Taxonomy of Machine Learning, and its Role in Developing AI Systems

Trung T. Pham, Ph.D.

`trung.t.pham@faa.gov`

05/13/2025

Federal Aviation Administration

Disclaimer

- **The objective of this technical presentation is only to provide an engineering perspective of the AI/ML technology in the effort to clarify what it can and cannot do, with examples in aviation that can foster conversation of the safe use of the technology.**

Agenda

- Introduction
- Taxonomy of Machine Learning
- Data Analytics with Unsupervised Machine Learning
- Developing AI Applications Through Supervised Machine Learning
- FAA’s Roadmap for AI Safety
- Data in the Context of Machine Learning
- Discussion

Introduction

- **The fast pace of AI applications being introduced to the market probably was due to the use of Machine Learning for developing them**

Introduction

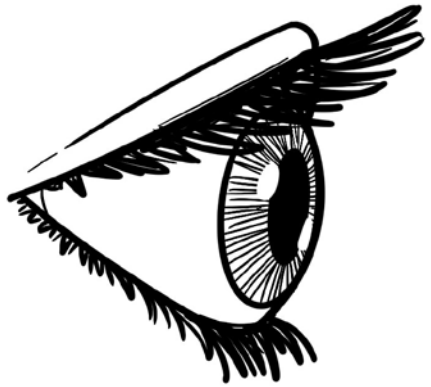
- The fast pace of AI applications being introduced to the market probably was due to the use of Machine Learning for developing them
  - **Machine Learning promises to shorten the development cycle**

Introduction

- The fast pace of AI applications being introduced to the market probably was due to the use of Machine Learning for developing them
  - Machine Learning promises to shorten the development cycle
  - **Machine Learning alleviates the need of engineering expertise for the developers (lowering the cost of development)**

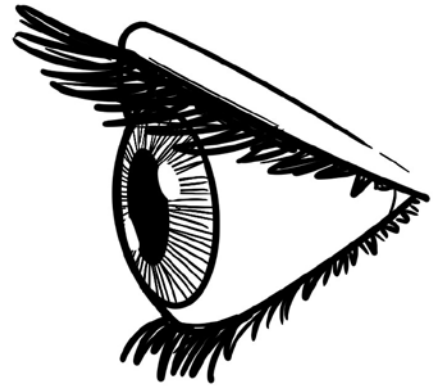
Introduction

ARTIFICIAL INTELLIGENCE



Public view of AI

AI Products	AI Related Technologies	SciFi AI Movies
• ChatGPT • Copilot • Alexa • Siri • Gemini • Deep Seek • Grammarly • etc.	• Large Language Model • Generative AI • Symbolic AI • Deep Learning • Reinforcement Learning • Artificial Neural Network • Computer Vision • etc.	• 2001: A Space Odyssey (1968) • Star Wars (1977) • Terminator (1984) • The Matrix (1999) • A.I. (2001) • I, Robot (2004) • Captain Marvel (2019) • etc.



Public view of AI

Taxonomy of Machine Learning

- **Machine Learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalise to unseen data, and thus perform tasks without explicit instructions.**

https://en.wikipedia.org/wiki/Machine_learning

Taxonomy of Machine Learning

- Machine Learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalise to unseen data, and thus perform tasks without explicit instructions.

https://en.wikipedia.org/wiki/Machine_learning

Generally speaking, Machine Learning is the use of computer (software) to capture useful information from data and transform it into some rules that can be reused later

Taxonomy of Machine Learning

- Machine Learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalise to unseen data, and thus perform tasks without explicit instructions.

https://en.wikipedia.org/wiki/Machine_learning

Generally speaking, Machine Learning is the use of computer (software) to capture useful information from data and transform it into some rules that can be reused later

- **Machine Learning is often thought of as mimicking human learning at computational machine**

Taxonomy of Machine Learning

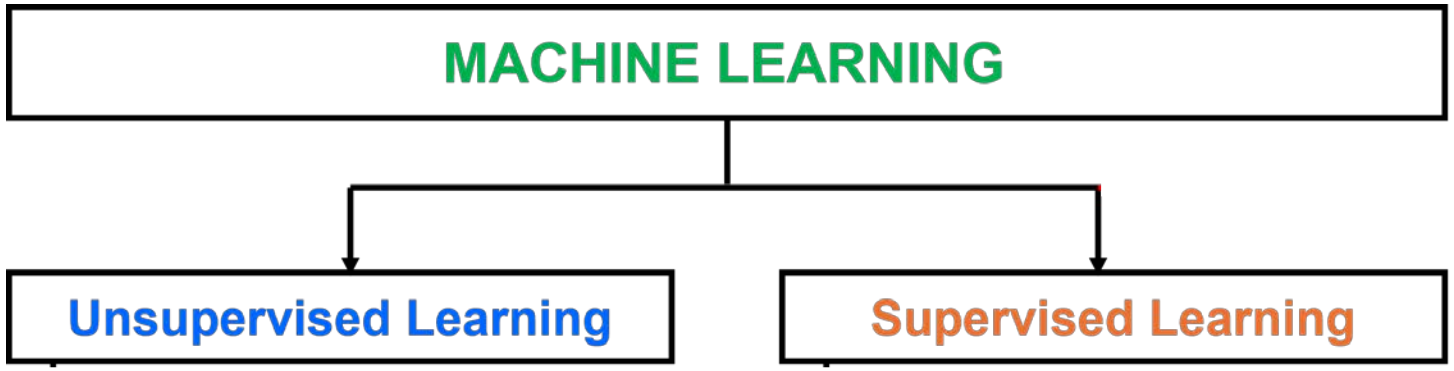
- Machine Learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalise to unseen data, and thus perform tasks without explicit instructions.

https://en.wikipedia.org/wiki/Machine_learning

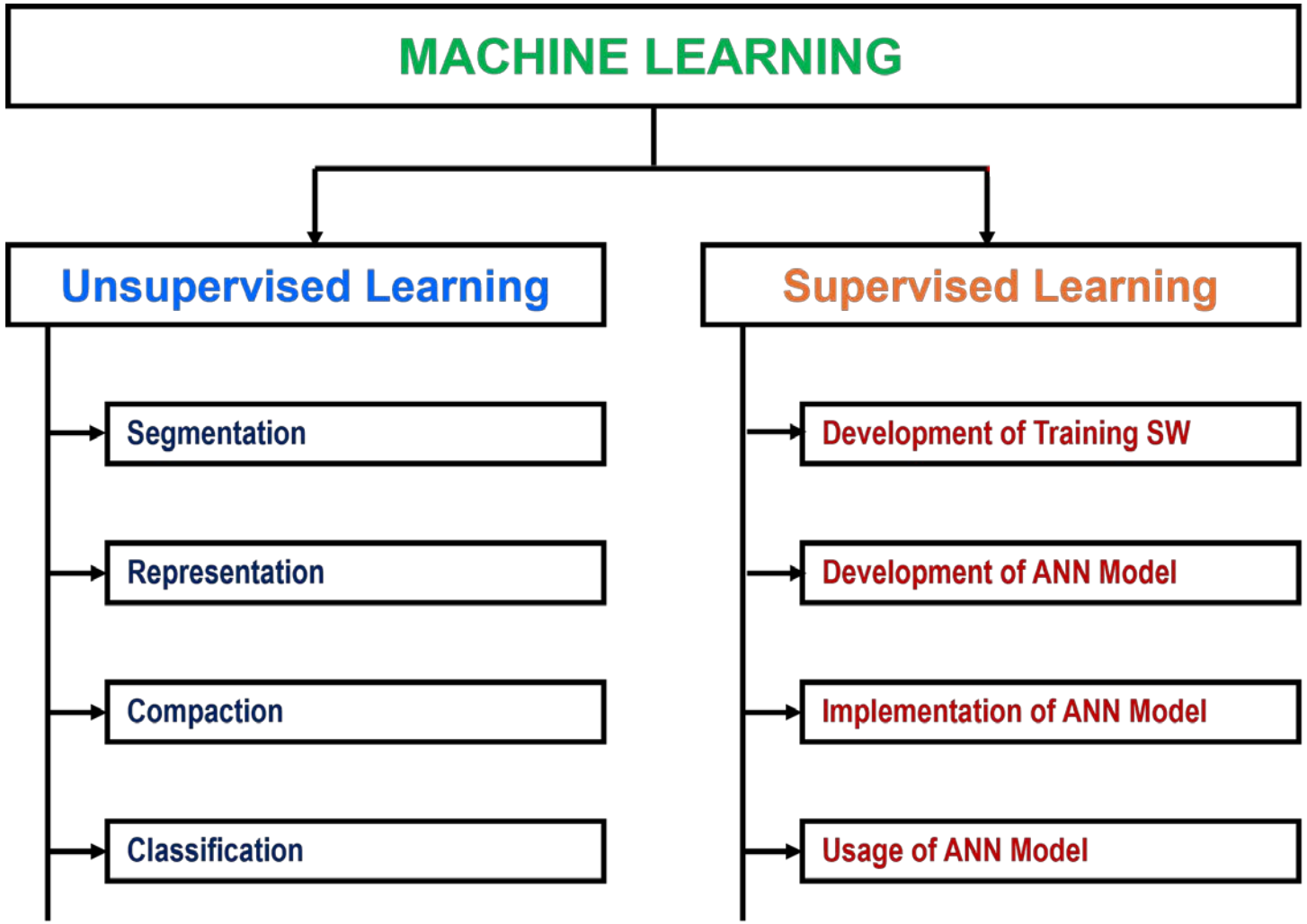
Generally speaking, Machine Learning is the use of computer (software) to capture useful information from data and transform it into some rules that can be reused later

- Machine Learning is often thought of as mimicking human learning at computational machine
- **Important Note: Machine Learning might appear to be similar to part of human learning, but the mechanism of doing it is completely different**

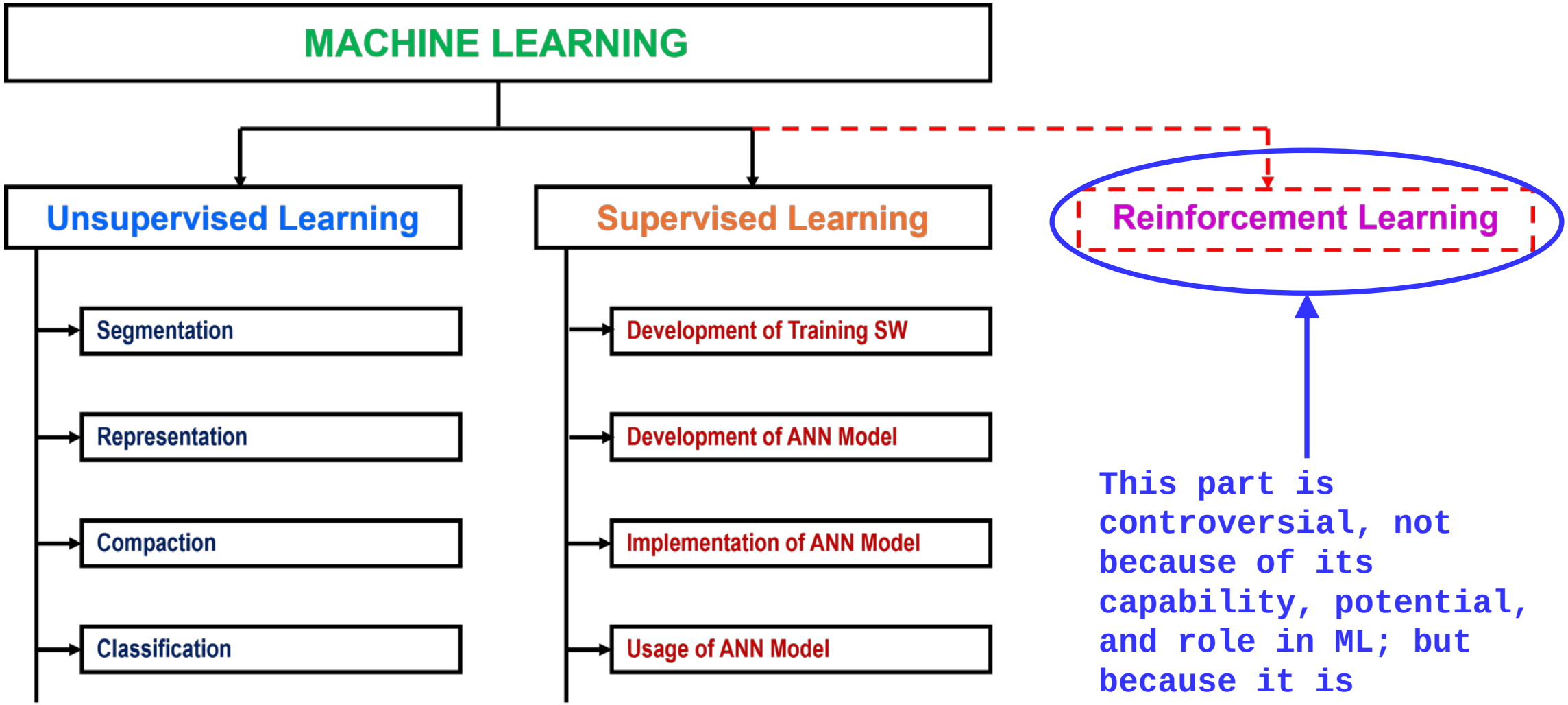
Taxonomy of Machine Learning



Taxonomy of Machine Learning



Taxonomy of Machine Learning



This part is controversial, not because of its capability, potential, and role in ML; but because it is incorrectly assumed to be one independent category of ML

Data Analytics with Unsupervised Machine Learning

- ***Data analytics*** **is the process of examining data to identify patterns, trends, and insights to support decision-making.**

Google AI Search

Data Analytics with Unsupervised Machine Learning

- **There are various methods of doing data analytics, one method is using Unsupervised Machine Learning to automatically discover the information (patterns, trends, and insights) embedded in the data**

Data Analytics with Unsupervised Machine Learning

- There are various methods of doing data analytics, one method is using Unsupervised Machine Learning to automatically discover the information (patterns, trends, and insights) embedded in the data
  - **real life example: Google is using Unsupervised Machine Learning (Data Mining) to discover the spending habit of people who use their search engine**

Data Analytics with Unsupervised Machine Learning

- There are various methods of doing data analytics, one method is using Unsupervised Machine Learning to automatically discover the information (patterns, trends, and insights) embedded in the data
  - real life example: Google is using Unsupervised Machine Learning (Data Mining) to discover the spending habit of people who use their search engine
  - **aviation example: to use Unsupervised Machine Learning to extract causal and situational data from various safety event reports (NTSB, ASRS, etc.)**

Will discuss this example later to illustrate the concept of Unsupervised Learning outlined earlier

Data Analytics with Unsupervised Machine Learning

1979 ORACLE DBMS™	1996 Google Search Engine™	2022 ChatGPT™
• organize data in a *structure* in a local file system	• data (in HTML format) are scattered across the Web	• data are scattered across the Web
• find data through *query* (SQL)	• find information on the Web based on *keyword*	• find information on the Web based on *humanlike conversations*
• generate *reports* in the form of charts and plots	• users must manually prepare *reports*	• generate *reports* in the form of comprehensive discussion

