



ER-27

ARTIFICIAL INTELLIGENCE IN AERONAUTICAL SAFETY-RELATED TAXONOMY

LEGAL NOTICE

This document is the exclusive intellectual and commercial property of EUROCAE. It is presently commercialised by EUROCAE. This electronic copy is delivered to your company/organisation for internal use exclusively. In no case it may be re-sold, hired, lent or exchanged outside your company. It may not be reproduced, in whole or in parts, without prior written permission by EUROCAE.

DECEMBER 2024

FOREWORD

This document was developed jointly by EUROCAE Working Group 114 - Artificial Intelligence and SAE G-34, and was approved by the EUROCAE Working Group 114 on 02 December 2024. The purpose of this document is to provide definitions for commonly used terms in the discussion of Artificial Intelligence (AI) applications and present a classification of AI techniques that will be discussed within other products of the EUROCAE WG-114/SAE G-34 Committee.

1. This document is technically identical to SAE AIR6987. Coordination has been achieved with SAE G-34.
2. EUROCAE, an international non-profit organisation, is the European leader in the development of worldwide recognised standards for aviation. This is achieved by utilising the expertise from our members and stakeholders across the global aviation community.
3. EUROCAE standards are developed following an open, transparent and consensus-based process, governed by process and quality management principles, which are clearly documented and is in line with the World Trade Organisation (WTO) Technical Barriers to Trade (TBT) Agreement, Annex 3 "Code of Good Practice for the Preparation, Adoption and Application of Standards".
4. EUROCAE standards and other deliverables are recommendations and best industry practices. EUROCAE is not an official body of the European governments. EUROCAE standards are intended to complement and support the regulatory and certification framework.

Whilst efforts are made during the standards-development process to avoid the inclusion of proprietary information and material, some of the elements of this document may be the subject of patent, copyright or other proprietary rights. EUROCAE shall not be held responsible for identifying any such rights. Trade names and similar terms used within this document do not constitute an endorsement by EUROCAE thereof.

5. Copies of this document may be obtained from:

EUROCAE

9 – 23 rue Paul Lafargue

93200 Saint-Denis

France

Telephone: +33 1 49 46 19 65

Email: eurocae@eurocae.net

Website: www.eurocae.net

TABLE OF CONTENTS

FOREWORD	i
TABLE OF CONTENTS	ii
TABLE OF FIGURES	ii
CHAPTER 1 SCOPE	3
CHAPTER 2 REFERENCES	4
2.1 APPLICABLE DOCUMENTS	4
2.2 DEFINITIONS	6
CHAPTER 3 CLASSIFICATIONS OF AI TECHNIQUES	11
3.1 MAIN CLASSIFICATION	11
3.1.1 Machine Learning	12
3.1.2 Symbolic AI	15
3.1.3 Numerical Analysis	18
3.2 SECONDARY CLASSIFICATION: FIELDS OF APPLICATION OF AI	22
3.2.1 Computer Vision	22
3.2.2 Natural Language Processing	22
3.2.3 Anomaly Detection [15], [49]	23
3.2.4 Optimization	24
CHAPTER 4 SUMMARY	26
CHAPTER 5 NOTES	27
5.1 REVISION INDICATOR	27
WG-114 MEMBERSHIP	28
IMPROVEMENT SUGGESTION FORM	40

TABLE OF FIGURES

Figure 3-1- AI classification	11
Figure 3-2- Machine learning diagram	13
Figure 3-3- Example of SNN [22] Reprinted with permission from [22]. © 2015 Michael A. Nielsen. ...	14
Figure 3-4 Example of DNN [22] Reprinted with permission from [22]. © 2015 Michael A. Nielsen. ...	14

CHAPTER 1

SCOPE

The objective of this document is to provide a classification of AI techniques that may be used in AI-based systems for aeronautical products. Aeronautical products include products in Airborne and Air Traffic Management (ATM)/ Air Navigation Systems (ANS) domains for crewed and uncrewed aircraft.

