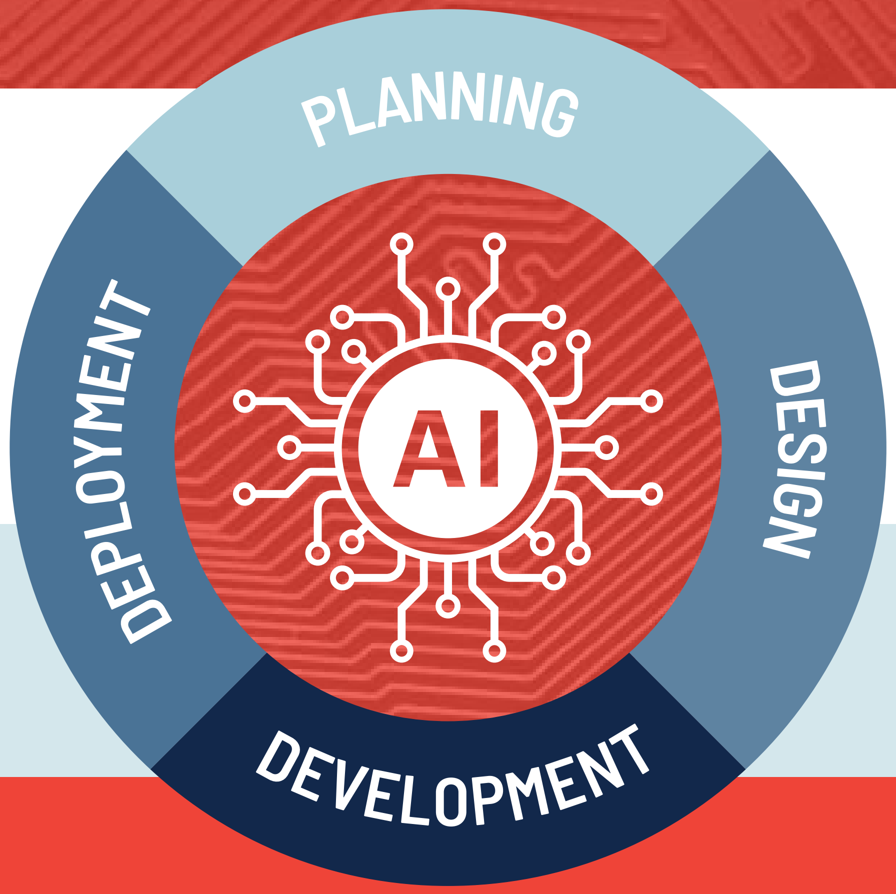
AT-A-GLANCE

AI Governance in Practice Report 2024

iapp

FTI CONSULTING | TECHNOLOGY



THE AI LIFE CYCLE

Organisations are responsible for managing risks and harms throughout the AI lifecycle by implementing effective AI governance controls.

See the full AI Governance in Practice Report 2024 for the most common and critical challenges in developing and deploying AI, and actionable remediation strategies.

CHALLENGES

Data

The quality of training, testing, validation and operational datasets used to develop and operate AI systems can generate risks.

Privacy and data protection

The inherent dependency of AI on data can conflict with fundamental privacy principles like data minimization and purpose specification.

Transparency, explainability and interpretability

Users and developers both struggle to explain or interpret the inner workings and outputs of AI systems.

Bias, discrimination and fairness

Bias can encode into AI throughout the system life cycle through the data, algorithms or humans involved in developing and deploying the system.

Security and robustness

Compromised security of AI systems could lead to a range of harms, from incorrect outputs to physical harm.

AI safety

Safety risks include alignment, security, malicious use and rogue behavior risks.

Copyright

Training data for generative AI may include copyrighted content, raising issues related to infringement and fair use.

PRACTICAL APPROACHES

Data:
- DATA GOVERNANCE
- DATA MANAGEMENT PLANS
- DATA LABELS
- CONFORMITY ASSESSMENTS

Privacy and data protection:
- INVENTORIES
- PRIVACY BY DESIGN
- RISK ASSESSMENTS

Transparency, explainability and interpretability:
- MODEL AND SYSTEM CARDS
- OPEN-SOURCE AI
- WATERMARKING

Bias, discrimination and fairness:
- PUBLIC-FACING INTERNAL AI ETHICS POLICIES
- BIAS TESTING
- IMPACT ASSESSMENTS
- HUMAN IN THE LOOP

Security and robustness:
- RED TEAMING
- SECURE DATA SHARING PRACTICES SUCH AS DIFFERENTIAL PRIVACY
- POST-MARKET MONITORING

AI safety:
- PROMPT ENGINEERING
- REPORTS AND COMPLAINTS
- SAFETY BY DESIGN
- INTERNAL AI SAFETY POLICIES
- TESTING AND EVALUATION

Copyright:
- OPT OUTS
- LIABILITY CONSIDERATIONS
- TECHNICAL GUARDRAILS SUCH AS CONTENT FILTERS, ABUSE DETECTION AND CLASSIFIERS

CROSS-CUTTING APPROACHES: RISK MANAGEMENT | TARGET OPERATING MODELS | POLICY AND PROCEDURE | COMPLIANCE ASSESSMENTS | TRAINING AND AWARENESS

IAPP disclaims all warranties, expressed or implied, with respect to the contents of this material, including any warranties of accuracy, merchantability, or fitness for a particular purpose. Nothing herein should be construed as legal advice.

Published September 2024.
iapp.org