



MONFORTE · AI RESEARCH DIVISION  
White Paper

---

# Ethical Foundations of AI Impact Assessment

*Before the Template*

**Resident Research Group**

Lead author: Michael Hall

October 2026

Publication edition

*Governing Intelligence. Securing Trust.*

## Notice and Disclaimer

**Purpose.** This white paper is published for general information and discussion. It sets out the views of the Monforte AI Research Division on ethical considerations in AI impact assessment as at the date of publication.

**Not professional advice.** Nothing in this paper constitutes legal, regulatory, compliance, audit, certification or other professional advice, and it should not be relied upon as such. Organizations should obtain advice appropriate to their own circumstances and jurisdictions before acting. Publication of this paper does not create a client, advisory or other professional relationship.

**Laws and standards.** References to legislation, regulations, case law and standards are summaries prepared for discussion and are not authoritative statements of their content. Legal requirements change, and their application depends on facts and jurisdiction. Readers should consult the official texts, including the current editions of ISO and ISO/IEC standards, which are available from the issuing bodies.

**Accuracy.** The information in this paper reflects the position as of October 2026. While reasonable care has been taken, no representation or warranty, express or implied, is given as to its accuracy, completeness or continued currency. To the extent permitted by law, Monforte Security LLC accepts no liability for any loss arising from reliance on this paper.

**Scenarios and examples.** The scenarios in Section 10 are hypothetical composites. They do not describe any named organization, product or individual. References to published cases, reports and organizations are drawn from public sources and imply no finding beyond what those sources state.

**No endorsement.** Reference to any framework, organization, product or publication does not imply endorsement by, or affiliation with, Monforte. Trademarks and standard designations remain the property of their respective owners.

**Copyright.** © 2026 Monforte Security LLC. All rights reserved. This paper may be quoted with attribution. Reproduction of the whole or substantial parts requires prior written permission.

# Contents

Executive Summary

1. Purpose and Scope

2. What Earlier Impact Regimes Teach

3. Why Ethics Comes Before the Template

4. Five Principles

5. Impact Domains

6. Acquired Systems and the Deployer's Responsibility

7. Independence of the Assessment

8. Tensions, Uncertainty and the Limits of Acceptance

9. From Principles to Practice

10. Illustrative Scenarios

11. Evidence That Ethics Changed the Decision

Annex A. Leadership Questions

Annex B. Crosswalk to Principal Frameworks

Annex C. References

## Executive Summary

The ethical questions raised by artificial intelligence are not new. For more than half a century, organizations have been required to assess the environmental, safety and privacy consequences of what they build, operate and deploy. The record of those regimes is instructive. Impact assessment has most often fallen short not for want of method, but because it exerted weak influence on decisions, was treated as a procedural requirement rather than an input to judgment, and was rarely followed up once approval was granted.

AI impact assessment is exposed to each of these failure modes, and to one more: an AI system may change after it is assessed, through model updates, vendor changes, drift, or extension to new purposes and populations.

This paper sets out the ethical foundations an organization requires before it completes an impact assessment template. It identifies five principles and the impact domains to which they apply, including workers, communities and the environment. It addresses two conditions that commonly weaken assessment in organizational practice: most AI is acquired rather than built, and assessments are frequently prepared by those who sponsor the system. It also sets out how an organization should treat uncertainty and where the authority to accept residual impacts ends.

**Central proposition.** An AI impact assessment has ethical value when it subjects a proposal to credible challenge and demonstrably informs an accountable decision. Changes to the proposal are strong evidence of influence; approval without change requires supported reasons. Follow-up that shows whether the conditions imposed actually worked is evidence of effectiveness.

The five principles are:

- **Necessity and proportionality** — whether a less intrusive means would serve the purpose, and whether the benefit, supported by evidence, justifies the intrusion that remains; including who receives the benefit and who bears the burden.
- **Human dignity, agency and contextual integrity** — what the system does to ordinary life when it moves information out of the setting in which it arose.
- **Power, procedural fairness and non-discrimination** — whether affected people can raise concerns, challenge consequential results and obtain accommodation, and whether the effects fall unevenly on particular groups.
- **Non-delegable accountability and human oversight** — who verifies an output before it causes harm, and who has the authority to change or stop the deployment.
- **Transparency and traceability** — what the system observes and infers, who uses the information, and what record exists of each use and its consequences.

# 1. Purpose and Scope

## 1.1 Purpose

This paper provides executives and the risk, human resources, procurement, legal and internal audit functions with a reference for the ethical judgments that an AI impact assessment must make. It is intended for use before an assessment template is completed, and again when the results of an assessment are presented for decision.

## 1.2 Scope

The paper applies to AI systems that organizations develop, acquire or use, including predictive, perception, generative and agentic systems. Its illustrations are drawn from workplace monitoring, public-space surveillance and employment screening because these settings expose the principles clearly. The principles apply more widely.

## 1.3 Limitations

The paper is not legal advice and does not describe a compliance method for any jurisdiction. Regulatory references reflect the position as of October 2026. Illustrative scenarios in Section 10 are composites and do not describe any named organization.

# 2. What Earlier Impact Regimes Teach

Formal assessment of consequences predates AI by decades. Environmental protection, occupational safety, privacy and human rights each developed a discipline for asking, before a decision is made, what an activity will do to people and to the world around them. Each moved, over time, from voluntary practice toward legal obligation. Each also accumulated evidence about how assessment fails.

| Regime                                    | Representative instruments                                                                                                                  | Relevance to AI impact assessment                                                                          |
|-------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| Environmental impact                      | National Environmental Policy Act (1969); EU EIA Directive 2011/92/EU as amended by 2014/52/EU; ISO 14001 environmental aspects and impacts | Pre-decision assessment of significant effects; the management-system precedent for "AI system impact"     |
| Machinery and robotics                    | Occupational Safety and Health Act (1970); Machinery Directive 2006/42/EC and Regulation (EU) 2023/1230; ISO 10218-1 and -2:2025            | Mandatory safety requirements for automation; machine-learning safety functions now regulated as high-risk |
| Privacy and data protection               | GDPR Art. 35 data protection impact assessment; Art. 22 automated decisions                                                                 | Assessment of high-risk processing, including systematic monitoring of public areas                        |
| Corporate responsibility and human rights | ISO 26000:2010; UN Guiding Principles on Business and Human Rights (2011); OECD Guidelines (2023); EU CSDDD as amended (2026)               | Due diligence to identify, prevent, mitigate and account for adverse impacts                               |

## 2.1 Environmental Impact Assessment

The National Environmental Policy Act of 1969 required United States federal agencies to prepare detailed statements on the environmental effects of major federal actions before taking them. In Europe, the EIA Directive, first adopted in 1985 and now codified as Directive 2011/92/EU, requires assessment of projects likely to have significant environmental effects before development consent is given. Directive 2014/52/EU added a requirement for measures to avoid or offset significant adverse effects and for procedures to monitor them.

