Responsible AI

AI Privacy & AI Justice

Definition of AI Privacy

Privacy in AI means giving people control over their own information—who collects it, how it’s used, stored, and shared. It’s about protecting personal data from misuse or unauthorized access.

Core Principles

- **Consent:** Always ask before collecting personal or sensitive data. Consent should be clear and informed.
- **Data Minimization:** Only collect the minimum amount of data needed. Avoid gathering unnecessary sensitive information.
- **Transparency:** Explain in simple terms what data is collected and how AI will use it.
- **Protection:** Use strong security (like encryption and access controls) to keep the data safe.
- **Control:** People should have rights to see, correct, or delete their data from AI systems.

Examples

- **Search history and ads:** When a search engine suggests ads based on your recent searches, it’s using your data. Privacy means you get to choose whether that happens.

- **Medical AI systems:** These often analyze health records. They must get clear permission and keep data secure to protect patients’ privacy.

Highlight Privacy Risks

- Collecting sensitive data (like health or financial details) without permission.
- Using data for purposes people never agreed to (for example, training AI without explicit consent).
- Data leaks—either by accident or through hacking

Customers: Is Your AI Protecting Your Privacy? Here's What to Ask

- What personal data is being collected, and why?
- How is my data stored, protected, and for how long?
- Will my data be shared or sold to third parties?
- Can I consent to data use, and can I withdraw consent?
- How can I access, correct, or delete my data?
- Is my data used to train AI systems, and can others benefit from it?
- What happens if there is a data breach? How will I be informed?

AI Privacy: Questions Organization Should Ask

- What types of personal or sensitive data does our AI system use?
- Is our data processing compliant with data privacy laws like GDPR, CCPA, or new local regulations?
- Who has access to the data within our organization and among vendors?
- Have we performed a Data Protection Impact Assessment (DPIA)?
- How do we minimize data collection and retention?
- What data security protocols are in place?
- Is customer data logically separated from other data and can it be deleted upon request?
- What processes exist for responding to data privacy incidents or breaches?

Definition of AI Justice

Justice in AI means fairness: ensuring that AI systems do not discriminate, are accessible to everyone, and that their decisions can be explained and challenged.

Core Principles

- **Fairness:** AI should not treat some people better or worse based on race, gender, income, age, etc.
- **Non-discrimination:** Algorithms must avoid biases that result in unfair outcomes.
- **Transparency:** AI decisions should be explainable—people should know how and why an AI made a decision.
- **Accountability:** There must be ways to appeal or question decisions made by AI.
- **Inclusivity:** AI should help, not exclude, marginalized or vulnerable groups.

Examples

- **Hiring tools:** If an AI system is used to screen job resumes, is it fair to everyone? If it tends to favor men over women, that’s an issue of justice.

- **Criminal justice:** If AI helps decide who gets bail, it must avoid using past biased data that could continue unfair treatment.

What Makes AI Justice Important?

- Unfair AI systems can amplify real-world discrimination.
- Lack of justice reduces trust and harms society as a whole.

AI Justice: Questions Customers Should Ask

- Is the AI system fair, or does it discriminate against certain groups?
- How does the AI’s decision-making process work? Can it be explained to me?
- What should I do if I think an AI system made an unfair or biased decision?
- Are there processes to contest or appeal AI decisions?

