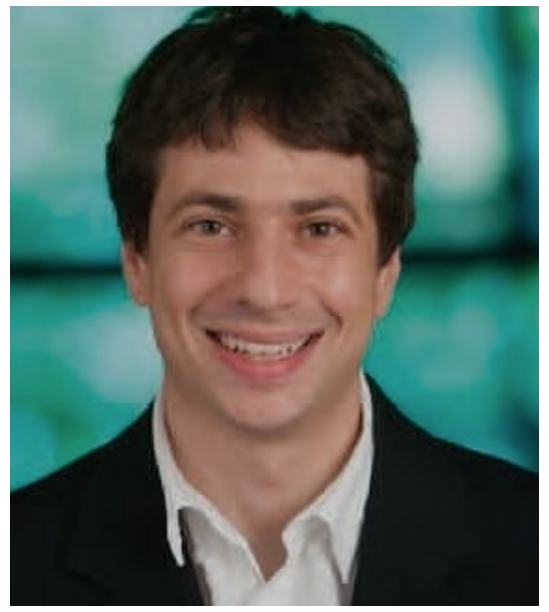
Decision-Theoretic Foundations of Conformal Prediction:

Optimal Uncertainty Quantification for Risk-Averse Agents

Hamed Hassani

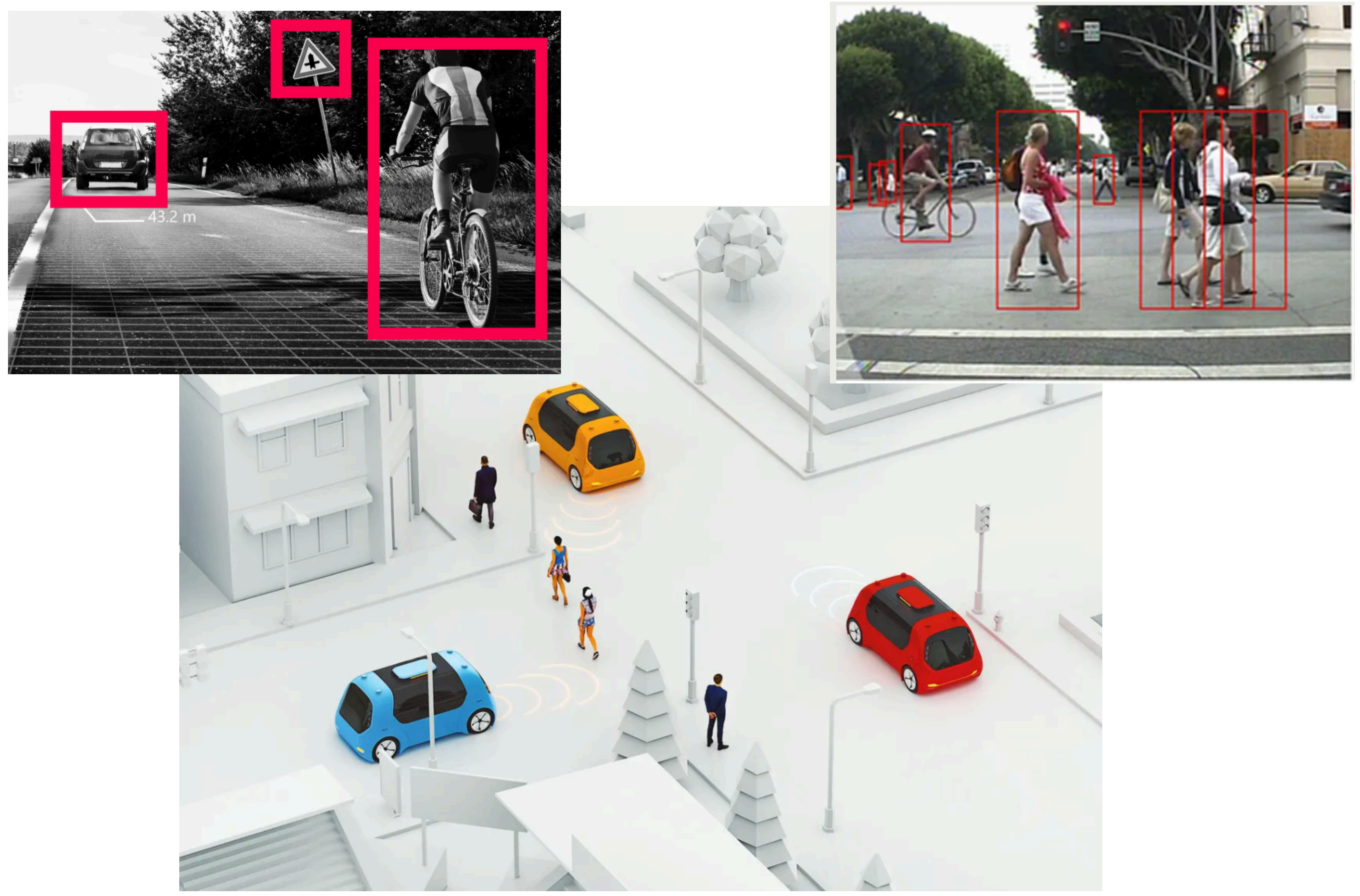
University of Pennsylvania

Joint work with: Shayan Kiyani, George Pappas, Aaron Roth

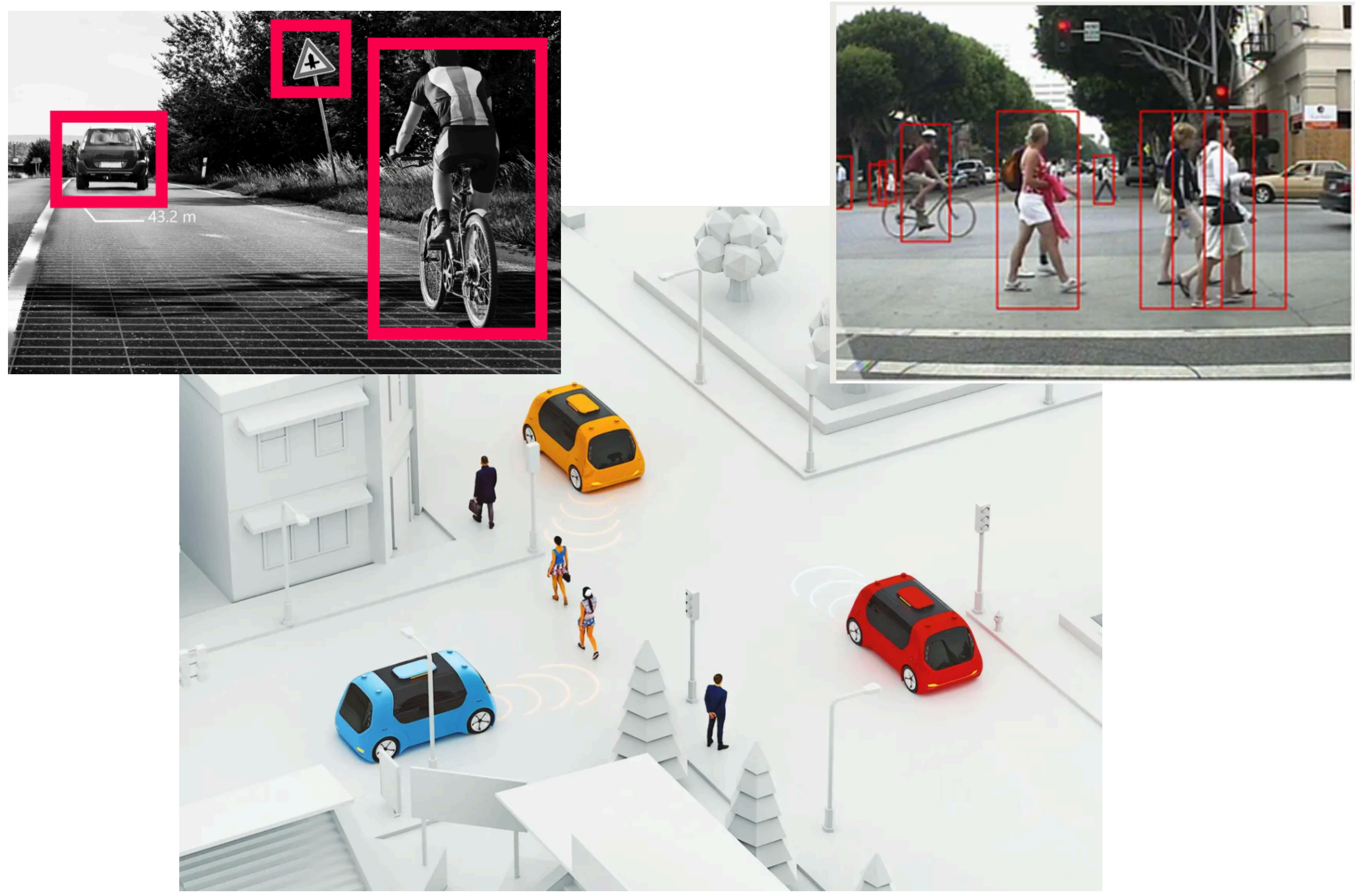


Predictions Shape Decisions

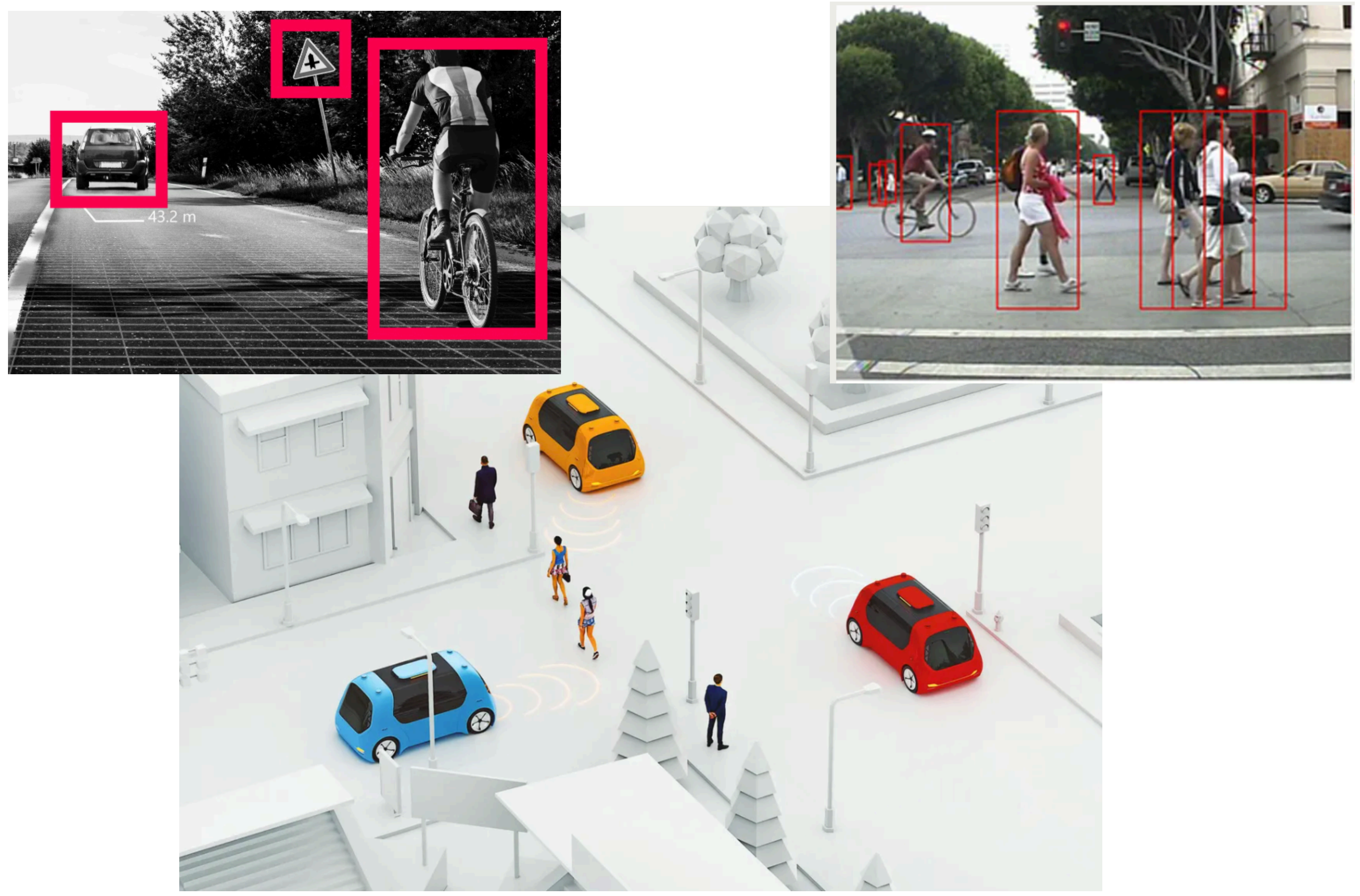
Predictions Shape Decisions



Predictions Shape Decisions

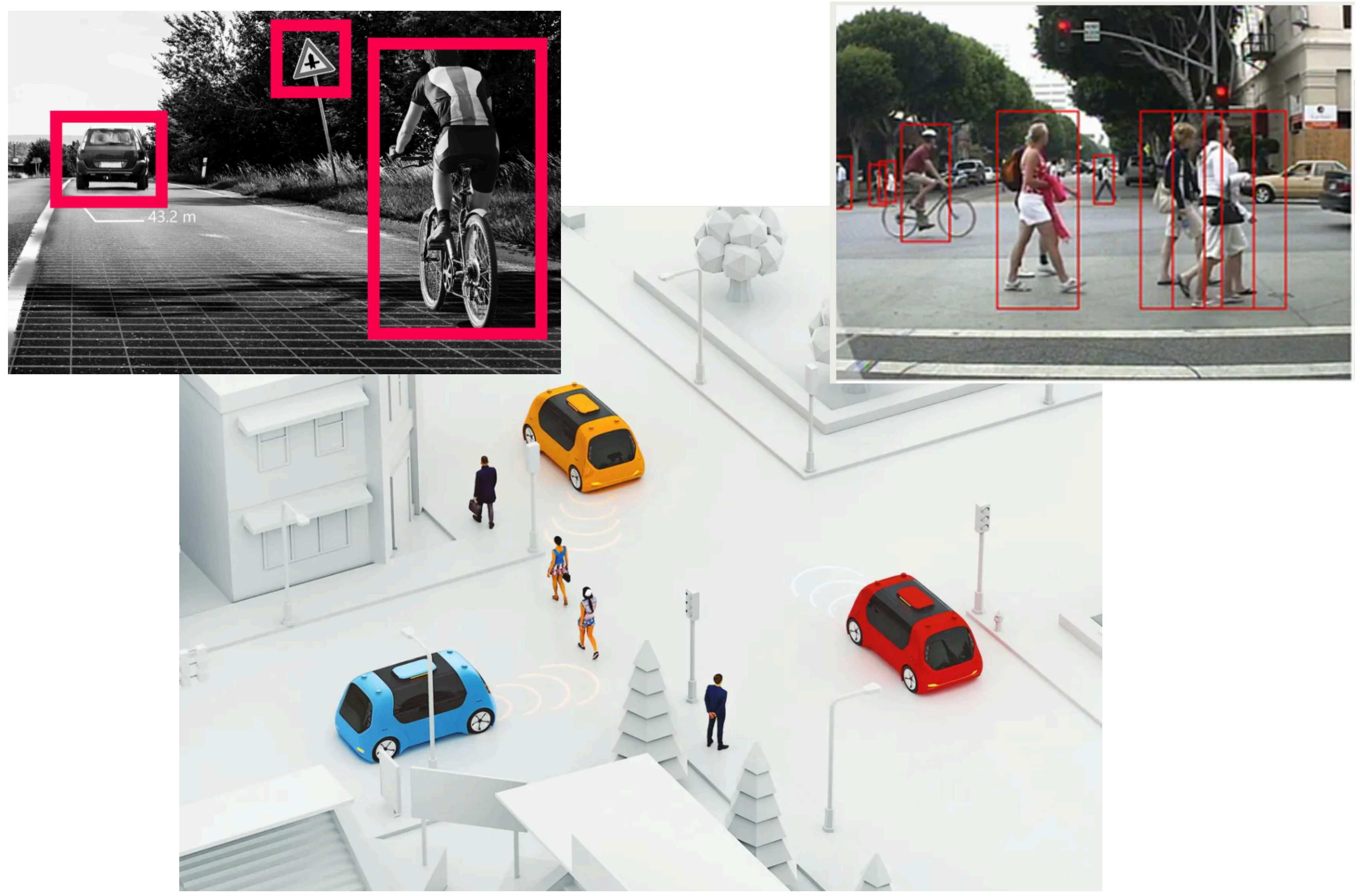


Predictions Shape Decisions



Predictions ⟶ Actions

Predictions Shape Decisions



Predictions $\longrightarrow$ Actions

Can we trust the predictions, and hence, the subsequent actions?

The Need for Precise Uncertainty Quantification (UQ)

The Need for Precise Uncertainty Quantification (UQ)

Tesla's robotaxi push hinges on 'black box' AI gamble

By **Norihiko Shirouzu** and **Chris Kirkham**

October 10, 2024 3:41 PM EDT · Updated 5 months ago

REUTERS

The Need for Precise Uncertainty Quantification (UQ)

Tesla's robotaxi push hinges on 'black box' AI gamble

By **Norihiko Shirouzu** and **Chris Kirkham**

October 10, 2024 3:41 PM EDT · Updated 5 months ago

REUTERS

An inside conversation:

Former Tesla Engineer: "Our end-to-end ML system is a complete black box—we have no insight into how it makes predictions, so if something goes wrong, we can't pinpoint the issue."

AI Researcher: "Right, and while computer vision is generally accurate, studies show it still misses about 3% of objects, which could mean failing to detect a pedestrian."

Former Tesla Engineer: "Exactly. There's always the risk of missing one of the infinite 'edge cases' on the road, making safety hard to guarantee."

The Need for Precise Uncertainty Quantification (UQ)

Tesla's robotaxi push hinges on 'black box' AI gamble

By Norihiko Shirouzu and Chris Kirkham

October 10, 2024 3:41 PM EDT · Updated 5 months ago

REUTERS

Incorrect AI Advice Influences Diagnostic Decisions

System developers must consider how AI explanation might impact reliance on AI advice

RSNA Radiological Society of North America

An inside conversation:

Former Tesla Engineer: "Our end-to-end ML system is a complete black box — we have no insight into how it makes predictions, so if something goes wrong, we can't pinpoint the issue."

AI Researcher: "Right, and while computer vision is generally accurate, studies show it still misses about 3% of objects, which could mean failing to detect a pedestrian."

Former Tesla Engineer: "Exactly. There's always the risk of missing one of the infinite 'edge cases' on the road, making safety hard to guarantee."

The Need for Precise Uncertainty Quantification (UQ)

Tesla's robotaxi push hinges on 'black box' AI gamble

By Norihiko Shirouzu and Chris Kirkham

October 10, 2024 3:41 PM EDT · Updated 5 months ago

REUTERS

Incorrect AI Advice Influences Diagnostic Decisions

System developers must consider how AI explanation might impact reliance on AI advice

RSNA Radiological Society of North America

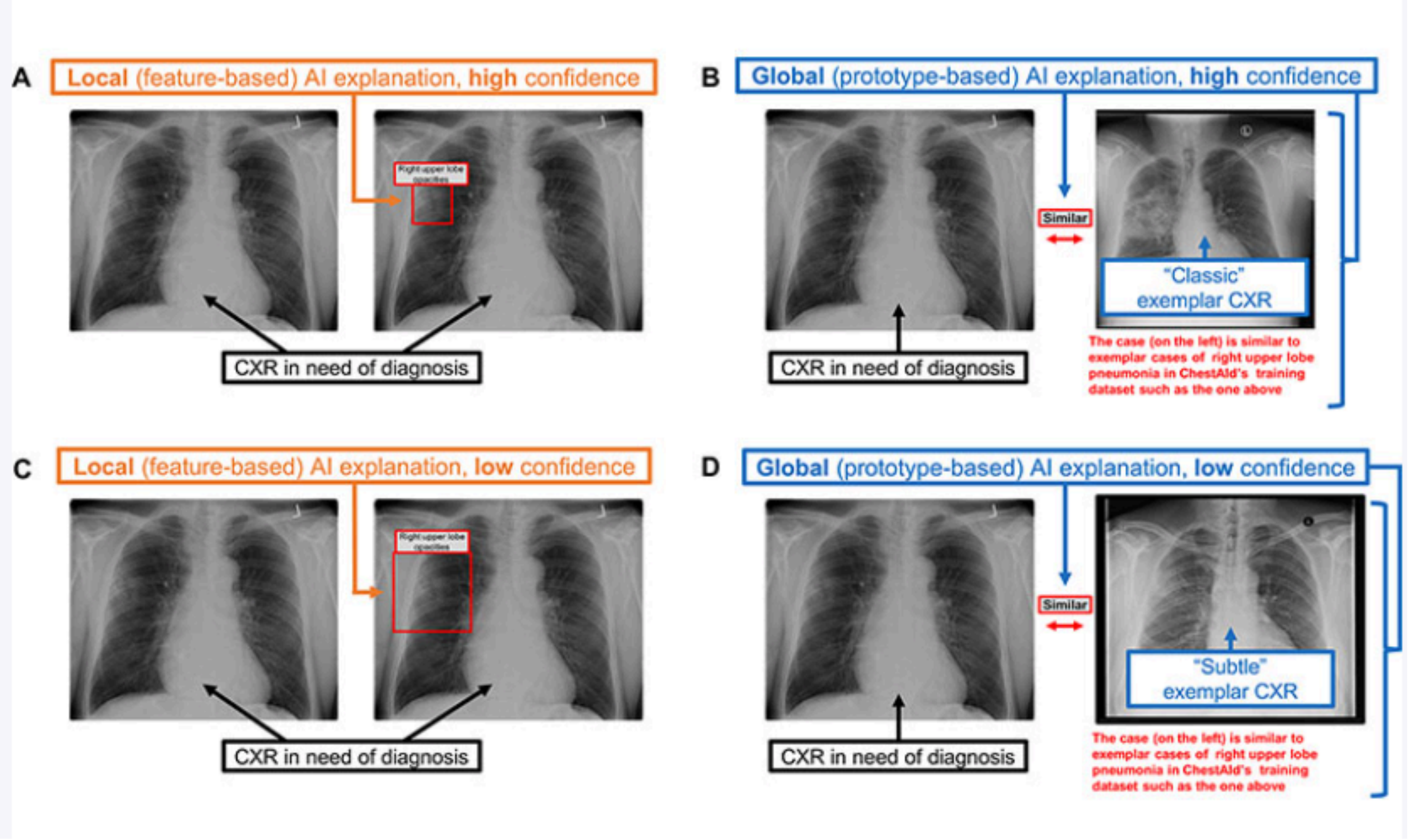
An inside conversation:

Former Tesla Engineer: “Our end-to-end ML system is a complete black box — we have no insight into how it makes predictions, so if something goes wrong, we can’t pinpoint the issue.”

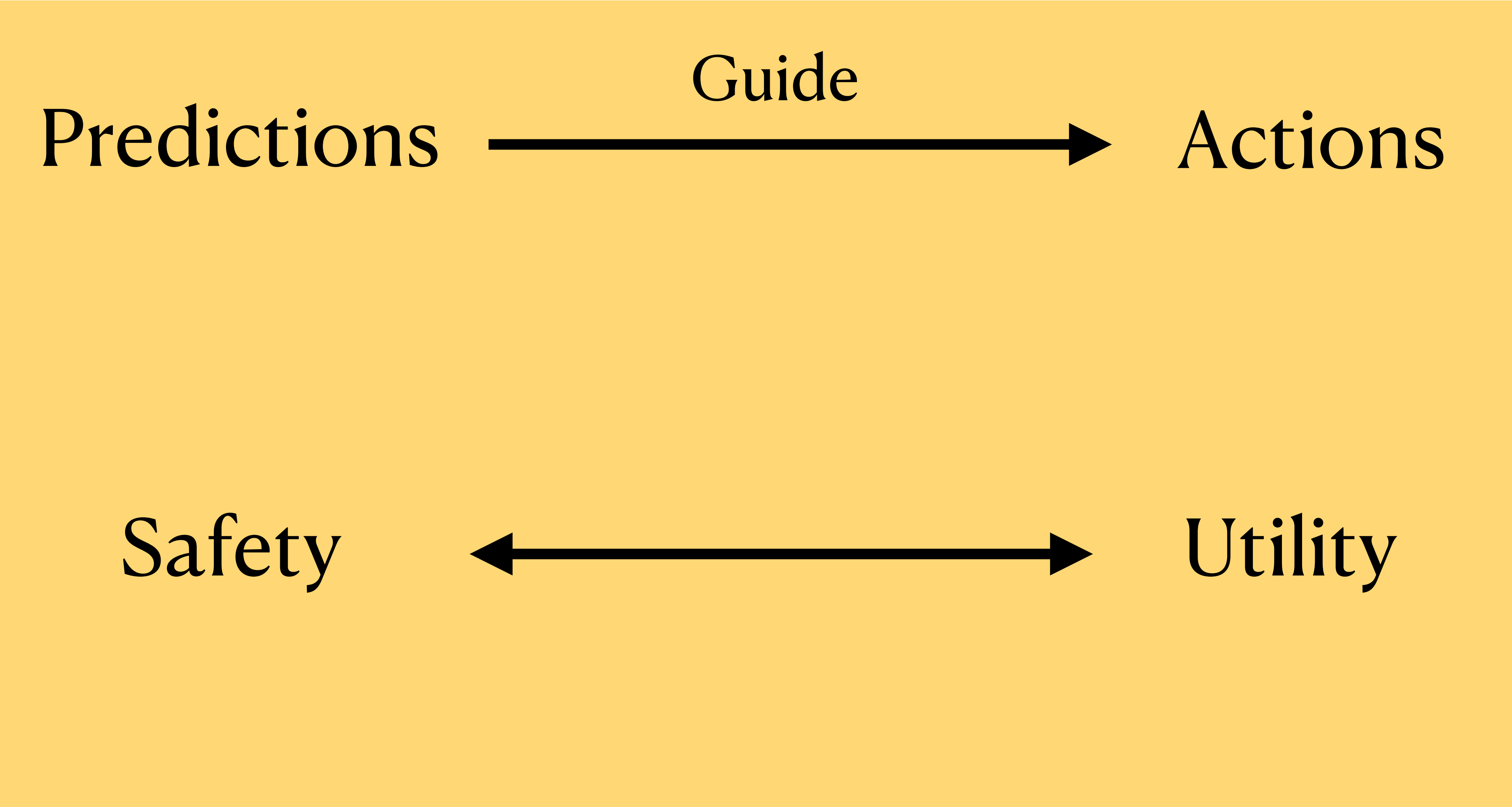
AI Researcher: “Right, and while computer vision is generally accurate, studies show it still misses about 3% of objects, which could mean failing to detect a pedestrian.”

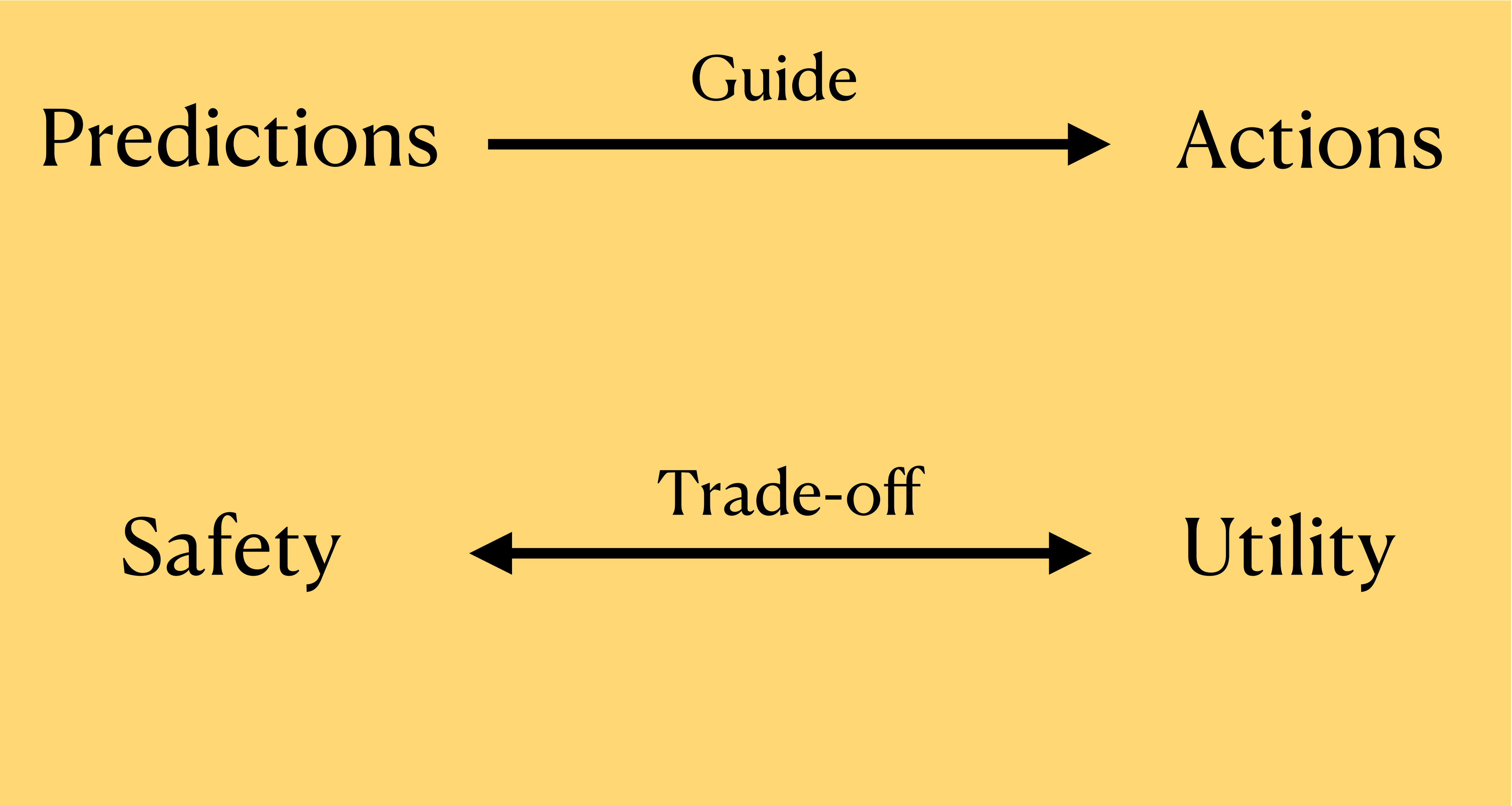
Former Tesla Engineer: “Exactly. There’s always the risk of missing one of the infinite ‘edge cases’ on the road, making safety hard to guarantee.”

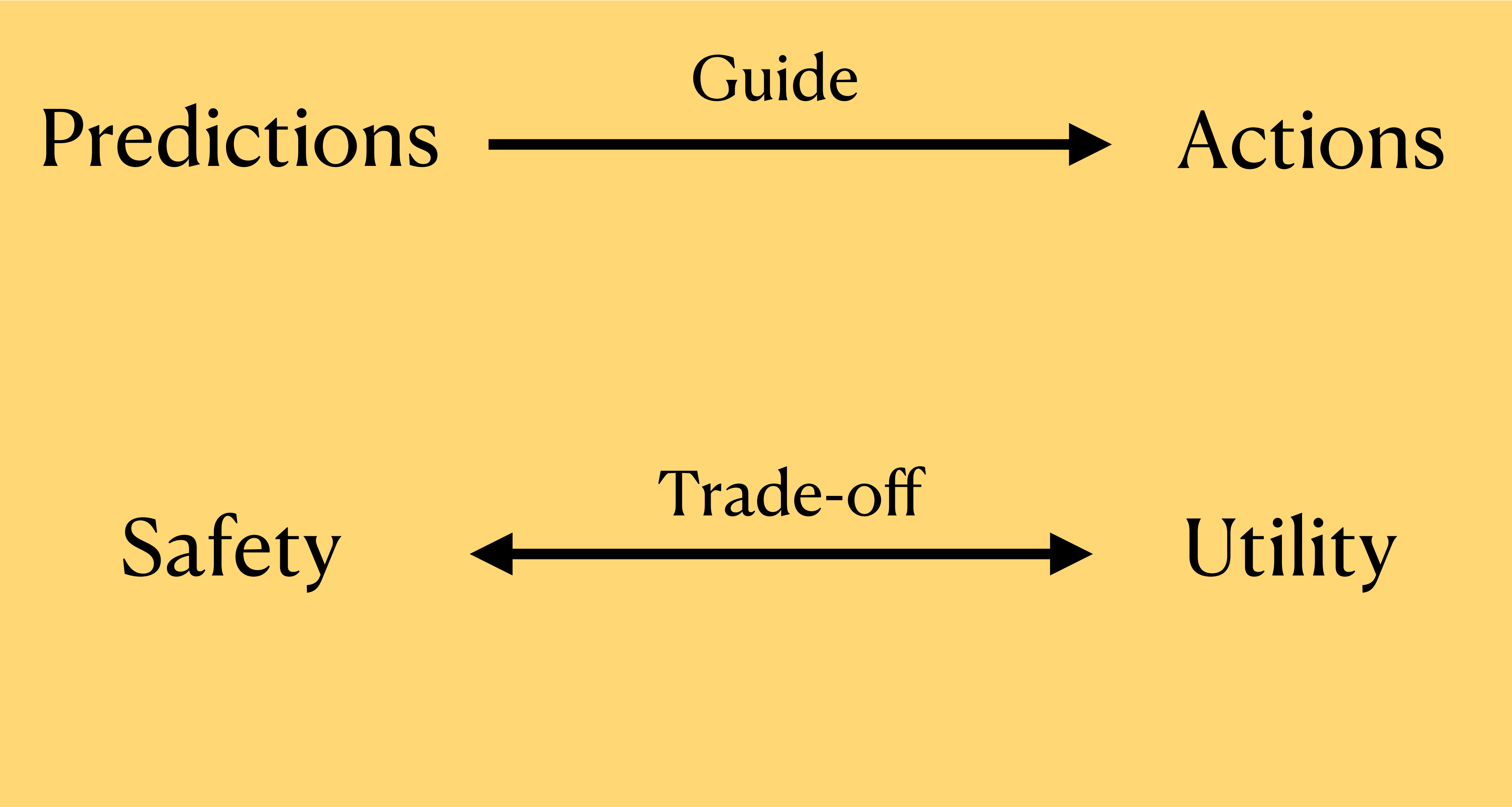
Trust in AI is a Double-Edged Sword



$$\text{Predictions} \xrightarrow{\text{Guide}} \text{Actions}$$


Guide
Predictions
Actions
Safety
Utility


Guide
Predictions
Actions
Trade-off
Safety
Utility



On the one hand:

Do nothing!

Predictions $\xrightarrow{\text{Guide}}$ Actions

Safety $\xleftrightarrow{\text{Trade-off}}$ Utility

On the one hand:

Do nothing! $\longrightarrow$ Ultimate safety

Predictions $\xrightarrow{\text{Guide}}$ Actions

Safety $\xleftrightarrow{\text{Trade-off}}$ Utility

On the one hand:

Do nothing! $\longrightarrow$ Ultimate safety

$\longrightarrow$ Zero utility

Predictions $\xrightarrow{\text{Guide}}$ Actions

Safety $\xleftrightarrow{\text{Trade-off}}$ Utility

On the one hand:

Do nothing! $\longrightarrow$ Ultimate safety

$\longrightarrow$ Zero utility

On the other hand:

Ignore safety $\longrightarrow$ Maximize utility

Predictions $\xrightarrow{\text{Guide}}$ Actions

Safety $\xleftrightarrow{\text{Trade-off}}$ Utility

On the one hand:

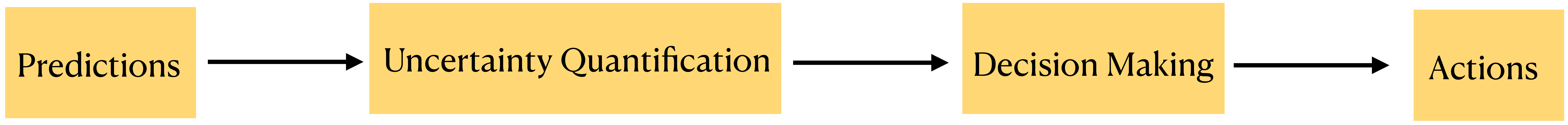
Do nothing! → Ultimate safety

→ Zero utility

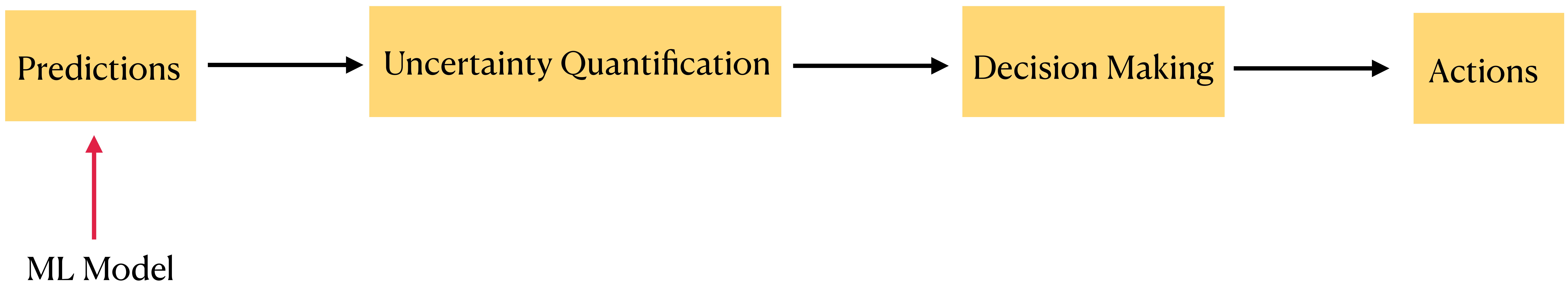
On the other hand:

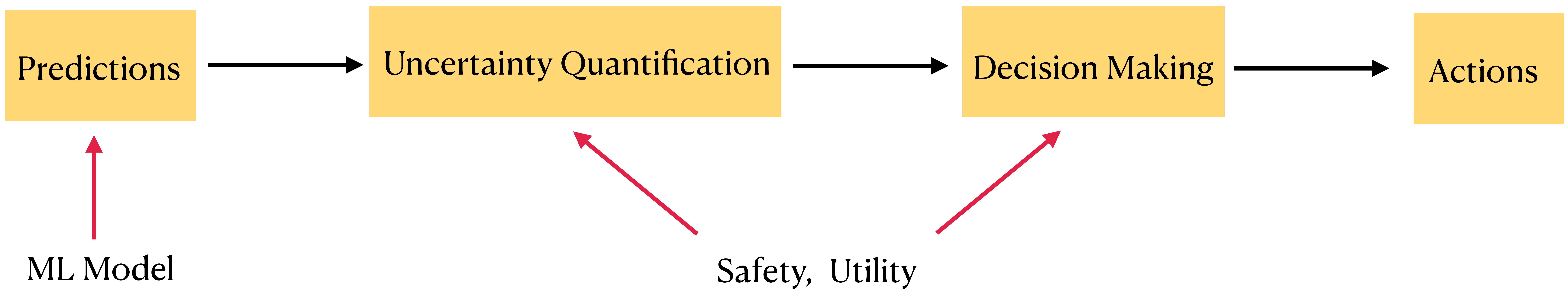
Ignore safety → Maximize utility

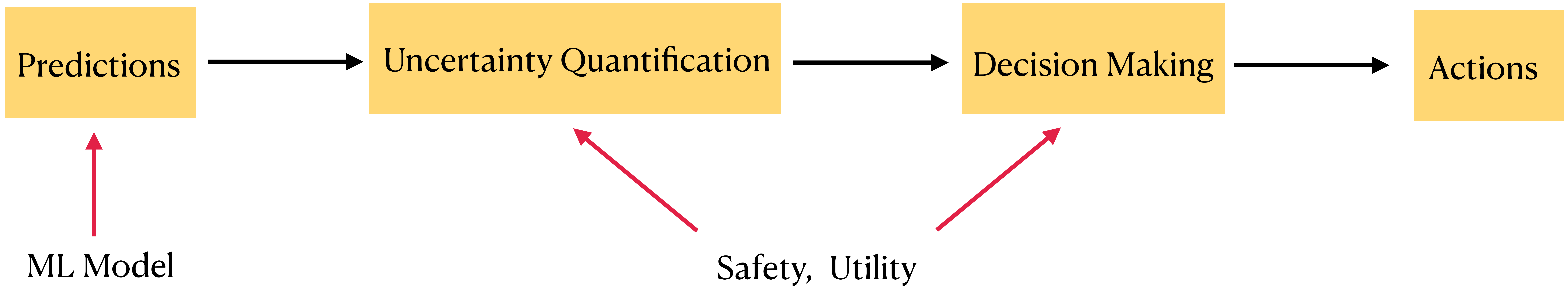
→ Disasterous outcomes



Predictions
Uncertainty Quantification
Decision Making
Actions


Predictions
Uncertainty Quantification
Decision Making
Actions
ML Model

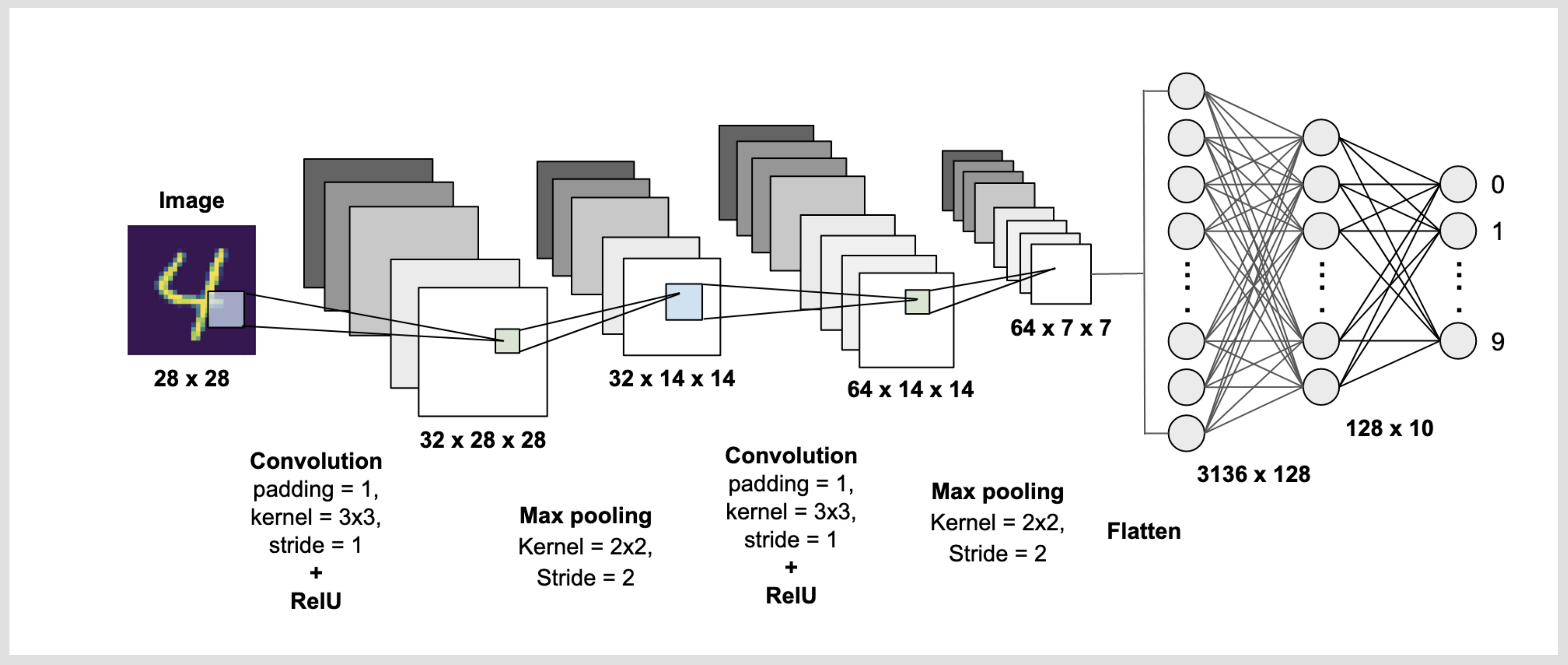

Predictions
Uncertainty Quantification
Decision Making
Actions
ML Model
Safety, Utility



What is the optimal interface between prediction and action that allows for navigating the trade-off between safety and utility in high-stakes applications?

