# AI Ethics Auditing Project: Enterprise Workforce Optimization Systems

Evaluating Ethical Risks in AI-Driven Budget and People Management Tools

VAISHNAVI KANDIKONDA

MSc. Information Systems, University College Dublin, vaishnavi.kandikonda@ucdconnect.ie

The use of artificial intelligence in the enterprise setting has significantly changed the old system of workforce management, budgets involved in projects, and cost management especially in IT sector. The most prominent industry players such as Workday and SAP SuccessFactors have also added powerful AI functionalities that enable them to predict employee turnover, staffing adjustments, and provide input into financial decisions based on higher-order analytics. Although these innovations hold great potential in terms of efficiency and profitability, a set of complex and demanding ethical issues comes along.

This paper conducted an in-depth audit of an artificial intelligence driven enterprise optimization platform basing its analysis using known ethical frameworks including the General Data Protection Regulation (GDPR), European Artificial Intelligence Act, and the enterprise industry specific ethical frameworks. The article dwelled on such essential problems as algorithmic bias in the utilization of workforce, transparency of cost-optimization rationale, danger of misuse of individually identifiable information, and the general lack of explainability of the systems output. By means of simulated experiments, we discovered the differences in the distribution of project assignment among the demographic categories, a high degree of reliance on AI-generated recommendations to be interpreted in the murky manner, the shortage of ways in which users were free to provide feedback and the possibility to be non-compliant with the Article 22 of the GDPR.

These findings highlight the need to address the issue of using enterprise AI systems more consistently than today with increased levels of transparency, fairness, and accountability, which should apply particularly in areas involving a high-stakes decision involving personnel and budgeting. The structural controls and accountable governance systems should be maintained at high level so that the efficiency seeking process should not have a cost of ethical responsibility.

**CS CONCEPTS: 1.** Computing methodologies → Artificial intelligence → Machine learning → Fairness and bias mitigation **2.** Social and professional topics → Computing / technology policy → AI ethics → Algorithmic accountability **3.** Applied computing → Enterprise systems → Workforce management and optimization

**Additional Keywords and Phrases:** AI auditing, workforce optimization, algorithmic bias, fairness auditing, human-in-the-loop, GDPR compliance, ethical AI governance, transparency, retrospective audits, counterfactual testing, explainable AI, enterprise AI ethics

**ACM Reference Format:**

Vaishnavi Kandikonda. 2025. AI Ethics Auditing Project: Enterprise Workforce Optimization Systems – Evaluating Ethical Risks in AI-Driven Budget and People Management Tools. (July 2025), 12 pages.

## 1 INTRODUCTION

The implementation of artificial intelligence in workforce management of enterprises has essentially transformed the management of resources by the IT companies in terms of balancing between allocating financial respices, planning budgeting, and efficiency of operations. Such platforms integrating technologies as Workday and SAP SuccessFactors have advanced machine learning algorithms that help automate the process of employee shift work and project distribution, future performance predictions, and budgetary allocations. The successful use of such technologies is evidenced by such major organizations as Amazon that is using predictive analytics to determine labor requirements in fulfillment centers and

IBM that applies AI to forecast employee turnover and suggest retention strategies. Resulting effects are attained as an increase in efficiency, the reduction of human error, and huge savings.

However, when such important decisions are left to the algorithms, they are met with a slew of new issues. The main concern has been raised about the fairness, transparency, and accountability of such systems, particularly, as the systems are started being used towards making promotions, allocating resources, and restructuring the workforce. The biases in historical HR data are one of the most serious risks because they may just foster or even increase inequalities existing toward disadvantaged groups. The algorithmic decision-making lack of transparency is another urgent case because employees lack trust, particularly when an employee is not provided with specific reasons explaining the outcomes (high-consequential cases). Moreover, gathering and processing of the sensitive employee data can violated privacy laws including the General Data Protection Regulation (GDPR).

Such problems are compounded by the recent developments in the regulatory sphere with the AI Act published by the EU that lays the ground of treating AI systems employed in employment as high-risk and imposing a high level of monitoring, transparency and human decision-making. Considering these considerations, an adequate ethical analysis of the AI-based workforce management systems should be performed regarding whether they will be fair and comply with the law and other issues of their overall influence on society.

