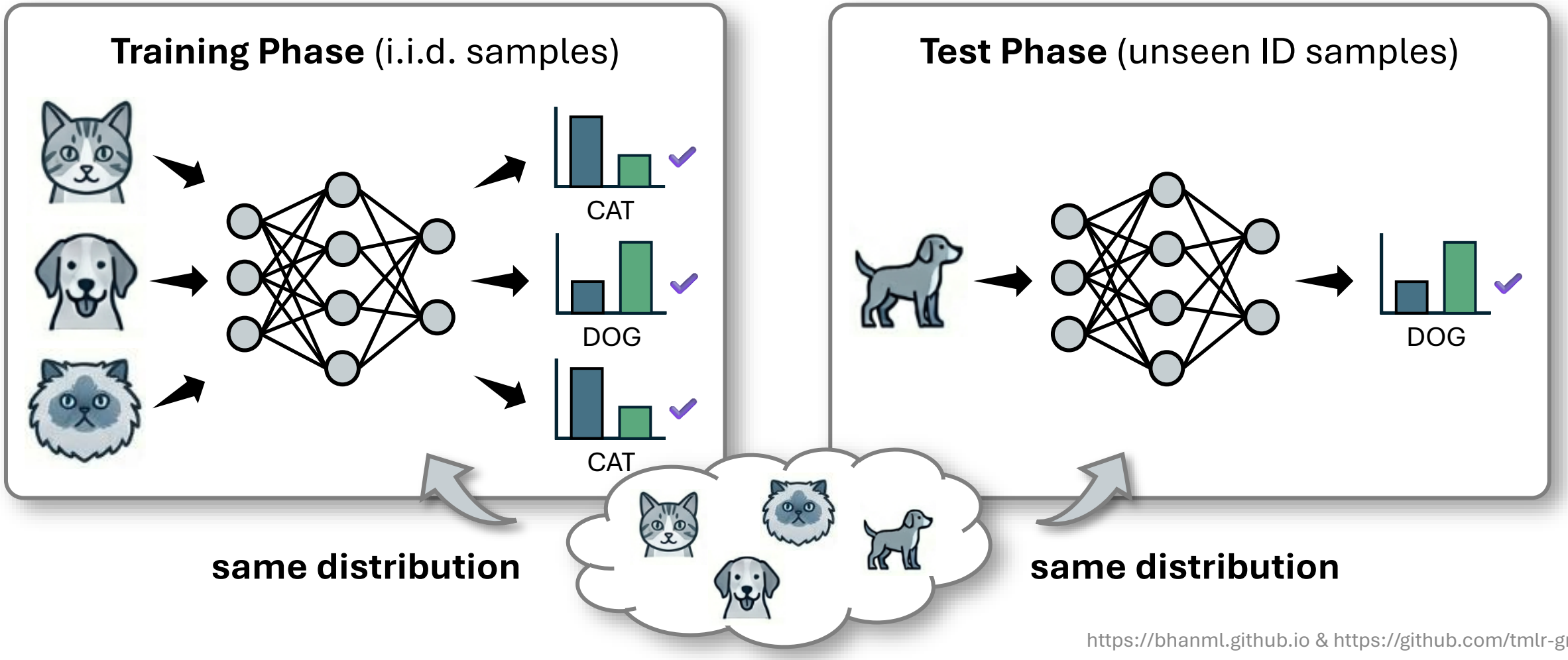
Handling Out-of-distribution Data in the Open World: Principles and Practices for Reliable AI

Dr. Jianing Zhu, Dr. Qizhou Wang, Dr. Yongqiang Chen, and Prof. Bo Han

Hong Kong Baptist University, TMLR Group
RIKEN Center for Advanced Intelligence Project (AIP)
University of Texas at Austin, VITA Group
Mohamed bin Zayed University of Artificial Intelligence
Carnegie Mellon University, CLeaR group

