

The Transformative Impact of Modern Generative AI and the Role of AutoML in AI's Advancement

Prof. Dr. Marius Lindauer



AI @ L3S: Excellent research in Europe, innovation in Niedersachsen



Research: Intelligent, reliable and responsible systems

- 4 European Research Training Groups (ITNs NoBIAS, KnowGraphs, ...)
- 8 ERC grants in the last 10 years (AutoML, ScienceGraph, Cont4Med, ...)
- 25+ L3S members, mainly from Hannover and Braunschweig
- 200+ postdocs and PhD students

Innovation and research in interdisciplinary groups

- Intelligent production
- Intelligent mobility
- Personalized medicine
- Digital education
- as well as biodiversity, media startups, quantum computing ...

- 15 million € annual budget (of which 2 million € basic funding)
- thus the largest AI research center in Lower Saxony

The L3S Research Center



Excellent Research in Europe, AI-driven Innovation for Lower Saxony

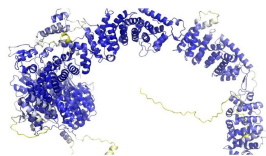


AI as Innovation Factor

AI wins Nobel Prizes!



Nobelpreis Chemie 2024



Nobelpreis Physik 2024

© Illustration: Niklas Elmehed © Nobel Prize Outreach 2024 (Ausschnitt)

Schreibe mir eine Webanwendung, die die Sparkasse Hannover als Landing Page für ihre Kunden nutzen kann



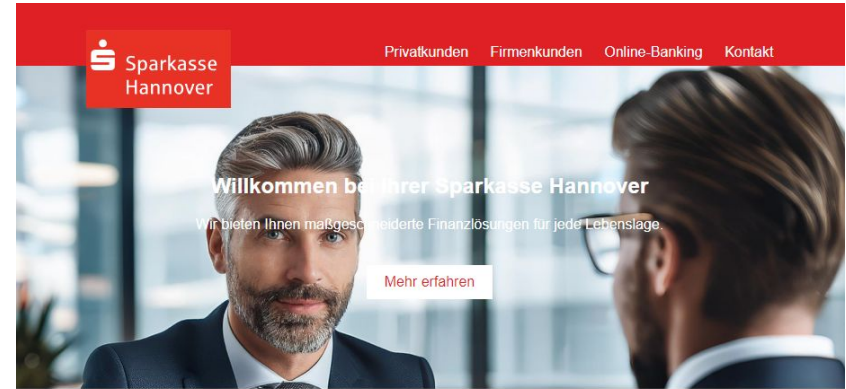
Gedanken zu 33 Sekunden ▾

Natürlich! Hier ist ein einfaches Beispiel für eine Landing Page für die Sparkasse Hannover. Bitte beachten Sie, dass dieses Beispiel grundlegend ist und für den Produktionseinsatz weitere Sicherheits- und Funktionsanpassungen erforderlich sind.

index.html

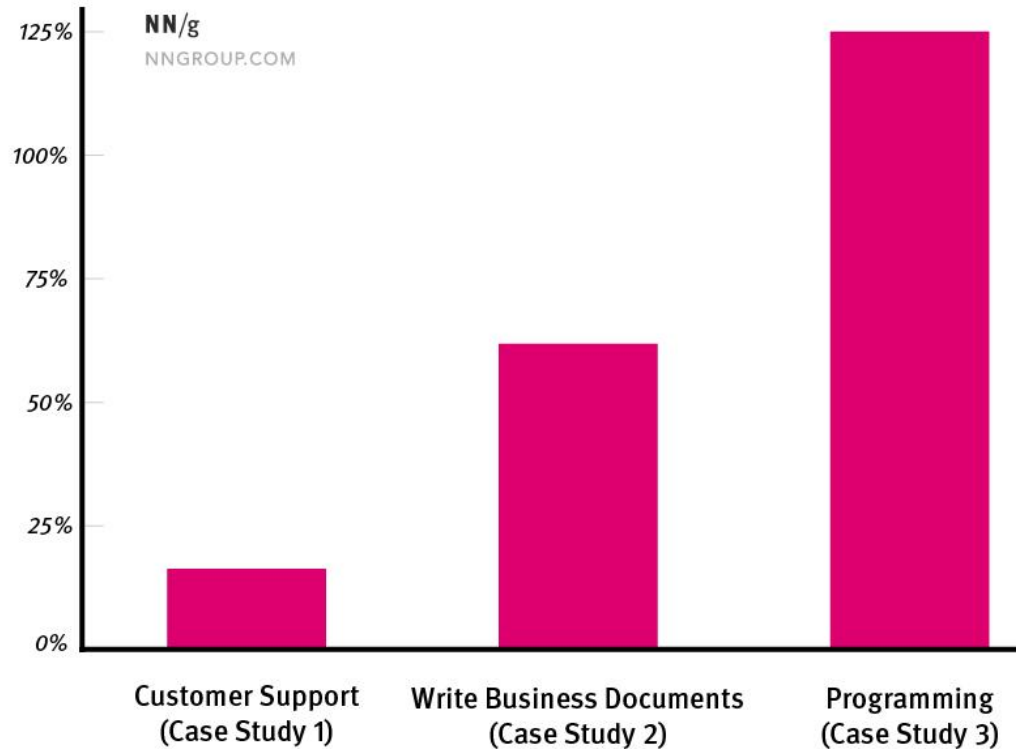
```
html Code kopieren

<!DOCTYPE html>
<html lang="de">
<head>
  <meta charset="UTF-8">
  <title>Sparkasse Hannover</title>
  <link rel="stylesheet" href="styles.css">
</head>
<body>
  <header>
    <div class="container">
      
      <nav>
        <ul>
          <li><a href="#">Privatkunden</a></li>
          <li><a href="#">Firmenkunden</a></li>
        </ul>
      </nav>
    </div>
  </header>
</body>
</html>
```



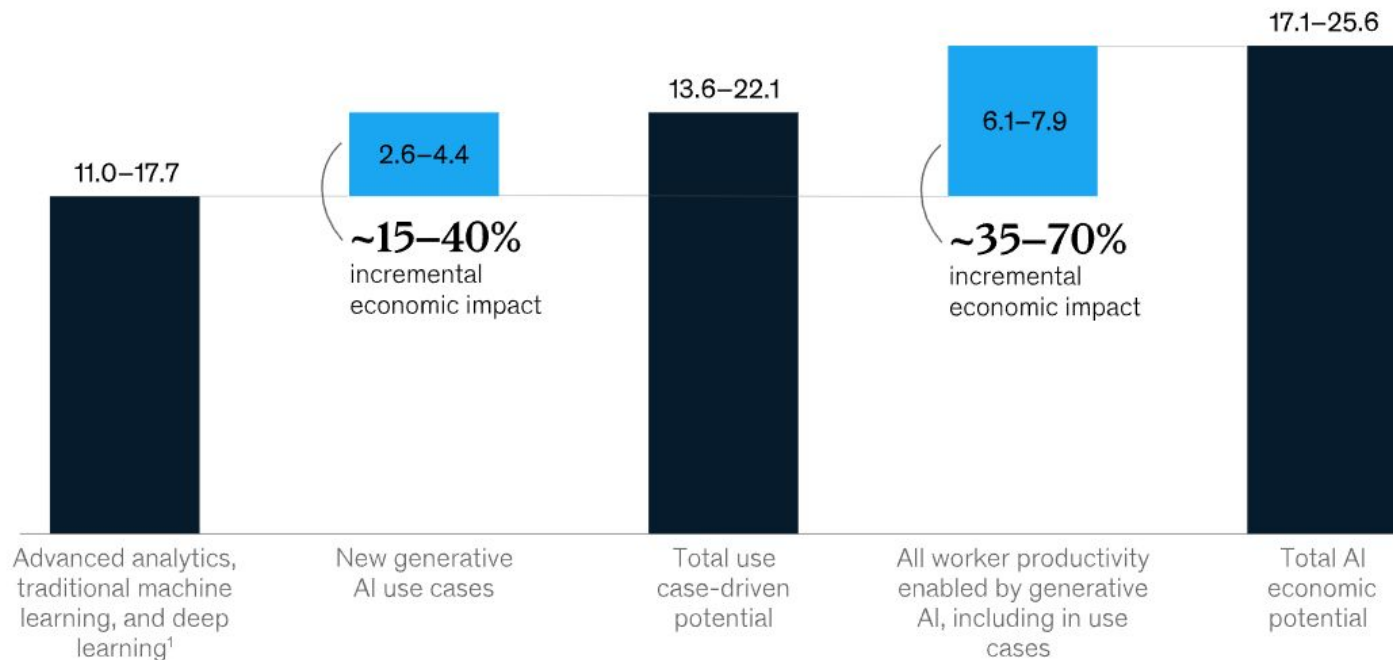
Quelle: ChatGPT o1 und Adobe Firefly 3

Productivity Increases with AI



Quelle: [NN/g. July 2023]

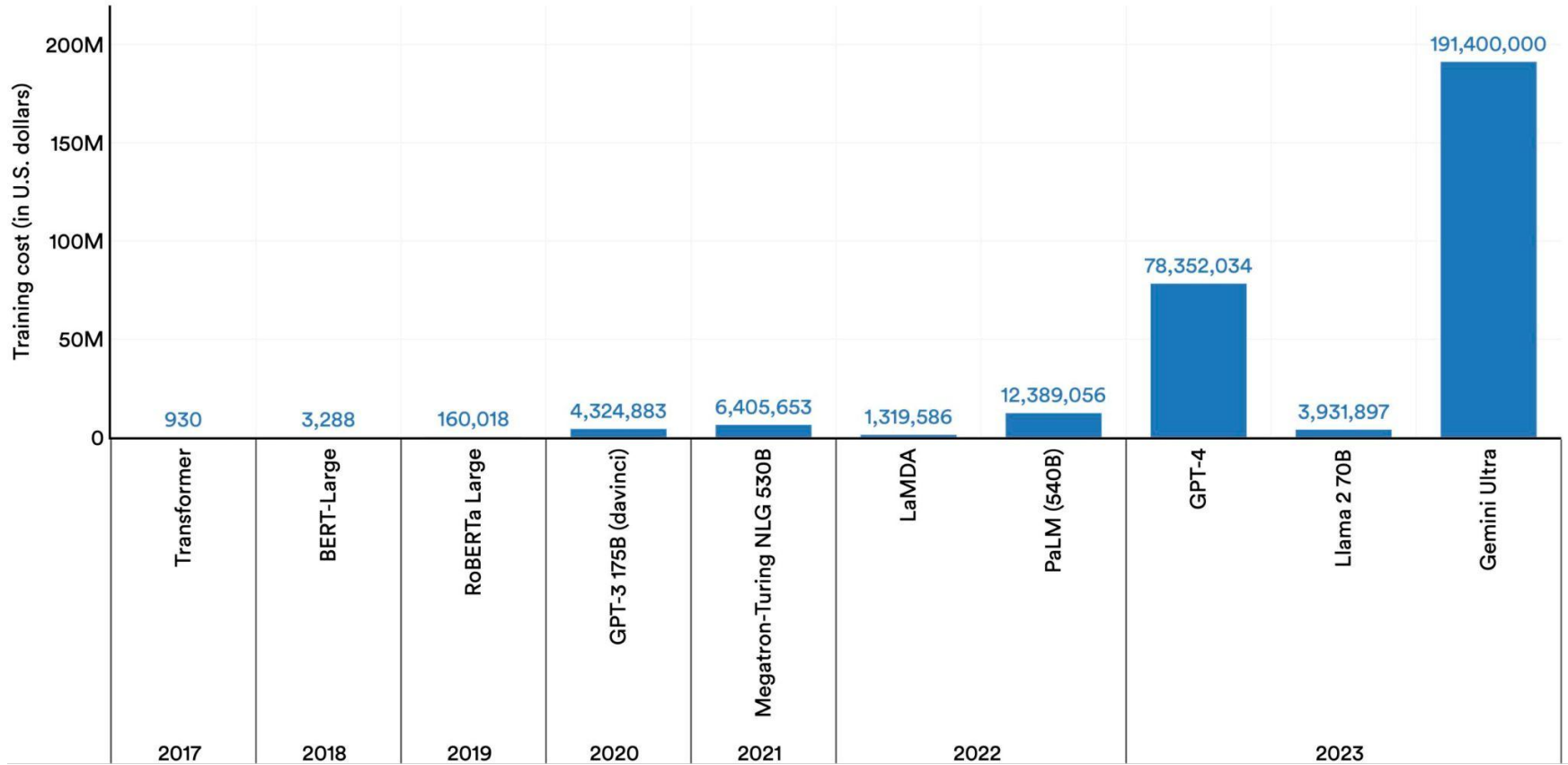
AI's potential impact on the global economy, \$ trillion



¹Updated use case estimates from "Notes from the AI frontier: Applications and value of deep learning," McKinsey Global Institute, April 17, 2018.

