MSECB

Webinar Power Point

Ethical AI: Addressing Bias and Transparency with ISO/IEC 42001

Webinars by MSECB

Varun Prasad

MSECB

Webinar Agenda

Presentation

Ethical AI: Addressing Bias and Transparency with ISO/IEC 42001.

Q&A

MSECB

MSECB Webinar

Background of Varun Prasad

Varun Prasad is an experienced IT audit and risk management professional with over 15 years in the field. Specializing in cloud security and privacy, he is an MSECB-approved auditor for ISO/IEC 27001, ISO 22301, and CSA STAR. With a Master's degree in Electrical Engineering from West Virginia University, Varun holds several certifications, including CISA, CISM, CCSK, CIPM, and PMP.

Throughout his career, he has worked with top consulting firms and multinational corporations across various industries, including technology, aerospace, and financial services. Varun is passionate about staying at the forefront of evolving technologies and is particularly interested in cloud security and the growing role of AI in cybersecurity.

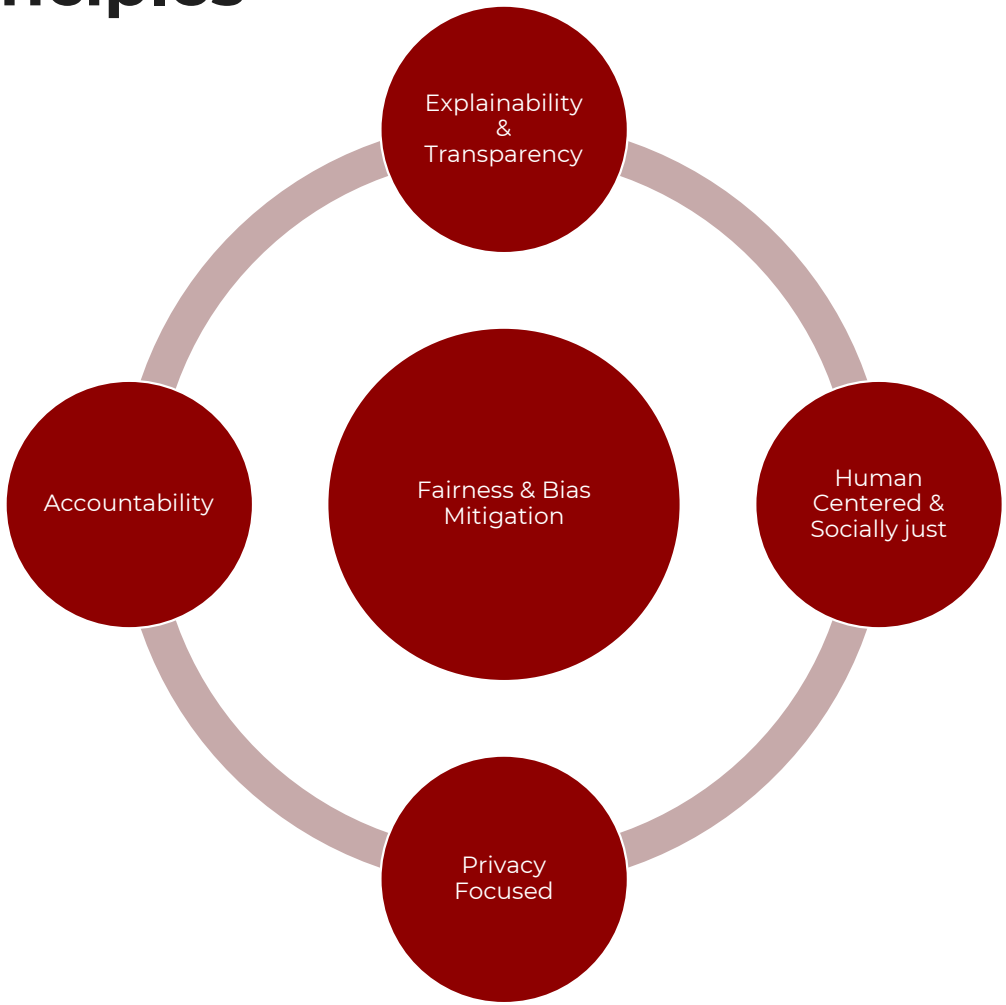
MSECB

Ethical AI – the what and why

- Set of principles and rules for development & deployment of AI technologies that promotes ethical, responsible and fair use of these technologies.
- AI systems introduce novel risks due to its in-built cognitive abilities.
- AI systems make or influences significant decisions with far-reaching consequences – provide or withdraw services.
- AI is perceived as a ‘black box’ eroding trust in society.
- Comply with wave of regulations.

