IPML

IMPRECISE PROBABILISTIC MACHINE LEARNING

Lecture 10: Reflections, Open Problems, and Outlook

Krikamol Muandet

6 February 2026

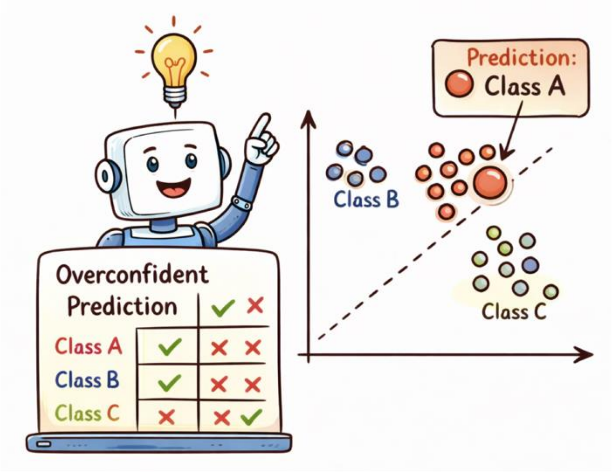
Outline

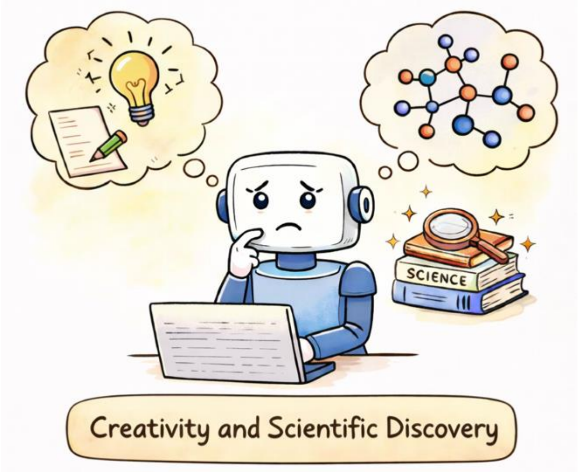
1. **Recap**: Imprecise Probability in Machine Learning
2. **Deep Learning and Foundation Models**:
   LLMs and Generative AI
3. **Societal and Regulatory Dimensions**:
   Fairness, Privacy, Ethics, and Safety
4. **Open and Unsettling Questions** in Imprecise Probability

Recap: Imprecise Probability in Machine Learning

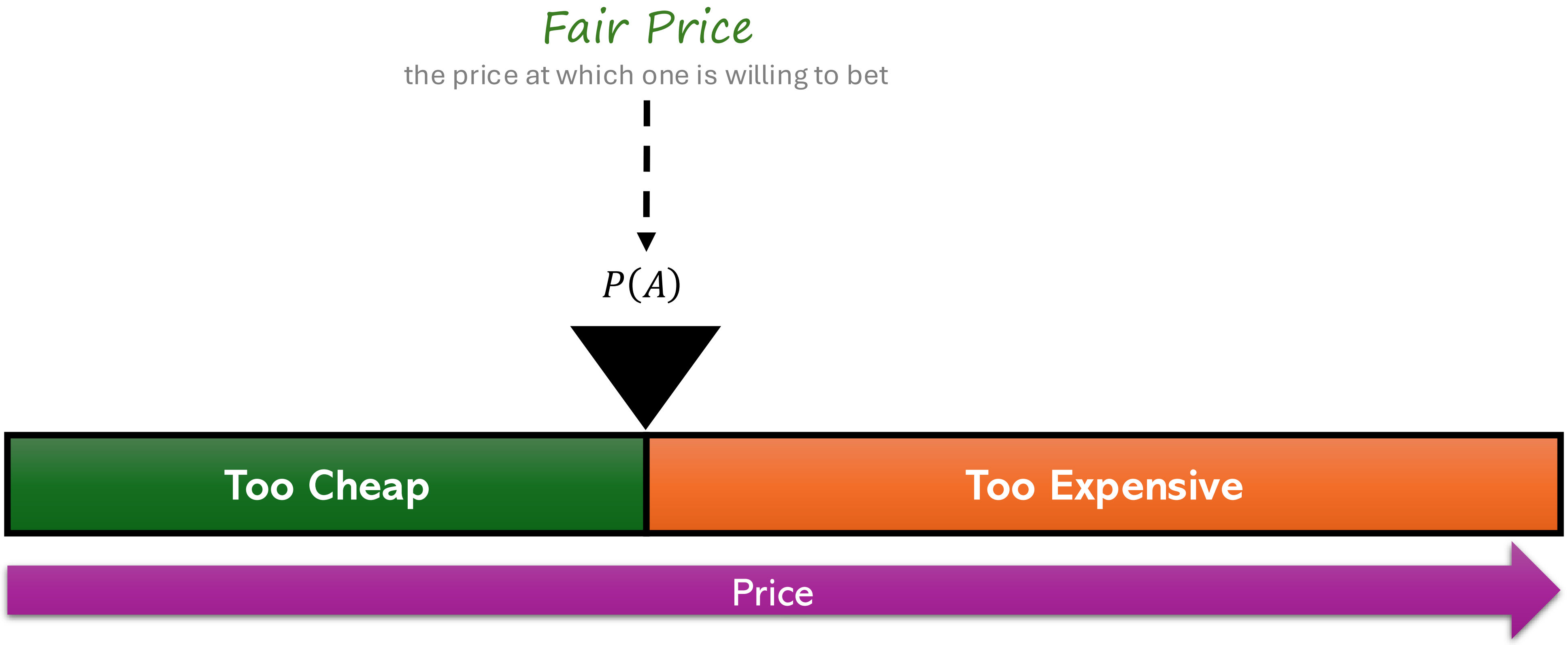
Why Imprecise Probability?

Precise probabilistic models inadequately capture real-world **uncertainty**.

