BRATISLAVA LAW REVIEW

PUBLISHED BY
COMENIUS UNIVERSITY BRATISLAVA
FACULTY OF LAW

p-ISSN 2585-7088
e-ISSN 2644-6359

THE IMPLEMENTATION OF BCTrustAI.SL INTO THE AUTOMATED PRACTICES OF DIGITAL LABOUR PLATFORMS TO ENSURE FAIRNESS, TRANSPARENCY AND ACCOUNTABILITY

Şaban İbrahim Göksal
Tallinn University of Technology,
Department of Law,
Ehitajate tee 5,
19086 Tallinn; Estonia
saban.goksal@taltech.ee
ORCID: 0009-0002-6253-9389

Kristi Joamets
Tallinn University of Technology,
Department of Law,
Ehitajate tee 5,
19086 Tallinn; Estonia
kristi.joamets@taltech.ee
ORCID: 0000-0002-6538-360X

Abstract: *Since digital labour platforms may infringe upon the rights of platform workers through automated decision-making and monitoring practices, the European Parliament and the Council has adopted the Directive (EU) 2024/2831 on improving working conditions in platform work (Directive 2024/2831). This directive seeks to foster fairness, transparency, and accountability, establishing four key requirements in its algorithmic management chapter: transparency, human oversight, human review, rights to information and consultation. However, due to the abstract nature of these provisions, meeting the normative expectations of the directive poses a challenge. This paper presents the implementation of the Blockchain-Based Trustworthy Artificial Intelligence Supported by Stakeholders-In-The-Loop Model (BCTrustAI.SL) into the automated decision and monitoring practices used by digital labour platforms. It aims to discuss theoretically the validation of the concept of BCTrustAI.SL, setting the stage for subsequent technical proofs of concept.*

Key words: *Digital Labour Platforms; Platform Workers; Automated Decision and Monitoring Mechanisms; Transparency; Human Oversight; Human Review; Information and Consultation; Provision of Information to Workers; Blockchain Technology; Trustworthy AI.*

Suggested citation:
Göksal, Ş. İ., Joamets, K. (2025). The Implementation of BCTrustAI.SL into the Automated Practices of Digital Labour Platforms to Ensure Fairness, Transparency and Accountability. *Bratislava Law Review*, 9(1), 9-26. https://doi.org/10.46282/blr.2025.9.1.949

Submitted: 13 November 2024
Accepted: 24 April 2025
Published: 08 July 2025

1. INTRODUCTION

The shift from mechanisation to digitalisation (Knell, 2021) driven by significant improvements in information and communication technologies (ICT), particularly artificial intelligence (AI), has given birth to the creation of new sectors, employment models, and operational practices (Feuerriegel et al., 2020). One sample is digital labour platforms, which employ platform workers and rely on AI algorithms for automated decision-making and monitoring practices to manage the workers (Kalamatiev and Murdzev, 2022; Uzunca and Kas, 2023). Although these algorithms are innovative, they cause significant mistrust due to unfairness, opaqueness and the lack of accountability. They may lead to power imbalances that favour platform owners and service requesters (van Doorn, 2017, pp. 905–910). Additionally, they can reinforce discrimination, as bias in the data or algorithm design impact minor groups (Muller, 2019, pp. 176–196). In response to these issues, the directive 2024/2831[1] represents a step in protecting the rights of platform workers on digital labour platform, it has been adopted on 23 October 2024.

The directive focuses on promoting fairness, transparency and accountability in these practices. Thus, its third chapter (Algorithmic Management) outlines four key requirements: ensuring transparency (Article 9), facilitating human oversight (Article 10), mandating human review (Article 11), and enhancing information and consultation rights (Article 13 and 14). Transparency is the key requirement for clear disclosures about data usage and decision-making processes that impact individuals (Toy, 2023). Human oversight refers to observation and evaluation of automated systems to protect individuals from adverse effects (Tóth et al., 2022). Human review ensures that decisions made by automated systems are open to challenge and scrutiny by humans, thereby providing a mechanism for accountability (Tóth et al., 2022). Finally, information and consultation are two rights mandate that affected individuals are kept informed and can participate in discussions regarding significant changes to automated systems, fostering a collaborative approach to governance and oversight (Lin et al., 2024). Nevertheless, these key requirements are too abstract to effectively meet the three aims established by the directive (Schmager and Sousa, 2021, p. 537, 541; Schmitz et al., 2022, pp. 795–796; Zicari et al., 2023, pp. 2–3; Hohma and Lütge, 2023, p. 905). Thus, Göksal and Solarte Vasquez designed BCTrustAI.SL as a socio-technical model to concretise the abstract elements of institutionalised trustworthy AI concepts (Göksal and Vasquez, 2024, pp. 1–3).

BCTrustAI.SL is a socio-technical model that operationalises the abstract trustworthiness elements—characteristics, principles, and key requirements—of AI systems applying blockchain technology. This model encompasses the characteristics of trustworthy AI *"robustness, ethicality and lawfulness, and human-centeredness."* It is founded on the principles including "*harm prevention, fairness, and human autonomy, which guide its application*." It is also built on the key requirements *"data protection, data governance, technical robustness and safety, transparency, accountability, diversity and non-discrimination, and human agency and oversight.* (Göksal and Solarte Vasquez, 2024, pp. 3–6)." It merges elements from the Blockchain Framework for Trustworthy AI (BF.TAI) (Nassar et al., 2020), BlockIoTIntelligence (Singh et al., 2020), and the Society-In-The-Loop (SITL) framework (Rahwan, 2018). The unique feature of the BCTrustAI.SL is that all AI algorithms can be integrated with this model during the training and data analysis phases, ensuring they operate within its framework. Consequently, this paper aims to implement the model into the most common automated decision or monitoring practice deployed by digital labour platforms, which causes mistrust among platform workers.

This paper maps the kind of practices employed by digital labour platforms and selects one of them, focusing on those that may be at risk of being associated with biased decision-making and invasive monitoring. Following this selection, the BCTrustAI.SL will be implemented into the selected practice to assess whether it meets the normative key requirements as outlined in the third chapter of the directive proposal. This is a theoretical implementation discussion that seeks to validate the BCTrustAI.SL concept in preparation for its technical proof of concept.