Conformal Prediction – A Promising Framework

Trained Prediction Models



trained model

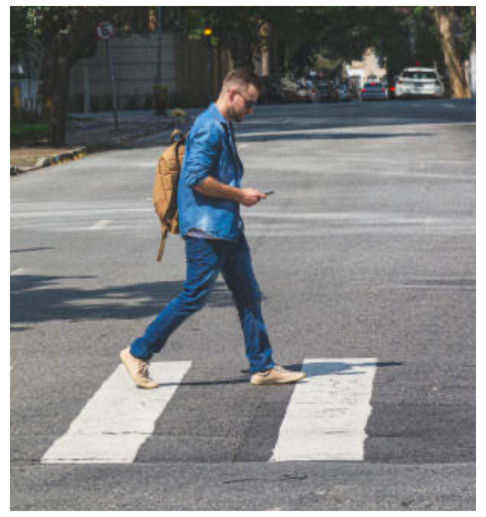
Trained Prediction Models

Trained Prediction Models

trained model

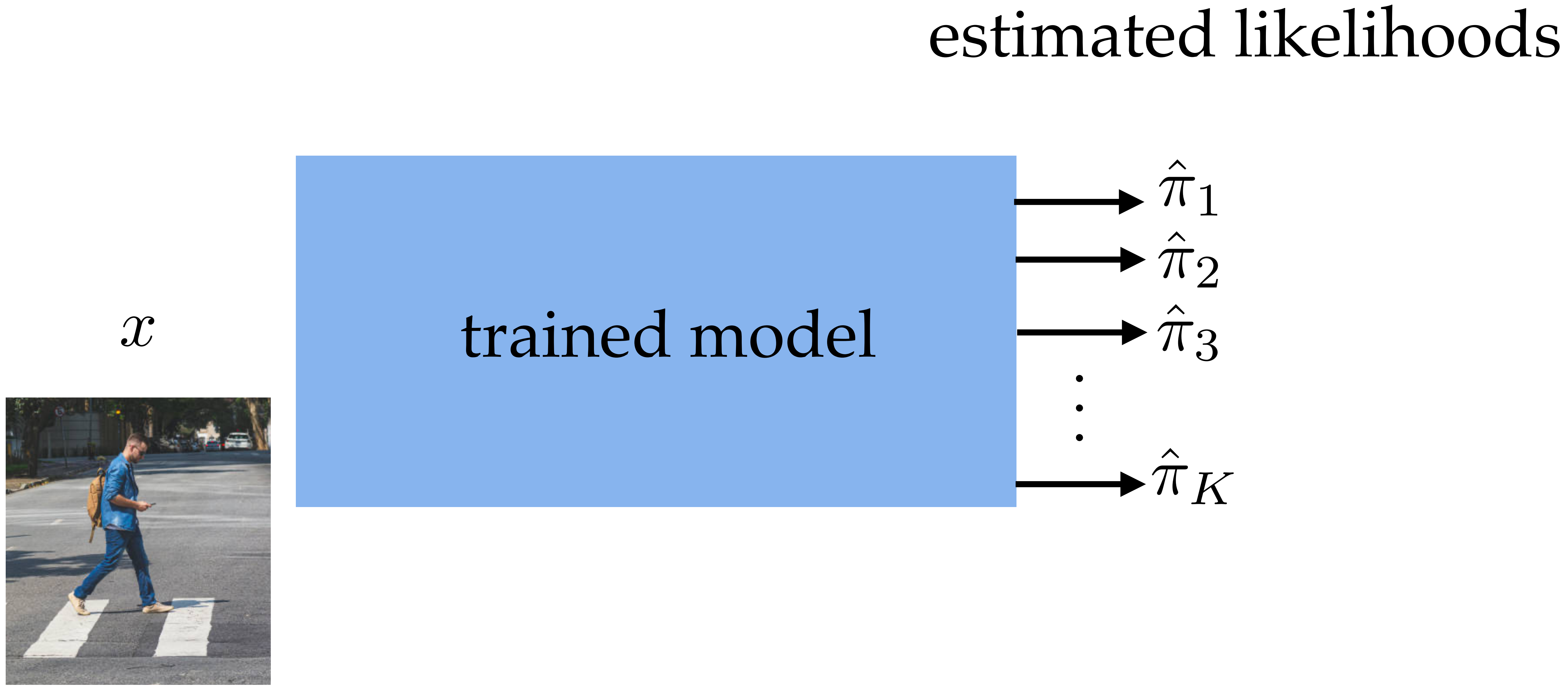
Trained Prediction Models

x



trained model

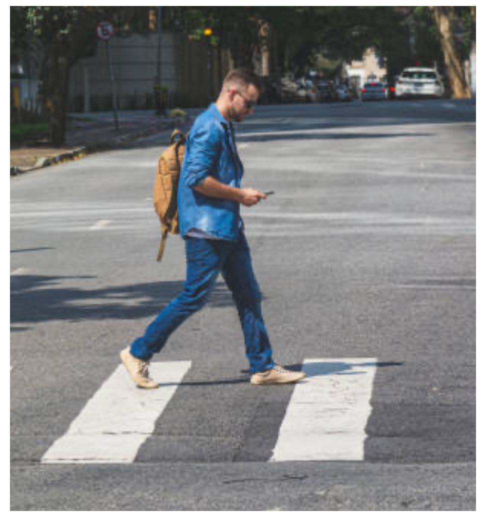
Trained Prediction Models



Trained Prediction Models

estimated likelihoods

x



trained model

$\hat{\pi}_1$ 0.01

$\hat{\pi}_2$ 0.04

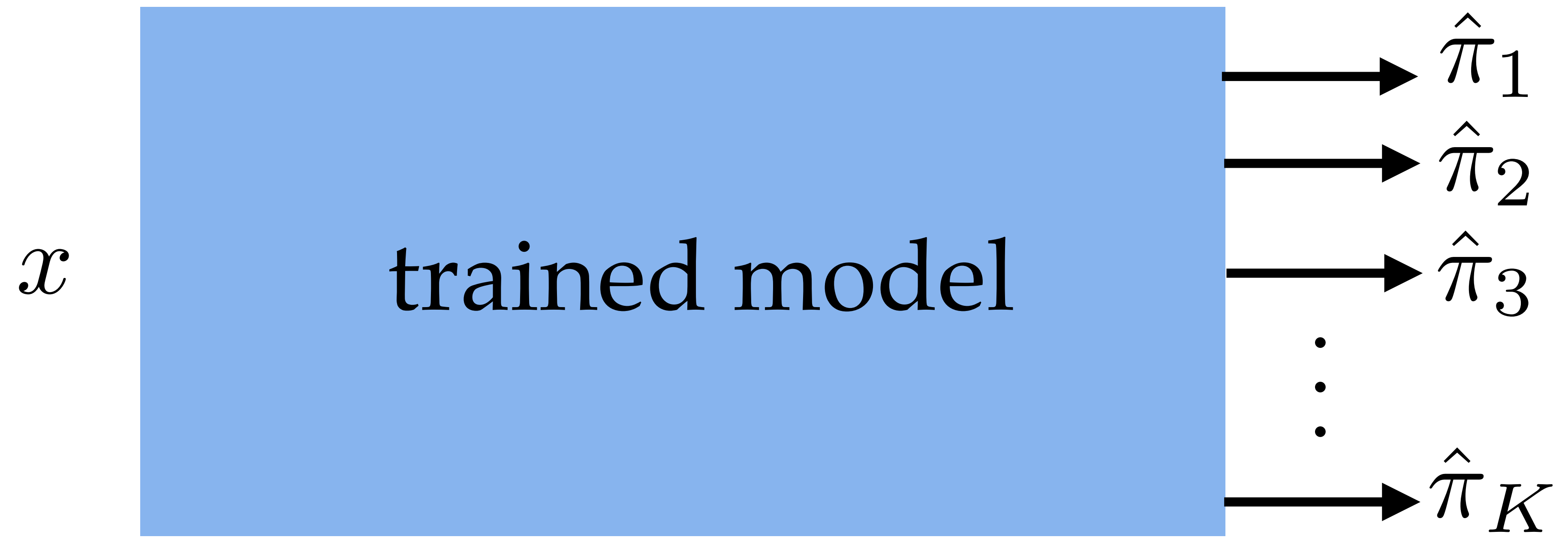
$\hat{\pi}_3$ 0.42

$\vdots$ $\vdots$

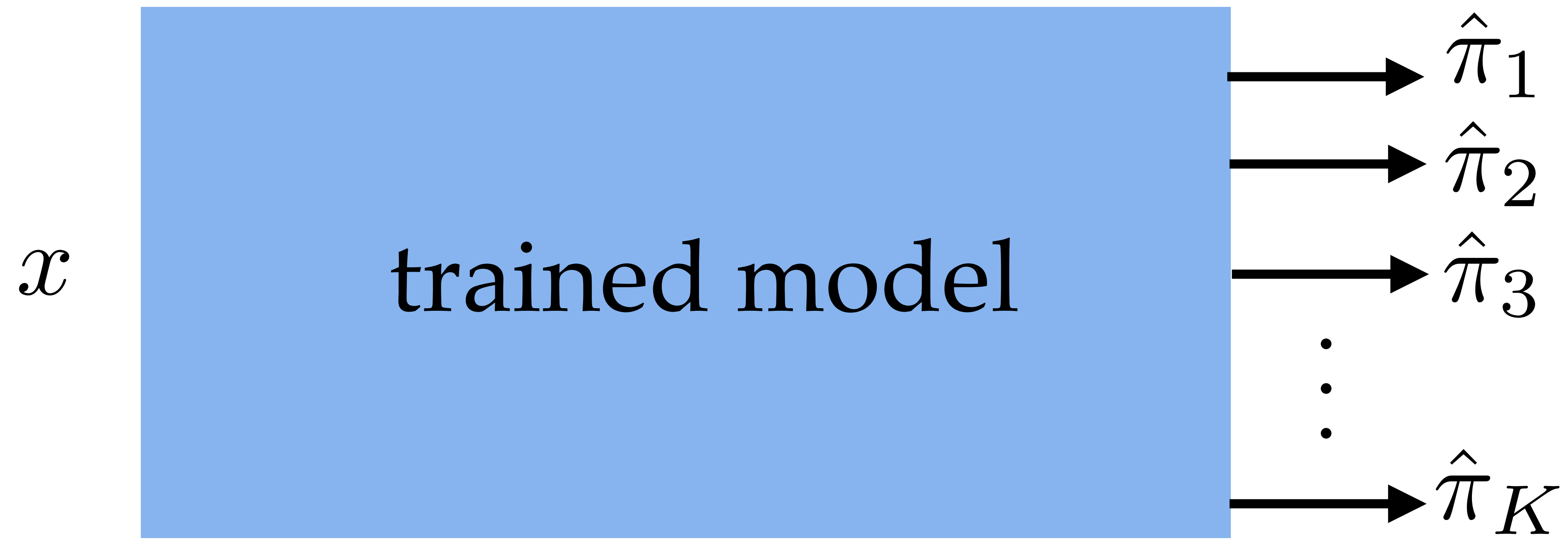
$\hat{\pi}_K$ 0.09

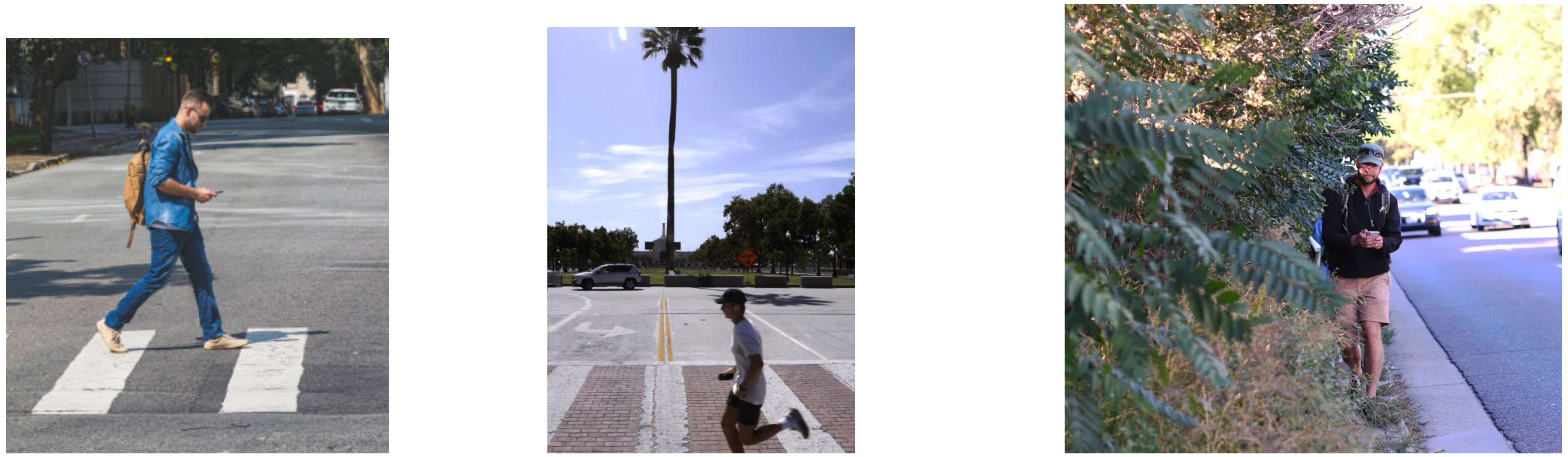
- prediction is then based on the class that has maximum likelihood

Trained Prediction Models



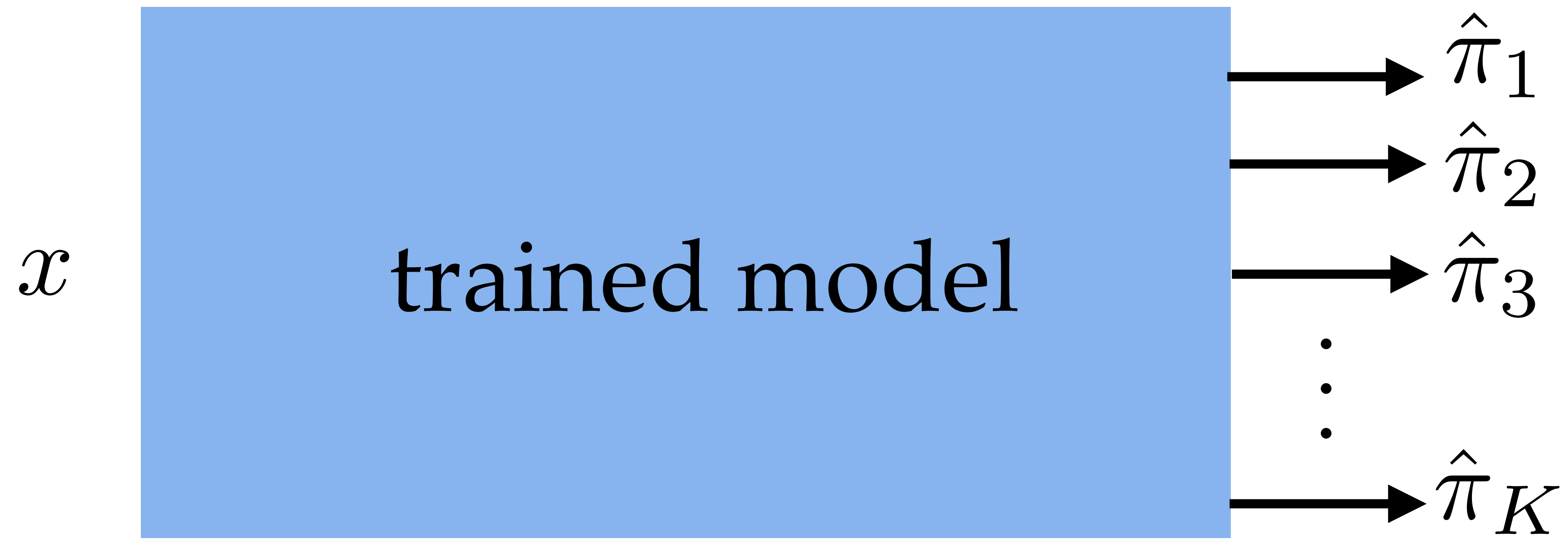
Trained Prediction Models

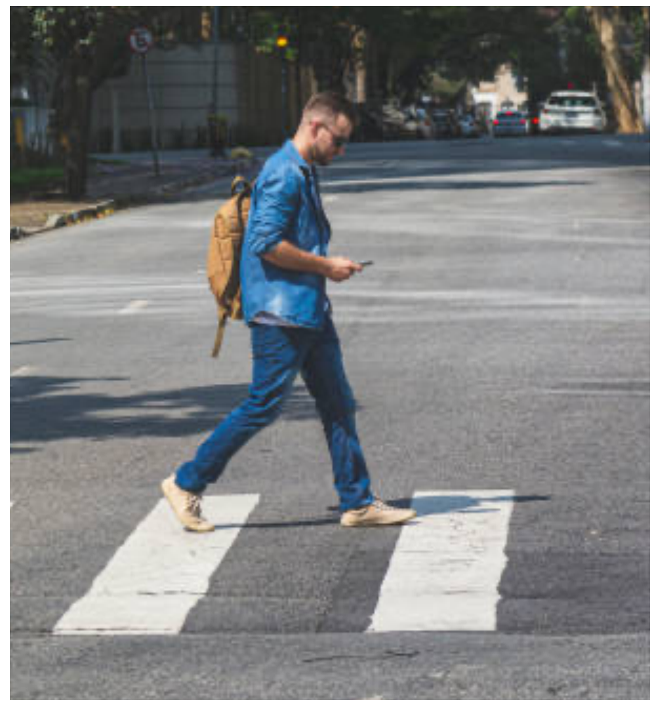




{ human 0.99 } { human 0.90, tree 0.08, bin 0.02 } { Tree 0.70, human 0.20, trash can 0.05 }

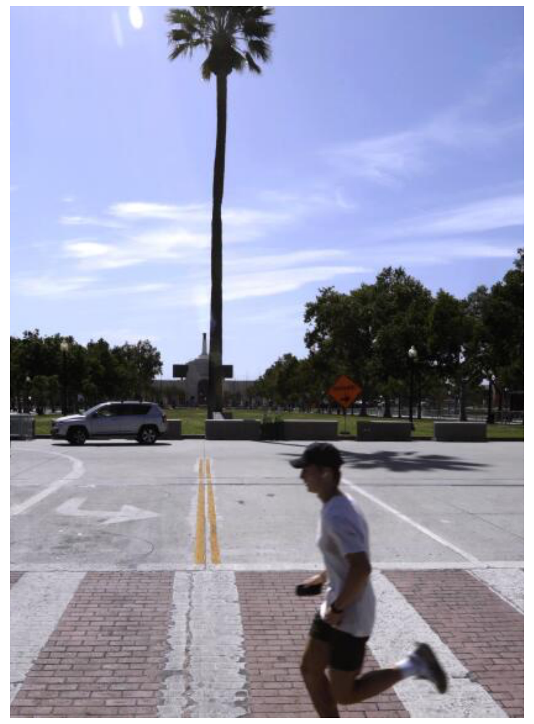
Trained Prediction Models





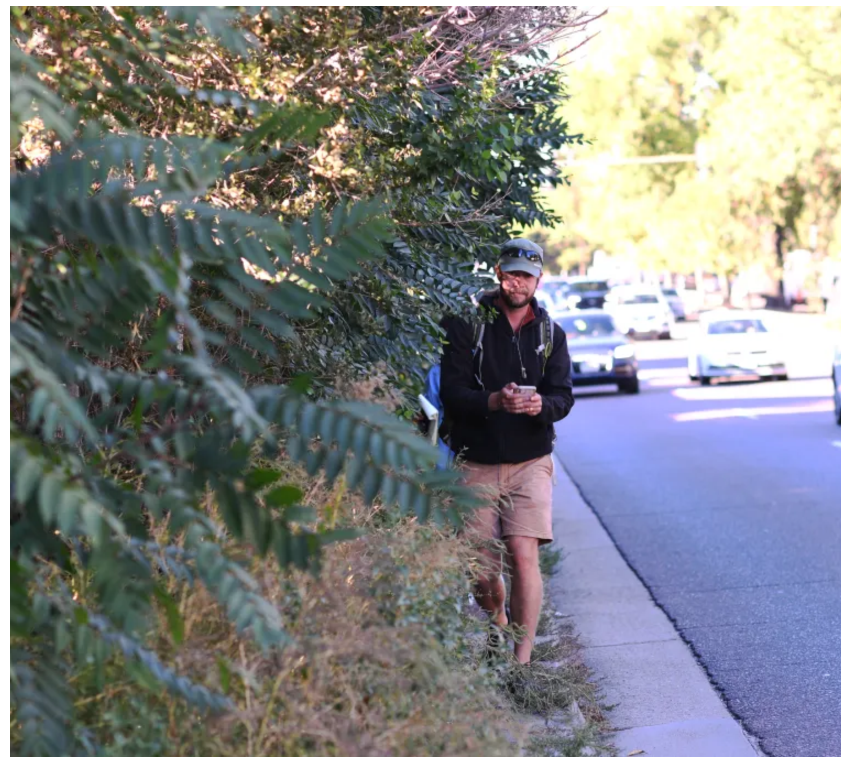
{ human }
0.99

easy



{ human , tree, bin }
0.90 0.08 0.02

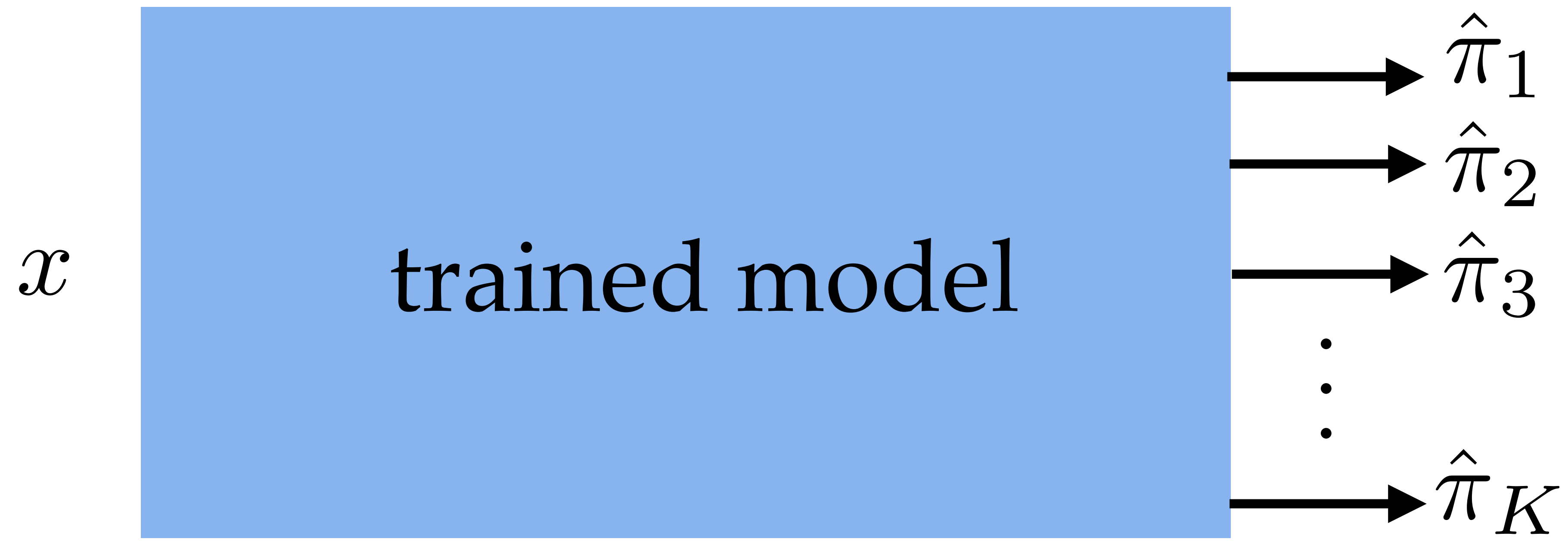
less easy

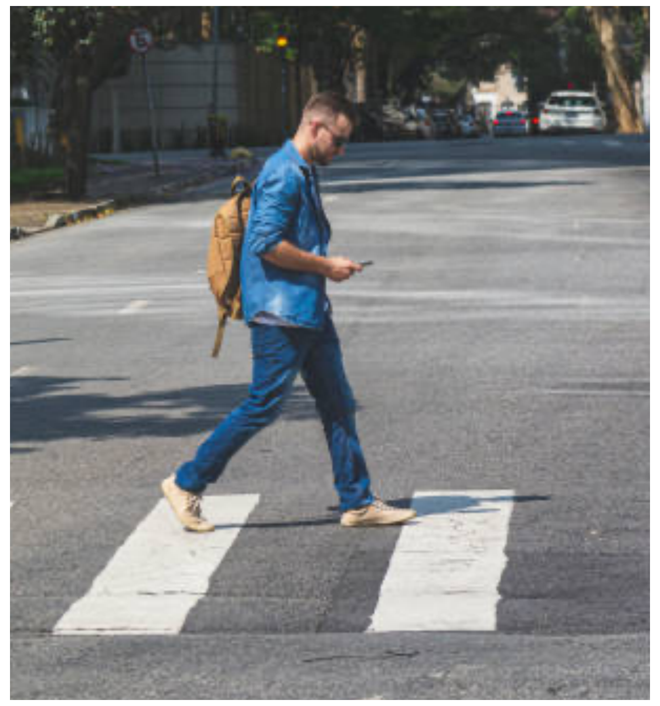


{ Tree, human, trash can }
0.70 0.20 0.05

difficult

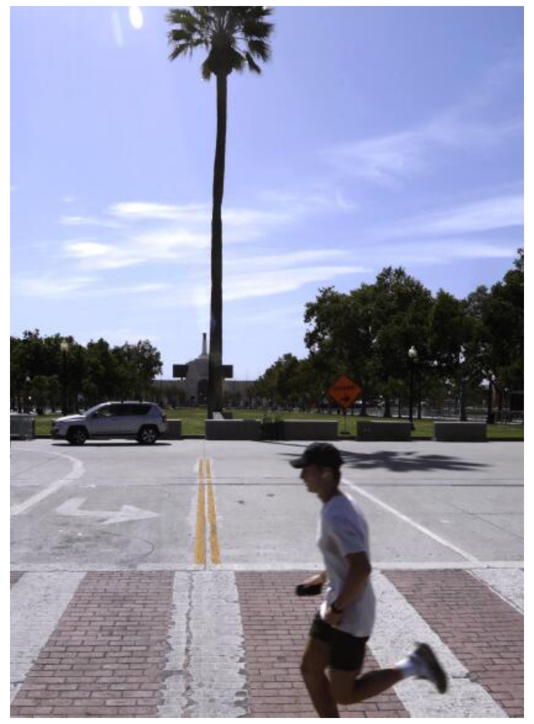
Trained Prediction Models





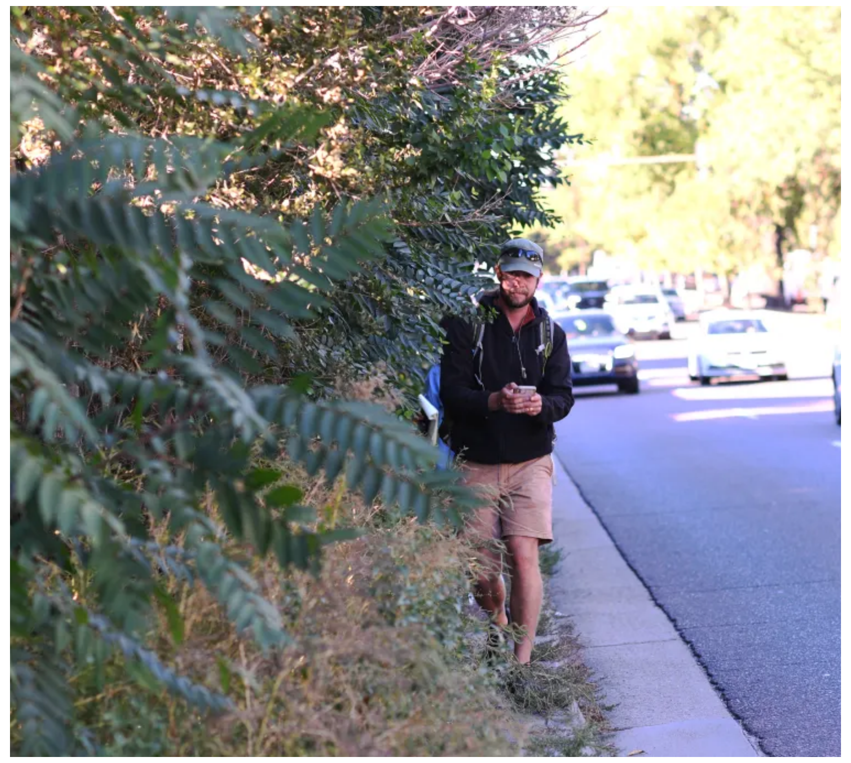
{ human 0.99 }

easy



{ human, tree, bin 0.90 0.08 0.02 }

less easy

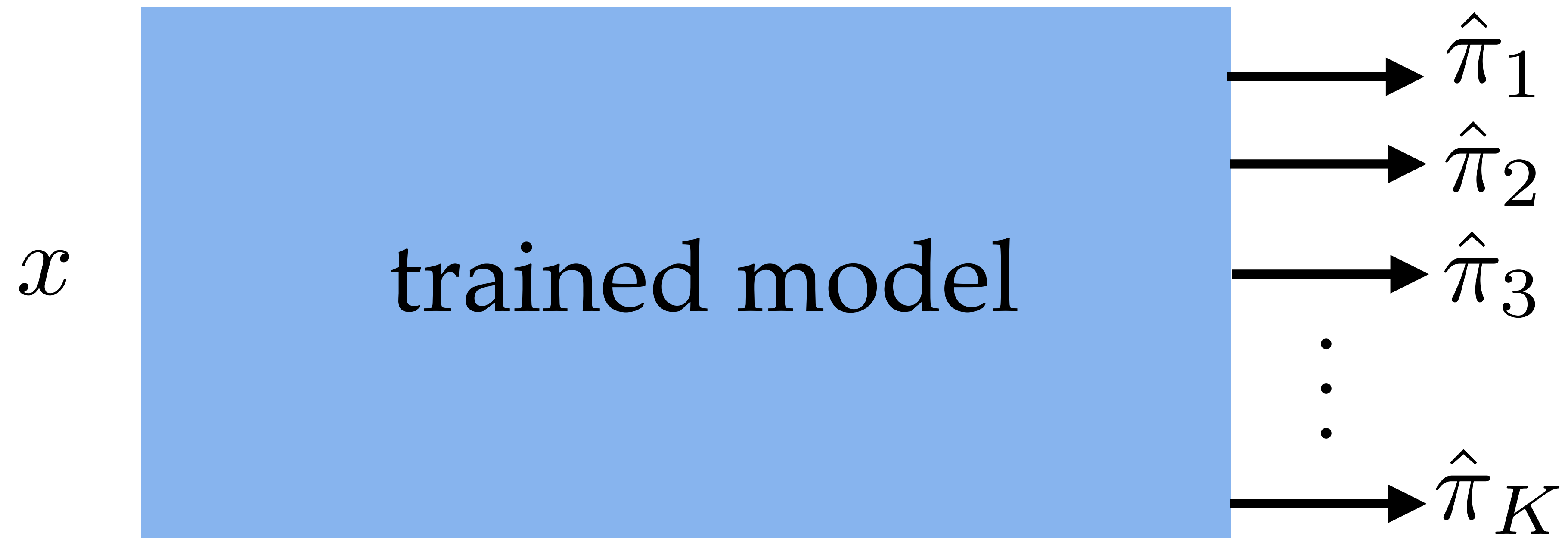


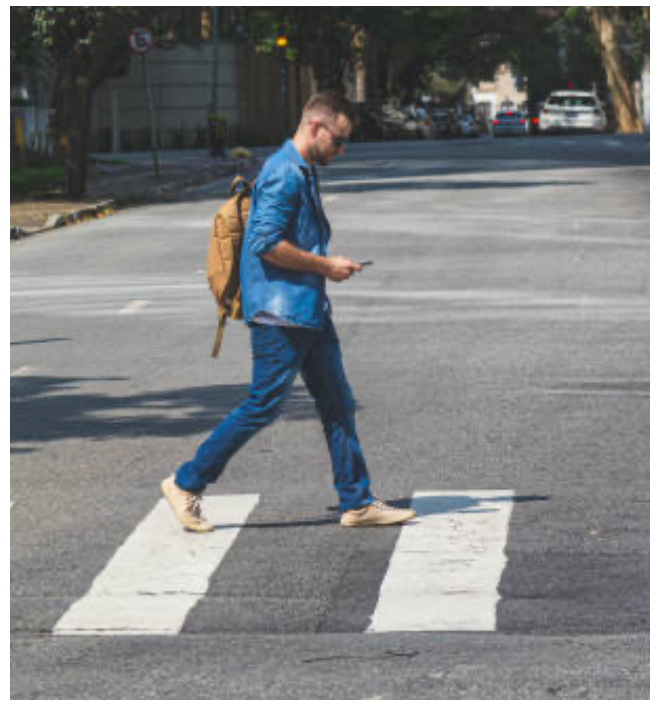
{ Tree, human, trash can 0.70 0.20 0.05 }

difficult

model's prediction:

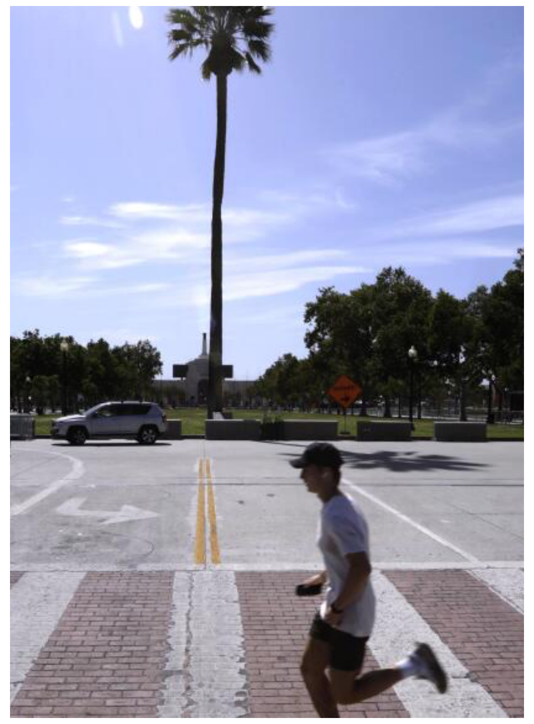
Trained Prediction Models





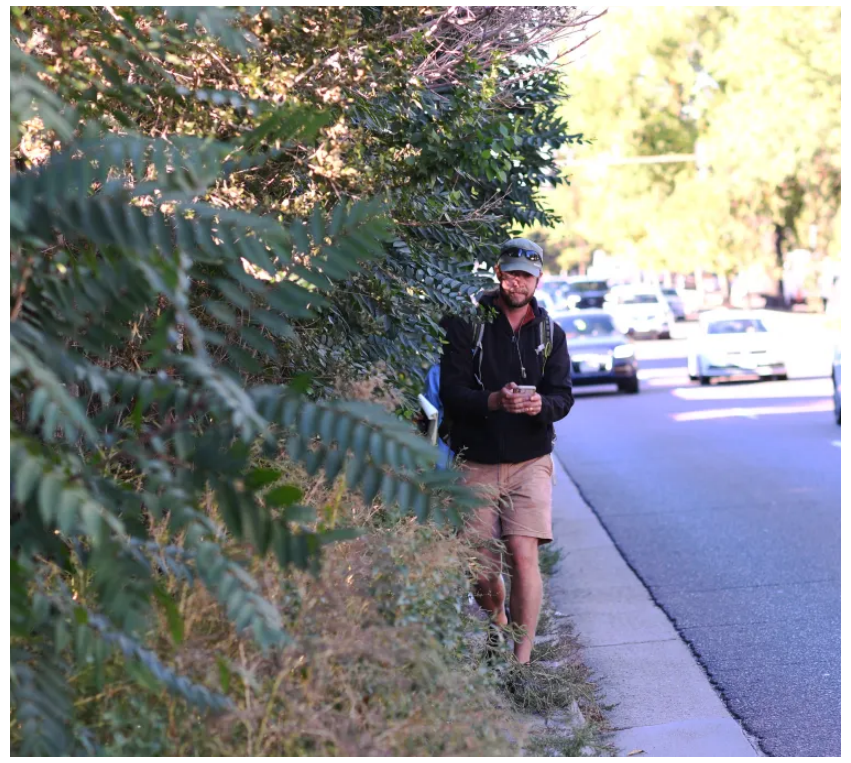
{ human }
0.99

easy



{ human , tree, bin }
0.90 0.08 0.02

less easy



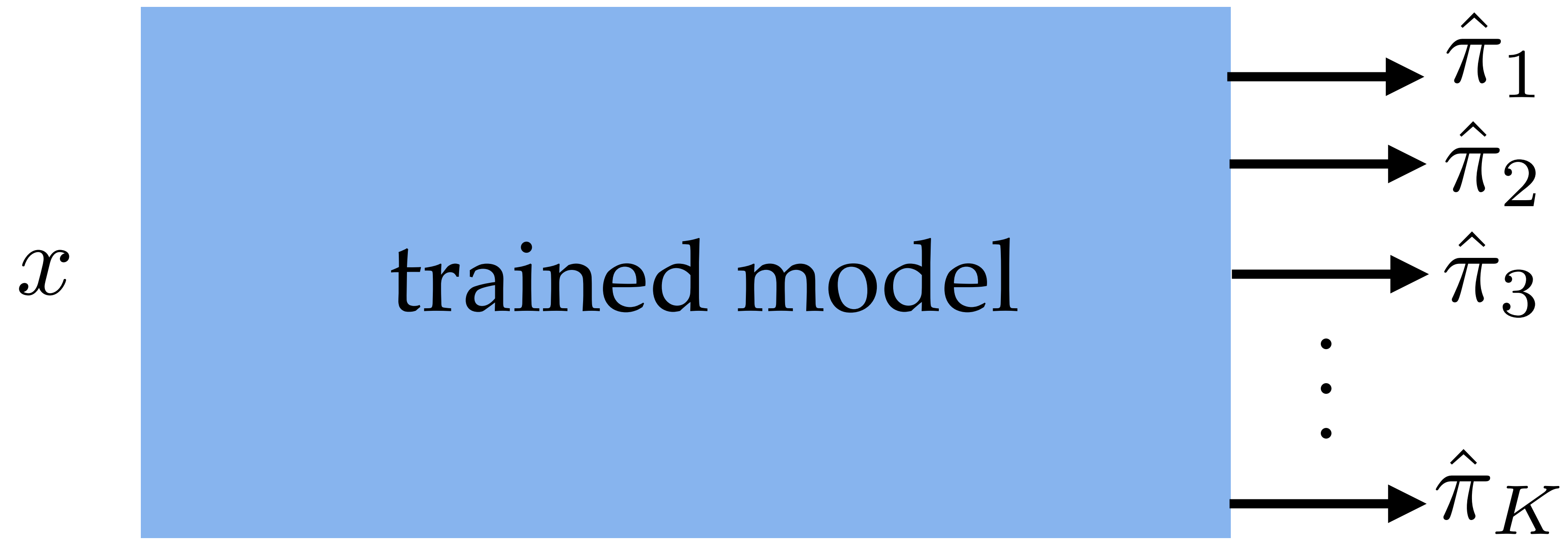
{ Tree, human, trash can }
0.70 0.20 0.05

