Domain Generalisation via Imprecise Learning

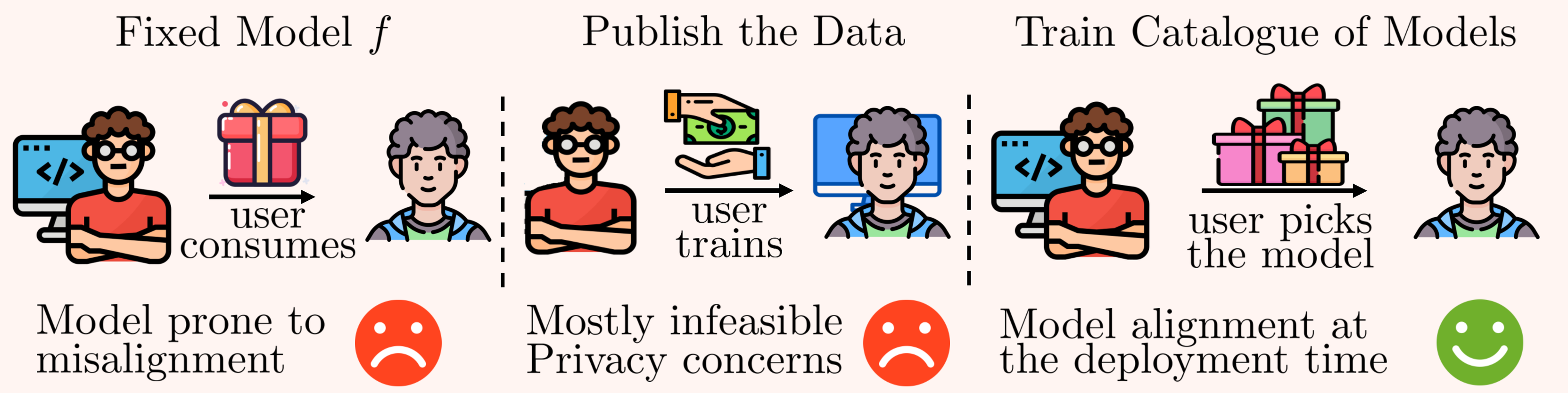
Anurag Singh [1] Siu Lun Chau [1] Shahine Bouabid [2] Krikamol Muandet [1]

[1]Rational Intelligence Lab, CISPA Helmholtz Center for Information Security, Saarbrücken, Germany [2]Department of Statistics, University of Oxford, Oxford, UK

