

STRUCTURAL LIMITS OF PLATFORM SELF-REGULATION: HUMAN RIGHTS DUE DILIGENCE IN THE SOCIAL MEDIA GOVERNANCE FRAMEWORK

Tamás Pongó*

1. Introduction

This article examines the evolving relationship between platform self-regulation and the European Union's emerging hard-law framework on corporate human rights due diligence. While major social media companies increasingly ground their internal governance architectures in internationally recognized soft-law instruments, most notably the UN Guiding Principles on Business and Human Rights (UNGPs) and the Global Network Initiative (GNI), the legal and practical significance of these frameworks remains contested. In particular, questions persist as to whether platform-designed mechanisms of transparency, risk assessment, and oversight are capable of delivering outcomes that align with the EU's increasingly robust expectations concerning accountability, remediation, and enforceability.

Against this background, the article advances the thesis that while major social media platforms formally embed their human rights commitments within the UNGPs, the OECD Guidelines, and sector-specific initiatives such as the GNI, these self-regulatory frameworks remain structurally constrained by voluntary compliance logics that prioritize procedural alignment over demonstrable human rights outcomes. By analyzing platform-level human rights due diligence practices through a comparative analytical matrix, the paper demonstrates how transparency, policy articulation, and internal governance mechanisms operate primarily as legitimacy-enhancing instruments rather than as effective safeguards against systemic rights-harm. The analysis, therefore, situates platform self-regulation within a broader regulatory ecosystem and argues that the EU's transition from soft-law expectations to hard-law obligations under the Corporate Sustainability Reporting Directive (CSRD) and the

* Senior lecturer (University of Szeged, Faculty of Law and Political Sciences) ORCID: <https://orcid.org/0009-0001-1727-5882>



Corporate Sustainability Due Diligence Directive (CSDDD) constitutes not merely a regulatory supplement, but a corrective response to the structural limitations of platform-driven human rights governance in the digital platform economy.

Methodologically, the article applies an analytical matrix derived from the UNGPs' due diligence cycle and maps it against selected platform practices, drawing on publicly available policies, oversight mechanisms, and case-based evidence. This matrix is then read in light of the CSRD and the CSDDD, allowing for a structured comparison between voluntary governance commitments and legally enforceable standards. By doing so, the article seeks to contribute to ongoing scholarly debates on regulatory hybridity, the limits of private ordering, and the role of EU law in recalibrating power asymmetries between platforms, users, and rights-holders.

2. The Business and Human Rights framework – UNGPs and GNI

The emergence of the Business and Human Rights (BHR) framework within the United Nations (UN) was not the result of a single regulatory act, but rather the outcome of a gradual institutional and conceptual development. A decisive step in this process was the creation of a dedicated mandate to examine the human rights implications of business activities, which led to the appointment of John Ruggie as the United Nations Secretary-General's Special Representative on business and human rights in 2005.¹ This mandate marked a shift away from earlier, fragmented approaches and laid the groundwork for a more coherent normative framework. The first substantive articulation of this framework occurred in 2008, when the United Nations endorsed the "Protect, Respect and Remedy" Framework. This tripartite model was subsequently operationalized through the UNGPs, elaborated under Ruggie's leadership. Although the UNGPs do not constitute binding international law, they have acquired considerable normative authority and now function as a global point of reference for assessing corporate responsibility in relation to human rights.²

¹ Florian Wettstein: CSR and the Debate on Business and Human Rights: Bridging the Great Divide. *Business Ethics Quarterly*, Vol. 22., N. 4. (2012) 739. <https://doi.org/10.5840/beq201222446>

² John Gerard Ruggie: Global Governance and "New Governance Theory": Lessons from Business and Human Rights. *Global Governance*, Vol. 20., N. 1. (2014) 11. <https://doi.org/10.1163/19426720-02001002>; Karin Buhmann – Roxana Olivera: Human rights and social media platforms: the corporate responsibility to respect human rights in regard to privacy infringement involving photo posting. *Australian Journal of Human Rights*, Vol. 26., N. 1. (2020) 127. <https://doi.org/10.1080/1323238X.2020.1802559>; Wettstein op. cit. 741.; Kun, Attila: Kellő gondosság alapú belső átvilágítás a munkajogban, mint a transznacionális vállalatok magatartásának új zsinórmértéke. *Miskolci Jogi Szemle*, 2017/2. 288.; John Gerard Ruggie: Business and Human Rights: The Evolving International Agenda. *American Journal of International Law*, Vol. 101., N. 4. (2007) 819–840. <https://doi.org/10.1017/S0002930000037738>; Chronowski, Nóra: Az alkotmány/bíraskodás potenciálja az üzleti szféra okozta emberi jog sérelmekkel szemben. *Alkotmánybírói Szemle*, 2023/1. 4.; András L. Pap – Nóra Chronowski – Zoltán Nemessányi: Corporate Human Rights Responsibility in Illiberal Regimes: The Example of the Ukrainian Refugee Crisis in Hungary. *Business and Human Rights Journal*, Vol. 9. (2024) 50. (<https://doi.org/10.1017/bhj.2024.5>) cites that France, Germany and Norway adopted laws and Switzerland put it on the agenda, unsuccessfully.