The following section provides a detailed background on the automated decision and monitoring practices to map and select one for the implementation and the directive with a specific focus on the four key requirements. The third section introduces and

[1] Directive (EU) 2024/2831 of the European Parliament and of the Council of 23 October 2024 on improving working conditions in platform work. http://data.europa.eu/eli/dir/2024/2831/oj

elaborates on BCTrustAI.SL. The fourth explains the selected practice with particular focus on its components, stakeholders and operational workflow and the implementation of BCTrustAI.SL into the practice. The final section concludes with a summary of the findings and contributions of this work.

2. EXPLORING OPERATIONAL AND REGULATORY LANDSCAPES OF AUTOMATED PRACTICES IN PLATFORM WORK

This section maps the automated decision and monitoring practices employed by digital labour platforms and selects the one for the implementation of the BCTrustAI.SL. It also examines the challenges associated with the chosen practice. Additionally, this section delves into the background of the directive.

2.1 Operational Landscape: Mapping the Automated Decision-Making and Monitoring Practices and Selection the Most Common One

In technology-driven work environments, automated decision-making and monitoring mechanisms are integral in the operational practices of digital labour platforms. As shown in Table 1, this part aims to map these practices. It identifies stakeholders and categorises the practices by colour codes: practices involving automated decision-making are shown in light grey, those with automated monitoring mechanisms are in dark grey, and practices that involve a mix of both are indicated in grey.

Table 1: Outline of Automated Decision-Making and Monitoring Practices Employed by the Digital Labour Platforms

Name	Direct Stakeholders	Reference
Task allocation practice	Platform operators, Workers, Customers, Technology developers, Data scientists, analysts	(Rozas et al., 2021) (Alasoini et al., 2023)
Conflict Resolution Practice	Platform operators, Workers, Customers, Technology developers, Legal advisors, Community groups and advocacy organisation	(Weiss, 2020) (Lee and Cui, 2024)
Platform Worker's Performance Evaluation Practice	Platform operators, Workers, Customers, Technology developers, Data scientists, analysts and engineers	(Rosenblat et al., 2017) (De Stefano and Aloisi, 2018, pp. 19–21) (Chan, 2019) (Alasoini et al., 2023)
Dynamic - Surge Pricing Practice	Platform operators, Workers, Clients or Customers, Technology developers, Economists and market analysts,	(Battifarano and Qian, 2019) (Yan et al., 2020) (Nunan and Di Domenico, 2022) (Cram et al., 2022) (Kopalle et al., 2023)

Name	Direct Stakeholders	Reference
Digital Tracking Practice	Platform operators, Workers, Clients or Customers, Technology developers, Privacy and data protection advocacy groups, Data engineer, Data scientists, analysts and engineers	(De Stefano and Aloisi, 2018, pp. 19–21) (Joyce and Stuart, 2021) (Hernandez et al., 2024)
Behavioural Influence Practice	Platform operators, Workers, Clients or Customers, Behavioural scientists, Economics, Ethics boards and advisory panels, Technology developers	(Wang et al., 2022) (Uzunca and Kas, 2023)

Source: Table prepared by the authors.

Task allocation practice refers to an automated process that systematically distributes tasks among workers using decision-making algorithms, assessing factors like availability, skill level, and past performance to optimise workflow (Alasoini et al., 2023). Conflict resolution practice is another automated process, and it addresses and resolves disputes among platform operators, workers, and clients, typically through predefined protocols and mediation strategies (Lee and Cui, 2024). Performance evaluation practice is defined as an automated procedure of assessing the performance of platform workers using quantitative and qualitative metrics such as task completion rate, quality of work, client feedback, and adherence to platform standards (De Stefano and Aloisi, 2018; Chan, 2019; Alasoini et al., 2023). Dynamic-Surge Pricing Practice refers to a method for service prices based on real-time demand and supply conditions through automated decision-making algorithms (Battifarano and Qian, 2019; Yan et al., 2020). Digital Tracking Practice is the systematic use of sensors and automated monitoring mechanism to monitor, collect, and analyse data on the activities, locations, and work patterns of workers (Joyce and Stuart, 2021; Hernandez et al., 2024). Finally, behavioural influence practice refers to the application of behavioural science techniques combined with automated decision-making and monitoring mechanisms to subtly guide and optimise worker actions and decisions, enhancing productivity and engagement (Wang et al., 2022; Uzunca and Kas, 2023).

Among the practices identified above, the platform worker's performance evaluation practice was selected for the implementation of the BCTrustAI.SL. It could generate mistrust due to its potential for discrimination (Grgurev and Radic, 2023) and intrusive surveillance (Mettler, 2024) via automated decision-making and monitoring mechanisms. These systems can exacerbate existing bias or data inaccuracy, leading to unjust evaluations and possible economic or reputational damage to workers (Jahanbakhsh et al., 2020; Fredman et al., 2020). The opaqueness of these processes and workers' inability to challenge or comprehend the rationale behind evaluations intensify feelings of mistrust and vulnerability (Chen et al., 2023). Along with biased decision making and lack of transparency, the invasive monitoring aspect of these systems involves excessively tracking worker activities and behaviours, which can feel overbearing and contribute to a lack of trust (Mettler, 2024). The platform worker's performance evaluation practice involves both automated decision-making and monitoring mechanisms and is prominently utilised across digital labour platforms (Dunn, 2020; Duggan et al., 2020).

2.2 Regulatory Landscape: Exploring the Background of the Directive and the Provisions of Its Algorithmic Management Sections

To better understand the regulatory framework surrounding digital labour platforms, it's crucial to delve into the Directive 2024/2831. Particular emphasis should be placed on the four key requirements. By examining these requirements, it is possible to study how they safeguard platform workers, and ensure that the development, deployment and use of automated decision and monitoring systems remains transparent, fair and accountable. Art. 9 mandates the disclosure of the use and specifics of automated monitoring and decision-making systems that impact their work conditions. It requires platforms to inform workers about the monitoring of their actions and the automated decisions that could affect their job security, earnings, and working conditions. This information must also be available to workers' representatives and labour authorities upon request. Art. 10 requires from the platforms monitoring and evaluating the impact of decisions made or supported by automated systems on workers' conditions, specially assessing potential risks to their health and safety. To safeguard these conditions, platforms should employ sufficiently trained and competent human personnel to oversee these systems. Article 11 establishes the right of platform workers to seek explanations and human review of decisions made by automated systems that significantly affect their working conditions. It mandates digital labour platforms to provide workers access to human contact to discuss the circumstances and rationale behind such decisions. Articles 13 and 14 determine that digital labour platforms must engage in open dialogues with platform workers' representatives—or directly with workers in the absence of representatives—regarding decisions related to the implementation or significant modification of automated monitoring and decision-making systems. The aim is to foster social dialogue in AI governance.