ICML 24 Spotlight

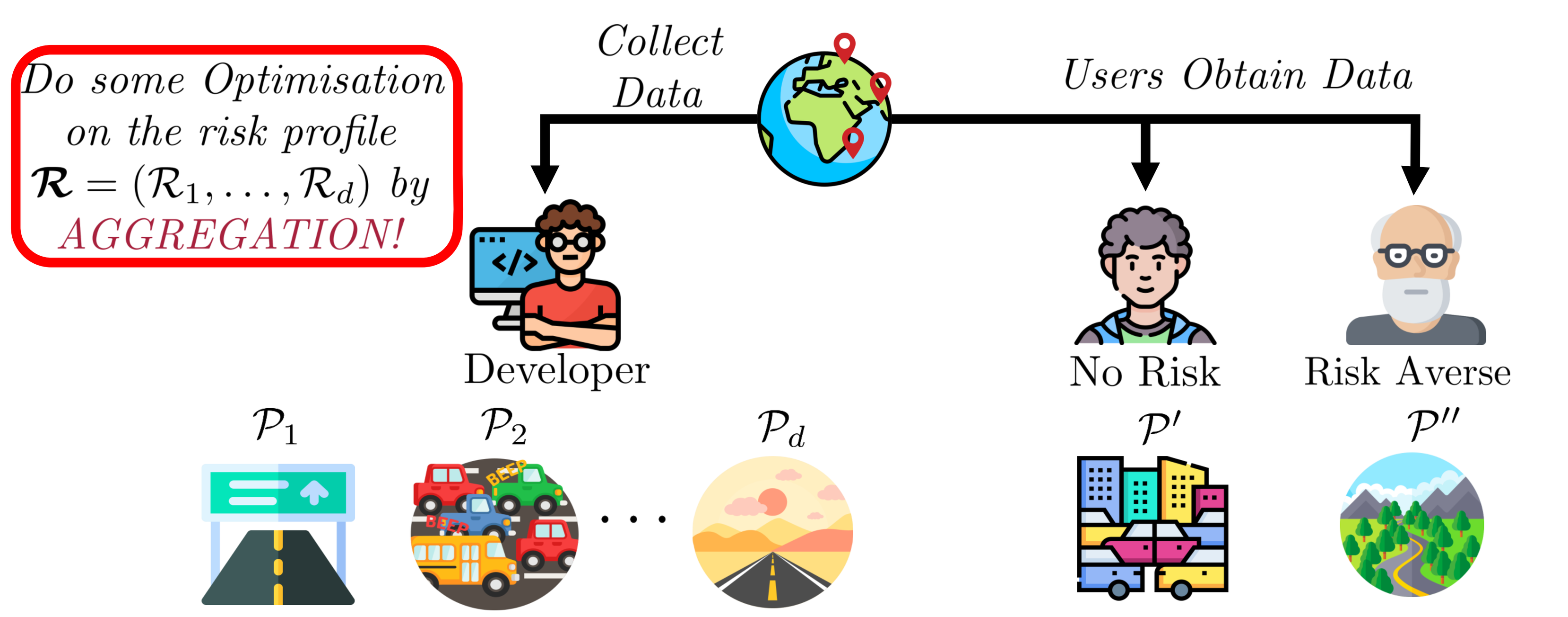
TL;DR

Users should pick the aligned model at deployment time!