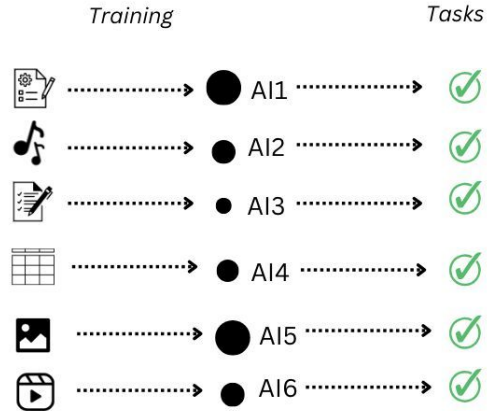
Estimated training cost of select AI models, 2017–23

Source: Epoch, 2023 | Chart: 2024 AI Index report



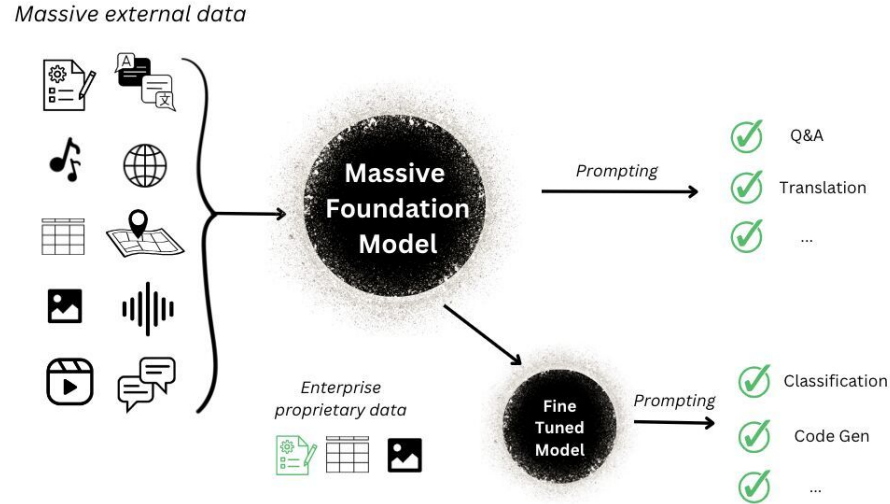
Credits [[Stanford AI Index](#)]

Traditional ML



- Individual siloed models
- Require task-specific training
- Lots of human supervised training

Foundation Models

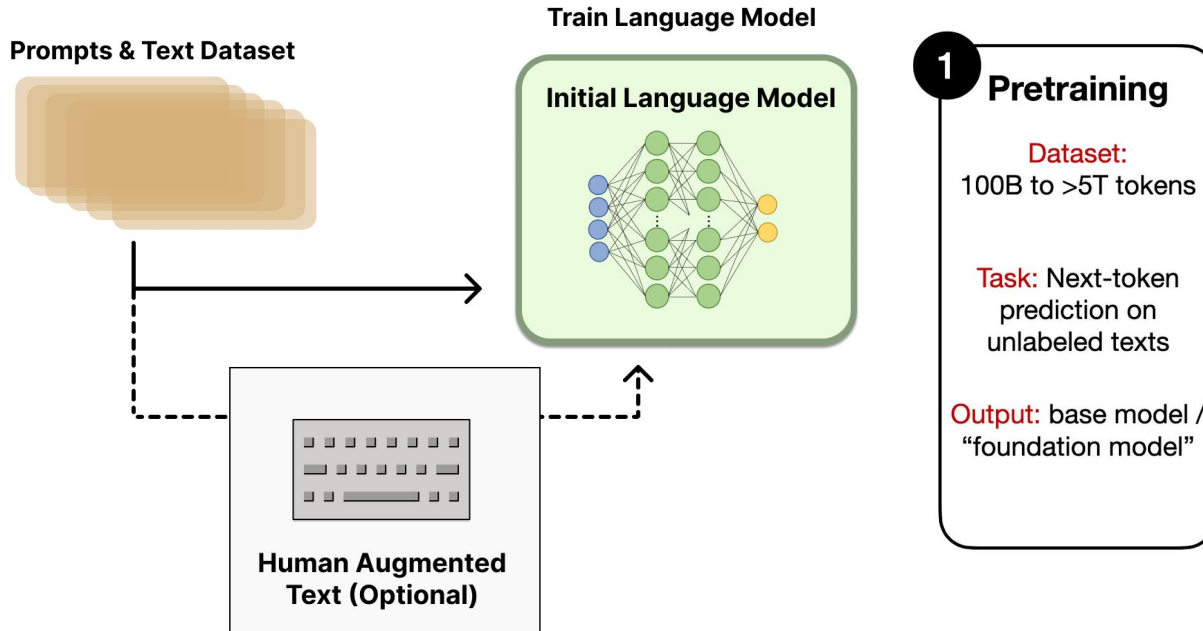


- Massive multi-tasking model
- Adaptable with little or no training
- Pre-trained unsupervised learning

Credits: [Armand Ruiz] [Conor Kelly](#)

How to train LLMs?

1. Self-supervised Pre-Training



Project Gutenberg (PG) is a volunteer effort to digitize and archive cultural works, as well as to "encourage the creation and distribution of eBooks." It was founded in 1971 by American writer Michael S. Hart and is the oldest digital **library**. Most of the items in its collection are the full texts of books or individual stories in the public domain. All files can be accessed for free under an open format layout, available on almost any computer. As of 3 October 2015, Project Gutenberg had reached 50,000 items in its collection of free eBooks.

Credits: [[Nathan Lambert et al.](#)]
[[Sebastian Raschka](#)]

2. Supervised (Instruction) Finetuning

2

Supervised finetuning

More **next-token
prediction**

Usually 1k-50k
instruction-response
pairs

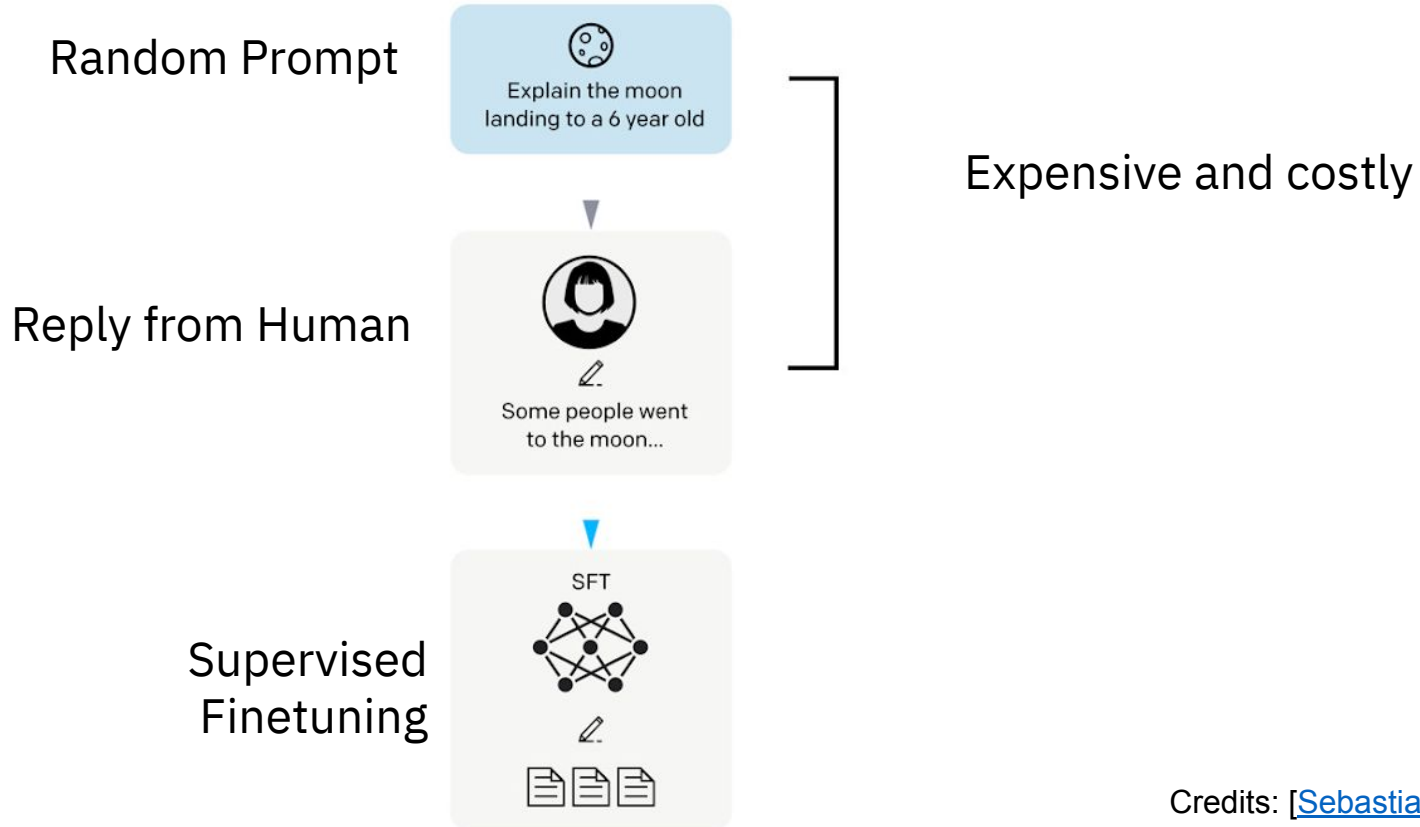
```
{
  "instruction": "Write a limerick about a
                pelican.",
  "input": "",
  "output": "There once was a pelican so fine,
            \nHis beak was as colorful as
            sunshine,\nHe would fish all day,\nIn
            a very unique way,\nThis pelican was
            truly divine!\n\n\n"
},

{
  "instruction": "Identify the odd one out from
                the group.",
  "input": "Carrot, Apple, Banana, Grape",
  "output": "Carrot\n\n"
},
```

Credits: [\[Sebastian Raschka\]](#)

3. Alignment Tuning:

1. Reinforcement Learning from Human Feedback

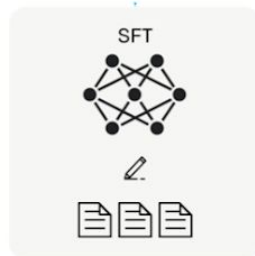


Credits: [\[Sebastian Raschka\]](#)

3. Alignment Tuning:

2. Reinforcement Learning from Human Feedback

LLM finetuned in step 1:



Random Prompt
Explain the moon landing to a 6 year old

Replies by LLM
A Explain gravity... B Explain war...
C Moon is natural satellite of... D People went to the moon...

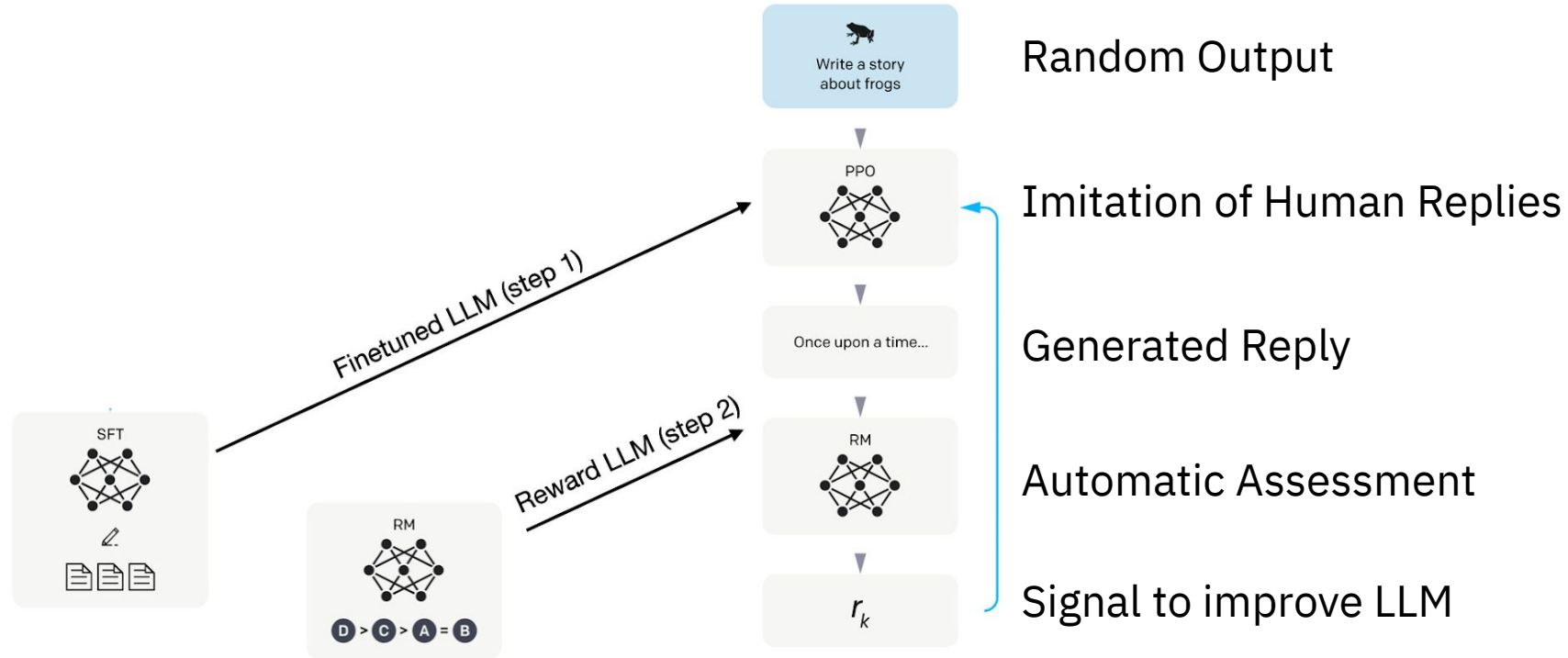
Sorted by Human according to quality
D > C > A = B