This document is:

- Intended to provide an understanding of the AI space, which will improve over time.
- Not intended to provide guidance, objectives, or safety considerations
- A scenario builder for AI technologies, in particular supervised learning

The publication of a taxonomy document for the aviation domain is an opportunity to support other AI standardization initiatives that will also publish taxonomy documents.

Disclaimer: This document provides content to support other products of the EUROCAE Working Group 114/SAE G-34.

CHAPTER 2

REFERENCES

2.1

APPLICABLE DOCUMENTS

- [1] OECD.AI/GPAI. (2024, May). *OECD AI principles overview*. <https://oecd.ai/en/ai-principles>.
- [2] EASA Artificial Intelligence (AI) Concept Paper Issue 2: Guidance for Level 1 & 2 machine learning applications. (2024). European Union Aviation Safety Agency.
- [3] Wilkinson, C., Lynch, J., Bharadwaj, R., and Woodham, K. (2016). Verification of adaptive systems [Report no. DOT/FAA/TC-16/4]. U.S. Department of Transportation, Federal Aviation Administration.
- [4] ISO/IEC 2382:2015, "Information technology - Vocabulary", 2015.
- [5] RTCA/EUROCAE DO-333/ED-216, "Formal Methods Supplement to DO-178C/ED-12C and DO-278A/ED-109A", 2011.
- [6] ISO/IEC 22989:2022, "Information technology - Artificial intelligence - Artificial intelligence concepts and terminology", 2022.
- [7] Vapnik, V.N. (1998). *Statistical learning theory*. Wiley.
- [8] Robert, N., Castro, R., and Aarti, S. (2009). *Statistical learning theory*. Rice University.
- [9] SAE/EUROCAE ARP4754B/ED-79B, "Guidelines for Development of Civil Aircraft and Systems", 2023.
- [10] RTCA/EUROCAE DO-178C/ED-12C, "Software considerations in airborne systems and equipment certification", 2011.
- [11] Russell, S. and Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- [12] Samuel, A.L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 211-229.
- [13] Mitchell, T.M. (1997). *Machine learning*. McGraw-Hill.
- [14] Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199-231.
- [15] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. The MIT Press.
- [16] Wolpert, D. and Macready, W.G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1).
- [17] Wolpert, D.H. (1996). The lack of A priori distinctions between learning algorithms. *Neural Computation*, 8(7), 1341-1390.
- [18] Murphy, K.P. (2012). *Machine learning: A probabilistic perspective*. The MIT Press.
- [19] Hastie, T., Tibshirani, R., and J. Friedman. (2009). *The elements of statistical learning* (2nd ed.). Springer.
- [20] Patel, A.A. (2019). *Hands-on unsupervised learning using Python*. O'Reilly.
- [21] Sutton, R.S. and Barto, A.G. (2018). *Reinforcement learning: An introduction* (2nd ed.). The MIT Press.
- [22] Nielsen, M.A. (2015). *Neural networks and deep learning*. Determination Press.
- [23] Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., and Shah, H. (2016). Wide & deep learning for recommender systems. In *DLRS 2016: Proceedings of the 1st workshop on deep learning for recommender systems*. Association for Computing Machinery.
- [24] Manna, Z. (1974). *Mathematical theory of computation*. Dover Publications.
- [25] OWL Working Group. (2004, February 10). *OWL web ontology language use cases and requirements*. W3C. <https://www.w3.org/TR/webont-req/#onto-def>.