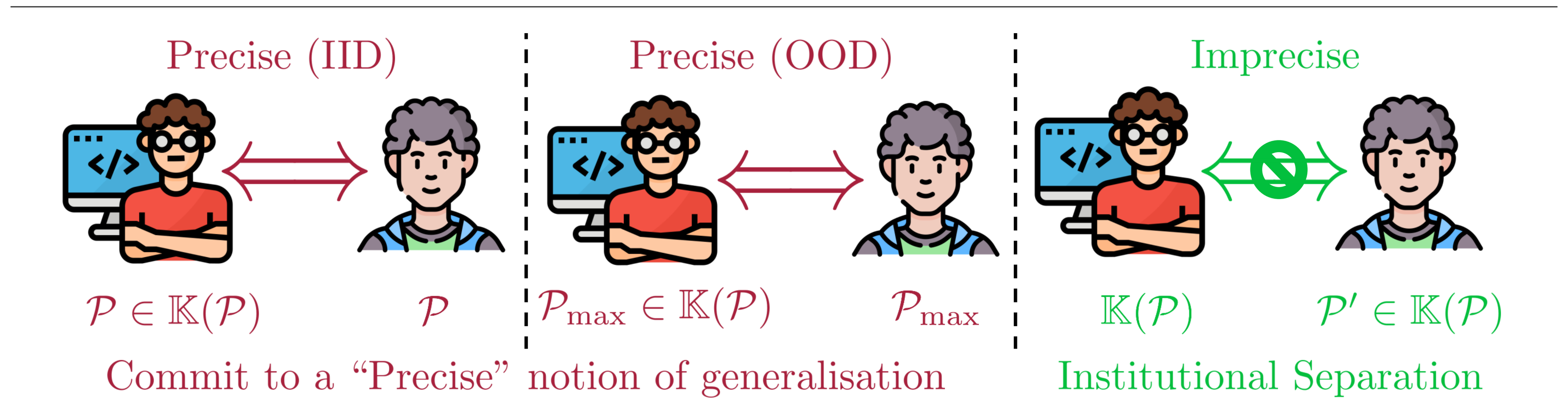


Domain Generalisation

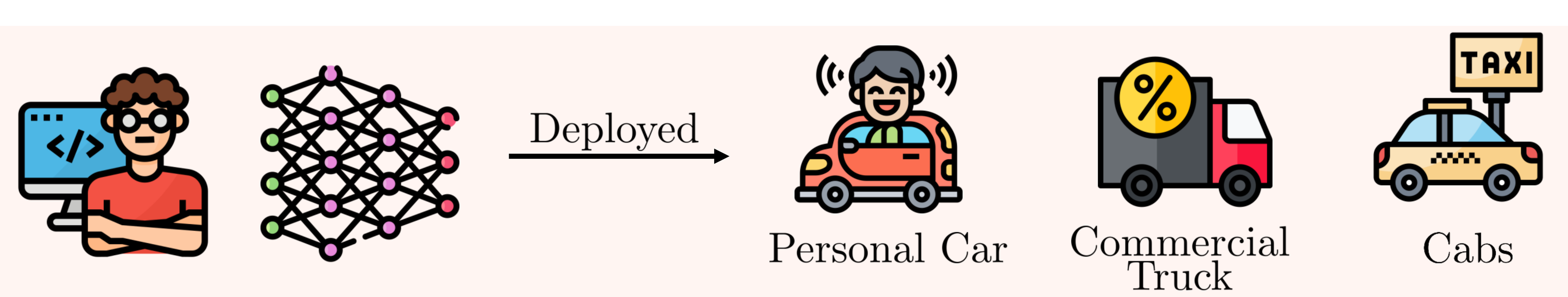
Given risk functional $\mathcal{R}: f \to \mathbb{E}_{\mathcal{P}}[\ell(f(X), y)]$, we learn Bayes optimal $f_{Bayes} := \arg\min_{f \in \mathcal{F}} \mathcal{R}(f)$



Why should we care about decision-making?



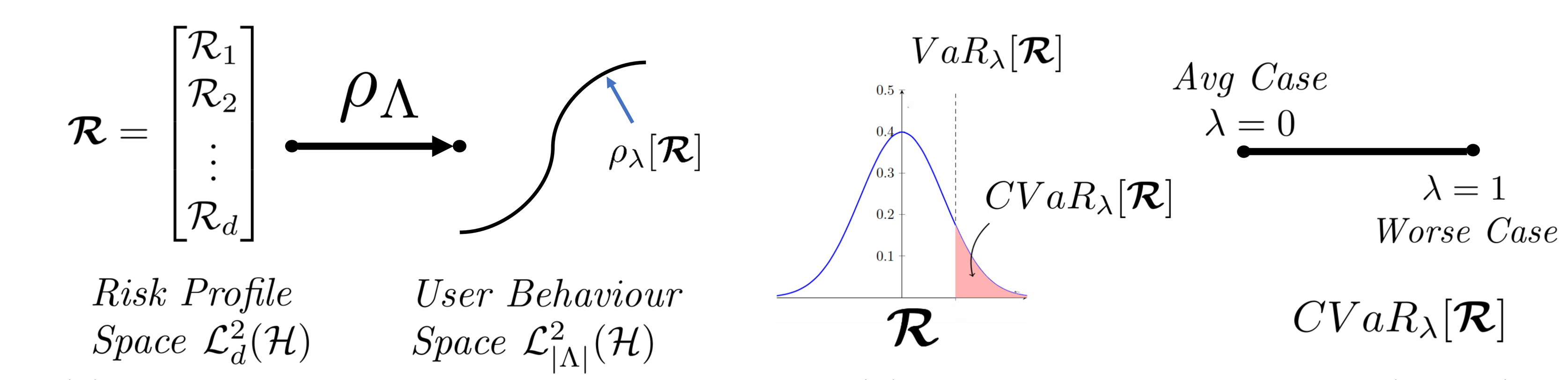
Thought Experiment



Should model act same in all vehicles in case of an accident?

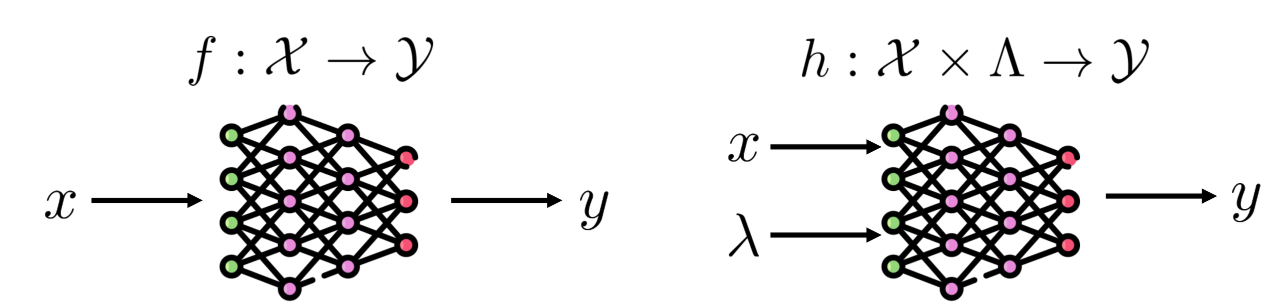
Problem Formulation

Aggregation Functions: Map risk profile $\mathcal{R}$ to objectives in user choice space Λ