Key technologies that revolutionized how data are handled

Data Analytics with Unsupervised Machine Learning

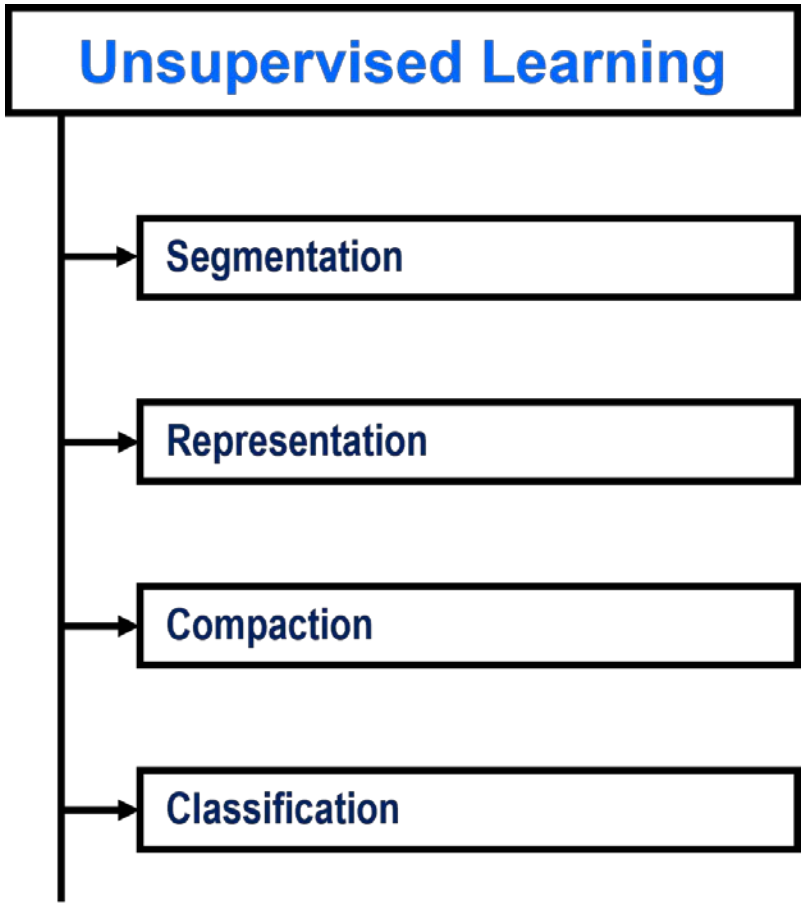
1979 ORACLE DBMS™	1996 Google Search Engine™	2022 ChatGPT™
• organize data in a *structure* in a local file system	• data (in HTML format) are scattered across the Web	• data are scattered across the Web
• find data through *query* (SQL)	• find information on the Web based on *keyword*	• find information on the Web based on *humanlike conversations*
• generate *reports* in the form of charts and plots	• users must manually prepare *reports*	• generate *reports* in the form of comprehensive discussion

Key technologies that revolutionized how data are handled

Large Language Model (LLM)

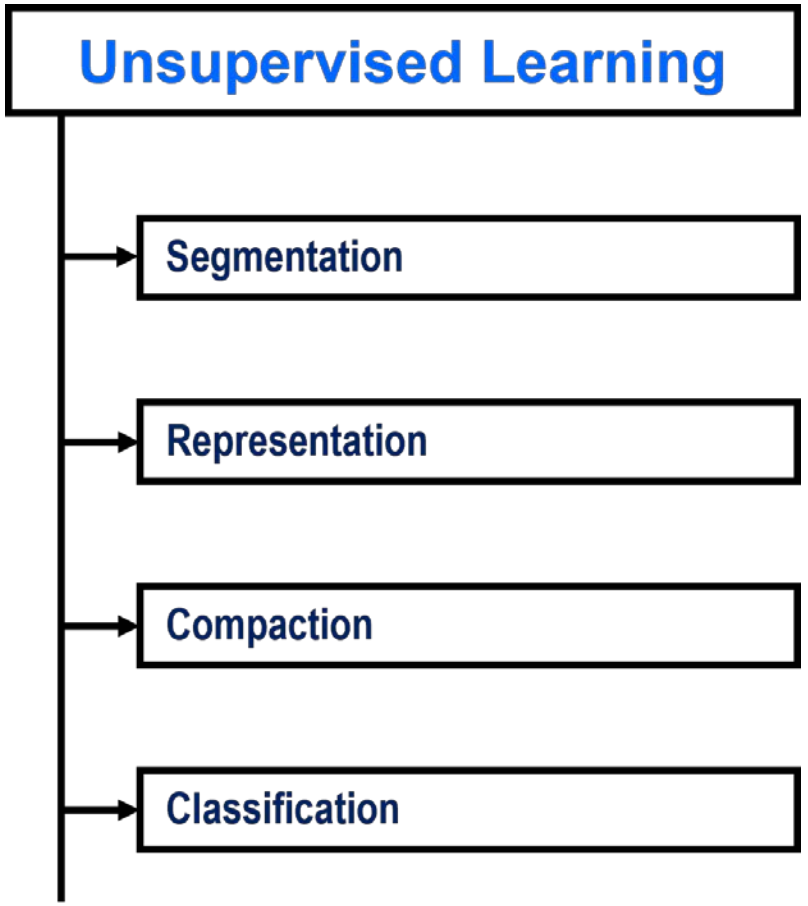
Data Analytics with Unsupervised Machine Learning

- **Unsupervised Learning consists of 4 steps**



Data Analytics with Unsupervised Machine Learning

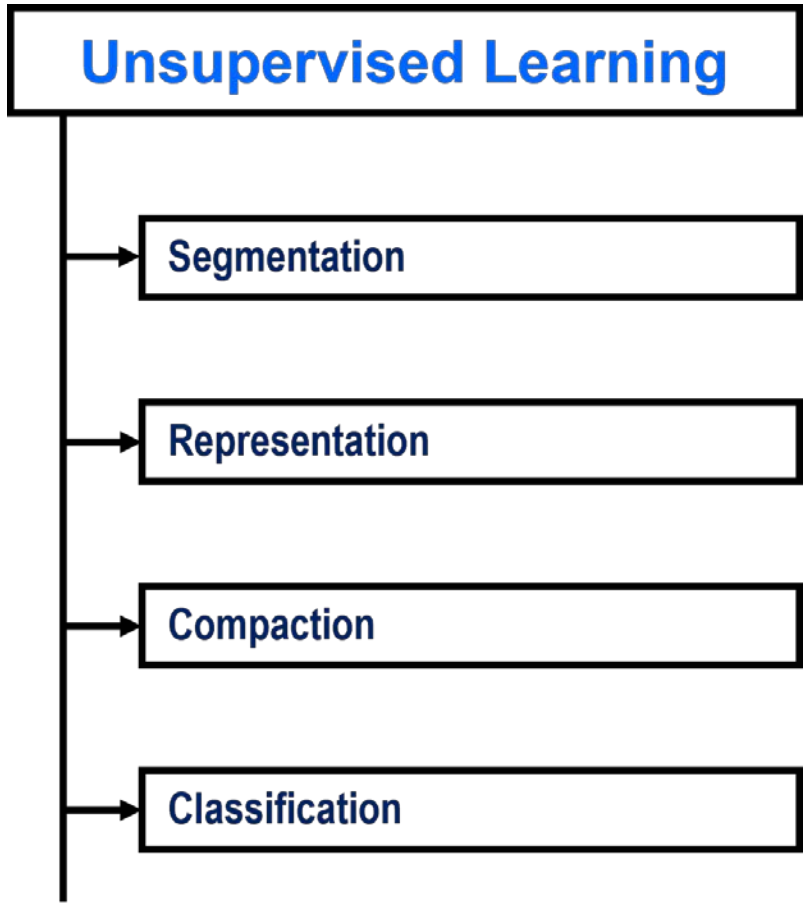
- **Unsupervised Learning consists of 4 steps**



Dividing a large set of data into smaller segments, each segment only contain data that are similar

Data Analytics with Unsupervised Machine Learning

- **Unsupervised Learning consists of 4 steps**

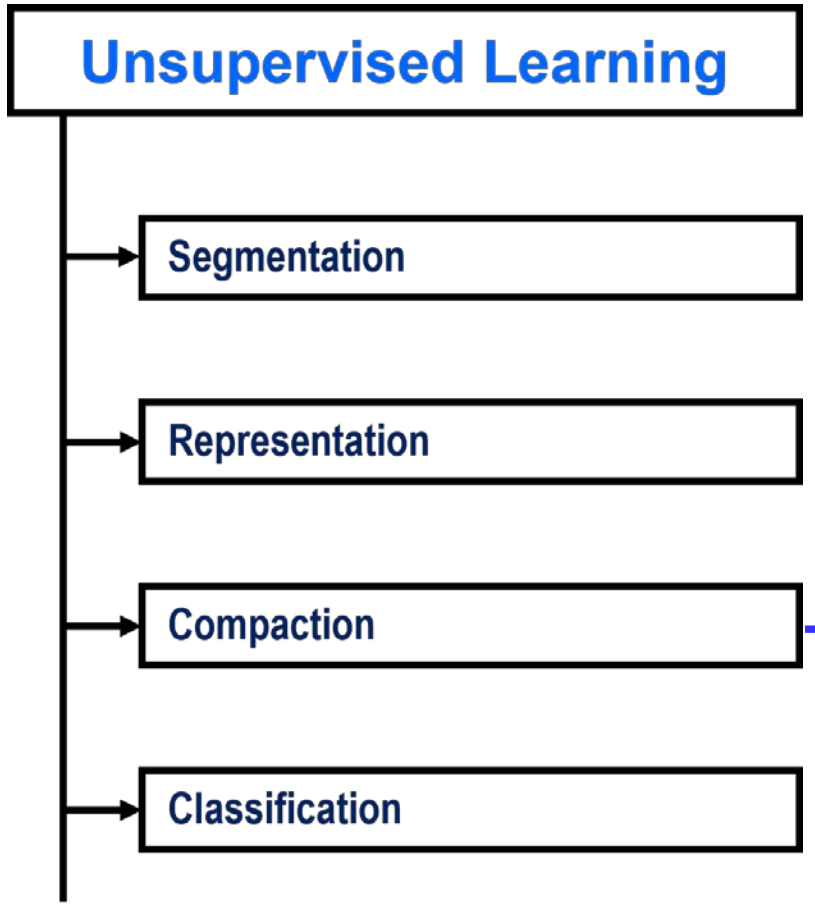


Dividing a large set of data into smaller segments, each segment only contain data that are similar

Extracting information (correlation, cause-and-effect relation, etc.) and put it in a understandable form

Data Analytics with Unsupervised Machine Learning

- **Unsupervised Learning consists of 4 steps**



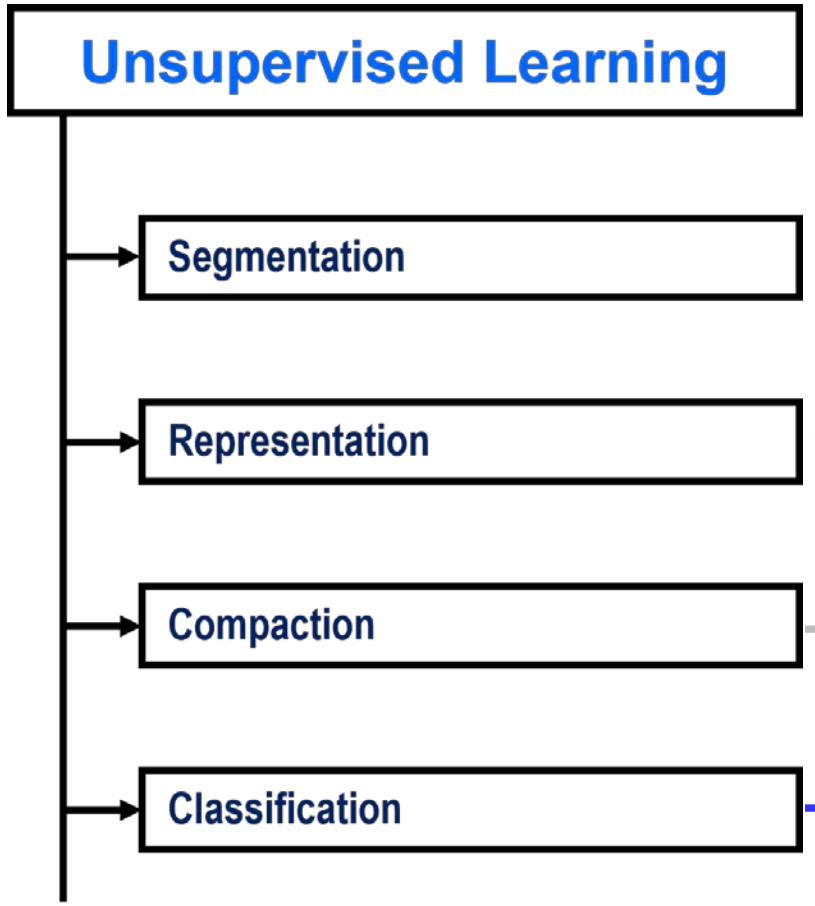
Dividing a large set of data into smaller segments, each segment only contain data that are similar

Extracting information (correlation, cause-and-effect relation, etc.) and put it in a understandable form

Reducing the format of the extracted information to a compact size (without redundancy, repetition, etc.)

Data Analytics with Unsupervised Machine Learning

- **Unsupervised Learning consists of 4 steps**



Dividing a large set of data into smaller segments, each segment only contain data that are similar

Extracting information (correlation, cause-and-effect relation, etc.) and put it in a understandable form

Reducing the format of the extracted information to a compact size (without redundancy, repetition, etc.)

Organizing the compact information so that it is reusable in recognizing similar situation when new data is encountered

Data Analytics with Unsupervised Machine Learning

- **aviation example: to use Unsupervised Machine Learning to extract causal and situational data from various safety event reports (NTSB, ASRS, etc.)**

Data Analytics with Unsupervised Machine Learning

- aviation example: to use Unsupervised Machine Learning to extract causal and situational data from various safety event reports (NTSB, ASRS, etc.)
  - **through the use of Large Language Model, all safety event reports are unified in the same keywords that can be analyzed**

Data Analytics with Unsupervised Machine Learning

- aviation example: to use Unsupervised Machine Learning to extract causal and situational data from various safety event reports (NTSB, ASRS, etc.)
  - through the use of Large Language Model, all safety event reports are unified in the same keywords that can be analyzed
  - **exploratory analysis is performed to identify relevant factors**

Data Analytics with Unsupervised Machine Learning

- aviation example: to use Unsupervised Machine Learning to extract causal and situational data from various safety event reports (NTSB, ASRS, etc.)
  - through the use of Large Language Model, all safety event reports are unified in the same keywords that can be analyzed
  - exploratory analysis is performed to identify relevant factors
  - **data mining is used find different groups of safety events based on their similarity**

Data Analytics with Unsupervised Machine Learning

- aviation example: to use Unsupervised Machine Learning to extract causal and situational data from various safety event reports (NTSB, ASRS, etc.)
  - through the use of Large Language Model, all safety event reports are unified in the same keywords that can be analyzed
  - exploratory analysis is performed to identify relevant factors
  - data mining is used find different groups of safety events based on their similarity
  - **data analytics is used to extract correlation and cause-and-effect relation**

Data Analytics with Unsupervised Machine Learning

- aviation example: to use Unsupervised Machine Learning to extract causal and situational data from various safety event reports (NTSB, ASRS, etc.)
  - through the use of Large Language Model, all safety event reports are unified in the same keywords that can be analyzed
  - exploratory analysis is performed to identify relevant factors
  - data mining is used find different groups of safety events based on their similarity
  - data analytics is used to extract correlation and cause-and-effect relation
  - **fault tree is built to represent various scenarios leading to each type of safety events**

Data Analytics with Unsupervised Machine Learning

- aviation example: to use Unsupervised Machine Learning to extract causal and situational data from various safety event reports (NTSB, ASRS, etc.)
  - through the use of Large Language Model, all safety event reports are unified in the same keywords that can be analyzed
  - exploratory analysis is performed to identify relevant factors
  - data mining is used find different groups of safety events based on their similarity
  - data analytics is used to extract correlation and cause-and-effect relation
  - fault tree is built to represent various scenarios leading to each type of safety events
  - **fault tree is used to provide early warning (precursors to safety events)**

Developing AI Applications Through Supervised Machine Learning

- **Modeling is the process of using data to train a model to do a task**

Developing AI Applications Through Supervised Machine Learning

- **For Artificial Neural Net (ANN) models used in Supervised Machine Learning, there are two different categories**

