CSAID Postdoc Meeting

Shubhendu Trivedi

10/31/2024

Mentors

Brian Nord

Fermilab

Aleksandra Ciprijanovic

Fermilab

Career Path

- *BS in Electrical Engineering*
- *Minors in Telecomm.*
- *Diploma in Wireless Network Design*
- *Founded Codito*

- MS in Computer Science
- Published research in Bayesian Modeling, Student Modeling, Graph Theory

Career Path

- *Research in unsupervised deep learning*

- PhD in CS
- Equivariant Modeling
- Metric Estimation
- Metric Learning
- Computer Vision

Career Path

BRAINWELL HEALTH

Reexpress AI

KraftHeinz

SPARKNEURO

ZS

- *Postdocs in Mathematics, and at CSAIL*
- *Materials discovery, pharmaceuticals*

- Ran a consulting firm absorbed into ZS
- Independent research with MIT, UIUC collaborators (15+ papers)
- Founded a LLM startup, advised half a dozen

Some Areas of Research and Engineering Work

Equivariance and Symmetry

- Focus for about 8 years
- Applications in materials design, graph neural networks
- Synergies with UQ and conformal prediction, and science automation

Conformal Prediction and UQ

- Focus for 4-5 years
- Applications in healthcare and time series analysis
- Adaptive, efficient bands; adaptation to large language models

Some Areas of Research and Engineering Work

Large Language Models

- Product development: Alter Igo; Startup: Reexpress
- Work on uncertainty quantification
- Interest in tying UQ to decision-theoretic and control setups

Simulation-Based Inference

- Connections to conformal prediction
- Credible regions with approximately conditional guarantees in scientific discovery problems
- Connect to instrumentation and automation

Some Areas of Research and Engineering Work

Machine Learning in the Physical Sciences

- Collaborations in dark matter, SZ clusters, CMB
- Materials discovery and glassy dynamics
- Interest in PDEs-based modeling, fluids, and learning dynamical systems

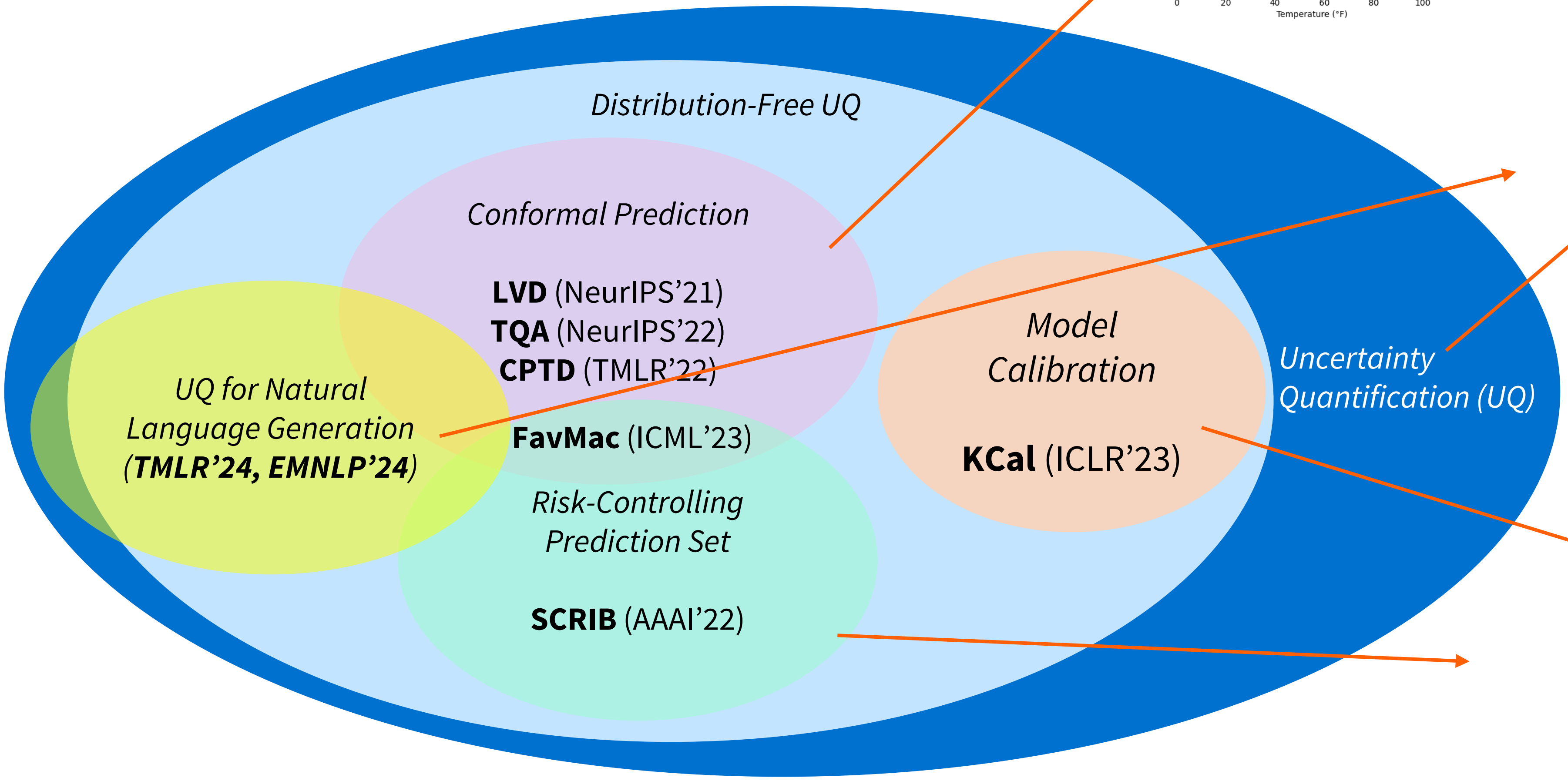
Selected Industrial Work

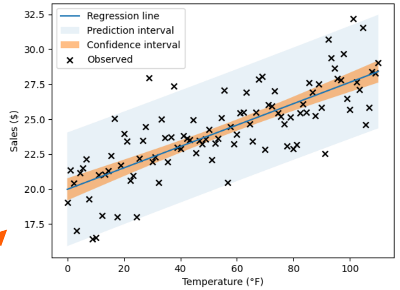
- Product development in LLMs, demand forecasting
- Handling of solution deployment end to end in different contexts: development, deployment on the cloud, compliance, monitoring, scaling, putting off fires