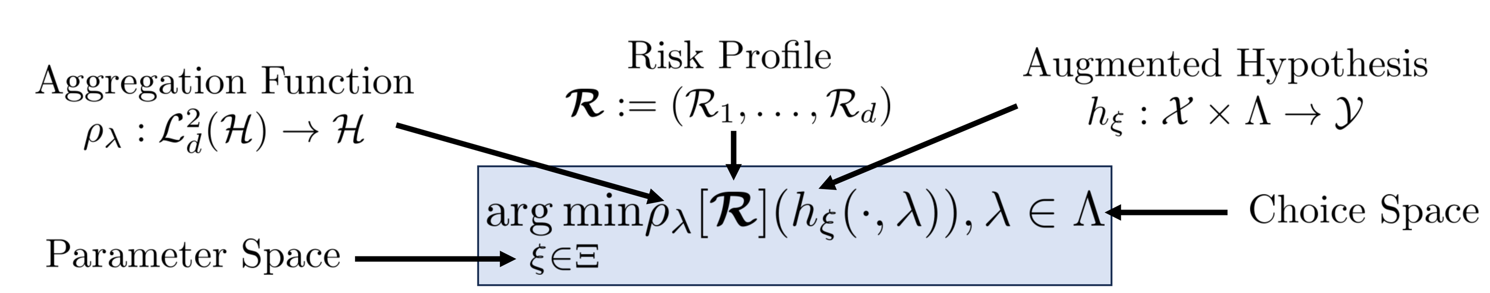
(a) Aggregation Function as Mapping (b) Conditional Value at Risk (CVaR)

Augmented Hypothesis: Conditions model on user choice space Λ



Imprecise Risk Optimisation (IRO)

Characterising the Optimal Catalogue:



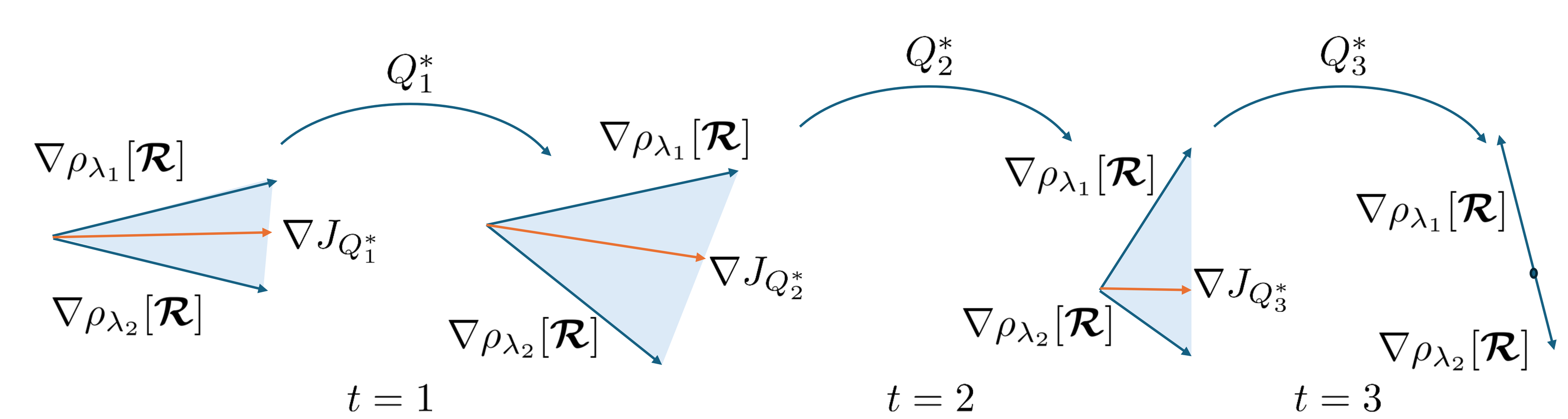
C-Pareto Optimality
$\rho_\lambda[\boldsymbol{\mathcal{R}}](h_1(\cdot, \lambda)) \le \rho_\lambda[\boldsymbol{\mathcal{R}}](h_2(\cdot, \lambda))$ for all $\lambda \in \Lambda$

