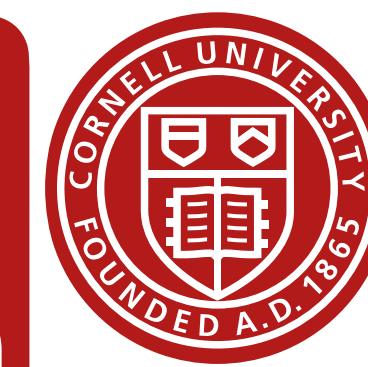


# N<sup>2</sup>: A Unified Python Package and Test Bench for Nearest Neighbor-Based Matrix Completion

Caleb Chin, Aashish Khubchandani, Harshvardhan Maskara, Kyuseong Choi, Jacob Feitelberg, Albert Gong, Manit Paul, Tathagata Sadhukhan, Anish Agarwal, Raaz Dwivedi



## Motivation

Nearest neighbor (NN) methods are effective tools for matrix completion applications but:

- There's no unified and extendable framework to consolidate NN methods for rapid experimentation and development.
- There's no standardized real-world benchmarks for matrix completion methods across multiple real-world datasets.

## Solution

N<sup>2</sup> is a **unified Python package** and **testbed** designed to:

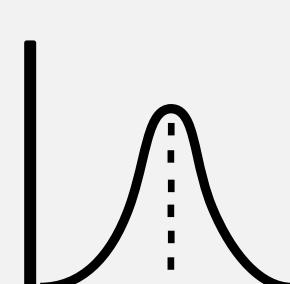
- Consolidate a broad class of NN-based methods through a **modular, extensible interface**.
- Stress-test matrix completion methods on **diverse real-world datasets** from healthcare and recommendation systems to causal inference and LLM evaluation.

## N<sup>2</sup> Framework

### DataType

Entry type:

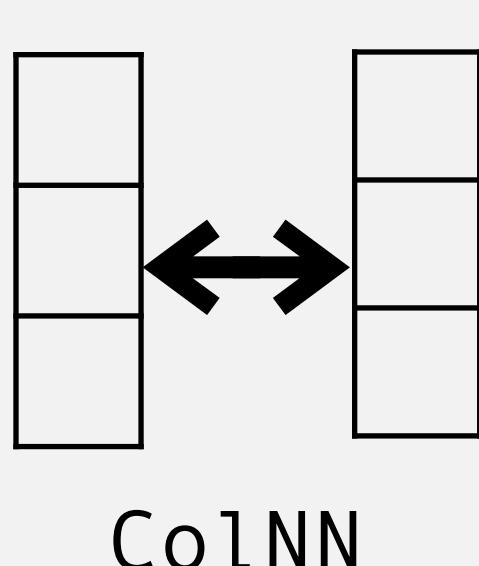
1234



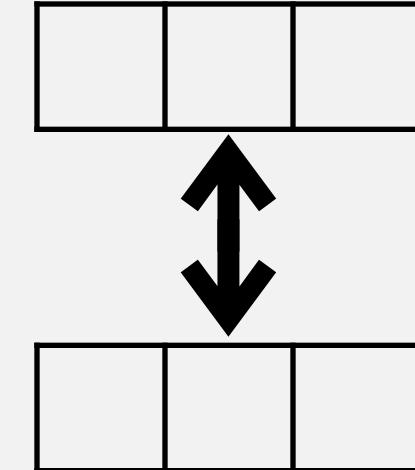
Define:

- DISTANCE
- AVERAGE

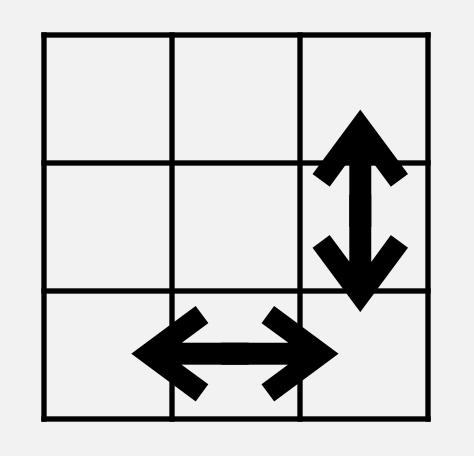
### EstimationMethod



ColNN



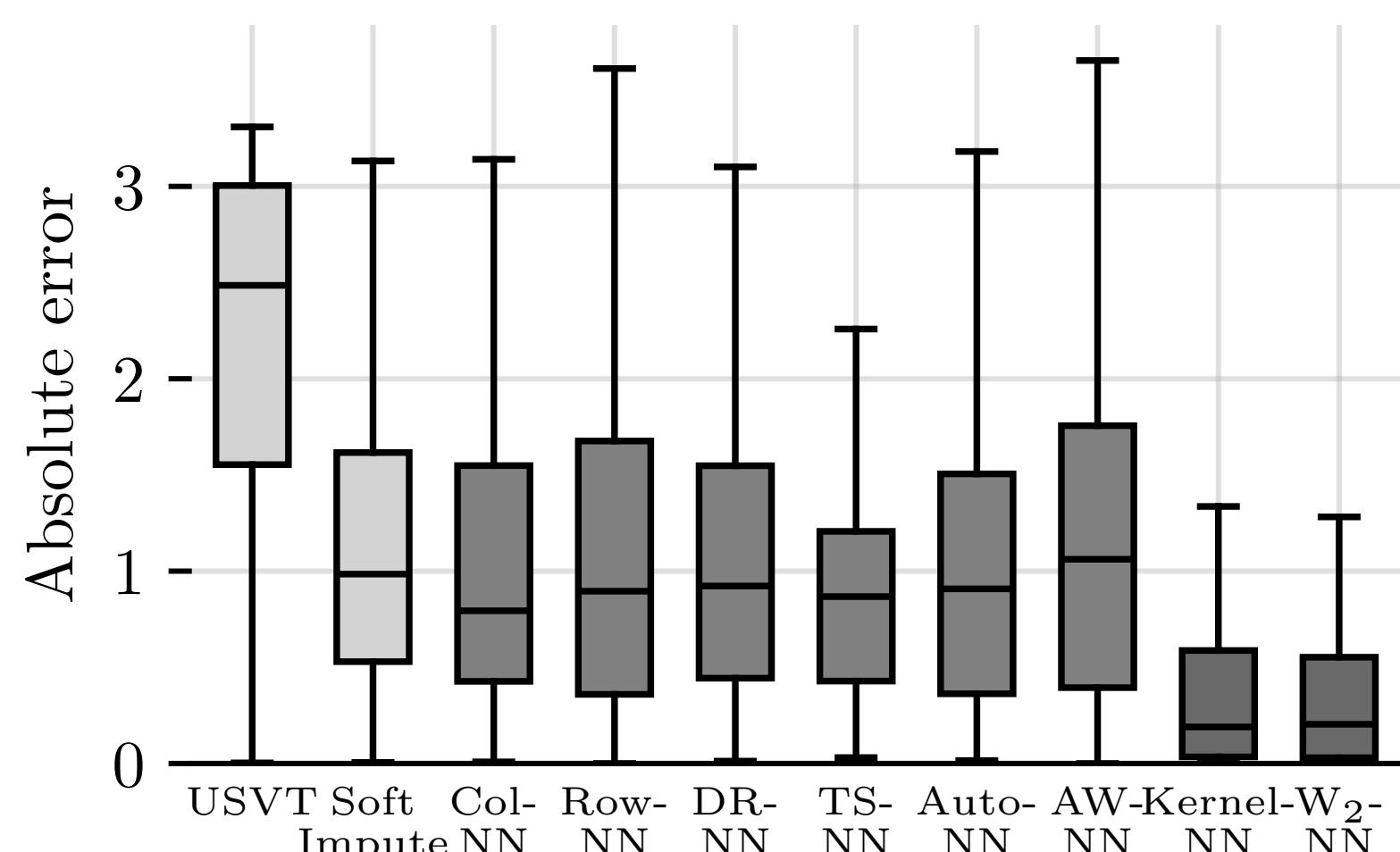
RowNN



