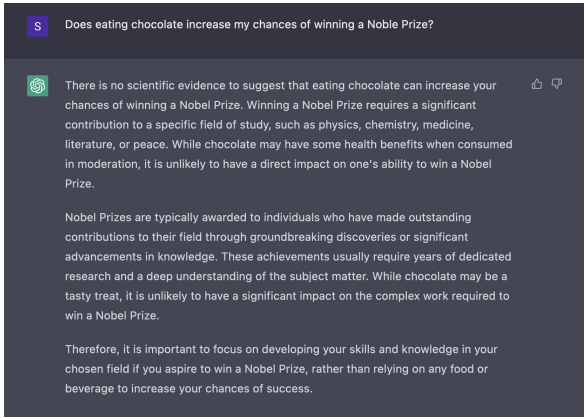
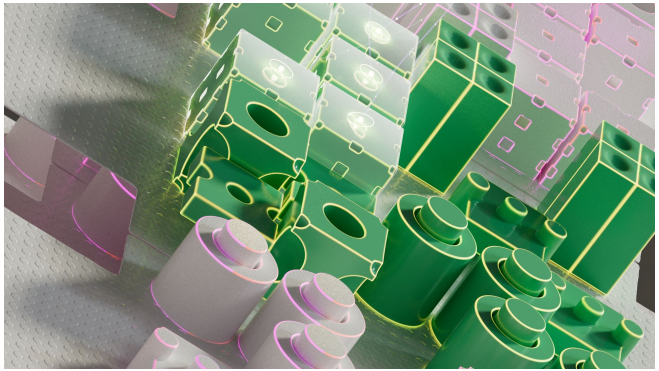
Causal Experimental Design

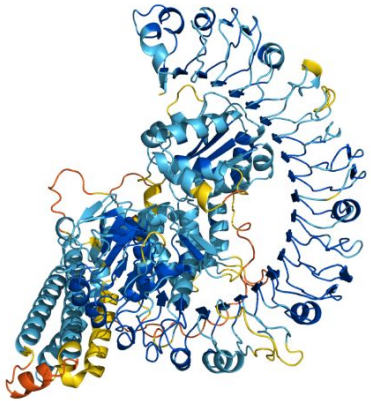
Stefan Bauer

st.bauer@tum.de

Success in AI

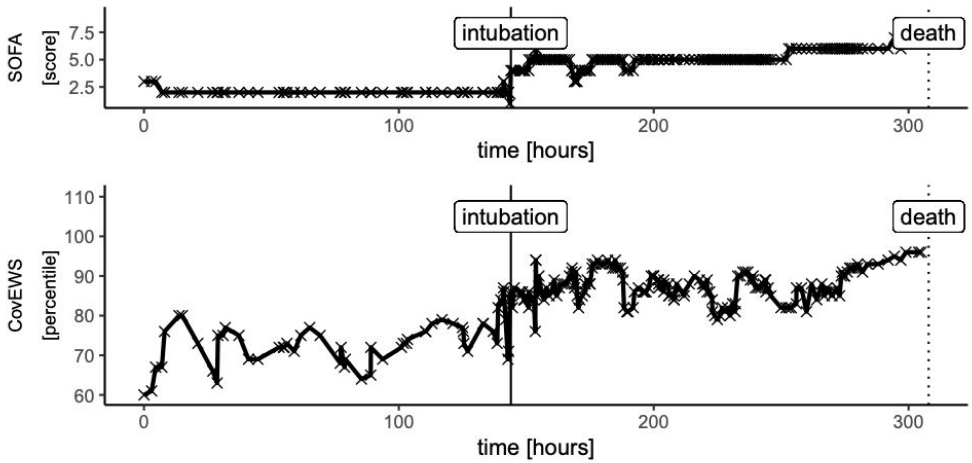


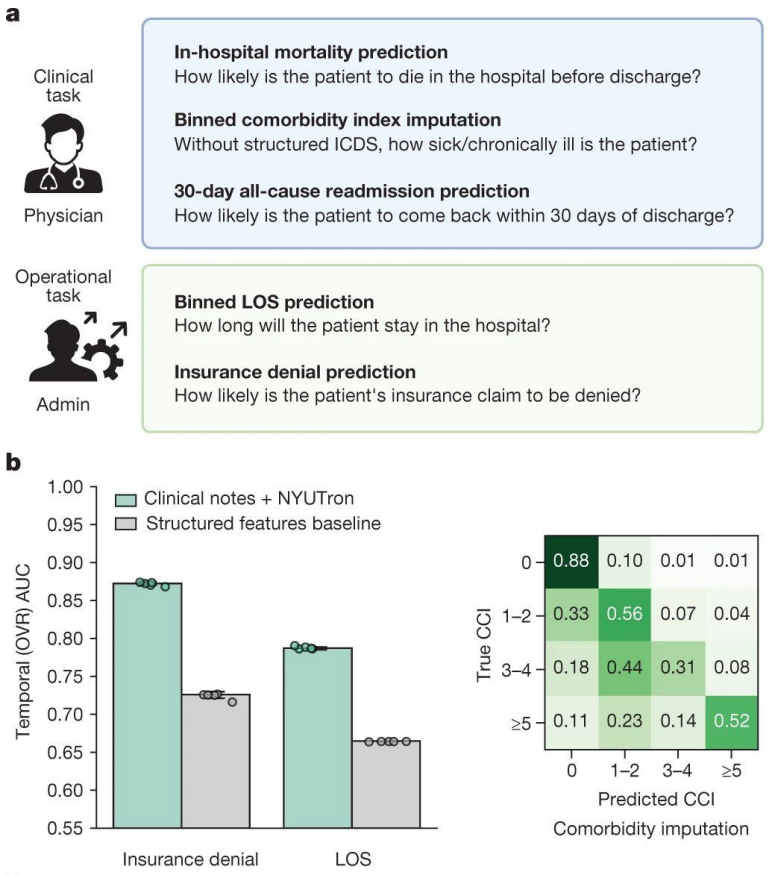




Clinically Applicable AI – Early Warning Systems

Continuous warning systems, to identify patients at risk early and reduce mortality, morbidity, length of stay, …



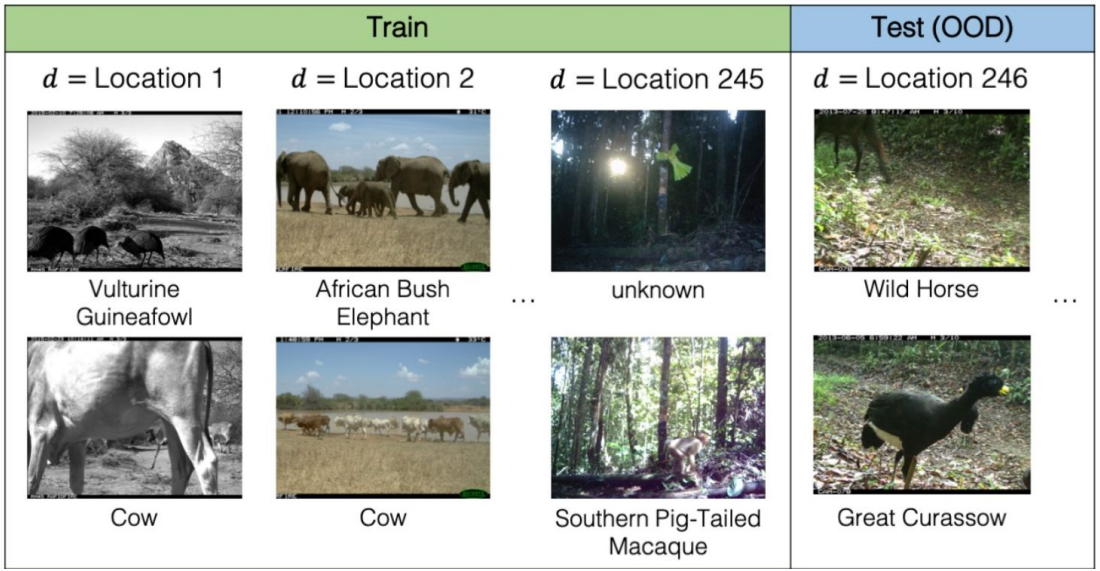


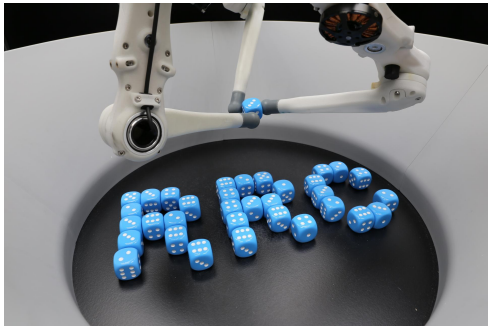
Henry, Katharine E., et al. "A targeted real-time early warning score (TREWScore) for septic shock." *Science translational medicine* (2015) P. Schwab, A. Mehrjou, S. Prabhoo, L. A. Celi, J. Hetzel, M. Hofer, B. Schölkopf, **S. Bauer** "Real-time Prediction of COVID-19 related Mortality using Electronic Health Records" Nature Communications (2020).

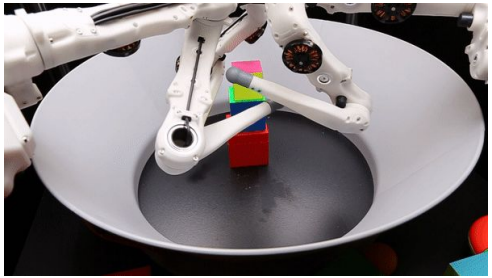
Jiang, L.Y., Liu, X.C., Nejatian, N.P. *et al.* Health system-scale language models are all-purpose prediction engines. *Nature* (2023).

