



Standard
Intelligence

White paper

Ungoverned AI Systems

*Enterprise Risk, Financial Exposure, and Probability
Sensitivity Analysis*

PUBLISHED

June 2026

WEB

standardintelligence.com

AUDIENCE

For enterprise risk owners and governance investors

CONTENTS

Summary	3
1. AI has become a top-tier institutional risk, and the exposure sits in ungoverned systems	4
2. The figures measure a risk surface, not a forecast of loss	5
3. Thirteen categories make up the surface	6
3.1 Data privacy and AI-facilitated security breaches	6
3.2 Algorithmic bias, discrimination and civil liability	7
3.3 Hallucination and high-consequence decision errors	7
3.4 Regulatory and compliance failures	8
3.5 Intellectual property infringement	9
3.6 Agentic AI and autonomous action failures	10
3.7 AI-enabled fraud, deepfakes and social engineering	10
3.8 Misinformation, synthetic media and market manipulation	11
3.9 Shadow AI and unsanctioned deployments	12
3.10 Supply chain and third-party AI risk	12
3.11 Workforce displacement, HR liability and labour litigation	13
3.12 Critical infrastructure and national security	14
3.13 Reputational damage and institutional trust	14
4. The aggregate surface runs to trillions of pounds a year, and tail risk sets its upper bound	15
5. Expected loss is concentrated in five categories	16
6. Governance leverage, not exposure, shows where investment pays	17
7. Sector, scale and institution type change which categories dominate	18
7.1 Corporations	19
7.2 Governments	19
7.3 Non-profit institutions	20
8. On our model, governance returns 8 to 22 pounds for each pound invested	20
Where to start, and what it costs	21
Basis and Standing	22
Sources	23

Summary

Figure	What it is
£1.6 trillion to £4.4 trillion	The modelled annual exposure surface from ungoverned AI, worldwide, in 2026
13	Risk categories assessed, each with a per-event cost range, an annual aggregate and a governance leverage rating
38% to 52%	The modelled reduction in expected annual loss from moving governance maturity from 30% to 70%

AI has become a top-tier institutional risk. In the 2026 Allianz Risk Barometer, a survey of risk management experts, artificial intelligence rose from tenth place to second, the biggest jump in that year’s ranking. The exposure sits in ungoverned systems: those without documented risk ownership, version control, audit trails or explainability, formal risk assessment, or institutional approval, or that operate outside data governance and privacy policy.

The figures measure a risk surface, not a forecast of loss. Standard Intelligence’s model gives each category a per-event cost range, an annual aggregate (the scale of loss put at risk across institutions worldwide in a year) and a probability-weighted expected value under three scenarios. Its inputs combine published surveys with structured expert judgement where data is sparse. The shadow AI expected values also exceed that category’s own annual aggregate (section 5), so the results are indicative rather than precise.

Thirteen categories make up the surface. They are data breaches, algorithmic bias, hallucination, regulatory non-compliance, intellectual property, agentic AI, fraud and deepfakes, misinformation, shadow AI, supply chain, workforce liability, critical infrastructure and reputational damage. Each has its own mechanism, evidence and legal exposure, and each is sized separately.

On our model, the aggregate surface is £1.6 trillion to £4.4 trillion a year. The largest single contributor to the upper bound is the tail risk of an AI-facilitated attack on critical infrastructure, an event of low probability and extreme severity.

Expected loss is concentrated in five categories. In the base case the largest expected losses are in shadow AI, AI-enabled fraud and deepfakes, data breaches, reputational damage and regulatory non-compliance. Agentic AI is smaller today but, on our model, the fastest-growing category.

Governance leverage, not exposure, should decide where governance money goes. Governance leverage is the proportional reduction in expected loss that an institution’s own investment can achieve. It is very high for regulatory non-compliance, agentic AI and critical infrastructure; moderate for fraud and data breaches, which external adversaries drive; and low to moderate for shadow AI, which employee behaviour drives. The largest exposures are therefore not always where investment reduces loss the most.

Regulatory exposure is partly positional. We expect enforcement attention to fall first on organisations with visible gaps, so an organisation's exposure depends on its governance relative to its peers as well as on the requirements themselves.

The profile differs by sector, scale and institution type. Financial services, healthcare, technology vendors, governments and non-profits each face a different mix of dominant categories, so no single governance template suits them all.

On our model, governance pays where leverage is high. Moving from 30% to 70% governance maturity cuts aggregate expected annual loss by roughly 38% to 52%, and each £1 invested in structured governance returns an estimated £8 to £22 in avoided loss. These are modelled ranges, not observed outcomes.

Where to start. Identify which systems meet the definition of ungoverned and which categories dominate the portfolio. Then put the cheap, high-leverage controls in place first (action logging, approval gates and sandboxing for agents; deepfake-resistant authentication for material commitments; approved tools good enough that employees stop looking elsewhere), and keep AI Act documentation and supplier evidence current from then on.

Basis and Standing, at the end, sets out which propositions are binding law, which are survey findings, and which are our own model and recommended controls.

1. AI has become a top-tier institutional risk, and the exposure sits in ungoverned systems

Adoption of AI is moving faster than governance, regulation and workforce readiness, and organisations increasingly see it as a source of operational, legal and reputational risk as well as a strategic opportunity. In the 2026 Allianz Risk Barometer, in which 3,338 risk management experts in 97 countries ranked their concerns, artificial intelligence rose to second place from tenth in 2025, “the biggest jump in this year’s ranking”. Allianz describes this as pushing AI “into the top tier of global risks for the first time”, and reports that “a rising share now views it as a distinct and increasingly complex risk category”, although 44% of respondents to its AI questions still see more benefits than risks for their industry.

What makes a system ungoverned. An AI system is treated here as ungoverned when one or more of these conditions applies:

- (a) it operates without documented risk ownership;
- (b) it lacks version control, audit trails or explainability mechanisms;
- (c) it was deployed without formal risk assessment;
- (d) it operates outside the boundaries of data governance or privacy policy;
- (e) it was adopted at user level without institutional approval, the pattern commonly called shadow AI.

Each condition is a property of a particular system, so applying the definition requires a view of the organisation's AI one system at a time.

Much of the risk arises after deployment. The Second International AI Safety Report, published in February 2026, reviews the science on the capabilities and risks of general-purpose AI; it was led by Yoshua Bengio and written by over 100 experts with the backing of over 30 countries and international organisations. It finds that “performance on pre-deployment tests does not reliably predict real-world utility or risk”, that AI agents “act autonomously, making it harder for humans to intervene”, and that because risk management measures have limitations, “they will likely fail to prevent some AI-related incidents”. Those findings point to how organisations connect AI to their processes, data and decisions, which is where governance operates.

Sizing the exposure from ungoverned systems needs a method that is explicit about what its figures mean.

2. The figures measure a risk surface, not a forecast of loss

The annual aggregate is a surface. For each of thirteen categories, the model estimates a per-event cost range (from a minor incident to a severe one), an at-risk population and an annual aggregate. The at-risk population is modelled from World Bank enterprise data, OECD statistics and publicly available AI adoption surveys as of the first quarter of 2026. The annual aggregate is the exposure surface: the scale of loss that ungoverned AI puts at risk across corporations, governments and non-profit institutions worldwide in a year, not a forecast of realised loss.

Expected value weights the surface by probability. Each category is run under three scenarios: low-risk (high governance maturity, strong enforcement, conservative adoption), base case (the current 2026 trajectory) and high-risk (accelerated ungoverned adoption, adversarial exploitation, regulatory arbitrage). Expected value is the institution-level annual event probability multiplied by the median event cost and the estimated annual event frequency.