Developing AI Applications Through Supervised Machine Learning

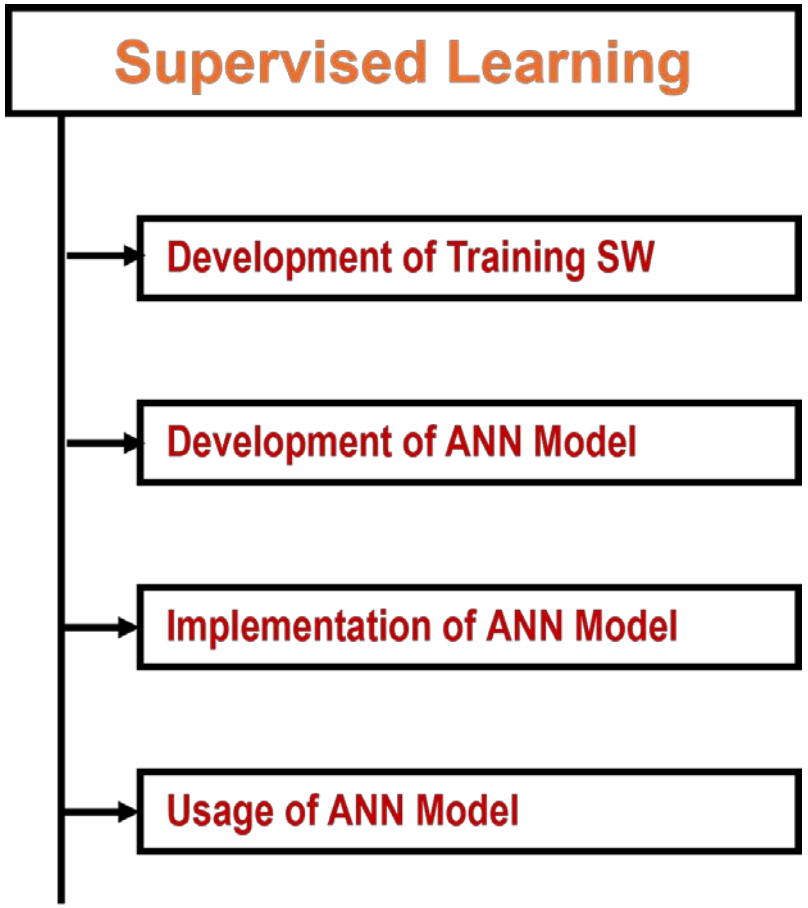
- For Artificial Neural Net (ANN) models used in Supervised Machine Learning, there are two different categories
  - **ANN with discrete output (zero or one) commonly used in classification problem (learning to identify if a new data that was never seen before belongs to one of the classes whose characteristics were extracted during training)**

Developing AI Applications Through Supervised Machine Learning

- For Artificial Neural Net (ANN) models used in Supervised Machine Learning, there are two different categories
  - ANN with discrete output (zero or one) commonly used in classification problem (learning to identify if a new data that was never seen before belongs to one of the classes whose characteristics were extracted during training)
  - **ANN with continuous output commonly used in regression problem (learning to mimic a system or process in order to predict its behavior)**

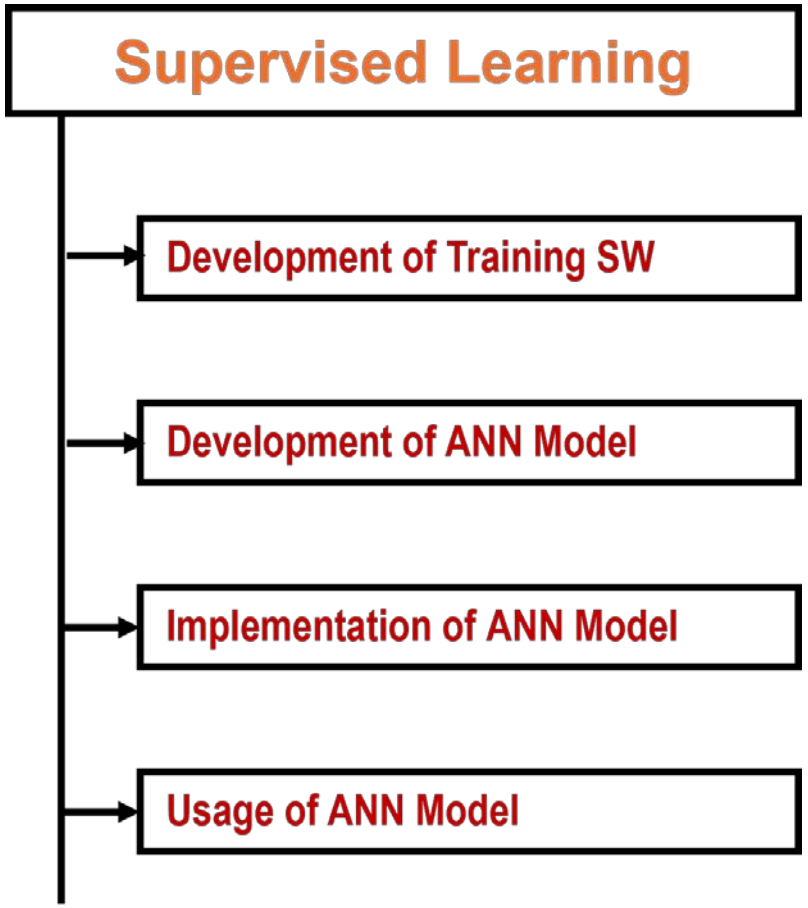
Developing AI Applications Through Supervised Machine Learning

- **Supervised Learning consists of 4 groups of who is doing what**



Developing AI Applications Through Supervised Machine Learning

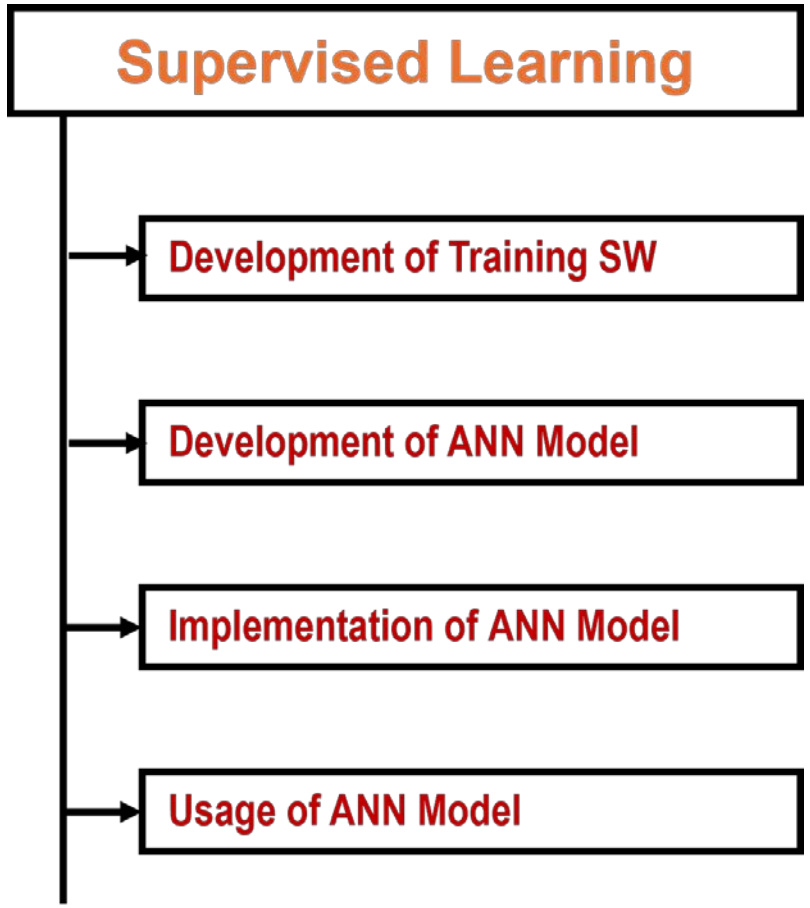
- **Supervised Learning consists of 4 groups of who is doing what**



Developers of Training SW are responsible for Learning Assurance

Developing AI Applications Through Supervised Machine Learning

- **Supervised Learning consists of 4 groups of who is doing what**

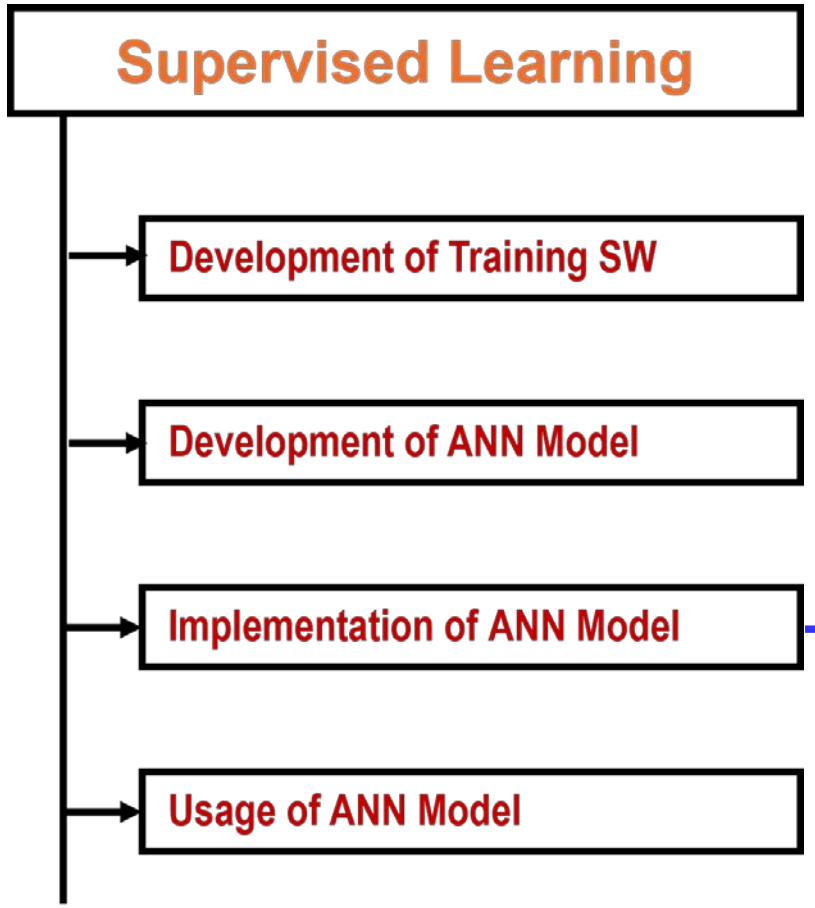


Developers of Training SW are responsible for Learning Assurance

Developers of ANN Models are responsible for Model's functionality and behavior

Developing AI Applications Through Supervised Machine Learning

- **Supervised Learning consists of 4 groups of who is doing what**



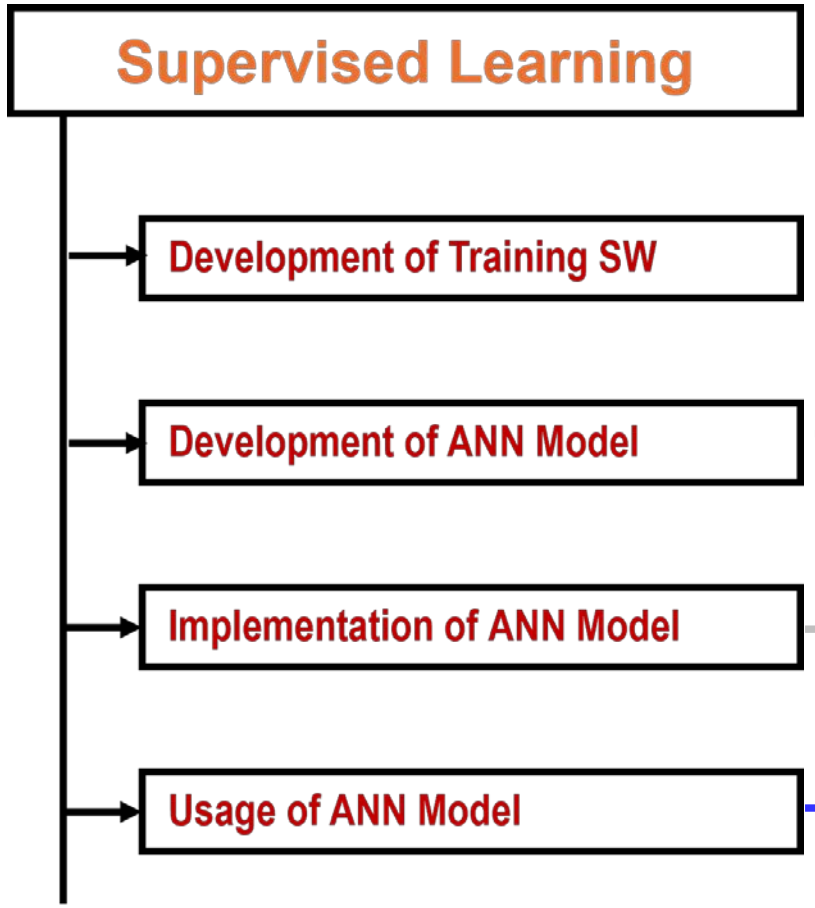
Developers of Training SW are responsible for Learning Assurance

Developers of ANN Models are responsible for Model's functionality and behavior

Implementers of ANN Models are responsible for the final Software that is put into an aircraft

Developing AI Applications Through Supervised Machine Learning

- **Supervised Learning consists of 4 groups of who is doing what**



Developers of Training SW are responsible for Learning Assurance

Developers of ANN Models are responsible for Model's functionality and behavior

Implementers of ANN Models are responsible for the final Software that is put into an aircraft

Applicants are responsible for the safety of an aircraft with AI systems in it

Developing AI Applications Through Supervised Machine Learning

- **aviation example: to use Supervised Machine Learning to develop AI systems**

Developing AI Applications Through Supervised Machine Learning

- aviation example: to use Supervised Machine Learning to develop AI systems
  - **recognize the location of a runway for guiding precision landing**

Developing AI Applications Through Supervised Machine Learning

- aviation example: to use Supervised Machine Learning to develop AI systems
  - recognize the location of a runway for guiding precision landing
  - **recommend an optimal cruising altitude**

