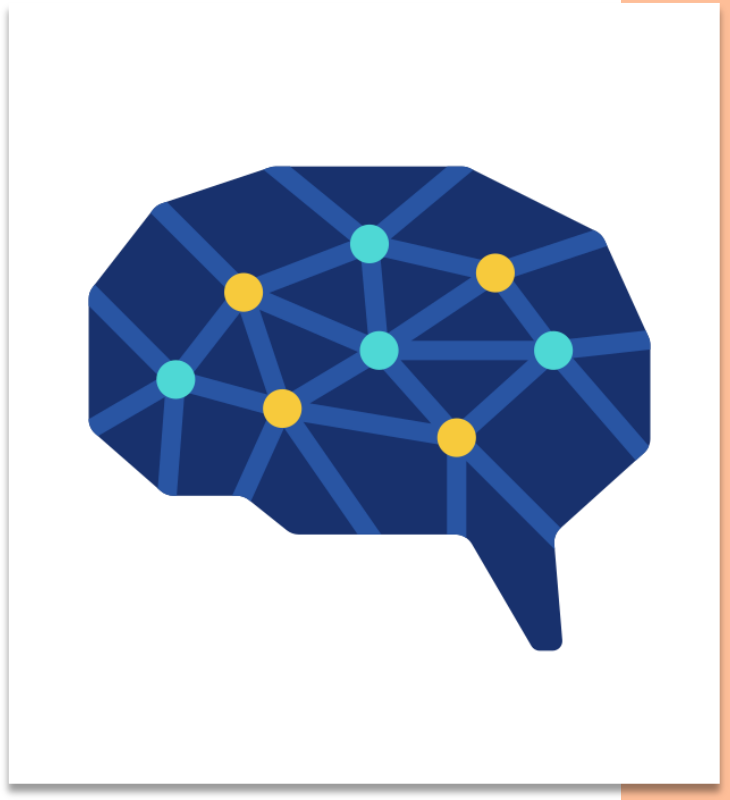
Security requirements for AI systems



19 February 2024

Zoe Sungchae Park (Ph. D. Candidate)
Senior Researcher, Soonchunhyang University
Korea (Rep. of)

1. Intro

- The potential impact of artificial intelligence……? “I won’t anything harm you.”

(Ref: https://www.youtube.com/watch?v=pv5zIIVla5g)

2. ITU-T X.sr-ai: Security requirement for AI systems (Q15/17)

[2022-2024] : [SG17] : [Q15/17]

[Declared patent(s)] - [Associated work]

Work item:	X.sr-ai
Subject/title:	Security requirements for AI systems
Status:	Under study
Approval process:	TAP
Type of work item:	Recommendation
Version:	New
Equivalent number:	-
Timing:	2026-09 (Medium priority)
Liaison:	ITU-T SG13, ISO/IEC JTC 1/SC 27, ISO/IEC JTC 1/SC 42, ETSI ISG-SAI
Supporting members:	Korea (Republic of), Soonchunhyang University, ETRI, KISA
Summary:	Artificial Intelligence (AI), and Machine Learning (ML) are increasingly being used by whole industries leveraging digital technologies. AI enables industries to provide differentiated products and services to customers and facilitates businesses in achieving more effective decision-making and higher operational performance. There is a need for identifying security threats arising from the increasing use of AI and the security requirements to effectively address them within AI systems for entire AI system lifecycle to ensure safe use and operation of AI systems. In order to more efficiently identify security threats and provide a set of detailed security requirements in addition to high-level mitigations described in ISO/IEC 27090 for addressing identified security threats, a completely understanding of AI system models and the lifecycle of AI systems is necessary. AI system lifecycle is composed of six stages: planning, data preparation, model design, model training/development, model deployment, model operation/monitoring & maintenance, which is modified from those provided by OECD (2022) [b-oecd]. This draft Recommendation is based on the AI system model defined by ISO/IEC 27091. Therefore, the draft Recommendation will provide a set of comprehensive detailed security requirements to address the security threats for organizations using and operating AI systems to effectively address the security threats in AI systems. This draft Recommendation can be used by organizations, which are involved in whole lifecycle of AI systems.
Comment:	-
Reference(s):	[SG17-TD1348R1-2/PLEN (2023-08)] Historic references: -
Contact(s):	Daeun Hyeon, Editor Jae Nam Ko, Editor Junhyung Park, Editor Sungchae Park, Editor Heung Youl Youm, Editor
ITU-T A.5 justification(s):	- [Submit new A.5 justification] See guidelines for creating & submitting ITU-T A.5 justifications