Reviewing institutionalised European Union (EU)'s Regulatory Framework on AI (Joamets and Vasquez, 2020, pp. 112, 115–122) shows that these four provisions highly align with the elements of the trustworthy AI concept proposed by EU (Göksal et al., 2025). The Ethics Guidelines, published in 2019 by the Independent High-Level Expert Group on Artificial Intelligence (AI HLEG)[2] , have institutionalised trustworthiness as a core attribute of AI at the EU level. Art. 9 primarily addresses issues of transparency in automated decision and monitoring mechanisms, it is one key requirement according to the Ethics Guidelines to ensure trustworthiness of AI. Transparency ensures that all elements of the AI process are accessible, understandable, and verifiable by users and other stakeholders (AI HLEG, 2019, p. 18). Articles 10, 11, 13 and 14 have a link with accountability and human agency and oversight, these are another two key requirements in the Ethics Guidelines. Accountability mandates that AI systems and their operators are responsible for the outcomes of their actions, ensuring that AI operations are justifiable, traceable, and that any adverse effects can be adequately addressed (AI HLEG, 2019, pp. 19–20). And, the key requirement of human agency and oversight, advocating for the active involvement of platform workers or their representatives in the automated decision-making and monitoring processes. Human agency and oversight ensure that AI systems incorporate human input and direction effectively, allowing humans to retain control over AI decisions and to intervene or override AI operations whenever necessary. This ensures that AI systems enhance human capacities without replacing them,

[2] Independent High-Level Expert Group on Artificial Intelligence Set Up By The European Commission. (2019). Policy and Investment Recommendations for trustworthy AI.

DOI: 10.46282/blr.2025.9.1.949

maintaining critical human control in sensitive and impactful AI applications (AI HLEG, 2019, pp. 15–16).

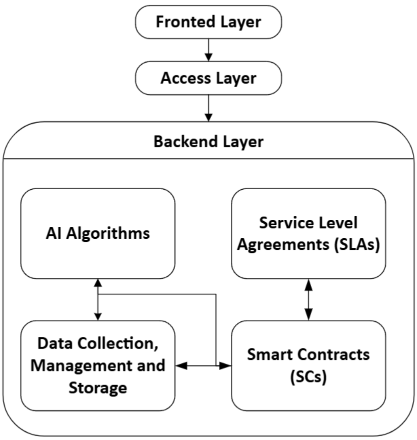
3. THE BLOCKCHAIN-BASED TRUSTWORTHY AI MODEL (BCTrustAI.SL)

This section presents the BCTrustAI.SL (Göksal and Solarte Vasquez, 2024), detailing its structure and functionality in three distinct parts. Firstly, the model's components will be introduced, subsequently, the workflow dynamics of the model, and finally, the discussion will explore how the model conforms and complies with the requirements of institutionalised trustworthy AI concept.

3.1 System Architecture

The BCTrustAI.SL's architecture is structured across two main tiers as seen in the Figure 1. The Backend Tier comprises five distinct intelligence layers: Access, Smart Contracts (SCs), Human Agency and Oversight (HAO), AI, and Data Governance. The Data Governance Layer is further segmented into two sub-layers, each containing four levels. The first, an Internet of Things (IoT) Device Sub-layer, handles data received and managed from IoT sources. The second, the Other Data Sub-layer, is responsible for data acquisition from various other sources through Web-based Platforms (WBPs).

Figure 1: Visualising the Blueprint of BCTrustAI.SL



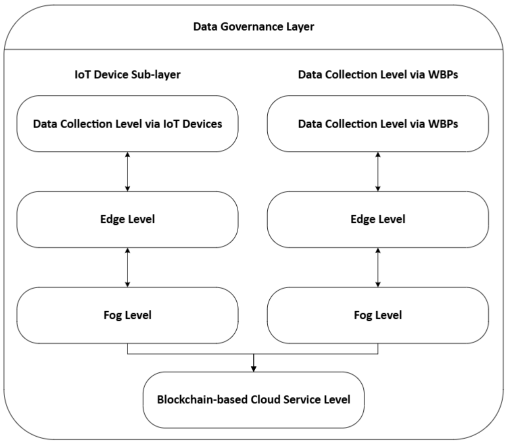
Source: The figure was created by the authors

Stakeholders interact with the system through the Frontend Tier with Decentralized Applications (DApps) component. The DApps should adhere to usability criteria, including heuristics for user-centric graphic design (Jang et al., 2020) and legally relevant attribution factors (Solarte-Vasquez and Nyman-Metcalf, 2017, p. 226–232). These interfaces allow stakeholders to select parameters and pre-coded Service Level Agreements (SLAs) for the HAO operations. In the Backend, the Access Layer serves as a crucial link between the two tiers, ensuring robust data connections between the DApps and the SCs. It supports multiple data transfer protocols to enhance system flexibility.

The SCs layer autonomously governs the system by executing data transactions based on parameters set by the SLAs. This layer has direct connections to the HAO, AI, and Data Governance layers. The HAO layer governs all operations related to AI governance mechanisms. From this layer, various protocols for impact assessments, user experience (UX) evaluations, and ethical compliance monitoring are initiated. The AI Layer contains algorithms for data processing and decision-making, including explainable AI (XAI) algorithms that clarify how these processes are conducted.

The Data Governance Layer oversees the storage and management of system data, which includes input, output, insights from training and analysis, as well as operational and raw data collected during various phases. This layer is organised hierarchically into levels and sub-layers specifically designed for data gathering as shown in Figure 2. One sub-layer specialises in IoT sources, while the other focuses on WBPs. Data collection is categorised into different levels: Data Collection, Edge, Fog, and a Blockchain-based Cloud Service levels. Data entry begins at the Data Collection Level within each sub-layer. The Edge Level houses intelligent base station nodes that collect and categorise external traffic data. The Fog Level consists of Fog Nodes essential for executing algorithms that manage resources, energy efficiency, and scalability in real-time on external traffic data. Finally, the common Blockchain-based Cloud Service Level, supported by its algorithmic intelligence, enhances system efficiency and fortifies the distributed storage with robustness and security through blockchain technology.