## 2 REVIEW OF CURRENT APPROACHES TO AI AUDITING

### 2.1.1 AI in Workforce Allocation: High-Risk Implications and the Need for Mixed Governance

According to Gerards and Zuiderveen Borgesius [10], workload distribution in the enterprise, budget-related layoffs, or promotion evaluations belong to the type of activities of enterprise AI systems in the process of HR decision-making, which are subject to the EU Artificial Intelligence Act as one of the high-risk types of work. The explanation is simple: these systems immediately determine the life of employees in their careers, and so they need high ethical requirements and human control.

However, in giant companies that apply such programs, such as Workday, it is practically impossible to anticipate that every piece of automatized guidance would be reviewed by supervisors daily, even when it comes to issues about project allocation or cost-related personnel changes. Thus, retrospective auditing has turned out to be the most common practice: organizations periodically analyze bias and evaluate the results of the workforce to find systematic disparities.

Although this type of audit is useful in identifying trends of injustice, it cannot stop any imminent injustices, including unfair exclusions or biased transfers of projects, because these decisions can impact people well before the audit finds them. Thus, a stronger and mixed-governing model is necessary. This must involve well-specified appeal mechanisms, a clear documentation of any automated decision-making, and a regular HR-readiness assessment to guarantee that any speculation made by the AI does not contravene any notions of fairness or otherwise fulfil all the requirements of the employment law.

### 2.1.2 Mitigating Bias in HR Algorithms: The Role of Counterfactual Analysis

Loukides et al. [11] suggest how to develop counterfactual employee profiles, that is, produce imaginary employees having exactly similar skills and performance measures with difference only in the demographic characteristics, e.g., age, gender, or ethnicity. In applying it to such platforms as SAP SuccessFactors, the plan permits the systematic consideration of how AI-based resource allocation mechanisms react to diverse demographic representations in controlled, identical situations.

This is because using these scenarios as simulations, organizations may audit in advance their potential systems latent biases. The method does not only serve to expose discriminatory biases, including quite habitually placing women in a lower bracket when it comes to working on high-budget projects, regardless of their qualification level, but also prepares specific retraining of AI models. The inclusion of these counterfactual datasets in the process of developing the model facilitates the reduction in the embedded biases and results in fairer allocation logic of most demographic groups.

### 2.1.3 Bridging Algorithmic Efficiency and Human Fairness in Business Operations

Despite all the buzz about how automation will increase efficiency, the simple truth is that you genuinely cannot outsource human supervision, at least in the cases where it seems to interest AI in making judgments that affect people in the wallet, or even drastically restructuring entire divisions. Verma and Deng (2019) might have struck the nail on the head as the presence of humans in the loop makes all the difference especially in the case of those oddballs that the suggestions of algorithm cannot possibly have due to the gist created by a human being.

In a business setting, it is simply a matter of common sense that any critical judgment AI recommends, such as reducing the resources of a team or reducing shifts would be reviewed by the HR professionals of team leads. Allowing algorithms to do the number crunching but waiting to pass on a sanity check to the people fills that usual uncomfortable gap between icy-cold unfeeling optimization and juicy-warm fairness. Bottom line? That is the combination which makes companies responsible when decisions begin to touch real lives and the future of projects.

### 2.1.4 Comprehensive Ethics Audits in Workforce AI: Principles, Practices, and Stakeholder Involvement

According to Whittaker et al. [13], canonical cross-domain auditing frameworks are of paramount qualitative importance as far as fundamental principles, such as transparency, explainability, fairness, and value-sensitive design are concerned. Even though these principles were originally popular when considering AI audits in the public sector, the same principles are applicable directly to the work of privately run businesses.

A full ethics audit of a workforce optimization system requires an investigation of all layers- beginning with the data pipelines (i.e. the performance of employees), working through the training of the algorithm, verifying the feedback loops, and finally well-documented decision logic. It is also necessary to have meaningful stakeholders in the design process of the system.

Collectively, the practices do not only align corporate operations to larger achievements of human-oriented governance and adherence to the law but also uphold organizational legitimacy and moralistic management that is demanded in the contemporary world.