Intelligent Machines that Generalize

iWildCam







https://real-robot-challenge.com/

Koh, Pang Wei, et al. "Wilds: A benchmark of in-the-wild distribution shifts." *International Conference on Machine Learning*. 2021.
Ahmed, Ossama, et al. "Causalworld: A robotic manipulation benchmark for causal structure and transfer learning." *arXiv preprint arXiv:2010.04296* (2020).
Bauer, Stefan, et al. "Real Robot Challenge: A Robotics Competition in the Cloud." *NeurIPS 2021 Competitions and Demonstrations Track*. PMLR, 2022.

Causal Structure Learning - A Quick Primer

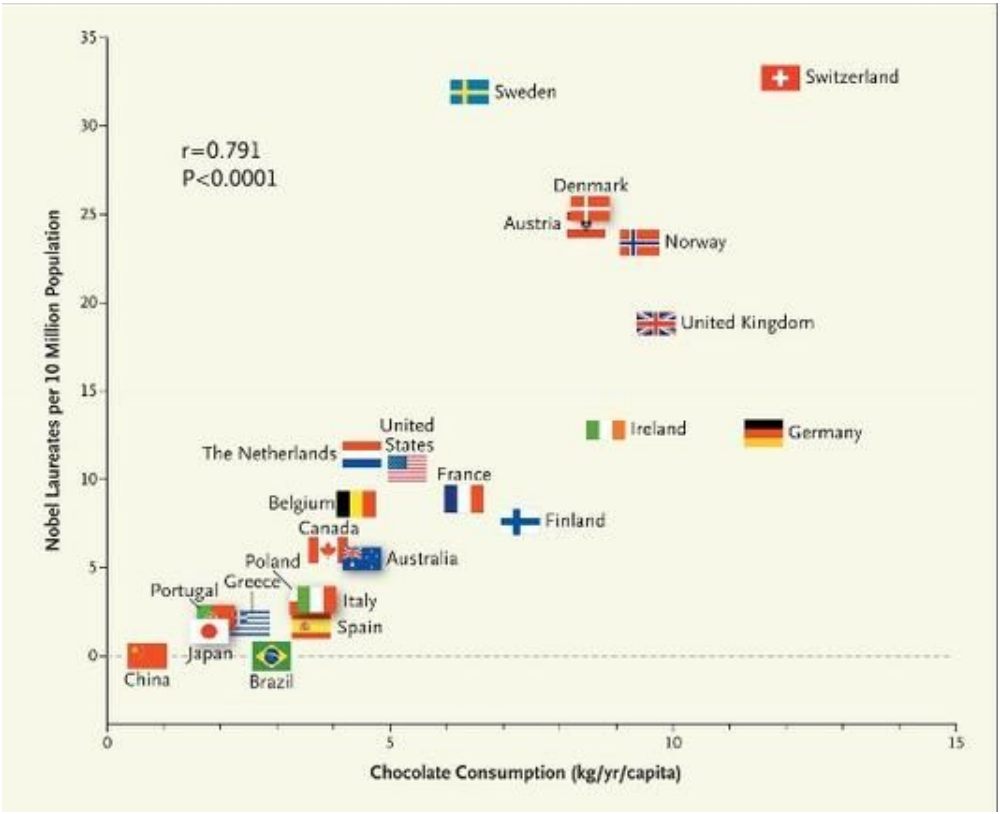


Image from: Messerli, F. H., et al. "Chocolate and Your Health." *N Engl J Med* 367.16 (2012): 1562-4.

Causal Structure Learning - A Quick Primer in 2 slides!

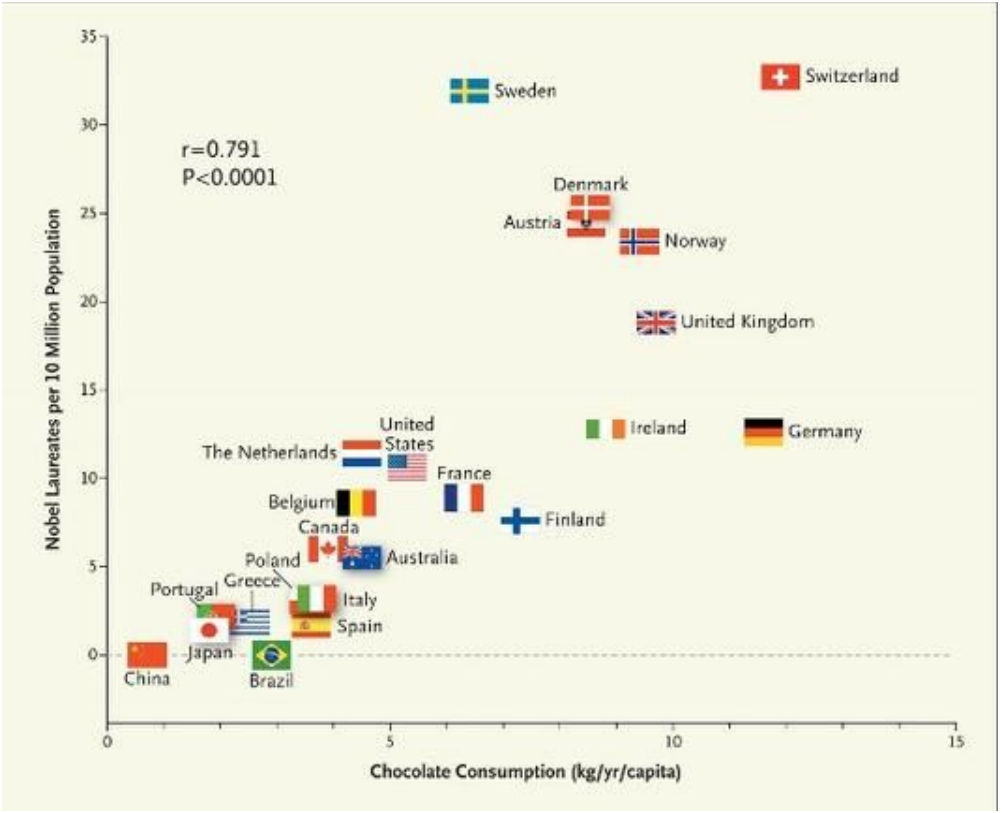
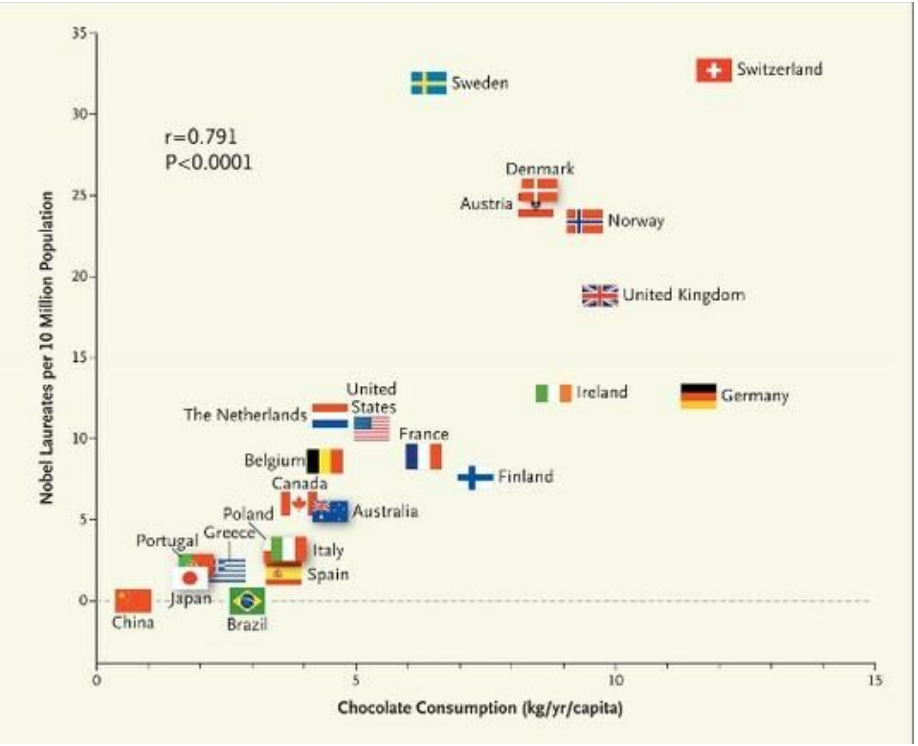


Image from: Messerli, F. H., et al. "Chocolate and Your Health." *N Engl J Med* 367.16 (2012): 1562-4.

- In low dimensions we can often differentiate between correlation and causation using domain knowledge
- Sweden seems to be an outlier.
- Swiss chocolate seems to be better German.

Causal question: What is the best prediction for #Nobel L. given chocolate consumption is set to 100?

Causal Structure Learning - A Quick Primer



Messerli F.H., et al. "Chocolate and Your Health." *N Engl J Med* (2012)

Golomb B.A., et al. “Chocolate habits of Nobel prizewinners.” *Nature* (2013)

S

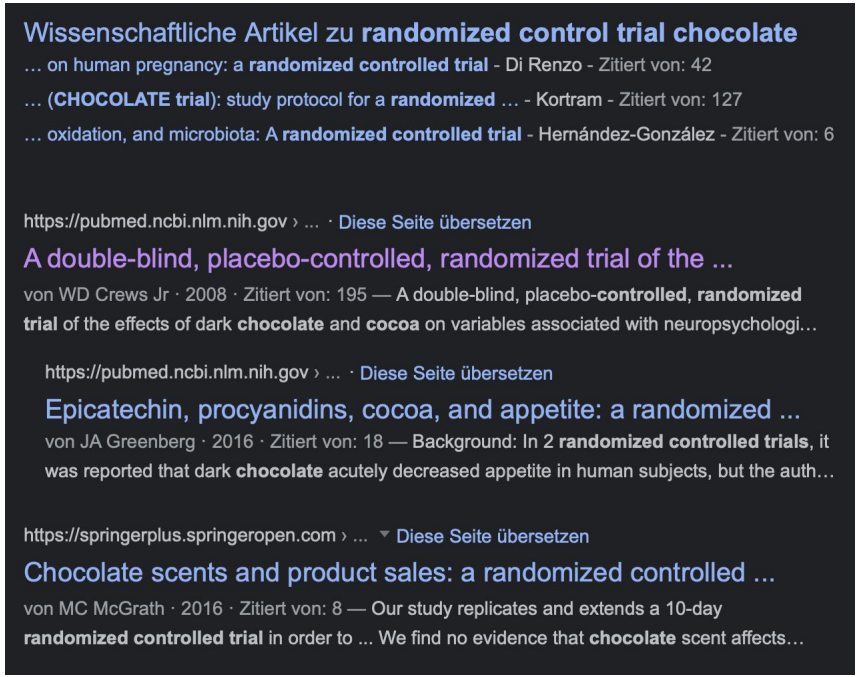
Does eating chocolate increase my chances of winning a Noble Prize?