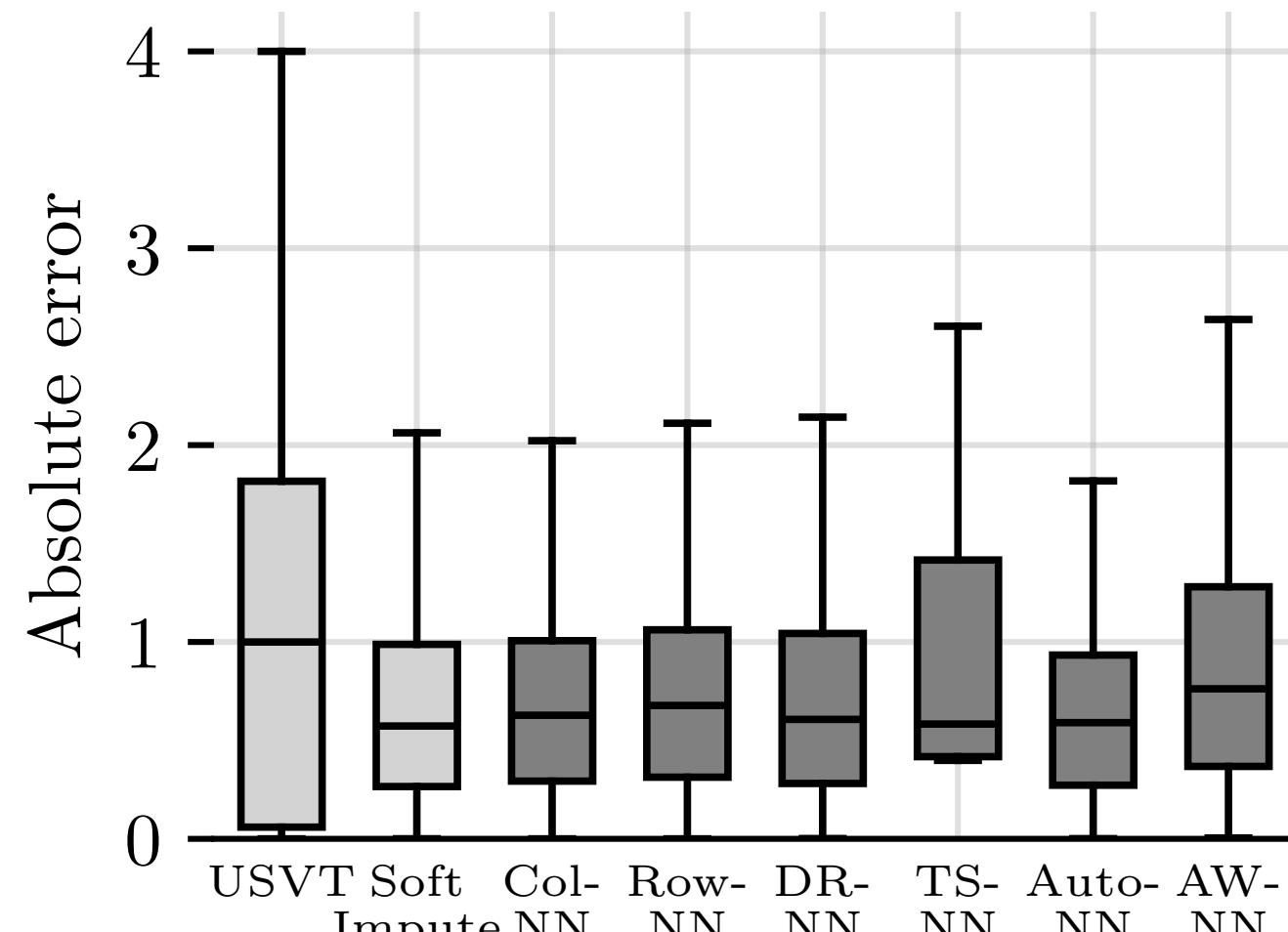
DRNN , TSNN , ...

## N<sup>2</sup>-Bench

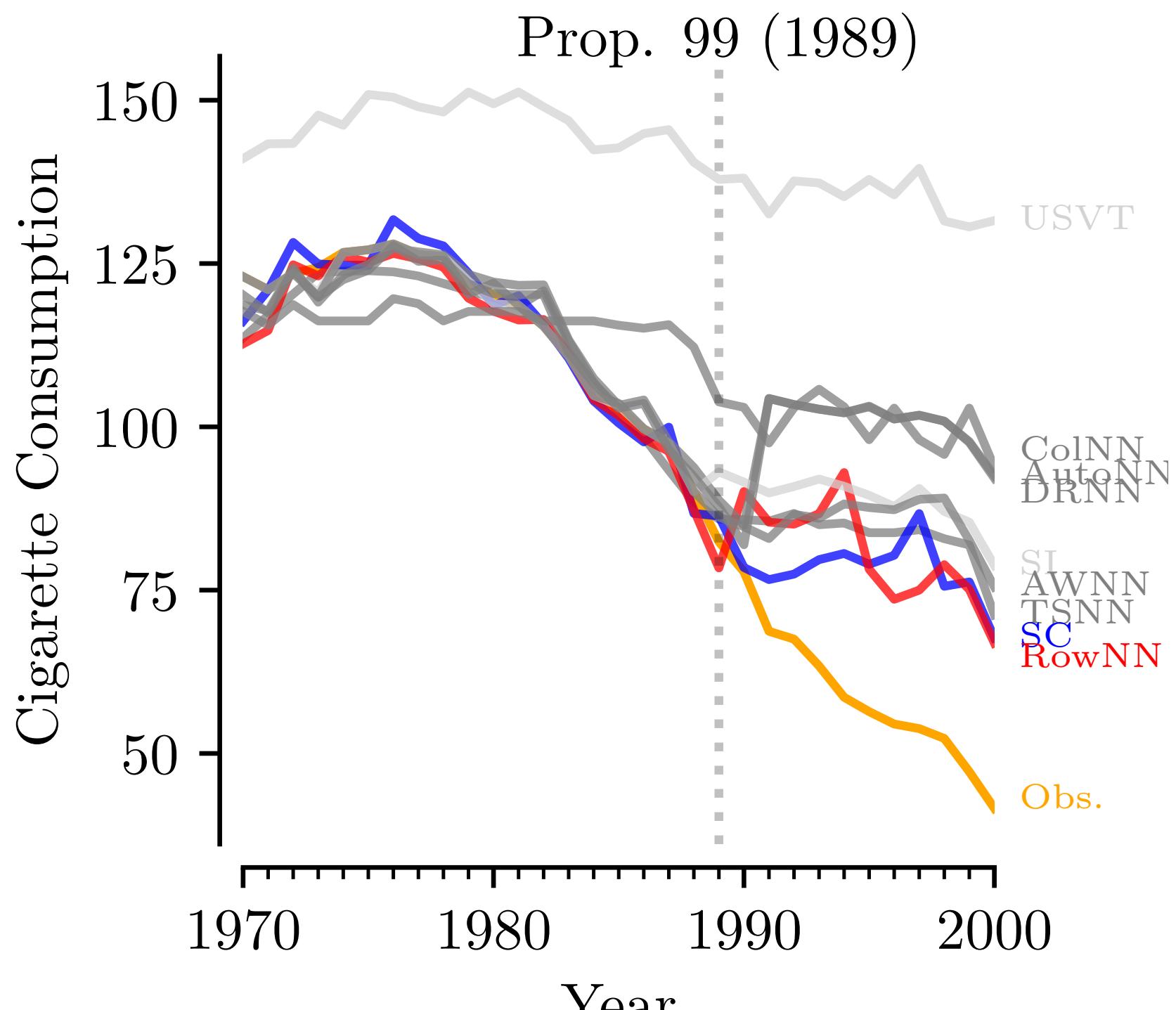
### Personalized Healthcare: Heartsteps Study