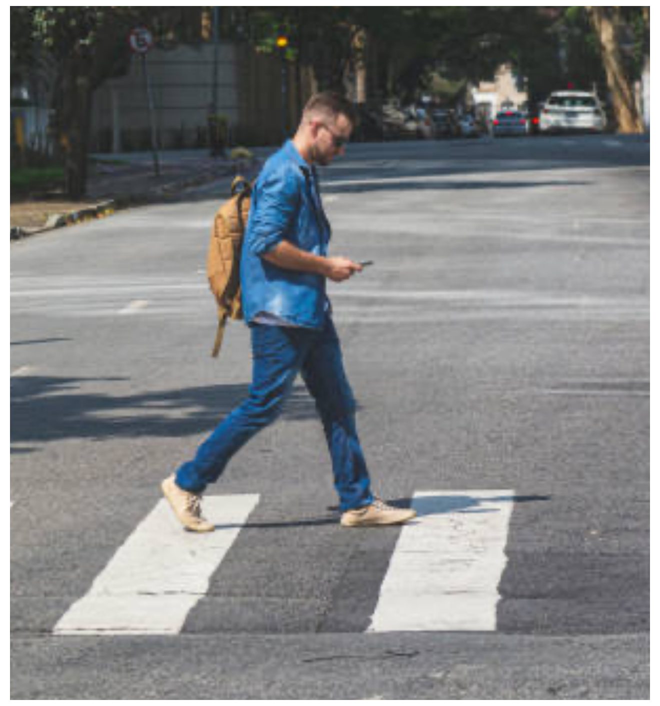
difficult

model's prediction: human

✅

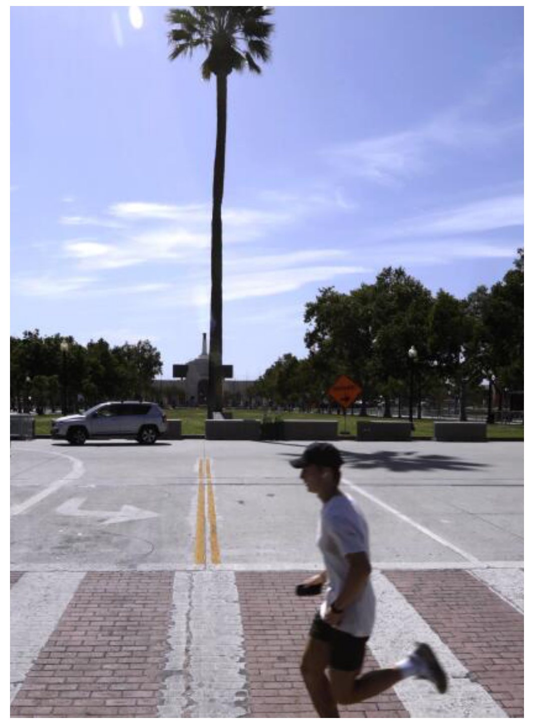
Trained Prediction Models





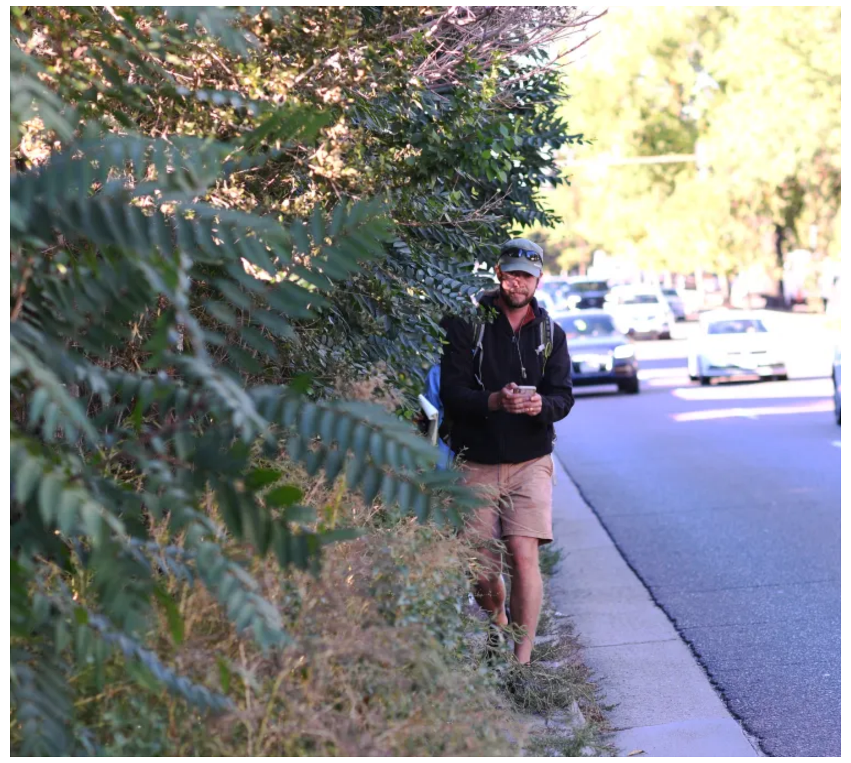
{ human }
0.99

easy



{ human, tree, bin }
0.90 0.08 0.02

less easy

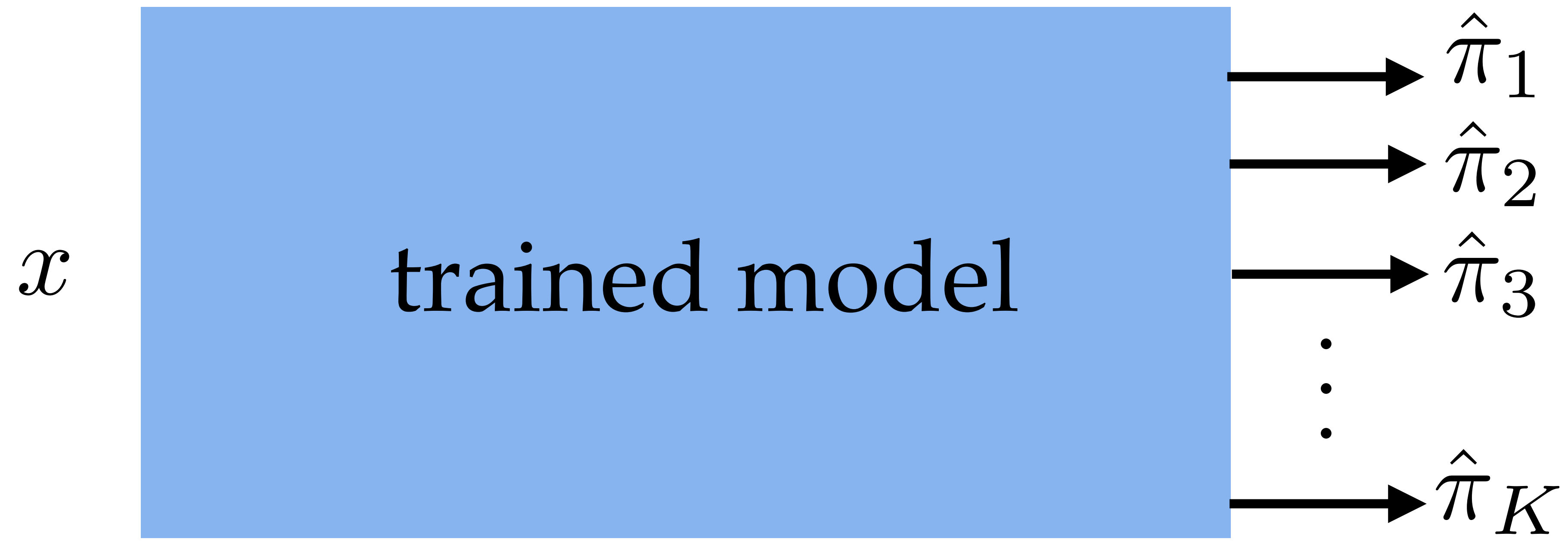


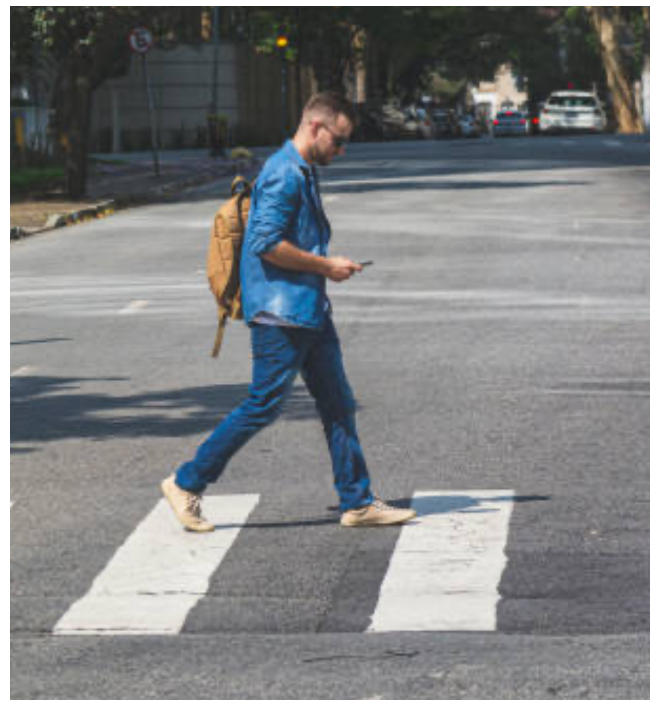
{ Tree, human, trash can }
0.70 0.20 0.05

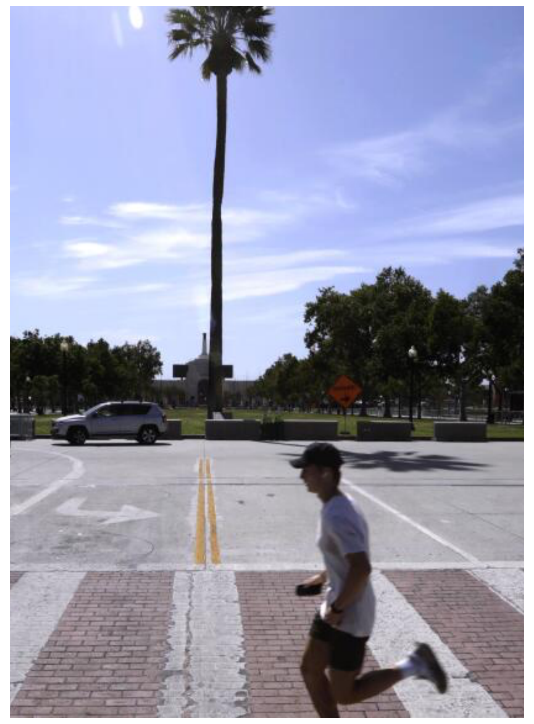
difficult

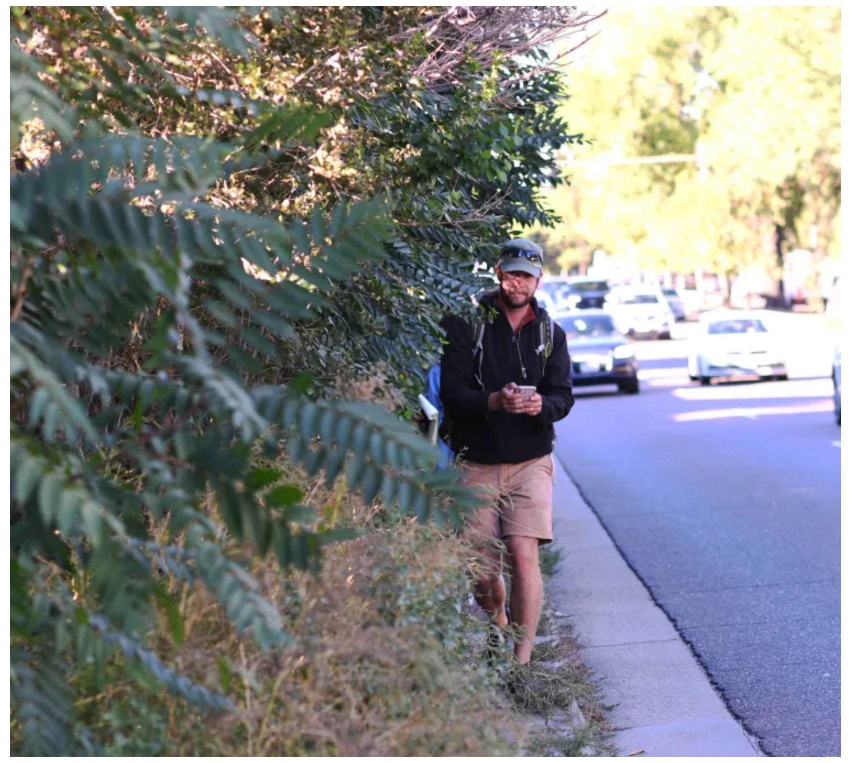
model's prediction: human ✅ human ✅

Trained Prediction Models



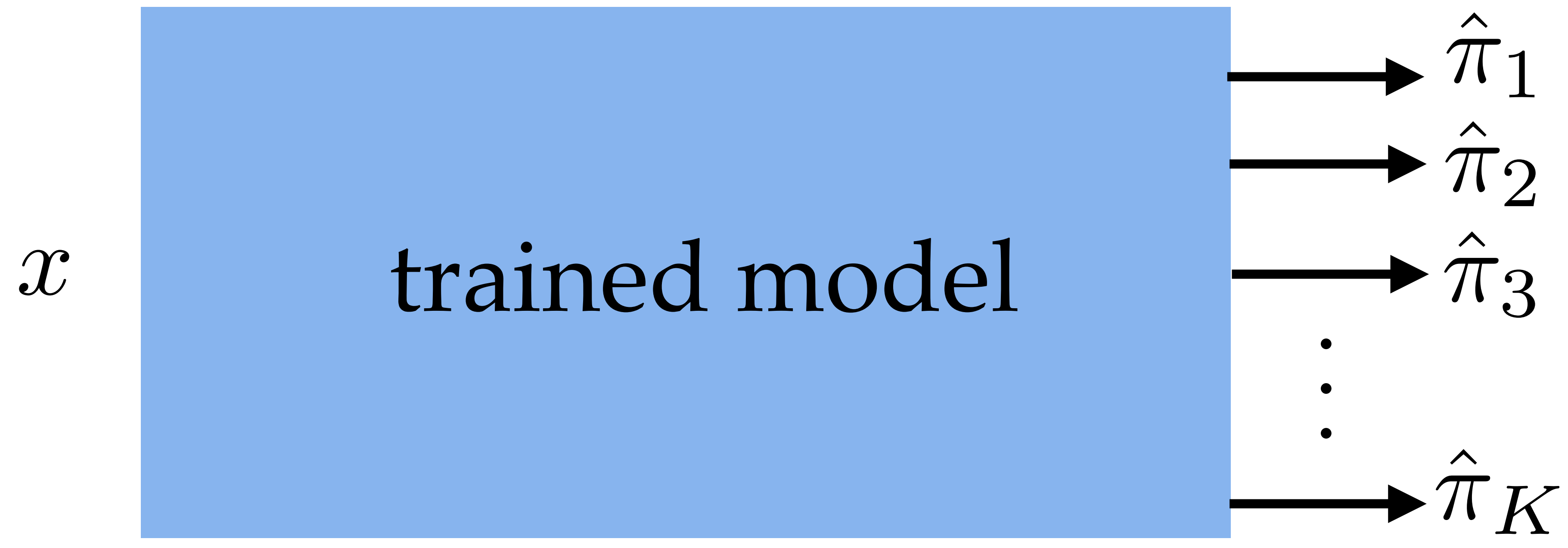


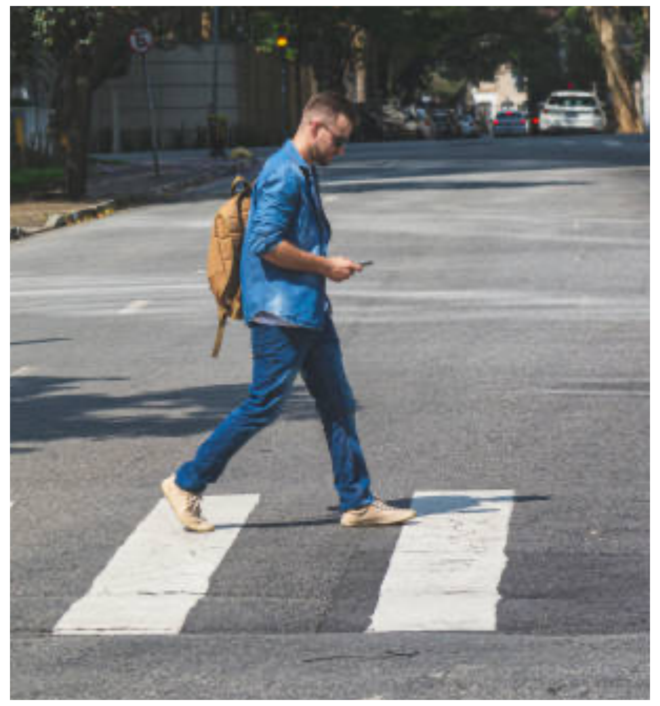




	{human 0.99}	{human 0.90, tree 0.08, bin 0.02}	{Tree 0.70, human 0.20, trash can 0.05}
	easy	less easy	difficult
model's prediction:	human	human	tree
	✅	✅	❌

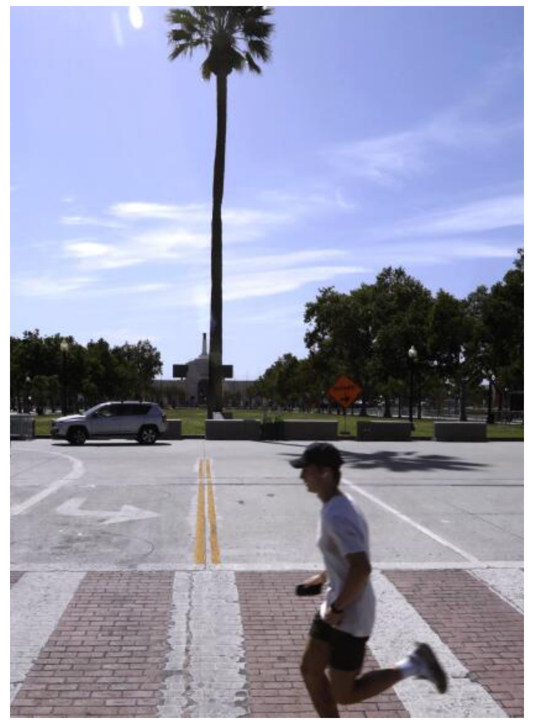
Trained Prediction Models





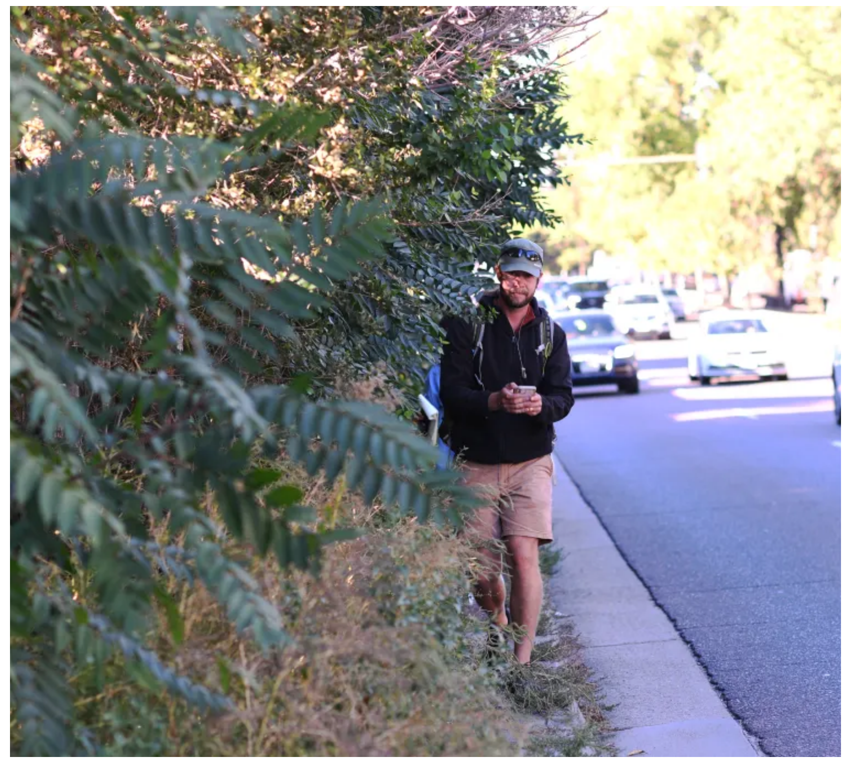
{ human 0.99 }

easy



{ human 0.90, tree 0.08, bin 0.02 }

less easy

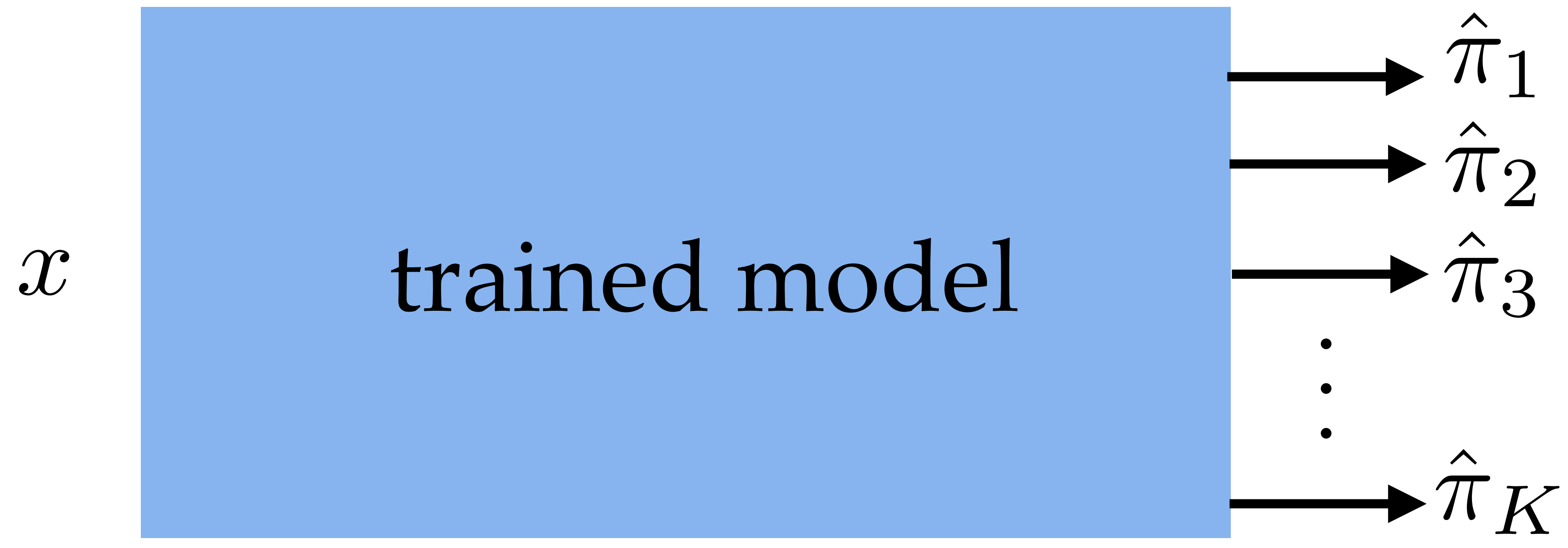


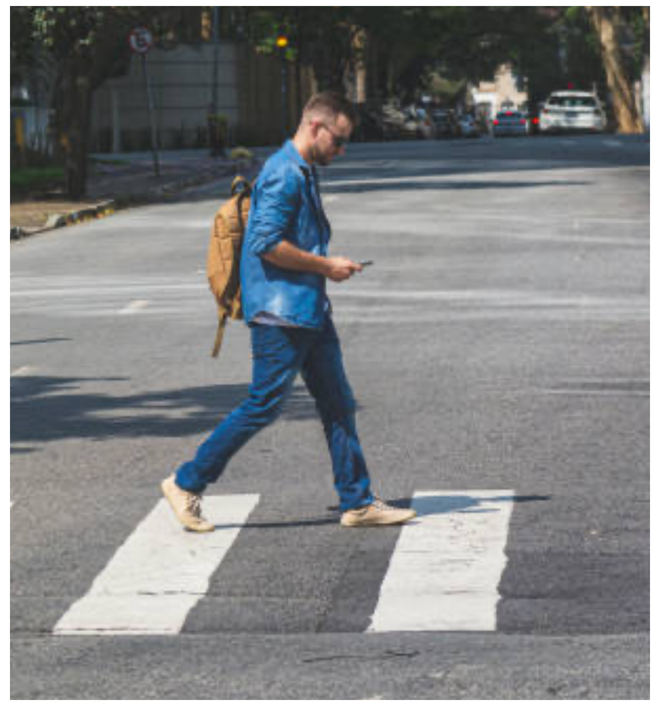
{ Tree 0.70, human 0.20, trash can 0.05 }

difficult

model's prediction: human ✅ human ✅ tree ❌

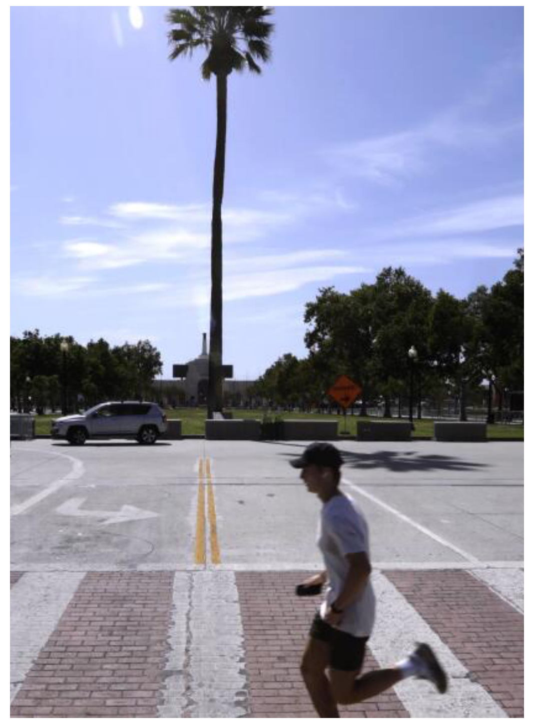
Trained Prediction Models





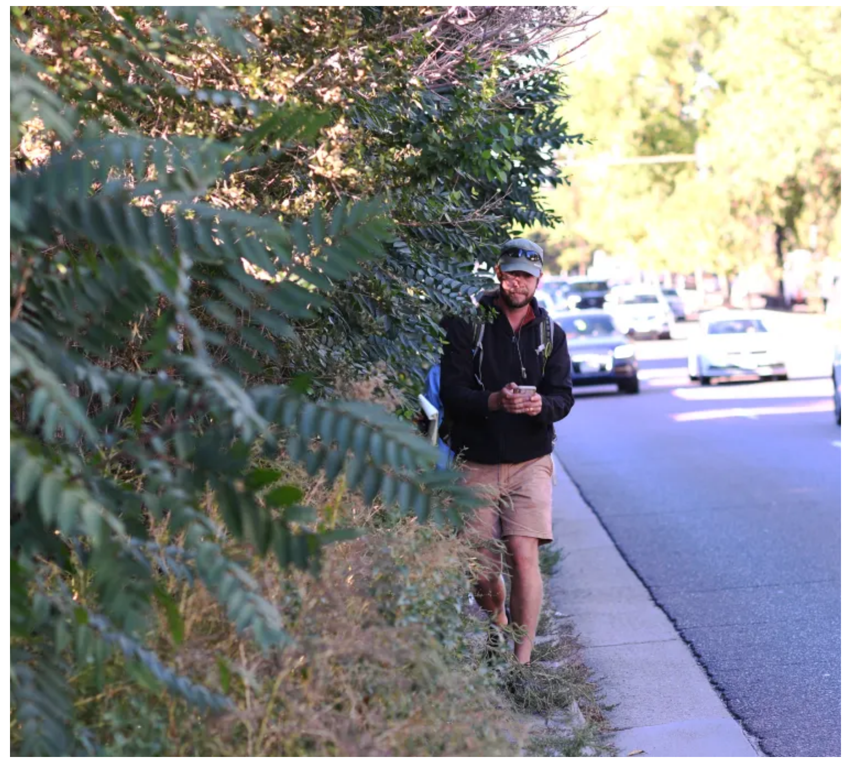
{ human 0.99 }

easy



{ human 0.90, tree 0.08, bin 0.02 }

less easy



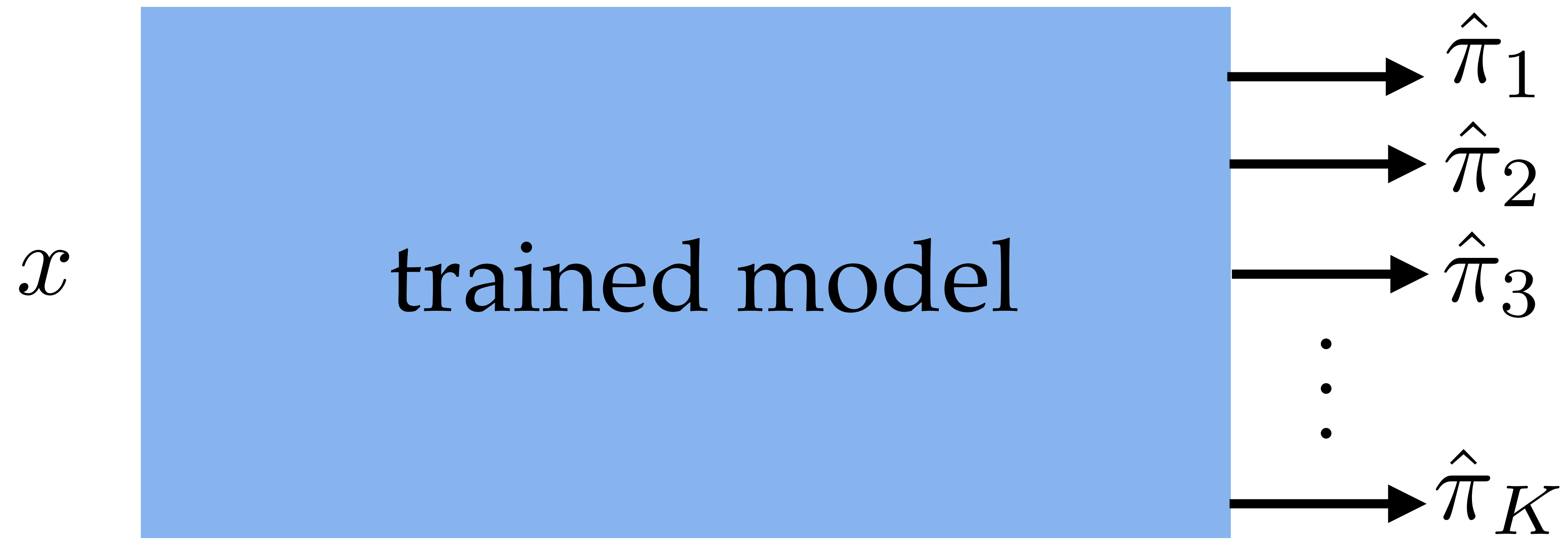
{ Tree 0.70, human 0.20, trash can 0.05 }

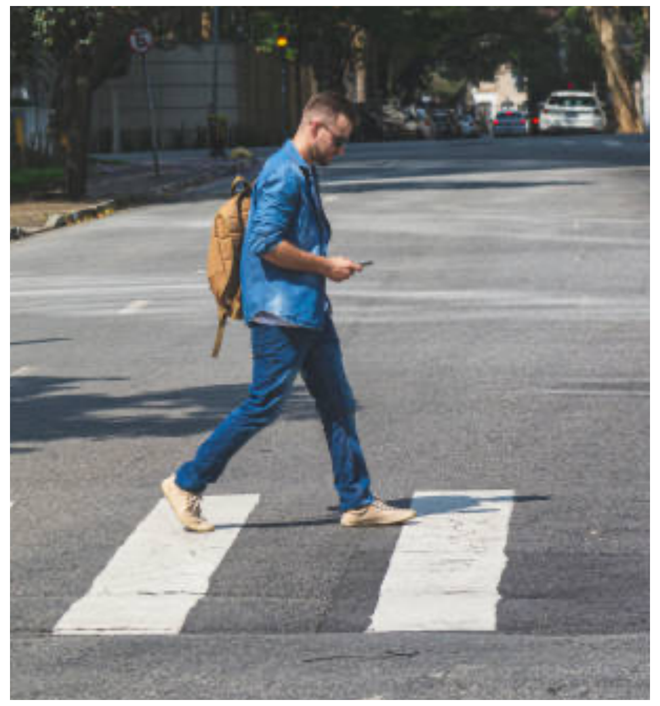
difficult

model's prediction:	human	human	tree
	✅	✅	❌

- the trained model provides estimated likelihoods (a notion of uncertainty)

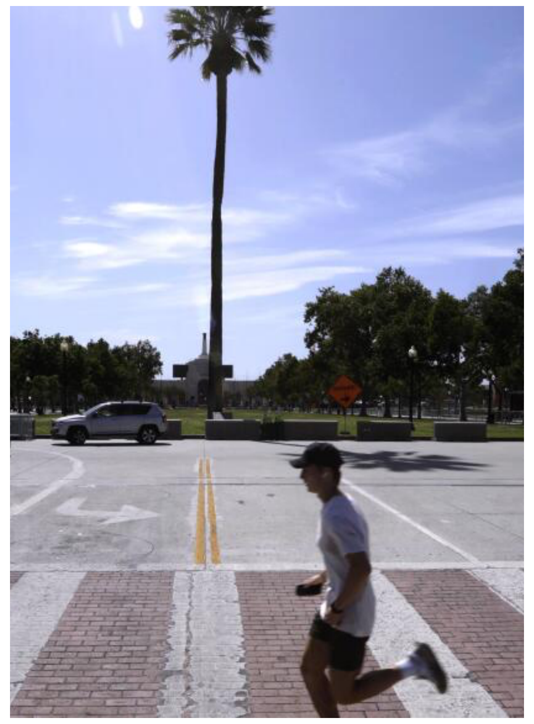
Trained Prediction Models





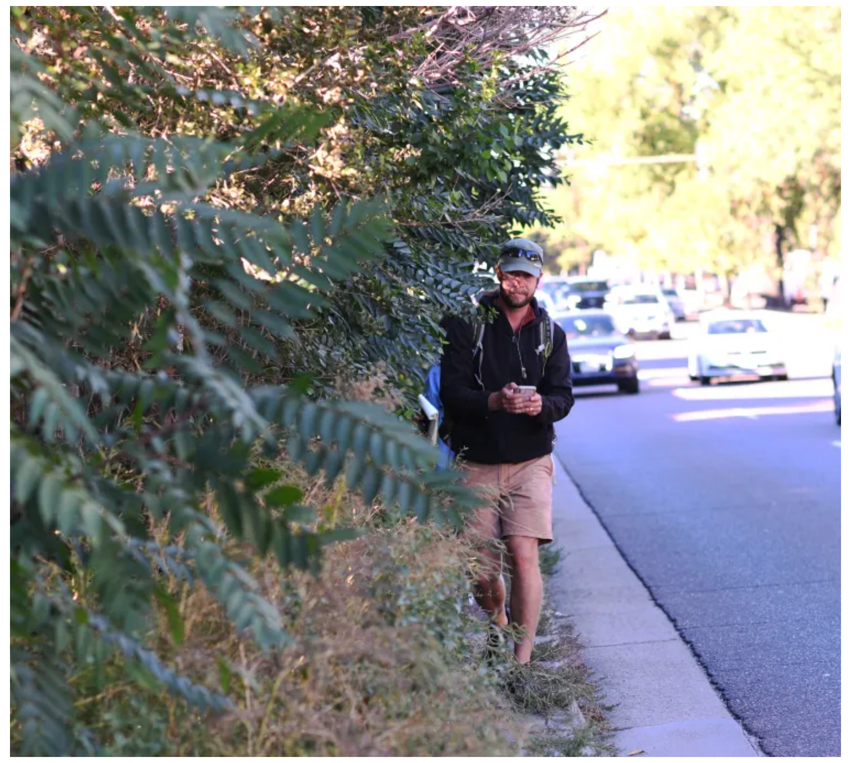
{ human 0.99 }

easy



{ human, tree, bin 0.90 0.08 0.02 }

less easy



{ Tree, human, trash can 0.70 0.20 0.05 }

difficult

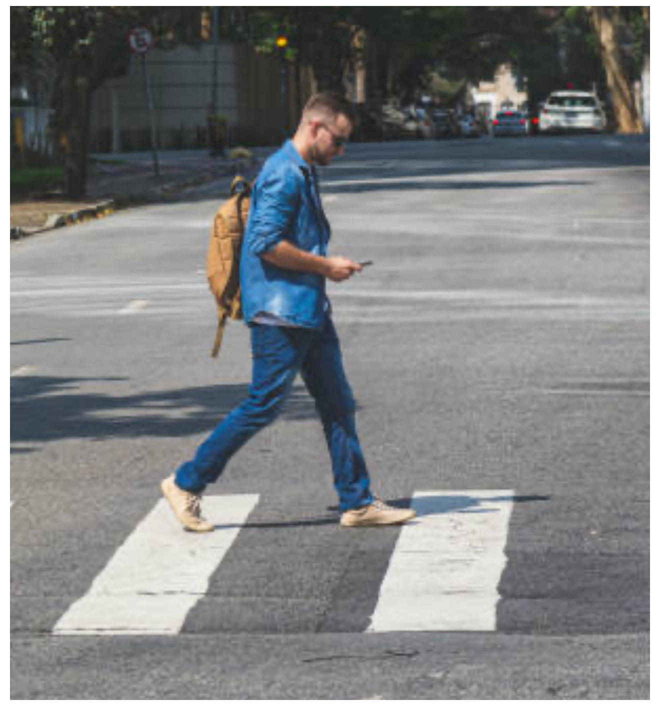
model's prediction: human ✅ human ✅ tree ❌

- the trained model provides estimated likelihoods (a notion of uncertainty)
- these likelihoods are informative but not always correct

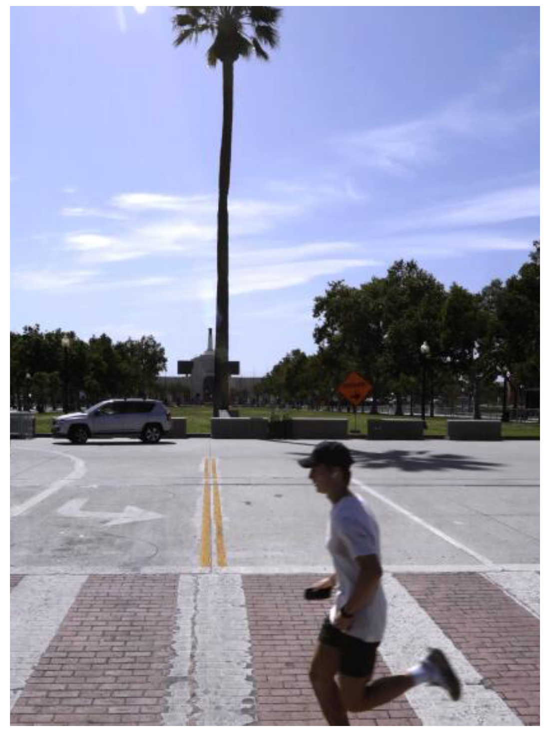
Trained Prediction Models

Trained Prediction Models

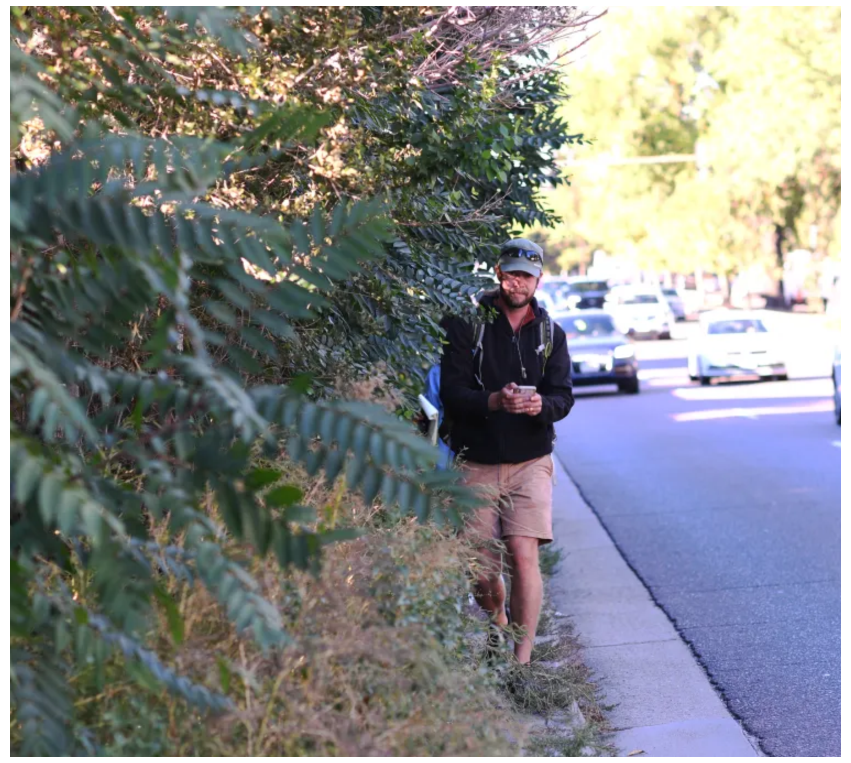
x



{ human } 0.99



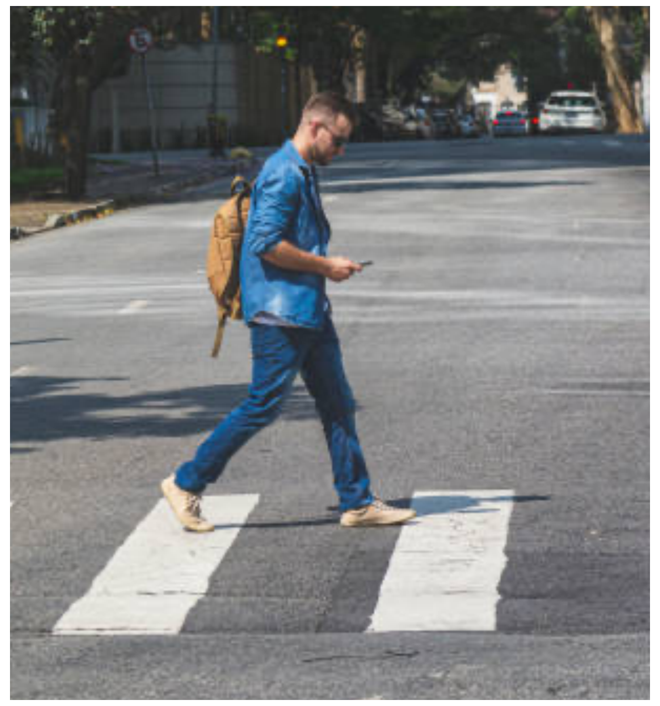
{ human, tree, bin } 0.90 0.08 0.02



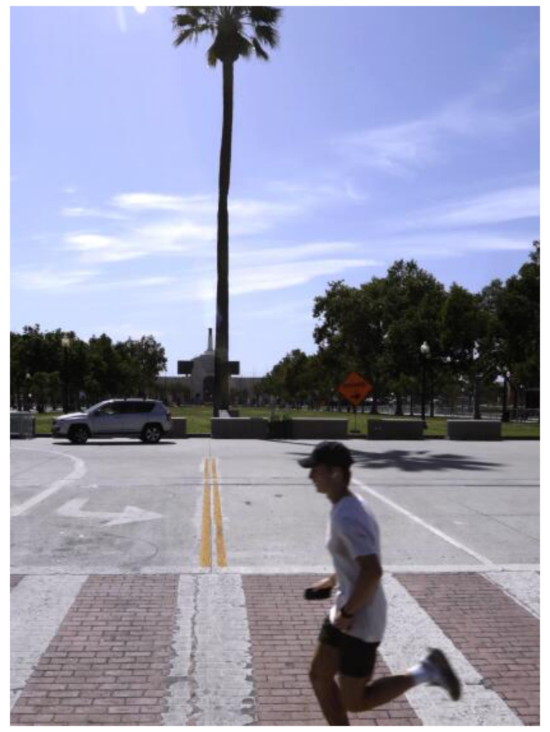
{ Tree, human, trash can } 0.70 0.20 0.05

