A Work Project, presented as part of the requirements for the Award of a Master's degree in Management from the Nova School of Business and Economics.

RESPONSIBLE ARTIFICIAL INTELLIGENCE IN HUMAN RESOURCE MANAGEMENT: STRATEGIES, SOLUTIONS AND FUTURE OUTLOOK ANALYSIS

EMILY FREIIN VON CRAILSHEIM

Work project carried out under the supervision of:

Professor Milton Jorge Correia de Sousa

17/12/2024

Abstract

This thesis explores the adoption and integration of Responsible Artificial Intelligence in human resource management, analyzing its current use and strategies to ensure ethical and effective application in human resource practices. A multi-level model reveals drivers like efficiency and cost saving, alongside barriers such as ethical concerns, skill gaps, and regulatory challenges. Additionally, to guide human resource professionals, a four-stage recommendation framework was developed, focusing on building foundational Artificial Intelligence knowledge, readiness audits, training, and piloting. Emphasizing transparency, collaboration, and alignment with regulatory frameworks, this study provides actionable strategies to balance technological innovation with the Human Resource Department's people-centric focus.

Keywords: business ethics, human resource management, artificial intelligence, responsible artificial intelligence

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Table of Contents

List of Abbreviations

Artificial Intelligence (AI)

Human resource (HR)

Human resource management (HRM)

Responsible Artificial Intelligence (Responsible AI)

Grounded theory (GT)

List of Tables

List of Figures

1. Introduction

Artificial Intelligence (AI) is rapidly reshaping industries and societies, establishing itself as one of the most important and relevant technological advancements of our time. From data-driven decision-making to automating complex tasks, AI promises significant improvements in efficiency, accuracy, and innovation across several fields. However, with its transformative power comes a responsibility to address new ethical, legal, and societal challenges. As AI continues to evolve, concerns about transparency, fairness, and accountability are becoming central to discussions around its responsible implementation (European Union 2024; Google 2018; UNESCO 2022).

These concerns are particularly significant in human resource management (HRM), where people are at the center of all decisions, and AI-driven systems can directly impact individuals, including their careers and personal data. While AI offers significant benefits in HRM, such as improved recruitment processes and workforce planning, it also introduces risks, including biased decision-making and potential privacy violations. These challenges have underscored the importance of implementing Responsible Artificial Intelligence (Responsible AI) frameworks, especially in human resources (HR), where ethical considerations are crucial to protect both employees and organizations.

Due to this urgency, we want to investigate the following research questions:

1. *How and to what extent do HR professionals adopt and integrate Responsible AI solutions in their organizations?*
2. *What strategies should be pursued to responsibly leverage AI opportunities in HR?*

Our thesis aims to understand how HR professionals adopt and integrate Responsible AI practices in their organizations. By exploring current adoption levels and identifying effective strategies, we seek to provide actionable insights that enable HR professionals to leverage AI

responsibly while addressing ethical, legal, and societal challenges. Through qualitative interviews, we have gathered insights into AI's role in HRM and the specific obstacles HR professionals face, which serve as the foundation for developing a practical model for implementing Responsible AI. In doing so, our work contributes to both academic research and provides HR professionals with concrete guidance to navigate the ethical complexities of AI in their field.

This paper explores the adoption and responsible implementation of AI in HRM, addressing both its potential benefits and critical challenges. Section 2 establishes the theoretical background, defining Responsible AI principles such as fairness, transparency, and accountability, and outlining key concerns including bias, data privacy, and employee well-being (Section 2.2). It also reviews existing frameworks like the EU AI Act and IEEE P7005, analyzing their applicability to HR and identifying gaps (Section 2.3). Section 3 presents empirical findings from 26 interviews with HR professionals and discusses these findings. Finally, Section 4 introduces a framework that synthesizes the findings and offers recommendations, outlining a four-stage model for HR professionals to responsibly integrate AI into their organizations while ensuring ethical oversight. The paper concludes with reflections on HR's evolving role, AI's long-term implications, and future research directions in Section 5.

2. Theoretical Background

The following section provides the theoretical foundation for understanding AI adoption and responsible implementation in HRM. It explores AI's role in HRM, highlighting its benefits and risks, and defines key principles of Responsible AI, including fairness, transparency, and accountability. Additionally, it examines concerns such as bias, data privacy, and employee well-being, along with existing frameworks for Responsible AI in HR. By synthesizing relevant literature, this chapter sets the stage for the empirical analysis, ensuring a comprehensive understanding of AI's impact on HRM.

2.1 Criteria for Responsible AI in HRM

To fully understand the implementation of Responsible AI in HRM, it is essential to establish the criteria that define its responsible application. Building on the foundational principles of trustworthiness and ethical standards explored earlier, the next section delves into the specific considerations and requirements for applying Responsible AI within the HR domain.

Trustworthiness is both a fundamental attribute and a desired outcome of Responsible AI and encompasses fairness, transparency, and accountability, while ethical standards focus on ethical principles, human-centric design, and sustainability. Together, these dimensions ensure that AI systems align with societal values, support employee well-being, and enhance organizational effectiveness. By adhering to these criteria, HRM can leverage the benefits of AI responsibly while mitigating its risks and fostering trust among all stakeholders. Each criterium is examined in both a general context and its specific application within HRM. This structured approach highlights the universal principles of Responsible AI while addressing the unique challenges and opportunities AI presents in HR practices.

Fairness is a cornerstone of Responsible AI, ensuring that AI systems produce equitable outcomes by addressing inherent biases in data and algorithms (European Union 2024; Google

2018; UNESCO 2022). This is particularly vital in HRM, where decisions regarding hiring, performance evaluation, and career progression must avoid discriminatory and subjective practices. For instance, recruitment algorithms trained on biased data can perpetuate inequalities, undermining workplace diversity (Tilmes 2022). Therefore, robust mechanisms to identify and mitigate algorithmic bias are essential for achieving fairness in HR applications. By prioritizing fairness, HRM can foster inclusivity and equitable opportunities for all employees.

Transparency plays a foundational role in fostering trust and accountability in Responsible AI (European Union 2024; Google 2018; UNESCO 2022). It enables stakeholders to understand how AI systems make decisions, providing clarity and insight into processes that impact individuals and organizations. In HRM, transparency is especially critical in sensitive areas such as recruitment and performance evaluations, where employee trust heavily depends on perceived fairness and clarity. Explainability complements transparency by ensuring that AI decisions are interpretable and understandable to all relevant parties (Ali et al. 2023). As a result, transparent HR systems foster confidence and mitigate mistrust among employees.

Accountability ensures that stakeholders take responsibility for the outcomes of AI systems, particularly in addressing unintended consequences such as errors, biases, or ethical violations (European Union 2024; Google 2018; Kaur 2024). Effective governance frameworks are necessary to uphold accountability, requiring all parties involved in AI development and deployment to share responsibility for its ethical and societal impacts (Kowald et al. 2024). Therefore, in HRM, accountability mechanisms should be in place to address the implications of AI-driven decisions on employee well-being and organizational fairness.

While trustworthiness focuses on operational principles like fairness, transparency, and accountability, ethical standards emphasize the broader alignment of AI systems with societal values, human-centric approaches, and sustainability. These standards ensure that AI not only

meets organizational goals but also contributes positively to societal and environmental well-being.

Responsible AI must align with ethical principles that prioritize human dignity, equity, and social good (European Union 2024; IBM 2024; UNESCO 2022). Hence, this alignment requires AI systems to reflect societal values and respect the rights and interests of all stakeholders. In HRM, ethical principles guide workforce practices, promoting fairness, inclusivity, and respect for individual autonomy (Greenwood and Freeman 2011). These principles are critical in fostering an ethical workplace culture that values employees as central to organizational success (Greenwood and Freeman 2011).

At the heart of Responsible AI lies a commitment to putting humans at the center of AI applications. This human-centric approach ensures that AI technologies enhance, rather than replace, human capabilities, fostering collaboration between humans and machines (Lase and Nkosi 2023). In HRM, this means leveraging AI tools to support decision-making, improve employee engagement, and enhance workforce management without undermining human contributions or creating undue anxiety about job displacement (Fenwick, Molnar, and Frangos 2024). A human-centric focus also promotes employee well-being by ensuring that AI systems are designed to serve, rather than disrupt, organizational values and individual needs.

Sustainability is another critical dimension of Responsible AI, integrating long-term societal and environmental considerations into its design and application (European Union 2024). Addressing the environmental impact of AI systems, such as energy consumption, is a core component of this principle. Concepts like “Green AI” emphasize the use of energy-efficient algorithms to reduce AI’s carbon footprint (Bolón-Canedo et al. 2024). These efforts align AI-driven HR processes with broader organizational goals for corporate social responsibility and environmental stewardship (Jia and Hou 2024).

2.2 Concerns related to Responsible AI in HRM

The adoption of AI in HRM introduces numerous concerns that extend beyond ethical dilemmas. The following section highlights four main concerns encompassing operational, technological, and organizational challenges that significantly influence the effectiveness, efficiency, and acceptance of AI-driven HR practices. A nuanced understanding of these issues is essential to enable HRM professionals to implement AI responsibly and effectively.

Firstly, the integration of AI disrupts established HR workflows, creating operational complexities and, therefore, resistance among employees and management (Calvard and Hislop 2022). Traditional processes, such as recruitment and performance evaluations, often require significant modifications to accommodate AI, leading to fears of inefficiency, increased workloads, and unfamiliarity with the technology (Calvard and Hislop 2022).

Secondly, AI technologies in HRM face limitations, particularly regarding their reliance on data quality (SHRM 2023). A key concern is the risk of algorithmic bias and inequity, as AI systems often depend on historical data that may perpetuate discriminatory practices in hiring, promotions, and performance management (Calvard and Hislop 2022; SHRM 2023). Fragmented, incomplete, or inaccurate HR data further exacerbate these risks, as AI requires robust, diverse, and high-quality data to function effectively (SHRM 2023). Biased data can disadvantage certain demographic groups, undermining efforts to promote diversity, equity, and inclusion while eroding trust in AI among employees and stakeholders.

Thirdly, the use of sensitive employee data by AI systems raises significant privacy, security, and ethical concerns (Ekuma 2023; Patel 2024). Processing personal information, such as performance metrics and behavioral patterns, increases the risk of data breaches and unauthorized access (Ekuma 2023; Patel 2024). Additionally, AI-driven surveillance can

infringe on privacy, fostering distrust, while ethical questions persist around the transparency and extent of data collection (Ekuma 2023; Patel 2024).

Fourthly, the automation of routine HR tasks has led to widespread concerns about job displacement and workforce anxiety (Cox 2023; Rotmann 2024). Employees, particularly those in roles centered on repetitive tasks, often fear that AI systems may render their positions obsolete (Cox 2023; Rotmann 2024). This anxiety can decrease morale, engagement, and trust in AI initiatives, fostering resistance to technological advancements (Cox 2023; Rotmann 2024). Furthermore, the uncertainty about how AI will reshape the future of work adds to employee stress, exacerbating concerns about job security (Cox 2023; Rotmann 2024).

2.3 Current Frameworks

Although numerous companies, scholars, intergovernmental organizations, and policy-making entities have developed frameworks to ensure AI is ethical, trustworthy, and responsible, these guidelines often lack specificity for the unique challenges and operational needs of HR management.

In the section below, Table 1 summarizes the most relevant regulatory frameworks for AI integration, highlighting their applicability to HR. These global initiatives establish foundational principles that shape standards and practices across organizations. However, despite their value as a starting point for understanding Responsible AI, they frequently fail to provide HR-specific guidance. Subsequently, Table 2 compares companies with established frameworks and guidelines for AI using aggregated criteria, showcasing their approaches and relevance to HR.

