Facultad de Administración y Dirección de Empresas

THE USE OF ARTIFICIAL INTELLIGENCE IN HUMAN RESOURCE DECISIONS: A LEGAL PERSPECTIVE

Carlota Navarro de la Riva

5º E-3 B

Madrid
Junio 2026

Abstract

Artificial intelligence is reshaping how organisations make decisions about their people — from screening candidates to evaluating performance and managing dismissals. This dissertation examines that transformation through a combined legal and managerial lens, asking where AI genuinely adds value to HR processes, where it generates legal risk, and what governance mechanisms organisations can implement to operate responsibly within the applicable framework.

The analysis draws on a systematic review of academic literature and six documented real-world cases — Amazon, Deliveroo Bologna, SCHUFA, Uber Amsterdam, HireVue, and Workday — each illustrating a distinct legal risk: algorithmic discrimination, fully automated decisions without genuine human oversight, and opacity. These risks are mapped against the European regulatory framework, comprising the GDPR, the EU AI Act, the LOPDGDD, and Law 15/2022, which together establish three legal barriers any organisation must respect when deploying AI in HR.

The central finding is that the risks of AI in HR are not inherent to the technology itself but to its ungoverned deployment. In every documented case, legal harm arose not because the tool was technically flawed but because governance steps were skipped before deployment. In response, this dissertation proposes a compliance protocol structured around four deployment phases — data collection and training, system design, implementation, and output review — preceded by a pre-deployment threshold question that the EU AI Act currently leaves unaddressed. The protocol translates existing legal obligations into concrete operational actions and identifies three regulatory gaps that require legislative or regulatory attention.

Keywords: artificial intelligence, human resource management, algorithmic discrimination, GDPR, EU AI Act, human oversight, compliance protocol

Index

List of Abbreviations

AI: Artificial Intelligence

AI Act: Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence

CHRO: Chief Human Resources Officer

CJEU: Court of Justice of the European Union

CV: Curriculum Vitae

DPIA: Data Protection Impact Assessment

EDPB: European Data Protection Board

EU: European Union

GDPR: General Data Protection Regulation (Regulation EU 2016/679)

HR: Human Resources

HRM: Human Resource Management

L&D: Learning and Development

LOPDGDD: Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales

ML: Machine Learning

NLP: Natural Language Processing

RPA: Robotic Process Automation

RQ: Research Question

SHRM: Strategic Human Resource Management

1. Introduction

1.1. Justification and relevance of the topic

The way organizations manage their people is changing in ways that were difficult to imagine a decade ago. Artificial intelligence has moved from a niche technological concept to an everyday business tool, and human resource management is one of the functions where its impact is most visible. Hiring decisions, performance evaluations, employee development, and even dismissal processes are increasingly being shaped or even entirely driven by algorithms. For anyone with an interest in how organizations work, and in the legal rules that govern them, this transformation raises questions that are practically urgent and worth discussing and investigating.

The scale of adoption makes this hard to ignore. Gong, Fan and Bartram (2025) document an exponential growth in AI-HRM research publications since 2019, and the organisational data points in the same direction: a Harris Poll survey cited by Budhwar et al. (2023) found that 69% of hiring managers in the United States were already using generative AI tools for HR-related tasks, while the workforce analytics market was projected to exceed one billion USD annually by 2022 (Zion Market Research, 2017, as cited in Cheng & Hackett, 2021). Yet adoption alone does not equal success: Chowdhury, Budhwar and Wood (2024) note that seven out of ten AI-related organisational projects fail — which raises an obvious question about whether companies are deploying these tools responsibly, or simply quickly.

What makes this topic particularly interesting from a business and legal standpoint is the tension it generates. On one hand, the efficiency case for AI in HR is genuinely strong: algorithms can process thousands of applications in the time it takes a recruiter to review a handful, performance data can be analyzed at a level of detail no human manager could achieve, and predictive models can identify employees at risk of leaving before they have given any explicit signal. On the other hand, these same tools carry risks that are far from trivial. The collapse of Amazon's internal recruitment algorithm in 2017 (abandoned after it was found to systematically downgrade applications from women, having been trained on a decade of historically male-dominated hiring data) is the most cited example, but it is far from isolated (Dastin, 2018).

The legal dimension adds a further layer of complexity that is often underestimated in purely technical discussions of AI. The EU's regulatory response to algorithmic decision-making has been building steadily since the General Data Protection Regulation (GDPR, Regulation EU 2016/679) came into force in 2018. Article 22 of the GDPR establishes the right not to be subject to decisions based solely on automated processing where those decisions produce legal or similarly significant effects — a provision directly applicable whenever an algorithm rejects a job candidate or generates a performance evaluation without meaningful human review. In Spain, this protection is incorporated and reinforced by Organic Law 3/2018 (LOPDGDD): Article 11 obliges employers to inform employees and candidates whenever their data will be used for profiling purposes and of their right to oppose automated decisions affecting them, Article 18 explicitly incorporates the rights established in Article 22 GDPR into the Spanish legal order, and Article 28 requires organizations to conduct a prior impact assessment whenever AI is used to predict or evaluate aspects of an employee's work performance, behaviour, or reliability. Beyond data protection, Law 15/2022 of 12 July on equal treatment and non-discrimination goes further still, explicitly prohibiting discrimination by algorithm in access to employment. Most significantly, the entry into force of Regulation (EU) 2024/1689 — the EU AI Act — on 1 August 2024 has placed HR AI systems squarely under a high-risk classification, attaching to them a set of enhanced compliance obligations — transparency, human oversight, technical documentation, and conformity assessments — and exposing non-compliant organizations to fines of up to €35 million or 7% of global annual turnover (EU AI Act, 2024, Art. 6, Annex III).

This dissertation approaches that tension from a dual perspective. Studying both Business Administration and Law makes it possible to analyse AI in HR not just as a management challenge, but also as a legal one. Neither dimension alone is sufficient. A purely managerial analysis of AI in HR risks underestimating the legal exposure it creates and a purely legal analysis risks losing sight of the business logic that drives adoption in the first place. The value of combining both is that it produces analysis with direct relevance for the people who actually have to make these decisions: HR managers, CHROs, and in-house counsel.

1.2. Research objectives and questions

The general objective of this dissertation is to examine the impact of AI adoption on human resource decision-making through a combined benefit/risk lens, grounded in both management theory and the applicable legal framework. The study aims to determine where AI genuinely adds value to HR processes, where it generates legal and organisational risks, and how companies can make the most of AI´s potential while keeping its legal risks in check.

This general objective is developed through three research questions that structure the analytical chapters of the work:

1. Descriptive question (RQ1): What is the current scope of AI adoption across the employee lifecycle? This question maps which HR decisions are today being automated or algorithmically supported, from talent acquisition through to offboarding, and at what level of intensity.
2. Explanatory question (RQ2): To what extent does AI adoption generate operational and decision-making efficiencies in HRM, from a legal and managerial perspective? This question evaluates the documented benefits of AI in HR and assesses the conditions under which they are realised.
3. Explanatory question (RQ3): To what extent does AI adoption generate legal risks in HR automation and decision-making, and what governance mechanisms can organisations implement to operate within the applicable legal framework? This question identifies the specific legal vulnerabilities created by AI deployment in HR and proposes the governance mechanisms through which they can be addressed.

1.3. Methodology

The methodology adopted in this dissertation combines two complementary approaches suited to a research question with both a normative and an empirical dimension.

The first is a systematic review of the academic literature on AI and HRM, drawing on a carefully selected body of peer-reviewed publications and primary legal sources. The bibliometric review by Gong, Fan and Bartram (2025), published in The International Journal of Human Resource Management, provides the structural backbone of the theoretical analysis, mapping the evolution of AI-HRM research from 1990 to 2024. This is complemented by Budhwar et al. (2023) on HRM in the age of generative AI, Chowdhury, Budhwar and Wood (2024) on strategic frameworks for generative AI in organisations, Malik, Budhwar and Kazmi (2023) on AI-assisted HRM, Cheng and Hackett (2021) on the critical review of algorithms in HRM, and Varma, Dawkins and Chaudhuri (2023) on the ethical assessment of AI in people management. On the legal side, the analysis draws on the primary EU and Spanish legislative texts — the GDPR, the EU AI Act, the LOPDGDD, and Law 15/2022 — together with academic commentary on their application to the employment context, including Goodman and Flaxman (2017) on algorithmic decision-making and the right to explanation, Parviainen (2022) on algorithmic recruitment under EU law, Abraha (2023) on data protection and algorithmic employment decisions, and Carter (2024) on discrimination and causal challenges in algorithmic hiring.