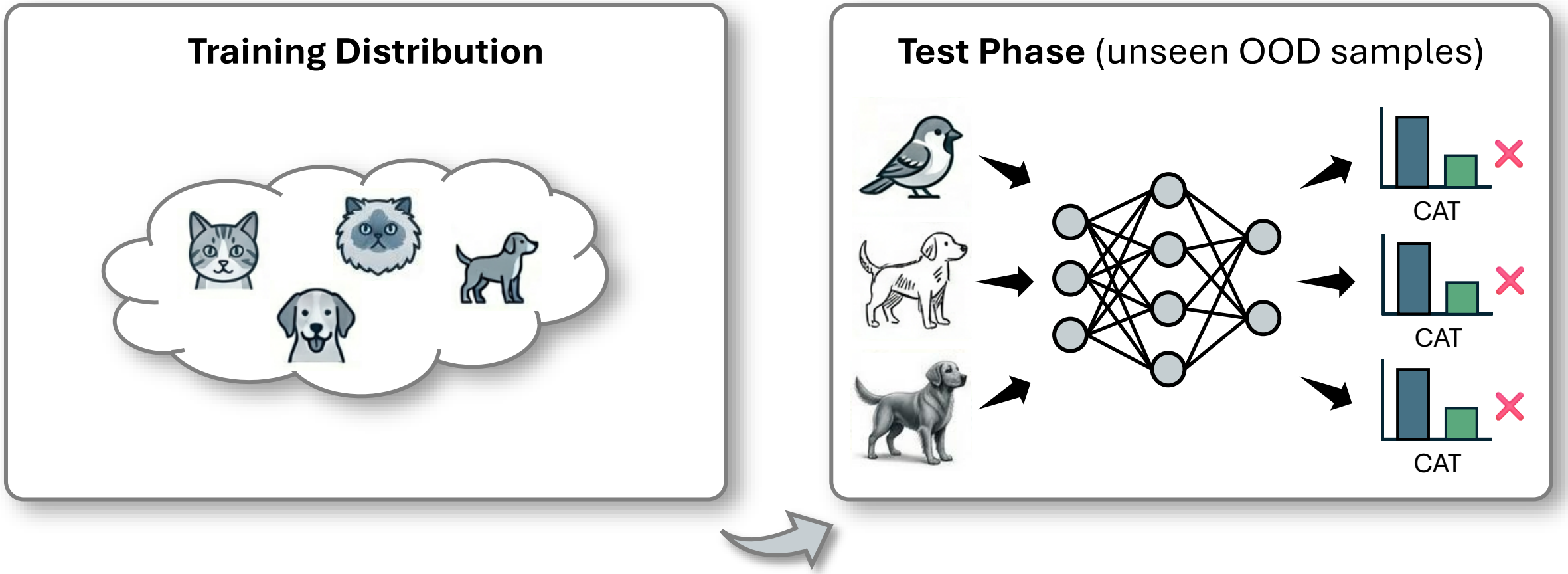
Model Generalization

Machine learning models are trained to generalize from a finite dataset to new, unseen i.i.d. samples drawn from the same distribution, i.e., **in-distribution (ID) data**.



Distribution Shifts

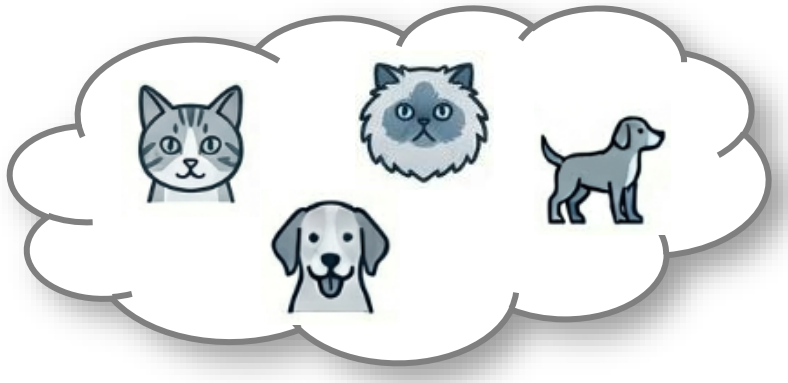
When test data distribution shifts, i.e., **out-of-distribution (OOD) data**, we can no longer guarantee the model performance (realistic dog and bird images are unseen).



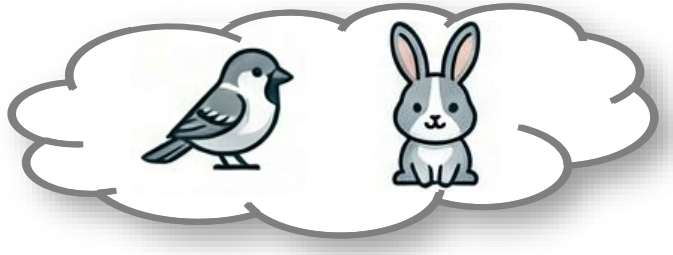
Distribution Shifts

When test data distribution shifts, i.e., **out-of-distribution (OOD) data**, we can no longer guarantee the model performance.

Training Distribution



Semantic Shift (different class, y shift)

