# PRINCIPLES AND POINTS OF FOCUS

recommending points of focus for enterprises to implement responsible AI principles and mitigate risks arising from AI adoption

## Principle 1

Inclusivity and Non-Discrimination — AI systems must be fair and inclusive, not fostering prejudices, discrimination or preference for an individual, a community or a group based on their sensitive attributes (e.g., race, gender, ethnicity).

| | Point of Focus |
|---|---|
| 1 | Set clear goals for diversity and inclusion |
| 2 | Create a diverse AI workforce |
| 3 | Ensure system fairness by countering various sources of bias and using representative data sets in AI modelling |
| 4 | Minimise the use of sensitive data |
| 5 | Ensure quality and correction of data annotations |
| 6 | Test the AI model with diverse user groups |

## Principle 2

Reliability and Safety — AI systems must produce consistent and reliable outputs in all scenarios. Appropriate grievance redressal mechanisms should be put in place to address cases of adverse impact.

| | Point of Focus |
|---|---|
| 1 | Ensure reproducibility of outcomes |
| 2 | Monitor feedback loops to the system |
| 3 | Conduct quality assurance checks using risk-based stress testing techniques |
| 4 | Choose appropriate algorithms for a problem |
| 5 | Conduct post-production monitoring to detect model or data drifts |
| 6 | Effectively decommission in the event of system failure |
| 7 | Place human supervisory control, wherever possible |
| 8 | Adopt MLOps |
| 9 | Implement appropriate grievance redressal and compensation mechanisms |

## Principle 3

Privacy — AI systems should respect user privacy. Users' right to know what data is collected, why it is collected and who has access to it should be protected. AI systems should not use the data for purposes other than what is stated.

| | Point of Focus |
|---|---|
| 1 | Minimize the use of personal data |
| 2 | Get user consent before collecting and storing personal data |
| 3 | Implement appropriate control measures to safeguard personal data collected |
| 4 | Leverage on-device processing, whenever possible |
| 5 | Consider machine learning techniques like federated learning |
| 6 | Be transparent about what data is collected from the user, how it will be used and who has access to it, to the extent possible |
| 7 | Avoid repurposing of personal data collected |
| 8 | Ensure that third or fourth party vendors implement all necessary data privacy controls |
| 9 | Set up mechanism for users to opt in and out from sharing personal data |
| 10 | Set up mechanisms for users to access, mange and delete data they generate |

## Principle 4

Security — AI system should be robust and secured against adversarial attacks and malicious use. Identifying and mitigating system vulnerabilities is critical.

| | Point of Focus |
|---|---|
| 1 | Implement protocols and measures against unauthorised access |
| 2 | Implement adequate controls to guard against adversarial attacks, data poisoning, model stealing, malicious use |
| 3 | Follow secure coding practices to reduce system vulnerabilities |
| 4 | Check for vulnerabilities in open source components and components provided by third party vendors |
| 5 | Ensure that third or fourth party vendors implement all necessary security controls |

**Principle 5**

Transparency — AI systems should be transparent about how they were developed, their processes, capabilities and limitations, to the extent possible.

| | Point of Focus |
|---|---|
| 1 | Ensure proper documentation throughout the model lifecycle to allow external scrutiny and audit |
| 2 | Allow visibility into training data - what data was used, source of the data, how it was processed, what features were used etc., to the extent possible |
| 3 | Maintain transparency about capabilities and limitations of the system |

**Principle 6**

Explainability — Users should be able to request explanations for significant decisions taken by AI systems. Explanations must be provided free of cost to the user and should contain human understandable summary of how the system arrived at a particular decision.

| | Point of Focus |
|---|---|
| 1 | Enable means for users to seek explanations for decisions significantly impacting them |
| 2 | Provide explanations for significant decisions taken by the system in non-technical, simple and intuitive language to maintain user trust |
| 3 | Avoid using complex algorithms that make model explanations difficult |

**Principle 7**

Accountability — Organisational structures and policies should be created to clarify who is accountable for the outcomes of AI systems. Human supervisory control of AI systems is recommended.

| | Point of Focus |
|---|---|
| 1 | Set up organisational structures and policies to identify people accountable for the outcomes of an AI system throughout its lifecycle |
| 2 | Establish clear roles and responsibilities of various stakeholders throughout the lifecycle of AI systems |

| | Point of Focus |
|---|---|
| 3 | Facilitate the adoption of responsible AI principles by third or fourth party vendors through training workshops, etc. |
| 4 | Establish human supervisory control over AI system, wherever possible |
| 5 | Seek user feedback |
| 6 | Take responsibility for the larger societal impact of the AI system |

**Principle 8**

Protection and Reinforcement of Positive Human Values — AI systems should be designed and operated such that they align with human values. AI should promote positive human values for the progress of humanity as a whole.

| | Point of Focus |
|---|---|
| 1 | Conduct critical review of proposed use-cases - anticipated benefits, harms, and overall impact on society |
| 2 | Do not pursue unethical use-cases |
| 3 | Pursue model alignment with the values and beliefs of the target user groups to the extent possible |
| 4 | Adopt human-centered AI design to provide optimal user experience |

**Principle 9**

Compliance — Throughout their lifecycle, AI systems should comply with all applicable laws, statutory standards, rules, and regulations – Organizations should be watchful of the evolving AI regulatory landscape and ensure compliance at all times.

| | Point of Focus |
|---|---|
| 1 | Proactively monitor and build organisational awareness around the evolving AI regulatory landscape |
| 2 | Ensure compliance with all applicable laws, rules, and regulations to avoid financial penalties and reputational loss |