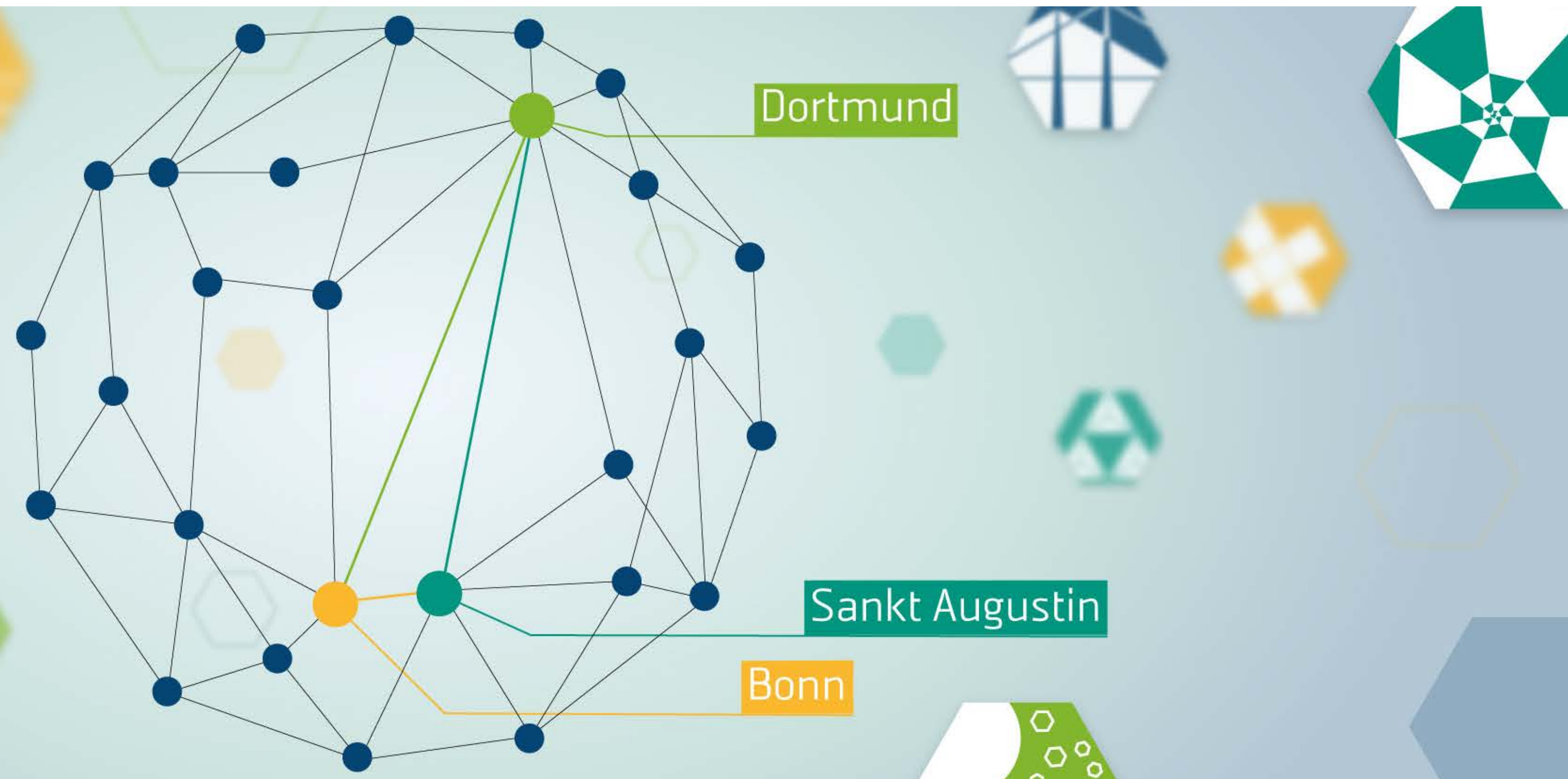


AI AND BIG DATA RESEARCH ISSUES AND THE NEW GERMAN COMPETENCE CENTRES FOR MACHINE LEARNING IN GERMANY



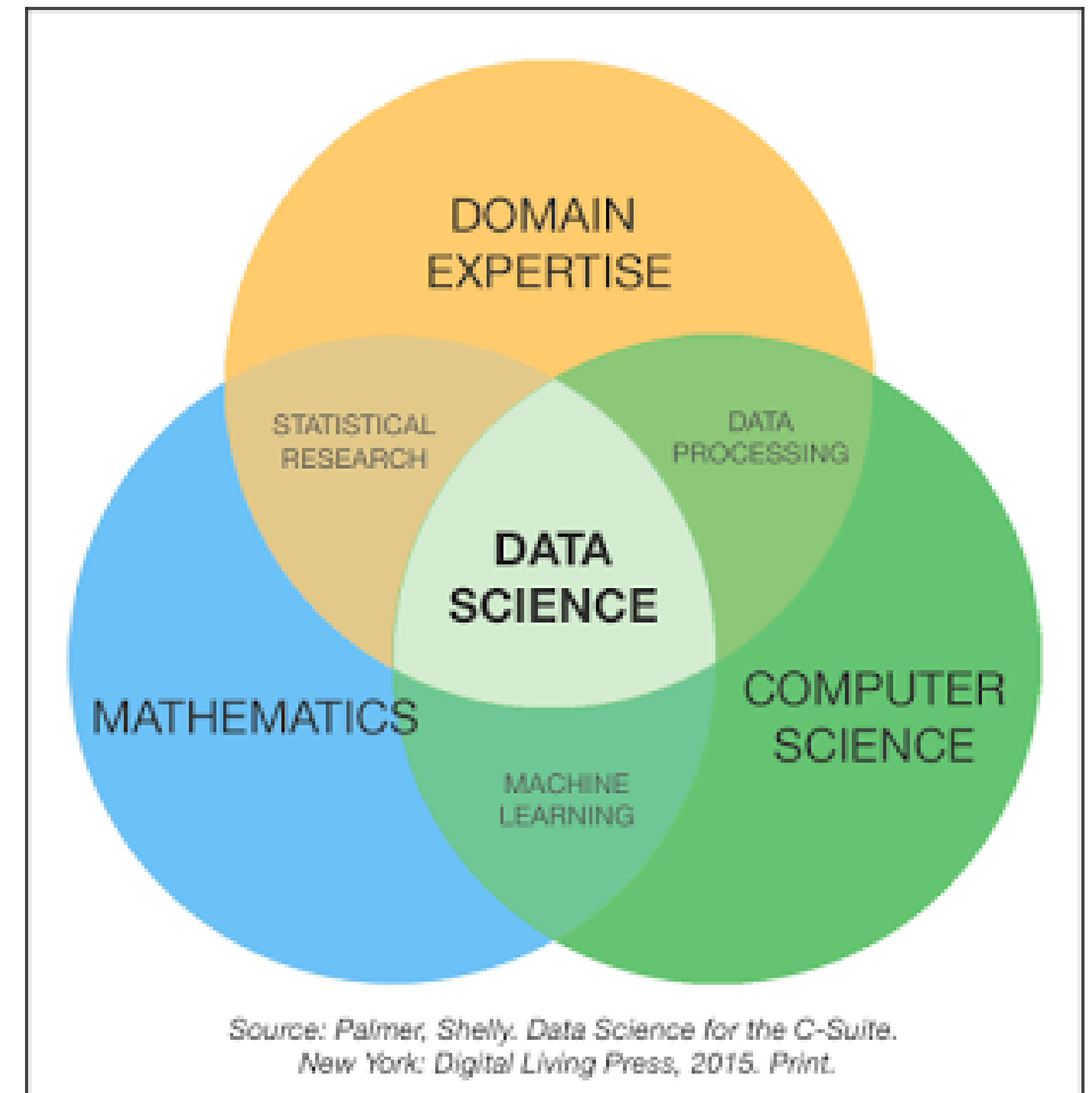
OVERVIEW

- Data Science and Artificial Intelligence
- Machine Learning
- Competence Centers in Germany
- Machine Learning Challenges



DATA SCIENCE

- Scientific experiments produce big data.
- Big data analysis delivers data summaries and predictions.
- Foundations are:
 - Sciences
 - Computer architectures: GPU, FPGA, Multicore, ...
 - Software frameworks: streams, Hadoop, SQL, ...
 - Algorithms
 - Data bases
 - Statistical methods