The UNGPs are structured around three interrelated pillars: “the state duty to protect against business-related human rights infringements, the corporate responsibility to respect human rights, and increased access to remedy for victims of business-related human rights infringements.”³ In the context of the information and communication technology sector, and social media platforms in particular, the latter two pillars are of central relevance. These dimensions of the UNGPs shift the analytical focus from state action to corporate conduct, emphasizing both preventive responsibilities and the availability of mechanisms capable of addressing rights infringements resulting from private governance practices. At the core of the second pillar lies the concept of human rights due diligence (HRDD), which requires business enterprises to identify, prevent, mitigate, and account for adverse human rights impacts associated with their operations.⁴ The scope of HRDD extends beyond direct corporate conduct and encompasses situations in which a company contributes to harm or is directly linked to adverse impacts through its business relationships, as reflected in Principle 17(a) of the UNGPs. Where such impacts arise, enterprises are expected to rely on their leverage to prevent or address the resulting harm. Leverage means “the influence a business enterprise has with an entity or entities involved in a human rights abuse...”⁵ At the same time, the notion of human rights due diligence remains conceptually open-ended. It is often used as an umbrella term that covers both human rights risk assessment and human rights impact assessment, despite the fact that these approaches serve distinct functions. While risk assessment primarily aims at identifying and managing potential negative consequences for the company itself, impact assessment entails a broader, rights-centered evaluation focused on preventing harm and integrating human rights considerations into strategic decision-making.⁶ This conceptual ambiguity has given rise to scholarly debate, with some authors arguing that the UNGPs implicitly prioritize risk management over a more robust, impact-oriented understanding of due diligence.⁷

The normative influence of the UNGPs has extended beyond the United Nations framework and has significantly shaped subsequent initiatives at both the international and regional levels. In particular, the human rights due diligence approach articulated in the UNGPs has been taken up and further elaborated by the Organisation for Economic Co-operation and Development (OECD) as well as by the European Union (EU). Within the OECD system, this influence is most clearly reflected in the updated 2023 edition of the OECD Guidelines for Multinational Enterprises on Responsible Business Conduct. In the General Policies section of the OECD Guidelines, enterprises are expressly expected to “respect the internationally recognized human rights of those

³ Buhmann–Olivera op. cit. 125.

⁴ Pap–Chronowski–Nemessányi op. cit. 45.

⁵ Buhmann–Olivera op. cit. 129.

⁶ Kendyl Salcito – Mark Wielga: What does Human Rights Due Diligence for Business Relationships Really Look Like on the Ground? *Business and Human Rights Journal*, Vol. 3. (2018) 114–115. <https://doi.org/10.1017/bhj.2017.31>

⁷ Salcito–Wielga op. cit. 115.

affected by their activities.”⁸ This obligation is operationalized through a risk-based due diligence approach, outlined in Sections II.A.11–13, which requires companies to identify, prevent, and mitigate adverse impacts arising from their operations. The scope of this approach closely mirrors the standards established by the UNGPs, as it encompasses harms that are caused by the enterprise itself, those to which it contributes, as well as impacts that are directly linked to its operations, products, or services through business relationships. The Commentary to the Guidelines further clarifies that due diligence is conceived as an ongoing process integrated into business decision-making and risk management systems. It is framed as a mechanism through which enterprises are expected not only to identify and address actual and potential adverse impacts, but also to account for the manner in which such impacts are managed.⁹ In this respect, the OECD Guidelines reinforce the second and third pillars of the UNGPs framework and seek to promote a coherent global standard for the implementation of human rights due diligence across different business sectors.

At the same time, the Guidelines acknowledge that due diligence obligations are inherently context-dependent. The nature and intensity of the required measures are influenced by factors such as the size of the enterprise, the severity and likelihood of the adverse impact, and the degree of the company’s involvement in the relevant activity.¹⁰ These considerations are reiterated and further developed in Chapter IV on Human Rights, which was itself a result of Ruggie’s mandate and the adoption of the UN Framework in 2008.¹¹ Beyond the Guidelines themselves, the OECD has also issued the Due Diligence Guidance for Responsible Business Conduct, a complementary instrument intended to assist enterprises in translating the UNGPs into operational practice.¹² The development of this Guidance was overseen by the OECD Working Party on Responsible Business Conduct and involved a broad multi-stakeholder process, bringing together OECD and non-OECD states alongside representatives of business, trade unions, and civil society.¹³ While the OECD has adopted a range of sector-specific due diligence instruments – for instance, in relation to minerals, agriculture, or the garment and footwear industries – no comparable guidance currently exists for social media companies. This regulatory gap is particularly striking given the human rights risks associated with platform-based business models and foreshadows some of the structural challenges addressed later in this paper.