The second approach is an analysis of six documented real-world cases. Amazon's scrapped recruitment algorithm illustrates historical bias in candidate screening (Dastin, 2018, as cited in Carter, 2024). The Deliveroo Bologna case illustrates blind discrimination in algorithmic task allocation (Filcams VGIL Bologna and others v. Deliveroo Italia SRL, Court of Bologna, RG 2949/2019, 2020, as cited in Carter, 2024). The SCHUFA ruling by the Court of Justice of the EU clarifies the scope of Article 22 GDPR in automated scoring systems (Case C-634/21, 2023). The Uber Amsterdam case illustrates the consequences of nominal human oversight in algorithmic dismissal (Amsterdam Court of Appeal, April 2023, as documented by Eurofound, 2023). HireVue illustrates the legal risks of biometric data processing in video interview analysis (EPIC, 2019). And Workday is an ongoing class action establishing vendor liability for algorithmic discrimination at scale (Mobley v. Workday, Inc., 740 F. Supp. 3d 796, N.D. Cal. 2024). Rather than serving purely as illustrations, these cases function as empirical material through which the theoretical and normative framework developed in earlier chapters is tested against concrete organisational reality.

Together, these two approaches produce a qualitative and exploratory study — one that builds on existing evidence and documented cases to identify patterns and draw conclusions with practical relevance.

1.4. Structure of the dissertation

The dissertation is organised into six chapters that follow a coherent analytical progression. Chapter 2 develops the theoretical framework, tracing the evolution of HRM and mapping the current scope of AI across the employee lifecycle, which directly addresses RQ1. Chapter 3 presents the regulatory framework applicable to AI in the workplace — the GDPR, the EU AI Act, the LOPDGDD, and anti-discrimination law — explained in terms of their practical implications for HR management. Chapter 4 analyses the benefits and efficiencies that AI generates across HR functions, addressing RQ2. Chapter 5 analyses the legal risks created by AI adoption in HR and illustrates each risk with documented real-world cases, addressing RQ3. Finally, Chapter 6 synthesises the findings, provides direct answers to the three research questions, proposes a compliance protocol as a concrete governance recommendation for organisations, and identifies the limitations of this study together with directions for future research.

2. Theoretical framework: artificial intelligence in human resource management.

2.1. The evolution of HRM: from administrative function to strategic partner

Understanding why AI matters in human resource management requires first understanding what HRM actually is today — and how different that is from what it used to be. The function has gone through a remarkable transformation over the past century, and tracing that evolution helps explain both the appeal of AI-assisted HR and the risks that come with it.

In its earliest form, what we now call HRM was simply personnel management. It was a largely administrative operation: payroll, attendance records, compliance with labour regulations, and the orderly hiring and dismissal of workers. As Kaufman (2014) traces in his historical review of the field, HRM has its roots in the personnel management practices that emerged during industrialisation. At the time, employers and employees were locked in an increasingly tense relationship (what was then widely referred to as the "Labour Problem") and organisations began creating dedicated personnel departments largely to manage that tension. The underlying logic was transactional. Employees were inputs to a production process, and the personnel department existed to manage those inputs as efficiently as possible. Employee development and wellbeing were, at best, secondary concerns.

The shift began in the latter decades of the twentieth century. Rising employment legislation, intensified competition for skilled labour, and growing evidence from organisational research all pushed the function in a new direction. Two pioneering books published in 1984 — Fombrun, Tichy and Devanna's *Strategic Human Resource Management* and Beer et al.'s *Managing Human Assets* — are widely credited with launching strategic HRM as a recognised academic and managerial discipline (Kaufman, 2014). The central idea was straightforward but transformative: human capital is not merely a cost to be controlled, but a potential source of sustained competitive advantage. As Jackson, Schuler and Jiang (2014) define it, strategic HRM encompasses the pattern of planned human resource deployments and activities intended to enable an organisation to achieve its goals — placing people management squarely within the domain of corporate strategy rather than administrative support.

The practical consequences were significant. HR departments were no longer expected simply to administer employment contracts. They were expected to attract, develop and retain talent in ways that directly served the organisation's strategic objectives. Workforce planning, talent management, performance management, and employee engagement all became core HR responsibilities. The CHRO became a regular presence in the executive team alongside the CEO and CFO. As Malik, Budhwar and Kazmi (2023) observe, this strategic elevation created the conditions for AI adoption in HR: once the function was understood as data-rich and decision-intensive with direct impact on organisational performance, the case for algorithmic tools became compelling.

It is worth noting, however, that this elevation also raised the stakes. When HR decisions (who to hire, how to evaluate performance, who to promote, who to dismiss) carry strategic weight, the consequences of getting them wrong are correspondingly serious. This is precisely the context in which algorithmic decision-making introduces both its greatest potential value and its most significant legal risk.

2.2. Artificial intelligence in organisations: types and applications

The term 'artificial intelligence' covers a broad and heterogeneous set of technologies, so it is worth being precise about what it means in an organisational context. For the purposes of this dissertation, AI is understood as a machine-based system that, for a given set of objectives, infers how to generate outputs — predictions, recommendations, decisions, or content — that influence real or virtual environments (EU AI Act, 2024, Art. 3(1)).

Within that broad category, Gong, Fan and Bartram (2025) show that AI in HR has evolved through distinct stages, from early expert systems in the 1990s designed to assist HR managers in candidate selection and scheduling, through the rise of big data analytics in the 2010s, to the sophisticated tools of today. Understanding that trajectory helps make sense of how varied the current landscape actually is.

At the more foundational level, machine learning algorithms identify patterns in large datasets and improve their outputs through experience — the technology behind CV screening tools and predictive attrition models. Deep learning, a more advanced branch, processes complex unstructured inputs through neural networks, enabling applications such as automated video interview analysis and speech recognition in candidate screening (Rodgers et al., 2023). Natural language processing allows machines to interpret and generate human language, powering CV reading tools, candidate chatbots, and sentiment analysis tools applied to employee surveys (Budhwar et al., 2023). Computer vision, used in video interview platforms to analyse facial expressions and non-verbal cues, sits at the more controversial end of the spectrum — a point returned to in Chapter 5.

People analytics platforms aggregate data from multiple organisational sources to generate performance dashboards and workforce forecasts. Cheng and Hackett (2021) note that the workforce analytics market was already projected to exceed one billion USD annually by 2022, reflecting the scale of organisational investment in data-driven HR. Algorithmic management takes this further: as Gong et al. (2025) document, particularly in gig economy contexts, algorithms now assign tasks, monitor performance in real time, and apply consequences automatically — with minimal or no human involvement. Generative AI represents the most recent and perhaps most disruptive development. Unlike earlier tools that predict or classify, generative AI creates original content in response to prompts. As Budhwar et al. (2023) document, it is already being used in HR to draft job advertisements, generate performance review templates, and produce personalised learning content at scale. Chowdhury, Budhwar and Wood (2024) note that this capability opens genuine new possibilities for HR personalisation — but also introduces new risks around accuracy, bias and accountability that earlier AI generations did not pose in the same way.

Finally, it is worth distinguishing all of the above from robotic process automation (RPA), which executes repetitive administrative tasks by following fixed, pre-programmed instructions. RPA cannot learn or adapt. This distinction matters legally: the EU AI Act regulates AI systems specifically and does not extend to rule-based automation, so HR departments need to be clear about which of their tools actually fall within the regulatory perimeter.

