Ralf **Herbrich**

APPLIED SCIENCE EXPERT & LEADER IN AI

✉ ralf@herbrich.me | ⌂ herbrich.me | ⌨ rherbrich74 | in ralf-herbrich-28a8324 | 🐦 @rherbrich | 🎓 Ralf Herbrich

Summary

Experienced AI and machine learning expert and organizational leader with more than two decades of industrial and academic experience. Research interests include approximate computing, Bayesian inference and decision making, game theory, information retrieval, natural language processing, computer vision, distributed systems, learning theory and knowledge representation and reasoning. Published over 80 conference and journal papers in these fields with over 15,000 citations and h-index 55. Experience in advisory and board roles through being a founding member of Partnership on AI. One of the inventors of the Drivatars™ system in the Forza Motorsport series as well as the TrueSkill™ ranking and matchmaking system in Xbox 360 Live. Co-invented the click-prediction technology used in bing's online advertising system (adPredictor) which is used widely in the industry.

I want my work to have a positive impact to many people's lives. I particularly enjoy the application of principled research to large-scale real-world problems. I enjoy learning new things every day and getting to work with (new) people. I do not accept existing scientific boundaries and think that the largest breakthroughs will be made at the intersection of existing disciplines.

Education

Amazon
LEADER OF LEADERS   *Seattle, US*
January 2015 – December 2015
- Senior executive leadership training program for high-performing Amazon employees (top 1%).

Microsoft
SENIOR LEADER BENCH   *Redmond, US*
October 2007 – October 2008
- Executive management program for high-performing Microsoft employees (top 4%).

Darwin College Cambridge
RESEARCH FELLOW   *Cambridge, UK*
October 2000 - October 2004
- Microsoft Research Fellow with focus on *philosophy of Science*

Australian National University
VISITING FELLOW   *Canberra, Australia*
February 2000 - April 2000
- Collaborative work Dr. R. C. Williamson on *PAC-Bayesian margin bounds and online-learning*

University of Bristol
VISITING FELLOW   *Bristol, UK*
February 1999 - April 1999
- Collaborative work Dr. C. Campbell and Prof. J. Shawe-Taylor on *Bayes point machines* and *PAC-Bayesian learning theory*

Technical University Berlin
PHD IN COMPUTER SCIENCE (DR. RER. NAT.)   *Berlin, Germany*
1997 - 2000
- PhD Thesis: *Learning Linear Classifiers: Theory and Algorithms*, supervised by Prof. Ulrich Kockelkorn
- Final grade: summa cum laude
- The thesis is also published as a book in the *Adaptive Computing and Machine Learning Series* of The MIT Press

Technical University Berlin
MSC IN COMPUTER SCIENCE (DIPLOM)   *Berlin, Germany*
1992 - 1997
- Master Thesis: *Segmentation with Gabor filters for the induction of structural classifiers on images*
- Final grade: excellent

Experience

Hasso Plattner Institute & University of Potsdam
PROFESSOR, ARTIFICIAL INTELLIGENCE & SUSTAINABILITY   *Potsdam, Germany*
2022 - today
- Basic research in the area of energy-efficient AI algorithms and the use of AI algorithms for energy storage solutions (e.g., batteries)
- Teaching of BSc courses in programming, mathematics and machine learning

betteries AMPS GmbH
MACHINE LEARNING & AI SCIENTIFIC ADVISOR   *Berlin, Germany*
2022 - today
- Estimation of the state of health of electric vehicles batteries and optimizing their lifetime during second life

DATAI Consulting DAC GmbH

Founder and CEO

Berlin, Germany

2022 - today

- Consulting of companies at all stages on application of data science and artificial intelligence for their business
- Talks and courses for applied science and machine learning for a broad audience
- Mentorship of C-level executives on technical and scientific leadership

Zalando

Senior Vice President, Builder Platform & AI

Berlin, Germany

2020 - 2021

- Leading Zalando Research, Centers of AI Excellence, Digital Foundation (Builder Infrastructure and Data Management & Intelligence) as well as Size & Fit and Style Advice (Zalon) and Zalando's internal skill development team Tech Academy.
- Company-wide sponsor for all applied science as well as sponsor and trainer for a range of mental health programs.