Learning of Reward model
RM
D > C > A = B

Credits: [\[Sebastian Raschka\]](#)

3. Alignment Tuning:

3. Reinforcement Learning from Human Feedback



Credits: [\[Sebastian Raschka\]](#)

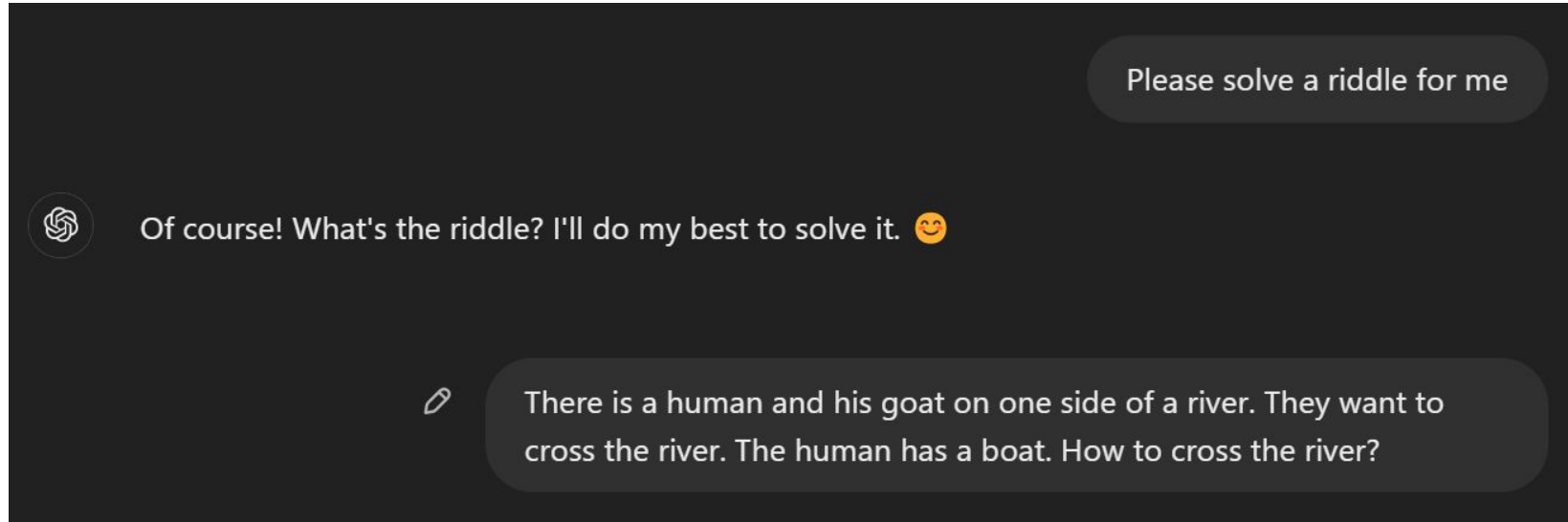
Further Improving LLMs ...

**In-Context
Learning**

Fine-Tuning

**Retrieval
Augmented
Generation (RAG)**

Agentic Systems



Inspired by: [[Marc-Oliver Gewalting. LinkedIn Mai 2024](#)]



Ah, this version adds a constraint where the human and the goat cannot fit in the boat together. In this case, the solution requires some creativity, assuming the boat can only carry either the human or the goat at a time:

Steps to Cross the River:

1. The human crosses the river alone, leaving the goat on the starting side.
2. The human leaves the boat on the far side and walks back to the starting side.
3. The human ties a rope to the goat and guides it across the river while the goat swims alongside the boat.

Now, both the human and the goat are safely across the river!

Inspired by: [[Marc-Oliver Gewaltig. LinkedIn Mai 2024](#)]

Automating Machine Learning

Why AutoML?

AutoML enables



More **efficient** research and development of ML applications

→ AutoML has been shown to outperform humans on subproblems



More **systematic** research and development of ML applications

→ no (human) bias or unsystematic evaluation



More **reproducible** research

→ since it is systematic!

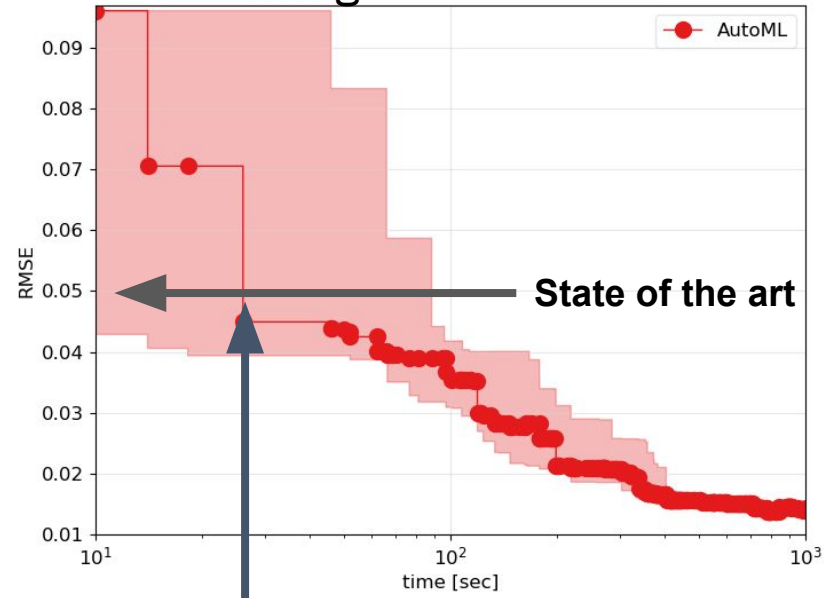


Broader use of ML methods

→ less required ML expert knowledge

→ not only limited to computer scientists

Digital Twin Task



**Better than state of the art
in less than 30sec!**

AutoML: Hyperparameter Optimization (HPO)



Home Installation Documentation Examples

Google Custom Search



sklearn.svm.SVC

```
class sklearn.svm.SVC(C=1.0, kernel='rbf', degree=3, gamma='auto_deprecated', coef0=0.0, shrinking=True, probability=False, tol=0.001, cache_size=200, class_weight=None, verbose=False, max_iter=1, decision_function_shape='ovr', random_state=None) [source]
```

C-Support Vector Classification.

The implementation is based on libsvm. The fit time complexity is more than quadratic with the number of samples which makes it hard to scale to dataset with more than a couple of 10000 samples.

The multiclass support is handled according to a one-vs-one scheme.

For details on the precise mathematical formulation of the provided kernel functions and how gamma, coef0 and degree affect each other, see the corresponding section in the narrative documentation: [Kernel functions](#).

Read more in the [User Guide](#).

Parameters: **C** : *float, optional (default=1.0)*

Penalty parameter C of the error term.

kernel : *string, optional (default='rbf')*

Specifies the kernel type to be used in the algorithm. It must be one of 'linear', 'poly', 'rbf', 'sigmoid', 'precomputed' or a callable. If none is given, 'rbf' will be used. If a callable is given it is used to pre-compute the kernel matrix from data matrices; that matrix should be an array of shape (n_samples, n_samples).

degree : *int, optional (default=3)*

Degree of the polynomial kernel function ('poly'). Ignored by all other kernels.

gamma : *float, optional (default='auto')*

Kernel coefficient for 'rbf', 'poly' and 'sigmoid'.

Current default is 'auto' which uses $1/n_{\text{features}}$, if `gamma='scale'` is passed then it uses $1/(n_{\text{features}} * X.\text{var}())$ as value of gamma. The current default of gamma, 'auto', will change to 'scale' in version 0.22. 'auto_deprecated', a deprecated version of 'auto' is used as a default indicating that no explicit value of gamma was passed.

coef0 : *float, optional (default=0.0)*

Independent term in kernel function. It is only significant in 'poly' and 'sigmoid'.

shrinking : *boolean, optional (default=True)*

Whether to use the shrinking heuristic.

probability : *boolean, optional (default=False)*

Whether to enable probability estimates. This must be enabled prior to calling fit, and will slow

SGD

```
CLASS torch.optim.SGD(params, lr=<required parameter>, momentum=0, dampening=0, weight_decay=0, nesterov=False, *, maximize=False, foreach=None, differentiable=False) [SOURCE]
```

Implements stochastic gradient descent (optionally with momentum).

Parameters:

- **params** (*iterable*) – iterable of parameters to optimize or dicts defining parameter groups
- **lr** (*float*) – learning rate
- **momentum** (*float, optional*) – momentum factor (default: 0)
- **weight_decay** (*float, optional*) – weight decay (L2 penalty) (default: 0)
- **dampening** (*float, optional*) – dampening for momentum (default: 0)
- **nesterov** (*bool, optional*) – enables Nesterov momentum (default: False)
- **maximize** (*bool, optional*) – maximize the params based on the objective, instead of minimizing (default: False)
- **foreach** (*bool, optional*) – whether foreach implementation of optimizer is used. If unspecified by the user (so foreach is None), we will try to use foreach over the for-loop implementation on CUDA, since it is usually significantly more performant. (default: None)
- **differentiable** (*bool, optional*) – whether autograd should occur through the optimizer step in training. Otherwise, the step() function runs in a torch.no_grad() context. Setting to True can impair performance, so leave it False if you don't intend to run autograd through this instance (default: False)

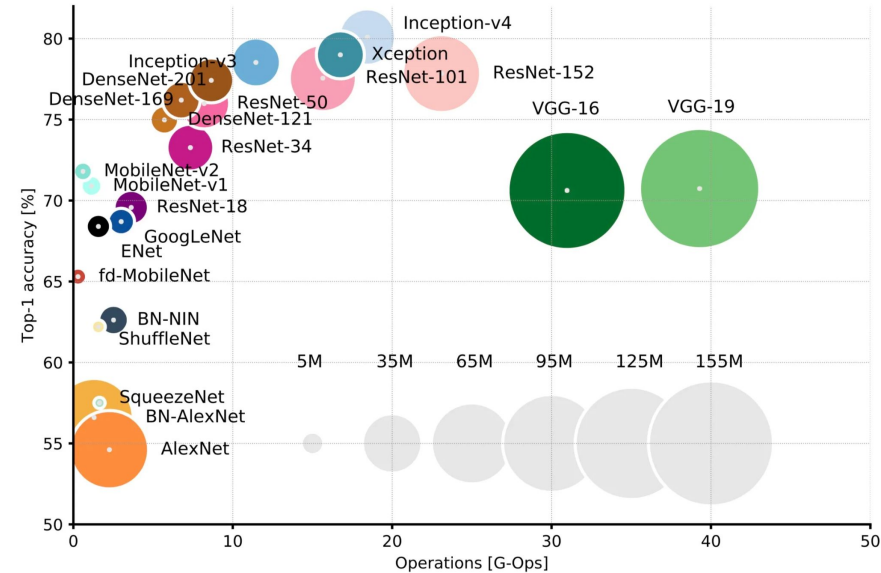
Architectures of Neural Networks

Many architectures exist and differ in

- Depth
- Resolution
- Width
- Operators
- Connections
- ...