AI Justice: Questions Organization Should Ask

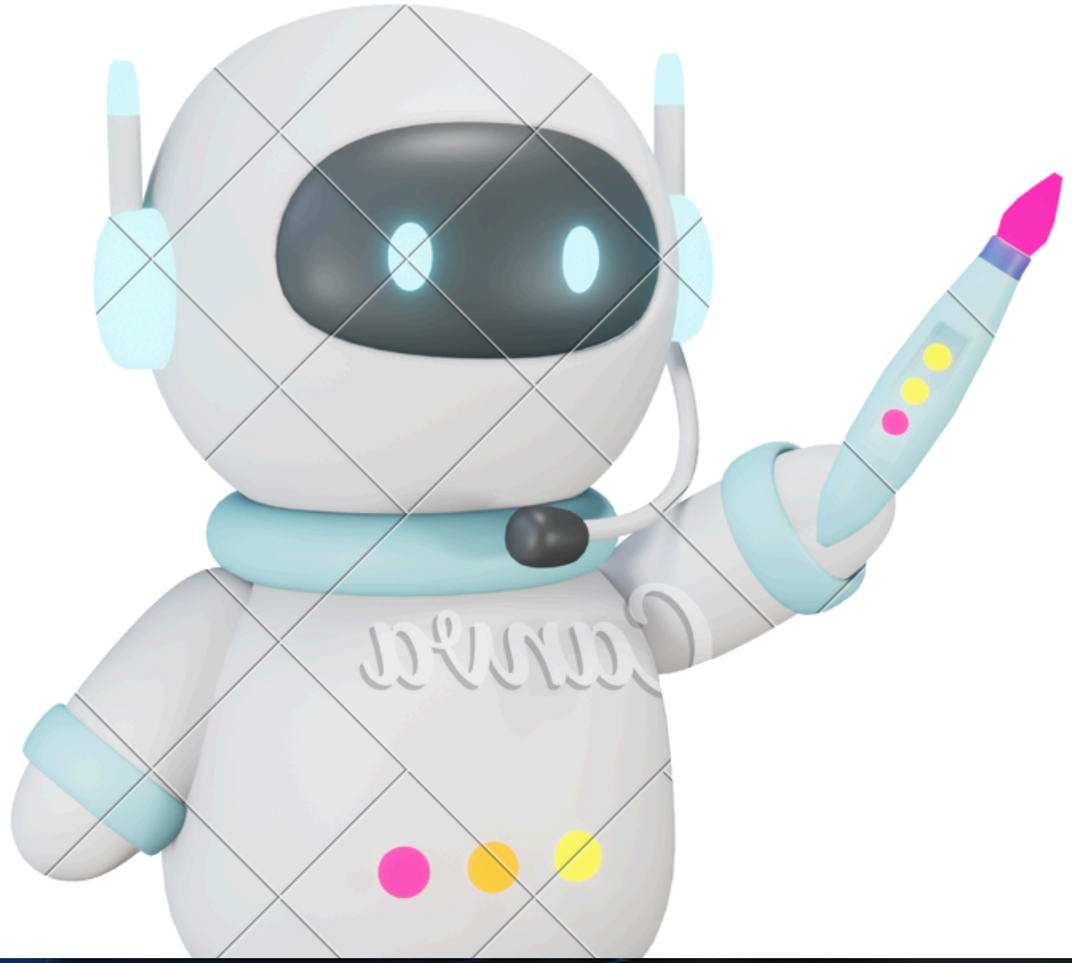
- How do we define and measure fairness in our AI models?
- What steps are taken to detect and mitigate bias in data and algorithms?
- Are our AI systems audited to detect and address discriminatory results?
- Do we have mechanisms for explainability and transparency in AI decisions?
- Have we involved diverse stakeholders to ensure our system doesn’t disadvantage particular populations?
- What processes are available to help users challenge or appeal AI decisions?
- How do our AI practices align with legal protections against discrimination and human rights principles?
- What mechanisms do we have to hold the team accountable for justice-related issues arising from AI use?

References

- https://owasp.org/www-project-ai-security-and-privacy-guide/
- https://www.scalefocus.com/blog/ai-security-and-privacy-guide
- https://www.youtube.com/watch?v=wg4BlOhiVaI
- https://sloanreview.mit.edu/article/putting-responsible-ai-into-practice/

Further Readings

- https://hai-production.s3.amazonaws.com/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf
- https://dl.acm.org/doi/fullHtml/10.1145/3657054.3657141
- https://www.amacad.org/publication/daedalus/toward-theory-justice-artificial-intelligence
- https://pmc.ncbi.nlm.nih.gov/articles/PMC9510536/



Thank You

Build Powerful AI.
Build with Ethics.
Build with Governance.
Build with Safety & Security.
Build with Privacy & Justice
Build Together.