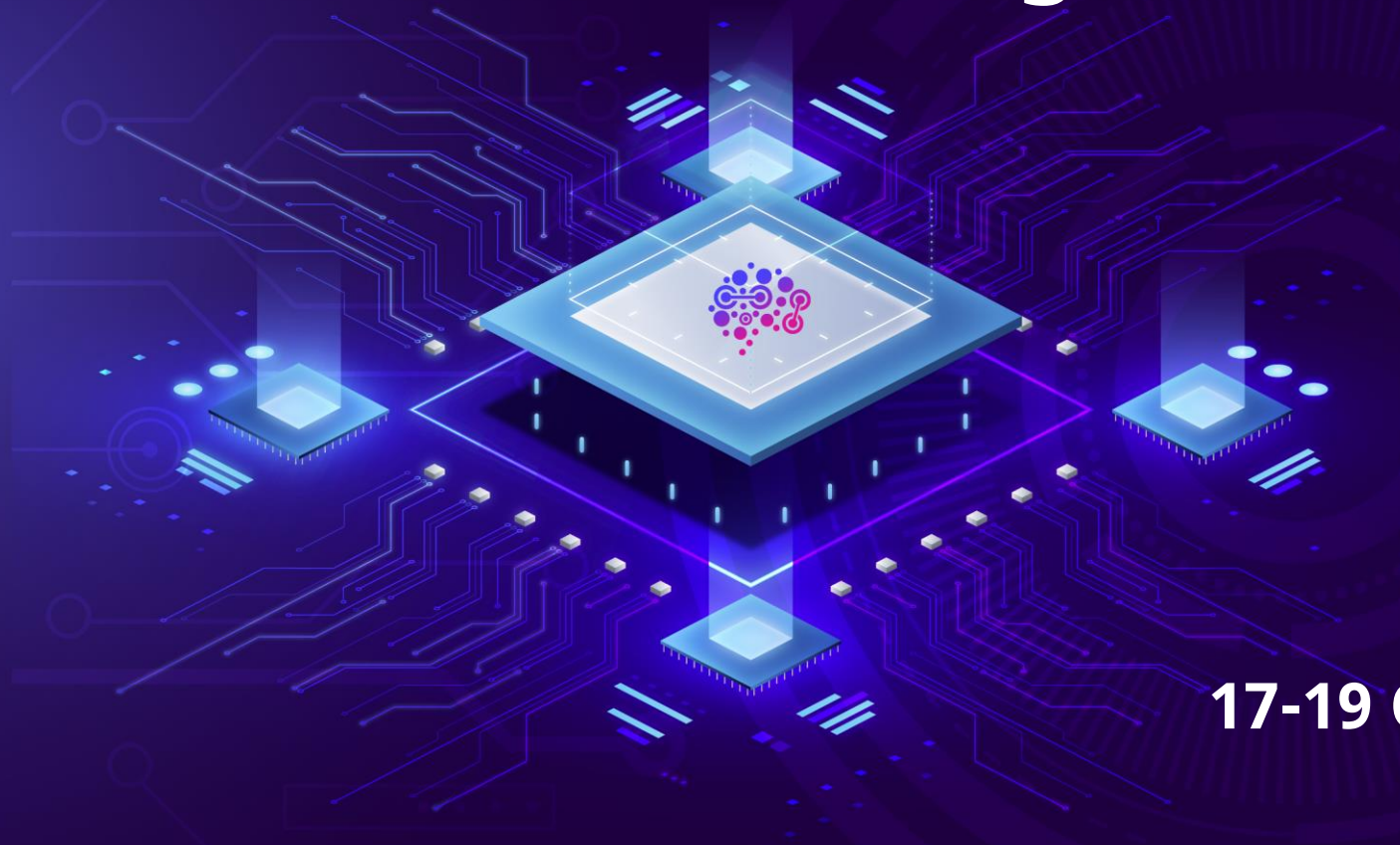
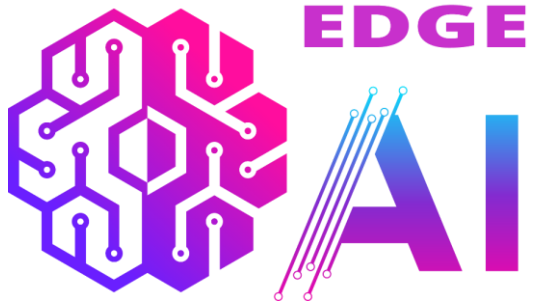
European Conference on EDGE AI Technologies and Applications - EEAI



Advancing emerging edge AI technologies and driving next generation intelligent applications.