Trained Prediction Models

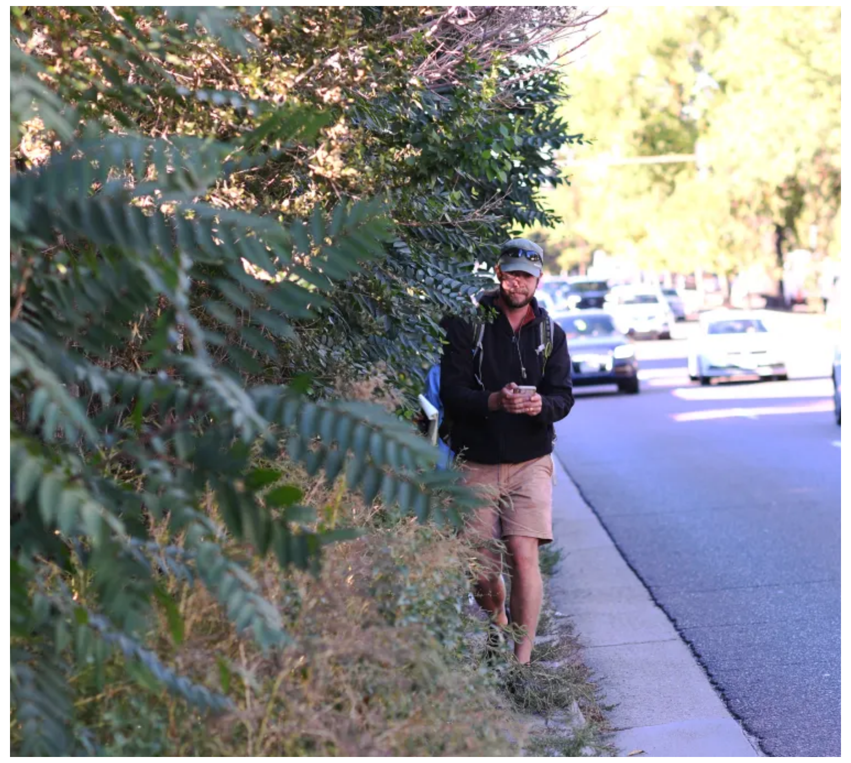
x



{ human 0.99 }



{ human, tree, bin 0.90 0.08 0.02 }

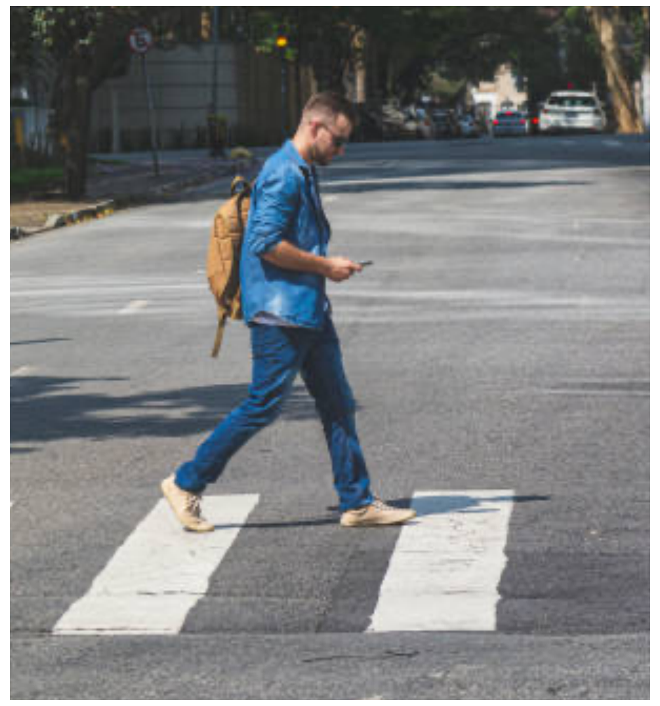


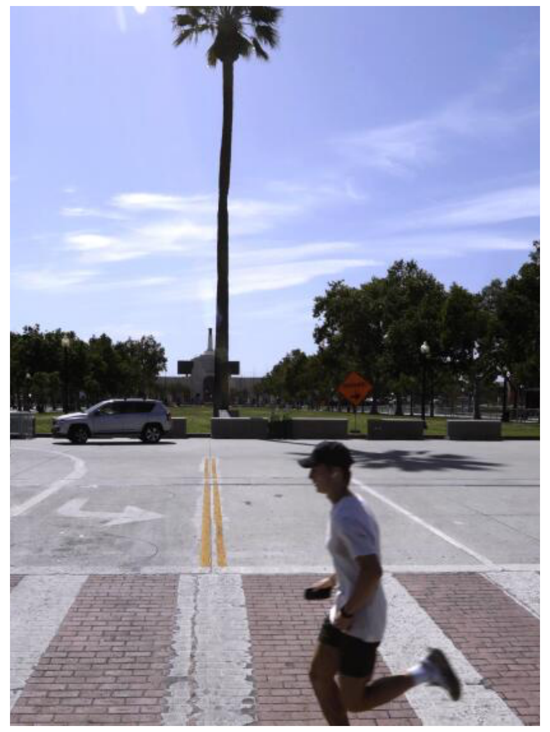
{ Tree, human, trash can 0.70 0.20 0.05 }

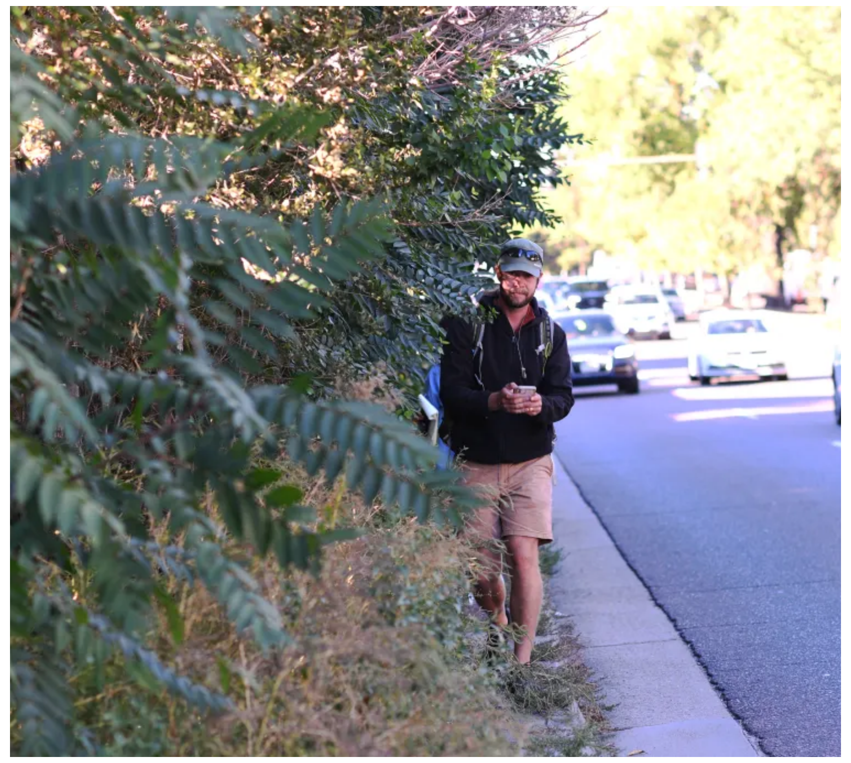
model's prediction:	human	human	tree
	✅	✅	❌

Trained Prediction Models

x





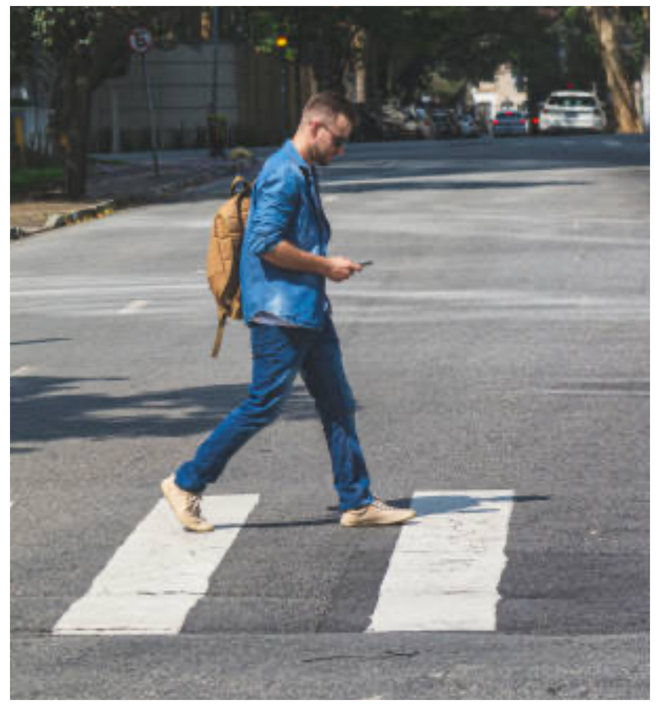
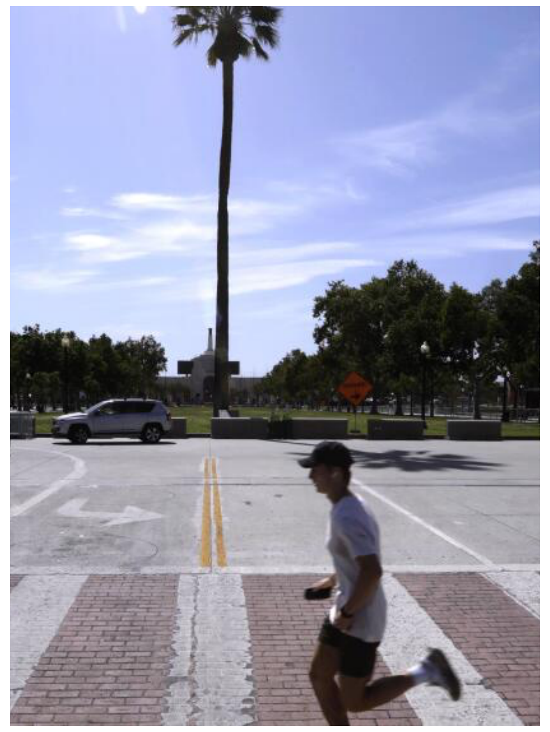
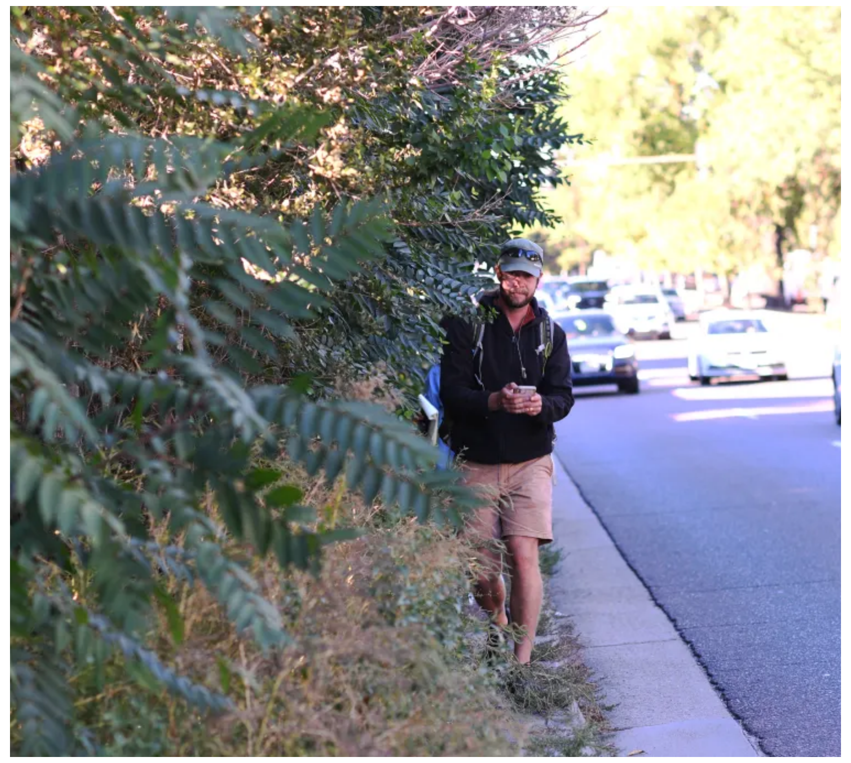


{ human 0.99 }

{ human 0.90, tree 0.08, bin 0.02 }

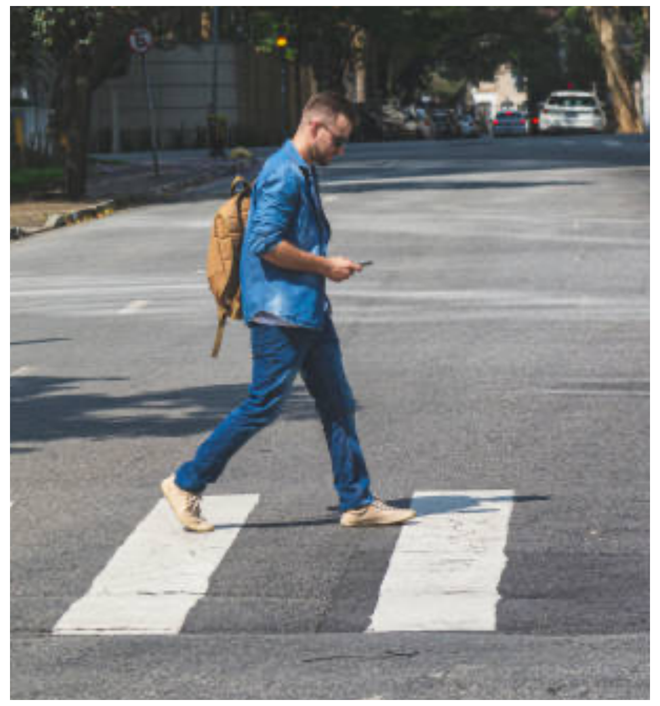
{ Tree 0.70, human 0.20, trash can 0.05 }

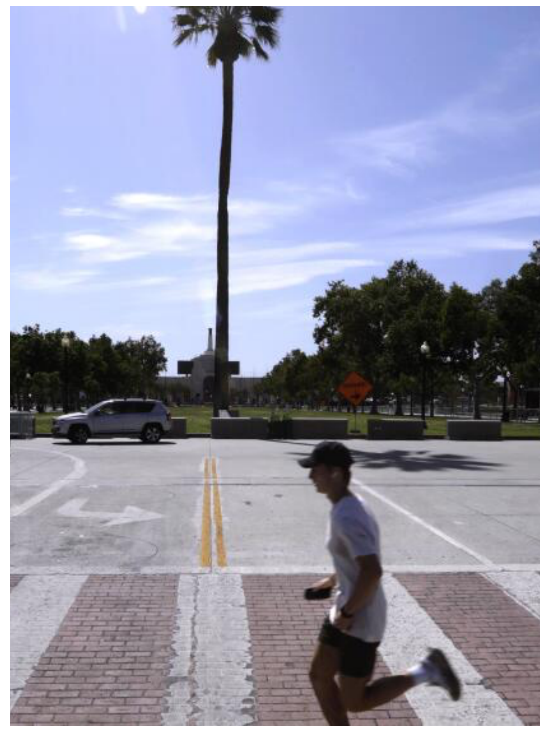
Trained Prediction Models

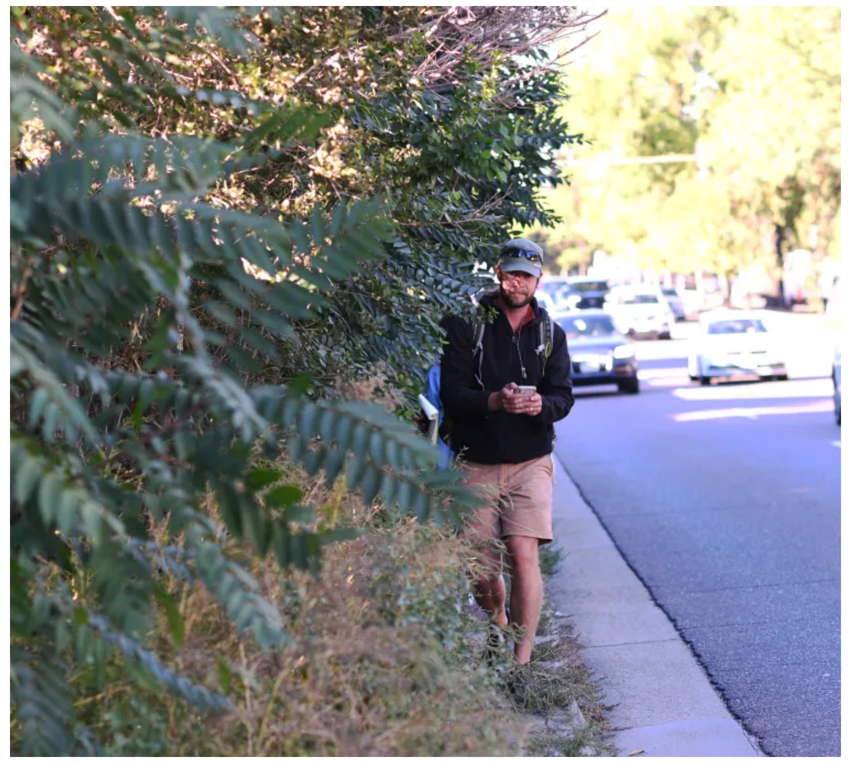
x			
	{ human } 0.99	{ human , tree, bin } 0.90 0.08 0.02	{ Tree, human, trash can } 0.70 0.20 0.05
$C(x)$ (model's conformal prediction)	{ human } ✅	{ human } ✅	{ tree , human } ✅

Trained Prediction Models

x







{ human } { human, tree, bin } { Tree, human, trash can }

0.99 — 0.90, 0.08, 0.02 — 0.70, 0.20, 0.05

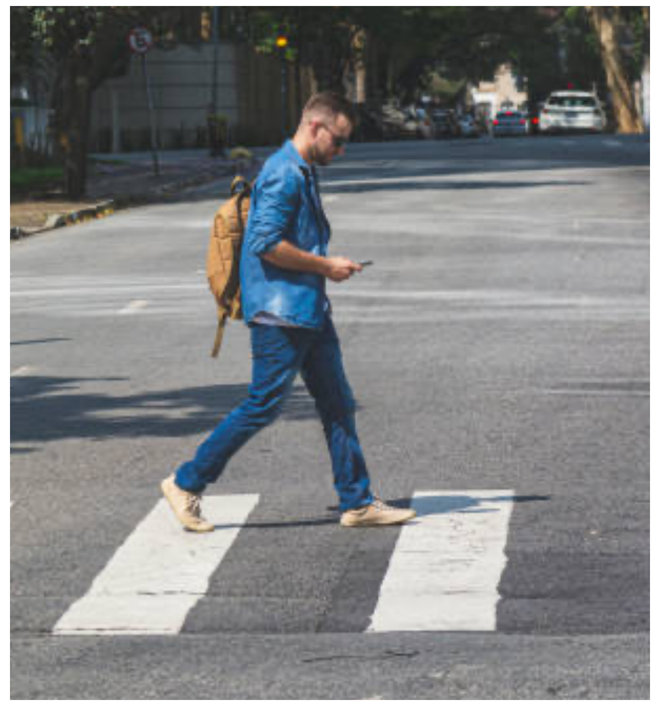
$C(x)$ { human } ✅ { human } ✅ { tree, human } ✅

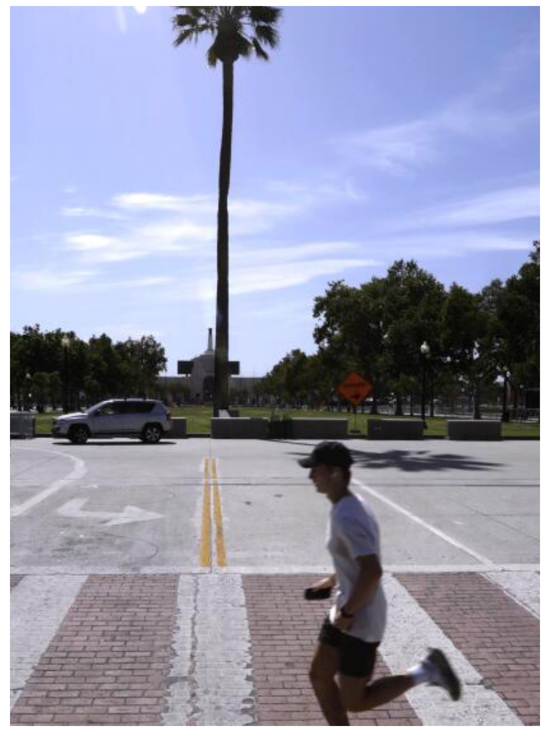
(model's conformal prediction)

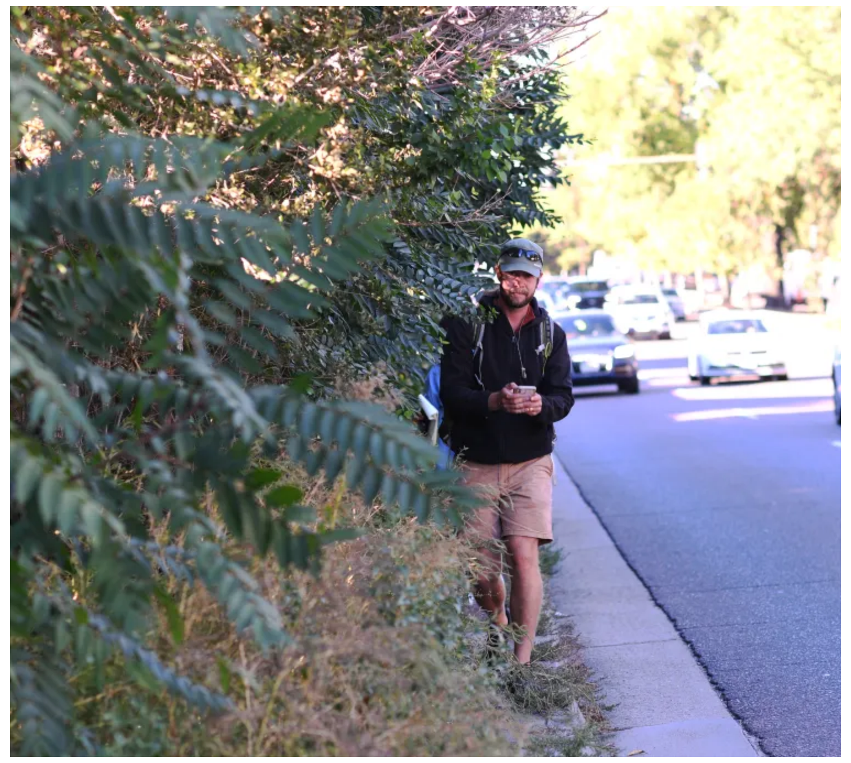
For every input x, CP will provide a set of labels, $C(x)$, that contains the true label with high probability

Trained Prediction Models

x







{ human } { human , tree, bin } { Tree, human, trash can }

0.99 | 0.90 0.08 0.02 | 0.70 0.20 0.05

$C(x)$

(model's conformal prediction)

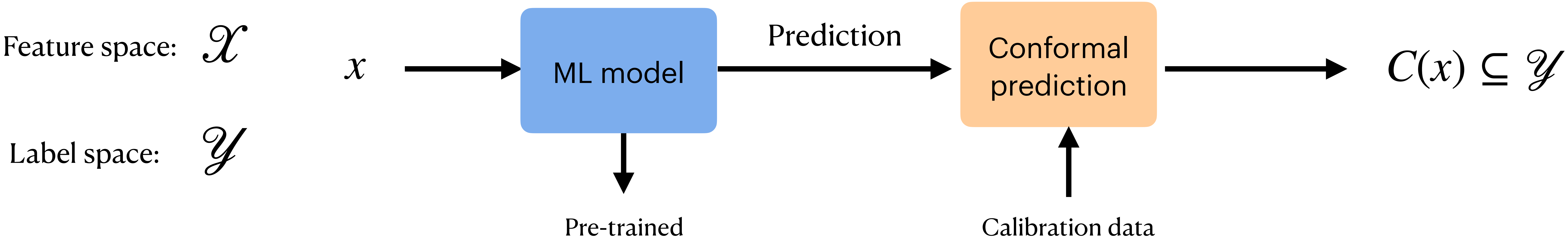
{ human } ✅ { human } ✅ { tree , human } ✅

For every input x, CP will provide a set of labels, $C(x)$, that contains the true label with high probability

$C(x)$ will be constructed using the trained model and some additional data

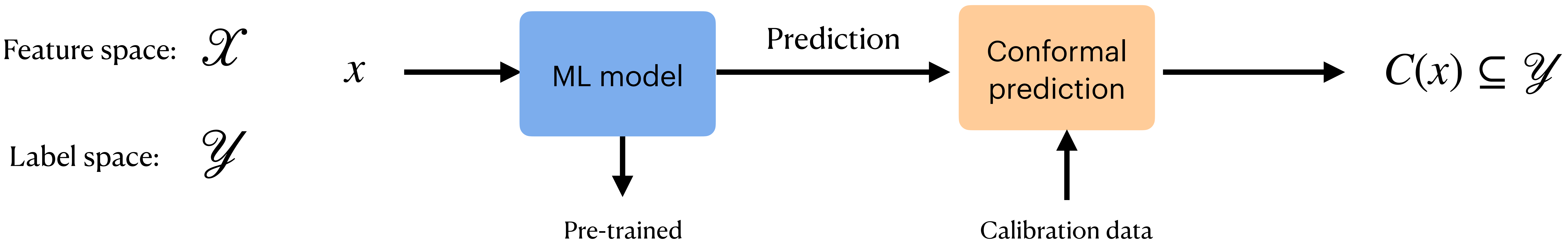
Conformal Prediction (CP)

From Predictions to Prediction Sets



Conformal Prediction (CP)

From Predictions to Prediction Sets



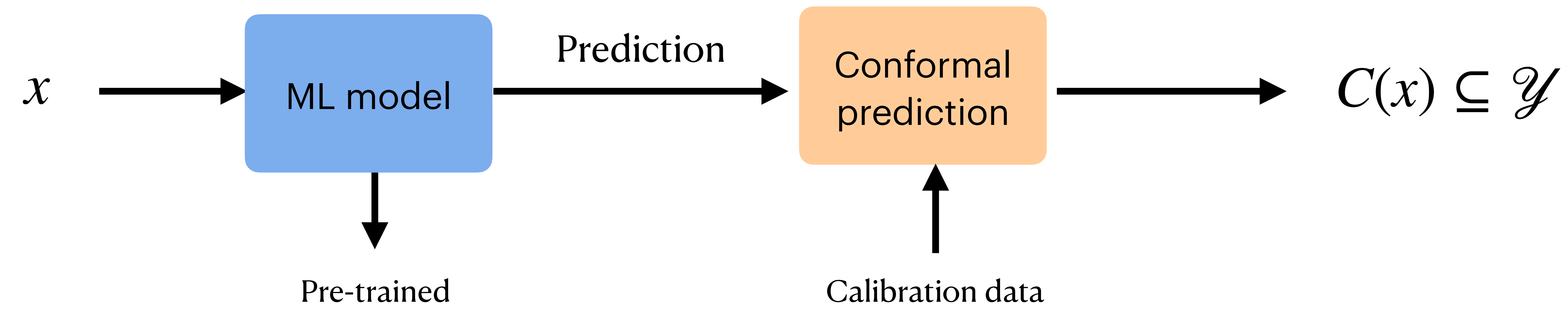
$$\Pr\{Y_{\text{test}} \in C(X_{\text{test}})\} \geq 1 - \alpha$$

Conformal Prediction (CP)

From Predictions to Prediction Sets

Feature space: $\mathscr{X}$

Label space: $\mathscr{Y}$



$$\Pr\left\{Y_{\text{test}} \in C(X_{\text{test}})\right\} \geq 1 - \alpha$$

$(X_{\text{test}}, Y_{\text{test}}) \sim \mathcal{D}$

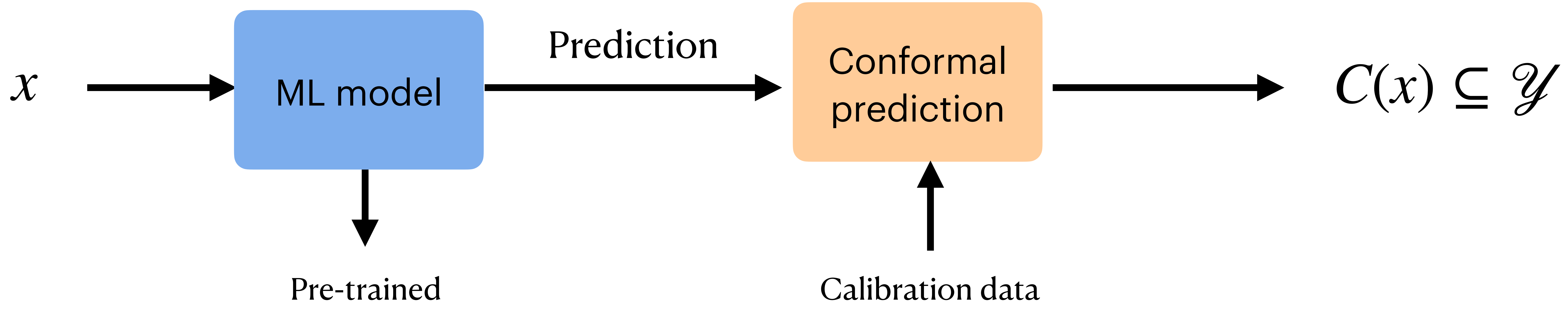
User-specified value; e.g. $1 - \alpha = 0.9$

Conformal Prediction (CP)

From Predictions to Prediction Sets

Feature space: $\mathscr{X}$

Label space: $\mathscr{Y}$



$$\Pr\{Y_{\text{test}} \in C(X_{\text{test}})\} \geq 1 - \alpha$$

$(X_{\text{test}}, Y_{\text{test}}) \sim \mathcal{D}$

User-specified value; e.g. $1 - \alpha = 0.9$

- This is called a marginal guarantee

Algorithms to Obtain Marginal Coverage

x

trained model

$\hat{\pi}_1$
$\hat{\pi}_2$
$\hat{\pi}_3$
$\vdots$
$\hat{\pi}_K$

{ `tree,` `trash can,` `human,` `motorcycle,` … }

`0.35` `0.35` `0.20` `0.05`

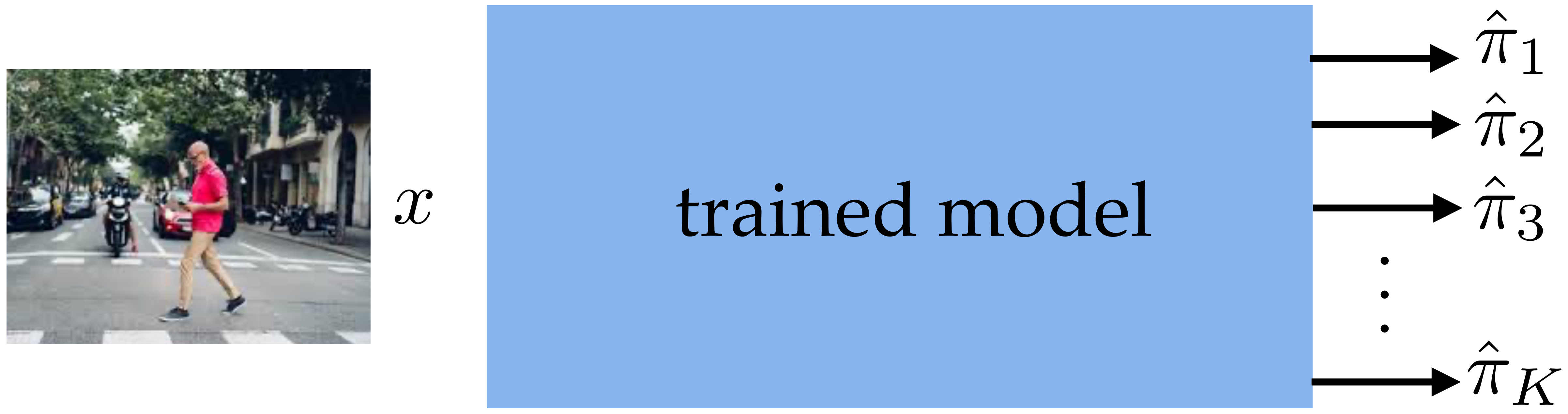
Algorithms to Obtain Marginal Coverage

{ tree, trash can, human, motorcycle, … }

0.35 0.35 0.20 0.05

- Prediction sets are constructed using a score and a threshold; Formally:

Algorithms to Obtain Marginal Coverage



{ tree, trash can, human, motorcycle, … }

0.35 0.35 0.20 0.05

- Prediction sets are constructed using a score and a threshold; Formally:

$$C(x) = \{y \in \mathcal{Y} : \quad S(x,y) \leq q\}$$

Algorithms to Obtain Marginal Coverage

{ `tree`, `trash can`, `human`, `motorcycle`, … }

`0.35` `0.35` `0.20` `0.05`

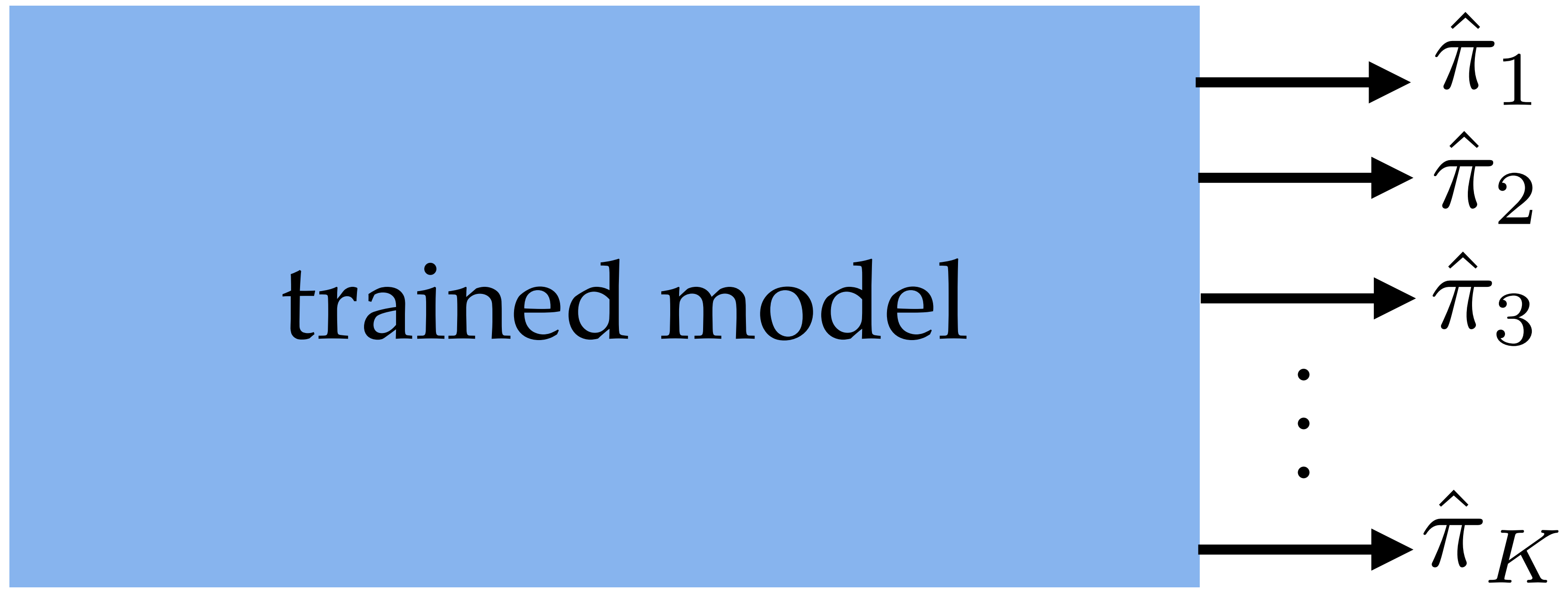
- Prediction sets are constructed using a score and a threshold; Formally:

the threshold

$$C(x) = \{y \in \mathcal{Y} : \quad S(x, y) \leq q\}$$

Algorithms to Obtain Marginal Coverage

x



{ `tree,` `trash can,` `human,` `motorcycle,` … }

`0.35` `0.35` `0.20` `0.05`

- Prediction sets are constructed using a score and a threshold; Formally:

the threshold

$$C(x) = \{y \in \mathcal{Y} : \quad S(x, y) \leq q\}$$

- How can we compute the threshold q from data?

