Ius Laboris UK Global HR Lawyers
LEWIS SILKIN

AI safety measures: a comparative chart

As technology continues to develop rapidly, legislators and regulators around the world are racing to keep up – or indeed catch up – by implementing measures to protect those whose interests are affected by AI systems.

Unsurprisingly there is no global standard on AI regulation. But we when we look at the existing and proposed safeguards intended to protect individuals where AI systems are introduced into the workplace, clear categories emerge.

Set out in the table below is an analysis of how measures in key existing and proposed legislation could be categorised on this basis.

This is by no means a comprehensive list – draft legislation is being debated and progressed around the world. For example, Canada's AI and Data Act envisages a risk based approach to regulation; and the US Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence supports the creation of standards for trustworthy AI

In the UK the new Labour government has indicated an intention to legislate to regulate those developing powerful AI tools. But whether wider regulation will be on the cards (reflecting the TUC draft bill or the new private member's bill) remains to be seen. It will be interesting to see the extent to which the "Brussels effect" sees the approach taken in the EU spread around the world.

Understanding the columns:

- **Consultation about use**: Is there a requirement to consult with employees/unions about the use of workplace AI systems?
- **Accountability and governance**: Are there specific roles, frameworks, policies or reporting structures required to manage the use of AI in the workplace?
- **Impact assessments**: Is there a requirement to undertake an impact assessment prior to use?
- **Auditing and monitoring**: What steps are required to ensure that the AI system is safe to use? Auditing meaning independent evaluation and monitoring meaning regular non-independent evaluation.
- **Transparency and explainability**: What level of openness is required about the use of AI and / or how the system's decisions are reached? Transparency meaning that the use and purpose of AI is disclosed. Explainability meaning that how an outcome was reached is disclosed.
- **Human oversight / intervention**: What human involvement is required? Human oversight meaning someone is required to oversee the operation of the system. Human intervention meaning that a human interposed between the AI system expressing preferences and the final decision.
- **Contestability**: Are measures required to enable an individual affected by an AI decision to be able to challenge it effectively?

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
UK Government White Paper: principles for regulators	Framework of five principles to guide regulators, rather than legislation: 1. Safety, security and robustness 2. Appropriate transparency and explainability 3. Fairness 4. Accountability and governance 5. Contestability and redress Focus is on regulating the use not the technology. Regulators' strategic responses published here. Following the UK's General Election on 4 July 2024 details are yet to emerge of the new government's approach to AI.	No reference	4th regulatory principle. Guidance states that "*[g]overnance measures should be in place to ensure effective oversight of the supply and use of AI systems, with clear lines of accountability established across the AI life cycle*". .	Not specifically referred to in principes but covered by assurance techniques and technical standards identified as key to supporting regulatory framework. See AI standards hub, roadmap for AI assurance and portfolio of AI assurance technologies .	1st regulatory principle refers to the need to continually identify, assess and manage risks throughout the AI life cycle. 4th regulatory principle also covers need for effective oversight.	2nd regulatory principle. Guidance states that "*AI systems should be appropriately transparent and explainable*", such that the other principles have meaningful effect (e.g. accountability). Parties affected by an AI system should have "*sufficient information about AI systems to be able to enforce their rights*".	4th regulatory principle. Guidance states that governance measures should "*ensure effective oversight of the supply and use of AI systems*"	5th regulatory principle. Guidance states that "*[w]here appropriate, users, impacted third parties and actors in the AI life cycle should be able to contest an AI decision or outcome that is harmful or creates material risk of harm.*" Regulators to implement proportionate measures to ensure outcomes of AI use are contestable where appropriate.