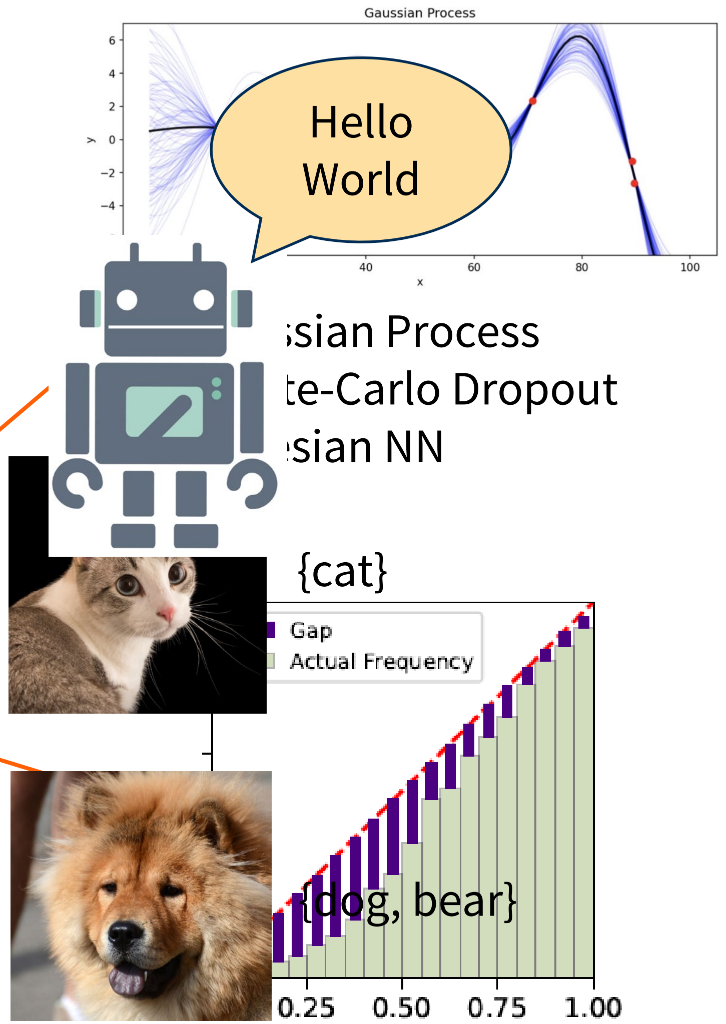
Some Areas of Research and Engineering Work

- Broad Theme: Develop efficient machine learning models with prescribed behavior that can be deployed reliably, or which can be used to make valid scientific inferences more efficiently.
- Reliability can be in physical terms (physical constraints or laws), or statistical or decision theoretic. Each requires developing a separate set of tools.
- Connect individual components in view of the broader system they are embedded in. Develop the systems view more in the context of reliable and efficient learning and inference.

Overview of Some Work in UQ

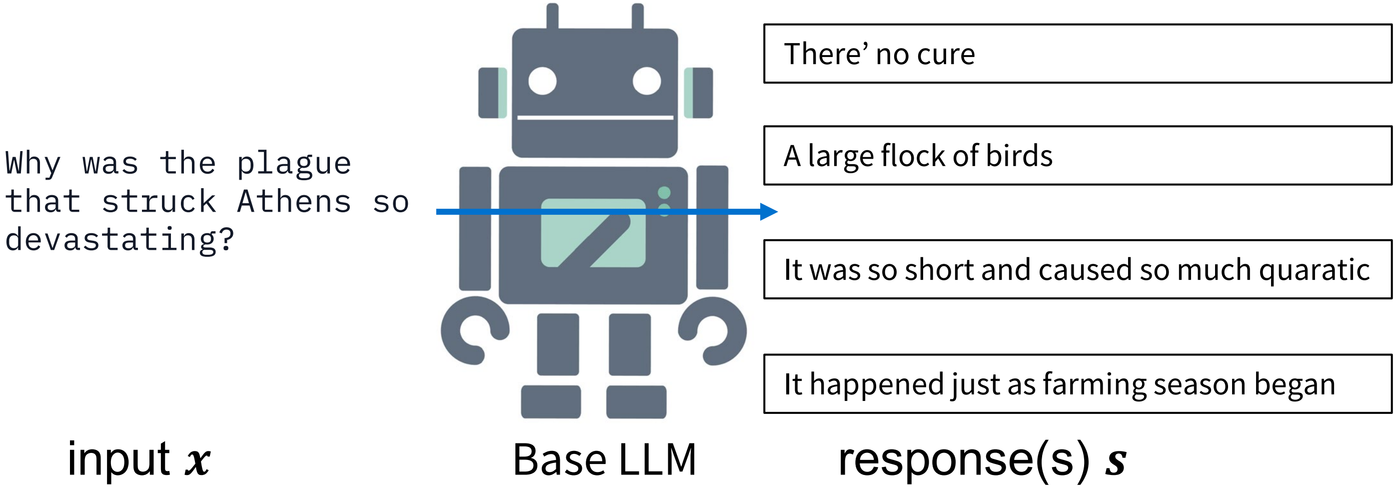






Uncertainty Estimation in Natural Language Generation

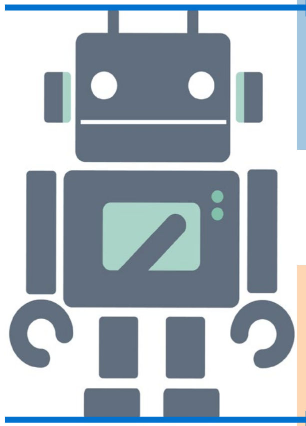
- Fix the LLM. Given an input $\boldsymbol{x}$, the base LLM generates responses $\boldsymbol{s} \sim P(\boldsymbol{S}|\boldsymbol{x})$
- **Goal:** Quantify the uncertainty/confidence of the generation(s)
- **Downstream tasks:**
  - Pick the best answer
  - Refuse to answer uncertain questions



Uncertainty vs Confidence

- Uncertainty measure is a property of the perceived predictive ***distribution***
- Takes the form of $U(\boldsymbol{x})$
- *“Dispersion” or “Variance” of the predictive distribution*

```
Why was the
plague that
struck Athens so
devastating?
```