There is no scientific evidence to suggest that eating chocolate can increase your chances of winning a Nobel Prize. Winning a Nobel Prize requires a significant contribution to a specific field of study, such as physics, chemistry, medicine, literature, or peace. While chocolate may have some health benefits when consumed in moderation, it is unlikely to have a direct impact on one's ability to win a Nobel Prize.

Nobel Prizes are typically awarded to individuals who have made outstanding contributions to their field through groundbreaking discoveries or significant advancements in knowledge. These achievements usually require years of dedicated research and a deep understanding of the subject matter. While chocolate may be a tasty treat, it is unlikely to have a significant impact on the complex work required to win a Nobel Prize.

Therefore, it is important to focus on developing your skills and knowledge in your chosen field if you aspire to win a Nobel Prize, rather than relying on any food or beverage to increase your chances of success.

https://chat.openai.com



Randomized Controlled Trial > Am J Clin Nutr. 2008 Apr;87(4):872-80.
doi: 10.1093/ajcn/87.4.872.

A double-blind, placebo-controlled, randomized trial of the effects of dark chocolate and cocoa on variables associated with neuropsychological functioning and cardiovascular health: clinical findings from a sample of healthy, cognitively intact older adults

Hariton, Eduardo, and Joseph J. Locascio. "Randomised controlled trials—the gold standard for effectiveness research." *BJOG: an international journal of obstetrics and gynaecology* (2018).

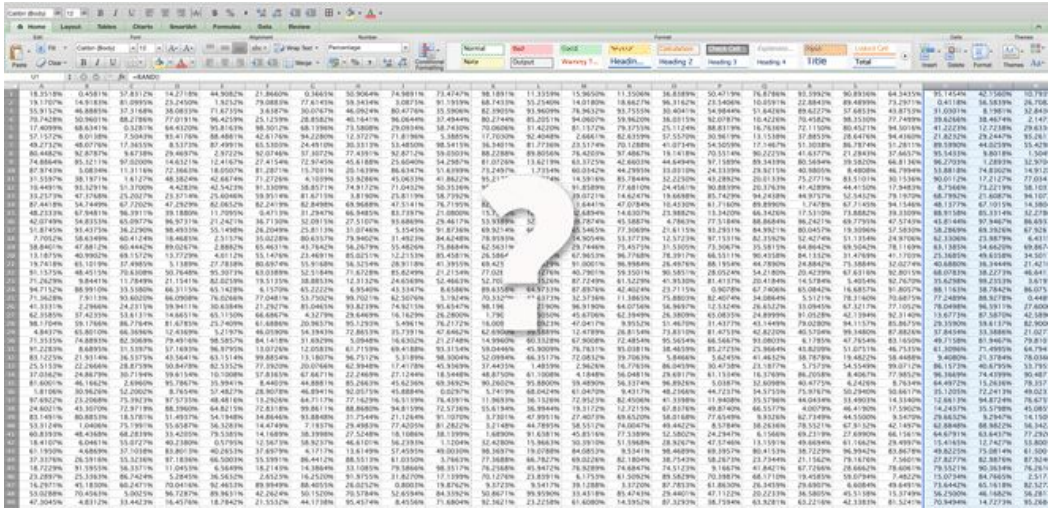
Prediction vs Causation

You have a big data scenario where you measure thousands of variables X.

Examples:

Predictive ML: Use X to predict Y e.g. *how well can we predict if customers churn / a patient needs to be transferred to ICU?*

Causal ML: What is the effect of a particular variable D (one element of X) on Y? *If we perform intervention on D, would the customer not churn / be able to leave the ICU?*



What happens if additionally we have more variables than observations? What if we know there are more variables than we can measure? What happens if we are interested in Z or T and not D and Y? …

Causal Structure Learning - A Quick Primer in 2 slides!

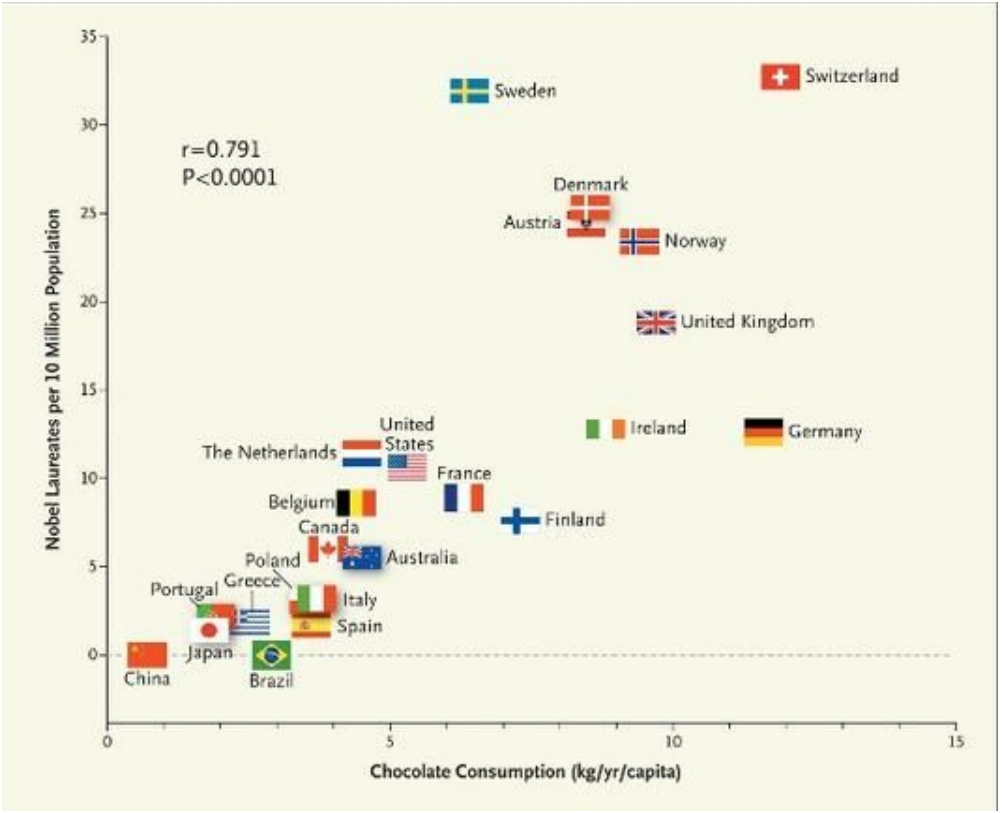
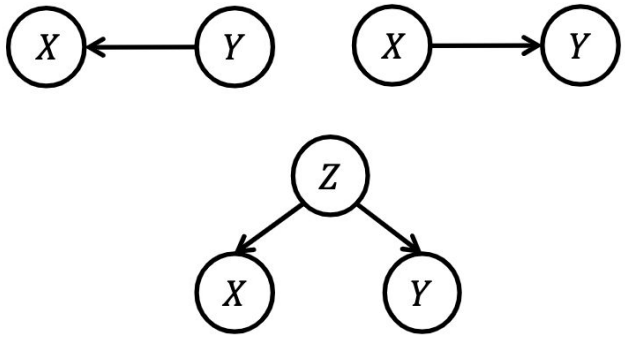


Image from: Messerli, F. H., et al. "Chocolate and Your Health." *N Engl J Med* 367.16 (2012): 1562-4.

Key problem - Many SCMs generate same distribution:



How we can learn the underlying causal model from data if we do not know it a priori - at scale and with ***few interventions***!

Causal Structure Learning

Goal: Learn a causal graph from data

Approaches:

- Constraint based
- Score based
- Restricted SCM

How to search efficiently?

# of Nodes	# of DAGs	# of Graphs
1	1	1
2	3	2^2
3	25	2^6
4	543	2^{12}
5	29281	2^{20}
6	3781503	2^{30}
7	1138779265	2^{42}
8	783702329343	2^{56}
9	1213442454842881	2^{72}
10	4175098976430598143	2^{90}

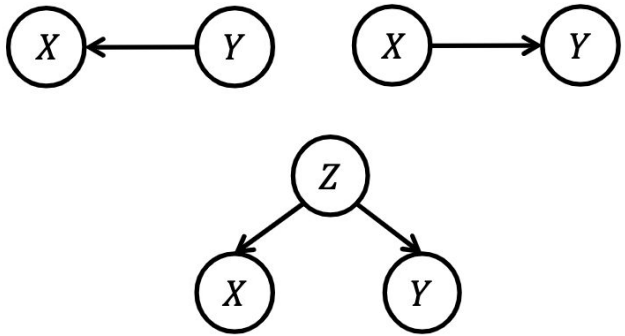
"I would rather discover one causal law than be King of Persia."
- Democritus, ~400 BC.