Organization	Regulatory Framework	Year of adoption	Key Message	HR relatedness	Applicability	Sources
Institute of Electrical and Electronics Engineers (IEEE)	**IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems**	April 2016	The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems aims to provide a guide for ethical design of AI technologies and to recommend standards for the development of ethically responsible autonomous systems.	The IEEE P7005 standard specifically focuses on the ethical use of AI in HR by providing guidelines for transparent, secure, and ethical handling of employee data. It aims to support employees in making informed decisions about their data while helping employers align with these practices for regular workflows.	Optional. Countries, organizations and companies can use it as a guideline.	(IEEE 2016; IEEE 2021; IEEE 2024)
Organization for Economic Co-operation and Development (OECD)	**OECD Principles on Artificial Intelligence**	May 2019	The OECD AI Principles are the first intergovernmental standards for AI. They aim to promote the development and use of an AI that is innovative, trustworthy and respects human rights and democratic values.	While the principles do not explicitly address the HR department, they can be applied. They suggest responsible use of AI to avoid discrimination, privacy protection, transparency in use of AI and accountability for decisions made by AI systems.	Optional, but additional to the 38 member countries, over 1000 policy initiatives in over 70 jurisdictions have been reported to follow the OECD AI principles.	(OECD 2019)
United Nations Educational, Scientific and Cultural Organization (UNESCO)	**UNESCO Recommendation on the Ethics of Artificial Intelligence**	November 2021	The UNESCO Recommendation on the Ethics of AI serves as the first global standard for ethical AI development and application, emphasizing respect for human rights, inclusive societies, diversity, and ecosystem sustainability, while providing member states with actionable tools for responsible AI implementation.	The recommendations do not explicitly address AI in HR, but principles like transparency, data protection, and human oversight are applicable to recruitment, development, and performance management processes.	Optional. Countries, organizations and companies can use it as a guideline.	(UNESCO 2022)
Global Partnership on Artificial Intelligence (GPAI)	**Scaling Responsible AI Solutions (SRAIS) Project**	December 2023	The SRAIS project supports AI teams worldwide in integrating responsible AI principles into scalable projects. The report identifies challenges like addressing bias, ensuring transparency, and safeguarding human rights, as well as structural issues like equitable access to data and computing resources. It offers recommendations for safe testing, equitable infrastructure access, and embedding responsible AI throughout the project lifecycle, emphasizing accountability during scaling.	The project does not focus on the HR department, but its principles of responsible AI can be applied to it. Key aspects include robust data governance to ensure employee data privacy and security, fairness to prevent discrimination in recruitment and performance evaluations, transparency to help employees understand and contest AI decisions, and a human-centered approach to support rather than replace HR professionals.	Optional. Countries, organizations and companies can use it as a guideline.	(GPAI 2023; GPAI 2024)
European Commission (EC)	**Ethics Guidelines for Trustworthy AI**	April 2019	Provide a framework for ethical, legal, and robust AI, emphasizing human autonomy, fairness, oversight, and data protection. Promote a human-centered, rights-focused approach to ensure AI benefits society.	The Ethics Guidelines for Trustworthy AI do not explicitly address the HR department but provide a general framework that can be applied to HR-related AI use cases. For example, AI systems in recruitment should avoid discrimination, promote diversity, ensure transparency, and include human oversight for fair decision-making. Similarly, AI tools for performance evaluation should deliver objective and fair assessments, operate transparently, and allow employees to challenge outcomes.	Optional. EU member states, organizations and companies can use it as a guideline.	(European Commission 2019)
European Union (EU)	**EU AI Act**	**Proposal:** April 2021 **Adoption:** June 2024 **Implementation:** Starting from February 2025	The EU AI Act is a regulatory framework designed to ensure the safe and trustworthy development and use of artificial intelligence within the European Union. It categorizes AI systems according to their riskiness and imposes strict requirements on their development and use to safeguard health, safety, and fundamental rights. Further, it bans specific harmful AI applications, such as biometric systems for inferring sensitive attributes like race or political opinions. Most importantly, it establishes a supervision framework involving national authorities, the European Commission, and a new AI Office. The European Commission will annually review whether the list of high-risk AI systems and prohibited AI practices needs to be updated. Compliance with the EU AI Act is enforced by national authorities, with support from the AI Office to ensure consistent application. Sanctions may be imposed for violations.	Under the EU AI Act, AI systems used in HR are classified as high-risk due to their potential to significantly affect careers and livelihoods, perpetuate discrimination, or infringe on fundamental rights such as data protection and privacy. To comply with the Act, such systems must meet strict requirements: they must utilize diverse, high-quality training data to avoid bias, ensure transparency and explainability so decisions can be understood and contested, maintain human oversight to prevent complete automation, and undergo fundamental rights impact assessments to identify and minimize risks.	Mandatory. Starting February 2025, the EU AI Act will ban high-risk practices like social scoring and biometric identification. By August 2026, full compliance with rules on risk assessments, data quality, and oversight is required. Authorities and providers have until 2030 for full implementation, with annual reviews and enforcement by national authorities supported by the AI Office.	(ArtificialIntelligenceAct.eu 2024a; ArtificialIntelligenceAct.eu 2024b; European Union 2024; European Commission 2024a; European Commission 2024b)

Table 1: Regulatory Frameworks

Among the analyzed frameworks from Table 1, the IEEE P7005 Standard and the EU AI Act stand out for their direct connection to HRM (European Union 2024; IEEE 2016; IEEE 2021). The IEEE P7005 Standard is part of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and focuses specifically on ethical and transparent data management within the HR domain (IEEE 2016; IEEE 2021). Its primary goals include protecting employee privacy, fostering informed decision-making, and aligning employer-employee data policies (IEEE 2016; IEEE 2021). By emphasizing the creation of safe data-sharing environments, it seeks to balance the operational needs of employers with the rights of employees (IEEE 2021). This targeted approach to HR processes makes the IEEE P7005 Standard particularly valuable in providing ethical guidance on data governance. However, it primarily addresses data-related

issues and does not fully tackle the broader operational complexities of AI integration in HR, such as algorithmic bias or capacity building for AI literacy (IEEE 2016; IEEE 2021).

In contrast, the EU AI Act represents a significant milestone as the first comprehensive legislation regulating AI systems globally, establishing legally binding obligations for the development, deployment, and oversight of AI systems (European Union 2024). Its relevance to HRM lies in its risk-based framework, which directly addresses the ethical, operational, and legal challenges associated with AI in sensitive sectors such as recruitment, promotion decisions, employee monitoring, termination decision, educational and vocational training as well as work allocation (ArtificialIntelligenceAct.eu 2024a; European Commission 2024a; European Union 2024). These HR-specific applications are classified as high-risk due to their potential impact on employment opportunities and livelihoods (ArtificialIntelligenceAct.eu 2024a; European Commission 2024a; European Union 2024). This classification imposes specific requirements on HR departments, including ensuring transparency, conducting risk assessments, maintaining human oversight, and implementing robust data governance practices (ArtificialIntelligenceAct.eu 2024a; European Union 2024). Hence, compliance is not optional for HR professionals, but a legal requirement.

One key component of the EU AI Act is its emphasis on fostering AI literacy, defined as the "skills, knowledge, and understanding" required by all stakeholders to responsibly develop, use, and oversee AI systems (European Union 2024, 49). For HR professionals, this could translate into targeted trainings to develop a working knowledge of AI technologies, including their functions, potential biases, and the ethical and legal standards governing their use.

Despite the valuable insights offered by frameworks like the IEEE P7005 Standard and the EU AI Act for HR, significant challenges remain. Many frameworks lack HR-specific guidance, requiring HR professionals to adapt broad principles for their specific needs.

Therefore, in addition to regulatory frameworks, selected frameworks and guidelines developed by scholars and companies have been analyzed to guide the responsible application of AI technologies. While both scholarly and corporate frameworks were analyzed, those developed by scholars were not included in the competing framework analysis (Table 2) as they do not fully meet the thesis's goal of providing HR-specific, practical, and actionable guidance for Responsible AI integration. For instance, Robert et al. (2020) and Verma et al. (2022) focus on fairness and ethics in HR but remain theoretical, lacking concrete implementation steps. Similarly, Anagnostou et al. (2022) and Schiff et al. (2020) offer general frameworks for Responsible AI but fail to address HR's roles in skill development, process adaptation, or leadership in Change Management. In addition, the authors Mökander and Floridi (2022) explore ethics-based auditing for AI governance, providing valuable insights for auditing and compliance but without specific applicability to HR contexts. Finally, Clarke (2019) and Bankins (2021) emphasize ethical guidelines or decision-making frameworks but overlook HR's unique operational challenges and its role as a driver of change.

In contrast, the frameworks developed by leading companies, analyzed in Table 2, are better suited to the thesis's objectives, as companies are often at the forefront of operationalizing AI (Staton 2024). This makes their frameworks particularly relevant for HR departments seeking actionable insights and tools to navigate AI integration.

The criteria used to compare these company frameworks were derived systematically by examining the elements highlighted by each company, summarizing them into key points to create a set of common, comparable criteria. This approach ensures a balanced and cohesive analysis across the companies, providing a clear understanding of their strengths and gaps in the context of HR. By focusing on company frameworks, the analysis emphasizes practical, actionable solutions tailored to the specific needs of HR departments, aligning closely with the thesis's goal of creating a user-friendly and effective framework for Responsible AI integration.

	Criteria / Company	Google AI Principles	Microsoft AI Principles	Deloitte AI Workforce Report	EY AI Principles	Workday AI in HR	IBM AI in HR	Cisco AI Framework	SAP AI Ethics Framework	Capgemini Code of Ethics for AI
Aggregated Criteria	Definition of Criteria for Responsible AI	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Framework / Guidelines for AI Implementation	⊖	✓	✓	⊖	✓	✓	✓	✓	⊖
	Governance and Oversight Structures	⊖	✓	✓	⊖	⊖	✓	✓	✓	✓
	Fairness and Bias Mitigation	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Explainability and Interpretability	⊖	✓	✓	✓	⊖	✓	✓	✓	✓
	Human-in-the-Loop Decision-Making	✓	✓	⊖	✓	⊖	✓	⊖	✓	✓
	Sustainability	⊖	⊖	✓	✓	⊖	⊖	⊖	✓	✓
	Risk Management and Compliance	✓	✓	✓	✓	⊖	✓	✓	✓	✓
	Auditability and Trust	⊖	✓	✓	⊖	⊖	✓	✓	✓	✓
	Data Transparency, Quality, and Accountability	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Scalability and Adaptability	⊖	⊖	✓	⊖	⊖	⊖	⊖	✓	✓
	HR-Specific Applications and Implications	⊖	⊖	⊖	⊖	✓	✓	⊖	⊖	✓
	Employee Well-being and Impact	⊖	✓	⊖	⊖	✓	⊖	⊖	✓	⊖
	Stakeholder Engagement and Inclusion	⊖	✓	⊖	✓	⊖	⊖	✓	✓	✓
	Cross-departmental Collaboration	⊖	⊖	⊖	⊖	⊖	⊖	✓	✓	✓
Lacking Criterium beyond Companies' Criteria	Holistic Approach to Introduction and Implementation of AI in HRM e.g. addressing skill development, change leadership, cultural alignment, and job security concerns	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖
Sources		(Google 2023)	(Microsoft 2022)	(Deloitte n. d.)	(EY 2023)	(Workday 2024c)	(Kohls 2024)	(Cisco 2024)	(SAP n. d.)	(Capgemini 2021)

Table 2: Competing Framework Analysis

The analysis from Table 2 reveals critical gaps in the aggregated criteria, underscoring the need for a new, holistic, and comprehensive framework tailored to Responsible AI implementation in HR. Key areas from Table 2, such as "HR-specific Applications and Implications", are notably underrepresented, reflecting a lack of focus on addressing the unique challenges and opportunities within the HR domain. "Employee Well-being and Impact", a vital aspect for fostering trust and acceptance of AI among employees, are similarly overlooked by most frameworks and guidelines. Further, little attention is given to "Stakeholder Engagement and Inclusion", which are essential for fostering multi-level collaboration. Similarly, "Cross-departmental Collaboration", a key driver for cohesive AI adoption within organizations and HR departments, seems to be often overlooked. While commonly mentioned criteria, such as "Fairness and Bias Mitigation" and "Governance and Oversight Structures", are essential, they fail to fully address the specific needs of HR departments aiming to integrate Responsible AI. Furthermore, some critical elements are lacking from current frameworks and guidelines in Table 2. These include skill development opportunities for HR teams, such as training in AI ethics, technical literacy, and the operational intricacies of AI tools. The frameworks also lack

emphasis on the changing role of HR in leading AI-driven Change Management initiatives and acting as a bridge between technical teams and employees. Additionally, the ethical and psychological impacts of AI on employees, such as concerns about job security and trust in automated decision-making, are inadequately addressed. The influence of AI on organizational culture and the need for its alignment with core values also remain unexplored.

These observations highlight the urgent need for a framework, which provides practical guidance and actionable steps for HR professionals to ensure the responsible and effective use of AI within HR departments.

3. Findings

Building on grounded research, this chapter presents key findings from 26 interviews with HR professionals (see Appendix 4). The results are structured into three dimensions, developed through a systematic analysis from first-order data (direct interview insights) to second-order data (researcher interpretations).

Each dimension provides a focused perspective on AI's impact on HRM, offering insights into its opportunities and the need for responsible implementation.

3.1 Barriers to Responsible AI Adoption

The dimension *"Limited awareness, clear guidelines, and expertise hinder the responsible adoption of AI in HR"* highlights the critical barriers preventing HR professionals from adopting AI responsibly (Figure 1). The findings highlight that limited awareness, lack of expertise, and absence of clear guidelines or standards are significant barriers to Responsible AI adoption in HR.

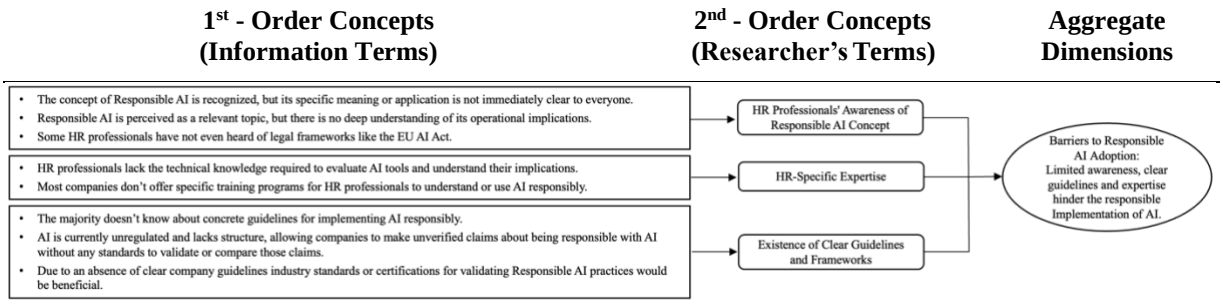

Figure 1: Data Structure for Barriers to Responsible AI Adoption

3.1.1 HR Professionals' Awareness of Responsible AI Concepts

The findings reveal that while the concept of Responsible AI is recognized as relevant within HR, a profound lack of understanding persists regarding its operationalization. Although some HR professionals associate Responsible AI with ethical AI usage in alignment with data security regulations, many struggle to articulate actionable steps for implementation. For

instance, IP 15 defines Responsible AI as using AI "*sensibly and not in a negative way*" but cannot specify how this could translate into organizational policies. This lack of understanding is not confined to HR but reflects broader organizational challenges. IP 24, an HR consultant, noted that in a project focused on defining Responsible AI practices for a company, there was *"hardly any knowledge about AI"* within the organization. To address this, external experts initially developed foundational content to help the company define what Responsible AI meant for them, including principles for data usage, AI development, and employee involvement. However, as the responsibility shifted to the company, it became clear that internal capacity and awareness were limited.

Moreover, awareness of formal legal frameworks, such as the EU AI Act, is alarmingly limited. Several HR professionals interviewed admitted either being unfamiliar with such regulations or failing to appreciate their implications. For example, IP 02 acknowledged hearing about Responsible AI but was unaware of its specific requirements or guidelines. This disconnect underscores a critical knowledge gap, leaving HR departments ill-prepared to navigate the legal and ethical dimensions of AI adoption.