At the level of the organization, ISO 14001 requires identification of the environmental aspects of activities, products and services and of their associated impacts. This is the closest management-system precedent for the AI system impact assessment required by ISO/IEC 42001. Organizations that operate an environmental management system already possess the discipline of asking what their operations do beyond their own boundary.

## 2.2 Mechanization, Robotics and Occupational Safety

Mechanization brought regulation of physical hazards. The Occupational Safety and Health Act of 1970 imposed a general duty on United States employers to provide workplaces free from recognized serious hazards. In the European Union, the Machinery Directive 2006/42/EC set essential health and safety requirements for machinery. Its successor, Regulation (EU) 2023/1230, applies from 20 January 2027 and classifies safety components with fully or partially self-evolving behaviour using machine-learning approaches as high-risk products requiring third-party conformity assessment. The international robot safety standards ISO 10218-1 and ISO 10218-2 were revised in 2025.

Mechanization also had economic consequences that impact assessment rarely addressed. Acemoglu and Restrepo (2020) found that the adoption of industrial robots in United States local labour markets was associated with reductions in employment and wages in the areas most exposed. The point for AI is not that automation should be resisted, but that its effects on work are foreseeable and assessable.

## 2.3 Privacy and Data Protection

Privacy impact assessment emerged as a practice in the 1990s and became a legal obligation in the European Union through Article 35 of the General Data Protection Regulation. A data protection impact assessment is required where processing is likely to result in a high risk to individuals, and expressly where processing involves systematic monitoring of a publicly accessible area on a large scale. Article 22 limits decisions based solely on automated processing that produce legal or similarly significant effects. Where such a decision is permitted because it is necessary for a contract or based on explicit consent, the controller must provide safeguards including at least the right to obtain human intervention, to express a point of view and to contest the decision.

## 2.4 Corporate Responsibility and Human Rights Due Diligence

ISO 26000:2010 provides guidance on social responsibility. The UN Guiding Principles on Business and Human Rights (2011) established a corporate responsibility to respect human rights through due diligence: identifying, preventing, mitigating and accounting for adverse impacts. The OECD Guidelines for

Multinational Enterprises, updated in 2023, carry the same expectation. The EU Corporate Sustainability Due Diligence Directive (EU) 2024/1760 converted it into law for the largest companies. Following amendments adopted in 2026, its scope is limited to companies with more than 5,000 employees and €1.5 billion in turnover, with compliance from July 2029. The narrowing illustrates a recurring pattern: obligations to assess impact are contested, and the legal floor moves.

## 2.5 Recurring Failure Modes

The literature on environmental impact assessment, the oldest of these regimes, documents four failure modes that bear directly on AI.

### Weak influence on decisions

Reviews of EIA effectiveness have found that its influence on planning decisions is often weak and that its findings are frequently marginalized in favour of other considerations (Jay et al., 2007). Cashmore et al. (2004) report survey evidence that nearly half of planning officers considered that EIA had made no difference to decisions, and note that affected communities may perceive assessment as a device to justify decisions already taken. An assessment completed after the commitment to deploy is subject to the same weakness.

### Procedure in place of substance

Cashmore et al. (2004) distinguish procedural effectiveness, whether the assessment steps were followed, from substantive effectiveness, whether the assessment changed outcomes. A completed template demonstrates the former. Only a change in the decision, or a reasoned decision not to change it, demonstrates the latter.

### Distance between assessment and decision

Where the proponent of a project commissions and controls its assessment, the assessment is exposed to the proponent's interest. The same structure arises when the business owner who sponsors an AI system also prepares its impact assessment. Section 7 addresses independence.

### Absence of follow-up

Arts, Caldwell and Morrison-Saunders (2001) and Morrison-Saunders and Arts (2004) identified the neglect of follow-up, the monitoring and evaluation of whether predicted impacts actually occurred, as a central weakness of EIA practice. Without follow-up, an assessment is a forecast that is never checked. The 2014 EU amendment to require monitoring of significant adverse effects responded to this weakness.

## 2.6 What AI Changes

The obligation to answer for the effects of what an organization deploys is unchanged. Four features of AI alter how it is discharged:

- **Scale and speed of inference.** A system can draw conclusions about many people continuously, and those conclusions can trigger action before anyone reviews them.
- **Observation at scale.** Privacy and human rights regimes already assess the observation and classification of people. AI extends the scale, granularity and inferential reach of that observation, so that conclusions about behaviour, identity or intent can be drawn continuously and at low cost.

- **Change after assessment.** Model updates, vendor changes, drift and extension to new purposes can alter a system's behaviour after it has been assessed.
- **Action as well as recommendation.** Agentic systems execute tasks with delegated authority. The point at which a human can intervene moves earlier, and the scope of authority delegated becomes part of the impact.

## 3. Why Ethics Comes Before the Template

Tegmark (2016) describes the central challenge of powerful technology as a race between its growing power and the wisdom with which it is managed, and argues that for the most consequential technologies that wisdom must be developed in advance rather than learned from mistakes. Within an organization, impact assessment is one of the mechanisms by which that wisdom is exercised before a capability is deployed. Russell (2019), discussed in Section 4.1, describes the technical side of the same concern.