Not all AI tools in HR operate at the same level of decision-making autonomy, and it is worth being clear about that distinction. Kellogg, Valentine and Christin (2020, as cited in Gong et al., 2025) identify a spectrum ranging from algorithms that recommend and support human decisions to those that replace them entirely. At the most autonomous end sit algorithmic management systems, which assign tasks, monitor performance and apply consequences in real time with no human involvement whatsoever. At the other end, tools like people analytics dashboards or generative AI content generators are designed to inform human judgement rather than substitute it. Most other AI types — machine learning scoring models, NLP-based screening tools, deep learning interview analysis — fall somewhere in between. They are formally positioned as decision-support tools, but in practice they can end up determining outcomes when human reviewers lack the time or incentive to meaningfully question what the algorithm has produced.

This is not a minor nuance. As Cheng and Hackett (2021) warn, the line between supporting a decision and making one is thinner than organisations tend to assume. And as Goodman and Flaxman (2017) argue, it is precisely that line — whether a human was genuinely involved in the decision or simply confirmed what the algorithm had already determined — that determines whether a system falls within the prohibition established by Article 22 of the GDPR. Understanding this spectrum of autonomy is essential before turning to the legal framework that governs it.

2.3. The scope of AI across the employee lifecycle

The clearest way to understand how far AI has penetrated HR is to trace it across the employee lifecycle: from the moment a candidate first applies to the moment they eventually leave. Gong, Fan and Bartram (2025) confirm that AI applications now exist at every single stage of that journey.

2.3.1. Talent acquisition and recruitment

Recruitment is where AI is most deeply embedded. Applicant tracking systems read and rank CVs automatically, chatbots handle initial candidate screening, and video interview platforms score candidates on verbal and non-verbal cues — all before a human recruiter has necessarily seen anything. Pessach et al. (2020, as cited in Gong et al., 2025) show that these systems genuinely reduce time-to-hire and cut costs by automating the most data-intensive parts of the selection process. The problem, as Cheng and Hackett (2021) warn, is that algorithms trained on historical data tend to replicate past hiring patterns — which in organisations with homogeneous workforces can mean systematically screening out candidates from underrepresented groups, not intentionally, but as a direct consequence of how the algorithm was trained.

2.3.2. Performance evaluation and people analytics

Once inside the organisation, employees are increasingly subject to AI-driven evaluation. People analytics platforms aggregate data from productivity tools, communication systems and engagement surveys to generate performance dashboards and flag patterns human managers might miss. Kellogg, Valentine and Christin (2020, as cited in Gong et al., 2025) document how predictive models can estimate an employee's likelihood of leaving months before they show any explicit sign of wanting to go, giving HR teams time to intervene. Useful as this is, the data collection it requires raises significant questions under the GDPR and the LOPDGDD, addressed in Chapter 3.

2.3.3. Learning and development

AI has made personalised learning possible at scale. Where traditional learning and development delivered the same programme to everyone, AI-powered platforms analyse each employee's skills, role and learning history to recommend tailored content. Budhwar et al. (2023) note that generative AI can now produce individualised career development pathways at a level of detail previously only possible through one-to-one coaching. Chowdhury, Budhwar and Wood (2024) add a note of caution: over-reliance on algorithmic recommendations risks reducing the human judgement and genuine manager-employee relationships that development programmes also depend on.

2.3.4. Compensation, engagement and workforce planning

AI is also present in compensation, where algorithms benchmark salaries and flag pay equity gaps, and in employee engagement, where sentiment analysis tools process survey responses in real time. In workforce planning, AI models forecast future headcount and skills needs. Perhaps the most striking application, however, is in job assignment and task allocation — particularly in gig economy platforms, where algorithmic management systems assign work, set deadlines, monitor performance and apply penalties entirely automatically, with no human manager in the loop. As Kellogg, Valentine and Christin (2020, as cited in Gong et al., 2025) document, this represents the most extreme form of AI-driven HR: the algorithm does not just support the decision, it is the decision. Taken together, Gong, Fan and Bartram (2025) describe all of this as a fundamental shift in HR from a reactive, judgement-based function to a predictive, data-driven one. The benefits of that shift are real. But so are the legal obligations and risks it generates — which is what the next two chapters are about.

3. Regulatory framework: the legal boundaries of AI in HR decision-making

3.1. Overview

The previous chapter mapped what AI does in HR. This chapter maps what the law says about it. The European regulatory framework applicable to AI in HR has been built in layers over the past decade, each responding to a different dimension of the same problem. Together, four instruments define what organisations can and cannot do when they use AI to make or support HR decisions. This chapter examines each one and closes by identifying, concretely, the legal barriers they establish.

3.2. GDPR Article 22: the right not to be subject to automated decisions

The General Data Protection Regulation (Regulation EU 2016/679), in force since 2018, is the foundational instrument. Article 22(1) establishes that individuals have the right not to be subject to a decision based solely on automated processing — including profiling — when that decision produces significant legal or similarly significant effects on them. In HR, this applies directly to automated decisions about hiring, dismissal, promotion, or performance evaluation.

Exceptions exist — the decision may be automated if necessary to perform a contract, authorised by law, or based on explicit consent — but even then the organisation must guarantee the individual's right to human intervention and to contest the outcome (GDPR, 2016, Art. 22(2)-(3)). As Goodman and Flaxman (2017) explain, Article 22 effectively creates a right to explanation: individuals are entitled to understand the logic behind decisions that affect them.

A key clarification came from the Court of Justice of the EU in SCHUFA (Case C-634/21, 2023), which confirmed that Article 22 applies even when a human formally signs off, if that person is simply confirming what the algorithm already determined. As Parviainen (2022) notes, this sets a high bar — a recruiter approving an algorithmically ranked shortlist without independently reviewing candidates does not satisfy the requirement. Article 35 adds that any systematic automated evaluation of personal aspects requires a prior Data Protection Impact Assessment (DPIA).

3.3. LOPDGDD: Spain's extension of GDPR protections

Organic Law 3/2018 (LOPDGDD) adapts the GDPR to Spanish law and extends it in three areas directly relevant to HR. Article 11 requires organisations to proactively inform individuals when their data will be used for profiling and of their right to oppose any automated decision affecting them. Article 18 formally incorporates Articles 21 and 22 of the GDPR into the Spanish legal order. Article 28 requires a documented risk assessment before deploying any tool that predicts or evaluates employee performance, behaviour, or reliability.

3.4. The EU AI Act: high-risk classification and its obligations

Regulation (EU) 2024/1689 — the EU AI Act — entered into force on 1 August 2024 and is the most far-reaching development in this space. It classifies AI systems according to the risk they pose, and under Article 6 and Annex III (point 4), AI systems used for recruitment, candidate evaluation, performance management, promotion, or termination are classified as high-risk.

This classification is not symbolic. It triggers binding obligations: risk management throughout the system's lifecycle, data quality controls, technical documentation, operational logging, transparency towards regulators and affected individuals, meaningful human oversight, and robustness requirements (EU AI Act, 2024, Arts. 9–15). Non-compliance carries fines of up to €35 million or 7% of global annual turnover (EU AI Act, 2024, Art. 99).

The Act also introduced a direct prohibition relevant to HR: since February 2025, emotion recognition and biometric categorisation systems in employment contexts are banned (EU AI Act, 2024, Art. 5) — with direct implications for video interview platforms that infer personality from facial expressions, examined in Chapter 5. As Sele and Chugunova (2024) document empirically, the human oversight requirement is not merely procedural — it responds to a real and documented tendency for reviewers to defer to algorithmic outputs even when they formally have the authority to override them.

3.5. Law 15/2022: the prohibition of algorithmic discrimination

Law 15/2022 of 12 July on equal treatment and non-discrimination is Spain's direct response to discriminatory AI in employment. Article 23 specifically addresses AI and automated decision-making, requiring bias minimisation, transparency, and impact assessments to identify potential discriminatory effects.

Critically, the law prohibits not only direct discrimination but also indirect discrimination. As Goodman and Flaxman (2017) demonstrate, an AI system can produce discriminatory outcomes without using protected characteristics as inputs — because variables like postcode, educational institution, or employment gaps may be correlated with race, gender, or socioeconomic origin. It is not legally sufficient to argue that a tool does not explicitly use protected data if its outputs disproportionately disadvantage protected groups.

