

Lifecycle processes and checklists for AI data

Jang-Yeol Kim, Jong-Gyun Choi

Korea Atomic Energy Research Institute(KAERI)

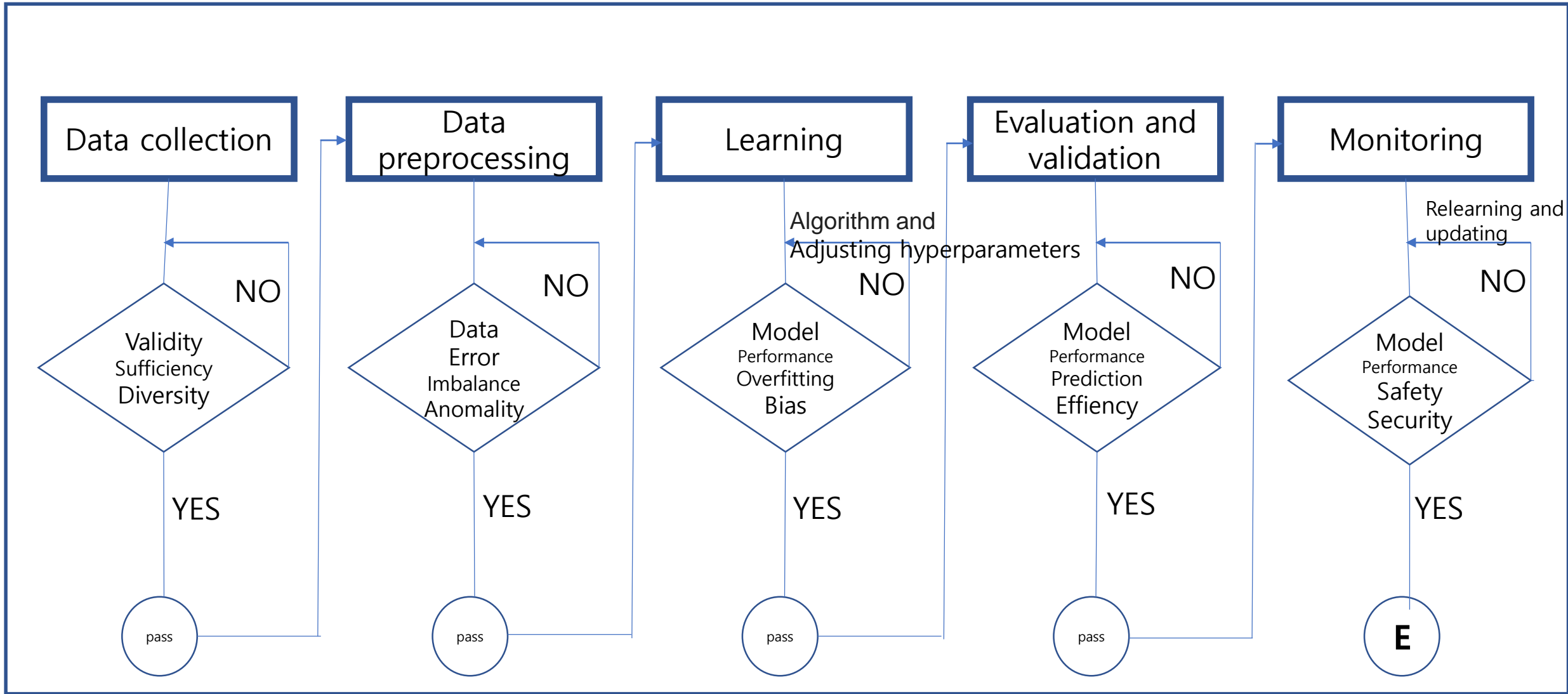
Table of Contents

- Relationship of Dataset and AI Model
- Basic processes of AI data lifecycle
- Expansion processes of AI data lifecycle
- Checklist for Data Requirement Definition Adequacy
- Checklist of Data Specification
- Data Supplier Criteria/Procedure
- Conclusion

