Australian Government
Department of Defence
Guided Weapons and
Explosive Ordnance Group

parari 2026 AUSTRALIAN EXPLOSIVE ORDNANCE SAFETY SYMPOSIUM

Proportional Approach to Weapon System Software Safety Engineering






Mr Warren Miller
Software Assurance
Desk Officer

Directorate of Engineering
P3M Services
GWEO

Defending Australia and its National Interests
www.defence.gov.au

Proportional Approach to Weapon System Software Safety Engineering

- AOP-52
  - What is AOP-52?
  - AOP-52 Major Update
  - AOP-52.x Standards Related Documents (SRDs)
- Artificial Intelligence / Machine Learning
  - AI vs Traditional Software
  - AI in Weapons Systems Context
  - Quick Primer in ML
  - Assurance Considerations

AOP-52

What is AOP-52?

- Allied Ordnance Publication (AOP)
- Published by the NATO Standardization Office (NSO)
- Released under NATO standardization agreement (STANAG) 4452 which was ratified by member nations and with partner nations invited to adopt
- Overarching requirement:
  - To achieve a reasonable level of assurance that munition related software and software-like devices will behave within the system context and operational environment with an acceptable level of risk.
- A risk based approach

AOP-52 Update

- AOP-52 has been updated from a guidance document to a standard.

- Aim.
  - The desired goal is mass adoption as requirements document.

- Desired outcomes.
  - Standardised processes and Objective Quality Evidence (OQE) artefacts.
  - Standardised product quality (robust, safe and secure software).
  - Eliminate/minimise recertification efforts following procurement between participating nations.

AOP-52 Update

- FROM
- Guidance on Software Safety Design and Assessment of Munition-related Computing Systems
- Edition B Version 1
- November 2016
- 208 pages

- TO
- Software Safety Requirements for Design and Assessment of Munition-related Computing Systems
- Edition C Version 1
- TBD 2026
- ~56 pages
- Guidance moved to Standards Related Documents (SRDs) that can be added/updated/removed with lower overhead than a standard

AOP-52 Update (contd)

- Retains:
  - Proportional risk-based approach. Where:
    - Hazard analysis reveals safety-significant software functions.
    - Software Criticality Index (SwCI) is a function of
      - Software Control Category (SCC) – the level of control authority or autonomy the software has over those functions, and
      - the credible accident severity.
    - SwCI is used to define required LoR tasks being commensurate with risk.
    - LoR tasks are additional and applied to a base software engineering process.

Software Control Category	Accident Severity Category			
	Catastrophic (1)	Critical (2)	Marginal (3)	Negligible (4)
1	SwCI 1	SwCI 1	SwCI 2	SwCI 3
2	SwCI 1	SwCI 2	SwCI 3	SwCI 4
3	SwCI 2	SwCI 3	SwCI 4	SwCI 4
4	SwCI 3	SwCI 4	SwCI 4	SwCI 4

Standards Related Documents (SRDs)

- Planned
  - AOP-52.1 Guidance on the use of Artificial Intelligence (AI) and Machine Learning (ML) in Munition-Related Computing Systems
  - AOP-52.2 Hardware considerations for Munition-Related Computing Systems
  - AOP-52.3 Guidance on Software Tool Qualification
  - AOP-52.4 Compliance Matrix for AOP-52

- AAS3P-01.2 Guidance on the S3 Assessment Testing of Loitering Munitions
  - Though not in the AOP-52.x suite, AAS3P-01.2 includes high level guidance for Loitering Munitions' software including the application of AI/ML.

AOP-52.1 Guidance on the use of Artificial Intelligence and Machine Learning in Munition-Related Computing Systems

- SRD AOP 52.1 tailored derivation from:
  - Machine Learning System Safety Engineering Guide with Implementation Process and Tasks Supporting MIL-STD-882E w/CHANGE 1 January 2026
  - MLSSEG provides the program system safety engineer with proposed system safety analysis processes, new level of rigor objectives, and associated technical information to identify hazards and manage the associated risks of DoW systems specifically containing supervised ML functionality. This guide also benefits project managers, design engineers, T&E engineers, specialty engineering areas, and system sustainment professionals in both government and industry by providing information to support the mitigating of system safety-related risks introduced during DoW system development, test, and use when supervised ML is present.

Artificial Intelligence (AI)

Artificial Intelligence (AI) and Machine Learning (ML)

- Most of us hear the term AI on a daily basis.
- Chatgpt. Copilot. Claude. Deepseek. Chatbots. Data centres. Surveillance systems. Self driving cars.
- We've been somewhat desensitised to the term.
- What does it really mean and what does it mean specifically in the context of weapon systems?

What is AI?