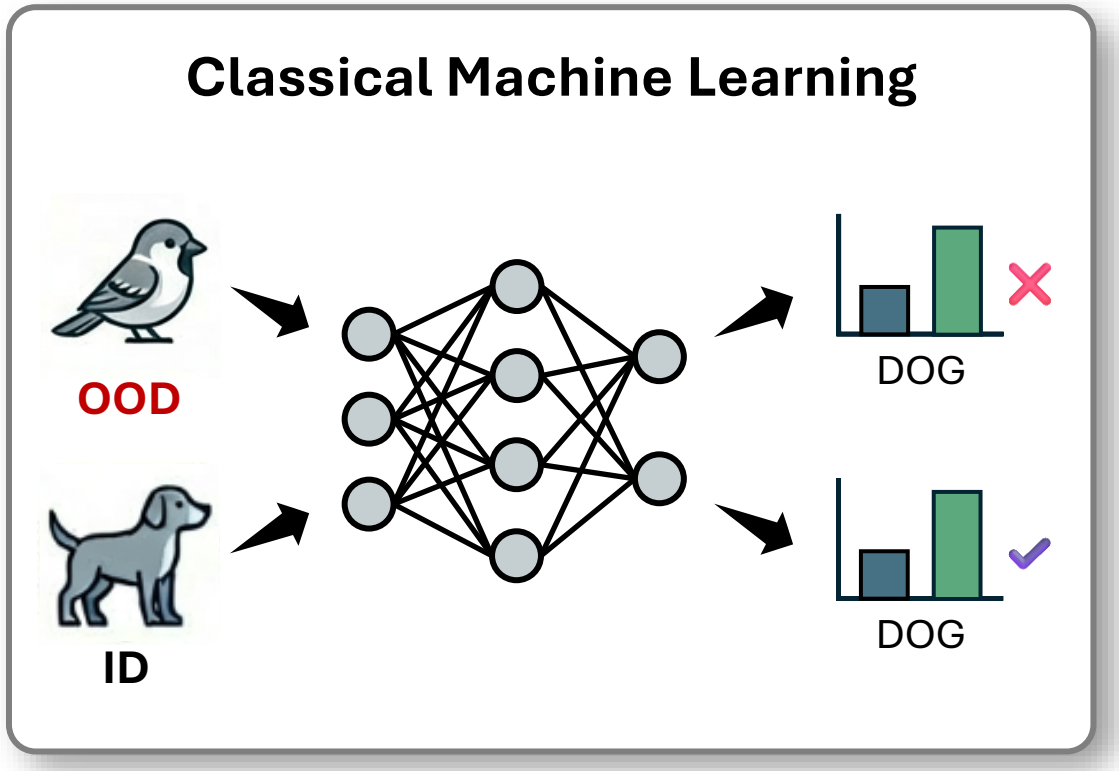


Covariate Shift (different style, x shift)

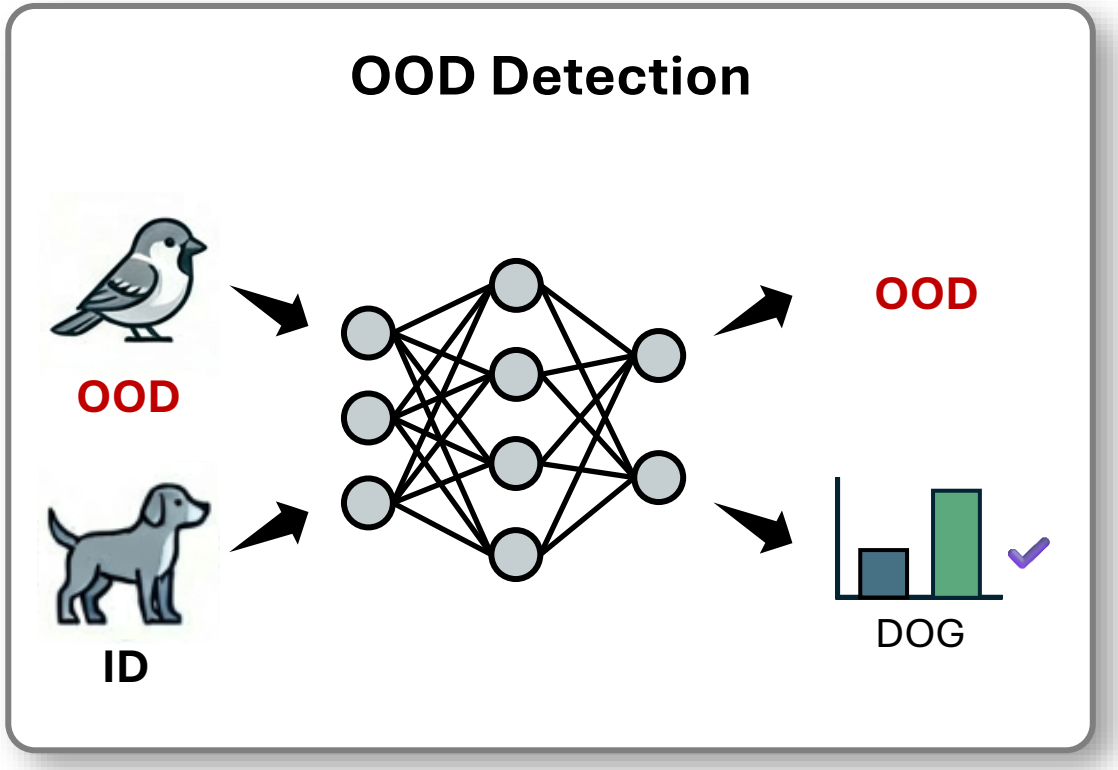
distribution shift

OOD Detection

Machine learning models should **detect semantic distribution shifts** and avoid making further label predictions (OOD generalization will consider covariate shifts).



No matter how strong the model is, **it will always make incorrect predictions**.



The model should **detect OOD data without further label predictions**.

OOD Detection in Autonomous Driving

Critical for road safety: Identifying **unknown scenarios or objects** to prevent accidents and ensure robustness.