### 3.1 Legal Baselines Vary

An organization that relies on the legal minimum will assess different things in different places for the same system. The EU Artificial Intelligence Act, Regulation (EU) 2024/1689, prohibits certain practices outright, including systems that infer emotions in the workplace or in educational institutions other than for medical or safety reasons (Art. 5(1)(f)). Following the Digital Omnibus, Regulation (EU) 2026/1744, obligations for high-risk systems listed in Annex III apply from 2 December 2027. The fundamental rights impact assessment in Article 27 applies only to public bodies, private entities providing public services, and deployers of certain credit and insurance systems. Most private-sector deployers therefore face no statutory obligation to assess fundamental-rights impacts, even for high-risk systems.

The United States has no general federal AI statute. Obligations arise from sectoral law and from state and local measures such as New York City Local Law 144 of 2021, which requires bias audits of automated employment decision tools that fall within its statutory scope. Elsewhere, the position is different again. Context, safeguards and available remedies can alter the effects of a system, but the organization's ethical responsibility for those effects does not lapse where legal requirements are weaker.

### 3.2 What the Standards Provide and What They Leave Open

ISO/IEC 42001:2023 requires an organization to define a process for assessing the potential consequences of AI systems for individuals, groups and societies (clause 6.1.4) and to perform such assessments at planned intervals or when significant changes are proposed (clause 8.4). ISO/IEC 42005:2025 is guidance: it offers a structure and suggestions for how an assessment might be conducted. ISO/IEC TR 24368:2022 surveys ethical and societal concerns. The NIST AI Risk Management Framework provides a MAP function for establishing context and identifying impacts.

None of these instruments determines whether a particular intrusion is justified, whether a burden is fairly distributed, or whether a residual impact is acceptable. Those are judgments. They require principles, and they require someone with the authority to make them.

### 3.3 Beyond Principles

AI ethics does not lack principles. Jobin, Ienca and Vayena (2019) identified 84 guideline documents converging on transparency, justice and fairness, non-maleficence, responsibility and privacy. Mittelstadt (2019) argued that principles alone cannot guarantee ethical AI, because AI development lacks the common aims, professional history, proven methods and accountability mechanisms that give principles force in fields such as medicine.

This paper therefore treats principles as questions an assessment must answer, record and act upon, and judges them by evidence of effect. Floridi and Cowls (2019) offer a comparable five-principle framework drawn from bioethics (beneficence, non-maleficence, autonomy, justice and explicability). The set used here is chosen for its operational character: each principle corresponds to a decision an organization can take.

## 4. Five Principles

### 4.1 Necessity and Proportionality

A legitimate purpose does not justify every means. Necessity and proportionality are two distinct tests. **Necessity** asks whether a less intrusive means, including a non-algorithmic one, would achieve the purpose to a comparable degree. **Proportionality** asks whether the benefit of the chosen means justifies the intrusion that remains.

Both tests depend on evidence that the benefit exists. Before deployment, some benefits will necessarily be estimates. The relevant distinction is between a supported estimate, with its uncertainty stated, and an unsupported assertion. Section 8.2 addresses how uncertainty should affect the decision. Russell (2019) identifies a related risk in the design of AI itself: systems built to optimize a fixed objective pursue it even where the objective is an incomplete proxy for what people actually want. In an organization, a metric such as monitored activity can stand in for the purpose it was meant to serve, such as protecting client data, and the assessment must test whether the proxy and the purpose coincide.

Proportionality also requires attention to distribution. An employer may obtain assurance while employees and their households bear the intrusion. A community may gain investigative capability while residents bear the risk of inappropriate searches. Neither distribution settles the decision, but both must be visible to those who make it.

European human rights law illustrates the balancing in employment. In *Bărbulescu v. Romania* (2017), the Grand Chamber of the European Court of Human Rights found a violation of the right to respect for private life where an employee's communications were monitored without adequate notice and without proper assessment of the justification and of less intrusive alternatives. In *López Ribalda and Others v. Spain* (2019), the Grand Chamber found no violation where covert video surveillance was limited in scope and duration and responded to reasonable suspicion of serious misconduct. The outcome turned on the balancing, not on the technology. In 2022, a Dutch court held that an employer's instruction to keep a webcam on throughout the remote working day was a disproportionate interference with private life and did not justify dismissal for refusal (Rechtbank Zeeland-West-Brabant, ECLI:NL:RBZWB:2022:5656).

**Questions the assessment must answer**

- ☐ What specific objective does the system serve, and what evidence shows that it achieves that objective?
- ☐ What alternatives, including non-algorithmic ones, were considered, and on what evidence were they rejected?
- ☐ What intrusion remains after the least intrusive effective option is chosen, and why is it justified?
- ☐ Who receives the benefit, and who bears the burden?

**4.2 Human Dignity, Agency and Contextual Integrity**

Nissenbaum (2004, 2009) argues that privacy is best understood not as secrecy but as contextual integrity: an information flow is appropriate when it conforms to the norms of the context in which the information arose. A system that moves information across contexts, from a private room into an employer's review queue, or from scattered sightings of a vehicle into a searchable history of movement, changes those norms even where each individual observation might be regarded as unremarkable. Aggregation is itself an impact.

Observation also changes behaviour. Penney (2016) found a significant decline in traffic to Wikipedia articles on privacy-sensitive topics following the 2013 revelations of government surveillance. In a workplace, a worker who performs well may still alter ordinary conduct under observation, with consequences for wellbeing and trust that no accuracy test or breach report records. Zuboff (2019) describes a commercial logic in which human experience is captured as behavioural data and converted into predictions; an organization deploying monitoring should be able to state whether the data it captures will serve that logic, through the vendor or otherwise.

The UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) places human dignity at the foundation of its values. In operational terms, dignity requires that the assessment consider people as participants in a setting, not as sources of telemetry.

**Questions the assessment must answer**

- ☐ In what context does the information arise, and what norms govern that context?
- ☐ Where will the information flow, and how does aggregation change what it reveals?
- ☐ How might the presence of the system change behaviour, wellbeing and trust, even where no adverse decision follows?
- ☐ Who else is captured, such as household members, passers-by or third parties?

**4.3 Power, Procedural Fairness and Non-Discrimination****Power**

Impact assessments often assume an equality of position that does not exist. European data protection authorities have concluded that employee consent is unlikely to be freely given because of the imbalance of power in employment (Article 29 Working Party, 2017; European Data Protection Board, 2020). Consent is unreliable in these conditions rather than impossible, but notice or acknowledgment does not establish that an arrangement is fair. A resident cannot readily avoid public roads. Where avoidance is impractical, the

legitimacy of the system must rest on its justification and safeguards, not on the acquiescence of those affected.