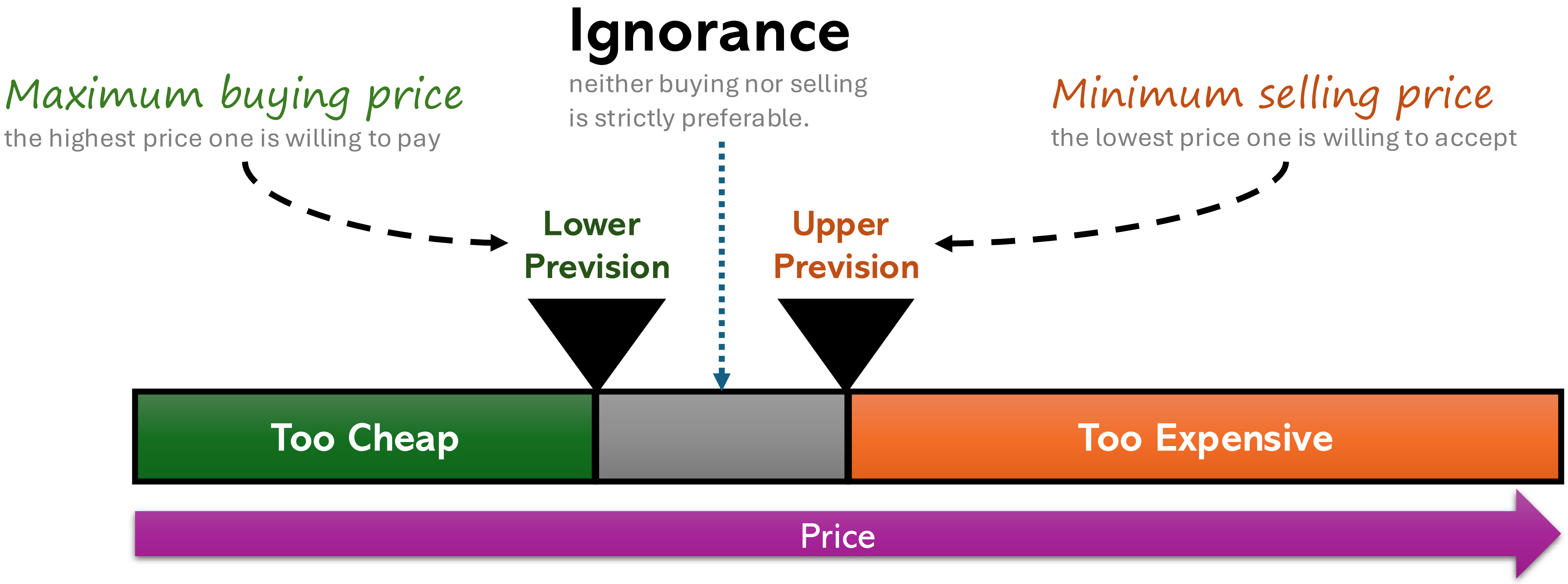




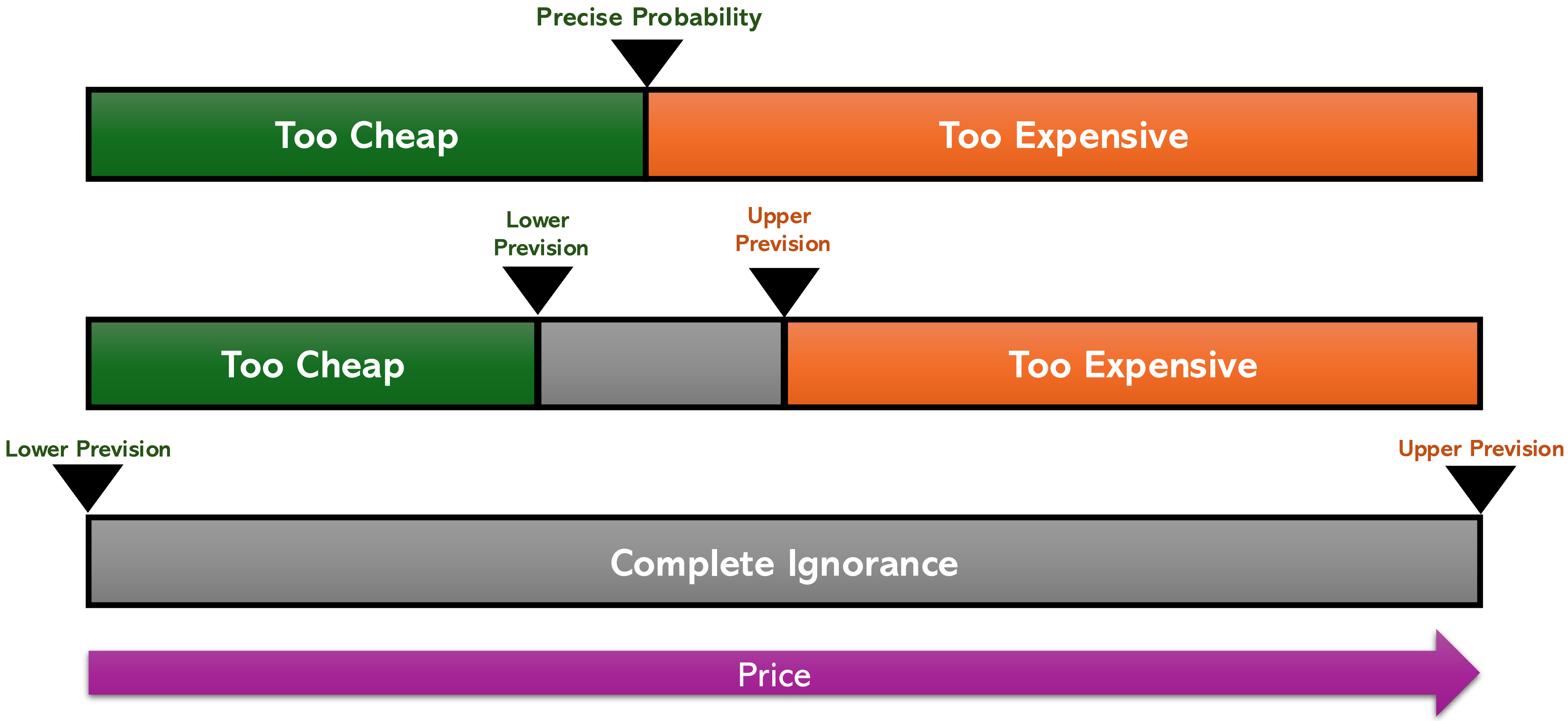
Betting Perspective



Betting Perspective

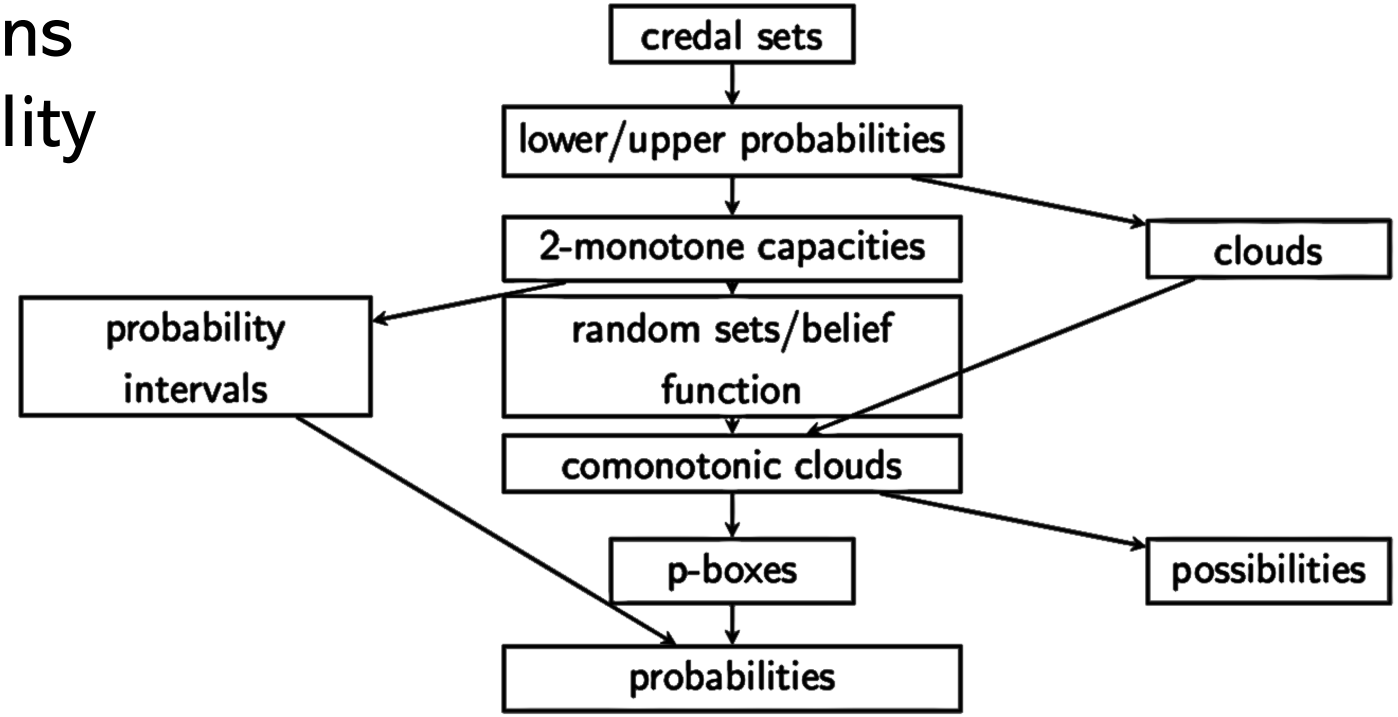


Being Precise about Imprecision



Imprecise Probabilistic Models

- Probability intervals
- Lower/upper previsions
- Lower/upper probability
- Possibility theory
- Belief functions
- Credal sets



Probability Intervals

- Let X be a random variable taking values in a *finite set* Ω
- A **probability interval** $(\underline{p}, \overline{p})$ is a pair of lower and upper probability mass functions satisfying
  1. $0 \leq \underline{p}_x \leq \overline{p}_x \leq 1$ for all outcome $x \in \mathcal{X}$ **(bounded)**
  2. $\sum_{x \in \mathcal{X}} \underline{p}_x \leq 1 \leq \sum_{x \in \mathcal{X}} \overline{p}_x$ **(proper)**
  3. $\underline{p}_x \geq 1 - \sum_{z \neq x} \overline{p}_z$ and $\overline{p}_x \geq 1 - \sum_{z \neq x} \underline{p}_z$ for all $x \in \mathcal{X}$ **(reachable)**

- The lower and upper probabilities for $S \subseteq \Omega$ can be defined as

$$\underline{P}(S) := \max\left\{\sum_{x \in S} \underline{p}_x, 1 - \sum_{x \in S^c} \overline{p}_x\right\}, \qquad \overline{P}(S) := \min\left\{\sum_{x \in S} \overline{p}_x, 1 - \sum_{x \in S^c} \underline{p}_x\right\}$$

Coherent Lower Prevision

A **gamble** f on $\mathcal{X}$ is a bounded real-valued map on $\mathcal{X}$ representing an *uncertain reward.* Two types of transaction involving a gamble f :