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
Artificial Intelligence (Regulation and Employment Rights) Bill 	Taskforce commissioned by the TUC has published a draft Bill which sets out a detailed legislative framework for regulating the use of AI in the workplace. Most use cases in the employment context will be high-risk.	Direct consultation with affected workers / employees required at least one month before high-risk decision making takes place. This must be repeated every 12 months.	An employer must maintain a register of AI systems used for high-risk decision making.	An employer can not undertake high-risk decision-making until a Workplace AI Risk Assessment has risk assessed an AI system in relation to health and safety, equality, data protection and human rights. While high-risk decision making continues, this must be repeated at least every year.	Auditing the AI system could rebut the burden of proof in a discrimination claim.	Right to personalised explanations for high-risk decisions which are or might reasonably be expected to be detrimental to employees, workers or jobseekers.	Employees, workers or jobseekers will be entitled to a right to human reconsideration of a high-risk decision.	In addition to the right to human reconsideration, breaches of duties such as the duties to consult and risk assess would be actionable in the employment tribunal, potentially giving rise to an injury to feelings award. Automatic unfair dismissal when the decision arises from the unfair reliance on high-risk AI decision making. Changes regarding the burden of proof strengthen discrimination protection.

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
Public Authority Algorithmic and Automated Decision-Making Systems Bill	Private Member's Bill introduced by a Liberal Democrat peer in the House of Lords on 9 September 2024. Applies to algorithmic and automated decision-making systems developed or procured by a public authority ("PA"). This is the first iteration of the Bill, and Private Member's Bills rarely become law. However, it could implement the UK's obligations under the Council of Europe's Framework Convention on artificial intelligence and human rights, democracy and the rule of law (the first international treaty on AI).	No reference	All systems must have logging capabilities to record events during its operation. Logs must be held for at least five years (subject to exceptions) and should record whether final decisions followed the system's recommendations.	PAs are required to complete an Algorithmic Impact Assessment (AIA) before deploying any algorithmic or automated decision-making system (subject to certain exceptions). The form of the AIA framework would be prescribed by the Secretary of State but requirements would include a detailed description of the system and a bias assessment.	PAs must develop processes to monitor outcomes to safeguard against unintended consequences and to validate data accuracy and relevance. PAs must also make arrangements to conduct regular audits and evaluations of these systems. Systems must be capable of scrutiny: PAs are prohibited from using systems where there are practical barriers preventing effective assessment or monitoring of the system (in relation to either individual outputs or overall performance).	Before using or procuring an algorithmic system, PA must complete (and then publish) an Algorithmic Transparency Record. Details will include a description of the system, its rationale, usage in decision-making, and information on human oversight Once in use, PA must notify the public when decisions are made using algorithmic systems and provide meaningful explanations to affected individuals about how decisions were made.	Employees involved in using these systems must be trained in their design, function, and risks. They should have the authority and competence to challenge the system's output. This oversight must be exercised in accordance with the principles set out in the Bill. These include non-discrimination, transparency and explainability, and accountability and governance.	An independent dispute resolution service would be available for challenging a decision made by a relevant system or to obtain redress for such a decision.

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
The EU AI Act	Risk based legislation. Came into force on 1 August 2024 with staggered implementation dates (most provisions in force by 2 August 2026). Most use cases in the employment context will be high-risk. The majority of obligations fall on providers but deployers must also satisfy a range of requirements. **Sanctions** Prohibited AI violations, up to 7% of global annual turnover or €35 million Most other violations, up to 3% of global annual turnover or €15 million Supplying incorrect information to authorities, up to 1% of global annual turnover or €7.5 million	**Deployer: 1)** Must inform workers' representatives that they will be subject to a high-risk AI system. **2)** Must inform individuals that the deployer plans to use a high-risk AI system to make or assist in making decisions relating to an individual.	**Provider: 1)** The system must ensure the automatic recording of events. **2)** Must establish, implement, document and maintain a risk and quality management systems. **3)** Must meet data governance requirements, including bias mitigation. **4)** Must draw up and maintain technical documentation. **5)** Comply with registration, record-keeping, logging and traceability obligations, as well CE conformity obligations. **Deployer: 1)** Retain automatically generated logs for at least 6 months.	**Deployer: 1)** Must use information from providers to carry out a data protection impact assessment (DPIA) (likely to be required for high-risk system), **2)** For certain deployers and certain high-risk systems, must undertake a fundamental rights impact assessment, assessing the system's impact on fundamental rights.	**Deployer: 1)** Monitor the operation of the system on the basis of instructions for use. **2)** Ensure input data in their control is relevant and sufficiently representative.	**Provider: 1)** System must be designed so that deployers can interpret the output, use it appropriately and the instructions must include information specified in the Act. **2):** Users must know that they are Interacting with an AI system and AI-generated content must be identifiable (limited risk systems). **Deployer: 1)** If AI generated decision results in legal or similarly significantly effects (most employment decisions), must provide a clear and meaningful explanation of the role of the AI system in the decision-making process and the main elements of the decision.	**Provider: 1)** System must be designed to allow for effective human oversight. **Deployer: 1)** Take certain appropriate technical and organisational measures and to assign human oversight. **2)** Assign someone to oversee the AI system who is trained, competent, and has the support and authority they need.	

