

# Leon Kloker

Institute of Computational and Mathematical Engineering ♦ Stanford University  
leonkl@stanford.edu ♦ (650) 441-4923 ♦ <https://leonkloker.github.io>

## EDUCATION

<b>Stanford University</b>	September 2022 – June 2024
<i>M.Sc. in Computational and Mathematical Engineering (current GPA 3.9/4.0)</i>	<i>Stanford, CA</i>
<b>University of Stuttgart</b>	October 2017 – August 2021
<i>B.Sc. in Simulation Technology. GPA 1.3 (inverted 4.0 scale, top of the class)</i>	<i>Stuttgart, Germany</i>

## INTERNSHIPS

<b>Scientific ML research intern @ Ansys</b>	June 2023 – September 2023
Developed transformer models in order to time integrate solution snapshots of the 2D incompressible Navier-Stokes equation in the turbulent regime as a benchmark. Also explored the intersection of Fourier Neural Operators and Transformers for Neural Operator Learning. (Tools: PyTorch, PDEs)	
<b>ML research intern @ Sandia National Laboratories</b>	December 2022 – March 2023
Investigated the performance of different model architectures, such as convolutional networks or graph transformers for predicting the probability of a successful trial for of a given single-outcome quantum computation circuit. (Tools: PyTorch Geometric, Quantum Computing)	

## RESEARCH PROJECTS

<b>CUDA parallel computing</b>	April 2023 – June 2023
<b>Course project.</b> A feedforward neural network was implemented from scratch in C++. Forward and backward pass were written as GPU kernels using CUDA with custom optimized kernels for several functions such as general matrix-matrix multiplication or softmax. The model was further parallelized by using MPI for data distribution. (Tools: CUDA, MPI, C++, Nvidia Nsights)	
<b>Computer Vision for precision oncology</b>	December 2022 – March 2023
<b>Integrative Imaging and Molecular Diagnostics lab @ Stanford Medicine.</b> Built ML algorithms for automated cell segmentation and classification in digitalized cancer tissue samples to discover biomarkers that can predict response to Immunotherapy. (Tools: PyTorch, Statistics)	
<b>Convection in saline groundwater</b>	October 2021 – July 2022
<b>Final thesis.</b> A mathematical model describing evaporation of saline water from a porous medium was developed to investigate the flow stability via a linear stability analysis and verify results with a direct numerical simulation as part of a RA position. (Tools: Matlab, PDEs)	

## PUBLICATIONS

***Solution approaches for evaporation-driven density instabilities in a slab of saturated porous media with Carina Bringedal. Physics of Fluids (Vol.34, Issue 9, 2022)***

## AWARDS AND FELLOWSHIPS

Scholarship of the German Academic Exchange Service (DAAD)	2022 – 2024
Simulation Technology valedictorian award	2021
Ferry Porsche Abitur Prize	2017
DPG Abitur Prize	2017

## ADDITIONAL WORK

One Semester of M.Sc. Simulation Technology at University of Stuttgart	2021 – 2022
Teaching Assistant: Engineering Mechanics 1-4, Machine Learning, Linear Algebra for Computing	2018 – 2023
Tennis and fitness coach at Tennis Club Grötzingen and Bernhausen	2018 – 2022

## SKILLS

Language: Fluent in German and English, proficient in French  
Technical: Python, C++, CUDA, MATLAB, Julia, Java, Git, Bash Script, Latex

## INTERESTS & RECENT CLASSES

Mathematical and probabilistic modeling and problem solving  
Using Machine Learning to investigate and solve impactful real-world problems  
Recent classes: *Advanced Software Engineering, Parallel Computing, LeanLaunchpad, Stochastic Methods, Deep Meta Learning, Machine Learning, applied PDEs, Numerical Linear Algebra, Optimization, Cryptocurrencies*

## **Relevant Machine Learning experience:**

- 1) Conducted a project on predicting the elevation map of a terrain from a single satellite image using conditional GANs. The models we employed were a U-Net as generator and a regular CNN as discriminator leading to state-of-the-art results.
- 2) Worked together with Sandia National Laboratory to develop Graph Neural Networks that estimate the probability of a successful trial when running a quantum circuit on a given quantum computer architecture. The most common model I worked with was a Graph Attention Network taking a specific graph embedding of the quantum circuit as input and predicting the probability of success as a single scalar.
- 3) Implemented a Multilayer Perceptron from scratch in C++ with self-written CUDA kernels for forward and backwards pass and Batch parallelization with MPI. Trained the network on ImageNet to verify its correctness and check for computation speed.
- 4) Interned at Ansys over summer to create Transformer Networks that can be used to integrate the solution of the Navier-Stokes equations in time. The physical solution at the previous timesteps is converted to a latent space with a convolutional encoder, then propagated in time autoregressively with a transformer and finally decoded back to the physical space with a convolutional decoder.
- 5) Worked in Prof Ruijiang Li's Stanford medicine lab to build random forests that predict the affiliation of a given patient to a certain treatment group using the values of spatial biomarkers derived from automated single-cell analysis of a digitalized tumor slide.
- 6) Used sequence models to predict SHAPE (reactivity) values of nucleotides in a given RNA sequence. We used a pre-trained BERT-style foundation model trained in a self-supervised manner on a large corpus of existing RNA data, appended our sequence models (GRU, LSTM, Transformer, 1d CNN) to the output of the foundation model and fine-tuned prediction head and the last layers of the foundation model on our own RNA data.