3.6. The three legal barriers

The four instruments together establish three concrete obligations that any organisation using AI in HR must respect.

The first barrier is that any HR decision that significantly affects an individual — hiring, dismissal, promotion, performance evaluation — must involve genuine human review, not just formal sign-off. A human who simply confirms what the algorithm already determined does not satisfy this requirement (GDPR Art. 22; SCHUFA, C-634/21, 2023; Parviainen, 2022).

The second barrier is transparency and the right to contest. Individuals must be proactively informed when AI is being used to make or support decisions about them, told the logic behind those decisions, and given a real opportunity to challenge the outcome. Silence or buried terms and conditions do not satisfy this obligation (GDPR Arts. 13–14; LOPDGDD Arts. 11, 18; Goodman & Flaxman, 2017).

The third legal barrier is no discriminatory outcomes, direct or indirect. AI tools must not produce outcomes that disproportionately disadvantage protected groups — even if they do not explicitly use protected characteristics as inputs. Active bias testing is legally required, not optional (Law 15/2022, Art. 23; GDPR Art. 9; Goodman & Flaxman, 2017).

Demonstrating compliance with all three requires prior documentation — a risk assessment conducted before the tool is deployed, maintained throughout its use, and available to regulators (GDPR Art. 35; LOPDGDD Art. 28; EU AI Act Arts. 9–15). This is not a fourth barrier but the procedural mechanism through which the three above are proven in practice.

4. The benefits of AI in HR decision-making

4.1. Efficiency and cost reduction

The most immediately visible benefit of AI in HR is operational efficiency. AI tools can process volumes of data and complete screening tasks at a speed and scale that is simply not achievable by human teams. Pessach et al. (2020, as cited in Gong et al., 2025) demonstrate that machine learning-based recruitment systems add measurable value by automating the most data-intensive parts of the selection process — tasks that previously consumed significant recruiter time. The result is a faster time-to-hire, lower recruitment costs, and the ability to reach a wider candidate pool without proportionally increasing headcount.

This efficiency gain extends beyond recruitment. In performance management, AI analytics platforms can process data from multiple sources simultaneously and surface patterns that would take human analysts considerably longer to identify. In L&D, automated systems can deliver personalised training at scale — something that previously required significant investment in individual coaching. As Malik, Budhwar and Kazmi (2023) note, it is precisely this capacity to deliver individualised HR practices at scale that makes AI strategically attractive to organisations: things that were once only possible for senior employees or high-potential groups can now, in principle, be extended to the entire workforce.

4.2. Objectivity and consistency

A second frequently cited benefit is that AI can reduce the role of human bias in HR decisions. Human decision-making in hiring and performance evaluation is well documented to be affected by factors that have nothing to do with job performance — the name on a CV, the interviewer's mood, the candidate's physical appearance, or simple cognitive fatigue (Cheng & Hackett, 2021). An algorithm, applied consistently, evaluates every candidate or employee against the same criteria every time.

Gong, Fan and Bartram (2025) document this as the effectiveness orientation of AI in HR: the technology improves the accuracy and objectivity of HR decisions by removing individual variation from the evaluation process. Pessach et al. (2020, as cited in Gong et al., 2025) show that AI tools can also integrate internal performance data with external labour market benchmarks, giving decision-makers a more complete and comparable picture than any individual manager could assemble manually.

It is important to be precise here, however. This objectivity is conditional. As Cheng and Hackett (2021) warn, an algorithm is only as objective as the data it was trained on — and if that data reflects historical patterns of bias, the algorithm will reproduce them systematically. The benefit of objectivity is real, but it depends entirely on the quality and fairness of the underlying data and on ongoing bias monitoring. Without those conditions, what looks like objectivity is actually bias at scale.

4.3. Predictive capability and proactive decision-making

Perhaps the most strategically significant benefit of AI in HR is its predictive capability — the ability to anticipate future outcomes rather than simply react to current ones. Predictive attrition models can identify employees at risk of leaving months before they show any explicit sign of wanting to go, giving HR teams time to intervene with targeted retention measures. Skills gap analysis tools can identify where the workforce will be underequipped for future business needs, enabling proactive investment in training rather than reactive crisis management. Succession planning algorithms can identify internal candidates for leadership roles based on performance trajectories rather than managerial intuition.

As Kellogg, Valentine and Christin (2020, as cited in Gong et al., 2025) document, this shift from reactive to predictive HR represents a fundamental change in how the function operates. Budhwar et al. (2023) note that generative AI extends this further — tools can now not only predict future needs but generate the content, plans, and recommendations needed to address them, from personalised development pathways to tailored onboarding programmes.

4.4. The condition: legal sustainability

The benefits described above are real and well documented. But they are only genuinely beneficial — from both a managerial and a legal perspective — when the three barriers identified in Chapter 3 are respected.

Without genuine human oversight, transparency towards affected individuals, and active bias monitoring, each efficiency gain carries a corresponding legal exposure that can offset it entirely.

The point is not that these tools should not be used. It is that the efficiency gain only translates into a net organisational benefit when the legal framework is respected. When it is not, the operational gain is offset — and frequently exceeded — by the legal, reputational and financial cost of non-compliance. As Chowdhury, Budhwar and Wood (2024) observe, seven out of ten AI-related organisational projects fail. Legal and governance failures are among the most common reasons why.

5. Legal risks of AI in HR decision-making

5.1. Overview

The previous chapter showed where AI genuinely adds value in HR. This one shows where that value comes at a legal cost. The risks examined here are not hypothetical, they have produced real court rulings, regulatory sanctions, and ongoing litigation. Each one maps onto the barrier framework established in Chapter 3, and each is illustrated by a documented case. Three distinct legal risks emerge: algorithmic discrimination, fully automated decisions without genuine human oversight, and opacity. Together they define the outer limits of what organisations can do with AI in HR without crossing into sanctionable territory.

5.2. Algorithmic discrimination

Discrimination is the most extensively documented legal risk of AI in HR, and it arises in distinct but related ways. The first is historical and representation bias, where an algorithm is trained on historical data that already reflects past discriminatory patterns or underrepresented groups and replicates them automatically. Another type of bias is technical and emergent bias, where bias arises from design constraints and the incapability for algorithms to adapt to changing societal norms like Köchling and Wehner (2020) mention.

Additionally there is blind discrimination, where the system produces discriminatory outcomes without anyone having noticed or intended it, because nobody was monitoring the outputs closely enough. Moreover, there is proxy discrimination, where apparently neutral variables such as postcode, university attended, or employment gaps produce outcomes that disproportionately disadvantage protected groups. As Goodman and Flaxman (2017) demonstrate, removing protected characteristics (like race, gender, age) from a dataset does not prevent this, because other variables (like postcode, employment gaps, college attended) may be statistically correlated with those characteristics and produce the same discriminatory effect through an indirect route.

This dissertation focuses on historical, proxy, and blind discrimination, and what connects all three is a common root: insufficient human judgement at the design and deployment stage. The algorithm was trained on biased data, deployed without adequate bias testing, and allowed to operate without ongoing monitoring — which is exactly what Barrier 3 and the procedural documentation requirement are designed to prevent.

Amazon's internal recruitment tool, abandoned in 2017, is the most cited example of historical bias in practice. The system was trained on ten years of CV data from a predominantly male technical workforce and learned over time to penalise CVs containing words associated with women ("women's college", "women's chess club"), while downgrading graduates of all-female universities. The bias was not intentional. It emerged directly from training data that reflected a decade of historically male-dominated hiring decisions, and Amazon could not guarantee the system would not find other discriminatory proxies even after attempting to correct it (Dastin, 2018, as cited in Carter, 2024).

Deliveroo Bologna illustrates blind discrimination in a different context. Deliveroo's algorithm scored drivers on participation and reliability, penalising those who had cancelled bookings or been less active, regardless of why. Drivers who had taken legitimate absences for illness, disability, or family reasons were treated identically to those who had simply chosen not to work. The Bologna Labour Tribunal found this to be indirectly discriminatory towards protected groups in 2020 (Filcams VGIL Bologna and others v. Deliveroo Italia SRL, Court of Bologna, RG 2949/2019, 2020, as cited in Carter, 2024). Notably, Deliveroo's refusal to explain how the algorithm worked was used against it by the court. A clear signal that opacity does not protect organisations from liability but actively compounds it.