### Recommendation Systems: MovieLens



### Counterfactual Inference: California Prop. 99

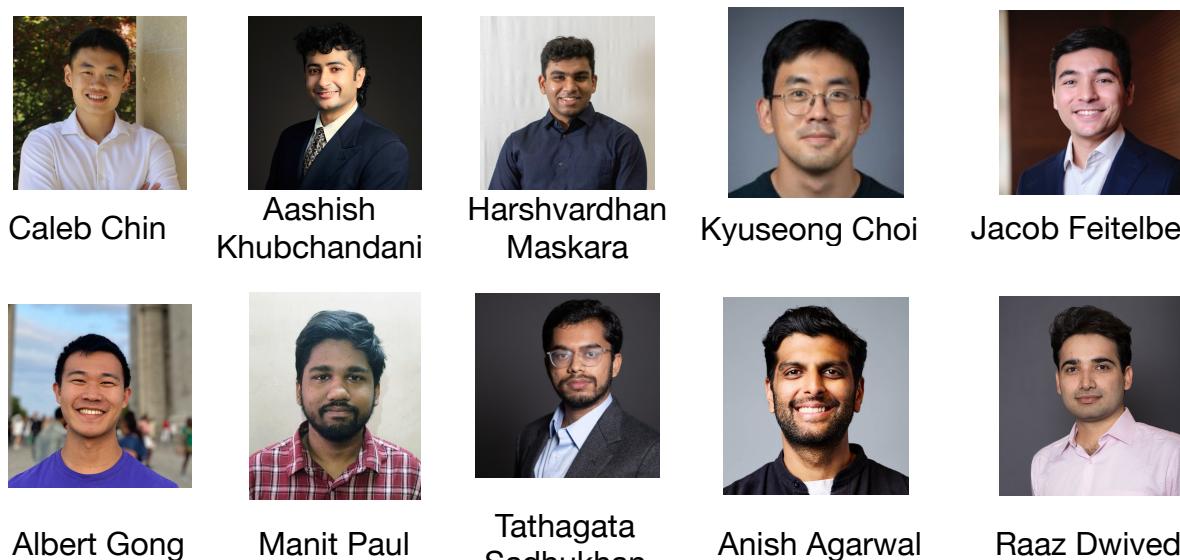


- Across-the-board reductions in error.
- Our findings lead us to introduce AutoNN, which automatically adjusts to the underlying noise level and debiases automatically.
- The modularity of the package and testbed allows for easy prototyping and benchmarking of new NN approaches.

Paper GitHub



# N<sup>2</sup>: A Unified Python Package and Test Bench for Nearest Neighbor-Based Matrix Completion



arxiv.org/abs/2506.04166



github.com/aashish-khub/NearstNeighbors

## Motivation

Nearest neighbor (NN) methods are effective tools for matrix completion but:

- A **unified framework** is needed to support rapid experimentation and development.
- **Standardized benchmarks** are needed for evaluating matrix completion methods across real-world scenarios.