Developing AI Applications Through Supervised Machine Learning

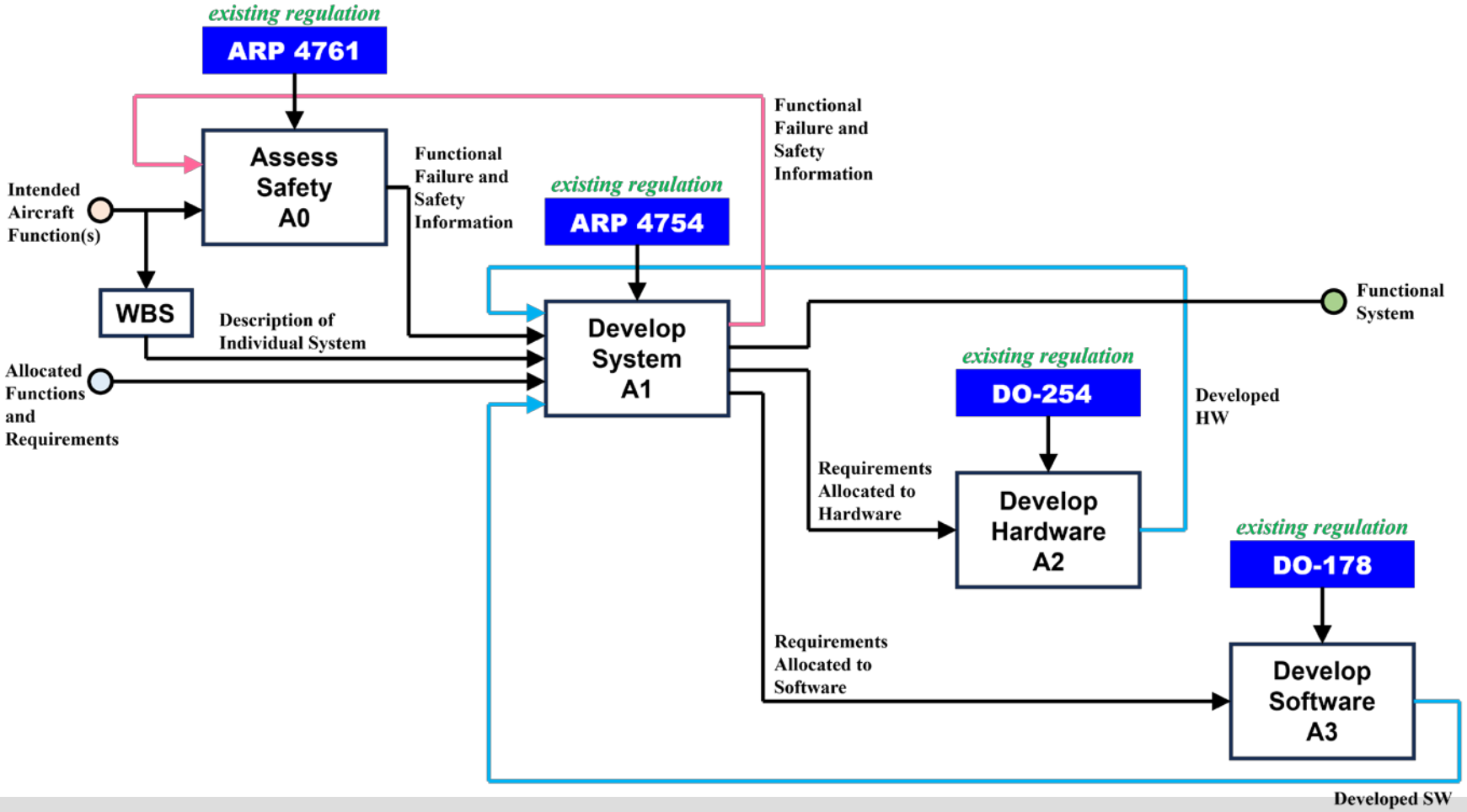
- aviation example: to use Supervised Machine Learning to develop AI systems
  - recognize the location of a runway for guiding precision landing
  - recommend an optimal cruising altitude
  - **generalize expert system into collision avoidance system**

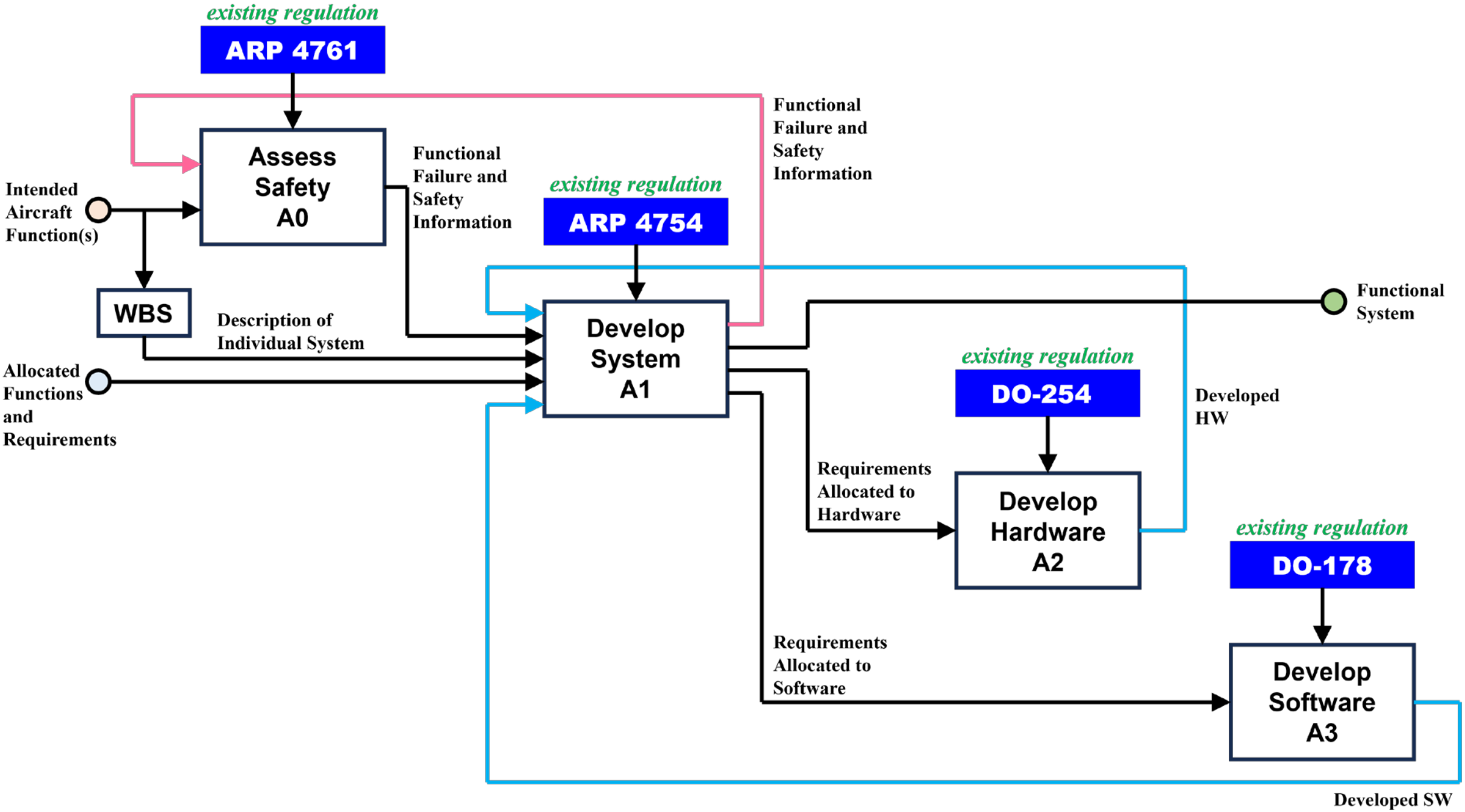
Developing AI Applications Through Supervised Machine Learning

- aviation example: to use Supervised Machine Learning to develop AI systems
  - recognize the location of a runway for guiding precision landing
  - recommend an optimal cruising altitude
  - generalize expert system into collision avoidance system
  - **monitor pilot's face to identify fatigue**

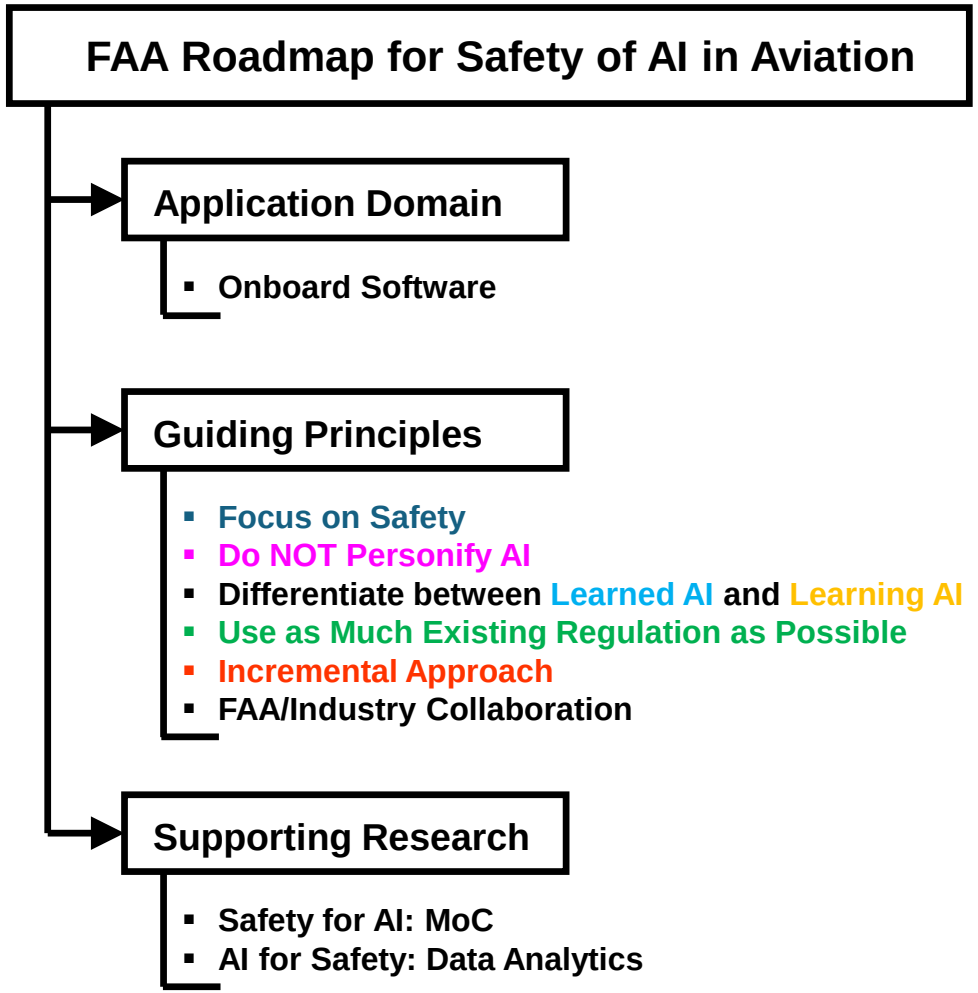
Developing AI Applications Through Supervised Machine Learning

- **Regulation side reviewing aviation examples**

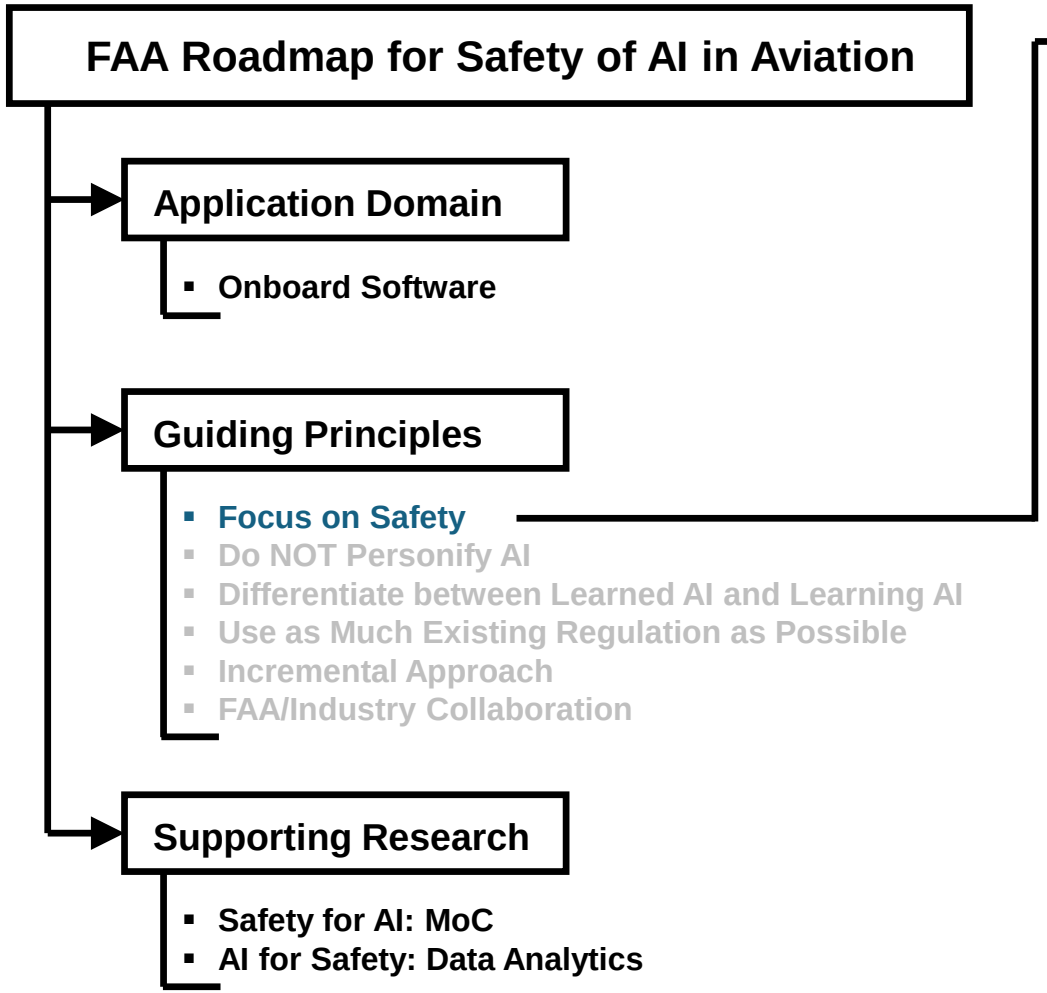



existing regulation
ARP 4761
Assess Safety A0
Intended Aircraft Function(s)
WBS
Description of Individual System
Allocated Functions and Requirements
Functional Failure and Safety Information
existing regulation
ARP 4754
Functional Failure and Safety Information
Develop System A1
Functional System
existing regulation
DO-254
Requirements Allocated to Hardware
Develop Hardware A2
Developed HW
existing regulation
DO-178
Requirements Allocated to Software
Develop Software A3
Developed SW

