, liberty, access to services, safeguarding status, enforcement action, eligibility, or restrictions on freedom — including decisions informed by intelligence, assessments, reports, or recommendations that contain AI-generated content.

This framework treats governance as a pre-condition of decision validity, not a downstream review function.

> **MINIMUM STANDARD**
>
> Before any AI-generated factual claim is included in an official record, the following must be in place as an irreducible minimum: (1) a provenance tag identifying the content as AI-derived; (2) independent verification from a non-AI source; (3) a recorded verification source; and (4) sign-off by a named officer. This is the floor. The principles below build upward from it.

Principle 1: Human Authority at the Boundary

AI may generate, draft, summarise, or analyse. It may not decide. Every AI output that will inform a public decision must pass through a human authority checkpoint before it enters the decision-making process. That checkpoint must involve a named individual who takes personal responsibility for the accuracy of the material they are passing forward.

Principle 2: Mandatory Provenance Marking

All AI-generated content must be marked as such at the point of creation. Documents containing AI-generated material must identify which elements are AI-derived and which are independently verified. This marking must be visible to every person in the decision chain, not just the originating author.

Principle 3: Verification Before Inclusion

No AI-generated factual claim may be included in any official document, report, assessment, or recommendation without independent verification from a non-AI source. The verification must be recorded and auditable. “The AI said so” is not evidence. It is a hypothesis requiring confirmation.

Principle 4: Competence as a Precondition

Deployment of AI tools must be preceded — not followed — by competence assurance. Every user must demonstrate understanding of the tool’s capabilities, limitations, and failure modes before being granted access. Annual refresher training must account for model updates and emerging risks. Competence is not awareness. It is the ability to identify when AI output is unreliable and to act accordingly.

Principle 5: Decision Audit Integrity

The audit trail for any public decision must record whether AI tools were used, what they produced, what was verified, what was discarded, and what weight was given to AI-derived material. This record must be available for inspection by oversight bodies, complaints investigators, courts, and the individuals affected by the decision.

Principle 6: Institutional Transparency

Public bodies must be able to state, at any point and without qualification, whether AI was used in any decision or process. Policies that permit AI use but discourage disclosure — whether through cultural pressure, reputational concern, or absence of recording requirements — are incompatible with public accountability.

Disclosure operates at three levels: (1) disclosure to decision-makers internally, which must occur in every case where AI-generated content informs a decision; (2) disclosure to affected persons where the decision engages their rights, which must occur where those rights may be affected; and (3) disclosure to oversight bodies on request, which must be available without qualification at any time.

Principle 7: Proportionality to Consequence

The level of governance must be proportionate to the consequence of the decision. An AI-assisted summary of a routine meeting requires less oversight than an AI-assisted intelligence dossier informing a decision to restrict individuals' rights. Public bodies must classify decisions by consequence and apply governance requirements accordingly. The higher the stakes, the higher the verification threshold.

Risk Categories for AI-Assisted Decisions

Risk	Description	Examples	Governance Required
HIGH	Decision directly affects individuals' rights, safety, welfare, or liberty	Intelligence assessments, social work assessments, benefits determinations, safeguarding decisions, court submissions	Full verification, provenance marking, statutory duty assessment, named sign-off, complete audit trail
MEDIUM	Decision informs or supports a rights-affecting process but is not the final determination	Briefing papers, background research, case summaries for review, internal policy analysis	Verification of factual claims, provenance marking, decision-maker informed of AI involvement
LOW	Administrative or formatting task with no direct impact on individuals' rights	Meeting summaries of known content, template generation, formatting, internal scheduling	Mark as [AI-DRAFTED], no further verification required

6. The PACE Question Nobody Has Asked

The Police and Criminal Evidence Act 1984 was enacted precisely because public confidence in policing had collapsed following a series of cases in which unreliable evidence had been used to justify decisions that affected people's rights. PACE established the

fundamental principle that evidence must be obtained properly and must be reliable. Where evidence is obtained improperly or is unreliable, it is inadmissible. Where reasonable suspicion is required, it must be grounded in objective facts, intelligence, or information — not assumption, generalisation, or fabrication.

PACE Code A requires that stop and search powers are exercised on the basis of "up-to-date and accurate intelligence or information."[9] The Code explicitly states that "reasonable suspicion can never be supported on the basis of personal factors. It must rely on intelligence or information about, or some specific behaviour by, the person concerned."