Standard Vehicle (Known)

Horse-Drawn Carriage (OOD)

Fallen Tree on Road (OOD)

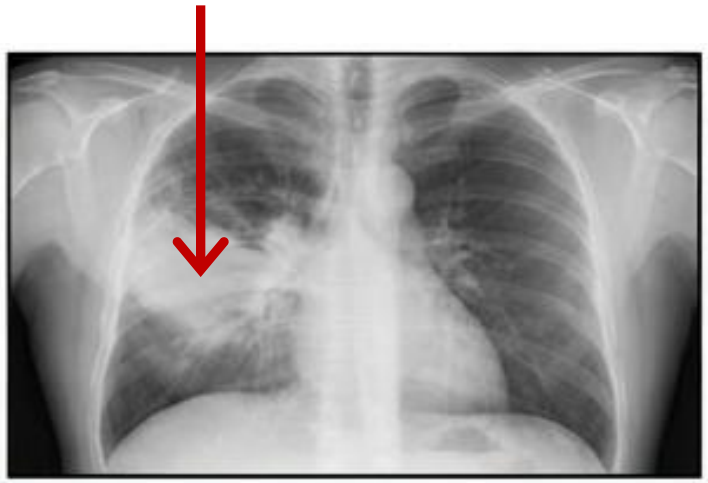
Models must recognize **novel objects** on the road to maintain safety boundaries.

The primary goal is to reliably flag OOD data, not to force **the vehicle prediction**.

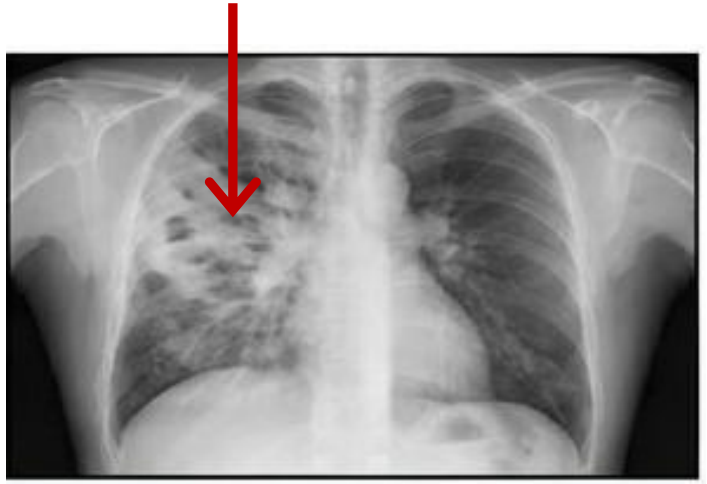
OOD Detection in Medical & Healthcare Systems

Critical for patient safety: Identifying **anomalies and unseen conditions** to prevent misdiagnosis and ensure reliability.

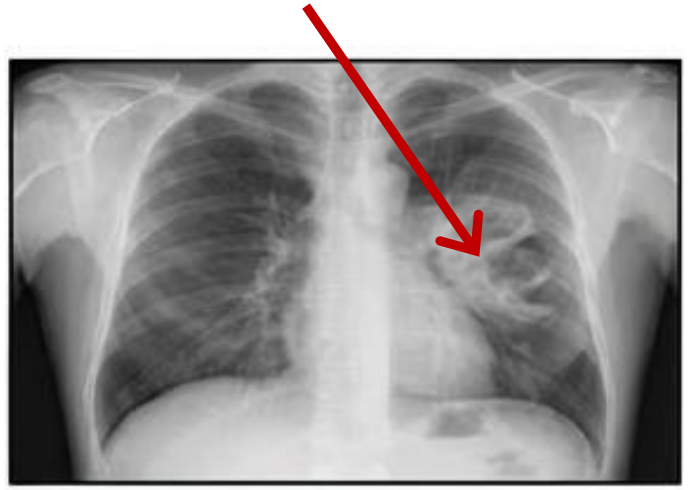
Common Pneumonia (Known)



Rare Tropical Disease (OOD)



Unseen Genetic Condition(OOD)



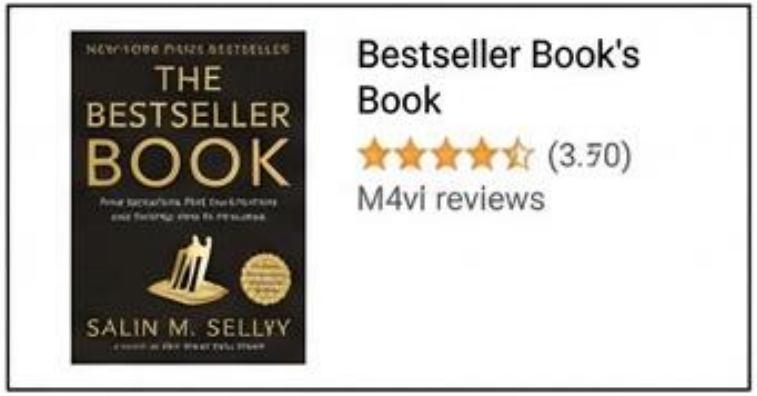
Models must recognize **unseen pathologies** to maintain the diagnostic integrity.

The primary goal is to reliably flag OOD cases for clinician review, not **automated diagnoses**.