Figure 2: Visualising Data Governance Layer of the BCTrustAI.SL



Source: The figure was created by the authors

3.2 Operational Processes

In the operation, the BCTrustAI.SL system is activated when stakeholders use the DApps at the Frontend Layer. Here, they set their preferences within the pre-defined SLAs to suit their objectives. This action triggers the corresponding SCs that oversee and execute all algorithmic processes, including data collection, training, and analysis. The SLAs function as customisable templates designed to accommodate the specific terms and conditions required for each operation that the SCs undertake. These templates, which are multimodal, should be crafted by developers drawing on expert knowledge to represent various types of operations.

In the Data Governance Layer, raw data initially enters from IoT Devices or WBPs. In the first sub-layer, a network of diverse IoT devices and sensors connected in nodes generates raw data, which is sent to the Edge Level. AI-enabled Base Station Nodes utilise this IoT device network to gather a second type of input consisting of external traffic data. Both this and the data from WBPs undergo algorithmic evaluation at the Edge Level, focusing on scalability, load balancing, and other crucial network considerations. The first raw data inputs from both sub-layers advance to their respective Fog Levels, where AI-enabled fog nodes analyse the internal traffic of this data in real time, addressing network challenges as they arise. Subsequently, all this input, along with the initial output from traffic analysis in each sub-layer, is forwarded to the Blockchain-based Cloud Service Level. Here, the data is sorted, pre-processed, and stored. By this stage, all data has undergone some processing and is no longer considered 'raw'. The outcomes from this phase, excluding traffic-related data, become the primary input for the AI Layer, where core AI functions, such as data analysis or Machine Learning (ML) training, are carried out.

Data moves continuously and synchronously between the AI-enabled Blockchain-Based Cloud Service and the predictors. The primary data input is relayed to the AI Layer for algorithm training and analysis. A preliminary output, along with insights generated by the predictors, is then sent back to the Cloud Service Level at the Data Governance Layer for both storage and further HAO activities involving expert stakeholders and users. Ultimately, this initial output and insights are forwarded to the HAO Layer for evaluation and possibly additional HAO actions.

The final stages involve stakeholders interacting with DApps to define SLAs using a unique set of templates tailored for post-analysis/training HAO activities. The goal is to activate the SCs that handle transactions across the different layers. Ultimately, stakeholders hold the decision-making power, choosing to approve, reject, or request enhancements. These steps form a feedback loop subsystem, which is a critical part of implementing and enhancing AI governance mechanisms. If the final outputs are rejected, the training phases will restart, and the HAO cycles will continue until further adjustments are deemed unnecessary.

3.3 The Model's Conformity and Compliance with the Four Key Requirements

BCTrustAI.SL achieves the practical implementation of the abstract requirements of institutionalised trustworthy AI framework by establishing a robust and complex infrastructure that incorporates chosen technologies and principled processes. Along with other elements of trustworthy AI, this model ensures conformity with the three trustworthy AI key requirements that are crucial for the compliance with the four provisions outlined by the directive that are transparency, accountability and human agency and oversight.

BCTrustAI.SL excels in operationalising transparency through a multi-layered approach that integrates blockchain technology and advanced algorithmic processes. This robust integration ensures that every transaction and decision made by the AI system is recorded on a decentralised ledger, enabling traceability, a foundational element of transparency. Beyond mere traceability, BCTrustAI.SL also places a strong emphasis on communication and explainability, crucial elements of transparency. The system employs XAI predictors to elucidate the reasoning behind its decisions, making these processes accessible and comprehensible to all stakeholders. Additionally, the model utilises communication protocols with its HAO layer to ensure solid human agency and oversight. This focus on traceability, explainability, and communication ensures that transparency in BCTrustAI.SL is not merely theoretical but a practical reality, deepening on the operational life of the AI system.

In terms of accountability, BCTrustAI.SL's architecture supports the identification of decision-making processes, which can be attributed to specific AI actions. This capability is crucial for maintaining the scrutiny and accountability of AI/ML applications. According to Felzmann et al. (2020, p. 3338), meeting transparency key requirement fosters a stronger accountability by ensuring that actions are both visible and justifiable to all stakeholders.

The model also facilitates extensive human agency and oversight by incorporating mechanisms such as human-in-the-loop (HITL) (Wu et al., 2022, p. 2), human-on-the-loop (HOTL) (Li et al., 2020), human-in-command (HIC) (Johnson, 2023), and stakeholders-in-the-loop (SITL) (Göksal and Solarte Vasquez, 2024) across its operational phases. These mechanisms ensure that human judgement is integral in training, monitoring, and retraining processes, enhancing the system's adherence to ethical standards and regulatory compliance. Together, these features establish BCTrustAI.SL not just as a tool, but as a partner in decision-making, elevating the standard for human-centred AI applications.

4. IMPLEMENTATION OF BCTrustAI.SL INTO THE AUTOMATED PLATFORM WORKER'S PERFORMANCE EVALUATION PRACTICE

This section will focus on the automated platform worker's performance evaluation practice, specifically its components, stakeholders, and operational dynamics. The implementation of the BCTrustAI.SL model into this practice will be explored.

4.1 Overview of Automated Performance Evaluation Practices

Automated Performance Evaluation (APE) Systems for platform workers encompass a series of interconnected *components* that are Data Collection and Algorithmic Evaluation (Waldkirch et al., 2021). Data Collection Mechanism refers to the systematic acquisition of specific, quantifiable information related to the activities and outcomes of platform workers. This component focuses on capturing a wide array of performance-related data points such as task completion times, customer ratings, adherence to schedules, and acceptance rates for assigned tasks. The scope and nature of the data collected are tailored to the particular requirements and operational nuances of the platform (Fredman et al., 2020). The Algorithmic Evaluation Mechanism processes the collected data to assess platform worker performance. This mechanism employs AI predictors to analyse the gathered metrics, calculating performance scores based on predefined criteria such as efficiency, reliability, customer satisfaction, and rule compliance. These predictors are designed to provide an evaluation and focus on measurable outcomes (Park and Ryoo, 2023).