The Safety Advisory Group decision in the West Midlands case was an administrative decision, not an exercise of PACE search or arrest powers. However, the PACE framework establishes principles of evidential reliability that extend beyond the specific powers it governs. Where policing decisions of any kind rest on intelligence, the standards of accuracy and reliability that PACE embodies represent the minimum governance expectation for a modern police force.

Apply this framework to the West Midlands case. The intelligence dossier contained a fabricated match generated by a probability model. That fabrication was not "up-to-date and accurate intelligence." It was not intelligence at all. It was a hallucination — plausible text generated by a system that has no concept of truth, no access to verified records, and no capacity to distinguish fact from invention. Under PACE principles, it should never have entered the decision chain.

The question nobody has yet asked is this: does AI-generated content constitute "intelligence" or "information" within the meaning of PACE and its associated governance principles?

If it does, then it must meet the same standards of accuracy and reliability that PACE imposes on all intelligence — and in the West Midlands case, it manifestly did not.

If it does not, then any decision resting on AI-generated content is, by definition, not grounded in intelligence or information as PACE understands it — and any power exercised on that basis may be unlawful.

Either way, PACE demands a governance response. The Act was designed for a world in which intelligence was gathered, recorded, and verified by human officers. It did not anticipate a world in which an officer could generate plausible-looking intelligence by typing a question into a probability engine and accepting whatever it produced. The Codes of Practice must be updated to address AI-generated content explicitly — not as an afterthought, but as a structural requirement for any force that deploys generative AI tools.

PACE was Parliament's answer to unreliable evidence in the 1980s. Four decades later, the source of unreliable evidence has changed. The principle has not. If anything, the risk is greater now — because AI fabrications do not look like fabrications. They look like facts. And without governance, they will be treated as facts. By officers. By Safety Advisory Groups. By courts. And by every institution that trusts the output of a machine that cannot tell the truth because it does not know what truth is.

7. The Statutory Framework Already Exists

The absence of a dedicated UK AI Act does not mean AI in public decision-making is unregulated. The West Midlands case engaged multiple existing statutory frameworks, none of which were complied with. Public bodies cannot treat AI as operating in a regulatory vacuum. The following legislation applies now, without amendment, to every public body using AI tools in decision-making.

The Equality Act 2010

Section 149 imposes the Public Sector Equality Duty, requiring public authorities to have due regard to the need to eliminate discrimination and advance equality of opportunity before implementing policies or taking decisions. Where AI-generated content informs a decision that disproportionately affects a community likely to share protected characteristics — as in the West Midlands case, where the ban disproportionately affected supporters of an Israeli football club, engaging potential race and ethnic origin considerations — the absence of an Equality Impact Assessment for the AI tool raises a serious question as to compliance with the PSED.[10] The Equality and Human Rights Commission has confirmed that algorithmic systems can perpetuate or amplify discrimination, and that even unintentional bias can breach the Act if it results in indirect discrimination. The Bridges v South Wales Police [2020] EWCA Civ 1058 judgment confirmed that the use of AI by public authorities is subject to equality duties and requires adequate safeguards.[11]

The Human Rights Act 1998

Article 8 ECHR (right to respect for private and family life), Article 6 (right to a fair hearing), and Article 14 (prohibition of discrimination) are all engaged where AI-generated content informs decisions affecting individuals' rights. The ban on travelling supporters engaged Article 11 (freedom of assembly) and Article 14 (discrimination). Where public authorities use AI outputs without verification, transparency, or audit, the procedural safeguards required by the Convention are absent. The Human Rights Act requires that any interference with Convention rights is lawful, necessary, and proportionate — a test that cannot be met when the evidential basis for the interference includes fabricated information.

UK GDPR and the Data Protection Act 2018

Article 22 of UK GDPR provides that individuals have the right not to be subject to decisions based solely on automated processing that produce legal or similarly significant effects. In the West Midlands case, human decision-makers used an AI-contaminated report, meaning Article 22 may not apply directly. However, Articles 13 and 14 impose broader transparency obligations: organisations must provide meaningful information about the logic involved in any processing that affects individuals, including mixed human-AI workflows. Where AI-generated content is incorporated into intelligence reports without provenance marking, the individuals affected cannot exercise their data protection rights because they do not know that AI was involved. AI-assisted evidential content triggers transparency duties even where Article 22 does not strictly apply. The Data (Use and Access) Act 2025, when fully commenced, will require that safeguards including the right to human intervention, the ability to contest decisions, and transparency about the logic and criteria used are in place for all automated decision-making.