Interestingly, while data privacy is often emphasized, it tends to overshadow broader ethical dimensions such as fairness, transparency, and sustainability. This narrow focus limits the ability of HR departments to adopt a holistic approach to Responsible AI. The absence of structural supports, such as ethics committees, exacerbates this issue. Only a few companies have established mechanisms to oversee ethical AI deployment, further emphasizing the disparity in Responsible AI governance across industries.

Interestingly, while data privacy is often emphasized, it tends to overshadow broader ethical dimensions such as fairness, transparency, and sustainability. This narrow focus limits the ability of HR departments to adopt a holistic approach to Responsible AI. The absence of

structural supports, such as ethics committees, exacerbates this issue. Only a few companies have established mechanisms to oversee ethical AI deployment, further emphasizing the disparity in Responsible AI governance across industries. IP 19 highlighted this gap, stating, *"Currently, we're facing several challenges regarding AI...we lack a specific code of conduct or ethical guidelines for its use."* This concern is echoed by various interviewees, who emphasize the need for clear policies and guidelines to ensure the responsible implementation of AI in HR.

3.1.2 HR Expertise

HR professionals bring valuable ethical and communicative strengths to the table but lack the technical expertise required for Responsible AI implementation. Their ethical orientation is evident in their concern for data privacy and fairness, core tenets of Responsible AI. For instance, IP 19 highlighted the risks of data misuse and algorithmic bias, demonstrating an awareness of potential pitfalls.

Additionally, IP 20 highlighted the importance of preparing employees both within HR and across the organization for the changes that AI will bring. Their department recently organized a launch event to introduce employees to three AI-driven use cases, addressing the need to familiarize the workforce with AI and its implications for their roles. This proactive approach underscores HR's responsibility not only to implement AI responsibly but also to manage the organizational change and employee engagement that such technologies entail.

However, technical gaps persist as a significant barrier. Many HR professionals lack a deep understanding of AI's underlying algorithms, capabilities, and limitations, leaving them ill-equipped to evaluate tools or address biases. IP 02 pointed out the need for HR professionals to understand *"the boundaries associated with AI tools,"* a skill that is currently underdeveloped. Without this technical acumen, they risk over-reliance on vendor assurances or failing to

identify ethical risks. IP 20 emphasized that AI systems must be embedded into existing workflows and processes to create real value, rather than functioning in isolation. This aligns with the broader need for HR professionals to develop technical expertise and collaborate closely with IT and data science teams.

The lack of targeted training exacerbates these challenges. Few organizations offer programs tailored to HR professionals, leaving them to navigate the complexities of AI implementation with insufficient support. This skills gap highlights the urgent need for capacity-building initiatives that equip HR professionals with the technical and ethical competencies required for Responsible AI integration. IP 20 mentioned that they plan to launch an event for all HR employees. This event not only aims to familiarize HR professionals with the functionalities of these tools but also to provide training on how they will affect daily workflows. The structured rollout of such tools, combined with training elements, reflects an understanding that the successful adoption of AI in HR requires both technical integration and workforce enablement. Similarly, IP 26 shared insights about their first AI workshops, which were well-attended and focused primarily on providing first insights, and practical use cases rather than delving into ethical discussions.

3.1.3 Existence of Clear Guidelines and Frameworks

The absence of clear and universally accepted guidelines remains a significant hurdle in Responsible AI implementation. Many organizations rely solely on broad regulatory frameworks, such as the GDPR or EU AI Act, which require significant contextual interpretation. Smaller companies, in particular, struggle with this ambiguity, as they lack the resources to develop tailored policies.

Some larger organizations have begun crafting internal ethical frameworks to address issues such as data privacy, fairness, and accountability. However, these efforts remain fragmented

and inconsistent across industries. IP 15 highlighted the challenge of measuring or verifying Responsible AI practices, pointing to the need for standardized guidelines or certification systems. This sentiment is echoed by other interviewees, such as IP 02 and IP 19, who advocated for a certification system to evaluate the sustainability and ethical implications of AI tools.

The unregulated nature of AI compounds these challenges, allowing companies to self-declare their practices as responsible without external validation. This lack of accountability undermines efforts to establish trust and transparency in AI adoption. The demand for certification systems and industry standards is growing, as these mechanisms could provide much-needed structure and clarity.

3.2 Unlocking Responsible AI Potential

The dimension *"The responsible use of AI in HR hinges on data protection and ethical considerations, yet it remains underutilized due to insufficient prioritization"* examines the perceived importance and priority of Responsible AI within organizations, along with the criteria HR professionals have established or consider important in its implementation (Figure 2).

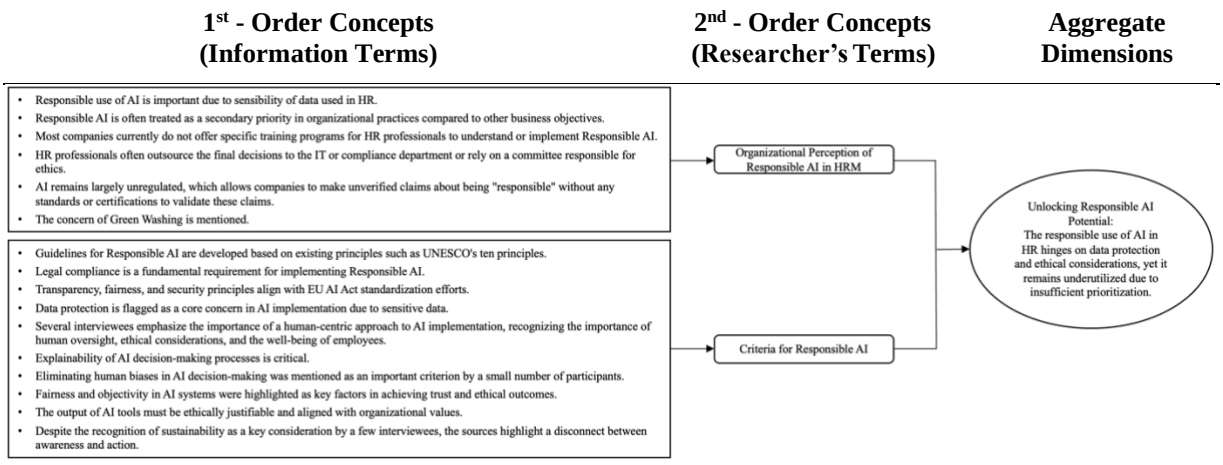

Figure 2: Data Structure for Unlocking Responsible AI Potential

3.2.1 Organizational Perception of Responsible AI in HRM

Responsible AI is frequently overshadowed by other organizational objectives. Interviewees described how the focus on immediate outcomes, such as optimizing workflows or achieving measurable financial gains, often relegates Responsible AI to a "nice-to-have" rather than an essential practice. This mindset reflects a broader organizational tendency to prioritize immediate needs over long-term, strategic initiatives like Responsible AI. Aspects such as transparency or sustainability, which are critical to responsible AI, are often deprioritized. As IP 01 explained, *"Aspects like Responsible AI, meaning whether the system is particularly transparent or sustainable, would then be of secondary importance."* That many HR professionals are not yet prioritizing Responsible AI is also represented by IP 11's statement: *"I don't think it's on our agenda yet because we're still focused on doing the basic tasks before we can even think about the next step. That's also why we haven't dealt with it so far. In the next one to two years, it just won't happen."*

However, there are encouraging examples of organizations attempting to integrate ethical considerations into their AI practices. For instance, IP 13 described a required "*mini training*" for employees registering to use an AI tool. This brief but mandatory training includes content on non-discriminatory language and the responsible handling of personal data, ensuring that employees are aware of critical ethical principles before accessing the tool. As IP 13 explained, this approach combines internal diversity guidelines with training to promote accountability and reinforce ethical practices at the outset. Although this training is brief, it demonstrates how companies can embed Responsible AI principles into everyday processes, fostering a culture of awareness and compliance.

Despite such efforts, most organizations still do not provide comprehensive training for HR professionals to understand or implement Responsible AI. Some HR professionals, especially those of large international companies, mentioned their reliance on external departments or

ethics committees to manage the complexities of AI adoption. Several interviewees noted that final decisions regarding AI systems were often made by IT teams or compliance officers without adequate input from HR professionals, who are closest to the human and ethical implications of these tools. This delegation reduces HR's control over AI implementation and creates a disconnect between technical decision-makers and the organizational values HR aims to uphold.

This disconnect is further complicated by the unregulated nature of AI in HRM. Some interviewees claim to implement AI responsibly without adhering to clear standards or certifications. This lack of accountability allows organizations to make unverified claims about their AI ethics, contributing to greenwashing. Greenwashing also highlights the broader systemic issue of unregulated AI use. Without industry standards or certifications, companies are free to define "Responsible AI" on their own terms, often tailoring the narrative to suit their business interests rather than prioritizing ethical considerations. As IP 13 put it, "*AI is currently so unregulated and unstructured that if I knew you were looking for a company that values responsibility, I could easily start a communication campaign claiming to be highly responsible with AI. However, there's no way to track or compare that claim to any standard*."

Another important role in many larger companies is played by the works council, which, as a representative of employees, is responsible for safeguarding their interests and well-being. This makes the topic of Responsible AI in HRM particularly relevant for the works council, as it directly aligns with their mandate to protect employees from potential risks associated with AI systems. For example, issues such as bias, transparency, and data privacy fall squarely within their purview. IP 13 emphasized the active role of their works council, highlighting its involvement in all discussions related to AI implementation. IP 13 told how the organization collaborated with work council to develop a framework agreement for specific AI tools, ensuring that employee-related implications were addressed. Additionally, the company is in

the process of creating a general works agreement on AI, designed to establish consistent standards for AI usage across various tools and minimize the need for separate regulations with each new system.

3.3 The Dawn of AI in HR

The dimension "*HRM is in its early stages of evolution, and the coming years will bring significant transformations, redefining HR's role and practices"* explores the evolving role of HR, emerging trends in responsible AI adoption, and the limitations of implementing AI responsibly within HR practices (Figure 3).

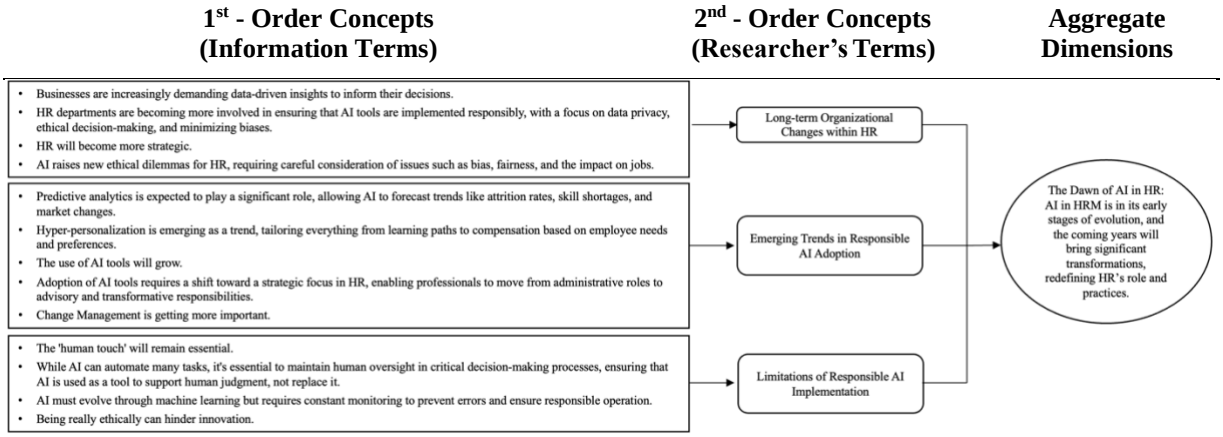

Figure 3: Data Structure for The Dawn of AI in HR

3.3.1 Long term Organizational Changes within HR

The integration of AI, especially when it's done responsibly, into organizational processes is reshaping the role of HR, transitioning it from an often still traditionally transactional function into a strategic partner in driving organizational success. Insights from interviews with HR professionals consistently highlighted this shift, showcasing how AI enables HR to redefine its priorities and contribute more effectively to long-term business objectives. As IP 16 aptly stated, *"We're witnessing a shift where machines are expected to take over more routine tasks, freeing humans to focus on more insightful and creative work."* This evolution positions HR as a proactive leader in shaping the future of work.

As AI becomes more integrated into HR processes, it introduces complex ethical challenges, such as algorithmic bias, fairness in decision-making, and the broader impact on job security. Many interviewees emphasized the critical role HR plays in navigating these dilemmas, ensuring that fairness and inclusivity remain guiding principles in AI implementation, also applying a more strategic role for HR in the future. This underscores the need for HR to not only oversee AI systems but also actively mitigate risks of bias and discrimination.

The growing reliance on AI also raises questions about the evolving responsibilities of HR professionals and the new competencies required to navigate these changes. As IP 07 pointed out, HR staff must not only understand the decisions but also the supporting information to interpret them effectively. While IT can assist, HR staff need the necessary competencies to use AI responsibly (IP 07).

Another recurring theme in the interviews was the increasing demand for data-driven insights from HR departments. Organizations are looking to HR to provide actionable intelligence that optimizes workforce planning, enhances employee engagement, and improves operational efficiency. AI empowers HR to analyze large volumes of workforce data, uncover patterns, and forecast trends.

3.2.2 Emerging Trends in Responsible AI Adoption

The interviews revealed a growing reliance on predictive analytics powered by AI in HR. Participants highlighted AI's ability to analyze workforce data and forecast trends such as attrition rates, skill shortages, and market changes. As IP 12, an HR consultant, noted, "*AI will become increasingly clever at predicting future trends.*". While some organizations have established a foundation for gathering employee data, others are still navigating digitalization and lack the necessary measures. These predictive capabilities enable HR to take a proactive

approach to workforce planning, helping organizations remain agile and prepared for future challenges.