Algorithms to Obtain Marginal Coverage

Algorithms to Obtain Marginal Coverage

- Recall:

$$C(x) = \{y \in \mathcal{Y} : \quad S(x, y) \leq q\}$$

Algorithms to Obtain Marginal Coverage

- Recall:

$$C(x) = \{y \in \mathcal{Y} : \quad S(x, y) \le q\} \quad \Rightarrow \quad \Pr\{Y \in C(X)\} = \Pr\{S(X, Y) \le q\}$$

Algorithms to Obtain Marginal Coverage

- Recall:

$$C(x) = \{y \in \mathcal{Y} : \quad S(x, y) \leq q\} \Rightarrow \Pr\{Y \in C(X)\} = \Pr\{S(X, Y) \leq q\} = 1 - \alpha$$

Algorithms to Obtain Marginal Coverage

- Recall:

$$C(x) = \{y \in \mathcal{Y} : \quad S(x, y) \leq q\} \quad \Rightarrow \quad \Pr\{Y \in C(X)\} = \Pr\{S(X, Y) \leq q\} = 1 - \alpha$$

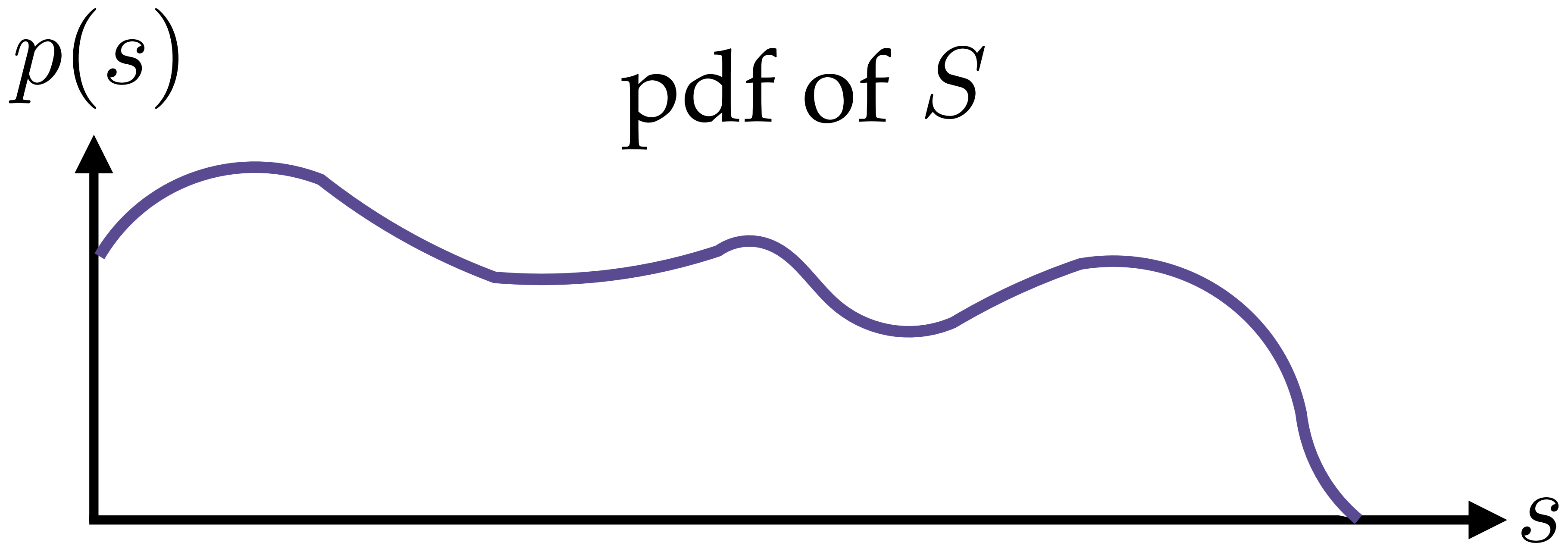
- Ideally:

Algorithms to Obtain Marginal Coverage

- Recall:

$$C(x) = \{y \in \mathcal{Y} : \quad S(x,y) \leq q\} \Rightarrow \Pr\{Y \in C(X)\} = \Pr\{S(X,Y) \leq q\} = 1 - \alpha$$

- Ideally:

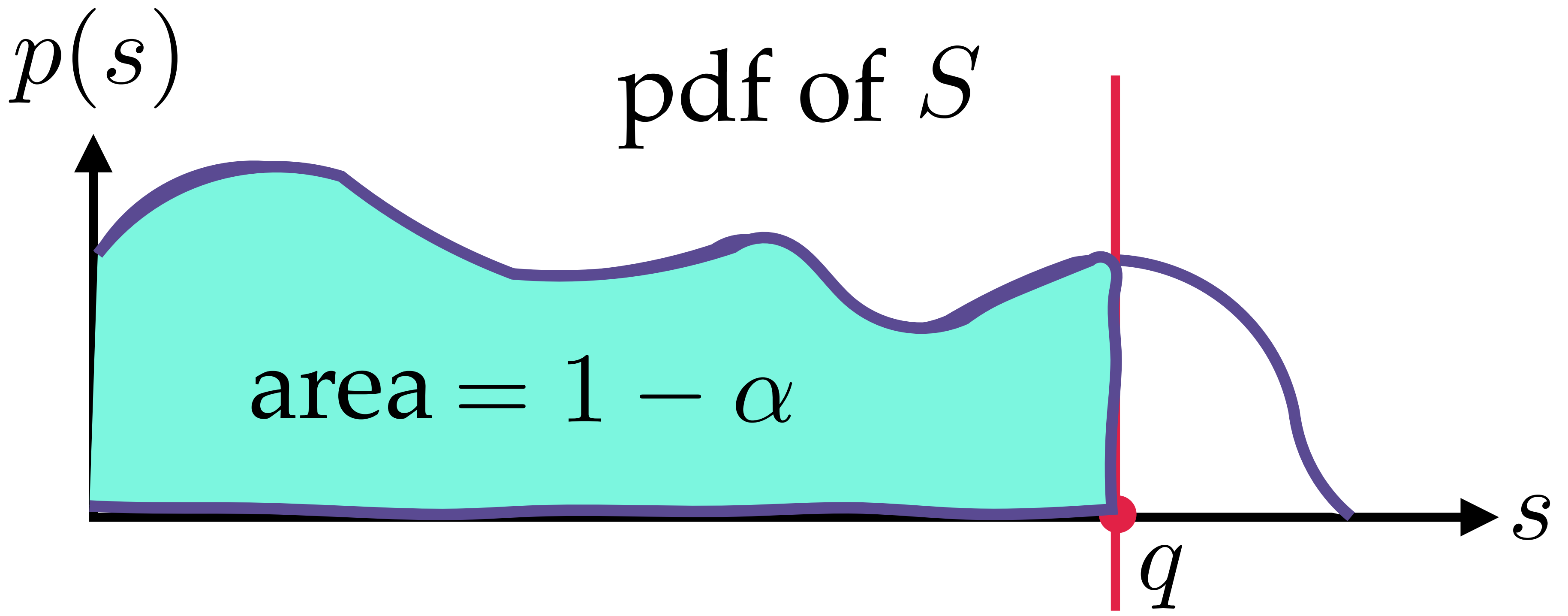


Algorithms to Obtain Marginal Coverage

- Recall:

$$C(x) = \{y \in \mathcal{Y} : \quad S(x,y) \leq q\} \Rightarrow \Pr\{Y \in C(X)\} = \Pr\{S(X,Y) \leq q\} = 1 - \alpha$$

- Ideally:

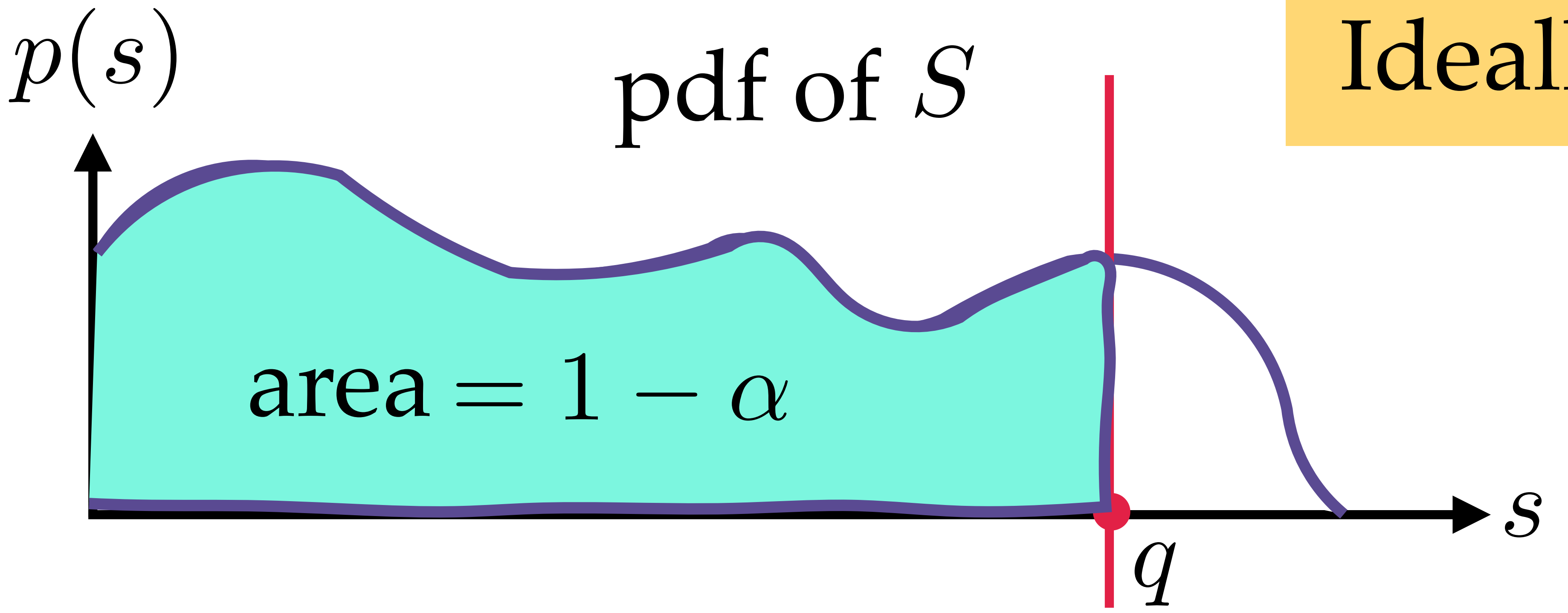


Algorithms to Obtain Marginal Coverage

- Recall:

$$C(x) = \{y \in \mathcal{Y} : \quad S(x,y) \leq q\} \Rightarrow \Pr\{Y \in C(X)\} = \Pr\{S(X,Y) \leq q\} = 1 - \alpha$$

- Ideally:



Ideally, q is the $(1-\alpha)$-quantile of the distribution of S

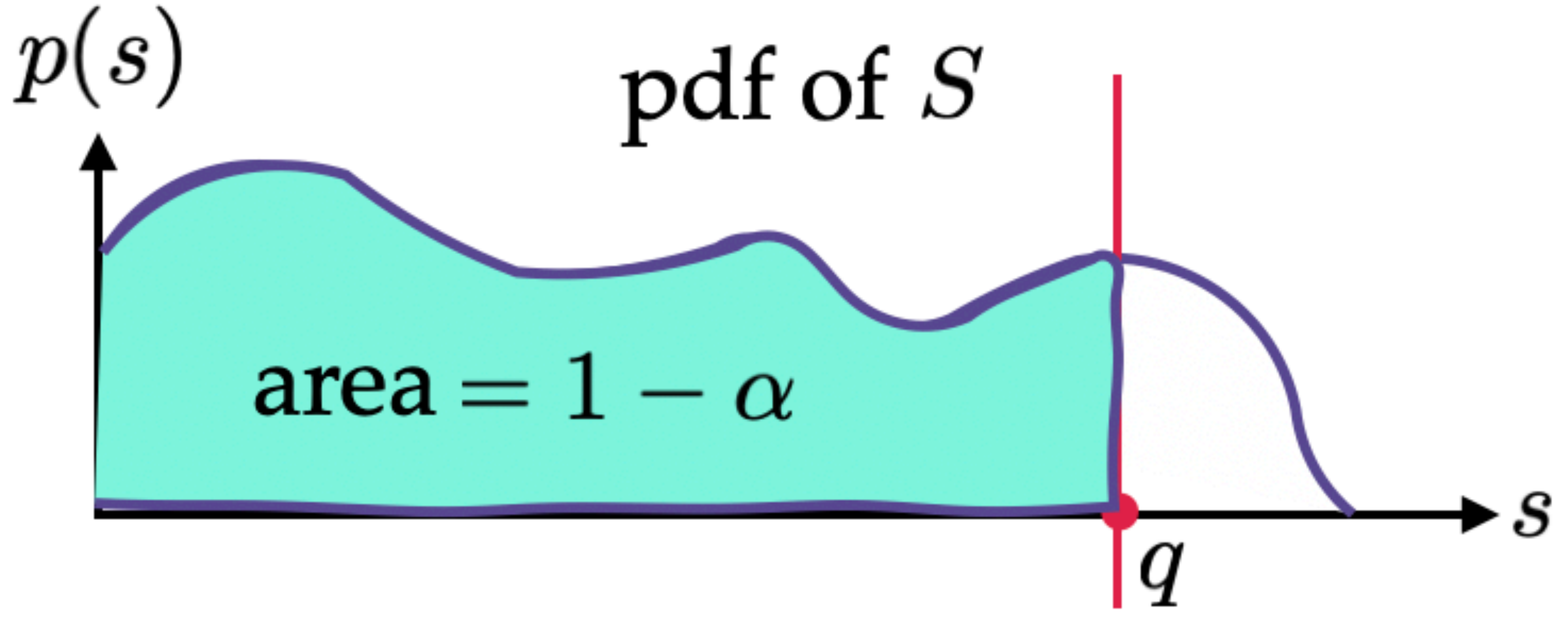
Conformal Prediction – A Promising Framework

How it works

Conformal Prediction – A Promising Framework

How it works

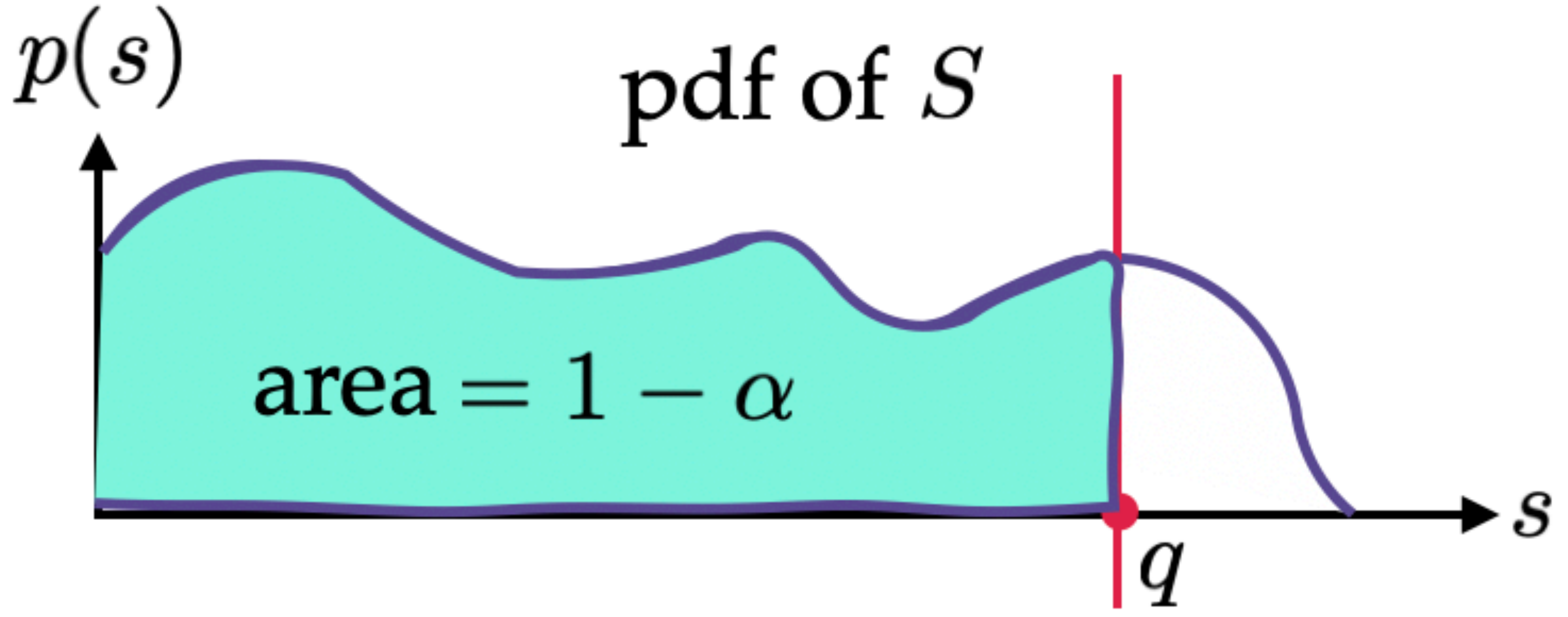
Input	Prediction	Ground truth	Non conformity score – $S(x, y) \in \mathbb{R}$
X_1	$f(X_1)$	Y_1	$S_1 = S(X_1, Y_1)$
X_2	$f(X_2)$	Y_2	$S_2 = S(X_2, Y_2)$
X_3	$f(X_3)$	Y_3	$S_3 = S(X_3, Y_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_n	$f(X_n)$	Y_n	$S_n = S(X_n, Y_n)$



Conformal Prediction – A Promising Framework

How it works

Input	Prediction	Ground truth	Non conformity score – $S(x, y) \in \mathbb{R}$
X_1	$f(X_1)$	Y_1	$S_1 = S(X_1, Y_1)$
X_2	$f(X_2)$	Y_2	$S_2 = S(X_2, Y_2)$
X_3	$f(X_3)$	Y_3	$S_3 = S(X_3, Y_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_n	$f(X_n)$	Y_n	$S_n = S(X_n, Y_n)$



$$S_1, S_2, S_3, \ldots, S_{n-1}, S_n$$

sort

$$S_{(1)} \leq S_{(2)} \leq S_{(3)} \leq \cdots \leq S_{(n-1)} \leq S_{(n)}$$

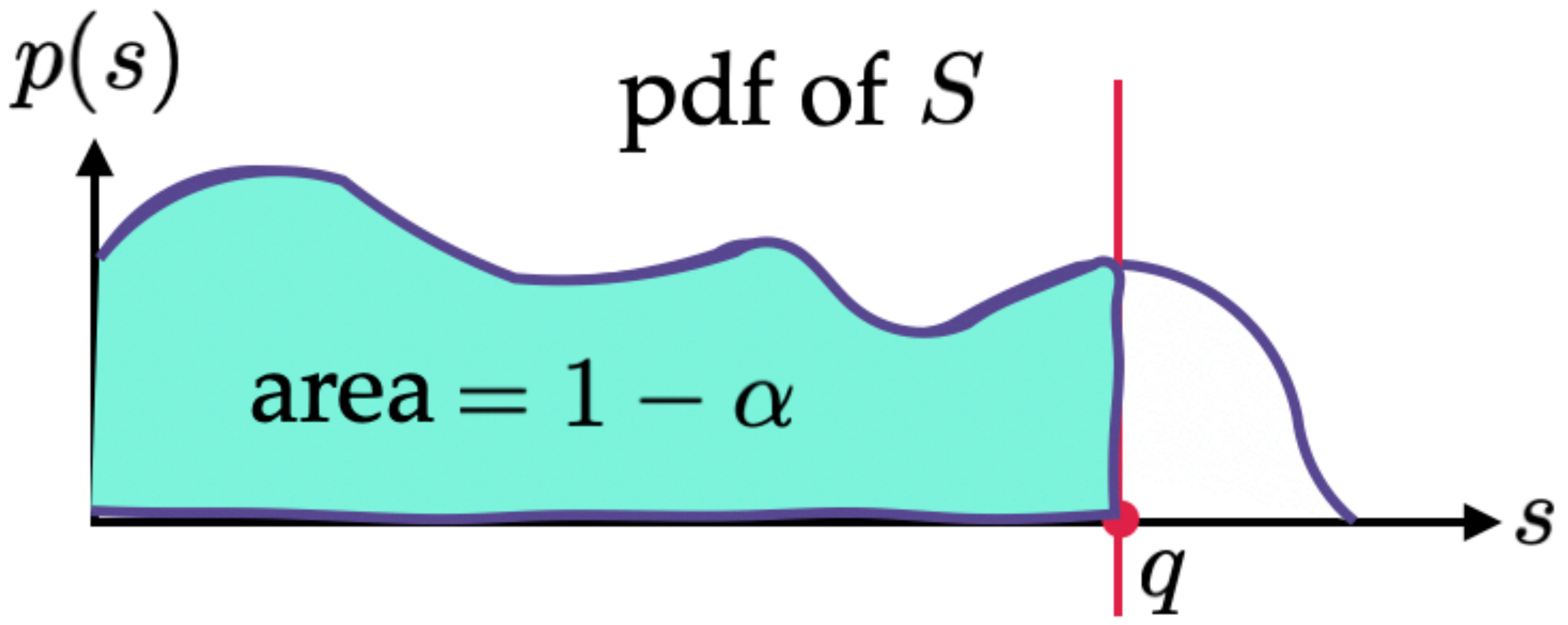
choose the value at position $\lceil (1-\alpha)(n+1) \rceil$

$$\hat{q} = S_{(\lceil (1-\alpha)(n+1) \rceil)}$$

Conformal Prediction – A Promising Framework

How it works

Input	Prediction	Ground truth	Non conformity score – $S(x, y) \in \mathbb{R}$
X_1	$f(X_1)$	Y_1	$S_1 = S(X_1, Y_1)$
X_2	$f(X_2)$	Y_2	$S_2 = S(X_2, Y_2)$
X_3	$f(X_3)$	Y_3	$S_3 = S(X_3, Y_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_n	$f(X_n)$	Y_n	$S_n = S(X_n, Y_n)$



$$\hat{q} = S_{(\lceil (1-\alpha)(n+1) \rceil)}$$

$$C(x) = \{y : S(x, y) \leq \hat{q}\}$$

Theorem: The prediction sets constructed by the Split-Conformal algorithm satisfy:

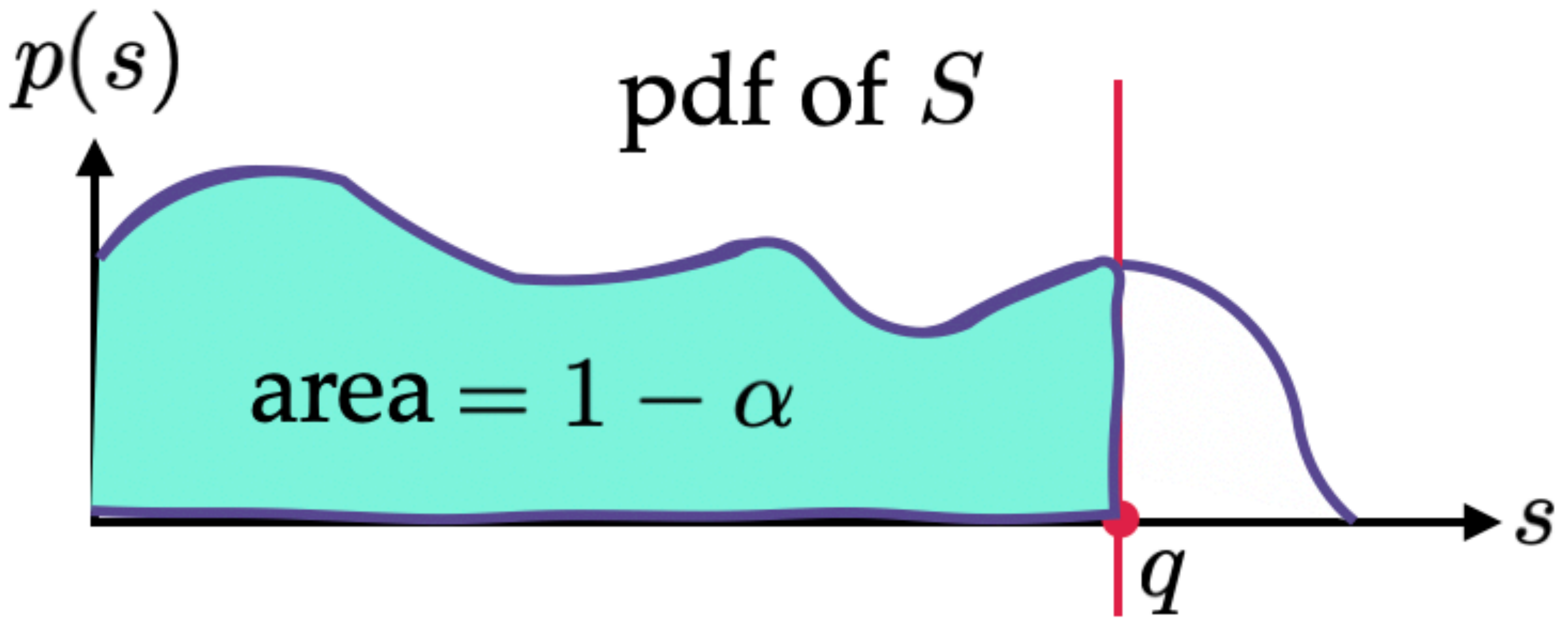
$$1 - \alpha \leq \Pr\{Y \in C(X)\} \leq 1 - \alpha + \frac{1}{n+1}$$

[Vovk, Gammerman, Saunders, 1999]

Conformal Prediction – A Promising Framework

How it works

Input	Prediction	Ground truth	Non conformity score – $S(x, y) \in \mathbb{R}$
X_1	$f(X_1)$	Y_1	$S_1 = S(X_1, Y_1)$
X_2	$f(X_2)$	Y_2	$S_2 = S(X_2, Y_2)$
X_3	$f(X_3)$	Y_3	$S_3 = S(X_3, Y_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_n	$f(X_n)$	Y_n	$S_n = S(X_n, Y_n)$



$$\hat{q} = S_{(\lceil (1-\alpha)(n+1) \rceil)}$$

$$C(x) = \{y : S(x, y) \leq \hat{q}\}$$

Theorem: The prediction sets constructed by the Split-Conformal algorithm satisfy:

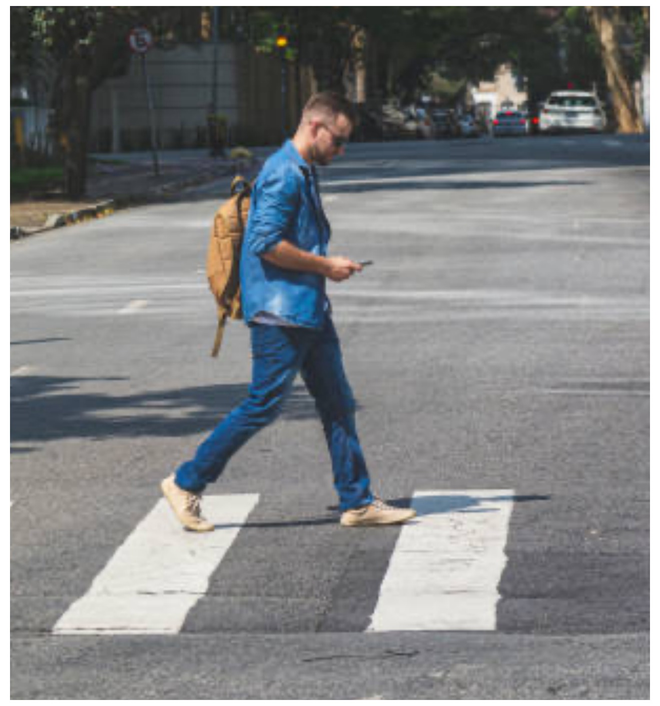
$$1 - \alpha \leq \Pr\{Y \in C(X)\} \leq 1 - \alpha + \frac{1}{n+1}$$

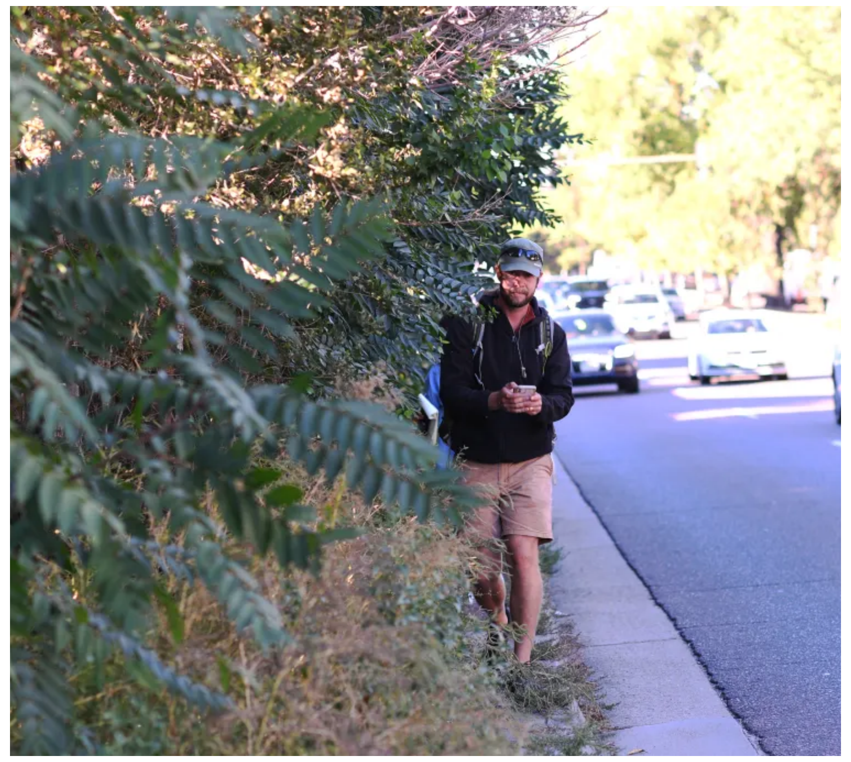
[Vovk, Gammerman, Saunders, 1999]

- Remarkably, this algorithm works no matter what the model is or what the data distribution is.

CP as a Framework for Uncertainty Quantification

x





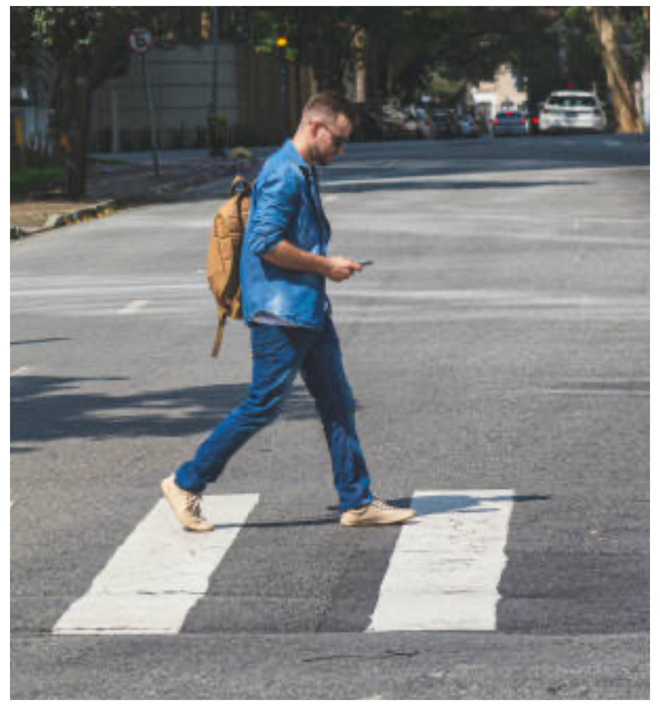
{ human 0.99 } { human, tree, bin 0.90 0.08 0.02 } { Tree, human, trash can 0.70 0.20 0.05 }

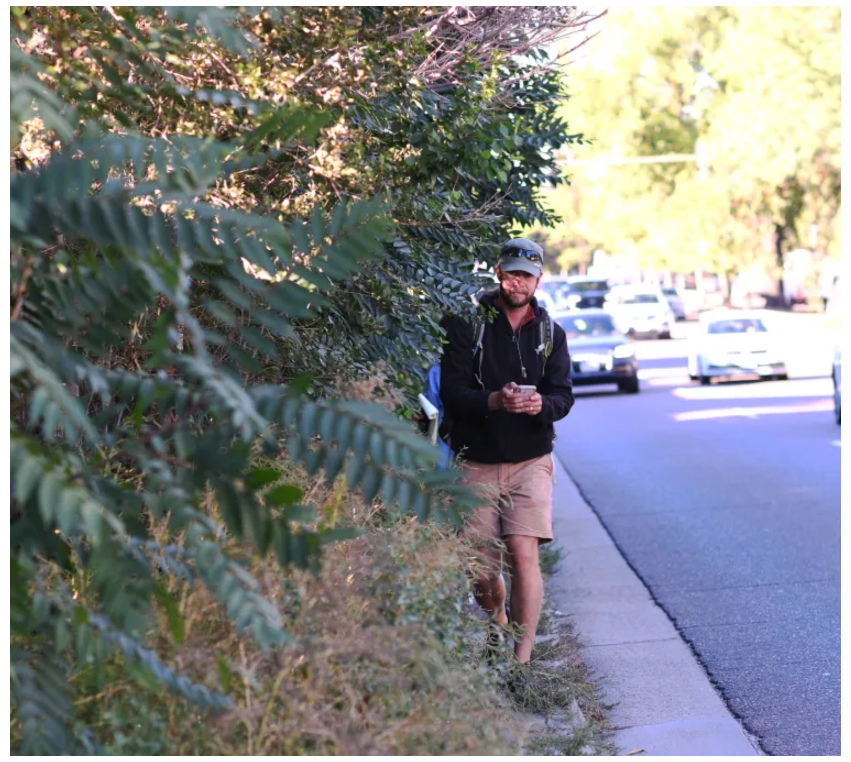
$C(x)$ { human } { human } { tree, human }

✅ ✅ ✅

CP as a Framework for Uncertainty Quantification

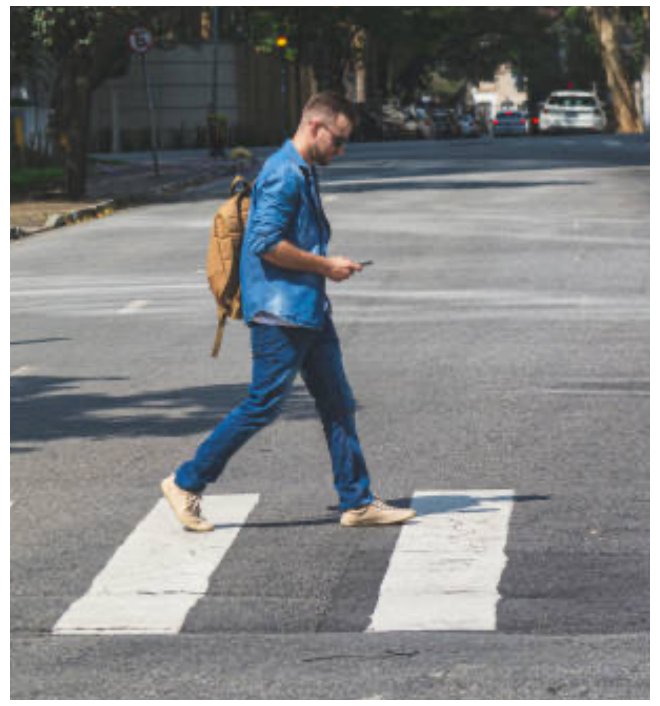
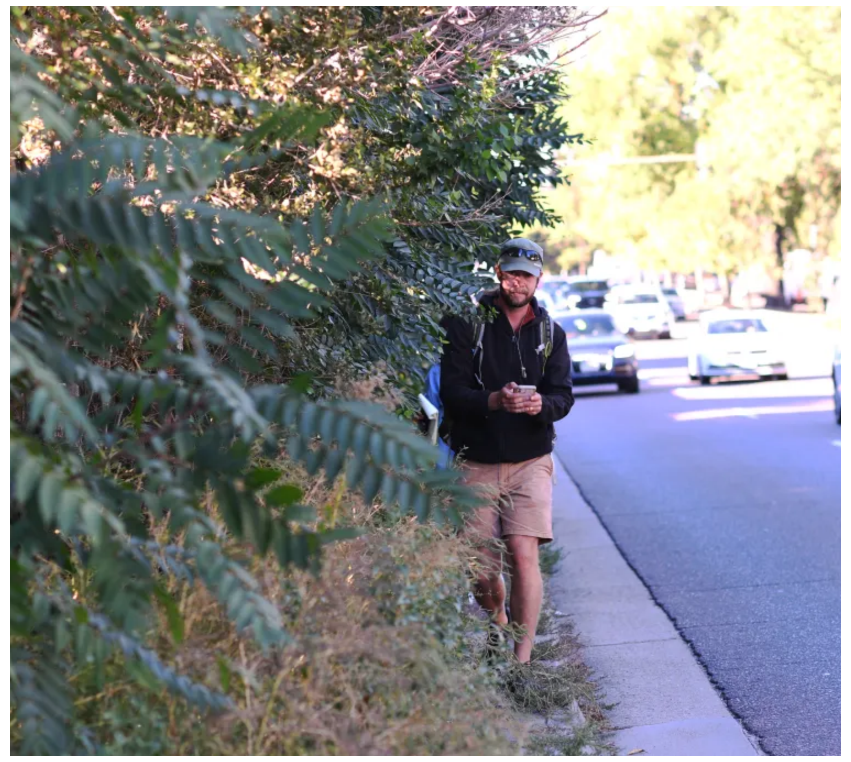
x





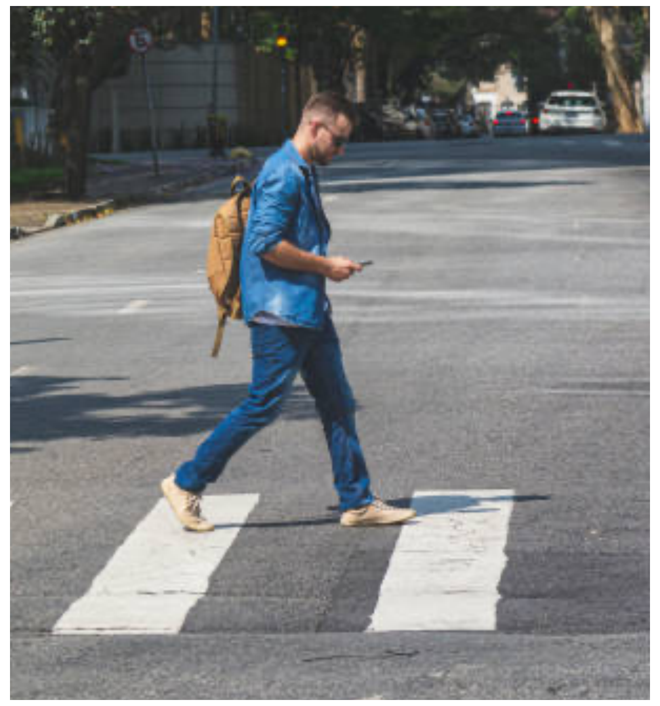
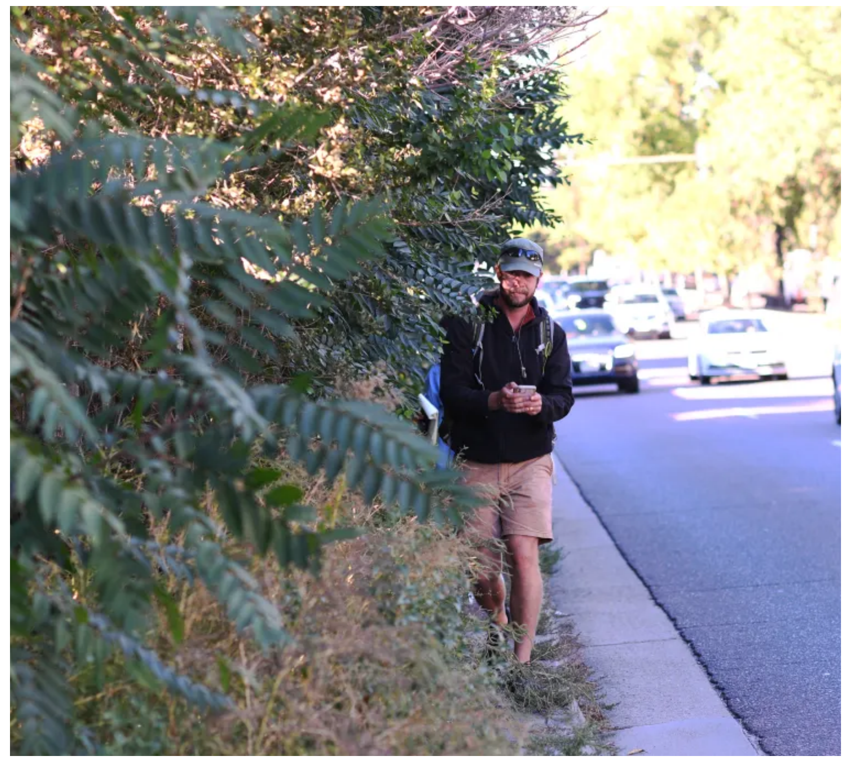
	{human 0.99}	{human 0.90, tree 0.08, bin 0.02}	{Tree 0.70, human 0.20, trash can 0.05}
$C(x)$	{ human }	{ human }	{ tree , human }
	✅	✅	✅
$\lvert C(x) \rvert$ size	1	1	2

CP as a Framework for Uncertainty Quantification

x			
	{ human } 0.99	{ human , tree, bin } 0.90 0.08 0.02	{ Tree, human, trash can } 0.70 0.20 0.05
$C(x)$	{ human } ✅	{ human } ✅	{ tree , human } ✅
$\|C(x)\|$ size	1	1	2

difficulty of the input ↑

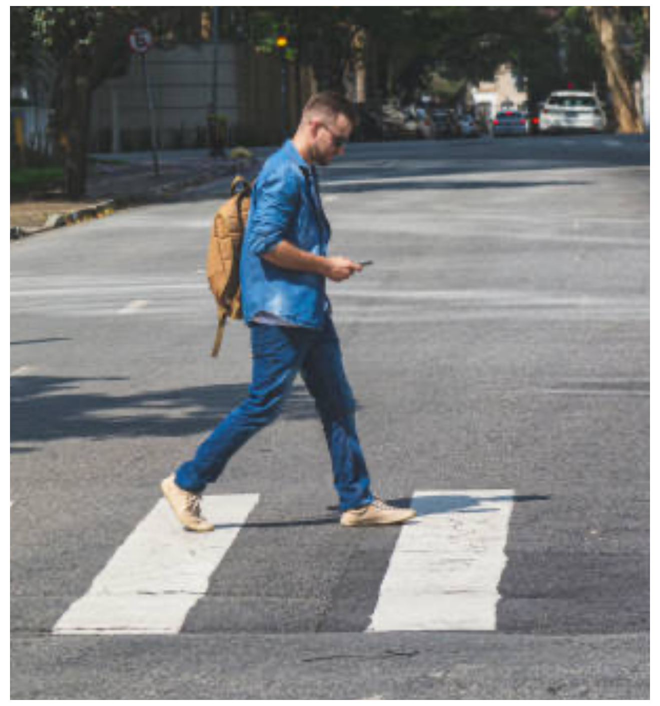
CP as a Framework for Uncertainty Quantification

x			
	{human 0.99}	{human 0.90, tree 0.08, bin 0.02}	{Tree 0.70, human 0.20, trash can 0.05}
$C(x)$	{ human }	{ human }	{ tree , human }
	✅	✅	✅
$\lvert C(x) \rvert$ size	1	1	2

difficulty of the input $\uparrow$ models' uncertainty about the label $\uparrow$