### Procedural fairness

Affected people need a credible way to raise concerns without retaliation, to challenge consequential results, and to request an exception or accommodation where appropriate, for example under disability law. These rights differ from an unconditional right to opt out of every security system, which is not a general principle. The line lies at the ability to contest a consequential outcome and to obtain an appropriate adjustment. Labour organizations have set out comparable expectations: the Trades Union Congress (2021, 2024) calls for consultation before high-risk AI is deployed at work, personalized explanation and human review of high-risk decisions, and the AFL-CIO (2025) has called for worker involvement in deployment and human review of automated employment decisions.

### Non-discrimination

Disparate impact is a well-documented harm of AI systems. The National Institute of Standards and Technology found that many face recognition algorithms produced false positive rates for some demographic groups that were ten to a hundred times higher than for others, depending on the algorithm (Grother et al., 2019). Disparate effects can arise from training data, design, the threshold at which an alert is raised, or the decision about where a system is deployed. Fairness criteria can also conflict. Where the prevalence of the predicted outcome differs between groups and prediction is imperfect, a risk score that is equally calibrated across groups cannot also equalize false positive and false negative rates between them (Chouldechova, 2017). An organization must therefore choose which fairness property it will prioritize and record why.

### Questions the assessment must answer

- ☐ Can affected people decline, and if not, what justifies the system without their consent?
- ☐ How can an affected person raise a concern, challenge a result and request an accommodation, and what protects them from retaliation?
- ☐ Has performance been tested across the groups the system will affect, and which fairness property has the organization chosen to prioritize?
- ☐ Does the pattern of deployment itself concentrate the burden on particular groups or places?

## 4.4 Non-Delegable Accountability and Human Oversight

A system produces outputs; it does not bear responsibility for them. Responsibility rests with the organization that decided how an output would be used. In *Green v. City and County of San Francisco* (2014), an automated licence plate reader misread a plate and flagged a car as stolen. Officers conducted a high-risk stop of the driver at gunpoint without visually confirming the plate. The United States Court of Appeals for the Ninth Circuit allowed her claims to proceed. The failure lay not in the camera alone but in the absence of verification between alert and action.

Human review is not a safeguard merely because it exists. Parasuraman and Manzey (2010) found that complacency and bias in the use of automated aids occur among experienced operators as well as novices and are not reliably overcome by practice alone. The EU AI Act requires that high-risk systems enable those overseeing them to remain aware of the tendency to over-rely on outputs (Art. 14(4)(b)), and requires

deployers to assign oversight to competent persons with the necessary authority (Art. 26). Oversight must therefore be designed: what is verified, against what evidence, by whom, with what authority to reject the output, and with what record of overrides.

Accountability also requires a named role with the authority to change or stop a deployment when unanticipated harms emerge.

### Agentic systems and bounded action

Agentic systems act rather than recommend: they call interfaces, change records, issue payments or lock accounts. Review after the fact cannot prevent an action already taken, and review of every action at machine speed tends to become nominal. The assessment of an agentic system must therefore examine the **action space** it is permitted: which actions it may take without approval, the maximum effect of a single action and of many actions in combination, whether each action can be reversed and how quickly, the rate at which actions can occur, and the conditions under which the system is automatically suspended. The point of human intervention must be placed before actions that are irreversible or exceed a defined threshold.

### Questions the assessment must answer

- ☐ What may happen because of an output, and what must a person verify before that step?
- ☐ Does the reviewer have the evidence, time, competence and authority to reject the output?
- ☐ Who has the authority to suspend or stop the deployment, and on what triggers?
- ☐ For agentic systems, what actions may the system take without human approval, what is the largest effect those actions can have, and can they be reversed?

## 4.5 Transparency and Traceability

Accountability depends on visibility. Internally, management must be able to reconstruct what the system collected, what it inferred, who accessed the information, what conclusion was drawn and what action followed. Externally, affected people need an understandable account of the system's purpose and of the decisions it may influence. Security-sensitive details may justify limits on disclosure; they do not remove the need to explain the scope and consequences of a deployment.

Traceability extends beyond the organization. Data may flow to the vendor, to other customers of the vendor, or into shared networks. Civil liberties organizations have documented how licence plate reader data, retained and shared across jurisdictions, can reveal patterns of movement and visits to sensitive places (American Civil Liberties Union, 2013; Electronic Frontier Foundation, 2023). The EU AI Act requires automatic logging for high-risk systems (Art. 12) and transparency toward people interacting with AI systems or exposed to synthetic content (Art. 50).

### Questions the assessment must answer

- ☐ What does the system observe, and what does it infer from what it observes?
- ☐ Who can access and search the information, where is it shared, and for how long is it retained?
- ☐ What record exists of each use, and could the organization reconstruct the path from output to action?
- ☐ What explanation do affected people receive, and is it intelligible to them?

## 5. Impact Domains

The principles apply across five domains. An assessment that considers only the individual user, or only the organization's own liability, will miss impacts that fall elsewhere.

### 5.1 Individuals

Effects on privacy, autonomy, health, opportunity and treatment by institutions, including those on people who never interact with the system directly but are observed or classified by it.

### 5.2 Workers

Beyond monitoring, AI changes the content of work. Tasks may be automated, skills may erode when judgment is delegated to a system, and the pace of work may be set by a metric. Acemoglu and Restrepo (2020) show that the employment effects of earlier automation were concentrated in particular labour markets. An assessment should consider changes in job content, the risk of deskilling, who bears the cost of transition, and how workers and their representatives were consulted. The EU AI Act requires employers to inform workers' representatives and affected workers before a high-risk AI system is used at the workplace (Art. 26(7)).

### 5.3 Communities and Groups

Effects that arise only in aggregate, such as the cumulative record produced by a network of sensors, or that fall disproportionately on particular groups, including marginalized communities and those with little ability to avoid the system.

### 5.4 Society

Effects on institutions, public trust, freedom of movement and association, and the information environment. Generative systems raise particular questions of disclosure, deception and the integrity of information.