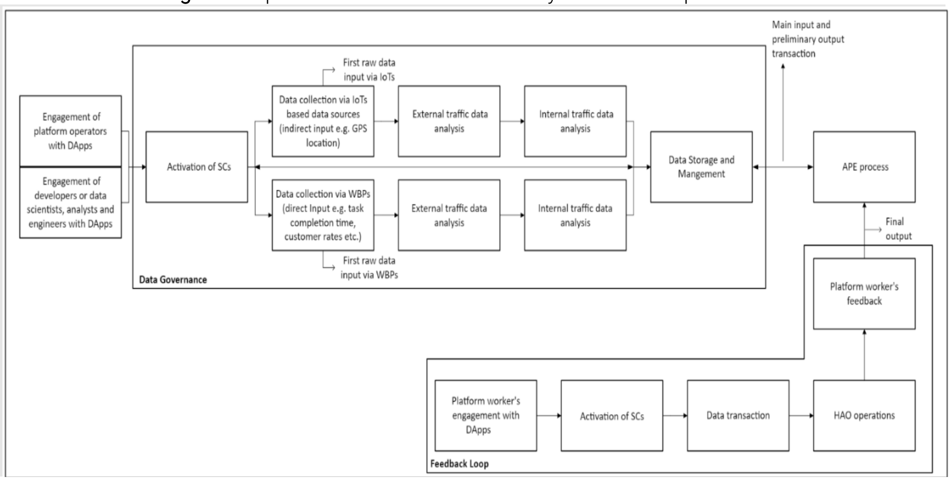
In the ecosystem of APE System, the stakeholders include platform workers, platform operators, customers, developers, and data scientists, analysts and engineers (Jahanbakhsh et al., 2020; Waldkirch et al., 2021). Platform workers are the individuals engaged in completing various tasks assigned through the platform, whose performance is directly evaluated based on metrics described above. They are essential to the operational dynamics of the platform as their performance data fuels the APE System, influencing not only their immediate job opportunities and earnings but also their long-term career prospects within the digital platform environment (Gallagher et al., 2023). Platform operators are entities or individuals who manage and oversee the functioning of the platform, responsible for setting the operational parameters and standards that define the working environment. They configure and maintain the performance evaluation system, ensuring that it aligns with broader business objectives and regulatory requirements. As key stakeholders, their role is critical in balancing the needs of customers and platform workers, while ensuring that the platform remains competitive and compliant with industry standards. Through their oversight, platform operators directly influence the effectiveness and integrity of the APE System, shaping the overall platform ecosystem (Harmon and Silberman, 2019). Customers refer to the end-users who utilise the platform to engage services offered by platform workers. They provide feedback and ratings based on their interactions and satisfaction with the services received. This input is important for the data collection mechanism of the performance evaluation system, as it directly influences the performance scores of platform workers (Lu et al., 2024). Developers and data scientists, analysts and engineers are two different actors in the APE System for the backend (Ahopelto, 2023). Developers are professionals who build, implement, and maintain the technical aspects of the APE System. They ensure the platform's infrastructure supports all functionalities related to performance evaluation, from data entry interfaces to the complex backend processes that handle data storage and processing. Their work is essential for the smooth operation and scalability of the system, as they address both immediate technical issues and long-term software enhancements (Shestakofsky and Kelkar, 2020). Data scientists, analysts and engineers are experts who specialise in analysing large sets of data and developing algorithms that process and interpret this information within the APE System. They apply statistical analysis and machine learning techniques to derive actionable insights from performance data, which inform the algorithmic evaluation mechanisms. Their contributions are critical in designing predictive models that assess worker performance fairly and accurately, ensuring that the system's outputs are both reliable and transparent (Basukie et al., 2020).

The *operational workflow* of the APE System as described by Waldkirch et al. (2021) is a sophisticated sequence that begins with the Data Collection Mechanism, where data about platform workers' activities and outcomes, such as task completion times, customer ratings, schedule adherence, and task acceptance rates, are gathered continuously. This mechanism uses various data sources and collection methods to ensure a holistic view of worker activities and outcomes. The primary source of data includes direct inputs from the platform's digital interface, where task completion times are logged, customer ratings are recorded, and worker adherence to schedules is monitored. Each interaction or transaction made by a worker is automatically tracked by the system, capturing details such as the time taken to complete tasks, the quality of service delivered as rated by customers, and the frequency and timeliness of task acceptance and completion (Rosenblat et al., 2017; Chan, 2019; Alasoini et al., 2023). In addition to these direct measures, the system also integrates indirect input such as GPS tracking for location-based tasks, which helps in monitoring route efficiency and punctuality in real-world settings. This is particularly relevant for delivery and transportation services where geographical movement and time efficiency are critical performance indicators (De Stefano and Aloisi, 2018). This data is then fed into the Algorithmic Evaluation Mechanism, where AI predictors analyse the metrics to assess each worker's performance. These predictors calculate performance scores based on criteria like efficiency, reliability, customer satisfaction, and compliance with rules (Harmon and Silberman, 2019). Throughout this process, developers and data scientists, analysts and engineers monitor and refine the technical and analytical aspects of the system, ensuring its accuracy, while platform operators use the insights generated to make informed decisions about operational changes, data updates and strategic directions (Wu et al., 2022).

4.2 The Implementation of BCTrustAI.SL with APE System

After the system is implemented, it will be activated when platform operators interact with the DApps at the Frontend Layer. Operators select the parameters from pre-defined SLAs related data collection and APE process. These initiate the corresponding SCs that execute all algorithmic operations. The SCs begin the data gathering and analysis needed to evaluate platform worker's performance, which takes place across two sub-layers: the Data Governance Layer and the AI Layer, the latter utilising AI and XAI predictors. The stages and operational flow of the implementation are illustrated in Figure 3.

Figure 3: Operational Processes of APE System after Implementation



Source: Original chart prepared by the authors.

When platform operators interact with the DApps to activate the related SCs, the data collection process begins. During this phase, the system gathers the raw data as well as internal and external traffic data for analysis. Subsequently, this collected raw data is analysed using AI and XAI predictors to evaluate the performance of platform workers. After the analysis, a preliminary output is generated.

In addition to platform operators, developers, data scientists, analysts, or engineers also need to engage with the DApps to oversee the data collection and analysis processes. The raw, external and internal traffic data from the collection process are collected and stored along with preliminary and final outputs from the analysis process in a Blockchain-Based Cloud Service. When the stakeholder activates a relevant SC, the necessary data is retrieved from the Blockchain-Based Cloud Service to carry out the oversight process.

Specifically, storing the raw data, preliminary outputs, and final outputs in the Blockchain-Based Cloud Service, and making them accessible to all stakeholders as needed, is critical to meeting the traceability aspect of the transparency key requirements (Mora-Cantallops et al., 2021). Along with the traceability aspect, ensuring the accessibility of communication for all stakeholders is crucial. This is because the user-centric design of the DApps will provide an interface that facilitates easy access to all necessary data regarding automated decision-making and monitoring processes, ensuring clear communication across stakeholders (Solarte-Vasquez and Hietanen-Kunwald, 2020, pp. 186–191). Furthermore, the aspect of explainability will be achieved through the use of XAI predictors located within the AI Layer. For instance, if a platform worker wants to understand the reasons behind a recent decrease in their performance rating, the XAI predictors can generate a detailed breakdown of the contributing factors, such as changes in customer satisfaction scores or task completion times (Hassija et al., 2024). This detailed feedback could allow the worker to identify specific areas for improvement and directly address any concerns about the fairness or accuracy of the evaluation process.