The inputs are partly judgement. They are calibrated against three sources: incident rates in IBM’s 2025 breach data and similar surveys; the Gartner survey of 360 organisations in the second quarter of 2025; and structured expert judgement where empirical data is sparse, particularly for agentic AI and critical infrastructure tail risk. Categories with stronger empirical grounding carry tighter ranges, and the wider ranges reflect genuine uncertainty about the inputs.

Governance leverage and governance maturity are our own measures. Governance leverage is the proportional reduction in expected value that institutional investment alone can achieve, holding external adversarial dynamics and enforcement intensity constant, rated on a five-level scale (very high, high, moderate, low to moderate, low). Governance maturity is a percentage input to the model, not a regulatory or industry standard, and it does not map onto the five named maturity levels used in Enterprise Shift or the TRACE maturity model. Its baseline is our calibration from the Gartner

survey, which Gartner reports only as a comparison of governance effectiveness with and without AI governance platforms; it is not a Gartner figure.

Currency and growth. Financial figures are in pounds sterling. USD source data is converted at a single fixed rate of £1 = \$1.27 for all years, so converted figures do not track exchange-rate movements, and AI Act penalty amounts are stated in euro as in the regulation. Growth rates are compound annual growth rates for 2026 to 2028 on our model.

3. Thirteen categories make up the surface

Each category has its own mechanism, evidence and legal exposure, and the model sizes each separately, with a per-event range and an annual aggregate.

3.1 Data privacy and AI-facilitated security breaches

AI connected to internal data widens the attack surface, by mechanisms conventional security does not cover. A retrieval-augmented system connected to a document store can, in response to crafted prompts, combine information from sources no single user is entitled to see together, where the system's own access is broader than any user's. An embedding pipeline creates a derivative data asset that data classification policies often do not reach. A model can memorise and reproduce parts of its training data (Carlini et al., 2021), so deleting the source records does not remove what a fine-tuned model has retained.

Ungoverned systems often lack data classification enforcement, output filtering and access controls. IBM's 2025 study of 600 organisations that suffered a breach between March 2024 and February 2025 found that 13% reported breaches of AI models or applications, and that 97% of those lacked AI access controls; 63% of the breached organisations had no AI governance policy or were still developing one.

In the same study the global average cost of a data breach fell to \$4.44 million, the first decline in five years, driven by faster identification and containment. The US average reached a record \$10.22 million, pushed up in part by higher regulatory fines. IBM's 2026 report puts the global average at \$4.99 million, a new record.

Regulatory exposure compounds the direct cost. A breach of personal data can attract GDPR fines of up to EUR 10 million or 2% of worldwide annual turnover for failure of the Article 32 security duty (Article 83(4)), and up to EUR 20 million or 4% where the integrity and confidentiality principle in Article 5(1)(f) is breached (Article 83(5)), whichever is higher in each case. Notification clocks often start before the scope of a breach is understood: 72 hours under GDPR Article 33, four business days after the incident is determined to be material under the SEC's Form 8-K Item 1.05, and sector frameworks such as DORA.

Where the breached system is a high-risk AI system, the AI Act adds concurrent exposure: the cybersecurity requirement of Article 15 is enforced through the provider's obligations in the EUR 15 million or 3% tier (Article 99(4)), from 2 December 2027 for Annex

III systems, subject to the Article 111(2) transition in section 3.4. The Act's EUR 35 million or 7% tier is reserved for prohibited practices, which a breach does not of itself engage.

Exposure (our model). Per-event £3.5 million to £8.5 million (global average; US incidents frequently exceed £8 million). Annual aggregate (2026): £310 billion to £580 billion globally.

3.2 Algorithmic bias, discrimination and civil liability

AI deployed without controls in consequential decisions can encode and amplify historical bias. The contexts include hiring, lending, life and health insurance underwriting, benefits allocation and criminal justice. Training data drawn from past decisions carries their demographic patterns, including patterns produced by discrimination; proxy variables such as postcode, employment gaps and name patterns can stand in for protected characteristics in ways that are statistically detectable but legally obscure; and retraining on the outcomes of earlier biased decisions can amplify the bias.

The AI Act treats most of these uses as high-risk. Recruitment and selection (Annex III point 4), public benefits (point 5(a)), creditworthiness (point 5(b)), life and health insurance pricing and risk assessment (point 5(c)), and law enforcement and the administration of justice (points 6 and 8) are listed, subject to the Article 6(3) derogation; insurance is covered only for life and health. Providers must examine training data for possible biases and take measures to detect, prevent and mitigate them (Article 10(2)(f) and (g)), tell deployers, where appropriate, how the system performs for specific persons or groups (Article 13(3)(b)(v)), and monitor the system after it reaches the market (Article 72). When these duties apply is set out in section 3.4.

In our experience, duties of this kind are new to most enterprises. Because post-market monitoring continues them for the life of the system, a pre-deployment fairness audit is where the obligation starts rather than where it ends.

US exposure arrives through a different route. Title VII, the ADEA, ECOA and the Fair Housing Act cover most high-risk contexts, with class actions as the dominant vehicle; the EEOC's 2023 technical assistance on AI under Title VII was removed from the EEOC website in January 2025. The EEOC's first settlement involving automated hiring software, iTutorGroup's \$365,000 settlement over automatic rejection of older applicants, sits at the lower end of the exposure curve. On our model the upper end is set by financial-services class actions calculated against the full affected applicant population.

Exposure (our model). Per-event £850,000 (minor regulatory finding) to £620 million (US federal class action with operational remediation). Annual aggregate: £95 billion to £210 billion globally.

3.3 Hallucination and high-consequence decision errors

Hallucination, a model producing fluent output that is false, causes material harm where generative output is used without validation. It matters most in clinical, legal, financial advisory and engineering contexts. It differs from software defect risk because it arises from probabilistic generation rather than from a fault that can be patched, and nothing in the output's form marks it out: a legal research tool presents a fabricated citation with the same fluency as a real one, and a clinical decision-support tool presents a hallucinated contraindication as it would accurate guidance.

The reviewer nominally in the loop often lacks the time or expertise to catch output that reads as plausible expert work. This is automation bias, the tendency to rely on automated output in place of one's own checking, which the human-factors literature reports in naive and expert users alike (Parasuraman and Manzey, 2010). In our assessment it may be harder to counter with generative systems, whose output resembles expert human work and may evade the cues reviewers use to flag suspect machine output.

The litigation pattern has begun to form. *Mata v. Avianca, Inc.* (S.D.N.Y. 2023) set the baseline for sanctions against lawyers who file AI-fabricated citations: Judge Castel imposed a \$5,000 Rule 11 sanction after counsel submitted a brief citing authorities generated by ChatGPT. Later decisions have applied the same principle to other filings.

In medicine, a study of the FDA's MAUDE database found 943 adverse event reports between 2010 and 2023 involving AI/ML-enabled devices, against around 950 such devices authorised by August 2024 (Babic et al., 2025); the FDA's published list had grown to 1,247 AI-enabled devices by July 2025. The reports concern AI/ML-enabled devices generally rather than hallucination specifically, and attribution is difficult because an AI error often combines with conventional clinical error rather than causing harm in isolation.

Exposure (our model). Per-event £120,000 (low-stakes administrative error) to £85 million (medical misdiagnosis with patient harm, full litigation cycle). Annual aggregate: £55 billion to £140 billion.

3.4 Regulatory and compliance failures

AI regulation is multiplying. Gartner projects that by 2030 fragmented AI regulation will quadruple and extend to 75% of the world's economies, and that spending on AI governance will reach \$492 million in 2026 and surpass \$1 billion by 2030.