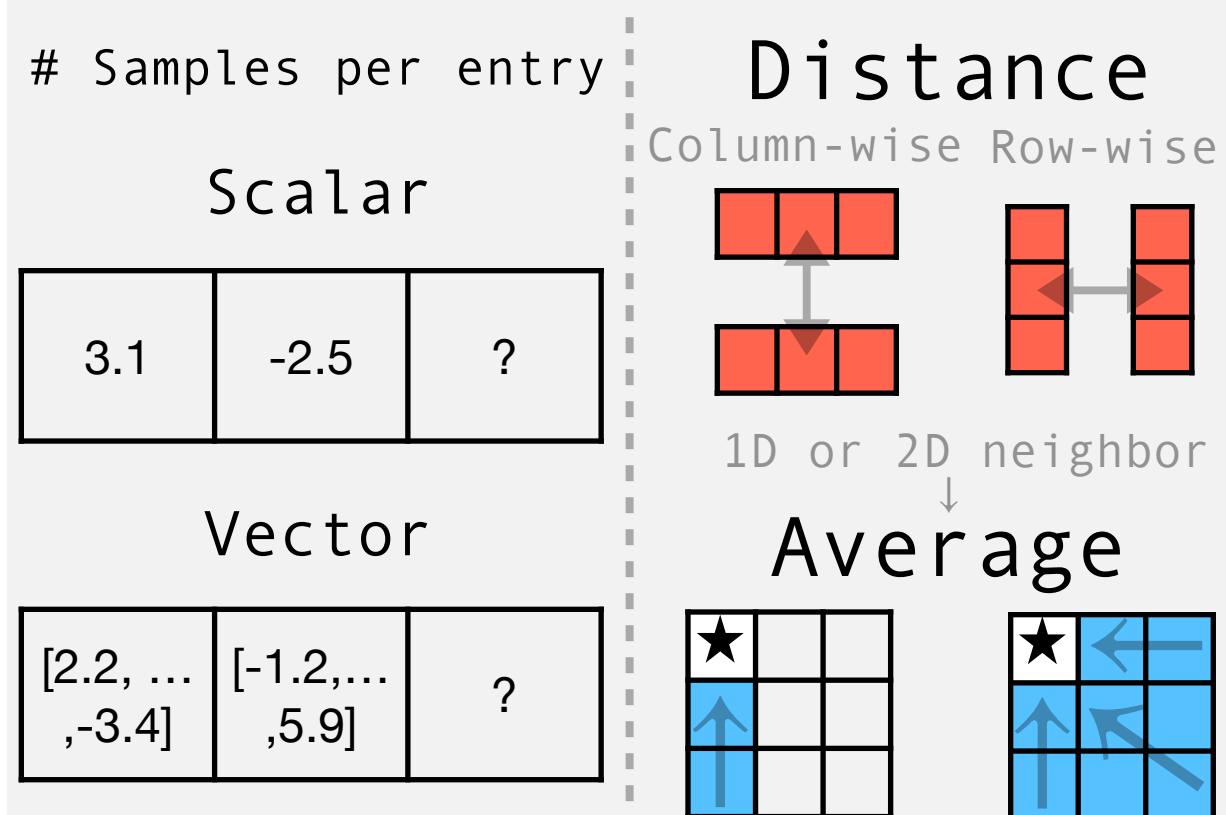
## Solution

N<sup>2</sup> is a **unified Python package** and **testbed** to:

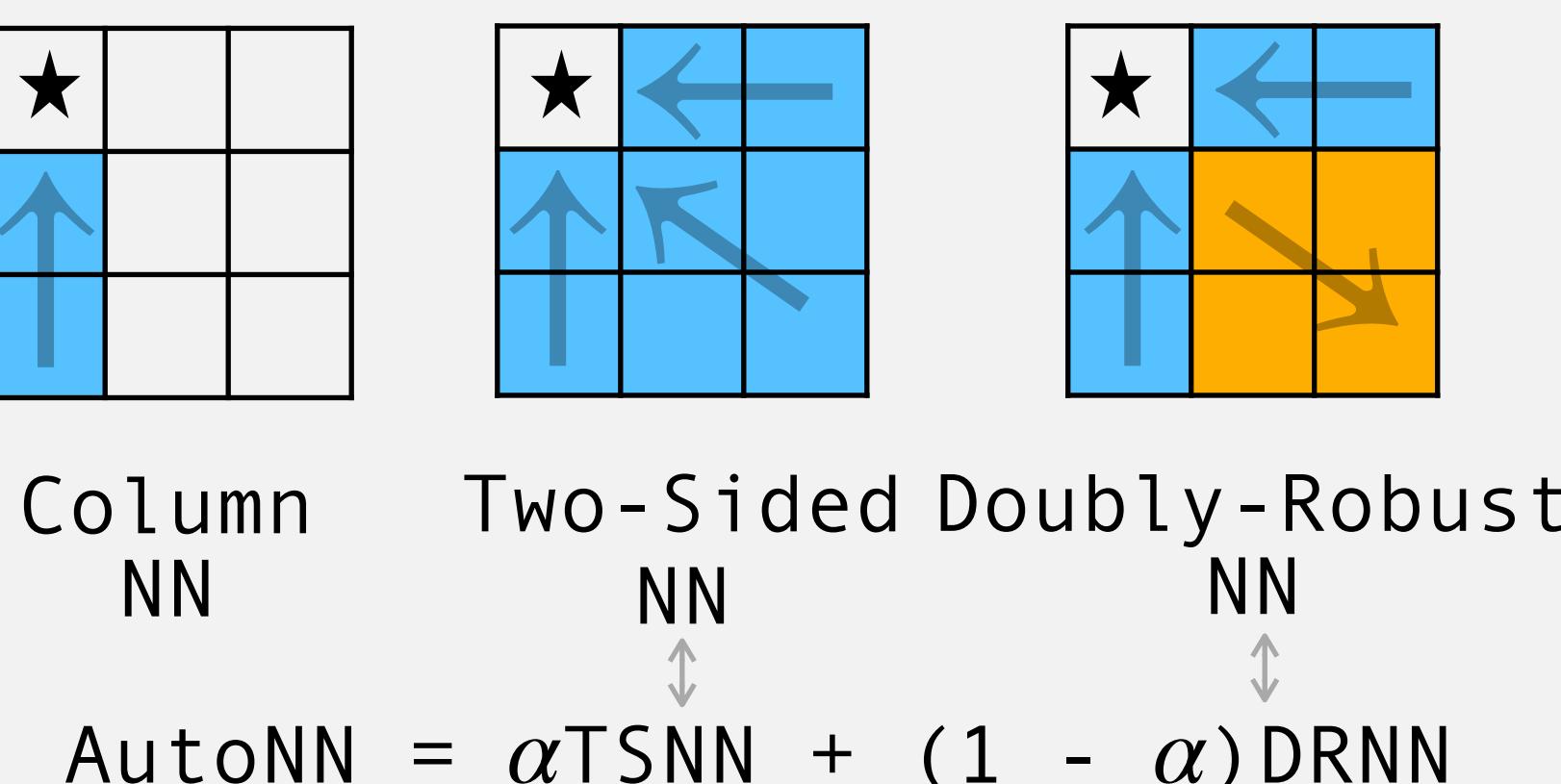
- Consolidate a broad class of NN-based methods through a **modular, extensible interface**.
- Stress-test matrix completion methods on **diverse real-world datasets**, spanning healthcare, recommendation systems, causal inference, and LLM evaluation.

