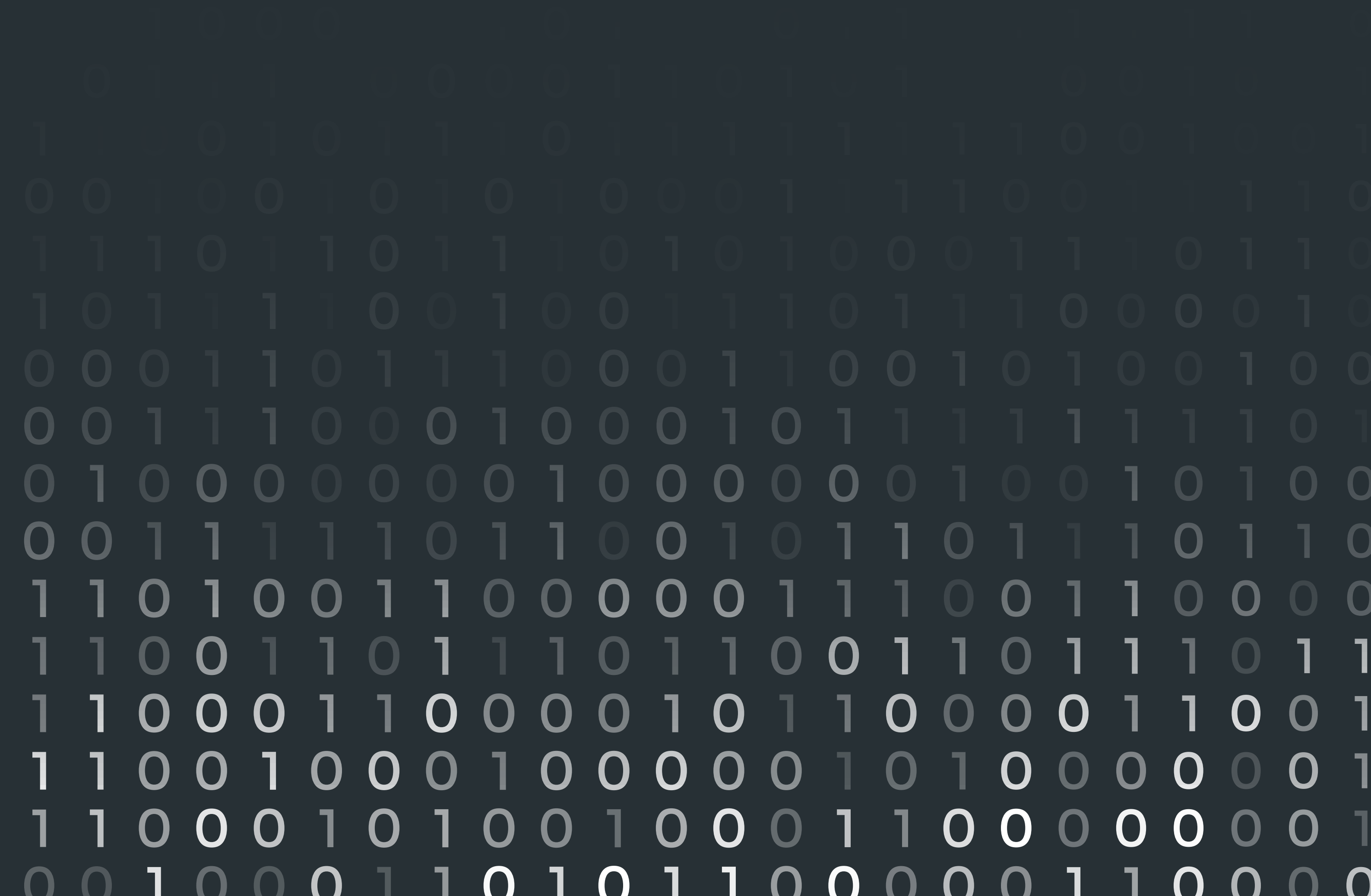
trowers & hamlins

Governance of AI in the Public Sector

A framework for safe, ethical and responsible deployment

Contents



Executive summary

The integration of artificial intelligence into public sector operations represents both an unprecedented opportunity and a significant governance challenge. Public bodies are increasingly adopting AI to improve service delivery, decision-making and productivity whilst facing a set of risks that demand clear frameworks, proactive oversight and strong organisational leadership.

This is fundamentally about building organisational capability rooted in governance, accountability and public trust. Ethical AI is not simply a technical project – it is a whole-organisation capability, rooted in governance, accountability and public trust.

This white paper addresses the critical governance issues associated with the safe and ethical rollout of AI in the public sector, with particular focus on:

- The UK regulatory regime and its principles-based approach.
- Government policy frameworks and constitutional interfaces.
- Internal governance structures and policy development.
- Public sector legal obligations and equality duties.
- Practical frameworks for risk-based AI governance.
- Essential procurement and contracting safeguards.

The challenges that AI presents to local government are not transient difficulties that will resolve as technology matures. Rather, they represent a fundamental shift in how public services are delivered and how authorities must approach governance and accountability.

Why AI governance matters for public bodies

We consistently observe that organisations are responding to accelerating service demands, pressures from legacy systems, increasing cyber risks, heightened expectations around fairness and equalities duties. There is also a growing need to demonstrate transparency to employees, customers and citizens.

AI is shifting from pilot experiments to core business capability. To realise the benefits from its use whilst protecting people and reputations, organisations must treat ethical AI as a whole-organisation capability.

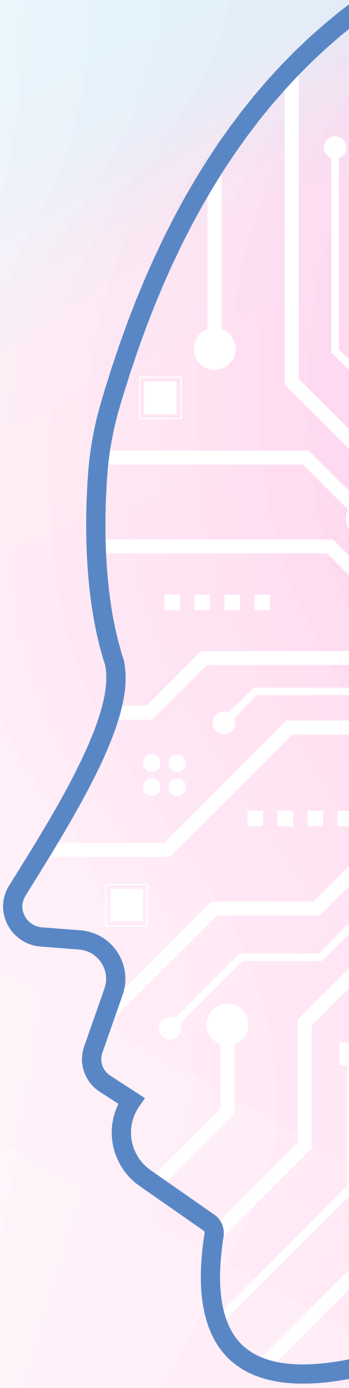
Public sector organisations face unique governance challenges:

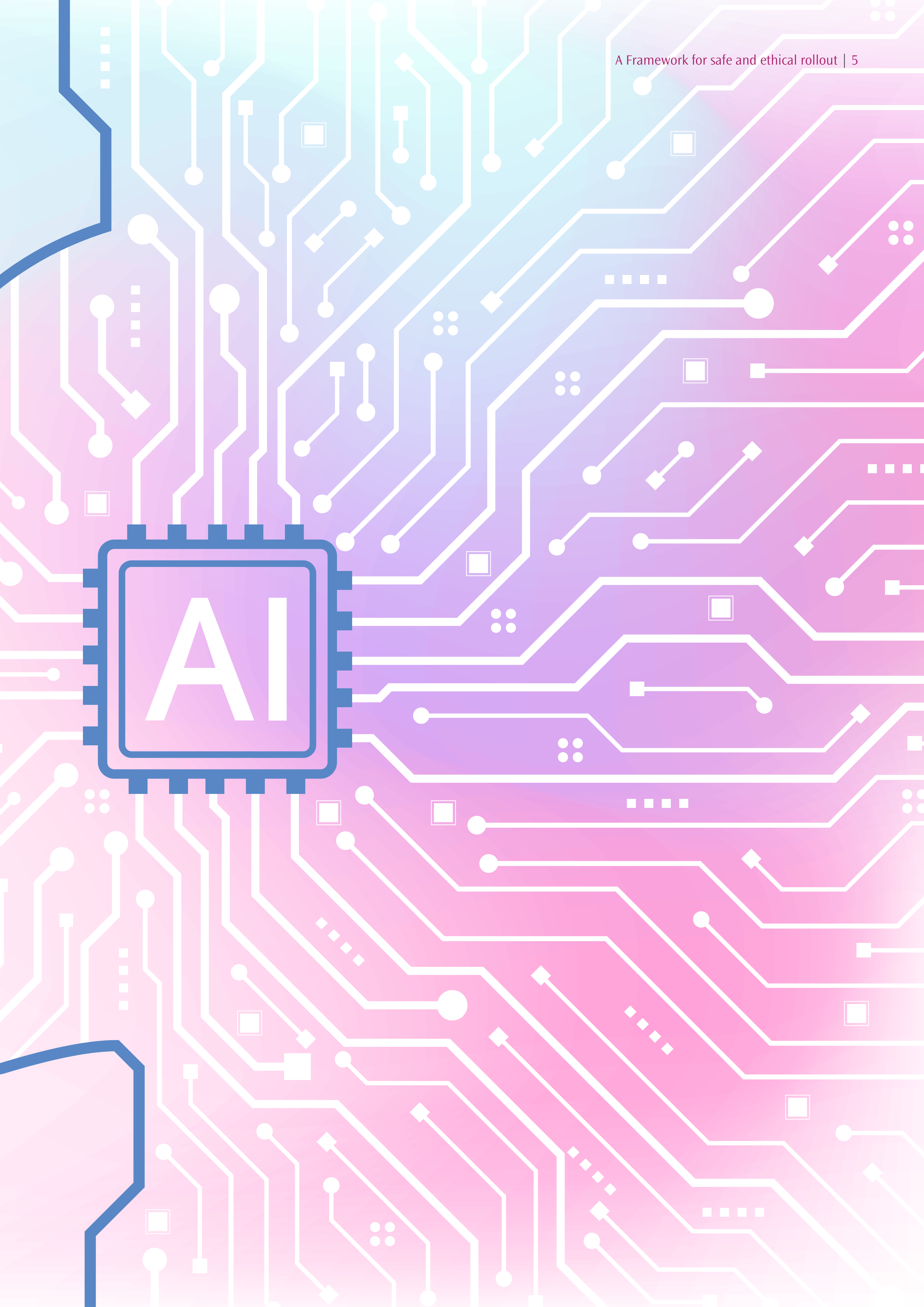
- **Constitutional duties:** Decisions remain subject to Wednesbury principles, procedural fairness duties and the legitimate expectation doctrines.
- **Public accountability:** Citizens have a legitimate expectation that public decisions will be fair, transparent and subject to challenge.
- **Equality obligations:** Public bodies subject to the Public Sector Equality Duty face particular risks from AI's well-documented algorithmic bias issues. AI systems are prone to algorithmic bias, as demonstrated by healthcare AI showing reduced accuracy for women and minority groups and a Department for Work and Pensions fraud detector disproportionately flagging certain nationalities.
- **Standard-setting role:** Public sector organisations are expected to embed AI principles into procurement and operations to build accountability and trust, as they are not just procuring AI for efficiency or cost savings, but demonstrating to the wider economy what responsible AI deployment looks like, and when public sector organisations procure AI systems with robust bias testing, meaningful explainability, and genuine human oversight, they are setting standards that influence private sector practice.

The whole-organisation capability model

To realise benefits whilst protecting people and reputations, organisations must treat ethical AI as a whole-organisation capability encompassing policy, procurement, risk, legal, technology and organisational change.

Effective AI governance cannot be siloed within a single department. Legal teams should lead AI governance by adopting risk-based approaches and encouraging collaboration across the organisation, positioning themselves as AI governance leaders rather than reactive advisers.




AI

The UK regulatory and policy landscape

The UK's principles-based approach

The UK's AI governance relies on five key principles:

1. Safety
2. Transparency
3. Fairness
4. Accountability
5. Contestability

Prescriptive AI Act, which adopts a risk based approach based on potential harm an activity or technology poses to health, safety, and fundamental rights.

Multi-regulator oversight

Rather than creating a single AI regulator, multiple existing regulators like ICO, Equality and Human Rights Commission, CMA, Ofcom and Medical & Healthcare Regulator oversee AI compliance in their specific sectors. For public bodies, this means understanding which regulators' guidance applies to your specific AI use case. The ICO will be concerned with data protection, the EHRC with equality and discrimination and other bodies will cover their respective area.

Flexibility and responsibility

The principles-based approach gives you flexibility to tailor AI governance to your specific context, but it also places responsibility on you to think critically about what these principles mean for your specific AI procurement and how you will ensure they are embedded throughout the contract lifecycle.

Buyers and operators should embed these principles into procurement specifications and contracts, map regulatory touchpoints at the use-case level and plan for continuous assurance as systems and risks evolve.

Key policy frameworks

Key documents such as the AI Regulation White Paper and National AI Strategy guide responsible AI innovation and ethics. The Government Digital Service has also published detailed guidance on responsible AI in government procurement that provides valuable frameworks public bodies should consider. The LGA also runs its AI Hub and published guides on buying AI responsibility and other guidance. The Algorithmic Transparency Recording Standard encourages public bodies to disclose their use of AI in decision-making and whilst compliance is currently voluntary, there is increasing expectation that public bodies will adopt it.

Essential policy documents:

- UK Government White Paper: A pro-innovation approach to AI regulation (March 2023)
- Implementing the UK's AI Regulatory Principles (Guidance for Regulators, February 2024)
- National AI Strategy (September 2021)
- Algorithmic Transparency Recording Standard (ATRS) Hub
- Government Digital Service guidance on responsible AI procurement
- LGA guidance on responsible AI procurement

Constitutional and public law considerations

Traditional public law principles apply

Decisions remain subject to Wednesbury principles, procedural fairness duties and the legitimate expectation doctrines.

AI deployment does not suspend constitutional safeguards. Public authorities must ensure that:

- Rationality: AI-informed decisions must be rational and based on relevant considerations.
- Procedural fairness: Individuals affected by AI decisions must have fair procedures, including the right to be heard and to understand the basis of decisions.
- Legitimate expectations: Where authorities have created legitimate expectations about how decisions will be made, AI deployment cannot circumvent those expectations without proper consultation and justification.

The transparency imperative

Transparency in AI-driven public decision-making is a legal obligation flowing from multiple statutory requirements. Yet, AI systems often operate as "black boxes". People must understand, question and challenge decisions affecting their lives.

Public bodies have heightened transparency obligations flowing from:

- Human Rights Act 1998 (right to fair trial and effective remedy).
- Freedom of Information Act 2000.
- Common law duties of procedural fairness.
- Increasingly, the Algorithmic Transparency Recording Standard.

The Public Sector Equality Duty

Public bodies subject to the Public Sector Equality Duty face particular risks from AI's well-documented algorithmic bias issues, with the legal framework being clear that the Public Sector Equality Duty, the Equality Act 2010 with its nine protected characteristics, and the GDPR fairness principle all apply – AI causing "unjust discrimination" violates GDPR.