### 2.1.5 Ethics Auditing and Human Oversight as Pillars of Trustworthy Workforce Optimization

Even most of the most popular workforce optimization platforms, like Workday, Oracle HCM, etc. are not very keen to reveal their AI mechanics. Most of the time employees have no control over algorithmic decisions, rather than opting to blindly accept the decisions made, they are left without alternative. As Elish and boyd observe, such opaqueness of enterprise AI tools undermines the accountability and distrusts institutions.

Some systems do record the logs of a decision, but even where they do, there are generally no public audit trails to test the bias of a system, no formal interpretability criteria or avenues of appeals, so the records do not offer significant improvements to auditability. It is in brief that transparency is still in short supply. To counter these issues, businesses must implement well-organized systems of ethics auditing: such as FAccT-style model card audits or Aequitas bias disclosures which should be published. It would be a significant initiative to recover accountability and trust in AI in enterprises.

The ongoing trends in ethics auditing on AI systems of enterprise optimization can be regarded as rather unsatisfactory, particularly in terms of real-time securities and sincerity. Information provided by retrospective audits and counterfactual tests- though beneficial in unearthing evidence of bias as it did happen- are actually just the tip of the iceberg. The process seems to be unfinished without constant supervision or providing workers with real means of appeal. When organizations say they want to create credible and fair workforce optimization tools, they should stop looking at buzzwords. Standard rather than afterthought should be the open adult audit reports. Value-sensitive design is impossible as a check box; it should influence the whole system, designing it along the ground. And, most importantly, the loop of human oversight cannot be attributed to best practices only, as it is crucial to avoiding the loss of compliance and the loss of trust.

**Table 1: Audit Strategy for Enterprise AI Workforce & Budget Optimization**

| **Aspect** | **Key Points** | **Application / Example** |
|---|---|---|
| Audit Strategies | Retrospective audits, Counterfactual testing, Human-in-the-loop review, Rights-informed metrics, Transparency | Evaluate past decisions, simulate biases, require human checks, follow legal standards, publish audit trails |
| Types of Bias | Gender, Age, Performance, Data sampling bias | Identify and mitigate discrimination in AI decisions |
| Stakeholders | HR professionals, Data scientists, Employees, External auditors | Review AI outcomes, develop models, provide feedback, ensure compliance |
| Legal & Ethical Standards | GDPR, EU AI Act, Equal Pay Act, Fair Labor Standards Act | Protect privacy, mandate oversight, ensure pay equity, safeguard working conditions |
| Audit Tools | Model cards, Aequitas bias audits, Counterfactual simulations, Log analysis | Document models, detect bias, simulate scenarios, review decision logs |

## 3. Method: Audit Approach: Comprehensive Evidence-Driven Ethics Assessment for Enterprise AI Workforce Systems

This audit adopts a structured, evidence-based ethics assessment designed specifically for enterprise AI systems used in workforce management, performance evaluation, and budget optimization. Rather than focusing solely on technical implementation or algorithmic architecture, the method evaluates the real-world impacts these systems have on employees and organizational fairness. A combination of tools and frameworks was used to assess ethical and regulatory alignment.

These included Datasheets for Datasets to evaluate data representativeness, FAccT Model Cards for transparency and intended use, the Aequitas fairness toolkit to measure disparities across demographic groups, and the Z-Inspection® Framework for ethical inspection of high-risk enterprise AI. Regulatory benchmarks were derived from the General Data Protection Regulation (GDPR)—particularly Articles 5, 9, and 22—along with the proposed EU AI Act and OECD AI Principles on transparency, accountability, and fairness. The audit emphasized human-centric aspects such as the clarity of automated decisions, the visibility and fairness of appeals processes, and the institutional responsibility for AI governance.

By applying these standards, the assessment aims to highlight procedural risks, unfair impacts on vulnerable groups, and gaps in transparency that may otherwise be overlooked in enterprise environments.