## N<sup>2</sup> Framework

### DataType



### EstimationMethod



```
from nsquared.estimation_methods import RowRowEstimator
from nsquared.data_types import Scalar
from nsquared.fitting_methods import LeaveBlockOutValidation

# Instantiate the NN modules
estimator = RowRowEstimator()
data_type = Scalar()
imputer = NearestNeighborImputer(estimator, data_type)

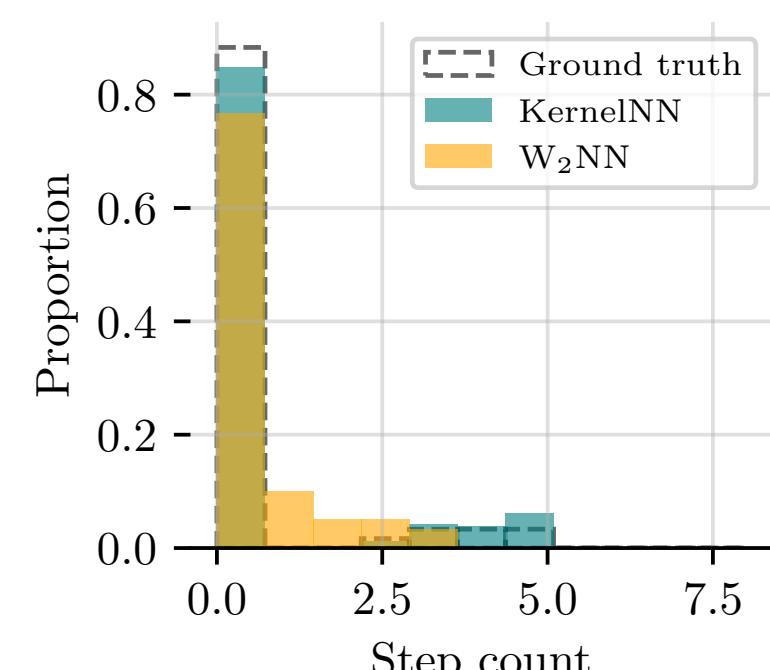
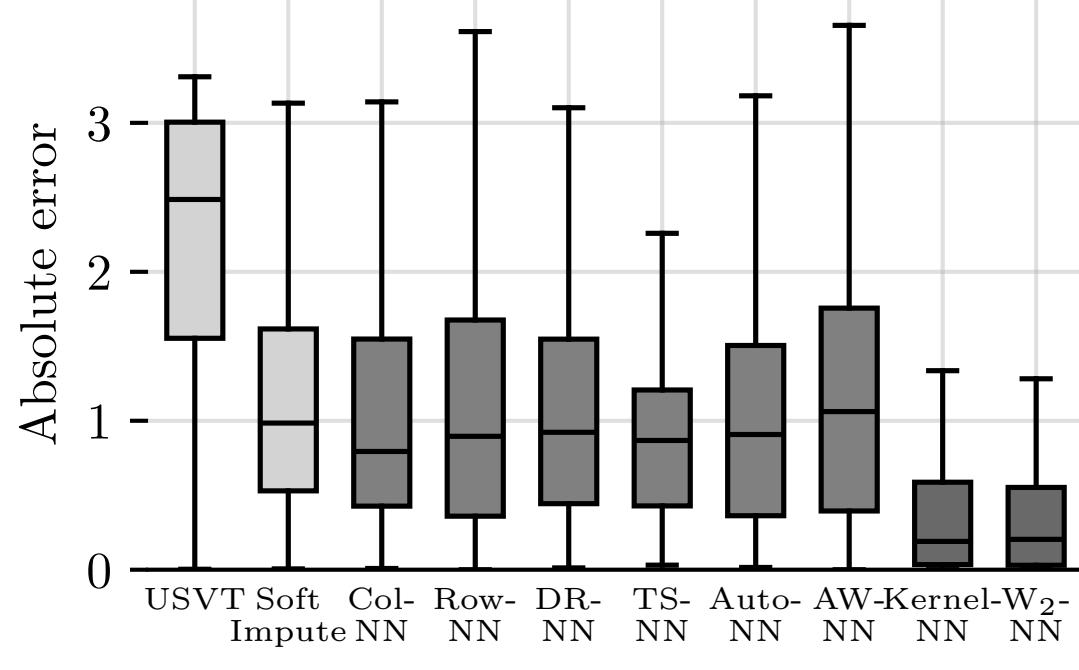
# Fit the NN threshold using cross validation
fitter = LeaveBlockOutValidation(
    block,
    distance_threshold_range=(0,1),
    n_trials=100,
    data_type=data_type
)
fitter.fit(data, mask, imputer)

# Impute the (row, col) value given data and mask matrix
imputed_value = imputer.impute(row, col, data, mask)
```