17-19 October 2023, Athens, Greece

European Conference on EDGE AI Technologies and Applications - EEAI



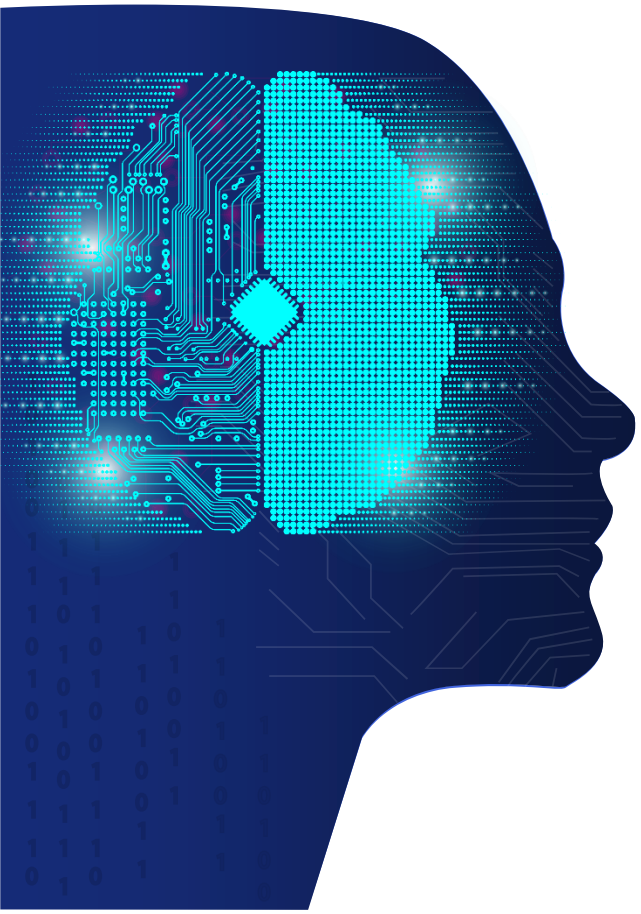
Explainability and Interpretability Concepts for Edge AI Systems

Angelo Genovese, Università degli Studi di Milano, Italy

17-19 October 2023 Athens, Greece

Copyright © 2023

Presentation Outline



- Motivations for explainability and interpretability
  - Focus on Edge computing
- Taxonomy of the latest methodologies
- Benchmarking approaches
- Challenges, open issues, and research trends

Copyright © 2023

Introduction

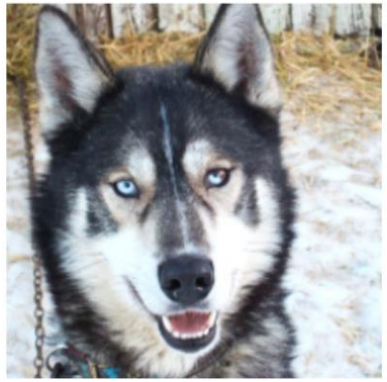
- Explainability
  - Understandable solutions
  - Communicate the results of AI/ML/DL technology to end users

- Interpretability
  - Model transparency
    - Users will understand precisely why and how a model generates predictions
    - Monitoring and scrutinising the inner mechanisms of AI/ML/DL methods
    - Interpreting the model's weights and features to determine the given output

Copyright © 2023

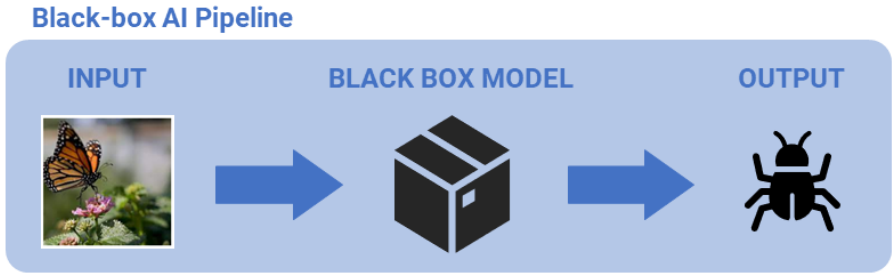
Motivations (1)