The Workday case shows proxy discrimination operating at a scale that makes it qualitatively different from individual human bias. Derek Mobley's 2023 class action alleged that Workday's AI screening tools discriminated against him on grounds of race, age, and disability across more than 100 job applications, using variables such as postcode, university attended, and group membership, variables that correlate with protected characteristics without using them explicitly (Mobley v. Workday, Inc., N.D. Cal., 2023). In July 2024, the court ruled that an AI vendor can be directly liable as an agent of employers for discriminatory outcomes its tools produce, even when it is not itself the employer (Mobley v. Workday, Inc., 740 F. Supp. 3d 796, N.D. Cal. 2024). Workday processed over 266 million applications in 2023 alone. As Köchling and Wehner (2020) highlight, this is what makes algorithmic bias so uniquely dangerous, because the same model processes every candidate, a single embedded bias does not affect one person, but millions, at a scale no human decision-maker could ever reach.

The common thread across all three forms of discrimination is the absence of adequate human judgement before and during deployment. But there is a deeper regulatory gap that the cases expose: the law prohibits discriminatory outcomes, yet it does not require organisations to proactively demonstrate that their algorithm does not discriminate before deploying it. Bias is discovered after the damage has been done, Amazon only detected the problem a year after launch, Deliveroo only after litigation was brought. And when affected individuals try to claim redress, they face the additional obstacle that accessing the algorithm's inner workings to prove the discrimination is practically impossible without the organisation's cooperation. Part of the difficulty lies in how bias enters the system in the first place. As Goodman and Flaxman (2017) demonstrate, simply removing protected characteristics from the dataset is not enough, because other variables may be statistically correlated with those characteristics and produce the same discriminatory effect through an indirect route. The challenge of selecting variables that are genuinely neutral, rather than apparently neutral, is technically complex and has no straightforward solution. As Köchling and Wehner (2020) note, this means the legal protection against algorithmic discrimination is largely reactive rather than preventive, which, given the scale at which these systems operate, is a significant structural weakness in the current framework.

5.3. Fully automated decisions and the human oversight gap.

GDPR Article 22 prohibits decisions based solely on automated processing when those decisions significantly affect an individual. In HR this covers hiring, dismissal, promotion, and performance evaluation. The provision is not an absolute ban, because exceptions exist, but even when an exception applies, organisations must guarantee genuine human intervention, the right to express one's point of view, and the right to contest the outcome (GDPR, 2016, Art. 22(3)). The operative word is genuine. As Abraha (2023) documents in his analysis of the GDPR's application to algorithmic management, many organisations formally comply with this requirement while substantively violating it. This can happen by placing a human in the process who reviews an algorithmically generated output for a matter of seconds and approves it without independent evaluation. As Sele and Chugunova (2024) demonstrate empirically, this tendency is not simply negligence: human reviewers consistently defer to algorithmic recommendations even when they formally have the authority to override them, because the algorithm's apparent objectivity creates a psychological tendency to trust its output over one's own judgement.

The SCHUFA case resolved the core legal question here definitively. SCHUFA, a German credit reference agency, generated automated creditworthiness scores that banks relied on to make lending decisions, and argued it was not subject to Article 22 because the bank, not the algorithm, made the final decision. The CJEU rejected this in December 2023, finding that Article 22 applies whenever an algorithmically generated score plays a determining role in a downstream decision, regardless of who formally approves it (CJEU, Case C-634/21, 2023). Although this case involves credit rather than employment, its implications for HR are direct. As Carter (2024) and Abraha (2023) both note, the same logic applies to any HR algorithm. Things like a candidate ranking, a performance score, an attrition prediction that effectively determines an outcome even when a human manager formally signs off, is an automated decision that requires human revision.

Uber Amsterdam then applied the same principle to an employment context and showed what the consequences look like in practice. Uber was automatically deactivating driver accounts through algorithms flagging poor performance or suspected fraud, and argued it had humans reviewing those flags before any deactivation. The Amsterdam Court of Appeal rejected this in April 2023, finding those reviews were on the facts "not much more than a purely symbolic act" (Amsterdam Court of Appeal, April 2023, as documented by Eurofound, 2023). When Uber failed to comply with the ruling, the court ordered €584,000 in penalties in October 2023, with €4,000 accruing per day of continued non-compliance.

Both cases expose the same unresolved regulatory gap: what does genuine human involvement actually look like in practice? The GDPR and the AI Act both require it, but neither defines it with operational precision. How much time must a reviewer spend? What criteria must they apply? Must they be able to explain their decision independently of the algorithm? As Abraha (2023) documents, courts and data protection authorities across the EU have reached conflicting answers, producing inconsistent levels of protection across member states. This is one of the most significant open questions in the current framework — and one the protocol proposed in the following chapter seeks to address directly.

5.4. Opacity and the right to explanation

The third legal risk is opacity: the inability of organisations to explain, in terms meaningful to the person affected, how an algorithmic HR decision was reached. This violates the transparency obligations of GDPR Articles 13–14 and LOPDGDD Article 11, and it undermines the practical enforceability of the rights established under Article 22. GDPR Article 15(1)(h) gives workers the right to access meaningful information about the logic involved in automated decisions affecting them, but as Abraha (2023) analyses in detail, what counts as meaningful is genuinely contested. Information that is too generic does not satisfy the requirement. Neither does information that is too technical for a non-specialist to understand. The EDPB has stated that meaningful information must be sufficient for the individual to understand the reasons for the decision and to exercise their rights, but the precise threshold remains ambiguous. As Cheng and Hackett (2021) document, HR algorithms are widely described by practitioners as black boxes precisely because their probabilistic, multi-variable logic is not easily translatable into a causal explanation a human can understand or a court can evaluate. Goodman and Flaxman (2017) argue that this is not merely a technical inconvenience but a fundamental legal problem: a company that cannot explain why its algorithm ranked a candidate last cannot satisfy its transparency obligations under the GDPR.

The deeper structural problem is that intellectual property and trade secret protections frequently limit what organisations are required to disclose about their algorithms. As Abraha (2023) documents, this creates a tension the current framework has not resolved: the law requires meaningful transparency, but organisations can invoke commercial confidentiality to resist disclosure, leaving affected individuals without the information they need to exercise their rights. Carter (2024) identifies this as one of the central obstacles to enforcing discrimination law against algorithmic HR systems, without access to the algorithm's logic, complainants cannot construct the causal explanation required to prove their case in court.

HireVue illustrates what opacity looks like in practice — and, unusually, how the regulatory framework has since closed the risk. HireVue's video interview platform generated employability scores from candidates' facial expressions and biometric data without any meaningful explanation of how those signals were used or weighted (EPIC, 2019). Candidates had no way of understanding why they had been rejected, let alone of contesting the outcome — a direct violation of the transparency obligations under GDPR Articles 13–14 and the right to explanation under Article 15(1)(h). Under GDPR Article 9, the processing of biometric data also required explicit legal justification that most employers deploying the tool had not met. The EU AI Act resolved this directly: since February 2025, emotion recognition systems in employment contexts are outright prohibited regardless of legal basis (EU AI Act, 2024, Art. 5). HireVue's original model would be flatly illegal in the EU today — making it, unlike the other cases examined in this chapter, an example of a risk the law has resolved rather than merely identified.

Opacity, in other words, is not just a legal risk in itself, it is what makes all the other legal risks harder to enforce against.

6. Conclusions

6.1. Answers to the research questions

This dissertation set out to examine the impact of AI adoption on HR decision-making through a combined legal and managerial lens. The three research questions that structured the analysis can now be answered directly.

The first research question was, what is the current scope of AI adoption across the employee lifecycle?