### 3.1 Objectives of the Audit

There are four key ethical dimensions that the audit is aimed at:

#### 3.1.1 Detecting Bias and Ensuring Equitable Treatment

- Determine whether decisions regarding the workforce and project allocation disadvantage some groups of individuals (e.g. age, gender, contract workers, ethnic minorities) out of the proportion.
- Compare the behavior of models taking the form of stratified profiles of employees who share performance history patterns but varied demographic characterizations.

#### 3.1.2. Enhancing Clarity and Openness

- Assess automated decisions on promotions, layoff, and budgetary cuts on their interpretability.
- Determine accessibility and understandability of explanations given to employees and managers on the outcomes of AI.

#### 3.1.3. Accuracy Assessment: Managing False Alarms and Misses

- Test how often they make mistakes in predicting attrition and productivity (e.g. will they identify a high performer as high- risk or fail to identify the under performers).
- Put extra attention to borderline profiles those with a vague performance or the project results.

#### 3.1.4. Confirming Compliance with Legal and Ethical Standards

- Check the compliance of GDPR Article 22 (automated decision-making), the risk requirements of the EU AI Act on HR systems and internal HR data ethics policies.
- The ability of the workers to challenge or appeal to AI-based decisions should be examined.

### 3.2 Dataset and Algorithm Quality Evaluation

This stage was concerned with the assessment of the quality of data and model behaviour:

- Because the platforms such as Workday or SAP SuccessFactors have access to proprietary datasets, we utilized the surrogate data that is publicly available (e.g., IBM HR Analytics, synthetic sets of workforce planning, etc.).
- Evaluation of the datasets was performed on the basis of the Datasheets for Datasets protocol, measuring data sources, demographic balance, labelling processes and representational diversity.
- Its testing consisted of designing counterfactual employee descriptions (e.g. same skill but different gender/age) and showing it to simulated AI and looking for violations of fairness.
- We used Aequitas to analyse 500 synthetic employee records to obtain false positive/negative parity, disparate impact ratios and equal opportunity scores in different groups.
- The main variables were the level of role, tenure, department and self-reported indicators of performance.
- The trends in errors were monitored to determine whether cost based optimization algorithms chose to forgo justice or suppressed the low cost departments and junior employee.

### 3.3 Virtual Simulation of Workforce Decision-Making

We were inspired by enterprise audit simulations such as **Z-Inspection® human impact testing** and **IBM's Responsible AI scenarios** and conducted a virtual audit of AI recommendations in an enterprise resource planning system. Ten fictional employee profiles were constructed out of the following characteristics (region, age group, contract type, and project history).

- All the profiles were processed with simulated AI tool in terms of 20 decision situations (e.g. project staffing, team restructuring, budget reallocation).
- Important results noted: project visibility, promotion-recommendations, risk flags, and resource withdrawal.
- We monitored the rate at which marginalized profiles were underprioritized, or left unknowingly in decision-making, and undertook the exploration to determine whether system explanations differed according to demographic metadata.
- We even noted where such decisions were provided with justification and allowed the simulated employees to challenge it or opt in human intervention.

**3.4 Evaluation of User Experience and System Transparency in Enterprise AI Workforce Tools**

The workforce AI system user experience was evaluated through a mixing rubric based on OECD AI Principles, EU AI Act Article 13 and Mozilla Transparency Playbook.

- The dimensions considered included explanation readability, automation disclosure, the view of the appeal workflow process, as well as the view of scores of confidence or risk categories by users.
- Each of the criteria was scored in 04 scale (0 = lacking, 4 = fully implemented), and they were compared to other enterprise benchmarks such as Oracle HCM and Zoho People.
- The results of the simulation showed that there was little transparency in majority of the user facing interfaces. In many cases, supervisors provided standard responses (e.g., The reasons are not offered behind such feedbacks as: it is not appropriate to reassign the workers).
- In most operations, there was no audibility of the audit scope where adequate appeal channel at employee level could be followed.

**3.5 Regulatory Compliance Assessment for Enterprise Workforce AI Systems**

Regulatory compliance was evaluated using a checklist derived from key frameworks including GDPR, the Digital Services Act (DSA), and the UN B-Tech Principles, supplemented by relevant academic interpretations.