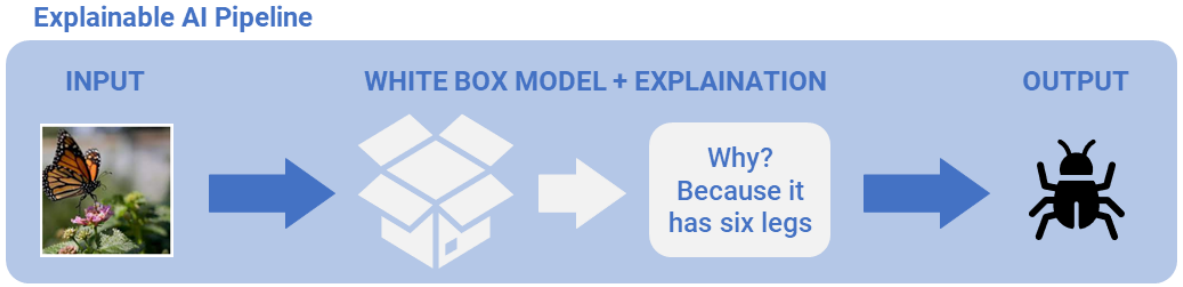
- XAI
  - Detect
    - Data bias
    - Incorrect features
    - Overfitting
    - …
- XAI in EdgeAI
  - *Informed decision-making*
  - Predictive analytics
  - Anomaly alerts