Image: https://en.wikipedia.org/wiki/Democritus

https://icml.cc/virtual/2022/tutorial/18442

What to do with a DAG?

- **Cause-Effect Estimation**
- **Imputation** (of missing but not randomly missing data)
- **Experimental Design**
  - Where we intervene matters!
  - Interventions are expensive
  - Some interventions might be impossible or unethical.

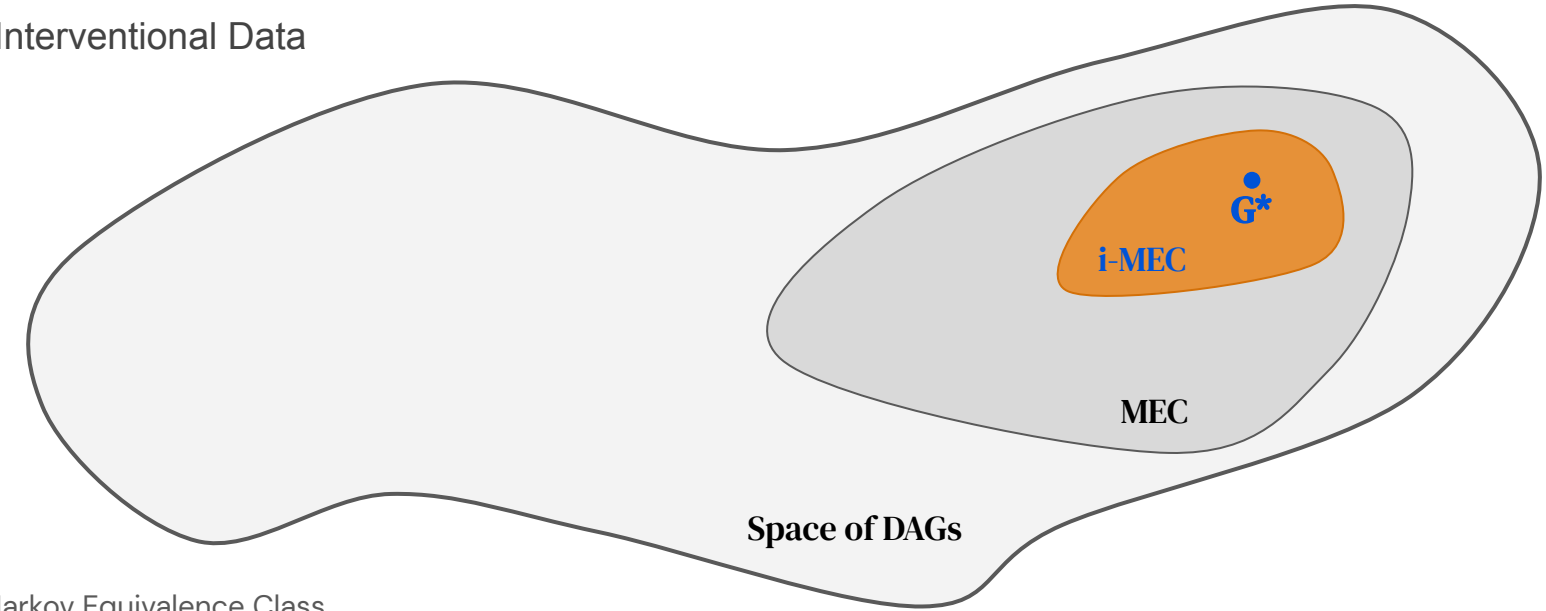


How we can learn the underlying causal model from data if we do not know it a priori - at scale and with ***few interventions***!

Geffner, Tomas, et al. "Deep end-to-end causal inference." *arXiv preprint arXiv:2202.02195* (2022).
Sharma, Amit, and Emre Kiciman. "DoWhy: An end-to-end library for causal inference." *arXiv preprint arXiv:2011.04216* (2020).

Causal Structure Learning

1. Restrictive Assumptions on the functional nature of the data
2. Interventional Data



MEC = Markov Equivalence Class
i-MEC = Interventional Markov Equivalence Class
G* = Ground Truth Graph

Jamie Robbins: “before you can pull a rabbit out of a hat, you have to put the rabbit in.”*

The ultimate goal of our collaboration was to develop discovery algorithms much more powerful than the FCI algorithm, the then and now current state of the art. The FCI algorithm is based on the assumption that observed data were generated by an unknown causal graph (that may include unmeasured (i.e. hidden) common causes of the observed variables) that satisfies the following faithfulness assumption: all conditional and unconditional independencies in the observed data joint distribution are due to causal structure, i.e. to missing arrows on the underlying causal graph.

We have gone from complete ignorance of the gene network to complete knowledge seemingly by magic. How is this possible? At this point in my talk, having pulled the Nobel rabbit out of the hat, I tell the audience I will explain the trick and show them how I used the generalized faithfulness assumption to place the rabbit in the hat.

*Recent Interviews with Heckman, Rubin, Pearl and Robbins in Journal of Observational Studies https://muse.jhu.edu/issue/48885

Scheines, Richard. "An introduction to causal inference." (1997).
Glymour, Clark, Kun Zhang, and Peter Spirtes. "Review of causal discovery methods based on graphical models." *Frontiers in genetics* (2019).

Bayesian Learning of Causal Graphs

- In order to do very few experiments, we need to plan interventions and select the *most informative ones*.
- **One** possible / popular answer: Get uncertainty quantification over causal structures learned from purely observational data.
- Then use this to efficiently design (fewer) experiments.
- Requires Bayesian Learning of causal graphs.

Bayesian Learning of Causal Graphs is a much harder problem (than just inferring one graph)!

Bayesian Learning of Causal Graphs

DAG - Bootstrap

- Frequentist
- Bootstrap samples and obtain graphs from any causal discovery alg.
- Estimate is normalised frequency count of bootstrapped graphs
- Does not have full support

N. Friedman, M. Goldszmidt, and A. J. Wyner. Data analysis with Bayesian networks: A bootstrap approach. *UAI* 1999.

MCMC

- Assume conjugate prior over parameters
- Obtain unnormalized posterior by marginalising over parameters
- Can sample from the posterior using MCMC
- Need graph search heuristics
- Slow mixing and convergence

D. Heckerman, C. Meek, and G. Cooper. A Bayesian approach to causal discovery. *Technical report,* 1997

Variational

$$\log p(\mathcal{D}) \geq \mathcal{L}(\psi \in \Psi) = \mathop{\mathbb{E}}_{q_\psi(\boldsymbol{\phi})} [\log p(\mathcal{D} \mid \boldsymbol{\phi})] - \mathrm{D}_{\mathrm{KL}}(q_\psi(\boldsymbol{\phi}) || p(\boldsymbol{\phi}))$$

- Penalty that promotes sparsity of the graphs (simpler hypotheses) but also DAGness
- How to define var. approximation

Annadani et al. "Variational causal networks: Approximate bayesian inference over causal structures." *arXiv* (2021).

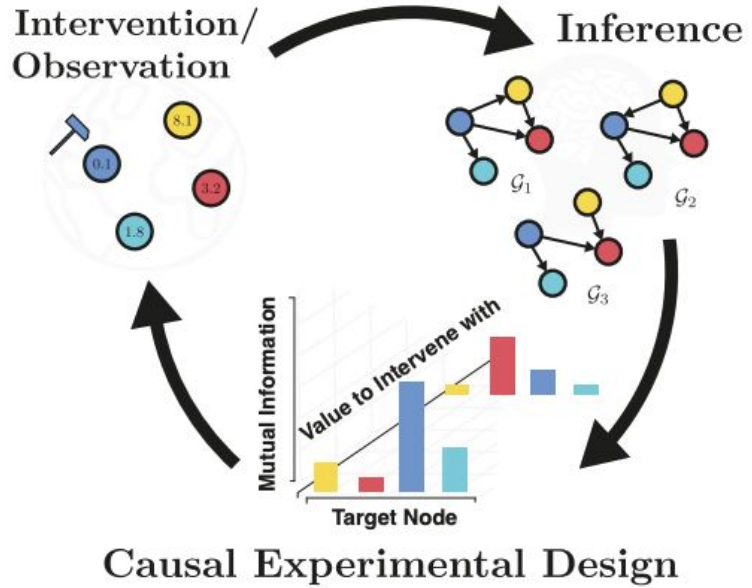
Lorch, et al. "Dibs: Differentiable bayesian structure learning." *NeurIPS* (2021)