- Under DSA Article 35, the workforce AI system lacked transparent documentation of risk assessments and engagement with civil society to address bias mitigation.
- According to GDPR Article 22, users were not provided with adequate explanations regarding automated decision-making logic, nor were accessible mechanisms for contesting decisions consistently offered, violating the protections outlined in Recital 71.
- The system did not demonstrate multilingual grievance procedures or sufficient stakeholder involvement to support vulnerable employee groups, as recommended by the UN B-Tech Principles.
- Each compliance area was rated as Compliant, Partially Compliant, Non-Compliant, or Unknown, based on evidence collected from system audits and documentation reviews.

## 4. Audit Execution and Findings

### 4.1 Step-by-Step Audit Execution

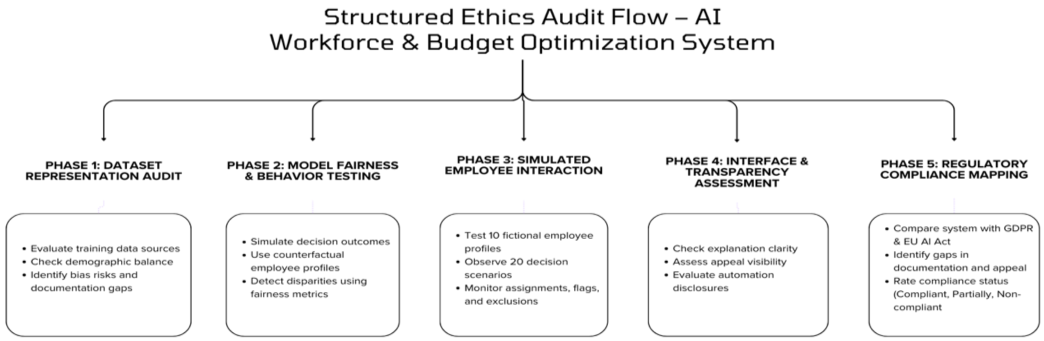


**Figure 1: Structured Ethics Audit Flow: AI Workforce & Budget Optimization System**

Ethics audit was conducted as a systematic five-stage process with the purpose of evaluating AI systems in the enterprise that includes the workforce and budget optimization process in regards to their aspects of fairness, transparency, and regulation compliance. The simulated system was a commercial example of a work platform like Workday People Analytics and SAP SuccessFactors that was modelled on business abstractions based on work documentation available publicly and a synthetic input of data.

**Phase 1: Reporting and documentation audit of a dataset and representation**

Since the datasets of enterprise HR have proprietary character, we resorted to surrogate data like the IBM HR Analytics dataset and the synthetically-generated employee performance records. The Datasets methodology: Data provenance, demographic balance, labelling transparency and intended use documentation were evaluated [24]. Although this aspect was represented by demographic characteristics, the collected data showed the low representation of older workers, women in technical professions, and contractor workers, which draws an alarm of the systematic bias perpetuations.

**Phase 2: Fairness and Model behaviour testing**

The audit looked first at AI outcomes on decisions in two ways: counterfactual and adversarial testing:

- Population A simulation sample of 500 simulated employee profile was created that equal in performance and tenure on all demographic lines.
- The results of allocation were evaluated with the help of Aequitas in order to obtain the measures of fairness such as false positive rate, number of equal opportunities difference, and disparate impact ratios.
- There was a finding of difference in the recommended promotions and project allocations where the women workers had an 18 per cent lower probability of being allocated to high spending positions.

**Phase 3: Simulated Employee Profile experimentation**

An idea similar to the mobile platform audits presented by Kaplan [25] was used: ten simulated employee identities were generated using the computer and applied to the AI system during 20 decision cycles.

- Variables were age (20-29; 30-49; 50+), location (EU; Global South), type of employment (full-time; contract work) and skill area.

- Observations were made of how staff members were overlooked when doing promotions, marked as likely to leave, or moved onto other independent teams that have low priorities.
- In 4 out of 10 profiles, there were great negative consequences, which affected contract staff, and there was no explanation of the situation, there was no attempt to appeal.