The exposure differs from conventional regulatory risk in three ways. The obligations are both horizontal and sectoral, creating a matrix no single internal function has historically owned. They apply to deployers as well as developers, which weakens the supplier indemnities enterprises have relied on to manage vendor risk. And they change faster than annual compliance cycles, which calls for continuous monitoring rather than periodic audit.

The AI Act applies in stages. It entered into force on 1 August 2024, and Regulation (EU) 2026/1744 (the Digital Omnibus on AI), in force since 27 July 2026, amended its timetable:

- the Article 5 prohibitions have applied since 2 February 2025, and the two prohibitions added by Regulation (EU) 2026/1744 apply from 2 December 2026;
- the Article 4 AI literacy duty applied from 2 February 2025; since 27 July 2026, Article 4 as replaced requires providers and deployers to take measures to support the development of AI literacy, without requiring them to guarantee any specific level of AI literacy of any individual;
- the penalty provisions have applied since 2 August 2025, and the Article 50 transparency obligations since 2 August 2026;
- high-risk obligations apply from 2 December 2027 for Annex III systems and 2 August 2028 for Annex I systems (Article 113);

- under Article 111(2), a high-risk system placed on the market or put into service before then is caught only if its design changes significantly, and systems intended for use by public authorities must comply by 2 August 2030.

The penalties are tiered. Prohibited practices, such as social scoring or biometric categorisation to infer characteristics such as race or political opinions, attract fines of up to EUR 35 million or 7% of worldwide annual turnover, whichever is higher (Article 99(3)). Most other operator obligations, including the provider obligations in Article 16 (which take in data governance), the deployer obligations in Article 26 and the transparency obligations in Article 50, attract up to EUR 15 million or 3% (Article 99(4)), and supplying incorrect, incomplete or misleading information to notified bodies or national competent authorities up to EUR 7.5 million or 1% (Article 99(5)); for SMEs each cap is whichever is lower (Article 99(6)).

The GDPR suggests how enforcement may develop. Its early years after May 2018 produced a few headline fines alongside a much larger number of unresolved investigations, and by 1 March 2024 cumulative GDPR fines had reached around EUR 4.48 billion. We expect AI Act enforcement to follow a comparable path with two amplifiers: the AI Office's direct authority over general-purpose models removes a layer of national variation, and the Annex IV technical documentation gives enforcers a structured evidential basis that GDPR investigators had to assemble from scratch.

Once high-risk obligations apply, a provider that does not have the technical documentation Article 11 requires, drawn up before placing on the market and “kept up-to-date”, will not meet its Article 16 obligations and falls within the Article 99(4) tier. As at 11 October 2026, no fine under the AI Act had been confirmed by any regulator or official source.

Exposure (our model). Per-event £180,000 (minor procedural violation, SME context) to £2.8 billion (systemic non-compliance, globally systemic institution). Annual aggregate: £160 billion to £380 billion.

3.5 Intellectual property infringement

An enterprise that uses infringing output can be liable for it. Ungoverned generative deployments can produce output that reproduces protected material without attribution or licence. Claims so far have mostly been brought against developers, but an enterprise that commercially uses infringing output can itself be liable, even where a third party trained the underlying model; output filtering and provenance checks reduce that risk.

The case law has addressed training, not outputs. Summary-judgment rulings in 2025 on fair use of copyrighted material in training, including the partial ruling in *Bartz v. Anthropic* and the ruling in *Kadrey v. Meta*, addressed the training stage without resolving whether outputs that reproduce protected expression infringe regardless of training-stage findings, and most high-profile cases remain in litigation. *Bartz* was later settled as a \$1.5 billion class action, preliminarily approved by Judge Alsup in September 2025 and finally approved on 20 July 2026.

Supplier indemnities can no longer be the primary IP control. The exposure is direct and, in our experience, the indemnity limits in standard AI vendor contracts are constrained, so governance of outputs becomes part of IP risk management. On our assessment exposure varies by content type: regulated text generation has the lowest volume but the highest per-event cost, image and video generation the highest volume and a more

variable profile, and music generation a difficult substantial-similarity test applied to melodic patterns that systems can reproduce without explicit copying.

Exposure (our model). Per-event £50,000 (single claim, licensing resolution) to £480 million (multi-plaintiff class action, systemic infringement). Annual aggregate: £28 billion to £95 billion.

3.6 Agentic AI and autonomous action failures

On our model agentic AI is the fastest-growing category, and its governance lags.

Agentic systems execute multi-step tasks without approval for each action. Deloitte's 2026 survey of 3,235 leaders found agentic AI usage "poised to rise sharply in the next two years", while "only one in five companies has a mature model for governance of autonomous AI agents".

Agentic systems take actions in the world (sending emails, executing trades, modifying files, calling external APIs, instructing other agents) rather than producing outputs that people then act on. The point of failure moves from the handoff between system and person, where a reviewer can intercept a flawed output, to the handoff between systems, where an erroneous action can propagate before anyone notices.

Multi-agent chains compound the error. Production deployments increasingly chain specialised agents (retrieval, reasoning, tool execution, verification), and the failures of a chain are not the sum of the failures of its parts. An error in one agent's output can propagate through agents that treat it as authoritative input, gathering distortion at each handoff, and a verification agent calibrated against typical upstream output may miss anomalies a domain expert would catch.

The largest tail risk sits where actions are hard to reverse: trading, supply chain orchestration, infrastructure control, and any context where an agent can commit the organisation's resources.

Incidents involving agentic and AI-integrated tools have already been reported. In July 2025 the founder of SaaSr reported that a Replit coding agent had deleted a live production database during a code freeze, invented data, and wrongly said that the data could not be recovered. In August 2024 researchers showed that Slack AI could be made, through indirect prompt injection, to exfiltrate content from private channels, and in 2025 the agentic-AI vendor Serviceaide disclosed that a database it held for a health system had been left publicly accessible, exposing more than 483,000 patient records.

Exposure (our model). Per-event £500,000 (contained automation error, data corruption) to £2.1 billion (agentic failure in trading, infrastructure or supply chain systems). Annual aggregate: £82 billion to £280 billion. On our model this category grows at 40% to 60% a year through 2028.

3.7 AI-enabled fraud, deepfakes and social engineering

Synthetic media has made convincing fraud cheap. Surfshark's count of media-reported incidents put deepfake-related losses at \$1.56 billion by October 2025, over \$1 billion of it in 2025 alone, against about \$130 million for 2019 to 2023 combined. The marginal cost of a convincing synthetic identity, voice or video has fallen sharply since 2022, bringing techniques that once needed specialist resources within reach of individual fraudsters using consumer tools.

A Deloitte Center for Financial Services forecast projects US generative-AI-enabled fraud losses rising from \$12.3 billion in 2023 to \$40 billion by 2027, a compound annual growth rate of 32%. In early 2024 a finance employee in the Hong Kong office of the engineering group Arup paid about \$25 million to criminals after a video call in which the colleagues and executives present were deepfakes. The case matters less as a new technique than as proof that the attack can succeed against a large multinational with established payment controls, because synthetic participants on a live call evaded the cues that would raise suspicion in a text or voice attempt.

Defence is asymmetric. IBM's 2025 study found that one in six breaches (16%) involved attackers using AI, most often for phishing (37%) and deepfake impersonation (35%). Detection is improving but remains in an arms race with generation technologies that carry larger commercial incentives, and out-of-band verification, callback procedures and multi-party authorisation remain effective only with operational discipline that erodes under time pressure, which deepfake fraud is designed to create.