(a) Husky classified as wolf

(b) Explanation

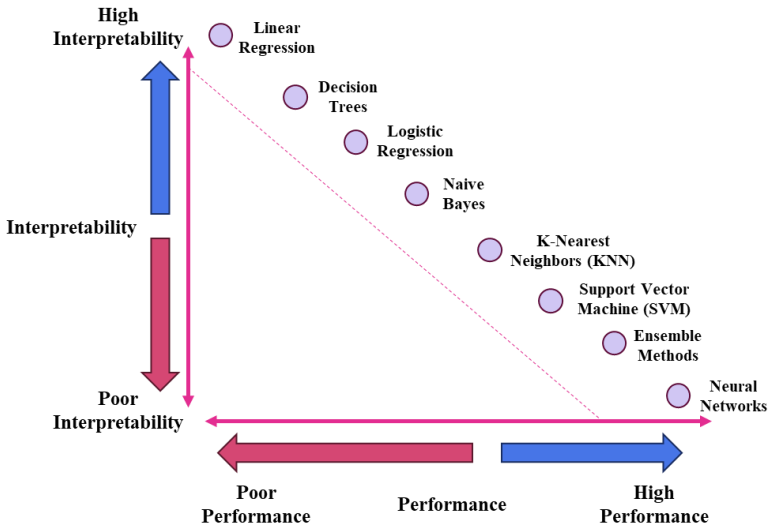


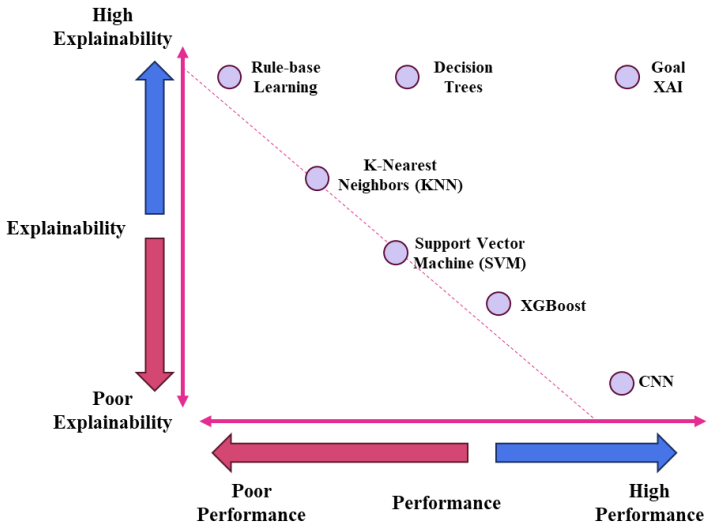


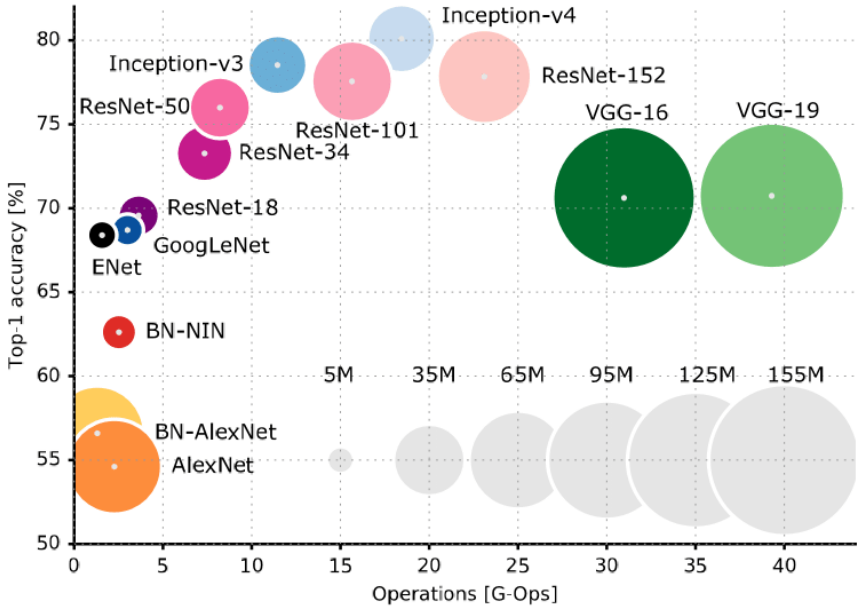
Copyright © 2023

Motivations (2)

- (Old) challenges
  - **Explainability and interpretability at the expense of performance**







Copyright © 2023

Motivations (3)

- **(New, edge-based)** challenges
  - Trade-off
    - Performance
    - Constrained resources at the edge
    - Model complexity
    - Power consumption
  - Lack of benchmarking and standardisation

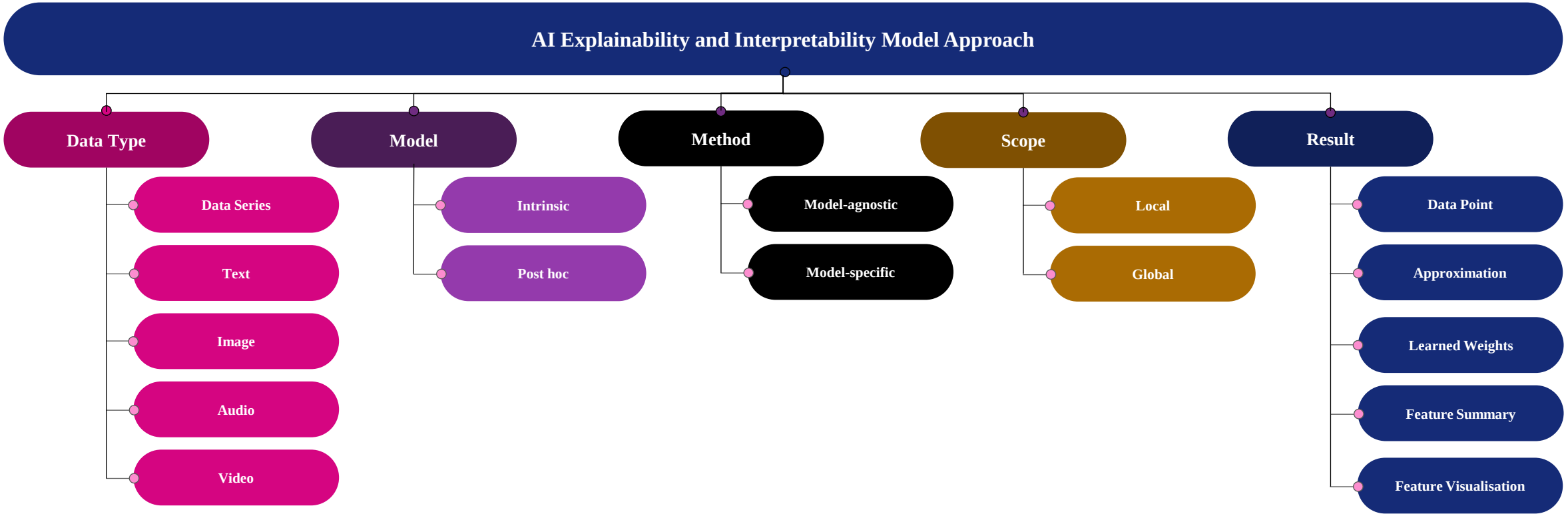
Copyright © 2023

Fields

- Agriculture
- Finance
- Industry and Autonomous Robots
- Medical Diagnosis
- Military and Security
- Recommendation Systems and Marketing
- Smart Cities

