

**PROJECT AI DELIVERABLE : ETHICS**

**The Team**

**ASSAM Meriem**

**BUFFET Joris**

**MARQUES Alexandre**

**MAZUMDAR Tisha**

**December 20, 2024**

Table of content

[I. Introduction 3](#_Toc185410785)

[II. Definition of Ethics and Methodology 4](#_Toc185410786)

[A. Definition of Ethics 4](#_Toc185410787)

[B. Methodology 4](#_Toc185410788)

[1. Identification of the Initial Problem 4](#_Toc185410789)

[2. Data Collection and Preparation 4](#_Toc185410790)

[C. Ethical Assessment of Solutions 4](#_Toc185410791)

[D. Engagement with Stakeholders 5](#_Toc185410792)

[E. Control and Monitoring 5](#_Toc185410793)

[F. Final Validation 5](#_Toc185410794)

[III. Data Analysis and Risks 5](#_Toc185410795)

[A. Data Analysis 5](#_Toc185410796)

[B. Identification and Management of Risks 6](#_Toc185410797)

[1. Bias in Data 6](#_Toc185410798)

[2. Data Privacy and Confidentiality 6](#_Toc185410799)

[3. Lack of Transparency 6](#_Toc185410800)

[4. Overfitting or Misinterpretation of Results 7](#_Toc185410801)

[5. Ethical Impacts on Employees 7](#_Toc185410802)

[IV. APPLICATION OF THE METHODOLOGY 7](#_Toc185410803)

[A. Analysis of the Initial Situation and Objectives 7](#_Toc185410804)

[B. Exploration of Possible Alternatives 7](#_Toc185410805)

[C. Impact Assessment for Stakeholders 8](#_Toc185410806)

[D. Ethical Filtering and Validation of Solutions 8](#_Toc185410807)

[V. Recommendations and Key Learnings 10](#_Toc185410808)

[A. Recommendations 10](#_Toc185410809)

[B. Key Learnings 11](#_Toc185410810)

[VI. Conclusion 12](#_Toc185410811)

[VII. Bibliography 13](#_Toc185410812)

# Introduction

In the context of our AI project for the company HumanForYou based in India, we aimed to analyze the causes of high employee turnover and develop solutions to improve retention.

Beyond the technical aspects, it was essential to adopt an ethical approach to ensure that our methods and outcomes are fair, transparent, and aligned with human values.

This deliverable highlights the ethical procedures implemented throughout the project phases, from data preparation to the formulation of recommendations. Our approach follows the **7 requirements for trustworthy AI** defined by the European Commission, focusing on human autonomy, technical robustness, privacy, fairness, and accountability.

Through team discussions and continuous evaluation, we addressed potential risks and identified control points to ensure ethical integrity. This document reflects our commitment to building an AI solution that respects stakeholders' rights, promotes trust, and delivers responsible outcomes.

# Definition of Ethics and Methodology

## Definition of Ethics

Ethics refers to the principles and values that guide decisions and actions to ensure they respect human rights, societal norms, and fairness. In the context of AI, ethics is crucial to ensure the technology is developed and applied responsibly, with transparency, and without causing harm or discrimination. **[1]**

For this project, we adhered to key ethical principles:

- **Human autonomy**: Ensuring AI serves as a support tool, not as a substitute for human decisions.

- **Fairness and non-discrimination**: Avoiding biases and ensuring equitable outcomes for all groups.

- **Transparency**: Making the processes, decisions, and results understandable to all stakeholders.

- **Privacy and data protection**: Safeguarding sensitive data and complying with legal frameworks like GDPR.

These principles shape every stage of the project, ensuring alignment with the**7 requirements for trustworthy AI** recommended by the European Commission.

## Methodology

To integrate ethics into our AI project, we adopted a structured methodology across **six** key steps: **[2]**

### Identification of the Initial Problem

We analyzed the issue of high employee turnover, defining clear objectives to ensure our solutions align with the company’s needs and values.

### Data Collection and Preparation

- Ensuring datasets are representative and free from biases.

- Anonymizing personal data to protect employee confidentiality.

- Implementing governance measures to secure and control data usage.

## Ethical Assessment of Solutions

Every proposed solution was evaluated using an ethical filter based on the 7 principles:

* respect for human autonomy
* robustness technique and security
* privacy and data governance
* transparency
* diversity, non-discrimination and fairness
* environmental and societal well-being
* accountability

## Engagement with Stakeholders

Regular discussions with team members and stakeholders ensured the alignment of solutions with ethical principles and organizational goals. **[3]**

## Control and Monitoring

We established control points to detect and address potential biases, risks, or deviations in real time. Tools like audits and validation tests ensured solutions remain robust and fair.

