International Labour Organization

ILO Working Paper 182

October / 2026

AI Ethics and the ILO's international labour standards system: exploring synergies

Authors / Jordi Agusti-Panareda, Daniel Samaan

© International Labour Organization 2026

Attribution 4.0 International (CC BY 4.0)

This work is licensed under the Creative Commons Attribution 4.0 International. See: https://creativecommons.org/licenses/by/4.0/. The user is allowed to reuse, share (copy and redistribute), adapt (remix, transform and build upon the original work) as detailed in the licence. The user must clearly credit the ILO as the source of the material and indicate if changes were made to the original content. Use of the emblem, name and logo of the ILO is not permitted in connection with translations, adaptations or other derivative works.

Attribution – The user must indicate if changes were made and must cite the work as follows: Agusti-Panareda, J., Samaan, D. *AI Ethics and the ILO's international labour standards system: exploring synergies*. ILO Working Paper 182. Geneva: International Labour Office, 2026.© ILO.

Translations – In case of a translation of this work, the following disclaimer must be added along with the attribution: *This is a translation of a copyrighted work of the International Labour Organization (ILO). This translation has not been prepared, reviewed or endorsed by the ILO and should not be considered an official ILO translation. The ILO disclaims all responsibility for its content and accuracy. Responsibility rests solely with the author(s) of the translation.*

Adaptations – In case of an adaptation of this work, the following disclaimer must be added along with the attribution: *This is an adaptation of a copyrighted work of the International Labour Organization (ILO). This adaptation has not been prepared, reviewed or endorsed by the ILO and should not be considered an official ILO adaptation. The ILO disclaims all responsibility for its content and accuracy. Responsibility rests solely with the author(s) of the adaptation.*

Third-party materials – This Creative Commons licence does not apply to non-ILO copyright materials included in this publication. If the material is attributed to a third party, the user of such material is solely responsible for clearing the rights with the rights holder and for any claims of infringement.

Any dispute arising under this licence that cannot be settled amicably shall be referred to arbitration in accordance with the Arbitration Rules of the United Nations Commission on International Trade Law (UNCITRAL). The parties shall be bound by any arbitration award rendered as a result of such arbitration as the final adjudication of such a dispute.

For details on rights and licensing, contact: rights@ilo.org. For details on ILO publications and digital products, visit: www.ilo.org/publns.

ISBN 9789220439449 (print), ISBN 9789220439456 (web PDF), ISBN 9789220439463 (epub), ISBN 9789220439470 (html). ISSN 2708-3438 (print), ISSN 2708-3446 (digital)

https://doi.org/10.54394/00035324

The designations employed in ILO publications, which are in conformity with United Nations practice, and the presentation of material therein do not imply the expression of any opinion whatsoever on the part of the ILO concerning the legal status of any country, area or territory

or of its authorities, or concerning the delimitation of its frontiers or boundaries. See: www.ilo.org/disclaimer.

The opinions and views expressed in this publication are those of the author(s) and do not necessarily reflect the opinions, views or policies of the ILO. During the preparation of this work, the authors utilized ChatGPT (OpenAI) for drafting assistance, language editing, and improving the clarity and structure of selected passages. Following the use of this tool, the authors carefully reviewed and edited all content as necessary, and assume(s) full responsibility for the document/publication's content. No material subject to intellectual property rights, including third party material, was used without acknowledgment and permission.

Reference to names of firms and commercial products and processes does not imply their endorsement by the ILO, and any failure to mention a particular firm, commercial product or process is not a sign of disapproval.

Information on ILO publications and digital products can be found at: www.ilo.org/research-and-publications

ILO Working Papers summarize the results of ILO research in progress, and seek to stimulate discussion of a range of issues related to the world of work. Comments on this ILO Working Paper are welcome and can be sent to samaan@ilo.org.

Authorization for publication: Caroline Fredrickson, Director, Research Department

ILO Working Papers can be found at: www.ilo.org/research-and-publications/working-papers

Suggested citation:
Agusti-Panareda, J., Samaan, D. 2026. *AI Ethics and the ILO's international labour standards system: exploring synergies*, ILO Working Paper 182 (Geneva, ILO). https://doi.org/10.54394/00035324

Abstract

Artificial intelligence (AI) introduces governance questions that increasingly affect labour markets, yet current AI ethics debates rarely engage systematically with the century-long normative tradition of the International Labour Organization (ILO). This paper conducts an exploratory comparative analysis between five widely cited AI ethics principles and the ethical foundations underlying the ILO's normative framework. It uses an interpretive mapping of these AI ethics principles against relevant ILO values, principles and fundamental rights, including those expressed in the ILO Constitution, the Declaration of Philadelphia and the Declaration on Fundamental Principles and Rights at Work (FPRW). The mapping identifies significant areas of convergence, especially in relation to the AI ethics principles of justice, beneficence, and non-maleficence, while also revealing divergences in focus concerning privacy, autonomy, and the collective dimensions of worker and employer representation. On this basis, the paper advances a two-fold argument: (i) that the ILO's normative framework provides a coherent and legitimate reference point for applying ethical AI principles in labour-related settings; and (ii) that AI ethics sheds light on emerging issues—such as data governance, explicability, and algorithmic management—that are relevant to the application of ILS in technologically evolving workplaces. By situating AI ethics within international labour law, the paper identifies pathways for mutual reinforcement between AI ethics and ILS, and points to the need for further work on labour-specific governance on AI.

The views expressed herein are those of the authors and do not necessarily reflect the views of the International Labour Organization.

About the authors

Jordi Agusti-Panareda, socio-legal researcher and Ethics Officer at the International Labour Organization (ILO) in Geneva.

Daniel Samaan is a Senior Economist at the Research and Statistics Department of the International Labour Organization (ILO) in Geneva.

Table of contents

Abstract 01

About the authors 01

Introduction 03

1 Methodology 05

2 AI Ethics and the world of work 07

3 Ethical foundations of the ILO's normative framework 10

4 Exploring the relationship between AI Ethics Principles and ILO's ethical framework 12

Areas of Convergence 12

Areas of divergence 14

5 Positioning AI Ethics in the Context of International Labour Standards 17

Conclusion 21

References 22

Acknowledgements 25

▶ Introduction

As AI continues to reshape economies, workplaces, and societies, policymakers have begun to pursue regulatory responses. Contrasting views, as illustrated by the differing approaches in two of the world's largest economic areas[1] illustrate the current debates on AI governance and the lack of consensus on the right balance between innovation and regulation. In many other jurisdictions, comprehensive AI legislation is still absent. Instead, governance often relies on "soft law" instruments such as voluntary guidelines, recommendations, national strategies, or international initiatives. Under the broad labels of "responsible AI," or "ethical AI" corporations, public institutions, and professional associations have issued codes of conduct and ethical frameworks, while non-profit organizations have published declarations aimed at guiding development and adoption. These efforts, although usually non-binding, have become influential reference points in policy debates and in some cases have laid the groundwork for future regulatory initiatives.

The debate on AI ethics and AI governance is not specifically targeted at or limited to addressing concerns related to AI's impact on the world of work but rather at broader questions about the social, political, and moral implications of AI on our societies. In both academic and policy literature, ethical AI is usually framed around identifying principles that would ensure AI systems operate in ways that are morally appropriate (e.g., Boddington, 2017; Bostrom, 2014; Floridi & Cowls, 2019; Yuste et al., 2017). This discourse about ethical or "responsible" AI is taking place in the theoretical space but is also heavily influenced by the emergence of legally non-binding AI ethics guidelines. Drafting of AI ethics guidelines is generally a cross-disciplinary undertaking that draws on expertise from fields such as economics, law, philosophy, and computer science. Because these initiatives emerge from diverse institutional settings—each governed by its own norms, priorities, and conceptions of what constitutes "ethical AI"—the resulting landscape is conceptually complex. This diversity makes it challenging to trace how distinct and sometimes conflicting perspectives converge into a shared understanding of ethical conduct in the development and use of AI.