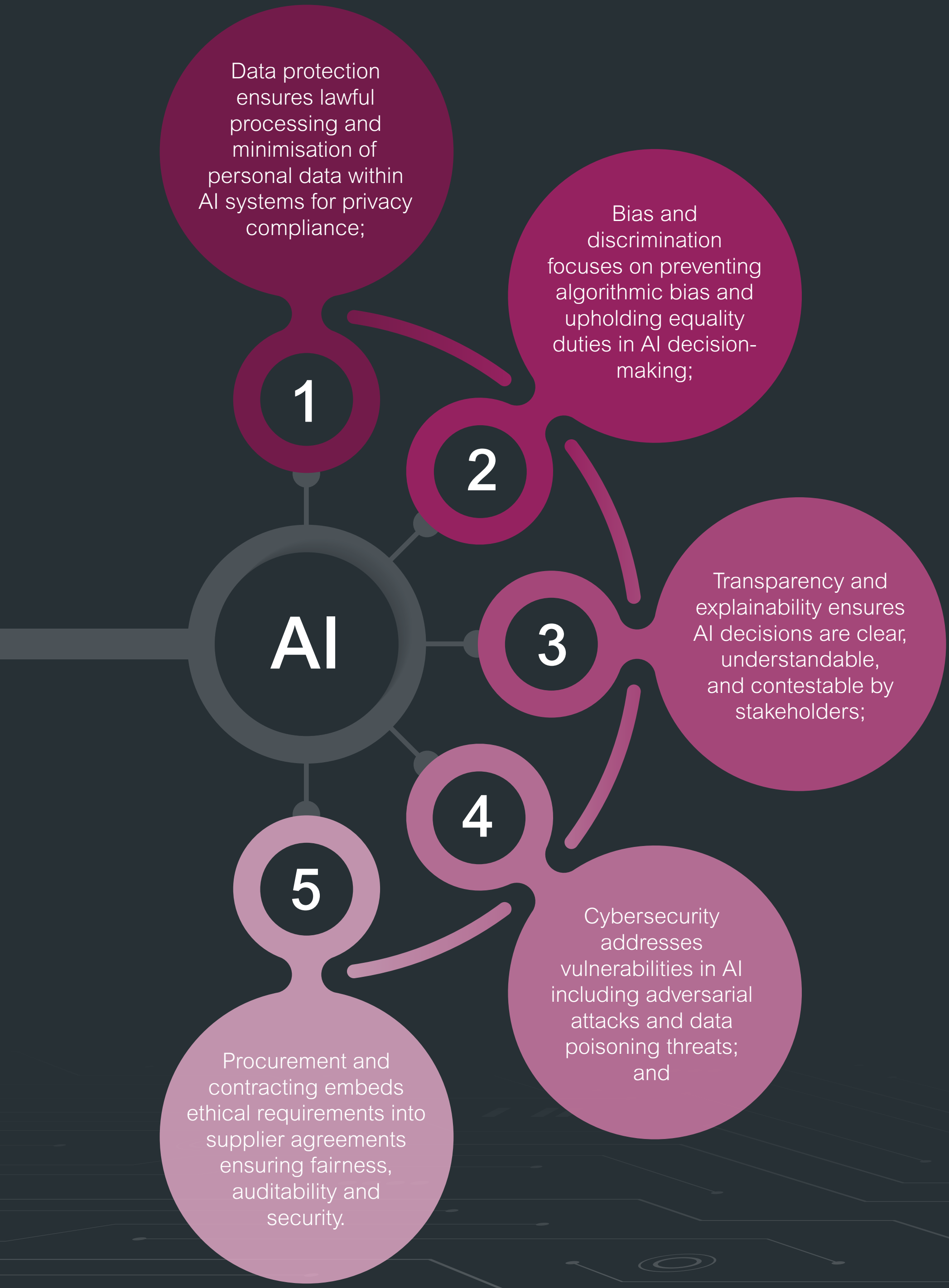
The Public Sector Equality Duty requires public authorities to have due regard to the need to:

- Eliminate unlawful discrimination, harassment and victimisation.
- Advance equality of opportunity between different groups.
- Foster good relations between different groups.

This duty applies to AI procurement and deployment, requiring public bodies to proactively assess and mitigate discriminatory impacts before systems are deployed.

Five critical domains of AI governance

The ethical use of AI in public services requires attention to five critical domains, each presenting distinct challenges and legal obligations:



Data protection ensures lawful processing and minimisation of personal data within AI systems for privacy compliance;

Bias and discrimination focuses on preventing algorithmic bias and upholding equality duties in AI decision-making;

Transparency and explainability ensures AI decisions are clear, understandable, and contestable by stakeholders;

Cybersecurity addresses vulnerabilities in AI including adversarial attacks and data poisoning threats; and

Procurement and contracting embeds ethical requirements into supplier agreements ensuring fairness, auditability and security.

1. Data protection

The challenge

The deployment of AI systems within local authorities introduces multifaceted risks that require systematic mitigation strategies. With data protection compliance at the forefront where the UK data protection regime creates stringent obligations, that many AI applications struggle to satisfy, including UK GDPR requirements for a lawful basis such as 'public task' and mandatory DPIAs for AI of processing personal data.

Robust DPIAs

These assessments cannot be perfunctory box-ticking exercises, as risks include unlawful data processing, superficial DPIAs, excessive data collection and lack of transparency. A DPIA is required to be robust describing the processing, assessing necessity and identifying risks.

Your DPIA must genuinely interrogate: what AI system is being deployed, what it does, and what data it processes; why the AI system needed, whether it is proportionate to the intended benefits, whether the same outcome be achieved with less intrusive means; and what are the risks to individuals, including discrimination or bias, inaccurate predictions leading to inappropriate interventions, privacy intrusion from processing sensitive data and a lack of transparency making it difficult for individuals to understand or challenge decisions.

Lawful basis

You need a clear lawful basis for AI-enhanced operations and whilst public task typically provides the foundation for most public sector functions, AI systems often require additional personal data or novel processing activities that stretch beyond traditional service delivery. This means public bodies must carefully assess whether their existing legal basis genuinely covers AI-enhanced operations or whether supplementary justification is required.

Data minimisation

The data minimisation principle is particularly challenging – AI must process only necessary data, challenging its need for large datasets to comply with minimisation. You must use the minimum data necessary and be able to justify why each data element is necessary for your AI system. AI systems often want as much data as possible to improve accuracy, but GDPR requires you to process only the minimum necessary.

Special category data

Special category data requires additional safeguards. This includes data revealing racial or ethnic origin, political opinions, religious beliefs, health data, biometric data and so on. You need both a lawful basis under Article 6 and a separate condition under Article 9. Many AI systems in social care, education, or health-related services will process special category data, so this is not a theoretical concern.

Apply enhanced safeguards for special category data; align retention, access controls and deletion schedules.

Practical requirements

For each AI system, authorities must:

- Complete comprehensive DPIAs before deployment.
- Establish a clear lawful basis under Article 6 GDPR.
- For special category data, identify additional Article 9 condition.
- Document necessity for each data element processed.
- Implement proportionate retention and deletion schedules.
- Establish strict access controls.
- Maintain clear data processing agreements with suppliers.

2. Bias and discrimination

The scale of the problem

AI systems are prone to algorithmic bias. Healthcare AI has shown reduced accuracy for women and minority groups. A DWP fraud detector disproportionately flagged certain nationalities. These are real failures with real consequences: legal liability, regulatory action, and harm to individuals.

The legal framework

The legal framework is clear, with the Public Sector Equality Duty, the Equality Act 2010 with its nine protected characteristics, and the GDPR fairness principle all applying – AI causing "unjust discrimination" violates GDPR.

Mitigation strategies

- **Data auditing** – Audit training data and models for representational imbalance; test for disparate impact across all protected characteristics.
- **Comprehensive bias testing** – Contracts must require comprehensive bias testing before deployment and regularly throughout the term, covering all nine protected characteristics – not just race and gender – using appropriate statistical methods with suppliers providing detailed bias testing reports.

  Mandate supplier bias testing prior to deployment and at defined intervals; require reporting and remediation triggers.

Ongoing monitoring

Just as important is ongoing monitoring of AI decisions by protected characteristic groups. Supply contracts should require regular reports analysing AI decisions by protected characteristic groups and statistical analysis showing whether outcomes differ across groups. When disparities emerge, public bodies must be ready to act, whether through algorithmic adjustment, additional human oversight or system suspension.

Combating automation bias

Even when humans are involved in reviewing AI decisions, you must combat automation bias through training and culture. Automation bias is the tendency to over-rely on automated systems and accept their recommendations without sufficient critical evaluation. This happens because humans assume AI is more accurate than it actually is, feel pressure to process decisions quickly, lack confidence to override "sophisticated" AI systems and performance metrics reward speed over quality

Public bodies must actively combat automation bias through training that emphasises human responsibility and accountability, performance metrics that value good decision-making rather than speed alone, a culture that empowers staff to override AI when appropriate, and regular reviews of decision quality not just the volume of decisions.