MSECB

Ethical AI Principles



MSECB

Types of Bias in AI

Data Bias	• Data selection bias • Sample bias
Algorithmic Bias	• Design flaws • Model training errors
System Output Bias	• Confirmation bias • Cognitive bias

MSECB

Bias in AI – Real world examples

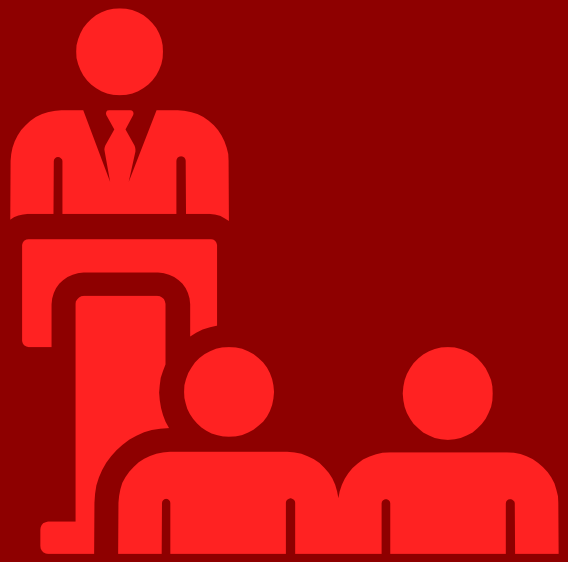
Amazon's AI Hiring Tool	COMPAS Criminal Justice tool	AI tool in healthcare
• Preferred male candidates • Representation bias	• Showed racial bias • Possible algorithmic bias • Erroneous data classification	• Racial bias towards persons of color. • Trained on insurance claims data.

MSECB

Overview of ISO/IEC 42001 Certification

- Introduced in Dec. 2023, world's first AI management system standard.
- Specifies the requirements and provides guidance for establishing, implementing, maintaining and continually improving an AI (artificial intelligence) management system (AIMS).
- Applicable for all organizations providing or using products or services that utilize AI systems.
- Helps organizations develop, use or provide AI systems responsibly to meet its objectives and applicable requirements.
- Addresses unique challenges and risks posed by AI like ethics/bias, transparency and explainability.
- Follows same structure as other ISO management system standards
  - Document the intended use of the AI system and the role of the organization.
  - Complete an AI risk assessment and a statement of applicability to denote the inclusion/exclusion of Annex A controls.
  - Requires the conduct of internal audits and management reviews.

MSECB

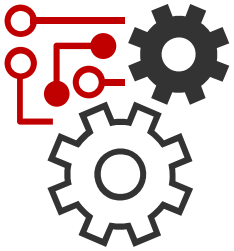


How ISO/IEC 42001 Certification promotes Ethical AI

- ✓ Establishing an AI ethics board.
- ✓ Incorporate Ethical AI principles within AI objectives.
- ✓ AI policies should cover specific rules and requirements for implementing Ethical AI concepts.
- ✓ Assign accountability within the organization for the implementation of AI principles.
- ✓ Periodic training for individuals involved in AI system development and operations.
- ✓ Repeat risk and system impact assessments at planned intervals.
- ✓ Internal audits to cover the effectiveness of controls in addressing bias and transparency.
- ✓ Include discussion points around Ethical AI requirements in management reviews.
- ✓ Implement applicable controls at each stage of the AI system lifecycle
    - Training data preparation
    - Training dataset diversity
    - Secure AI system development
    - Model validation
    - Model monitoring

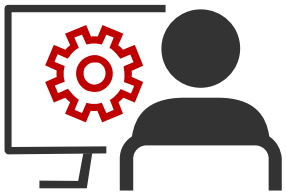
MSECB

Risk Assessments vs System Impact Assessments



OCONSIDERATIONSRGANIZATIONAL

- Complexity and relevance
- AI expertise
- Technology readiness
- Data governance
- Accountability and compliance

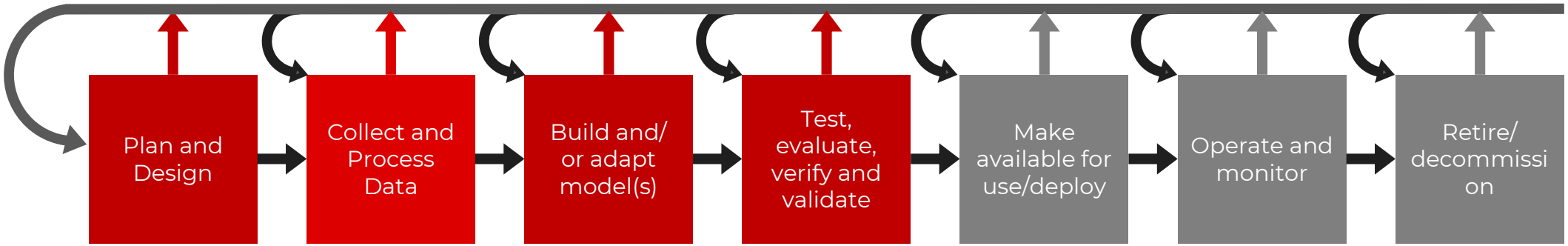


SYSTEMIC CONSIDERATIONS

- Model or use case level
- Technical aspects for safety and security
- Impact on users of the system – impact on individuals, groups, communities and organization
- Evaluate the applicability and risks of AI principles