1. Accepting to **buy** f for a price $\mu \Leftrightarrow$ *accepting the gamble* $f - \mu$
2. Accepting to **sell** f for a price $\lambda \Leftrightarrow$ *accepting the gamble* $\lambda - f$

Lower prevision $\underline{P}(f)$ represents *supremum* acceptable buying price for f:

$$\underline{P}(f) \coloneqq \sup\{\boldsymbol{\mu} \in R : f - \boldsymbol{\mu} \in \mathcal{D}\}$$

Upper prevision $\overline{P}(f)$ represents *infimum* acceptable selling price for f:

$$\overline{P}(f) \coloneqq \inf\{\boldsymbol{\lambda} \in R : \boldsymbol{\lambda} - f \in \mathcal{D}\}$$

Remark: When $f = I_A$, $\underline{P}(I_A) \coloneqq \underline{P}(A)$ is called a *coherent lower probability.*

Possibility Theory

- A **possibility distribution** $\pi : \Omega \to [0,1]$ represents an agent's *epistemic state* of the actual state in Ω.
  - $\pi(s) = 0 \Longrightarrow$ state s is rejected as **impossible**.
  - $\pi(s) = 1 \Longrightarrow$ state s is **totally possible (plausible)**.
- Two extreme forms of partial knowledge in possibilistic framework:

Complete knowledge

$\pi(s_0) = 1$ and $\pi(s) = 0$ for all $s \neq s_0$

Complete ignorance

$\pi(s) = 1$ for all $s \in S$

Possibility and Necessity Measures

From a possibility distribution π, we can derive the **possibility** and **necessity** measures:

Possibility (upper probability)

$$\Pi(A) \coloneqq \bar{P}(A) \coloneqq \max_{x \in A} \pi_x$$

Necessity (lower probability)

$$N(A) \coloneqq \underline{P}(A) \coloneqq \min_{x \notin A} (1 - \pi_x)$$

"Does event A occur?"