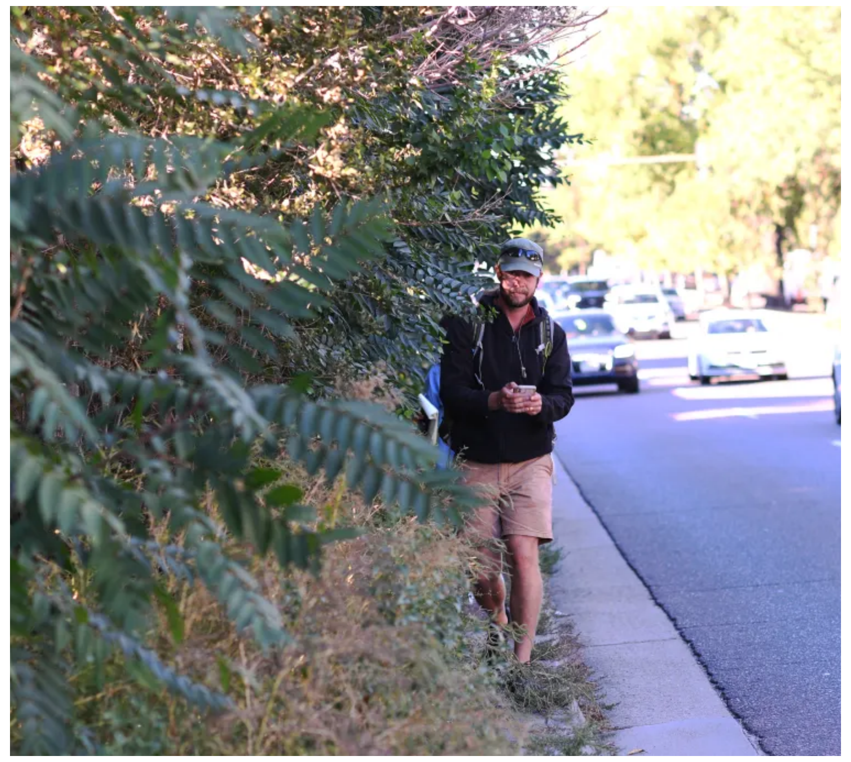
CP as a Framework for Uncertainty Quantification

x



{ human 0.99 }

{ human 0.90, tree 0.08, bin 0.02 }



{ Tree 0.70, human 0.20, trash can 0.05 }

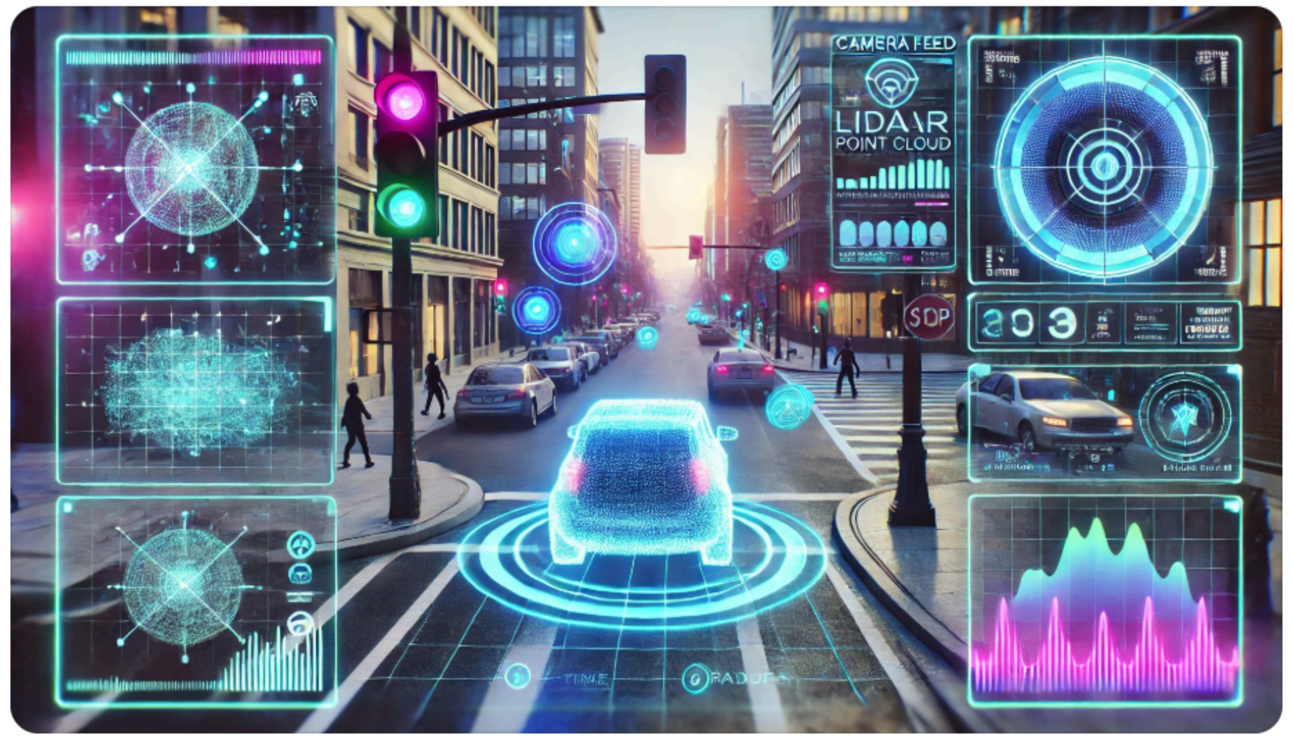
$C(x)$	{ human } ✅	{ human } ✅	{ tree , human } ✅
$\|C(x)\|$ size	1	1	2

difficulty of the input ↑ models' uncertainty about the label ↑ size of the prediction set ↑

Revisiting CP as a framework for UQ

Autonomous vehicles

x



$C(x)$

{Tree, Traffic light, Street sign, Utility pole}

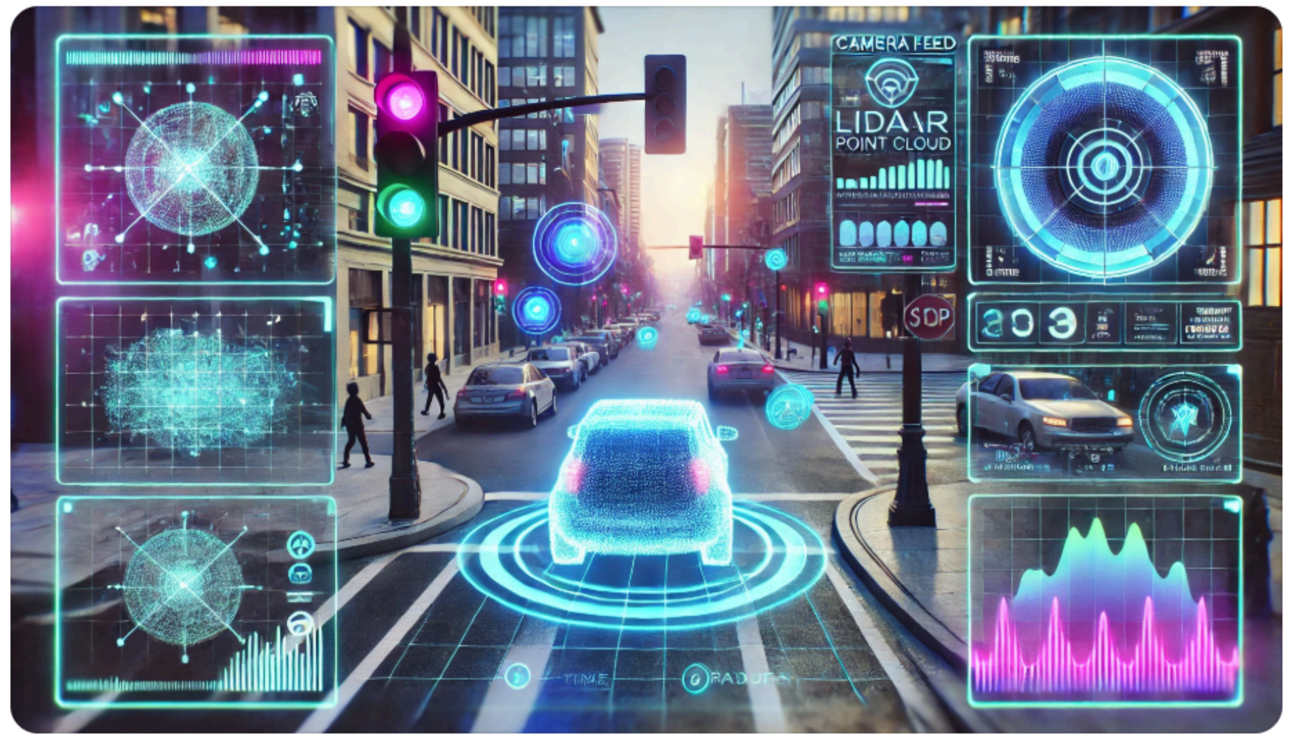
{Pedesterian, Tree}

[Related work: Risk-controlling CP , Angelopoulos et al '22, Lindermann et al '23]

Revisiting CP as a framework for UQ

Autonomous vehicles

x



$C(x)$ {Tree, Traffic light, Street sign, Utility pole}   {Pedesterian, Tree}

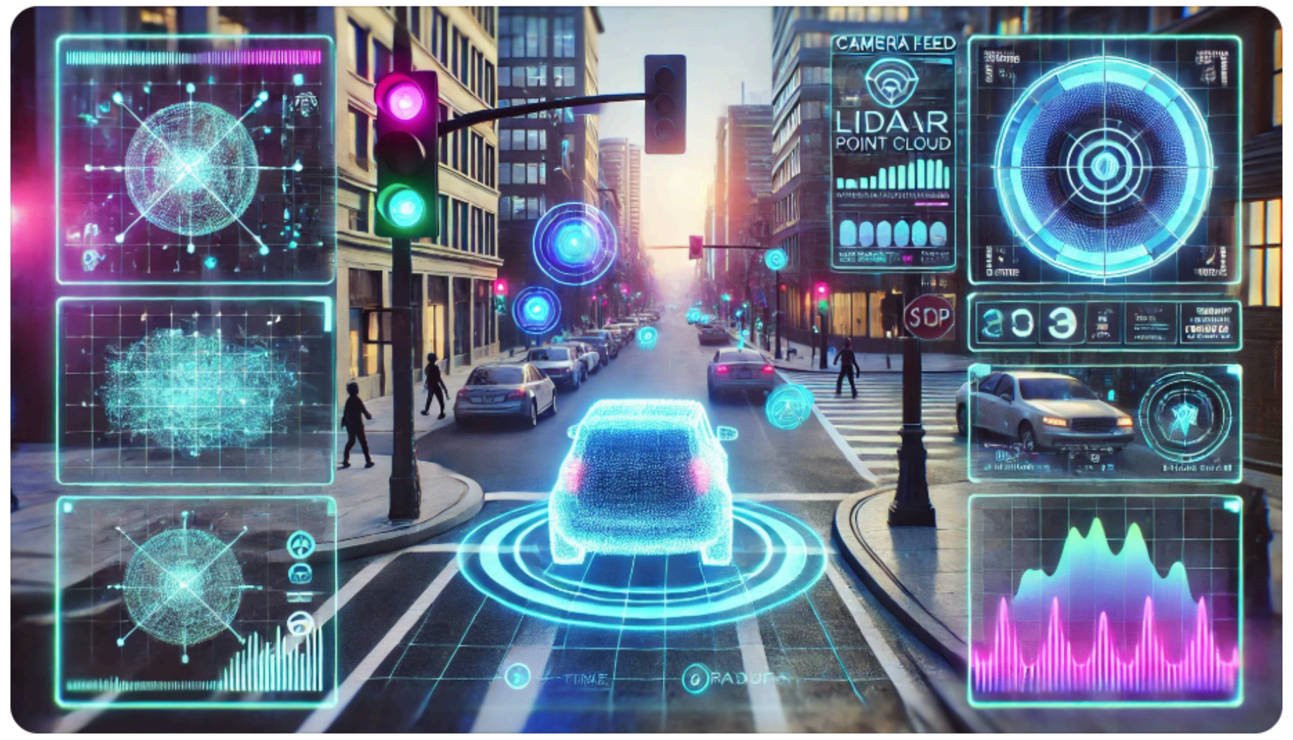
$|C(x)|$ 4   2

[Related work: Risk-controlling CP , Angelopoulos et al '22, Lindermann et al '23]

Revisiting CP as a framework for UQ

Autonomous vehicles

x

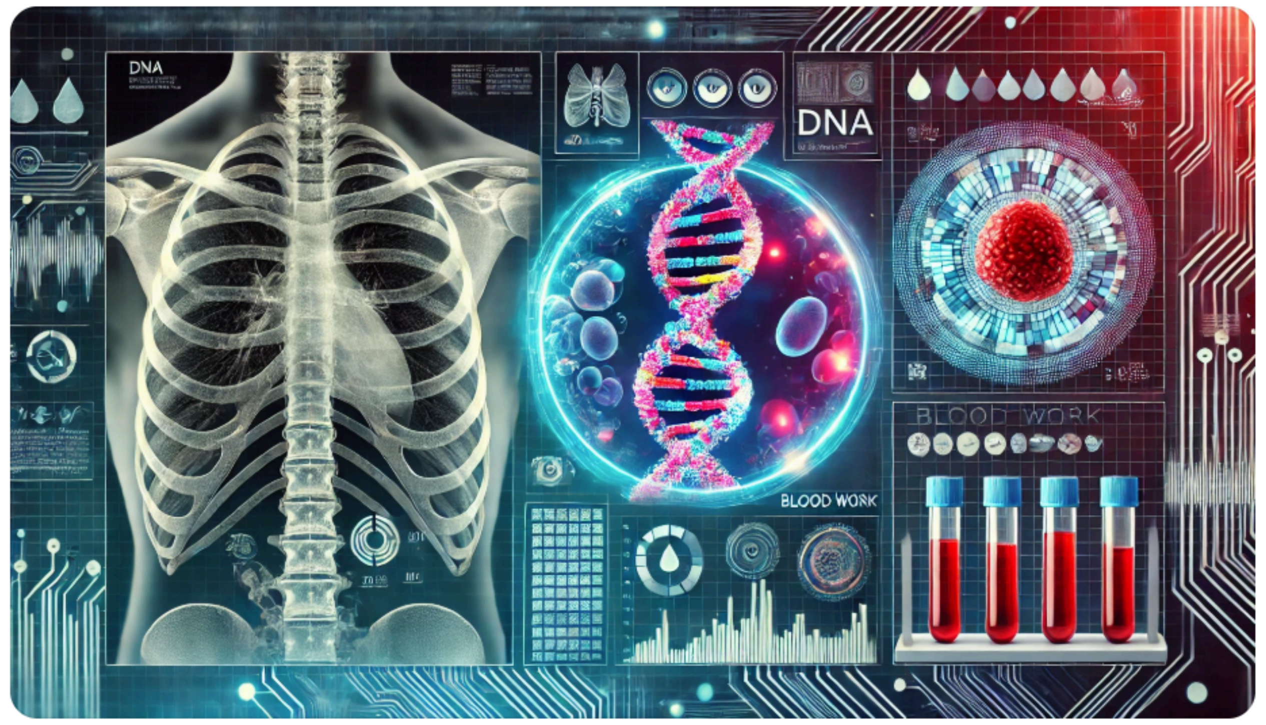


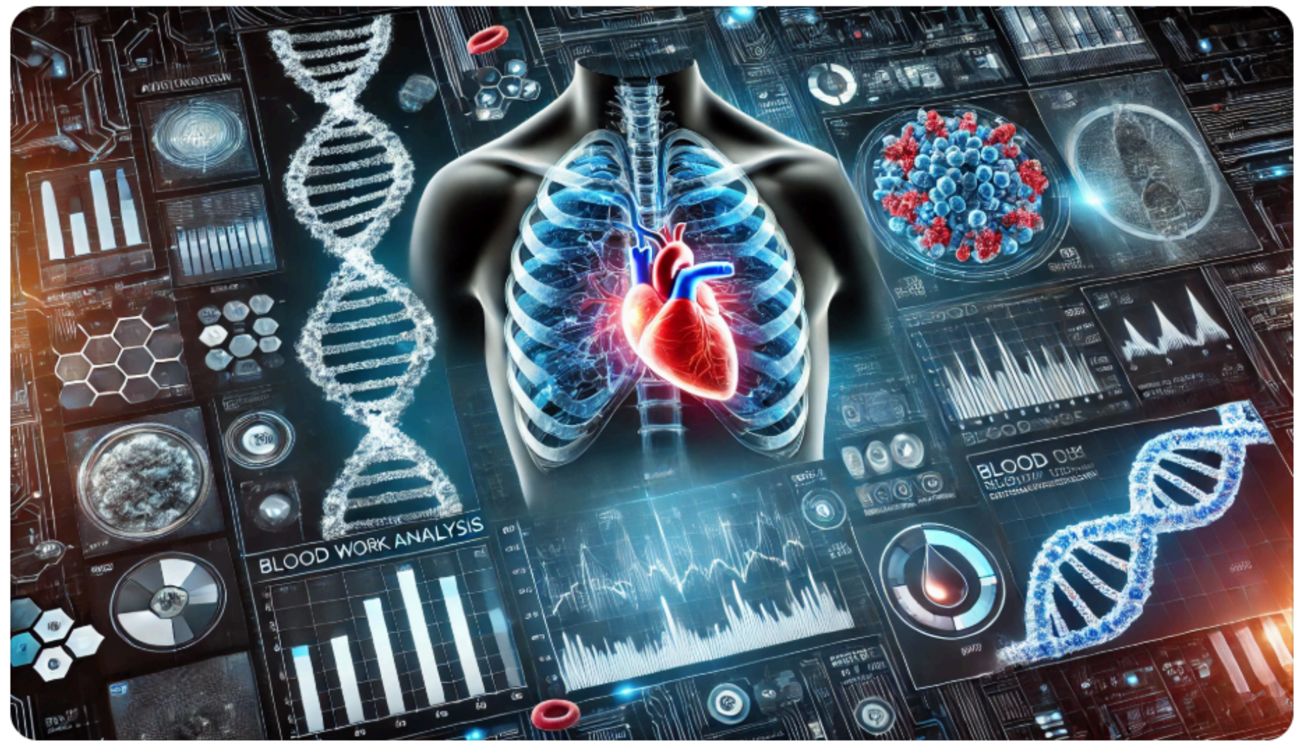
$C(x)$ {Tree, Traffic light, Street sign, Utility pole} {Pedesterian, Tree}

$|C(x)|$ 4 2

Clinical medicine

x





$C(x)$ {Cold, Flu, Covid, Allergies} {Cold, Lung cancer}

$|C(x)|$ 4 2

[Related work: Risk-controlling CP , Angelopoulos et al '22, Lindermann et al '23]

Revisiting CP as a framework for UQ

$C(x)$

$\{\text{Pedesterian, Tree}\}$

UQ should be decision informed!

[Related work: Risk-controlling CP , Angelopoulos et al '22, Lindermann et al '23]

What are prediction sets good for?

What are prediction sets good for?

What are prediction sets good for?

1) What kind of downstream decision making process make prediction sets the correct notion of UQ?

What are prediction sets good for?

1) What kind of downstream decision making process make prediction sets the correct notion of UQ?

"Well designed" prediction sets are a sufficient statistic for risk averse decision makers who wish to optimize their value at risk.

What are prediction sets good for?

1) What kind of downstream decision making process make prediction sets the correct notion of UQ?

"Well designed" prediction sets are a sufficient statistic for risk averse decision makers who wish to optimize their value at risk.

2) What is the optimal policy that a risk averse decision maker should use to map prediction sets to actions?

What are prediction sets good for?

1) What kind of downstream decision making process make prediction sets the correct notion of UQ?

"Well designed" prediction sets are a sufficient statistic for risk averse decision makers who wish to optimize their value at risk.

2) What is the optimal policy that a risk averse decision maker should use to map prediction sets to actions?

A simple max min policy is an optimal map from prediction sets to actions.

What are prediction sets good for?

1) What kind of downstream decision making process make prediction sets the correct notion of UQ?

"Well designed" prediction sets are a sufficient statistic for risk averse decision makers who wish to optimize their value at risk.

2) What is the optimal policy that a risk averse decision maker should use to map prediction sets to actions?

A simple max min policy is an optimal map from prediction sets to actions.

3) How can we derive prediction sets that are optimal for such decision makers?

What are prediction sets good for?

1) What kind of downstream decision making process make prediction sets the correct notion of UQ?

"Well designed" prediction sets are a sufficient statistic for risk averse decision makers who wish to optimize their value at risk.

2) What is the optimal policy that a risk averse decision maker should use to map prediction sets to actions?

A simple max min policy is an optimal map from prediction sets to actions.

3) How can we derive prediction sets that are optimal for such decision makers?

We will drive an explicit characterization of these sets over expectation and also provide a finite sample approximation of them.

Decision Making Pipeline

Decision Making Pipeline

$\mathcal{X}$ $\mathcal{Y}$ $\mathcal{A}$

Decision Making Pipeline

$\mathscr{X}$ $\mathscr{Y}$ $\mathscr{A}$

Decision Making Pipeline

$\mathscr{X}$ $\mathscr{Y}$ $\mathscr{A}$

Patients data including X-ray, bloodwork, etc.

Decision Making Pipeline

$\mathscr{X}$

Patients data including X-ray, bloodwork, etc.

$\mathscr{Y}$

Diagnosis

$\mathscr{A}$

Decision Making Pipeline

$\mathcal{X}$

Patients data including X-ray, bloodwork, etc.

$\mathcal{Y}$

Diagnosis

$\mathcal{A}$

Treatment

Decision Making Pipeline

$\mathcal{X}$

$\mathcal{Y}$

$\mathcal{A}$

Patients data including X-ray, bloodwork, etc.

Diagnosis

Treatment

Decision Making Pipeline

	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{A}$
	Patients data including X-ray, bloodwork, etc.	Diagnosis	Treatment
	Sensory inputs including LiDAR, vision, etc.		

Decision Making Pipeline

	$\mathscr{X}$	$\mathscr{Y}$	$\mathscr{A}$
	Patients data including X-ray, bloodwork, etc.	Diagnosis	Treatment
	Sensory inputs including LiDAR, vision, etc.	Position of objects, pedestrian, cars, etc, in the next second.	

Decision Making Pipeline

	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{A}$
	Patients data including X-ray, bloodwork, etc.	Diagnosis	Treatment
	Sensory inputs including LiDAR, vision, etc.	Position of objects, pedestrian, cars, etc, in the next second.	Car inputs

Decision Making Pipeline

	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{A}$
	Patients data including X-ray, bloodwork, etc.	Diagnosis	Treatment
	Sensory inputs including LiDAR, vision, etc.	Position of objects, pedestrian, cars, etc, in the next second.	Car inputs
			

Decision Making Pipeline

	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{A}$
	Patients data including X-ray, bloodwork, etc.	Diagnosis	Treatment
	Sensory inputs including LiDAR, vision, etc.	Position of objects, pedestrian, cars, etc, in the next second.	Car inputs
	Historical trends, news, etc.		

Decision Making Pipeline

	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{A}$
	Patients data including X-ray, bloodwork, etc.	Diagnosis	Treatment
	Sensory inputs including LiDAR, vision, etc.	Position of objects, pedestrian, cars, etc, in the next second.	Car inputs
	Historical trends, news, etc.	State of the stock market for tomorrow.	

Decision Making Pipeline

	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{A}$
	Patients data including X-ray, bloodwork, etc.	Diagnosis	Treatment
	Sensory inputs including LiDAR, vision, etc.	Position of objects, pedestrian, cars, etc, in the next second.	Car inputs
	Historical trends, news, etc.	State of the stock market for tomorrow.	Portfolio design

Decision Making Pipeline

	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{A}$
	Patients data including X-ray, bloodwork, etc.	Diagnosis	Treatment
	Sensory inputs including LiDAR, vision, etc.	Position of objects, pedestrian, cars, etc, in the next second.	Car inputs
	Historical trends, news, etc.	State of the stock market for tomorrow.	Portfolio design
⋮			

Decision Making Pipeline

	$\mathscr{X}$	$\mathscr{Y}$	$\mathscr{A}$
	Patients data including X-ray, bloodwork, etc.	Diagnosis	Treatment
	Sensory inputs including LiDAR, vision, etc.	Position of objects, pedestrian, cars, etc, in the next second.	Car inputs
	Historical trends, news, etc.	State of the stock market for tomorrow.	Portfolio design
⋮			

Utility map: $u(a, y) \in \mathbb{R}$ → It captures the preference of decision maker.

Decision Making Pipeline

$\mathcal{X}$

$\mathcal{A}$

Patients data including X-ray, bloodwork, etc.

Treatment

Sensory inputs including LiDAR, vision, etc.

Car inputs

Historical trends, news, etc.

Portfolio design

⋮

Utility map: $u(a, y) \in \mathbb{R}$ → It captures the preference of decision maker.

The order of things: $X \in \mathcal{X} \to a \in \mathcal{A} \to u(a, Y)$

Decision Making Pipeline

$\mathscr{X}$

$\mathscr{A}$

Patients data including X-ray, bloodwork, etc.

Treatment

Instead, relying on (uncertain) predictions.

Sensory inputs including LiDAR, vision, etc.

Car inputs

Historical trends, news, etc.

Portfolio design

$\vdots$

Utility map: $u(a, y) \in \mathbb{R}$ → It captures the preference of decision maker.

The order of things: $X \in \mathscr{X} \to a \in \mathscr{A} \to u(a, Y)$

Risk Averse Decision Making

Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Safety guarantee:

$$\Pr[u(a(X), Y) \geq \nu(X)] \geq 1 - \alpha$$

Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Risk Averse Decision Policy Optimization (RA-DPO):

$$\underset{a(\cdot),\, \nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr[u(a(X), Y) \geq \nu(X)] \geq 1 - \alpha$$

Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Risk Averse Decision Policy Optimization (RA-DPO):

$$\underset{a(\cdot),\, \nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr[u(a(X), Y) \geq \nu(X)] \geq 1 - \alpha$$

Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Risk Averse Decision Policy Optimization (RA-DPO):

$$\underset{a(\cdot),\, \nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr[u(a(X), Y) \geq \nu(X)] \geq 1 - \alpha$$

Value at risk!

Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

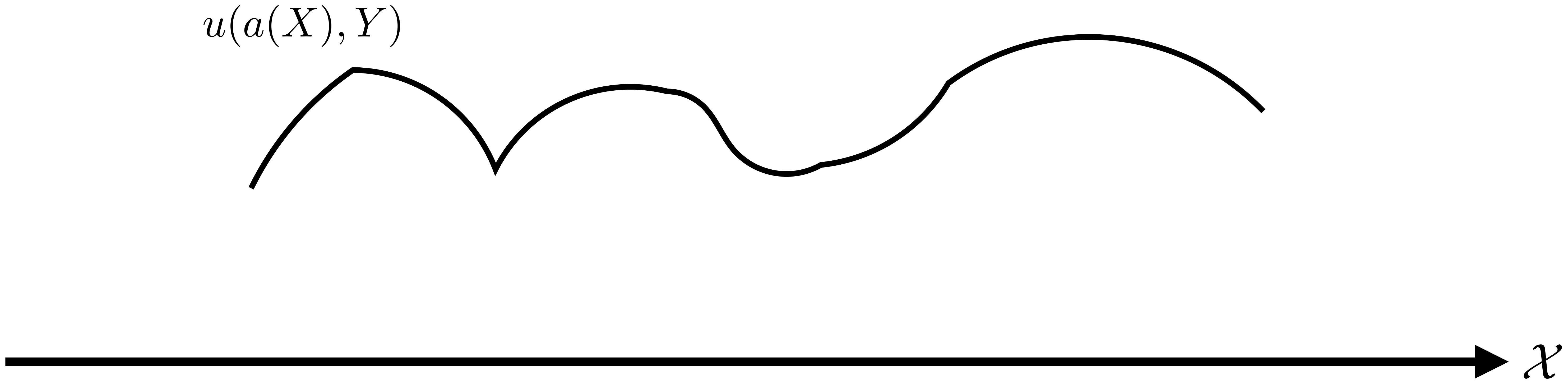
$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Risk Averse Decision Policy Optimization (RA-DPO):

$$\underset{a(\cdot),\,\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr[u(a(X),Y) \geq \nu(X)] \geq 1-\alpha$$

Value at risk!



Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

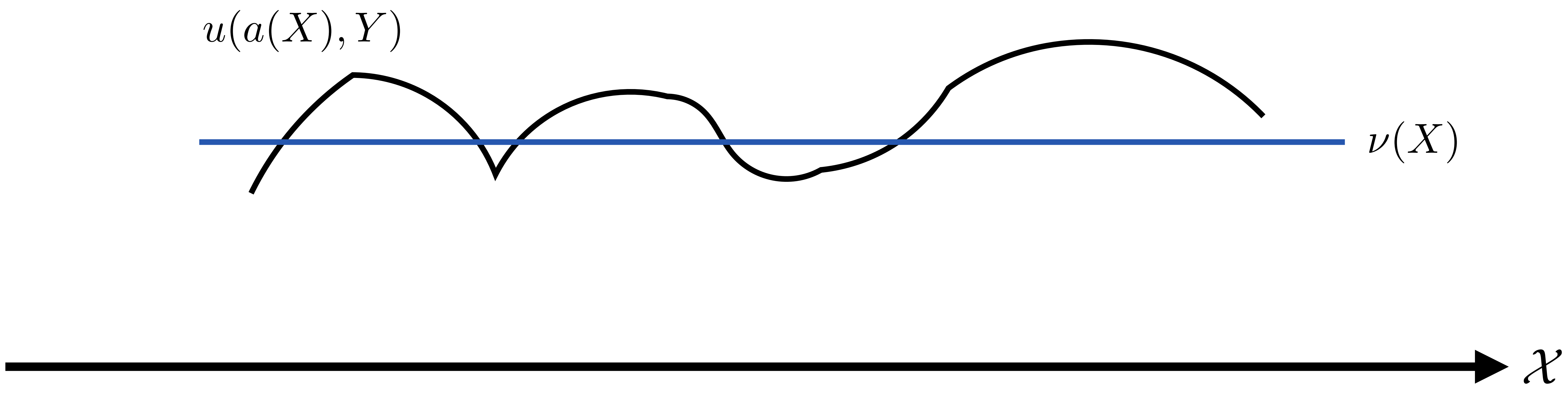
$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Risk Averse Decision Policy Optimization (RA-DPO):

$$\underset{a(\cdot),\,\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr[u(a(X),Y) \geq \nu(X)] \geq 1-\alpha$$

Value at risk!



Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

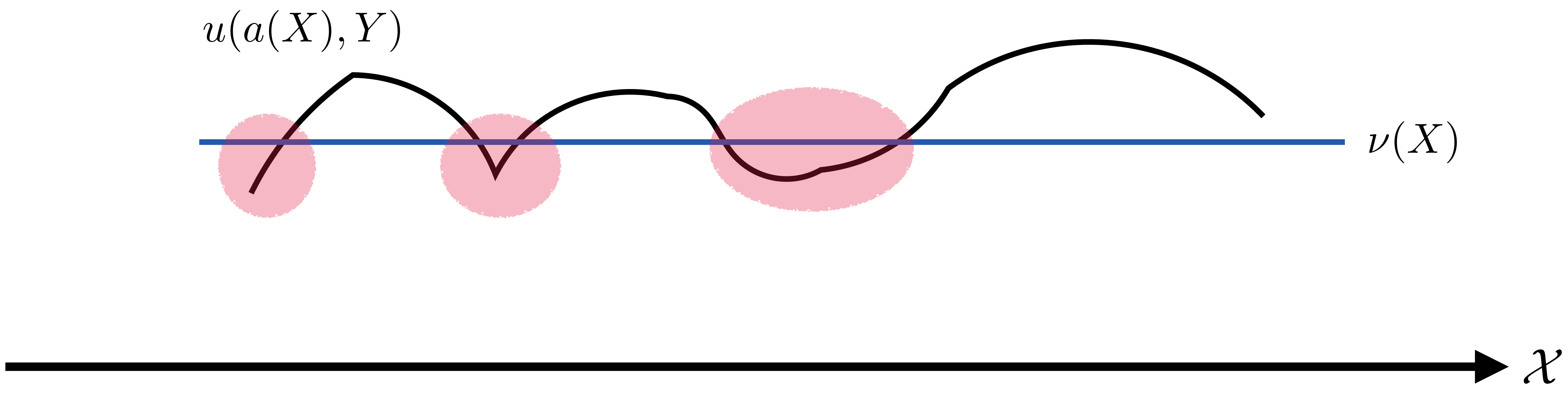
$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Risk Averse Decision Policy Optimization (RA-DPO):

$$\underset{a(\cdot),\, \nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr[u(a(X), Y) \geq \nu(X)] \geq 1 - \alpha$$

Value at risk!



Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

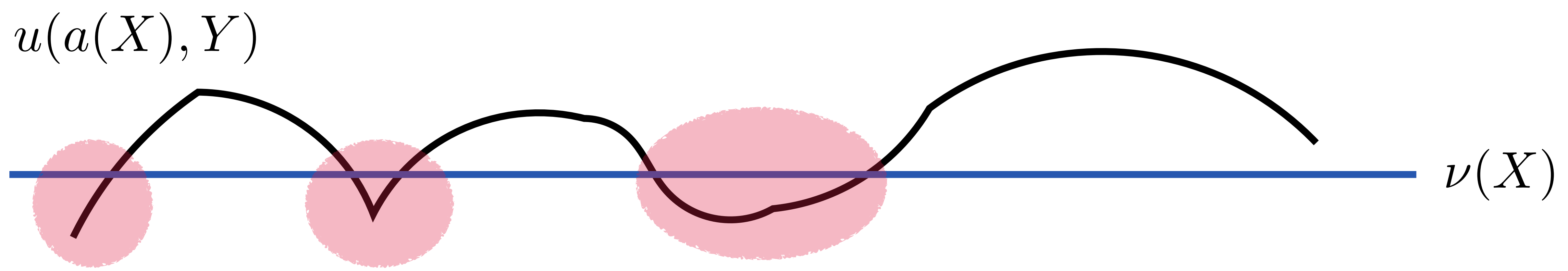
$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Risk Averse Decision Policy Optimization (RA-DPO):

$$\underset{a(\cdot),\, \nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr[u(a(X), Y) \geq \nu(X)] \geq 1 - \alpha$$

Value at risk!



$$\Pr[u(a(X), Y) \geq \nu(X)] \geq 1 - \alpha$$

$\mathcal{X}$

Risk Averse Decision Making

Action policy

$$a(\cdot) : \mathcal{X} \to \mathcal{A}$$

Utility certificate

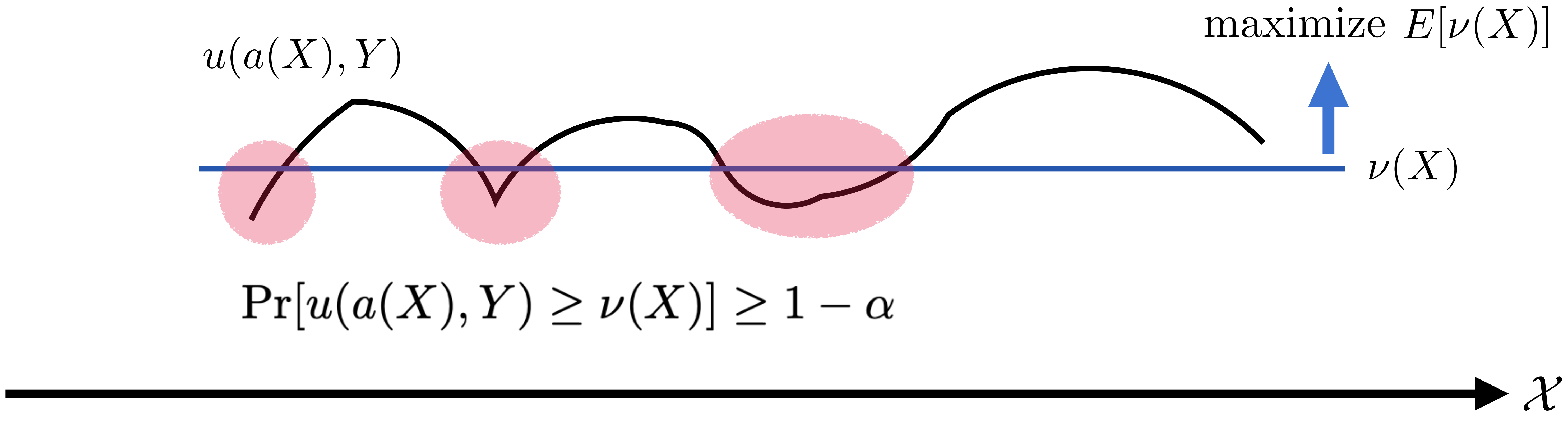
$$\nu(\cdot) : \mathcal{X} \to \mathbb{R}$$

Risk Averse Decision Policy Optimization (RA-DPO):

$$\underset{a(\cdot),\,\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr[u(a(X),Y) \geq \nu(X)] \geq 1-\alpha$$

Value at risk!



Risk Averse Decision Making

A prediction set perspective

Risk Averse Decision Making

A prediction set perspective

$$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y} \qquad\qquad \Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$$

Risk Averse Decision Making

A prediction set perspective

$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$

$$\Pr_{(X,Y)}\left[Y \in C(X)\right] \geq 1 - \alpha$$

Goal: $\pi(.) : 2^{\mathcal{Y}} \to \mathcal{A}$

Risk Averse Decision Making

A prediction set perspective

$$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$$

$$\Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$$

Goal: $\pi(.) : 2^{\mathcal{Y}} \rightarrow \mathcal{A}$

Risk Averse Decision Making

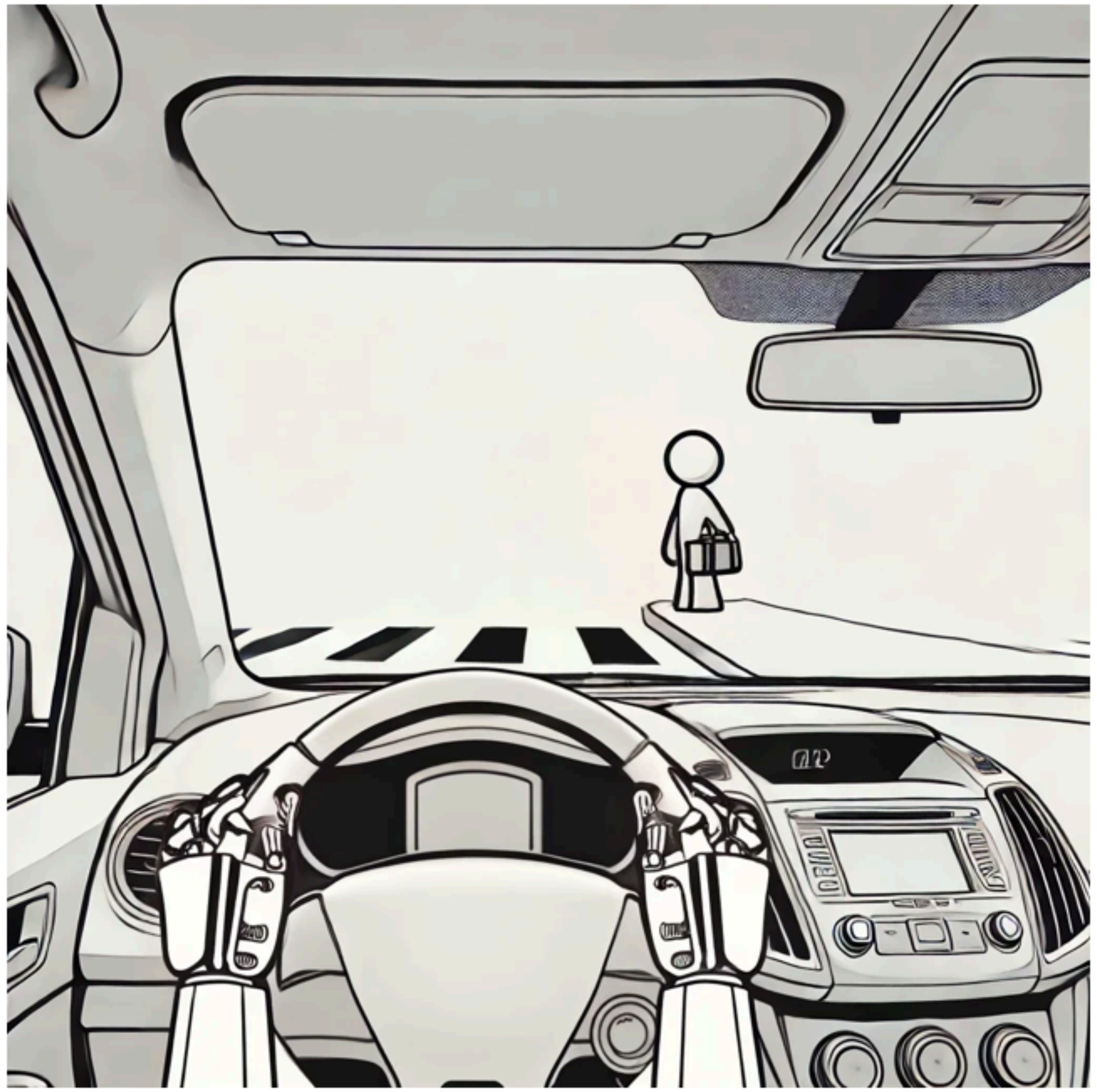
A prediction set perspective

$$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y} \qquad \Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$$

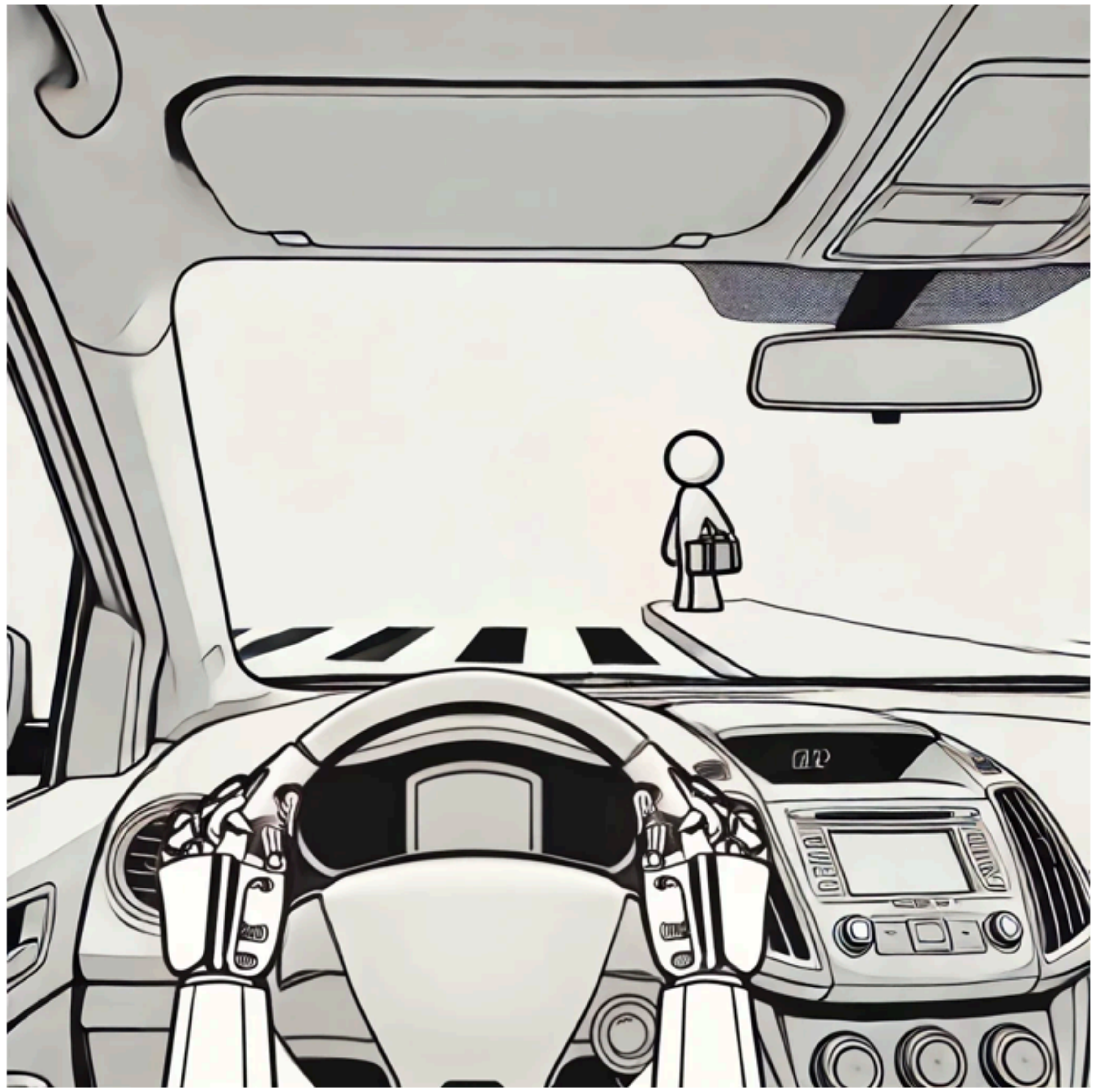
Goal: $\pi(.) : 2^{\mathcal{Y}} \to \mathcal{A}$

What policy π do humans typically choose?

