



面向未知世界的机器学习

赵知临

个人介绍

赵知临 副教授/博士生导师 国家级高层次青年人才

- ❑ 2012.09 - 2016.06 中山大学 计算机学院 学士
- ❑ 2016.09 - 2018.06 中山大学 计算机学院 硕士
- ❑ 2018.08 - 2022.08 悉尼科技大学 工程与信息技术学院 博士
- ❑ 2022.08 - 2024.12 悉尼科技大学 / 麦考瑞大学 博士后



代表性成果
(第一作者)



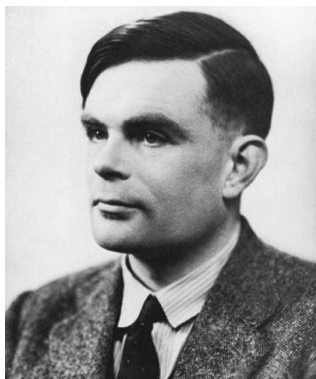
期刊与会议名称	篇数
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)	3
Annual Conference on Neural Information Processing Systems (NeurIPS)	2
Artificial Intelligence Journal (AIJ) / Machine Learning Journal (MLJ)	2
IEEE Transactions on Neural Networks and Learning Systems (TNNLS)	2
Association for the Advancement of Artificial Intelligence (AAAI)	1
International Joint Conference on Artificial Intelligence (IJCAI)	1
Transactions on Machine Learning Research (TMLR)	1

