CITMA The Chartered Institute of Trade Mark Attorneys

Part B – Critical risk assessment

Part B of this series examines the key technical, operational, and regulatory risks that must be assessed when Trade Mark Attorneys and IP professionals adopt AI-based tools. As AI systems become increasingly integrated into legal workflows, understanding and managing these risks is essential to protecting clients, maintaining professional standards, and ensuring the reliability of outputs used in legal decision-making. This section provides a structured overview of the critical areas firms should evaluate as part of any AI procurement or governance process.

The guidance covers five core components of risk assessment:

- **Hallucination prevention measures**
  AI tools can produce false, incomplete, or fabricated information, commonly known as "hallucinations". In legal practice, such inaccuracies pose a significant risk, as they can mislead professionals or form an incorrect basis for advice. Firms should ensure that vendors have robust mechanisms for reducing hallucinations, such as confidence scoring, internal validation layers, fact-checking modules, and human-in-the-loop review processes. Vendors must be able to demonstrate how often hallucinations occur, how they are detected, and how erroneous outputs are corrected. Effective hallucination controls are fundamental to maintaining professional integrity and ensuring safe reliance on AI-generated outputs.

- **Explainability and auditability**
  Legal professionals must be able to understand, challenge, and justify the outputs generated by AI tools. This requires systems that offer clear explainability, meaning users can see how inputs were processed and why particular outputs were generated. Vendors should also provide audit functionality, enabling firms to trace decisions, review system behaviour, and retain evidence for regulatory inquiries, court proceedings, or internal risk reviews. Without meaningful explainability and audit trails, an AI system may expose the firm to compliance failures, reputational damage, or difficulties in defending professional decisions.

- **Data security, privacy, and compliance**
  AI systems frequently process sensitive, regulated, or confidential information. Firms must therefore verify that vendors implement strong security controls, including encryption, access management, intrusion detection, secure hosting environments, and clearly defined data-retention and deletion policies. Vendors should be fully compliant with the GDPR and any other data protection regimes relevant to the jurisdictions in which the firm operates. Firms must also assess whether vendors use subcontractors, where data is stored geographically, and how cross-border transfers are managed. Strong data governance ensures that AI adoption does not inadvertently introduce privacy breaches or regulatory non-compliance.

- **Global data isolation and segregation**
  Multinational legal practices require careful handling of client and matter data, particularly where data residency, cultural norms, or regulatory frameworks differ across regions. Firms should evaluate whether vendors maintain rigorous data isolation, preventing data from one client, office, or jurisdiction from being accessed or inferred by another. This includes analysing how the vendor handles multi-region deployments, model hosting, and access controls. Proper data segregation protects confidentiality, reduces the risk of inadvertent disclosure, and supports compliance with international privacy requirements.

- **Due-diligence documentation and independent assurance**
  A mature and trustworthy AI vendor should be able to provide detailed documentation supporting the safety, robustness, and reliability of their tools. This may include model testing results, risk assessments, security audits, performance metrics, and evidence of compliance with industry-recognised standards such as ISO 27001 (information security) or ISO 42001 (AI management systems). Such documentation enables firms to conduct thorough internal assessments, benchmark vendor quality, and demonstrate responsible procurement to regulators, clients, and internal governance teams.

What is an ISO?
The **International Organisation for Standardisation (ISO)** develops globally recognised standards that define best practice across a wide range of industries. These standards set out agreed requirements, guidelines, and specifications designed to support consistency, safety, quality, and interoperability. ISO standards are developed through international expert consensus and, while voluntary, are widely adopted as benchmarks for commercial, regulatory, and operational excellence.

A full list of ISO Standards can be found at https://www.iso.org/home.html

Top 5 tips for critical risk assessment

1. **Implement robust controls to detect and prevent hallucinations**
   AI systems may occasionally generate inaccurate, fabricated, or misleading outputs. Vendors should therefore demonstrate clear, tested mechanisms for minimising hallucinations, such as internal verification layers, confidence scoring, cross-referencing methods, and human-in-the-loop safeguards. Firms should assess how the vendor identifies errors, flags uncertainties, and prevents unverified information from influencing decisions. Effective hallucination mitigation is essential for maintaining professional integrity and client trust.

2. **Require clear explainability and full auditability of outputs**
   Legal professionals must be able to understand, verify, and justify the basis of any AI-assisted output. Choose vendors who can provide transparent explanations of how their models operate, including input pathways, decision logic, and model limitations. Systems should produce audit trails that allow firms to trace how specific outputs were

generated and to demonstrate defensibility in regulatory, client, or court contexts. Lack of explainability represents a material risk in professional practice.

3. **Rigorously assess data security, privacy, and compliance standards**
   Strong data protection practices are fundamental to risk management. Firms should confirm that the vendor adheres to recognised cybersecurity and privacy frameworks, including encryption standards, access controls, retention schedules, and deletion procedures. Compliance with GDPR and other relevant regional regulations must be demonstrable. Vendors should be transparent about their data-processing activities, including subcontractors, hosting arrangements, and any third-party integrations.

4. **Evaluate global data isolation and jurisdictional safeguards**
   Where firms operate across multiple offices or jurisdictions, the ability to isolate data at a regional, client, or matter level is critical. Vendors should provide clear information on how data is segregated to prevent cross-border exposure or inadvertent sharing between clients or practice groups. Firms should assess whether the supplier's technical architecture and governance model support compliance with differing regulatory regimes and confidentiality obligations worldwide.

5. **Request comprehensive due-diligence and technical assurance documentation**
   Trusted vendors should supply detailed evidence of system robustness, including:

   - risk assessments and testing results
   - model performance and accuracy metrics
   - bias evaluations
   - change-management processes
   - incident-response procedures
   - relevant certifications (e.g., ISO 27001, ISO 42001) This documentation enables firms to conduct informed evaluations and to benchmark the vendor's reliability and readiness to support professional-grade AI use.