Samaan et al. (in press) provide an extensive empirical analysis of AI ethics guidelines to date. Their study reviews 245 documents issued by a wide range of actors and applies systematic content analysis supported by natural language processing (NLP) and large language models (LLMs) to examine how ethical principles are articulated and contextualized, which themes dominate, and to what extent the world of work is considered. The analysis identifies seven recurring thematic clusters: autonomous systems, AI policy, data protection, financial applications, responsible AI, health, and governance. Notably, labour markets or the "world of work" do not appear as a distinct topic across the guidelines. Still, many guidelines raise concerns about employment, working conditions, or labour relations, underscoring the salience of work-related issues in the ethics discourse. When evaluating the extent to which AI ethics guidelines align with established labour standards, Samaan et al. (in press) identify a significant gap. The study shows that very

[1] At the time of writing this paper, in Europe, the EU AI Act (European Parliament & Council of the European Union, 2024) represents the most comprehensive effort so far, introducing a risk-based framework that will have implications well beyond the EU. Some voices, particularly from industry, have recently raised criticism of the Act and warned of "overregulation" that may stifle innovation. The United States initially took a markedly different path. Following the rescission of the 2023 Executive Order on "Safe, Secure, and Trustworthy AI," the federal government adopted a more innovation-led and less precautionary stance, emphasizing national competitiveness and the removal of barriers to AI development. More recently, however, the United States has articulated a National AI Legislative Framework that sets out federal priorities in areas such as child protection, community impacts, intellectual property, free speech, innovation, and workforce development. While the practical implications of this framework remain to be seen, it suggests a more affirmative federal policy agenda than the earlier deregulatory shift alone, while still differing significantly from the EU's binding risk-based model.

few guidelines explicitly reference the International Labour Organization (ILO) or its International Labour Standards (ILS), and only a small number implicitly touch on issues aligned with the ILO's Fundamental Principles and Rights at Work (FPRW), even though AI ethics guidelines acknowledge labour-related risks. Recent scholarship has started examining workplace AI through established human-rights and labour-law frameworks, including the role of the ILO and collective bargaining, but the relationship between contemporary AI ethics principles and the ethical foundations of the ILO's normative framework remains underexplored (Santoro & Stone, 2026).

Considering these findings, we believe that ethical and legal commitments embedded in international labour law and reflected in the ILO's normative and supervisory practice remain insufficiently integrated into the ongoing debate about AI governance and AI ethics. A crucial question left open by the existing literature, which we attempt to address in this paper, is: how do AI ethics principles (nascent, heterogenous and nonbinding) relate to the values and principles underlying the ILO's normative framework (anchored on a Constitution ratified by 186 states, highly developed through myriad ILS and with over a century of supervisory practice)? More specifically, we seek to address the following gap: despite the widespread issuance of AI ethics guidelines, their relationship to established ILO principles and ILS remains underexplored. Much of the ethics discourse proceeds in isolation from the century-long normative architecture developed by the ILO, in which constitutional values, principles and rights are developed through Conventions, Recommendations (and other soft law instruments), with well-established supervisory mechanisms and institutional practice. While the ILO has no officially designated ethical framework, and its normative system cannot be equated with or reduced to an ethical framework, we argue that the ILO's normative system embodies implicit ethical foundations that can be meaningfully compared with contemporary AI ethics.[2]

We examine the relationship between AI ethics principles and the ILO's ethical framework with a dual purpose: (i) to reflect on what the ILO can draw from AI ethics principles as it considers how to address the specific challenges and transformations of the world of work resulting from AI; and (ii) to consider what discussions on AI ethics can, in turn, learn from the ILO's social justice, decent work, and human-centred corpus and approaches.[3] As we also aim to position AI ethics within the field of international labour law, this paper seeks to contribute to a normative approach to aligning AI ethics with labour standards.

[2] Throughout the paper, we use the term "ILO's ethical framework" strictly in this analytical sense: that the ILO's normative system embodies implicit ethical foundations that can be analytically reconstructed as a framework for comparative purposes. We do not suggest that the ILO has officially designated an "ethical framework", nor that its normative system can be equated with or reduced to such a framework. Rather, the "ILO's ethical framework" is an analytical construct based on our interpretation of the values, principles, rights and other normative commitments expressed in the ILO's constitutional texts, declarations, international labour standards and related supervisory practice.

[3] As to the reflection within the ILO on harnessing artificial intelligence for decent work see the recently published report of the Director General to the 114th Session of the International Labour Conference.

▶ 1 Methodology

The aim of this article is to examine the potential gaps and existing synergies between commonly invoked ethical principles for the development and use of AI and the ILO's normative system. Such an analysis requires several methodological choices. There exists no single framework that clearly defines AI ethics, nor is there agreement on whether AI ethics should be grounded in principlism, deontology, virtue ethics care ethics, etc.

For a meaningful comparison, we synthesize an "AI ethics framework" from a wide range of AI ethics documents, which often use different terminology and operate at different levels of abstraction, referring, for example, to principles, values, or rights, or combining these concepts. The formulation of principles constitutes the most visible and recurrent organizing vocabulary. In practice, institutional AI ethics guidelines have emerged rapidly over the last years and although AI ethics as a field is not limited to principlism, the guideline literature is overwhelmingly principle-based (Jobin et al., 2019; Samaan et al., in press). These developments have prompted us to base the AI ethics framework for our analysis on principles. Floridi and Cowls (2019) provide one of the most influential principlist syntheses in AI ethics, and their framework is broadly consistent with meta-analyses of AI ethics guidelines that find convergence around recurring mid-level principles, even while emphasizing continuing disagreement about how those principles should be interpreted and implemented in practice. We therefore use Floridi and Cowls (2019) as the operative AI ethics framework for this analysis.

On the ILO side, we observe a richer and much more complex normative framework, which we discuss in detail in Section 4. ILO instruments and supervisory practice encompass values, principles, rights and obligations, with principles playing an important role in articulating the Organization's foundational commitments.[4] For the purposes of this paper, we analyze and interpret these materials to identify the implicit ethical foundations of the ILO's normative system, drawing in particular on its Constitution, the Declaration of Philadelphia, the Declaration on Fundamental Principles and Rights at Work, ILS and related supervisory practice. We refer to these analytically identified ethical foundations as the "ILO ethical framework" and compare them with the selected AI ethics principles to identify similarities and differences.

Methodologically, the paper therefore undertakes an exploratory interpretive mapping of the selected AI ethics principles against ethical commitments identified within the ILO's normative system, in order to identify areas of convergence, divergence and complementarity, without claiming conceptual equivalence or exhaustiveness, or seeking to derive an ethics of labour from AI ethics.

Before turning to the analysis, two limitations should be noted. First, Floridi and Cowls (2019) represent only one among several prominent principlist or values-based approaches to AI ethics, including the European Commission High-Level Expert Group on Artificial Intelligence's (2019) Ethics Guidelines for Trustworthy AI, the OECD's (2019) AI Principles, UNESCO's (2021) Recommendation on the Ethics of Artificial Intelligence, IEEE's (2019) Ethically Aligned Design, and WHO's (2021)

[4] The distinction between values, principles and rights, further developed in Section 4, was also discussed during the preparation of the 1998 FPRW Declaration. Francis Maupain, Special Advisor, described values as widely shared concepts of moral order, principles as their manifestation in action in a particular context, and rights as an active acknowledgement in law of the principles. See ILC, 86th Session, Report of the Committee on the Declaration of Principles, para. 73.

sectoral guidance on AI in health. Our choice of Floridi and Cowls as the operative framework is therefore methodological rather than exhaustive, and the use of another framework might produce a somewhat different mapping.