Administrative Law and Judicial Review

Public law principles of fairness, rationality, and procedural propriety apply to all public authority decisions. A decision informed by AI-fabricated evidence is vulnerable to judicial review on grounds of irrationality — a decision-maker who relies on false information has failed to take into account relevant considerations and has taken into account irrelevant ones. The duty to give reasons, the duty to act fairly, and the principle of Wednesbury reasonableness all require that the evidential basis for public decisions is reliable. AI-generated content that has not been verified does not meet this standard.

The Algorithmic Transparency Recording Standard

While currently voluntary, the ATRS represents an emerging public sector standard requiring public bodies to publish details of algorithmic and AI systems used in decision-making. The Public Authority Algorithmic and Automated Decision-Making Systems Bill [HL], introduced in September 2024, proposes mandatory requirements including impact assessments, transparency records, logging capabilities, and compliance with the Equality Act and Human Rights Act. Whether or not this Bill becomes law, it signals the direction of travel — and the West Midlands case demonstrates precisely why such requirements are necessary.

The Combined Effect

No single Act governs AI in public decision-making. But taken together, PACE, the Equality Act, the Human Rights Act, UK GDPR, and administrative law principles create a comprehensive framework that the West Midlands Police failed to engage with at any point. The governance gap is not in the law. It is in the institutions that have adopted AI tools without mapping them against the statutory duties they already owe.

Every public body in the United Kingdom already has the legal obligations necessary to govern AI responsibly. What is missing is the institutional awareness that these obligations apply to AI outputs in exactly the same way they apply to every other form of evidence, intelligence, and information that informs public decisions.

8. The Wider Implications

The West Midlands Police case involved policing. But the governance failures it exposed are not specific to policing. They are structural features of how AI has been adopted across the public sector: rapidly, cheaply, without training, without governance, and without the institutional frameworks to catch errors before they become decisions.

The same Copilot tool is available to social workers writing assessments, to teachers preparing reports, to housing officers making allocation decisions, to benefits assessors determining entitlements, and to planning officers evaluating applications. In each of these contexts, an AI hallucination that passes unchecked into an official document has the potential to affect the rights and welfare of real people.

The West Midlands case was detected because the fabricated match was verifiably false — a binary fact that could be checked. But many AI errors are not binary. They are errors of emphasis, of framing, of omission, of contextual misunderstanding. A social work

assessment that underweights a risk factor. A housing report that mischaracterises a household's circumstances. A benefits decision that incorporates an inaccurate summary. These errors may never be detected because they are plausible enough to pass through human review, particularly when the reviewer is overworked, time-pressured, and trusting of the tool.

This is the real risk. Not the spectacular, headline-generating fabrication of a football match. But the quiet, invisible, plausible errors that accumulate across thousands of decisions in hundreds of public bodies, affecting millions of people who will never know that AI was involved in the process that determined their outcome.

9. Conclusion

The West Midlands Police Copilot failure is not an AI story. It is a governance story. The technology did exactly what it was designed to do — generate plausible text based on probabilistic patterns. The failure was entirely human: the failure to verify, the failure to mark, the failure to train, the failure to audit, and the failure to tell the truth.

These are not complex problems. They are not expensive to solve. They require policy, not technology. They require leadership, not legislation. They require the recognition that AI is a tool — powerful, useful, and fundamentally incapable of accountability. Accountability belongs to the humans who deploy it, the institutions that adopt it, and the governance frameworks that constrain it.

Every public body in the United Kingdom that currently has Microsoft Copilot, or any generative AI tool, available to its staff should ask itself one question: if a Parliamentary Committee asked us tomorrow whether we use AI and how we govern it, could we answer honestly, completely, and without embarrassment?

If the answer is no, the governance framework does not yet exist. And every decision made in that gap carries the risk of becoming the next West Midlands.

APPENDIX A: AI Output Verification Checklist

For use by any public sector officer before including AI-generated content in an official document, report, assessment, or recommendation.

Before including any AI-generated content, complete all steps below. Retain this checklist as part of the decision audit trail.