FAA’s Roadmap for AI Safety

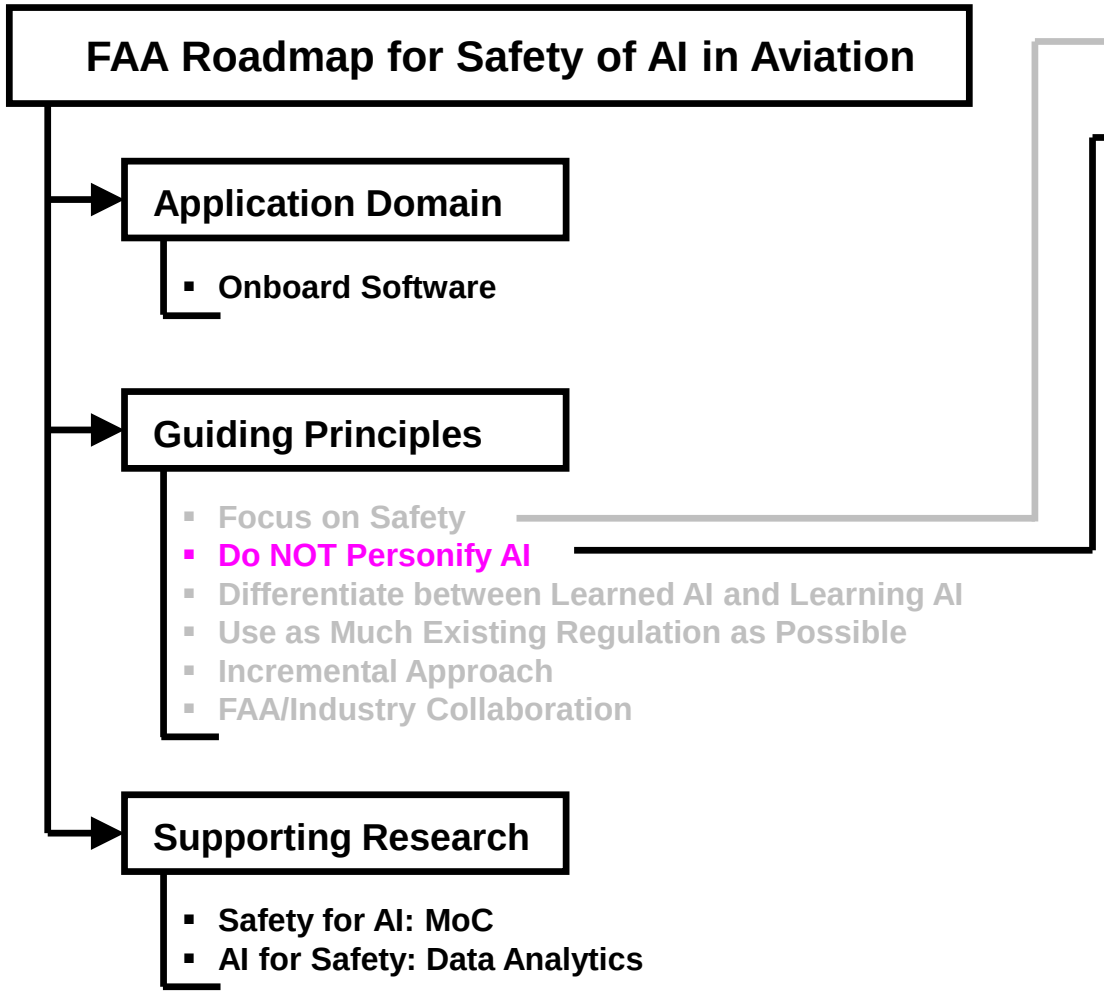


FAA’s Roadmap for AI Safety



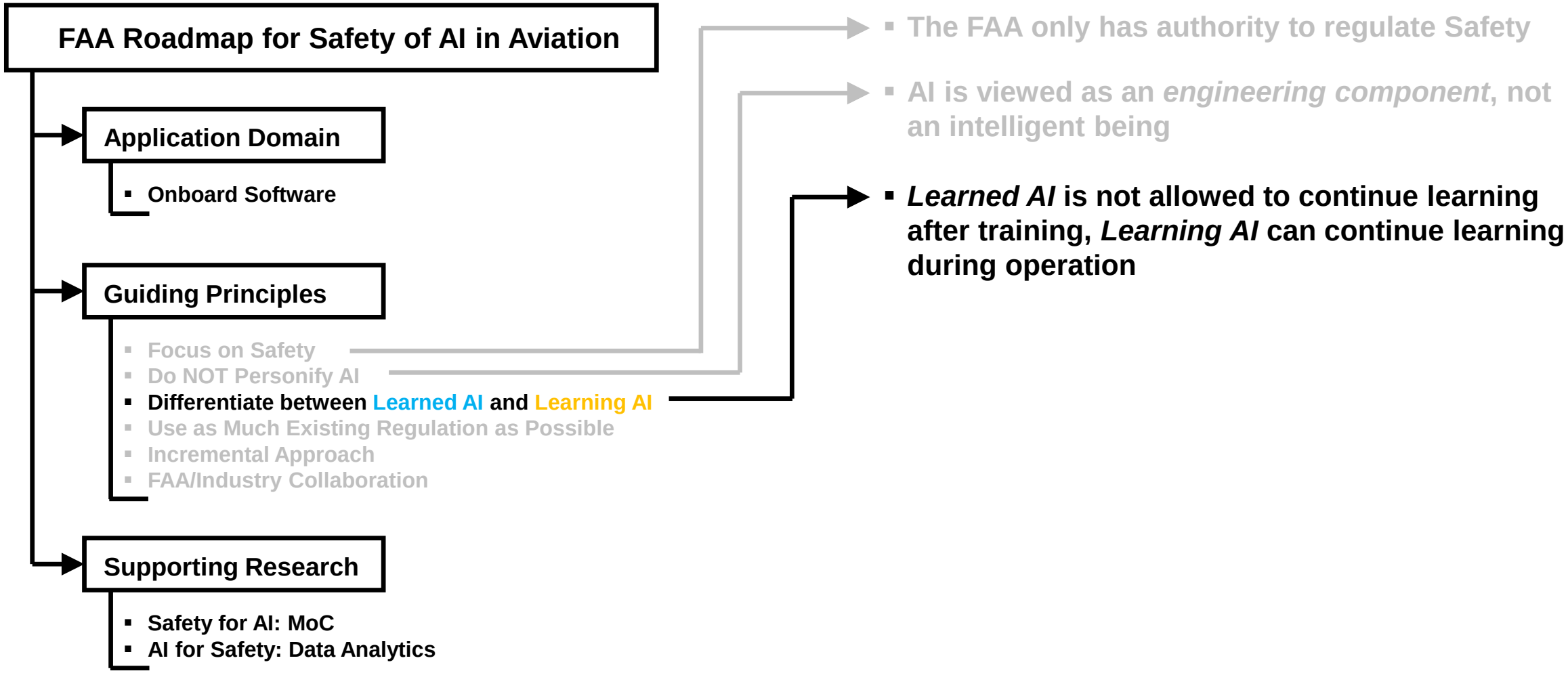
- The FAA only has authority to regulate Safety
- AI is viewed as an *engineering component*, not an intelligent being

FAA’s Roadmap for AI Safety

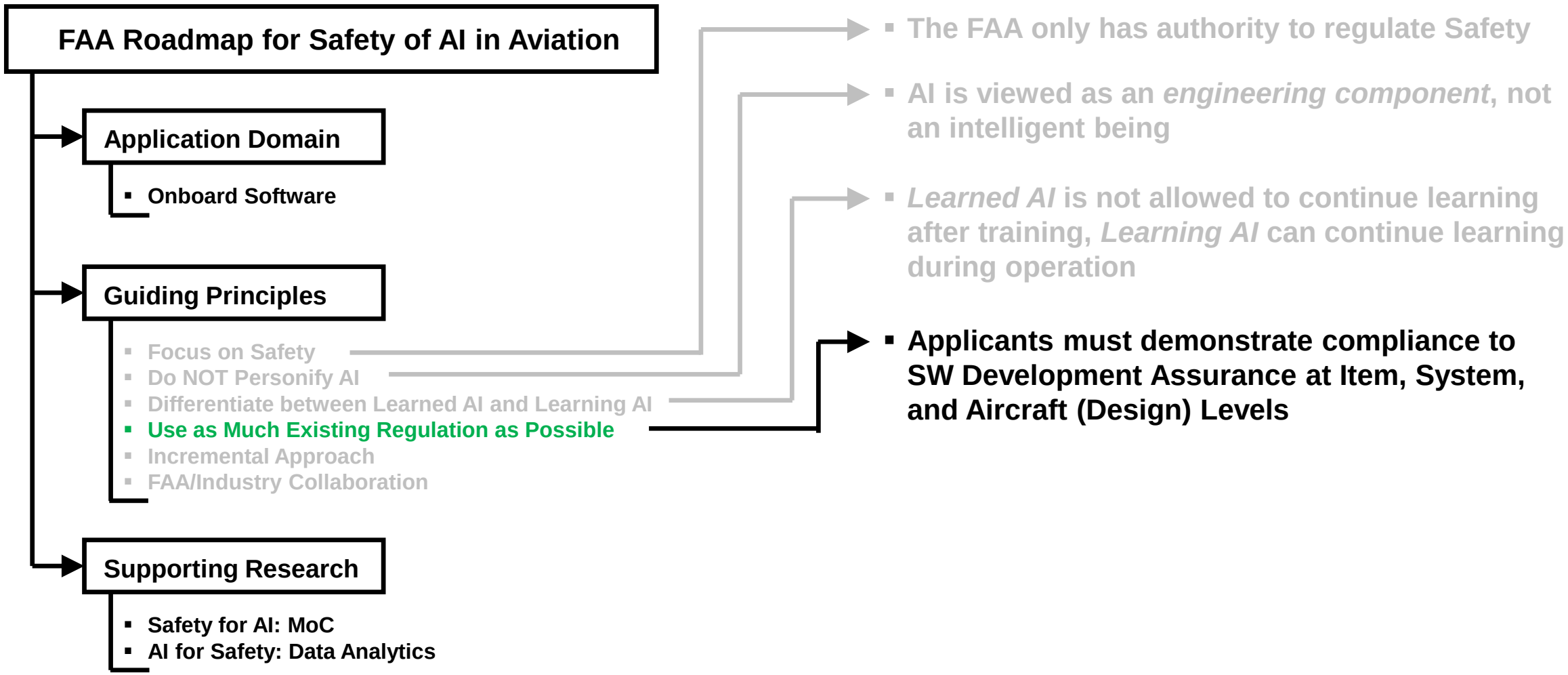


- The FAA only has authority to regulate Safety
- **AI is viewed as an *engineering component*, not an intelligent being**