### 5.5 Environment

AI has a material footprint. The International Energy Agency (2025) estimated data centre electricity use at about 415 TWh in 2024, around 1.5 per cent of global electricity consumption, and projected that it would more than double to around 945 TWh by 2030, with AI a principal driver. Water use and hardware lifecycles add to the footprint. The OECD AI Principles include protection of natural environments within their first principle, ISO/IEC 42001 lists environmental impact among potential AI-related organizational objectives (Annex C), and the EU AI Act requires providers of general-purpose AI models to document known or estimated energy consumption (Annex XI).

For most deployers, the footprint is incurred through a vendor or cloud provider and is therefore invisible unless requested. A deployer rarely controls the energy used to train a model, but it does control how the model is used: continuous analysis of a video stream consumes far more than scheduled processing in batches, and a large general-purpose model may be used where a smaller one would serve. These choices belong in the necessity test. The environmental cost of a system belongs in the proportionality judgment

alongside its other costs. The environmental benefits of use, such as more efficient logistics, belong there too, and both should be supported by evidence rather than assumption.

## 6. Acquired Systems and the Deployer's Responsibility

Most organizations deploy AI systems built by others. The provider controls the design; the deployer controls the purpose, the context, the configuration, the population affected and the consequences attached to an output. The EU AI Act reflects this division by imposing distinct obligations on deployers (Art. 26). Ethically, the deployer cannot transfer its responsibility for the effects of its own deployment to the vendor.

Procurement is therefore where many of the decisions that an impact assessment should inform are actually made. An organization acquiring an AI system should obtain:

- a statement of intended purpose and known limitations;
- performance evidence across the groups the system will affect;
- a description of data flows, onward sharing, retention and the vendor's own use of customer data;
- logging sufficient to reconstruct each use;
- notification of material changes to the model or service;
- contractual rights to audit, to restrict configuration and to terminate; and
- information on energy consumption where available.

A vendor's description of available controls is not evidence of how a particular deployment operates. The assessment must examine the configuration and rules the organization has actually adopted.

## 7. Independence of the Assessment

The environmental record shows that an assessment controlled by the proponent of a project is exposed to the proponent's interest. Three roles should be distinguished: the **sponsor**, who proposes the system and expects its benefits; the **assessor**, who evaluates its impacts; and the **approver**, who decides whether and on what conditions it proceeds. In a small organization one person may hold more than one role, but the decision record should show who held each and how conflicts of interest were managed.

The assessor requires competence in the system, its context and the principles, and access to affected people. Participatory approaches, such as the algorithmic impact assessment process piloted by the Ada Lovelace Institute (2022) for access to a national medical imaging platform, show how affected parties can contribute to an assessment rather than merely receive its conclusions.

Independence depends on structure rather than intention. At a minimum:

- approval should not rest with the sponsoring business owner alone;
- the assessor should not report to the sponsor for the purposes of the assessment;
- the assessor and affected parties should have a right of escalation to a body independent of the sponsor, such as a risk committee or audit committee; and

- dissenting views should be entered in a register kept with the decision record, together with the response to each, rather than resolved silently.

## 8. Tensions, Uncertainty and the Limits of Acceptance

### 8.1 Tensions Between Principles

The principles will not always point in the same direction. Common tensions include:

| Tension                        | Illustration                                                                                            |
|--------------------------------|---------------------------------------------------------------------------------------------------------|
| Security and privacy           | Stronger monitoring may reduce data loss while increasing intrusion into private life.                  |
| Transparency and security      | Full disclosure of detection logic may enable evasion; limited disclosure may impede challenge.         |
| Oversight and efficiency       | Mandatory human verification slows decisions and adds cost; its absence shifts risk to affected people. |
| Competing fairness criteria    | Equal error rates across groups and calibrated scores cannot both be achieved where base rates differ.  |
| Benefit and environmental cost | A larger model may perform better while consuming substantially more energy and water.                  |

Tensions cannot be eliminated, but they can be governed. The decision record should state the tension, the option chosen, the reasons, the person who decided and with what authority, and any conditions attached. A tension resolved silently is indistinguishable from one that was never noticed.

### 8.2 Uncertainty and Evidence Quality

Assessments are rarely made with complete evidence. A vendor may be unable to show how its system performs in the intended population; a benefit may be projected rather than observed; consultation may produce conflicting accounts. Uncertainty does not suspend judgment, but it must have a defined consequence. The assessment should record:

- the quality of the evidence and its relevance to the actual deployment, as distinct from the vendor's test conditions;
- the assumptions that remain unverified, and any findings that conflict;
- the severity of potential harm, whether it could be reversed, and how uncertain its likelihood is; and
- the response chosen: further testing, a restricted pilot with defined success and suspension criteria, deferral, or decline.

The greater the potential severity and the less reversible the harm, the less uncertainty an organization can responsibly accept. A restricted pilot is a legitimate response to uncertainty only if it is designed to resolve it and has a defined end.

## 8.3 Limits on What May Be Accepted

A named approver must accept the residual impacts of a deployment, but a signature cannot make every impact acceptable. Without limits, proportionality can become a justification for trading serious harm to one group against benefits elsewhere. Decisions should therefore follow an order:

1. **Exclude** unlawful uses and uses that breach the organization's redlines. These are not weighed.
2. **Prevent or reduce** the remaining harms, including by choosing a less intrusive design.
3. **Balance** the trade-offs that remain permissible, recording the tension and the reasons for the choice.
4. **Accept** remaining impacts only within the approver's defined authority, with conditions and review triggers. Impacts that are severe, irreversible or concentrated on people with little power to avoid them should require a higher level of authority.