The modelled aggregate below covers all AI-enabled fraud and social engineering across institutions worldwide, and it is far larger than the losses so far reported or forecast: Surfshark counts only media-reported deepfake incidents, and Deloitte's forecast covers US generative-AI-enabled fraud alone. It should be read as a surface in the sense section 2 defines. The base-case expected value in section 5 (£389 billion) is likewise an indicative model output: it is about twelve times Deloitte's 2027 US forecast converted at the fixed rate in section 2 (about £31 billion), and the model has not been reconciled to reported loss data.

Exposure (our model). Per-event £95,000 (failed but costly business email compromise) to £120 million (synthetic identity fraud ring, multi-institution). Annual aggregate: £320 billion to £680 billion. On our model this is the second fastest-growing category, after agentic AI.

3.8 Misinformation, synthetic media and market manipulation

Generative AI changes the volume of disinformation more than its quality. Synthetic content can move financial markets, undermine public health campaigns, disrupt elections and damage institutional reputations. Skilled operators have always been able to produce persuasive disinformation; generative AI lowers the cost of each further variant, letting coordinated campaigns run at a scale where pattern detection or human review becomes uneconomic.

Market manipulation is hard to govern because it exploits the channels of legitimate market information. On 22 May 2023 an AI-generated image of an explosion near the Pentagon, spread by a verified account that falsely presented itself as linked to Bloomberg News, briefly moved markets, with the S&P 500 dipping about 0.3% before recovering. Adjacent vectors that exploit the same trust, such as fabricated regulatory statements or synthetic executive audio attached to earnings communications, are plausible extensions of demonstrated capability, though documented high-profile institutional incidents remain few.

Response time shapes the reputational damage. An attack rebutted within hours can be contained, while the same attack identified after a day of viral spread can do lasting damage, because the initial content reaches audiences who never see the rebuttal.

Exposure (our model). Per-event £1.2 million (contained attack, rapid rebuttal) to £8.5 billion (coordinated campaign triggering regulatory action, share-price collapse, or electoral interference with downstream effects). Annual aggregate: £75 billion to £195 billion.

3.9 Shadow AI and unsanctioned deployments

Shadow AI is common among breached organisations, and it is associated with higher breach costs. Shadow AI is AI adopted by employees, departments or contractors without institutional approval, security review or data governance. Of the breached organisations in IBM's 2025 study, one in five reported a breach due to shadow AI, and those with high levels of shadow AI saw average breach costs \$670,000 higher than others.

Shadow AI is one route by which AI escapes the organisation's records, but not the only one. AI embedded in purchased platforms, and pilots that become production without review, also bypass the channels compliance watches, so measures of employee adoption count only part of the unrecorded estate.

The drivers are largely benign. Employees meet friction in approved tooling, find consumer tools that remove it, and adopt them because the productivity benefit is immediate while the governance risk is diffuse. Treating shadow AI mainly as an enforcement problem misreads that dynamic: reducing friction in approved tooling matters at least as much as monitoring for unauthorised use.

The data loss runs through routes conventional data loss prevention does not monitor. An employee who pastes a strategy document into a consumer assistant for summarisation triggers no egress alert, because the document goes through an authorised browser session to an authorised cloud service; the loss occurs at the point of submission. Where shadow AI was involved, IBM found customer personal data compromised in 65% of incidents, against 53% across all breaches, and intellectual property, though compromised less often, carries the highest cost per record.

What separates shadow AI from earlier shadow IT is that data submitted to an AI service may be used to train models, depending on the tool's terms. Enterprise tiers commonly exclude customer data from training, but consumer tiers may use it by default, and the employee making the submission rarely reviews the terms. In regulated industries the result can be a compliance breach and a permanent loss of control over the confidentiality of the information.

Exposure (our model). Per-event £85,000 (contained data exposure) to £45 million (large-scale IP or personal data leakage through an unsanctioned tool, with regulatory enforcement). Annual aggregate: £140 billion to £310 billion.

3.10 Supply chain and third-party AI risk

Organisations that buy AI-enabled products inherit their suppliers' governance failures. In IBM's 2025 study, supply chain compromise was among the most costly breach vectors, averaging \$4.91 million and taking the longest to identify and contain, and compromised apps, APIs and plug-ins were among the most common routes into AI systems.

The dependency graph runs through layers organisations have not historically inventoried. A SaaS feature may draw on a third-party foundation model, training data

from further parties, embedding services, vector databases, fine-tuning datasets and prompt templates that change over time, and a failure at any layer can create exposure for the deploying organisation, often with limited contractual recourse.

The AI Act puts duties on deployers that depend on the provider's information.

Article 26 requires deployers of high-risk systems to use them in line with the instructions for use, ensure that input data they control is relevant and sufficiently representative, monitor operation, and inform the provider and the market surveillance authority of serious incidents (Article 26(1), (4) and (5)). Each of these depends on documentation that providers may treat as proprietary.

Contractual mitigation transfers imperfectly. In our experience supplier indemnities are commonly capped at the contract value or a multiple of it, which is rarely enough to cover a class action settlement, and regulatory fines may not be recoverable under an indemnity at all.

Audit rights, where granted, often exclude inspection of training data, model architecture and development process, which are the elements conformity assessment requires. A deployer is entitled to the provider's Article 13 instructions for use; the Annex IV technical documentation is kept for the national competent authorities (Articles 11 and 18), so a deployer obtains it only by contract. Requiring suppliers to deliver structured Annex IV documentation as a contractual deliverable is gaining ground in regulated industries but is far from universal.

Exposure (our model). Per-event £250,000 (third-party error, contained remediation) to £380 million (systemic third-party failure across the supply chain, multi-party litigation). Annual aggregate: £65 billion to £160 billion.

3.11 Workforce displacement, HR liability and labour litigation

AI-driven restructuring without governance exposes employers under three existing regimes.

They are collective consultation under national labour law, such as the UK's Trade Union and Labour Relations (Consolidation) Act 1992 (TULRCA), the EU's Collective Redundancies Directive and their equivalents; individual unfair dismissal protection; and anti-discrimination law. The AI Act adds to them: AI used for recruitment, promotion, termination, task allocation or monitoring of workers is high-risk under Annex III point 4, and an employer must inform workers' representatives and affected workers before a high-risk AI system is used at the workplace (Article 26(7)) from 2 December 2027, subject to the Article 111(2) transition in section 3.4.

In our assessment the highest-risk context is not algorithmic dismissal in isolation but the use of AI in restructuring where the employer cannot explain why individual employees were selected. Where a system scored employees against criteria the employer cannot fully explain or defend as objectively related to the role, or where consultation did not disclose the system's role, a tribunal may treat unexplained AI involvement as evidence of an unfair process rather than a neutral element of the decision.

Collective redundancy exposure arises at population level. An employer proposing to dismiss as redundant 20 or more employees at one establishment within 90 days or less must consult, and must disclose in writing "the proposed method of selecting the employees who may be dismissed" (TULRCA section 188(1) and (4)(d)); where AI-assisted selection is used, we recommend that the disclosure explain the system's role

and the criteria it applied. The Employment Rights Act 2025 adds an alternative threshold measured across the whole employer, to be set by regulations under a new section 195A, but as at 11 October 2026 that change was in force only for the purpose of making those regulations. Failure to comply can lead to a protective award of up to 180 days' pay for each affected employee (section 189(4), as amended by the Employment Rights Act 2025 with effect from 6 April 2026).

Exposure (our model). Per-event £40,000 (individual unfair dismissal with an AI decision element) to £180 million (large-scale collective redundancy with AI governance failure and class action). AI Act fines for non-compliant high-risk employment AI reach EUR 15 million or 3% of worldwide annual turnover, whichever is higher (Article 99(4)), and EUR 35 million or 7% where the use is prohibited, such as inferring employees' emotions from biometric data (Article 5(1)(f)). Annual aggregate: £32 billion to £88 billion.