Ius Laboris UK Global HR Lawyers
LEWIS SILKIN

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
The AI Pact 	The AI Pact is a voluntary, non-binding scheme set up by the EU Commission to encourage early implementation of the measures contained in the EU AI Act. Signatories agree to make three core commitments and then select others depending on their area of activity. They are then invited to report against the pledges 12 months after signature. On 25 September 2024, the three core commitments were announced: • Adopting an AI governance strategy to foster the uptake of AI in the organisation and work towards future compliance with the AI Act • Identifying and mapping AI systems likely to be categorised as high-risk under the AI Act • Promoting AI awareness and literacy among staff, ensuring ethical and responsible AI development	**Deployer:** may choose to inform workers' representatives and affected workers when deploying workplace AI systems.	**Provider: 1)** may choose to put processes in place to identify possible known and reasonably foreseeable risks to health, safety and fundamental rights that could follow from the use of relevant AI systems throughout their lifecycle. **2)** may choose to develop policies to ensure high-quality training, validation and testing datasets for relevant AI systems. **3)** may choose to implement logging features to allow for traceability appropriate for the intended purpose of the system. **4)** may choose to inform deployers of appropriate use, capabilities, limitations and potential risks of AI systems. .	**Provider** may implement policies and processes aimed at mitigating risks associated with the use of relevant AI systems, in line with the relevant obligations and requirements envisaged in the EU AI Act, to the extent feasible. **Deployer**: may choose to carry out a mapping of known and reasonably foreseeable possible risks to fundamental rights of persons and groups of individuals that may be affected through the use of relevant AI systems.	**Provider:1)** may choose when developing all or certain AI systems to implement logging features to allow traceability appropriate for the intended purpose of the system.	**Provider 1)**: may choose to design AI intended to directly interact with individuals so they are informed they are interacting with an AI system. **2)** may choose to design gen AI systems so AI generated content is marked and detectable as such & provide means for deployers to clearly & distinguishably label AI generated content, unless the text has been subject to human review & a natural or legal person holds editorial responsibility for its publication. **Deployer: 1)** may choose to inform individuals when a decision made about them is prepared, recommended or taken by an AI system. **2)** may choose to ensure that individuals are informed when they are directly interacting with an AI system. **3)** may choose to clearly & distinguishably label AI generated content, unless the text has been subject to human review and a natural or legal person holds editorial responsibility for its publication.	**Providers and Deployers**: may choose to implement concrete measures to ensure human oversight over the operation of high-risk AI systems as defined by the EU AI Act.	

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
UK/EU GDPR 	The GDPR is a Regulation that requires businesses to protect the personal data and privacy of EU citizens. **Sanctions** GDPR - up to €20 million or 4% annual global turnover UK GDPR – up to £17.5 million or 4% of annual global turnover		The accountability principle means as well as being responsible for complying with data protection law, you must also demonstrate that compliance in any AI system that processes personal data.	A Data Protection Impact Assessment (DPIA) is required for processing that includes innovative technologies or the novel application of existing technologies. Therefore a DPIA should be carried out before an AI system is deployed	Undertaking a DPIA, as well as the need to define procedures for ongoing compliance supervision and AI system audits. Also, regular checks to ensure AI systems function as anticipated/required.	Individuals must be informed if their personal data is going to be used in an AI system. This information should be provided at the point of data collection. Individuals also need meaningful information about logic and consequences of any solely automated decisions.	The automated decision making provision applies where there is solely automated decision-making that has legal or similarly significant effects. Often referred to as "human in the loop", for human review to be meaningful, it should come after the automated decision has taken place and it must relate to the actual outcome.	Individuals must be able to contest a decision in a timely manner in order for the processing to be fair and compliant. Appropriate measures must be put in place to ensure individuals can exercise their rights.