## Assignment description:

As the available data is basically limited to 6 sample populations of varying sizes, I decide to opt for a linear regression model (optionally with L1 or L2 regularization).

Since the amount of cells per population as well as their activation percentages are known, I augment the data by creating all possible non-empty combinations of populations, calculating their activations as a weighted average of the population sizes. By doing this, the amount of samples increases from 6 to  $2^6 - 1$ .

For each of these samples, the mean, standard deviation, skew and kurtosis of all 12 measurements is calculated.

In order to check which features are useful for predicting the activation of the population as well as which form of regularization and regularization parameter yield the best performing model, I perform a variation of 6-fold cross validation. Here, given a held-out population, all combinations of non-held-out populations are used as a train set (size  $2^5 - 1$ ) and the resulting model is validated only on the held-out population (size 1), as validating on combinations with populations used in training would leak train data into the validation set. To find the best model, different regularization parameters for L1 and L2 regularization as well as different combinations of features are used.

The model with the lowest validation MAE is unregularized linear regression using only the means of the 12 measurements as features. This model has an average MAE of 6.39 when predicting the activation percentage of the left-out population during cross-validation.

Moreover, binning the predicted activation percentages of the linear model into 4 classes (0-13%, 13-35%, 35-56% and 56-100%) yields a 100 accuracy when tested on all  $2^6 - 1$  combinations of cell populations.

To further improve the interpretability and generalizability of the model, one could consider which of 12 features are actually pertinent for the model performance, further pruning any features that don't improve validation performance. Moreover, a different type of model such as a (boosted) decision tree could be employed.

Leland Stanford Jr. University  
Stanford, CA 94305  
USA

Graduate Unofficial Transcript - Detailed

Name : Kloker, Leon Hannes  
Student ID : 06531892

Print Date: 01/15/2024

\* Worksheet - For office use by authorized Stanford personnel \*

----- Academic Program -----

Program : Comput & Math Engineering  
09/26/2022 : Computational and Mathematical Engineering (MS)  
Active in Program

----- Beginning of Academic Record -----

2022-2023 Autumn

Course	Cmpt	Title	Attempted	Earned	Grade
CME 211	LEC	SOFTWARE DEVELOPMENT FOR SCIENTISTS AND ENGINEERS Andreas Santucci	3.00	3.00	A
CME 302	LEC	NUMERICAL LINEAR ALGEBRA Eric Darve	3.00	3.00	B+
CS 230	LEC	DEEP LEARNING Andrew Ng; Kian Katanforoosh	3.00	3.00	A+
GR Term GPA	3.866	Term Totals	9.00	9.00	
GR Cum GPA	3.866	Cum Totals	9.00	9.00	

2022-2023 Winter

Course	Cmpt	Title	Attempted	Earned	Grade
CME 291	INS	MASTER'S RESEARCH Kari Hanson	3.00	3.00	A
CME 308	LEC	STOCHASTIC METHODS IN ENGINEERING Peter Glynn	3.00	3.00	A
CME 364A	LEC	CONVEX OPTIMIZATION I Stephen Boyd	3.00	3.00	A
GR Term GPA	4.000	Term Totals	9.00	9.00	
GR Cum GPA	3.933	Cum Totals	18.00	18.00	

2022-2023 Spring

Course	Cmpt	Title	Attempted	Earned	Grade
CME 213	LEC	INTRODUCTION TO PARALLEL COMPUTING USING MPI, OPENMP, AND CUDA Eric Darve	3.00	3.00	B+
CME 500	COL	DEPARTMENTAL SEMINAR Sanjay Saigal	1.00	1.00	S
ENGR 245	LEC	THE LEAN LAUNCHPAD: GETTING YOUR LEAN STARTUP OFF THE GROUND Steve Blank; Steve Weinstein	3.00	3.00	A
MATH 237A	LEC	TOPICS IN FINANCIAL MATH: MARKET MICROSTRUCTURE AND TRADING ALGORITHMS George Papanicolaou	3.00	3.00	A
GR Term GPA	3.766	Term Totals	10.00	10.00	
GR Cum GPA	3.877	Cum Totals	28.00	28.00	

2022-2023 Summer

Course	Cmpt	Title	Attempted	Earned	Grade
CME 390	INS	CURRICULAR PRACTICAL TRAINING Gianluca Iaccarino	1.00	1.00	S
GR Term GPA	0.000	Term Totals	1.00	1.00	
GR Cum GPA	3.877	Cum Totals	29.00	29.00	

2023-2024 Autumn

Course	Cmpt	Title	Attempted	Earned	Grade
CME 303	LEC	PARTIAL DIFFERENTIAL EQUATIONS OF APPLIED MATHEMATICS Leonid Ryzhik	3.00	3.00	A
CS 251	LEC	CRYPTOCURRENCIES AND BLOCKCHAIN TECHNOLOGIES Dan Boneh	3.00	3.00	A
CS 330	LEC	DEEP MULTI-TASK AND META LEARNING Chelsea Finn	3.00	3.00	A
CS 337	SEM	AI-ASSISTED CARE Amit Kaushal; Arnold Milstein Devanshi Dash; Ehsan Adeli Fei-Fei Li; Kevin Schulman	1.00	1.00	CR
GR Term GPA	4.000	Term Totals	10.00	10.00	
GR Cum GPA	3.908	Cum Totals	39.00	39.00	

Information must be kept confidential and must not be disclosed to other parties without the written consent of the student.