3.12 Critical infrastructure and national security

This category combines two kinds of risk. First-order risks are AI failures within infrastructure operations, which conventional safety engineering, extended to AI-specific failure modes, can govern. Second-order risks are attacks on infrastructure that use AI capability from outside, where the threat actor is external and defence depends on cross-sector coordination that few jurisdictions have institutionalised.

The population of obligated entities has grown. Beyond the traditional operators of essential services in energy, transport, water, healthcare, banking and financial market infrastructure, NIS2 in the EU, CIRCIA in the US and equivalents in the UK, Japan and Australia can reach a mid-sized supplier to an essential-service operator, a logistics provider or a software vendor whose products run in critical environments, and an organisation may fall within them without having noticed. Their incident-reporting duties apply to AI-related incidents as to any other (NIS2 is in force; CIRCIA will apply once its final rule takes effect).

Autonomous vulnerability discovery raises the urgency. Systems that analyse codebases, identify previously unknown vulnerabilities and produce working exploit code were reported through 2025 and 2026 in research and adversarial settings. In our assessment the asymmetry is sharp, because protective AI capability is concentrated among a small number of well-resourced defenders while a much larger population of operators lacks it.

Exposure (our model). Per-event £8 million (localised infrastructure fault, AI-assisted) to £380 billion (coordinated AI-facilitated attack on interconnected critical national infrastructure). Annual aggregate: £120 billion to £850 billion. The base-case expected value, about £140 billion (section 5), sits near the low end of that range; the upper bound is set almost entirely by tail-risk scenarios.

3.13 Reputational damage and institutional trust

In our assessment AI failures carry higher public salience than other technology failures. We expect a breach exposing a given number of records to draw less attention than an AI bias incident affecting fewer people, because the bias incident raises concern about systemic decision-making rather than an isolated security failure. That asymmetry rewards organisations that make their governance visible, through public commitments, transparency reports and third-party audits, and penalises those that manage it as a purely internal matter.

The trust that AI failure erodes most directly is trust in the organisation's expert judgement. When a hospital's AI triage system produces a publicly disclosed error, the damage can extend to the hospital's clinical decision-making generally, and when a bank's lending system is found discriminatory, to its risk management generally; that generalisation produces the upper-bound values in this category.

We expect recovery to be slower than from comparable non-AI incidents, partly because the system often stays in operation during the recovery period and keeps producing outputs that are scrutinised against the original failure.

Exposure (our model). Per-event £2.5 million (contained incident, effective crisis communications) to £4.2 billion (existential reputational event for a major institution). Annual aggregate: £180 billion to £460 billion.

4. The aggregate surface runs to trillions of pounds a year, and tail risk sets its upper bound

On our model the thirteen categories together produce an annual exposure surface of £1,662 billion at the low end and £4,428 billion at the high end. Critical infrastructure, through its tail-risk profile, is the largest single contributor to the upper bound (£850 billion) and to the gap between low and high (£730 billion of £2,766 billion), and tail-risk scenarios beyond the high case would be higher still. As section 2 sets out, the figures measure the institutional risk surface, not realised loss.

Aggregate annual exposure by risk category (our model)

Risk category	Per-event range	Annual low (£bn)	Annual high (£bn)	Annual growth 2026 to 2028
Data privacy and security breaches	£3.5m to £8.5m	310	580	18%
Algorithmic bias and discrimination	£850k to £620m	95	210	24%
Hallucination and decision errors	£120k to £85m	55	140	31%
Regulatory and compliance failures	£180k to £2.8bn	160	380	28%
IP infringement and copyright	£50k to £480m	28	95	22%
Agentic AI failures	£500k to £2.1bn	82	280	52%
AI-enabled fraud and deepfakes	£95k to £120m	320	680	45%
Misinformation and market manipulation	£1.2m to £8.5bn	75	195	29%
Shadow AI and unsanctioned use	£85k to £45m	140	310	19%
Supply chain and third-party AI risk	£250k to £380m	65	160	33%

Risk category	Per-event range	Annual low (£bn)	Annual high (£bn)	Annual growth 2026 to 2028
Workforce and HR liability	£40k to £180m	32	88	17%
Critical infrastructure	£8m to £380bn	120	850	38%
Reputational damage	£2.5m to £4.2bn	180	460	21%
Total		1,662	4,428	29% (simple average)

5. Expected loss is concentrated in five categories

Probability-weighting changes the order. In the base case the largest expected losses are in shadow AI (£390 billion), AI-enabled fraud and deepfakes (£389 billion), data breaches (£324 billion), reputational damage (£182 billion) and regulatory non-compliance (£148 billion). Agentic AI, at £68 billion, is smaller today but grows fastest on our model, at 52% a year.

Expected value sensitivity by scenario (our model)

Category	Low EV (£bn)	Base EV (£bn)	High EV (£bn)	Spread (£bn)	Governance leverage
Data privacy and breaches	135	324	417	282	Moderate
Algorithmic bias	18	48	70	52	High
Hallucination and errors	58	117	138	80	Moderate
Regulatory non-compliance	65	148	206	141	Very high
IP and copyright	22	41	60	38	High
Agentic AI failures	20	68	123	103	Very high
AI fraud and deepfakes	226	389	479	253	Moderate
Misinformation and markets	51	117	180	129	High
Shadow AI	243	390	434	191	Low to moderate
Supply chain AI risk	20	51	76	56	High
Workforce and HR liability	26	41	62	36	High
Critical infrastructure	50	140	275	225	Very high
Reputational damage	97	182	243	146	Moderate

One category's figures do not reconcile. The shadow AI base-case and high-risk expected values (£390 billion and £434 billion) sit above that category's annual aggregate in section 4 (£140 billion to £310 billion), the scale of loss the model puts at risk in that category. These figures, and the portfolio results that draw on them, should be read as indicative rather than precise. At the top of its aggregate range (£310 billion) shadow AI would still rank among the five largest categories, though third rather than first.

The spread measures size as well as leverage. The spread is the distance between the low-risk and high-risk expected values, in pounds, so it grows with the size of a category as well as with the scope for governance to change it. Data breaches have the widest spread (£282 billion) but only moderate leverage, while workforce liability has a narrow spread (£36 billion) and high leverage. Leverage, which is proportional, is the measure to use in deciding where to invest.

6. Governance leverage, not exposure, shows where investment pays

Leverage and headline exposure point to different places. The leverage rating identifies where investment produces the largest proportional reduction in expected loss, which is not always where the headline figures that dominate boardroom discussion point. Of the thirteen categories, eight show very high or high leverage, four moderate and one low to moderate.

Very high: agentic AI, regulatory non-compliance and critical infrastructure.

Agentic AI shows the steepest probability-to-loss curve on our model. A ten-percentage-point rise in governance maturity reduces expected value by around £35 billion a year, about half the category's base-case expected value of £68 billion, which makes agentic AI the single highest-return target in 2026. That is an indicative model output, steep compared with the portfolio result in section 8 (a 38% to 52% reduction for a 40-point rise), and it depends on the non-linear curve described next.

The return is non-linear: the first ten points of maturity produce the largest absolute reduction, because they address the failure modes that are easiest to remediate (action logging, approval gates, sandboxed execution), while later improvements address harder ones (error propagation through orchestration, emergent behaviour, alignment drift).

Regulatory non-compliance is rated very high because we assume enforcement probability falls as governance maturity rises, which makes this risk partly positional, measured against the governance of peers as well as against the requirements. We expect regulators to direct attention towards organisations with visible gaps, both because they present the strongest cases and because action against them has the greatest deterrent effect, so organisations with documented Annex IV files, conformity records, post-market monitoring evidence and incident-reporting arrangements should face lower enforcement probability even where their deployments resemble those of less-governed competitors. As at 11 October 2026 no fine under the AI Act had been confirmed by any regulator or official source to test this assumption.

