

AI risk controls: what to put in place

Practical controls for putting AI governance into action.

1 Autonomy in decision-making

- ☐ Define clear accountability and human oversight roles for AI decisions
- ☐ Establish policies for human-in-the-loop or human-on-the-loop controls
- ☐ Set predefined conditions or thresholds for when human intervention is mandatory (high level of risk, low confidence score, ethical or sensitive decisions, etc.)

2 Transparency and explainability

- ☐ Use inherently interpretable models when possible
- ☐ Ensure that AI models are designed and implemented to support post-hoc explainability techniques when complex models are used
- ☐ Define explainability expectations based on the system's risk profile and require model cards detailing purpose, limitations, and interpretability approach

3 Access control misalignment between source data and AI outputs

- ☐ Consider using policy-aware or context-aware Retrieval-Augmented Generation (RAG) to constrain the information retrieved and used in AI-generated outputs based on user roles and rights
- ☐ Use model segregation or data tagging to restrict output access
- ☐ Audit AI outputs regularly for unauthorised data disclosures
- ☐ Avoid applying AI models to highly sensitive data unless effective security enforcement mechanisms are in place

4**Hallucination and mistakes in generated outputs**

- ☐ Establish human review and validation processes for critical AI outputs
- ☐ Continuously monitor AI outputs for errors and inconsistencies
- ☐ Train staff and users to critically assess and verify AI-generated outputs before acting on them, especially in high-risk or decision-making contexts
- ☐ Develop escalation protocols for detected mistakes

5**Bias and fairness risks**

- ☐ Use diverse and representative training datasets
- ☐ Involve diverse stakeholders in model review
- ☐ Perform bias audits on training data and AI models
- ☐ Establish remediation plans for identified biases

6**Data availability and quality for AI training**

- ☐ Estimate data sufficiency and quality before model training when possible
- ☐ Implement data cleansing and augmentation practices
- ☐ Monitor model performance for signs of data-related issues

7**Operational and lifecycle risks**

- ☐ Implement continuous monitoring for model drift and performance degradation
- ☐ Develop robust model update, retraining, and retirement procedures
- ☐ Establish AI-specific incident response plans
- ☐ Maintain version control and documentation for AI models

8**Extended supply chain and third-party AI governance risks**

- ☐ Conduct due diligence on AI governance practices of suppliers
- ☐ Include AI governance requirements in contracts
- ☐ Monitor and audit third-party AI usage regularly
- ☐ Establish incident reporting agreements
- ☐ Define and enforce policies specifying data types or sensitivity levels that are prohibited from being processed or stored in third-party or cloud environments

9**Legal, regulatory, and compliance risks**

- ☐ Maintain up-to-date knowledge of applicable AI and data regulations and standards
- ☐ Integrate regulatory requirements into the design, procurement, and deployment of AI systems
- ☐ Document compliance measures and maintain audit trails for AI system operations
- ☐ Train staff regularly on relevant legal and ethical requirements

10

Organisational readiness and skills

- ☐ Provide AI literacy and ethics training
- ☐ Build organisational awareness, clear accountability, open communication, and leadership commitment to ensure responsible and ethical AI use
- ☐ Align AI initiatives with business strategy
- ☐ Establish clear communication channels for AI governance

Watch on demand

Catch our latest webinar, ***AI: The governance gap nobody is talking about***, and learn how to assess your current governance, spot the gaps, and weigh the value of AI against its risks.

[**Watch now >**](#)