Scalarised Objectives
$J_Q(h) = \mathbb{E}_{\lambda \sim Q}\left[\rho_\lambda[\boldsymbol{\mathcal{R}}](h(\cdot, \lambda))\right]$ where $Q \in \Delta(\Lambda)$

We pick distribution Q^* which does **C-Pareto Improvement**. See, MGDA (Desideri, 2012)

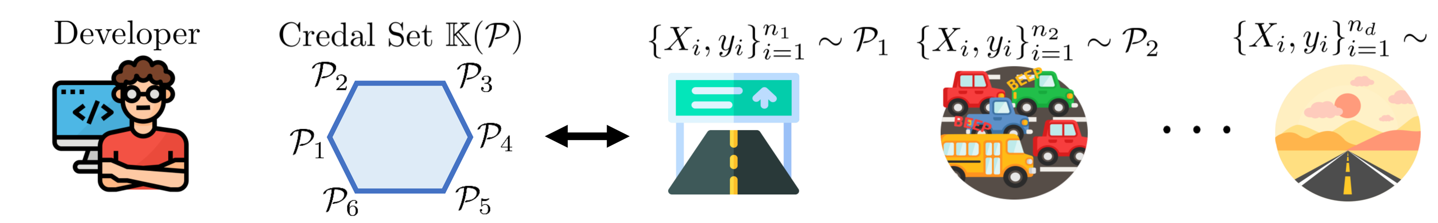
$$Q^*_t \in \arg\min_{Q \in \Delta(\Lambda)} \left\| \nabla_{\xi_{t-1}} J_Q(h_{\xi_{t-1}}) \right\|_2 \quad h_{\xi'} := h_\xi - \nabla_{\xi_t} J_{Q^*_t}(\xi_t)$$

Can Q^*_t not just be uniform?

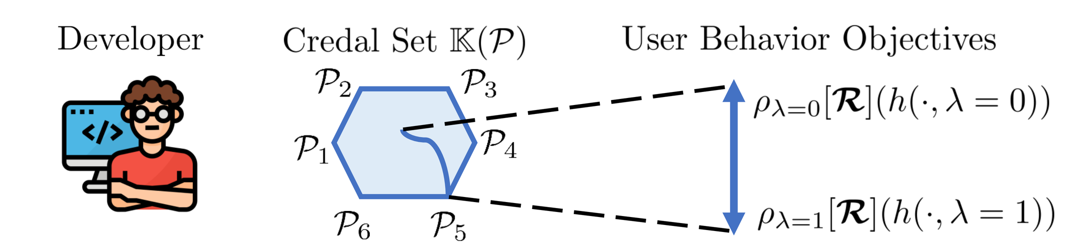


Summary of Imprecise Learning Framework

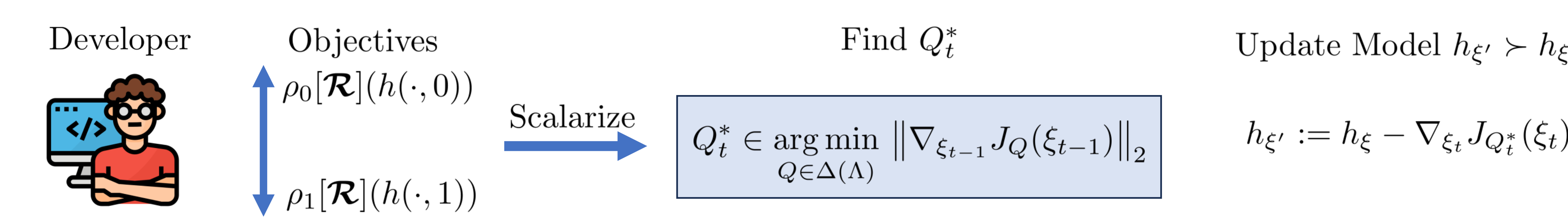
Step 1: Developer represents their uncertainty with credal set



Step 2: Map credal set to user behaviour choice space Λ and pick hypothesis class $\mathcal{H} \subseteq \mathcal{H}_\Lambda$