Another key development identified in the interviews is the rise of hyper-personalization as a transformative trend in HR. AI now allows organizations to tailor processes to the unique needs and preferences of individual employees, from designing customized learning paths to creating personalized compensation packages. IP 12 described this as a significant shift, and IP 14 elaborated, suggesting that *"AI could be used to identify employee weaknesses and suggest personalized training programs,"* demonstrating how AI can empower employees and foster professional growth during this transition. This trend not only enhances employee engagement and satisfaction but also aligns personal development with organizational objectives, creating a mutually beneficial dynamic.

With the growing use of AI tools, effective Change Management has become crucial for successful adoption. HR professionals emphasized that success depends not just on the technology but on its integration into organizational culture. Transparent communication, addressing employee concerns, and aligning with organizational goals are critical for fostering trust and acceptance of AI-driven changes. IP 18 noted that while change strategies in AI transformation build on years of digitalization, they are becoming more personal as technology evolves. This complexity demands closer support as employees face new challenges in adapting to advanced tools. IP 18 also posed a key question: *"When will we reach the point where human value in certain tasks is questioned because AI is already performing them better?"* While much focus remains on routine tasks and pattern recognition, more advanced scenarios will emerge, requiring proactive preparation. By integrating these forward-looking considerations into Change Management strategies, HR can support employees in adapting to current AI capabilities while preparing for future developments. This approach fosters adaptability,

resilience, and trust, ensuring continued employee engagement throughout the AI adoption process.

As AI tools continue to evolve, HR's role is being redefined, shifting from traditional administrative responsibilities to more strategic functions. By moving beyond operational tasks, HR is positioning itself as a critical driver of organizational growth and resilience and stepping into the role of a business partner. However, many interviewees also highlight that the human element will stay essential in HR departments and AI should mainly, as IP 02 for example states "*serve as a complementary tool, taking over specific, repetitive tasks, but not replacing the human element entirely."*

3.3.3 Limitations of Responsible AI Implementation

The interviews highlight that, while AI holds significant potential in HRM, its adoption remains highly complex, with discussions around how to implement it responsibly still in the early stages for many companies. Organizations are just beginning to evaluate the extent to which AI can be responsible, while encountering limitations in terms of the potential for Responsible AI to be truly responsive in its areas of application.

One critical limitation of Responsible AI, as highlighted in the interviews, is the inherent trade-off between ethical responsibility and innovation. While strict adherence to ethical standards is essential, it can also slow the pace of AI adoption. This challenge was particularly evident in discussions about how different regions approach AI regulation. Interviewees noted that organizations in highly regulated regions like the EU face greater hurdles in staying competitive compared to their counterparts in less-regulated areas such as the USA or China. As IP 24 observed, *"The EU often has strict rules, while countries like the USA and China operate more freely. This creates economic disadvantages for the EU. At the same time, we cannot assess the long-term consequences of the unregulated use of AI."*

Furthermore, several interviewees cautioned against relying solely on AI without human oversight, particularly in processes like recruitment or performance evaluations. This reflects a broader hesitation to entrust AI with full ethical decision-making responsibilities. It reinforces the continued need for human responsibility in guiding and monitoring AI systems to ensure decisions remain fair, contextual, and aligned with organizational values.

This ongoing reliance on human oversight illustrates that, while AI can support decision-making and enhance efficiency, it is not yet capable of navigating the nuanced ethical and social implications inherent in HR processes. As such, the role of HR professionals remains pivotal—not only in managing AI-driven systems but also in upholding the ethical integrity of decisions made within these systems.

Furthermore, HR professionals consistently highlighted the irreplaceable value of the human touch and their “gut feeling” while HR professionals emphasized that empathy, interpersonal communication, and context-driven decision-making are essential aspects of employee engagement and organizational culture. These human qualities, they argued, cannot be replicated by even the most advanced AI tools.

4. Discussion

4.1 The Model

By synthesizing qualitative data from our findings, we identified six aggregate dimensions that are crucial for understanding the adoption and integration of Responsible AI solutions in HR. Each dimension highlights key factors influencing the extent to which HR professionals adopt Responsible AI technologies. Together, these six dimensions form an interconnected framework (Figure 4; see p. 101) that provides a nuanced understanding of our first research question "*How and to what extent do HR professionals adopt and integrate Responsible AI solutions in their organizations".*

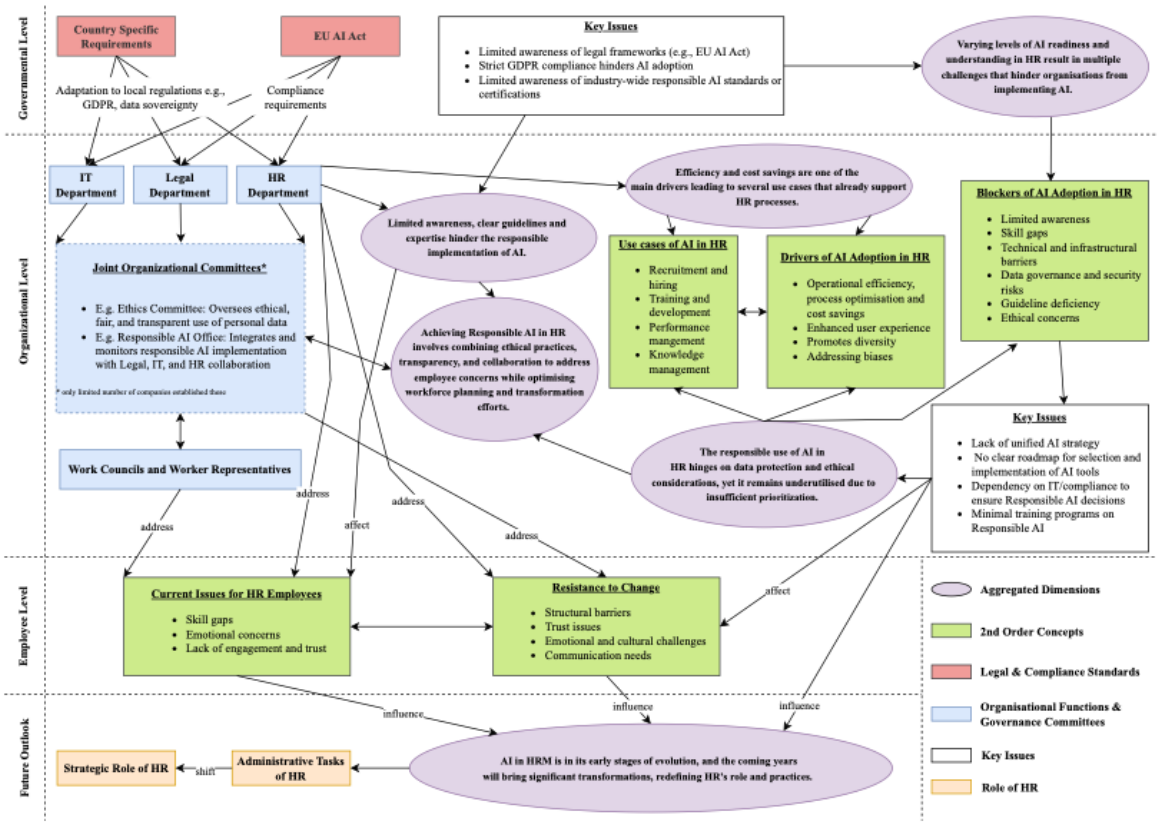

Figure 4: Adoption and Integration of Responsible AI by HR Professionals in Organizations

The first dimension, *"There are varying levels of AI readiness and understanding in HR result in multiple challenges that hinder organisations from implementing AI"*, highlights the fragmented landscape of AI adoption across organizations. Many HR departments face

challenges due to uneven knowledge and preparedness, which manifest as skill gaps, insufficient strategic planning, and limited exposure to successful real-world applications. This lack of readiness results in technical and structural barriers, as represented in the framework, where challenges like limited awareness and infrastructure misalignment are prominent blockers to implementation. This dimension highlights the need for capacity-building and strategic roadmaps, highlighting that readiness gaps hinder the adoption of Responsible AI solutions.

"Efficiency and cost savings are one of the main drivers leading to several uses cases that already support HR processes". This second dimension demonstrates how operational benefits motivate organizations to implement AI in key processes such as recruitment, training, and performance management. The framework illustrates these drivers through practical use cases, showcasing how AI facilitates process optimization, knowledge management, and enhanced user experiences. This dimension reflects how HR professionals adopt AI primarily to achieve tangible economic benefits, focusing on automation and efficiency. However, the focus on operational benefits often overshadows long-term ethical goals, highlighting a tension between economic gains and responsible AI integration.

The third dimension, *"Limited awareness, clear guidelines and expertise hinder the responsible Implementation of AI"*, underscores the obstacles created by insufficient regulatory frameworks and a lack of targeted training for HR professionals. While the concept of Responsible AI is widely recognized, its operational implementation is hindered by unclear industry standards and the absence of structured governance. This is reflected in the framework, where issues like the lack of unified AI strategies and minimal training programs hinder progress. HR professionals often rely on IT or legal teams for compliance, exposing skill gaps. This highlights a disconnect between theory and practice, limiting Responsible AI adoption, which remains underutilized due to insufficient focus on data protection and ethics.

This fourth dimension, *"The responsible use of AI in HR hinges on data protection and ethical considerations, yet it remains underutilized due to insufficient prioritization"*, reflects a critical tension within organizations, where legal compliance and ethical concerns, such as GDPR adherence and fairness, are recognized but not prioritized. The framework ties these considerations to ethical transparency and data governance issues, emphasizing the importance of aligning AI practices with regulatory standards and ethical principles. Despite their importance, these factors are often secondary, leading to cautious, compliance-driven AI adoption and highlighting HR's challenge in balancing ethics with operational goals. *"Achieving Responsible AI in HR involves combining ethical practices, transparency, and collaboration to address employee concerns while optimizing workforce planning and transformation efforts"*. This fifth dimension forms a core pillar of the framework, integrating cross-functional collaboration, transparency, and trust-building measures. Ethical oversight committees and joint organizational structures are critical to addressing employee concerns and fostering trust in AI-driven decisions. This dimension showcases HR professionals' role in aligning workforce strategies with AI to ensure fairness and inclusivity, while also fostering responsible adoption by mitigating resistance through proactive communication and engagement.

Finally, the sixth dimension is, *"AI in HRM is in its early stages of evolution, and the coming years will bring significant transformations, redefining HR's role and practices"*. As shown in the framework, HR's evolving role will shift from administrative tasks to strategic advisory functions, driven by advancements in predictive analytics and hyper-personalization. This dimension underscores the gradual integration of AI by HR professionals, emphasizing its potential to transform practices despite current challenges in readiness and ethical prioritization.

4.2 Recommendations for HR Professionals

Based on our review of existing literature and insights gathered from HR professionals during interviews, we have developed a four-stage framework consisting of strategies to responsibly leverage AI opportunities in HR (Figure 5). This model aims to answer our second research question "*What strategies should be pursued to responsibly leverage AI opportunities in HR"*. It provides a clear roadmap for HR professionals to navigate the complexities of AI adoption while ensuring alignment with ethical principles and organizational goals.

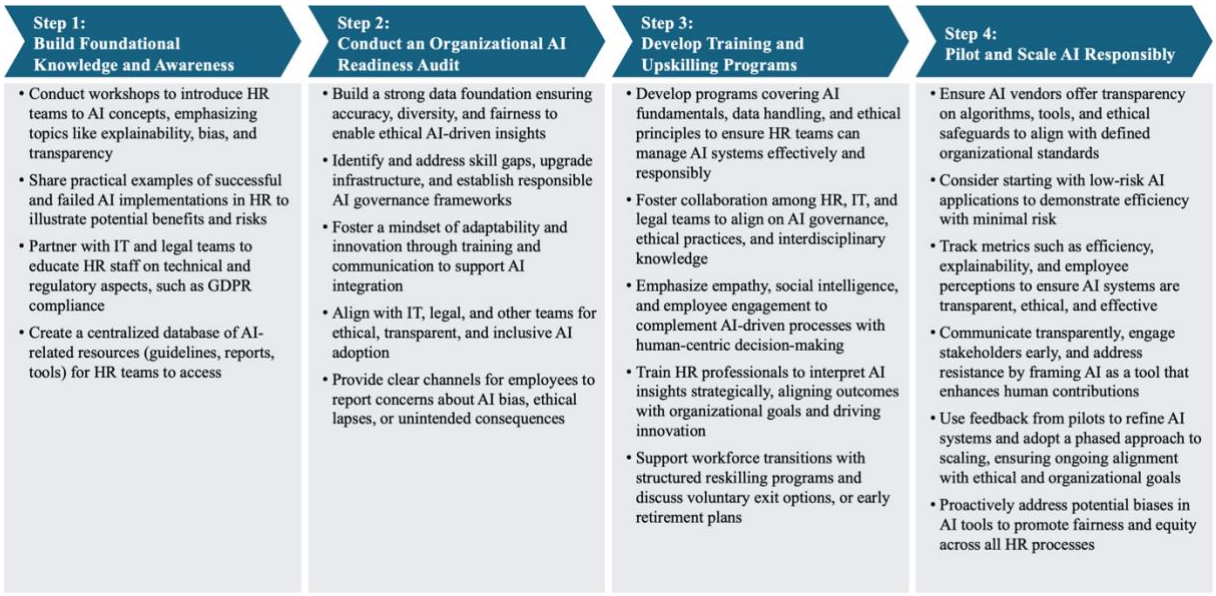

Figure 5: Four-stage Framework: Strategies to Responsibly Leverage AI Opportunities for HR Professionals

Step 1: Build Foundational Knowledge and Awareness

Building foundational knowledge is a critical first step for HR teams to navigate the complexities of AI adoption responsibly. It ensures that HR professionals are equipped to understand AI's potential, evaluate its risks, and align its use with organizational and ethical standards. This foundational understanding can be achieved through structured learning opportunities designed to enhance awareness and foster collaboration.