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
NY City Law: Local Law 144 of 2021 	This prohibits employers and employment agencies from using an automated employment decision tool (AEDT) in New York City unless they ensure a bias audit was done and required notices provided. **Sanctions** Civil penalty of not more than $500 for a first violation (including each additional violation occurring on the same day as the first violation). Civil penalty of not less than $500 nor more than $1,500 for each subsequent violation. Each day the AEDT tool is used in violation gives rise to a separate violation and civil penalty.	No reference	No reference	No reference	Mandatory annual independent audit of the AEDT. Summary of the most recent results must be published. Independent auditor must (as a minimum) evaluate calculations of selection or scoring rates and the impact ratio across sex categories, race/ethnicity categories, and intersectional categories.	Employers and employment agencies must notify employees and job candidates who are residents of New York City that they are using an AEDT and the job qualifications or characteristics the AEDT will assess.	No reference	Compliance is enforced by the NYC Department of Consumer and Worker Protection. Claims of discrimination involving an AEDT can be made to the NYC Commission on Human Rights.

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
Illinois AI Video Interview Act (820 ILCS 42/1) 	This Act applies to all employers that use AI tools to analyse video interviews of candidates for positions based in Illinois.	No consultation obligations but the applicant's consent must be sought by employers to be evaluated by AI before the video interview. Employers may not use AI to evaluate a video interview without consent.	Video interviews must only be shared with those whose expertise or technology is required to evaluate the interview. In other words, video interviews may be shared with vendors providing the AI used to analyse the interview.	No reference	Where AI analysis of a video interview is used to select candidates for in person interview, demographic data must be collected and reported as follows: (a) race and ethnicity of those not invited to in person interview; and (b) race and ethnicity of candidates who are hired. Data must be reported to the Department of Commerce and Economic Opportunity annually. Data must be analysed to determine if the data discloses racial bias in the use of AI and results reported to the Governor and General Assembly.	Employers must inform candidates before analysis takes place that AI is being used to assess their suitability for the role, how the AI works, and which characteristics will be used in the evaluation. The candidate's consent is required before any analysis takes place.	Annual report prepared by Department of Commerce and Economic Opportunity analysing data collected in order to determine if the data discloses racial bias in the use of AI. Findings reported to the Governor and General Assembly.	No reference

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
Colorado: Senate Bill 205 	This Act is the first comprehensive piece of AI legislation in the US but does not come into force until 1 February 2026. The Act may still be subject to scrutiny and amendment to 'fine tune' the provisions. Like the EU AI Act, the majority of obligations apply to high-risk AI systems (which may include systems used to make decisions about employment or employment opportunities). **Sanctions** The Act gives the Attorney General exclusive authority to enforce it. The majority of violations of the Act amount to deceptive trade practices under the Colorado Consumer Protection Act which may carry penalties of up to $20,000 per violation.		**Deployers and Developers**: Requirement to use reasonable care to safeguard consumers from the risks of algorithmic discrimination (whether known or reasonably foreseeable).	**Deployers**: When deploying high-risk systems deployers must undertake impact assessments in respect of those systems at least once a year or within 90 days of any intended and substantial modification to any such system.	**Deployers and Developers**: Varying reporting requirements to report certain confirmed or likely instances of algorithmic discrimination to the Attorney General within 90 days. Developer's reporting obligations also include providing such information to deployers that are known or other developers of the particular system. **Deployers:** High-risk AI will need to have a risk management policy and program, and an annual assessment to check they do not cause algorithmic discrimination.	**Developers and Deployers**: Requirement to provide public notices online which includes certain information about their AI. Deployers also have transparency obligations which may require an additional pre or post use notice be provided where certain decisions will be/ have been made using high risk AI systems. **Developers**: Also required to provide both those deploying high-risk AI systems as well as other developers who may be substantially and intentionally modifying their systems with certain transparency information.		**Deployers**: Where a high-risk system leads to an adverse consequential decision, deployers must provide notice which – among other things – provides the consumer with the chance to correct personal data and appeal the AI system's decision.