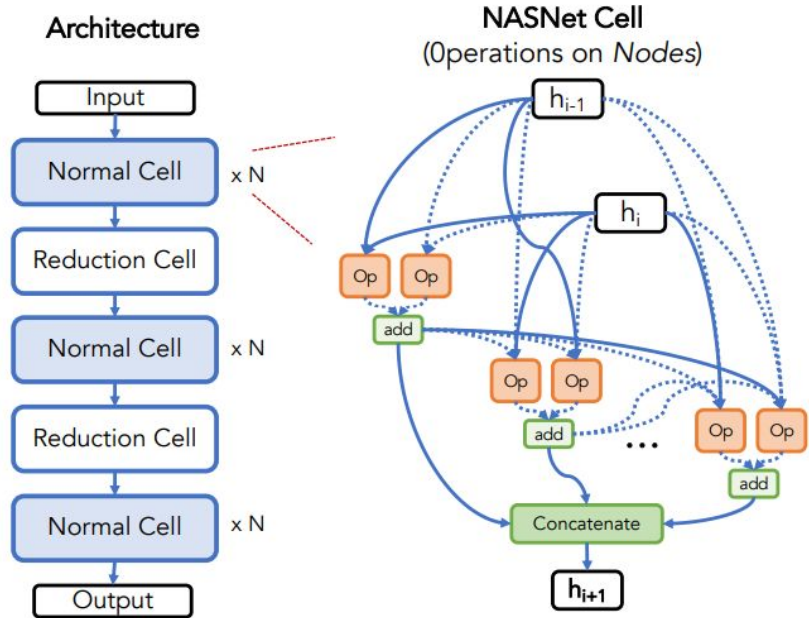
Already on a single dataset (e.g. ImageNet), it is **not obvious** which architecture to choose

- On **different** datasets → different architectures
- On **similar** datasets → scaled versions of known architectures (e.g. ImageNet and Cifar10)



Source: [[Culurciello et al. 2018](#)]

AutoML: Neural Architecture Search (NAS)

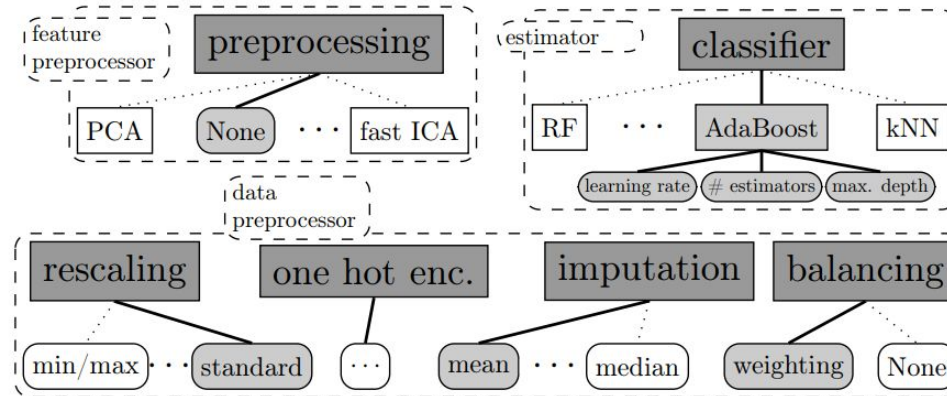


[White et al. 2023]

- Determine the best architecture of a neural network for a given dataset
- Includes
 - Macro-structure
 - Cell structure
 - Operators
 - ...

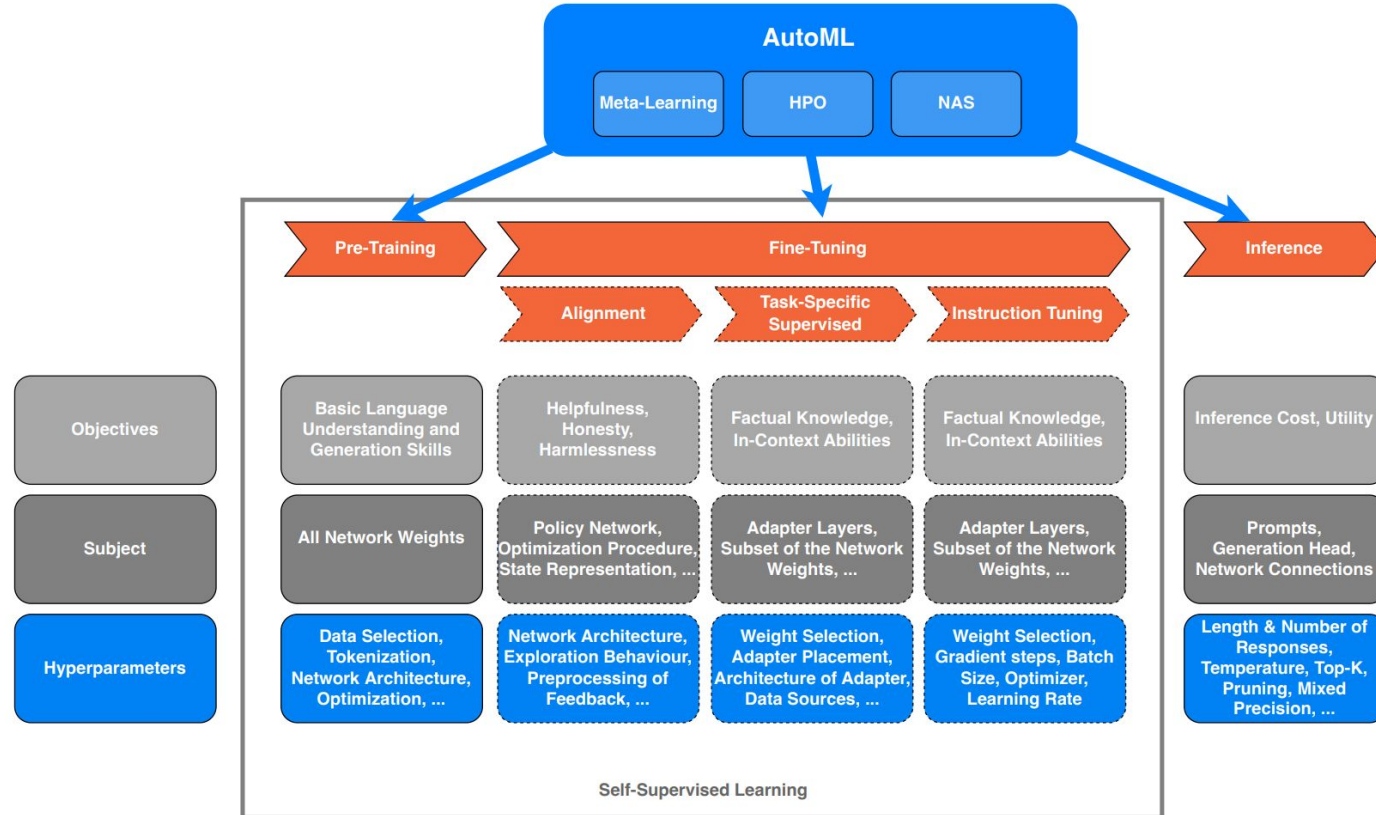
AutoML: ML Pipeline Design

- Eventually, we want to design entire ML pipelines
- Consisting of
 - Model class (e.g., DNN vs XGBoost)
 - Preprocessing of the data
 - Data augmentation
 - Feature engineering
 - Postprocessing of the predictions ...

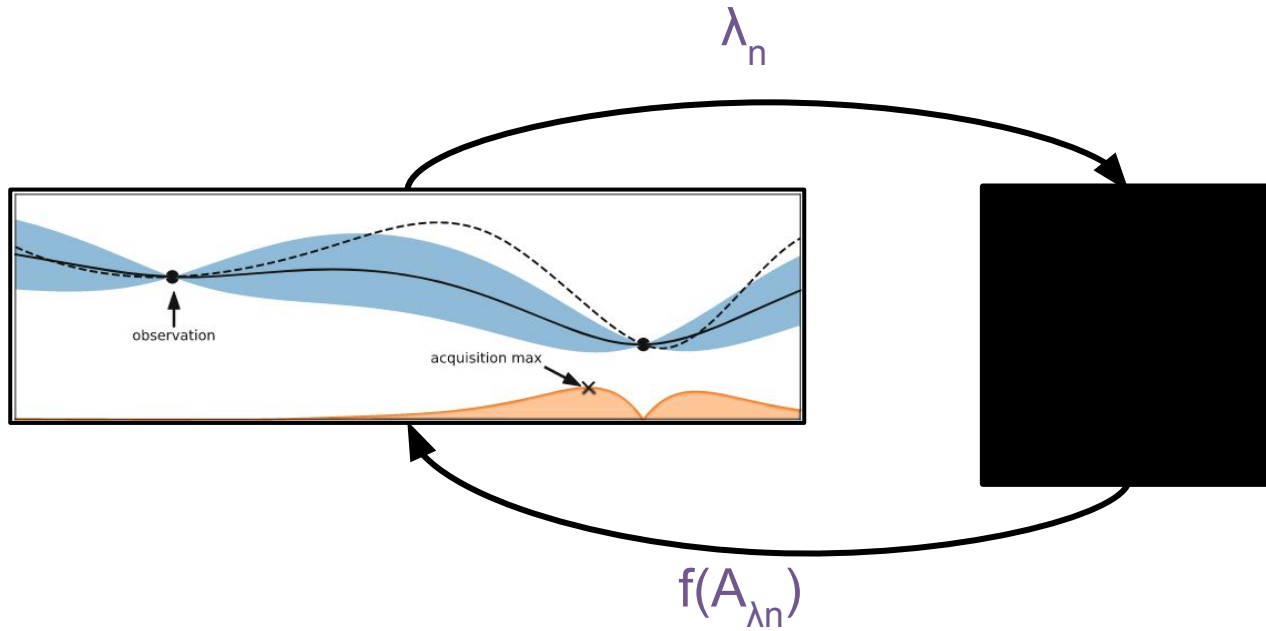


[Feurer et al. 2015]

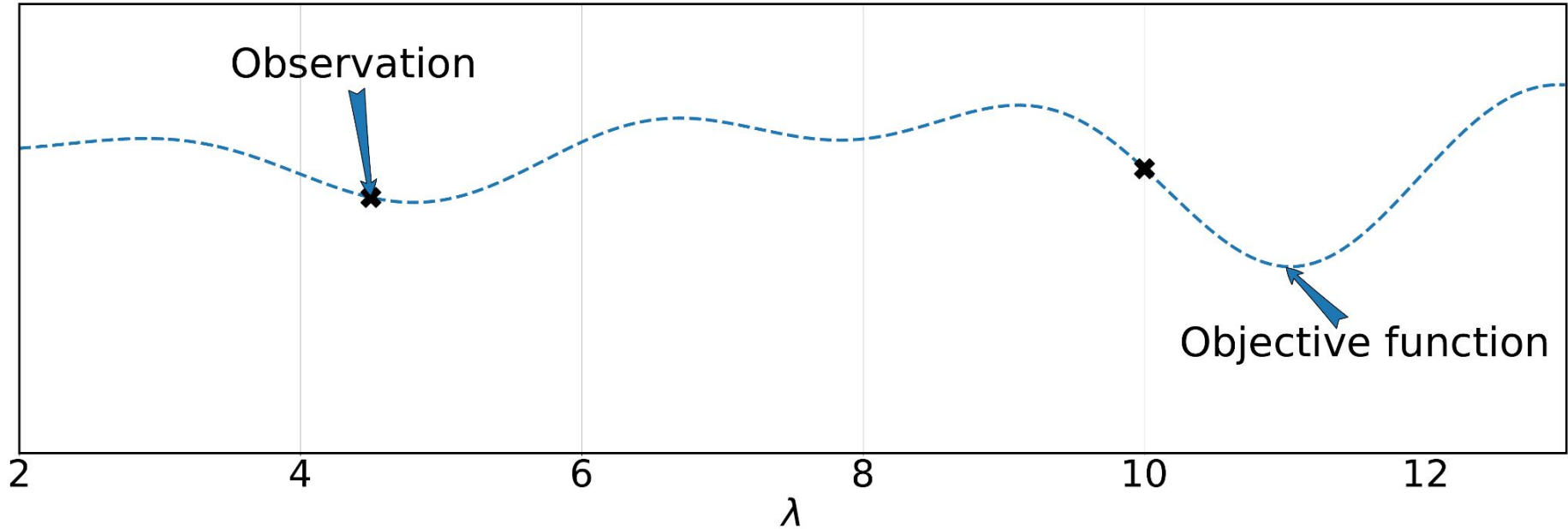
AutoML in the Age of LLMs [Tornede et al. TMLR'24]