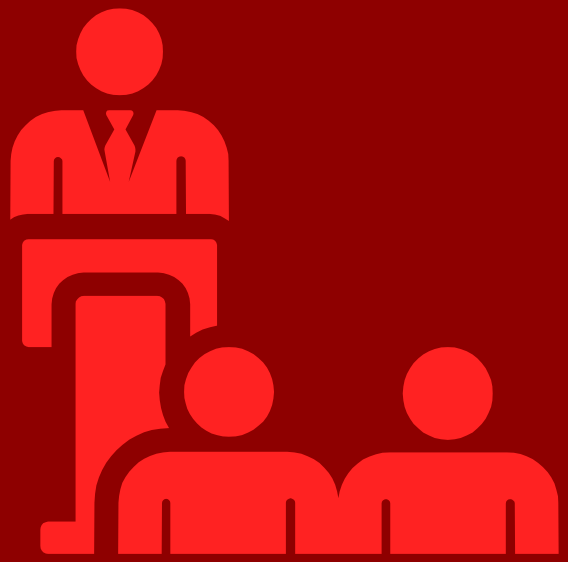
MSECB

AI System Lifecycle (OECD)



Source: www.oecd.ai/en/ai-principles

MSECB



Data Governance Controls

Define & Implement Data Management Practices

- Data management practices to address topics like:
  - privacy and security implications due to the use of data;
  - accuracy and integrity of the data.
  - transparency and explainability aspects including data provenance and the ability to provide an explanation of how data are used for determining an AI system's output; representativeness of training data compared to operational domain of use.

Data Acquisition & Preparation

- Implement processes for acquisition and selection of data used in AI systems.
  - Data sources and categories of data.
  - Characteristics of data source and attributes
  - Data cleansing
  - Data subject access rights

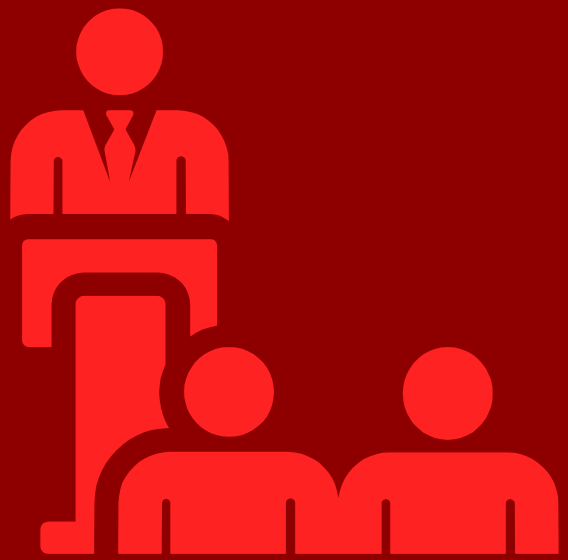
Data Quality

- Ensure data used in AI system lifecycle meets quality requirements.
  - Consider impact of bias on system performance and system fairness.
  - Training data is always representative of user population.

Data Provenance

- Document a process for recording the provenance of data.
- Ensure transparency and accountability.

MSECB



Model Design & Development

Objectives for Responsible Development

- Identify objectives that affect the AI system design and development including Ethical AI principles.
- Document steps to incorporate the objectives into every stage of system development from requirements specification, data acquisition, data conditioning, model training, verification and validation.

Model Desing & Development Process

- Similar to SDLC, document the lifecycle stages; approval and testing requirements; training data expectation and rules; release criteria; usability and change control.
- Secure coding requirements.

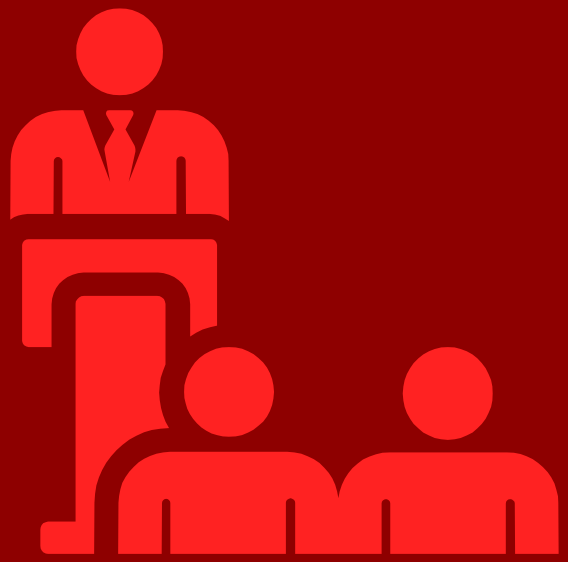
AI System Requirements & Specifications

- Include details around why the system is developed; model and data requirements.
- AI principles to be considered in the system.