First registration in the WP: 2023-09-20 23:29:57

Last update: 2023-09-20 23:48:52

1. Scope
2. References
3. Definitions
4. Abbreviations and acronyms
5. Conventions
6. Overview
   - 6.1 AI system lifecycle
   - 6.2 Stakeholders
   - 6.3 AI system model
   - 6.4 Choosing the machine learning model
7. Security threats and associated risks
8. Security requirements

3. AI system lifecycle

- Planning: The planning stage involves assessing the scope, success metric, and feasibility of the AI application. The business should be understood and how to use machine learning to improve the current business process should be identified.
- Data preparation: The data preparation stage is further divided into four parts: data collection and labelling, data cleaning, data processing, and data management.
- Model design: In this stage, all the information from the planning phase is used to build and train a machine learning model, tracking model metrics, ensuring scalability and robustness, and optimizing storage and compute resources.
- Model training/development: The model on a test dataset is tested and the error in the predictions are identified. It follows industrial, ethical, and legal frameworks for building AI solutions. The model for robustness on random and real-world data is tested and the model inferences is proven to be fast enough to bring the value.
- Model deployment: AI models are deployed to the current system. A computer vision model into the current system is deployed.
- Operation/Monitoring and Maintenance: After deploying the model to production the AI system should be monitored and improved.

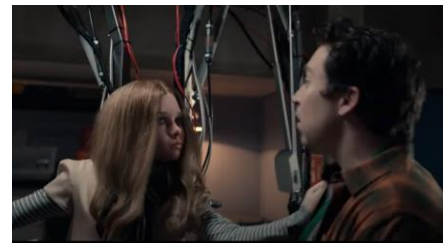
[Lifecycle of AI systems]

4. Stakeholders

- When considering the lifecycle of an AI system, the following stakeholders are involved in the planning, design, development, deployment, operation, and maintenance of an AI system.
  - ✓ **Organization management**: AI system's strategy, goals, resources, ethics, evaluation. Lead AI adoption aligned with business objectives.
  - ✓ **AI model architects**: Choose algorithms, design & fine-tune models, prioritize scalability & interpretability, and stay updated on AI advancements. Collaborate to turn business problems into AI solutions.
  - ✓ **AI data scientists/engineers**: Prepare and manage data for models (cleaning, pipelines, features). Analyze, select data, handle bias and data protection. Build the foundation for strong, ethical AI models.
  - ✓ **AI developers**: Build and deliver AI solutions. Implement models, create APIs, deploy and optimize, test and debug, maintain infrastructure. Bridge the gap between models and real-world use.
  - ✓ **AI service operators**: Deploy and manage AI services. Monitor, update, maintain & secure AI systems. Ensure compliance & handle customer support. Keep AI running smoothly in the real world.