- [26] OWL Working Group. (2012, December 11). *OWL 2 web ontology language document overview (second edition)*. W3C. <https://www.w3.org/TR/owl2-overview/>.
- [27] Bellini, P., Mattolini, R., and Nesi, P. (2000). Temporal logics for real-time system specification. *ACM Computing Surveys*, 32(1), 12-42.
- [28] Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. The MIT Press.
- [29] Gruber, T.R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220.
- [30] Sikos, L.F. (2015). *Mastering structured data on the semantic web: From HTML5 microdata to linked open data*. Apress.
- [31] Horrocks, I., Sattler, U., and Baader, F. (2007). Description logics. In *Foundations of Artificial Intelligence*. Elsevier.
- [32] Arp, R., Smith, B., and Spear, A.D. (2015). *Building ontologies with basic formal ontology*. The MIT Press.
- [33] Heleber, J., Fisher, M., and Perez-Lopez, A. (2009). *Semantic web programming*. Wiley.
- [34] Wong, W., Liu, W., and Bennamoun, M. (2012). Ontology learning from text: A look back and into the future. *ACM Computing Surveys*, 44(4), 1-36.
- [35] Casellas, N. (2012). *Linked legal data: A SKOS vocabulary for the Code of Federal Regulations*. IOS Press.
- [36] Mandal, S., Gandhi, R., and Siy, H. (2015, August 25). *Semantic web representations for reasoning about applicability and satisfiability of federal regulations for information security* [Conference paper]. 2015 IEEE Eighth International Workshop on Requirements Engineering and Law (RELAW), Ottawa, ON, Canada.
- [37] Sanya, I.O. and Shehab, E.M. (2015). A framework for developing engineering design ontologies within the aerospace industry. *International Journal of Production Research*, 53(8), 2383-2409.
- [38] CRYSTAL - Critical System Engineering Acceleration. (2013). *State of the art for healthcare ontology* [Report no. D407.010].
- [39] Aikins, J.S., Kunz, J.C., Shortliffe, E.H., and Fallat, R.J. (1982). PUFF: An expert system for interpretation of pulmonary function data. *Computers and Biomedical Research*, 16(3), 199-208.
- [40] Besl, P.J. and McKay, N.D. (1992). Method for registration of 3-D shapes. *Proc. SPIE 1611, Sensor Fusion IV: Control Paradigms and Data Structures*.
- [41] LaValle, S.M. and Kuffner, J.J., Jr. (2001). Randomized kinodynamic planning. *The International Journal of Robotics Research*, 20(5), 378-400.
- [42] Karaman, S. and Frazzoli, E. (2011). Sampling-based algorithms for optimal motion planning. *The International Journal of Robotics Research*, 30(7), 846-894.
- [43] Fischler, M.A. and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.
- [44] Hart, P.E., Nilsson, N.J., and Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2), 100-107.
- [45] Jackson, P. (1985). *Introduction to artificial intelligence*. Dover Publications.
- [46] Moravec, H. and Elfes, A. (1985, March 25-28). *High resolution maps from wide angle sonar* [Conference paper]. 1985 IEEE International Conference on Robotics and Automation, St. Louis, MO, USA.
- [47] Harris, C.G. and Stephens, M. (1988). A combined corner and edge detector. In C.J. Taylor (Ed.), *Proceedings of the Alvey Vision Conference* (pp. 23.1-23.6). Alvey Vision Club.
- [48] Bhattacharyya, S., Cofer, D., Musliner, D.J., Mueller, J., and Engstrom, E. (2015). *Certification considerations for adaptive systems* [Report no. NASA/CR-2015-218702]. National Aeronautics and Space Administration.

[49] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1-58.

[50] Holstein, W.K., Tanenbaum, M., Eilon, S., and Ackoff, R.L. (2024). Operations research. In *Encyclopedia Britannica*. <https://www.britannica.com/topic/operations-research>.

2.2

DEFINITIONS

ARTIFICIAL INTELLIGENCE (AI): Artificial intelligence is machine-based technology that, for explicit or implicit objectives, infers,¹ from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment (derived from [1]).

ARTIFICIAL NEURAL NETWORK (ANN) or NEURAL NETWORK (NN): A computational graph that consists of connected nodes ("neurons") that define the order in which operations are performed on the input. Neurons are connected by edges parameterized by weights (and biases). Neurons are organized into layers, specifically an input layer, several intermediate layers, and an output layer (extracted from [2]²).