As established in the theoretical framework developed in section 2.3, AI has penetrated every stage of the employee lifecycle. As Gong, Fan and Bartram (2025) confirm in their bibliometric review, applications exist from talent acquisition — where algorithms screen CVs, rank candidates, and conduct automated video interviews — through performance evaluation, learning and development, compensation, and workforce planning, all the way to dismissal decisions in gig economy contexts where algorithmic management systems operate without any human involvement. The EU AI Act classifies all of these applications as high-risk under Article 6 and Annex III, which means the entire employee lifecycle falls within the regulatory perimeter. This is precisely why the compliance protocol proposed in section 6.3 is structured around the deployment phases of the AI tool rather than the stages of the employee lifecycle — because all stages are equally subject to the same legal obligations.

The second research question is to what extent does AI adoption generate operational and decision-making efficiencies in HRM?

The efficiency case for AI in HR is genuine and well documented. Machine learning systems reduce time-to-hire and lower recruitment costs by automating the most data-intensive screening tasks (Pessach et al., 2020, as cited in Gong et al., 2025). People analytics platforms surface workforce patterns at a level of granularity no human manager could achieve manually, enabling proactive retention and succession decisions (Kellogg, Valentine & Christin, 2020, as cited in Gong et al., 2025). Generative AI democratises personalised learning and development by producing individualised career pathways at scale previously only achievable through one-to-one coaching (Budhwar et al., 2023). And AI-driven compensation analysis identifies pay equity gaps that manual benchmarking routinely misses. These benefits are real. But as Chapter

4 established, they are only genuinely beneficial — from both a managerial and a legal perspective — when the three barriers identified in Chapter 3 are respected.

Lastly, the third research question was to what extent does AI adoption generate legal risks in HR automation and decision-making, and what governance mechanisms can organisations implement to operate within the applicable legal framework?

The benefits described above are only genuinely beneficial when the three barriers identified in Chapter 3 are respected. Without genuine human oversight, transparency towards affected individuals, and active bias monitoring, each efficiency gain carries a corresponding legal exposure that can offset it entirely. As Chowdhury, Budhwar and Wood (2024) observe, seven out of ten AI-related organisational projects fail — and legal and governance failures are among the most common reasons why.

6.2. The benefit/risk balance

The analysis conducted across Chapters 4 and 5 produces a clear overall picture: AI in HR generates genuine operational value, but that value is conditional. It is conditional on whether the organisation has governed the tool's deployment in a way that respects the legal framework — and the cases examined in Chapter 5 show that most organisations that got into legal trouble did not fail because their tools were technically flawed. They failed because they skipped the governance steps that would have caught the problem before it caused harm.

Amazon did not intend to discriminate against women. Deliveroo did not intend to penalise disabled drivers. Workday did not design its system to exclude older workers. In every case the harm emerged from the absence of adequate oversight at the right moment — before the system was deployed, not after. This is the central finding of the benefit/risk analysis: the risks are not inherent to AI itself, they are inherent to ungoverned AI. A tool that is genuinely efficient and a tool that is legally compliant are not opposites — but achieving both requires deliberate action at each stage of the deployment process.

6.3. A compliance protocol for AI deployment in HR

The protocol proposed here is addressed to HR managers, CHROs, and in-house counsel who are responsible for deciding whether and how to implement AI tools in their organisations. It is structured around four deployment phases of the AI tool — rather than the stages of the employee lifecycle — because, as established in RQ1, the EU AI Act classifies all HR AI applications as high-risk, meaning the same compliance obligations apply regardless of which HR function is involved.

The protocol draws on the deployment framework implicit in EU AI Act Articles 9–15, the bias typology of Köchling and Wehner (2020) and Raghavan et al. (2020), the autonomy spectrum identified by Kellogg, Valentine and Christin (2020, as cited in Gong et al., 2025), the empirical findings of Sele and Chugunova (2024) and Rosenthal-von der Pütten and Sach (2024) on human oversight behaviour, and the transparency analysis of Abraha (2023) and Goodman and Flaxman (2017).

The protocol proposed here does not merely restate existing legal obligations — it translates them into operational steps and fills three governance gaps that the current framework leaves open: the absence of a pre-deployment decision threshold, the lack of an operational autonomy classification, and the unresolved distribution of responsibility between deployers and vendors. Before the four deployment phases are engaged, a pre-deployment threshold question is proposed as Phase 0 — addressing a gap the EU AI Act leaves open by not establishing whether deployment should proceed at all.

a. Phase 0: Pre-deployment threshold

Before any of the four phases below are engaged, organisations must first ask a more fundamental question: should this AI system be deployed at all? The EU AI Act classifies HR AI as high-risk and establishes what must be done once deployment is decided — but it does not establish a threshold for deciding whether deployment is appropriate in the first place. This gap is identified in recent literature: as Haque (2025) observes, existing AI regulations provide only indirect oversight of recruitment and fail to address the specific ethical risks of algorithmic hiring, making sector-specific frameworks necessary.

This dissertation proposes that before signing any vendor contract, the responsible CHRO or in-house counsel should be able to answer yes to three questions: Does the organisation have the human oversight capacity to genuinely review algorithmic outputs, not just formally approve them? Is there sufficient confidence that the training data is representative of the candidate or employee population the system will evaluate? And can the organisation explain, in plain language, how the system influences decisions affecting individuals? If the answer to any of these is no, deployment should not proceed — not because the law explicitly prohibits it, but because the conditions for legal compliance cannot be met.

b. Phase 1 Data: collection and training

This is where bias enters the system — before anyone has used it and before anyone has been harmed by it. As Köchling and Wehner (2020) document, historical bias, representation bias, technical bias, and emergent bias all originate in the training data. An algorithm trained on a decade of homogeneous hiring decisions will learn to replicate those patterns with mathematical precision, as Amazon discovered to its cost.

The compliance obligation at this phase, grounded in EU AI Act Article 10, is data governance — and it needs to happen before the contract is signed, not after the tool is already running. In practice this means four concrete steps. First, request the vendor's training data documentation: what historical datasets were used, and were protected groups adequately represented in them? As Raghavan et al. (2020) identify, this is the foundational question of any bias assessment — and if the vendor cannot answer it, that is already a red flag. Second, require an independent bias audit as a condition of purchase — not a vendor self-assessment, but an external audit with documented impact ratios across protected groups. This is precisely what New York City's Local Law 144 (2023) mandates for automated employment decision tools: independent auditing before deployment, with published results. Third, conduct a Data Protection Impact Assessment under GDPR Article 35 and LOPDGDD Article 28 before the tool goes live, documenting the risks identified and the measures taken to address them. Fourth, keep all of this documentation available for regulators, as EU AI Act Articles 9 and 11 require for high-risk systems. If the vendor cannot support any of these steps, the system should not be deployed — under EU AI Act Articles 9 and 10, the deployer shares responsibility for data governance and cannot rely solely on the vendor's assurances.

A fifth step addresses a risk that the current framework has identified but not resolved: vendor liability. As Mobley v. Workday (2024) establishes, an AI vendor can be directly liable as an agent of the employer for discriminatory outcomes — but as Jones Walker (2025) documents, most vendor contracts simultaneously cap liability, exclude compliance warranties, require the employer to indemnify the vendor against discrimination claims, and limit audit rights that would allow the employer to examine the algorithm for bias. The deploying organisation therefore bears legal responsibility for outcomes it cannot examine, using training data it cannot audit. To address this gap, this dissertation proposes that any vendor contract for a high-risk HR AI system should include as minimum conditions: an independent bias audit clause with published impact ratios across protected groups, explicit compliance warranties covering anti-discrimination obligations under Law 15/2022 and GDPR Article 9, shared liability provisions for discriminatory outcomes, and unrestricted audit access to training data documentation upon regulatory request. Without these contractual protections, the organisation assumes full legal exposure for a system it did not build and cannot fully inspect.

c. Phase 2: System design

This is where the organisation decides how much of the HR decision the algorithm will make and how much will be left to human judgement. As Kellogg, Valentine and Christin (2020, as cited in Gong et al., 2025) identify, there are three fundamentally different configurations: replacing, where the algorithm decides without human involvement; restricting, where the algorithm filters and the human only sees what the algorithm permits; and recommending, where the algorithm scores but the human retains full visibility and decision authority.