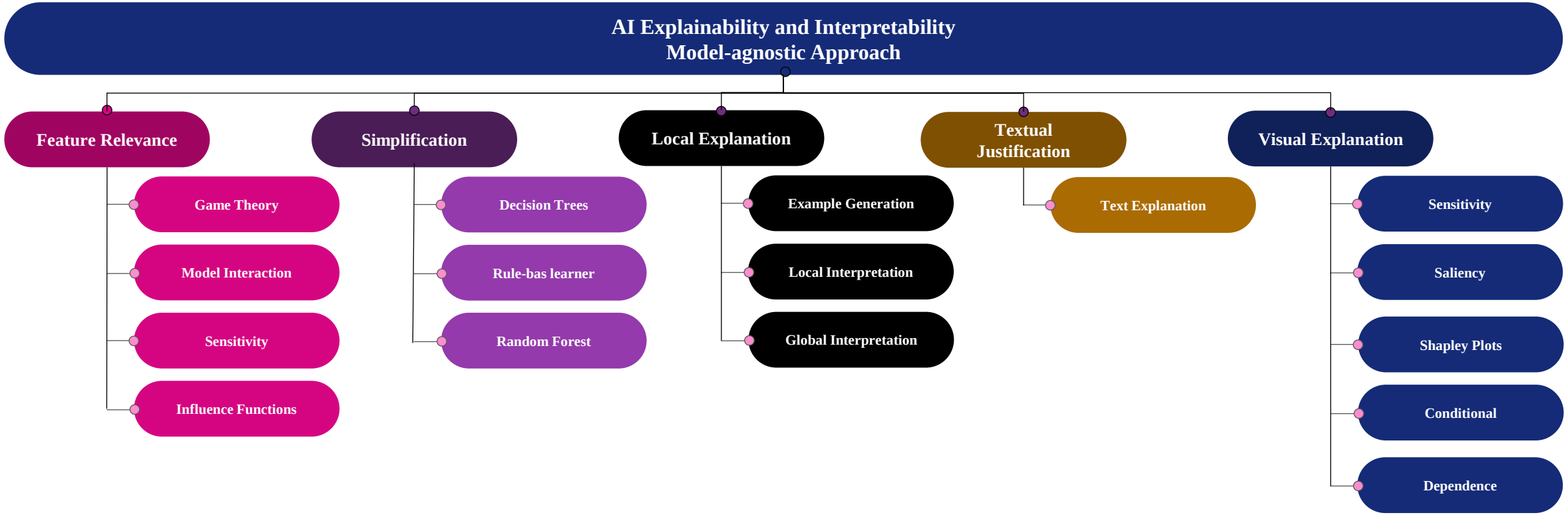
Copyright © 2023

Taxonomy (1)



Copyright © 2023

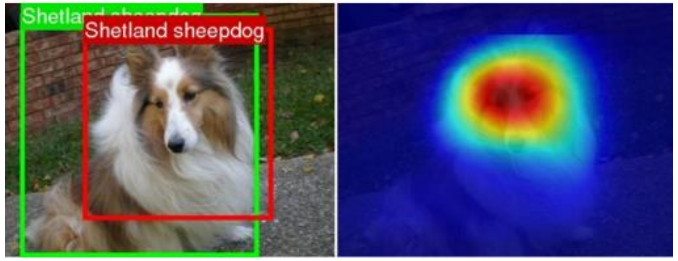
Taxonomy (2)

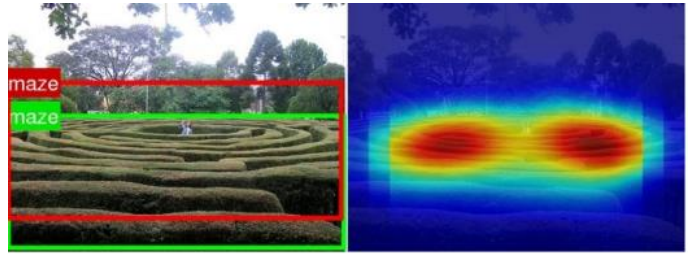


Copyright © 2023

Benchmarking

- Quantitative
  - Synthetically generated data with known properties
  - Measures
    - Fidelity, stability, deletion, insertion, monotonicity, running time
- Qualitative
  - Questionnaires