ATTRIBUTE: Information representing a property or characteristic of a subject as recorded in structured data, such as a table, spreadsheet, or database. In a database context, attributes are commonly referred to as "fields", whereas in a spreadsheet or table context, attributes are commonly referred to as "columns". For instance, structured data representing a person may have attributes such as name or age.

BIAS (IN DATA): Bias in data refers to the presence of systematic errors, skewed representations, or imbalances in a dataset that can unfairly influence outcomes or predictions.

BIAS (IN ML MODEL): Bias in an ML model indicates a systematic error in the ML model's outputs, often stemming from biased data, flawed assumptions during the development, or oversimplification of the ML model.

CLASSIFICATION (MACHINE LEARNING): Method that maps input data into one of a number of discrete classes (extracted from [3]).

CLUSTERING: Method that identifies similarities between features of the data and groups data items into clusters with similar features. It is a common example of unsupervised learning (derived from [3]).

CONVOLUTIONAL NEURAL NETWORK (CNN): A specific type of deep neural networks that are particularly suited to process image data, based on convolution operators (extracted from [2]³).

DATA DRIFT: Phenomenon in which a machine learning system encounters data during operation whose distribution evolves over time. The predictive capabilities of the system are compromised as in-service data values diverge from those used in model training.

DATA-DRIVEN AI: An approach focusing on building a system that can learn a function based on having been trained on a large number of examples (extracted from [2]³).

DATASET (MACHINE LEARNING): Samples of data used for various development phases of the ML model, i.e., training, validation, and test datasets.

TRAINING DATASET: Dataset used as input during the training process to determine and assign values to the ML model's parameters (i.e., weights).

VALIDATION/DEVELOPMENT DATASET: Dataset used during the training process to evaluate the ML model. The validation dataset can be used to select the type of model and to tune the model's hyperparameters. Therefore, it is not used to perform verification activities.

¹ Refer to <https://artificialintelligenceact.eu/recital/12/> for more information about inference.

² Copyright European Union Aviation Safety Agency, 2003-2024.

³ Copyright European Union Aviation Safety Agency, 2003-2024.

TEST DATASET: Dataset used to verify that the final ML model satisfies the desired properties. The test data are not seen during the training process.

DECISION MAKING: Decision making is the selection of a course of action based on available information.

DEEP LEARNING (DL): A specific type of machine learning based on the use of large neural networks. These networks are designed to learn and represent data at multiple levels of abstraction, allowing them to automatically extract features and patterns from raw data (derived from [2]).

EXPERT SYSTEM: Knowledge-based system that provides support for solving problems in a particular domain or application area by drawing inferences from a knowledge base developed from human expertise. The term “expert system” is sometimes used synonymously with “knowledge-based system” but should be taken to emphasize expert knowledge. Some expert systems are able to improve their knowledge base and develop new inference rules based on their experience with previous problems. Expert systems usually rely on explicit, symbolic encoding of knowledge (extracted from [4]⁴).

EXPLAINABILITY (AI): Capability to provide the human with understandable and relevant information on how an AI/ML application is coming to its results (derived from [2]).

EXPLAINABLE AI (XAI): An emerging field of study that aims to make the workings of AI systems transparent, such that humans can see process and decision flows. Note that this may not be the same as having AI systems be logically understood by humans, although the goal of many Explainable AI projects is often aligned to that outcome.

FORMAL LANGUAGE: In logic, mathematics, computer science, and linguistics, a formal language is defined as a set of words comprised of letters from a specific alphabet, together with syntactical and semantical rules to construct well-formed sentences.

FORMAL METHODS: Formal methods are mathematically based techniques for the specification, development, and verification of software aspects of digital systems (extracted from [5]).

FORMAL LOGIC: Study of the structure and forms of valid argument without regard to the meaning of the terms in the argument (extracted from [4]⁴).