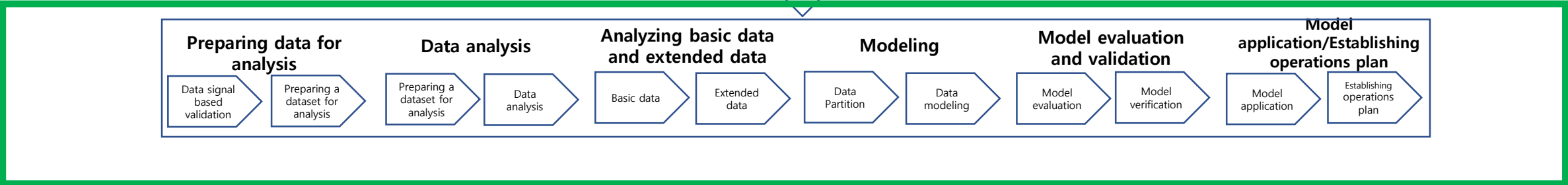
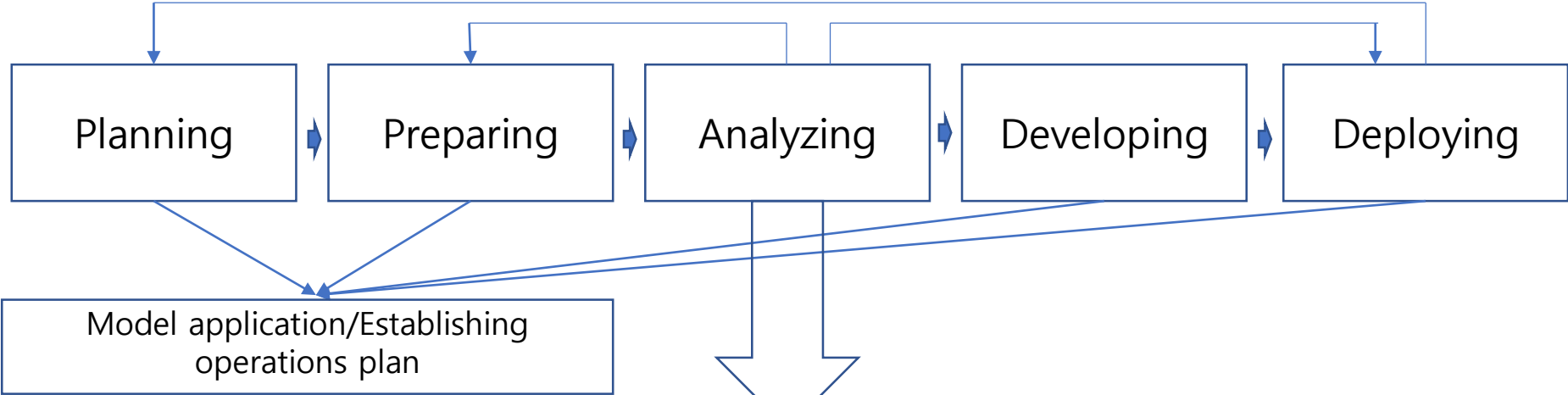
Relationship of Dataset and AI Model

- Relationship between data and artificial intelligence (AI)
 - AI model requires to create an dataset
 - AI needs good quality big data to learn and make right decision makings
 - > Trying to MODEL mapping: modify, and improve/optimize models that fit your data

Basic process of AI data lifecycle



Expansion process of AI data lifecycle



Checklist for Data Requirement Definition Adequacy

Requirement Definition Adequacy

1. Have data processing requirements been appropriately managed?

1.1 Are the preparing of data processing targets (requirements) and acceptance criteria without omissions?

1.2 Are the acceptance criteria (inspection methods, pass/fail criteria) specific and clear?

Checklist of Data Specification

Design and implementation data adequacy	2. Do the data processing requirements and data contents of the data specification each other? 2.1 Is the processed data implemented according to the processing methods and procedures included in the requirements and acceptance criteria? 2.2 Is the processed data implemented according to the data type, format, and columns (attributes) defined in the data specification?
Adequacy of data duplication check	3. Is the processed data not duplicated with publicly available data? 3.1 Was the source data collected independently? 3.2 Is the processed data not duplicated with public data provided by AI hubs homepages and public data portals?
Adequacy of Data licenses	4. Do you have a valid license for the dataset you are processing? 4.1 Does the dataset use data that is already publicly available?

Data Supplier Criteria/Procedure

- **Preparation for data business**
 - Data Provider education (ethics education, Ombudsman education)
 - Personal information protection education
- **Data business title (000): Requirements and acceptance criteria**
 - Data processing methods and procedures (supplier self-diagnosis)
 - Data Product Specification
 - Processed data
- **Source of original data, self-check for data duplication (self-diagnosis of supplier)**
 - [AI Hub](https://aihub.or.kr) (https://aihub.or.kr)
 - [Public data portal](https://data.go.kr) (https://data.go.kr)
- **Defect repair and maintenance plan**
- **Inspection confirmation**
 - (1) Inspection Criteria
 - (2) Acceptance Criteria
 - (3) Data Product Specification

Conclusion

○ Among the topics of interest in data-driven artificial intelligence, this paper “newly defines each phase of the life cycle on artificial intelligence,” and lifecycle definition is also essential for verification and verification.

- Verification and validation of AI systems
- Functional classification of AI systems
- Life cycle of AI system
- Security of AI systems
- Integration of operator and AI
- Risk analysis of AI systems etc.