# Exercise 03

**Group:** Jellis Hurtekant, Beppe Verheijden, Stephen Bombeke, Evarist Verstraete, Serafim Ciobanu, Los Elvis

**Goal:** We have been hired by a company, which is launching a couple of AI products – AI powered recruitment tool, AI-driven personalized healthcare platform, and AI for fraud detection in finance. We have to check the compliance with global regulations and ethicality.

**Questions to answer:**

**Ethical Guidelines:**  
• a. What core ethical principles will the company follow in its AI development?  
• b. How will these principles be applied to the AI systems the company develops (e.g., fairness, privacy, transparency)?

**AI Governance Committee:**  
• a. Who should be included in the AI governance committee (e.g., roles from various departments)?  
• b. What will the committee's responsibilities be?  
**Policies and Protocols:**  
• a. What specific policies should be implemented to ensure data privacy, fairness, and accountability?  
• b. How will the company ensure that the AI models are free of bias? What protocols will be in place for regular auditing?

**Regulatory Compliance:**  
• a. Which regulations (e.g., GDPR, EU AI Act) should TechSphere Inc. comply with, considering its global customer base?  
• b. How will you ensure ongoing compliance with these regulations as AI technology evolves?  
**Monitoring and Evaluation:**  
• a. How will TechSphere Inc. monitor and evaluate its AI systems to ensure they are functioning ethically over time?  
• b. What performance indicators should the company track (e.g., bias reduction, compliance rates)?

## Ethical Guidelines

First of all, it is essential to mention the necessary principles of ethicality that the company needs to comply with.

The company is supposed to ensure transparency of the AI systems, as these are supposed to tackle some serious problems like **fraud detection** and **recruitment**. The technics that are used for various decision-making in such cases should be as accurate as possible, as this decides the different outcomes for people and not only. Fairness is another crucial principle that comes out of the other principle, as yet again, it is very crucial to minimize the **unfair** choices that the system might undertake. Safety and privacy are the other 2 crucial principles that should be complied with in the case of “TechSphere” company, as the company will most probably tackle crucial medical data for example, and in that case any move can get the system to make a **wrong decision,** while privacy should also be primordial, due to the **data** that is fed to the systems for trainings, and the outcomes of the systems, which again is related to medical information of patients and etc. This type of data should be tackled as **securely** as possible.

To ensure that **fairness** is implemented into the systems, for example the recruitment tool and fraud detection should be fed with as **much various data** as possible. That will ensure that the bias is getting limited to the minimum and hence provide more accurate results in the end. To implement **transparency**, the system should explain the **decision** that has been taken towards any of the results, and why something has been decided. That way, the transparency will be enhanced, and users will have clearer view of the decision making process. **Privacy** wise implementation, the system of healthcare for example, should make sure that the **data is encrypted** with consideration of the best practices, and the user’s data should be tackled with ultimate consent. It should also be noted that **data minimization** will help massively in such cases as fraud detection and recruitment, as there is no need for all the possible personal data to give a result. Considering the privacy implementations, the **safety** of the systems and data will enhance the usage of the AI, and further safety is supposed to be enhanced by various risk mitigations.

## AI Governance Committee

### Who should be included in the AI governance committee (e.g., roles from various departments)?

The AI governance committee should hire people from diverse backgrounds to cover all departments:

* **Leaders:** Senior executive and product managers for each AI product
* **Legal Experts**: Lawyers and privacy officers to ensure compliance with laws like GDRP.
* **Ethics Advocates:** Someone to focus on fairness, transparency and social impact.
* **Tech Specialists:** AI engineers, data scientists and IT security professionals to handle the technical side.
* **Industry Experts**: Specialists in healthcare and finance to ensure regulatory compliance.
* **HR and Inclusion:** HR reps to ensure unbiased recruitment and diversity advocates to prevent discrimination.
* **External voices:** independent ethics reviewers and customer advocates to bring outside perspectives.

### What will the committee's responsibilities be?

The committee will guide how AI is developed and used responsibly by:

* **Setting Rules:** Creating policies to ensure data privacy and legal compliance
* **Auditing AI:** Regulary checking for bias, accuracy and ethical guidelines
* **Managing Risks:** identyfing potential issues, like unfair hiring or healthcare errors and proventing them.
* **Handling Incidents:** Responding to any problems quickly, such as data breaches or ethical complaints
* **Educating Teams:** training employees on AI ethics and best practices
* **Improving Continuously:** Updating practices as technology and regulations evolve.

By combining technical expertise, the committee ensures the company’s AI tools are trustworthy.

Policies and Protocols  
a. What specific policies should be implemented to ensure data privacy, fairness, and accountability?

* **Data Privacy:** they should adopt robust data governance frameworks aligned with international standards like GDPR. Policies must define how data is collected, stored, processed, and shared to ensure user consent and data minimization. Encryption and anonymization should be mandatory for sensitive data processing.
* **Fairness**: Implement algorithmic fairness guidelines requiring continuous assessment of AI models for equitable outcomes across different demographic groups. Establish protocols to identify and mitigate disparate impacts during development and post-deployment phases.
* **Accountability**: Develop clear accountability structures within the organization. Use AI transparency tools to ensure explainable outputs and maintain logs for auditing AI decisions, especially in high-stakes use cases.

b. How will the company ensure that the AI models are free of bias? What protocols will be in place for regular auditing?

* **Bias Mitigation** protocols: conduct comprehensive bias testing during model development and deployment using diverse datasets representative of the target population.

Apply bias mitigation techniques such as re-weighting data, adversarial debiasing, or introducing fairness constraints during training

* **Regular Audits:** schedule periodic algorithmic audits to evaluate bias and ethical compliance.  
  Document audit results and corrective measures to create feedback loop for continuous improvement.
* **Stakeholder Involvement**: Include input from diverse stakeholder groups, including ethicists, domain experts, and affected communities, to improve the inclusiveness and fairness of the AI models

Regulatory Compliance  
• a. Which regulations (e.g., GDPR, EU AI Act) should TechSphere Inc. comply with, considering its global customer base?  
GDPR, EU AI Act, HIPAA, FCRA, FINRA, Sarbanes-Oxley Act, PIPEDA, DPDP Act, LGPD, PIPL, OECD AI Principles, UNESCO AI Ethics Recommendations  
• b. How will you ensure ongoing compliance with these regulations as AI technology evolves?   
To ensure ongoing compliance, TechSphere Inc. will monitor regulatory changes, conduct regular audits, train staff, maintain robust data governance, and design adaptable systems that align with evolving laws.

## Monitoring and Evaluation

To ensure that the system’s performance can achieve the needed levels for the company and provide meaningful results for the users. To achieve maximum performance, it is necessary to implement systems for auditing and monitoring. There are auditing tools that can track decisions and detect anomalies in the system’s behavior. This can ensure continuous improvement and analysis. It might also be useful to implement deviation detection, which can be considered an anomaly, and ensure that the system stays on the same line. Bias testing can also be implemented, which will lower the chances of AI models being yet again out of line. Tools that handle this for us are Microsoft Fairlearn and Google What-If.

Other things that should be considered in such AI systems are compliant with various regulations like GDPR and EU AI Act. Ensuring continuous checks is essential for the company providing services, as this will make the systems more trustworthy, and overall safer to use.