Against this background, the analysis must turn to sector-specific instruments capable of translating general business and human rights standards into operational guidance for technology companies. Given that the present research focuses on social media companies, particular attention is warranted with regard to the companies’ corporate human rights policies (if they have one), which explicitly refer not only to

⁸ OECD Guidelines for Multinational Enterprises, 2008, II. General Policies A.2.

⁹ Ibid. Commentary 15.

¹⁰ Ibid. Commentary 19.

¹¹ Ruggie (2014) op. cit. 11.

¹² OECD Due Diligence Guidance for Responsible Business Conduct, 2018, Foreword.

¹³ Ibid.

the UNGPs but also to the GNI. This reference signals an attempt to embed platform governance within a more specialized normative framework tailored to the human rights risks characteristic of the digital environment.

The GNI was developed together with the UN framework since the GNI founders launched the cooperation in 2008.¹⁴ The GNI constitutes a multi-stakeholder framework with a heterogeneous membership, bringing together technology companies (Yahoo!, Microsoft, Google), civil society organizations (Human Rights Watch, Human Rights First, Amnesty International), investors (Calvert, Boston Common Asset Management), and academic actors (Berkman Center at Harvard). “As Yahoo!, Microsoft, and Google discovered, developing agreed-upon principles and guidelines to direct behavior, co-drafted by a range of stakeholders, is a practical step toward maintaining the public’s trust while preserving organizational business models and protecting human rights.”¹⁵ At the core of this framework are the GNI’s General Principles on Freedom of Expression and Privacy. As stated in the preamble, these Principles are grounded in the main international human rights instruments and are explicitly informed by the UN’s “Protect, Respect and Remedy” Framework, the UNGPs, as well as the OECD Guidelines for Multinational Enterprises.¹⁶ In this sense, the GNI can be understood as a sector-specific articulation of the broader BHR regime, which initially developed within the UN, subsequently reinforced through OECD instruments, and further operationalized for the technology sector with a particular focus on freedom of expression and privacy.

Beyond the adoption of its General Principles, the GNI has also developed Implementation Guidelines intended to support member companies in translating normative commitments into internal processes and practices.¹⁷ These Guidelines explicitly encourage companies to conduct human rights due diligence in line with the requirements of the UNGPs¹⁸, and provide illustrative examples of situations in which a human rights impact assessment should form part of that process.¹⁹ The emphasis on implementation reflects an effort to move beyond abstract commitments and to address the practical challenges of integrating human rights considerations into platform governance structures.

At the same time, it should be recalled that the influence of the UNGPs has not remained confined to voluntary or sector-specific initiatives. As noted above, their normative logic has also informed developments at the regional level, most notably

¹⁴ Dunstan Hope: Global Network Initiative: An Ethical Compass for International Information and Communications Firms in the Internet Age. *Global Business and Organizational Excellence*, Vol. 30., N. 3. (2011) 9. <https://doi.org/10.1002/joe.20373>

¹⁵ Hope op. cit. 7.

¹⁶ Global Network Initiative Principles. <https://globalnetworkinitiative.org/gni-principles/>

¹⁷ Global Network Initiative Implementation Guidelines. <https://globalnetworkinitiative.org/implementation-guidelines/>

¹⁸ Ibid. 2.4.

¹⁹ Ibid. 2.6.

within the EU. Through the adoption of two directives²⁰ aimed at strengthening corporate responsibility for human rights, the EU has sought to progressively transform human rights due diligence from a predominantly soft-law expectation into a set of binding legal obligations. This emerging shift from voluntary standards to enforceable rules provides an important contextual backdrop for assessing the role and limits of self-regulatory frameworks such as the GNI in the governance of social media platforms.

Following an extended legislative process lasting several years, the EU adopted the CSRD as a central element of its sustainability and corporate governance agenda.²¹ The Directive was formally signed on 14 December 2022 and entered into force on 5 January 2023. Its regulatory architecture reflects a deliberate alignment with the normative logic of the UNGPs, most notably through three interrelated design choices that integrate human rights due diligence into corporate disclosure and governance structures. First, the CSRD is built around the concept of double materiality and an expanded value-chain perspective, requiring companies to report not only on financial risks and opportunities, but also on the actual and potential impacts of their activities on human rights and the environment.²² Second, the Directive relies on the European Sustainability Reporting Standards, within which standards S1–S4 impose specific disclosure obligations relating to the identification, management, and mitigation of human rights impacts.²³ These requirements closely mirror the due diligence process set out in Principle 17 of the UNGPs. Third, the CSRD introduces governance-related safeguards, including mandatory external assurance and an explicit responsibility for company directors to oversee the sustainability statement, thereby strengthening internal accountability mechanisms.²⁴

A further cornerstone of the EU's approach to business and human rights is the CSDDD, which may be understood as the hard-law counterpart to the UNGPs framework within the European legal order.²⁵ Adopted on 13 June 2024 and entering into force on 25 July 2024, the CSDDD codifies, in Articles 5–11, the core elements of the human rights due diligence cycle. It obliges companies to identify, prevent, mitigate, and, where necessary, bring to an end adverse human rights and environmental impacts arising from their own operations, those of their subsidiaries,

²⁰ Directive of the European Parliament and of the Council of 14 December 2022 Amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting. OJ L 322, 16.12.2022, pp. 15–80. (Hereinafter: Directive 2022/2464.); Directive of the European Parliament and of the Council of 13 June 2024 On corporate sustainability due diligence and amending Directive (EU) 2019/1937 and Regulation (EU) 2023/2859. OJ L, 2024/1760, 5.7.2024.