FUZZY LOGIC or FUZZY-SET LOGIC: Non-classical logic in which facts, inference rules, and quantifiers are given certainty factors (extracted from [4]⁴).

GENERALIZATION CAPABILITY: Ability of a learning algorithm to produce models that perform well on unseen data encountered during the operational phase. In this context, “unseen data” is from the same distribution as the training dataset (extracted from [2]⁵).

GENETIC ALGORITHM: An algorithm that simulates natural selection by creating and evolving a population of individuals (solutions) for optimization problems (extracted from [6]⁶).

HYPERPARAMETER (ML MODEL): The parameters whose values control the learning process and determine the values of ML Model Parameters. For example, number of hidden nodes and layers are some of the hyperparameters of neural network models.

INFERENCE (MACHINE LEARNING): The process of feeding the machine learning model with an input and computing its output (extracted from [2]⁵).

KNOWLEDGE: Collection of facts, events, beliefs, and rules, organized for systematic use (extracted from [4]⁴).

⁴ Republished with permission of the International Electrotechnical Commission (IEC), from ISO/IEC 2382:2015; permission conveyed through Copyright Clearance Center, Inc.

⁵ Copyright European Union Aviation Safety Agency, 2003-2024.

⁶ Republished with permission of the International Electrotechnical Commission (IEC), from ISO/IEC 22989:2022; permission conveyed through Copyright Clearance Center, Inc.

KNOWLEDGE ACQUISITION: Process of locating, collecting, and refining knowledge and converting it into a form that can be further processed by a knowledge-based system. Knowledge acquisition normally implies the intervention of a knowledge engineer, but it is also an important component of machine learning (extracted from [4]⁴).

KNOWLEDGE BASE; K-BASE: Database that contains inference rules and information about human experience and expertise in a domain. In self-improving systems, the knowledge base additionally contains information resulting from the solution of previously encountered problems (extracted from [4]⁷).

KNOWLEDGE ENGINEERING: Discipline concerned with acquiring knowledge from domain experts and other knowledge sources and incorporating it into a knowledge base. The term “knowledge engineering” sometimes refers particularly to the art of designing, building, and maintaining expert systems and other knowledge-based systems (extracted from [4]⁷).

KNOWLEDGE-BASED SYSTEM: Information processing system that provides for solving problems in a particular domain or application area by drawing inferences from a knowledge base. The term “knowledge-based system” is sometimes used synonymously with “expert system”, which is usually restricted to expert knowledge (extracted from [4]⁷).

LEARNING: Any of the processes followed to extract information from observations/data in the aim of perceiving, understanding, and characterizing physical phenomena or abstract constructs, pursuing a broad spectrum of objectives like predicting (e.g., problem solving), mimicking (e.g., speaking), replicating (e.g., physics law experimentation), etc. In the context of the present taxonomy, learning refers to the processes and techniques involved in statistical learning methods [7], [8].

MACHINE LEARNING (ML): The branch of AI concerned with the development of algorithms that allow computers to evolve behaviors based on observing data and making inferences on this data (extracted from [2]⁸).

ML MODEL: A model characterized by parameters determined by a data-driven machine learning process to satisfy one or more specific requirements.

MODEL: Abstract representation of a given set of aspects of a system/subsystem/item that has unambiguous, well-defined syntax and semantics (derived from [9]).

NATURAL LANGUAGE PROCESSING (NLP): A sub-field of artificial intelligence, concerned with computational techniques for analyzing, representing, and processing natural (human) language texts and voices, at one or more levels of linguistic analysis, such as phonetics, syntax, semantics, and discourse, for the purpose of achieving human-like language abilities for a range of applications, including natural language understanding and generation, and speech recognition and generation.

OPTIMIZATION: Methods for finding the optimal solution to a problem, according to a defined objective or cost function. It is important to note that there can be multiple optima in an optimization problem, including local and global optima. A local optimum is a solution that is optimal with respect to the surrounding values of the decision variables, while a global optimum is the overall optimal solution for the entire range of decision variables.

