

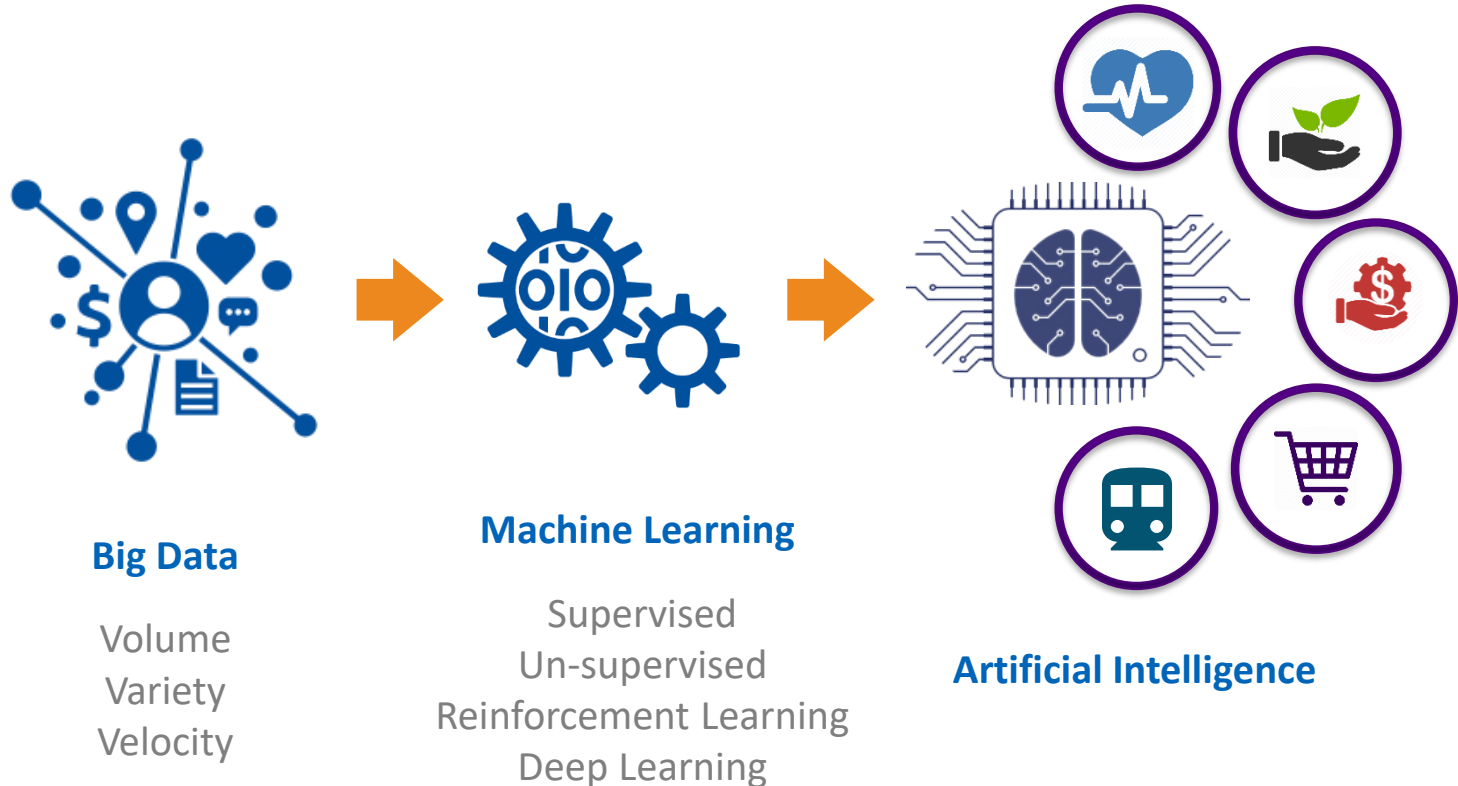
# World Standards Day

Technical Report

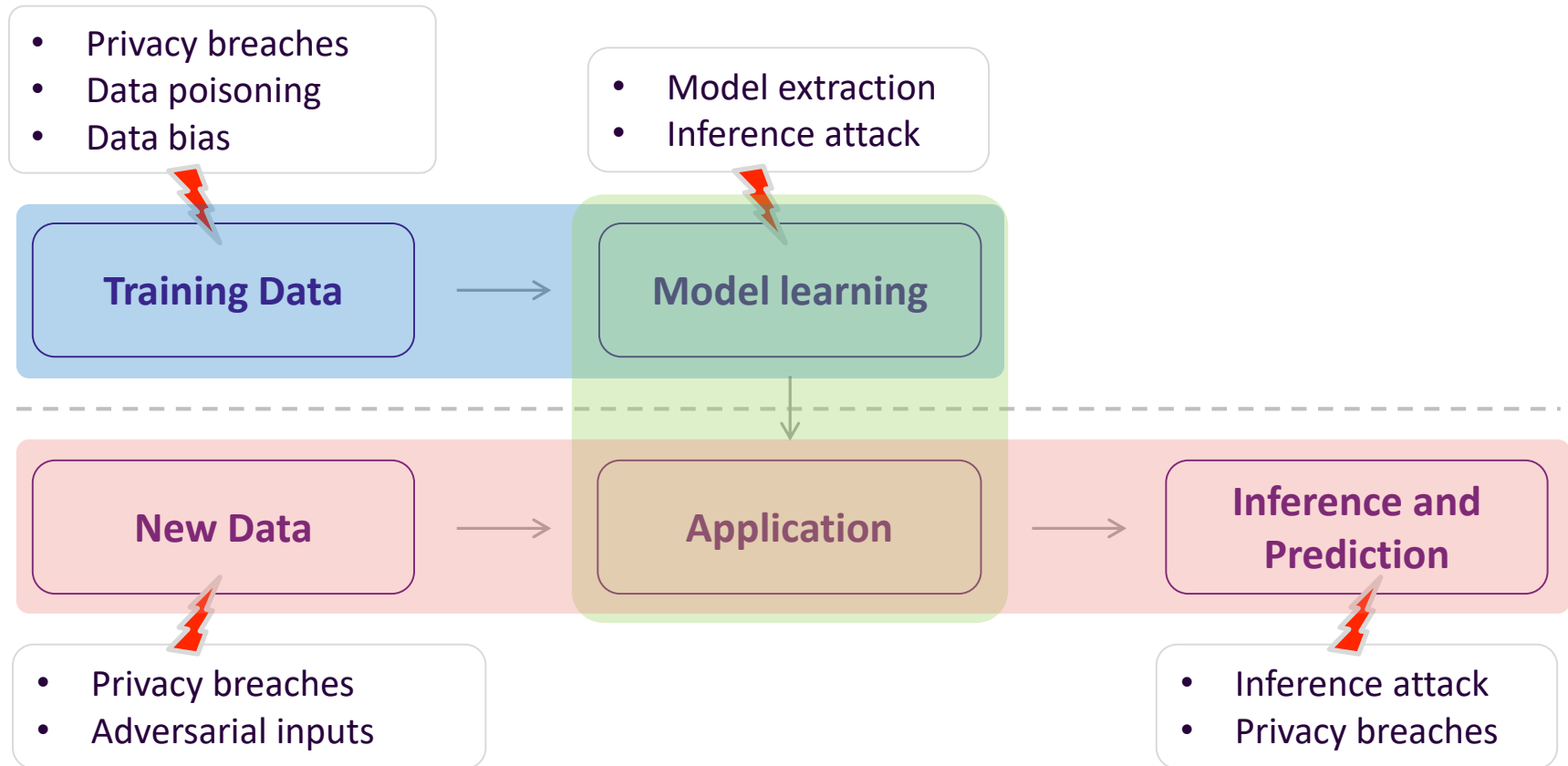
# Artificial Intelligence and Big Data

Saharnaz DILMAGHANI

# Big Data, Machine Learning (ML), and Artificial Intelligence (AI)



# Data Privacy and Trustworthiness Threats



# Attack Examples



Recovered image



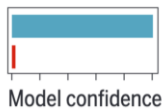
Training set image

Fredrikson, Matt, Somesh Jha, and Thomas Ristenpart. "Model inversion attacks that exploit confidence information and basic countermeasures." Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. ACM, 2015.

Original image

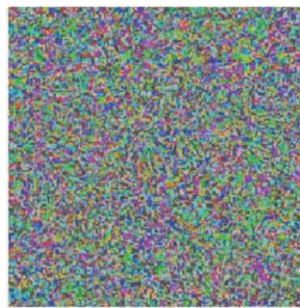


Benign  
Malignant



Dermatoscopic image of a benign melanocytic nevus, along with the diagnostic probability computed by a deep neural network.

Adversarial noise



Perturbation computed by a common adversarial attack technique. See (7) for details.

+ 0.04 ×

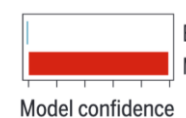
=

Adversarial example