## 9. From Principles to Practice

The following practices connect the principles to the requirements of ISO/IEC 42001. ISO/IEC 42005, the NIST AI RMF MAP function, the UNESCO Ethical Impact Assessment tool (UNESCO, 2023) and data protection impact assessment practice under GDPR Article 35 each offer comparable structures for carrying them out.

| Practice                                | Content                                                                                                                                                                                                 | Supports (ISO/IEC 42001)         |
|-----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| Threshold screen                        | Before model metrics are reviewed: legitimacy of purpose, necessity, foreseeable affected parties, and whether any redline is engaged                                                                   | Cl. 6.1.4; Cl. 8.4               |
| Affected-party mapping and consultation | Identification of those affected beyond users, including third parties, workers and their representatives, and marginalized groups; consultation that permits concerns to be raised without retaliation | Cl. 4.2; Cl. 6.1.4               |
| Organizational redlines                 | Capabilities the organization will not deploy regardless of commercial convenience, recorded separately from capabilities the law already prohibits                                                     | Cl. 5.2; Cl. 6.2                 |
| Decision record                         | Findings, alternatives considered, conditions imposed, residual impacts accepted, approver and authority, review date                                                                                   | Cl. 8.4 (documented information) |
| Follow-up and review triggers           | Monitoring of whether predicted impacts occurred; reassessment on change of purpose, population, vendor model, performance, complaints, incidents or law                                                | Cl. 8.4; Cl. 9.1; Cl. 10         |

### 9.1 Organizational Redlines

Redlines express an organization's own ethical floor. Examples an organization might adopt include continuous video capture inside employees' homes, keystroke logging used to score individual performance, and persistent location profiling of individuals outside working time. Emotion inference in the workplace is already prohibited in the European Union except for medical or safety reasons; an organization operating elsewhere may choose to adopt the same line. Redlines should be carried into procurement: where a vendor

cannot disable a capability that breaches a redline, the product should be excluded at the tender stage rather than configured after purchase. A redline is credible only if the organization can show that a proposed system was declined or modified because of it.

## 9.2 Perspectives the Assessment Should Hear

Different constituencies frame AI ethics differently, and each identifies impacts the others may miss. Workers and their representatives emphasize autonomy, the pace and content of work, and the right to contest automated management. Civil society organizations emphasize aggregate harms, structural bias and chilling effects that are invisible in individual transactions. Regulators emphasize the legal floor in their jurisdiction. Technical specialists emphasize the gap between the objective a system optimizes and the outcome the organization intends. An assessment informed by only one of these perspectives is incomplete.

# 10. Illustrative Scenarios

The following scenarios are hypothetical composites. They show how the principles translate into decisions; they do not describe any named organization. The first is set out as a complete decision record.

## 10.1 Monitoring Remote Workers: A Worked Decision Record

**Proposal.** A services firm handling client data proposes a system that captures a webcam image of each remote employee every few minutes during work on client files and uses facial verification to confirm that only the authorized employee is present. The claimed benefit is prevention of unauthorized viewing of client information.

The record below shows the complete chain from proposal to follow-up. All details are hypothetical.

| Element                 | Record                                                                                                                                                                                                                                                                                                                                                                                            |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Roles                   | Sponsor: Head of Client Operations. Assessor: AI governance and privacy lead, reporting to the Chief Risk Officer for this purpose. Approver: Chief Risk Officer. Escalation: Risk Committee.                                                                                                                                                                                                     |
| Evidence examined       | Incident history for the past three years; existing controls (multi-factor sign-in, endpoint data loss prevention, role-based access); vendor documentation; a consultation with affected employees and their representatives.                                                                                                                                                                    |
| What remained uncertain | The incremental protection that periodic imagery adds over existing controls was not quantified; no recorded incident involved an unauthorized viewer. The vendor could not provide verification error rates for home lighting conditions or across demographic groups.                                                                                                                           |
| Redline and legal check | Not unlawful in the operating jurisdiction. The organization's redline prohibits continuous video capture in homes; frequent periodic capture was judged to approach that line.                                                                                                                                                                                                                   |
| Ethical findings        | Intrusion into the home, including capture of household members (Section 4.2). Monitoring imposed as a condition of employment (4.3). Risk of uneven false mismatches between groups, untested (4.3). No defined path from an alert to any employment consequence (4.4). Employees reported effects on wellbeing and the use of shared rooms.                                                     |
| Decision                | Original design declined. Approved instead: facial re-verification at sign-in and on access to designated high-sensitivity client files only; no image retained after a successful match; session lock, without image capture, where an additional face is detected; mismatch images retained for 30 days for review. Approved as a 90-day pilot in one team, conditional on the vendor supplying |

| Element                     | Record                                                                                                                                                                                                                                                                                        |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                             | demographic performance data before any wider use.                                                                                                                                                                                                                                            |
| Reasons                     | Necessity: event-based verification addresses the stated risk at the points where it arises.<br>Proportionality: periodic imagery added unquantified benefit at substantial intrusion.<br>Uncertainty: the pilot is designed to measure error rates in the actual population.                 |
| Dissent                     | The employee representative objected to any imaging at sign-in. Response recorded: partially accepted; a non-biometric verification route is available to any employee on request, without justification.                                                                                     |
| Challenge and accommodation | A mismatch cannot lead to any disciplinary step without human review and an interview with the employee. Accommodation requests go to Human Resources, not to the sponsor.                                                                                                                    |
| Follow-up                   | Monthly review of mismatch rates by group, accommodation requests, complaints and incidents. Suspension triggers: a disparity in mismatch rates above the defined threshold, any adverse action taken without review, or a vendor model change. Reassessment at 90 days before any extension. |
| Influence and effectiveness | Influence: the design, scope and conditions changed. Effectiveness: to be established by pilot data on error rates, incidents and employee experience.                                                                                                                                        |

## 10.2 Licence Plate Reader Network

**Proposal.** A local authority, together with private property owners, proposes a network of automated licence plate readers whose data would be searchable across participating agencies.

**Assessment.** Necessity requires a defined investigative purpose and evidence of the network's contribution to it. Proportionality is engaged chiefly by retention and sharing: the value of a single observation may be modest, while the capacity to search observations across places and dates creates a record of movement. Misreads are foreseeable, as *Green* illustrates, so verification before any enforcement action is essential. In the European Union, systematic large-scale monitoring of a public area would require a data protection impact assessment.

**Decisions the assessment might produce.** Prompt deletion of reads that match no hotlist; a requirement to record a case reference and purpose for every search; restrictions on sharing beyond defined agencies; mandatory visual confirmation before a stop; periodic audit of searches; and public reporting of use.