Critical infrastructure has one of the widest spreads (£225 billion) but a low base probability (28% on our model), the classic tail-risk profile of low frequency and extreme severity. Its leverage is very high despite the external threat, because on our assessment investment in AI-assisted vulnerability discovery, automated patch management and segmented operational technology materially reduces the exploitable surface. The difficulty is distributional: the institutions facing the largest exposure are not always those with the largest governance budgets, and cross-sector dependencies create externalised risk no single operator can fully address.

High: algorithmic bias, IP, misinformation, supply chain and workforce liability.

Algorithmic bias combines high leverage with a narrow spread (£52 billion). Bias-testing protocols and audit methods have had a decade of development in financial services, which in our assessment makes this the most operationally tractable governance domain even though its headline exposure is lower.

Moderate: fraud and deepfakes, data breaches, hallucination and reputational damage.

In these categories investment reduces exposure but cannot remove it, because part of each lies outside the institution's control: adversaries drive fraud and data breaches, while hallucination arises from probabilistic generation and reputational damage from public reaction.

Alfraud and deepfakes is the most externally driven category. Institutions can reduce the chance that each attempt succeeds through authentication architecture, training and anomaly detection, but they cannot reduce attack volume, which adversarial economics outside their control sets, and the defensive ceiling depends on how fast defences spread across the target population rather than on any single organisation's investment.

Data breaches have the largest absolute spread (£282 billion), reflecting both a high baseline probability and real scope for investment to reduce incidence. Leverage is limited above by attack vectors the institution has not yet seen, and below by the regulatory floor of minimum compliance that every institution must meet to operate.

Low to moderate: shadow AI. *Shadow AI* has a high base-case probability (88% on our model) and the lowest leverage, because human behaviour rather than system design drives it. The rating may understate what is achievable where an organisation pairs detection with reduced friction in approved tooling. The decisive variable is whether approved tooling is capable and accessible enough that employees do not look elsewhere; where it is, incidence falls to a structural floor that detection alone cannot reach.

7. Sector, scale and institution type change which categories dominate

No single governance template suits every organisation, so investment should be calibrated to the categories that dominate the organisation's own deployment portfolio.

7.1 Corporations

Large enterprises face the broadest surface. Data breaches, regulatory compliance, agentic AI failures and fraud are their highest-priority categories, but the profile differs by sector. Financial services face elevated regulatory and fraud exposure, the largest agentic tail risk in trading and lending, and heavy compliance overlap because AI Act obligations sit alongside DORA, MiFID II and MAR.

Healthcare and life sciences face elevated exposure to hallucination and decision errors, because clinical and diagnostic deployments carry the longest tail of patient-harm liability, with data-privacy exposure amplified where HIPAA, the GDPR's Article 9 rules on special-category data and national health AI frameworks meet.

Manufacturing and infrastructure operators carry concentrated tail risk in critical infrastructure but lower exposure in consumer-facing categories. Technology and SaaS companies face the highest exposure in agentic failures, IP infringement and supply chain risk, because their products embed AI that customers then deploy under obligations the provider has limited ability to verify. For retail and consumer services, fraud and reputational exposure are highest.

Scale changes the profile further. Large multinationals face the broadest jurisdictional surface but usually have the resources for proportionate governance, so their per-event costs are highest while their exposure relative to revenue is often lower. Mid-market firms are large enough to attract scrutiny and litigation, but typically lack the compliance capacity or buying power to demand structured Annex IV documentation; small enterprises bear a higher governance cost as a share of revenue and are most likely to rely on consumer tools that create shadow AI exposure, though their per-event costs are bounded by size.

7.2 Governments

Governments face lower fraud exposure but far higher critical infrastructure and misinformation exposure. They operate or regulate critical infrastructure, which places them at the front line of second-order adversarial risk, and they are the principal targets of synthetic media, particularly during elections. Their use of automated decision-making in welfare, criminal justice, immigration and tax concentrates exposure from unlawful or discriminatory decisions with high political salience.

Two cases show the scale. Australia's Robodebt scheme led to an A\$112 million class settlement approved by the Federal Court in 2021, alongside some A\$1.76 billion of debts withdrawn, and a further A\$548.5 million settlement agreed in 2025 and approved in 2026. The Dutch SyRI welfare-fraud risk system was halted by The Hague District Court in 2020 as incompatible with the right to private life under Article 8 of the European Convention on Human Rights.

The governance response differs because government separates policy-making, regulation and operational deployment. Policy-makers regulate capabilities that change faster than the legislative cycle, and so rely on standards bodies (NIST, ISO, CEN-CENELEC) for substantive content. Regulators must monitor deployments across sectors with limited specialist capacity, and operational deployers face procurement requirements (data sovereignty, security clearance, accessibility, audit transparency) that differ materially from commercial AI procurement.

Cross-government coordination has grown since 2023 through the G7 Hiroshima AI Process, the OECD AI Principles and AI Incidents Monitor, the network of AI safety institutes, and arrangements such as the EU-US Trade and Technology Council. None has yet produced binding cross-jurisdictional incident response, which remains the most acute gap for transnational threats, particularly synthetic media, where an attack on one jurisdiction can originate in another.

7.3 Non-profit institutions

In our assessment non-profits carry the lowest absolute exposure but the highest proportional impact. They often rely on consumer tools, which creates elevated shadow AI risk with little internal capacity to manage it, and they frequently work in sensitive contexts (humanitarian assistance, mental health support, legal aid, advocacy for marginalised groups) where AI failures cause harms that are severe but hard to quantify. They hold concentrated beneficiary data that is a valuable target, with privacy obligations at least as stringent as those of commercial enterprises, and their public-interest mission amplifies reputational exposure, because donor and beneficiary trust can fall quickly even after a contained incident.

The difficulty is structural as well as financial. Non-profit boards are typically drawn from mission-related expertise rather than technology governance, board-level AI literacy is often lower than in commercial enterprises, and the routes to developing it are less mature. Procurement budgets rarely stretch to enterprise tooling with governance features, which drives reliance on consumer alternatives whose terms may allow training on submitted data and offer limited recourse, and international non-profits face added complexity across the AI Act's extraterritorial reach, US state frameworks and beneficiary countries that may have no framework at all.

Three responses are emerging. Sector-wide procurement consortia give individual non-profits access to enterprise tooling at affordable scale with shared due diligence; peer-learning networks let organisations without dedicated compliance staff share assessment methods, incident playbooks and policy templates; and philanthropic funders are beginning to make governance maturity a grant condition, creating pressure for investment the organisation's own budget could not sustain.

8. On our model, governance returns 8 to 22 pounds for each pound invested

On our model, structured governance returns £8 to £22 in avoided loss for every £1 invested, depending on organisation size and sector. Structured governance here means risk-ownership frameworks, conformity assessments, model inventories, audit trails and data classification enforcement. This is a modelled range, not an observed outcome.

Across the portfolio, moving from 30% to 70% governance maturity reduces aggregate expected annual loss on our model by roughly 38% to 52%, with category-level reductions varying by leverage class. Governance maturity is the percentage input defined in section 2.

Third-party evidence points the same way, but measures something narrower.

Gartner projects that effective governance technologies could reduce regulatory expenses by around 20%, and its survey of 360 organisations in the second quarter of 2025 found that organisations that deployed AI governance platforms were 3.4 times more likely to achieve high effectiveness in AI governance than those that did not. Neither figure measures avoided loss.