OOD Detection in Recommender Systems

Critical for user trust: Identifying inputs **anomalous user-term interactions** to improve robustness and ensure fairness.

Popular Item (Known)



Niche Item (OOD)

Inappropriate Input (OOD)

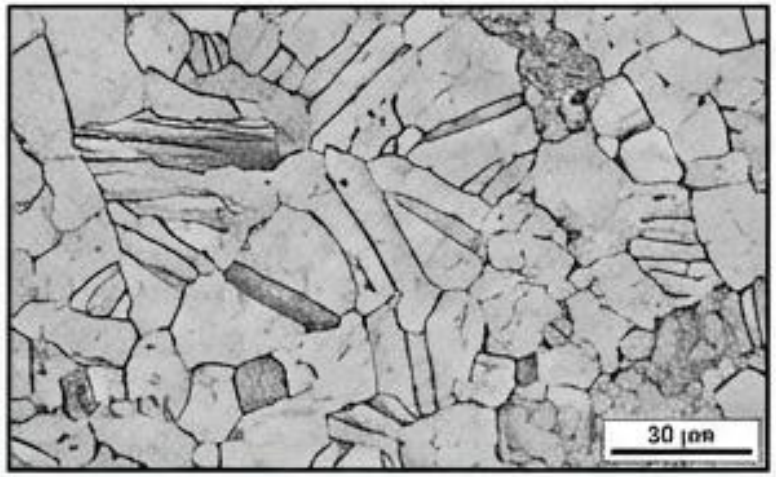
Models must recognize **novel user interests and item types** to maintain system integrity.

The primary goal is to reliably flag interactions for **review or adaptive recommendations.**

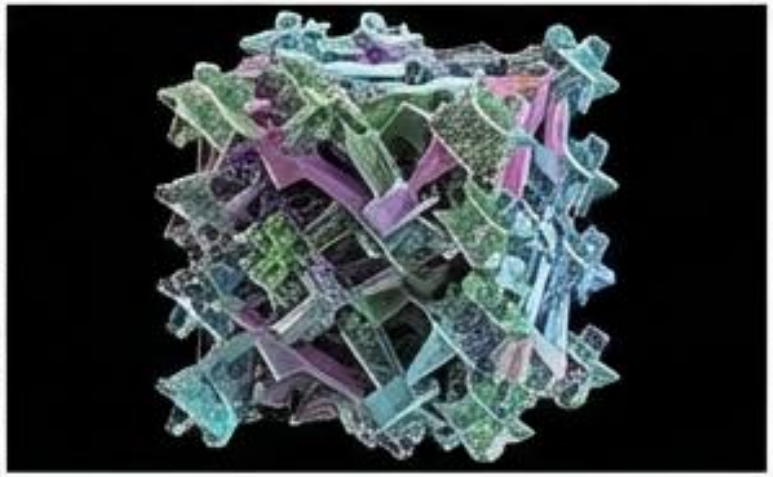
OOD Detection in Scientific Discovery

Critical for breakthroughs: Identifying anomalous data to **catalyse new theories and novel experiments**.

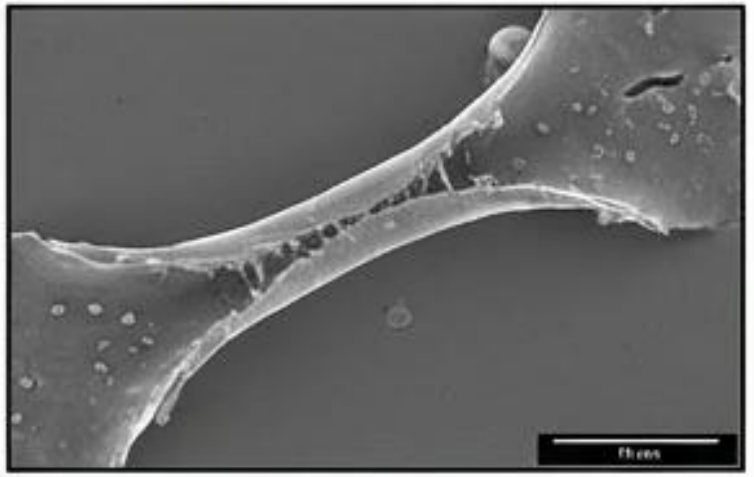
Common Alloy (Known)



Superconductor (OOD)



Bio-inspired Polymer (OOD)