## 10.3 AI-Assisted Applicant Screening

**Proposal.** An employer proposes a system that ranks job applicants and generates summaries of their applications for recruiters.

**Assessment.** Non-discrimination requires testing for disparate selection rates across groups; in New York City, a bias audit would be required by law if the tool fell within the scope of Local Law 144. Generated summaries may introduce errors or omit relevant information, and recruiters may defer to rankings, so automation bias must be addressed in the design of review. Applicants require notice, an explanation they can understand, a means to contest, and an accessible alternative where the format disadvantages applicants with disabilities. In the European Union, AI systems used for recruitment are high-risk, with obligations applying from December 2027.

**Decisions the assessment might produce.** Human review of every rejection recommended by the system; periodic disparity testing with a defined threshold for suspension; disclosure to applicants; an alternative assessment route; and prohibition of the use of inferred characteristics unrelated to the role.

## 10.4 Customer-Service Agent with Delegated Authority

**Proposal.** A retailer proposes an AI agent that resolves customer enquiries and is authorized to issue refunds and change account details without staff involvement.

**Assessment.** The principal impacts arise from the actions the agent can take rather than from what it observes. A single erroneous refund is minor; many erroneous refunds at machine speed, or changes to account ownership induced by a customer manipulating the agent, are not. Refunds can be reversed; a changed delivery address or payment detail may enable fraud that cannot. Customers must know they are dealing with an AI system and must be able to reach a person. Each action needs a record of what the agent did and on what basis.

**Decisions the assessment might produce.** Refund authority capped per transaction and per customer per period, with an aggregate daily limit; changes to payment details, delivery addresses and account ownership reserved for staff approval; automatic suspension when refund volume or value departs from a defined baseline; a named owner with authority to switch the agent off; disclosure to customers and a route to a person; and rollback procedures tested before launch.

## 11. Evidence That Ethics Changed the Decision

The history of impact assessment shows that a completed assessment is not the same as an effective one. The test of AI ethics in an organization is therefore a test of evidence. An assessment has ethical value when it subjects a proposal to credible challenge and demonstrably informs an accountable decision. A change to the proposal is strong evidence of influence. Approval without change is legitimate where the reasons are supported and recorded. Influence is not yet effectiveness: a condition imposed at approval shows that the assessment mattered, and only follow-up shows whether the condition worked.

That evidence takes recognizable forms:

- a condition imposed on use;
- a scope narrowed or a capability removed;
- an alternative tested and adopted;
- a deployment deferred or declined;
- an accommodation granted;
- a residual impact accepted in writing by a named person within defined authority, on stated grounds and conditions;
- an approval without change, supported by recorded reasons that address the findings; and
- follow-up showing whether predicted impacts occurred, whether conditions worked, and what was done when they did not.

Cosmetic change is not evidence of influence, and an organization whose assessments never alter a proposal should examine the assessment process before concluding that every proposal was sound. The ethical questions raised by AI are old ones. The responsibility they impose comes before any template: an organization must understand, and be prepared to answer for, the effects of the systems it chooses to use.

## Annex A. Leadership Questions

The following questions are intended for the executive or committee that approves a deployment.

### Purpose and evidence

- ☐ What problem does the system address, and what evidence shows that it addresses it?
- ☐ What less intrusive alternatives were tested, and why were they rejected?

### People affected

- ☐ Who is affected, including people who never interact with the system, and how were they consulted?
- ☐ Who receives the benefit, and who bears the burden?
- ☐ Has performance been tested across the groups the system will affect?

### Control

- ☐ What may happen because of an output, and what must a person verify first?
- ☐ How can an affected person challenge a result or obtain an accommodation?
- ☐ Who can suspend or stop the deployment, and on what triggers?

### Visibility

- ☐ What does the system observe and infer, who can use the information, where is it shared and for how long is it kept?
- ☐ Could the organization reconstruct the path from any output to the action that followed?

### Decision

- ☐ What did the assessment change in the proposal?
- ☐ Which residual impacts are being accepted, by whom, on what grounds and under what conditions?
- ☐ When, and on what events, will the assessment be revisited?

## Annex B. Crosswalk to Principal Frameworks

The table indicates thematic correspondence between each principle and principal frameworks. The frameworks differ in force: the EU AI Act imposes legal requirements within its scope; ISO/IEC 42001 contains requirements for organizations that adopt or certify to it; ISO/IEC 42005 and the NIST AI RMF are guidance; the OECD and UNESCO instruments are non-binding recommendations to states. Correspondence does not mean equivalence. For example, Article 50 transparency obligations and NIST explainability do not by themselves establish the operational traceability described in Section 4.5.

| Principle                           | OECD AI Principles (2024) — recommendation    | UNESCO Recommendation (2021) — recommendation                        | NIST AI RMF 1.0 — guidance                           | ISO/IEC 42001 requirement / 42005 guidance       | EU AI Act — legal requirement                          |
|-------------------------------------|-----------------------------------------------|----------------------------------------------------------------------|------------------------------------------------------|--------------------------------------------------|--------------------------------------------------------|
| Necessity and proportionality       | Human rights and democratic values            | Proportionality and do no harm                                       | MAP: context and impacts                             | 42001 cl. 6.1.4, 8.4; 42005 benefits and harms   | Art. 5 prohibitions; Art. 27 FRIA (limited scope)      |
| Dignity, agency, context            | Human rights; privacy                         | Human dignity; right to privacy                                      | Privacy-enhanced                                     | 42001 Annex C (privacy)                          | Art. 5(1)(f) emotion inference                         |
| Power, fairness, non-discrimination | Fairness; inclusive growth                    | Fairness and non-discrimination                                      | Fair, with harmful bias managed                      | 42001 Annex C (fairness); 42005 affected parties | Art. 10 data governance; Art. 26(7) worker information |
| Accountability and oversight        | Accountability                                | Human oversight and determination; responsibility and accountability | GOVERN; accountable and transparent                  | 42001 cl. 5, 8.4                                 | Art. 14 human oversight; Art. 26 deployers             |
| Transparency and traceability       | Transparency and explainability               | Transparency and explainability                                      | Explainable and interpretable                        | 42001 Annex C (transparency)                     | Art. 12 logging; Art. 50 transparency                  |
| Environment (domain)                | Sustainable development; natural environments | Environment and ecosystem flourishing                                | MEASURE 2.12 environmental impact and sustainability | 42001 Annex C (environmental impact)             | Annex XI energy consumption (GPAI)                     |

The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225, opened for signature 5 September 2024) addresses each principle and adds obligations on remedies and procedural safeguards.