Train reviewers to counter automation bias and align performance metrics to decision quality rather than speed.

3. Transparency and explainability

Legal obligations

Transparency in AI-driven public decision-making is a legal obligation flowing from multiple statutory requirements. Yet AI systems often operate as "black boxes". People must understand, question and challenge decisions affecting their lives.

What explainability requires

Explainability requirements must provide meaningful explanations of how specific outcomes were reached, case-specific explanations identifying which factors influenced decisions and technical outputs translated into accessible language. For high-stakes decisions affecting individual rights – such as benefit eligibility determinations or child protection risk assessments – authorities must be able to provide meaningful explanations of how specific outcomes were reached. This extends beyond generic descriptions of algorithmic functioning to case-specific explanations.

Require plain English explanations of decision logic and the case-specific factors being considered and ensure any limitations are clearly documented.

Contractual requirements

When procuring AI systems, embed transparency requirements into contracts from the outset. Demand transparency clauses requiring explanations of how AI works, with explanations in plain English within specified timeframes. Your contracts should require suppliers to provide documentation explaining the AI's decision-making logic, identify which factors the AI considers and how they're weighted, provide case-specific explanations when requested and translate technical outputs into accessible language.

Overcoming the "proprietary" excuse

One of the most common obstacles is suppliers claiming they cannot provide transparency because the technology is "proprietary". But transparency and commercial confidentiality can coexist. You need an understanding of how AI reaches decisions, factors considered, how inputs are weighted. Suppliers can protect specific code, proprietary algorithms and training methodologies. Safeguard these through appropriate confidentiality arrangements.

Do not accept "It's proprietary" as an excuse for lack of transparency. What you legitimately need includes understanding how the AI reaches decisions, what factors it considers, how it weighs different inputs, what data it uses and how it can be audited for fairness and accuracy. What suppliers legitimately want to protect includes specific code, proprietary algorithms, training methodologies and competitive advantages. These commercial interests can be safeguarded through appropriate confidentiality arrangements like NDAs, restricted access to technical documentation, secure audit environments and confidentiality rings for sensitive reviews.

Overcome 'black box' constraints via audit rights and confidentiality arrangements (e.g., NDA, secure audit rooms).

Practical mechanisms for balancing transparency and confidentiality:

- Non-disclosure agreements for technical reviewers.
- Restricted access to technical documentation.
- Secure audit environments for independent experts.
- Confidentiality rings for sensitive reviews.
- Escrow arrangements for source code.

4. Cybersecurity

Cybersecurity measures for AI systems extend beyond conventional IT security protocols. AI systems face unique threats such as adversarial attacks, data poisoning and model extraction targeting proprietary data.

1. Adversarial attacks

Adversarial attacks are particularly concerning for AI systems. These are deliberate attempts to manipulate AI outputs by feeding it carefully crafted inputs designed to fool the system. For example, an adversarial attack on a fraud detection system might involve structuring transactions in ways that evade detection whilst still being fraudulent. Likewise, an adversarial attack on an image recognition system might involve subtle modifications to images that are imperceptible to humans but cause the AI to misclassify them.

2. Data poisoning

Data poisoning involves corrupting the training data used to develop the AI model, causing it to learn incorrect patterns or biases. If an attacker can inject malicious data into your training dataset, they can manipulate how the AI behaves. This is particularly concerning if you're using AI systems that continue to learn from new data after deployment – so-called "online learning" systems.

3. Model extraction

Model extraction involves attackers repeatedly querying an AI system and analysing its outputs to reverse-engineer the underlying model, effectively stealing the intellectual property and public investment in its development. This technique can also enable adversaries to infer sensitive information about the training data – potentially including personal data about citizens – and to identify vulnerabilities that can be exploited through subsequent adversarial attacks.

Essential security measures

You must implement encryption, strict access controls and audit trails to safeguard AI data and system interactions effectively. This includes encryption for data at rest and in transit, access controls that limit system interaction to authorised personnel and audit trails that document all system queries and outputs.

Harden AI systems against adversarial inputs, data poisoning and model extraction; encrypt data in transit and at rest; implement strict access controls and audit trails; conduct regular penetration testing and vulnerability management.

Procurement and ongoing assurance

Regular penetration testing, vulnerability assessments and procurement requirements ensure ongoing cybersecurity compliance. Procurement processes offer a critical opportunity to embed security requirements. When commissioning AI systems, public bodies should demand comprehensive security documentation, including penetration testing results, vulnerability assessments and incident response protocols. Contracts must clearly allocate responsibility for security breaches and establish service level agreements that mandate prompt patching and updates.

5. Procurement and contracting

This critical domain is addressed comprehensively in Section 6 below.

Building a risk-based governance framework

The risk-based imperative

Not all AI is created equal. High-stakes decisions affecting individual rights require rigorous safeguards – comprehensive DPIAs, extensive bias testing, meaningful human oversight, robust explainability. Lower-risk applications can be managed with lighter-touch governance.

A robust strategy aligns governance to risk and impact.

Creating an AI inventory

Cataloguing AI systems and classifying them by risk level is fundamental to tailored oversight and governance. Preparatory steps that public bodies can take now will ease the transition as regulatory requirements crystallise. Conducting an AI inventory represents an essential starting point, cataloguing all AI systems currently in use or under consideration, their purposes, risk levels and compliance status. Many public bodies lack comprehensive awareness of AI deployment across their organisations, with individual departments procuring systems without central oversight.

Your AI inventory should capture: what AI systems you’re using, what they do, what data they process, who the supplier is, what the risk level is, whether a DPIA has been conducted, whether bias testing has been done, who is responsible for oversight and when the system was last reviewed.

Catalogue AI systems (purpose, data, supplier, risk, DPIA status, oversight owner).

Essential inventory elements:

Element	Details Required
System identification	Name, supplier, version, deployment date
Purpose and function	What the system does, what decisions it informs or makes
Data processing	Types of data processed, volume, retention periods
Risk classification	High/medium/low based on rights impact
Compliance status	DPIA completed, bias testing conducted, approvals obtained
Oversight	Named accountable owner, review frequency
Human oversight	Level and nature of human involvement in decisions

Risk classification

Risk classification should consider;

- The impact on individuals – whether it affects their rights, access to services, or life opportunities;
- The sensitivity of data processed;
- The degree of automation and if there is meaningful human oversight or is it fully automated;
- The scale of deployment – how many people are affected; and
- The potential for discrimination or bias.

A risk-based approach means that AI systems affecting fundamental rights or public safety. For example, those involved in social care decision-making, education assessments, or benefit determinations, face heightened regulatory requirements. Lower-risk applications, such as appointment scheduling or routine enquiries, can be managed with lighter-touch governance.

Classify risk (rights impact, data sensitivity, level of automation, scale, bias potential).

Risk classification factors:

1. High-risk systems
(requiring maximum governance):

- Directly affect fundamental rights (liberty, family life, fair trial).
- Process special category data at scale.
- Fully or substantially automated decision-making.
- Affect vulnerable populations (children, elderly, disabled persons).
- High potential for discriminatory impact.
- Limited transparency or explainability.

2. Medium-risk systems
(requiring proportionate governance):

- Affect service access or quality.
- Process personal data (non-special category).
- Recommendations reviewed by humans.
- Moderate scale of impact.
- Some potential for bias.

3. Lower-risk systems
(requiring light-touch governance):

- Administrative or operational support.
- Limited personal data processing.
- Human-controlled with AI assistance.
- Limited individual rights impact.
- Low discrimination potential.

Governance structures

Developing internal policies with approval processes and accountability structures ensures responsible AI management. This provides a framework that can be adapted as regulatory requirements evolve, establishing approval processes for AI procurement and deployment, mandating risk assessments and impact evaluations and creating clear accountability structures.

Establish governance (policy, approvals, monitoring cadence, escalation routes).

