


Chapter 17

Methods, Techniques, and Application of Explainable Artificial Intelligence

Ankur Dumka


Women Institute of Technology, India

Vaibhav Chaudhari

 <https://orcid.org/0000-0001-5781-2147>

Nutanix Technologies India Pvt. Ltd., India

Anil Kumar Bisht

 <https://orcid.org/0000-0002-5629-8809>

MJP Rohilkhand University, India

Ruchira Rawat

Graphic Era University, India

Arnav Pandey

St. Joseph's Academy, India

ABSTRACT

With advancement in machine learning, the use of machine learning has been increased, and explainable artificial intelligence (XAI) has emerged as an area of research and development for addressing the opacity and complexity of machine learning models. This chapter has proposed the overview of the current state of explainable artificial intelligence with highlighting its significance, disadvantages, and its potential applications in different fields. This chapter explores several explainable artificial techniques ranging from post-hoc methods like SHAP, LIME to decision tree and rule-based systems. This chapter also focusses on complexity and interpretability of a model.

DOI: 10.4018/979-8-3693-2351-9.ch017

INTRODUCTION

Artificial intelligence can be explained as simulation of intelligence of humans in machine which are programmed in such a manner that they can learn and think like a human. Artificial intelligence helps in enabling machine to perform task by developing algorithms and thus avoid the use of human intelligence. The various task involved in artificial intelligence are problem solving, perception, understanding of languages, learning and decision making.

Artificial intelligence can be categorized into following two types as:

1. Narrow AI or Weak AI
2. General AI or Strong AI

Narrow AI: These AI techniques are developed to perform some specific task or a narrow range of tasks just like recognition of image, translation of language etc.

General AI: These AI techniques are having ability of learning, understanding and applying knowledge to broader range of tasks similar like human intelligence.

Machine learning is a sub-set of artificial intelligence which trained algorithms on large datasets in order to recognize patterns and make predictions. Deep learning is also sub-set of artificial intelligence which involves artificial intelligence neural network which are inspired by function of human brain. AI can be used in different fields like healthcare, finance, robotics, natural language etc.

AI can be divided into following types as:

Figure 1 categorizes AI into three parts as:

1. Symbolic AI
2. Statistical AI
3. Explainable AI

Symbolic AI: symbolic AI models are explainable by design. They focus on encoding knowledge and reasoning rules to perform task.

Statistical AI: these AI relies on learning from data to make prediction or decisions.

Explainable AI: Explainable AI models are models with inner logic that can clearly be described in human language.

Figure 1. Categorization of AI



16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/methods-techniques-and-application-of-explainable-artificial-intelligence/346584

Related Content

Data Science for Industry 4.0

Indraneel Dabhade (2023). *Encyclopedia of Data Science and Machine Learning* (pp. 58-71).

www.irma-international.org/chapter/data-science-for-industry-40/317438

Deep Learning Models for Physiological Data Classification of Children During Computerized Auditory Tests: Deep Learning-Based Emotion Recognition in Child-Computer Interaction

Duygun Erol Barkana, Itr Kakç, Hatice Kose, Elif Toprak, Selma Ylarand Dilara Demirpençe Seçinti (2022). *Handbook of Research on New Investigations in Artificial Life, AI, and Machine Learning* (pp. 43-59).

www.irma-international.org/chapter/deep-learning-models-for-physiological-data-classification-of-children-during-computerized-auditory-tests/296800

Development of a Charge Estimator for Piezoelectric Actuators: A Radial Basis Function Approach

Morteza Mohammadzaheri, Mohammadreza Emadi, Mojtaba Ghodsi, Issam M. Bahadur, Musaab Zarogand Ashraf Saleem (2020). *International Journal of Artificial Intelligence and Machine Learning* (pp. 31-44).

www.irma-international.org/article/development-of-a-charge-estimator-for-piezoelectric-actuators/249251

Integrated Regression Approach for Prediction of Solar Irradiance Based on Multiple Weather Factors

Megha Kambleand Sudeshna Ghosh (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-12).

www.irma-international.org/article/integrated-regression-approach-for-prediction-of-solar-irradiance-based-on-multiple-weather-factors/294105

Autonomous Last Mile Shuttle ISEAUTO for Education and Research

Raivo Sell, Mairo Leier, Anton Rassõlkinand Juhan-Peep Ernits (2020). *International Journal of Artificial Intelligence and Machine Learning* (pp. 18-30).

www.irma-international.org/article/autonomous-last-mile-shuttle-iseauto-for-education-and-research/249250