Second, non-principlist approaches could also be compared with the ILO's normative framework and may in some respects be closer to its legal and institutional architecture. These include human-rights-based and legal-normative approaches, such as the Council of Europe's (2024) *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, OHCHR's (2025) guidance on human rights due diligence for digital technology use, human-rights impact assessment approaches (Mantelero & Esposito, 2021), and recent labour-specific rights-based analyses of AI governance (Santoro & Stone, 2026). Other relevant alternatives include governance- and process-oriented frameworks such as the NIST AI Risk Management Framework, Responsible Innovation and Value Sensitive Design, as well as broader philosophical approaches including virtue ethics, care ethics, capability approaches and deontological ethics. These alternatives fall outside the scope of the present exploratory analysis but could yield different points of convergence and divergence with the ILO framework.

▶ 2 AI Ethics and the world of work

The volume of AI ethics guidelines has expanded rapidly in recent years. Early systematic reviews were carried out by Floridi and Cowls (2019) and Jobin et al. (2019). Clarke (2019) examined guidelines from 30 organizations and distilled them into a catalogue of 50 "responsible AI" principles, ranging from robustness and resilience to human safety and well-being. Algorithm Watch reported more than 170 guidelines in its Global Inventory by 2021 (Deshpande & Sharp, 2022). Khan et al. (2022), drawing on a systematic literature review, analyzed 27 studies and identified 22 principles alongside 15 key challenges, emphasizing transparency, privacy, accountability, and fairness as priorities for AI system designers. In another recent review, Corrêa et al. (2023) catalogued 17 ethical principles.

Taken together, these reviews highlight both the diversity of AI ethics guidelines in terms of scope and content, and the frequent absence of clearly defined objectives or purposes of AI ethics. The most recent review, conducted by Samaan et al. (in press), is also the first to analyze AI ethics guidelines explicitly from a labour market perspective.[5] Their study confirms that the dominant ethical principles remain those outlined by Floridi and Cowls (2019) and Jobin et al. (2019) but raises additional questions about how such principles should be interpreted in the specific context of labour markets. This challenge is not unique to work-related issues; similar concerns have been voiced in other domains such as healthcare, public administration, and financial services, where high-level AI ethics principles have been criticized for their limited operational usefulness. For example, Hagendorff (2020) and Munn (2023) criticize AI ethics guidelines more broadly for their lack of operationalization and for failing to translate principles into concrete actions or consequences. More generally, AI ethics guidelines have been criticized as strategic instruments to deflect regulatory pressure by emphasizing abstract principles and technical fixes rather than addressing concrete governance needs (e.g., Greene et al., 2019; Mittelstadt, 2019).

This criticism regarding lack of concrete consequences rests on two issues. First, the fact that ethical principles are invoked means that, by their very nature, they remain general guidelines for the formulation of more specific rules (Beauchamp & Childress, 2013), without laying out the consequences. Second, when expressed in the form of institutional guidelines lacking any enforcement or supervisory mechanism, such principles clearly lack the capacity to oblige actors to any specific behavior, unlike for example legal norms that carry binding force.

Our focus in this paper is not on the status of AI ethics guidelines as voluntary or binding instruments. Rather, we are interested in the ethical principles for AI that these documents put forward, and how such principles relate to ILO's normative framework, in particular its values and principles . The various empirical reviews of AI ethics guidelines serve primarily as an academic resource from which we extract insights into which principles are currently being debated in academic, policy, and public discourse. We also recognize that treating all sources of AI ethics guidelines as equivalent—whether originating from governments, intergovernmental organizations, corporations, NGOs, or academia—risks obscuring important differences in their political relevance and practical impact (Mantelero & Esposito, 2021). Although certain principles or values may appear frequently across documents, their recurrence alone does not capture the

[5] See also the related ILO blog Samaan, D. (2025, 14 November). https://www.ilo.org/resource/article/governing-ai-world-work-review-global-ethics-guidelines

varying weight they carry depending on institutional origin. For example, principles repeatedly articulated in intergovernmental texts are likely to exert greater regulatory influence than those cited with equal frequency in academic or NGO publications.

For the methodological reasons set out in Section 2, we rely on Floridi and Cowls' framework as the primary reference point for our discussion, supplemented to a limited extent by Jobin et al. (2019). Floridi and Cowls (2019) provided one of the first convergence exercises, identifying five core principles for ethical AI: (i) beneficence, (ii) non-maleficence, (iii) autonomy, (iv) justice, and (v) explicability, understood both as intelligibility and as accountability. Their framework builds directly on Beauchamp and Childress's (2013) *Principles of biomedical ethics*, a canonical work in applied ethics that established the four widely recognized principles of beneficence, non-maleficence, autonomy, and justice. By adapting these principles to the context of artificial intelligence and adding "explicability" as an enabling principle, Floridi and Cowls provide one of the most coherent theoretical foundations for AI ethics. In addition to this conceptual contribution, they conducted an empirical review of six "high-profile" AI ethics initiatives to extract and substantiate these principles. Jobin et al. (2019), in turn, complement this approach with an empirical synthesis mapping the convergence of ethical principles across a broader corpus of 84 guidelines.[6]

Floridi and Cowls (2019) offer explanations that clarify how the five principles should be understood in the context of AI. **Beneficence** refers to the expectation that the creation of AI technology is beneficial to humanity, i.e. promoting the well-being of people and preserving their dignity. **Non-maleficence** is not the logical equivalent of beneficence. Non-maleficence cautions against the potentially negative consequences of overusing or misusing (i.e. intentional harm), or misoperation of AI technologies. Floridi and Cowls (2019) subsume the prevention of infringements on personal privacy, hence "privacy" and control over personal data, as a main concern under maleficence, but acknowledge that other harms can relate to this principle. **Autonomy** is understood as the power to decide (to decide).[7] In the context of AI, maintaining autonomy appears contradictory since adopting AI and its smart agency means willingly cede decision-making power to machines. Thus, the principle of autonomy in the context of AI requires finding a balance between decision-making power of machines and humans respectively, or that humans should retain the power to decide which decisions to take, and any delegation should remain overridable, i.e. the power to decide to decide again. The last of the four classic bioethics principles that is being proposed for AI ethics is justice. **Justice** as ethical principle typically refers to distribution of benefits and resources as well as eliminating barriers of unjustified unequal access (discrimination). As fifth ethical principle Floridi and Cowls (2019) propose **Explicability** as an enabler of the other principles through intelligibility and accountability. Beneficence and non-maleficence require that we can assess the societal benefits and harms of AI; autonomy requires that choices about delegating decisions be based on understanding how AI would act in our place; and

[6] In their empirical review of 84 guidelines from the public and private sectors, they found a convergence around eleven ethical principles, several of which partly overlap with Floridi and Cowls' framework. These are: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity. From this list, one can immediately see overlap with four of Floridi's principles (underlined in the preceding list), as well as "transparency," which Floridi subsumes under explicability (together with accountability and responsibility). Of the remaining six principles, Floridi argues that "privacy" (or protecting privacy) belongs to non-maleficence because one of the biggest menaces of AI is to infringe privacy. "Responsibility" connects to transparency and accountability. The other four concepts—trust, dignity, sustainability, and solidarity—can also be integrated into Floridi's framework, which we discuss in more detail below. (Human) dignity and sustainability are to be preserved or maintained through beneficence. Solidarity belongs to justice. Trust is related to explicability, which incorporates accountability and responsibility.

[7] In light of the distinction drawn in Section 2 between values and principles, some of these terms, taken in isolation, may also be read as values. In Floridi and Cowls' framework, however, they are used as principles in the sense of general action-guiding norms. The relevant distinction lies less in the word itself than in its function: where a term primarily denotes a desirable condition, it may function as a value; where it is used to guide judgment, justify decisions, and orient conduct, it functions as a principle.

justice requires mechanisms to hold developers and users, or deployers accountable for harmful outcomes. Without explicability, these other principles cannot be meaningfully applied.