Essential governance elements:

1. AI governance policy

- Defines what constitutes AI within the organisation.
- Sets out risk classification methodology.
- Establishes approval thresholds and authorities.
- Mandates assessments required before deployment.
- Defines ongoing monitoring requirements.

2. Approval processes

- Low-risk: Departmental approval with notification to central governance.
- Medium-risk: Cross-functional review (legal, IT, DPO, service lead).
- High-risk: Executive approval following comprehensive assessment and external review where appropriate.

3. Accountability structures

- Named executive sponsor for AI governance.
- Cross-functional AI governance board.
- Designated owners for each AI system.
- Clear escalation routes for concerns.
- Regular reporting to leadership and elected members.

4. Escalation routes

- Technical issues → IT security team → Chief Information Officer.
- Bias or discrimination concerns → Equality lead → Chief Executive.
- Data protection concerns → Data Protection Officer → Information Commissioner.
- Constitutional concerns → Legal team → Monitoring Officer.

Human oversight requirements

Embedding human oversight and training programmes addresses automation bias and promotes responsible AI use. Human oversight must be meaningful, not perfunctory. This means human reviewers must have sufficient information about how the AI works and what factors it considered, adequate time to conduct proper review – not just rubber-stamping AI decisions, genuine authority to override AI decisions without penalty and a culture that empowers them to do so with performance metrics that value good decision-making not just speed.

Design human-in-the-loop controls and empower reviewers to override recommendations.

For high-stakes decisions, AI should recommend and humans should decide. The human must have genuine authority to override the AI. This should include sufficient information to make an informed decision, adequate time to conduct proper review and a culture that empowers them to exercise independent judgement.

Requirements for meaningful human oversight:

- **Information:** Human reviewers receive explanation of AI reasoning, factors considered, and confidence levels.
- **Time:** Adequate time allocated for review (performance metrics don't penalise thoroughness).
- **Authority:** Clear power to override without requiring justification beyond professional judgement.
- **Culture:** Organisation values quality decisions over speed; overrides are tracked as quality indicators, not performance failures.
- **Training:** Reviewers understand AI limitations, potential biases, and their own accountability.

Training and organisational capability

Training is vital. Staff using AI systems must understand how they work and their limitations, potential biases and how to identify them, when and how to override AI recommendations, and their own responsibilities and accountability. This training must be ongoing, not a one-off exercise, because AI systems evolve and new risks emerge.

Training programme elements:

For all staff:

- Awareness of where AI is used in the organisation.
- General understanding of AI capabilities and limitations.
- How to escalate concerns.

For AI system users:

- How their specific system works.
- Known limitations and failure modes.
- Potential biases and how to identify them.
- When and how to override recommendations.
- Individual accountability for decisions.

For senior leaders:

- Strategic implications of AI deployment.
- Governance responsibilities.
- Risk oversight and escalation.
- Public accountability and transparency obligations.
- For procurement and legal teams:

Technical understanding sufficient for effective contracting:

- Essential contractual protections.
- Risk assessment methodologies.
- Supplier due diligence.

Continuous monitoring

Ongoing monitoring for bias and compliance requires collaboration across legal, IT, and operational teams. AI systems can drift over time – their performance can degrade, new biases can emerge, or they can become less accurate as the real-world environment changes. You need continuous monitoring for drift, bias, performance degradation, and security vulnerabilities.

Monitor for drift, fairness and performance; refresh DPIAs and security testing.

Monitoring framework:

Performance monitoring:

- Accuracy and error rates.
- Processing times.
- System availability.
- User satisfaction.

Fairness monitoring:

- Decision outcomes by protected characteristic.
- Disparate impact analysis.
- Bias testing at regular intervals.
- Complaint and challenge rates.

Security monitoring:

- Access logs and anomalous queries.
- Attempted attacks or manipulation.
- Vulnerability assessments.
- Incident responses.

Compliance monitoring:

- DPIA currency and accuracy.
- Training completion rates.
- Override rates and quality.
- Regulatory developments.

Procurement, contracting and supplier management

The Procurement Act 2023 opportunities

The Procurement Act 2023 provides opportunities including the Competitive Flexible Procedure for complex AI contracts, early pre-market engagement to understand and enhanced transparency requirements.

The Competitive Flexible Procedure is particularly valuable because it allows negotiation with suppliers, which is critical when procuring technology where requirements may need refinement through dialogue. The procedure can be structured in stages, eliminating suppliers who do not meet your requirements.

For AI contracts, this means you can have initial enables early discussions on capability, and later negotiations on bias testing, explainability, and governance before final selection.

Using flexible procedures also helps engage the market early to assess data provenance, fairness and explainability.

Pre-market engagement

Pre-market engagement is now encouraged. For AI contracts, you need dialogue on training data quality and provenance, bias mitigation approaches and track record, explainability capabilities and limitations and auditability looking at whether you can actually inspect how the system works. Soft-market testing is now easier under the new regime but must comply with transparency and equal treatment principles.

Key pre-market engagement questions:

Training data:

- What data was used to train the AI?
- What is the provenance and quality of training data?
- Does training data include representational balance across protected characteristics?
- How is training data refreshed?

Bias mitigation:

- What bias testing has been conducted?
- What methodologies were used?
- What were the results across all protected characteristics?
- What remediation has been undertaken?
- What is the supplier’s track record in bias mitigation?

Explainability:

- Can the system provide case-specific explanations?
- In what format and at what level of detail?
- What are the limitations of explainability?
- What technical methods are used (e.g., LIME, SHAP)?

Auditability:

- Can the authority audit the system’s functioning?
- What documentation will be provided?
- What intellectual property constraints exist?
- How can commercial confidentiality be balanced with transparency?

Essential contract terms

Your contracts must include essential terms covering: comprehensive bias testing across all nine protected characteristics, non-discrimination warranties, mandatory DPIAs before deployment, transparency and explainability clauses, human oversight requirements, ongoing performance monitoring and bias audits, clear exit provisions covering IP ownership and data portability and audit rights to inspect AI decision-making.

Include essential terms: DPIA completion; non-discrimination warranties; bias testing; transparency; human oversight; audit rights; exit and data portability.

Data protection clauses

Data Protection: Supplier conducts and shares comprehensive DPIA before deployment, processes data only per your written instructions, implements data minimisation with clear justification for each data element, maintains clear retention schedules with automated deletion and provides immediate breach notification.

Required provisions:

- Supplier completes comprehensive DPIA before deployment and shares with authority.
- Data processing only on documented written instructions.
- Justification for necessity of each data element (data minimisation).
- Clear retention schedules with automated deletion.
- Sub-processor approval and management.
- Immediate breach notification (within hours, not days).
- Cooperation with supervisory authorities.
- Data protection impact on pricing and service levels.

Bias and discrimination clauses

Bias and Discrimination: Non-discrimination warranties, comprehensive bias testing before deployment and regularly throughout the term covering all nine protected characteristics, representative training data with documentation of data sources, ongoing fairness monitoring with regular reporting, disparate impact thresholds triggering review and indemnities for discrimination claims.

Required provisions:

- Warranty that the system complies with Equality Act 2010.
- Comprehensive bias testing before deployment covering all nine protected characteristics.
- Defined methodologies for bias testing (e.g., disparate impact analysis, confusion matrices by group).
- Representative and balanced training data with documentation of sources.
- Regular bias testing throughout contract term (e.g., quarterly, annually).
- Supplier provides detailed bias testing reports.
- Defined disparate impact thresholds triggering mandatory review and remediation.
- Authority right to suspend system if unacceptable bias identified.
- Indemnities for discrimination claims resulting from system bias.
- Remediation obligations and timescales.

Transparency and explainability clauses

Transparency: Explainability requirements with explanations in plain English within specified timeframes, audit rights to inspect the model's functioning including access to technical documentation, and limitations disclosure – the supplier must be upfront about what the AI cannot do.

Required provisions:

- Case-specific explanations in plain English.
- Defined timeframes for providing explanations (e.g., within 48 hours of request).
- Explanation must identify factors considered and their relative weights.
- Audit rights to inspect model functioning, including access to technical documentation.
- Confidentiality arrangements to protect legitimate IP whilst enabling transparency.
- Limitations disclosure: supplier must document what the system cannot do, known failure modes, and circumstances where output may be unreliable.
- Regular reporting on system performance and limitations.
- Publication-ready descriptions of system functioning for ATRS compliance.