²¹ Katrin Hummel – Dominik Jobst: An Overview of Corporate Sustainability Reporting Legislation in the European Union. *Accounting in Europe*, Vol. 21., N. 3. (2024) 320–355. <https://doi.org/10.1080/17449480.2024.2312145>

²² Recital 27, and Article 19a of Directive 2022/2464.

²³ Preamble, and Chapter 6a of Directive 2022/2464.

²⁴ Recital 62, Article 19a, and Article 34 of Directive 2022/2464.

²⁵ Nicolas Bueno – Nadia Bernaz – Gabrielle Holly – Olga Martin-Ortega: The EU Directive on Corporate Sustainability Due Diligence (CSDDD): The Final Political Compromise. *Business and Human Rights Journal*, Vol. 9., N. 2. (2024) 296. <https://doi.org/10.1017/bhj.2024.10>

and their established business relationships. In contrast to the aspirational language of the UNGPs, the CSDDD employs mandatory formulations, thereby transforming due diligence expectations into legally enforceable obligations.

Taken together, the UNGPs and the OECD Guidelines articulate normative expectations, whereas the CSRD and the CSDDD embed those expectations within the framework of EU company law. HRDD thus evolves from a predominantly reputational or best-practice commitment into a compliance-driven requirement, supported by audit obligations, supervisory oversight at the national level, and, under the CSDDD, potential exposure to civil liability. The two directives operate through complementary regulatory logics: the CSRD exerts a transparency-driven pull by requiring public disclosure of due diligence policies, processes, and outcomes, while the CSDDD creates a substantive push by mandating the performance of due diligence to an adequate standard. In this combined form, the directives give concrete effect to the second and third pillars of the UNGPs within the EU internal market.

The UNGPs and the OECD Guidelines operate as soft-law instruments, so they formulate expectations regarding corporate conduct, while leaving compliance largely dependent on reputational considerations, market responses, and peer pressure. Their normative force thus derives less from legal enforceability than from their capacity to shape corporate behavior through external scrutiny and internalization. The EU, by contrast, has chosen to translate these same expectations into binding legal obligations through the adoption of the CSRD and the CSDDD. In this regard, the CSDDD is of particular significance, as it introduces not only procedural requirements but also a substantive duty of conduct with respect to human rights due diligence. The regulatory logic of the two directives is complementary. The CSRD operates primarily through transparency, exposing shortcomings in corporate practice to public and stakeholder review, whereas the CSDDD relies on coercive mechanisms by requiring companies to perform due diligence to a legally defined standard, supported by sanctions and the possibility of civil liability. Taken together, the two directives reconfigure the relationship between voluntary standards and legal obligation. They transform the aspirational commitments articulated in the UNGPs and the OECD Guidelines into an enforceable European framework that enhances accountability and strengthens access to remedy. At the same time, this framework explicitly situates corporate human rights management within the broader objectives of sustainability and climate transition, thereby linking business and human rights considerations to the EU's evolving model of corporate governance. Against this normative background, the following section turns to the position of social media platforms within the international human rights framework, with the aim of clarifying their institutional role and the implications this role entails for corporate responsibility and accountability.

3. Methodology and Analytical Framework

This article employs a qualitative, doctrinal–normative research methodology, combining structured document analysis with comparative legal assessment. Its objective is not to measure platform behavior empirically, but to evaluate the extent to which the self-regulatory human-rights frameworks of major social media platforms

are normatively compatible with the expectations articulated in international soft-law instruments and emerging EU hard-law obligations. The analysis is grounded in a systematic examination of publicly available corporate human rights policies, transparency reports, and related governance documents issued by selected social media platforms. These materials are assessed against the normative benchmarks established by the UNGPs, the GNI Principles and Implementation Guidelines, and the EU's CSR and CSDDD. This approach reflects the central role that policy articulation, disclosure, and procedural commitments play within contemporary platform governance and self-regulatory models.

The platforms included in the analysis were selected on the basis of their market significance, the top three social media companies with the highest number of monthly active users. Particular attention is paid to platforms that publicly reference the UNGPs and/or the GNI in their corporate human rights policies, as these references indicate an express commitment to the business and human rights regime and thus render the platforms suitable subjects for normative compatibility assessment.