As previously indicated, the actual application of these ethical principles in different contexts remains debated and often unclear. This is equally true for the world of work, where the implications of abstract AI ethics principles are largely unexplored. It is a reality, however, that AI is increasingly being used in the economy to replace human labor, to increase workers' and organizational efficiency, or to oversee and manage human labor (Del Rio Chanona et al., 2025). Would an AI that is developed and employed in accordance with the five ethical principles ensure that desired labor market outcomes can be achieved? In their review of institutional and organizational AI ethics guidelines, Samaan et al. (in press) explore to what extent such guidelines refer to world-of-work-related matters. In other words, the question is whether organizations that employ AI ethics guidelines do so because they believe – possibly among other reasons – that the use of AI may raise ethical concerns in labor market contexts. The answer is yes for more than 60 percent of the reviewed guidelines: AI ethics is (also) discussed in the context of the world of work. The ILO and its standards and principles, however, are rarely mentioned (in less the 20 percent of the guidelines).

The gap is particularly striking given that the governance of labour has long been shaped by international labour law with its well-developed normative system under the auspices of the ILO. The ILO's framework, grounded in its Constitution, including the Declaration of Philadelphia, and reaffirmed through the ILO's Declaration on Fundamental Principles and Rights at Work, articulates values, principles and rights that have been consolidated into international labour standards as well as guidance by the ILO supervisory bodies. In the next section, we turn to these normative elements and the ethical foundations they collectively embody, not only to examine their relevance for addressing the challenges posed by AI in the workplace, but also to assess whether and how they might provide a complementary perspective to the existing AI ethics discourse.

3 Ethical foundations of the ILO's normative framework

The ILO is a value-based organization, established in 1919 on the wake of World War I to advance social justice as a precondition for universal peace.[8] The ILO's normative framework follows a well-established structure ranging from general values (e.g. freedom, social justice, etc.) to principles (as manifestations of such values into action in a specific context - e.g. freedom of association, non-discrimination, etc.) and rights and obligations (as a concrete manifestation of these principles in law, notably as set out in international labour Conventions - e.g. the right of employers and workers without distinction whatsoever to create and join organizations for the defence of their interests). The ILO's normative framework is further concretized through Recommendations, guidelines, codes of practice and other soft law instruments, as well as in the guidance provided by its supervisory organs.[9]

The ILO's core values and principles, as well as the ethical mandate of the Organization, are set out in the ILO Constitution, in particular its Annex containing the Declaration concerning the aims and purposes of the International Labour Organisation – known as the Declaration of Philadelphia.[10] On the basis of such constitutional provisions, ILO values and principles have been further affirmed through Resolutions of the International Labour Conference – the ILO's general tripartite assembly. Of particular relevance in this regard is the ILO Declaration on Fundamental Principles and Rights of work, which recalls that all ILO member states, whether or not they have ratified the relevant Conventions, have the obligation to respect promote and realize the following fundamental principles and rights at work (FPRW):[11]

a) freedom of association and the effective recognition of the right to collective bargaining;

b) the elimination of all forms of forced or compulsory labour;

c) the effective abolition of child labour;

d) the elimination of discrimination in respect of employment and occupation; and

e) a safe and healthy working environment.

More broadly, the ILO's axiological framework is grounded on a human centred pursuit of social justice. As such the ILO Constitution, in its Philadelphia Declaration, affirms as the Organization's fundamental objective that "all human beings, irrespective of race, creed or sex, have the right to pursue both their material well-being and their spiritual development in conditions of freedom and dignity, of economic security and equal opportunity". To attain it, the Declaration adds that "all national and international policies and measures, in particular those of an economic and financial character, should be judged in this light and accepted only in so far as they may be held to promote and not to hinder the achievement of this fundamental objective;" and declares that "it is a responsibility of the ILO to examine and consider all international economic

8 See preamble to the ILO Constition. For a deeper analysis on the ILO as an axiological values-based organization see Agusti-Panareda (2023).

9 For an overview of the ILO's international labour standards and supervisory system see Rules of the Game: An introduction to the standards-related work of the International Labour Organi...

10 See 1946 Declaration of Philadephia annexed to the ILO Constition.

11 See ILO's Declaration on Fundamental Principles and Rights at Work (as amended in 2022)

and financial policies and measures in the light of this fundamental objective".[12] In a similar vein, the ILO Centenary Declaration calls for a human-centred approach to the future of work.[13] It is on the basis of this ethical framework, codified into international law through the ILO Constitution, that the Organization carries out its standards setting and supervisory work and other activities to promote decent work and social justice.

[12] Furthermore the Declaration of Philadephia sets out as obligation of the ILO "to further amont the nations of the world programmes which will achieve: (a) full employment and the raising of standards of living; (b) the employment of workers in the occupations in which they can have the satisfaction of giving the fullest measure of their skill and attainments and make their greatest contribution to the common well-being; (c) the provision, as a means to the attainment of this end and under adequate guarantees for all concerned, of facilities for training and the transfer of labour, including migration for employment and settlement; (d) policies in regard to wages and earnings, hours and other conditions of work calculated to ensure a just share of the fruits of progress to all, and a minimum living wage to all employed and in need of such protection; (e) the effective recognition of the right of collective bargaining, the cooperation of management and labour in the continuous improvement of productive efficiency, and the collaboration of workers and employers in the preparation and application of social and economic measures; (f) the extension of social security measures to provide a basic income to all in need of such protection and comprehensive medical care; (g) adequate protection for the life and health of workers in all occupations; (h) provision for child welfare and maternity protection; (i) the provision of adequate nutrition, housing and facilities for recreation and culture; and (j) the assurance of equality of educational and vocational opportunities."

[13] See the ILO Centenary Declaration for the Future of Work, 2019.

▶ 4 Exploring the relationship between AI Ethics Principles and ILO's ethical framework

The five commonly cited AI ethics principles and the ethical foundations embodied in the ILO's normative system, are fundamentally different in nature. As explained in Section 2, however, they can be meaningfully compared at the level of principles in order to identify areas of convergence, divergence and possible complementarity. The ILO framework constitutes a coherent and longstanding body of international law supported by a sophisticated supervisory system, whereas AI ethics principles represent a more recent and still emerging set of normative orientations. AI ethics principles are broad in scope, seeking to provide general guidance on how AI should be designed and used across all areas of society. The ILO's framework, by contrast, is more context-specific but also far richer and more developed in terms of applicability and operationalization. Although the two operate at different levels of abstraction, both express ethical orientations and commitments to human well-being, dignity, and fairness. Comparing these frameworks, therefore remains meaningful for understanding how they relate to one another, where they converge or diverge, and what synergies may be drawn for the governance of AI in the world of work.

Areas of Convergence

AI ethics principles and the ethical foundations embodied in the ILO's normative framework, including its values, principles, rights and other standards converge in important respects. Both frameworks seek to guide social and technological transformations in our societies in a way that they have a positive impact and no negative consequences on human life, well-being and dignity. To illustrate this convergence, we note how some of the most widespread AI ethics principles discussed above relate to core ILO values, principles and standards, while recognizing that the comparison is not exhaustive and that particular ILO values and fundamental principles and rights may be related to more than one AI ethics principle.[14]

• **Justice:** There exists a strong overlap between the AI ethics principle of justice and the overall mandate of the ILO to pursue social justice.[15] The Declaration of Philadelphia provides several articles that formulate principles which resonate with the AI ethical principle of promoting justice.[16] The AI ethical principle of justice in the context of the world of work can therefore be filled with more concrete labour-specific content by referring to the respective articles in the Declaration and their development through ILS. It is also worth recalling in this respect that the AI ethics justice principle includes for example not only fairness and procedural justice, but also equality

[14] Our analysis of areas of convergence does not aim to establish conceptual equivalence, but rather to explore how principles derived from a common morality (as expressed by Floridi) may be instantiated within a particular morality (as embodied in the ILO's normative framework). Practically, this comparison involves identifying the ethical principles underpinning the Philadelphia Declaration and related ILO instruments and examining how they align with the five AI ethics principles—beneficence, non-maleficence, autonomy, justice, and explicability. This interpretive mapping highlights both shared moral foundations and the distinct institutional purposes that shape their respective expressions.