To begin, workshops on key AI concepts should introduce HR teams to the fundamental principles of AI, including explainability, bias, and transparency. These sessions provide HR professionals with the tools to critically evaluate AI systems, understanding their implications

for processes such as recruitment, performance evaluation, and workforce planning. Workshops empower HR to approach AI with confidence, ensuring decisions are guided by both technological insights and ethical considerations.

In addition to workshops, practical case studies are an effective way to highlight real-world applications of AI in HR. Sharing examples of both successful and problematic implementations illustrates the potential benefits of AI-driven tools, such as improved efficiency and decision-making, while emphasizing the risks of ethical lapses. For example, cases of biased recruitment algorithms or opaque decision-making processes act as warnings, underscoring the critical importance of Responsible AI practices.

Cross-functional collaboration is another key component of foundational knowledge building. Partnering with IT and legal teams ensures HR professionals understand the technical and regulatory contexts of AI adoption, such as compliance with data protection laws like GDPR. This collaboration aligns AI implementation with organizational policies, ensuring that HR operates within a broader framework of accountability and governance.

Organizations should also establish a Centralized AI Resource Hub and Platform to support learning and collaboration. This hub would provide HR teams with access to essential AI-related materials, including guidelines, tools, reports, and training resources, fostering a consistent and well-informed approach to AI. Beyond serving as a repository, the platform should facilitate active collaboration by allowing employees to share ideas for improving HR functions, discuss best practices, and raise concerns about AI applications. Importantly, the hub should act as a reporting channel for ethical issues or perceived biases in AI systems. Employees could document and report potential ethical violations, ensuring that these concerns are transparently addressed. By embedding this mechanism, the platform reinforces the

organization's commitment to ethical AI use while fostering trust and accountability among stakeholders.

Step 2: Organizational AI Readiness Audit

Preparing for the responsible and effective adoption of AI requires HR professionals to conduct a thorough Organizational AI Readiness Audit. This process ensures the organization has the necessary data foundation, team readiness, and cultural alignment to implement AI tools ethically and effectively. By addressing potential challenges early, HR can create a roadmap for sustainable and Responsible AI integration.

A critical first step is for HR to establish a strong data foundation for people-related information, ensuring accuracy, diversity, and completeness. High-quality data allows organizations to leverage AI for trends like data prediction and personalization, driving informed decision-making and tailored employee experiences. At the same time, HR must focus on fairness and inclusivity, addressing any gaps to prevent biased outcomes. Collaborative efforts across HR and IT teams are essential to enhance and structure data for AI-driven processes, ensuring the organization benefits from reliable and ethical insights.

With a strong data base in place, HR should assess team readiness by conducting structured surveys or interviews with HR, IT, and legal teams. This helps identify technical, skill, and cultural gaps that may hinder AI adoption, such as insufficient familiarity with AI tools or outdated infrastructure. Addressing these gaps requires a roadmap that includes training programs, infrastructure upgrades, and process adjustments. Additionally, HR must evaluate whether the organization has a framework for Responsible AI, including clear guidelines and access to internal or external ethical experts. If absent, HR should advocate for the creation of such frameworks and ensure their involvement in shaping governance practices.

Beyond technical readiness, fostering a culture of learning and adaptability is critical. The introduction of AI requires a mindset shift across the organization, encouraging employees to embrace continuous learning and innovation. HR plays a pivotal role in promoting this culture through training initiatives and communication strategies that highlight AI as an enabler of growth and improvement.

Finally, cross-departmental collaboration is vital for ensuring cohesive AI adoption. HR must engage IT, legal, and other teams early to align priorities and ensure all perspectives are considered. By coordinating these efforts, HR ensures that AI systems are implemented effectively while adhering to principles of fairness, transparency, and inclusivity.

Step 3: Developing Training and Upskilling Programs

This stage in the process addresses the skill gaps identified in both literature and our interviews, ensuring HR professionals are prepared to manage AI solutions effectively while aligning them with organizational and ethical goals. The primary purpose of developing training and upskilling programs is to empower HR teams to confidently navigate the complexities of AI adoption. To achieve this, training initiatives must go beyond technical know-how and address the broader interdisciplinary knowledge needed for successful implementation. Programs should focus on key areas such as AI fundamentals, data handling, and ethical principles, ensuring HR professionals understand not just how AI systems work, but also how to manage their ethical and societal implications. Without this technical acumen, HR professionals risk over-reliance on vendor assurances or the inability to identify ethical risks. For instance, training on algorithm functionality and data management helps HR teams evaluate AI systems critically, while ethical training emphasizes principles such as fairness, accountability, and transparency in decision-making.

A key action in this phase is the development of cross-functional workshops that bring together HR, IT, and legal teams. These workshops foster collaboration and ensure alignment between departments, enabling a holistic understanding of AI's potential and its associated challenges. By integrating technical and legal expertise with HR's domain knowledge, organizations create a foundation for responsible and effective AI adoption. Cross-functional training also ensures that all teams are on the same page when it comes to AI governance and ethical considerations.

While technical and ethical skills are essential, the importance of soft skills such as empathy and social intelligence is becoming increasingly evident in the AI-driven workplace. As AI takes over routine administrative tasks, HR professionals must focus on higher-value activities, such as building relationships, fostering employee engagement, and addressing complex workplace dynamics. Empathy and social skills will play a vital role in bridging the gap between AI-driven insights and human-centric decision-making, ensuring employees feel supported and valued in an increasingly automated workplace. Additionally, the ability to think strategically is becoming increasingly important for HR professionals. With AI providing actionable insights, HR teams need to develop the skills to interpret data strategically and align AI-driven outcomes with broader organizational goals. This involves not only understanding the data but also using it to drive innovation, improve workforce planning, and enhance employee experiences.

In addition to these efforts, organizations should develop a structured plan for managing potential job changes or reductions associated with AI adoption. Such plans should focus on reskilling employees for new roles within the organization while exploring voluntary exit options or early retirement programs as alternatives. Integrating these strategies into upskilling programs reflects a commitment to ethical workforce management, balancing efficiency with social responsibility.

Step 4: Pilot and Scale AI Responsibly

Implementing AI in HR requires a cautious and structured approach to ensure that systems are effective, ethical, and aligned with organizational values. Piloting AI provides an opportunity to evaluate its impact, refine processes, and build trust before full-scale implementation.

We recommend beginning by selecting pilot areas that deliver significant benefits without exposing the organization to high ethical or operational risks. For example, AI can be piloted in tasks such as scheduling or automating responses to common employee queries. These areas are ideal for demonstrating AI's potential to improve efficiency while maintaining a low-risk profile. However, avoid deploying AI in sensitive functions such as diversity tracking or hiring decisions during the pilot phase, as these require more thorough validation to ensure fairness and ethical compliance.

To evaluate the effectiveness of the pilot, set up monitoring protocols to track key metrics such as efficiency, Responsible AI criteria, and employee perceptions. Monitoring should especially assess whether AI decisions are explainable and transparent, enabling users to understand how outcomes are reached. Explainability is crucial for building trust, as employees and professionals are more likely to embrace AI tools that operate transparently and ethically.

Change Management plays a crucial role in the successful adoption of Responsible AI, as employees and stakeholders may initially be hesitant to embrace new technologies. To address these challenges, HR professionals should focus on three key strategies. First, communicate early and often by clearly articulating the purpose of AI initiatives, their objectives, and the benefits they bring to employees and the organization. Open and transparent communication reduces uncertainty, and fosters trust among all parties involved. Second, engage stakeholders by involving employees, professionals, IT, and works councils from the outset. This inclusive approach helps build a sense of ownership and collaboration, ensuring that all

perspectives are considered. Finally, HR professionals should proactively address resistance by understanding employee concerns, such as fears of job displacement or algorithmic biases. Framing AI as a tool that enhances, rather than replaces, human contributions can help alleviate anxieties and encourage acceptance. Together, these Change Management strategies create a supportive environment for the ethical and effective implementation of AI in HR.

A successful implementation depends on gathering feedback from those interacting with the AI system, including employees, professionals, and other stakeholders. Use surveys, interviews, or focus groups to understand user experiences and identify potential issues. Use the feedback and monitoring data collected to address biases, improve functionality, and ensure alignment with organizational goals. Collaborate with IT teams and vendors to refine the AI tool, focusing on areas like algorithm transparency, decision accuracy, and user-friendliness. Once the pilot demonstrates success, scale the AI system incrementally. A phased approach reduces the risks associated with rapid deployment and allows for ongoing adjustments. Scaling AI responsibly requires ongoing evaluation to ensure the system remains effective and ethical as its use expands.

4.3 Critical Reflections and Implications

This section critically reflects on the findings through the lens of the research questions, engaging with existing literature and posing critical questions to deepen the understanding of Responsible AI adoption in HR. These questions are intentionally formulated as open-ended, providing future researchers with direction for further investigation while offering HR professionals thought-provoking insights to reflect on and address in their practices. The analysis highlights the profound transformations AI is bringing to HR, necessitating both structural and strategic adjustments while exploring its ethical implications.

This research contributes to the literature by addressing the prioritization of Responsible AI in HR, an area that remains underexplored, especially in practice. It extends prior studies by emphasizing the enduring relevance of empathy, emotional intelligence, and interpersonal skills, even as AI enhances efficiency. The findings challenge the narrative of AI as a purely technical tool, advocating for its integration in ways that prioritize both ethical principles and organizational goals.

Interesting findings from our interviews, which were not prominently highlighted in the existing literature, include the generational differences in attitudes toward AI in HR. While older HR professionals often expressed more skepticism regarding the extent to which AI can be effectively integrated into HR practices, younger HR professionals showed more enthusiasm but seemed to be tempered particularly by the overwhelming nature of AI technologies and a lack of time to engage deeply with the topic.

While numerous applications and use cases of AI across all areas of HRM are widely discussed in theory, our interviews painted a different picture. For instance, in theory, continuous monitoring is often highlighted as a means to ensure timely and actionable feedback, fostering a culture of accountability and improvement (Bankins et al. 2022; Nocker and Sena 2019). However, participants in our study expressed skepticism and reluctance to adopt AI in areas such as performance management, citing concerns about privacy intrusion. This raises the question of whether such skepticism stems from a lack of knowledge or reflects a gap between theory and practice, as certain areas may not yet be feasible for AI implementation.

One of the central insights of this research is the need for HR to embrace a more strategic role in response to AI-driven transformation. As AI automates routine tasks, HR professionals are increasingly positioned to shift from operational responsibilities to becoming key drivers of

organizational strategy. This aligns with findings by Floridi and Cowls (2019), who emphasize the potential of AI to enhance efficiency and decision-making.

Our interviews also identify HR's unique role as a mediator between technical teams and employees as a key opportunity for advancing Responsible AI. Unlike frameworks that focus on leadership or IT, this research highlights HR's capacity to bridge technical expertise with organizational values and employee needs. By doing so, HR can ensure that AI adoption aligns with ethical and human-centric principles. This dual role also positions HR as a strategic driver, leveraging AI for data-driven workforce planning, identifying skill gaps, and creating inclusive policies that balance efficiency with employee well-being. However, this raises a critical question: To what extent does HR possess the necessary influence and organizational authority to balance the technical priorities of AI development with the ethical and human-centric principles required to foster trust and inclusivity, particularly in hierarchical or cost-focused environments? The ability of HR to fulfill this mediating role depends on its organizational standing and access to decision-making processes. In many organizations, HR may lack the strategic authority or resources to challenge technical priorities or advocate for long-term ethical considerations, potentially undermining its capacity to ensure balanced AI integration.

The growing reliance on AI in decision-making raises important ethical challenges. How can HR prevent AI from unintentionally reinforcing biases, even in systems designed to be neutral? And how can accountability be maintained as AI-driven insights increasingly influence decisions? These challenges call for robust governance frameworks and clear accountability structures to ensure fairness, transparency, and trust in AI-integrated HR practices.

Our interviews highlight the critical importance of human-centric strengths, such as empathy and emotional intelligence, as distinct contributions that HR professionals bring to AI-augmented workplaces. This finding resonates with existing literature, which emphasizes the

irreplaceable value of human connection (Floridi and Cowls, 2019). Similarly, intergovernmental recommendations, such as those from UNESCO (2022), advocate for capacity-building in ethical AI practices. However, this research underscores the need to integrate these human-centric competencies into HR's evolving identity alongside technical expertise. This raises critical questions: Are organizations making sufficient investments in nurturing these uniquely human skills alongside advancing technical capabilities? If not, could the growing emphasis on technology risk devaluing the interpersonal abilities that are central to HR's role and impact?

The findings emphasize the challenge of balancing efficiency with ethical responsibility, particularly in workforce planning. HR professionals are tasked with designing systems that optimize operations while respecting employees' rights, privacy, and well-being. This dual focus aligns with the OECD AI Principles' emphasis on fairness and transparency. However, this study goes further by identifying the role of works councils in fostering this balance, ensuring that workforce planning aligns with both organizational objectives and employee concerns. This raises critical questions: To what extent do works councils and HR departments currently collaborate in AI implementation, and are these partnerships sufficient to safeguard employee rights? How can organizations prevent ethical considerations from being sidelined when operational pressures demand rapid AI adoption?

A key opportunity highlighted in the interviews is the potential to use Responsible AI, including sustainable AI, as part of an employee branding strategy. This insight is absent in existing AI in HRM literature, which focuses primarily on ethics and fairness, highlighting the originality of this study. By adopting transparent and environmentally conscious AI systems, organizations can not only align with long-term values but also position themselves as ethical and forward-thinking employers. This dual benefit enhances both internal practices and external reputation. In an era where employees increasingly seek alignment with their personal values, Responsible

AI offers HR a powerful tool for attracting and retaining top talent. However, integrating sustainable AI into branding strategies raises critical questions: Are organizations willing to trade revenue for the ethical and environmental benefits of sustainable AI? How can HR professionals effectively advocate for this priority amid cost-focused leadership?