PERCEPTRON: A perceptron is a neuron defined by a linear function that sums up its inputs, weighted by a set of values and bias, and the output is modulated by an activation function.

RECURRENT NEURAL NETWORK (RNN): A type of neural network that involves directed cycles in memory (extracted from [2]⁸).

⁷ Republished with permission of the International Electrotechnical Commission (IEC), from ISO/IEC 2382:2015; permission conveyed through Copyright Clearance Center, Inc.

⁸ Copyright European Union Aviation Safety Agency, 2003-2024.

REGRESSION: A set of statistical techniques to postulate the mathematical relationship of one or more dependent variables to one or more independent variables and use such a relationship for statistical inference or prediction (derived from [3]).

REINFORCEMENT LEARNING: A process of learning in which the algorithm rewards positive results and punishes for a negative result, enabling it to improve over time. The learning system is called an agent.

ROBUSTNESS (SYSTEM): Ability of a system to maintain its level of performance under all foreseeable conditions (derived from [2] and [6]).

ROBUSTNESS (SOFTWARE): The extent to which software can continue to operate correctly despite abnormal inputs and conditions (extracted from [10]).

RULE-BASED SYSTEM: Knowledge-based system that draws inferences by applying a set of if-then rules to a set of facts following given procedures (extracted from [4]⁹).

SEMANTIC NETWORK (SEMANTIC NET): Concept-based knowledge representation in which objects or states appear as nodes connected with links that indicate the relationships between various nodes (extracted from [4]⁹).

SEMI-SUPERVISED LEARNING: The process of learning in which the algorithm is capable of learning from data that is partially labeled.

SUPERVISED LEARNING: The process of learning a function that maps an input to an output based on labeled training dataset.

SYMBOLIC AI: The branch of AI capturing knowledge and deriving decisions through explicit representation and rules.

SYMBOLIC LOGIC or MATHEMATICAL LOGIC: Discipline in which valid argument and operations are dealt with using an artificial language designated to avoid the ambiguities and logical inadequacies of natural languages (extracted from [4]⁹).

TRAINING: The process of optimizing the parameters (weights) of an ML model given a data set and a task to achieve on that data set. For example, in supervised learning, the training data consists of input (e.g., an image)/output (e.g., a class label) pairs, and the ML model “learns” the function that maps the input to the output, by optimizing its internal parameters (extracted from [2]¹⁰).

UNSUPERVISED LEARNING: The process of learning a function from a non-labeled dataset, by adapting the model to increase accuracy of the algorithm, based upon criteria that define the intended function objectives or constraints.

VARIANCE (MACHINE LEARNING): An error from sensitivity to small fluctuations in the training set. High variance can cause a learning algorithm to model the random noise in the training data, rather than the intended outputs (overfitting; extracted from [2]¹⁰).

VARIANCE (STATISTICS): A standard calculation in statistics that indicates the “spread” of a data distribution while reducing the sensitivity to outliers. Essentially, it tells us how far a set of numbers are spread out from their average value. More formally, variance is the expectation of the squared deviation of a random variable from its mean.

WEIGHT (MACHINE LEARNING): A parameter within a machine learning model that transforms input data. The inputs get multiplied by a weight value. (Note: Both bias and weights are learnable parameters inside the model).

WEIGHT (SYMBOLIC AI): A parameter within Semantic Networks which indicates the strength of the relationship between the nodes.

WELL FORMEDNESS: The quality of a clause, word, or other linguistic element that conforms to the grammar of the language of which it is a part. Well-formed words or phrases are grammatical, meaning they obey all relevant rules of grammar. In contrast,

⁹ Republished with permission of the International Electrotechnical Commission (IEC), from ISO/IEC 2382:2015; permission conveyed through Copyright Clearance Center, Inc.