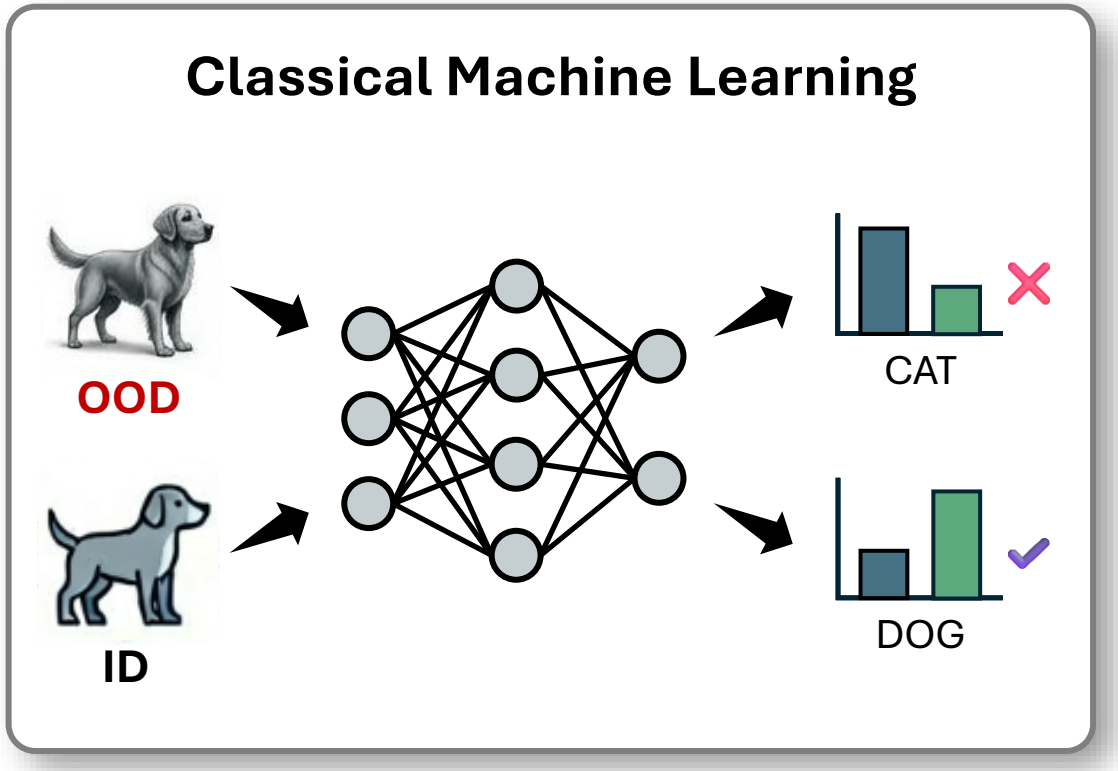


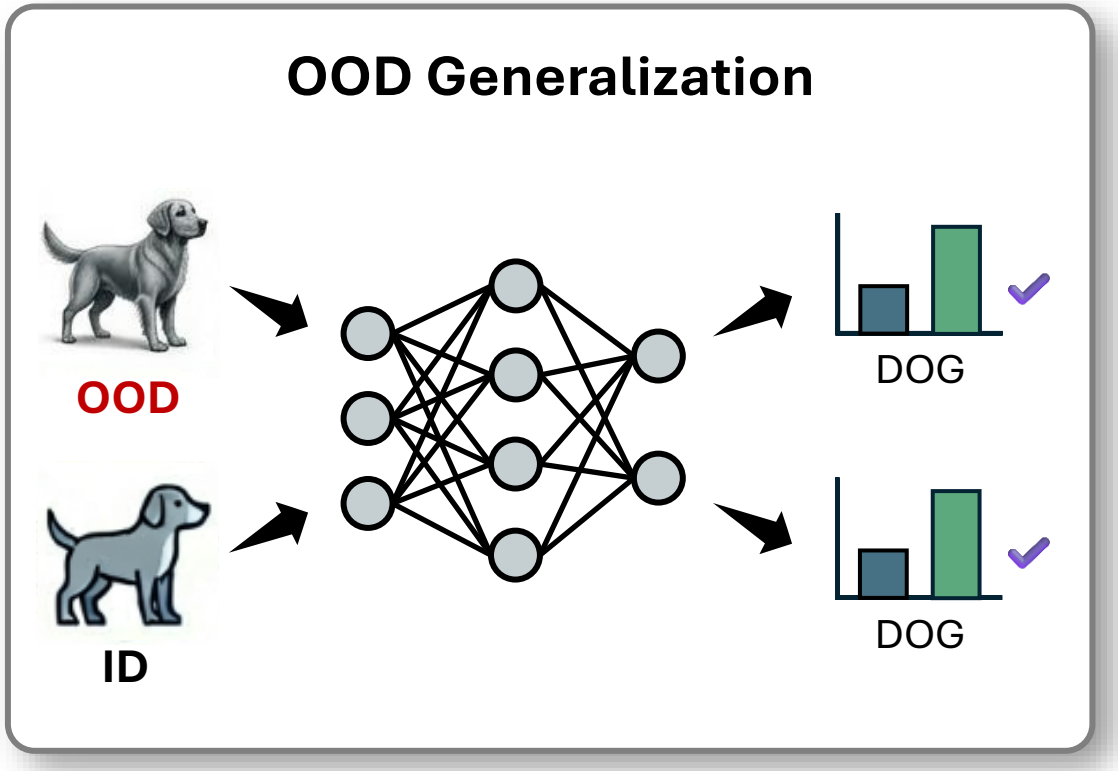
Models must flag **anomalous results and unexpected** patterns to guide research.

The primary goal is to highlight data for further investigation, not to force **fit existing categories**.