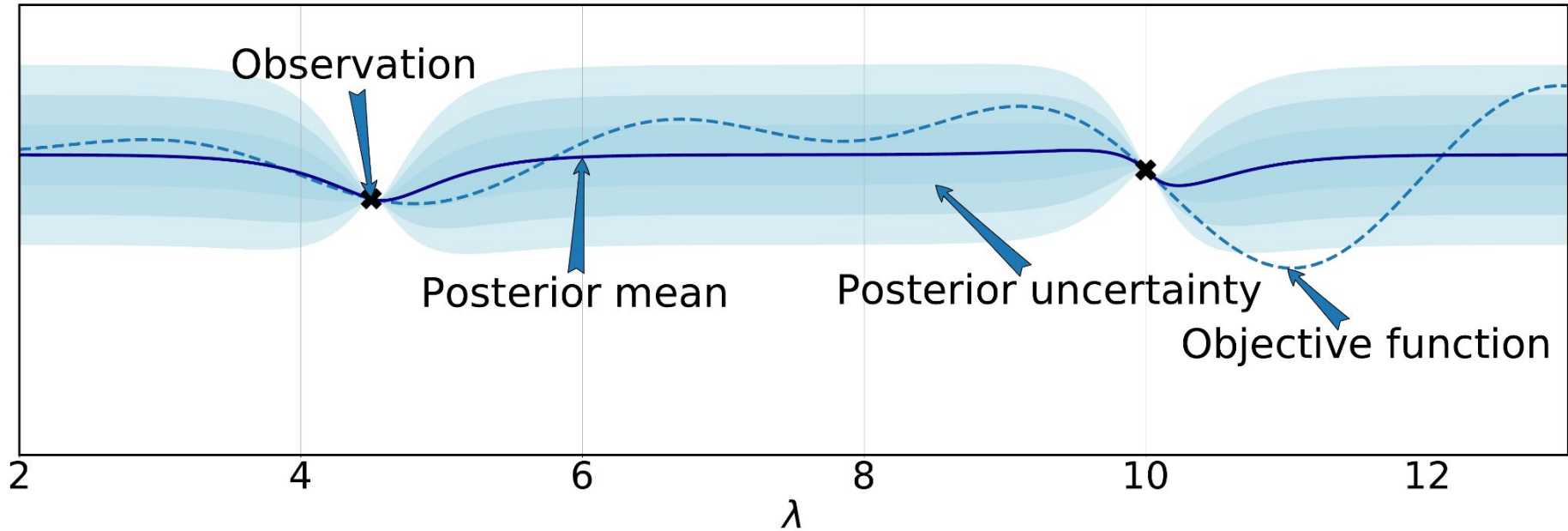
Model-based Optimization



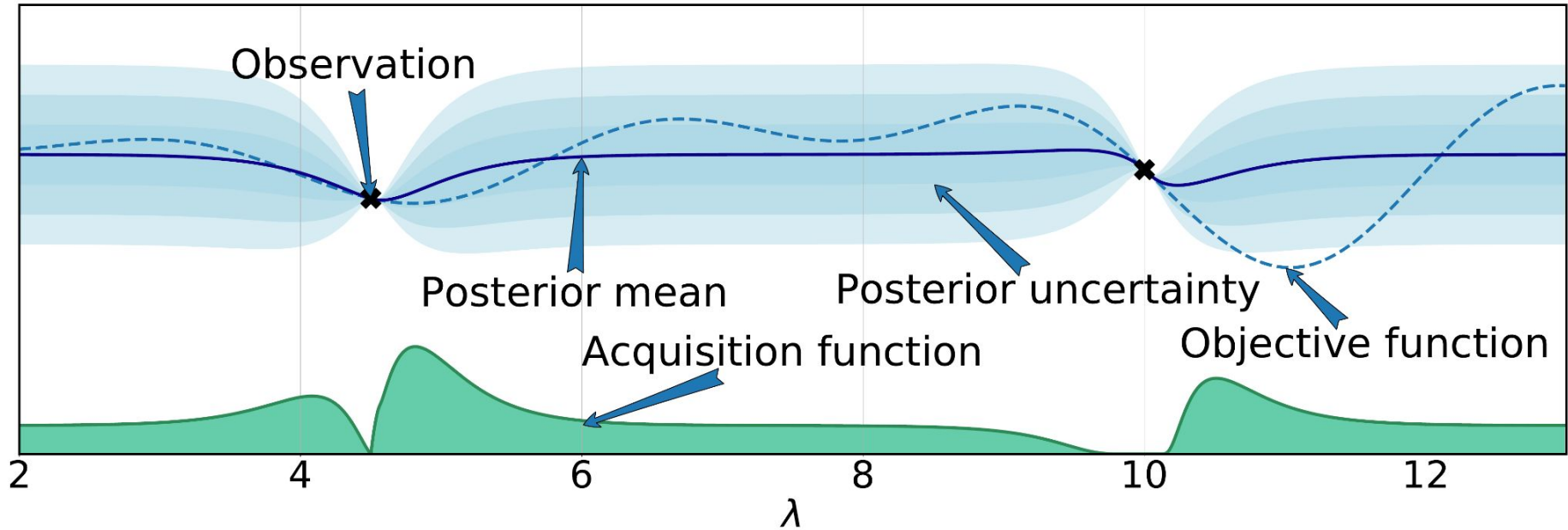
Bayesian Optimization in a Nutshell



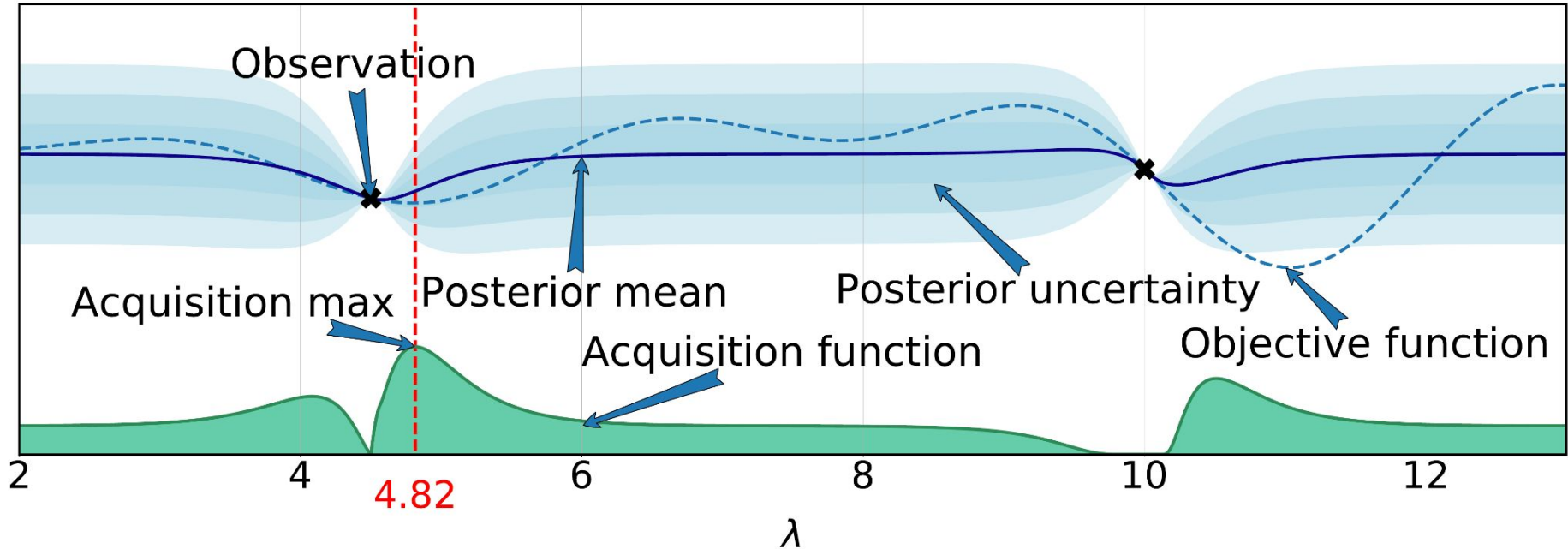
Bayesian Optimization in a Nutshell



Bayesian Optimization in a Nutshell



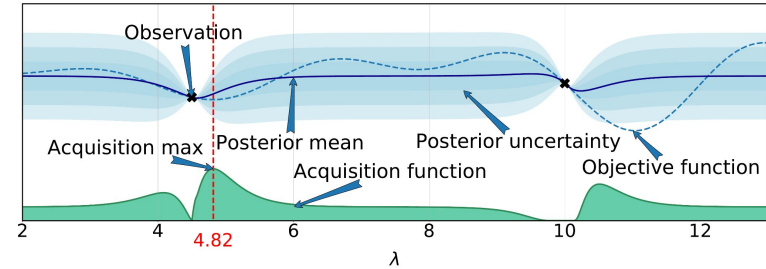
Bayesian Optimization in a Nutshell



Bayesian Optimization in a Nutshell

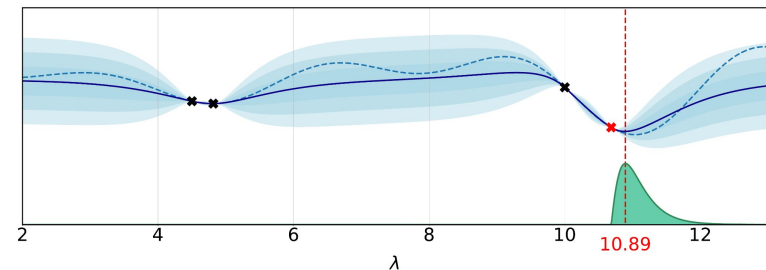
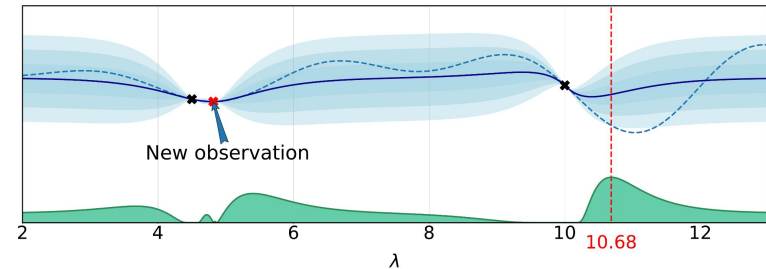
General approach

- Fit a **probabilistic model** to the collected function samples $\langle \lambda, c(\lambda) \rangle$
- Use the model to guide optimization, trading off **exploration vs exploitation**



Popular approach in the statistics literature since Mockus et al. [1978]

- Efficient in **#function evaluations**
- Works when objective is **nonconvex**, **noisy**, has **unknown derivatives**, etc.
- Recent **convergence results**
 [Srinivas et al. 2009; Bull et al. 2011; de Freitas et al. 2012; Kawaguchi et al. 2015]



Bayesian Optimization: Pseudocode

BO loop

Require: Search space Λ , cost function c , acquisition function u , predictive model \hat{c} , maximal number of function evaluations T

Result : Best configuration $\hat{\lambda}$ (according to \mathcal{D} or \hat{c})

- 1 Initialize data $\mathcal{D}^{(0)}$ with initial observations
- 2 **for** $t = 1$ **to** T **do**
- 3 Fit predictive model $\hat{c}^{(t)}$ on $\mathcal{D}^{(t-1)}$
- 4 Select next query point: $\lambda^{(t)} \in \arg \max_{\lambda \in \Lambda} u(\lambda; \mathcal{D}^{(t-1)}, \hat{c}^{(t)})$
- 5 Query $c(\lambda^{(t)})$
- 6 Update data: $\mathcal{D}^{(t)} \leftarrow \mathcal{D}^{(t-1)} \cup \{ \langle \lambda^{(t)}, c(\lambda^{(t)}) \rangle \}$