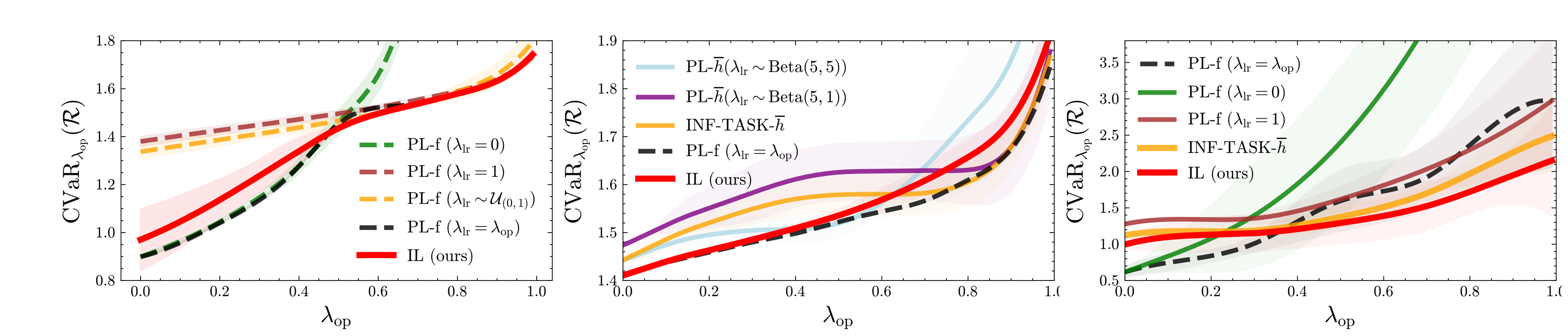
Step 3: Find $Q^*_t \in \Delta(\Lambda)$ that does Pareto Improvement for model update



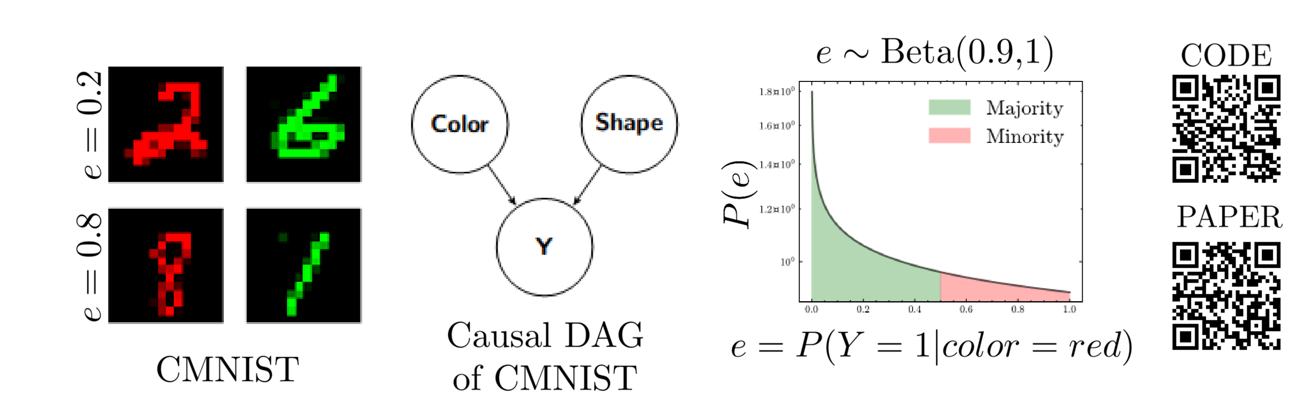
Step 4: At deployment users can consume model $h(\cdot, \lambda)$ with their choice of $\lambda \in \Lambda$

Experiments and Simulations

Imprecise vs Precise Learners under Institutional Separation:



Comparison with other Domain Generalisation Methods:



Objective	Algorithm	Test Environments based on $(Y = 1 \mid \text{color} = \text{red}) = e$ 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Regret
Average Case	ERM	96.1	87.1	78.0	72.1	65.8	59.2	51.8	47.1	39.9	33.6	28.3	72.7
Worse Case	GrpDRO	54.1	55.6	58.1	595	61.5	64.5	66.3	69.1	70.5	73.9	75.5	46.9
Invariance	IRM	72.0	72.0	72.0	72.0	72.1	69.7	69.3	69.9	69.2	69.7	67.7	32.3
Imprecise	**IRO (Ours)**	**95.8**	**87.2**	**78.8**	**68.9**	**69.4**	**69.5**	70.8	**70.1**	**70.0**	**70.4**	**70.3**	29.7

anurag.singh@cispa.de