Bayesian Optimal Experimental Design for Causal Discovery

Bayesian Optimal Experimental Design framework to identify where to intervene to quickly identify the causal structure

$$\arg\max_{j,v} \{\mathrm{I}(\mathbf{Y}; \mathbf{\Phi} \mid \{(j, v)\}, \mathcal{D})\}$$

outcomes latent causes experiment

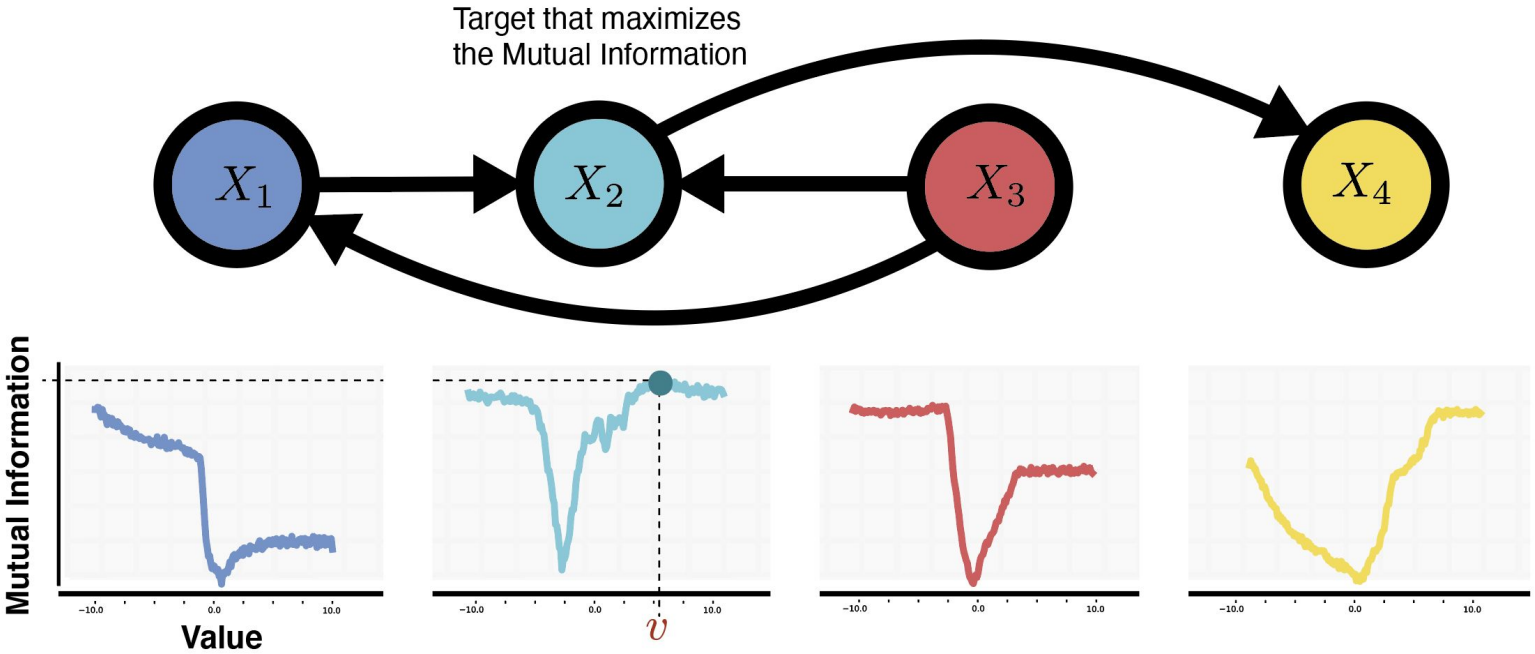


Problem: How to solve the optimization problem and select intervention node (j, discrete) and value (v, continuous)?

Dennis V Lindley. On a measure of the information provided by an experiment. The Annals of Mathematical Statistics, pages 986–1005, 1956.
Chaloner, Kathryn, and Isabella Verdinelli. "Bayesian experimental design: A review." *Statistical science* (1995).

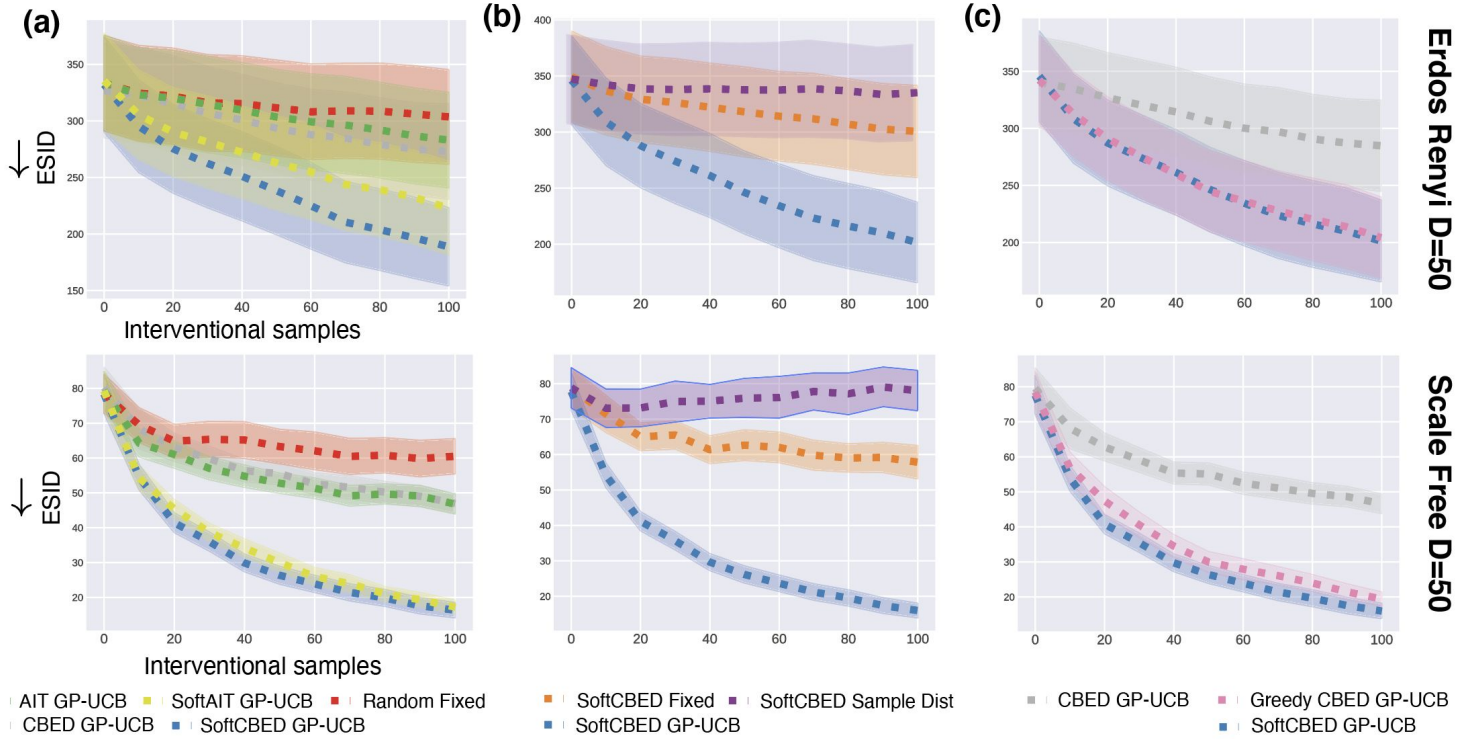
Bayesian Optimal Experimental Design

$$\mathrm{U}_{\mathrm{BOED}}(\xi) \triangleq \mathrm{I}(\mathbf{Y}; \theta \mid \xi, \mathcal{D}) = \mathop{\mathbb{E}}_{p(\mathbf{y}|\theta,\xi)p(\theta|\mathcal{D})} [\log p(\mathbf{y} \mid \xi, \mathcal{D}) - \log p(\mathbf{y} \mid \theta, \xi, \mathcal{D})]$$

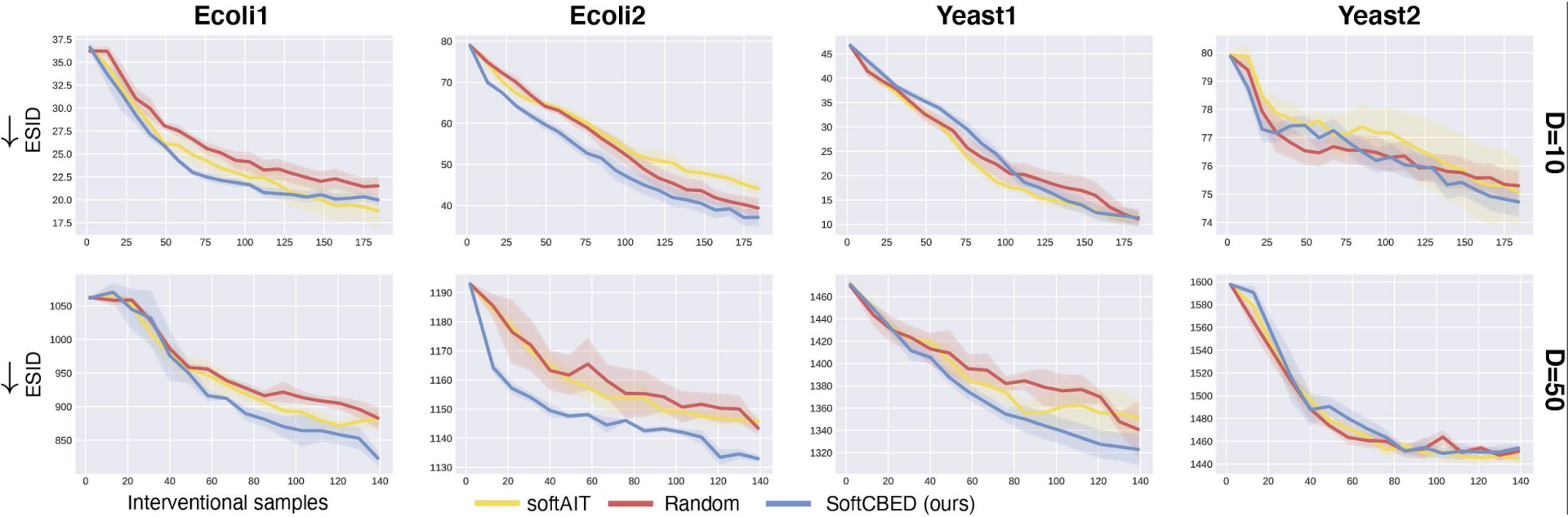


Tigas, Panagiotis, Annadani, Yashas et al. "Interventions, where and how? experimental design for causal models at scale." *NeurIPS* (2022)
Houlsby, Neil, et al. "Bayesian active learning for classification and preference learning." *arXiv preprint arXiv:1112.5745* (2011).