- $\Pi(A)$ evaluates to what extent A is **consistent with π**
- $\mathrm{N}(A)$ evaluates to what extent A is **certainly implied by π**

Possibility and Necessity Measures

Formally, for a random variable X taking values in a set of outcomes Ω, a possibility measure satisfies:

1. **Non-negativity**: $\pi_x \coloneqq \Pi(X = x) = \Pi(\{x\}) \geq 0$ for all outcomes $x \in S$
2. **Maxitivity**: $\Pi(A) = \max_{x \in A} \pi_x$ for all events $A \subseteq \mathcal{X}$
3. **Normed**: $\Pi(\Omega) = 1$

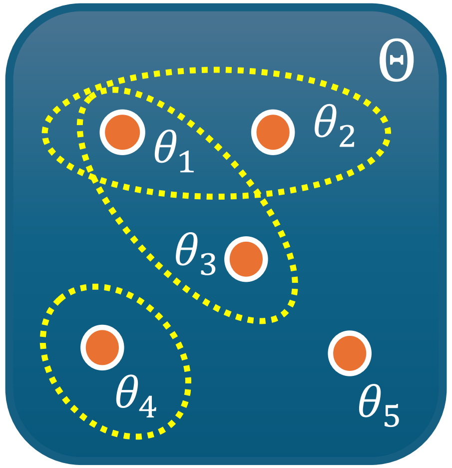
- The ***possibility-necessity duality***: $N(A) = 1 - \Pi(A^c)$.
- Generally, $\Pi(\Omega) = \mathrm{N}(\Omega) = 1$ and $\Pi(\emptyset) = \mathrm{N}(\emptyset) = 0$.

Belief Functions

- A **frame of discernment** $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$
- A basic probability assignment, or **mass function**, over Θ is a set function $m\colon 2^{\Theta} \to [0,1]$ such that

$$m(\emptyset) = 0, \qquad \sum_{A \subseteq \Theta} m(A) = 1$$

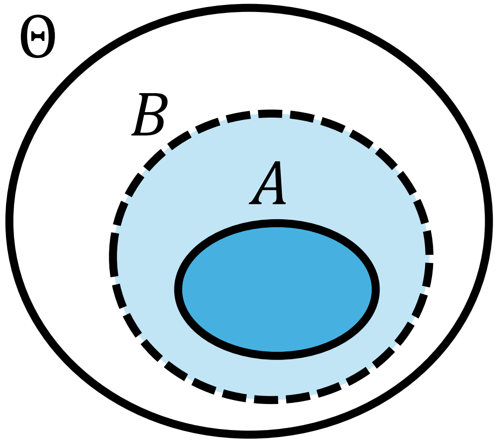
- The mass $m(A)$ measures the **belief** committed exactly to A.
- A set $A \in 2^{\Theta}$ for which $m(A) > 0$ is called a **focal set** of m.
- The union of focal sets is called a **core** of m.

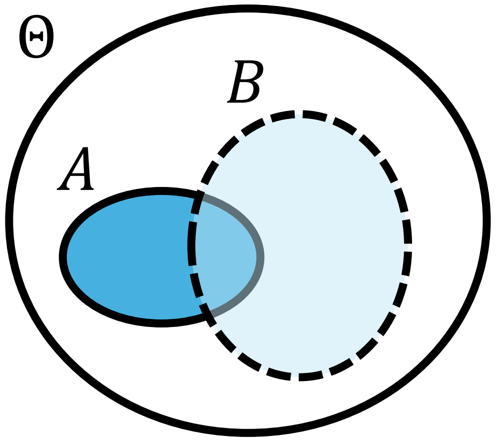


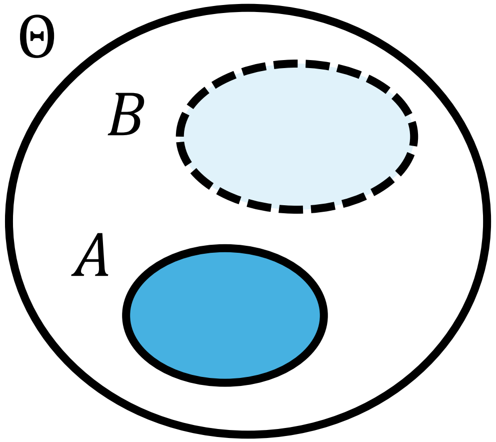
$A_1 = \{\theta_1, \theta_2\}$
$A_2 = \{\theta_1, \theta_3\}$
$A_3 = \{\theta_4\}$
$m(A_1) = 0.40$
$m(A_2) = 0.35$
$m(A_3) = 0.25$

Certainty and Possibility

- Suppose our evidence tells us that for some $A \in \Theta$, $X \in A$ **for sure and nothing more**, represented by the logical mass function m_A.
- What can we say about the proposition "$X \in B$" for some $B \subseteq \Theta$?







Certain: If $A \subseteq B$, we know for sure that $X \in B$. It supported/implied by the evidence.

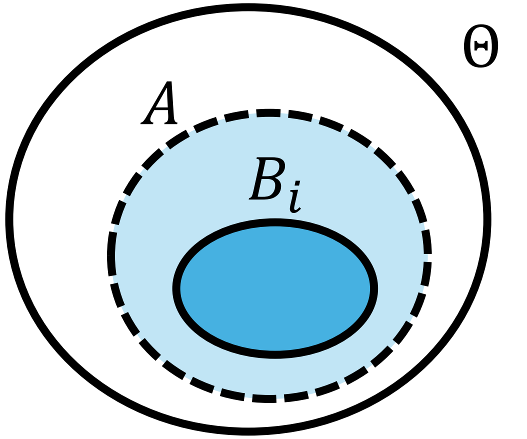
Possible: If $A \cap B \neq \emptyset$, we cannot exclude that $X \in B$. It is consistent with the evidence.

Impossible: If $A \cap B = \emptyset$, the proposition "$X \in B$" is impossible It is inconsistent with the evidence.

Belief Function

- For an arbitrary mass function m with focal sets, $B_1, B_2, \dots, B_n$, the proposition $X \in A$ is **supported by the evidence** whenever $B_i \subseteq A$.
- The **total degree of belief** supporting the fact that the true state is in $A \subseteq \Theta$ is captured by a **belief function** Bel induced by m as:

$$Bel(A) = \sum_{B \subseteq A} m(B)$$
$$= P(\{\omega \in \Omega : \Gamma(\omega) \subseteq A\})$$



- Elementary properties of belief functions:

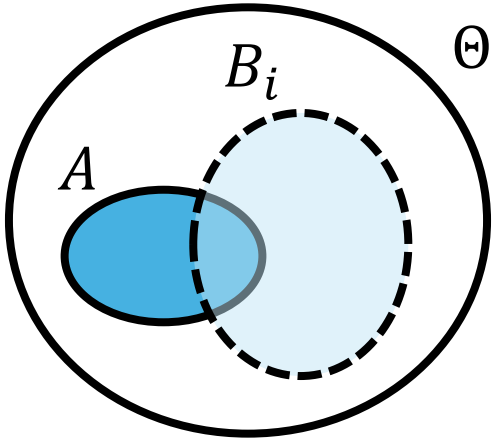
$$Bel(\emptyset) = 0, \qquad Bel(\Theta) = 1$$