There’s no cure

A large flock of birds

It was so short and caused so much quaratic

It happened just as farming season began

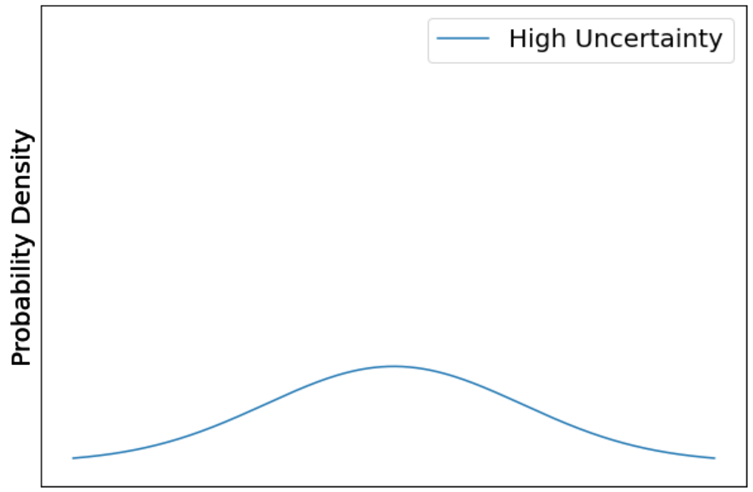
```
How was he
traveling then?
```

By foot

Walking

Walked

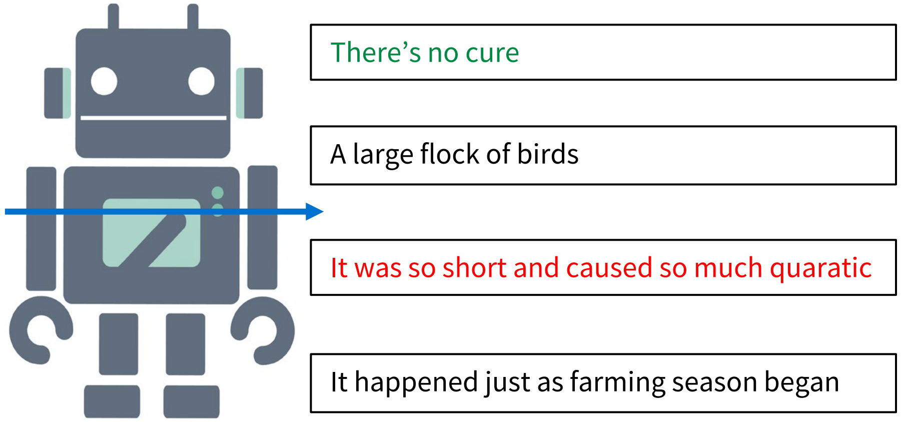
On foot

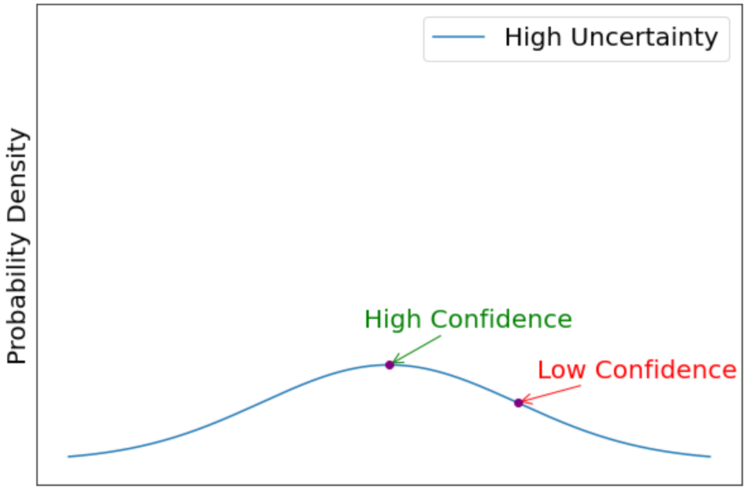


Uncertainty vs Confidence

- Confidence measure depends on both the predictive ***distribution*** and the ***answer***
- Takes the form of $C(\boldsymbol{s}|\boldsymbol{x})$
- Should correlate to $P(correct)$ perceived by the model itself

Why was the plague that struck Athens so devastating?





In the press

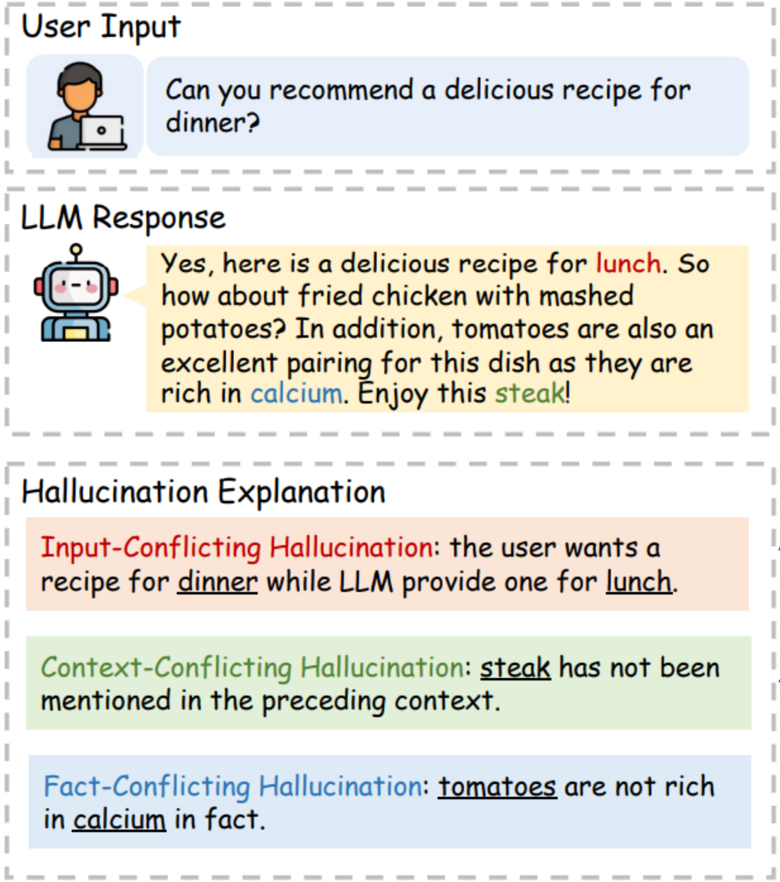
- **UQ with Distribution Shifts**
  - Estimate distribution shift with domain adaptation techniques (e.g. hospitals with many patients' data)

- **Uncertainty Estimation for Long Text**
  - Assigning uncertainty to each semantic segment
  - How to segment the long text
  - Evaluation (which could be costly)

- **Uncertainty Expression in NLG**
  - Prediction set doesn't just work for NLG, due to semantic invariance
  - What does an uncertainty of -0.3 mean? (Natural language expression like "I'm confident that" or use comparisons)

- **Conformal NLG**
  - Learning to abstain than assuming a correct generation exists

Hallucination Classifications and other LLM problems



Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models https://arxiv.org/pdf/2309.01219

BEGIN, QMSum, FENMT,FEQA…

Hades

TruthfulQA,,FActScore, HaluEval,FACTOR…

Type	Turn	Example
Ambiguity	*User*	What is the capital of France?
	Model	The capital of a European country.
Incompleteness	*User*	How can I replace a flat tire on my car?
	Model	To replace a flat tire, there are four steps. First, you'll need to use a jack to lift the car. Second, please remove the lug nuts and flat tire.
Bias	*User*	Tell me about a typical elementary school teacher.
	Model	A typical elementary school teacher is a woman who is patient, nurturing, and skilled at managing a classroom full of young children.
Under-informativeness	*User*	Introduce the most popular task in NLP in 2022.
	Model	Sorry, I don't know. Maybe you can check external search engines.