Reaching a basic baseline costs little relative to AI spend. We estimate the investment required to reach a basic governance baseline at around 0.8% to 1.4% of annual AI-related spend, and compliance personnel costs for AI Act readiness at \$150,000 to \$250,000 per FTE for specialised roles, with two to five FTEs for large organisations. The 0.8% to 1.4% is the modelled cost of reaching a minimum baseline, not a recommended governance budget: the 15 to 20% of the AI budget that Enterprise Shift recommends for large enterprises is a target for the full allocation to governance, so the two figures should not be compared directly.

Where to start, and what it costs

The exposure surface measures what ungoverned AI puts at risk; how much of it becomes loss depends on three variables. They are the pace at which agentic adoption outruns governance maturity, the course of adversarial capability (particularly deepfakes and synthetic identity), and the consistency with which the EU AI Act, US sector frameworks and other domestic regimes are enforced. Only the first is within an organisation's control, which is why the order of work matters.

First: find the ungoverned systems and the dominant categories. Apply the five conditions in section 1 to each AI system the organisation runs, including AI embedded in purchased platforms and tools adopted without approval, and identify which of the thirteen categories dominate the portfolio. That view sets the priorities for everything that follows.

Next: the cheap, high-leverage controls. For agentic AI, put basic action logging, human approval gates for high-stakes operations, and sandboxed execution in place, since on our model these first increments produce the largest reduction. For material commitments, establish deepfake-resistant authentication: out-of-band verification, callback procedures on independently held contact details, and multi-party authorisation. For shadow AI, make approved tooling capable and accessible enough that employees stop looking elsewhere, and pair it with detection; enforcement alone pushes use further out of view.

Then, for as long as the systems run: documentation and supplier evidence.

Treat conformity assessment and the Annex IV technical documentation as continuing obligations, since Article 11(1) requires the documentation to be kept up to date, and build deployer-side evidence for third-party AI by contracting for the documentation that Article 13 does not provide. On our model reaching a basic baseline costs around 0.8% to 1.4% of annual AI-related spend, with two to five specialised FTEs in a large organisation.

On timing. The AI Act's high-risk obligations apply from 2 December 2027 for Annex III systems and 2 August 2028 for Annex I systems, subject to the Article 111(2) transition, and as at 11 October 2026 no fine under the AI Act had been confirmed by any regulator or official source. Most of the exposure described here, however, does not wait for the AI Act: breaches, fraud, litigation, employment claims and reputational damage arise under rules that already apply.

The categories with the largest absolute exposure are not always those where investment reduces expected loss most. An organisation that has mapped its systems against the thirteen categories and the leverage ratings can say where its governance spend does most, and why it spends there rather than where the headline figures point.

This paper is a modelled assessment for planning and budgeting, and is not legal, financial or actuarial advice. Its exposure figures are modelled estimates of an institutional risk surface, not predictions of realised loss, individual exposure will vary materially, and the legal position is stated as at 11 October 2026. Take qualified professional advice before acting on it.

Basis and Standing

Binding law. The AI Act (Regulation (EU) 2024/1689) as amended by Regulation (EU) 2026/1744, in force 27 July 2026: Article 4 (as replaced), Article 5, Articles 10, 11, 13, 15, 16, 18, 26, 50 and 72, Article 99 on penalties, Article 111(2) on systems already on the market, and Article 113 on application dates (2 December 2027 for Annex III high-risk systems, 2 August 2028 for Annex I). GDPR Articles 5(1)(f), 32, 33 and 83(4) and (5). In the UK, TULRCA sections 188 and 189, with the maximum protective award raised to 180 days by the Employment Rights Act 2025 from 6 April 2026. In the US, the SEC's 2023 cybersecurity disclosure rule (Form 8-K Item 1.05); Title VII, the ADEA, ECOA and the Fair Housing Act are named but not analysed. NIS2 is in force; CIRCIA's final rule had not taken effect as at 11 October 2026.

Case law and settlements. Mata v. Avianca (sanctions for AI-fabricated citations); EEOC v. iTutorGroup (consent decree); Bartz v. Anthropic (class settlement, finally approved 20 July 2026) and Kadrey v. Meta (fair use in training); Prygodicz v Commonwealth (No 2) and the later Knox settlement (Robodebt); NJCM v The State of the Netherlands (SyRI). They illustrate exposure in each category; none decides how the AI Act applies.

Regulator guidance. None is relied on. The EEOC's 2023 technical assistance on AI under Title VII is mentioned only as removed.

Not a regulatory metric. The exposure model is Standard Intelligence's own: the per-event ranges, at-risk populations, annual aggregates, growth rates, scenario probabilities, expected values, governance leverage ratings, the governance maturity percentage, the £8 to £22 return, the 38% to 52% reduction, the 0.8% to 1.4% baseline cost and the FTE estimate. So is the definition of an ungoverned system. The recommended controls (action logging, approval gates, sandboxing, deepfake-resistant

authentication, friction reduction for shadow AI, contractual Annex IV deliverables, disclosure of an AI system's role in redundancy consultation) are our design; the documentation control reflects Article 11(1), supplier evidence reflects the gap between Article 13 and Articles 11 and 18, and the consultation disclosure reflects TULRCA section 188(4)(d).

Surveys and research, and what they do and do not support. The Allianz Risk Barometer measures the concerns of risk experts, not losses. IBM's breach studies are vendor research on organisations that suffered a breach (600 in the 2025 study); several of their detailed figures are taken here from secondary summaries, noted in Sources. Gartner's figures are projections and a relative-likelihood survey result; Surfshark's count covers media-reported deepfake incidents only; Deloitte's figures are a survey of leaders and a forecast. Babic et al. (2025), the International AI Safety Report, Parasuraman and Manzey (2010) and Carlini et al. (2021) support the specific points for which they are cited. None of these sources validates the model's aggregates, which rest partly on structured expert judgement, and some category figures are indicative outputs that do not reconcile with each other or with reported losses (sections 3.7, 5 and 6).

Our experience. That AI Act duties of the bias-testing kind are new to most enterprises (section 3.2), that indemnity limits in standard AI vendor contracts are constrained (section 3.5) and that supplier indemnities are commonly capped at the contract value or a multiple of it (section 3.10) rest on Standard Intelligence's engagements, not on a published survey.

Do not read the headline as a forecast. The £1.6 trillion to £4.4 trillion is a measure of the risk surface, and it should not be used to allocate governance spend; the leverage ratings are the basis for allocation.

Enforcement position. As at 11 October 2026, no fine under the AI Act had been confirmed by any regulator or official source.

Sources

Legislation

1. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024, consolidated version of 27 July 2026. <https://eur-lex.europa.eu/eli/reg/2024/1689/2026-07-27/eng>
2. Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI), OJ L, 2026/1744, 24.7.2026; in force 27 July 2026. <https://eur-lex.europa.eu/eli/reg/2026/1744/oj>
3. Regulation (EU) 2016/679 (General Data Protection Regulation), OJ L 119, 4.5.2016, p. 1, Articles 5, 32, 33 and 83. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
4. Trade Union and Labour Relations (Consolidation) Act 1992, sections 188 and 189, as amended by the Employment Rights Act 2025. <https://www.legislation.gov.uk/ukpga/1992/18/section/188>

[gov.uk/ukpga/1992/52/section/188](https://www.gov.uk/ukpga/1992/52/section/188); <https://www.legislation.gov.uk/ukpga/1992/52/section/189>

5. US Securities and Exchange Commission (2023) Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure, final rule, adopted 26 July 2023 (Form 8-K Item 1.05). <https://www.sec.gov/newsroom/press-releases/2023-139>