AI System Design

- Include architecture details including hardware and software component.
- Key details must cover machine learning approach and learning algorithms and type of machine learning models to be used.
- Model training.

MSECB

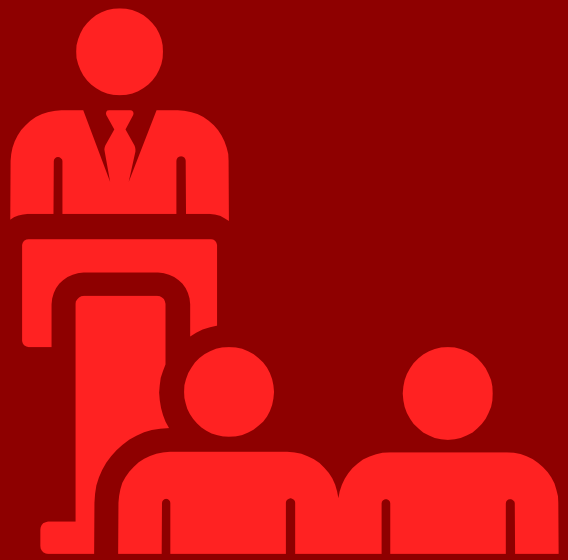


Model Verification & Validation

AI System Testing

- Evaluation plan to cover the following:
  - Selection of test data and requirements to ensure it's representative of the user base.
  - Reliability and safety requirements of the AI system, including acceptable error rates for the AI system performance;
  - Responsible AI system development and use objectives;
  - Operational factors such as quality of data, intended use, and sandboxing.
- Key metrics and acceptable deviations.
- Types of model testing strategies include fairness test; robustness and accuracy; adversarial AI testing and red teaming.

MSECB



Model Monitoring

AI System Deployment

- Document a deployment plan or checklist that ensures:
  - Verification and validation objectives including KPI are met.
  - Human-in-the-loop

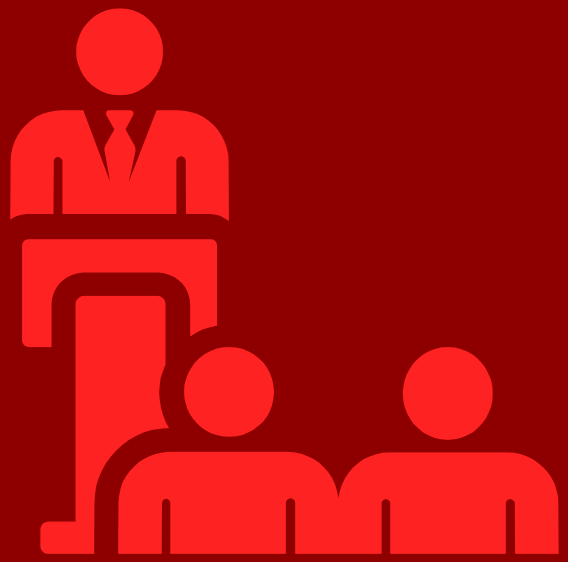
AI System Operations & Monitoring

- Covers system and performance monitoring, repairs, updates and support.
  - Monitor for general errors or failures.
  - Monitor AI observability metrics like accuracy and precision of outputs; bias; data drift; explainability
  - Identify AI-specific information security threats like data poisoning, model inversion attacks etc.
  - React and responds to alerts.

AI System Logging

- Ensure logging of the AI system at various phases to enable traceability and facilitate troubleshooting.

MSECB



The Future of ISO/IEC 42001 in operationalizing Ethical AI

- ❑ The growing need for AI governance – provides a baseline framework.
- ❑ First-of-its-kind standard for AI that companies can get certified with.
- ❑ Build trust and increase confidence with customers and stakeholders.
- ❑ Helps address vendor risk assessment requirements.
- ❑ Demonstrate commitment to Responsible AI principles.
- ❑ Helps comply with key regulations like the EU AI Act and other US federal (proposed) and state legislations.
- ❑ Leverage early adopter advantage.

MSECB

Q&A

MSECB

Thank you for your attention!

Stay updated!

+1 (450) 328-1227

info@msecb.com

www.msecb.com