- Basic definitions:
  - Artificial Intelligence (AI) is a branch of computer science focused on building systems capable of performing tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, understanding language, and recognizing patterns.
  - Software/firmware/complex electronics functionality based on algorithm training of collected data, acquired human knowledge, or direct feedback that creates a model/mathematical function to perform a task e.g. assess or categorize an object and/or determine a numeric solution.
- Instead of relying on strict step-by-step instructions, AI systems can analyse massive amounts of data to identify patterns and learn from experience.

Common Branches of AI




```
Artificial Intelligence (AI)
│
├── Symbolic AI
│   ├── Expert Systems
│   └── Fuzzy Logic
│
├── Learning-Based AI
│   ├── Machine Learning
│   │   └── Neural Networks
│   │       └── Deep Learning
│   ├── Evolutionary Computation
│   └── Swarm Intelligence
│
├── Perception AI
│   ├── Natural Language Processing (NLP)
│   └── Computer Vision
│
├── Embodied AI
│   └── Robotics
│
└── Human-like Reasoning
    └── Cognitive Computing
```

Traditional vs AI/ML Software

- Traditional Software
  - Rule-based
  - Deterministic
  - Developed through explicit programming
  - Behaviour traceable directly to source code
- AI/ML Software
  - Data-driven
  - Potentially probabilistic
  - Patterns learned rather than explicitly programmed
  - Difficult to completely trace from source code alone (some operate as “black boxes”)
- AI improves the ability of systems to perceive, interpret, adapt, and make decisions in complex environments where traditional software relies on fixed rules and pre-programmed logic.

Inherent Limitations and Risks of AI

- Verification and Certification
  - Traditional software: Can often be fully tested.
  - AI: May behave unexpectedly in edge cases. Difficult to exhaustively validate.
- Explainability
  - Traditional systems: Clear decision logic.
  - AI systems: Some deep-learning models operate as “black boxes”.
- Adversarial Vulnerability
  - AI may be fooled by: Carefully crafted inputs, sensor spoofing and data poisoning
- Ethical and Legal Concerns
  - Military organizations generally seek meaningful human control over engagement decisions because accountability remains a human responsibility; and International humanitarian law requires distinction and proportionality assessments.

Hybrid Architecture Optimal

- Traditional Software
  - Safety-critical functions
  - Flight control
  - Fuzing
  - Weapons release logic
- AI Components
  - Perception
  - Classification
  - Threat assessment
  - Decision support
  - Navigation assistance
- AI is likely to act as an advanced sensing and decision-support layer while conventional software continues to govern critical safety and control functions. This combines the adaptability of AI with the predictability and certifiability of traditional engineered systems.

Artificial Intelligence (AI) Applications – Munitions Related

- Application examples extend across:
  - guided missile systems
  - loitering munitions systems
  - unexploded ordnance (UXO) detection systems
  - intelligence, surveillance and reconnaissance (ISR) operations
  - reliability surveillance
  - ammunition stockpile management

Key Areas Where AI Can Improve Weapon Systems

- Target Recognition and Classification
- Multi-Sensor Data Fusion
- Guidance, Navigation and Control
- Countermeasure Resistance
- Predictive Mission Planning
- Autonomous Search and Surveillance
- Maintenance and Reliability
- Sensor and Image Processing
- Training and Simulation

Target Recognition and Classification

- Traditional software:
  - Uses fixed signatures
  - May struggle with camouflage, weather, or unusual viewing angles

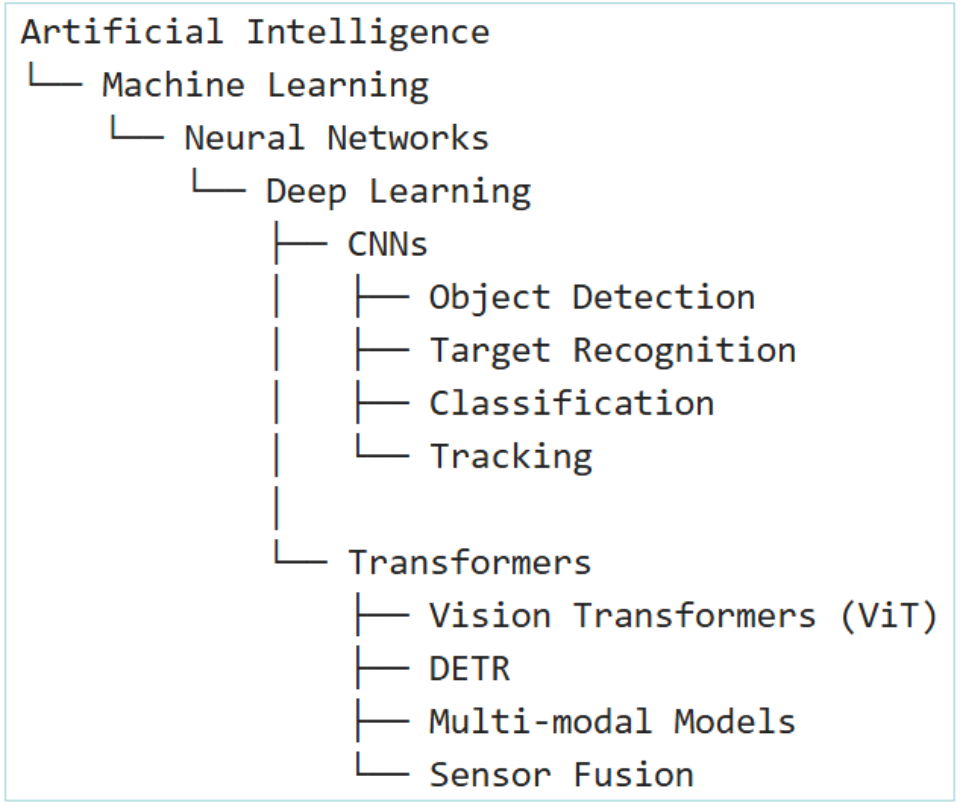
- AI:
  - Identifies vehicles, missile launchers, aircraft, ships, drones, and other objects across varying conditions
  - Learns subtle features across thousands of examples