Cybersecurity clauses

Cybersecurity: Robust security architectures meeting specified standards, regular penetration testing with results shared, vulnerability assessments and prompt patching, incident response protocols with defined timeframes and clear allocation of responsibility for breaches.

Required provisions:

- Security architecture meeting defined standards (e.g., Cyber Essentials Plus, ISO 27001).
- Encryption for data at rest and in transit.
- Strict access controls with multi-factor authentication.
- Comprehensive audit trails.
- Regular penetration testing (at least annually) with results shared with authority.
- Vulnerability assessments and prompt patching (defined SLAs).
- Protection against adversarial attacks, data poisoning, and model extraction.
- Incident response protocols with defined notification timeframes.
- Clear allocation of responsibility and liability for security breaches.
- Insurance requirements.
- Right to audit security controls.

Human oversight clauses

Human Oversight: For decisions significantly affecting individuals, meaningful human review before implementation or available on request, AI provides recommendations not final decisions for high-stakes matters and human reviewers having sufficient information and authority to override.

Required provisions:

- AI has an advisory rather than decisional role in critical matters.
- Human reviewers receive sufficient information to make informed decisions, including explanation of AI reasoning.
- System design enables human override without technical barriers.
- Performance metrics for human reviewers value decision quality, not just speed.
- Training provisions for human reviewers.
- Monitoring of override rates and patterns.
- Escalation procedures when human reviewers consistently override AI.

Performance monitoring and audit clauses

Required provisions:

- Regular reporting on system performance (accuracy, error rates, processing times).
- Monitoring data disaggregated by protected characteristics.
- Authority right to conduct audits (or appoint independent auditors).
- Supplier cooperation with audits, including access to technical staff.
- Defined SLAs with consequences for failure.
- Continuous improvement obligations.
- Authority right to require system modifications if performance or fairness issues identified.

Exit and transition clauses

You also need contract exit planning looking at what happens to the AI model at the end of the contract – who owns it and whether they can continue using it? What about your data – can you extract it in a usable format? Can you port it to a different system? Build in post-contract audits as well. You need the right to audit the supplier's performance after the contract ends, particularly for AI deployments where issues like bias or discrimination may only become apparent over time.

Required provisions:

- Clear intellectual property ownership provisions.
- Authority rights to continue using system or transition to alternative.
- Data extraction rights in usable, portable formats.
- Assistance with transition to replacement system.
- Post-contract audit rights for defined period.
- Retention of evidence and documentation.
- Ongoing liability for issues discovered post-contract.

Supplier evaluation

For evaluation, you need evidence of bias testing across all protected characteristics, explainability capabilities and demonstrations, ethics compliance and governance structures, and security certifications and track record.

Practical applications and risk profiles

From automated planning application assessments to predictive models for social care interventions, AI systems promise efficiency gains and enhanced service delivery.

The key point is that each of these applications requires governance tailored to its risk level and impact on individuals. High-stakes decisions affecting individual rights require the most rigorous safeguards – comprehensive DPIAs, extensive bias testing, meaningful human oversight, robust explainability.

Lower-risk applications can be managed with lighter-touch oversight – perhaps a simpler risk assessment, basic bias testing and periodic review rather than continuous monitoring.

Illustrative use cases within the public sector and corporate environments include; planning assessments, social care risk modelling, customer service chatbots, fraud detection, maintenance prioritisation and demand forecasting. Each of these requires tailored governance proportional to its level of risk.

1. High-risk applications

Social care decision support

Social Care: Predictive models assist in intervention needs, resource optimisation and vulnerability assessments for social care. These are high-stakes applications requiring rigorous bias testing, transparency, and human oversight.

The consequences of getting these wrong, either false positives leading to unnecessary interventions or false negatives missing vulnerable individuals, can be severe. Some authorities are using AI to predict which children are at risk of harm, which elderly residents may need care services or which families may benefit from early intervention.

Risk factors:

- Affects fundamental rights (family life, Article 8 ECHR)
- Processes special category data (health, ethnicity)
- Impacts vulnerable populations (children, elderly, disabled persons)
- High potential for discriminatory impact
- Serious consequences of error (harm to individuals, legal liability)

Governance requirements:

- Comprehensive DPIA with ongoing review
- Extensive bias testing across all protected characteristics before deployment and regularly throughout use
- Independent ethics review
- Meaningful human oversight: AI recommends, humans decide
- Human reviewers have full information, adequate time, and authority to override
- Case-specific explanations available on request
- Ongoing monitoring of outcomes by protected characteristic
- Clear complaints and challenge procedures
- Regular reporting to elected members
- Public transparency via ATRS

Planning and development decisions

Planning and Development: AI supports automated application assessments, building compliance checks and impact analyses to streamline development processes.

Some public bodies are exploring AI to assist with planning application assessments, checking building compliance against regulations and analysing the potential impact of developments. These systems can process applications more quickly and identify issues that might be missed in manual review, but they require careful oversight to ensure consistency with planning policy and to avoid bias in decision-making. For example, you need to ensure the AI doesn't systematically favour or disfavour applications in certain postcodes or from certain types of applicants.

Risk factors:

- Affects property rights and economic interests
- Potential for indirect discrimination (e.g., by postcode, applicant type)
- Impacts on public trust in planning system
- Subject to statutory appeal rights

Governance requirements:

- DPIA addressing necessity and proportionality
- Bias testing for indirect discrimination (by postcode, applicant characteristics)
- Human review before decisions issued
- Clear explanations of how AI informed decision
- Transparency about AI use in planning process
- Ongoing monitoring of decision patterns
- Regular bias audits

2. Medium-risk applications

Service delivery chatbots

Service Delivery: Chatbots and fraud detection systems improve enquiries handling, appointment scheduling and security in services. Many public bodies are deploying chatbots to handle routine resident enquiries, freeing up staff for more complex cases.

Risk factors:

- Affects service access and quality
- Potential for differential service experience
- May process personal data
- Limited but real rights impact

Governance requirements:

- DPIA for personal data processing
- Testing for bias in language understanding (accents, dialects, non-native speakers)
- Clear escalation to human advisers
- Monitoring of chatbot performance and user satisfaction
- Regular review of interactions for bias or service failures
- Transparency about AI use
- Human oversight of novel or complex queries

Fraud detection

AI is also being used for appointment scheduling, optimising service delivery routes for things like waste collection or home care visits and detecting potential fraud in benefit claims or procurement. These applications are generally lower-risk but still require governance. For example, your chatbot needs to be able to handle diverse accents and dialects without bias and your fraud detection system needs to avoid disproportionately flagging certain demographic groups.

Risk factors:

- May disproportionately affect certain groups
- Consequences of false positives (wrongful accusation, investigation)
- Potential reputational harm
- Processing of personal and potentially sensitive data

Governance requirements:

- DPIA addressing proportionality
- Bias testing to ensure system doesn’t disproportionately flag protected groups
- Human review before investigation initiated
- Clear thresholds and criteria
- Transparency about fraud detection methods (within security constraints)
- Ongoing monitoring of flagging rates by demographic
- Appeals process for those wrongly flagged

3. Lower-risk applications

Operational optimisation

Operations: AI enables maintenance prioritisation, resource allocation, and demand forecasting for efficient operations. AI can help prioritise maintenance requests based on urgency and impact, allocate resources efficiently across different services, and forecast demand for services to support planning. For example, AI might predict when roads will need resurfacing based on traffic patterns and weather data or forecast demand for social care services based on demographic trends.