PARADIGM OF EMPIRICAL SCIENCE

Medicine

- Personalized therapy based on genetic data
- Explanation of processes, e.g. cancer

Physics

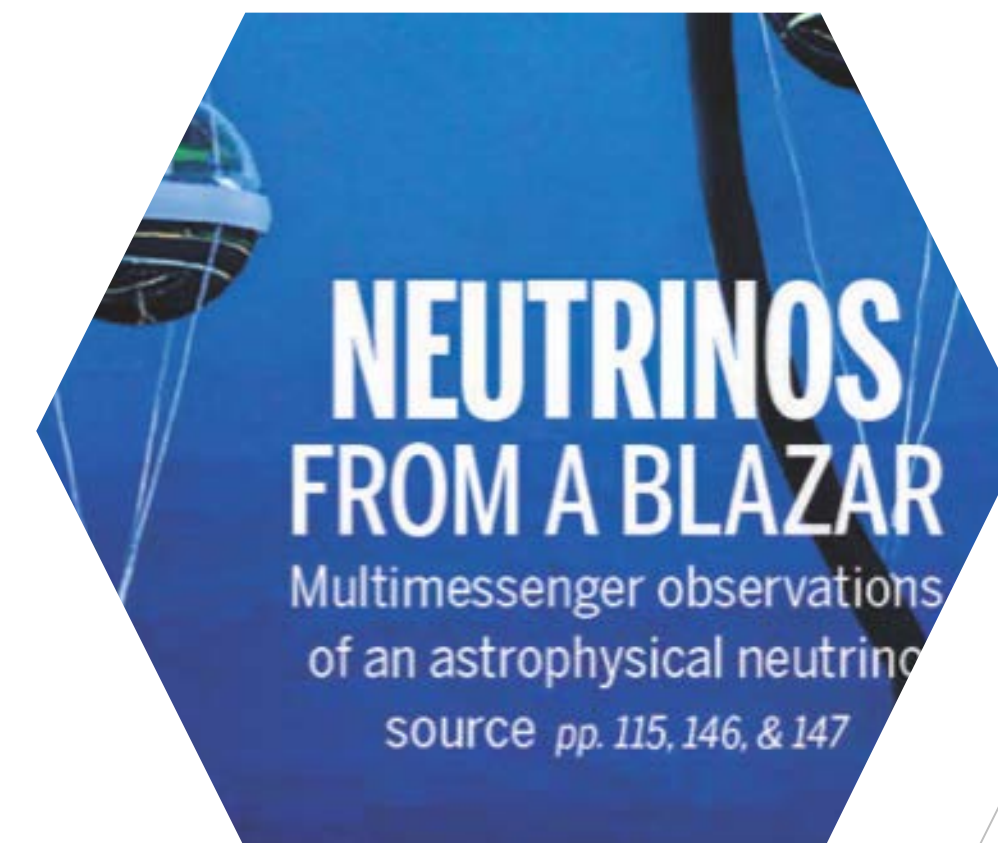
- Discovery of stars in other galaxies
- Explanation of phenomena based on heterogeneous measurements, e.g. gravitations waves, black holes

Linguistics

- Meaning is language use

Sociology

- Social networks show social behavior



Following the observation of a high energy neutrino in IceCube, a black hole at the center of a distant galaxy in the constellation of Orion was observed by telescopes, including MAGIC.

ARTIFICIAL INTELLIGENCE

- Natural Language Processing
- Planning
- Robotics
- Reasoning and Inference
- Information Retrieval, World Wide Web
- German National AI Research Institute is continuously funded since 1988.

➤ Orthogonal mission:
Everything becomes smarter through **learning!**



MACHINE LEARNING IS ...