Results on synthetic graphs



Results on In Silico dataset DREAM



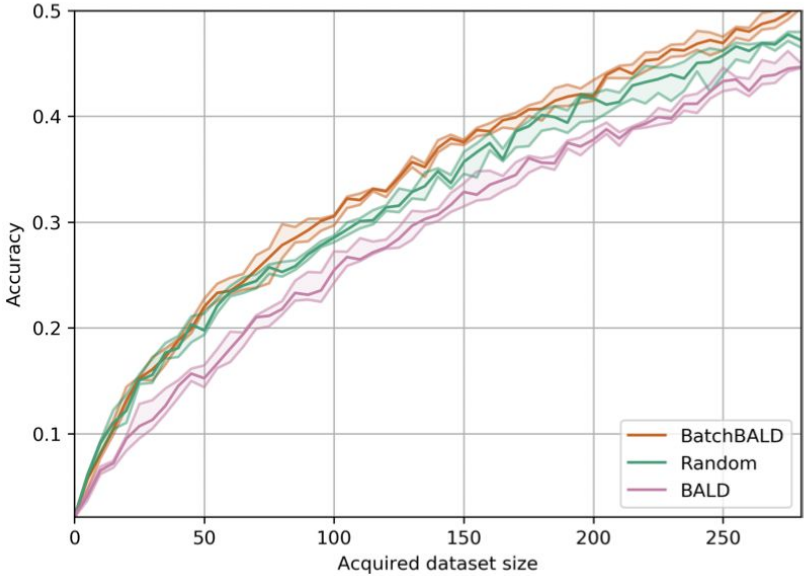
GeneDisco - A Benchmark & Community Challenge for Experimental Design in Drug Discovery

Task 1 - Maximize target discovery rate: Selecting gene targets for interventional experiments in order to maximize the discovery rate of interesting targets (“target discovery rate”).

Task 2 – Maximize model performance: Selecting gene targets for interventional experiments in order to maximize the performance of a machine-learning model trained on the data selected.

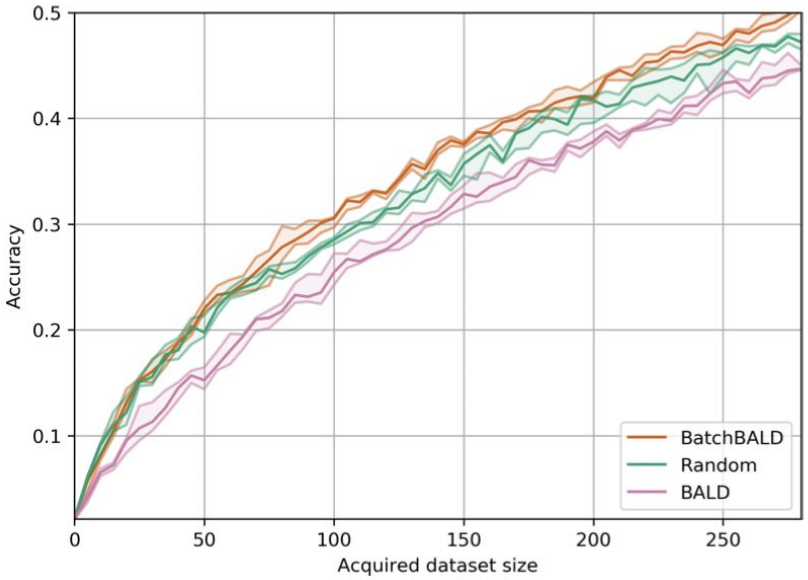
A. Mehrjou, A. Soleymani, P. Notin, A. Jesson, Y. Gal, S. Bauer and P. Schwab: „GeneDisco: A Real World Experimental Design Benchmark for Batch Active Learning for Drug Discovery“. International Conference on Learning Representations (2022).

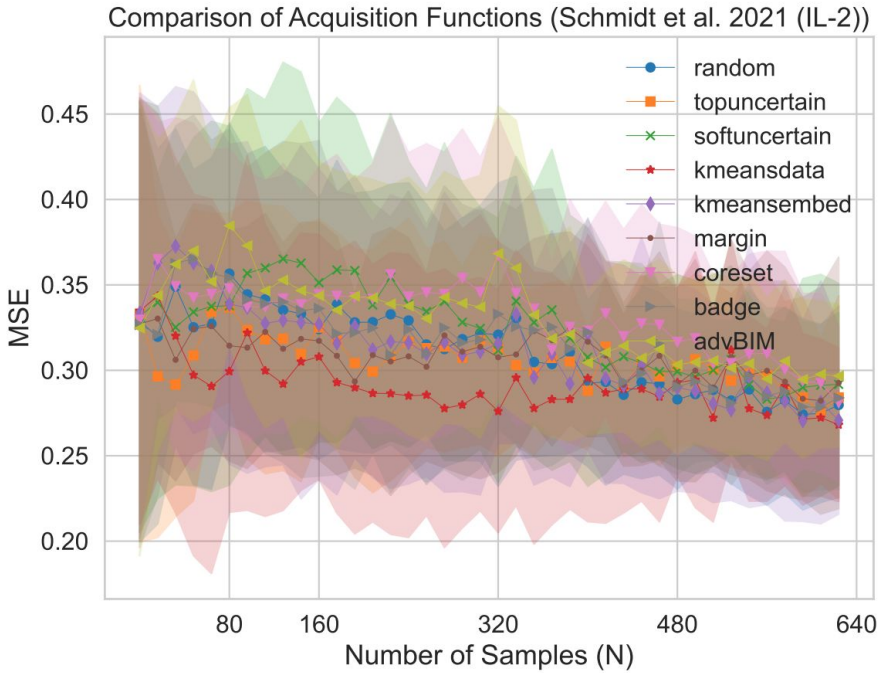
Design for Batched Experiments (Clean Data)



Houlsby, Neil, et al. "Bayesian active learning for classification and preference learning." *arXiv preprint arXiv:1112.5745* (2011).
Kirsch, A. et al. "Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning." *NeurIPS (*2019).
Mehrjou, A. et al. „GeneDisco: A Real World Experimental Design Benchmark for Batch Active Learning for Drug Discovery“. ICLR (2022).

Design for Batched Experiments (Real World)



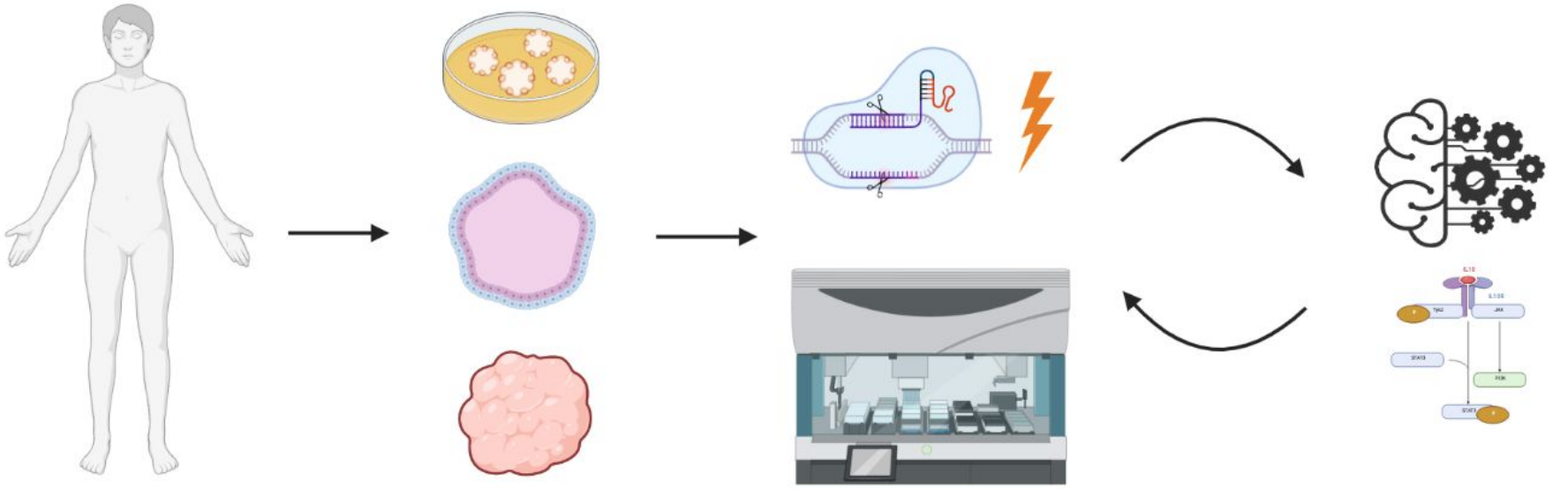


Houlsby, Neil, et al. "Bayesian active learning for classification and preference learning." *arXiv preprint arXiv:1112.5745* (2011).

Kirsch, A. et al. "Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning." *NeurIPS (*2019).

Mehrjou, A. et al. „GeneDisco: A Real World Experimental Design Benchmark for Batch Active Learning for Drug Discovery“. ICLR (2022).

Potential Future: Interactive Learning for Personalized Therapies



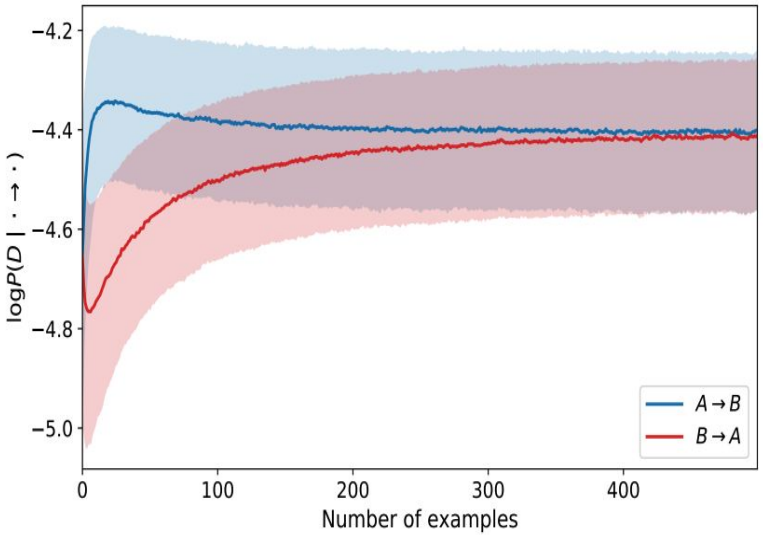
Ji, Yuge, et al. "Machine learning for perturbational single-cell omics." *Cell Systems* (2021).
Kornauth, Christoph, et al. "Functional Precision Medicine Provides Clinical Benefit in Advanced Aggressive Hematological Cancers and Identifies Exceptional Responders." *Cancer discovery* (2021).