Partnership on AI (PAI)

Founding Member of the Board of Directors

San Francisco, US

2016 - 2020

- Represented Amazon as one of the founding members of PAI together with senior colleagues from Microsoft, Google/DeepMind, Facebook, and IBM, as well as representatives of civil society organisations.
- Helped PAI define a long-term strategy and was Head of the Finance Committee.

Amazon

Director of Machine Learning (Core ML & Core AI)

Berlin, Germany

2012 - 2019

- Leading Amazon's worldwide machine learning team working on forecasting, machine translation, machine learning services, robotics, computer vision and machine learning education and community programs.
- Leading four acquisitions including BodyLabs and supported acquisition of Safaba.
- Site-lead and Managing Director for Amazon Development Center Germany GmbH growing from 0 to nearly 1,000 employees in 7 years.

Facebook

Engineering Manager

Menlo Park, US

2011 - 2012

- Built a large-scale, real-time, distributed probabilistic inference system that was used to drive the click-through prediction for all advertisements
- Worked on efficient-to-evaluate value models to allow optimizing for long-term economic value.

Microsoft/Microsoft Research

Principal Development Manager, Director, Principal Researcher, PostDoc

Cambridge, UK

2000 - 2011

- Developed frequentist guarantees for Bayesian algorithms, algorithmic luckiness, and the usefulness of unlabelled data (PostDoc: 2000 - 2001).
- Worked in the areas of Bayesian inference, reinforcement learning, application of Bayesian methods in computer games (Researcher 2001 - 2005).
- Co-invented the Drivatars systems used in the Forza Motorsport racing game series (Researcher: 2001-2004).
- Co-invented and developed TrueSkill, a ranking and matchmaking system for Microsoft's online gaming system Xbox Live with about 20M subscriptions. (Senior Researcher: 2007)
- Started and co-led the Applied Games group (Team Lead: 2006-2008) and Online Services and Advertising group. (Team Lead: 2008-2009)
- Co-invented and developed AdPredictor, a Click-through rate prediction system powering Paid Search advertising in Microsoft's search engine Bing with over 1B page views per month. (Principal Researcher: 2009)
- Built a 10-person applied science team (Director: 2009-2011).
- Application of computational intelligence technologies for agile development of new social experiences (Director: 2009-2011).
- Prototyping and enabling personalized experiences across Microsoft's Online Services Division, including Bing Mobile, Bing News, Bing Web and AdCenter (Principal Development Manager: 2011).

Pixfind.com Net Technologies

Co-founder

Berlin, Germany

1999 - 2000

- Developed text-based image retrieval algorithms for an early professional photography marketplace application.
- Helped secure 2.5M Euro funding for Pixfind.com Net Technologies.

Technical University of Berlin

Teaching Assistant

Berlin, Germany

1997 - 2000

- Teaching of Statistics (tutorials), theory of Machine Learning (seminars) and business mathematics (lecture).
- Supervised Diploma thesis of Yang Gang: *A comparison of learning algorithms on automobile recording data*.

Technical University of Berlin

Research Assistant

Berlin, Germany

1996 - 1997

- Non-teaching position in the project "Modern Methods of ML". Implemented a fully functional logical learning system ($\approx$30 000 LOC).

Skills

Computer Science	Algorithms, Distributed Systems, Databases, Programming Languages
Programming	C, C++, C#, Python, Java, OCaml, F#
Leadership	Mechanism Design, Strategy development, Operational planning
Languages	German, English

Services

Year	Role	Location
2018 -	**Editor**, Harvard Data Science Reviews	*Virtual*
2010 -	**Series Editor**, Machine Learning & Pattern Recognition	*Chapman & Hall*
2022	**General Co-Chair**, AI-ML Systems	*Bangalore, India*
2020	**Workshops Co-Chair**, MLSys	*Austin, US*
2018	**Demonstrations Co-Chair**, NeurIPS	*Montreal, US*
2017	**Workshop Co-Chair**, NIPS	*Long Beach, US*
2016	**Workshop Chair**, NIPS	*Barcelona, Spain*
2015	**Tutorial Chair**, NIPS	*Montreal, Canada*
2008	**Demonstrations Chair**, NIPS	*Vancouver, Canada*
2006	**Program Committee Member**, COLT	*Pittsburgh, US*
2005	**Workshops Co-Organizer "Learning to Rank"**, NIPS	*Vancouver, Canada*
2005	**Program Committee Member**, COLT	*Bertinoro, Italy*
2003	**Track Co-chair "Support Vector Machines and Kernel Methods"**, ICANN	*Istanbul, Turkey*
2002	**Program Committee Member**, COLT	*Sydney, Australia*
1999	**Workshops Co-Organizer "Using Unlabelled Data For Supervised Learning"**, NIPS	*Denver, US*
1998	**Workshops Co-Organizer "Learning on Relational Data Representations"**, NIPS	*Denver, US*