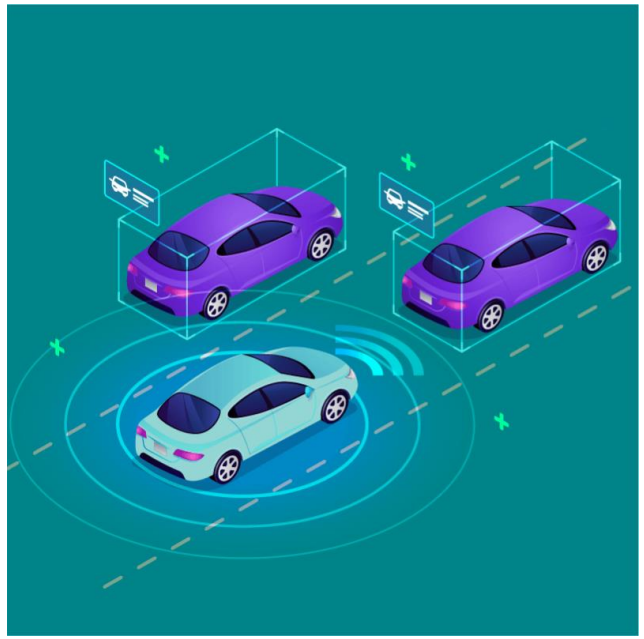


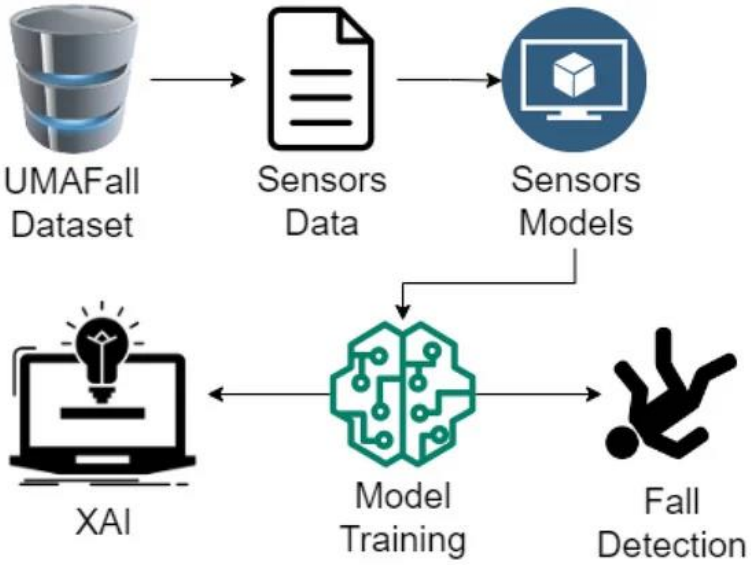


Copyright © 2023

Edge AI Explainability and Interpretability

- Specific applications
  - Autonomous driving
    - XAI for LiDAR
    - Driver behavior
  - Health care
  - IoT
    - Traffic control
    - Smart buildings
    - Environmental monitoring
    - Industrial control





Copyright © 2023

Challenges and Open Issues

- Integration of edge and XAI
  - Holistic edge AI design framework for the whole edge AI system
- Trade-off
  - Model complexity, interpretability, and explainability
  - AI models that are powerful and provide accurate results
    - Yet sufficiently simple to be understandable
- Lack of expertise
  - Lack of standard, best practices

Copyright © 2023

Conclusion

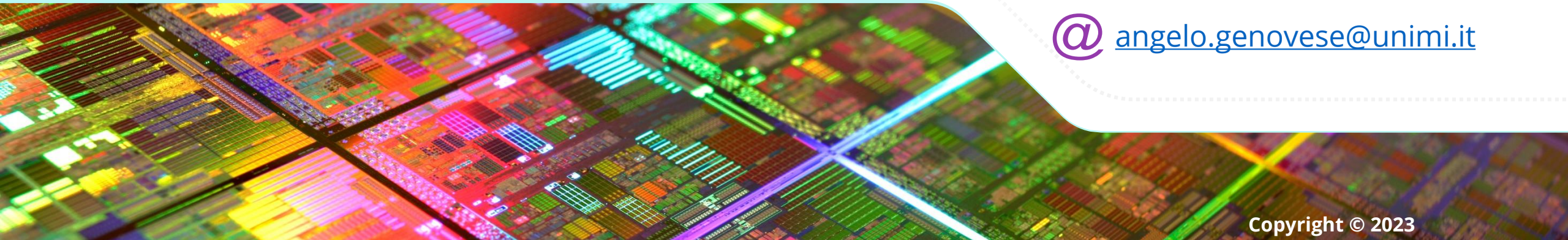
- Motivation for explainability and interpretability
- Taxonomy of approaches
- Benchmarking methods
- Challenges and issues of Edge-based XAI

Copyright © 2023

EDGE AI

Thank You

For your attention



@ angelo.genovese@unimi.it

Copyright © 2023