Plausibility Function

- The proposition $X \in A$ is **consistent with the evidence** whenever $B_{\mathrm{i}} \cap A \neq \emptyset$.
- The **total sum of belief** that are not in contradiction with $A \subseteq \Theta$ is captured by a **plausibility function** Pl induced by m as:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - Bel(A^{\mathrm{c}})$$

- The number $Pl(A)$ is called the **plausibility** of A.
- Elementary properties:
    - $Pl(\emptyset) = 0, \quad Pl(\Theta) = 1$
    - For all $A \subseteq \Theta$, $Bel(A) \leq Pl(A)$
    - For any $A \subseteq \Theta$, $Pl(A) = 1 - Bel(\mathrm{A}^{\mathrm{c}})$



Special Cases

- If the mass function m is Bayesian,

$$Bel(A) = Pl(A) = P(A) \text{ for } A \subseteq \Theta.$$

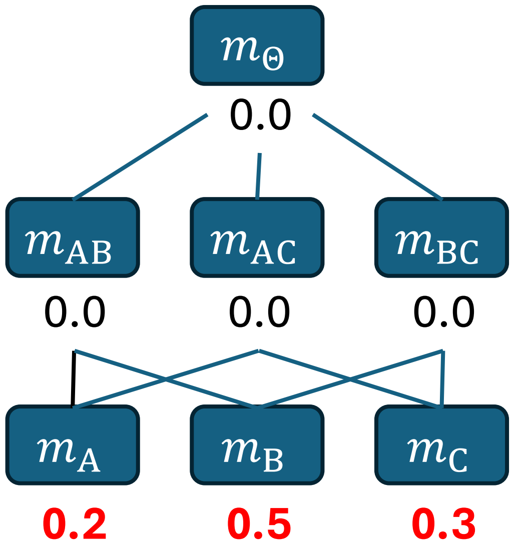
- If the focal sets of m are nested, then m is said to be **consonant**:
  - Pl becomes a **possibility measure**, i.e.,

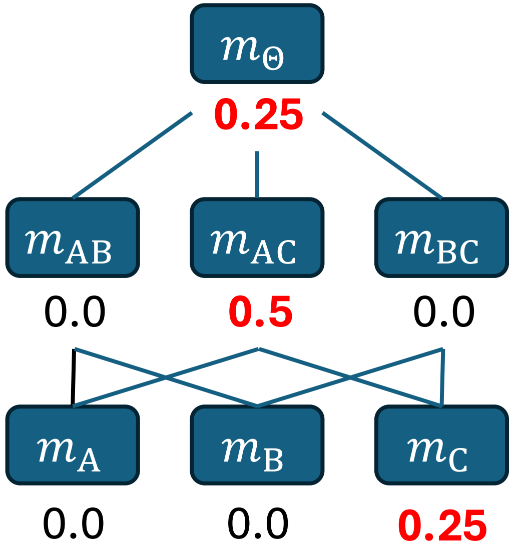
$$Pl(A \cup B) = \max\big(Pl(A), Pl(B)\big),\ \forall A, B \subseteq \Theta$$

  - Bel becomes a **necessity measure**, i.e.,

$$Bel(A \cap B) = \min\big(Bel(A), Bel(B)\big), \forall A, B \subseteq \Theta$$

- The **contour function** $pl(\theta) = Pl(\{\theta\})$ corresponds to the possibility distribution (membership function).





Credal Set

- A credal set $\mathcal{K}$ on a finite outcome space Ω is a **closed**, **convex** subset of the probability simplex:

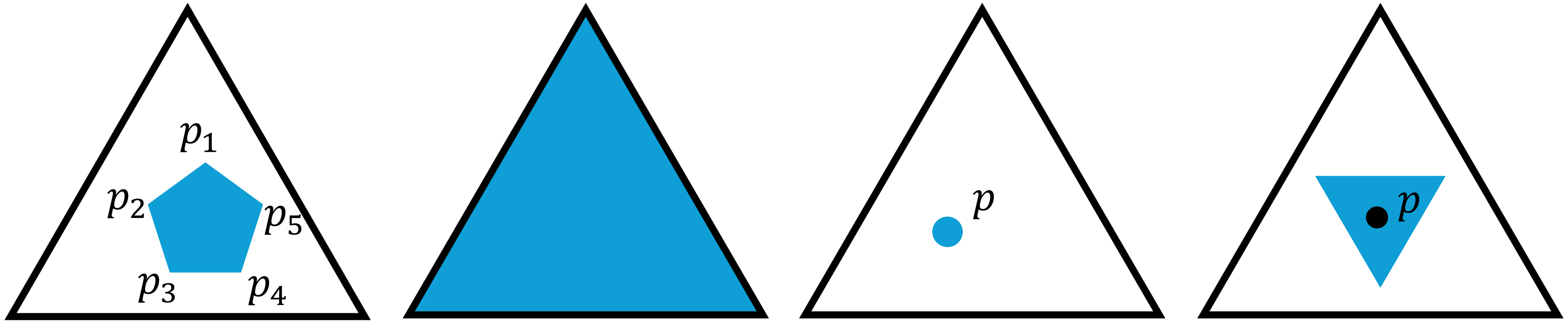
$$\mathcal{K} \subseteq \Delta^{|\Omega|-1} = \left\{ \boldsymbol{p} \in R^{|\Omega|} : p_i \geq 0, \sum_{i=1}^{|\Omega|} p_i = 1 \right\}$$

- **Convexity** ensures that any mixture of admissible probability measures is itself admissible, i.e., $p, q \in \mathcal{K} \Longrightarrow \alpha p + (1 - \alpha) q \in \mathcal{K}$ for any $\alpha \in [0,1]$.
- By the **Krein–Milman theorem**, the credal set $\mathcal{K}(\Omega)$ can be equivalently described by its **extreme points** $\text{ext}[\mathcal{K}(\Omega)]$.