Risk Averse Decision Making



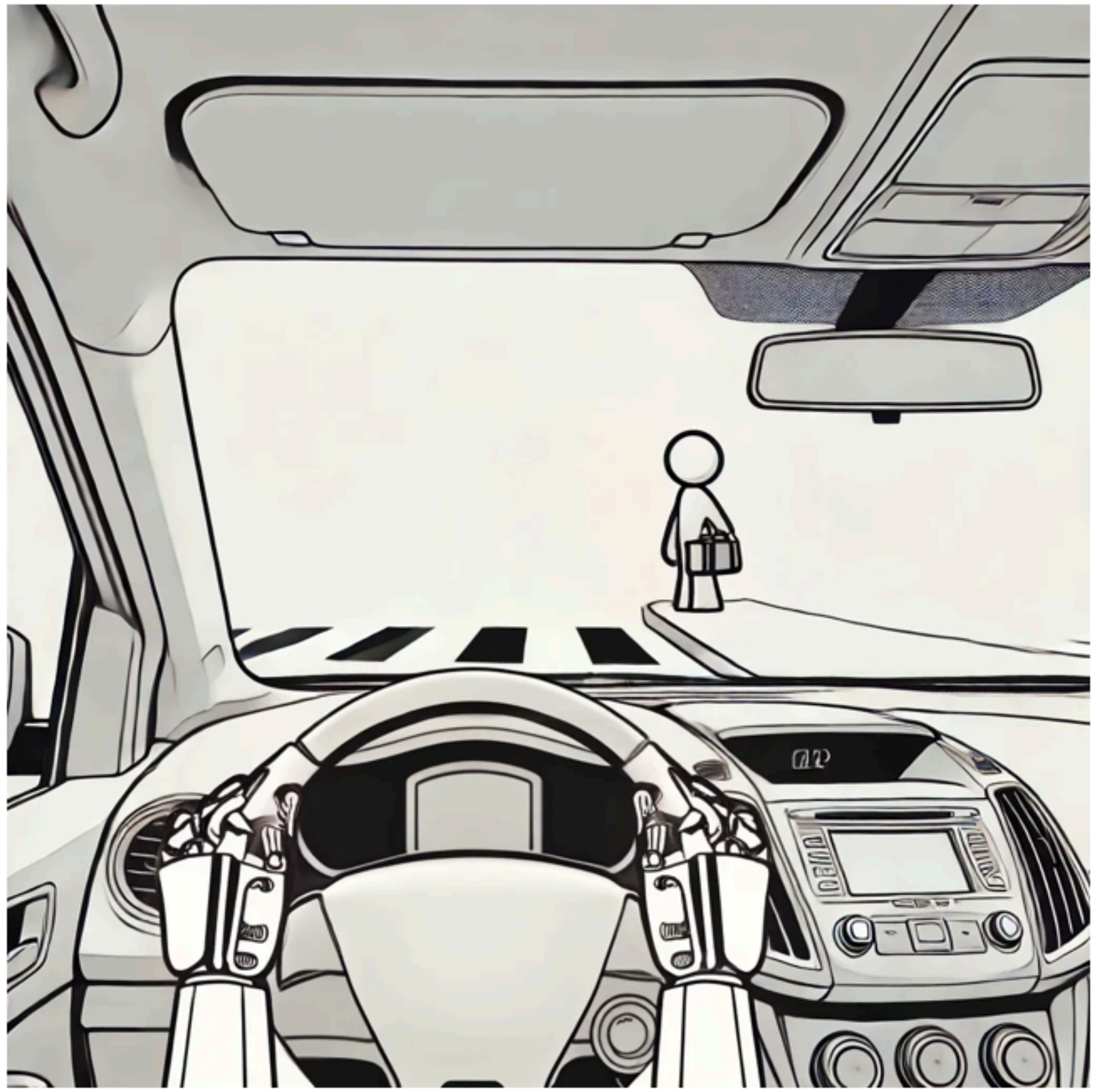
Risk Averse Decision Making



- The pedestrian seems to be standing still

- But there is a small chance that they walk

Risk Averse Decision Making

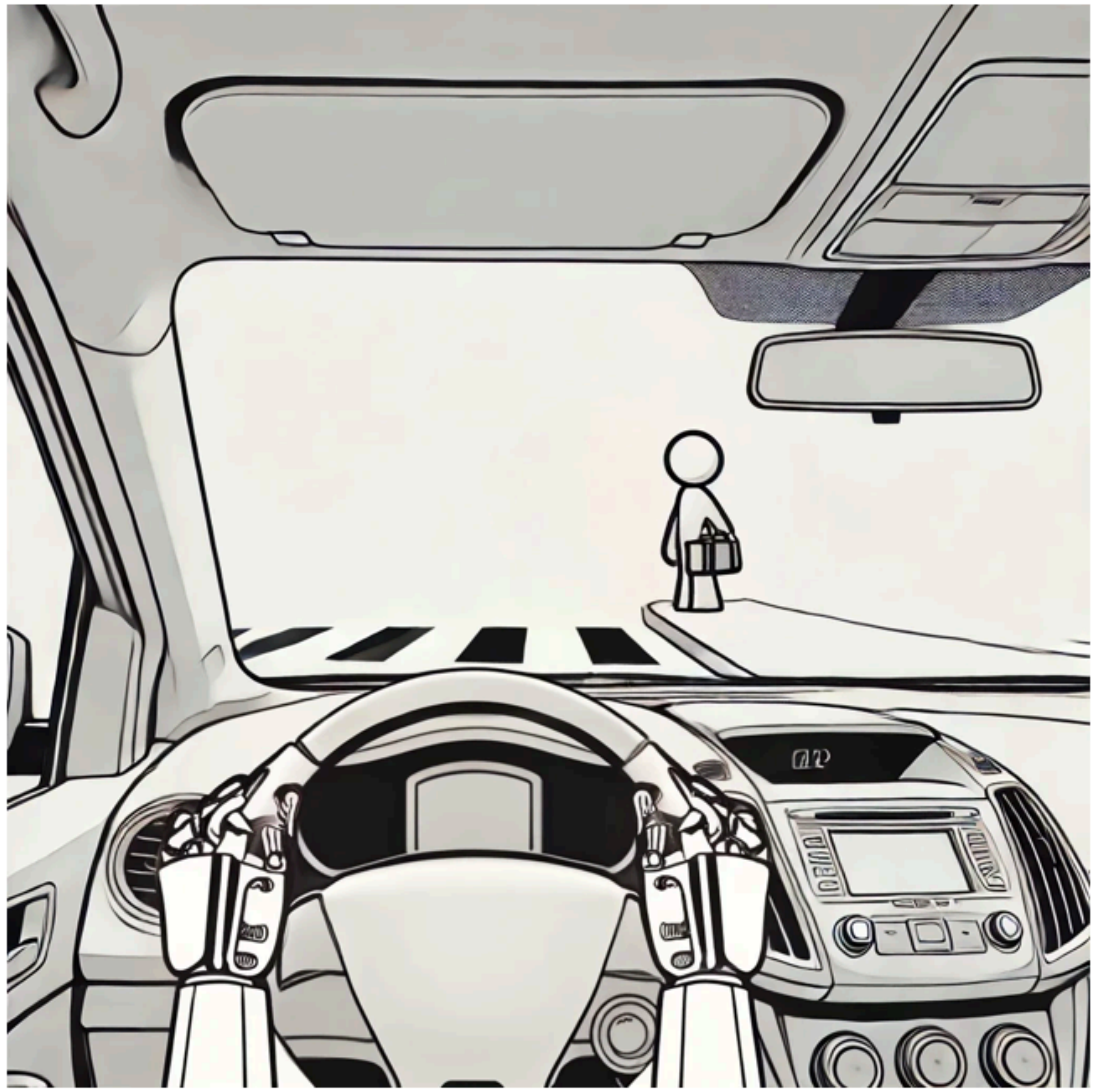


- The pedestrian seems to be standing still

- But there is a small chance that they walk

$$C(x) = \{\text{standing}, \text{walking}\}$$

Risk Averse Decision Making



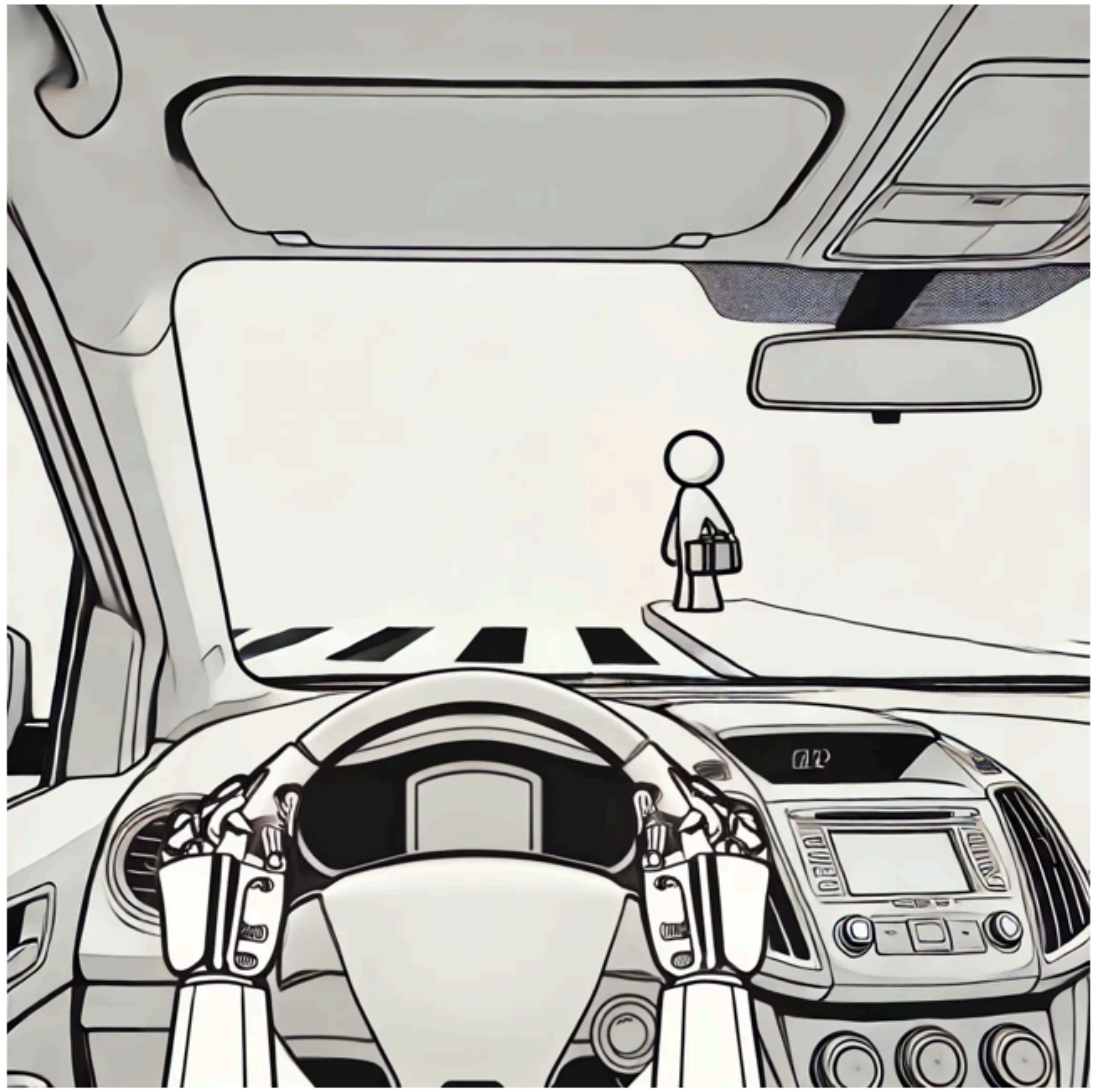
- The pedestrian seems to be standing still
- But there is a small chance that they walk

$$C(x) = \{\text{standing}, \text{walking}\}$$

- To be safe, we act according to the worst-case

Risk Averse Decision Making

x → ML model —Prediction→ Conformal prediction —$C(x)$→ Decision maker → action $= \pi(C(x))$



- The pedestrian seems to be standing still

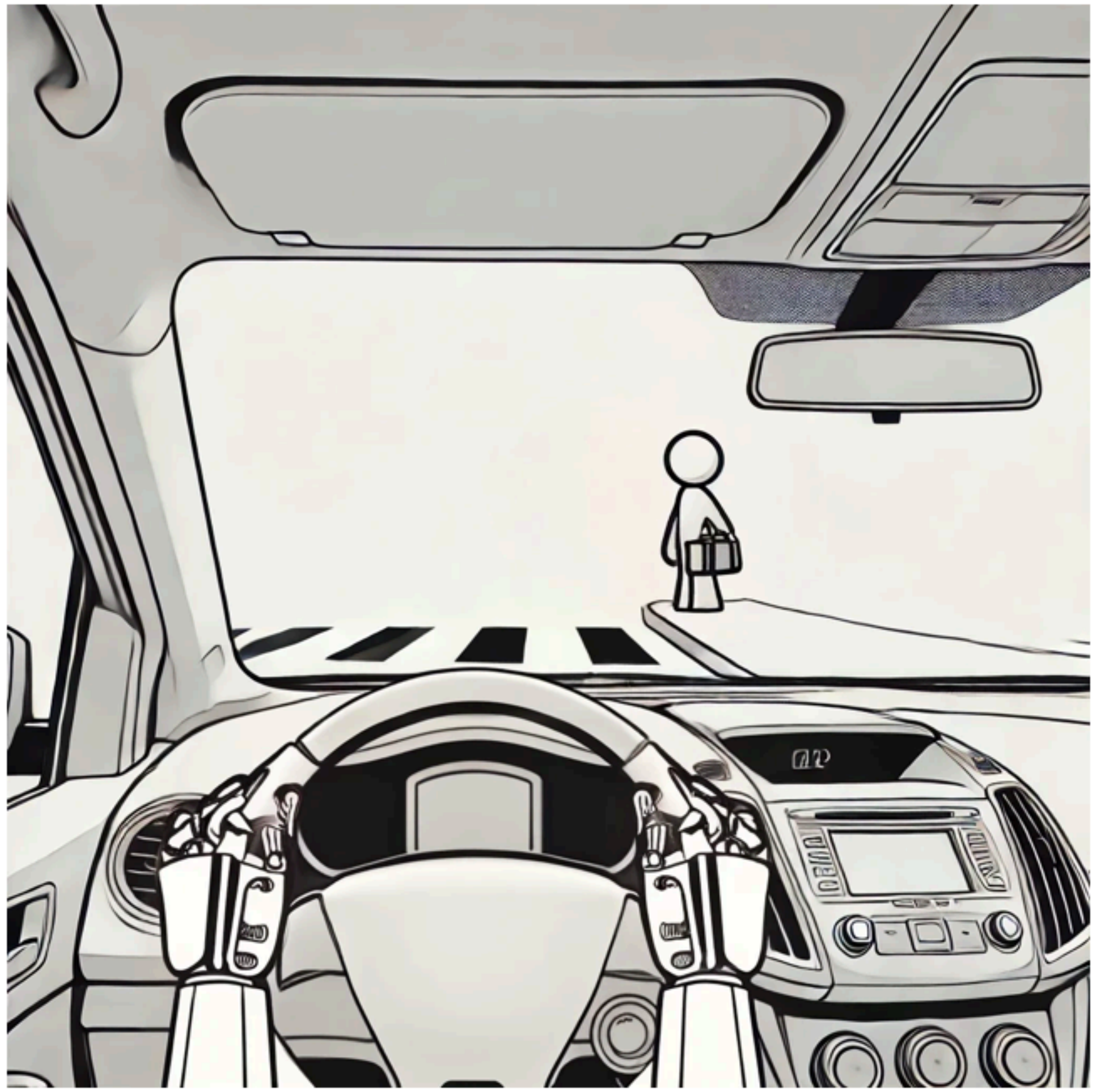
- But there is a small chance that they walk

$$C(x) = \{\text{standing}, \text{walking}\}$$

- To be safe, we act according to the worst-case

$$\pi^*(x) = \arg\max_{a \in \mathscr{A}} \min_{y \in C(x)} u(a, y)$$

Risk Averse Decision Making



- The pedestrian seems to be standing still

- But there is a small chance that they walk

$$C(x) = \{\text{standing}, \text{walking}\}$$

- To be safe, we act according to the worst-case

$$\pi^*(x) = \arg\max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$$

We can show that this is provably optimal!

Risk Averse Decision Making

Risk Averse Decision Making

$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$ $\qquad$ $\Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$

Risk Averse Decision Making

$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$

$$\Pr_{(X,Y)}\left[Y \in C(X)\right] \geq 1 - \alpha$$

Goal: $\pi(.) : 2^{\mathcal{Y}} \to \mathcal{A}$

Risk Averse Decision Making

$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$ $\qquad \Pr_{(X,Y)}\left[Y \in C(X)\right] \geq 1 - \alpha$ $\qquad$ Goal: $\pi(.) : 2^{\mathcal{Y}} \rightarrow \mathcal{A}$

Risk Averse Decision Making

$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$

$$\Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$$

Goal: $\pi(.) : 2^{\mathcal{Y}} \to \mathcal{A}$

Let Ω be the set of all distributions satisfying the above inequality.

Risk Averse Decision Making

$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$

$$\Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$$

Goal: $\pi(.) : 2^{\mathcal{Y}} \to \mathcal{A}$

Let Ω be the set of all distributions satisfying the above inequality.

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X), Y) \geq \nu(X)\Big] \geq 1 - \alpha,$$

Risk Averse Decision Making

$$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y} \qquad \Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha \qquad \text{Goal: } \pi(.) : 2^{\mathcal{Y}} \to \mathcal{A}$$

Let Ω be the set of all distributions satisfying the above inequality.

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X), Y) \geq \nu(X)\Big] \geq 1 - \alpha,$$

$$\nu^*(\pi, p) = \underset{\nu(.)}{\text{Maximize}} \quad \mathop{\mathbb{E}}_{X \sim p(x)}\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr_{X,Y \sim p(x,y)}[u(\pi(C(X)), Y) \geq \nu(X)] \geq 1 - \alpha$$

Risk Averse Decision Making

$$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y} \qquad \Pr_{(X,Y)}[Y \in C(X)] \geq 1-\alpha$$

Goal: $\pi(.) : 2^{\mathcal{Y}} \to \mathcal{A}$

Let Ω be the set of all distributions satisfying the above inequality.

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X), Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$$\nu^*(\pi, p) = \underset{\nu(.)}{\text{Maximize}} \quad \mathbb{E}_{X\sim p(x)}\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr_{X,Y\sim p(x,y)}[u(\pi(C(X)), Y) \geq \nu(X)] \geq 1-\alpha$$

$$\underset{\pi}{\text{Maximize}}\ \underset{p\in\Omega}{\text{Minimize}} \quad \nu^*(\pi, p)$$

Risk Averse Decision Making

$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$

$$\Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$$

Goal: $\pi(.) : 2^{\mathcal{Y}} \rightarrow \mathcal{A}$

Let Ω be the set of all distributions satisfying the above inequality.

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X), Y) \geq \nu(X)\Big] \geq 1 - \alpha,$$

$$\nu^*(\pi, p) = \underset{\nu(.)}{\text{Maximize}} \quad \underset{X \sim p(x)}{\mathbb{E}}\Big[\nu(X)\Big]$$

$$\text{subject to} \quad \Pr_{X,Y \sim p(x,y)}[u(\pi(C(X)), Y) \geq \nu(X)] \geq 1 - \alpha$$

over all data distributions compatible with C

$$\underset{\pi}{\text{Maximize}} \; \underset{p \in \Omega}{\text{Minimize}} \quad \nu^*(\pi, p)$$

Risk Averse Decision Making

$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$ $\qquad \Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$ $\qquad$ Goal: $\pi(.) : 2^{\mathcal{Y}} \to \mathcal{A}$

Let Ω be the set of all distributions satisfying the above inequality.

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\left[\nu(X)\right],$$

$$\text{subject to} \quad \Pr\left[u(a(X), Y) \geq \nu(X)\right] \geq 1 - \alpha,$$

$$\nu^*(\pi, p) = \underset{\nu(.)}{\text{Maximize}} \quad \mathbb{E}_{X \sim p(x)}\left[\nu(X)\right]$$

$$\text{subject to} \quad \Pr_{X,Y \sim p(x,y)}[u(\pi(C(X)), Y) \geq \nu(X)] \geq 1 - \alpha$$

over all policies that only depend on C $\qquad$ over all data distributions compatible with C

$$\underset{\pi}{\text{Maximize}} \; \underset{p \in \Omega}{\text{Minimize}} \quad \nu^*(\pi, p)$$

Risk Averse Decision Making

$\forall x \in \mathcal{X}, \quad C(x) \subseteq \mathcal{Y}$

$$\Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$$

Goal: $\pi(.) : 2^{\mathcal{Y}} \to \mathcal{A}$

Let Ω be the set of all distributions satisfying the above inequality.

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\left[\nu(X)\right],$$

$$\text{subject to} \quad \Pr\left[u(a(X),Y) \geq \nu(X)\right] \geq 1-\alpha,$$

$$\nu^*(\pi, p) = \underset{\nu(.)}{\text{Maximize}} \quad \mathbb{E}_{X\sim p(x)}\left[\nu(X)\right]$$

$$\text{subject to} \quad \Pr_{X,Y\sim p(x,y)}[u(\pi(C(X)),Y) \geq \nu(X)] \geq 1-\alpha$$

over all policies that only depend on C

over all data distributions compatible with C

$$\underset{\pi}{\text{Maximize}}\ \underset{p\in\Omega}{\text{Minimize}} \quad \nu^*(\pi, p)$$

Proposition: Let $\pi^*(x)$ be the optimal solution to the above objective. Then we have,

$$\pi^*(x) = \arg\max_{a\in\mathcal{A}} \min_{y\in C(x)} u(a, y)$$

Risk Averse Decision Making

A prediction set perspective

Risk Averse Decision Making

A prediction set perspective

Risk averse action policy: $a_{\mathrm{RA}}\big(C(x)\big) = \arg\max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$ Risk averse utility certificate: $\nu_{\mathrm{RA}}\big(C(x)\big) = \max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$

Risk Averse Decision Making

A prediction set perspective

Risk averse action policy: $a_{\mathrm{RA}}\big(C(x)\big) = \arg\max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$

Risk averse utility certificate: $\nu_{\mathrm{RA}}\big(C(x)\big) = \max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$

$$\Pr_{(X,Y)}[Y \in C(X)] \geq 1 - \alpha$$

Safety Guarantee: $\Pr_{(X,Y)}[u(a_{\mathrm{RA}}(X), Y) \geq \nu_{\mathrm{RA}}(X)] \geq 1 - \alpha$

Risk Averse Decision Making

A prediction set perspective

Risk averse action policy: $a_{\mathrm{RA}}\big(C(x)\big) = \arg\max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$

Risk averse utility certificate: $\nu_{\mathrm{RA}}\big(C(x)\big) = \max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$

$$\Pr_{(X,Y)} [Y \in C(X)] \geq 1 - \alpha$$

Safety Guarantee: $\Pr_{(X,Y)} [u(a_{\mathrm{RA}}(X), Y) \geq \nu_{\mathrm{RA}}(X)] \geq 1 - \alpha$

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X), Y) \geq \nu(X)\Big] \geq 1 - \alpha,$$

Risk Averse Decision Making

A prediction set perspective

Risk averse action policy: $a_{\mathrm{RA}}\big(C(x)\big) = \arg\max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$

Risk averse utility certificate: $\nu_{\mathrm{RA}}\big(C(x)\big) = \max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$

$$\Pr_{(X,Y)} [Y \in C(X)] \geq 1 - \alpha$$

Safety Guarantee: $\Pr_{(X,Y)} [u(a_{\mathrm{RA}}(X), Y) \geq \nu_{\mathrm{RA}}(X)] \geq 1 - \alpha$

RA-DPO:

$$\begin{aligned} &\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} && \mathbb{E}_X\Big[\nu(X)\Big], \\ &\text{subject to} && \Pr\Big[u(a(X), Y) \geq \nu(X)\Big] \geq 1 - \alpha, \end{aligned}$$

Risk Averse Conformal Prediction Optimization (RA-CPO):

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\nu_{\mathrm{RA}}(C(X))\right] := \mathbb{E}_X\left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha.$$

Risk Averse Decision Making

A prediction set perspective

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X), Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

Risk Averse Decision Making

A prediction set perspective

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}}\ \min_{y\in C(X)}\ u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

Theorem: RA-DPO and RA-CPO are equivalent.

What are prediction sets good for?

What are prediction sets good for?

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$$\equiv$$

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

What are prediction sets good for?

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$$\equiv$$

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

1) What kind of downstream decision making process make prediction sets the correct notion of UQ?

2) What is the optimal policy that a risk averse decision maker should use to map prediction sets to actions?

3) How can we derive prediction sets that are optimal for such decision makers?

What are prediction sets good for?

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$
$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$\equiv$

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$
$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

1) What kind of downstream decision making process make prediction sets the correct notion of UQ?

✓ "Well designed" prediction sets are a sufficient statistic for risk averse decision makers who wish to optimize their value at risk.

2) What is the optimal policy that a risk averse decision maker should use to map prediction sets to actions?

✓ A simple max min policy is an optimal map from prediction sets to actions.

3) How can we derive prediction sets that are optimal for such decision makers?

What are prediction sets good for?

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$\equiv$

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

1) What kind of downstream decision making process make prediction sets the correct notion of UQ?

✓ "Well designed" prediction sets are a sufficient statistic for risk averse decision makers who wish to optimize their value at risk.

2) What is the optimal policy that a risk averse decision maker should use to map prediction sets to actions?

✓ A simple max min policy is an optimal map from prediction sets to actions.

3) How can we derive prediction sets that are optimal for such decision makers?

We will drive an explicit characterization of these sets over expectation and also provide a finite sample approximation of them.

?

How to Optimize Prediction Sets?

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y) \right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha$$

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}}\ \min_{y\in C(X)}\ u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

No convexity or concavity property!

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y) \right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha$$

No convexity or concavity property!

High-level idea: - A different parametrization

- Convex program, strong duality

- Calibrate the dual parameter using data

How to Optimize Prediction Sets?

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y) \right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha$$

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y) \right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha$$

A different parametrization:

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y) \right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha$$

A different parametrization:

Conditional coverage assignment: $t(x) = \Pr\left[Y \in C(X) \,|\, X = x\right]$

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

A different parametrization:

Conditional coverage assignment: $t(x) = \Pr\left[Y \in C(X) \,|\, X = x\right]$

Optimal risk averse utility for a fixed coverage assignment:

$$\boldsymbol{\theta}(x,t) = \max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{a}(x,t) = \arg\max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y) \right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha$$

A different parametrization:

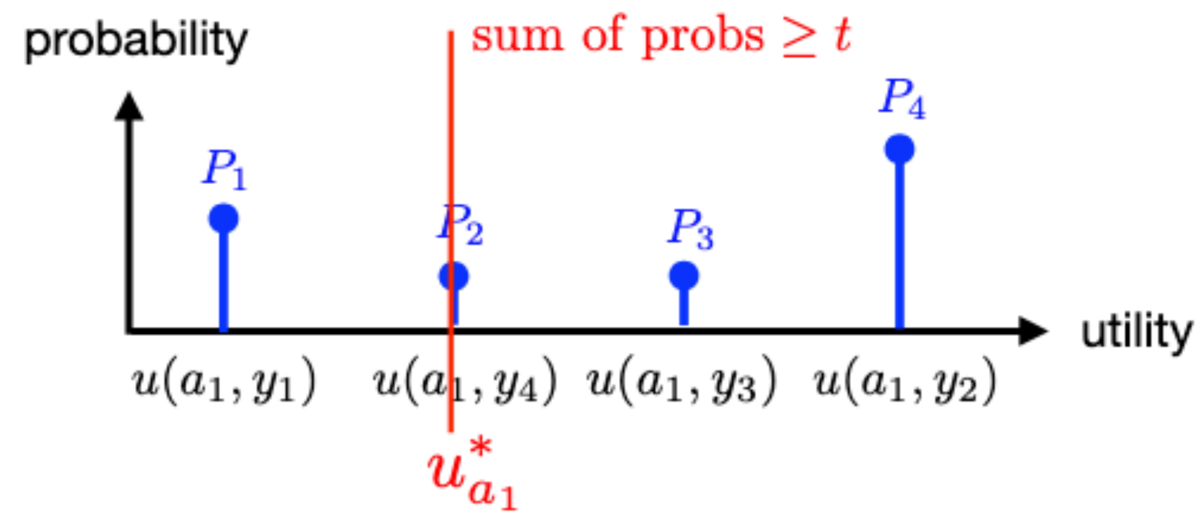
Conditional coverage assignment: $t(x) = \Pr\left[Y \in C(X) \,|\, X = x\right]$

Optimal risk averse utility for a fixed coverage assignment:

$$\boldsymbol{\theta}(x, t) = \max_{a \in \mathcal{A}} \text{quantile}_{1-t} \left[u(a, Y) \mid X = x\right]$$

$$\boldsymbol{a}(x, t) = \arg\max_{a \in \mathcal{A}} \text{quantile}_{1-t} \left[u(a, Y) \mid X = x\right]$$

action a_1:



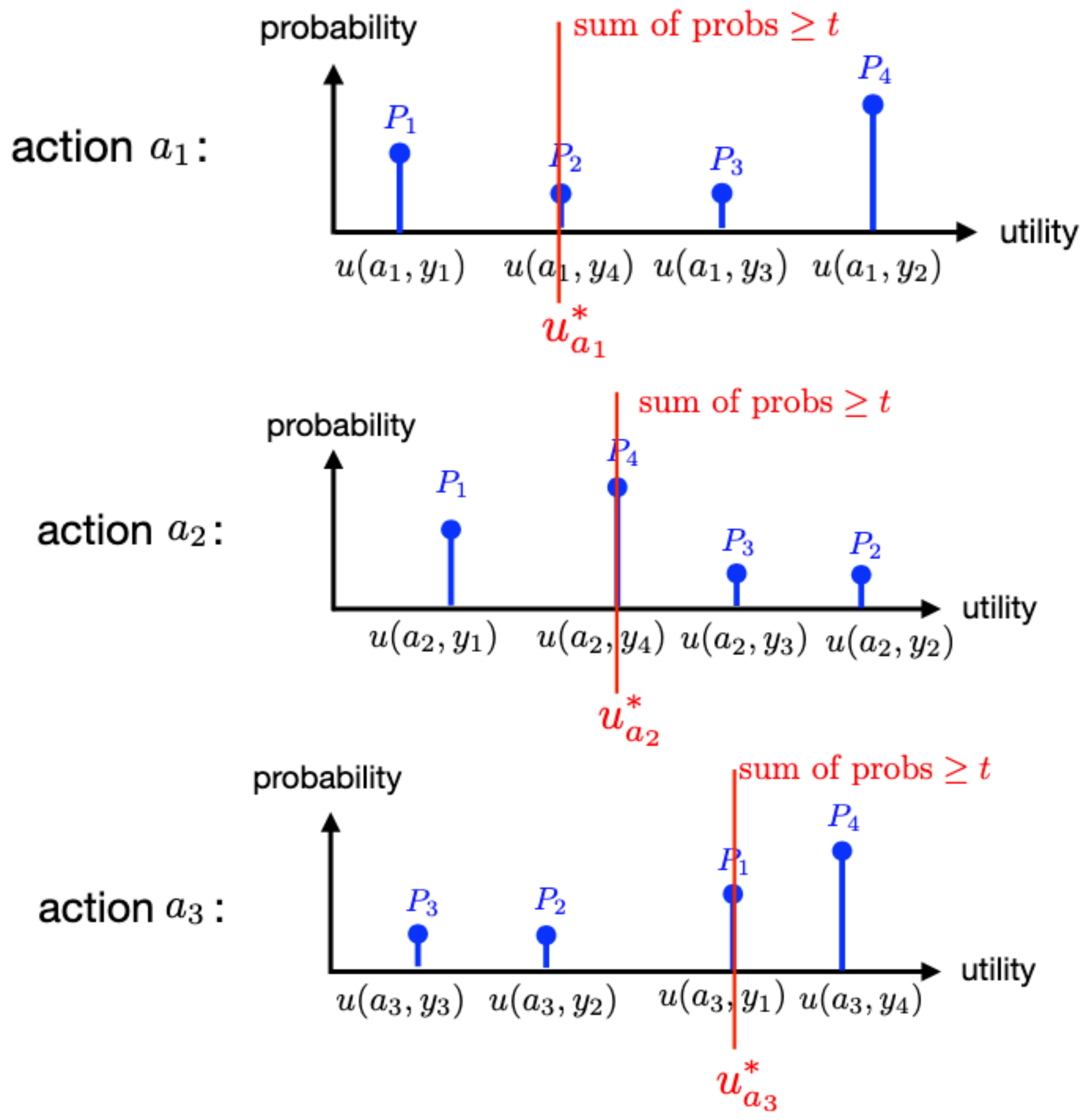
How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y) \right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha$$

A different parametrization:



Conditional coverage assignment: $t(x) = \Pr\left[Y \in C(X) \,|\, X = x\right]$

Optimal risk averse utility for a fixed coverage assignment:

$$\boldsymbol{\theta}(x, t) = \max_{a \in \mathcal{A}} \text{quantile}_{1-t} \left[u(a, Y) \mid X = x \right]$$

$$\boldsymbol{a}(x, t) = \arg\max_{a \in \mathcal{A}} \text{quantile}_{1-t} \left[u(a, Y) \mid X = x \right]$$

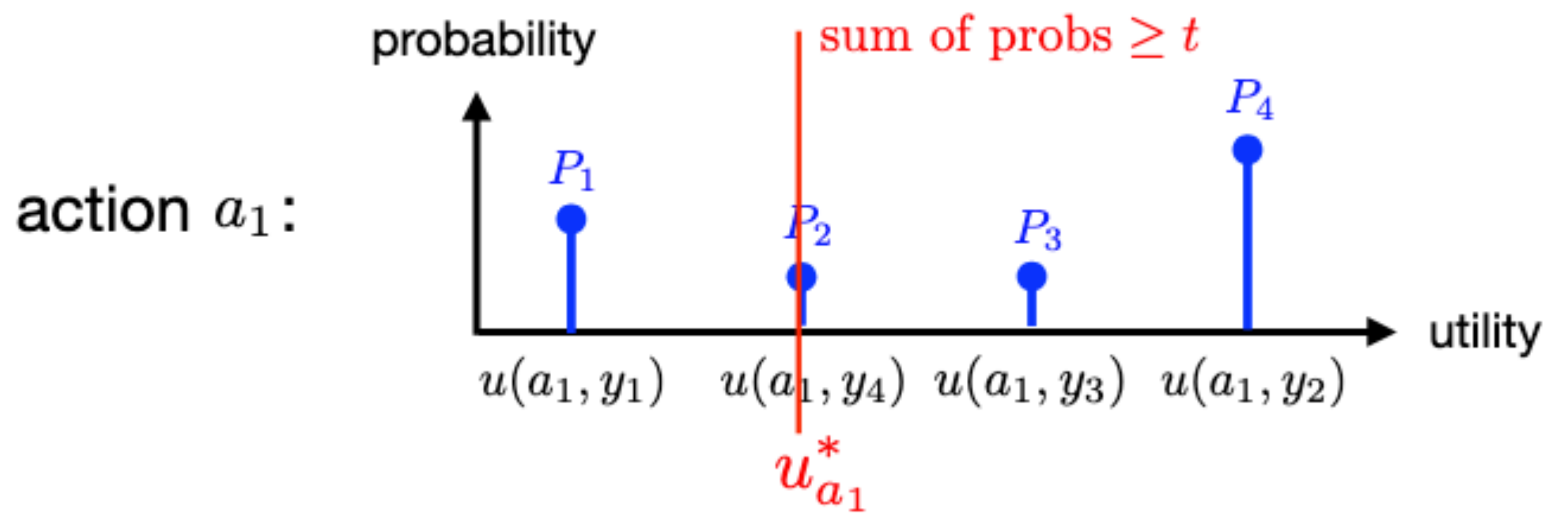
How to Optimize Prediction Sets?

RA-CPO:

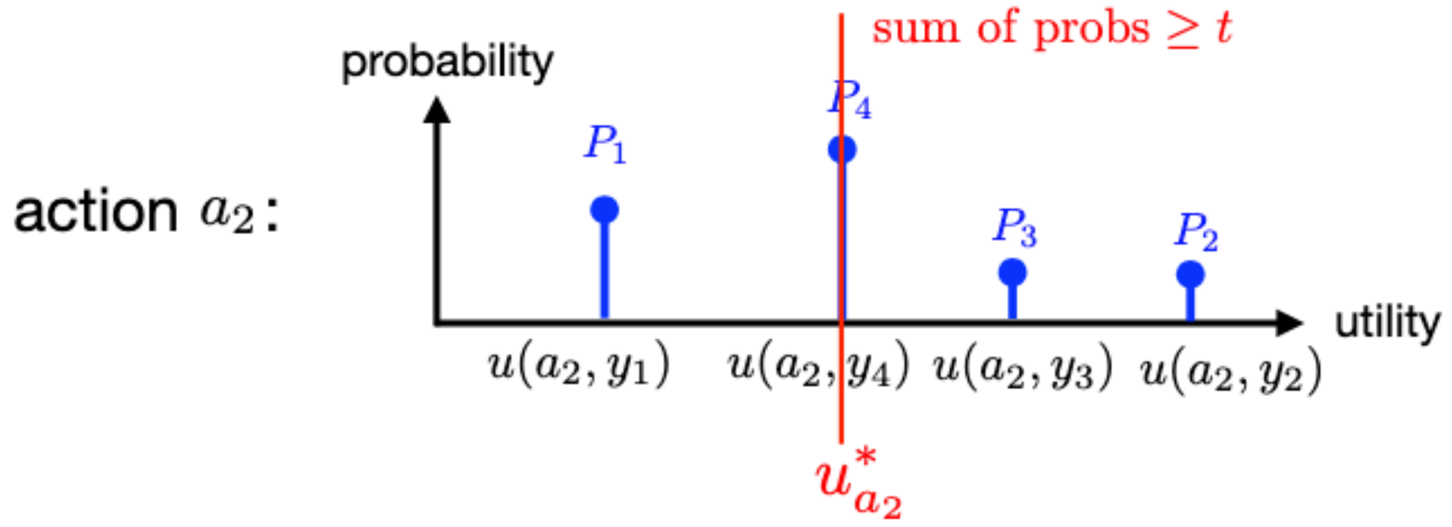
$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y) \right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha$$

A different parametrization:



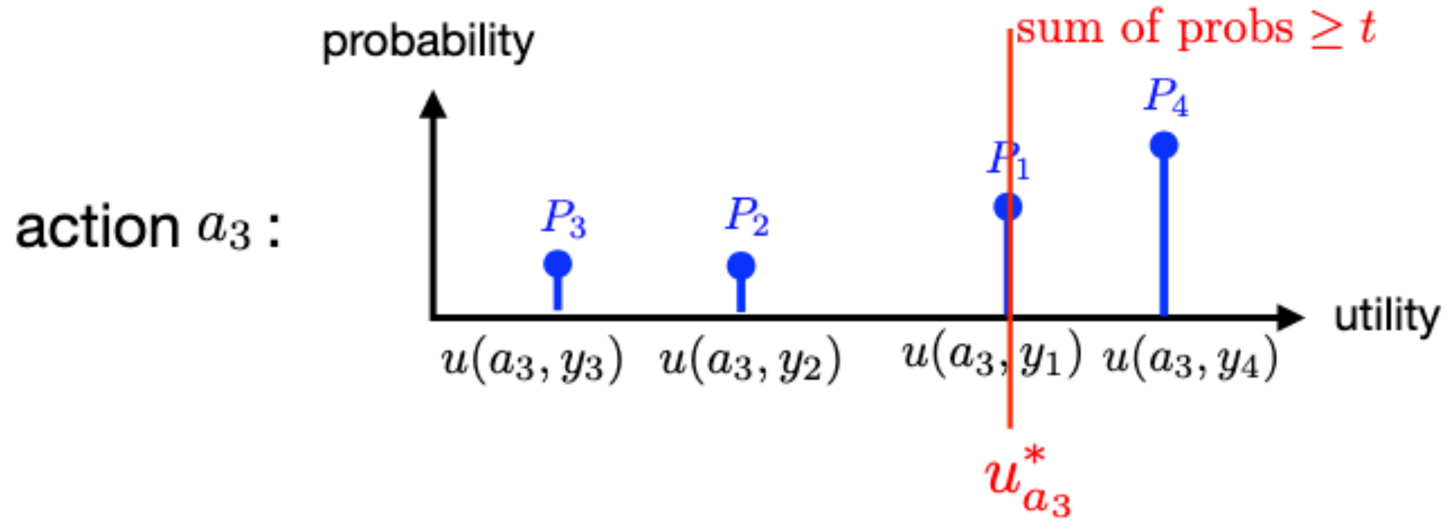
Conditional coverage assignment: $t(x) = \Pr\left[Y \in C(X) \,|\, X = x\right]$



Optimal risk averse utility for a fixed coverage assignment:

$$\boldsymbol{\theta}(x, t) = \max_{a \in \mathcal{A}} \text{quantile}_{1-t} \left[u(a, Y) \mid X = x \right] = \max\{u^*_{a_1}, u^*_{a_2}, u^*_{a_3}\} = u^*_{a_3}$$

$$\boldsymbol{a}(x, t) = \arg\max_{a \in \mathcal{A}} \text{quantile}_{1-t} \left[u(a, Y) \mid X = x \right] = a_3$$



How to Optimize Prediction Sets?