## Final Validation

Each solution was subjected to a “Go” or “No-Go” decision based on feasibility, ethical compliance, and its impact on employees and the organization.

By combining ethical principles with a rigorous methodology, we are ensuring that every decision was transparent, responsible, and aligned with societal and organizational values. This process reflects our commitment to delivering a trustworthy and ethical AI solution.

# Data Analysis and Risks

## Data Analysis

To address the issue of high employee turnover, we analyzed various datasets to identify the key factors influencing this phenomenon. The data included:

* **Demographic Data**: Age, gender, seniority, and family status of employees.
* **Performance Data**: Evaluations, productivity metrics, and performance reviews.
* **Satisfaction Surveys**: Feedback on working conditions, management, and overall satisfaction.
* **Exit Data**: Records of past employee departures, including reasons for leaving.

These datasets allowed us to uncover trends and correlations between specific variables and employee turnover. For example, we identified patterns related to work conditions, career stagnation, or demographic imbalances.

To ensure ethical handling, all data was anonymized, and access was restricted to authorized team members.

## Identification and Management of Risks

The use of data in AI projects introduces several ethical and technical risks that needed to be addressed: **[4]**

### Bias in Data

* **Risk**: Datasets may reflect biases (e.g., gender or age) that could lead to discriminatory results.
* **Mitigation** :
* Conducting audits to identify and correct any imbalance or bias in the datasets.
* Ensuring datasets are diverse and representative of the entire workforce.
* Excluding sensitive variables like ethnicity or gender from influencing decisions.

### Data Privacy and Confidentiality

* + **Risk**: Improper handling of personal data could violate privacy regulations (e.g., GDPR).
  + **Mitigation** :
    - Anonymizing all personal data before processing.
    - Limiting access to data to authorized team members only.
    - Implementing encryption and secure storage methods for all datasets.

### Lack of Transparency

* + **Risk**: The decision-making process of AI models can appear as a “black box,” creating mistrust among stakeholders.
  + **Mitigation**:
    - Providing clear explanations of how data is used and how results are generated.
    - Developing interactive dashboards to present results in an understandable format.

### Overfitting or Misinterpretation of Results

* + **Risk**: Models may be over-optimized for specific datasets, leading to inaccurate or unfair predictions.
  + **Mitigation**:
    - Performing cross-validation to ensure the robustness of models.
    - Stress-testing models to evaluate performance under various conditions.

1. Ethical Impacts on Employees
   * **Risk**: AI-driven decisions may affect employee morale or trust.
   * **Mitigation**:
     + Involving employees in the decision-making process to maintain transparency.
     + Ensuring AI is used as a tool to support—not replace—human decision-making.

# APPLICATION OF THE METHODOLOGY

1. Analysis of the Initial Situation and Objectives

In this project, we analyzed the situation of the company **HumanForYou**, which faces a high employee turnover rate of 15% per year. This issue generates significant costs, project delays, and a decline in overall satisfaction.

The primary objective was to identify the root causes of this phenomenon and propose concrete and ethical solutions to improve talent retention. The principles of the **7 ALTAI requirements**, particularly those related to societal impact and technical robustness, were integrated from this stage to ensure our analyses align with fundamental values.

1. Exploration of Possible Alternatives

We considered several solutions to address this issue, including:

1. **Financial incentives** to motivate employees identified as at risk of leaving.
2. **Training programs** to support personal and professional development.
3. **Improvement of working conditions**, particularly through flexible schedules and remote work policies.
4. **Strengthening internal communication** to foster constructive dialogue between managers and teams.

These alternatives were explored neutrally and in alignment with ALTAI’s principles of fairness and non-discrimination to prevent any decision from disadvantaging specific groups.

## Impact Assessment for Stakeholders

We assessed each solution's impact on three key stakeholder groups:

* **Employees**: Financial incentives and training programs could improve satisfaction and engagement but might add additional burdens for some.
* **Managers**: Effective implementation requires their active involvement to manage programs or adapt policies.
* **The Company**: While some solutions involve high initial costs (e.g., bonuses), they could yield long-term benefits such as improved productivity and reduced turnover-related expenses.