Part of Computer Science

- Builds upon and contributes to theoretical and technical computer science

Based on Data

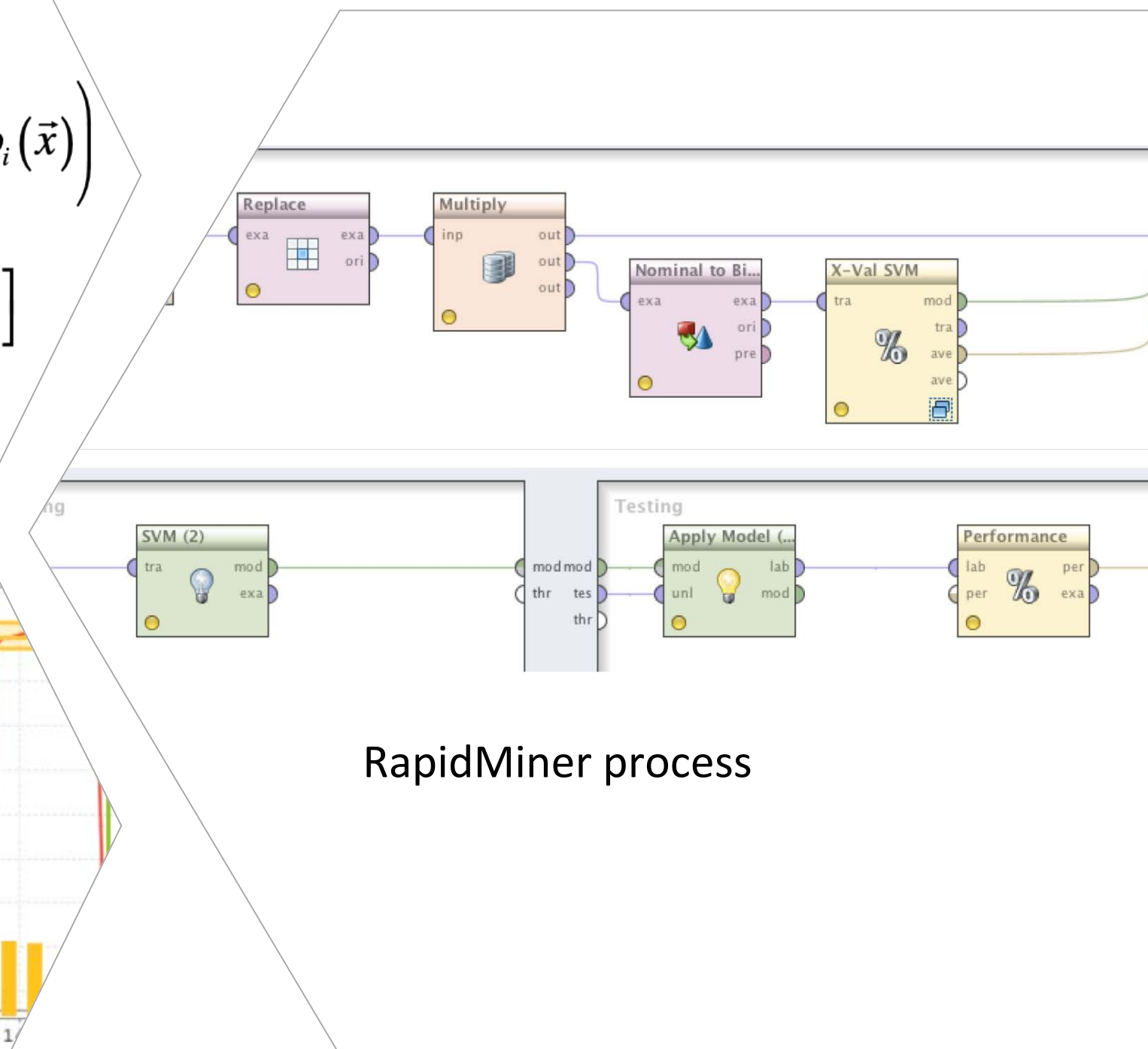
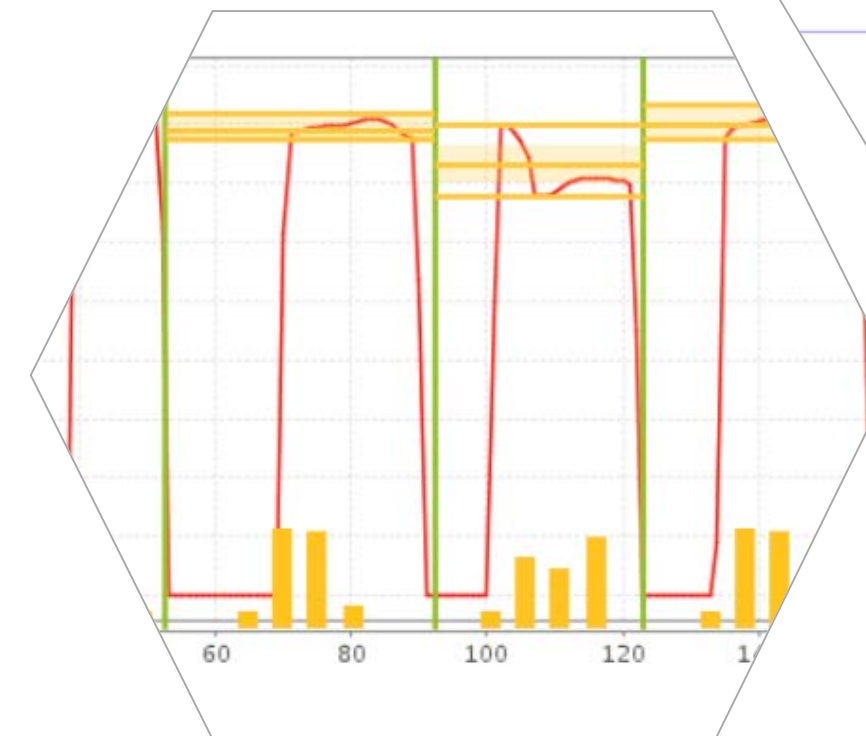
- Data streams mining, data summaries
- Distributed data analysis, federated learning
- Storage and curation of small and big data