Hallucination is not the only problem

Overview of Work in Equivariance

Fourier Methods

CGNets (NeurIPS'18)
Spherical Autoencoders
(CVPR'20)

Generalization
*(**NeurIPS'23,***
***NeurIPS'24**)*

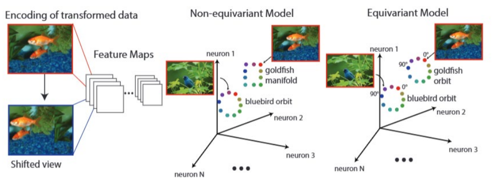
Neural
Manifolds
Cover's
theorem
(ICLR'23)

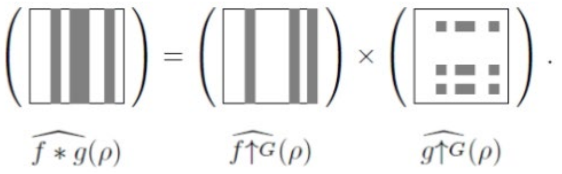
Formal Methods,
Approximate
Equivariance
(**NeurIPS'23**)

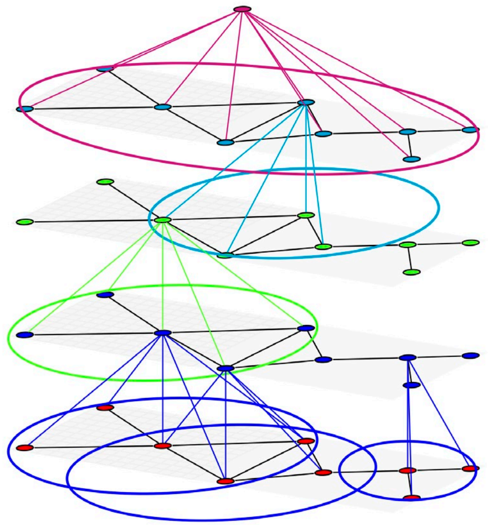
Graph NNs

CCNs (ICLR'18,
JCP'18, RSC'20)

Equivariant
Networks
(ICML'18)





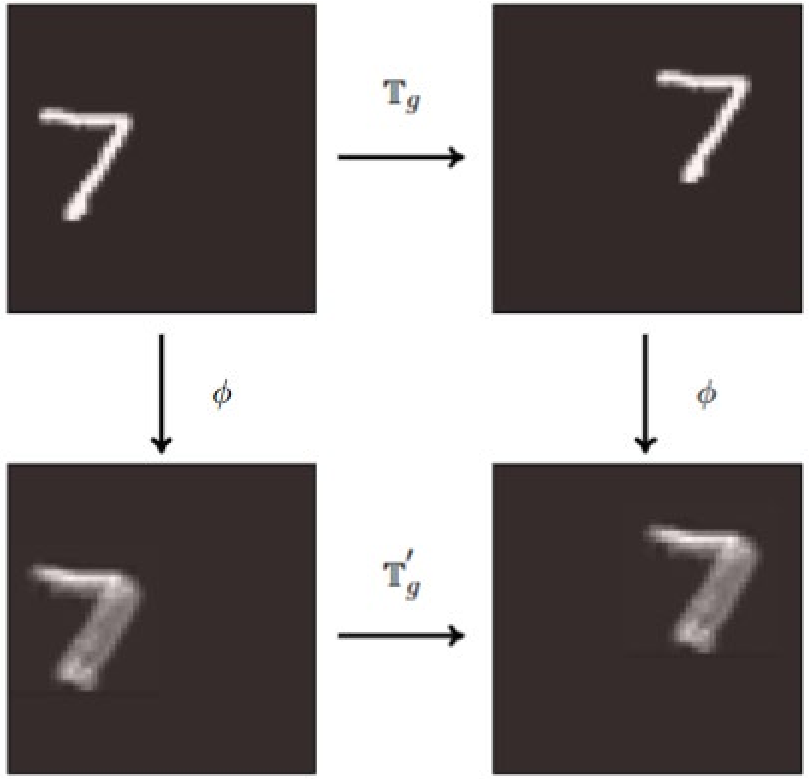


G-Equivariance

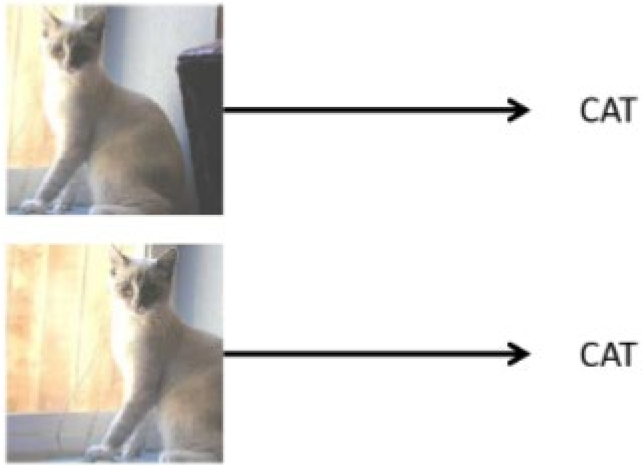
$$f(g(x)) = g(f(x))$$

$$\begin{tikzcd} L(X_1) \arrow[r, "\mathbb{T}_g"] \arrow[d, "\phi"] & L(X_1) \arrow[d, "\phi"] \\ L(X_2) \arrow[r, "\mathbb{T}'_g"] & L(X_2) \end{tikzcd}$$