人工智能中“未知世界”的探索历程



1950. 智能判别

图灵提出图灵测试，首次用“在未知输入下的反应”来定义机器智能。



1988. 不确定性建模

Pearl系统化提出贝叶斯网络，为AI应对不完全信息和因果未知性提供数学工具。

1995. 开放世界假设

在知识表示领域提出“开放世界假设”，标志AI开始面向真实世界的知识建模。

2002. 开放集识别

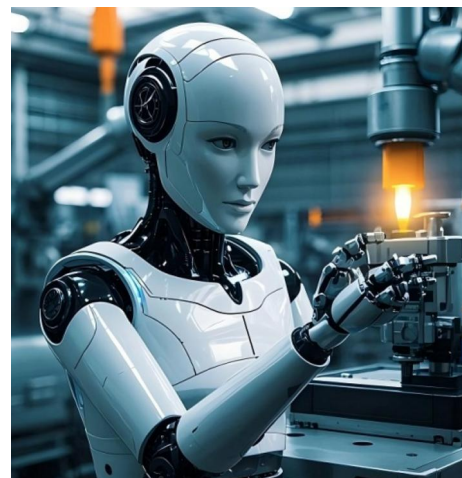
开放集识别概念提出，机器学习开始处理“未见类别”的识别问题。

2016. 外分布检测

深度学习下的外部分布检测方法兴起，揭示AI系统对未知数据缺乏鲁棒性。

2025. 开放环境泛化

面向具身智能与多智能体协作，AI迈向能在开放世界中自主适应与泛化的新时代。



探索未知世界的智能学习闭环框架



高风险场景下的智能泛化应用

在天气预测、医学诊断、具身智能等不确定性强的任务中验证模型的鲁棒性与安全性。



外分布机器学习理论

聚焦未知分布的界定、不确定性建模与泛化理论，奠定面向开放世界学习的理论基础。

大模型驱动的泛化学习算法

发展生成模型、多智能体机制与大模型指令泛化方法，实现面向未知任务的自主建模与推理。

理论机制建构

● 外分布界定

构建形式化测度方法，明确训练与测试分布之间的偏移类型和边界。

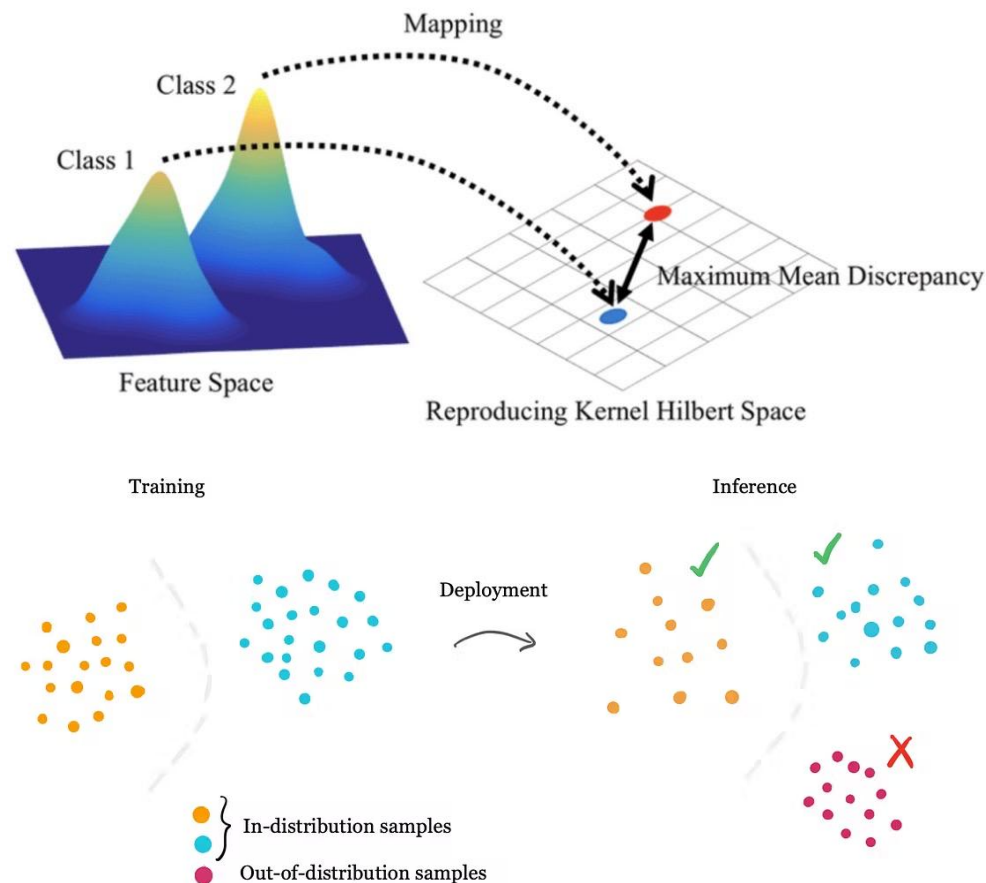
● 外分布检测

设计判别准则与不确定性指标，实现未知样本与异常输入的识别。

● 外分布利用

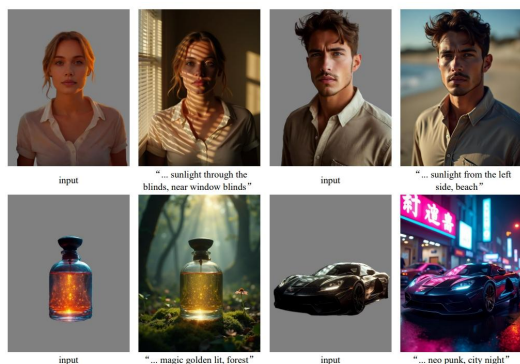
探索利用外分布样本提升模型泛化与迁移能力的理论机制。

目标：刻画未知世界的数学结构与推理边界，为模型设计与学习算法提供基础认知框架。



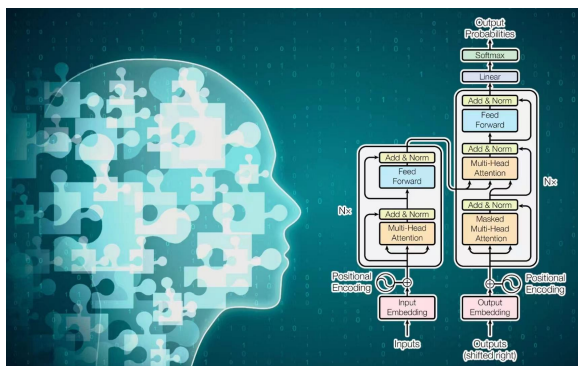
算法模型设计

目标：构建可感知未知、可泛化任务、可协作适应的智能学习系统。



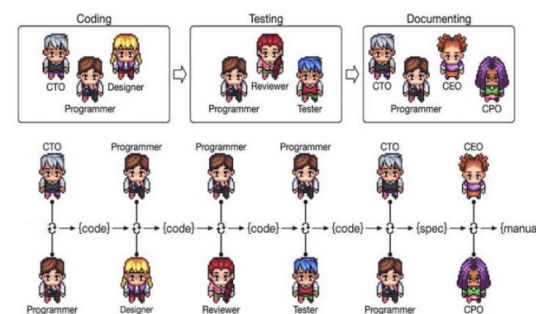
生成式内容生成

利用扩散模型等生成方法对未知分布建模，支持异常检测、类别外推与数据增强。



大模型的任务泛化

通过大语言模型与多模态模型实现任务重组与知识迁移，提升系统对未知任务的理解与响应能力。

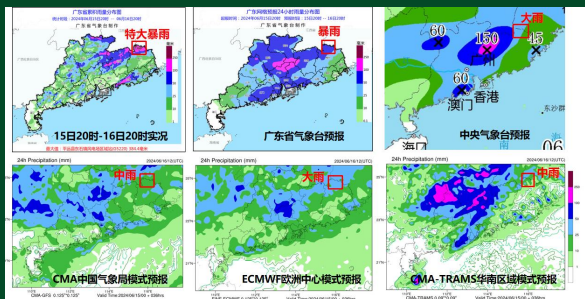


多智能体协作与对抗

设计具备协作、对抗与语言通信能力的智能体群体，实现开放环境下的自主适应与任务泛化。

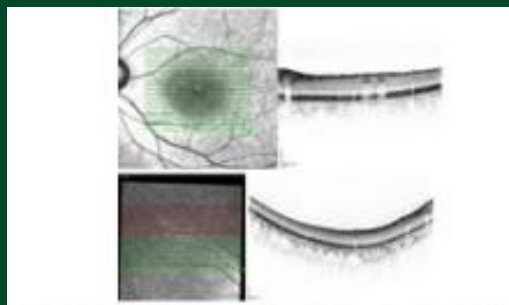
关键应用验证

目标：在高不确定性、强开放性的现实任务中验证理论与算法的有效性与鲁棒性。



极端天气预测

泛化强、具不确定性提示的气象模型，用于早期极端天气的预警与策略调整。



医学诊断泛化

高置信辅助诊断系统，应对跨设备、跨人群、跨病种不一致问题，提供可靠异常提示。



具身智能迁移

基于多模态感知与多智能体协作，实现开放环境中的任务适应与策略泛化。



欢迎加入
iSEE Lab