Target Recognition and Classification (contd)

- Examples:
  - Autonomous loitering munitions
  - Missile seekers
  - Counter-UAS systems
  - ISR platforms
- Benefits:
  - Faster detection
  - Reduced false alarms
  - Improved target identification accuracy
  - Reduced cognitive load

Target Recognition and Classification AI ML Models

- Munitions embedded AI models that can rapidly and accurately identify a target's unique signature using sensor data in real time:
  - Visual (electro optical)
  - Acoustic (combination of noises and vibrations)
  - Magnetic (magnetic anomaly detection – localised disturbance an object makes in the Earth's magnetic field)
  - Radar (reflected RF)
  - Infrared (heat)
  - RF (electromagnetic emissions)



Target Recognition and Classification AI ML Models

- Munitions embedded AI models that can rapidly and accurately identify a target's unique signature using sensor data in real time:
  - Visual (electro optical)
  - Acoustic (combination of noises and vibrations)
  - Magnetic (magnetic anomaly detection – localised disturbance an object makes in the Earth's magnetic field)
  - Radar (reflected RF)
  - Infrared (heat)
  - RF (electromagnetic emissions)



Machine Learning

Supervised

Machine Learning Algorithm vs Model

ML Algorithm	ML Model
The method or procedure used to learn from data.	The result produced after the algorithm has been trained on data - a trained output that can make perform inference (make predictions).
A set of mathematical rules and training processes.	A learned representation containing parameters, weights, and patterns.
Exists before and used during training.	Exists after training and used during deployment.
Example: Convolutional Neural Network (CNN).	Trained CNN model.

Training Process:

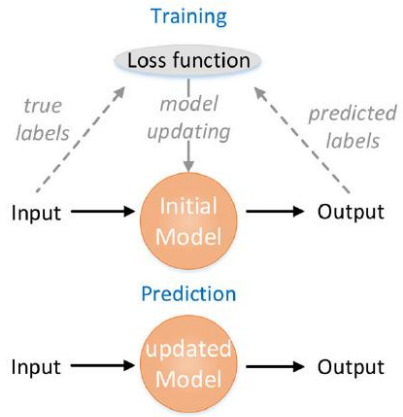
Data + Algorithm + Learning → Model

Inference Process:

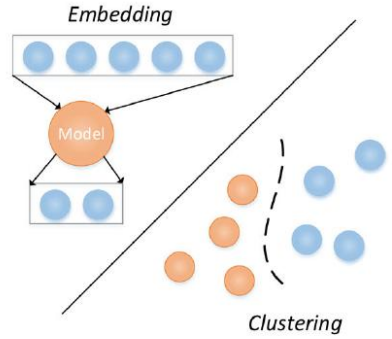
Model + New Data → Prediction

Machine Learning (ML) Types

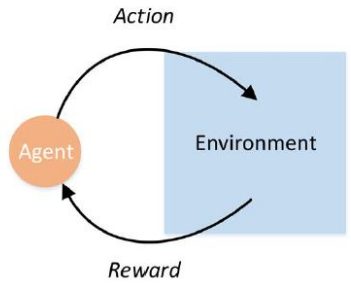
Supervised learning



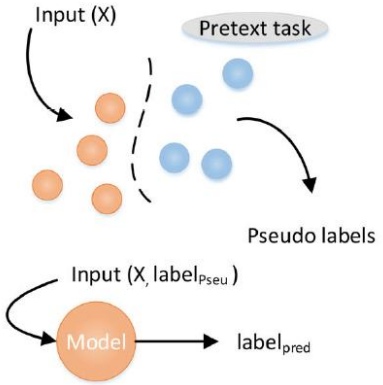
Unsupervised learning



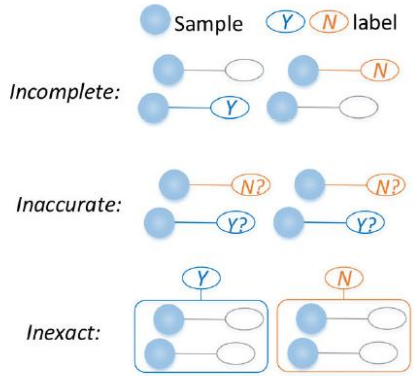
Reinforcement learning



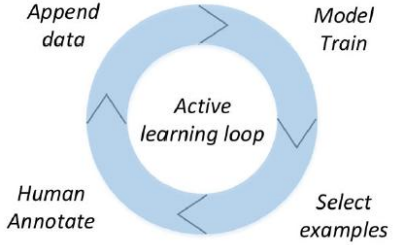
Self-supervised learning



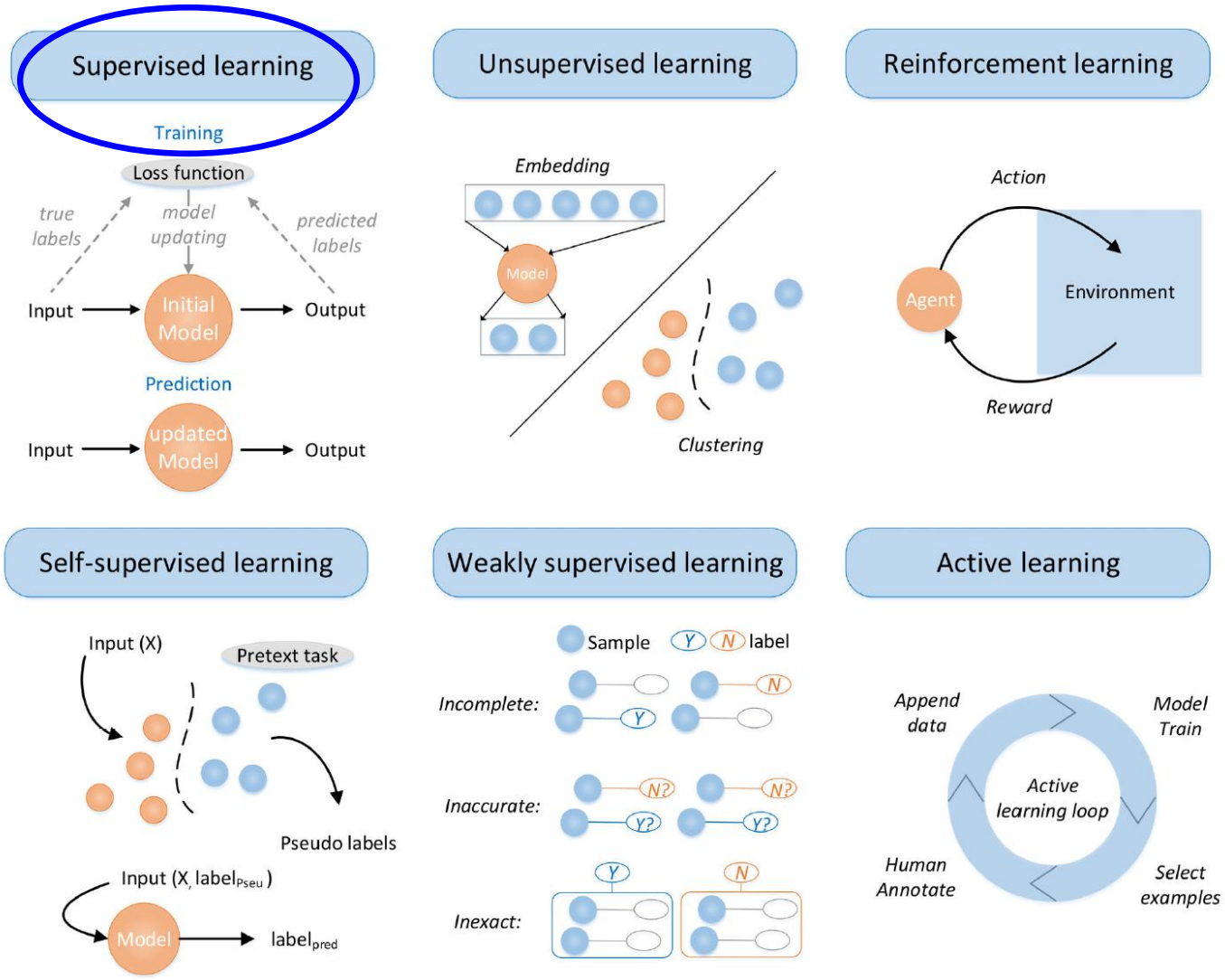
Weakly supervised learning



Active learning



Machine Learning (ML) Types



ML Supervised Learning Types

- Regression
  - Used when the output is a continuous numerical value (e.g., forecasting house prices or predicting stock trends). Common algorithms include Linear Regression and Gradient Boosting.