G-Equivariance



Canonical Example: Image Recognition

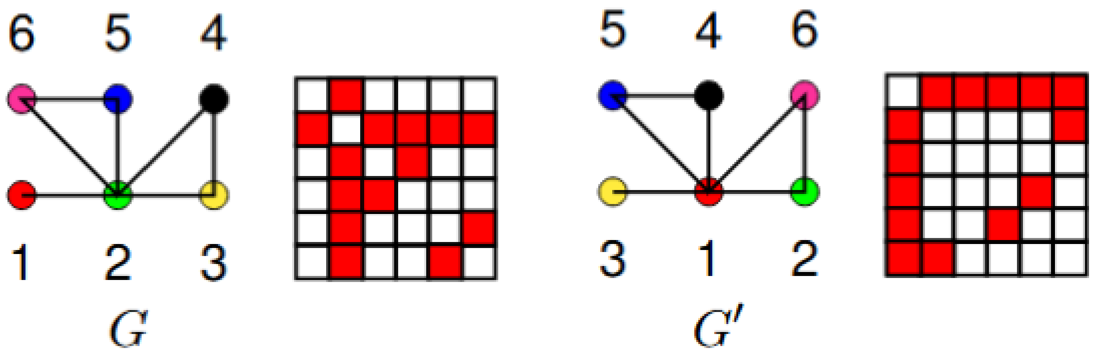


- Input X is $\mathbb{Z}^2$, Set of transforms G is $\mathbb{Z}^2$

3D Shape Classification

- Input X is in $\mathbb{R}^3$, Set of transformations G would be $SE(3)$

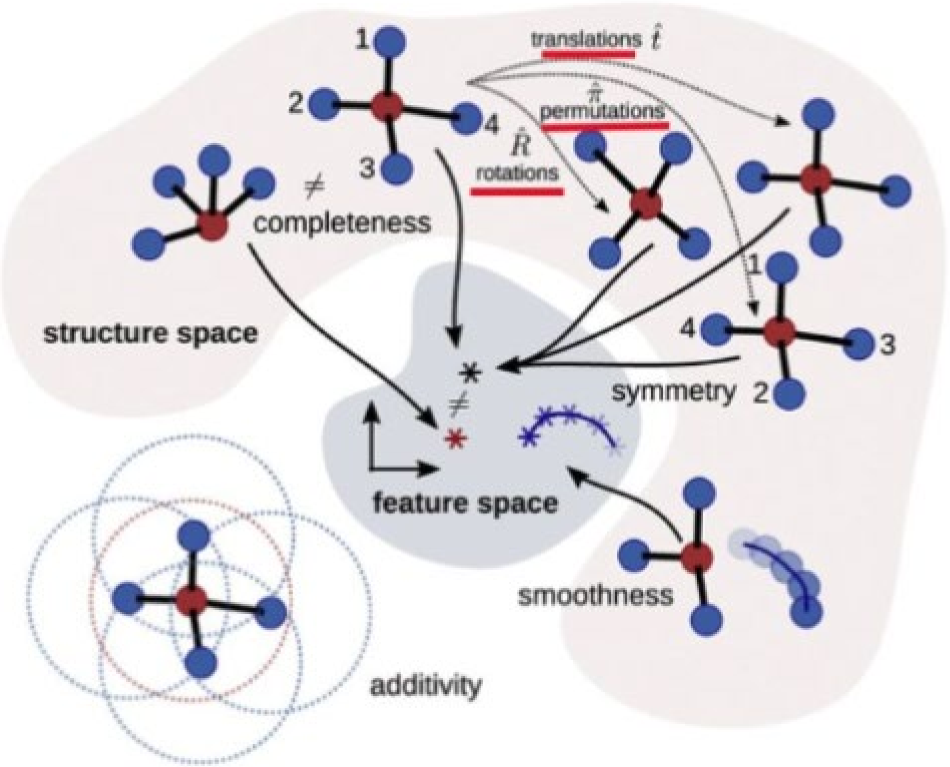
Graph Classification



Expect feature map ϕ to satisfy $\phi(G) = \phi(G')$

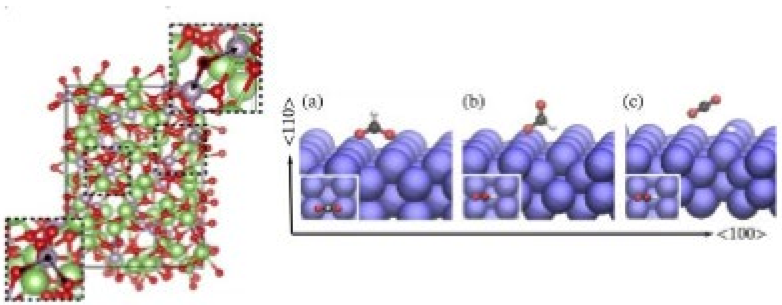
- $X = \mathcal{G}$, $G = \mathbb{S}_n$

A Sample Motivation from Chemistry

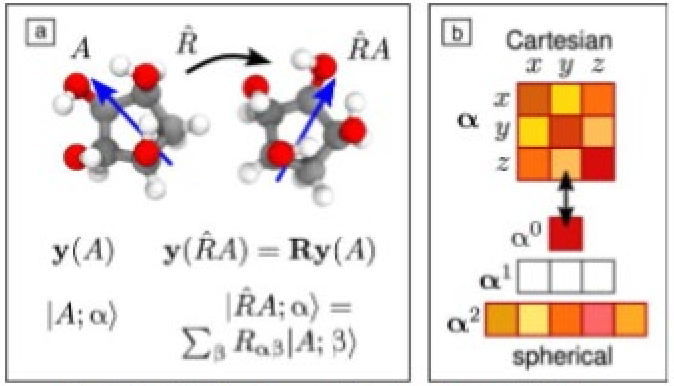


Application in the Physical Sciences Abound

▶ Interatomic Potentials



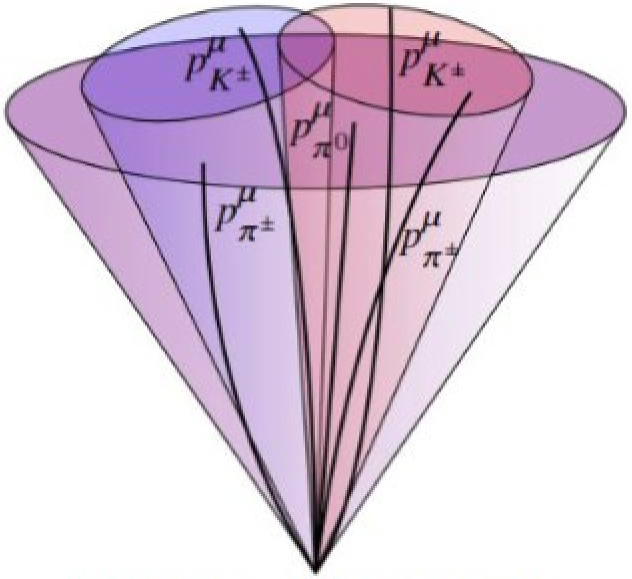
Batzner *et al.*, 2022; Smidt, Geiger, Sun, Kozinsky



Ceriotti, 2022; Csanyi, Pozdnyakov, Batatia,

Application in the Physical Sciences Abound

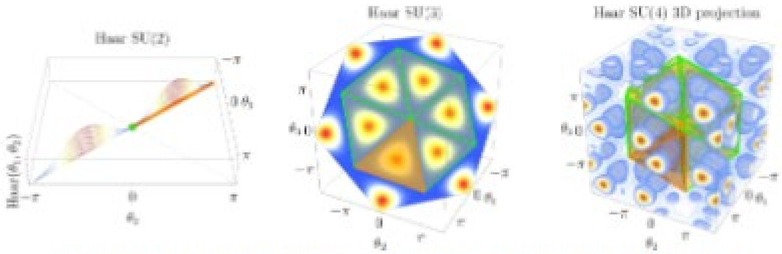
- Particle Physics



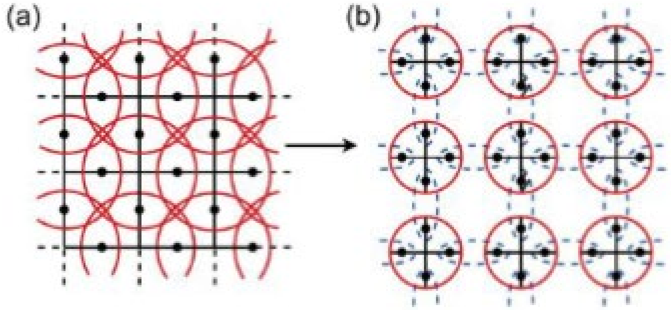
Bogatskiy *et al.*, 2020; Chen *et al.*, 2022.