Bayesian Optimization + X

**Multi-fidelity
Optimization**

Meta-Learning

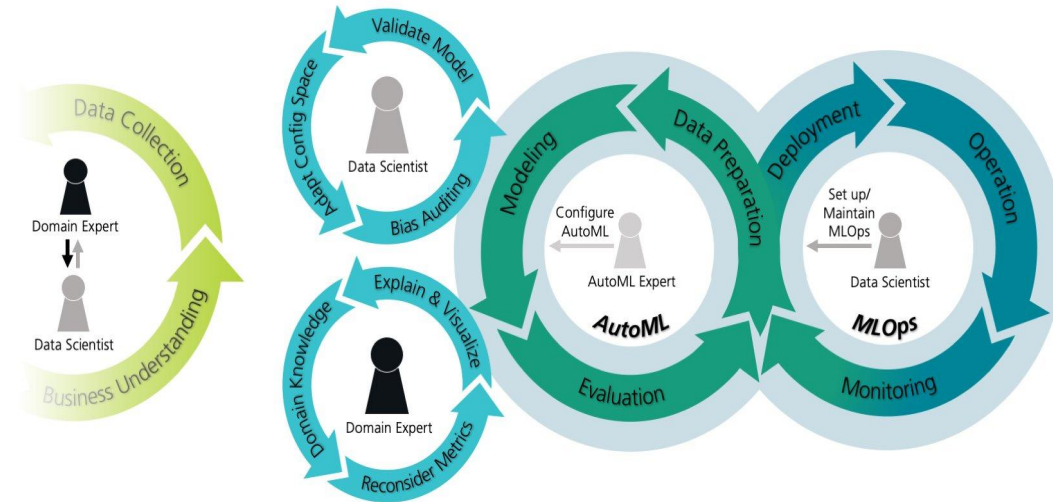
**Multi-Objective
Optimization**

User Priors

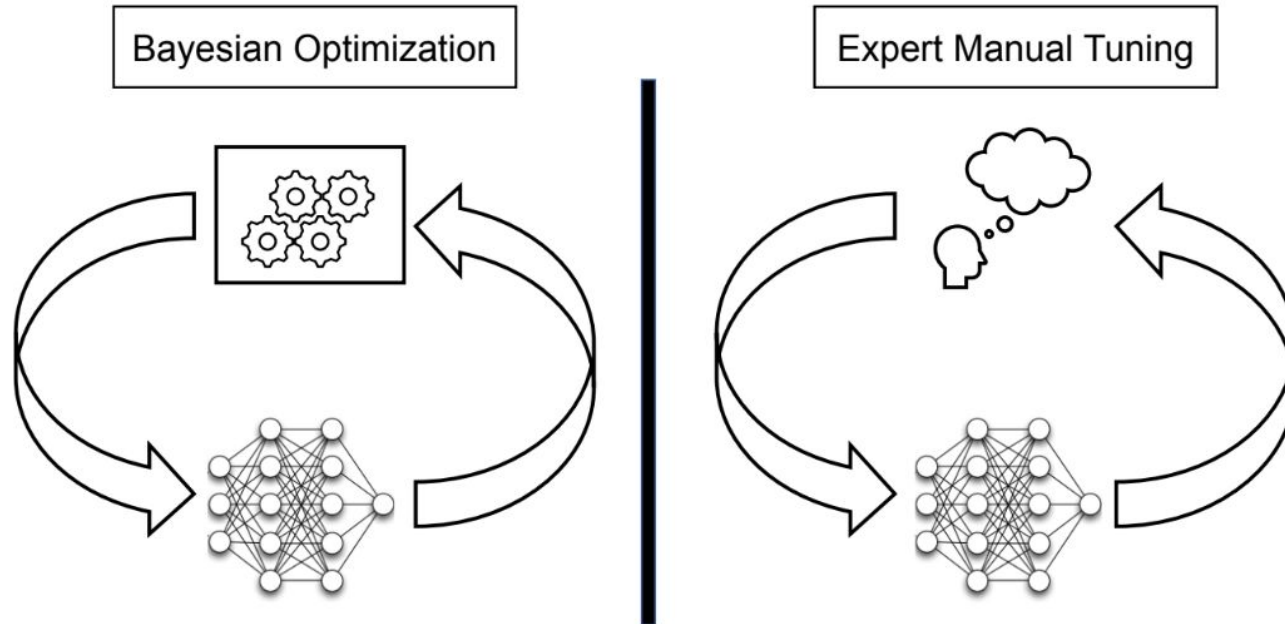
Human-Centered AutoML

Human-Centered AutoML [Lindauer et al. ICML'24]

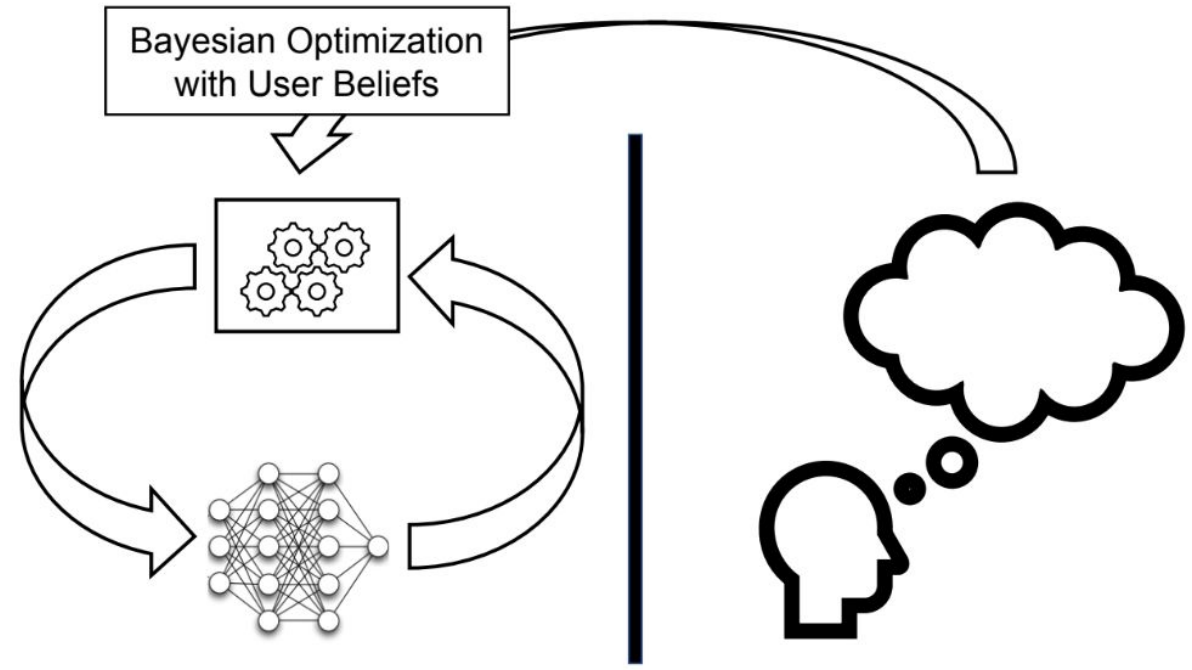
- More **efficient** by leveraging human expertise
- More **trustworthy** by considering human expectations
- More **customizable** by building approaches for different user groups
- Better integrable into existing **data science workflows**
- Less wasted time in failed experiments because of **interactivity**
- **Empowering humans** instead of making them dependent

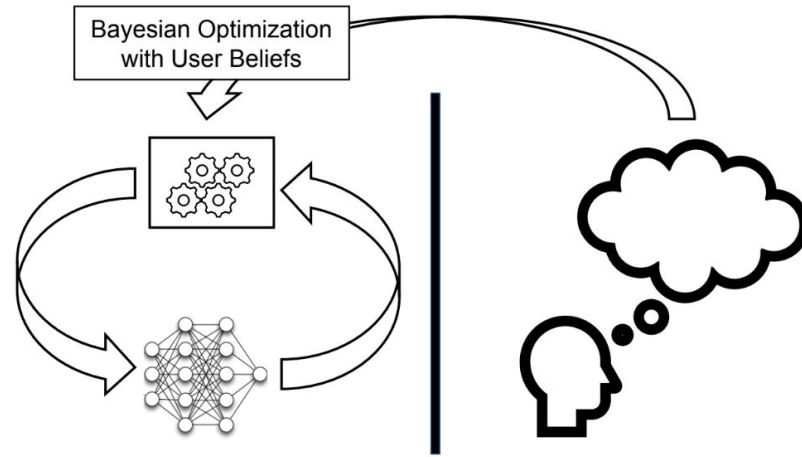


Bayesian Optimization vs Manual Tuning



Bayesian Optimization with Expert Knowledge



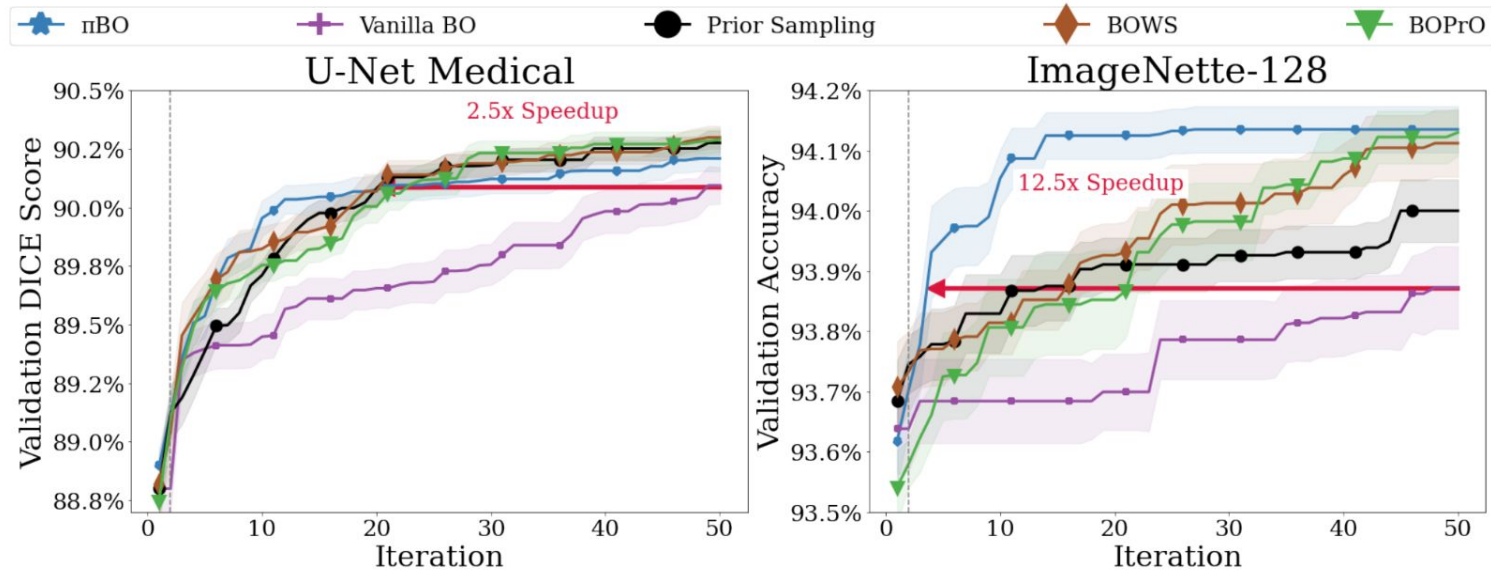


$$\mathbf{x}_n \in \arg \max_{\mathbf{x} \in \mathcal{X}} \alpha(\mathbf{x}, \mathcal{D}_n) \pi(\mathbf{x})^{\beta/n}$$

Acquisition Function

User Prior

Speed of forgetting user prior

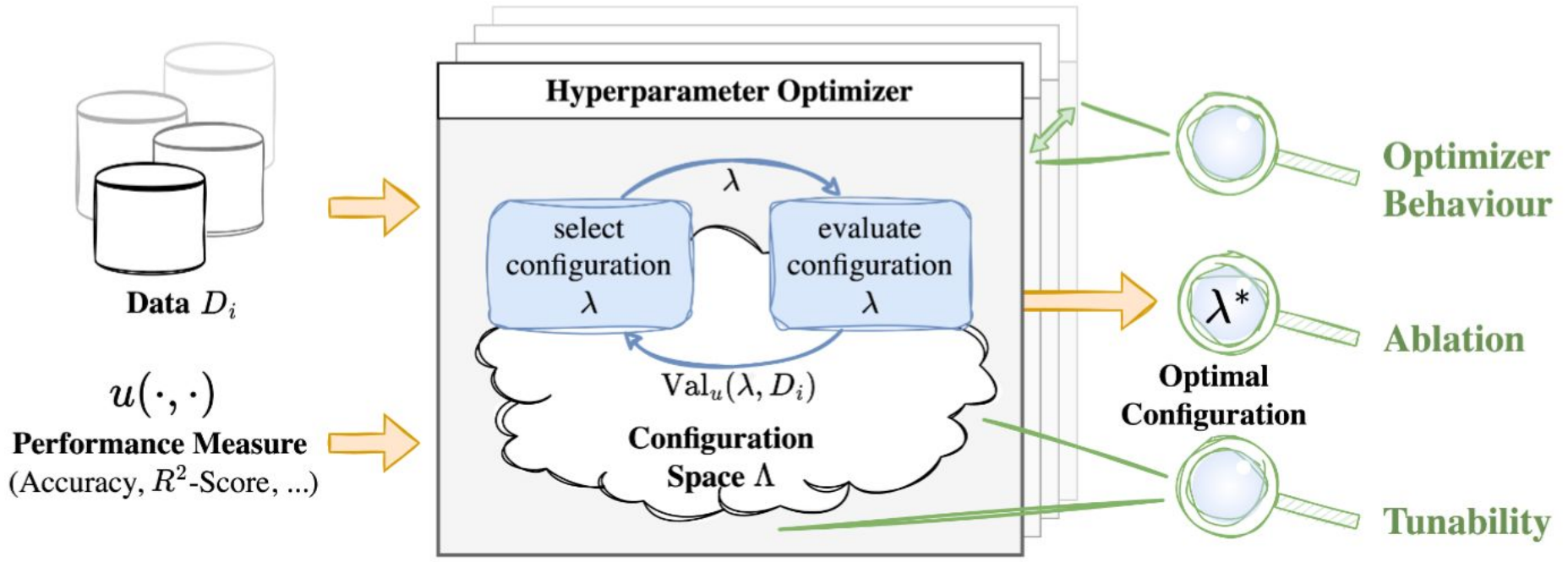


- Uses expert knowledge to speed up Bayesian Optimization
- Robust also against wrong beliefs
- Substantially speeds up AutoML
- Follow up with PriorBand [\[Mallik et al. NeurIPS'23\]](#)