[15] Floridi's interpretation of this principle focuses strongly on AI enabling or at least being compatible with prosperity (i.e. material well-being), sharing benefits, access to benefits, solidarity and a just distribution.

[16] See Declaration of Philadelphia Sections I(c), II(a), III(a), III(d), and III(j), on dimensions relating to social and distributive justice and social opportunity and which address e.g. the tension between poverty and prosperity, full employment and raising living standards, equal opportunity, and "a just share of the fruits of progress to all". Sections Articles II(a) and III(a) may also be read as relevant for beneficence.

and distributive justice, both forms of justice are also prominently featured in the Declaration of Philadelphia.

• **Beneficence**: This AI principle relates to promoting the well-being of people, not only materially but also spiritually, understanding that human dignity be preserved. The Declaration of Philadelphia mentions e.g. the right of all human beings to pursue their spiritual development in conditions of freedom and dignity.[17] Hence, from an ILO perspective, beneficence in the world of work can be understood as the promotion of both the material well-being and the spiritual development of human beings. Related principles from the FPRW include the elimination of all forms of forced or compulsory labour and the effective abolition of child labour, both of which (while could also be mapped under non-maleficence) address concrete violations of human dignity, freedom, and the right to spiritual development[18]. The beneficence principle in its material sense is also fundamental for the ILO's ILS, as these seek to promote full employment, establish social protection and security measures and raise standards of living[19].

• **Autonomy:** This AI ethical principle relates to the division of decision-making power between humans and machines and the need to ensure meaningful human oversight. This connects to questions of agency, organization, and representation in the ILO system. The ILO framework, while it seeks to safeguard the core of individual autonomy (e.g. through calling for the eradication of forced or compulsory labour), emphasizes a collective dimension of autonomy through the capacity of workers and employers to create and join organizations and to bargain collectively (see, for example, ILS on freedom of association, collective bargaining, and worker representation).[20] In the context of the world of work, autonomy as an AI ethical principle is therefore not only about ensuring that humans retain ultimate decision-making authority over AI systems, but also about determining which humans (e.g. workers or employers) exercise control over different aspects of AI-mediated processes.

• **Non-maleficence**: There exists an equivalent in the ILO framework in the principle to protect the life and health of all workers.[21] A safe and healthy working environment is one of the ILO's FPRW. This ethical principle is also given effect through ILS setting safeguards, rights and procedures to avoid any pernicious effects of work activities – e.g. working time standards, protection against occupational risks, etc. ILO occupational safety and health (OSH) instruments increasingly recognize psychological and psychosocial harm which can be potentially linked to AI-related risks such as pervasive monitoring, intrusive data practices, or forms of algorithmic management. AI ethical principles emphasize the preservation of privacy as a core element of non-maleficence[22], a concern that has been less central in the ILO system, but which could gain more prominence in the context of current technological transformations in the world of work.

• **Explicability:** In the field of AI ethics this principle concerns the ability to understand, explain and hold to account the decision-making of AI systems. It encompasses elements such as

17 See Declaration of Philadelphia Sections II(a), III(b) and III(c) on spiritual development and material well-being dimensions.

18 One might alternatively argue that these two FPRW principles fall under *non-maleficence*, understood as the avoidance of harm. However, the terms *abolition* and *elimination* suggest an active moral commitment to improving the status quo, rather than a mere abstention from causing harm. We also hold that the ethical objections to child and forced labour are grounded more deeply in violations of human dignity, freedom, and the right to spiritual development than in concerns over material well-being. Naturally, we do not claim that our interpretation is the only possible one.

19 See Declaration of Philadelphia Sections II(a), III(a), III(f), and III(i) on material wellbeing, welfare and protection dimensions.

20 See Declaration of Philadelphia Sections I(b) (freedom of expression and association), I(d) (free discussion) and III(e) (collective bargaining and worker participation), as well as the principle of freedom of association and the effective recognition of the right to collective bargaining in the FPRW.

21 See e.g. Declaration of Philadelphia Sections III(g) and III(h).

22 Floridi refers to threats to privacy as one of the most important harms that AI could cause. The ILO has issued a Code of Conduct on Data Privacy (1997), before the emergence of AI in its currently existing form. See Hendrickx (2022).

accountability, trust, and intelligibility. Within the ILO framework, explicability can be linked to the transparency, contestability and accountability required for the application and enforcement of labour norms, including through social dialogue, meaningful tripartite consultation, collective bargaining, labour administration or labour inspection. In this respect, explicability underscores the importance of enforcement and supervisory mechanisms. Moreover, because it calls for understanding the reasons behind AI-supported decisions, explicability is crucial for the effective operation of other ILS, particularly when identifying and addressing issues such as bias or discrimination at work.

A common concern aligns the goals of both, the ILO's normative framework and AI ethics: to safeguard human dignity and freedom; address the risks of abuse and dehumanization that can emerge due to the concentration (or loss) of power (economic, technological, etc.); and mitigate AI's related risks to control or disenfranchise human lives, and threaten their agency.[23]

Areas of divergence

General Differences

AI ethics principles and the ILO's ethical and normative framework are neither interchangeable nor fully comparable frameworks. The two operate at different levels of abstraction. AI ethics principles are highly general, aiming to provide broad guidance on how AI should be designed and used across all areas of society. By contrast, ILO principles and rights are elaborated through a coherent and longstanding legal corpus, in particular through Conventions, Recommendations, and a rich supervisory practice, making them far more context-specific and operationalizable in the world of work. In this sense, AI ethics (in Floridi's formulation) can be understood, in Beauchamp and Childress's terminology, as articulating common morality: mid-level ethical commitments intended to guide conduct across diverse domains. The ILO framework could then be interpreted as corresponding to a particular morality: a domain-specific codification of normative commitments within the institutional and legal setting of labour relations. A second general difference concerns emphasis. Drawing on the preceding analysis, the ILO's normative framework and its ethical foundations — in particular the Declaration of Philadelphia and the Declaration on Fundamental Principles and Rights at Work — if appraised from the standpoint of the five AI ethics principles, could be seen as placing its strongest normative weight on justice and beneficence, with somewhat less explicit development or more limited treatment of AI-specific ethical concerns on non-maleficence, autonomy and explicability (though they do cover labour-specific concerns relating to those principles, as acknowledged in the previous section on convergence). AI ethics distributes emphasis differently: it focuses more on autonomy and non-maleficence (especially privacy-related harms).[24]

Building on these general contrasts, several more specific areas of divergence can be identified.

[23] See for example OECD report on Assessing potential future AI risks, benefits and policy imperatives, November 2024.

[24] This assessment reflects our interpretation of and comparison with the principles articulated in the Declaration of Philadelphia and the FPRW Declaration, rather than the full corpus of ILS. It therefore involves an element of analytical simplification. Furthermore, we are referring to those principles as understood in the AI ethics context, but acknowledge that other understandings may be possible – for example if one were instead to define non-maleficence as precluding any negative outcomes from labor activities then FPRW would be deemed as more prominently embodying such principle.