Causality, Deep Learning and Generalization

$$P_{A\to B}(A,B) = P_{A\to B}(A)P_{A\to B}(B \mid A)$$
$$P_{B\to A}(A,B) = P_{B\to A}(B)P_{B\to A}(A \mid B)$$

Incorrect Knowledge Factorization leads to poor Transfer!

- **Right** factorization of joint distribution leads to fast adaptation to new problem
- With the **wrong** factorization a change in P(cause) influences all modules and all modules need to be adapted!

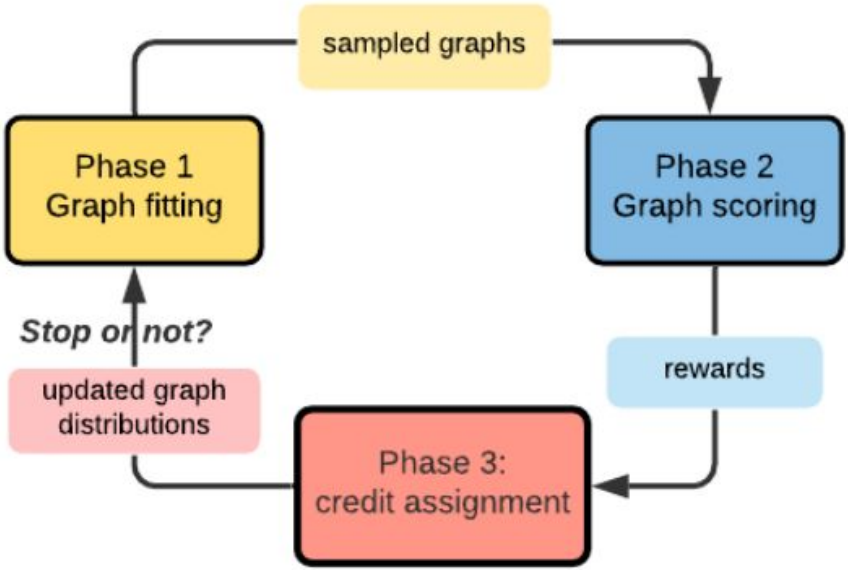


Idea: Use the speed of adaptation as a learning signal to identify the causal model.

Bengio et al. “A Meta-Transfer Objective for Learning to Disentangled Causal Mechanisms”, 2019.
Scherrer et al. "Deep Learning for Causality" *upcoming* , basis for https://icml.cc/virtual/2022/tutorial/18442

Neural Causal Models

- Use **adaptation rate** as score
- Iteratively **Functional** and **Structural** Fitting
  - Functional parameters → Conditional Relationships
  - Structural parameters → (Soft) Adjacency Matrix
- Use **neural networks** to learn causal relationships
  - Continuous optimization
  - Don't need to iterate through all DAG



R. Ke, O. Bilaniuk, A. Goyal, **S. Bauer**, H. Larochelle, C. Pal und Y. Bengio. „Learning Neural Causal Models from Unknown Interventions", arxiv (2020).
Bengio et al. "A Meta-Transfer Objective for Learning to Disentangled Causal Mechanisms", 2019.

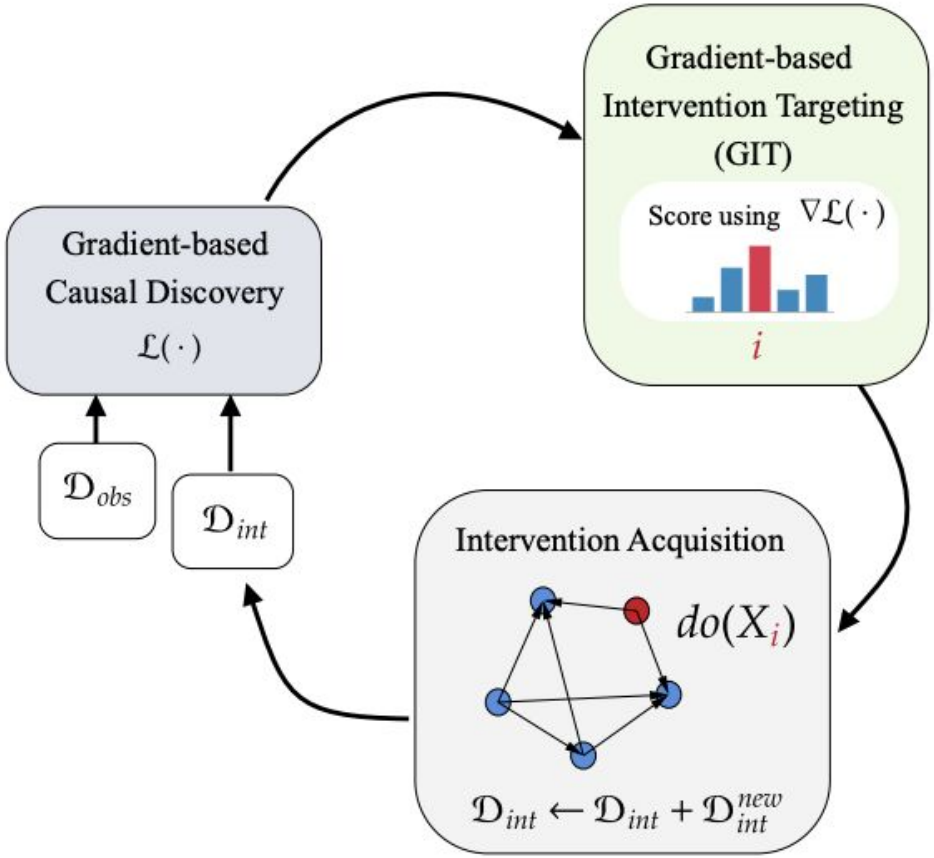
Upcoming Review of Neural Causal Models from Nino Scherrer, underlying basis for https://icml.cc/virtual/2022/tutorial/18442

Neural Causal Models for Experimental Design

- Drop the *(optimal)* mutual information scores

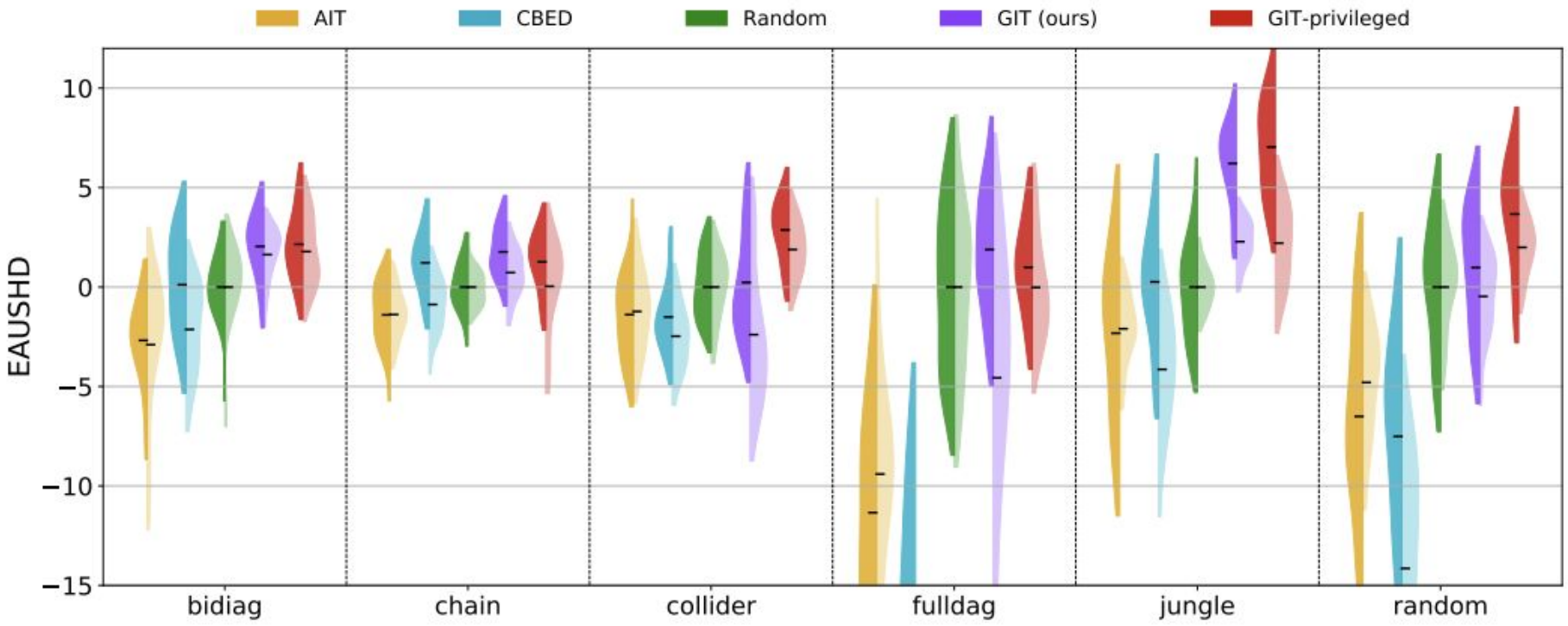
$$\arg\max_{j,v} \{\mathrm{I}(\mathbf{Y}; \mathbf{\Phi} \mid \{(j, v)\}, \mathcal{D})\}$$

- Learn the intervention targets using gradient information alone!
- Multiple different gradients possible depending on underlying NCM.