Step	Action	Done	Initials	Date
1	Identify the AI tool used (name, version, deployment type)	☐		
2	Record the prompt or query entered	☐		
3	Record the full AI output received	☐		
4	For each factual claim, identify a non-AI source that independently confirms it	☐		
5	Record the verification source for each factual claim	☐		
6	Remove or flag any claim that cannot be independently verified	☐		
7	Mark all AI-derived content with provenance indicator [AI-ASSISTED]	☐		
8	Confirm decision-maker informed which elements are AI-derived	☐		
9	Assess statutory duties engaged: PACE / Equality Act / HRA / UK GDPR / Sector-specific	☐		
10	Confirm AI-derived content meets standard required by relevant legislation	☐		
11	Sign off: all content verified, marked, and decision-makers informed	☐		

Signed: _____ Date: _____

Role/Rank: _____ Organisation: _____

Data Retention Note

Completed checklists, recorded prompts, and AI outputs retained under this protocol should be held in accordance with the organisation's existing data retention schedule and UK GDPR requirements. Where prompts or outputs contain personal data, retention must be proportionate to the decision they informed. Checklists supporting high-risk decisions (see risk categorisation table) should be retained for the same period as the decision record itself. Organisations should consult their Data Protection Officer on retention periods for AI audit material.

APPENDIX B: AI Provenance Marking Standard

Recommended marking convention for public sector documents containing AI-generated or AI-assisted content.

Document-Level Marking

Every document containing AI-generated or AI-assisted content must include a disclosure statement identifying the AI tool used, confirming that AI-derived elements are marked, and providing a reference to the verification record.

Paragraph-Level Marking

[AI-ASSISTED] — Content generated by AI and subsequently verified by a named officer.

[AI-DRAFTED] — Structural or narrative content drafted by AI where factual accuracy is not the primary concern.

[AI-UNVERIFIED] — AI-generated content included for reference only, not independently verified.

Prohibited Categories

The following must never be generated by AI, regardless of verification: direct witness statements or testimony; forensic evidence summaries; medical or clinical opinions; legal advice or legal conclusions; risk assessment judgements requiring professional expertise; any content purporting to represent the words, views, or professional opinion of a named individual.

APPENDIX C: AI Competence Framework for Public Sector Staff

Minimum competence requirements before any officer or staff member may use generative AI tools in an official capacity.

Level 1: AI Awareness (All Staff)

Competence	Evidence of Achievement
Can explain in plain language what generative AI is	Written or verbal assessment
Understands AI generates probabilistic text, not verified facts	Written or verbal assessment
Can define “hallucination” in the AI context	Written or verbal assessment
Understands AI output may be plausible but factually wrong	Written or verbal assessment
Knows the organisation’s AI usage policy	Signed acknowledgement
Knows how to report suspected AI-related errors	Signed acknowledgement

Level 2: AI User (Authorised Staff)

Competence	Evidence of Achievement
All Level 1 competences achieved	Certification
Can demonstrate the AI Output Verification Checklist	Practical assessment
Can correctly apply the Provenance Marking Standard	Practical assessment
Can identify statutory duties engaged by AI-assisted decisions	Written assessment
Understands distinction: AI-assisted admin vs AI-generated evidence	Written assessment
Completed supervised AI exercise with verification, marking, disclosure	Supervisor sign-off

Level 3: AI Governance Lead (Senior Staff)

Competence	Evidence of Achievement
All Level 1 and Level 2 competences achieved	Certification
Can design and implement AI governance policy	Policy document
Can conduct or commission EIA for AI deployment	EIA document
Understands PACE, Equality Act, HRA, UK GDPR, admin law re: AI	Written assessment
Can advise senior leadership on AI-related risk	Advisory report
Can respond to Parliamentary, regulatory, or FOI enquiries on AI	Practical assessment

Maintains organisation's AI audit trail and provenance records	Ongoing

All staff at Levels 1–3 must complete annual refresher training accounting for tool updates, new case law, lessons learned, and policy changes.

APPENDIX D: AI Governance Decision Flowchart

For use at the point of decision to determine whether AI-generated content may be included in an official process.

Step 1

Is the content being used for an administrative or formatting purpose only?

- YES → Mark as [AI-DRAFTED]. No further verification required. Proceed to Step 6.
- NO → Continue to Step 2.

Step 2

Does the content contain factual claims?

- YES → Continue to Step 3.
- NO → Mark as [AI-DRAFTED]. Proceed to Step 6.

Step 3

Can each factual claim be independently verified from a non-AI source?

- ALL VERIFIED → Mark as [AI-ASSISTED]. Record verification sources. Proceed to Step 4.
- PARTIALLY → Remove unverified claims. Mark accordingly. Proceed to Step 4.
- **NONE VERIFIED → DO NOT INCLUDE in official document. Record for audit only. STOP.**

Step 4

Does the decision this content will inform engage statutory duties?

- YES → Complete statutory duty assessment. Confirm content meets required standard. Proceed to Step 5.
- NO → Proceed to Step 5.