Specific Differences

Another important area of divergence concerns the ethical principle of autonomy, which is central in the context of AI. By definition, autonomous AI systems assume a degree of decision-making capacity that was previously reserved for humans. Hence, autonomy in the context of AI entails finding and maintaining the appropriate balance between human and machine agency, and the ability to redefine that balance when necessary. In this sense, AI ethics primarily frames autonomy as individual autonomy, the capacity of a person to make choices. In contrast, the ILO ethical framework, while it acknowledges the importance of individual autonomy (distinguishing between admissible forms of subordination under the employment relationship, and unacceptable limitations to individual autonomy, such as forced labour, and recalling that "labour is not a commodity"[25]) it further emphasizes collective autonomy. Within the world of work, collective autonomy is institutionalized through the principles of freedom of association and collective bargaining, which enable workers and employers to collectively pursue their interests through their organizations. These principles also foster the determination of employment conditions and labor norms through autonomous social dialogue rather than unilateral determination. Conversely, autonomy as an AI ethics principle brings to the forefront aspects of individual autonomy affected by AI that hitherto have not received as much attention in the ILO framework.

We note another specific divergence related to non-maleficence and privacy concerns under AI ethics. AI ethics principles, in Floridi's interpretation, highlight harms that are specific to AI as a data-driven technology, most notably through the principle of non-maleficence in relation to privacy. For AI systems to operate effectively, they must collect and process large amounts of data, often revealing information about the "state of the world"[26], including actions, intentions, and decisions that would otherwise remain private. While such data use may reduce traditional information asymmetries and improve decision-making, it can also redistribute informational power by allowing those who collect and control data to observe or infer actions, characteristics and intentions that were previously private. This can significantly alter the balance of economic power by concentrating information in the hands of those who collect and exploit the data using AI. One of the principal harms of AI, therefore, lies in its potential to expose intentions or behaviours previously protected from scrutiny, including in the workplace. This dimension is less developed in the ILO's FPRW framework. Although non-maleficence is present through occupational safety and health (OSH) standards, and these now also address psychosocial risks, the ILO has historically placed less emphasis on privacy as a labour right. Moreover, preserving informational privacy is not only a matter of preventing harm but could also be a prerequisite for maintaining workers' and employers' autonomy, e.g. revealing or controlling private information through AI could weaken bargaining power or prevent collective action. Control or misuse of information could also lead to procedural harm, e.g., opaque decision-making that denies the chance to contest an outcome.

[25] The role and emphasis on individual autonomy in the ILO's normative framework and its ethical foundations would need more detailed analysis than we can undertake in this paper. A contemporary interpretation of the principle that "labour is not a commodity" in the ILO Constitution would be grounded on the recognition that the marketable service of human labour work cannot be separated from the human being. One could argue that this view implies that a certain autonomy of decision-making power has to remain with the human being performing the work (fully acknowledging that labour in an employment relationship also implies giving up autonomy). Often understood as a response to the dehumanizing tendencies of industrial mechanization (although the phrasing originally may have come from the 1914 US Clayton Act to exclude labour unions from antitrust law), this concern may take on renewed relevance as AI, i.e. another type of mechanization, becomes increasingly embedded in work, work processes, and labour relations.

[26] The term 'state of the world' comes from economic decision-making theory. Because economic actors cannot directly observe all relevant conditions, they form expectations, often based on subjective probabilities, about which state the world is in and make decisions accordingly.

A further divergence emerges where some ILO principles and rights have no or only partial ethical equivalence in existing AI ethics frameworks. As pointed out by Samaan et al. (in press), ILO's principles are not explicitly reflected among the five AI ethical principles discussed.[27] Our mapping exercise has shown that most of the ILO's principles and rights can be related to one or more of the five AI ethics principles. Freedom of association can be linked to autonomy as a form of self-determination, although it reflects a collective rather than an individual form of autonomy. For the right of collective bargaining, however, we do not see a clear ethical equivalent in existing AI ethics frameworks. This right embodies collective and institutional dimensions of human agency that are largely absent from the predominantly individualistic orientation of most AI ethics principles. Its omission thus reveals a deeper divergence between AI ethics' focus on individual moral agents and the ILO's recognition of social actors and collective representation in the world of work.

[27] And also not featured in the vast majority of existing AI ethics guidelines.

▶ 5 Positioning AI Ethics in the Context of International Labour Standards

In this section, we explore how AI ethics and the ILO's normative framework and its ethical foundations can support one another in governing the transformation of work. The discussion proceeds in two steps: first, by identifying how the ILO's rights-based and tripartite framework can provide orientation and coherence for emerging approaches to AI governance in labour-related contexts; and second, by examining how AI ethics principles can inform, complement, or prompt reflection on the application of ILS, and where appropriate also enhance and help updating the application of International Labour Standards in light of technological change driven by AI.

A first point of synergy stems from the fact that ILO membership already entails a set of shared ethical and legal commitments. By joining the Organization, Member States are bound by the ILO Constitution, including its Declaration of Philadelphia, and any Conventions they ratify, under the guidance provided by the Organization's supervisory mechanisms. Furthermore, regardless of ratification, as recalled by the Declaration on FPRW, Member States also have the obligation to respect, promote and realize the principles concerning the fundamental rights. This normative framework is to guide how transformations in the world of work, including technological changes, should be approached. As our analysis in the previous section illustrates, many of the values expressed in contemporary AI ethics debates already resonate with ethical principles embedded in the ILO's foundational instruments. For ILO members, therefore, the development and application of AI ethics in labour contexts naturally needs to be understood in light of their pre-existing ILO commitments, which provide a durable reference point for assessing the opportunities and risks associated with new technologies.

In this sense, existing ILO norms offer an important framework for thinking about how AI ethics principles should apply in labour contexts. Because the ILO's foundational commitments have been internationally agreed and are supported by dedicated supervisory mechanisms, they provide a level of clarity, legitimacy, and durability that most contemporary AI ethics frameworks typically do not yet possess. This does not diminish the value of AI ethics initiatives as such; rather, it suggests that their use in the world of work is most coherent when interpreted in a manner that remains consistent and can draw synergies with the established rights and principles already recognized by ILO constituents. From this perspective, ILO standards help define the parameters within which ethical AI in the world of work can be meaningfully discussed and implemented.

The ILO's normative framework can, for example, complement AI ethics frameworks by acknowledging the importance of guaranteeing enabling rights such as freedom of association and collective bargaining, as crucial means in the pursuit of, for instance, the shared principle of justice. The ILO ethical framework can thus contribute to AI ethics through a human centred approach that also encompasses its collective dimension. This is most pertinent at a time where the main priority in the field of AI ethics is to ensure democratic control over its governance.[28] More specifically complementing such mandate, the ILO is endowed with extensive technical expertise in the field of labour, and a convening power to debate with and integrate constituents from the most varied fields of economic activity. The Organization's myriad fora, and varied instruments

[28] See for example Sarah Kreps and Doug Kriner, How AI Threatens Democracy, Journal of Democracy, 24(4), October 2023, pages 122-31.

and application mechanisms, provide a most suitable setting for a rich discussion on the multiple impacts of AI in the field of labour. For example, starting with a sectoral approach, a review could show the impact of AI technologies on the application of FPRW and other ILS by area of economic activity.[29] Based on the ensuing impact findings, the ILO could both provide general guidance, including through its governance and supervisory bodies, and consider standard setting or review measures.[30]

The implications of AI ethics principles for the application and continued development of ILS are multiple. Some stem directly from the areas of divergence identified in the previous section, i.e. cases in which AI ethics principles do not have, or have only partial, equivalents in ILO's normative framework. Most notably, this concerns the AI ethics principle of non-maleficence in relation to privacy. As discussed above, AI ethics views violations of informational privacy, which is also considered as a human right, as a central potential harm of data-driven technologies. Privacy concerns arising from AI therefore represent an area where insights from AI ethics could strengthen the development or interpretation of ILS in the age of AI. A related point concerns autonomy: while collective autonomy is deeply embedded in the ILO framework, AI ethics focuses predominantly on individual autonomy. This difference suggests that evolving forms of digital autonomy and human oversight may warrant further reflection in labour contexts.