Just like the communication aspect of transparency, providing a user-centric interface through the DApps will successfully meet the key requirement of human oversight by allowing all stakeholders to interact with the system effectively (Sharp et al., 2021). This ensures that stakeholders can monitor and evaluate the impacts of automated decisions on platform workers. For example, platform operators can use these interfaces to track real-time data flows and AI decision logs, which include details on how decisions such as task assignments or performance evaluations are made. This capability could enhance immediate human intervention when necessary to correct or modify AI decisions that may adversely affect platform workers, thereby supporting a more dynamic and responsive governance model (Hadzovic et al., 2024).

Ensuring transparency and facilitating human review inherently strengthens accountability—a key requirement tightly interwoven with transparency (Felzmann et al., 2020, p.3338). Consequently, this aligns closely with the Directive 2024/2831, which proposes specific measures for human review to ensure accountability within digital labour platforms. For instance, platform workers could challenge and seek clarification on automated decisions that impact their work schedules and compensation. If a worker notices a sudden decrease in allocated tasks or an unexpected change in pay rates, they can request a human review of the algorithm’s decision. And the authorised human can review the related automated process on the row data, the parameters of predictors and preliminary and final outputs by activating related SC to receive data via DApps.

With the implementation, besides data collection and analysis, the APE System will also include a feedback loop as mandated by the directive under Article 13 and 14. Platform workers may engage with the DApps, activating the relevant SC. This SC will then manage the HAO operation, providing feedback on the preliminary outputs generated post-analysis. This feedback will inform the creation of the final outputs, completing in that way the APE process. For example, platform workers will be able to directly interact with the data that has been used to evaluate them, querying the system for further details or discrepancies they perceive in their ratings. They can submit corrections or provide additional context, which may alter how their performance is ultimately assessed and reported. This capability ensures that the evaluation process remains dynamic and interactive, rather than static and unilateral (Lin et al., 2020).

5. CONCLUSION

Advancements in AI and ICT have led to catalysing digital labour platforms that use AI for decision-making and monitoring. However, concerns regarding fairness, transparency, and accountability in these practices have eroded stakeholder trust. In response, the European Commission introduced the Directive 2024/2831, which outlines four key requirements to enhance these aspects, but the directive's key requirements are often considered too vague for practical application on digital labour platforms. Thus, this paper explained that the implementation of BCTrustAI.SL into the selected automated decision-making and monitoring mechanism (Platform Worker's Performance Evaluation Practice) operationalises the directive's provisions and validates the model conceptually.

BCTrustAI.SL was designed as a socio-technical proposal that conforms and complies with the elements of the institutionalised trustworthy AI framework, encapsulating the four key requirements specified in the Directive 2024/2831. Following its implementation, the APE System has the capacity to enhance fairness, transparency and accountability:

- becoming transparent as Art. 9 mandates. Because it is founded on blockchain technology, every transaction and decision made by the AI system is recorded on a decentralised ledger, enabling traceability. It also uses XAI predictors to clarify decision-making processes, making them understandable to the stakeholders. Additionally, it uses communication protocols in its HAO layer to facilitate robust stakeholder interactions.
- becoming accountable and ensures human review as Article 11 requires. The implemented APE System, based on blockchain architecture, supports the identification and attribution of decision-making processes to specific AI actions.
- allowing human oversight and information and consultation as required in Article 10, 13 and 14 by promoting extensive human interaction through various AI governance mechanisms during operational phases. These mechanisms integrate human judgment in training, monitoring, and retraining processes, boosting the system's conformity and compliance with ethical standards and regulations.

With the implementation, the APE System may have potential for significant advancements in conformity and compliance with the directive's four key requirements. However, the underlying infrastructure of blockchain technology currently lacks the capacity to fully support the BCTrustAI.SL model's integration into automated platform worker performance evaluation practices. Thus, it is essential to focus on enhancing the capacity of blockchain technology to accommodate larger datasets and facilitate broader stakeholder interaction. This advancement requires targeted research and development efforts. Additionally, there is a need to concentrate on the practices and algorithms employed by digital labour platforms. By focusing more on these areas in academic research, there is an opportunity to improve the effective implementation.

BIBLIOGRAPHY:

Ahopelto, H. (2023). *Governance mechanisms and metrics for digital platform workers: Multiple case study of digital labour platforms*. https://osuva.uwasa.fi/handle/10024/15420

Alasoini, T., Immonen, J., Seppänen, L. and Känsälä, M. (2023). Platform workers and digital agency: Making out on three types of labor platforms. *Frontiers in Sociology*, 8, https://doi.org/10.3389/fsoc.2023.1063613

Basukie, J., Wang, Y., and Li, S. (2020). Big data governance and algorithmic management in sharing economy platforms: A case of ridesharing in emerging markets. *Technological Forecasting and Social Change*, *161*, 120310, https://doi.org/10.1016/j.techfore.2020.120310

Battifarano, M. and Qian, Z. (S.) (2019). Predicting real-time surge pricing of ride-sourcing companies. *Transportation Research Part C: Emerging Technologies*, 107, 444–462, https://doi.org/10.1016/j.trc.2019.08.019

Chan, N. K. (2019). The Rating Game: The Discipline of Uber's User-Generated Ratings. *Surveillance & Society*, 17(1/2), 183–190, https://doi.org/10.24908/ss.v17i1/2.12911

Chen, J., Luo, J., Hu, W. and Ma, J. (2023). Fit into work! From formalizing governance of gig platform ecosystems to helping gig workers craft their platform work. *Decision Support Systems*, 174, 114016, https://doi.org/10.1016/j.dss.2023.114016

Cram, W. A., Wiener, M., Tarafdar, M. and Benlian, A. (2022). Examining the Impact of Algorithmic Control on Uber Drivers' Technostress. *Journal of Management Information Systems*, 39(2), 426–453, https://doi.org/10.1080/07421222.2022.2063556

De Stefano, V., and Aloisi, A. (2018). European legal framework for digital labour platforms (JRC Science for Policy Report No. JRC112243). Publications Office of the European Union. https://doi.org/10.2760/78590