OOD Generalization

Machine learning models should **general well to data with covariate distribution shifts** and still produce correct label predictions.





If the model is strong enough, **it can make right predictions**.

The model should **generalize to unseen OOD data with right label predictions.**

OOD Generalization in Autonomous Driving

Beyond OOD detection that merely flags unseen objects, autopilot needs to generalize to **unseen scenarios** to prevent accidents and ensure safety under different conditions.

Standard Vehicle (Known)

Horse-Drawn Carriage (OOD)

Fallen Tree on Road (OOD)

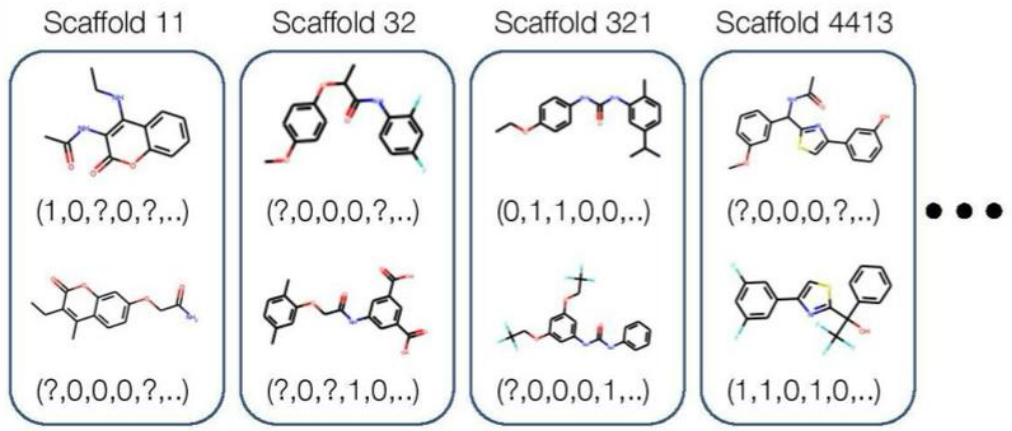
Training Environment

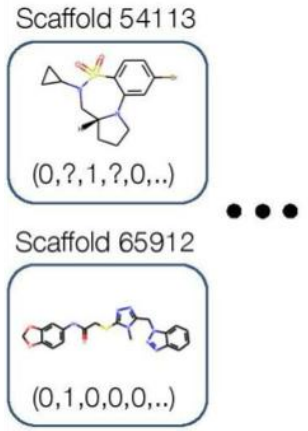
OOD Test Environments

Waymo Open Challenge

OOD Generalization in Drug Discovery

Critical for drug discovery: During the screening of candidate drugs, identifying **critical functional groups** that characterize the important biochemical properties of drugs, and avoid **spurious correlations of scaffolds** that take a large part of the molecule.





MolPCBA in wilds benchmark

ID molecules

OOD molecules with different scaffolds

OOD Generalization in Fairness

Critical for satellite imagery: In the analysis of images taken for different regions and subpopulations, immune to the spurious correlations associated with demographics.

Train			Test	
				
2002 / Americas	2009 / Africa	2012 / Europe	2016 / Americas	2017 / Africa
shopping mall	multi-unit residential	road bridge	recreational facility	educational institution

FMoW in wilds benchmark

Building/Land Type Classification

	Train			Test	
Satellite image (x)					
Country / Urban-rural (d)	Angola / urban	Angola / rural	Angola / urban	Kenya / urban	Kenya / rural
Asset index (y)	0.259	-1.106	2.347	0.827	0.130

PovertyMap in wilds benchmark

Poverty Estimation

https://bhanml.github.io & https://github.com/tmlr-group

OOD Generalization in Multimodal Alignment

Critical for hallucination mitigation of multimodal foundation models: Aligning concepts from multiple modalities and avoid multimodal spurious correlations.

Ice Bear in Snow (common) CLIP ACCU: 80.25

Ice Bear in Grass (counter) CLIP ACCU: 9.17

Tutorial Organization

30 mins — **Opening, Background, and Key Insights** (Prof. Bo Han)
Introducing the tutorial goals and real-world motivations.

1 hour — **Post-hoc OOD Detection** (Dr. Jianing Zhu)
*Perspectives on enhance OOD detection through better **logits**, **embedding**, and **distribution**.*

15 mins

1 hour — **OOD Detection Learning** (Dr. Qizhou Wang)
*Perspectives on enhance OOD detection through better **representation**, **calibration**, and **classification**.*

15 mins

1 hour — **OOD Generalization** (Dr. Yongqiang Chen)
*Perspectives on enhance OOD generalization through causality, from better **representation**, **optimization**, to **benchmarking**.*

30 mins — **Q&A**

About Me

Prof. Bo Han

- **Current Roles**

  Associate Professor in Machine Learning and **Director** of TMLR group at HKBU.

  BAIHO Visiting Scientist of Imperfect Information Learning Team at RIKEN AIP, hosted by Prof. Masashi Sugiyama.

- **Professional Service & Recognition**

  Senior Area Chair for NeurIPS and ICML, **Area Chair** for ICLR, UAI, and AISTATS.