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \min_{y \in C(X)} u(a, y) \right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1 - \alpha$$

$$\equiv$$

$$\underset{t:\mathcal{X} \to [0,1]}{\text{maximize}} \quad \mathbb{E}_X \left[\boldsymbol{\theta}(X, t(X)) \right]$$

$$\text{subject to:} \quad \mathbb{E}_X \left[t(X) \right] \geq 1 - \alpha.$$

How to Optimize Prediction Sets?

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

$$\equiv$$

$$\underset{t:\mathcal{X}\to[0,1]}{\text{maximize}} \quad \mathbb{E}_X\left[\boldsymbol{\theta}(X, t(X))\right]$$

$$\text{subject to:} \quad \mathbb{E}_X\left[t(X)\right] \geq 1-\alpha.$$

Having t^* we have,

$$C^*(x) = \left\{y \in \mathcal{Y} : \ u(\boldsymbol{a}(x, t^*(x)), y) \geq \boldsymbol{\theta}(x, t^*(x))\right\}.$$

How to Optimize Prediction Sets?

How to Optimize Prediction Sets?

What the risk averse agent cares about

RA-DPO:

$$\underset{a(\cdot),\,\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$\equiv$

An equivalent conformal optimization

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}}\ \min_{y\in C(X)}\ u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

$\equiv$

An equivalent parametrization

$$\underset{t:\mathcal{X}\to[0,1]}{\text{maximize}} \quad \mathbb{E}_X\left[\boldsymbol{\theta}(X,t(X))\right]$$

$$\text{subject to:} \quad \mathbb{E}_X\left[t(X)\right] \geq 1-\alpha.$$

How to Optimize Prediction Sets?

What the risk averse agent cares about

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$
$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$\equiv$

An equivalent conformal optimization

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$
$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

$\equiv$

An equivalent parametrization

$$\underset{t:\mathcal{X}\to[0,1]}{\text{maximize}} \quad \mathbb{E}_X\left[\boldsymbol{\theta}(X, t(X))\right]$$
$$\text{subject to:} \quad \mathbb{E}_X\left[t(X)\right] \geq 1-\alpha.$$

What is t^*?

How to Optimize Prediction Sets?

What the risk averse agent cares about

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$\equiv$

An equivalent conformal optimization

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

$\equiv$

An equivalent parametrization

$$\underset{t:\mathcal{X}\to[0,1]}{\text{maximize}} \quad \mathbb{E}_X\left[\boldsymbol{\theta}(X,t(X))\right]$$

$$\text{subject to:} \quad \mathbb{E}_X\left[t(X)\right] \geq 1-\alpha.$$

What is t^*?

No convexity or concavity property!

How to Optimize Prediction Sets?

What the risk averse agent cares about

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$\equiv$

An equivalent conformal optimization

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

$\equiv$

An equivalent parametrization

$$\underset{t:\mathcal{X}\to[0,1]}{\text{maximize}} \quad \mathbb{E}_X\left[\boldsymbol{\theta}(X,t(X))\right]$$

$$\text{subject to:} \quad \mathbb{E}_X\left[t(X)\right] \geq 1-\alpha.$$

What is t^*?

No convexity or concavity property!

A one dimensional characterization:

$$\boldsymbol{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\}$$

How to Optimize Prediction Sets?

What the risk averse agent cares about

RA-DPO:

$$\underset{a(\cdot),\nu(\cdot)}{\text{maximize}} \quad \mathbb{E}_X\Big[\nu(X)\Big],$$

$$\text{subject to} \quad \Pr\Big[u(a(X),Y) \geq \nu(X)\Big] \geq 1-\alpha,$$

$\equiv$

An equivalent conformal optimization

RA-CPO:

$$\underset{C(.)}{\text{Maximize}} \quad \mathbb{E}_X\left[\max_{a\in\mathcal{A}} \min_{y\in C(X)} u(a,y)\right]$$

$$\text{subject to} \quad \Pr[Y \in C(X)] \geq 1-\alpha$$

$\equiv$

An equivalent parametrization

$$\underset{t:\mathcal{X}\to[0,1]}{\text{maximize}} \quad \mathbb{E}_X\left[\boldsymbol{\theta}(X, t(X))\right]$$

$$\text{subject to:} \quad \mathbb{E}_X\left[t(X)\right] \geq 1-\alpha.$$

What is t^*?

No convexity or concavity property!

A one dimensional characterization:

$$\mathbf{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\}$$

Theorem: Let $t^*(x)$ be the optimal solution. Then, there exists a $\beta^* \geq 0$ such that,

$$t^*(x) = \mathbf{g}(x,\beta^*)\,.$$

Further, β^* is a solution to the following equation $\mathbb{E}_X[\mathbf{g}(X,\beta^*)] = 1-\alpha$

Finite Sample Algorithm

Essential components

Finite Sample Algorithm

Essential components

$$\boldsymbol{\theta}(x,t) = \max_{a \in \mathcal{A}} \text{quantile}_{1-t} \left[u(a,Y) \mid X = x \right]$$

$$\boldsymbol{a}(x,t) = \arg\max_{a \in \mathcal{A}} \text{quantile}_{1-t} \left[u(a,Y) \mid X = x \right]$$

Finite Sample Algorithm

Essential components

$$\boldsymbol{\theta}(x,t) = \max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{a}(x,t) = \arg\max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\}$$

$$t^*(x) = \boldsymbol{g}(x,\beta^*)$$

$$\mathbb{E}_X[\boldsymbol{g}(X,\beta^*)] = 1 - \alpha$$

Finite Sample Algorithm

Essential components

$$\boldsymbol{\theta}(x,t) = \max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{a}(x,t) = \arg\max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\}$$

$$t^*(x) = \boldsymbol{g}(x,\beta^*)$$

$$\mathbb{E}_X[\boldsymbol{g}(X,\beta^*)] = 1 - \alpha$$

$$C^*(x) = \left\{y \in \mathcal{Y} :\ u(\boldsymbol{a}(x,t^*(x)),y) \geq \boldsymbol{\theta}(x,t^*(x))\right\}$$

Finite Sample Algorithm

Essential components

$$\boldsymbol{\theta}(x,t) = \max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right] \qquad \boldsymbol{a}(x,t) = \arg\max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\} \qquad t^*(x) = \boldsymbol{g}(x,\beta^*) \qquad \mathbb{E}_X[\boldsymbol{g}(X,\beta^*)] = 1-\alpha$$

$$C^*(x) = \left\{y \in \mathcal{Y}:\ u(\boldsymbol{a}(x,t^*(x)),y) \geq \boldsymbol{\theta}(x,t^*(x))\right\}$$

The only quantity that we have to approximate: $\text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$

Finite Sample Algorithm

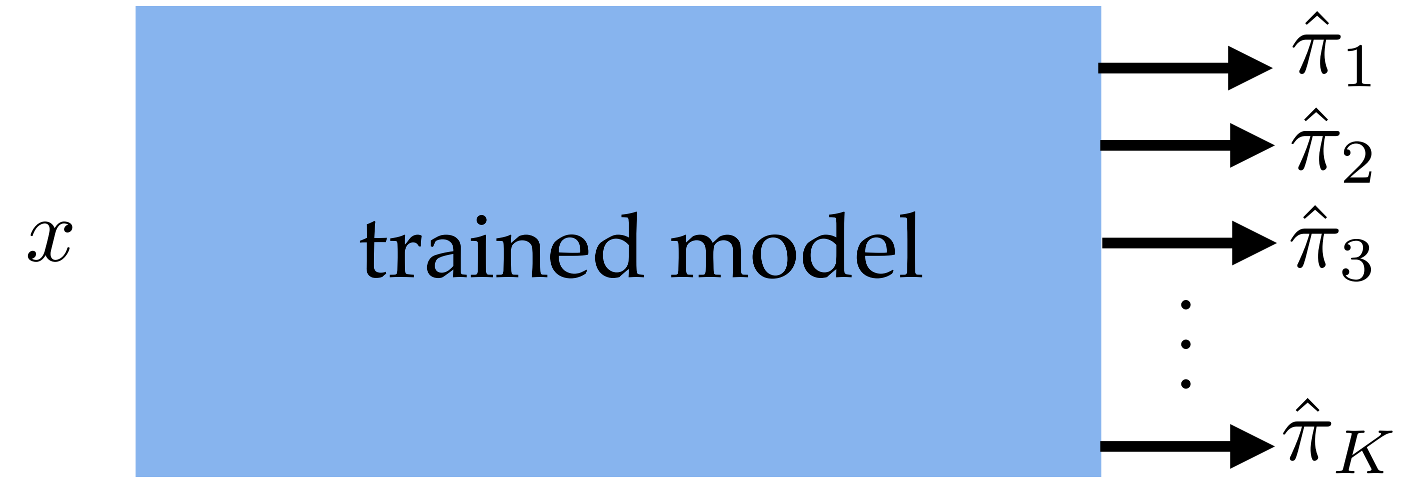
Essential components

$$\boldsymbol{\theta}(x,t) = \max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right] \qquad \boldsymbol{a}(x,t) = \arg\max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\} \qquad t^*(x) = \boldsymbol{g}(x,\beta^*) \qquad \mathbb{E}_X[\boldsymbol{g}(X,\beta^*)] = 1-\alpha$$

$$C^*(x) = \left\{y \in \mathcal{Y} :\ u(\boldsymbol{a}(x,t^*(x)),y) \geq \boldsymbol{\theta}(x,t^*(x))\right\}$$

The only quantity that we have to approximate: $\text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$



Finite Sample Algorithm

Essential components

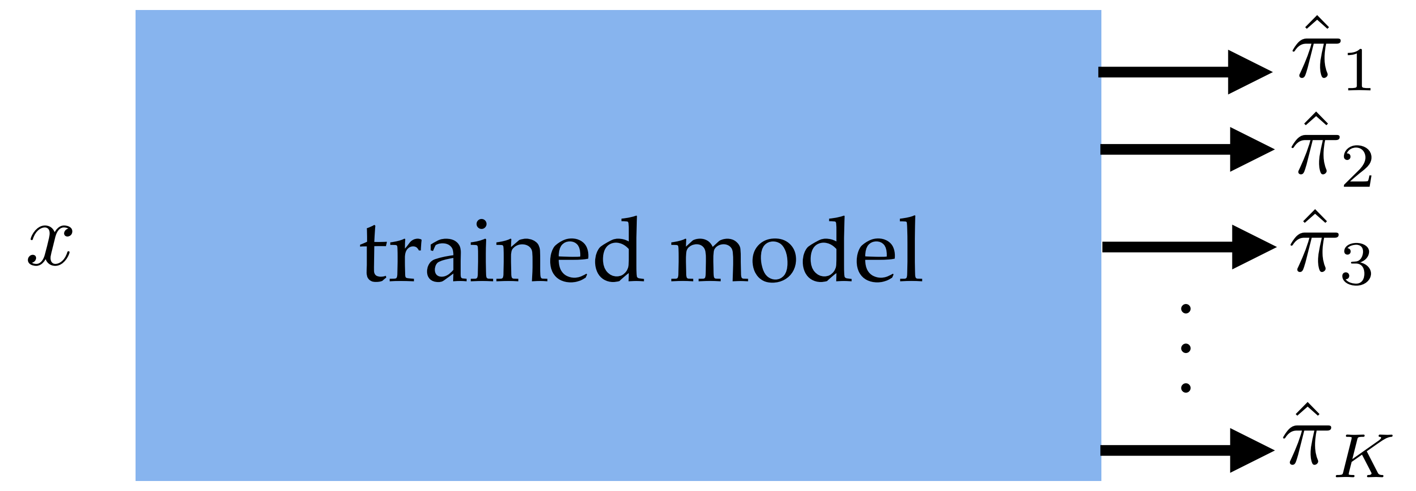
$$\boldsymbol{\theta}(x,t) = \max_{a \in \mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right] \qquad \boldsymbol{a}(x,t) = \arg\max_{a \in \mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{g}(x,\beta) = \arg\max_{s \in [0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\} \qquad t^*(x) = \boldsymbol{g}(x,\beta^*) \qquad \mathbb{E}_X[\boldsymbol{g}(X,\beta^*)] = 1 - \alpha$$

$$C^*(x) = \left\{ y \in \mathcal{Y} : \ u(\boldsymbol{a}(x, t^*(x)), y) \geq \boldsymbol{\theta}(x, t^*(x)) \right\}$$

The only quantity that we have to approximate: $\text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$

Approximate by: $\text{quantile}_{1-t}\left[u(a,Y) \mid Y \sim \hat{\pi}(x)\right]$



Finite Sample Algorithm

Essential components

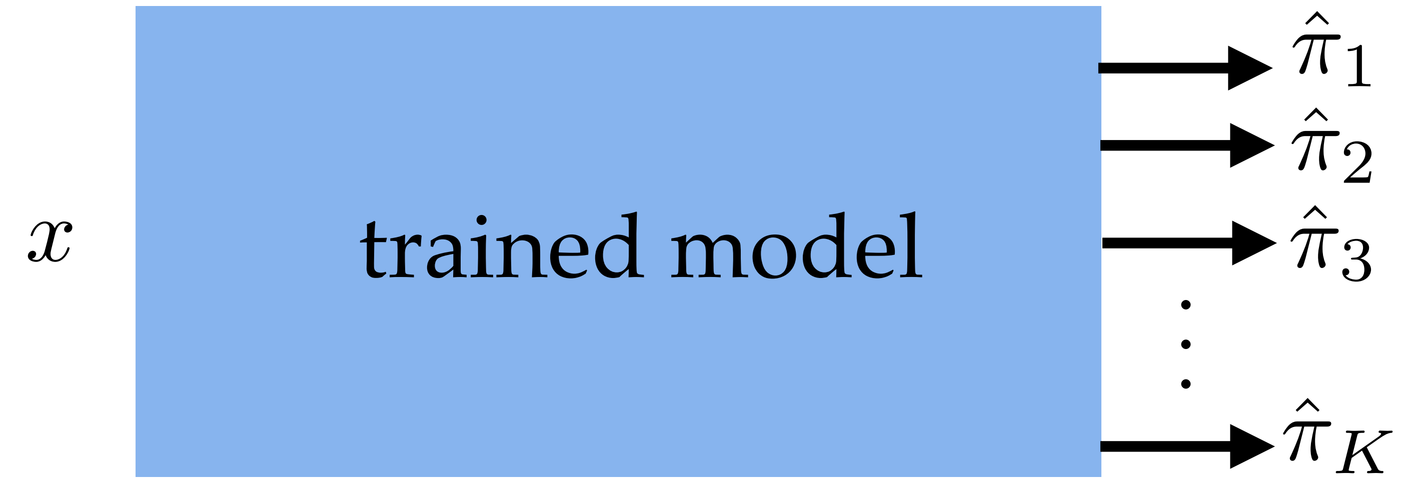
$$\boldsymbol{\theta}(x,t) = \max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right] \qquad \boldsymbol{a}(x,t) = \arg\max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\} \qquad t^*(x) = \boldsymbol{g}(x,\beta^*) \qquad \mathbb{E}_X[\boldsymbol{g}(X,\beta^*)] = 1-\alpha$$

$$C^*(x) = \left\{y \in \mathcal{Y} :\ u(\boldsymbol{a}(x,t^*(x)),y) \geq \boldsymbol{\theta}(x,t^*(x))\right\}$$

The only quantity that we have to approximate: $\text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$

Approximate by: $\text{quantile}_{1-t}\left[u(a,Y) \mid Y \sim \hat{\pi}(x)\right]$



$$\boldsymbol{\theta}(x,t) \quad \boldsymbol{g}(x,\beta) \quad \boldsymbol{a}(x,t) \longrightarrow \hat{\boldsymbol{\theta}}(x,t) \quad \hat{\boldsymbol{g}}(x,\beta) \quad \hat{\boldsymbol{a}}(x,t)$$

Finite Sample Algorithm

Essential components

$$\boldsymbol{\theta}(x,t) = \max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X=x\right] \qquad \boldsymbol{a}(x,t) = \arg\max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X=x\right]$$

$$\boldsymbol{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\} \qquad t^*(x) = \boldsymbol{g}(x,\beta^*) \qquad \mathbb{E}_X[\boldsymbol{g}(X,\beta^*)] = 1-\alpha$$

$$C^*(x) = \left\{y \in \mathcal{Y} :\ u(\boldsymbol{a}(x,t^*(x)),y) \geq \boldsymbol{\theta}(x,t^*(x))\right\}$$

$$\boldsymbol{\theta}(x,t) \quad \boldsymbol{g}(x,\beta) \quad \boldsymbol{a}(x,t) \longrightarrow \hat{\boldsymbol{\theta}}(x,t) \quad \hat{\boldsymbol{g}}(x,\beta) \quad \hat{\boldsymbol{a}}(x,t)$$

Finite Sample Algorithm

Essential components

$$\boldsymbol{\theta}(x,t) = \max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{a}(x,t) = \arg\max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\}$$

$$t^*(x) = \boldsymbol{g}(x,\beta^*)$$

$$\mathbb{E}_X[\boldsymbol{g}(X,\beta^*)] = 1 - \alpha$$

$$C^*(x) = \left\{y \in \mathcal{Y} :\ u(\boldsymbol{a}(x,t^*(x)),y) \geq \boldsymbol{\theta}(x,t^*(x))\right\}$$

$$\boldsymbol{\theta}(x,t) \quad \boldsymbol{g}(x,\beta) \quad \boldsymbol{a}(x,t) \longrightarrow \hat{\boldsymbol{\theta}}(x,t) \quad \hat{\boldsymbol{g}}(x,\beta) \quad \hat{\boldsymbol{a}}(x,t)$$

$$\hat{C}(x;\beta) = \left\{y \in \mathcal{Y} :\ u\left(\hat{\boldsymbol{a}}(x,\,\hat{\boldsymbol{g}}(x,\beta)),y\right) \geq \hat{\boldsymbol{\theta}}(x,\,\hat{\boldsymbol{g}}(x,\beta))\right\}$$

Finite Sample Algorithm

Essential components

$$\boldsymbol{\theta}(x,t) = \max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right] \qquad \boldsymbol{a}(x,t) = \arg\max_{a\in\mathcal{A}} \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$$

$$\boldsymbol{g}(x,\beta) = \arg\max_{s\in[0,1]} \left\{\boldsymbol{\theta}(x,s) + \beta s\right\} \qquad t^*(x) = \boldsymbol{g}(x,\beta^*) \qquad \mathbb{E}_X[\boldsymbol{g}(X,\beta^*)] = 1-\alpha$$

$$C^*(x) = \left\{y \in \mathcal{Y}:\ u(\boldsymbol{a}(x,t^*(x)),y) \geq \boldsymbol{\theta}(x,t^*(x))\right\}$$

$$\boldsymbol{\theta}(x,t) \quad \boldsymbol{g}(x,\beta) \quad \boldsymbol{a}(x,t) \longrightarrow \hat{\boldsymbol{\theta}}(x,t) \quad \hat{\boldsymbol{g}}(x,\beta) \quad \hat{\boldsymbol{a}}(x,t)$$

$$\hat{C}(x;\beta) = \left\{y \in \mathcal{Y}:\ u\left(\hat{\boldsymbol{a}}(x,\ \hat{\boldsymbol{g}}(x,\beta)),y\right) \geq \hat{\boldsymbol{\theta}}(x,\ \hat{\boldsymbol{g}}(x,\beta))\right\}$$

We calibrate β to get the correct coverage!

Risk Averse Calibration (RAC)

Algorithm:

Risk Averse Calibration (RAC)

Algorithm:

1: **Input:** Miscoverage level α, Calibration samples $\{(X_i, Y_i)\}_{i=1}^n$, Test covariate X_{test}.

2: **for each** $y \in \mathcal{Y}$**:**

$$\hat{\beta}_y = \operatorname*{argmin}_{\beta \in \mathbb{R}} \beta \quad \text{subject to:} \quad \frac{1}{n+1}\{\sum_{i=1}^{n} [Y_i \in \hat{C}(X_i; \beta)] + \mathbf{1}[y \in \hat{C}(X_{\text{test}}; \beta)]\} \geq 1 - \alpha.$$

3: **Output:**

$$C_{\text{RAC}}(X_{\text{test}}) = \{y \in \mathcal{Y} \mid y \in \hat{C}(X_{\text{test}}; \hat{\beta}_y)\}.$$

Risk Averse Calibration (RAC)

Algorithm:

1: **Input:** Miscoverage level α, Calibration samples $\{(X_i, Y_i)\}_{i=1}^n$, Test covariate X_{test}.

2: **for each** $y \in \mathcal{Y}$**:**

$$\hat{\beta}_y = \operatorname*{argmin}_{\beta \in \mathbb{R}} \beta \quad \text{subject to:} \quad \frac{1}{n+1}\Big\{\sum_{i=1}^{n} [Y_i \in \hat{C}(X_i; \beta)] + \mathbf{1}[y \in \hat{C}(X_{\text{test}}; \beta)]\Big\} \geq 1 - \alpha.$$

3: **Output:**

$$C_{\text{RAC}}(X_{\text{test}}) = \{y \in \mathcal{Y} \mid y \in \hat{C}(X_{\text{test}}; \hat{\beta}_y)\}.$$

Theorem: Assume that the calibration samples $\{(X_i, Y_i)\}_{i=1}^n$ and $\{(X_i, Y_i)\}_{i=1}^n$ are exchangeable. Then, we have

$$\Pr\left[Y_{\text{test}} \in C_{\text{RAC}}(X_{\text{test}})\right] \geq 1 - \alpha,$$

over the randomness of the test and calibration data.

Risk Averse Calibration (RAC)

Algorithm:

1: **Input:** Miscoverage level α, Calibration samples $\{(X_i, Y_i)\}_{i=1}^n$, Test covariate X_{test}.

2: **for each** $y \in \mathcal{Y}$**:**

$$\hat{\beta}_y = \operatorname*{argmin}_{\beta \in \mathbb{R}} \beta \quad \text{subject to:} \quad \frac{1}{n+1}\left\{\sum_{i=1}^{n} [Y_i \in \hat{C}(X_i; \beta)] + \mathbf{1}[y \in \hat{C}(X_{\text{test}}; \beta)]\right\} \geq 1 - \alpha.$$

3: **Output:**

$$C_{\text{RAC}}(X_{\text{test}}) = \left\{y \in \mathcal{Y} \mid y \in \hat{C}(X_{\text{test}}; \hat{\beta}_y)\right\}.$$

Theorem: Assume that the calibration samples $\{(X_i, Y_i)\}_{i=1}^n$ and $\{(X_i, Y_i)\}_{i=1}^n$ are exchangeable. Then, we have

$$\Pr\left[Y_{\text{test}} \in C_{\text{RAC}}(X_{\text{test}})\right] \geq 1 - \alpha,$$

over the randomness of the test and calibration data.

$$\Pr\left[u\left(a_{\text{RA}}\left(C_{\text{RAC}}(X_{\text{test}})\right), Y_{\text{test}}\right) \geq \nu_{\text{RA}}(C_{\text{RAC}}(X_{\text{test}}))\right] \geq 1 - \alpha$$

Risk Averse Calibration (RAC)

Algorithm:

1: **Input:** Miscoverage level α, Calibration samples $\{(X_i, Y_i)\}_{i=1}^n$, Test covariate X_{test}.

2: **for each** $y \in \mathcal{Y}$**:**

$$\hat{\beta}_y = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}}\, \beta \quad \text{subject to:} \quad \frac{1}{n+1}\Big\{\sum_{i=1}^{n} [Y_i \in \hat{C}(X_i;\beta)] + \mathbf{1}[y \in \hat{C}(X_{\text{test}};\beta)]\Big\} \geq 1-\alpha.$$

3: **Output:**

$$C_{\text{RAC}}(X_{\text{test}}) = \{y \in \mathcal{Y} \mid y \in \hat{C}(X_{\text{test}}; \hat{\beta}_y)\}.$$

Theorem: Assume that the calibration samples $\{(X_i, Y_i)\}_{i=1}^n$ and $\{(X_i, Y_i)\}_{i=1}^n$ are exchangeable. Then, we have

$$\Pr\left[Y_{\text{test}} \in C_{\text{RAC}}(X_{\text{test}})\right] \geq 1-\alpha,$$

over the randomness of the test and calibration data.

$$\Pr\left[u\left(a_{\text{RA}}\left(C_{\text{RAC}}(X_{\text{test}})\right), Y_{\text{test}}\right) \geq \nu_{\text{RA}}(C_{\text{RAC}}(X_{\text{test}}))\right] \geq 1-\alpha$$

Importantly,

Predictive model performance $\uparrow$

Risk Averse Calibration (RAC)

Algorithm:

1: **Input:** Miscoverage level α, Calibration samples $\{(X_i, Y_i)\}_{i=1}^n$, Test covariate X_{test}.

2: **for each** $y \in \mathcal{Y}$**:**

$$\hat{\beta}_y = \operatorname*{argmin}_{\beta \in \mathbb{R}} \beta \quad \text{subject to:} \quad \frac{1}{n+1}\left\{\sum_{i=1}^{n} [Y_i \in \hat{C}(X_i;\beta)] + \mathbf{1}[y \in \hat{C}(X_{\text{test}};\beta)]\right\} \geq 1-\alpha.$$

3: **Output:**

$$C_{\text{RAC}}(X_{\text{test}}) = \{y \in \mathcal{Y} \mid y \in \hat{C}(X_{\text{test}}; \hat{\beta}_y)\}.$$

Theorem: Assume that the calibration samples $\{(X_i, Y_i)\}_{i=1}^n$ and $\{(X_i, Y_i)\}_{i=1}^n$ are exchangeable. Then, we have

$$\Pr\left[Y_{\text{test}} \in C_{\text{RAC}}(X_{\text{test}})\right] \geq 1-\alpha,$$

over the randomness of the test and calibration data.

$$\Pr\left[u\left(a_{\text{RA}}\left(C_{\text{RAC}}(X_{\text{test}})\right), Y_{\text{test}}\right) \geq \nu_{\text{RA}}(C_{\text{RAC}}(X_{\text{test}}))\right] \geq 1-\alpha$$

Importantly,

Predictive model performance $\uparrow$ $\quad \text{quantile}_{1-t}\left[u(a,Y) \mid Y \sim f_x\right] \longrightarrow \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$

Risk Averse Calibration (RAC)

Algorithm:

1: **Input:** Miscoverage level α, Calibration samples $\{(X_i, Y_i)\}_{i=1}^n$, Test covariate X_{test}.

2: **for each** $y \in \mathcal{Y}$**:**

$$\hat{\beta}_y = \underset{\beta \in \mathbb{R}}{\text{argmin}}\, \beta \quad \text{subject to:} \quad \frac{1}{n+1}\{\sum_{i=1}^{n}[Y_i \in \hat{C}(X_i;\beta)] + \mathbf{1}[y \in \hat{C}(X_{\text{test}};\beta)]\} \geq 1-\alpha.$$

3: **Output:**

$$C_{\text{RAC}}(X_{\text{test}}) = \{y \in \mathcal{Y} \mid y \in \hat{C}(X_{\text{test}};\hat{\beta}_y)\}.$$

Theorem: Assume that the calibration samples $\{(X_i, Y_i)\}_{i=1}^n$ and $\{(X_i, Y_i)\}_{i=1}^n$ are exchangeable. Then, we have

$$\Pr\left[Y_{\text{test}} \in C_{\text{RAC}}(X_{\text{test}})\right] \geq 1-\alpha,$$

over the randomness of the test and calibration data.

$$\Pr\left[u(a_{\text{RA}}\left(C_{\text{RAC}}(X_{\text{test}})\right), Y_{\text{test}}) \geq \nu_{\text{RA}}(C_{\text{RAC}}(X_{\text{test}}))\right] \geq 1-\alpha$$

Importantly,

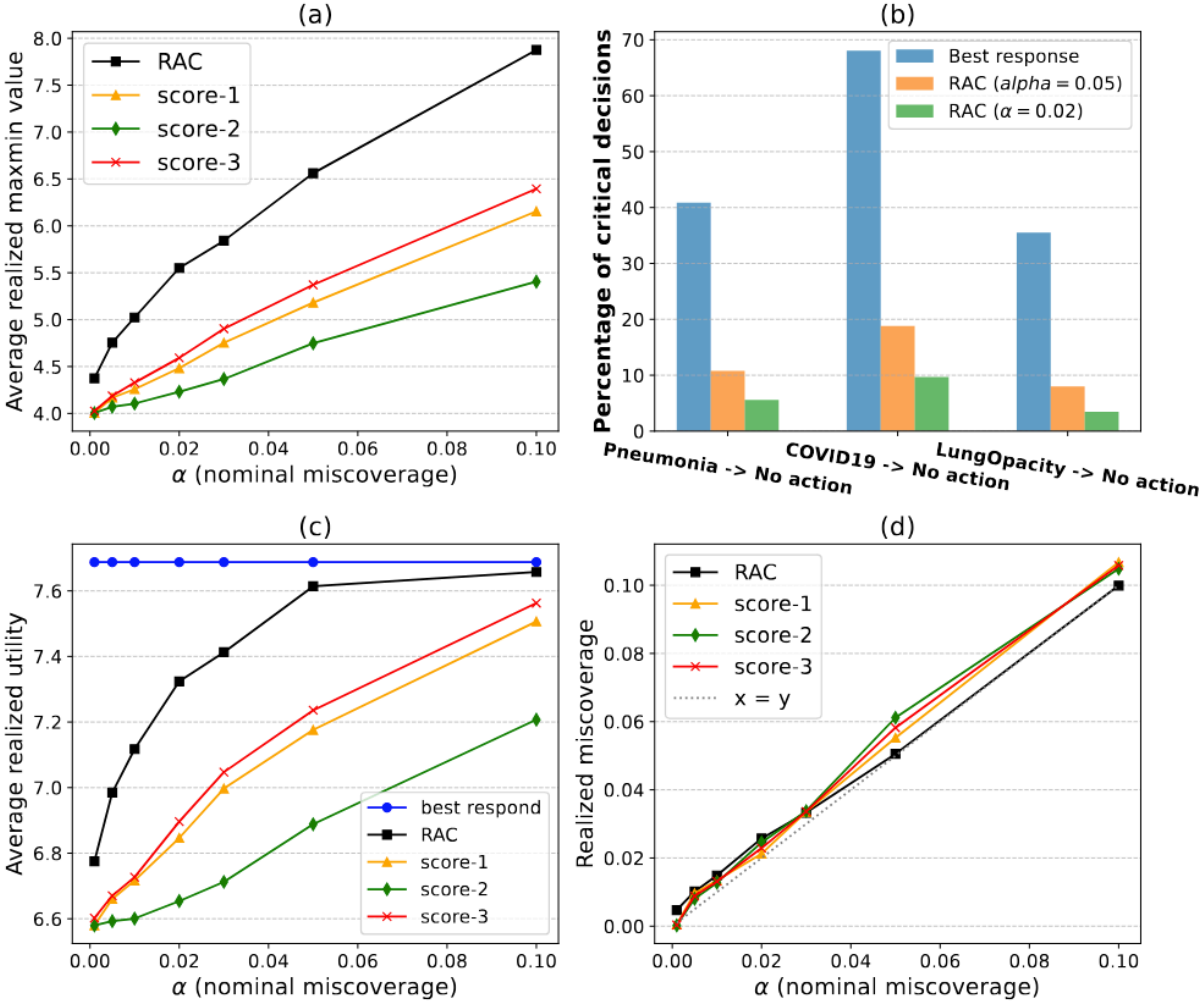
Predictive model performance $\uparrow$ $\quad \text{quantile}_{1-t}\left[u(a,Y) \mid Y \sim f_x\right] \longrightarrow \text{quantile}_{1-t}\left[u(a,Y) \mid X = x\right]$ $\quad$ RAC utility performance $\uparrow$

Clinical Medicine

Labels: Normal, COVID, Lung opacity, Pneumonia.

Actions: ‘No action’, Antibiotics, Quarantine, Additional Testing

True Label	No Action	Antibiotics	Quarantine	Additional Testing
Normal (0)	10	2	2	4
Pneumonia (1)	0	10	3	7
COVID-19 (2)	0	3	10	8
Lung Opacity (3)	1	4	4	10



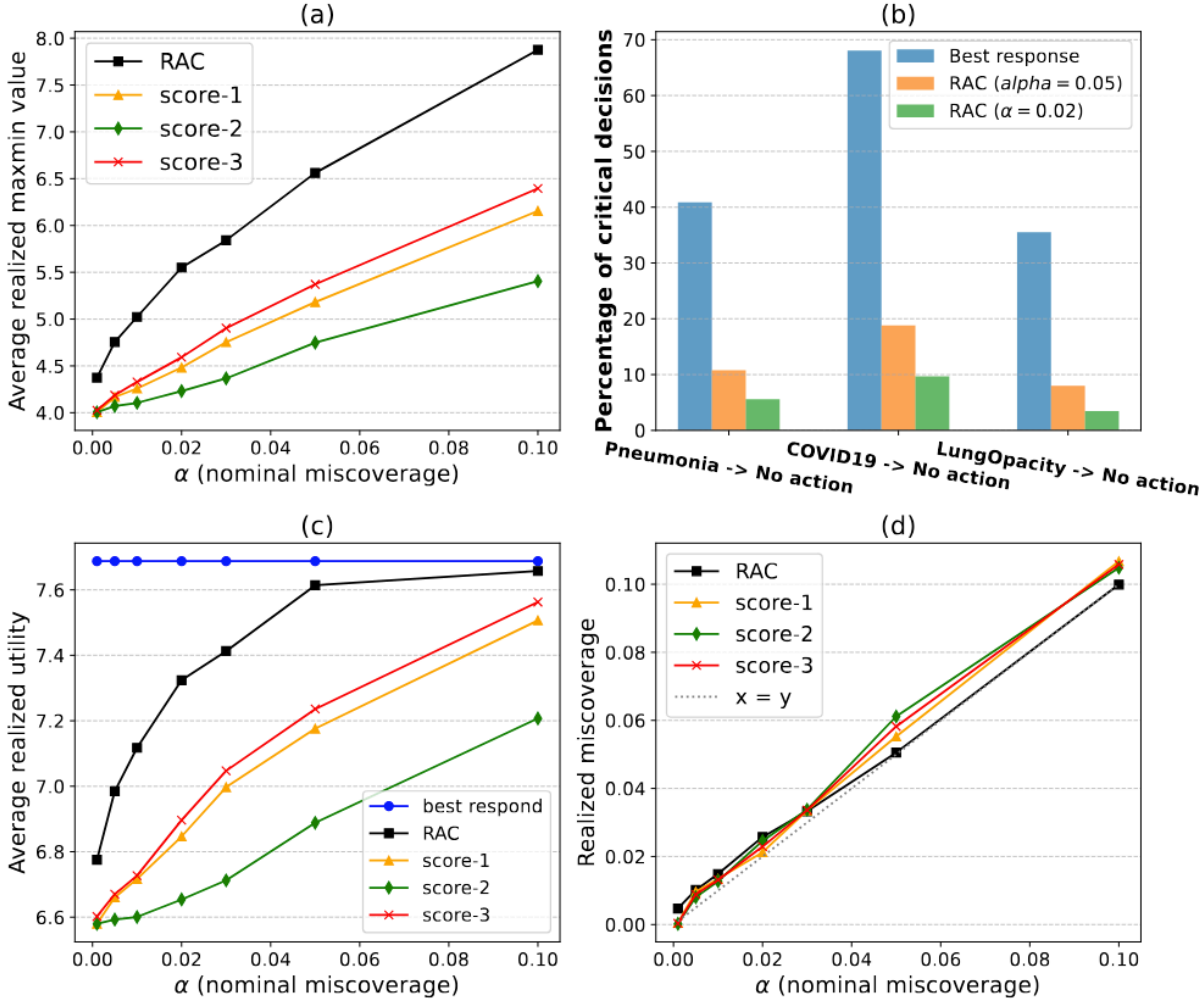
Clinical Medicine

Labels: Normal, COVID,
Lung opacity,
Pneumonia.

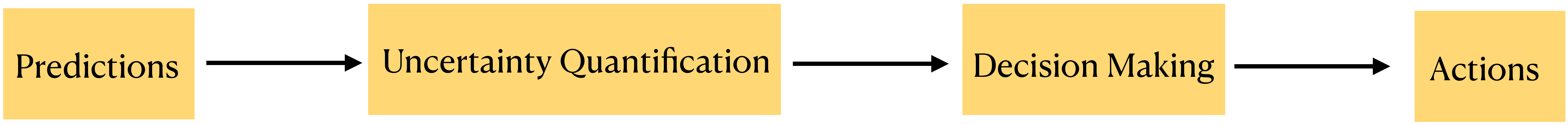
Actions: ‘No action’,
Antibiotics, Quarantine,
Additional Testing

True Label	No Action	Antibiotics	Quarantine	Additional Testing
Normal (0)	10	2	2	4
Pneumonia (1)	0	10	3	7
COVID-19 (2)	0	3	10	8
Lung Opacity (3)	1	4	4	10

How to design utility? Preference functions?
How to handle multiple sequential decisions?



Risk-averse decsion making (safety vs utility):



Risk-averse decsion making (safety vs utility):

Risk-averse decsion making (safety vs utility):

Risk-averse decsion making (safety vs utility):

Risk-averse decsion making (safety vs utility):

Risk-averse decsion making (safety vs utility):

Predictions $\longrightarrow$ Prediction sets $\xrightarrow{C(x)}$ Max-min $\longrightarrow$ Actions

$$t^*(x) = \mathbf{g}(x, \beta^*)$$

$$\pi^*(x) = \arg\max_{a \in \mathscr{A}} \min_{y \in C(x)} u(a, y)$$

$$C^*(x) = \left\{ y \in \mathcal{Y} : \; u(\boldsymbol{a}(x, t^*(x)), y) \geq \boldsymbol{\theta}(x, t^*(x)) \right\}$$

Risk-averse decsion making (safety vs utility):

Predictions → Prediction sets $\xrightarrow{C(x)}$ Max-min → Actions

$$t^*(x) = \mathbf{g}(x, \beta^*)$$

$$\pi^*(x) = \arg\max_{a \in \mathscr{A}} \min_{y \in C(x)} u(a, y)$$

$$C^*(x) = \left\{ y \in \mathcal{Y} : \; u(\boldsymbol{a}(x, t^*(x)), y) \geq \boldsymbol{\theta}(x, t^*(x)) \right\}$$

Risk-neutral decsion making (utility maximization):

Predictions → Uncertainty Quantification → Decision Making → Actions

Risk-averse decsion making (safety vs utility):

Predictions $\longrightarrow$ Prediction sets $\xrightarrow{C(x)}$ Max-min $\longrightarrow$ Actions

$$t^*(x) = \mathbf{g}(x, \beta^*)$$

$$\pi^*(x) = \arg\max_{a\in\mathcal{A}} \min_{y\in C(x)} u(a, y)$$

$$C^*(x) = \left\{ y \in \mathcal{Y} :\ u(\boldsymbol{a}(x, t^*(x)), y) \geq \boldsymbol{\theta}(x, t^*(x)) \right\}$$

Risk-neutral decsion making (utility maximization):

Predictions $\longrightarrow$ [ ] $\longrightarrow$ [ ] $\longrightarrow$ Actions

Risk-averse decsion making (safety vs utility):

$$\text{Predictions} \longrightarrow \text{Prediction sets} \xrightarrow{C(x)} \text{Max-min} \longrightarrow \text{Actions}$$

$$t^*(x) = \boldsymbol{g}(x, \beta^*)$$

$$\pi^*(x) = \arg\max_{a \in \mathscr{A}} \min_{y \in C(x)} u(a, y)$$

$$C^*(x) = \left\{ y \in \mathcal{Y} : \; u(\boldsymbol{a}(x, t^*(x)), y) \geq \boldsymbol{\theta}(x, t^*(x)) \right\}$$

Risk-neutral decsion making (utility maximization):

$$\text{Predictions} \longrightarrow \text{Calibration} \longrightarrow \text{Best-response} \longrightarrow \text{Actions}$$

[Related work: Foster & Vohra ‘98 , Zhao et al ’21, Noarov et al ’23]

Risk-averse decsion making (safety vs utility):

Predictions → Prediction sets $\xrightarrow{C(x)}$ Max-min → Actions

$$t^*(x) = \mathbf{g}(x, \beta^*)$$

$$\pi^*(x) = \arg\max_{a \in \mathcal{A}} \min_{y \in C(x)} u(a, y)$$

$$C^*(x) = \left\{ y \in \mathcal{Y} : \ u(\boldsymbol{a}(x, t^*(x)), y) \geq \boldsymbol{\theta}(x, t^*(x)) \right\}$$

Thank You!

Risk-neutral decsion making (utility maximization):

Predictions → Calibration → Best-response → Actions

[Related work: Foster & Vohra ‘98 , Zhao et al ’21, Noarov et al ’23]