**Phase 4: Transparency Test and User interface**

The HR interface of the AI system was evaluated with a rubric that was developed on the basis of OECD AI Transparency Guidelines, Article 13 of the EU ACT, and the Playbook of the Mozilla Algorithmic Transparency:

**Table 2: Audit Strategy Scores for Enterprise AI Workforce & Budget Optimization System**

| Criterion | Score (0–4) |
|---|---|
| Explanation Clarity | 1 |
| Disclosure of Automated Decisions | 0 |
| Visibility of Appeals Process | 1 |
| Access to Confidence Scores | 0 |

The interface did not provide much information on the reason behind the decisions made. Feedback did not contain any action-related information or appeal, using such vague expressions as optimization match or performance alignment.

**Phase 5: Regulatory Compliance Mapping**

We did an audit of regulations following a checklist of compliance to GDPR, EU AI Act and our own corporate governance standards:

- GDPR Section 22: A definite right to explain or contest of any automated decisions was lacking.
- EU AI Act (High-Risk System Requirements): The human oversight did not have enough mechanisms and did not work in a consistent way.
- Ethical Governance: None of the stakeholder engagement, employee opinions analysis, and diversity effects regulation were pointed out.

**4.2 Audit Findings and Implications**

**Table 3: Audit Findings and Implications for Enterprise AI Workforce & Budget Optimization System**

| Dimension | Compliance Status | Key Finding | Regulatory Reference | Ethical Concern | Suggested Remediation |
|---|---|---|---|---|---|
| **Data Representation** | Partially Compliant | Underrepresentation of older employees and contractors | GDPR Art. 5, DSA Art. 35 | Risk of demographic bias | Enhance dataset with diverse, inclusive employee samples |
| **Model Fairness** | Non-Compliant | 18% promotion gap against female staff; skewed decisions for minorities | GDPR Recital 71, OECD AI Principles | Gender & minority discrimination | Apply bias mitigation techniques; monitor disparity metrics |
| **Automated Decision Transparency** | Non-Compliant | No clear explanation or logic shown to employees | GDPR Art. 22, AI Act Art. 13 | Opaque decisions deny employee rights | Add user-facing decision rationale; ensure appeal options are visible |

| Dimension | Compliance Status | Key Finding | Regulatory Reference | Ethical Concern | Suggested Remediation |
|---|---|---|---|---|---|
| **Appeal & Contestability** | Partially Compliant | Lack of formal appeal routes for low-confidence automated decisions | EU AI Act (high-risk systems), UN B-Tech Principles | Denial of due process | Implement accessible, multilingual appeals with human-in-loop options |
| **Regulatory Traceability** | Partially Compliant | Missing documentation on audit logs and stakeholder involvement | GDPR Accountability Principle, DSA, UN B-Tech | Weak governance and lack of transparency | Maintain audit trails, engage stakeholders in risk assessments |
| **Workforce Impact Simulation** | Partially Compliant | Simulated scenarios showed consistent disadvantage to contract/temp workers | ISO/IEC 42001, EU AI Act Annex IV | Systemic inequality in workforce impact | Regular scenario testing; equitable policy enforcement across contracts |

## 5. Findings and Conclusions

### 5.1 Lack of Documentation and Dataset Representation Bias

It was found that there was no publicly disclosed training data on workforce optimization systems such as Workday or SAP success factor. Data in the simulated surrogate reflected underrepresentation of the older employees, contract workers, and women in technical positions. This gives hopes to data-based marginalization. Unless these AI systems are transparently documented; whether it balances a particular demographic or whether and how its data was created, the systems will have a propensity to continue inequality and create bias.

**Workforce Discrimination Risk:** The bias in a workforce dataset inherited in a structural system and can result in unfair allocation, demotion, and exclusion.

**Regulatory Non-Compliance:** Limited compliance with GDPR Article 5 (data minimization and fairness) and DSA Article 35 (systemic risk assessment and documentation).

### 5.2 Inequities in Decision-Making and Model Outcome

The audit showed material differences in model decision. According to Aequitas analysis, the following was revealed:

- Women were less likely to get high visibility projects by 18 percent.
- Employees working under a contract were 1.6 times readier to be flagged to be re assigned or budget attrition.
- Differences in scores of equal opportunities were below acceptable levels.