Proper functioning

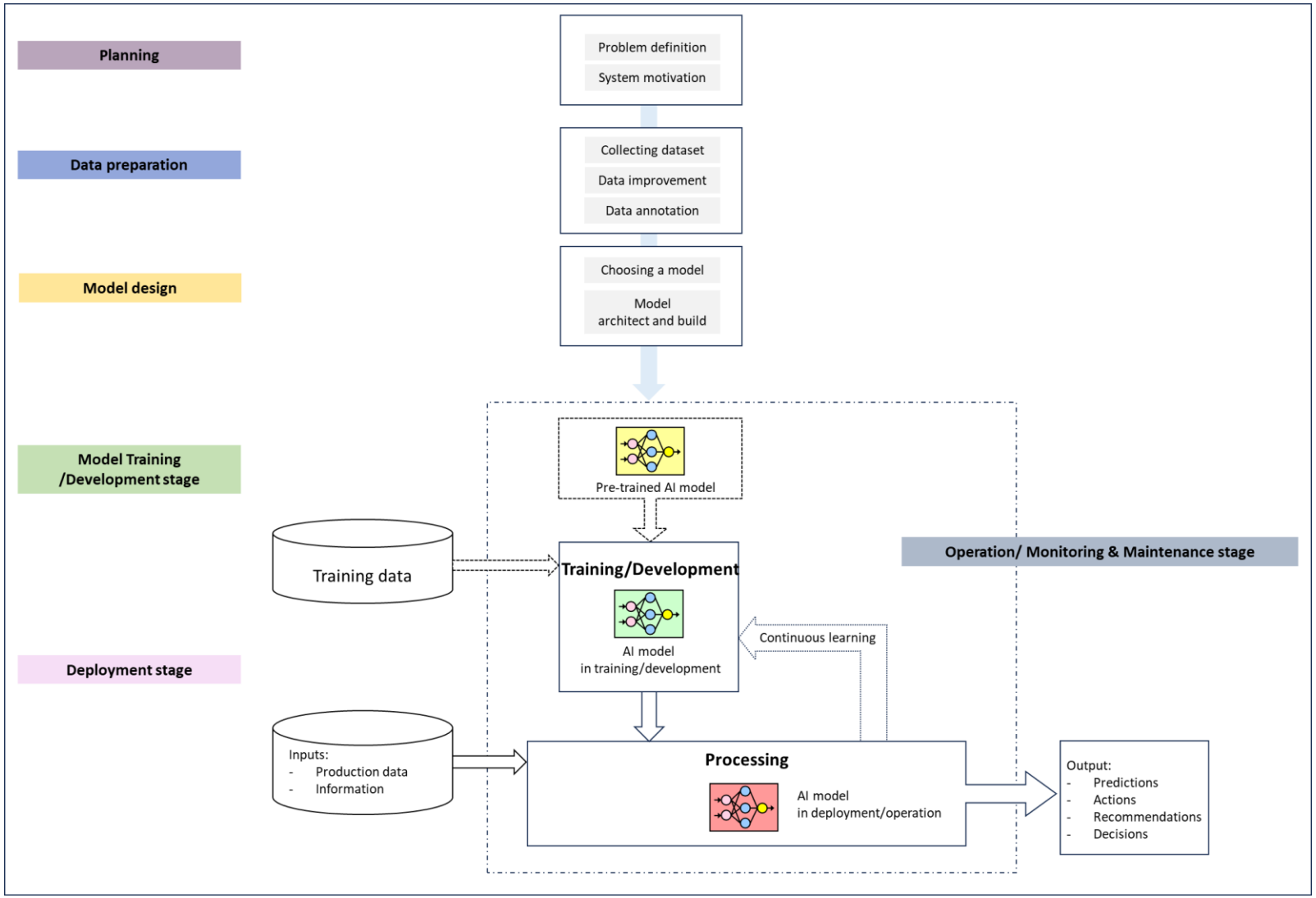
Malfunctioning



(Ref: A Movie M3GAN, 2023)