## Annex C. References

- Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, 128(6), 2188–2244. <https://doi.org/10.1086/705716>
- Ada Lovelace Institute. (2022). *Algorithmic impact assessment: A case study in healthcare*. <https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/>
- AFL-CIO. (2025, October 15). *AFL-CIO launches 'Workers First Initiative on AI'* [Press release]. <https://aflcio.org/node/9902>
- American Civil Liberties Union. (2013). *You are being tracked: How license plate readers are being used to record Americans' movements*. <https://www.aclu-md.org/app/uploads/2018/06/071613-aclu-alprreport-opt-v05.pdf>
- Article 29 Data Protection Working Party. (2017). *Opinion 2/2017 on data processing at work* (WP 249).
- Arts, J., Caldwell, P., & Morrison-Saunders, A. (2001). Environmental impact assessment follow-up: Good practice and future directions – findings from a workshop at the IAIA 2000 conference. *Impact Assessment and Project Appraisal*, 19(3), 175–185. <https://doi.org/10.3152/147154601781767014>
- Bărbulescu v. Romania [GC], No. 61496/08 (European Court of Human Rights, 5 September 2017).
- Cashmore, M., Gwilliam, R., Morgan, R., Cobb, D., & Bond, A. (2004). The interminable issue of effectiveness: Substantive purposes, outcomes and research challenges in the advancement of environmental impact assessment theory. *Impact Assessment and Project Appraisal*, 22(4), 295–310. <https://doi.org/10.3152/147154604781765860>
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153–163. <https://doi.org/10.1089/big.2016.0047>
- Council of Europe. (2024). *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225).
- Directive 2011/92/EU of the European Parliament and of the Council of 13 December 2011 on the assessment of the effects of certain public and private projects on the environment, as amended by Directive 2014/52/EU.
- Directive (EU) 2024/1760 of the European Parliament and of the Council on corporate sustainability due diligence, as amended by Directive (EU) 2026/470.
- Electronic Frontier Foundation. (2023). *Automated license plate readers (ALPRs)*. Street-Level Surveillance. <https://sls.eff.org/technologies/automated-license-plate-readers-alprs>
- European Data Protection Board. (2020). *Guidelines 05/2020 on consent under Regulation 2016/679* (Version 1.1).
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Green v. City and County of San Francisco*, 751 F.3d 1039 (9th Cir. 2014).
- Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face recognition vendor test (FRVT) part 3: Demographic effects* (NISTIR 8280). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.IR.8280>
- International Energy Agency. (2025). *Energy and AI*. <https://www.iea.org/reports/energy-and-ai>
- International Organization for Standardization. (2010). *Guidance on social responsibility* (ISO 26000:2010).
- International Organization for Standardization. (2015). *Environmental management systems — Requirements with guidance for use* (ISO 14001:2015).
- International Organization for Standardization. (2025). *Robotics — Safety requirements — Part 1: Industrial robots* (ISO 10218-1:2025); *Part 2: Industrial robot applications and robot cells* (ISO 10218-2:2025).
- ISO/IEC. (2022). *Information technology — Artificial intelligence — Overview of ethical and societal concerns* (ISO/IEC TR 24368:2022).

- ISO/IEC. (2023). *Information technology — Artificial intelligence — Management system* (ISO/IEC 42001:2023).
- ISO/IEC. (2025). *Information technology — Artificial intelligence (AI) — AI system impact assessment* (ISO/IEC 42005:2025).
- Jay, S., Jones, C., Slinn, P., & Wood, C. (2007). Environmental impact assessment: Retrospect and prospect. *Environmental Impact Assessment Review*, 27(4), 287–300. <https://doi.org/10.1016/j.eiar.2006.12.001>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- López Ribalda and Others v. Spain [GC], Nos. 1874/13 and 8567/13 (European Court of Human Rights, 17 October 2019).
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Morrison-Saunders, A., & Arts, J. (Eds.). (2004). *Assessing impact: Handbook of EIA and SEA follow-up*. Earthscan.
- National Environmental Policy Act of 1969, 42 U.S.C. § 4321 et seq.
- National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). <https://doi.org/10.6028/NIST.AI.100-1>
- New York City Local Law 144 of 2021 (automated employment decision tools).
- Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), 119–157.
- Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.
- Occupational Safety and Health Act of 1970, 29 U.S.C. § 651 et seq.
- OECD. (2019, revised 2024). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449).
- OECD. (2023). *OECD guidelines for multinational enterprises on responsible business conduct*.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410. <https://doi.org/10.1177/0018720810376055>
- Penney, J. W. (2016). Chilling effects: Online surveillance and Wikipedia use. *Berkeley Technology Law Journal*, 31(1), 117. <https://doi.org/10.15779/Z38SS13>
- Rechtbank Zeeland-West-Brabant. (2022, September 28). ECLI:NL:RBZWB:2022:5656.
- Regulation (EU) 2016/679 (General Data Protection Regulation).
- Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery.
- Regulation (EU) 2024/1689 (Artificial Intelligence Act), as amended by Regulation (EU) 2026/1744.
- Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.
- Tegmark, M. (2016, January 7). *The wisdom race is heating up*. Future of Life Institute. <https://futureoflife.org/nuclear/the-wisdom-race-is-heating-up/>
- Trades Union Congress. (2021). *Dignity at work and the AI revolution: A TUC manifesto*. <https://www.tuc.org.uk/research-analysis/reports/dignity-work-and-ai-revolution>
- Trades Union Congress. (2024). *Artificial Intelligence (Regulation and Employment Rights) Bill* [Draft bill]. <https://www.tuc.org.uk/ai-reports>
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- UNESCO. (2023). *Ethical impact assessment: A tool of the Recommendation on the Ethics of Artificial Intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000386276>

United Nations. (2011). *Guiding principles on business and human rights*. Office of the High Commissioner for Human Rights.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.