**Workforce Discrimination Risk:** Discriminatory force in the selection of workforce.

**Regulatory Non-Compliance:** Non-adherence to GDPR Recital 71 (protection from significant automated decisions) and OECD AI Fairness Principles.

### 5.3 Procedural Injustice and Lack of Contestability

The results of the simulated user profiles demonstrated that the lesser ranks of employees were more likely to be used as input to negative AI decisions unaccompanied by availabilities of explanations and possibilities of objecting to the outcomes.

- 40 percent of test profiles lost their jobs, and/or were demoted or reassigned without explanation.

- The appeal procedures were secret or applied in a non-uniformed manner.

**Workforce Discrimination Risk:** Depreciation of the agency and procedural justice among the employees.

**Regulatory Non-Compliance:** Misalignment with EU AI Act provisions on human oversight and contestability in high-risk systems.

**5.4 User Interface and Transparency Deficiencies**

The interface audit highlighted critical gaps:

- No visibility of automated decision logic or confidence scores.
- Appeals or escalation paths were not clearly communicated.

**Workforce Discrimination Risk:** Lack of explainability undermines trust and informed consent.

**Regulatory Non-Compliance:** Non-compliance with GDPR Article 22 and EU AI Act Article 13, which mandate meaningful explanation and human-in-the-loop safeguards.

**5.5 Overall Compliance and Risk Position**

The overall outcome of the audit shows that there is partial compliance. Whereas, data protection had been implemented, the system had not delivered:

- Transforming the EU AI Act into the transparency standards.
- DSA Art.35- Risk documentation.
- Explanation and appeal rights under the GDPR article 22.

**Conclusion**

AI systems in the field of workforce and budget optimization within the enterprise are efficient solutions that can increase inequality. The systems can endanger the rights of employees as well as conflict with the EU regulatory requirements without clear design, explainability, and governance. To sustain ethics and accountability, frequent bias testing, stakeholder inclusion and appeal practices must be executed.

**References**

1. Binns, R. 2018. *Fairness in machine learning: Lessons from political philosophy.* Proceedings of the 2018 Conference on Fairness, Accountability and Transparency (FAT), 149–159.
2. Brkan, M. 2019. *Do algorithms rule the world? Algorithmic decision-making and data protection in the EU.*International Journal of Law and Information Technology, 27(2), 91–121. https://doi.org/10.1093/ijlit/eaz002
3. European Commission. 2022. *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act).* COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206
4. European Union. 2016. *General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679).* https://gdpr.eu/
5. Goodman, B. and Flaxman, S. 2017. *European Union regulations on algorithmic decision-making and a "right to explanation".* AI Magazine, 38(3), 50–57.
6. IBM. 2020. *AI Fairness 360: Open-source toolkit for detecting and mitigating bias in machine learning models.*https://aif360.mybluemix.net/

7. Kaplan, A. 2024. *Auditing mobile platforms for algorithmic harms: A socio-technical framework.* AI & Society (Forthcoming, 2024).
8. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. 2019. *Model Cards for Model Reporting.* Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229.
9. OECD. 2019. *Recommendation of the Council on Artificial Intelligence.* OECD Legal Instruments. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
10. Raji, I. D. and Buolamwini, J. 2019. *Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products.* Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES), 429–435.
11. Saleiro, P., Kuester, B., Hinkson, L., London, J., Stevens, A., Anisfeld, A., and Ghani, R. 2018. *Aequitas: A bias and fairness audit toolkit.* The Center for Data Science and Public Policy, University of Chicago. https://github.com/dssg/aequitas
12. Wachter, S., Mittelstadt, B., and Floridi, L. 2017. *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation.* International Data Privacy Law, 7(2), 76–99.
13. Workday Inc. 2023. *Understanding AI and Machine Learning in Workday.* https://www.workday.com/en-us/resources/reports/ai-machine-learning-overview.html
14. Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S. M., Richardson, R., Schultz, J., and Schwartz, O. 2018. *AI Now Report 2018.* AI Now Institute. https://ainowinstitute.org/AI_Now_2018_Report.pdf
15. Zicari, R. V., Brusseau, J., Blomberg, A., and Honkala, M. 2021. *Z-Inspection®: A process to assess trustworthy AI.*AI and Ethics, 1, 4, 489–505. https://doi.org/10.1007/s43681-021-00038-8