Complex architectures

- Long pipelines
- Embedded processes
- Meta-learning

MRF :

$$p(\vec{x}) = \frac{1}{Z(\vec{\theta})} \exp\left(\sum_i \theta_i \phi_i(\vec{x})\right)$$
$$= \exp\left[\langle \vec{\theta}, \phi(\vec{x}) \rangle - A(\vec{\theta})\right]$$



RapidMiner process

OVERVIEW

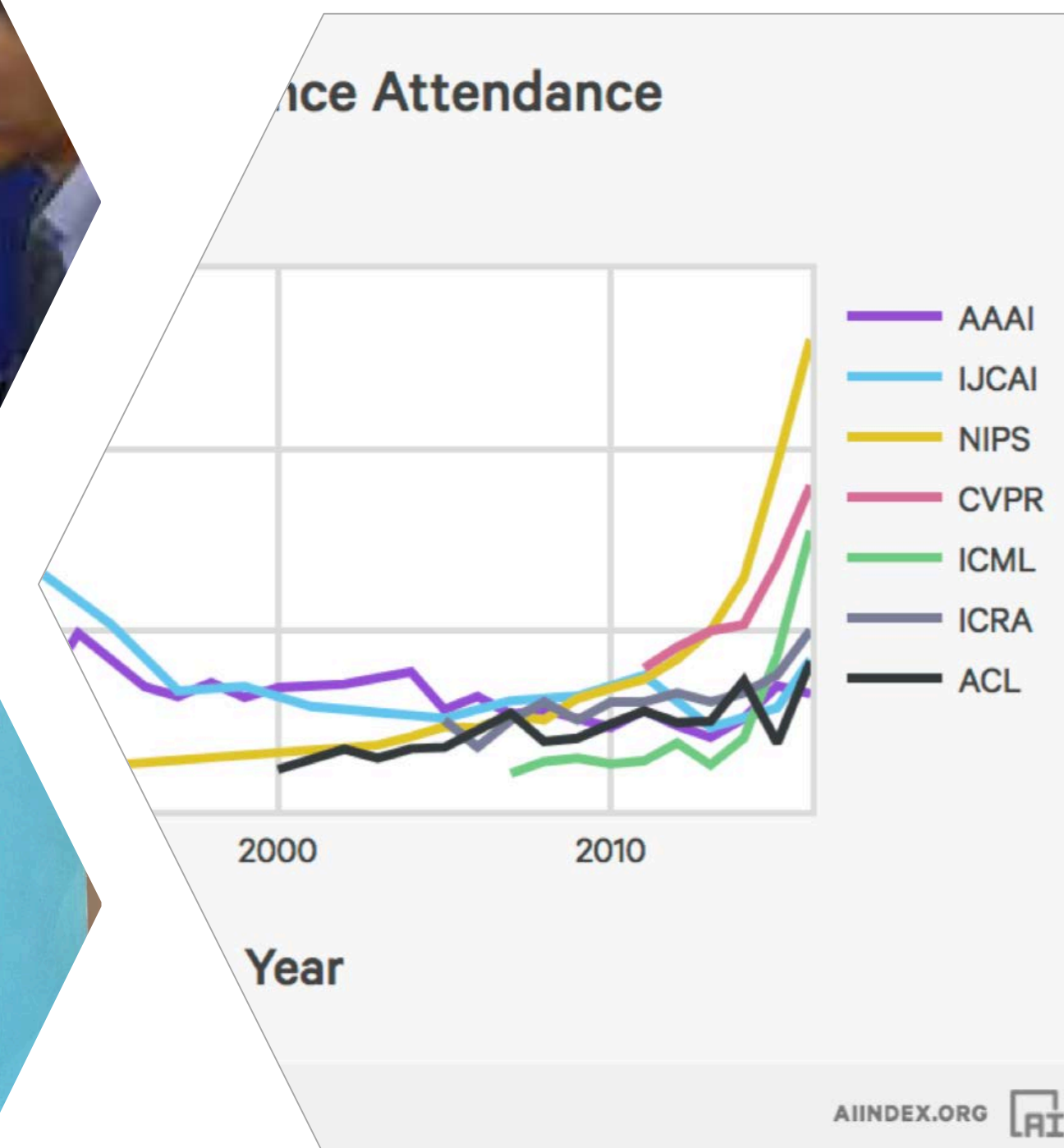
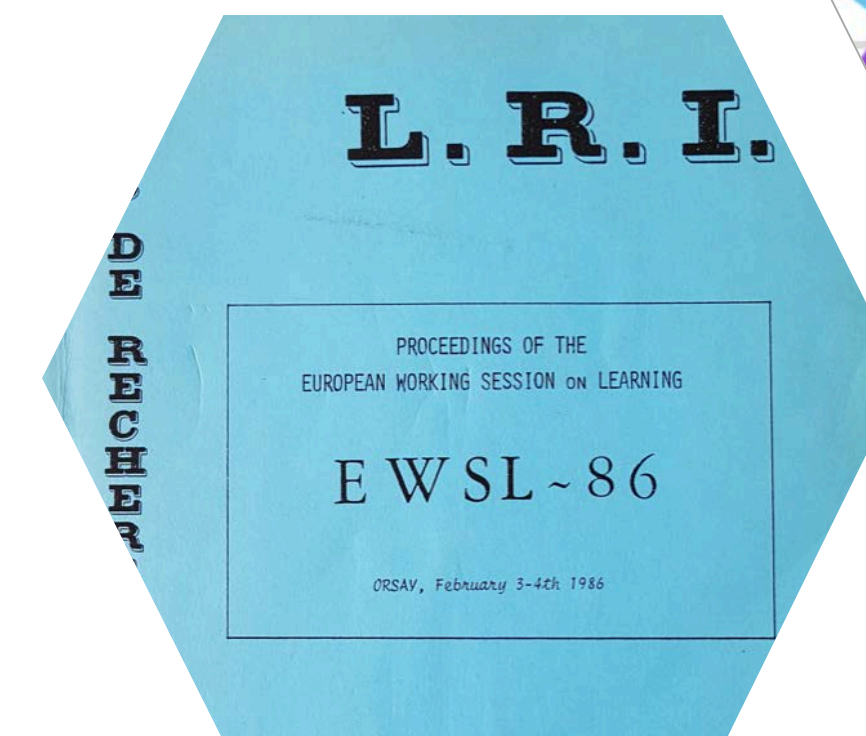
- Data Science and Artificial Intelligence
- Machine Learning
- Competence Centers in Germany
- Machine Learning Challenges