Special Cases

- A finitely generated credal set (FGCS) $\mathcal{K} := \mathrm{ConvexHull}(\{p_1, \ldots, p_n\})$
- A vacuous credal set $\mathcal{K}^S := \{\, p \in \Delta^{|\Omega|-1} \colon p(S) = 1\}$ for some $S \in \mathcal{E}$.
- A vacuous credal set $\mathcal{K} := \Delta^{|\Omega|-1}$ (complete ignorance)
- A singleton credal set $\mathcal{K} := \{p\}$ (precise belief)
- A linear-vacuous or ϵ-contamination credal set:

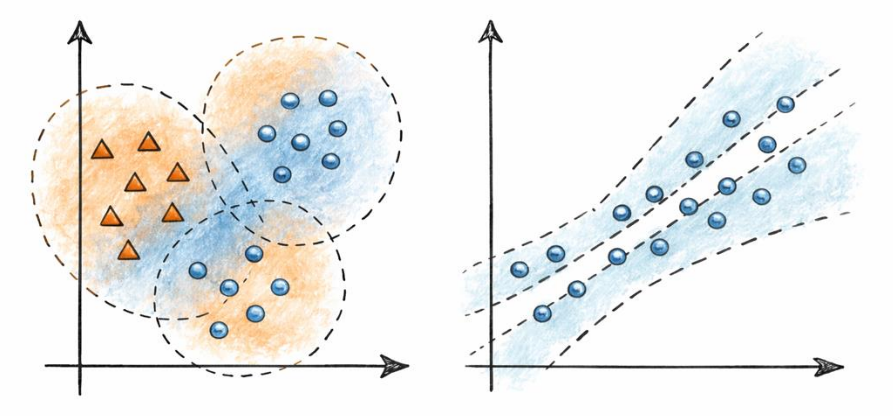
$$\mathcal{K}_{p,\epsilon} := \left\{(1-\epsilon)p + \epsilon q \,:\, q \in \Delta^{|\Omega|-1}\right\}$$



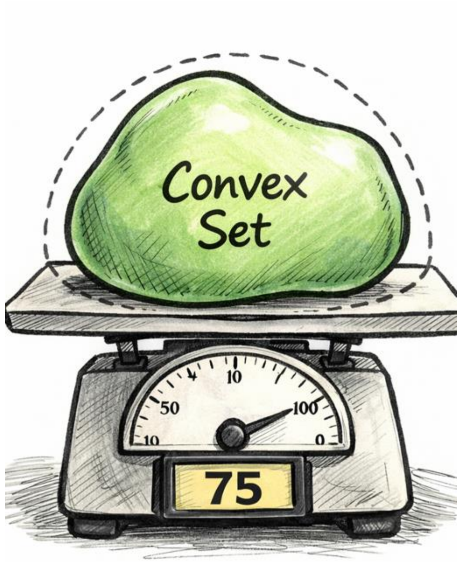
Decision Rules for Imprecise Probability

Decision Rule	Description
Statewise Admissibility	Remove actions that don't make sense **regardless of the probability**
Interval Dominance	Remove actions that have **worst best** case than some other's **worst** case
Γ-maximin	Prepare for the **worst** case
Γ-maximax	Aim for the **best** case
Maximality	Remove actions that are **strictly dominated** in expected utility over the entire credal set.
E-Admissibility	Is it the **best** action **at least in one** case?