Selected Publications

CRISP: A Probabilistic Model for Individual-Level COVID-19 Infection Risk Estimation Based on Contact Data

arxiv

Ralf Herbrich, Rajeev Rastogi, Roland Vollgraf, *2020*

- Present a probabilistic graphical model for COVID-19 infection spread through a population based on the SEIR model. The micro-level model keeps track of the infection state for each individual at every point. Developed a Monte Carlo EM algorithm to infer contact-channel specific infection transmission probabilities using block-Gibbs sampling to draw samples of the latent infection status of each individual over the entire time period of analysis.

Practical Lessons from Predicting Clicks on Ads at Facebook

SIGKDD Conference on Knowledge Discovery and Data Mining

Xinran He, Junfeng Pan, Ou Jin, Tianbing Xu, Bo Liu, Tao Xu, Yanxin Shi, Antoine Atallah, Ralf Herbrich, Stuart Bowers, Joaquin Quiñonero Candela *2014*

- Introduce a model which combines decision trees with logistic regression, outperforming either of these methods on its own by over 3% on the task of click-through rate prediction of ads. Picking the optimal handling for data freshness, learning rate schema and data sampling improve the model slightly, though much less than adding a high-value feature, or picking the right model to begin with. Describes the production system for click-through rate prediction on advertisements at Facebook in 2012.

Sociable Killers: Understanding Social Relationships in an Online First-Person Shooter Game

ACM Conference on Computer Supported Cooperative Work

Yan Xu, Xiang Cao, Abigail Sellen, Ralf Herbrich, Thore Graepel *2011*

- Describes what social relationships mean under the context of online First-Person Shooter (FPS) games, how these relationships influence game experience, and how players manage them. Combines qualitative interview and quantitative game log data, and finds that despite the gap between the non-persistent game world and potentially persistent social relationships, a diversity of social relationships emerge and play a central role in the enjoyment of online FPS games.

Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine

International Conference on Machine Learning

Thore Graepel, Joaquin Quiñonero Candela, Thomas Borchert, Ralf Herbrich *2010*

- Describes a Bayesian click-through rate prediction algorithm used for Sponsored Search in Microsoft's Bing search engine. The algorithm is based on a probit regression model that maps discrete or real-valued input features to probabilities which maintains Gaussian beliefs over weights of the model and performs Gaussian online updates derived from approximate message passing. Scalability of the algorithm is ensured through a principled weight pruning procedure and an approximate parallel implementation. This algorithm is used in many large-scale click-through rate systems today (e.g., Amazon's Gateway).

Matchbox: Large Scale Online Bayesian Recommendations

International Conference on World Wide Web

David Stern, Ralf Herbrich, Thore Graepel *2009*

- Presents a probabilistic model for generating personalised recommendations of items to users of a web service. The proposed system makes use of content information in the form of user and item meta data in combination with collaborative filtering information from previous user behavior in order to predict the value of an item for a user. Efficient inference is achieved by approximate message passing involving a combination of Expectation Propagation (EP) and Variational Message Passing. System was used in the Xbox Live Marketplace.

TrueSkill Through Time: Revisiting the History of Chess

Pierre Dangauthier, Ralf Herbrich, Tom Minka, Thore Graepel

- Extension of the Bayesian skill rating system TrueSkill to infer entire time series of skills of players by smoothing through time instead of filtering. Presents an analysis of the skill curves of important players in the history of chess over the past 150 years. Results include plots of players' lifetime skill development as well as the ability to compare the skills of different players across time. Results indicate that (a) the overall playing strength has increased over the past 150 years, and (b) that modelling a player's ability to force a draw provides significantly better predictive power.

Trueskill™: A Bayesian Skill Rating System

Ralf Herbrich, Tom Minka, Thore Graepel

- Describes the Bayesian skill rating system TrueSkill, a generalisation of the well known Chess rating system Elo to multiple teams while also tracking posterior uncertainty in the skill estimates. TrueSkill was later implemented and used in the blockbuster video game Halo 3 as well as more widely in Microsoft's XBox Live online gaming service.