Combined image of nevus and attack perturbation and the diagnostic probabilities from the same deep neural network.

Benign  
Malignant



Finlayson, Samuel G., et al. "Adversarial attacks on medical machine learning." Science 363.6433, 2019.



# Gap Analysis

## Identification

- Definitions and terminology
- Computational approaches
- Applications and use cases
- Privacy and security threats

- 
- ✓ ISO/IEC 2382
  - ✓ ISO/IEC WD 22989
  - ✓ ISO/NP TR 23347
  - ✓ ISO/IEC AWI 24372
  - ✓ ISO/NP 3534-5
  - ✓ ISO/IEC WD 23053
  - ✓ ISO/IEC NP TR 24030
  - ✓ ISO/IEC TR 20547-2
  - ✓ ISO/IEC PDTR 24028



# Gap Analysis

## Measurements

- Metrics
- Mitigation Strategies
- Risk quantification

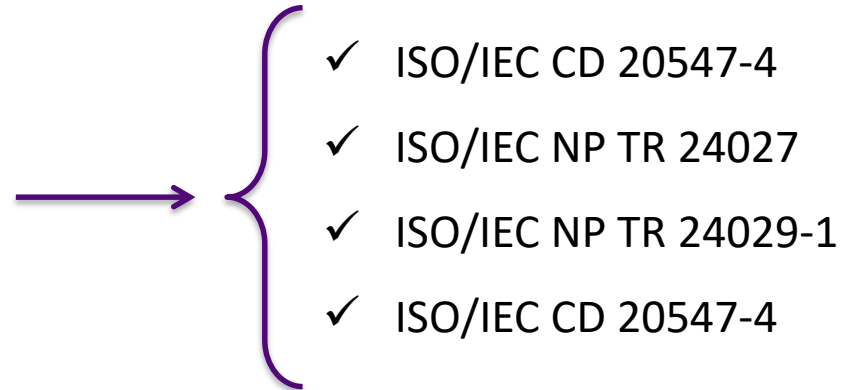
- 
- ✓ ISO/IEC PDTR 24028
  - ✓ ISO/IEC PDTR 24028
  - ✓ ISO/IEC NP TR 24027
  - ✓ ISO/IEC NP TR 24029-1
  - ✓ ISO/IEC 20889
  - ✓ ISO/IEC TR 27103
  - ✓ ISO/IEC 18033-6



# Gap Analysis

## Implementation

- Tests and Evaluation
- Constraints
- Implementation





# Thank You!

saharnaz.dilmaghani@uni.lu