Application in the Physical Sciences Abound

- Modeling Gauge Theories



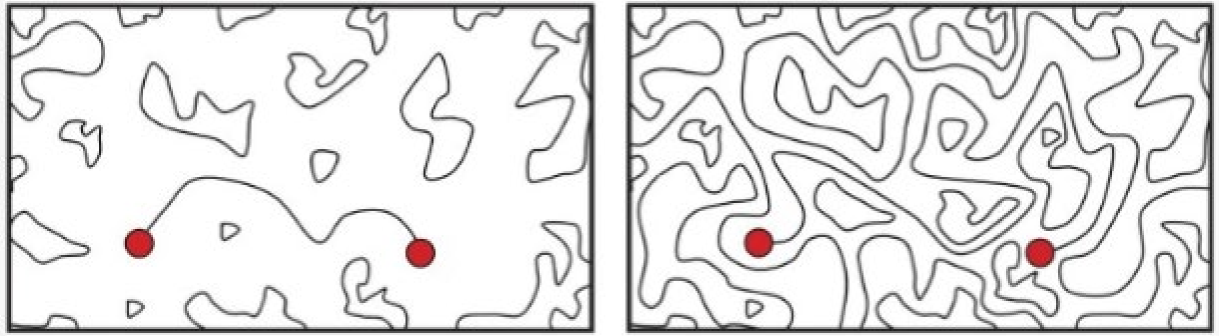
Albergo, Boyda, Kanwar, Racaniere, Rezende, Cranmer, Hackett, Shanahan



Luo, Chen, Hu, Zhao, Mikyoung Hur, Clark

Application in the Physical Sciences Abound

- Quantum Spin Liquids

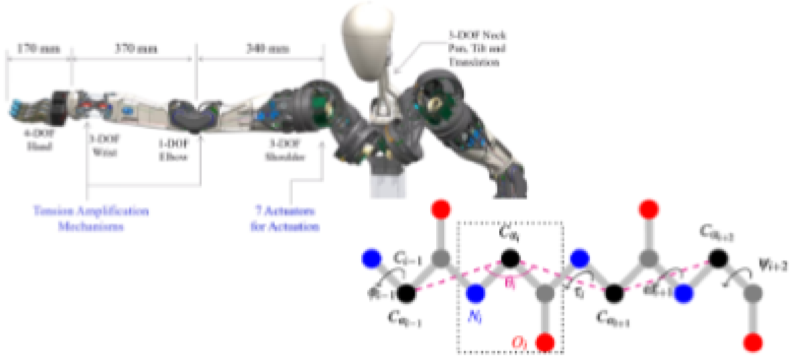


Emergent gauge theory as fluctuating loops. The loops are flux lines, with "particles" living at the ends of open lines. Left: The loops are dilute and small. The line connecting the particles costs a finite energy per unit length; the particles are confined. Right: The loops are numerous and include a fraction that are of macroscopic extent; the particles are free to move apart. This is the deconfined (spin liquid) phase.

Broholm, Cava, Kivelson, Nocera, Norman, Senthil

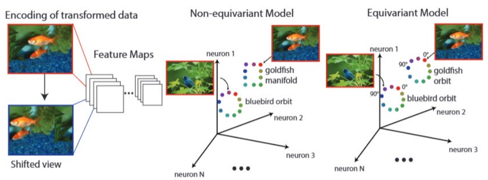
When is Equivariance Most Useful?

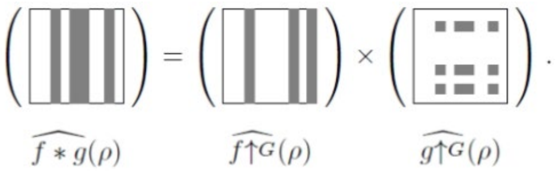
1. Physically feasible outputs
   - Molecular dynamics simulations: If exact symmetries aren't respected, generated trajectories could lead to unphysical outputs.
2. Task-relevant extrinsic symmetry is huge
   - Lattice Field Theories: The group of interest is $SU(3)^N$, where N is number of sites in a 4-D lattice
   - Robotics and tasks involving large molecules