More broadly, AI ethics principles draw attention to themes that have not yet been systematically addressed across the ILO normative system including the already mentioned privacy and data rights, algorithmic transparency, explicability, individual aspects of digital autonomy, prediction and surveillance asymmetries, and certain forms of algorithmic management that may generate psychosocial risks not fully captured by traditional OSH approaches. This does not necessarily indicate deficiencies in existing ILS.[31] Rather, it highlights emerging issues that may inform how ILS are interpreted and applied in technologically evolving work environments. In this sense, AI ethics principles can complement, not replace, the obligations of ILO Members by offering additional perspectives that help ensure that longstanding labour rights remain effective under new conditions.

A concrete illustration of such complementarity arises with the principles of explicability and justice. Being able to understand and explain how AI systems make their decisions, as well as preventing unfair biases, is of crucial importance not only to allow access to information relevant to workers and their representatives (for example, to enable meaningful collective bargaining or to prevent and address discrimination) but also to ensure human agency at the workplace, an aim central to the ILO's human-centred approach, and to better monitor the application of ILS.

Explicability entails being able to understand and justify workplace decisions that rely on AI inputs. Justice, as an AI ethics principle, calls for fairness, equality, and the absence of bias in AI systems. Applying these principles in the workplace can help identify systemic issues affecting AI systems used at work, such as algorithmic biases, which could compromise the observance of ILS. For instance, where AI systems contribute to performance monitoring that leads to dismissal, explicability and justice require transparency regarding the processes and rationales used by AI systems to recommend certain decisions, so that these can be assessed for legal validity and

[29] For example, the ILO recently held a tripartite (sectorial) discussion in April 2026 titled:" Technical Meeting on Challenges and Opportunities for Promoting Decent Work, Productivity and a Just Transition Arising from Artificial Intelligence in the Manufacturing Industry."

[30] For this purpose, the ILO has, among other means, the Standards Review Mechanism.

[31] We are not examining this question in this paper. As an example of a normative gap analysis in the field of the related field of the platform economy see document to the ILO Governing Body GB.347/POL/1. The ILO recently adopted the Decent Work in the Platform Economy Convention, 2026 (No. 193)

to ensure that AI is not used as a veneer for practices contrary to fundamental principles and rights at work and other ILS.

To demonstrate the relevance of ethical and legal questions that the deployment of AI systems can entail in the world of work, we refer to actual discussions on the implementation of AI assisted talent recruitment platforms. The following illustration is based on a real world example that eventually resulted in an legal case in an ILO member state still pending at the time of writing: The complaint alleges that a recruiting platform assembled and evaluated applicant information drawn not only from résumés and employer-provided data, but also from third-party online sources and inferred personal characteristics, and then generated proprietary "match scores" and rankings for hiring purposes. Such information can be provided to and used by employers to evaluate and hire or reject candidates. The complainant further alleges that applicants were not given a meaningful opportunity to know what information had been used about them, to review or correct that information, or to contest adverse outcomes, including any discriminatory treatment. The case has not yet been decided, and no view is taken here on the ultimate merits of the parties' claims and defenses. The example shows in legally concrete terms why the AI ethics principles of explicability and justice matter in labour-related contexts for workers, employers and courts: where AI systems influence access to employment, effective protection of workers' rights may require transparency regarding the data sources, inferred features, scoring logic, and grounds for adverse decisions, so that potential bias, discrimination, privacy violations, and other unfair effects can be identified and challenged.

The case is also notable because the challenge is framed through (national) consumer-reporting law, rather than through a settled labour-law framework specifically tailored to AI-assisted hiring. This suggests that, in the current legal landscape, some of the most acute AI-related hiring concerns such as opacity, data aggregation, inferred characteristics, notice, and contestability, are still being litigated through adjacent legal regimes that were not originally designed for AI. From an ILO perspective, a close normative concern applicable to the dispute would be equality of opportunity and treatment in employment. In that sense, while AI ethics identifies the relevant concerns (at a high level of abstraction), the ILO's ethical and normative framework helps specify how those concerns should be understood in the world of work, in particular through its emphasis on dignity, fairness, and non-discrimination in employment and occupation. For example, the Discrimination (Employment and Occupation) Convention, 1958 (No. 111) provides a relevant benchmark to assess whether the system's data, inferred characteristics or learned correlations allegedly used by the recruitment system produce a distinction, exclusion or preference based on race, colour, sex, religion, political opinion, national extraction or social origin, with the effect of nullifying or impairing equality of opportunity or treatment in access to employment, as protected by the Convention. In order to undertake this assessment the application of the principle of explicability becomes paramount.

The principles of explicability and justice thus underscore the importance of transparency concerning the use of AI systems in employment decisions—such as hiring, performance evaluation, supervision, or dismissal and help detect and challenge practices that could curtail the effective application of FPRW and other ILS.[32] Similarly, applying these principles to pervasive (or potentially invasive) workplace monitoring enables the identification and contestation of excessive or unlawful surveillance, and may support the development of necessary safeguards to protect human privacy, freedom, and dignity, including through ILS. This does not imply that AI tools

[32] For a labour and human centred analysis on how technology is never neutral and it must be governed, and emphasizing how AI systems may be increasingly used for managerial decision-making see for instance Aloisi and De Stefano (2022).

should necessarily be approached as potentially biased against ILS, since they may also be designed and applied with a deliberate aim to promote compliance with FPRW and other ILS - for example developing AI recruitment systems that include all pertinent safeguards and other best practices to avoid discrimination or assist humans becoming aware of biases and preventing discrimination – with explicability calling for the necessary insights to assess tools in this regard.

In summary, there is a need for further research on the interplay and synergies between these two ethical frameworks.[33] Moreover, to ensure coherence and draw from the comparative advantage of the ILO tripartite system, the development of AI ethics principles and their application in the world of work could synergize with and defer to ILO instruments and the guidance and convening power of the ILO governance and supervisory bodies.[34]

[33] A further perspective, which extends beyond our core comparative analysis but aligns with the ILO's human centered approach concerns the role of human intelligence as such, and particular in the world of work. The promotion of human intelligence, is largely absent from AI ethics debates. Agustí-Panareda and Agustí-Cullell (2025) argue that human intelligence is largely shaped and exercised in the field of professional education and labour. Technological, AI induced change may affect the development, exercise, and preservation of human cognitive capacities. From an ILO standpoint, rooted in the Philadelphia Declaration's emphasis on both material well-being and "spiritual development", the cultivation of human intelligence through work, learning, and professional development could be seen as essential for safeguarding the continued relevance and value of human capabilities in the workplace. See also Maupain, 2005, Revitalization Not Retreat: The Real Potential of the 1998 ILO Declaration for the Universal Protection of Workers' Rights, The European Journal of International Law Vol. 16 no.3.

[34] As international value projects are often pursued through a multiplicity of organizational frameworks, referral mechanisms are key to ensure synergistic coherence. See for example, as to the field of trade and labour (where trade agreements often refer to FPRW and other ILS), LeClercq, D., Agustí-Panareda, J., & Ebert, F. C. (2021).

▶ Conclusion

Our paper examines how widely used AI ethics principles intersect with the International Labour Organization's ethical and normative framework. Its analysis shows that both frameworks share concerns for human dignity, fairness, and well-being, but differ in scope, orientation and institutional grounding: AI ethics, at least in its dominant principlist formulations, remains relatively abstract and predominantly individual-centred, whereas the ILO corpus provides a legally grounded, collectively oriented framework tailored to the world of work. The comparison of the five AI ethics principles onto ILO's normative framework, focusing on its ethical foundations, in particular its core values and fundamental principles and rights, reveals important areas of convergence, particularly with respect to justice, beneficence, and non-maleficence, but also significant divergences. Most notably, AI ethics gives more prominence to privacy, explicability, and certain forms of individual digital autonomy, while the ILO framework places stronger normative weight on collective autonomy, freedom of association, collective bargaining, and the broader institutional conditions of social justice in labour relations. The absence of a clear equivalent in mainstream AI ethics for some of these collective and institutional dimensions is one of the paper's central findings.