- Classification
  - Used when the output is a category (e.g. classifying a tumor as malignant or benign, or detecting email spam, or identifying the presence of a particular object in an image).

ML Supervised Learning Types

- Regression
  - Used when the output is a continuous numerical value (e.g., forecasting house prices or predicting stock trends). Common algorithms include Linear Regression and Gradient Boosting.

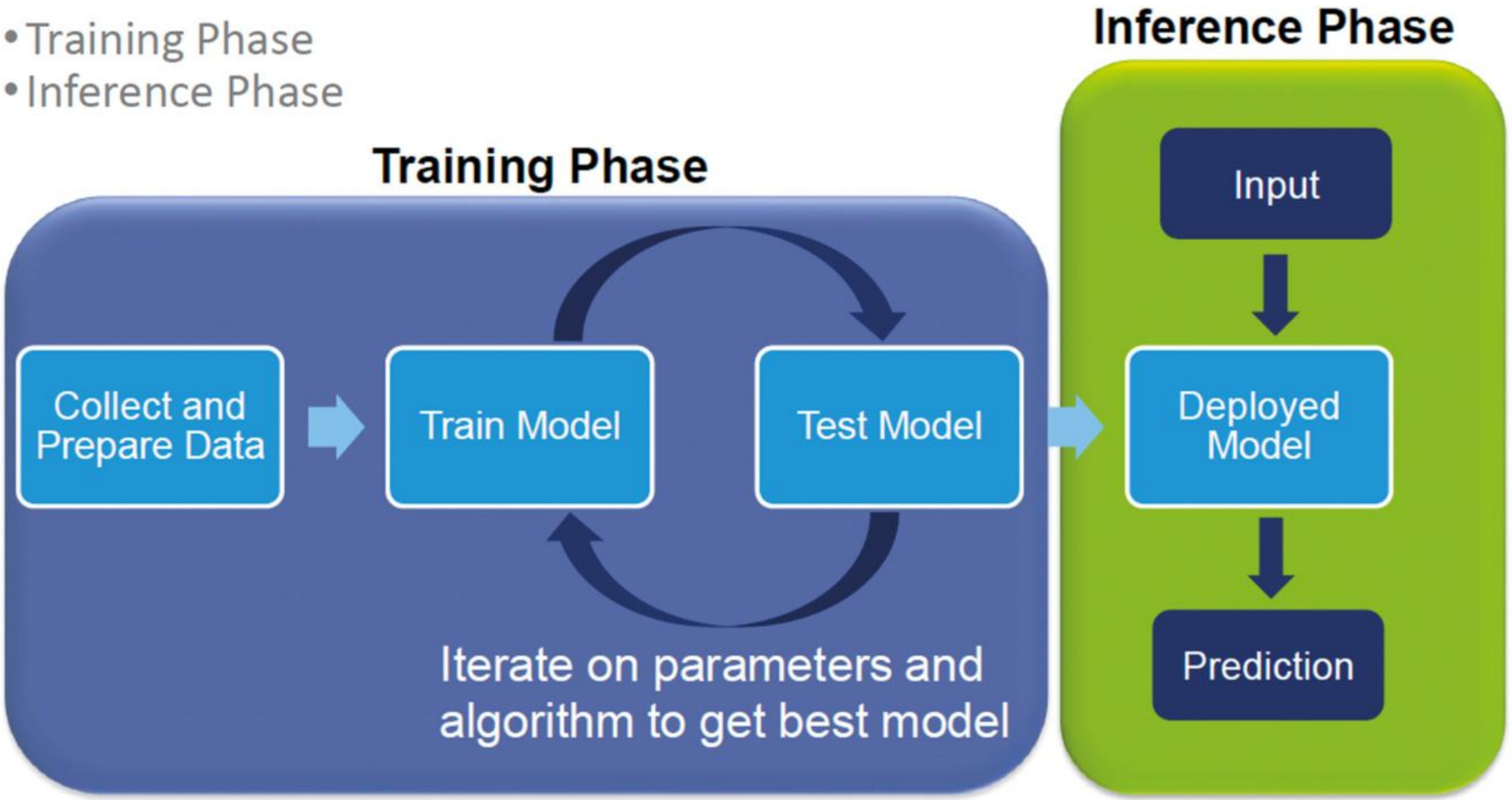
- Classification
  - Used when the output is a category (e.g. classifying a tumor as malignant or benign, or detecting email spam, or identifying the presence of a particular object in an image).

