.addepto

Trustworthy AI Checklist for 2024

Safety & Security

- [ ] Design AI systems with robust security measures.
- [ ] Implement secure development practices, data encryption, access controls, and vulnerability assessments.
- [ ] Employ encryption techniques and access controls to safeguard sensitive information.
- [ ] Implement swift responses for identifying and mitigating security threats.

Respect for Privacy

- [ ] Uphold privacy standards and protect user data from unauthorized access.
- [ ] Only use personal data for agreed-upon purposes.
- [ ] Implement privacy-preserving techniques such as Differential Privacy and federated learning.

Transparency & Explainability

- [ ] Provide insights into AI decision-making processes and ensure transparency in algorithms and data usage.
- [ ] Enable users to understand and challenge AI-generated outcomes.

.addepto

Trustworthy AI Checklist for 2024

Reliability

- [ ] Ensure seamless, consistent, and safe AI system functioning.
- [ ] Use techniques like adversarial training and error-handling mechanisms for robustness.

Fairness

- [ ] Design AI tools to treat all individuals fairly and without bias.
- [ ] Employ fairness-aware algorithms, bias detection tools, and fairness metrics.
- [ ] Train AI models on diverse data to improve accuracy and reliability.

Accountability & Responsibility

- [ ] Implement audit trails, logging mechanisms, and transparency reports.
- [ ] Encourage collaboration and foster a culture of responsible ownership throughout the AI lifecycle.