Resistance to change emerges as a critical barrier, driven by fears of job displacement and mistrust of AI systems. Existing literature acknowledges organizational resistance broadly but often underplays the emotional and cultural dimensions specific to HR. This study adds depth to this understanding by emphasizing the discomfort employees feel about delegating traditionally human-centered tasks to algorithms. Transparency and communication are highlighted as key tools for overcoming resistance, echoing Rai (2020)'s argument for explainability as a trust-building measure. However, further questions arise: How can HR professionals balance the need for transparency with the complexity of AI systems that may be difficult to fully explain?

The findings reveal significant skill gaps and low AI literacy among HR professionals, presenting a major challenge to the adoption of Responsible AI. These gaps highlight the importance of HR professionals embracing their role as change agents in digital transformation, as described by Ferdinan (2021) and Farndale and Vidovic (2021).

Aligned with this perspective, our research highlights the need for HR-specific training in areas like algorithmic bias and ethical data governance. The step-by-step guidance model developed in this study offers actionable pathways to address these skill gaps. However, the interviews revealed that many HR professionals lack the time and capacity to actively engage with AI and its responsible implementation due to heavy workloads. While only a few organizations currently offer training programs, those that do report high participation rates, reflecting the topic's relevance and urgency. This raises critical questions: Are organizations prepared to

invest in comprehensive training programs that address both technical and ethical aspects of AI? How can HR professionals balance the time required for upskilling with their existing responsibilities?

From a practical perspective, the status quo model serves as a diagnostic tool for identifying barriers, while the step-by-step guidance model provides actionable strategies to address them. Together, these models offer HR professionals a roadmap for navigating resistance, closing skill gaps, and integrating AI responsibly.

AI in HR should not replace the human element but rather strengthen it. Building on Storey's (1995) assertion that HRM plays a pivotal role in securing a competitive edge through a committed and capable workforce, it becomes evident that the effective integration of AI has the potential to enhance this workforce even further. This discussion affirms that the future of HR lies in its ability to integrate AI strategically while preserving its commitment to human connection. By redefining its role, balancing ethical considerations with operational demands, and advocating for sustainable AI practices, HR can navigate the complexities of AI-driven transformation. Yet, critical questions persist: Are organizations truly ready to prioritize ethical AI over short-term gains? Can HR lead this transformation without systemic support and investment? The answers to these questions will shape the path forward for HR as it evolves to meet the challenges and opportunities of the AI era.

4.4 Limitations

While this thesis provides valuable insights into the adoption and implementation of Responsible AI in HRM, several limitations must be acknowledged. The scope of this research is shaped by the characteristics of the organizations involved. The participating companies varied in size, with employee numbers ranging from approximately 500 up to 50,000. While this diversity provided a broad perspective, most of the organizations were large enterprises,

often exceeding 10,000 employees. This focus intentionally does not reflect the practices or challenges faced by smaller organizations, such as those with fewer than 500 employees, as our target group has been defined to include larger organizations for reasons previously outlined in our methodology.

We aimed to interview HR professionals with a comprehensive understanding of all HR processes within their organizations. However, this was not always feasible. In some instances, the focus had to be restricted to specific HR functions or areas, potentially limiting the breadth of insights. This limitation introduces a degree of uncertainty regarding whether the perspectives provided by these interviewees fully reflect the overall reality within their respective organizations.

Furthermore, despite efforts to design unbiased questions and triangulate data where possible, some level of subjectivity in participants' interpretations of Responsible AI, as well as potential interviewer influence, could not be entirely eliminated.

Additionally, the study examined European companies, with a significant focus on German firms, which limits the representativeness of the findings. German companies operate within unique regulatory and cultural frameworks, including the presence of works councils, which play a significant role in shaping HR policies and the implementation of new technologies like AI. While this context offers valuable insights into German and European practices, it may not be representative of organizational structures or regulatory landscapes in other regions, such as North America or Asia.

Moreover, the emphasis on works councils and European labor relations provides a specific perspective that might not apply universally, particularly in countries where such structures are less common or entirely absent. This focus highlights the need for further research into how AI adoption in HR is influenced by varying organizational and cultural contexts globally.

While over 26 interviews were conducted with HR professionals, the sample is not statistically representative. As such, the findings are more indicative of trends and best practices rather than generalizable conclusions applicable across all industries or regions. The qualitative nature of the research, conducting semi-structured interviews, provided rich and nuanced insights but also introduced potential biases.

Another significant limitation lies in the rapidly evolving nature of AI technology and its regulatory frameworks. This thesis represents a snapshot of current practices and attitudes but may become outdated as new advancements and legal requirements emerge. Moreover, the study primarily focuses on specific HR functions, such as recruitment, training, performance management, and strategic workforce planning. Other areas, like employee well-being or labor relations, received less attention, which may narrow the comprehensiveness of the findings.
While the thesis offers recommendations for implementing Responsible AI in HR, it does not thoroughly address the subsequent steps for effectively managing AI once employed. This includes handling ongoing challenges such as monitoring AI performance, addressing ethical concerns in real-time, and ensuring that AI systems remain aligned with organizational goals and compliance requirements. These areas should be explored in future research to provide a more holistic understanding of Responsible AI adoption and utilization in HRM.

Access to detailed internal data on AI systems and their decision-making processes was often restricted due to confidentiality concerns. This limited the depth of technical analysis and prevented a more thorough evaluation of specific AI tools used in HR. Additionally, there is a potential self-selection bias, as participants who agreed to be interviewed may have been more positively inclined toward AI adoption, potentially skewing the findings toward a more favorable outlook on Responsible AI.

Despite these limitations, the findings contribute valuable knowledge to the field and underscore the importance of further research on Responsible AI in HRM, particularly in understanding how organizations can effectively manage AI technologies in the long term.

5. Conclusion and Future Outlook

This thesis examined the introduction and integration of Responsible AI in HRM, highlighting both opportunities and challenges associated with AI adoption. Drawing on 26 interviews with HR professionals and a systematic review of literature, the study bridges the gap between academic insights and practical applications. While AI offers significant potential to improve efficiency, decision-making, and workforce transformation, its integration into HR requires balancing technological advancements with ethical governance and human-centered values.

To address the first research question *"How and to what extent do HR professionals adopt and integrate Responsible AI solutions in their organizations?"* the study identified six interconnected dimensions that capture the current state of AI adoption and pathways for responsible implementation.

The first two dimensions focus on general AI adoption in HR. The first dimension highlights foundational barriers, such as knowledge gaps, skill shortages, and inadequate infrastructure, which hinder AI readiness and limit progress. The second dimension emphasizes that operational benefits, like efficiency and cost savings, drive AI adoption, with applications already enhancing recruitment, scheduling, and workforce planning. However, HR's relationship-driven processes, such as mentorship and employee engagement, underline the importance of maintaining a "human touch." AI should complement, not replace, these areas to preserve trust, nuance, and personal connection.

The third dimension through the sixth dimension focus on responsible AI implementation, addressing the need for ethical practices and structured governance. The third dimension reveals

that limited awareness, guidelines, and expertise hinder HR professionals from addressing ethical risks like bias and transparency gaps. The fourth dimension underscores the importance of prioritizing fairness, transparency, and data protection but notes that these principles often take a backseat to short-term operational goals. The fifth dimension highlights the role of cross-functional collaboration—between HR, IT, legal teams, leadership, and works councils—in mitigating risks, fostering transparency, and addressing employee concerns. Transparent communication about AI's role and limitations is essential for building trust and reducing resistance. Finally, the sixth Dimension situates AI in HR within a broader transformation process, emphasizing the need for HR professionals to balance innovation with ethical responsibility as AI continues to evolve.

These dimensions collectively demonstrate that AI adoption in HR is an iterative process, requiring alignment between technical advancements and human-centric values. The study also identifies a tension between efficiency-focused adoption in the second dimension and the underprioritization of ethical considerations in the fourth dimension, emphasizing the need for HR to take a proactive leadership role in bridging this gap.

In response to the second research question, *"What strategies should be pursued to responsibly leverage AI opportunities in HR?"* a four-stage framework was developed. This framework outlines actionable steps for HR professionals, beginning with building foundational AI knowledge and awareness to address existing skill gaps. Conducting organizational AI readiness audits helps identify opportunities and barriers, while implementing training and upskilling programs equips HR teams with the necessary technical and ethical competencies. Finally, piloting and scaling AI tools responsibly ensures alignment with organizational values and minimizes risks. Together, these strategies enable HR professionals to adopt AI ethically and effectively, driving workforce transformation while maintaining trust and accountability.

The study contributes to existing Responsible AI frameworks by applying them to HRM, providing a grounded perspective on both adoption challenges and ethical pathways. The findings also expand strategic HRM theories by positioning HR as a mediator between technology, ethics, and human-centric values.

AI is poised to redefine the world of work, offering both immense opportunities and significant challenges for HR. It is essential for HR professionals to approach its implementation with caution and responsibility, prioritizing ethical considerations such as fairness, inclusivity, and transparency. Identifying appropriate use cases is crucial to maximizing AI's positive impact while minimizing potential harm. AI should not replace the human element in HR but rather complement it, recognizing the importance of human touch and authentic connections that are essential in managing and supporting people. While AI empowers HR professionals to optimize processes, make data-driven decisions, and unlock new opportunities for talent development and engagement, it should serve as a tool to enhance—not substitute—the empathy, trust, and interpersonal relationships that define effective HR practices. Future research should explore the long-term impacts of AI on workforce dynamics, including employee well-being, labor relations, and organizational culture. Additionally, examining how smaller organizations and diverse sectors address Responsible AI would provide further insights into scalable governance practices. Developing industry-wide standards and certifications for Responsible AI will remain essential to fostering trust and accountability across industries.

In conclusion, this thesis demonstrates that the adoption and integration of Responsible AI in HRM requires a human-centered commitment to balance innovation with ethics. The coming years will determine whether AI serves as a tool for empowerment or exacerbates inequalities—an outcome that depends on the leadership, responsibility, and vision of HR professionals.

References

Ali, Sajid, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. "Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence." *Information Fusion* 99 (2023): 101805. https://doi.org/10.1016/j.inffus.2023.101805.

Anagnostou, Marianna, Olga Karvounidou, Chrysovalantou Katritzidaki, Christina Kechagia, Kyriaki Melidou, Eleni Mpeza, Ioannis Konstantinidis, Eleni Kapantai, Christos Berberidis, Ioannis Magnisalis, and Vassilios Peristeras. "Characteristics and Challenges in the Industries Towards Responsible AI: A Systematic Literature Review." *Ethics and Information Technology* 24, no. 1 (2022): 37–50. https://doi.org/10.1007/s10676-022-09634-1.

ArtificialIntelligenceAct.eu. "High-Level Summary of the EU AI Act." Accessed November 14, 2024a. https://artificialintelligenceact.eu/high-level-summary/.

ArtificialIntelligenceAct.eu. *Full Text of the EU AI Act.* Accessed November 14, 2024b. https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AI-Act-FullText.pdf.

Bankins, Sarah. "The Ethical Use of Artificial Intelligence in Human Resource Management: A Decision-Making Framework." *Ethics and Information Technology* 23, no. 4 (2021): 841–54. https://doi.org/10.1007/s10676-021-09619-6.

Bankins, Sarah, Patrick Formosa, Yannick Griep, and Denise Richards. "AI Decision Making with Dignity? Contrasting Workers' Justice Perceptions of Human and AI Decision Making in a Human Resource Management Context." *Information Systems Frontiers* 24, no. 3 (2022): 857–75. https://doi.org/10.1007/s10796-021-10223-8.

Bolón-Canedo, Verónica, Laura Morán-Fernández, Brais Cancela, and Amparo Alonso-Betanzos. "A Review of Green Artificial Intelligence: Towards a More Sustainable Future." *Neurocomputing* 599 (2024): 128096. https://doi.org/10.1016/j.neucom.2024.128096.

Capgemini. *Our Code of Ethics for AI.* Paris: Capgemini, 2021. https://www.capgemini.com/wp-content/uploads/2021/06/Capgemini_Code_of_Ethics_for_AI_2021_EN.pdf.

Calvard, Thomas, and Donald Hislop. "Artificial Intelligence and the Changing Nature of Work: The Benefits, Challenges, and Risks of AI for Jobs, Workers, and Organizations." *The International Journal of Human Resource Management* 33, no. 16 (2022): 3317–36. https://doi.org/10.1080/09585192.2022.2035161.

Cisco. *The Responsible AI Framework.* San Jose, CA: Cisco Systems, Inc., 2024. https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-responsible-artificial-intelligence-framework.pdf.

Clarke, Roger. “Principles and Business Processes for Responsible AI.” *Computer Law & Security Review* 35 (2019): 410–22. https://doi.org/10.1016/j.clsr.2019.04.007.

Council of the European Union. “AI Explained: Uses and Impact.” *Consilium.* Accessed October 11, 2024. https://www.consilium.europa.eu/en/policies/ai-explained/.

Cox, Josie. “AI Anxiety: The Workers Who Fear Losing Their Jobs to Artificial Intelligence.” *BBC Worklife,* April 18, 2023. Accessed October 11, 2024. https://www.bbc.com/worklife/article/20230418-ai-anxiety-artificial-intelligence-replace-jobs.

Deloitte. “Trustworthy AI™.” *Deloitte Development LLC.* Accessed November 14, 2024. https://www2.deloitte.com/us/en/pages/deloitte-analytics/solutions/ethics-of-ai-framework.html.

Ekuma, Kelechi. “Artificial Intelligence and Automation in Human Resource Development: A Systematic Review.” *Human Resource Development Review* 23, no. 2 (2023): 199–229. https://doi.org/10.1177/15344843231224009.

European Commission. *Ethics Guidelines for Trustworthy AI,* April 8, 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

European Commission. “Regulatory Framework for Artificial Intelligence.” Accessed November 14, 2024a. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

European Commission. *Prohibitions and Definition Survey 2024.* Accessed November 14, 2024b. https://ec.europa.eu/eusurvey/runner/Prohibitions-and-Definition-Survey-2024.