5. AI system model

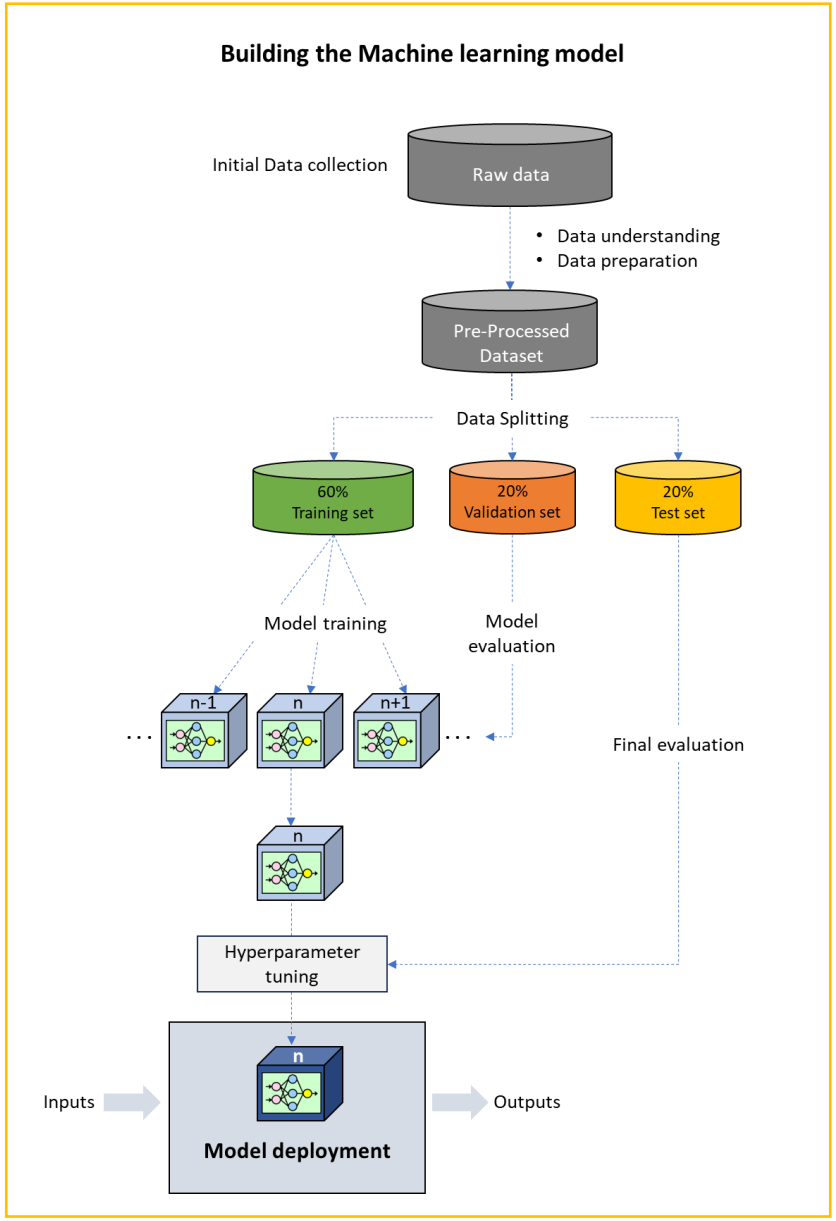
- AI system model based on ISO/IEC 27091 and ETSI GR SAI 005 considering its lifecycle.



6. How to choose the machine learning model

- Training: the process of a model learning from data. The model learns from the given data to perform a specific task.
- Evaluation: the process of measuring the performance of a model. Evaluation is typically performed using the training data. Evaluation can be used to see how much the model's performance has improved.
- Testing: the process of validating the performance of a model. Testing is typically performed using data that is different from the training data. Testing can be used to see how well the model performs on new data.

Initial data collection	Model evaluation
Understanding data Data preparation	Hyperparameter tuning
Data splitting	Final evaluation
Model training	Model deployment



7. Security threats and associated risks

- The security threats and associated risks can impact all stages of the AI lifecycle.

Security threats	Planning	Data preparation	Model design	Model training /development	Model deployment	Operation /Monitoring &Maintenance
• Algorithm bias discrimination			x	x		
• Code security vulnerability				x		
• Unbalanced training data		x		x		
• Training data poisoning		x		x	x	
• Evasion(input manipulation) attack		x		x	x	
• Membership inference		x		x		
• Model extraction				x	x	
• Model inversion		x		x		
• Unauthorized access, use, disclosure, disruption, modification, or destruction to training dataset				x	x	
• Weak AI algorithm to input datasets			x	x	x	

7. Security threats and associated risks

- The security threats and associated risks can impact all stages of the AI lifecycle.