Replacing and restricting both violate Barrier 1 — they should simply not be used for any HR decision that significantly affects an individual. Only recommending is potentially compliant, but with a critical caveat. Sele and Chugunova (2024) demonstrate empirically that human reviewers follow algorithmic recommendations more closely than they should, and are least likely to intervene precisely when the algorithm is least accurate. Rosenthal-von der Pütten and Sach (2024) found that 60% of reviewers in a hiring context did not detect algorithmic bias even when it was present. Choosing a recommending configuration is necessary — but it is not sufficient on its own.

The compliance obligation at this phase, grounded in EU AI Act Article 14, comes down to one concrete action: configure the system to show the reviewer all candidates, not just those the algorithm ranked highly. A reviewer who only sees the top ten never has the opportunity to catch what the algorithm filtered out — and a system that structurally prevents that review is not a recommending system, it is a restricting system in disguise. A fair objection at this point is that requiring the reviewer to see all candidates appears to defeat the purpose of AI adoption — if the reviewer must look at everything anyway, where is the efficiency gain? The answer is that the algorithm still does the most time-consuming work: reading, parsing, and scoring hundreds of applications. What the reviewer does is exercise final judgement on a preorganised, pre-scored pool — which is substantially faster than starting from scratch, while still satisfying the genuine oversight requirement. As Pessach et al. (2020, as cited in Gong et al., 2025) document, the efficiency gains of AI in recruitment come from automating the data-intensive screening stage — not from removing the human from the decision.

To operationalise this distinction, this dissertation proposes a three-tier autonomy classification for HR AI tools. Green-tier systems operate in a recommending configuration and present the reviewer with full candidate or employee visibility — these are potentially compliant with Barrier 1 subject to the Phase 4 requirements below. Amber-tier systems restrict visibility or weight algorithmic outputs so heavily that genuine override is practically unlikely — these require structural redesign before deployment. Red-tier systems replace human judgement entirely — these should not be used for any HR decision that significantly affects an individual, regardless of efficiency gains. This classification is not derived from the EU AI Act, which does not define autonomy tiers operationally, but from the empirical evidence of Sele and Chugunova (2024) and Kellogg, Valentine and Christin (2020, as cited in Gong et al., 2025) on how human oversight actually behaves in practice.

d. Phase 3: Implementation.

This is the first phase in which the tool affects real people — candidates submitting applications, employees being evaluated or monitored. Before any individual is subject to an AI-assisted HR decision, they must be informed. As Abraha (2023) analyses in detail, GDPR Articles 13–14 and LOPDGDD Article 11 require organisations to proactively disclose that AI is being used, what data is being processed, for what purpose, and what rights the individual has — including the right to oppose any automated decision that significantly affects them.

In practice this means three concrete actions. First, include a specific AI disclosure notice at the point of application or onboarding — not buried in a general privacy policy, but presented clearly and separately before the individual submits any data. The notice must explain in plain language what the AI system does, what data it processes, and how it influences the decision, as well as an explicit statement of the individual's right to oppose and how to exercise it.

Second, keep a record of when and how each individual was informed — because if a candidate or employee later challenges the process, the organisation must be able to demonstrate that the transparency obligation was met before the decision was made, not retrofitted afterwards.

Under GDPR Articles 13–14 and LOPDGDD Article 11, the timing of disclosure is not optional — it must precede the processing, not follow it.

Third, the disclosure notice must meet a minimum content threshold to be legally effective. It is not sufficient to inform candidates that 'AI is used in our recruitment process'. As Goodman and Flaxman (2017) establish, meaningful transparency requires that the individual understands what data is being processed, how it influences the decision, and what they can do about it. In practical terms this means the notice should specify: the type of AI system used and its function in the decision, the categories of data it processes, the weight it carries relative to human review, and the concrete steps to exercise the right to oppose. A notice that does not cover these elements does not satisfy GDPR Articles 13–14 or LOPDGDD Article 11 — regardless of when it was delivered

e. Phase 4: Output review and decision

This is the phase at which legal exposure is highest and governance failures are most common. As both SCHUFA (2023) and Uber Amsterdam (2023) demonstrate, the presence of a human in the process is not sufficient — what matters is whether that human exercised genuinely independent judgement. A reviewer who confirms what the algorithm determined is not a genuine human override; it is automation with an extra step.

The compliance obligation at this phase, grounded in GDPR Article 22 and EU AI Act Article 14, translates into three concrete actions. First, give the reviewer access to the full picture — not just the algorithmic score but the underlying data that generated it, so they can actually evaluate the output rather than simply receive a number. A score without context cannot be meaningfully reviewed. Second, require the reviewer to record their own reasoning in writing — an assessment that stands independently of the algorithm's conclusion, in terms that could be communicated to the affected individual if they ask for an explanation. As Goodman and Flaxman (2017) argue, this is the only practical way to satisfy the right to explanation under GDPR Article 15(1)(h) — if the reviewer's notes simply restate the algorithm's output, the organisation has no real explanation to give. Third, establish a genuine contestation process — one that routes a challenge to a different human reviewer who conducts a fresh evaluation from scratch, rather than an automated acknowledgement or a confirmation of the original decision. As Abraha (2023) documents, the right to contest is only meaningful if it can actually change the outcome, and that requires a process designed to make that possible.

6.4. Regulatory gaps and recommendations

The protocol above addresses what organisations can do within the current framework. But three gaps remain that organisations cannot close on their own and that require legislative or regulatory action.

The first is the absence of a clear operational definition of genuine human involvement. Both the GDPR and the EU AI Act require it, but neither defines with sufficient precision what it looks like in practice. As Abraha (2023) documents, courts and data protection authorities across the EU have reached conflicting answers, producing inconsistent protection. The EDPB or the EU legislature should issue specific guidance defining minimum standards for human oversight in high-risk HR AI systems — including minimum review time, documentation requirements, and criteria for independence.

The second gap concerns enforcement. Even where the legal framework is clear, its practical impact depends on whether regulators are actively monitoring compliance. As Abraha (2023) notes, only three of the twelve European data protection authorities surveyed by the Future of Privacy Forum featured employment as a strategic priority. This means that organisations deploying AI in HR can, in practice, operate with limited regulatory scrutiny — not because what they are doing is legal, but because nobody is systematically checking. A robust legal framework is only as effective as its enforcement, and in the employment context that enforcement is currently uneven across member states.

The third is the unresolved tension between transparency obligations and commercial confidentiality. As Carter (2024) identifies, organisations routinely invoke trade secret protections to resist disclosing how their algorithms work — which makes it practically impossible for affected individuals to prove discrimination even when it has occurred. Legislative clarification is needed on the limits of commercial confidentiality as a defence against transparency obligations in high-risk AI systems.

6.5. Limitations and future research

This dissertation has several limitations that should be acknowledged. The analysis is based on a systematic review of existing literature and documented cases rather than primary empirical research — which means the findings reflect the state of published evidence rather than direct organisational observation. The case studies examined are predominantly from large organisations with significant public visibility; smaller organisations deploying AI tools may face different practical challenges that are not captured here.

The protocol proposed in section 6.3 is grounded in the current EU and Spanish legal framework as it stands in 2025–2026. The EU AI Act's high-risk obligations for HR systems are not yet fully in force, and the regulatory landscape will continue to evolve. The protocol should be understood as a compliance framework for the current moment, not a permanent solution.

Three directions stand out for future research. First, the empirical evidence on how humans actually behave when reviewing algorithmic outputs — produced in laboratory settings by Sele and Chugunova (2024) and Rosenthal-von der Pütten and Sach (2024) — needs to be tested in real organisational HR contexts, to better understand what oversight designs genuinely reduce automation bias in practice. Second, the relationship between employers and AI vendors requires further legal clarity: as the Workday case illustrates, the distribution of responsibility between the organisation that deploys a tool and the company that built it is still being worked out in the courts, and clearer governance frameworks are needed. Third, the opacity problem remains largely unsolved — the law requires organisations to explain their algorithms in meaningful terms, but how to do that without revealing commercially sensitive technical information is a question that sits at the intersection of law and computer science, and one that neither discipline has yet answered satisfactorily.