To enable a structured and comparable evaluation, the article applies a dimensional analytical matrix. The matrix translates the abstract requirements of the UNGPs, GNI Principles, and EU due diligence legislation into a set of analytically distinct, yet interrelated dimensions. These dimensions include, *inter alia*, the explicit recognition of international human rights standards, the scope and depth of HRDD processes, stakeholder engagement mechanisms, governance and accountability structures, transparency and reporting practices. Each dimension reflects a core element of the UNGPs' second and third pillars and is further informed by the procedural and substantive expectations codified in the CSRD and the CSDDD. The matrix thus functions as a normative bridge between soft-law standards and hard-law obligations, allowing for a systematic assessment of whether, and to what extent, platform self-regulation anticipates, aligns with, or falls short of binding EU requirements. The matrix does not seek to rank platforms or to quantify compliance. Rather, it serves as an analytical tool to identify structural patterns, normative gaps, and recurring tensions within platform self-regulatory frameworks. The article acknowledges the inherent limitations of policy-based analysis. Corporate human rights policies and governance documents do not necessarily reflect operational practice, nor do they provide insight into internal enforcement, resource allocation, or decision-making dynamics. Nevertheless, given that self-regulation in the platform context is predominantly articulated through publicly accessible policies and procedural commitments, document-based normative analysis remains a suitable and widely accepted method for evaluating the structure and orientation of platform human rights governance.

4. Application of the comparative analytical matrix to social media platforms

The research focuses on the three most dominant market players based on monthly active users. As a result, Meta, Google, and TikTok will be analyzed through the lens of the comparative analytical matrix. According to the data, three Meta-owned platforms

have the highest number of monthly active users, i.e., Facebook, Instagram, and WhatsApp, while YouTube (owned by Google) is in fourth place, followed by TikTok.²⁶

4.1. Meta

With more than three billion monthly active users on Facebook and around three billion on Instagram, Meta occupies a structurally central position in the global social media ecosystem.²⁷ Given this scale, the manner in which Meta conceptualizes and operationalizes its human rights responsibilities carries particular normative significance within the BHR framework.

Meta's Corporate Human Rights Policy (HRP) explicitly situates the company within the BHR regime. The policy affirms Meta's commitment to the UNGPs, a range of core UN human rights instruments, as well as the GNI Principles and Implementation Guidelines.²⁸ Building on these normative references, the HRP identifies five key operational areas, namely (i) human rights policies; (ii) human rights due diligence; (iii) access to remedy; (iv) oversight, governance, and accountability; and (v) the protection of human rights defenders. These areas correspond closely to the core expectations articulated under the second and third pillars of the UNGPs. From an HRDD perspective, Meta's framework appears, at least at the level of policy design, to align with UNGPs standards. According to the HRP, Meta "*identify and prioritize the most salient human rights issues in each context using the UNGP framework of likelihood and severity.*"²⁹ The company further indicates that it develops strategies to avoid, prevent, and mitigate adverse human rights impacts and conducts human rights impact assessments as part of its due diligence processes.³⁰ Taken together, these commitments reflect a procedural understanding of HRDD consistent with Pillar Two of the UNGPs. Meta's framework also incorporates a range of grievance and redress mechanisms. These include internal whistleblowing channels, complaint-handling procedures, and avenues to appeal content moderation decisions.³¹ Most notably, Meta has established the Oversight Board as an independent body empowered to review selected content moderation cases and issue policy advisory opinions. In the design of this mechanism, Meta explicitly relied on the UNGPs effectiveness criteria for non-judicial grievance mechanisms, thereby seeking to reinforce access to remedy and accountability within its self-regulatory architecture.

However, when assessed through a more critical analytical lens, Meta's human rights governance model reveals important structural limitations. While the HRP articulates a comprehensive procedural framework, existing research has highlighted discrepancies

²⁶ Thuy Duong: Most popular social networks worldwide as of February 2025, by number of monthly active users. *Statista*, Mar 11, 2026. <https://tinyurl.com/yky943ak>

²⁷ *Ibid.*

²⁸ Meta Corporate Human Rights Policy Commitments.

²⁹ HRP 02 Conduction Human Rights Due Diligence and Disclosure.

³⁰ *Ibid.*

³¹ HRP 03 Providing Remedies for Human Rights Impact.

between policy commitments and their implementation in practice.³² For instance, the Oversight Board, which provides users with accessible remedies, does not serve as a forum for redress, as it selects to decide only 0.00235% of complaints submitted by users, compared to Meta's 14,95% acceptance rate.³³ Moreover, recent developments in Meta's content moderation strategy further illustrate the fragility of policy-based human rights protections. In January 2025, Meta announced the termination of third-party fact-checking in favor of a Community Notes model relying on user reporting.³⁴ This shift represents a significant recalibration of enforcement responsibilities and raises concerns regarding the consistency and effectiveness of human rights risk mitigation. These concerns are partially reflected in Meta's own transparency data. According to the Q1 2025 report, the prevalence of bullying and harassment content on Facebook increased from 0.06–0.07 per cent to 0.07–0.08 per cent, while violent and graphic content rose to approximately 0.09 per cent.³⁵ Meta attributes these changes to increased sharing of violating content and ongoing efforts to reduce enforcement errors, yet the data nonetheless suggests limits to the preventive and remedial capacity of the existing governance framework.