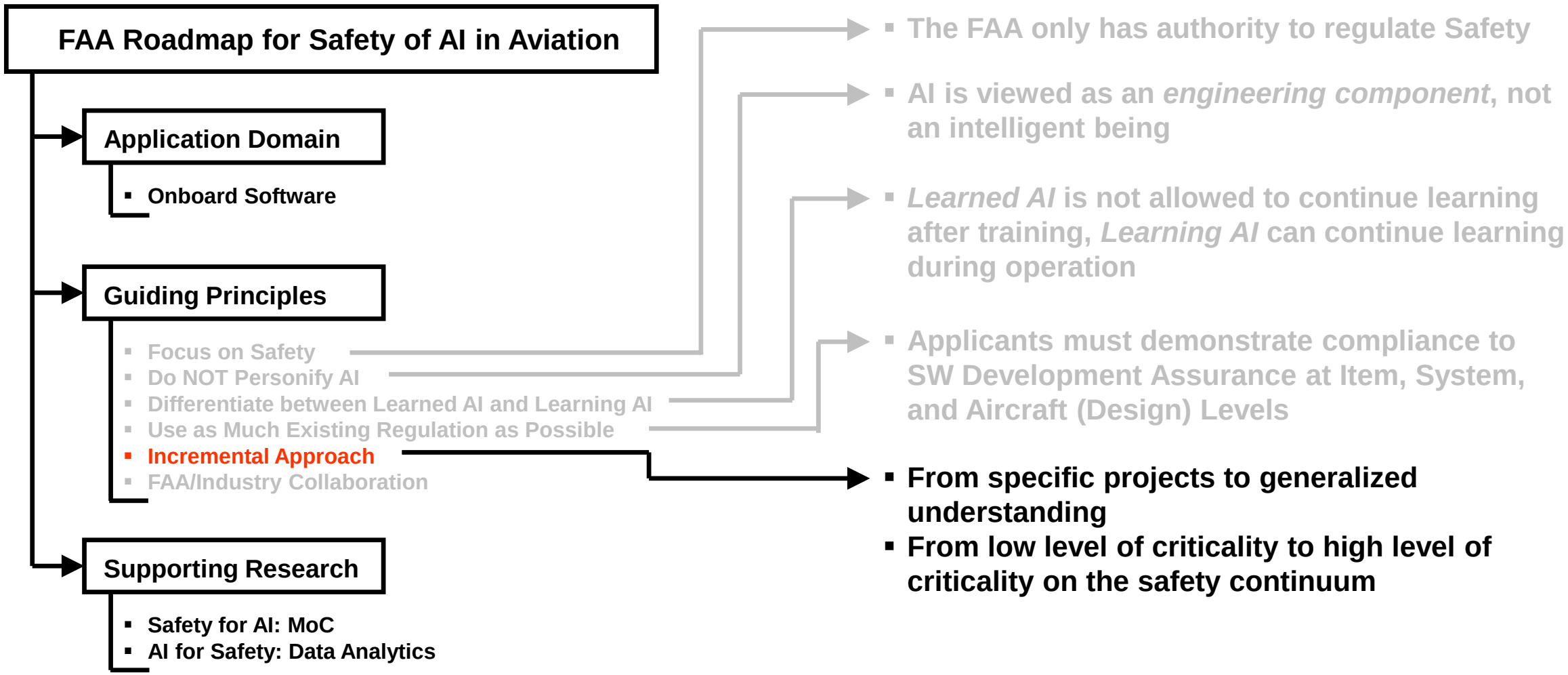
FAA's Roadmap for AI Safety



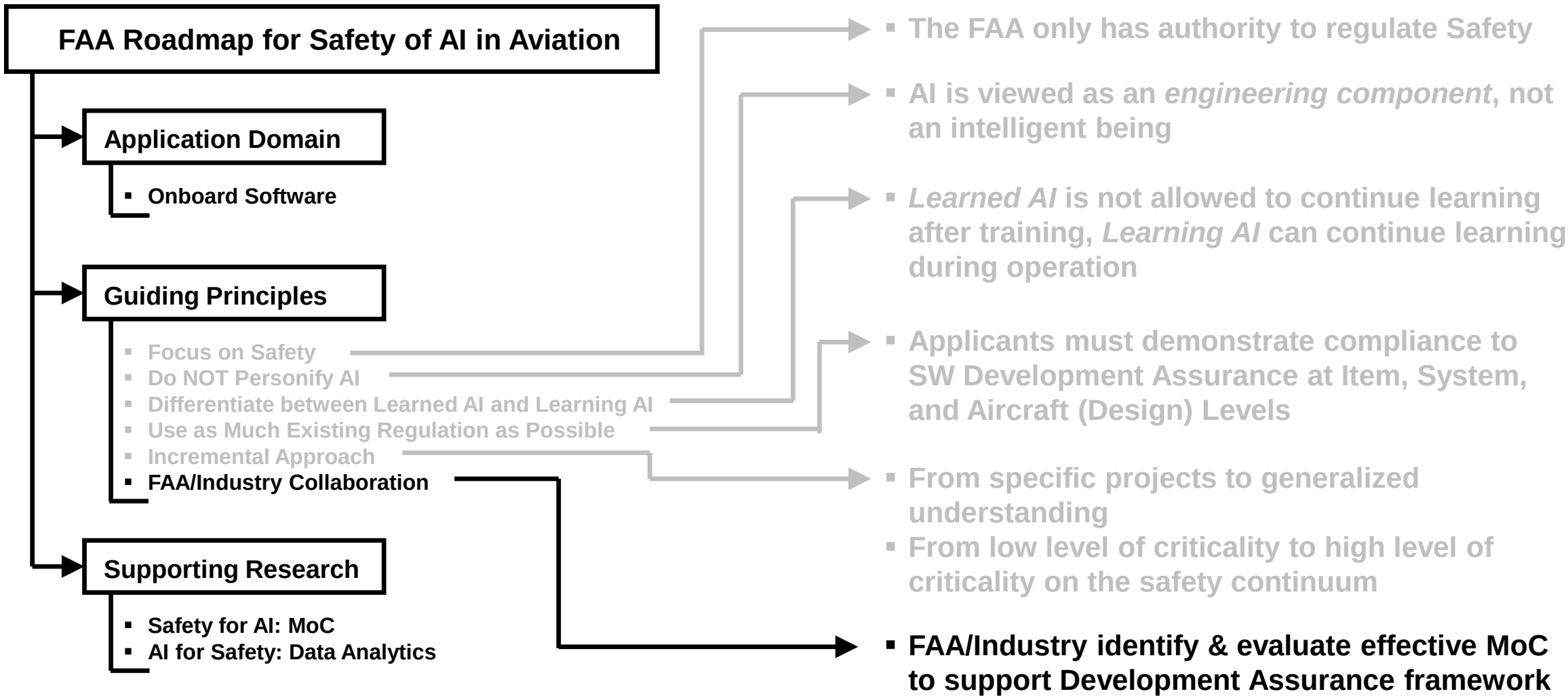
FAA's Roadmap for AI Safety



FAA's Roadmap for AI Safety



FAA's Roadmap for AI Safety

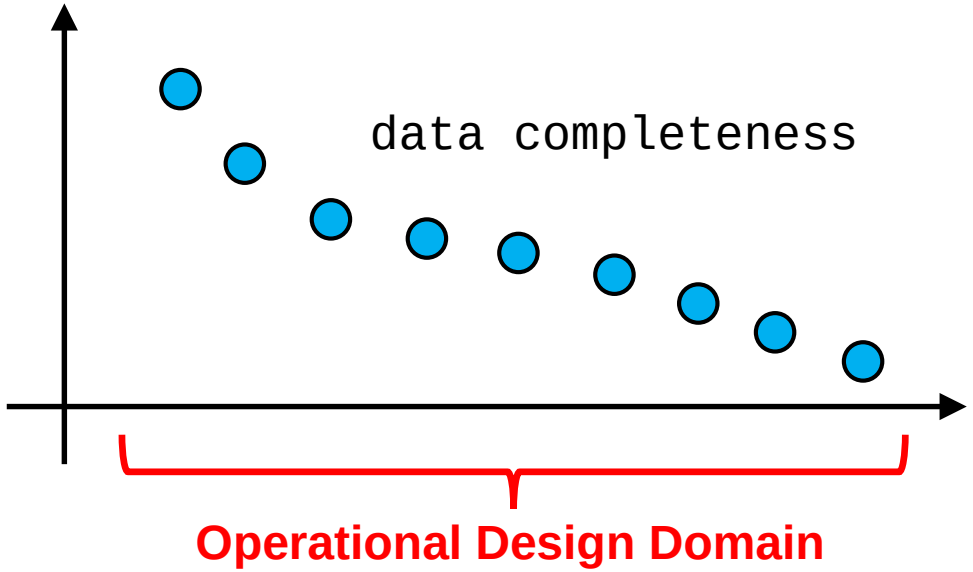


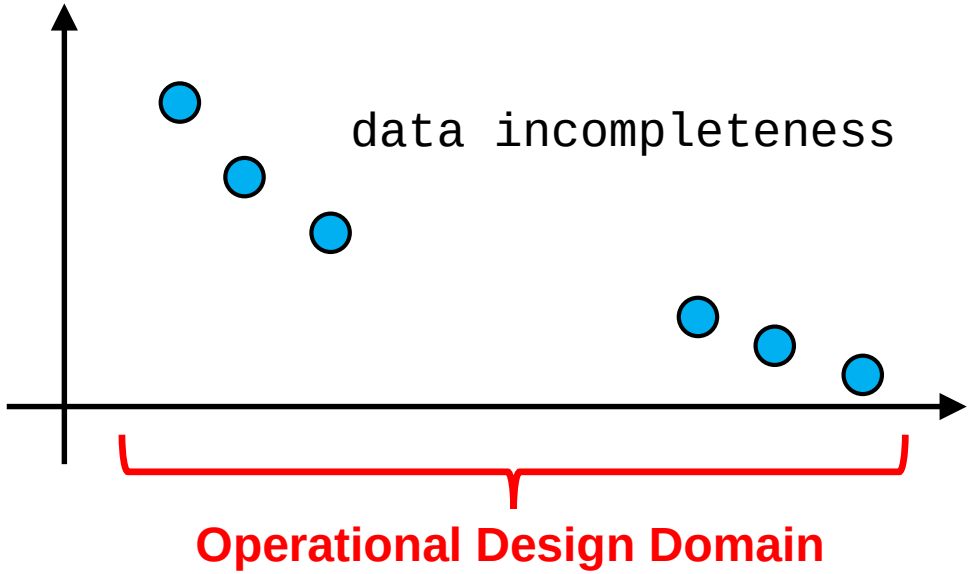
Data in the Context of Machine Learning

- **Data is the major factor driving the development work in Machine Learning, therefore it is important and needs special consideration**

Data in the Context of Machine Learning

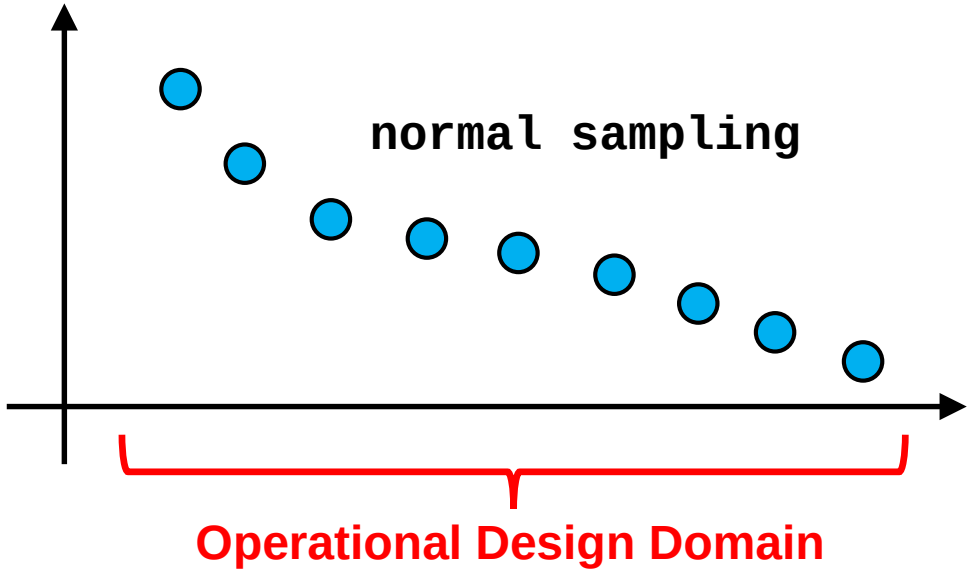
- Data is the major factor driving the development work in Machine Learning, therefore it is important and needs special consideration
  - **data quality assurance**
    - **data completeness in an operational design domain is important to assure the reliability of a model or of an analysis**

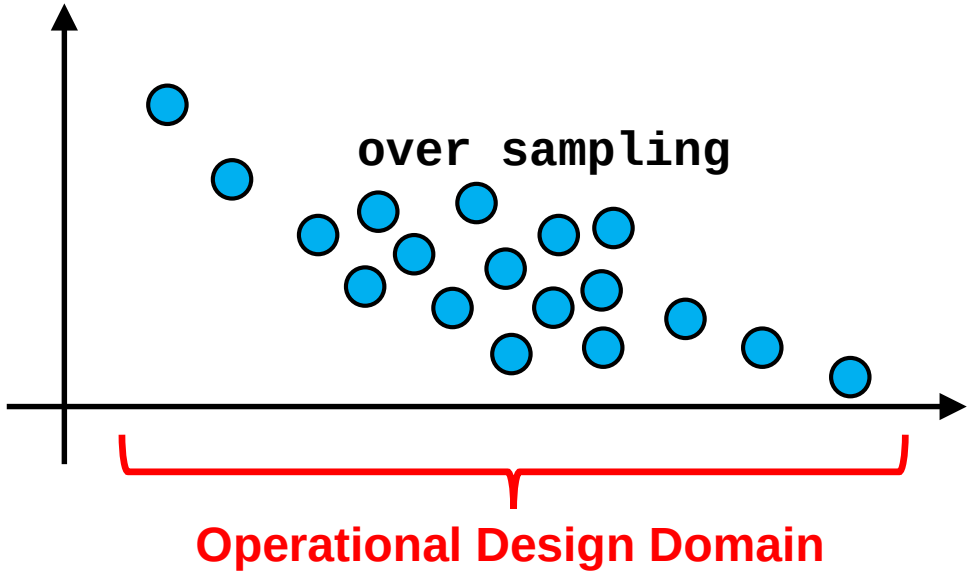




Data in the Context of Machine Learning

- Data is the major factor driving the development work in Machine Learning, therefore it is important and needs special consideration
  - **data redundancy and its implications**
    - **data distribution in an operational design domain will shift a model or add bias to an analysis**





Data in the Context of Machine Learning

- Data is the major factor driving the development work in Machine Learning, therefore it is important and needs special consideration
  - data quality assurance
    - data completeness in an operational design domain is important to assure the reliability of a model or of an analysis
  - data redundancy and its implications
    - data distribution in an operational design domain will shift a model or add bias to an analysis
  - **data quantification (on a uniform scale) and its effects**
    - **unifying data on the same scale will eliminate bias in the analysis or shift in a model**

Data in the Context of Machine Learning

- Data is the major factor driving the development work in Machine Learning, therefore it is important and needs special consideration
  - data quality assurance
    - data completeness in an operational design domain is important to assure the reliability of a model or of an analysis
  - data redundancy and its implications
    - data distribution in an operational design domain will shift a model or add bias to an analysis
  - data quantification (on a uniform scale) and its effects
    - unifying data on the same scale will eliminate bias in the analysis or shift in a model
  - **cybersecurity concerning for data**
    - **data integrity must be protected for future audit**

Discussion

- **Machine Learning, even though has been around for years, has been introduced into the AI landscape recently due to the advancement of computer technology**

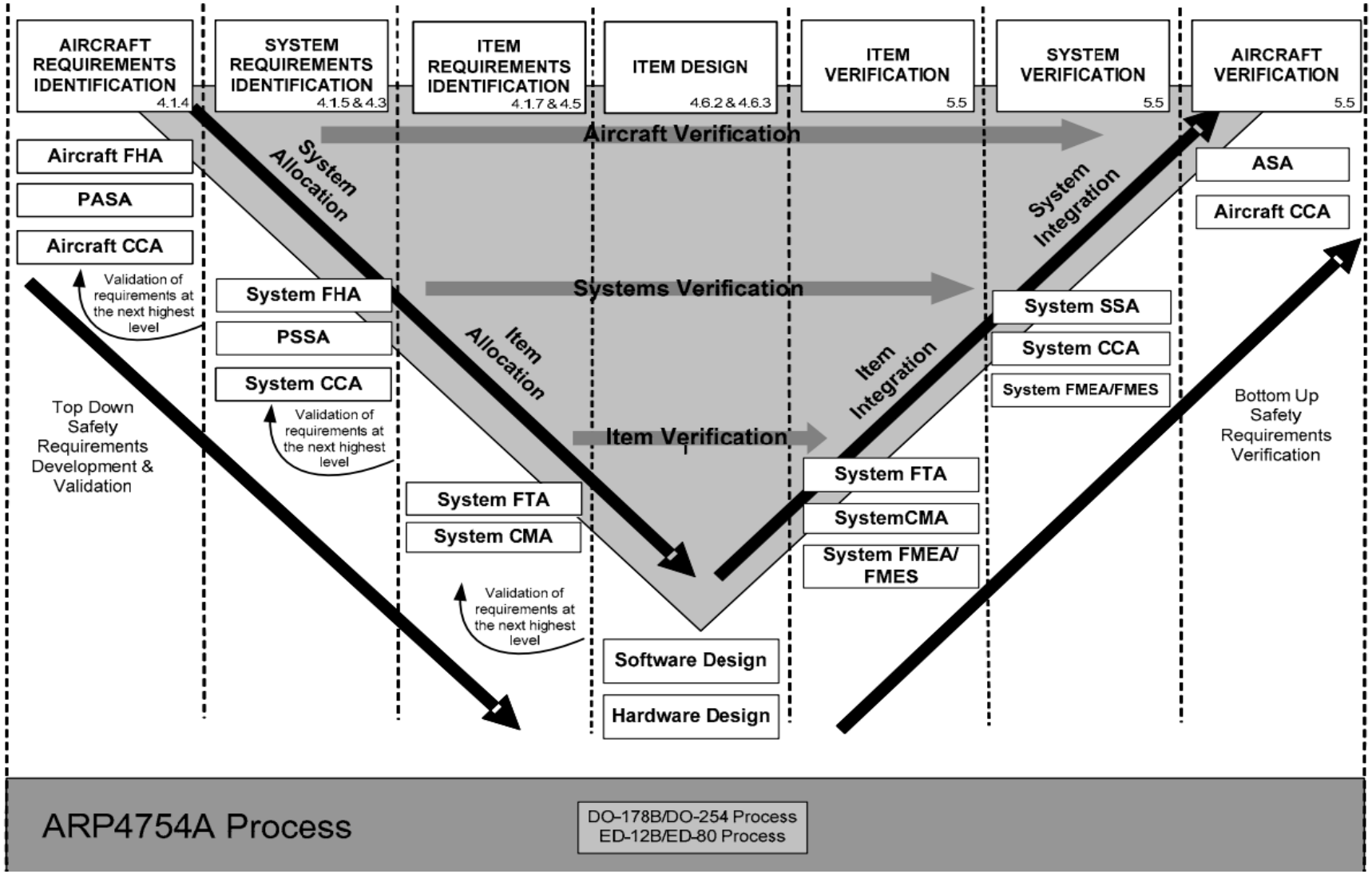
Discussion

- Machine Learning, even though has been around for years, has been introduced into the AI landscape recently due to the advancement of computer technology
  - **Machine Learning is often applied in various disciplines, solving various types of problems (classification, regression modeling, etc.)**

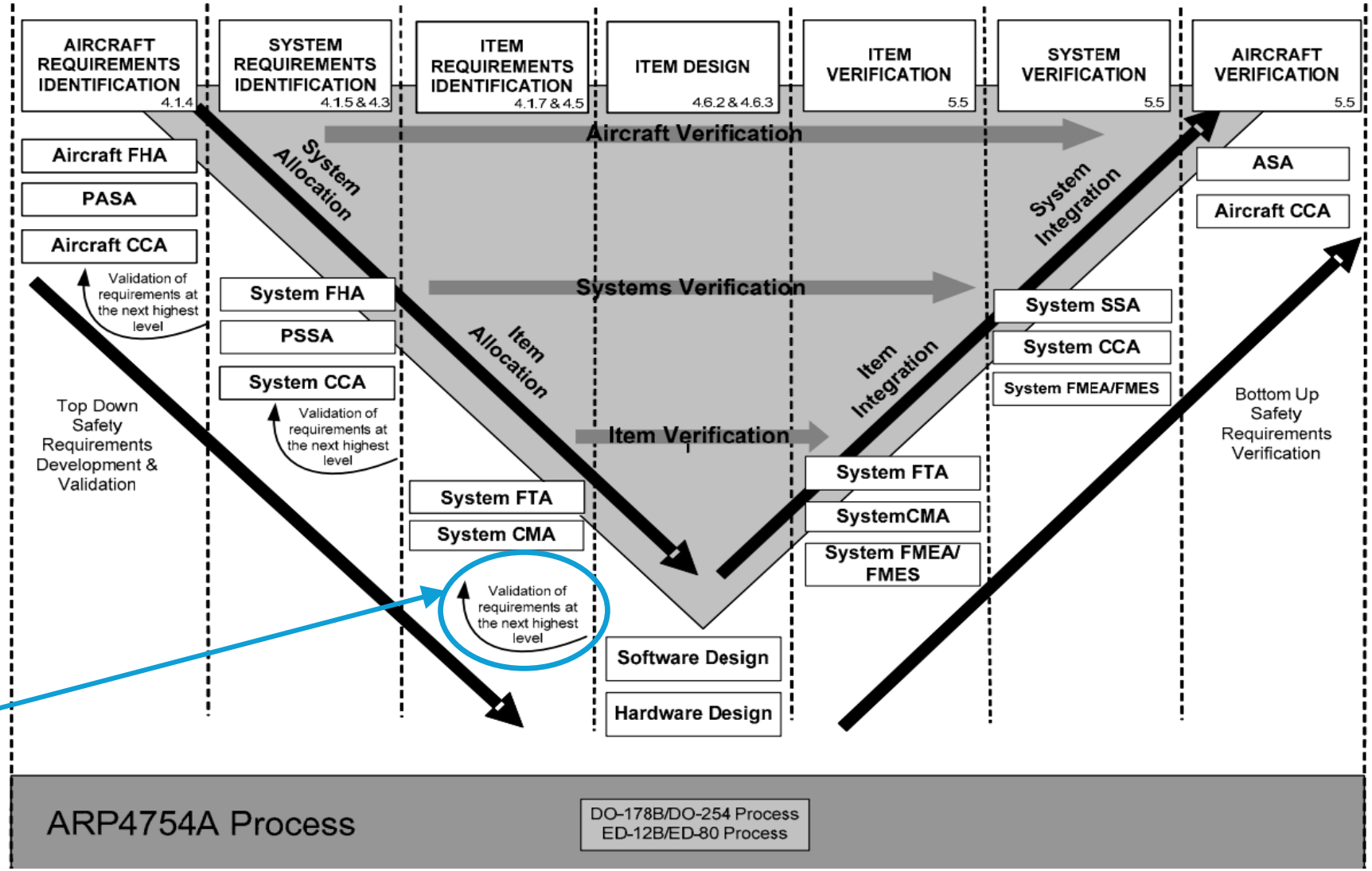
Discussion

- Machine Learning, even though has been around for years, has been introduced into the AI landscape recently due to the advancement of computer technology
  - Machine Learning is often applied in various disciplines, solving various types of problems (classification, regression modeling, etc.)
  - **A major issue that needs to be addressed is: the final form (an AI system or the analytical results) still need to be verified according to regulation and to comply with new consideration (safety, trustworthiness, ethics, etc.) according to how it is used**