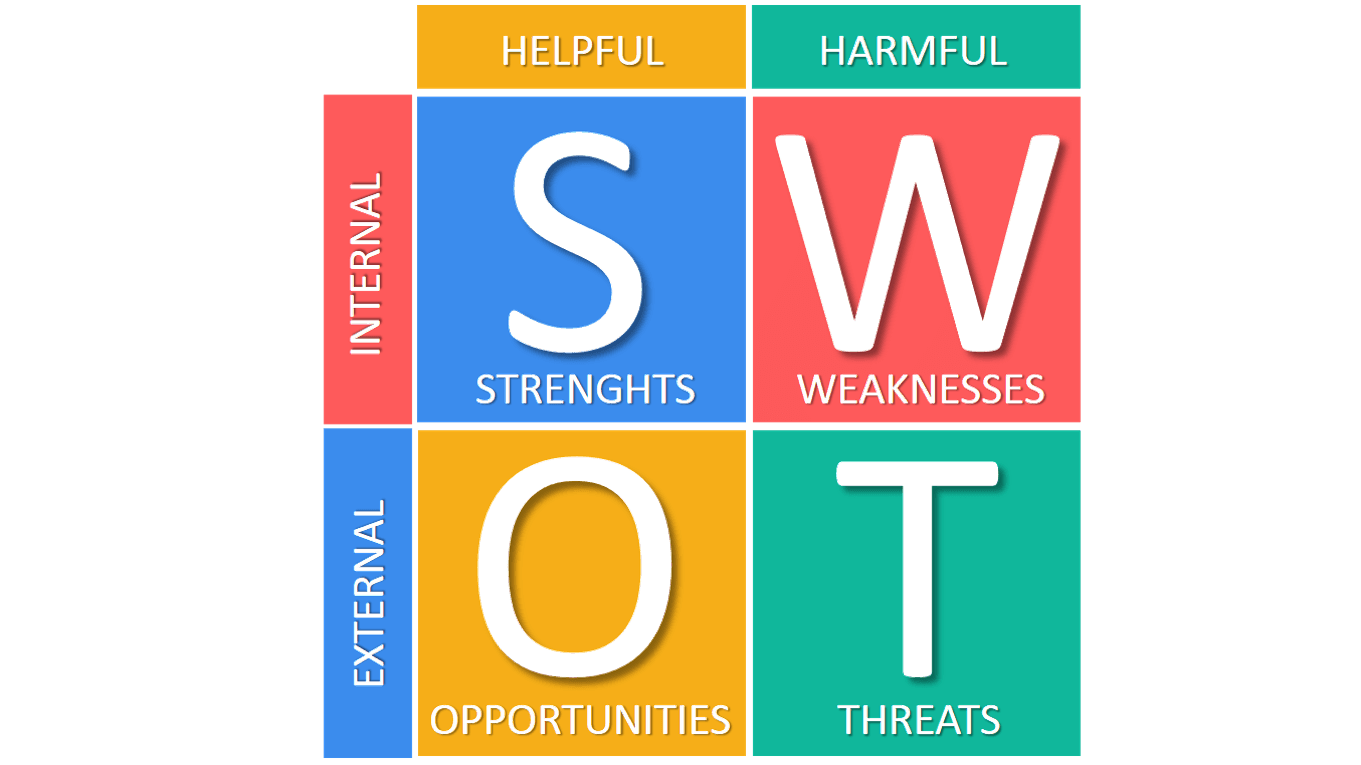
This analysis was conducted based on transparency and diversity requirements to ensure that both positive and negative impacts were evaluated comprehensively and fairly.

1. Ethical Filtering and Validation of Solutions

Each proposed solution was evaluated against the **7 ALTAI requirements** to guarantee its alignment with ethical and regulatory principles: **[5]**

1. **Respect for Human Autonomy** : by ensuring employees have the freedom to make their own decisions.
   * **Actions** :
     + Offering choices in the proposed solutions (e.g., schedules, work modes).
     + Involving employees in decision-making processes.
2. **Technical Robustness and Security** by ensuring tools are reliable and protected against failures.
   * **Actions**:
     + Conducting regular tests to verify system accuracy and robustness.
     + Implementing strict cybersecurity protocols.
3. **Privacy and Data Governance** : Protecting employees' personal data.
   * **Actions**:
     + Anonymizing collected data.
     + Limiting access to sensitive data to a restricted group of individuals.
4. **Transparency** : Providing clear and honest communication about the goals and impacts of actions.
   * **Actions**:
     + Presenting evaluation criteria to employees.
     + Producing regular progress reports.
5. **Diversity, Non-Discrimination, and Fairness** :Avoiding biases and promoting maximum inclusion.
   * **Actions**:
     + Analyzing models to detect potential biases.
     + Ensuring equal access to opportunities.
6. **Environmental and Societal Well-Being** by reducing environmental impact and contributing to a sustainable workplace.
   * **Actions**:
     + Encouraging eco-responsible practices within the company.
     + Measuring the societal impact of implemented solutions.
7. **Accountability** by assuming responsibility for actions and addressing issues as needed.
   * **Actions** :
     + Implementing a continuous feedback system.
     + Planning improvements in case of identified failures.