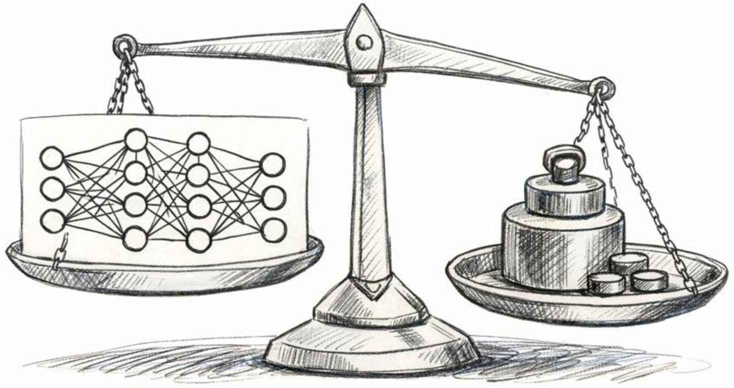
Applications in Machine Learning



Classification and Regression



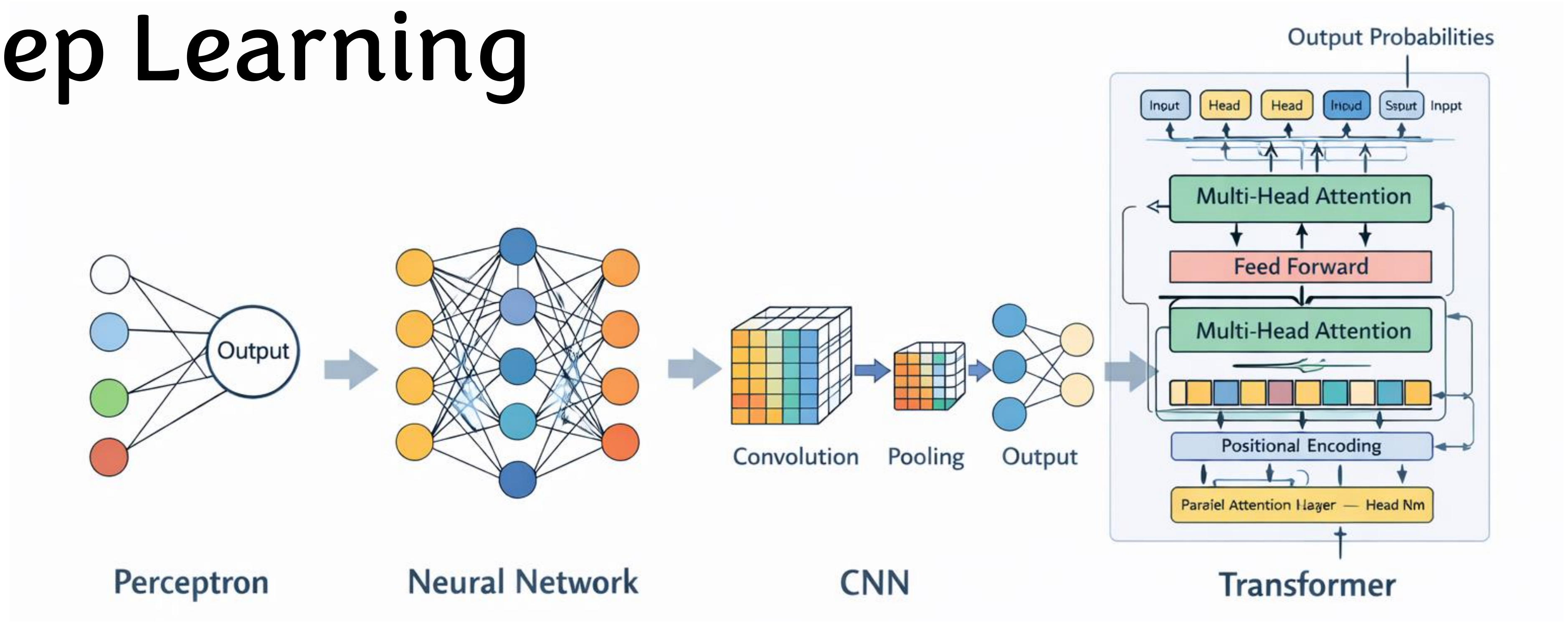
Uncertainty Quantification



Conformal Prediction and Calibration

Deep Learning and Foundation Models:
LLMs and Generative AI

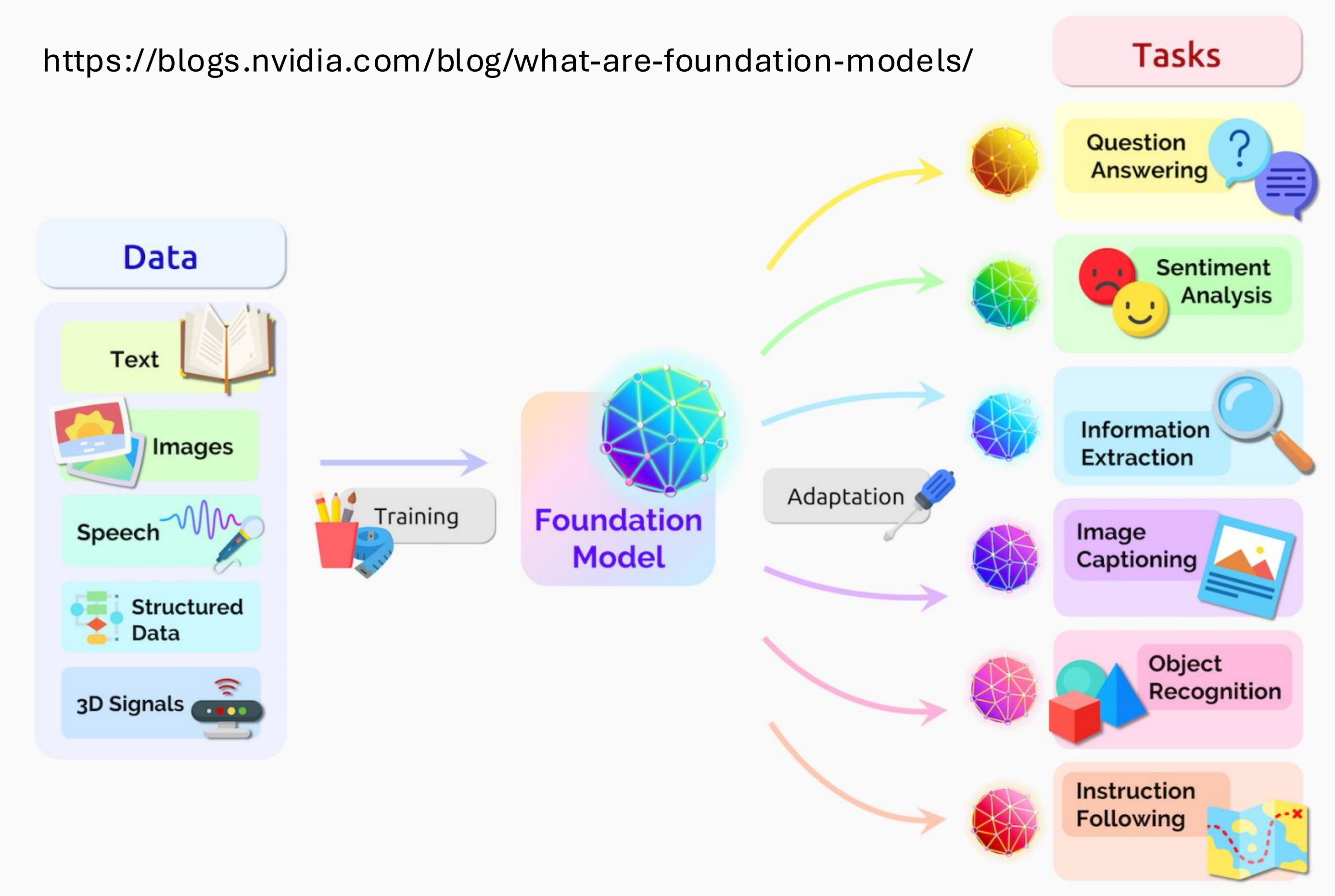
Deep Learning



- Overcoming **structural uncertainty** via inductive biases.
- Credal Bayesian deep learning (CBDL) [Caprio et al., TMLR 2024]

Foundation Models

- LLMs, vision–language models, multimodal FMs
- **Catastrophic forgetting** – *finetuning implies previous knowledge can be safely overwritten.*
- Model multiplicity