Risk factors:

- Limited direct impact on individual rights
- Administrative and operational efficiency focus
- Minimal personal data processing

Governance requirements:

- Light-touch DPIA if personal data processed
- Basic risk assessment
- Human oversight of significant resource allocation decisions
- Periodic review of performance and accuracy
- Monitoring for unintended consequences (e.g., systematic underinvestment in certain areas)

Implementation: a governance roadmap

Phase 1: Assessment and inventory (Months 1-3)

Objectives:

- Understand current AI deployment across the organisation
- Identify gaps in governance
- Establish baseline for improvement

Actions:

1. Conduct AI inventory

- Survey all departments for current and planned AI use
- Document each system (purpose, data, supplier, risk level)
- Identify systems lacking DPIAs or bias testing

2. Assess current governance

- Review existing policies and procedures
- Identify accountability gaps
- Review procurement practices
- Assess staff awareness and capability

3. Classify systems by risk

- Apply risk classification framework
- Prioritise high-risk systems for immediate attention
- Identify systems requiring retrospective assessment

4. Establish governance structures

- Designate executive sponsor
- Create cross-functional AI governance board
- Define approval authorities and processes
- Establish escalation routes

Phase 2: Policy and framework development (Months 3-6)

Objectives:

- Create comprehensive AI governance policy
- Develop risk assessment and approval processes
- Establish training programmes
- Create procurement guidance

Actions:

1. Develop AI governance policy

- Define what constitutes AI
- Set out risk classification methodology
- Establish approval thresholds
- Define assessment requirements
- Set ongoing monitoring standards

2. Create assessment templates

- DPIA templates tailored for AI
- Bias testing requirements and methodologies
- Security assessment frameworks
- Explainability assessment criteria

3. Develop procurement guidance

- Essential contract terms library
- Supplier evaluation criteria
- Pre-market engagement guidance
- Due diligence checklist

4. Design training programmes

- Awareness training for all staff
- Specialist training for AI users
- Procurement and legal team training
- Leadership briefings

Phase 3: Remediation and compliance (Months 6-12)

Objectives:

- Bring existing systems into compliance
- Implement monitoring frameworks
- Embed new governance in practice
- Build organisational capability

Actions:

1. Remediate high-risk systems

- Complete retrospective DPIAs
- Conduct comprehensive bias testing
- Implement or enhance human oversight
- Review and strengthen contracts
- Address gaps identified

2. Implement monitoring frameworks

- Performance monitoring dashboards
- Fairness monitoring by protected characteristic
- Security monitoring and incident response
- Compliance tracking

3. Roll out training

- Deliver awareness training organisation-wide
- Specialist training for AI users and reviewers
- Procurement and legal team upskilling
- Leadership briefings on governance responsibilities

4. Establish transparency practices

- Publish AI register (ATRS compliance)
- Develop plain-English explanations of AI use
- Create challenge and complaints procedures
- Engage with residents and stakeholders

Phase 4: Continuous improvement (Ongoing)

Objectives:

- Maintain compliance as systems evolve
- Adapt to regulatory developments
- Learn from experience
- Build leading practice

Actions:

1. Regular reviews

- Quarterly governance board meetings
- Annual system reviews
- Regular bias audits
- DPIA refreshes as systems change

2. Monitoring and reporting

- Performance dashboards reviewed monthly
- Fairness monitoring analysed quarterly
- Annual report to leadership and elected members
- Public transparency reporting

3. Continuous learning

- Track regulatory developments
- Learn from incidents and near-misses
- Benchmark against other public bodies
- Engage with professional networks (e.g., LGA AI Hub)

4. Adaptation

- Refine policies based on experience
- Update contracts for new procurements
- Enhance training as risks evolve
- Strengthen governance where gaps identified

Key recommendations

For leadership and elected members

Legal teams should lead AI governance by adopting risk-based approaches and encouraging collaboration across the organisation. Position yourselves as AI governance leaders rather than reactive advisers. This means proactive engagement with AI initiatives from inception, participation in procurement decisions and ongoing oversight of deployed systems. It requires developing collaborative relationships with IT departments, data protection officers, service delivery teams and elected members. AI governance cannot be siloed within legal teams, but legal expertise must inform every stage of AI deployment.

Recommendations:

- Treat AI governance as strategic priority: Ethical AI is not simply a technical project – it is a whole-organisation capability, rooted in governance, accountability and public trust.
- Appoint executive sponsor: Designate a senior executive with responsibility and accountability for AI governance.
- Establish governance structures: Create cross-functional AI governance board with representation from legal, IT, data protection, equality, service delivery and finance.
- Invest in capability: Invest in upfront assessments and continuous monitoring to prevent costly failures.
- Ensure transparency: Being open about AI use demonstrates accountability and builds public confidence. Publish information about what AI systems you are using, what they do, how they work, and how residents can challenge decisions. The Algorithmic Transparency Recording Standard provides a framework for this.
- Maintain public law compliance: Remember that decisions remain subject to Wednesbury principles, procedural fairness duties and the legitimate expectation doctrines.

For legal teams

Legal teams should lead AI governance by adopting risk-based approaches and encouraging collaboration across the organisation. Position yourselves as AI governance leaders rather than reactive advisers.

Recommendations:

- Lead, don't follow: Participate in AI initiatives at the planning stage, not after systems have already been acquired.
- Build technical literacy: Develop sufficient technical understanding to ask the right questions and identify risks.
- Champion transparency: Do not accept "It's proprietary" as an excuse for lack of transparency. Push for meaningful explainability, balanced with appropriate confidentiality protections.
- Embed equality duties: Ensure Public Sector Equality Duty compliance is assessed for every AI system.
- Strengthen contracts: Use contract as a primary mechanism to embed governance requirements and allocate risks.
- Monitor and enforce: Ongoing contract management to ensure suppliers deliver on commitments.

For procurement teams

Recommendations:

- Engage early: Pre-market engagement is now encouraged. For AI contracts, you need dialogue on training data quality and provenance, bias mitigation approaches and track record, explainability capabilities and limitations, and auditability – simply whether you can actually inspect how the system works.
- Use flexible procedures: The Competitive Flexible Procedure is particularly valuable for complex AI contracts. It allows you to negotiate with suppliers, which is critical when you're procuring cutting-edge technology where requirements may need to be refined through dialogue.
- Mandate essential terms: Include essential terms: DPIA completion; non-discrimination warranties; bias testing; transparency; human oversight; audit rights; exit and data portability.
- Evaluate ethics rigorously: For evaluation, you need evidence of bias testing across all protected characteristics, explainability capabilities and demonstrations, ethics compliance and governance structures, and security certifications and track record.
- Plan for exit: Build in post-contract audits. You need the right to audit the supplier's performance after the contract ends, particularly for AI deployments where issues like bias or discrimination may only become apparent over time.

For service delivery teams

Recommendations:

- Understand your accountability: Human reviewers remain accountable for decisions, even when AI-informed.
- Exercise critical judgement: Authorities must actively combat automation bias through multiple mechanisms. Training must emphasise human responsibility and accountability. Performance metrics should value good decision-making not just speed. An organisation's culture must empowers staff to override AI when appropriate and regular reviews should assess decision quality not just decision volume.
- Demand explainability: Insist on understanding how AI reaches recommendations before relying on them
- Report concerns: Escalate concerns about bias, accuracy or inappropriate use immediately
- Maintain human centrality: For high-stakes decisions, AI should recommend and humans should decide.

For data protection officers

Recommendations:

- Scrutinise DPIAs: These assessments cannot be perfunctory box-ticking exercises. Ensure they genuinely interrogate necessity, proportionality and risks.
- Challenge data maximisation: AI systems often want as much data as possible to improve accuracy, but GDPR requires you to process only the minimum necessary.
- Protect special category data: Many AI systems in social care, education, or health-related services will process special category data, so this is not a theoretical concern. Ensure both Article 6 and Article 9 conditions are satisfied.
- Maintain oversight: Regular reviews of AI data processing, not just at deployment.
- Collaborate widely: Work closely with legal, IT and service teams to embed data protection throughout AI lifecycle.

For IT and security teams

Recommendations:

- Address AI-specific threats: AI systems face unique threats such as adversarial attacks, data poisoning and model extraction targeting proprietary data.
- Implement robust controls: You must implement encryption, strict access controls and audit trails to safeguard AI data and system interactions effectively.
- Test regularly: Regular penetration testing, vulnerability assessments and procurement requirements ensure ongoing cybersecurity compliance.
- Monitor continuously: AI systems can drift over time and their performance can degrade. New biases can emerge or they can become less accurate as the real-world environment changes. You need continuous monitoring for drift, bias, performance degradation, and security vulnerabilities.
- Support transparency: Enable audit and explainability whilst protecting security.