HYPERSHAP: SHAPLEY VALUES AND INTERACTIONS FOR HYPERPARAMETER IMPORTANCE [Wever et al under Review]

Hyperparameter Optimization Pipeline

HyperSHAP



Shapley-Games for HPO I

Definition 1 (HPI Game - Ablation). *The Ablation HPI game is defined as a tuple*

$$G_A = (\mathcal{N}, \lambda^0, \lambda^*, D, \nu),$$

consisting of a player set \mathcal{N} , a reference HPC λ^0 , an HPC of interest λ^ , a dataset D , and a value function ν . Given a coalition $S \subseteq \mathcal{N}$, we construct an intermediate HPC λ^S from λ^0 and λ^* as:*

$$\lambda^S = \begin{cases} \lambda_i^* & \text{if } i \in S \\ \lambda_i^0 & \text{else} \end{cases}$$

Then, the value function $\nu : 2^{\mathcal{N}} \rightarrow \mathbb{R}$ is defined as

$$\nu(S) = \text{VAL}_u(\lambda^S, D) .$$

Definition 2 (HPI Game - Tunability). *The tunability HPI game is defined as a tuple*

$$G_T = (\mathcal{N}, \lambda^0, \Lambda, \mathcal{D}, \nu),$$

consisting of a set of players \mathcal{N} , a reference HPC $\lambda^0 \in \Lambda$, a hyperparameter configuration space Λ , a collection of datasets $\mathcal{D} = \{D_1, D_2, \dots, D_M\}$, and a value function ν .

Given a coalition S , we construct a hyperparameter configuration space Λ^S from the original Λ as a subspace $\Lambda^S = \times_{i \in S} \Lambda_i$ and the value function is defined as

$$\nu(S) = \bigoplus_{i=1}^M \text{VAL}_u \left(\arg \max_{\lambda \in \Lambda^S} \text{VAL}_u(\lambda, D_i), D_i \right),$$

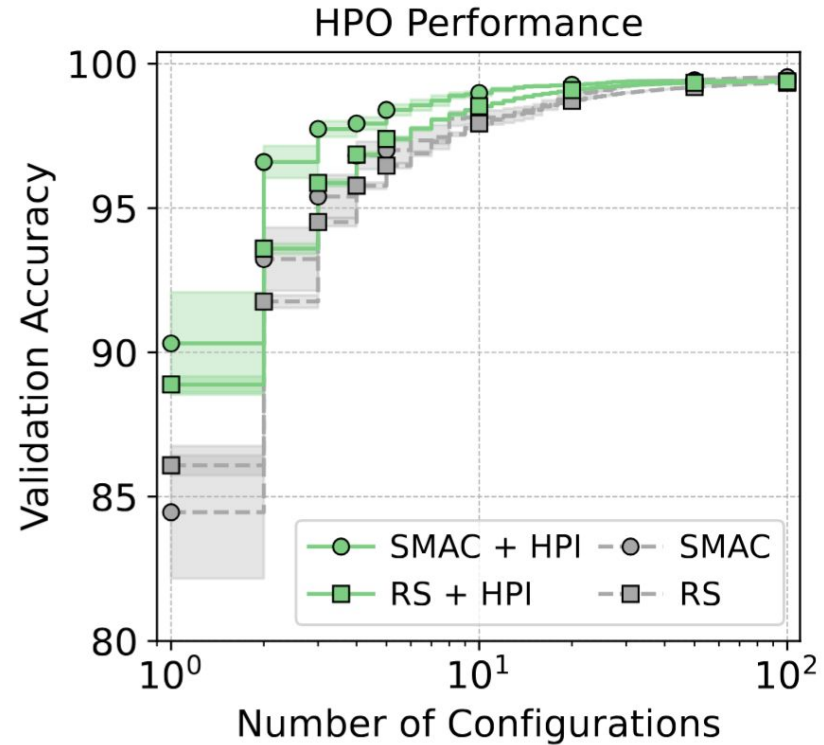
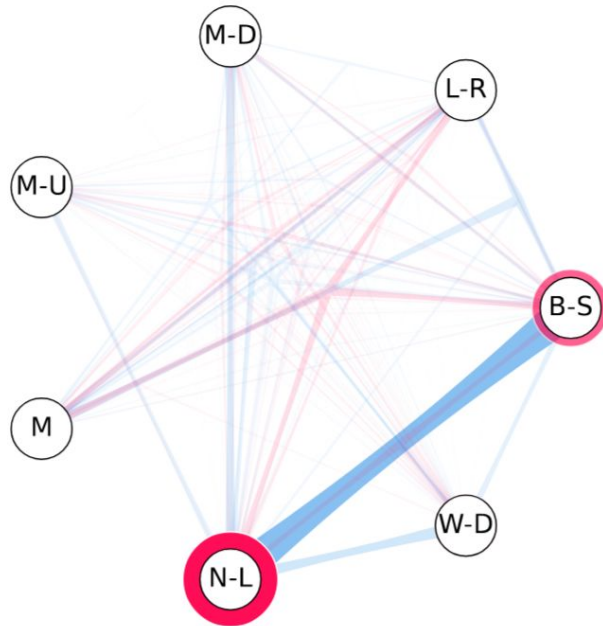
Definition 4 (HPI Game - Optimizer Bias). *The Optimizer Bias HPI game is defined as a tuple*

$$G_O = (\mathcal{N}, \Lambda, \lambda^0, \mathcal{O}, \mathcal{D}, \nu),$$

where \mathcal{N} , Λ , λ^0 , \mathcal{D} , and the construction of Λ^S are as in Definition 2, \mathcal{O} the hyperparameter optimizer of interest, and a value function ν . Then, the value function is defined as

$$\nu(S) = \bigoplus_{i=1}^M \left[\text{VAL}_u (\mathcal{O}(D_i, \Lambda^S), D_i) - \text{VAL}_u \left(\arg \max_{\lambda \in \Lambda^S} \text{VAL}_u(\lambda, D_i), D_i \right) \right].$$

Use of Shapley-Explanations for HPO



The next generation of ML?

(not from my group!)

TabPFN: Pre-Trained Transformers for Inductive Learning of Tabular Data

[[Hollmann et al. ICLR'23](#)]

Done once, offline

Sample synthetic datasets D_i
from prior: $D_i \sim p(\mathbf{D})$

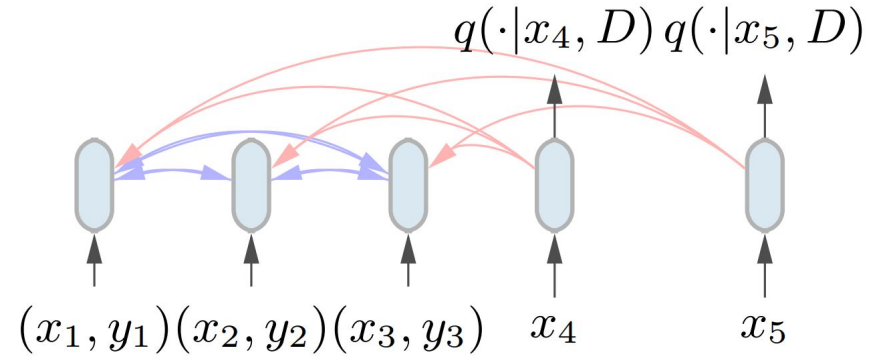
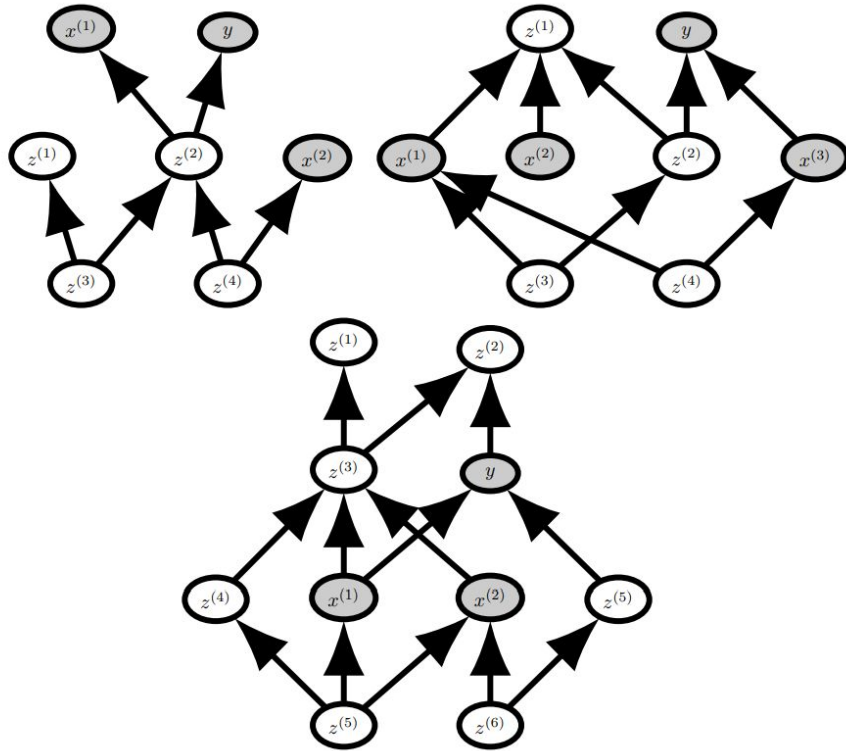
Train TabPFN q_θ on synthetic
datasets $\{D_1, \dots, D_n\}$

Done per real-world dataset, online

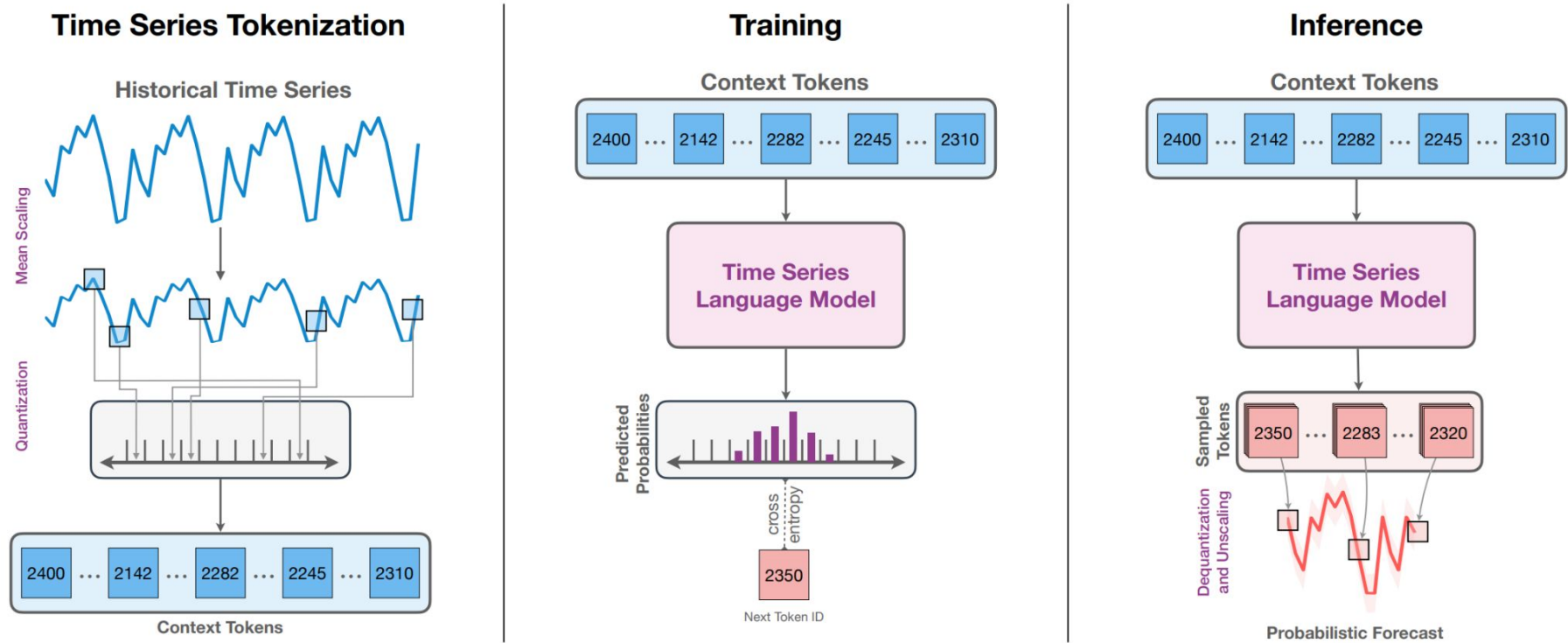
Real-world training dataset D_{real}
and test point x_{test}