MACHINE LEARNING HAS MANY MODELS

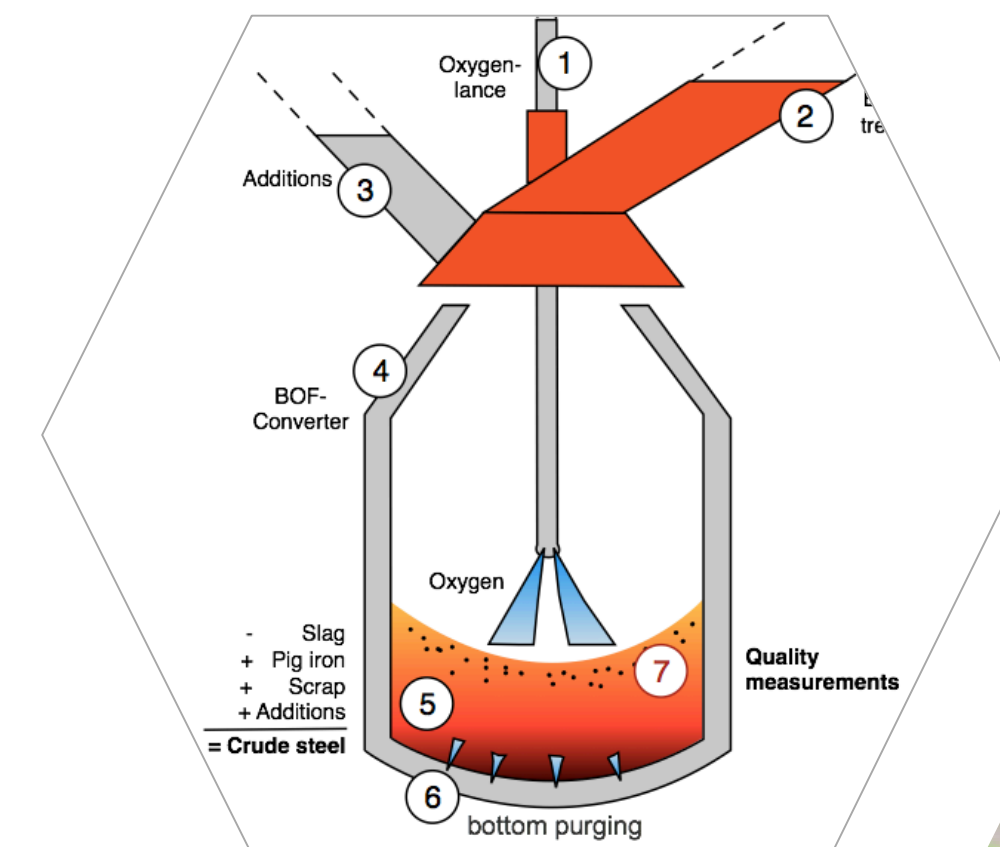
- Induction of Decision Trees
- Support Vector Machine
- Clustering
- Probabilistic Graphical Models
- Frequent Itemset Mining
- Reinforcement, Q Learning
- Neural Networks
- Feature extraction, Feature Selection, Convolutional Neural Networks
- Time Series Classification, Clustering, Prediction

2016 alphaGo
defeats Lee Sedol



APPLICATION AREAS

- Industrial Production
 - Predictive Maintenance
 - Quality Prediction
 - Model Predicted Control
- Logistics and Modern Mobility
 - Congestion Prognosis
 - Smart, multi-modal Routing
 - Shelf management, tracking of goods
- Medicine and Health
- Information Extraction and Knowledge Graphs