Bayesian Pattern Ranking for Move Prediction in the Game of Go

David Stern, Ralf Herbrich, Thore Graepel

- A probabilistic approach to the problem of learning to predict moves in the board game of Go from game records of expert players which can serve as an efficient stand-alone Go player and a move selector for game tree search. The system is trained on 181,000 expert games and shows excellent prediction performance as indicated by its ability to perfectly predict the moves made by professional Go players in 34% of test positions. A similar algorithm was used for move selection in game tree search in AlphaGo.

Kernel Methods for Measuring Independence

Arthur Gretton, Ralf Herbrich, Alex Smola, Olivier Bousquet, Bernhard Schölkopf

- First introduction of the constrained covariance and the kernel mutual information to measure the degree of independence of random variables. The performance of the kernel dependence functionals in measuring independence is verified in the context of independent component analysis.

Learning to Fight

Thore Graepel, Ralf Herbrich, Julian Gold

- Applications of reinforcement learning to the problem of finding good policies for a fighting agent in a commercial computer game. One of the first applications of neural network function approximators for learning the value function. Demonstrates that the learning agent finds interesting (and partly near-optimal) policies in accordance with the reward functions provided.

Invariant Pattern Recognition by Semidefinite Programming Machines

Thore Graepel, Ralf Herbrich

- Develop a new framework for learning linear classifiers under known transformations based on semidefinite programming. The Semidefinite Programming Machine (SDPM) is able to find a maximum margin hyperplane when the training examples are polynomial trajectories instead of single points. The solution is found to be sparse in dual variables and allows to identify those points on the trajectory with minimal real-valued output as virtual support vectors.

A PAC-Bayesian Margin Bound for Linear Classifiers

Ralf Herbrich, Thore Graepel

- Presents a bound on the generalisation error of linear classifiers in terms of a refined margin quantity on the training sample using a PAC-Bayesian framework. The proof is based on geometrical arguments in the space of linear classifiers. The bound was an exponential improvement of the so-far tightest margin bound.

Algorithmic Luckiness

Ralf Herbrich, Robert Williamson

- Classical statistical learning theory studies the generalisation performance of machine learning algorithms rather indirectly. This paper is one of the first to study learning algorithms more directly and in a way that allows to exploit the serendipity of the training sample. Demonstrate how the resulting framework relates to the VC, luckiness and compression frameworks, together with an application of this framework to the maximum margin algorithm for linear classifiers which results in a bound that exploits the margin, the sparsity of the resultant weight vector, and the degree of clustering of the training data in feature space.

Learning Kernel Classifiers: Theory and Algorithms

Ralf Herbrich

- Textbook which contains all the derivations and proofs of basic kernel algorithms including support vector machines, relevance vector machines, Gaussian processes as well as PAC, PAC-Bayesian and Algorithmic Stability bounds.

Bayes Point Machines

RALF HERBRICH, THORE GRAEPEL, COLIN CAMPBELL *2001*

- Describes the Bayes point machine as a (set of) algorithm(s) which approximates the Bayes-optimal decision by the centre of mass of version space. Experimentally, Bayes point machines consistently outperform support vector machines on both surrogate data and real-world benchmark data sets. Demonstrate that the real-valued output of single Bayes points on novel test points is a valid confidence measure and leads to a steady decrease in generalisation error when used as a rejection criterion.

A Generalized Representer Theorem

BERNHARD SCHÖLKOPF, RALF HERBRICH, ALEX SMOLA *2001*

- Generalizes Whaba's classical representer theorem to a larger class of regularizers and empirical risk terms, and give a self-contained proof utilizing the feature space associated with a kernel. The result shows that a wide range of problems have optimal solutions that live in the finite dimensional span of the training examples mapped into feature space, thus enabling to carry out kernel algorithms independent of the (potentially infinite) dimensionality of the feature space.

Unbiased Assesment of Learning Algorithms

TOBIAS SCHEFFER, RALF HERBRICH *1997*

- Many researchers conduct experiments on benchmark datasets, but since most learning algorithms have domain-specific parameters, it is a popular custom to adapt these parameters to obtain a minimal error rate on the test set. Demonstrate that this results in a bias and used a communication channel model to quantify this bias. One of the first works to quantify the bias of wrong experimentation.