Security threats	Planning	Data preparation	Model design	Model training /development	Model deployment	Operation /Monitoring &Maintenance
• Algorithm unexplainability				X		
• Adversarial attack				X		
• Contaminated or distorted data				X	X	X
. . .						

8. Security requirements(1/3)

- The following security requirements are identified to address security threats for AI systems.

6-stage AI system lifecycle	Security requirements
Planning	• Management is required to identify all potential security threats that may arise throughout the AI systems lifecycle and analyse their ramifications. • Management is required to put into place measures to address threats identified and to eliminate, prevent or mitigate their consequence resulted from threats, etc.
Data preparation	• AI data scientists and engineers are required to identify abnormal data by checking whether the data is normal or abnormal. • AI data scientists and engineers are required to prepare countermeasures against such security threats as data poisoning and evasion, etc. • AI data scientist and engineers are required to protect training dataset from unauthorized access, use, disclosure, disruption, modification, or destruction. • AI data scientist and engineers are required to prepare proper datasets for training and evaluation, which includes cleaning, formatting, and normalizing datasets, etc.
Model design	• AI model architects are recommended to consider risk management process taking into account whole lifecycle of AI systems which consist of risk identification, risk analysis, risk evaluation, and risk treatment. • AI model architects are recommended to identify security threats and vulnerabilities inherent to the AI system they are designing and establish appropriate security countermeasures to address them, etc.

8. Security requirements(2/3)

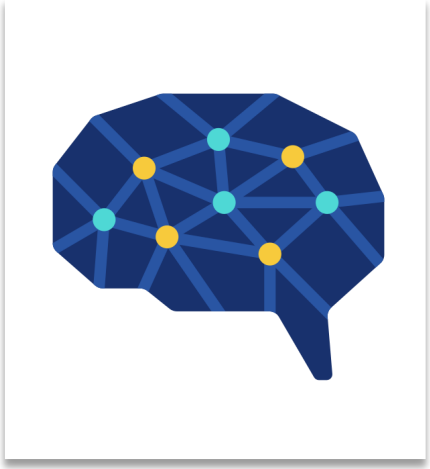
- The following security requirements are identified to address security threats for AI systems.

6-stage AI system lifecycle	Security requirements
Model training/Development	• Developers are required to check for security threats and vulnerabilities in the open-source libraries they are using. • Developers are required to apply mitigation techniques against model extraction attacks. • Developers are required to apply defenses against model evasion attacks. • Developers are required to check the operational stability of open-source libraries included in the AI system. • Data scientists are required to store training data in secure environments, such as encrypted databases, with restricted access to authorized persons only. It is recommended for them to regularly audit access logs and monitor data usage to detect suspicious activity to manage the data securely, etc.
Deployment	• AI system operators are recommended to put in place risk management process, which consists of risk identification, risk analysis, risk evaluation, and risk treatment continuously and repeatedly for deployment and operation stage of the AI system lifecycle to address the security threats identified, etc.

8. Security requirements(3/3)

- The following security requirements are identified to address security threats for AI systems.

6-stage AI system lifecycle	Security requirements
Operation/Monitoring & Maintenance	• AI service operators are recommended to provide explanations to guide the correct use of AI services. AI service operators are required to explain to AI service users the purpose, intent, role, and scope of the AI services they provide, and to disclose information such as how the decisions of AI services affect service users. • AI system operators are recommended to continuously monitor the learning datasets and models where an AI system is designed and developed to learn during operation. • AI system operators are recommended to define the information to be logged and monitored at each stage of the system to track the impact of AI system inference results that may be caused by functional aspects such as the construction of AI models, datasets, and the system itself, as well as human factors such as AI system operators and users, etc.



Thank you

hyyoum@sch.ac.kr
zoesc.park@sch.ac.kr