Obtain $q_\theta(y_{test}|x_{test}, D_{real})$
with a single forward pass

Data Modelling in TabPFN [\[Hollmann et al. ICLR'23\]](#)

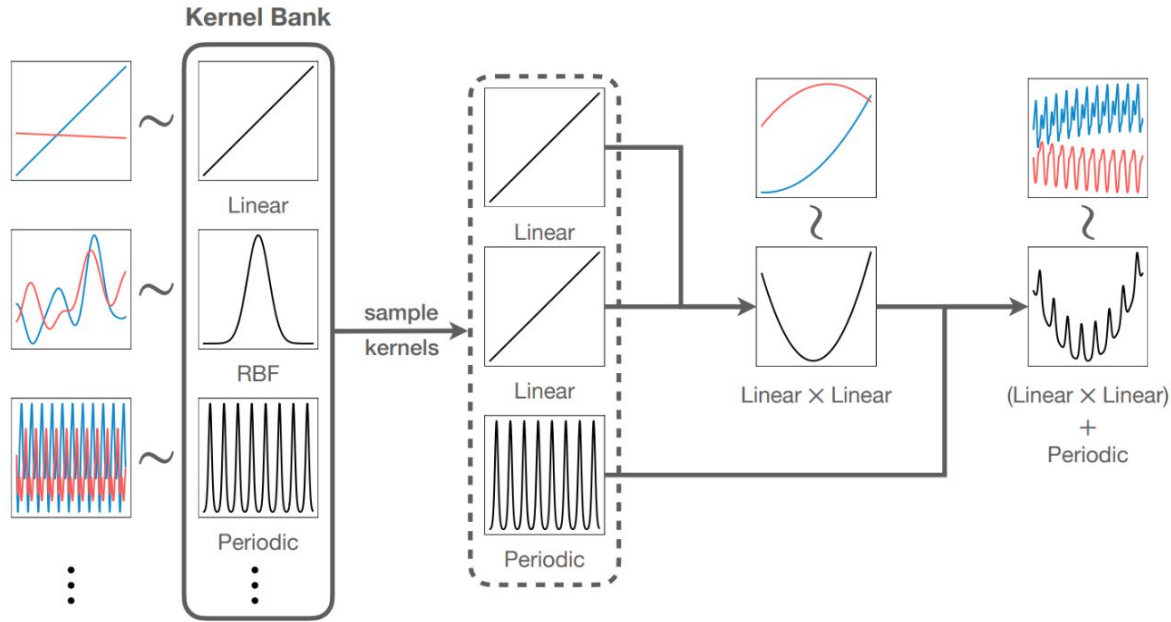


Forecasting [Ansari et al. TMLR'24]

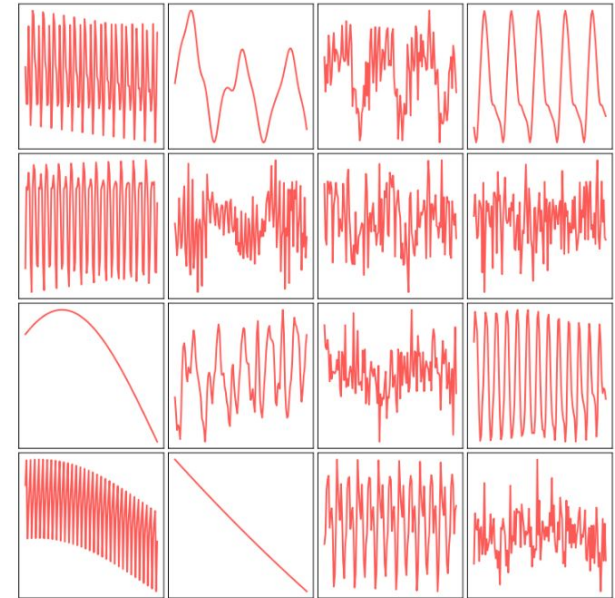


Data Modelling for Timeseries

[Ansari et al. TMLR'24]



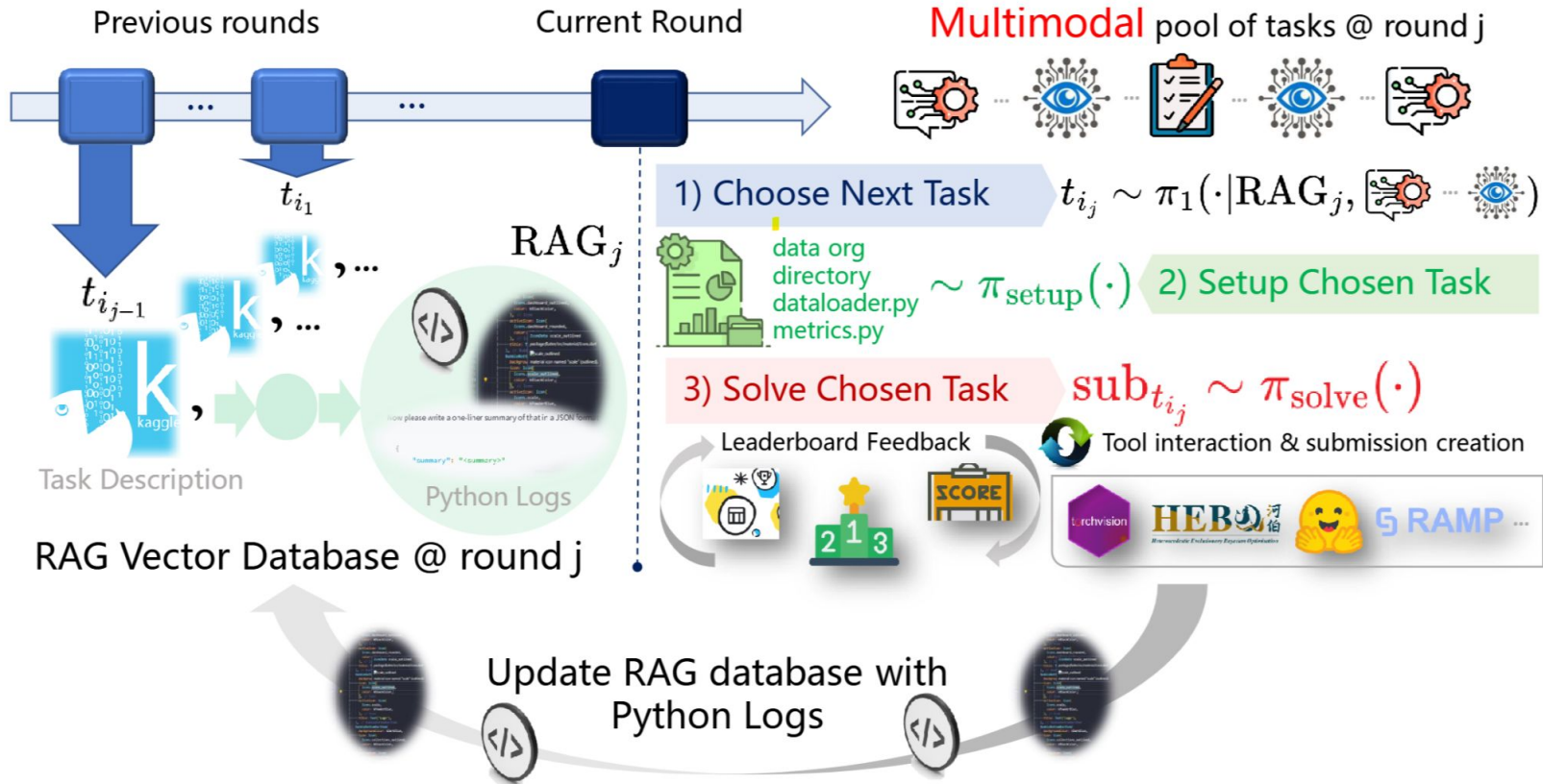
(a) KernelSynth



(b) Synthetic samples from KernelSynth

Agent K: The next Kaggle Grandmaster?

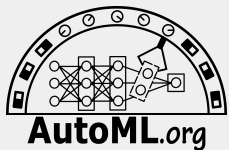
[Grosnit et al. 2024]



Take-Aways

1. Foundational models are the next / current wave of AI advances.
2. Although LLMs provide nice text interfaces, AutoML is still key for best performance on specific applications.
3. AutoML can take care of a multitude of design decisions, including hyperparameters, neural architectures and entire pipelines.
4. Foundational models for data modalities (such as tables or time series) **or** agentic LLM-systems promise the next generation of (Auto)ML.

Find Us



 [automl-org](https://www.linkedin.com/company/automl-org)

 [automl](https://github.com/automl)

 [@AutoML_org](https://www.youtube.com/@AutoML_org)



 [luh-ai](https://www.linkedin.com/company/luh-ai)

 [LUH-AI](https://github.com/luh-ai)

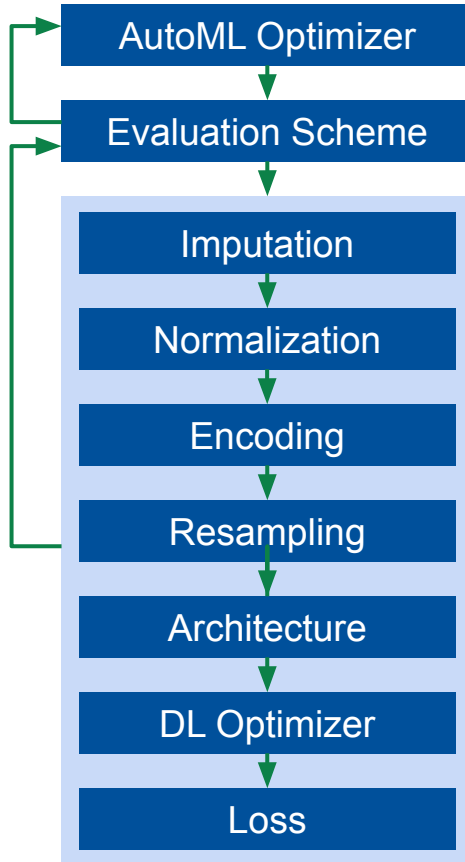
 [@luh-ai](https://www.youtube.com/@luh-ai)



Funded by:



AutoDL: Joint NAS & HPO



1. Deep Learning includes complex pipelines
2. NAS & HPO need to go hand in hand

→ Auto-PyTorch [\[Zimmer et al. 2021\]](#)

→ Auto-PyTorch for Time Series Forecasting [\[Deng et al. 2022\]](#)

```

# initialise Auto-PyTorch api
api = TabularClassificationTask()

# Search for an ensemble of machine learning algorithms
api.search(
    X_train=X_train,
    y_train=y_train,
    X_test=X_test,
    y_test=y_test,
    optimize_metric='accuracy',
    total_walltime_limit=300,
    func_eval_time_limit_secs=50
)

# Calculate test accuracy
y_pred = api.predict(X_test)
  
```

CATEGORY	PROMPT	PEDAGOGICAL PRINCIPLES	EXAMPLE PROMPT IN PAPER
SIMULATION	Role-playing with AI feedback	Practicing and applying knowledge	Negotiation simulator
SIMULATION	Goal-playing with AI feedback	Practicing applying frameworks in new situations	Help a fictional character develop goals; help a fictional character self-distance
CRITIQUE	Critique a scenario	Structuring knowledge. Critical thinking and protégé effect	Critique a scenario about groupthink
TEACH	Teach the AI	Teaching others is a powerful learning technique	Teach the AI about a subject you know well
CO-CREATE	Co-create a case	Break the illusion of explanatory depth. Structuring knowledge. Retrieval	Work with the AI to create a case for peer review by another student
MENTOR AND COACH	Reflection coach	Reflection is critical to learning	NA
MENTOR AND COACH	Integration agent	Creating connections and interleaving concepts	Helps students integrate two concepts
TUTOR	Tutor	Tutoring is an effective technique for improving learning	Provides structured, interactive tutoring support

- LLMs allows for 24/7 tutors
→ always available
- You can ask whatever you want
- Education will be personalized
- How students will learn and what they need to learn will substantially change.
- **Risk:** If we (=users) don't know anymore whether the output of LLMs is correct or not, we will get into huge trouble.

Credits: [[Ethan Mollick](#) 2023]