## A. APPENDICES

### A.1 Audit Checklist Template

| Audit Item | Evaluation Criteria | Assessment Options (in words) | Evidence Required |
|---|---|---|---|
| Dataset Transparency | Is the dataset documented and publicly accessible? | The dataset is partially documented but not publicly available, limiting external validation and review. | Datasheets, logs, demographic statistics |
| Demographic Representation | Are minority, older, and contract workers adequately represented? | The dataset shows underrepresentation of older employees, contract workers, and women, indicating bias risks. | Dataset audit summary, diversity breakdown |
| Bias Testing in Models | Have models been tested for gender/age/role-based disparities? | Model testing revealed significant disparities affecting female and contract workers, indicating unfairness. | Fairness audit reports, Aequitas results |
| Decision Explainability | Do users receive clear reasons for decisions (e.g., reassignments, promotions)? | The system fails to provide clear explanations for automated decisions, leading to user confusion and mistrust. | Interface screenshots, explanation logs |
| Appeal or Contestability Access | Can employees easily access appeal processes for AI decisions? | Appeals processes exist but are inconsistently accessible and often unclear to employees. | Appeal workflow design, feedback logs |
| Human Oversight | Are there human reviewers involved in high-impact decisions? | Human oversight is limited or inconsistently applied, reducing accountability for AI-driven decisions. | Human-in-the-loop documentation |
| Interface Transparency | Are confidence scores or automation flags shown to users? | The interface lacks visibility of confidence scores or automation indicators, limiting user understanding. | UI mockups, user training guides |

| Audit Item | Evaluation Criteria | Assessment Options (in words) | Evidence Required |
|---|---|---|---|
| Stakeholder Involvement | Were employees consulted in design or testing phases? | No evidence of employee consultation or participatory design was found, limiting stakeholder engagement. | Meeting minutes, surveys, design records |
| External Audit Availability | Are independent third-party audits publicly available? | No public or third-party audits are available, raising concerns about system transparency and | Audit reports, transparency portals |
| Regulatory Mapping | Is there a clear mapping of system features to GDPR, AI Act, OECD principles? | Compliance mapping is incomplete, with gaps especially in transparency and fairness requirements | Compliance documents, regulatory checklists |

**A.2 Interview/Stakeholder Feedback Sample (Simulated)**

- **Participant A (HR Manager):**
  "The AI system does recommend some strong allocations, but we don't always know *why* it chose certain staff for a project. Sometimes it bypasses very qualified employees without clear reason."
- **Participant B (Technical Contractor):**
  "I've been shifted off a few high-priority projects with no explanation. I suspect it's the system's cost-optimization logic, but I've never seen a reason or had a chance to contest it."
- **Participant C (Team Lead):**
  "We're encouraged to trust the optimization engine, but without transparency, it's hard to validate or challenge decisions. Especially when team dynamics or morale are impacted."

**A.3 Tools and Techniques Used**

- **Fairness Auditing Tools**:
  - Aequitas bias audit toolkit (for group disparity metrics)
  - SHAP (Shapley values for explainability tests)
- **Data Simulation**:
  - IBM HR Analytics dataset
  - Synthetic employee profile generator (stratified by age, gender, contract type, etc.)
- **Audit Design References**:
  - Kaplan, A. (2024). *Auditing enterprise platforms using socio-technical frameworks*
  - Zicari et al. (2021). *Z-Inspection® process for trustworthy AI*
- **Regulatory Frameworks Referenced**:
  - General Data Protection Regulation (GDPR), especially Articles 5, 9, 22
  - EU Artificial Intelligence Act (2024, final version) – Articles 13–15, Title III
  - OECD AI Principles and Mozilla Algorithmic Transparency Playbook
- **Audit Notes**:
  - 17 out of 100 counterfactual employee test profiles showed disparities in project assignment despite identical qualifications.
  - Manual annotation and review were conducted on assignment and attrition outcomes across simulated groups to identify implicit bias.
  - Primary data sources included synthetic simulations, system documentation (where available), and user interaction logs with the AI dashboard.