7. Declaration of use of AI

Declaration of Use of Generative Artificial Intelligence Tools in Final Degree Dissertations

Hereby, I, Carlota Navarro de la Riva, student of Law and Business Administration and Management at Universidad Pontificia Comillas, in presenting my Final Degree Dissertation entitled "The Use of Artificial Intelligence in Human Resource Decisions: A Legal Perspective", declare that I have used the Generative Artificial Intelligence tool ChatGPT or other similar open-code GAI tools only in the context of the activities described below:

1. Research idea brainstorming: Used to generate and outline possible research areas.
2. Critical: To find counter-arguments to a specific thesis I intend to defend.
3. References: Used in conjunction with other tools, such as Science, to identify preliminary references which I subsequently contrasted and validated.
4. Multidisciplinary studies: To understand perspectives from other communities on topics of a multidisciplinary nature.
5. Literary style and language corrector: To improve the linguistic and stylistic quality of the text.
6. Synthesiser and disseminator of complex books: To summarise and understand complex literature.
7. Reviewer: To receive suggestions on how to improve and refine the work at different levels of rigour.
8. Translator: To translate texts from one language to another.

I affirm that all information and content presented in this work are the product of my own research and individual effort, except where otherwise indicated and corresponding credits have been given (I have included appropriate references in the dissertation and have made explicit the purposes for which ChatGPT or other similar tools have been used). I am aware of the academic and ethical implications of presenting non-original work and I accept the consequences of any violation of this declaration.

Date: 03/06/2026

Signature: 

Bibliography

Academic papers and books

Abraha, H. (2023). Regulating algorithmic employment decisions through data protection law. *European Labour Law Journal*, *14*(2), 172–191. https://doi.org/10.1177/20319525231167317

Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., Boselie, P., Cooke, F. L., Decker, S., DeNisi, A., Dey, P. K., Guest, D., Knoblich, A. J., Malik, A., Paauwe, J., Papagiannidis, S., Patel, C., Pereira, V., Ren, S., … Varma, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, *33*(3), 606–659. https://doi.org/10.1111/1748-8583.12524

Carter, C. (2024). Why the algorithmic recruiter discriminates: The causal challenges of data-driven discrimination. *Maastricht Journal of European and Comparative Law*, *31*(3), 333–359. https://doi.org/10.1177/1023263X241248474

Cheng, M. M., & Hackett, R. D. (2021). A critical review of algorithms in HRM: Definition, theory, and practice. *Human Resource Management Review*, *31*(1), 100698. https://doi.org/10.1016/j.hrmr.2019.100698

Chowdhury, S., Budhwar, P., & Wood, G. (2024). Generative artificial intelligence in business: Towards a strategic human resource management framework. *British Journal of Management*, *35*(4), 1680–1691. https://doi.org/10.1111/1467-8551.12824

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation." *AI Magazine*, *38*(3), 50–57. https://arxiv.org/abs/1606.08813

Gong, Y., Fan, D., & Bartram, T. (2025). Integrating artificial intelligence and human resource management. *The International Journal of Human Resource Management*. https://doi.org/10.1080/09585192.2024.2440065

Haque, A. (2025). AI is hiring you: algorithmic power, ethical risks, and the future of recruitment. *International Journal of Organization Theory & Behavior*. https://doi.org/10.1108/IJOTB-02-2025-0045

Jackson, S. E., Schuler, R. S., & Jiang, K. (2014). An aspirational framework for strategic human resource management. *Academy of Management Annals*, *8*(1), 1–56. https://scholarship.libraries.rutgers.edu/view/pdfCoverPage?instCode=01RUT_INST&filePid=13643492370004646&download=true

Kaufman, B. E. (2014). The historical development of American HRM broadly viewed. *Human Resource Management Review*, *24*(3), 196–218. https://www.researchgate.net/publication/261599451_The_historical_development_of_American_HRM_broadly_viewed

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, *13*, 795–848. https://doi.org/10.1007/s40685-020-00134-w

Malik, A., Budhwar, P., & Kazmi, B. A. (2023). Artificial intelligence (AI)-assisted HRM: Towards an extended strategic framework. *Human Resource Management Review*, *33*(1), 100940. https://doi.org/10.1016/j.hrmr.2022.100940

Parviainen, P. (2022). Algorithmic recruitment: Legal challenges and perspectives for enabling the use of automated systems in recruitment under EU law. *European Labour Law Journal*, *13*(2), 225–248. https://doi.org/10.1177/20319525221093815

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469–481. https://arxiv.org/abs/1906.09208

Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2023). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review*, *33*(1), 100925. https://doi.org/10.1016/j.hrmr.2022.100925

Rosenthal-von der Pütten, A. M., & Sach, A. (2024). Michael is better than Mehmet: Exploring the perils of algorithmic biases and selective adherence to advice from automated decision support systems in hiring. *Frontiers in Psychology*, *15*, 1416504. https://doi.org/10.3389/fpsyg.2024.1416504

Sele, D., & Chugunova, M. (2024). Putting a human in the loop: Increasing uptake, but decreasing accuracy of automated decision-making. *PLOS ONE*, *19*(2), e0298037. https://doi.org/10.1371/journal.pone.0298037

Varma, A., Dawkins, C., & Chaudhuri, K. (2023). Artificial intelligence and people management: A critical assessment through the ethical lens. *Human Resource Management Review*, *33*(1), 100923. https://doi.org/10.1016/j.hrmr.2022.100923

Legal instruments

General Data Protection Regulation. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Union*, L 119/1. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. *Boletín Oficial del Estado*, núm. 294, de 6 de diciembre de 2018. https://www.boe.es/eli/es/lo/2018/12/05/3

Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación. *Boletín Oficial del Estado*, núm. 167, de 13 de julio de 2022. https://www.boe.es/eli/es/l/2022/07/12/15

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

Case law

Amsterdam Court of Appeal. (2023, April). *Uber B.V. and Others v. Drivers* [Robo-Firing cases]. As documented by Eurofound (2023). https://apps.eurofound.europa.eu/platformeconomydb/amsterdam-court-of-appeal-rules-uber-violated-data-protection-rights-110150

Court of Justice of the European Union. (2023, December 7). *SCHUFA Holding AG v. Datenschutz Hessen* (Case C-634/21). https://curia.europa.eu/juris/document/document.jsf?docid=280426

Filcams CGIL Bologna and Others v. Deliveroo Italia SRL. (2020, December 31). Court of Bologna, RG 2949/2019. As cited in Carter, C. (2024). Why the algorithmic recruiter discriminates. *Maastricht Journal of European and Comparative Law*, *31*(3), 333–359. https://app.vlex.com/search/jurisdiction:ES,GB+country_jurisdiction:UK_ND,UK_NI,UK_SC,UK_EW/Why+the+algorithmic+recruiterdiscriminates%3A+The+causalchallenges+of+data-drivendiscrimination/vid/1069633161

Mobley v. Workday, Inc. (2024). 740 F. Supp. 3d 796 (N.D. Cal. 2024). https://caselaw.findlaw.com/court/us-dis-crt-n-d-cal/116378658.html

Institutional and press sources

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Electronic Privacy Information Center. (2019, November 6). In the matter of HireVue, Inc.: Complaint and request for investigation, injunction, and other relief. Federal Trade Commission. https://epic.org/wp-content/uploads/privacy/ftc/hirevue/EPIC_FTC_HireVue_Complaint.pdf

Eurofound. (2023). *Amsterdam Court of Appeal rules Uber violated data protection rights*. European Foundation for the Improvement of Living and Working Conditions. https://apps.eurofound.europa.eu/platformeconomydb/amsterdam-court-of-appeal-rules-uber-violated-data-protection-rights-110150

Jones Walker LLP. (2025). *AI vendor liability squeeze: Courts expand accountability while contracts shift risk*. https://www.joneswalker.com/en/insights/blogs/ai-law-blog/ai-vendor-liability-squeeze-courts-expand-accountability-while-contracts-shift-r.html

New York City. (2021). *Local Law 144 of 2021: Automated Employment Decision Tools*. In force July 2023. https://rules.cityofnewyork.us/rule/automated-employment-decision-tools-2/