Overall, Meta's human rights policy architecture demonstrates a high degree of formal alignment with the UNGPs and the sector-specific standards developed under the GNI. At the same time, the emphasis placed on policy articulation, transparency, and procedural design contrasts with persistent challenges in impact prevention, remediation, and accountability. This tension illustrates the broader structural mismatch between platform self-regulation and the increasingly substantive human rights due diligence expectations emerging under EU hard-law instruments.

4.2. Google

As YouTube operates as part of the Google corporate structure, Google's human rights commitments provide the relevant normative and institutional framework for assessing platform governance. At the level of formal policy articulation, Google explicitly

³² Working Group on the issue of human rights and transnational corporations and other business enterprises et al.: Letter to Mark Zuckerberg, Switzerland, 2024, The UN Working Group on business & human rights requests Meta address human rights concerns regarding its operations in OPT. *Business and Human Rights Centre*, 18 April 2024 – 29 Aug 2024. <https://tinyurl.com/yjcb8kzw>; or Oversight Board proved is not an effective grievance mechanism Tamás Pongó – Renáta Kálmán – Réka B. Szaniszló – Norbert Tribl: Oversight Board – An effective tool to protect free speech? *Masaryk University Journal of Law and Technology*, 2025., Vol. 19., N. 2. (2025) 151–178. <https://doi.org/10.5817/MUJLT2025-2-3> 151–178.; Criticism concerning the Oversight Board interpretation methodology, see Andreas Kulick: Meta's Oversight Board and Beyond – Corporations as Interpreters and Adjudicators of International Human Rights Norms. *The Law and Practice of International Courts and Tribunals*, Vol. 22. Issue 1. (2022) 161–193. <https://doi.org/10.1163/15718034-12341496>

³³ Pongó–Kálmán–Szaniszló–Tribl op. cit. 171.

³⁴ Joel Kaplan: More Speech and Fewer Mistakes. *Meta*, January 7, 2025. <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>

³⁵ Integrity Reports, First Quarter 2025. *Meta*, May 29, 2025. <https://transparency.meta.com/reports/integrity-reports-q1-2025/>

anchors its human rights responsibilities within the international BHR regime. The company publicly affirms its commitment to the Universal Declaration of Human Rights, the UNGPs, and the GNI, of which it is a founding member.³⁶ Beyond these declaratory commitments, Google has established dedicated governance structures to operationalize its human rights responsibilities. The Google Human Rights Program (GHRP) and the Human Rights Executive Council (HREC) are tasked with overseeing the integration of human rights considerations across the company's operations. According to publicly available information, the GHRP applies HRDD processes, including human rights impact assessments, and cooperates "*with internal stakeholders in areas such as data governance, content policies, and supply chain.*"³⁷ This internal coordination is intended to support the identification, prioritization, and management of actual and potential human rights impacts throughout the product lifecycle.

Transparency and external accountability constitute further elements of Google's governance framework. Through its Transparency Center, the company regularly publishes detailed transparency reports, enabling users, regulators, and policymakers to gain insight into its content moderation practices, data requests, and enforcement decisions. In addition, Google has been subject to repeated independent assessments conducted by the GNI. The most recent publicly available GNI assessment (2021/2022)³⁸ positively evaluates the institutional design and functioning of the GHRP and the HREC, highlighting, *inter alia*, the provision of targeted human rights training for employees and the embedding of due diligence considerations into internal processes.³⁹ Google's due diligence approach extends beyond training and policy formulation. Human rights considerations are reportedly integrated into the development of new products and features, as well as into acquisition processes, where contractual terms aligned with freedom of expression and privacy principles are applied. "Where this ongoing HRDD and review of potential human rights issues may surface products or jurisdictions with higher risk to human rights, Google may use a formal Human Rights Impact Assessment."⁴⁰ Risk mitigation is thus presented as a continuous, multi-layered process encompassing identification, mitigation planning, and ongoing review.

At first glance, this governance architecture positions Google as a paradigmatic example of corporate alignment with the UNGPs and sector-specific standards under the GNI. However, this assessment warrants careful contextualization. The positive evaluations cited above largely derive from a multistakeholder initiative in which

³⁶ Google: Human Rights. <https://about.google/company-info/human-rights/>

³⁷ Ibid.

³⁸ "Given the growing number of GNI member companies, as well as changes to legal and regulatory frameworks in several key jurisdictions, GNI's Board has decided that the next (fifth) assessment cycle will run until mid-2026, with independent assessments of GNI companies taking place on a rolling basis beginning in the second half of 2024 and running through the end of 2025." Global Network Initiative: Revisions to GNI's Fifth Cycle of Independent Company Assessments. *GNI News*, November 16, 2023. <https://tinyurl.com/3pe9pwdd>

³⁹ Ibid. 28.

⁴⁰ Ibid.

Google itself plays a constitutive role, raising questions about the structural limits of peer-based accountability mechanisms.