Duggan, J., Sherman, U., Carbery, R. and McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. Human *Resource Management Journal*, 30(1), 114–132, https://doi.org/10.1111/1748-8583.12258

Dunn, M. (2020). Making gigs work: Digital platforms, job quality and worker motivations. *New Technology, Work and Employment*, 35(2), 232–249, https://doi.org/10.1111/ntwe.12167

Felzmann, H., Fosch-Villaronga, E., Lutz, C. and Tamò-Larrieux, A. (2020). Towards Transparency by Design for Artificial Intelligence. *Science and Engineering Ethics*, 26(6), 3333–3361, https://doi.org/10.1007/s11948-020-00276-4

Feuerriegel, S., Dolata, M. and Schwabe, G. (2020). Fair AI. *Business & Information Systems Engineering*, 62(4), 379–384, https://doi.org/10.1007/s12599-020-00650-3

Fredman, S., du Toit, D., Graham, M., Howson, K., Heeks, R., van Belle, J.-P., Mungai, P. and Osiki, A. (2020). Thinking Out of the Box: Fair Work for Platform Workers. *King's Law Journal*, 31(2), 236–249, https://doi.org/10.1080/09615768.2020.1794196

Gallagher, C., Gregory, K., and Karabaliev, B. (2023). Digital worker inquiry and the critical potential of participatory worker data science for on-demand platform workers. *New Technology, Work and Employment*. https://doi.org/10.1111/ntwe.12286

Göksal, Ş. İ. and Solarte Vasquez, M. C. (2024). The Blockchain-Based Trustworthy Artificial Intelligence Supported by Stakeholders-In-The-Loop Model. *Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration*, 32(2), Article 2, https://doi.org/10.46585/sp32022083

Göksal, Ş. İ., Vasquez, M. C. S. and Chochia, A. (2025). The EU AI Act's Alignment within the European Union's Regulatory Framework on Artificial Intelligence. *International and Comparative Law Review*, *24*(2), 25–53, https://doi.org/10.2478/iclr-2024-0017

Grgurev, I. and Radic, I. B. (2023). Indirect Discrimination of Platform Workers. *Zbornik Pravnog Fakulteta u Zagrebu*, 73(2–3), 233–260.

Hadzovic, S., Becirspahic, L. and Mrdovic, S. (2024). It's time for artificial intelligence governance. *Internet of Things*, 27, 101292, https://doi.org/10.1016/j.iot.2024.101292

Harmon, E. and Silberman, M. S. (2019). Rating Working Conditions on Digital Labor Platforms. *Computer Supported Cooperative Work (CSCW)*, 28(5), 911–960, https://doi.org/10.1007/s10606-018-9313-5

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M. and Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, 16(1), 45–74, https://doi.org/10.1007/s12559-023-10179-8

Hernandez, R. H. (Lindy), Song, Q., Kou, Y. and Gui, X. (2024). 'At the end of the day, I am accountable': Gig Workers' Self-Tracking for Multi-Dimensional Accountability Management. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–20, https://doi.org/10.1145/3613904.3642151

Hohma, E. and Lütge, C. (2023). From Trustworthy Principles to a Trustworthy Development Process: The Need and Elements of Trusted Development of AI Systems. *AI*, 4(4), Article 4, https://doi.org/10.3390/ai4040046

Jahanbakhsh, F., Cranshaw, J., Counts, S., Lasecki, W. S. and Inkpen, K. (2020). An Experimental Study of Bias in Platform Worker Ratings: The Role of Performance Quality and Gender. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13, https://doi.org/10.1145/3313831.3376860

Jang, H., Han, S. H. and Kim, J. H. (2020). User Perspectives on Blockchain Technology: User-Centered Evaluation and Design Strategies for DApps. *IEEE Access*, 8, 226213–226223, https://doi.org/10.1109/ACCESS.2020.3042822

Joamets, K. and Vasquez, M. C. S. (2020). Regulatory Framework of the Research-Based Approach to Education in the EU. *TalTech Journal of European Studies*, 10(3), 109–136, https://doi.org/10.1515/bjes-2020-0024

Johnson, J. (2023). Automating the OODA loop in the age of intelligent machines: Reaffirming the role of humans in command-and-control decision-making in the digital age. *Defence Studies*, *23*(1), 43–67, https://doi.org/10.1080/14702436.2022.2102486

Joyce, S. and Stuart, M. (2021). Digitalised management, control and resistance in platform work: a labour process analysis. In: Haidar, J. and Keune, M. (ed.), *Work and Labour Relations in Global Platform Capitalism* (pp. 158–584). Cheltenham: Edward Elgar Publishing.

Kalamatiev, T. and Murdzev, N. (2022). The Notion of Digital Labour Platforms and the European Incentive for Improvement of the Working Conditions of the Platform Workers. *Harmonius: Journal of Legal and Social Studies in South East Europe*, 2022, 207–248.

Knell, M. (2021). The digital revolution and digitalized network society. *Review of Evolutionary Political Economy*, 2(1), 9–25, https://doi.org/10.1007/s43253-021-00037-4

Kopalle, P. K., Pauwels, K., Akella, L. Y. and Gangwar, M. (2023). Dynamic pricing: Definition, implications for managers, and future research directions. *Journal of Retailing*, 99(4), 580–593, https://doi.org/10.1016/j.jretai.2023.11.003

Lee, W. K. and Cui, Y. (2024). Should Gig Platforms Decentralize Dispute Resolution? *Manufacturing & Service Operations Management*, 26(2), 519–536, https://doi.org/10.1287/msom.2022.0398

Li, N., Adepu, S., Kang, E. and Garlan, D. (2020). Explanations for human-on-the-loop: A probabilistic model checking approach. *Proceedings of the IEEE/ACM 15th*

International Symposium on Software Engineering for Adaptive and Self-Managing Systems, 181–187, https://doi.org/10.1145/3387939.3391592

Lin, J., Ma, Z., Gomez, R., Nakamura, K., He, B. and Li, G. (2020). A Review on Interactive Reinforcement Learning From Human Social Feedback. *IEEE Access*, 8, 120757–120765, https://doi.org/10.1109/ACCESS.2020.3006254

Lin, J., Tomlin, N., Andreas, J. and Eisner, J. (2024). Decision-Oriented Dialogue for Human-AI Collaboration. *Transactions of the Association for Computational Linguistics*, 12, 892–911, https://doi.org/10.1162/tacl_a_00679

Lu, Y., Yang, M. M., Zhu, J., and Wang, Y. (2024). Dark side of algorithmic management on platform worker behaviors: A mixed-method study. *Human Resource Management*, *63*(3), 477–498, https://doi.org/10.1002/hrm.22211

Mettler, T. (2024). The connected workplace: Characteristics and social consequences of work surveillance in the age of datification, sensorization, and artificial intelligence. *Journal of Information Technology*, 39(3), 547–567, https://doi.org/10.1177/02683962231202535

Mora-Cantallops, M., Sánchez-Alonso, S., García-Barriocanal, E. and Sicilia, M.-A. (2021). Traceability for Trustworthy AI: A Review of Models and Tools. *Big Data and Cognitive Computing*, 5(2), Article 2, https://doi.org/10.3390/bdcc5020020

Muller, Z. (2019). Algorithmic Harms to Workers in the Platform Economy: The Case of Uber. *Columbia Journal of Law and Social Problems*, 53(2).