## N<sup>2</sup>-Bench

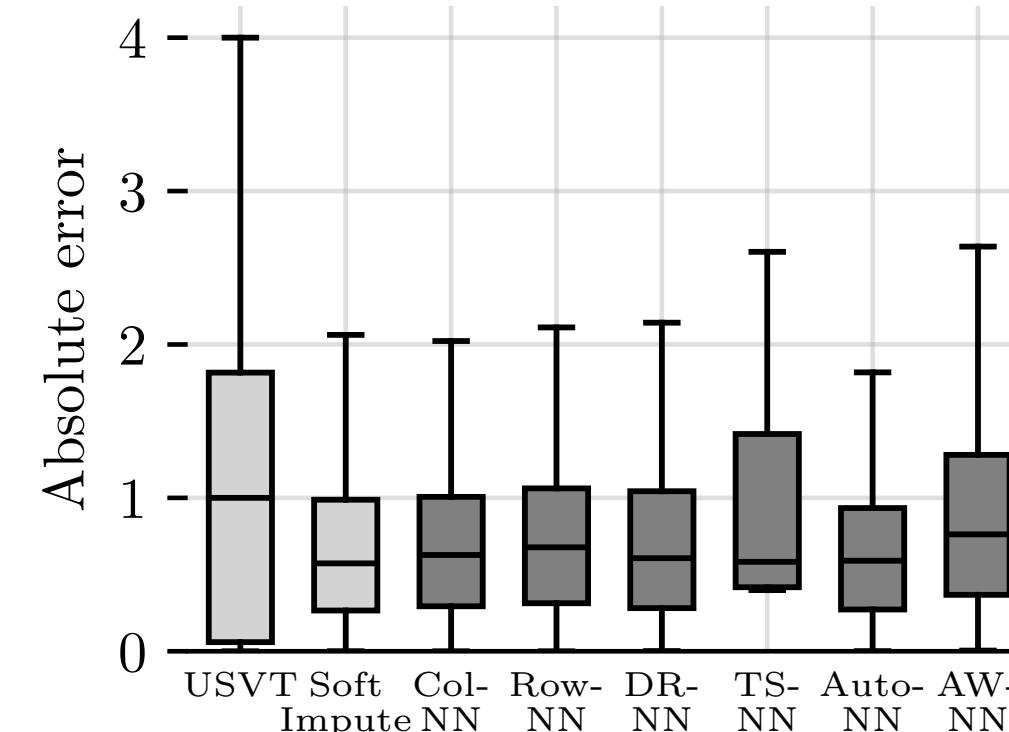
### Personalized Healthcare: HeartSteps

Task: Predict mean participant step count during the hour after a nudge is sent (37 x 200)