Discussion

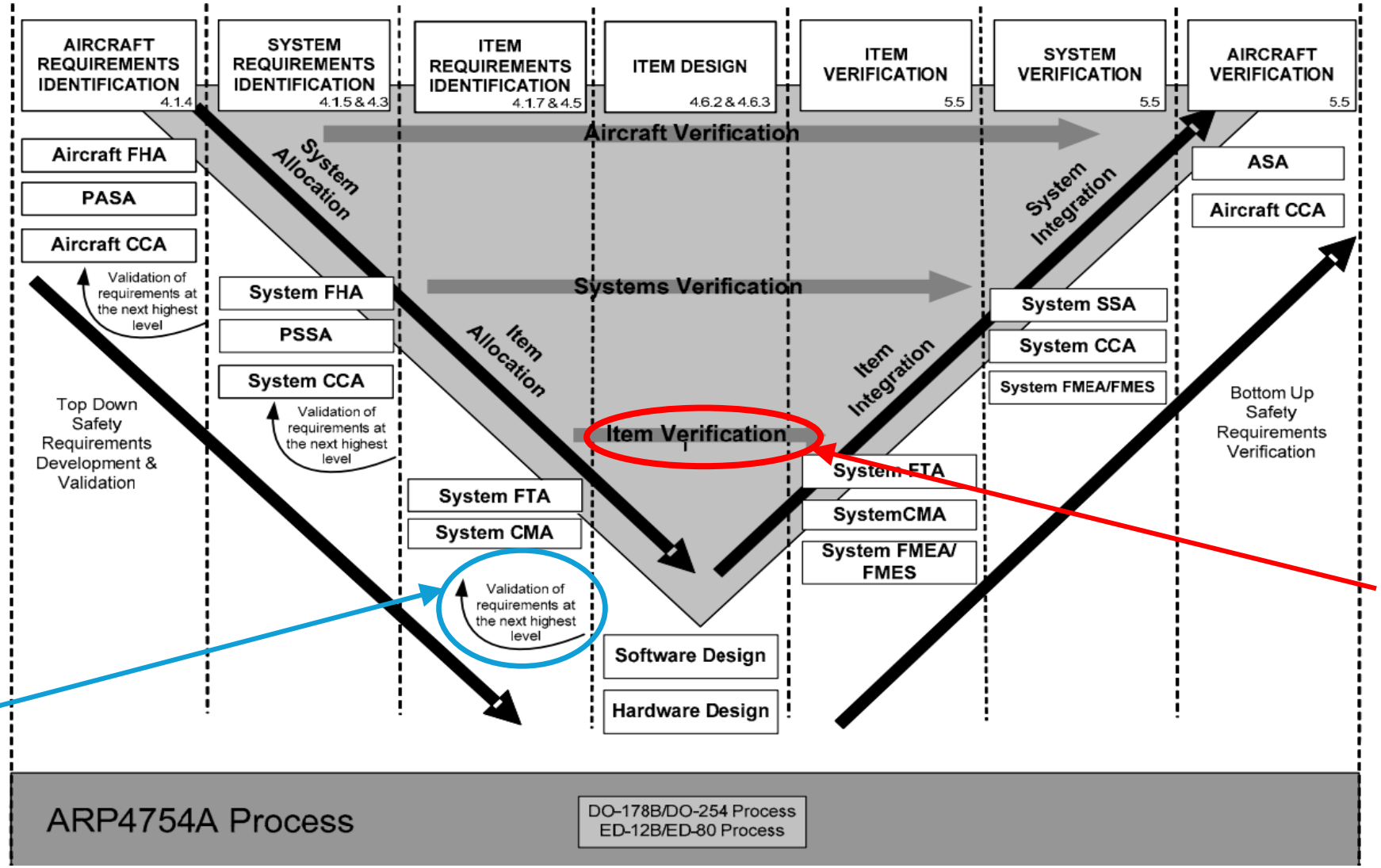

AIRCRAFT REQUIREMENTS IDENTIFICATION 4.1.4
SYSTEM REQUIREMENTS IDENTIFICATION 4.1.5 & 4.3
ITEM REQUIREMENTS IDENTIFICATION 4.1.7 & 4.5
ITEM DESIGN 4.6.2 & 4.6.3
ITEM VERIFICATION 5.5
SYSTEM VERIFICATION 5.5
AIRCRAFT VERIFICATION 5.5
Aircraft Verification
Aircraft FHA
PASA
Aircraft CCA
System Allocation
Validation of requirements at the next highest level
System FHA
PSSA
System CCA
Systems Verification
Item Allocation
Item Verification
Item Integration
System Integration
System SSA
System CCA
System FMEA/FMES
ASA
Aircraft CCA
Top Down Safety Requirements Development & Validation
Validation of requirements at the next highest level
System FTA
System CMA
Validation of requirements at the next highest level
System FTA
SystemCMA
System FMEA/ FMES
Bottom Up Safety Requirements Verification
Software Design
Hardware Design
ARP4754A Process
DO-178B/DO-254 Process
ED-12B/ED-80 Process

Discussion



Item requirements are allocated from system level, independent of the developer's choice of how to develop it

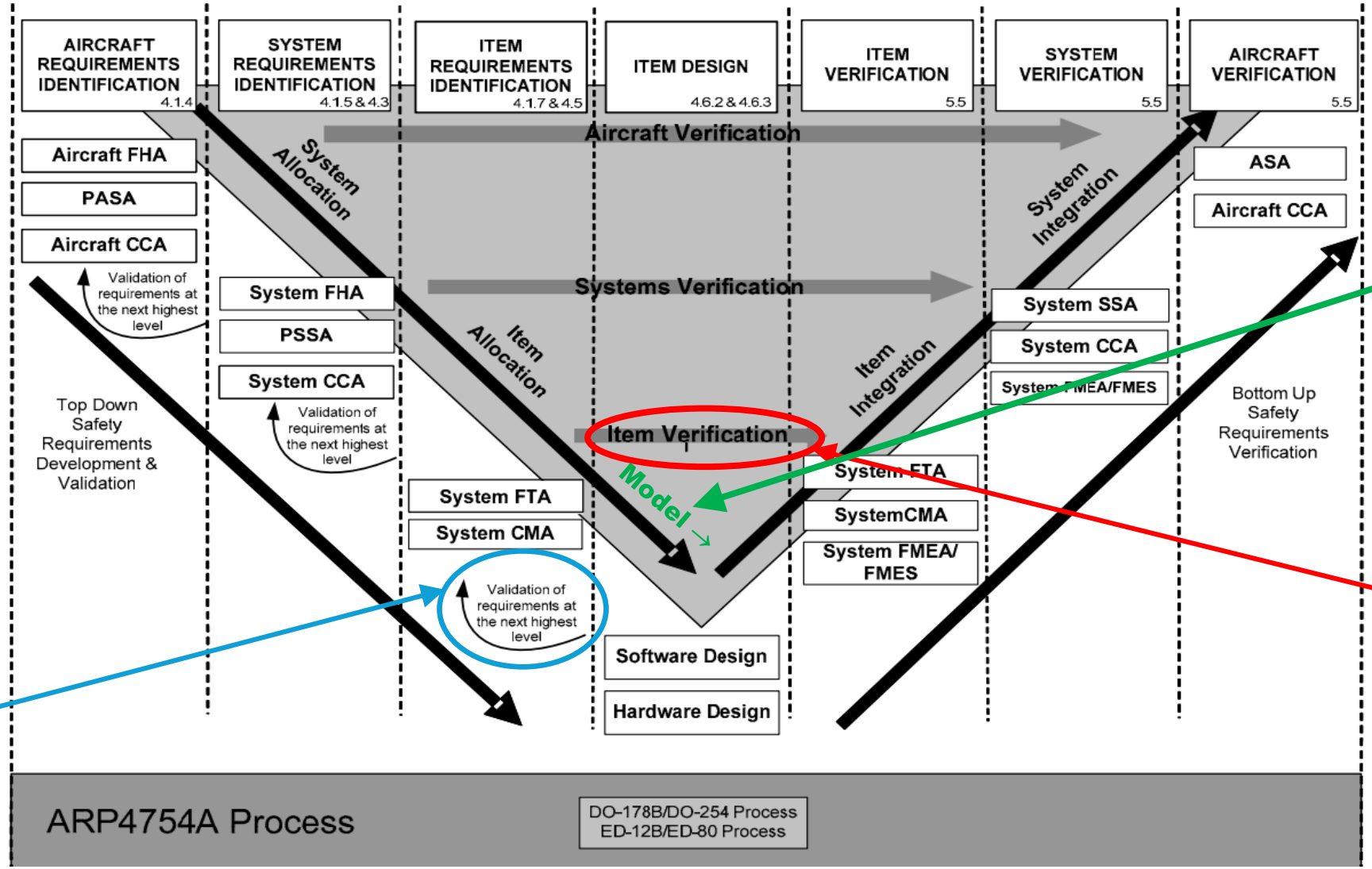
Discussion



Item requirements are allocated from system level, independent of the developer's choice of how to develop it

Verification of item's requirements should also be independent of the developer's choice of how to develop it

Discussion

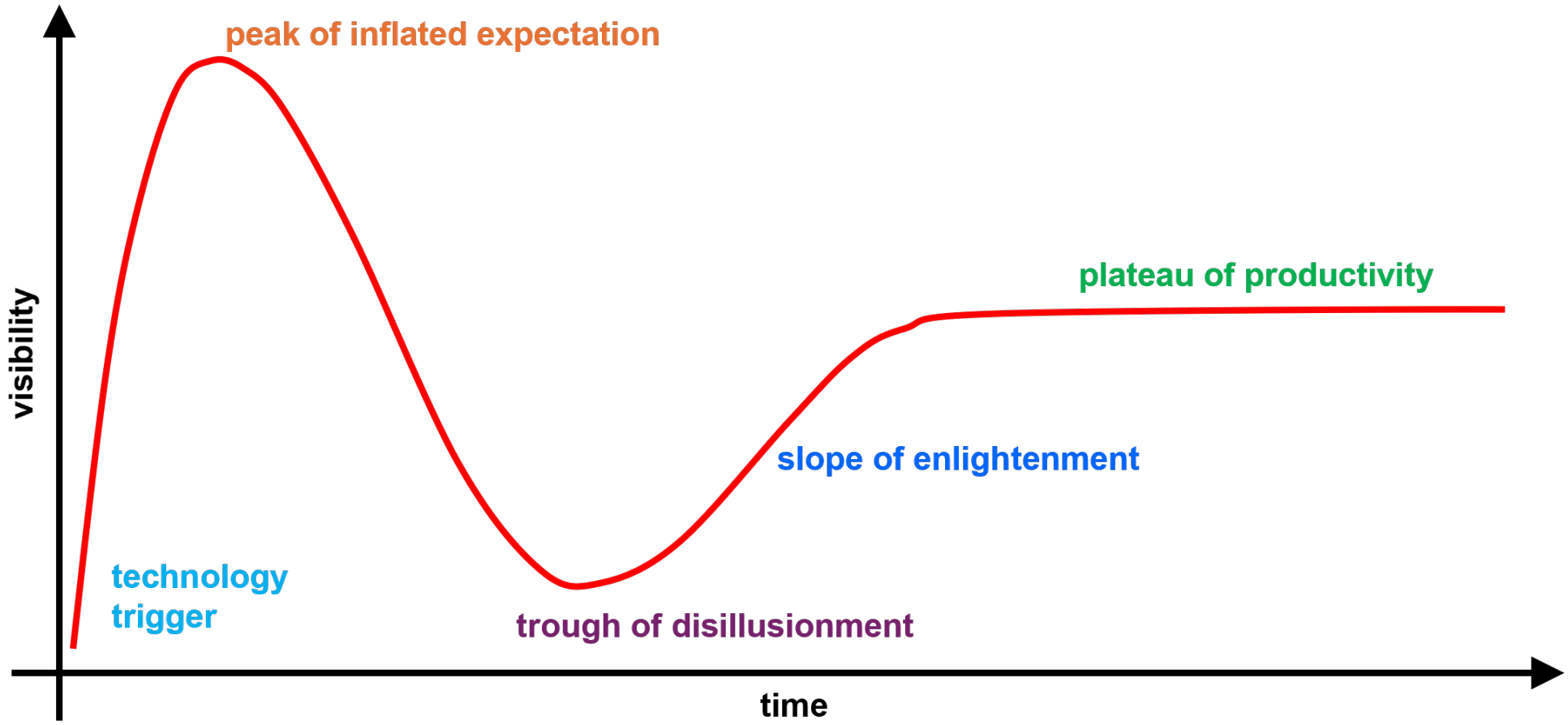


Item requirements are allocated from system level, independent of the developer's choice of how to develop it

Verification of model's specifications has been the FAA's effort to use as much existing regulation and guidance as possible

Verification of item's requirements should also be independent of the developer's choice of how to develop it