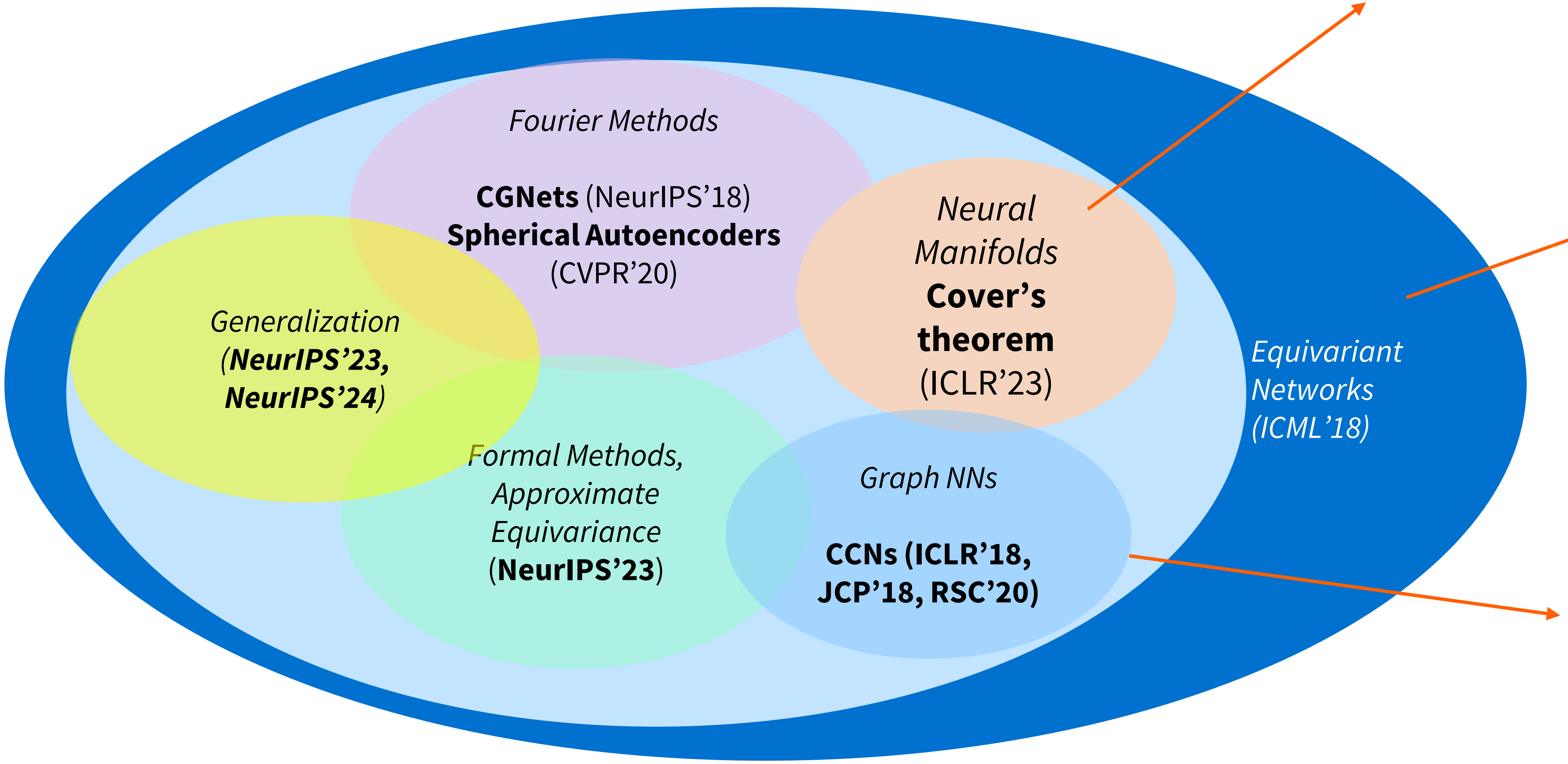


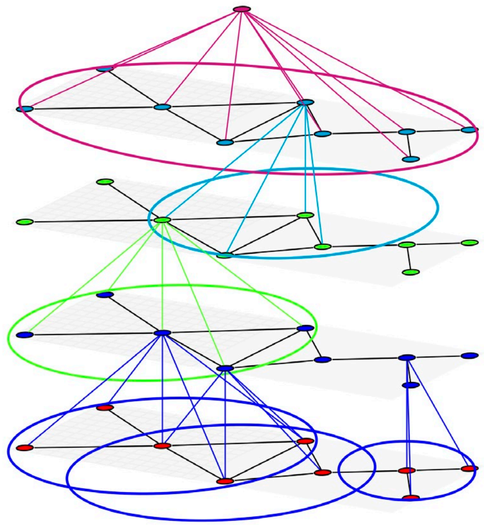
3. Intrinsic (local) symmetries are large e.g. crystals, fluids