¹⁰ Copyright European Union Aviation Safety Agency, 2003-2024.

a form that violates some grammar rule is ill-formed and does not constitute part of the language.

WELL-FORMED FORMULA (WFF): Finite sequence of symbols from a given alphabet that is part of a formal language.

CHAPTER 3

CLASSIFICATIONS OF AI TECHNIQUES

The goal of this document is to have a common knowledge base of AI terminology that will be used in other related documents. Not only will AI techniques be explained, but they will also be mapped into categories.

At first, different references which include a classification of AI techniques (see [2], [6], [11]) were analyzed. In a second step, seven approaches to the classification of AI techniques were proposed and considered, each one with a particular focus, such as mathematical decomposition or explainability. Depending on the selected criterion used to cluster the techniques, the resulting classification significantly changes.

Finally, two distinct classifications were selected because they were considered correct (i.e., with limited overlaps between the different categories), complete (i.e., covering all techniques), and complementary to each other. They are presented in this section. The main classification splits artificial intelligence into three different categories: machine learning, symbolic AI, and numerical analysis. The secondary classification is also proposed to take into account the different fields of application: computer vision, natural language processing, anomaly detection, and optimization.

3.1

MAIN CLASSIFICATION

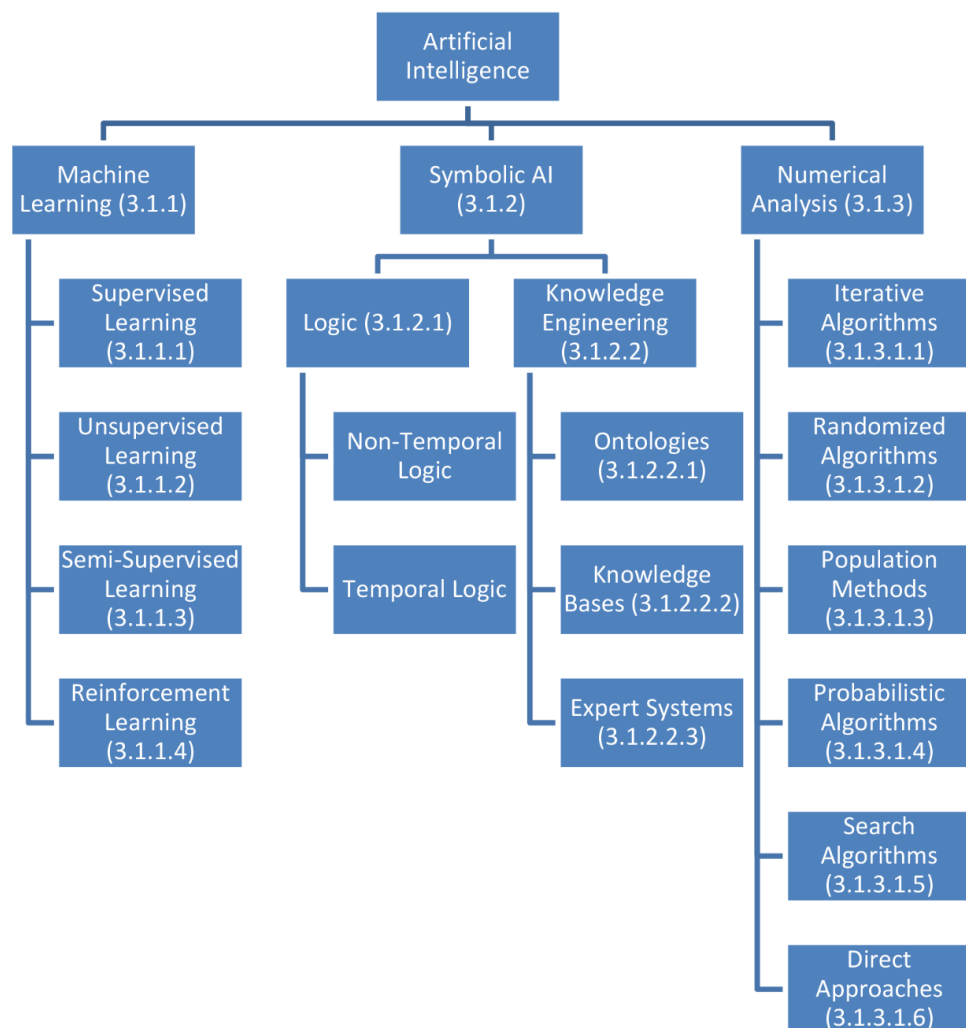


FIGURE 3-1- AI CLASSIFICATION

3.1.1 Machine Learning

Arthur Samuel [12] defined machine learning (ML) in 1959 as “the field of study that gives computers the ability to learn without being explicitly programmed”. Three decades later, Tom Mitchell [13] proposed a more actionable definition of ML: “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks on T , as measured by P , improves with experience E ”.

Examples of ML applications are document classification, email spam filtering, image classification, semantic segmentation, handwriting recognition, face detection and recognition, regression, dimensionality reduction (by generating embeddings), clustering, anomaly detection, (multi-)agent planning, speech recognition, sentiment analysis, and autonomous vehicle navigation.

The distinction between the algorithmic nature of ML and traditional statistical modeling is extensively explored in [14]. The distinction also appears in [15], where Goodfellow, Bengio, and Courville state that “machine learning is essentially a form of applied statistics with increased emphasis on the use of computers to statistically estimate complicated functions and decreased emphasis on proving confidence intervals around these functions”.