At the same time, the paper has argued that the relationship between the two frameworks is not merely one of contrast. The ILO's rights-based architecture offers a coherent reference point for assessing how AI should be governed in labour contexts. Conversely, AI ethics highlights emerging issues, such as data governance, algorithmic transparency, surveillance symmetries and psychosocial risks, that may warrant renewed attention in the interpretation and application of labour standards. In particular, concerns relating to privacy, explicability, and data-driven opacity in hiring and management show how AI ethics can help specify what the effective application of labour standards may require under new technological conditions, in synergy with the rights-based and supervisory architecture of international labour law.

These findings point to several avenues for further research and policy development. Further work is needed on how AI-related concerns such as privacy, explicability, and digital autonomy should be considered in labour-specific norms and practices, and on how existing international labour standards are to be applied in response to AI-mediated transformations of work. More generally, the analysis suggests that AI ethics in the world of work should be developed not in isolation, but in closer dialogue with the ILO's normative framework, its tripartite governance, and its supervisory institutions. Together, these questions point to the need for sustained interdisciplinary dialogue as AI continues to impact the world of work.

References

Agustí-Panareda, J. (2023). International axiologies for social justice at the International Labour Organization: Value-based perspectives and ways forward. In B. Langille & A. Trebilcock (Eds.), *Social justice and the world of work: Possible global futures: Essays in honour of Francis Maupain* (pp. 55–64). Hart Publishing. https://doi.org/10.5040/9781509961283.ch-005

Agustí-Panareda, J., & Agustí-Cullell, J. (2025). Responding to the challenge of AI: Retrieving human intelligence through labour. *International Labour Review*, *164*(4). https://doi.org/10.16995/ilr.23798

Aloisi, A., & De Stefano, V. (2022). Your boss is an algorithm: Artificial intelligence, platform work and labour. Hart Publishing. https://doi.org/10.5040/9781509953219

Beauchamp, T. L., & Childress, J. F. (2013). *Principles of biomedical ethics* (7th ed.). Oxford University Press.

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Springer International Publishing. https://doi.org/10.1007/978-3-319-60648-4

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law & Security Review*, *35*(4), 410–422. https://doi.org/10.1016/j.clsr.2019.04.007

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, *4*(10), 100857. https://doi.org/10.1016/j.patter.2023.100857

Council of Europe. (2024). Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. https://edoc.coe.int/en/artificial-intelligence/11926-council-of-europe-framework-convention-on-artificial-intelligence-and-human-rights-democracy-and-the-rule-of-law.html

Del Rio Chanona, M., Ernst, E., Merola, R., Samaan, D., & Teutloff, O. (2025). *AI and jobs: A global review of theory, estimates, and evidence* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2509.15265

Deshpande, A., & Sharp, H. (2022). Responsible AI systems: Who are the stakeholders? *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 227–236. https://doi.org/10.1145/3514094.3534187

European Commission High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission. https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf

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

Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. *Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2019.258

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Hendrickx, F. (2022). *Protection of workers' personal data: General principles* (ILO Working Paper 62). International Labour Organization. https://doi.org/10.54394/VBKR9991

IEEE. (2019). *Ethically aligned design: General principles* [PDF]. IEEE Standards Association. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_general_principles_v2.pdf

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

Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., & Akbar, M. A. (2022). Ethics of AI: A systematic literature review of principles and challenges. In Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering (pp. 383–392). Association for Computing Machinery. https://doi.org/10.1145/3530019.3531329

Kreps, S., & Kriner, D. (2023). How AI threatens democracy. *Journal of Democracy*, *34*(4), 122–131. https://doi.org/10.1353/jod.2023.a907693

LeClercq, D., Agustí-Panareda, J., & Ebert, F. C. (2021). ILO labor standards and trade agreements: A case for consistency. *Comparative Labor Law & Policy Journal*, *36*.

Mantelero, A., & Esposito, M. S. (2021). An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. *Computer Law & Security Review*, *41*, 105561. https://doi.org/10.1016/j.clsr.2021.105561

Maupain, F. (2005). Revitalization not retreat: The real potential of the 1998 ILO Declaration for the universal protection of workers' rights. *European Journal of International Law*, *16*(3), 439–465. https://doi.org/10.1093/ejil/chi125

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

Munn, L. (2023). The uselessness of AI ethics. *AI and Ethics*, *3*(3), 869–877. https://doi.org/10.1007/s43681-022-00209-w

OECD. (2019). *OECD principles on artificial intelligence*. Organisation for Economic Co-operation and Development. https://www.oecd.org/en/topics/ai-principles.html

OECD. (2024). Assessing potential future artificial intelligence risks, benefits and policy imperatives (OECD Artificial Intelligence Papers, No. 27). OECD Publishing. https://doi.org/10.1787/3f4e3dfb-en

OHCHR. (2025). Office of the United Nations High Commissioner for Human Rights: Human rights due diligence for digital technology use: Guidance of the Secretary-General. https://www.ohchr.org/en/documents/tools-and-resources/human-rights-due-diligence-digital-technology-use-guidance

Samaan, D. (2025, November 14). Governing AI in the world of work: A review of global ethics guidelines. International Labour Organization. https://www.ilo.org/resource/article/governing-ai-world-work-review-global-ethics-guidelines

Samaan, D., Cabiddu, F., Di Lauro, S., & Tursunbayeva, A. (in press). Governing AI in the world of work: An international review of 245 ethics guidelines. *European Management Journal*. Preprint available at https://doi.org/10.31234/osf.io/wsjuq_v1

Santoro, M. A., & Stone, B. D. (2026). Artificial intelligence, human rights, and the future of work: Global governance and international organizations in the age of AI. *Journal of Human Rights*. https://doi.org/10.1080/14754835.2026.2658703

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000381133/PDF/381133eng.pdf.multi.page=3

United Nations Human Rights. (2025). Steps towards implementing human rights due diligence for digital technology use: International Labour Organization procurement case study. Case study, Human Rights Due Diligence for Digital Technology Guidance of the Secretary-General. https://www.ohchr.org/sites/default/files/2025-09/ohchr-case-study-ilo.pdf

WHO. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. https://iris.who.int/server/api/core/bitstreams/f780d926-4ae3-42ce-a6d6-e898a5562621/content

Yuste, R., Goering, S., Arcas, B. A. Y., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., Friesen, P., Gallant, J., Huggins, J. E., Illes, J., Kellmeyer, P., Klein, E., Marblestone, A., Mitchell, C., Parens, E., Pham, M., Rubel, A., Sadato, N., . . . Wolpaw, J. R. (2017). Four ethical priorities for neurotechnologies and AI. *Nature*, *551*(7679), 159–163. https://doi.org/10.1038/551159a

Acknowledgements

The authors are grateful to Xavier Beaudonnet, Ekkehard Ernst, Caroline Fredrickson, Claire La Hovary, Francis Maupain, Erich Prem, Uma Rani, and Morgan Williams, for their thoughtful comments, feedback, and discussions on earlier drafts and the ideas developed in this paper.

Advancing social justice, promoting decent work

The International Labour Organization is the United Nations agency for the world of work. We bring together governments, employers and workers to improve the working lives of all people, driving a human-centred approach to the future of work through employment creation, rights at work, social protection and social dialogue.

Contact details

Research Department (RESEARCH)

International Labour Organization
Route des Morillons 4
1211 Geneva 22
Switzerland
T +41 22 799 6530
research@ilo.org
www.ilo.org/research

ISBN 9789220439456