Step 5

Has the decision-maker downstream been informed which elements are AI-derived?

- YES → Proceed to Step 6.
- NO → Inform the decision-maker before the content is used.

Step 6

Complete AI Output Verification Checklist (Appendix A). Retain as part of decision audit trail. PROCEED.

APPENDIX E: Model AI Governance Policy for Public Bodies

Template policy for adoption by any public authority. Adapt to local context.

1. Purpose

This policy establishes the governance framework for the use of artificial intelligence tools by staff in the course of their official duties.

2. Scope

This policy applies to all staff, contractors, and agency workers who use AI tools in any capacity that may inform, support, or contribute to decisions affecting the rights, safety, or welfare of individuals.

3. Principles

3.1 Human Authority: AI may assist but not determine. Every AI output that informs a decision must be reviewed, verified, and approved by a named individual.

3.2 Verification: No AI-generated factual claim may be included in any official document without independent verification from a non-AI source.

3.3 Provenance: All AI-generated or AI-assisted content must be marked using the Provenance Marking Standard or equivalent.

3.4 Transparency: The use of AI must be disclosable at any point without qualification.

3.5 Competence: No member of staff may use AI tools for official purposes without first achieving the required competence level.

3.6 Proportionality: The level of governance must be proportionate to the consequence of the decision it informs.

3.7 Audit: A complete audit trail must be maintained for all AI-assisted work.

4. Prohibited Uses

AI tools must not be used to: generate content purporting to be professional opinion or witness testimony of a named individual; produce risk assessments or safeguarding decisions without full professional oversight; create content for courts or Parliamentary committees without full AI disclosure; replace professional judgement in statutory assessments; or process personal data constituting automated decision-making under Article 22 UK GDPR without required safeguards.

5. Responsibilities

All Staff: Comply. Report errors. Complete training.

Line Managers: Ensure compliance and verify competence.

AI Governance Lead: Maintain policy, training, audit trail, and external enquiry response.

Senior Leadership: Resource governance adequately. Establish transparency culture.

6. Breach

Failure to comply may constitute a disciplinary matter. Governance failures resulting in harm will be referred to the appropriate professional body or oversight mechanism.

7. Review

Annual review, or sooner if required by changes to AI tools, legislation, or case law.

Approved by: _____ Date: _____

Position: _____

Sources and Notes

[1] Summary drawn from public reporting by The Guardian, Reuters, BBC News, and the published record of the Home Affairs Select Committee hearings, December 2025 and January 2026.

[2] Birmingham Safety Advisory Group, meeting ahead of Aston Villa v Maccabi Tel Aviv, Europa League, 6 November 2025.

[3] The fabricated fixture — Maccabi Tel Aviv v West Ham United — was generated by Microsoft Copilot and did not correspond to any match in UEFA, FA, or domestic football records. Initially attributed to "social media scraping" then "a Google search" before Copilot was identified.

[4] The SAG recommended, and the force implemented, a ban on travelling Maccabi Tel Aviv supporters. The match took place on 6 November 2025 with no away supporters present.

[5] Home Affairs Select Committee, oral evidence December 2025 and 6 January 2026. Written correction from the Chief Constable dated 12 January 2026.

[6] Home Secretary Shabana Mahmood, public statement following the findings of HMICFRS (His Majesty's Inspectorate of Constabulary and Fire & Rescue Services), led by Sir Andy Cooke, HM Chief Inspector of Constabulary. The inspectorate's review identified multiple inaccuracies in the force's intelligence reporting and a pattern consistent with confirmation bias.

[7] Chief Constable Craig Guildford resigned 16 January 2026. PCC Simon Foster made a voluntary IOPC referral.

[8] West Midlands Police ethics panel session, approximately one month before the intelligence failure.

[9] PACE Code A, paragraph 2.2. Code A governs stop and search and requires reasonable suspicion grounded in objective intelligence or information.

[10] The PSED requires "due regard" — a procedural obligation to consider equality implications before decisions. Whether an absence of EIA constitutes a breach depends on circumstances, but disproportionate impact without any equality consideration raises a strong inference of non-compliance.

[11] R (Bridges) v Chief Constable of South Wales Police [2020] EWCA Civ 1058. The Court of Appeal held that automated facial recognition by police engaged the PSED and required adequate legal framework and safeguards.

LOGOS Bound is undertaking a programme of research into AI governance across public services. Professionals and practitioners with relevant experience will be contacted directly and invited to contribute.