### Recommendation Systems: MovieLens

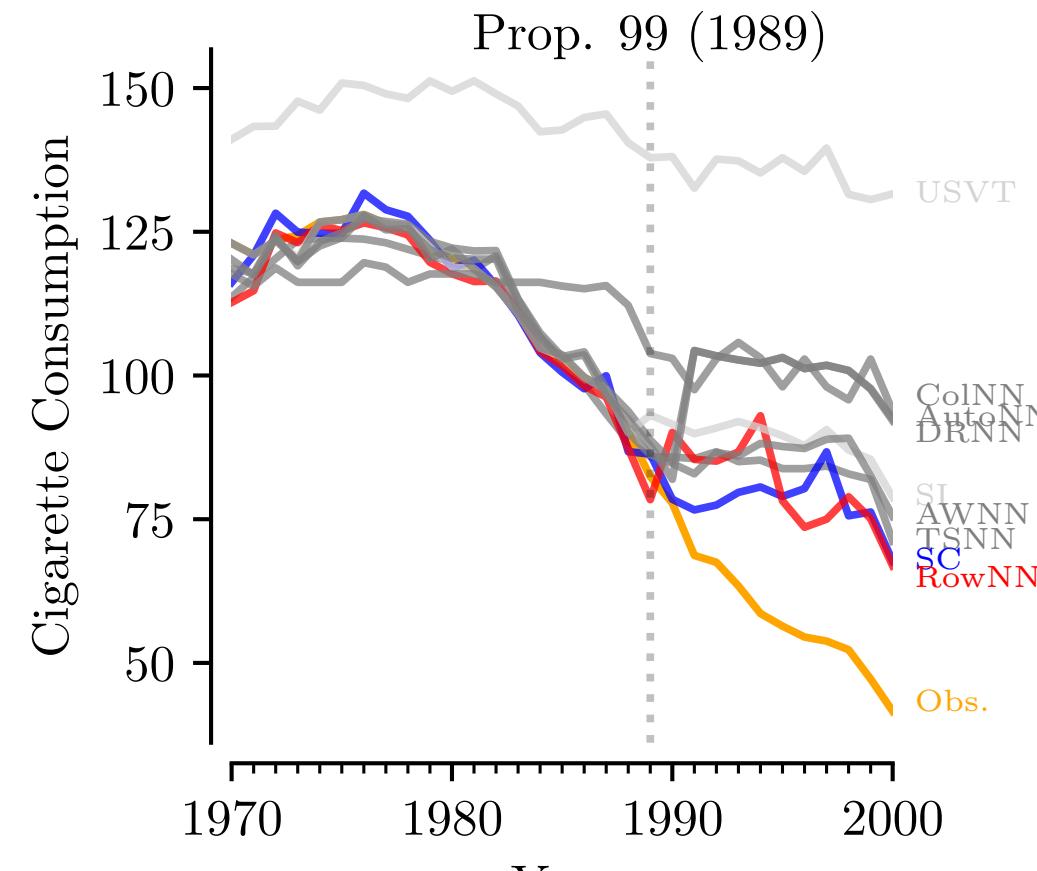
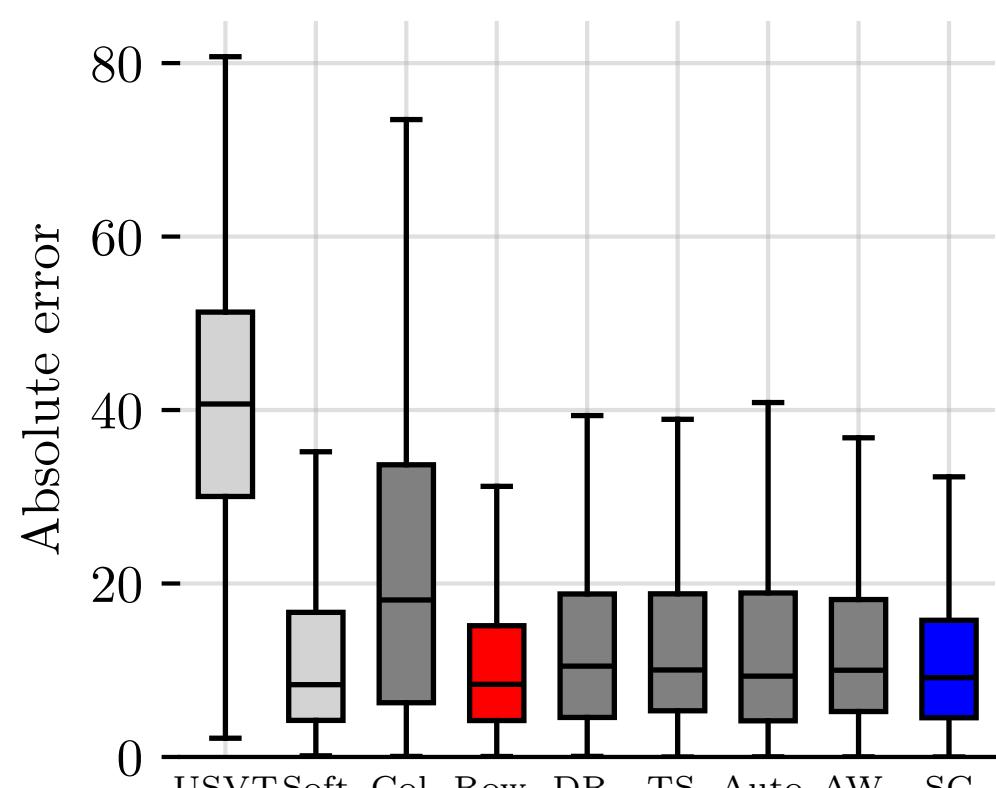
Task: Predict unobserved ratings (6040 x 3952)



1. NN-based techniques perform on-par or better than classical methods in real-world scenarios.
2. N<sup>2</sup> and N<sup>2</sup>-Bench enable easy prototyping and benchmarking of new NN methods.
3. We introduce AutoNN, which adjusts to the underlying noise level and debiases automatically.

### Counterfactual Inference: California Prop. 99

Task: Predict cigarette consumption had a tobacco tax not been enacted (39 x 31)



### Efficient LLM Evaluation: PromptEval

Task: Predict model performance on MMLU tasks without performing all evaluations (15 x 57)