  Associate Editor for IEEE TPAMI, MLJ, and JAIR, **Editorial Board** for JMLR and MLJ.

  ACM Distinguished Speaker and **IEEE Senior Member**.

About Members

- **Dr. Jianing Zhu**

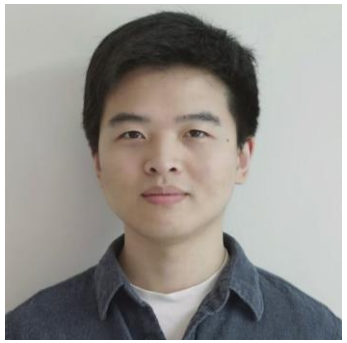
  Postdoctoral Fellow in the VITA group at UT Austin.

  Achieved PhD from the TMLR Group at HKBU.

- **Dr. Qizhou Wang**

  Postdoctoral Fellow in the Imperfect Information Learning Team at RIKEN AIP, working with Prof. Masashi Sugiyama.

  Achieved PhD from the TMLR Group at HKBU.



- **Dr. Yongqiang Chen**

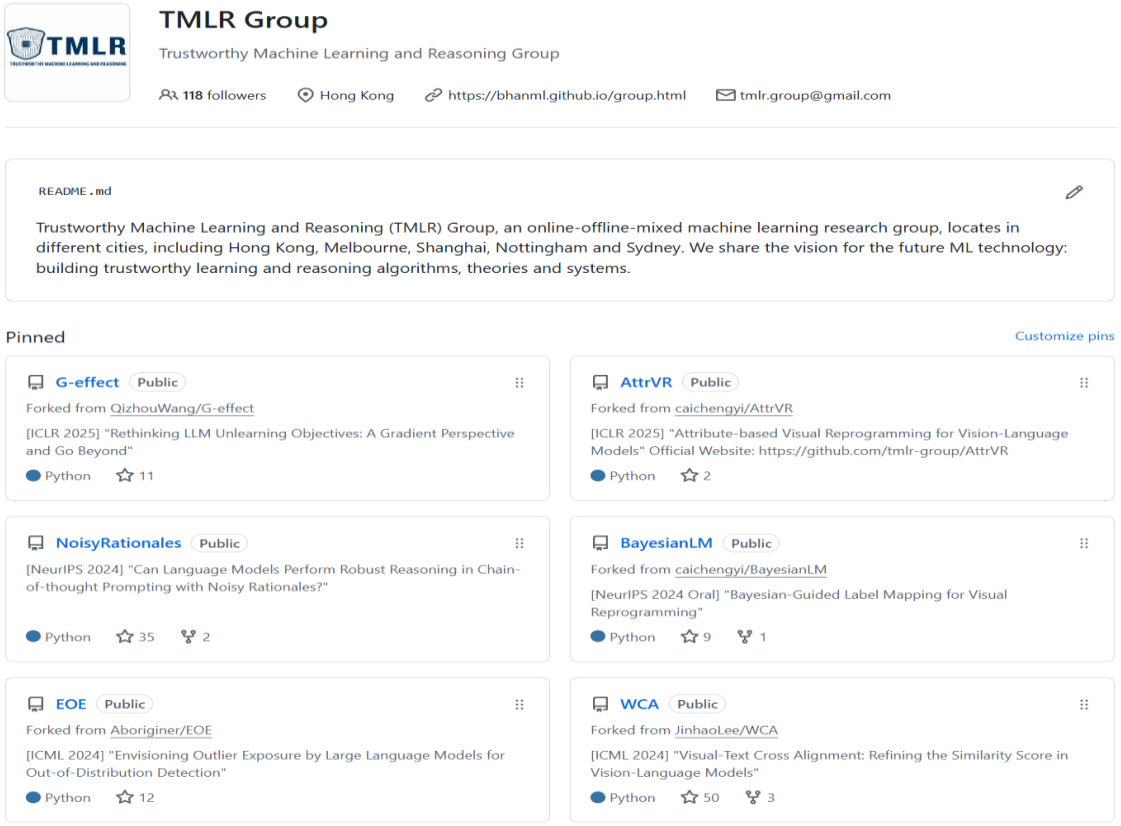
  Postdoctoral Fellow at CLeaR Group with Prof. Kun Zhang.

  Achieved PhD in CSE at CUHK in 2024, advised by Prof. James Cheng.

TMLR Group

TMLR Group, an online-offline-mixed **machine learning** research group, locates in different cities, including Hong Kong, Melbourne, Shanghai, Nottingham and Sydney.

We are welcoming the **synergetic collaboration** between yours and HKBU TMLR!!



- Research Twitter:
  - https://x.com/tmlrgroup
- Research RedNote:
  - https://www.xiaohongshu.com/user/profile/646ee4b900000000110010b6
- Research Blog:
  - https://www.jiqizhixin.com/columns/TMLRGroup

https://bhanml.github.io & https://github.com/tmlr-group

Thank you for attending!

🔗 https://bhanml.github.io

🔗 https://github.com/tmlr-group