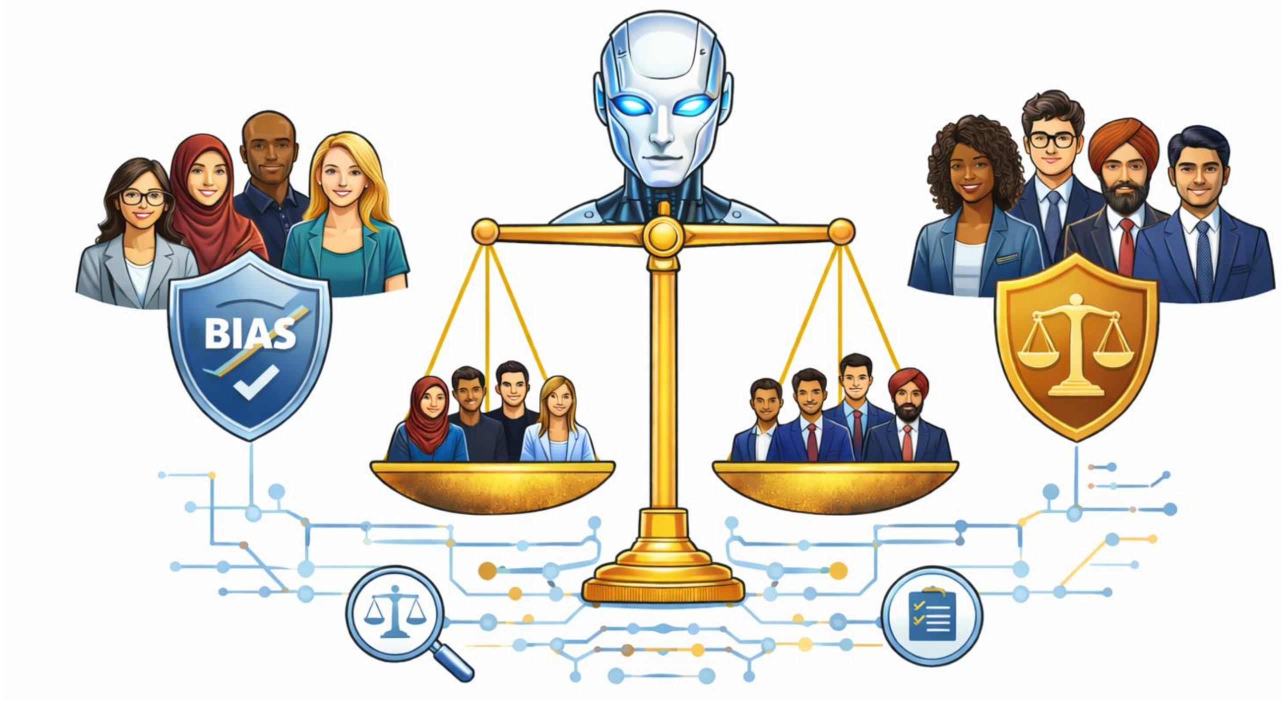


https://openai.com/index/why-language-models-hallucinate/


GENUINE CREATIVITY IN GENERATIVE AI

Societal and Regulatory Dimensions: Fairness, Privacy, Ethics, and Safety

Algorithmic Fairness



Real-world deployment exposes models to **socio-economic ambiguity** unseen during training.

Differential Privacy

- For a data space $\mathcal{X}$ and a space of summary statistics $\mathcal{T}$, define a **data-release mechanism** as

$$M\colon \mathcal{X} \times [0,1] \to \mathcal{T}$$

- A distribution on the *seed* U (e.g., $U \sim \mathrm{Unif}[0,1]$) induces a probability on the summary statistic $t = M(x, U) \sim P_x$.
- An attacker is tasked with inferring x based on observing a draw $t = M(x, U) \sim P_x$.
- Pure ϵ-differential privacy as a **Lipschitz continuity** condition and the **interval of measures**.

$$d_{\mathrm{MULT}}(P_x, P_{x'}) \leq \epsilon\, d(x, x')$$

James Bailie, Ruobin Gong. **Differential privacy: general inferential limits via intervals of measures**, *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*, PMLR 215:11-24, 2023.

AI Safety, Alignment, and Regulation

Overconfident models are **hard to align**, whereas uncertainty-aware models are easier to constrain.

AI4Science

Scientific discovery demands **exploration** beyond the boundaries of existing knowledge.

Open and Unsettling Questions
in Imprecise Probability

Update Rules for Imprecise Probability

Generalised Bayes Rule

$$\underline{P}(A \mid B) = \inf_{P \in \mathcal{K}} \frac{P(A \cap B)}{P(B)}$$

$$\overline{P}(A \mid B) = \sup_{P \in \mathcal{K}} \frac{P(A \cap B)}{P(B)}$$

Dempster's Rule of Conditioning

$$\underline{P}(A \mid B) = 1 - \overline{P}(A^c \mid B)$$

$$\overline{P}(A \mid B) = \frac{\sup_{P \in \mathcal{K}} P(A \cap B)}{\sup_{P \in \mathcal{K}} P(B)}$$

Geometric Rule

$$\underline{P}(A \mid B) = \frac{\inf_{P \in \mathcal{K}} P(A \cap B)}{\inf_{P \in \mathcal{K}} P(B)}$$

$$\overline{P}(A \mid B) = 1 - \underline{P}(A^c \mid B)$$

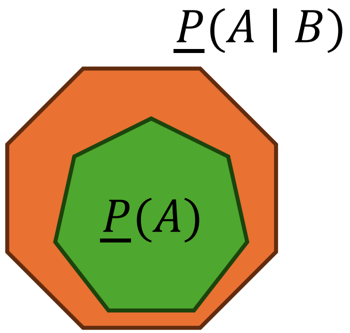
$\mathcal{K}$ is a closed, convex set of probability measures on Ω

Dilation, Contraction, and Sure Loss

Let $\mathcal{B}$ be a partition of Ω and $\mathcal{K}$ a credal set on Ω:

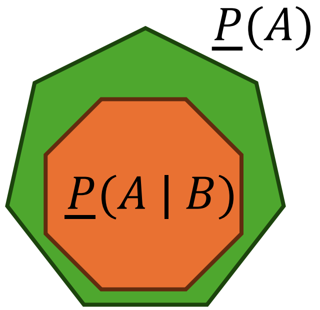
- **Dilation**: $\mathcal{B}$ strictly dilates A under the update rule if

$$\sup_{B\in\mathcal{B}} \underline{P}(A\mid B) < \underline{P}(A) \leq \overline{P}(A) < \inf_{B\in\mathcal{B}} \overline{P}(A\mid B)$$



- **Contraction**: $\mathcal{B}$ strictly contracts A under the update rule if

$$\underline{P}(A) < \inf_{B\in\mathcal{B}} \underline{P}(A\mid B) \leq \sup_{B\in\mathcal{B}} \overline{P}(A\mid B) < \overline{P}(A)$$



- **Sure loss**: $\mathcal{B}$ incurs sure loss in A under the update rule if either

$$\inf_{B\in\mathcal{B}} \underline{P}(A\mid B) > \overline{P}(A) \quad \text{or} \quad \sup_{B\in\mathcal{B}} \overline{P}(A\mid B) < \underline{P}(A)$$

Ruobin Gong. Xiao-Li Meng. "**Judicious Judgment Meets Unsettling Updating: Dilation, Sure Loss and Simpson's Paradox.**" Statist. Sci. 36 (2) 169 - 190, May 2021. https://doi.org/10.1214/19-STS765


Unknown
Known

Recommended Reading

- Caprio, M., Dutta, S., Jang, K. J., Lin, V., Ivanov, R., Sokolsky, O., & Lee, I. (2024). **Credal Bayesian Deep Learning**. *Transactions on Machine Learning Research (TMLR)*.
- James Bailie, Ruobin Gong. **Differential Privacy: General Inferential Limits via Intervals of Measures**. *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*, PMLR 215:11-24, 2023.
- Ruobin Gong. Xiao-Li Meng. **Judicious Judgment Meets Unsettling Updating: Dilation, Sure Loss and Simpson's Paradox**. Statist. Sci. 36 (2) 169 - 190, May 2021.