Surveys, forecasts and research

6. Allianz Commercial (2026) Allianz Risk Barometer 2026: Artificial Intelligence. <https://commercial.allianz.com/news-and-insights/expert-risk-articles/allianz-risk-barometer-2026-ai.html>
7. IBM (2025) Cost of a Data Breach Report 2025, IBM Security and Ponemon Institute; press release, 30 July 2025 (600 organisations, breaches March 2024 to February 2025). <https://newsroom.ibm.com/2025-07-30-ibm-report-13-of-organizations-reported-breaches-of-ai-models-or-applications,-97-of-which-reported-lacking-proper-ai-access-controls>
8. IBM (2026) Cost of a Data Breach Report 2026 (global average \$4.99 million). <https://www.ibm.com/reports/data-breach>
9. Secondary summaries of the IBM 2025 report, for figures not in the IBM press release: Baker Donelson, Ten key insights from IBM's Cost of a Data Breach Report 2025 (AI-assisted attacks by type; regulatory fines in the US average). <https://www.bakerdonelson.com/ten-key-insights-from-ibms-cost-of-a-data-breach-report-2025>; Kiteworks, How Shadow AI Costs Companies \$670K Extra (supply chain compromise cost and time; personal data and intellectual property figures). <https://www.kiteworks.com/cybersecurity-risk-management/ibm-2025-data-breach-report-ai-risks/>
10. Deloitte AI Institute (2026) State of AI in the Enterprise 2026: The Untapped Edge (3,235 leaders surveyed August to September 2025). <https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html>
11. Deloitte Center for Financial Services (2024) Generative AI is expected to magnify the risk of deepfakes and other fraud in banking, 29 May. <https://www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html>
12. Gartner (2026) Global AI Regulations Fuel Billion-Dollar Market for AI Governance Platforms, press release, 17 February. <https://www.gartner.com/en/newsroom/press-releases/2026-02-17-gartner-global-ai-regulations-fuel-billion-dollar-market-for-ai-governance-platforms>
13. International AI Safety Report (2026) Second International AI Safety Report, led by Yoshua Bengio, 3 February 2026. <https://internationalaisafetyreport.org>
14. Surfshark Research (2025) AI Drives Deepfake Losses to \$1.56 Billion, 21 October (media-reported incidents). <https://surfshark.com/research/chart/ai-deepfake-losses>
15. Boris Babic, I. Glenn Cohen, Ariel Dora Stern, Yiwen Li and Melissa Ouellet, "A general framework for governing marketed AI/ML medical devices", npj Digital Medicine 8, 328 (2025), 31 May. <https://www.nature.com/articles/s41746-025-01717-9>
16. US Food and Drug Administration, Artificial Intelligence-Enabled Medical Devices list (1,247 devices at the 10 July 2025 update, as reported by the FDA Law Blog). <https://www.fda.gov/medical-devices/artificial-intelligence-enabled-medical-devices/list-artificial-intelligence-enabled-medical-devices>; <https://www.thefdalawblog.com/2025/07/fdas-latest-lists-for-digital-health-technologies/>

17. Raja Parasuraman and Dietrich H. Manzey, "Complacency and Bias in Human Use of Automation: An Attentional Integration", *Human Factors*, vol. 52(3) (2010), pp. 381-410 (subscription access). <https://journals.sagepub.com/doi/10.1177/0018720810376055>
18. Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski and others, "Extracting Training Data from Large Language Models", *USENIX Security Symposium* (2021). <https://arxiv.org/abs/2012.07805>
19. CMS (2024) GDPR Enforcement Tracker Report 2024, executive summary (fines to 1 March 2024). <https://cms.law/en/lux/publication/gdpr-enforcement-tracker-report-2024/executive-summary>

Incidents

20. Fortune (2025) "AI coding tool Replit wiped database, called it a catastrophic failure", 23 July. <https://fortune.com/2025/07/23/ai-coding-tool-replit-wiped-database-called-it-a-catastrophic-failure>; The Register (2025), 22 July. https://www.theregister.com/2025/07/22/replit_saastr_response/
21. PromptArmor (2024) Data Exfiltration from Slack AI via indirect prompt injection, August. <https://promptarmor.com/resources/data-exfiltration-from-slack-ai-via-indirect-prompt-injection>
22. HIPAA Journal (2025) Serviceaide data breach (483,126 Catholic Health patients). <https://www.hipaajournal.com/serviceaide-data-breach/>
23. University of Hong Kong Data Protection Office (2024) Engineering giant Arup falls victim to deepfake scam, 17 May (reporting Arup's confirmation to CNN). <https://dpo.hku.hk/news/engineering-giant-arup-falls-victim-to-deepfake-scam>
24. Associated Press (2023) Fact Focus: fake image of Pentagon explosion briefly sends jitters through stock market, 23 May (as republished by West Hawaii Today). <https://www.westhawaiiitoday.com/2023/05/23/nation-world-news/fact-focus-fake-image-of-pentagon-explosion-briefly-sends-jitters-through-stock-market/>

Cases and settlements

25. *Mata v. Avianca, Inc.*, 678 F.Supp.3d 443 (S.D.N.Y. 2023). Sanctions order, Castel J., 22 June 2023. <https://www.law.berkeley.edu/wp-content/uploads/2025/12/Mata-v-Avianca-Inc.pdf>
26. *EEOC v. iTutorGroup, Inc.*, No. 1:22-cv-02565 (E.D.N.Y.). Consent decree approved 2023; \$365,000 settlement. <https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit>
27. *Bartz v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal.). \$1.5 billion class settlement, preliminary approval by Alsup J., 25 September 2025; final approval by Martinez-Olguin J., 20 July 2026. <https://www.jurist.org/news/2026/07/judge-approves-record-1-5-billion-settlement-involving-anthropic/>
28. *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417 (N.D. Cal.). Summary judgment on fair use in AI training.
29. *Prygodicz v Commonwealth of Australia* (No 2) [2021] FCA 634, 11 June 2021 (A\$112 million settlement; approximately A\$1.763 billion of debts withdrawn). <https://www.judgments.fedcourt.gov.au/judgments/Judgments/fca/single/2021/2021fca0634>
30. *Knox v Commonwealth of Australia*, VID982/2024 (Federal Court of Australia), with *Prygodicz*, VID1252/2019: notice of proposed settlement of A\$548,500,000. <https://www.servicesaustralia.gov.au/sites/default/files/2025-11/notice-of-proposed-settlement.pdf>

31. NJCM and Others v The State of the Netherlands, The Hague District Court, 5 February 2020, C/09/550982/HA ZA 18-388, ECLI:NL:RBDHA:2020:1878 (SyRI).

Regulator guidance

32. EEOC, technical assistance “Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964” (EEOC-NVTA-2023-2), 18 May 2023, removed from the EEOC website in January 2025; archived copy. https://data.aclum.org/storage/2025/01/E0CC_www_eeoc_gov_laws_guidance_select-issues-assessing-adverse-impact-software-algorithms-and-artificial.pdf

All Article, Annex and Recital references are to Regulation (EU) 2024/1689 (the AI Act), as amended by Regulation (EU) 2026/1744, unless otherwise stated.



Standard
Intelligence

GET IN TOUCH

Map your exposure to the leverage that
reduces it.

EMAIL

hello@standardintelligence.com

WEB

standardintelligence.com

LOCATION

Manchester, United Kingdom

AI Use: AI tools were used in research, editorial review and drafting support. All sources were retrieved and checked against the published record by the author, who holds editorial responsibility for the contents and for the conclusions drawn. This disclosure is made voluntarily. To the extent the Article 50(4) obligation in Regulation (EU) 2024/1689 applies to this paper as text published to inform the public on matters of public interest, the exception for content that has undergone human review and for which a person holds editorial responsibility applies.