Core Types of Classification Algorithms

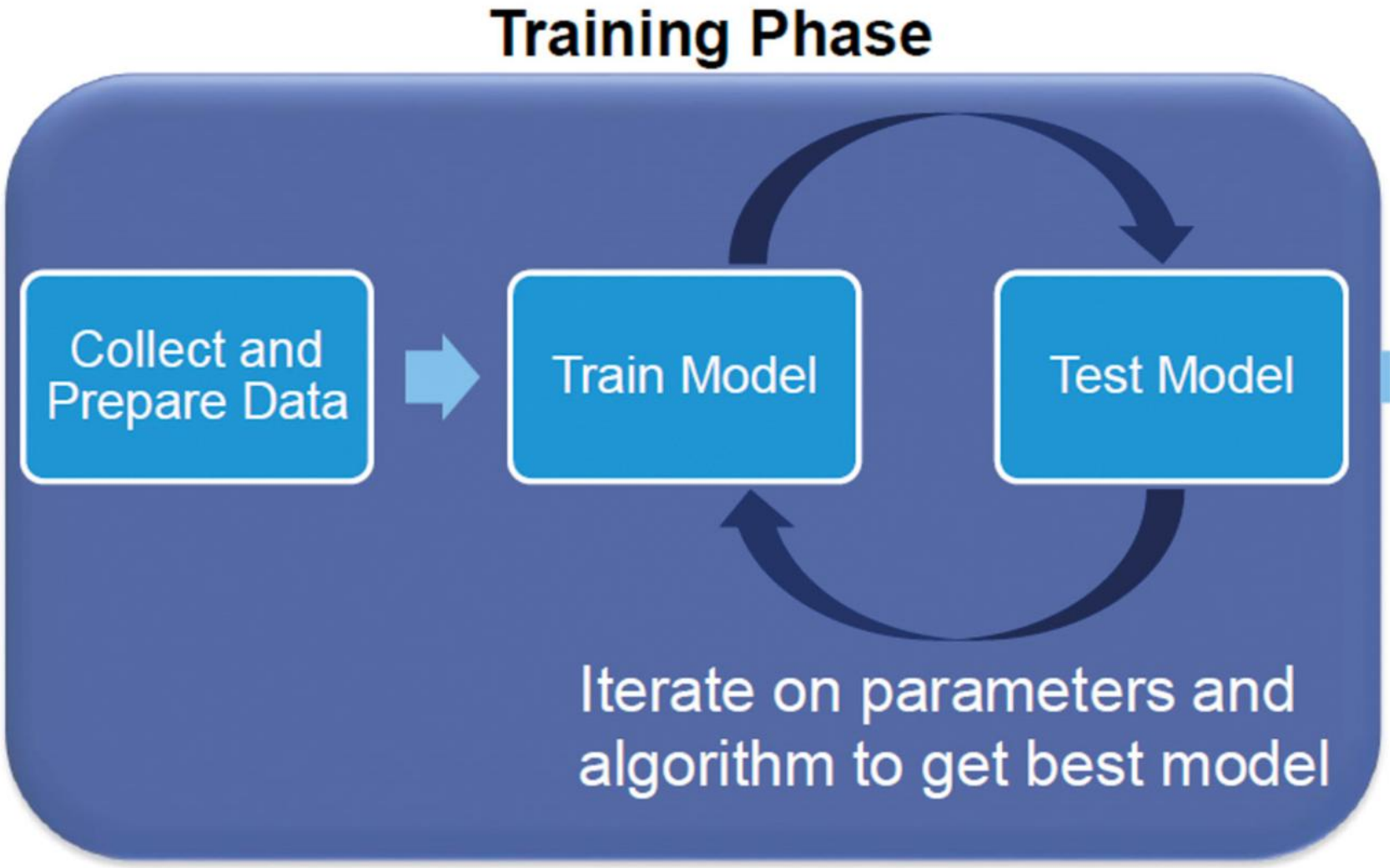
- Binary Classification: Sorting data into exactly two contrasting categories (e.g. UXO or No UXO)
- Multiclass Classification: Assigning data to one class out of three or more possible categories (e.g. identifying whether an image shows a Car, Truck, or Tank)
- Multilabel Classification: Permitting an individual data point to be assigned multiple target labels simultaneously (e.g. tag an image as Tracked + Armoured + Turreted)