These concerns are further reinforced by recent policy developments. In early February 2026, media reports revealed that Alphabet, Google's parent company, had revised its internal AI guidelines, effectively lifting a prior prohibition on the use of artificial intelligence for weapons and surveillance purposes.⁴¹ Although the company justified this shift by reference to national security considerations⁴², the decision signals a significant recalibration of previously articulated human rights constraints. When viewed alongside earlier policy reversals by other major platforms, such as Meta's abandonment of third-party fact-checking, this development underscores a broader structural vulnerability of soft-law-based commitments. While platform governance frameworks may align with international human rights standards at the level of policy design, their resilience appears contingent upon political and economic context, thereby exposing the limits of self-regulation in the absence of binding legal obligations. In this sense, Google does not merely exemplify an individual instance of limited corporate enforcement but rather exposes a structural limitation inherent in voluntary platform governance within the BHR framework. While the UNGPs encourage corporate responsibility through internal policies, reporting mechanisms, and stakeholder engagement, the absence of externally enforceable obligations leaves meaningful accountability largely dependent on corporate discretion. This governance gap illustrates why self-regulatory models, even when normatively sophisticated, remain institutionally fragile – an imbalance that recent EU initiatives, most notably the CSDDD and the broader CSR regulatory turn, explicitly seek to correct by translating due diligence expectations into legally binding duties accompanied by public oversight and sanctioning mechanisms.

4.3. TikTok

TikTok constitutes the final platform examined in this article. Following the transformation of musical.ly into TikTok in 2018, the Chinese platform rapidly reshaped the social media ecosystem through its short-form, highly engaging video format. Today, TikTok ranks among the three largest social media companies worldwide in terms of monthly active users, which renders its approach to human rights responsibilities particularly salient.⁴³ Accordingly, this section assesses TikTok's HRDD framework on the basis of its formal commitments and recent regulatory and policy developments.

⁴¹ Global: Google's shameful decision to reverse its ban on AI for weapons and surveillance is a blow for human rights. *Amnesty International News*, 6 February 2025. <https://tinyurl.com/3rx7tcz6>; Lucy Hooker – Chris Wallace: Concern over Google ending ban on AI weapons. *BBC*, 5 February 2025. <https://www.bbc.com/news/articles/cy08lnqx2zjo>

⁴² James Manyika – Demis Hassabis: Responsible AI: Our 2024 report and ongoing work. *News from Google*, Feb 04, 2025. <https://blog.google/innovation-and-ai/products/responsible-ai-2024-report-ongoing-work/>

⁴³ Duong op. cit.

TikTok's Transparency Center includes a dedicated section on *Upholding human rights*, in which the company explicitly anchors its operations to international human rights standards. The declaration refers to key international instruments, including the Universal Declaration of Human Rights, the Convention on the Rights of the Child, and the UNGPs.⁴⁴ At a formal level, these references position TikTok squarely within the global BHR regime. The company commits itself to identifying, assessing, and addressing salient human rights risks, complying with domestic legal frameworks in the jurisdictions where it operates, and safeguarding users' freedom of expression and privacy. Human rights considerations are presented as an integral lens in policy development, product design, and risk mitigation, supported by employee and business-partner training as well as stakeholder engagement with external experts and human rights organizations.

In line with the UNGPs, TikTok has commissioned external human rights impact assessments and incorporated international human rights standards – most notably the ICCPR three-part test – into its content governance framework.⁴⁵ These measures suggest that HRDD is, at least structurally, operationalized within the company's internal processes. At the same time, the governance and oversight architecture of this framework remains opaque. Publicly available information merely indicates that “key leaders” are responsible for oversight, without providing further institutional detail. This lack of transparency raises concerns regarding accountability and internal control, which are core components of effective HRDD.

The gap between formal commitment and practical implementation becomes even more pronounced when TikTok's recent regulatory challenges are examined. In June 2025, a briefing published by the European Parliament identified several systemic risks and instances of potential non-compliance with EU law. Among the most persistent concerns is the possibility of data transfers to China, which could jeopardize users' data protection and privacy. As the briefing states, “[a] central worldwide concern is whether TikTok transfers data to China, potentially compromising user data protection and privacy.”⁴⁶ These concerns have escalated beyond data protection into the realm of national security, leading EU institutions (the European Parliament, the European Commission, and the Council of the European Union) to ban TikTok from corporate devices as early as 2023.⁴⁷ Additional risks have been identified in relation to the protection of minors, particularly the misuse of TikTok Live “to lure minors into performing sexual acts, rewarding them with virtual gifts that can be exchanged for 5 EPRS | European Parliamentary Research Service money (known as ‘giftbaiting’).”⁴⁸

⁴⁴ TikTok Transparency Center: Upholding human rights. <https://tinyurl.com/5awzspuw>

⁴⁵ Ibid.

⁴⁶ European Parliament: TikTok and EU regulation: Legal challenges and cross-jurisdictional insights. *Briefing*, 26-06-2025. 3. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2025\)775837](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2025)775837)

⁴⁷ Ibid. 4.