Nassar, M., Salah, K., ur Rehman, M. H. and Svetinovic, D. (2020). Blockchain for explainable and trustworthy artificial intelligence. *WIREs Data Mining and Knowledge Discovery*, *10*(1), 1340, https://doi.org/10.1002/widm.1340

Nunan, D., and Di Domenico, M. (2022). Value creation in an algorithmic world: Towards an ethics of dynamic pricing. *Journal of Business Research*, 150, 451–460, https://doi.org/10.1016/j.jbusres.2022.06.032

Park, S. and Ryoo, S. (2023). How Does Algorithm Control Affect Platform Workers' Responses? Algorithm as a Digital Taylorism. *Journal of Theoretical and Applied Electronic Commerce Research*, 18(1), Article 1, https://doi.org/10.3390/jtaer18010015

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, *20*(1), 5–14, https://doi.org/10.1007/s10676-017-9430-8

Rosenblat, A., Levy, K. E. C., Barocas, S. and Hwang, T. (2017). Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination. *Policy & Internet*, 9(3), 256–279, https://doi.org/10.1002/poi3.153

Rozas, D., Saldivar, J. and Zelickson, E. (2021). The platform belongs to those who work on it! Co-designing worker-centric task distribution models. *Proceedings of the 17th International Symposium on Open Collaboration*, 1–12, https://doi.org/10.1145/3479986.3479987

Schmager, S. and Sousa, S. (2021). A Toolkit to Enable the Design of Trustworthy AI. In: Stephanidis, C. et al. (eds.), *HCI International 2021—Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence* (pp. 536–555). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-90963-5_41

Schmitz, A., Akila, M., Hecker, D., Poretschkin, M. and Wrobel, S. (2022). The why and how of trustworthy AI: An approach for systematic quality assurance when working with ML components. *At - Automatisierungstechnik*, 70(9), 793–804, https://doi.org/10.1515/auto-2022-0012

Sharp, M., Dadfarnia, M., Sprock, T. and Thomas, D. (2021). Procedural Guide for System-Level Impact Evaluation of Industrial Artificial Intelligence-Driven Technologies: Application to Risk-Based Investment Analysis for Condition Monitoring Systems in Manufacturing. *Journal of Manufacturing Science and Engineering*, 144(7), 071008–071022, https://doi.org/10.1115/1.4053155

Shestakofsky, B., and Kelkar, S. (2020). Making platforms work: Relationship labor and the management of publics. *Theory and Society*, *49*(5), 863–896. https://doi.org/10.1007/s11186-020-09407-z

Singh, S. K., Rathore, S. and Park, J. H. (2020). BlockIoTIntelligence: A Blockchain-enabled Intelligent IoT Architecture with Artificial Intelligence. *Future Generation Computer Systems*, 110, 721–743, https://doi.org/10.1016/j.future.2019.09.002

Solarte-Vasquez, M. C. and Hietanen-Kunwald, P. (2020). Transaction design standards for the operationalisation of fairness and empowerment in proactive contracting. *International and Comparative Law Review*, 20(1), 180–200, https://doi.org/10.2478/iclr-2020-0008

Solarte-Vásquez, M. C. and Nyman-Metcalf, K. (2017). Smart Contracting: A Multidisciplinary and Proactive Approach for the EU Digital Single Market. *TalTech Journal of European Studies*, 7(2), 208–246, https://doi.org/10.1515/bjes-2017-0017

Tóth, Z., Caruana, R., Gruber, T. and Loebbecke, C. (2022). The Dawn of the AI Robots: Towards a New Framework of AI Robot Accountability. *Journal of Business Ethics*, 178(4), 895–916, https://doi.org/10.1007/s10551-022-05050-z

Toy, T. (2023). Transparency in AI. *AI & SOCIETY,* 39*,* 2841–2851, https://doi.org/10.1007/s00146-023-01786-y

Uzunca, B. and Kas, J. (2023). Automated governance mechanisms in digital labour platforms: How Uber nudges and sludges its drivers. *Industry and Innovation*, 30(6), 664–693, https://doi.org/10.1080/13662716.2022.2086450

van Doorn, N. (2017). Platform labor: On the gendered and racialized exploitation of low-income service work in the 'on-demand' economy. *Information, Communication & Society*, 20(6), 898–914, https://doi.org/10.1080/1369118X.2017.1294194

Waldkirch, M., Bucher, E., Schou, P. K. and Grünwald, E. (2021). Controlled by the algorithm, coached by the crowd – how HRM activities take shape on digital work platforms in the gig economy. *The International Journal of Human Resource Management*, 32(12), 2643–2682, https://doi.org/10.1080/09585192.2021.1914129

Wang, C., Chen, J. and Xie, P. (2022). Observation or interaction? Impact mechanisms of gig platform monitoring on gig workers' cognitive work engagement. *International Journal of Information Management*, 67, 102548, https://doi.org/10.1016/j.ijinfomgt.2022.102548

Weiss, M. (2020). The platform economy: The main challenges for labour law. In: Méndez, L. M. (ed.), *Regulating the Platform Economy*. *International Perspectives on New Forms of Work* (pp. 11–20). Oxon: Routledge.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T. and He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, 135, 364–381, https://doi.org/10.1016/j.future.2022.05.014

Yan, C., Zhu, H., Korolko, N. and Woodard, D. (2020). Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)*, 67(8), 705–724, https://doi.org/10.1002/nav.21872

Zicari, R. V. et al. (2023). How to Assess Trustworthy AI in Practice. *Digital Society*, 2(3), 35, https://doi.org/10.1007/s44206-023-00063-1