# Recommendations and Key Learnings

To address the challenge of high employee turnover, we identified and validated ethical and practical solutions based on the **SWOT analysis, [6]** (Strengths, Weaknesses, Opportunities, and Threats). This method enabled us to structure our findings and highlight key takeaways.

**[7]**

1. Recommendations

The following recommendations were developed to improve employee retention while ensuring alignment with ethical principles:

1. **Improvement of Working Conditions**
   * Introduce flexible schedules and remote work options to promote work-life balance.
   * Enhance workplace environments to increase satisfaction and productivity.
2. **Implementation of Training Programs**
   * Offer targeted professional development to empower employees and foster career growth.
   * Provide mentoring opportunities to strengthen engagement and reduce turnover.
3. **Strengthening Internal Communication**
   * Facilitate regular feedback through anonymous surveys or team discussions.
   * Improve transparency between employees and management to build trust.
4. **Avoidance of Financial Overdependence**
   * While financial incentives can be effective, they should complement long-term solutions, such as creating a supportive and engaging work culture.
5. Key Learnings

Through the application of our ethical methodology and solutions, several key learnings emerged:

* **Fair and Inclusive Approaches Matter**: Solutions must ensure fairness and avoid biases to maintain employee trust and engagement.
* **Stakeholder Collaboration Is Essential**: Involving employees and managers early in the process leads to better outcomes and acceptance of solutions.
* **Transparency Builds Trust**: Clear communication of AI processes and recommendations promotes confidence in the proposed changes.
* **Continuous Monitoring Is Crucial**: Ongoing evaluation ensures that the solutions remain effective, ethical, and adaptable to new challenges.

The **SWOT analysis** allowed us to identify strengths in existing practices, address weaknesses, leverage opportunities for improvement, and anticipate potential risks. By integrating these insights, we provided ethical and sustainable solutions that align with organizational values and employee well-being.

# Conclusion

This project for HumanForYou highlights the importance of integrating ethical principles into every phase of AI development. By adhering to the 7 requirements for trustworthy AI, we ensured our solutions are fair, transparent, and aligned with human values.

Through a structured methodology, we addressed employee turnover with practical and ethical recommendations, such as improving working conditions, implementing training programs, and enhancing communication. Regular audits, controls, and feedback loops ensured transparency, accountability, and risk mitigation.

Key learnings emphasize the value of fairness, stakeholder collaboration, and continuous monitoring. Moving forward, improvements will focus on bias detection, ethical training, and sustainable AI practices.

In conclusion, this project establishes a responsible foundation for ethical AI solutions, benefiting both the organization and its employees.

# Bibliography

[1]: IBM. (n.d.). *AI Ethics.* Retrieved from <https://www.ibm.com/think/topics/ai-ethics>

[2]: Springer. (2024). *Ethics in Artificial Intelligence: A Comprehensive Review.* Retrieved from <https://link.springer.com/article/10.1007/s13347-024-00710-6#Sec2>

[3]: The Alan Turing Institute. (n.d.). *The Future of AI Ethics.* <https://www.youtube.com/watch?v=viyXQhPAqME&ab_channel=TheAlanTuringInstitute>

[4] : Teradata. (n.d.). *The Importance of AI Ethics.* Retrieved from <https://www.teradata.com/insights/ai-and-machine-learning/importance-of-ai-ethics#:~:text=Bias%E2%80%94one%20of%20the%20biggest,insights%20will%20also%20be%20biased>.

[5] : European Commission. (n.d.). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment.* Retrieved from <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

[6] : Business Queensland. (n.d.). *SWOT Analysis: A Strategic Planning Tool.* Retrieved from <https://www.business.qld.gov.au/running-business/planning/swot-analysis#:~:text=A%20SWOT%20analysis%20helps%20you%20assess%20internal%20factors%20that%20might,results%20from%20the%20SWOT%20analysis>.

[7]: Business-to-You. (n.d.). *SWOT Analysis Explained.* Retrieved from <https://www.business-to-you.com/swot-analysis/>