European Union. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 19 June 2024 on Artificial Intelligence and Amending Certain Union Legislative Acts. Official Journal of the European Union.* https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

EY. “EY’s Commitment to Developing and Using AI Ethically and Responsibly.” *EY Global,* September 28, 2023. https://www.ey.com/en_it/insights/ai/principles-for-ethical-and-responsible-ai.

Fenwick, Ali, Gabor Molnar, and Piper Frangos. “The Critical Role of HRM in AI-Driven Digital Transformation: A Paradigm Shift to Enable Firms to Move from AI Implementation to Human-Centric Adoption.” *Discover Artificial Intelligence* 4 (2024): Article 34. https://doi.org/10.1007/s44163-024-00125-4.

Google. “AI Principles.” Mountain View, CA: Google LLC, 2018. https://ai.google/principles/.

Google Cloud. “What Is Artificial Intelligence?” *Google Cloud.* Accessed October 1, 2024. https://cloud.google.com/learn/what-is-artificial-intelligence.

Greenwood, Michelle, and R. Edward Freeman. “Ethics and HRM: The Contribution of Stakeholder Theory.” *Business and Professional Ethics Journal* 30, no. 3–4 (2011): 269–92. https://doi.org/10.5840/bpej2011303/413.

IEEE. *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, December 12, 2016. https://standards.ieee.org/wp-content/uploads/import/documents/other/ec_about_us.pdf.

IEEE. *IEEE Standard for Transparent Employer Data Governance.* IEEE Std 7005-2021, November 19, 2021. https://ieeexplore.ieee.org/document/9618905.

IEEE. “IEEE Membership.” Accessed November 14, 2024. https://www.ieee.org/membership/index.html.

Jia, Xiao, and Yanghong Hou. “Architecting the Future: Exploring the Synergy of AI-Driven Sustainable HRM, Conscientiousness, and Employee Engagement.” *Discover Sustainability* 5, no. 30 (2024). https://doi.org/10.1007/s43621-024-00214-5.

Kaur, Sharda. “Embracing Responsible AI: A Comprehensive Guide and Call to Action.” *Educator Developer Blog*,December 11, 2024. https://techcommunity.microsoft.com/blog/educatordeveloperblog/embracing-responsible-ai-a-comprehensive-guide-and-call-to-action/4355454.

Kowald, Dominik, Sebastian Scher, Viktoria Pammer-Schindler, Peter Müllner, Kerstin Waxnegger, Lea Demelius, Angela Fessl, Maximilian Toller, Inti Gabriel Mendoza Estrada, Ilija Šimić, Vedran Sabol, Andreas Trügler, Eduardo Veas, Roman Kern, Tomislav Nad, and Simone Kopeinik. “Establishing and Evaluating Trustworthy AI: Overview and Research Challenges.” *arXiv* (2024). https://arxiv.org/pdf/2411.09973.

Microsoft. *Microsoft Responsible AI Standard, v2.* Redmond, WA: Microsoft Corporation, 2022. https://www.microsoft.com/en-us/ai/responsible-ai.

Mökander, Jakob, and Luciano Floridi. “Operationalising AI Governance through Ethics-Based Auditing: An Industry Case Study.” *AI and Ethics* 3, no. 3 (2022): 451–68. https://doi.org/10.1007/s43681-022-00171-7.

Nocker, Manuela, and Vania Sena. “Big Data and Human Resources Management: The Rise of Talent Analytics.” *Social Sciences* 8, no. 10 (2019): 273. https://doi.org/10.3390/socsci8100273.

OECD. *OECD Principles on Artificial Intelligence.* Paris: OECD, 2019. https://www.oecd.org/en/topics/ai-principles.html.

Patel, Rishika. “With the Rise of AI-Powered Recruitment: What Data Privacy and Ethical Considerations Should HR Teams Follow?” *TecHRseries*, October 4, 2024. https://techrseries.com/featured/with-the-rise-of-ai-powered-recruitment-what-data-privacy-and-ethical-considerations-should-hr-teams-follow/.

Rai, Arun. “Explainable AI: From Black Box to Glass Box.” *Management Science* 66, no. 3 (2020): 1–10. https://doi.org/10.1007/s11747-019-00710-5.

Robert, Laura P., Charles Pierce, Lauren Marquis, Soeun Kim, and Reem Alahmad. "Designing Fair AI for Managing Employees in Organizations: A Review, Critique, and Design Agenda." *Human–Computer Interaction* 35, no. 5–6 (2020): 545–75. https://doi.org/10.1080/07370024.2020.1735391.

Rotman, David. "People Are Worried That AI Will Take Everyone's Jobs. We've Been Here Before." *MIT Technology Review*, January 27, 2024. Accessed December 11, 2024. https://www.technologyreview.com/2024/01/27/1087041/technological-unemployment-elon-musk-jobs-ai/.

Schiff, David, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, and Michael Lennon. "Principles to Practices for Responsible AI: Closing the Gap." *arXiv*, 2006.04707 (2020). https://doi.org/10.48550/arXiv.2006.04707.

Staton, Bethan. "Employers Look to AI Tools to Plug Skills Gap and Retain Staff." *Financial Times*, November 7, 2024. https://www.ft.com/content/9cf58a76-5245-4cdf-9449-239e90077eb5.

Storey, John. *Human Resource Management: A Critical Text.* London: Routledge, 1995.

Tilmes, Nicholas. "Disability, Fairness, and Algorithmic Bias in AI Recruitment." *Ethics and Information Technology*24, no. 2 (2022): 1–13. https://doi.org/10.1007/s10676-022-09633-2.

UNESCO. *Recommendation on the Ethics of Artificial Intelligence.* SHS/BIO/PI/2021/1. Paris: UNESCO. 2022. https://unesdoc.unesco.org/ark:/48223/pf0000381137.

Verma, Prikshat, Parth Patel, Arup Varma, and Alessio Ishizaka. "A Multi-Stakeholder Ethical Framework for AI-Augmented HRM." *International Journal of Manpower* 43, no. 1 (2022): 226–44. https://doi.org/10.1108/IJM-03-2021-0118.

Appendix

Appendix 1: Overview Interviewees

Interview Person	Role	Years of Experience	Industry	Number of Employees	Country
IP 01	Senior Level HR	~20 years	Astronomical Research Organization	~750	Germany
IP 02	Talent & Leadership Development Manager	~20 years	Retail and Consumer Goods	~50,000 (Group)	Portugal
IP 03	CHRO	~30 years	Navigation Software Provider	~500	Hungary
IP 04	Head of HR	~17 years	Enterprise Software Leader	~25,300	Germany
IP 05	Expert Talent Acquisition & Branding	~15 years	Media and Broadcasting Company	~12,800	Germany
IP 06	HR Digitalisation Consultant	~15 years	Global Battery Cell Manufacturer	~1,500	Germany
IP 07	Transformation and Talent, HR BP Research & Development	~28 years	Pharmaceutical and Life Sciences Company	~22,000	Germany
IP 08	HR Strategy and Project Manager	~5 year	Elevator and Escalator Manufacturer	~4,300	Germany
IP 09	Senior HR Change Management Consultant	~1 year	IT and Consulting Firm	~6,000	Germany
IP 10	HR Business Partner	~35 years	Banking and Financial Services Provider	~6,000	Germany
IP 11	Director People and Culture	~16 years	ICT Services Provider	~700	Germany
IP 12	Senior Advisor in the EMEIA People Consulting Center of Excellence for Digital HR	~34 years	Global Professional Services Firm	~1,000 – 2,000	Portugal
IP 13	HR Expert	~6 years	Insurance Provider	~17,200	Germany
IP 14	Talent Acquisition Partner	~4 years	Process Mining and Optimization Company	~2,800	Germany
IP 15	HR Business Partner	~6 years	Audit and Consulting Services Firm	~15,200	Germany
IP 16	Senior Manager in HR Consulting	~10 years	Management Consulting and Professional Services Firm	~1,001 – 5,000	Portugal
IP 17	Head of People and Services	~16 years	Procurement and Supply Chain Consultancy	~500	Germany
IP 18	Senior Manager for Employee Eyesperience & HR	~19 years	Digital Transformation and IT Consulting Firm	~16,000	Germany
IP 19	HR Manager	~15 years	Food Distribution and Retail	~33,000	Portugal
IP 20	HR Manager	Anonymous	Anonymous	~37,000	Germany
IP 21	Ethical Committee Member	~20 years	Enterprise Software and Cloud Solutions Leader	~25,300	Germany
IP 22	HR Manager and Recruiting Expert	~20 years	Healthcare Provider	~1,800	Germany
IP 23	HR Business Partner	~24 years	Risk Management and Insurance Brokerage	~1,400	Germany
IP 24	Change Management Manager	~3 years	Global Professional Services Firm	~11,100	Germany
IP 25	Talent Acquisition and Sourcing	~4 years	HR and Payroll Software Provider	~1,001 – 5,000	Germany
IP 26	Member of the Board Human Resources	~28 years	German Automotive Producer	~13,000	Slovakia

Appendix 2: Interview Guide

Introduction

We are conducting a study to understand how HR professionals are adopting responsible AI in their processes and the associated challenges and opportunities.

- **Research context:** We are using Grounded Theory to analyze the Status Quo and develop theories & new frameworks based on the experiences of interviewees.
- Consent for transcribing the interview

1. Current Use and Non-Use of AI in HR

Main Question: How is AI currently being used in HR processes within your organization?

- How is AI currently being used in HR processes within your organization?
- **If not using AI:** What are the reasons for non-adoption (e.g., cost, complexity, ethical concerns)? What would be needed to consider AI adoption in the future?

2. Motivations and Drivers for AI Adoption

Main Question: What are the main drivers and challenges for adopting AI in your HR processes?

- Are there specific business objectives tied to AI adoption in HR?
- Do you see AI as enhancing fairness, transparency, or achieving business goals?

3. Challenges, Ethical Concerns, and Regulatory Compliance

Main Question: How does your organization ensure AI in HR aligns with ethical standards, and how do you manage regulatory guidelines?

- What challenges or concerns have you encountered in integrating AI into HR?
- How does your organization ensure compliance with regulatory frameworks and guidelines for AI in HR?

4. The Role of Human Intelligence and Future of AI in HR

Main Question: How do you see the balance and future evolution of human judgment and AI collaboration in HR decision-making processes?

- How do you see the role of human decision-making vs. AI in HR processes?
- How do you foresee AI evolving in HR over the next few years?

Closing:

- Is there anything else you would like to share regarding AI in HR that we haven't covered?

Appendix 3: Expert Insights - Questionnaire AI Act for HR use-cases

António Novais:

Unbabel | Center for Responsible AI | BridgeAI Portugal

Iakovina Kindylidi:

Vieira de Almeida & Associados | Center for Responsible AI | BridgeAI Portugal

Introduction

The AI Act (Regulation (EU) 2024/1689) is the first comprehensive legal framework on artificial intelligence (AI). The AI Act is characterized by its horizontal application, meaning it applies uniformly across all sectors, imposing different obligations on various stakeholders within the AI value chain regardless of the specific industry. Nonetheless, most obligations are posed to:

- The "provider" of the AI system, which is the company that develops and places on the market the AI system under its name/trademark.
    - Example: A software company develops an AI-powered resumé screening tool for recruitment purposes and offer it to businesses across the EU as a SaaS solution
- The "deployer", which can be a company/professional that uses an AI system created and provided by others and integrates in their own processes/products.
    - Example: A large bank purchases the AI resumé screening tool and implements it in their HR department to assist with initial candidate screening for job openings

Moreover, its extraterritorial reach ensures that it applies to AI systems outside the EU, provided they are placed on the EU market, or their outputs are used within the EU.

The AI Act adopts a risk-based approach, categorizing AI systems according to their potential impact on health, safety, and fundamental rights. AI systems that violate EU fundamental rights and values are deemed to pose "unacceptable risks" and are prohibited. High-risk AI systems—product- or service-based—offer significant benefits but also substantial risks, making them the primary focus of the Act. These high-risk systems are regulated based on their applications rather than the underlying technology, with specific use cases, such as AI in HR tasks, being identified as areas where risks to citizens demand stringent obligations.

The Act also addresses risks associated with AI systems that interact with people or generate content, such as impersonation, manipulation, or deception. Providers and deployers of these systems must comply with specific transparency requirements. Additionally, the Act introduces specific rules for providers of general-purpose AI models, such as large language models (LLMs), which can be used to develop various AI systems.

Based on that qualification of risk, the Regulation introduces different obligations for the stakeholders involved in the value chain of those AI models or systems.

Topic 1: Relevance of the AI Act in HR

- Q: How do you assess the relevance of the AI Act for HR departments (or in general for organizations), especially regarding the adoption and use of responsible AI solutions?
- Q: What legal challenges or compliance risks could arise from the AI Act for HR departments?

Answer: The AI Act will significantly impact HR departments and strategies across organizations, regardless of sector. It directly addresses the use of AI in workplace contexts in two key scenarios:

i. Prohibition on emotion recognition systems in the workplace: AI systems designed to infer emotions in the workplace are prohibited under the Act, with compliance required by February 2, 2025. Organizations using or developing such systems must remove them from the market or cease their use by this date. The prohibition stems from concerns about the lack of scientific robustness of these systems and the risk of discriminatory and intrusive outcomes in asymmetric relationships, such as those between employers and employees. Exceptions are permitted for medical or safety-related applications, but further guidance is needed to clarify the boundaries of lawful use. To address this, the European Commission has initiated stakeholder consultations to collect insights and issue practical guidelines on prohibited use cases.

ii. High-risk AI systems in employment, workers management and access to self-employment: AI systems intended to be used for (a) recruitment or selection of natural persons, in particular to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates; as well as AI intended to be used to (b) make decisions affecting terms of the work-related relationships, the promotion and termination of work-related contractual relationships, (c) to allocate tasks based on individual behaviour or personal traits or characteristics or to monitor and (d) evaluate the performance and behaviour of persons in such relationships, are high risk. Providers and deployers of high-risk AI systems in HR must adhere to obligations such as risk management system throughout the AI lifecycle, data governance of the personal and non-personal data used for the training and use of the system, and implement human supervision measures, among others. Deployers are also required to notify employee representatives before deploying high-risk AI systems. At a national level, this aligns

with recent amendments to the Portuguese Labour Code, which introduced similar obligations. Most obligations under the AI Act become enforceable on August 2, 2026, providing organizations ample time to align their operations. However, organizations should begin preparing their internal AI governance frameworks to ensure compliance with the AI Act and related legal requirements based on their role in the AI value chain (provider or deployer).