⁴⁸ Ibid. 5–6.

Most recently, in early February, the European Commission issued preliminary findings concluding that TikTok's addictive design features breach the Digital Services Act.⁴⁹ According to the Commission, the platform systematically encourages prolonged use, fails to adequately address indicators of compulsive behavior, especially among minors, and applies ineffective risk mitigation tools, including screen-time management and parental control mechanisms.⁵⁰ On this basis, "the Commission considers that TikTok needs to change the basic design of its service."⁵¹ While these findings remain preliminary and TikTok retains its right to defense, they are the outcome of a formal proceeding initiated in February 2024 and are grounded in extensive evidentiary material, including internal company data, assessment reports, TikTok's responses, and scientific research.

Taken together, TikTok's case closely mirrors that of Google. At the level of corporate policy and formal commitment, the company appears to embrace the UNGPs and the broader BHR framework in a structured and comprehensive manner. Yet regulatory practice and enforcement proceedings reveal persistent deficiencies in implementation. This reinforces the conclusion already drawn in relation to Google that self-regulatory human rights commitments, when detached from binding legal enforcement, remain inherently fragile. The TikTok case thus further substantiates the argument that effective corporate HRDD cannot be sustained through soft-law instruments alone, but requires binding regulatory frameworks, such as the CSRD and the CSDDD, that translate voluntary human rights commitments into enforceable duties of conduct, institutional oversight, and legal accountability.

5. Conclusions

The paper proceeded from the thesis that while major social media platforms formally embed their human rights commitments within the business and human rights regime, in particular the UNGPs, the OECD Guidelines, and sector-specific initiatives such as the GNI, these self-regulatory frameworks remain structurally constrained by voluntary compliance logics that prioritize procedural alignment over demonstrable human rights outcomes. By applying a comparative analytical matrix to platform-level HRDD, the article has shown that policy-making, training programs, transparency reports, and internal governance mechanisms frequently operate as legitimacy-

⁴⁹ Commission preliminarily finds TikTok's addictive design in breach of the Digital Services Act. *Press release*, Feb. 6, 2026. https://ec.europa.eu/commission/presscorner/detail/en/ip_26_312; Laura Cress: TikTok told to change 'addictive design' by EU or face massive fines. *BBC*, 6 February 2026. <https://www.bbc.com/news/articles/cr7j7n3151mo>; TikTok accused of breaching EU digital safety rules. *DigWatch*, 6 Feb 2026. <https://dig.watch/updates/tiktok-accused-of-breaching-eu-digital-safety-rules>; EU: European Commission must urgently enforce landmark law to stop TikTok's addictive harms. *Amnesty International*, 6 February 2026. <https://www.amnesty.org/en/latest/news/2026/02/eu-commission-tiktok-investigation/>

⁵⁰ European Commission (2026) op. cit.

⁵¹ Ibid.

enhancing instruments rather than as effective safeguards capable of preventing or remedying systemic rights-harm.

Part 2 mapped the contemporary BHR ecosystem, outlining the normative architecture of the UNGPs, the OECD Guidelines, the GNI Principles and Implementation Guidelines, and the recent EU directives. Having clarified the regulatory environment, Part 3 introduced the comparative analytical matrix as a methodological tool designed to assess not only formal commitment, but also operationalization, oversight, transparency, and remedial capacity. The matrix was subsequently applied to Meta, Google, and TikTok, as the platforms with the highest global monthly active user bases and therefore the greatest potential human rights impact. The findings reveal a consistent pattern. At the level of formal policy commitments and reporting practices, all examined companies align themselves with the BHR framework and invoke the UNGPs logic of HRDD. However, when examined through the lenses of impact mitigation, accountability, and access to remedy, significant deficiencies emerge. Meta's Oversight Board, despite its innovative design, does not function as a broadly accessible remedial mechanism for ordinary users. Alphabet has lifted its previous ban on certain AI applications capable of causing harm, thereby weakening earlier safeguards. The European Commission's preliminary findings on TikTok indicate that the platform's design requires fundamental revision to comply with the DSA.

Taken together, these case studies and recent events demonstrate a broader structural insight: soft-law commitments, absent external enforcement and binding standards of conduct, lack sufficient deterrent force in situations where human rights protection competes with market incentives, geopolitical pressure, or product-growth strategies. The gradual transformation of HRDD expectations into binding obligations through the CSRD and the CSDDD, therefore, marks not merely a regulatory refinement, but a paradigmatic shift in the governance of corporate human rights responsibility within the European legal order. By embedding due diligence within enforceable company law frameworks, supplemented by transparency duties, supervisory oversight, and potential civil liability, the EU seeks to convert reputational accountability into legal accountability. Whether this transformation will generate measurable improvements in practice will ultimately depend on the consistency and intensity of enforcement. Nevertheless, the transition from voluntary alignment to legally mandated conduct constitutes a decisive step toward closing the structural gap identified throughout this article, namely the distance between normative commitment and operational reality in platform-based human rights governance.