Supervised Machine Learning Process

- Training Phase
- Inference Phase

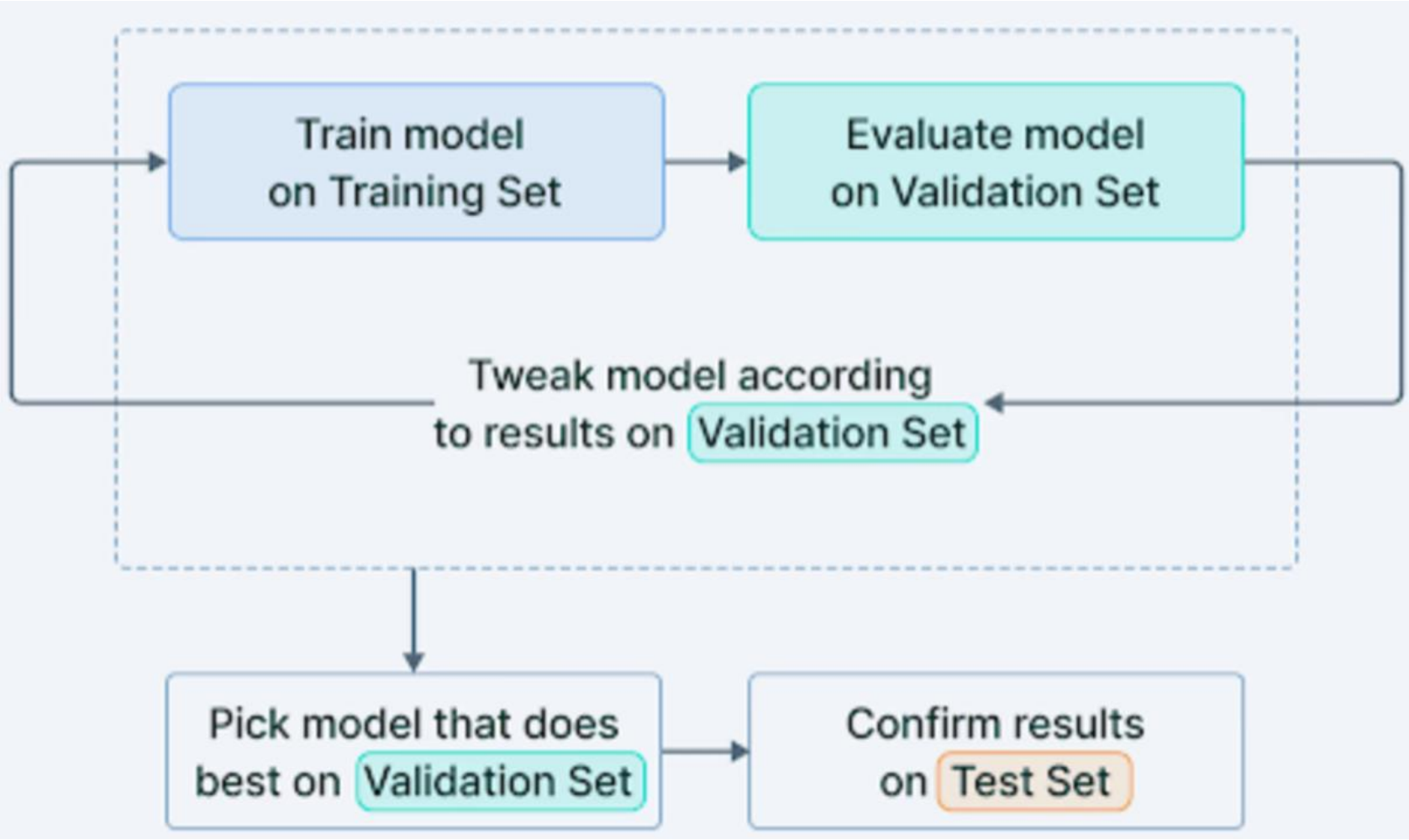


Supervised Machine Learning Process contd



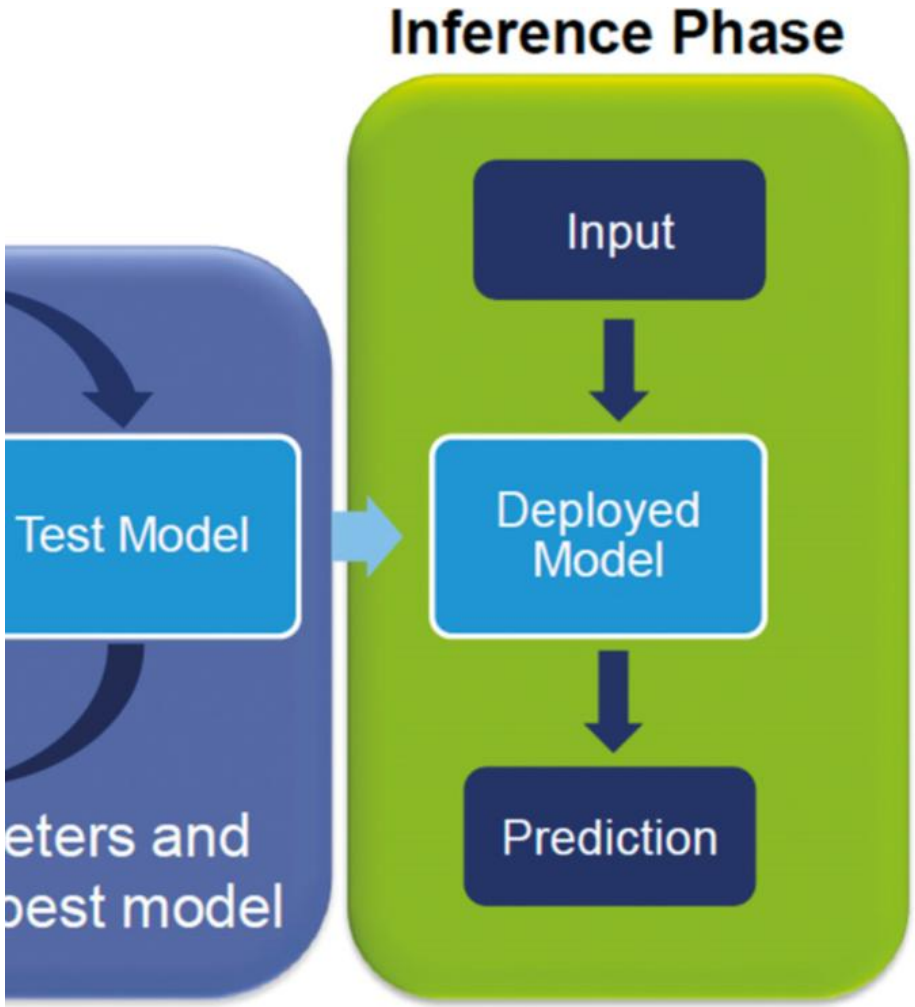
1. Data Collection and Preparation
2. Data Splitting
3. Model Selection
4. Training and Validation
5. Model Evaluation
6. Deployment and Monitoring