Olko, Mateusz, et al. "Trust your∇: Gradient-based intervention targeting for causal discovery." *arXiv* (2022).

Simplest implementation of *none-optimal* approach already competitive!



Olko, Mateusz, et al. "Trust your∇: Gradient-based intervention targeting for causal discovery." *arXiv* (2022).

Challenge - What to do with multi-modal data?

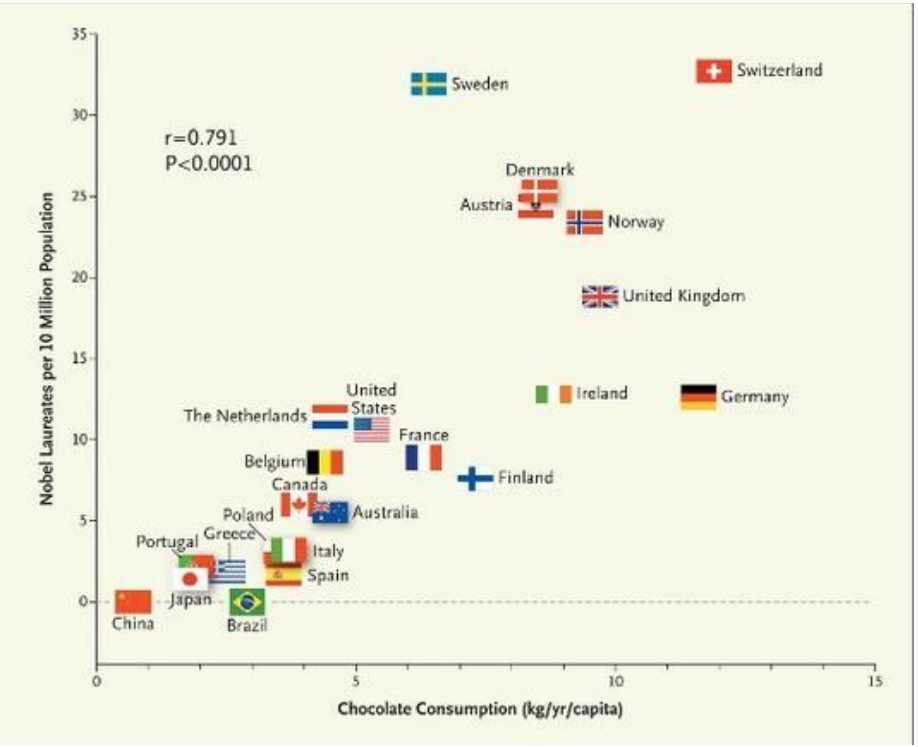
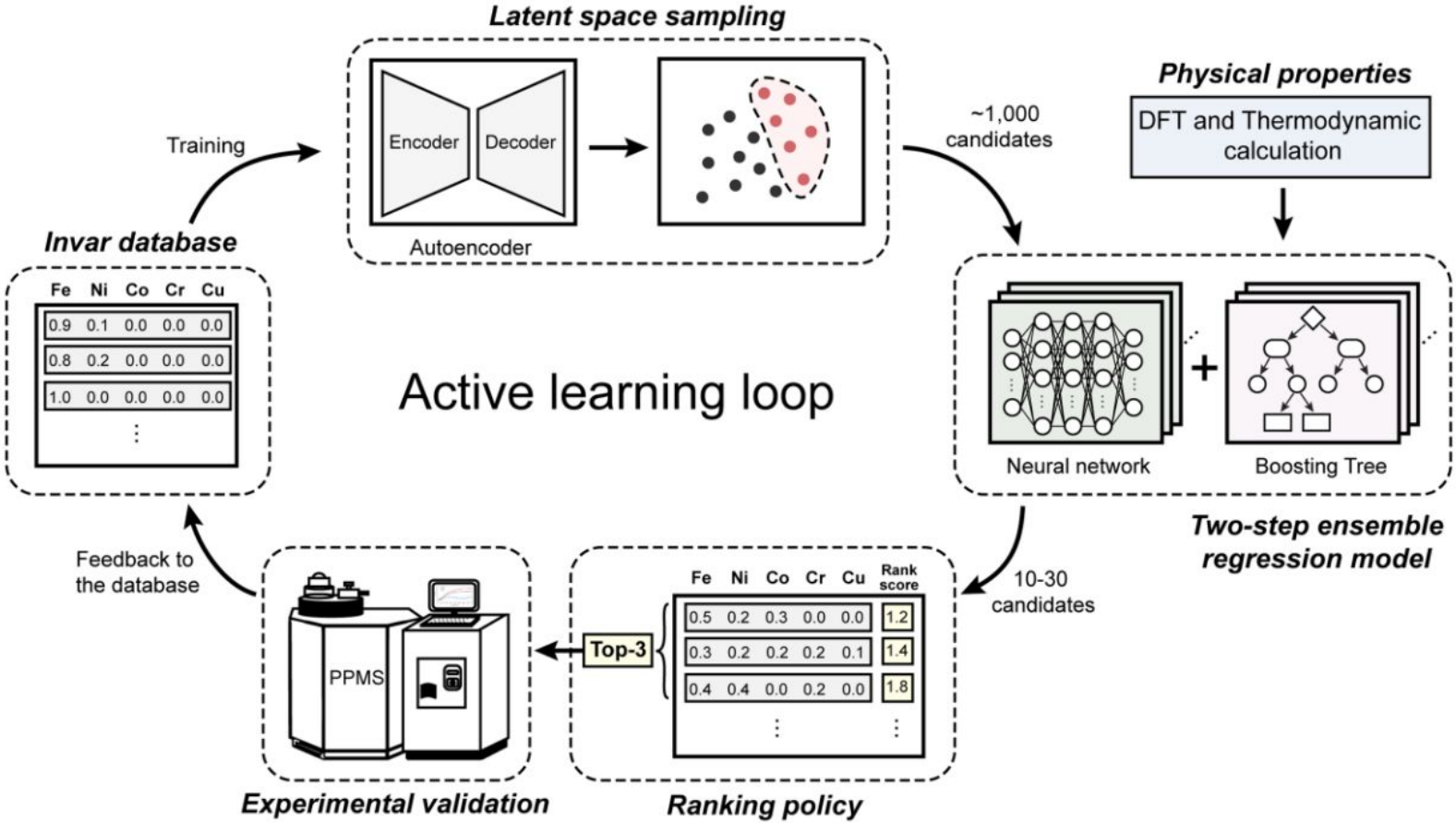


Image from: Messerli, F. H., et al. "Chocolate and Your Health." *N Engl J Med* (2012).

https://twitter.com/ManuelKRausch1/status/1518268003258875904 from David Hunter @cyclingmole

Structured Spaces for Materials Discovery - Invar Alloys



Z. Rao, P-Y. Tung, R. Xie, Y. Wei, H. Zhang, A. Ferrari, T. Klaver, F. Körmann, P. Sukumar, A. Kwiatkowski da Silva, Y. Chen, Z. Li, D. Ponge, J. Neugebauer, O. Gutfleisch, **S. Bauer** and D. Raabe. „Machine learning enabled fast high-entropy alloy discovery - a case study on novel INVAR alloys“. Science (2022).

Summary

- Full causal discovery from observational data is too difficult or requires very strong (unreasonable?) assumptions.
- Still hard to evaluate learned causal graphs especially in real-world.
- Need to evaluate causality wrt. downstream tasks e.g. experimental design (rather than just graph recovery metrics).
- Need inductive biases (e.g. simulations) or interventions for multi-modal data.

st.bauer@tum.de

Thank you!

- Patrick Schwab
- Arash Mehrjou

- **Piotr Milos**
- **Łukasz Kuciński**
- **Mateusz Olko**
- **Michał Zając**
- **Aleksandra Nowak**

- **Yashas Annadani**
- **Nino Scherrer**

- Yarin Gal
- Pascal Notin
- Andrew Jesson
- Desi R. Ivanova
- **Panagiotis Tigas**
- **Adam Foster**

A. Mehrjou, A. Soleymani, P. Notin, A. Jesson, Y. Gal, S. Bauer and P. Schwab „GeneDisco: A Real World Experimental Design Benchmark for Batch Active Learning for Drug Discovery“. ICLR (2022).

N. Scherrer, O. Bilaniuk, Y. Annadani, A. Goyal, P. Schwab, B. Schölkopf, MC. Mozer, Y. Bengio, S. Bauer and R. Ke. „Learning Neural Causal Models with Active Interventions“, arXiv (2021).

P. Tigas, Y. Annadani, A. Jesson, B. Schölkopf, Y. Gal und S. Bauer. „Interventions, Where and How? Experimental Design for Causal Models at Scale“. NeurIPS (2022).

O. Mateusz, M. Zając, A. Nowak, N. Scherrer, Y. Annadani, S. Bauer, L. Kucinski, P. Milos. "Trust your∇: Gradient-based intervention targeting for causal discovery." *arXiv (*2022).

Y. Annadani, J. Rothfuss, A. Lacoste, N. Scherrer, A. Goyal, Y. Bengio und S. Bauer. „Variational Causal Networks: Approximate Bayesian Inference over Causal Structures“, arXiv (2021).

T. Deleu, A. Góis, C. Emezue, M. Rankawat, S. Lacoste-Julien, S. Bauer and Y. Bengio. „Bayesian Structure Learning with Generative Flow Networks“. UAI (2022).

http://www.disentanglement-challenge.com/
https://real-robot-challenge.com/
https://github.com/rr-learning/
http://mlss.tuebingen.mpg.de/2020/index.html

CIFAR HELMHOLTZ MUNICH