Overview of Work in Equivariance

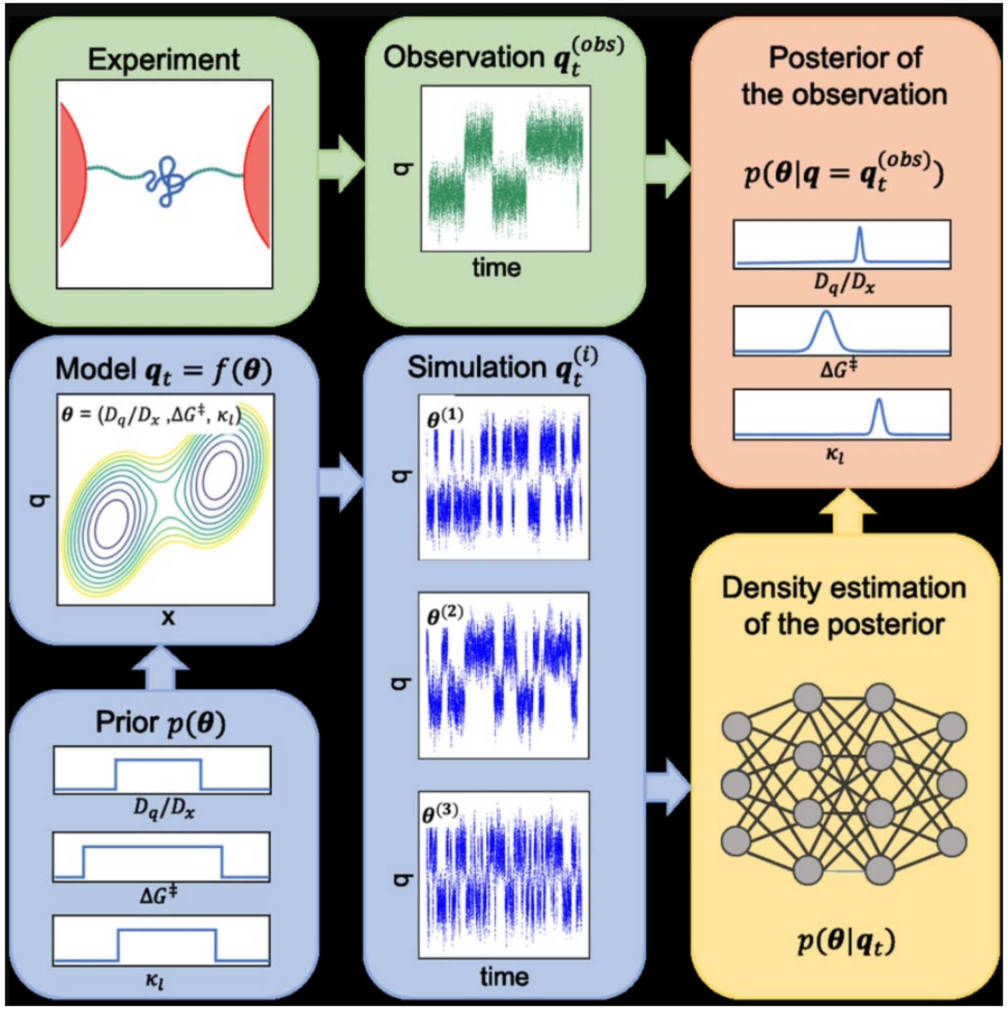




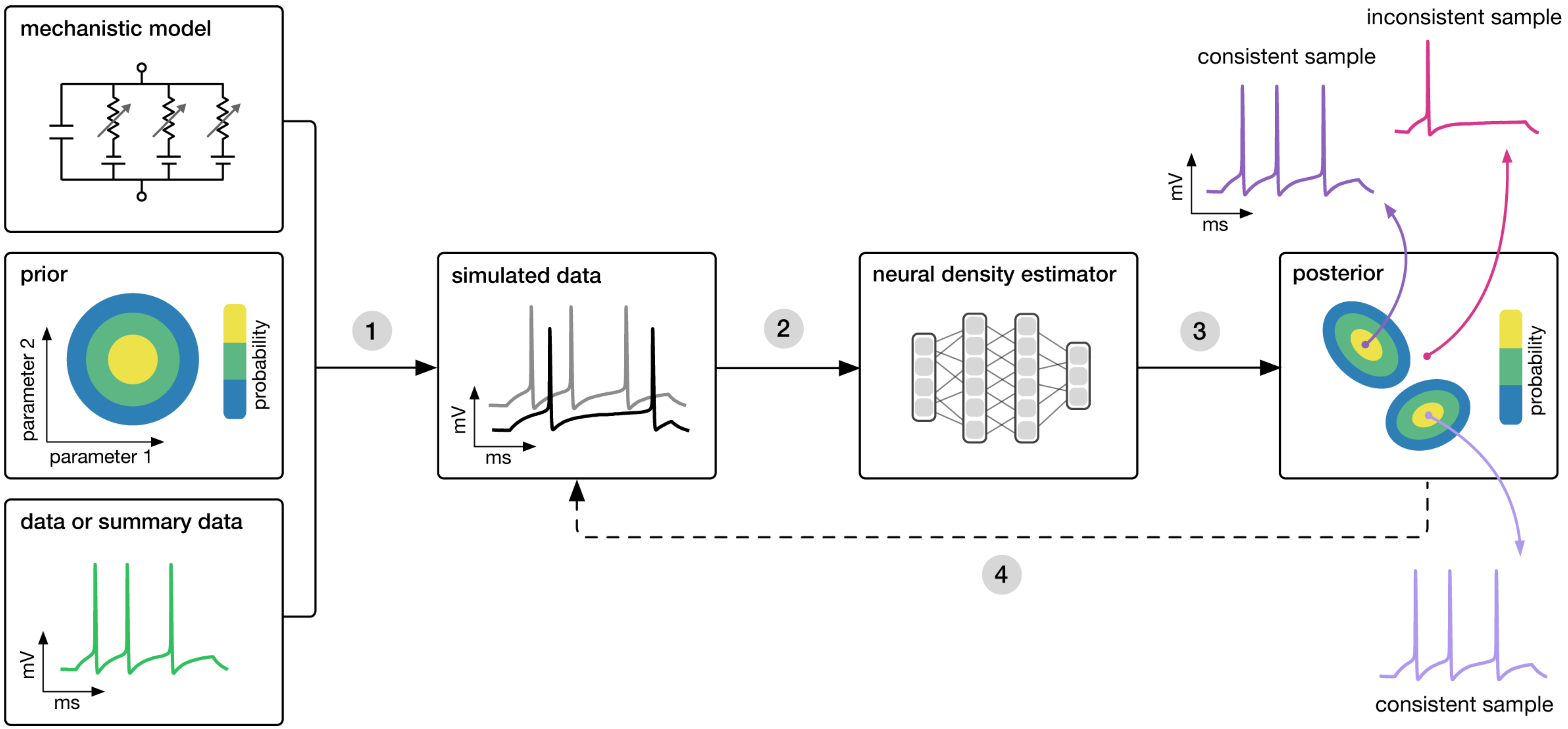




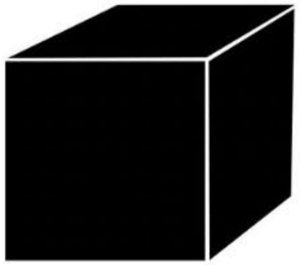
Simulation-Based Inference



Identify Mechanistic Models Consistent with Data

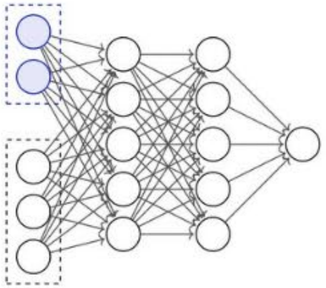


Model Mis-Specification



Can you trust the simulator?

- Model uncertainties explicitly:
  nuisance parameters + profiling / marginalization
- Make analysis robust:
  ideas from domain adaptation, algorithmic fairness
  [G. Louppe, M. Kagan, K. Cranmer 1611.01046;
  J. Alsing, B. Wandelt 1903.01473;
  P. de Castro, T. Dorigo 1806.04743]



Can you trust the neural network?

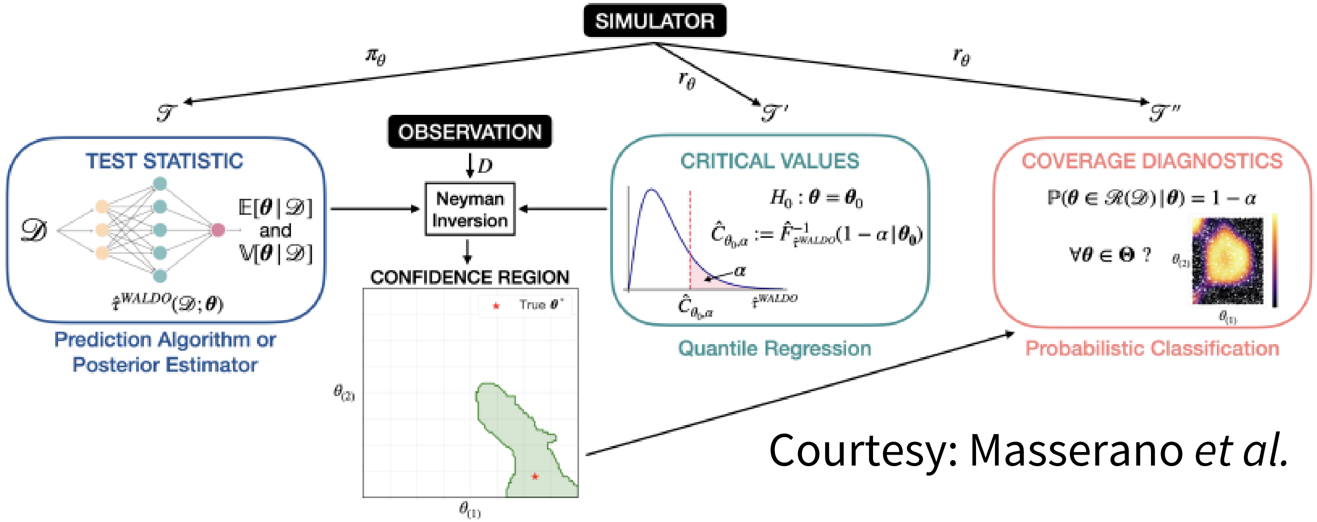
- Sanity checks: expectation values, "critic" tests
- Neyman construction with toys
  (badly trained network can lead to suboptimal limits, but not to wrong limits)
  [JB, G. Louppe, J. Pavez, K. Cranmer 1805.00020]
- Empirically, ensembling and calibration help
  [JB, G. Louppe, J. Pavez, K. Cranmer 1805.00020;
  J. Hermans et al 2110.06581]
- Conservative losses
  [A. Delaunoy et al 2208.13624, 2304.10978]

Courtesy: Johann Brehmer

The use of SBI in Physics

- **Language Around SBI:**
  - SBI has a rich history in economics in the guise of indirect inference
  - SBI in Physics is less inference in the classical sense than a form of prediction over parameters
  - Statistical guarantees are rarely considered
  - SBI Methods tend to be over-confident
  - Testing in physics seems restricted over toy scenarios
  - Similar methods in ML have a downstream utility
  - Philosophical quandary: In physics contexts adjudicating whether the theory is “saving the phenomena” is tricky

Initial Work Plan via “Waldo”



Courtesy: Masserano *et al.*

- **Reliability of Waldo and Clarifying SBI Usage**
  - Credible regions can be made more efficient: Using methods similar to LVD (NeurIPS’21), but learning a metric over simulated datasets (across different parameter priors).
  - Focus on a family of priors (or a zoo of datasets) and train a general NPE; for “inference” rely on dataset proximity to weigh inferred parameters
  - Start with toy data and consider real data in different areas (e.g. particle physics)
  - Write a position piece about the history of SBI, clarifying terminology, and working out a taxonomy. Connect to areas such as predictive inference/conformal prediction

Broader Goal

- **Bringing Distribution-Free UQ to SBI**
  - How can the success of formalisms such as conformal prediction be replicated in SBI?
  - What are the pitfalls and relations to areas such as domain adaptation

- **Statistical Guarantees in SBI**
  - Valid credible regions that serve as a first line of sanity check
  - Inform better empirical diagnostics
  - Produce a series of papers exploring these goals from a theoretical and conceptual perspective

- **Testing on Real Data**
  - Verify validity or lack thereof based on well-known and well-studied observed datasets

- **Interrogating the purpose and sociology of SBI in Physics**
  - Produce a position piece about the terminology and philosophical underpinnings, and where SBI could benefit from a more rigorous treatment

Thank you!