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
Singapore	Patchwork of frameworks and/or guidelines created by statutory/regulatory bodies instead of legislating. For example, the Model AI Governance Frameworks for both generative and 'traditional' AI, and the Personal Data Protection Commission has issued Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems.		Accountability is included in the compilation of ethical principles within the Model Framework for 'Traditional' AI. Further, accountability is one of the nine dimensions under the Model AI Governance Framework for Generative AI.		Auditability is included in the compilation of ethical principles within the Model Framework for 'Traditional' AI. Testing and assurance is also one of the nine dimensions under the Model AI Governance Framework for Generative AI.	One of the two guiding principles under the Model Framework for 'Traditional' AI is for AI decisions to be explainable, transparent, and fair. Explainability and transparency are both also included in the compilation of ethical principles collated within the framework. Additionally, transparency around content provenance is one of the dimensions under the Model AI Governance Framework for Generative AI.	A guiding principle under the Model Framework for 'Traditional' AI is for systems to be human centric.	The responsibility, accountability, and transparency principle, included in the compilation of ethical principles within the Model Framework for 'Traditional' AI, proposes making available avenues of redress for adverse individual or societal effects of an algorithmic decision system.

	Overview	Consultation about use	Accountability and governance	Impact assessments	Auditing and monitoring	Transparency and explainability	Human oversight / intervention	Contestability
Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and Rule of Law	Designed to be a synthesis of '*common general principles and rules*' to address challenges arising throughout the AI lifecycle – from design through to decommission. So far signatories to the Convention include Andorra, Georgia, Iceland, Norway, Moldova, San Marino, the UK, the US, Israel, and the EU. Signatories will apply the Convention to the activities within the lifecycle of an AI system undertaken by public authorities (or private actors on their behalf). Signatories and regulators must also address the risks and impacts arising from activities of private actors (i.e. unrelated to public authorities) but they can elect to apply the principles of the Convention or take '*other appropriate measures*' to achieve the same goal. Signatories must establish an independent oversight mechanism to oversee compliance, raise awareness, stimulate informed public debate, and carry out consultations on how AI should be used.		Signatories must adopt or maintain measures to ensure accountability and responsibility for adverse impacts on human rights, democracy, and the rule of law resulting from activities within the lifecycle of AI systems.	Signatories must also adopt or maintain measures for the identification, assessment, prevention, and mitigation of risks posted by AI systems. The measures may be differentiated as appropriate but shall include documenting risks, actual and potential impacts, and the risk management approach.	Further measures for the identification, assessment, prevention, and mitigation of risks which signatories must adopt or maintain shall include monitoring for risks and adverse impacts to human rights, democracy, and the rule of law. Equally, where appropriate, such measures shall require testing of AI systems before they are made available and again when significantly modified.	Each signatory must adopt or maintain measures which ensure there is adequate transparency requirements in place, tailored to the specific contexts / risks (including the identification of AI generated content). Additionally, as part of their obligation to ensure there are effective remedies for violations of human rights, signatories must adopt or maintain measures to ensure that information regarding the AI systems and their use which could significantly affect human rights is documented and, where appropriate, made available / communicated to affected people. Moreover, signatories must seek to ensure that important questions raised about AI systems are considered through public discussion and multistakeholder consultation.	Signatories are also required to either adopt or maintain measures to ensure there is adequate oversight requirements in place, again tailored to the specific contexts and risks.	The transparency information provided in connection with AI systems and their use which have the potential to significantly affect human rights, must be sufficient for people to be able to contest: (a) decisions made or substantially informed by the use of the system; and (b) the use of the system itself.

Ius Laboris UK Global HR Lawyers
LEWIS SILKIN