The ML field is composed of multiple learning strategies and algorithm families with no discernible advantage among alternatives if one makes absolutely no assumption about the training data (No Free Lunch theorem [15], [16], [17]). ML algorithms can be typically classified into major subcategories (see Figure 3-1): supervised learning (see 3.1.1.1), unsupervised learning (see 3.1.1.2), semi-supervised learning (see 3.1.1.3), reinforcement learning (see 3.1.1.4). Other emerging learning strategies such as self-supervised learning are not presented in this classification. The different learning strategies are summarized in 3.1.1.5. Finally, a focus on neural networks and deep learning is provided in 3.1.1.6.

3.1.1.1 Supervised Learning [18], [19], [11]

In supervised learning, the training set (i.e., experience E) fed to the ML algorithm includes the desired solutions (aka labels) for each instance data. The comparison of algorithm predictions against labeled data drives the learning process. Typical supervised learning tasks include regression and classification.

3.1.1.2 Unsupervised Learning [18], [20], [11]

In unsupervised learning, the training set does not include labeled data, and the learning process is then guided by algorithms able to search for distinctive characteristics among datasets. Unsupervised algorithms learn underlying structures in the training set and use these patterns to make predictions. Typical unsupervised learning tasks include clustering, anomaly detection, visualization, dimensionality reduction, and association rule learning.

3.1.1.3 Semi-Supervised Learning [20], [11]

In semi-supervised learning, the ML algorithm is capable of learning from data that is partially labeled. Most semi-supervised learning algorithms use unsupervised learning approaches to improve supervised learning solutions in dealing with problems such as insufficient labeled data, the curse of dimensionality, feature engineering, outliers, and data drift.

3.1.1.4 Reinforcement Learning [21], [11]

In reinforcement learning, the ML algorithm (agent) learns what to do by interacting with the environment and observing its reaction to the agent's actions. The learning objective is to determine an action-selection policy that maximizes a numeric reward signal. The agent must discover which actions yield the most reward by a trial-and-error process. The trial-and-error search and the learner's need to consider delayed rewards are the two most important distinguishing features of reinforcement learning. A classic example of reinforcement learning is a game, where the objective, such as the safe landing of a spacecraft without crashing, is well known. The algorithm uses repeated attempts to determine the optimal set of maneuvers and fuel usage to best achieve the objective.