Supervised Machine Learning Process contd



1. Data Collection and Preparation
2. Data Splitting
3. Model Selection
4. Training and Validation
5. Model Evaluation
6. Deployment and Monitoring

Supervised Machine Learning Process contd



1. Data Collection and Preparation
2. Data Splitting
3. Model Selection
4. Training and Validation
5. Model Evaluation

6. Deployment and Monitoring

AI ML Assurance Considerations

Considerations – General

- Conceptually there is a difference between traditional software and an AI model from an assurance perspective. Traditional software behaviour is deterministic. In computer science a deterministic algorithm is an algorithm that, given a particular known input, will always produce a known output. When deployed, an AI model will be presented with data it hasn't yet experienced and hence the output will be unknown, and limited by its algorithm and inherent accuracy from the training phase. Considerations:
  - How good / representative / voluminous is the data set?
  - How well was the data set split between training/validation/testing sets?
  - How much training/validation/testing is optimal? (under/over-trained model)
  - How to handle safety-significant functions performed by the model?
  - How to handle AI specific information security requirements?
  - Operational Design Domain (ODD) vs operational environment

Considerations - AI-Assisted vs AI-Enabled Weapons

- The critical difference between an AI-assisted weapon and an AI-enabled weapon lies in the role of the human in the "kill chain" (Find, Fix, Track, Target, Engage, Assess).
- AI-assisted weapons rely on models to process data and recommend actions, but a human must authorise the strike.
- AI-enabled weapons can select and engage targets entirely on their own without human intervention.

Considerations – Safety

- Similar process to that used for software:
  - Proportional risk-based approach. Where:
    - Hazard analysis reveals safety-significant ML functions.
    - Safety Function Criticality Index (SFCI) is a function of:
      - the Function Control Category (FCC) - the level of control authority or autonomy the ML has over those functions
      - and the credible accident severity risk.
    - SFCI is used to defined required LoR tasks being commensurate with risk.
    - LoR tasks are additional and applied to a base machine learning process.

Function Control Category	Accident Severity Category			
	Catastrophic (1)	Critical (2)	Marginal (3)	Negligible (4)
1	SFCI 1	SFCI 1	SFCI 3	SFCI 4
2	SFCI 1	SFCI 2	SFCI 3	SFCI 4
3	SFCI 2	SFCI 3	SFCI 4	SFCI 4
4	SFCI 3	SFCI 4	SFCI 4	SFCI 4
5	SFCI 5	SFCI 5	SFCI 5	SFCI 5

Considerations – Security

- DoD Artificial Intelligence Cybersecurity Risk Management Tailoring Guide 14 July 2025 Version 2
- This guidance, establishes the DoD cybersecurity risk management tailoring guidance for the acquisition, development, use, sustainment, monitoring, and disposal of artificial intelligence (AI) systems. DoD system owners should use this tailoring guidance to plan for and tailor the control mitigations related to the cybersecurity of AI systems.
- Includes:
  - System security objectives for AI systems
  - Security requirements for AI systems
  - Authorisation considerations



DoD Artificial Intelligence Cybersecurity Risk Management Tailoring Guide

14 July 2025

Version 2

CLEARED
For Open Publication
Aug 07, 2025
Department of Defense
OFFICE OF PREPUBLICATION AND SECURITY REVIEW

Unclassified

Concluding Remarks

- AOP-52
  - What is AOP-52?
  - AOP-52 Major Update
  - AOP-52.x Standards Related Documents
- Artificial Intelligence / Machine Learning
  - AI vs Traditional Software
  - AI in Weapons Systems Context
  - Quick Primer in ML
  - Assurance Considerations

Questions