Non-compliance with the obligations mentioned above may lead to very high fines. In relation to scenario (i), if a provider or deployer of an emotion recognition system in the workplace continue to make available on the market or continue to use that system after 2 February 2025, it will be subject to fines rising up to 35 million euros or 7% of the worldwide annual turnover of the preceding financial year, whichever is higher. In the case of scenario (ii), non-compliance with the obligations applicable to high-risk AI system providers and deployers may lead to fines up to 15 million or 3% of the worldwide annual turnover of the preceding financial year, whichever is higher.

Although implementation of the AI Act is still in its early stages, depending on the role that organizations assume in the AI value chain (i.e., provider or deployers) will need to start preparing their internal AI governance to ensure compliance with the AI Act and the applicable legal framework. As first step, organisation should start by mapping the AI systems in use and understanding the risk level of these systems will be a key first step. In HR, it is common for organizations or HR departments to be unaware of the use of AI tools. For instance, a software-based automated solution that has been in use for years may have undergone updates to incorporate advanced AI systems without recognition or understanding by the organization. Moreover, until now, assessing the risks of AI systems was fragmented: some companies invested in risk assessment while others, fell on *ethics-washing* practices, while the

majority took a reactive and limited approach by tackling, for example, data protection and privacy-related concerns while not addressing other risks to health, safety, and fundamental rights of AI solutions.

Topic 2: Implementation and Compliance Strategies

- Q: What steps should companies take to ensure their AI solutions in HR comply with the AI Act requirements?
- Q: Are there specific guidelines or best practices you would recommend to HR managers to meet the regulatory standards of the AI Act?

Answer: The first key step is to map the AI use case, confirm whether it is subject to the AI Act, and classify its risk level. If the use case falls within the prohibited list, the AI system should be removed from the market or discontinued as soon as possible. For other use cases, the organization should identify its role in the AI value chain to determine its obligations, as well as those of its partners. This includes conducting a risk assessment to evaluate the severity and likelihood of potential damage to health, safety, and fundamental rights. It should also identify the relevant technical and organizational risk management measures for any residual risks after initial mitigation efforts.

However, it is important to note that the risk assessment should be comprehensive, not limited to compliance with the AI Act, but also considering applicable national and European frameworks. In HR, for instance, the analysis should also address risks related to data protection and labour law. Moreover, risk assessments should not be a one-off exercise, especially in AI systems, which can evolve over time. Therefore, risk assessments should be revisited at key stages of the AI lifecycle.

After this initial assessment, organizations will need to implement AI governance to ensure compliance with applicable obligations. A one-size-fits-all approach is not suitable for AI governance, but typically, these activities include identifying an internal multidisciplinary team responsible for human supervision, updating policies and contracts, creating new internal processes, and aligning AI Act obligations with other regulatory requirements.

To facilitate the implementation of these measures and compliance with the AI Act, the European Commission will publish relevant guidelines, and European Standards Organisations (such as CEN and CENELEC) are developing harmonized standards.

To ensure the success of this risk management approach, digital literacy is essential. Beyond being an obligation under the AI Act, AI literacy helps demystify the complex regulation, promote a responsible AI culture, and address common misconceptions surrounding the Act. AI literacy also fosters collaboration: (a) internal collaboration across different departments, including HR, is key to meeting obligations and managing risks; (b) external collaboration with stakeholders in the AI ecosystem is crucial for exchanging information, sharing experiences, and staying updated on ethical practices and state-of-the-art research.

Topic 3: Ethics and Bias in AI Systems for HR

- Q: How does the AI Act address ethical aspects, such as discrimination or bias in AI systems, particularly in HR processes like recruitment and performance evaluation?
- Q: What legal mechanisms exist to ensure AI systems are used ethically and without bias in HR processes?

Answer: The AI Act incorporates the foundational work of the High-level Expert Group on AI (AI HLEG), created in 2018 by the European Commission. The AI HLEG was an autonomous body, consisting of 52 members, many of which coming from the technology industry in 2018

to support the Commission's objectives on AI regulation set out in its Communications dated 25th April 2018 and 7th December 2018. The Commission's vision was to build upon three fundamental pillars: (i) amplifying both public and private investments in AI to enhance its adoption, (ii) proactively addressing socio-economic transformations, and (iii) establishing a fitting ethical and legal structure to reinforce European values. In this regard, the AI HLEG published in 2019 the Trustworthy AI Guidelines, setting out the cumulative requirements for a trustworthy AI: (i) lawful, meaning that it complies with the applicable laws and regulations; (ii) ethical, meaning that it observes ethical principles and values, and (iii) technically and socially robust. For these components to materialize, a set of core ethical principles, as well as seven requirements based on technical and non-technical methods, should be met. Following the public consultation on its Trustworthy AI guidelines, AI HLEG published in 2020 its Assessment List for Trustworthy AI (ALTAI), a self-assessment checklist, in order to further provide guidance to stakeholders of the AI ecosystem when assessing the ethical principles outlined in the guideline.

In the context of HR processes, ethical principles and fundamental rights are safeguarded by recognizing that AI tools used for emotion recognition in the workplace pose unacceptable risks to individual rights and European values. Furthermore, the AI Act classifies the use of AI in HR processes as high risk, as discussed in the response to Topic 2 above.

While the AI Act outlines various obligations aimed at protecting health, safety, and fundamental rights in AI system deployment, certain mechanisms are especially relevant for managing errors and biases. Importantly, the AI Act acknowledges the presence of data and historical biases in algorithms and seeks to minimize these biases where eradication is not feasible, considering the specific use cases.

To manage these risks in the HR process, the following mechanisms apply to providers and deployers of AI systems:

Obligations of Providers	Obligations of Deployers
Data Governance: Ensure quality datasets for training/testing that are relevant, representative, error-free, and complete[1]. More specifically, the datasets should have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used. As such, to detect possible biases in the datasets, providers can exceptionally process special categories of personal data if there are subject to adequate protection and a series of requirements are met.	**Data Quality[2] and Data Protection[3]:** Ensure input data is relevant and representative for the system's purpose if controlled by deployers. Moreover, deployers should use provider information to carry data protection impact assessments, where applicable.
Human Oversight: Enable effective oversight proportional to the system's risks and autonomy level[4].	**Human Oversight:** Assign qualified individuals for oversight, ensuring they have training, authority, and support[5].
Quality Management System: Implement a documented quality management system for regulatory compliance, design, risk management, and monitoring[6].	**Employees Involvement:** Inform employees and representatives prior to workplace deployment of high-risk AI systems[7].
Transparency and Information Provision: Provide accessible and understandable instructions for deployers[8].	**Right to Explanation:** Inform individuals when subject to decision-making AI systems, except in certain law enforcement cases[9].
Corrective Actions and Information Duty: Take corrective actions for non-compliance and inform relevant parties and authorities of risks or incidents[10].	**Fundamental Rights Impact Assessment:** Conduct assessments of fundamental rights impacts before deploying certain high-risk AI systems[11].

[1] See Article 10 of the AI Act.
[2] See Article 26(4) of the AI Act.
[3] See Article 26(9) of the AI Act.
[4] See Article 14 of the AI Act.
[5] See Article 26(2) and (3) of the AI Acy.
[6] See Article 17 of the AI Act.
[7] See Article 26(7) of the AI Act.
[8] See Article 13 of the AI Act.
[9] See Article 26(11) and Article 86 of the AI Act.
[10] See Article 20 of the AI Act.
[11] See Article 27 of the AI Act.

Topic 4: Long-Term Impact of the AI Act on HR Strategy

- Q: What long-term impact do you foresee for HR strategy as the AI Act becomes more established? How should companies prepare for regulatory changes in responsible AI?

Answer: Considering that we are in the early stages of implementing the AI Act, it is not possible to confidently predict the long-term impact on HR strategy. However, it is unlikely that the AI Act will reduce the adoption of AI systems in HR processes or diminish the advantages of their use cases. On the contrary, the AI Act is expected to foster trust in AI applications within HR, potentially improving decision-making and addressing human errors and biases. As such, itis important to look at risk as a spectrum and manage it. Shifting from a reactive to a proactive approach towards AI risks will be a key for creating AI systems that are not just innovative but also responsible.

Topic 5: Importance of Human Oversight and Transparency

- Q: To what extent does the AI Act mandate human oversight and transparency in the use of AI in HR? What challenges might this pose in practice?
- Q: What measures can HR departments take to ensure transparency and human control in line with the AI Act's requirements?

Answer: Human oversight and transparency throughout the AI lifecycle are key obligations in the Act, particularly for providers and deployers of high-risk AI systems, as mentioned in Topic 3.

When it comes to human oversight of high-risk AI systems, these systems must be designed with consideration for their risks, autonomy level, and context to ensure effective human oversight. This is crucial to minimize risks to health, safety, or fundamental rights during use,

including in cases of foreseeable misuse. Given the technical complexity of AI, the oversight team must understand the system's capabilities, monitor for anomalies, and address performance issues by overriding or halting the system when necessary.

For this to be effective, the teams responsible for human oversight—including representatives from HR departments when dealing with HR systems—must possess sufficient AI literacy. This includes training on the rules set out in the AI Act, as well as an awareness of automation bias (over-reliance on AI outputs), to monitor and intervene in the AI system's operations effectively.

In terms of transparency, the obligations for providers and deployers of high-risk AI systems differ. Providers are required to ensure that AI systems are designed to enable transparency for deployers, while deployers must fulfil two types of transparency obligations. First, internal transparency refers to the obligation to inform employees and involve their representatives before deploying AI systems, particularly in HR contexts. Second, external transparency involves informing individual users who may be adversely affected by decisions made using the outputs from high-risk AI systems. These transparency obligations complement existing legal frameworks, such as the GDPR.

Specifically, providers of high-risk systems must ensure the AI system is designed to provide sufficient transparency, enabling deployers to understand and use the system's outputs correctly. Providers should share comprehensive, clear, and accessible instructions, including illustrative examples, to help deployers use the system safely and comply with their own obligations, including transparency-related requirements.

On the other hand, deployers of these systems must inform workers and their representatives prior to deployment, in addition to complying with other applicable laws. This includes

providing details about the workers subject to the AI system's use, to protect their fundamental rights and complement existing labour rights. Furthermore, deployers must inform individuals who may be significantly impacted by decisions made using AI outputs, such as those in HR for recruitment or performance evaluation, about the role of the AI system in the decision-making process. This ensures individuals can exercise their rights effectively.

In light of the above, it is clear that the human oversight and transparency obligations outlined in the AI Act can be time- and resource-intensive for organizations. Promoting AI literacy, including legal literacy, is therefore essential to support the activities undertaken by organizations, especially HR departments. This process requires a proactive and collaborative approach from organizations, as discussed in Topic 1.

Topic 6: General Question

- Q: How important would you say the implementation of the AI Act is in establishing responsible AI practices?
- Q: Do you believe the Act alone provides sufficient guidelines to ensure responsible AI use across companies, or is additional action needed?

Answer: The AI Act marks a pivotal moment in shaping the future of AI, as it represents the first comprehensive regulatory initiative for AI. The AI Act aims to foster the development and adoption of safe AI, safeguard the fundamental rights of EU citizens, and position the EU as a global leader in AI regulation. The Act highlights that responsible innovation can serve as a differentiator, offering competitive advantages to organizations, building trust in AI, and protecting rights and freedoms.

Adopting a principle-based approach, similar to other product safety regulations under the New Legislative Framework, the AI Act is supported by implementation guidelines from the European Commission. Additionally, the European Standards Organisations (CEN and CENELEC) are developing harmonized standards to aid compliance for stakeholders across the AI ecosystem.

The global influence of the AI Act is evident, with international, regional, and national initiatives following its example. Among these, the Council of Europe Framework Convention on Artificial Intelligence, Human Rights, and the Rule of Law stands out as the first legally binding treaty in the field. Adopted on 17 May 2024 and opened for signature on 5 September 2024, this landmark treaty sets a global standard for AI governance, ensuring alignment of the AI lifecycle with human rights, democratic principles, and the rule of law.

Appendix 4: Interview Transcripts

Due to the length of the interview transcripts and concerns regarding space and data size, and in consultation with our supervisor, the transcripts are not attached to this document. However, they are available upon request, except for the interview with IP 20.

Transcript IP 1

[available upon request]

Transcript IP 2

[available upon request]

Transcript IP 3

[available upon request]

Transcript IP 4

[available upon request]

Transcript IP 5

[available upon request]

Transcript IP 6

[available upon request]

Transcript IP 7

[available upon request]

Transcript IP 8

[available upon request]

Transcript IP 9

[available upon request]

Transcript IP 10

[available upon request]

Transcript IP 11

[available upon request]

Transcript IP 12

[available upon request]

Transcript IP 13

[available upon request]

Transcript IP 14

[available upon request]

Transcript IP 15

[available upon request]

Transcript IP 16

[available upon request]

Transcript IP 17

[available upon request]

Transcript IP 18

[available upon request]

Transcript IP 19

[available upon request]

Transcript IP 20

[NOT available due to company's data protection rules]

Transcript IP 21

[available upon request]

Transcript IP 22

[available upon request]

Transcript IP 23

[available upon request]

Transcript IP 24

[available upon request]

Transcript IP 25

[available upon request]

Transcript IP 26

[available upon request]