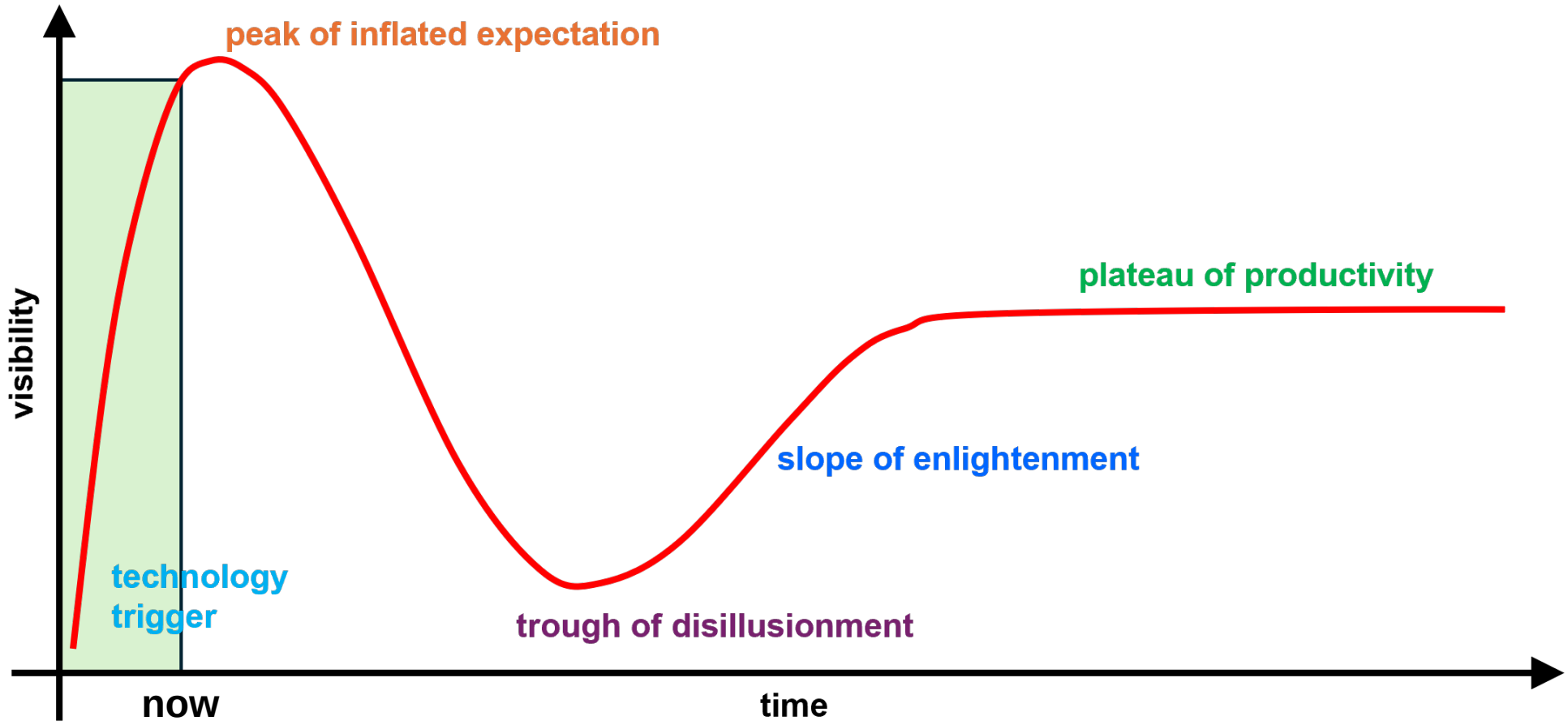
Discussion



*Source: https://en.wikipedia.org/wiki/Gartner_hype_cycle

Gartner Hype Cycle of a New Technology

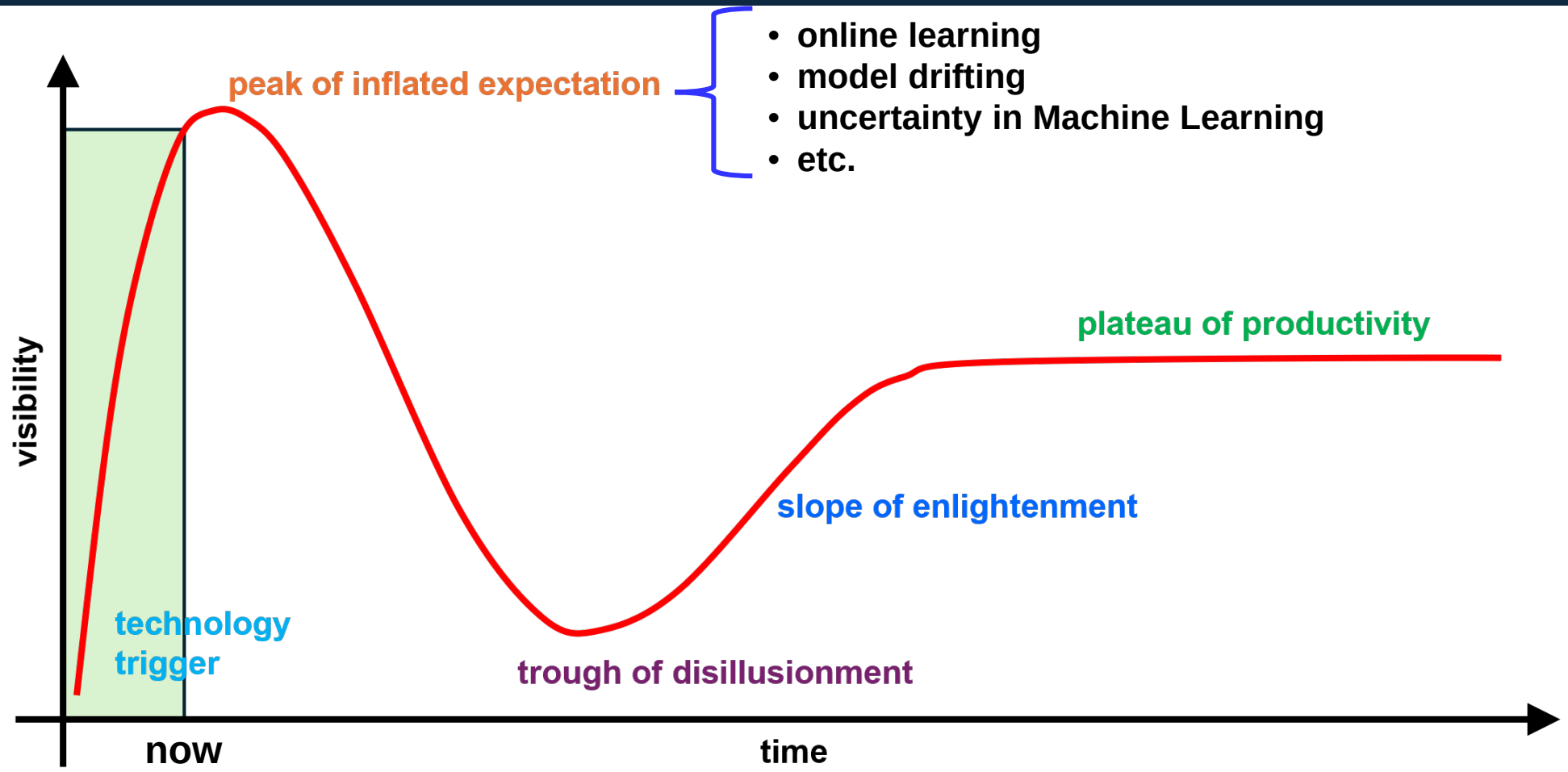
Discussion



*Source: https://en.wikipedia.org/wiki/Gartner_hype_cycle

Gartner Hype Cycle of a New Technology

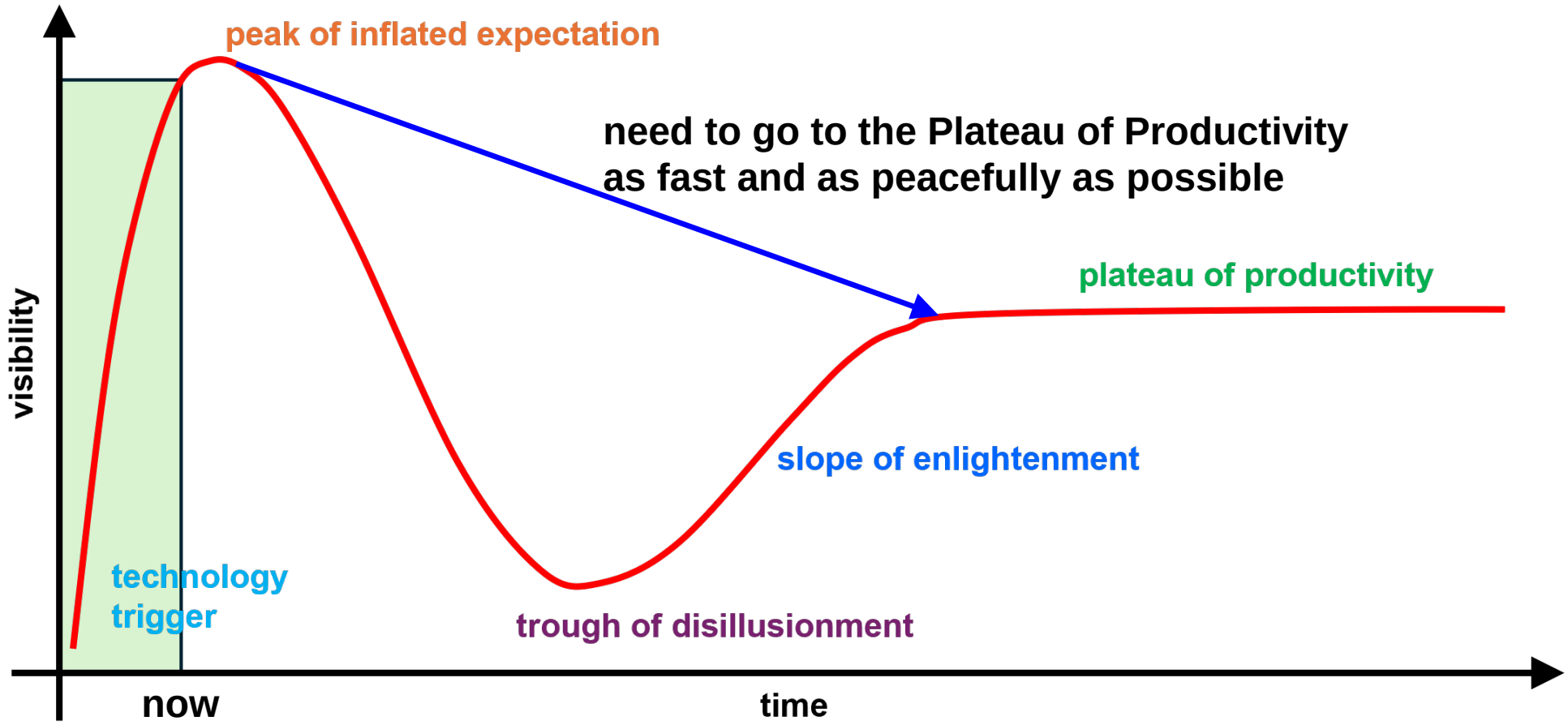
Discussion



*Source: https://en.wikipedia.org/wiki/Gartner_hype_cycle

Gartner Hype Cycle of a New Technology

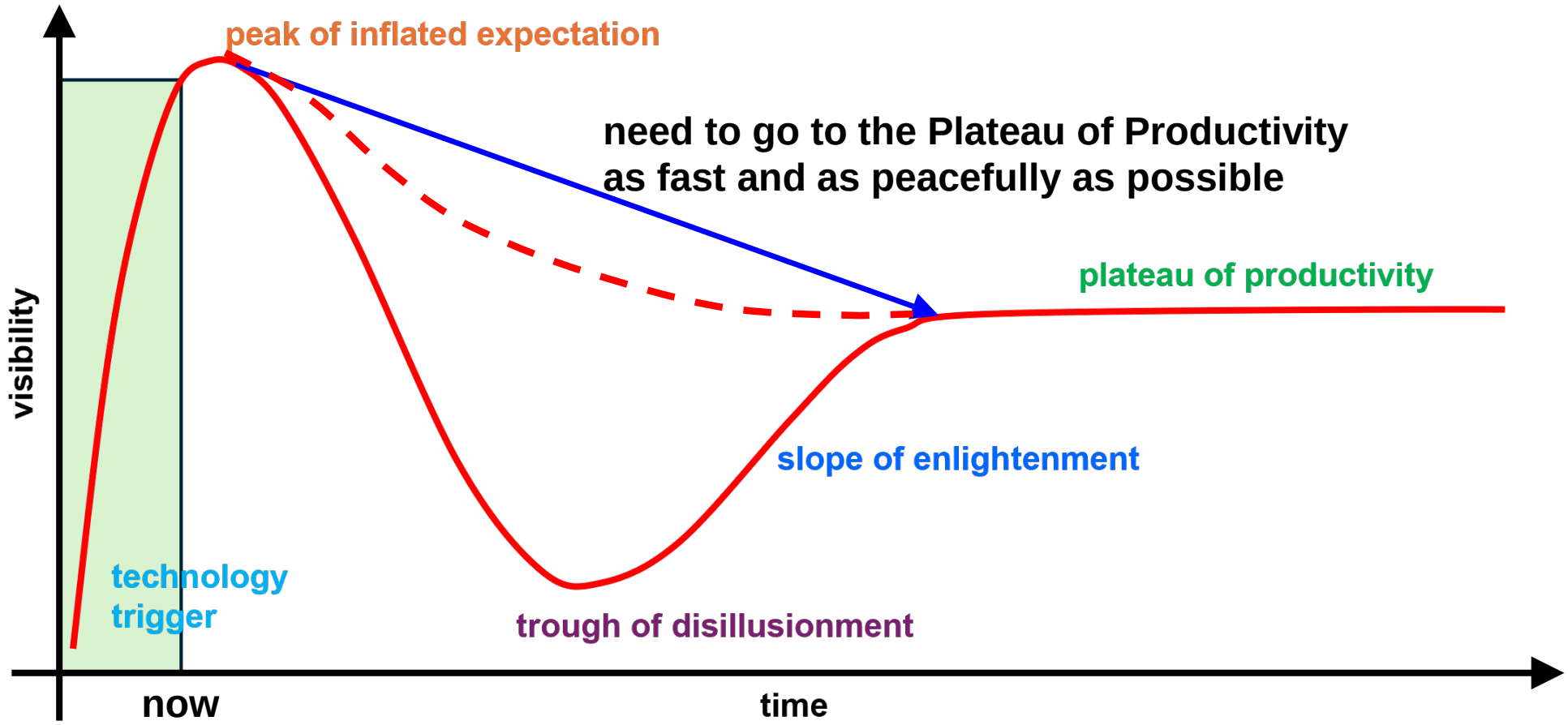
Discussion



*Source: https://en.wikipedia.org/wiki/Gartner_hype_cycle

Gartner Hype Cycle of a New Technology

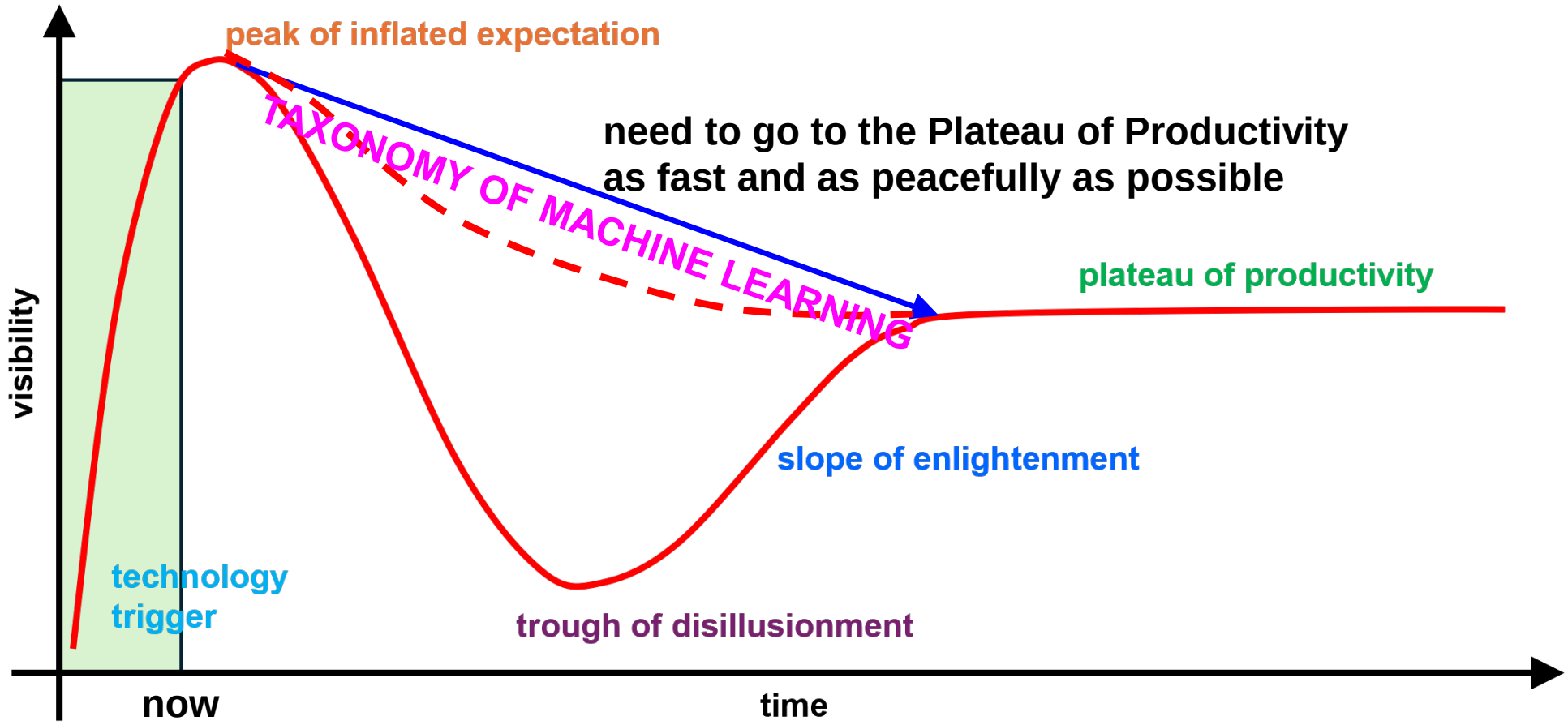
Discussion



*Source: https://en.wikipedia.org/wiki/Gartner_hype_cycle

Gartner Hype Cycle of a New Technology

Discussion



*Source: https://en.wikipedia.org/wiki/Gartner_hype_cycle

Gartner Hype Cycle of a New Technology

Thank you very much

Merci beaucoup

Muchas gracias

Cám ơn rất nhiều

非常感謝

Grazie mille

Muito obrigado

Moltes gràcies

आपका बहुत बहुत धन्यवाद

Taxonomy of Machine Learning, and its Role in Developing AI Systems

Trung T. Pham, Ph.D.

`trung.t.pham@faa.gov`

05/13/2025

Federal Aviation Administration