Conclusion

The challenges that AI presents to the public sector are not transient difficulties that will resolve as technology matures. Rather, they represent a fundamental shift in how public services are delivered and how authorities must approach governance and accountability.

AI governance frameworks must evolve as technology and regulation develop. What works today may not work tomorrow. Stay informed about regulatory developments, learn from other public bodies' experiences and be prepared to adapt your approach.

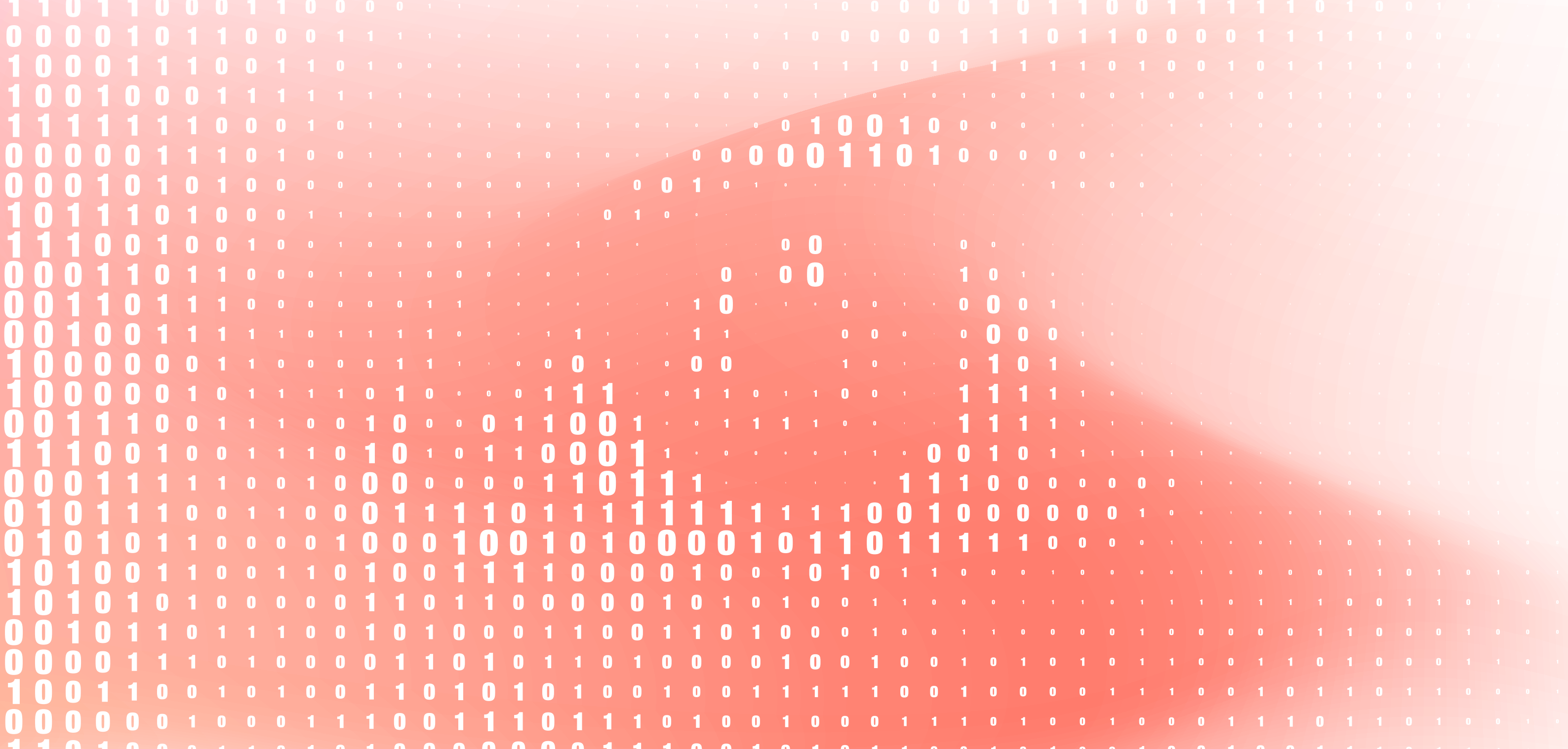
The safe and ethical rollout of AI in the public sector requires:

- Strategic leadership treating AI governance as an organisational priority.
- Robust frameworks that are risk-based, proportionate and adaptable.
- Clear accountability with named owners and escalation routes.
- Proactive compliance with data protection, equality and public law obligations.
- Meaningful transparency building public trust through openness.
- Continuous vigilance monitoring performance, fairness and security.
- Whole-organisation capability embedding ethics throughout the AI lifecycle.

Prevention is better than cure. Rigorous upfront assessment through DPIAs, bias testing, and security reviews prevents costly failures down the line. It is much easier to get things right from the start than to fix problems after deployment, particularly when those problems may involve discrimination against vulnerable residents or data breaches affecting thousands of people. Being open about AI use demonstrates accountability and builds public confidence.

Public sector organisations have both an obligation and an opportunity: the obligation to deploy AI responsibly, protecting citizens' rights and maintaining public trust; and the opportunity to demonstrate to the wider economy what responsible AI looks like in practice.

Ethical AI is not simply a technical project, it is a whole-organisation capability, rooted in governance, accountability and public trust.



AI Governance Checklist

Category	Essential Actions
Governance and Strategy	Catalogue AI systems (inventory) and classify by risk; set approval processes and assign accountable owners
Data Protection	Complete DPIAs; justify lawful basis and data minimisation; apply safeguards for special category data; retention and deletion
Bias and Discrimination	Audit data; mandate supplier bias testing; monitor outcomes; train staff to counter automation bias
Transparency and Explainability	Include plain-English explainability clauses; audit rights; publish AI use; document limitations and routes to challenge
Cybersecurity	Encrypt data; strict access controls; audit trails; pen-testing and vulnerability management; incident response
Procurement and Contracting	Use flexible procedures; pre-market engagement; essential terms: DPIA, fairness, transparency, oversight, audit, exit
Human Oversight and Training	Meaningful human review for high-stakes decisions; ongoing training on AI risks, overrides and accountability
Continuous Improvement	Monitor for drift, bias and performance; refresh DPIAs; track regulatory updates and adapt governance

References and further reading

- UK Government White Paper: A pro-innovation approach to AI regulation (March 2023)
— https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
- Implementing the UK's AI Regulatory Principles (Guidance for Regulators, February 2024)
— https://assets.publishing.service.gov.uk/media/65c0b6bd63a23d0013c821a0/implementing_the_uk_ai_regulatory_principles_guidance_for_regulators.pdf
- Algorithmic Transparency Recording Standard (ATRS) Hub
— https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub
- ICO Guidance on AI and Data Protection; Explaining Decisions made with AI (March 2023)
— https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/
- National AI Strategy (Sept 2021)
— https://www.gov.uk/government/publications/national-ai-strategy
- LGA/LOTI: Responsibly buying AI (April 2025)
— https://www.local.gov.uk/publications/responsible-buying-how-build-equality-data-protection-your-ai-commissioning
- Procurement Act 2023: Competitive Flexible Procedure (April 2024)
— https://www.gov.uk/government/publications/the-official-procurement-act-2023-e-learning/module-4-competitive-flexible-procedure
- Trowers & Hamlins: Ethics of AI in the workplace (December 2025)
— https://www.trowers.com/insights/2025/december/ethics-of-ai-in-the-workplace

Contact

For further guidance on AI governance in the public sector, please contact:

Amardeep Gill
Partner, National Head of Public Sector
t +44 (0)7917 507675
e agill@trowers.com

Louis Sebastian
Partner
t +44 (0)7725 102031
e lsebastian@trowers.com

Matt Whelan
Senior Associate
t +44 (0)7980 963980
e mwhelan@trowers.com

This white paper provides guidance on governance issues associated with AI deployment in the public sector. It does not constitute legal advice. Organisations should seek specific legal advice tailored to their circumstances before making decisions about AI procurement or deployment.

trowers & hamlins

trowers.com