MACHINE LEARNING

Large Field

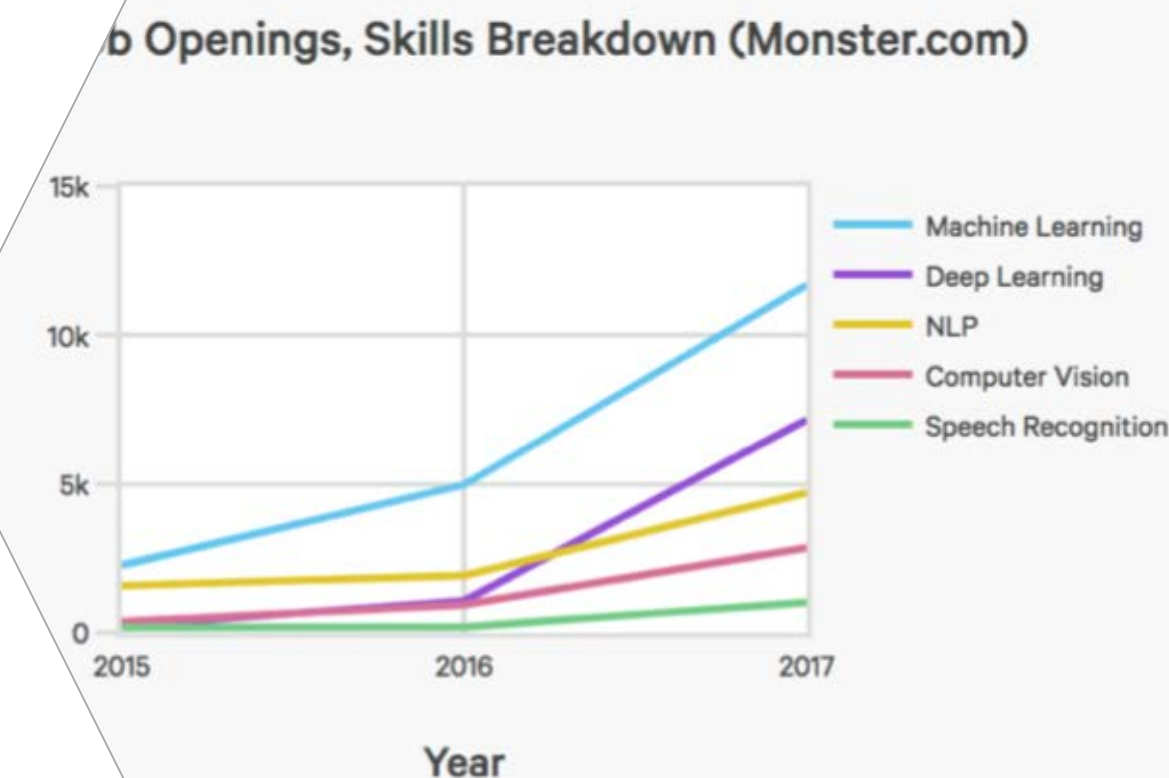
- Model classes
- Learning tasks
- Complex processes

Further Research is Needed

Key to

- Data driven sciences
- Smart economy

Need of Machine Learning Specialists:
research, start-ups, applications!



Germany needs 85000 academics with machine learning and big data skills.

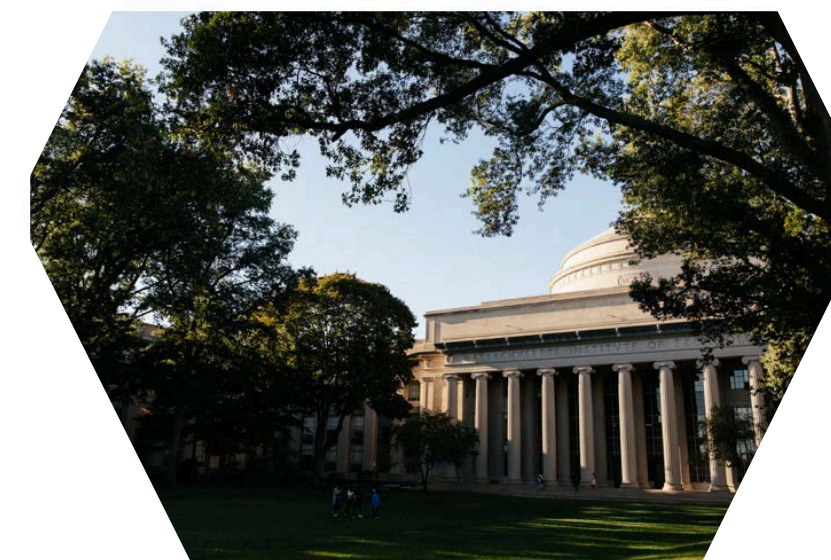
DEMANDS

- International Competition
- Scalability of Research
- Attracting Excellence
- Innovation
- Transfer, Novel Business Models

How to address these demands?

China:
\$150 Billion AI
Germany: €4 Billion
new technologies
France:
€ 1,5Billion AI

*MIT Plans College for Artificial
Intelligence, Backed by \$1 Billion*

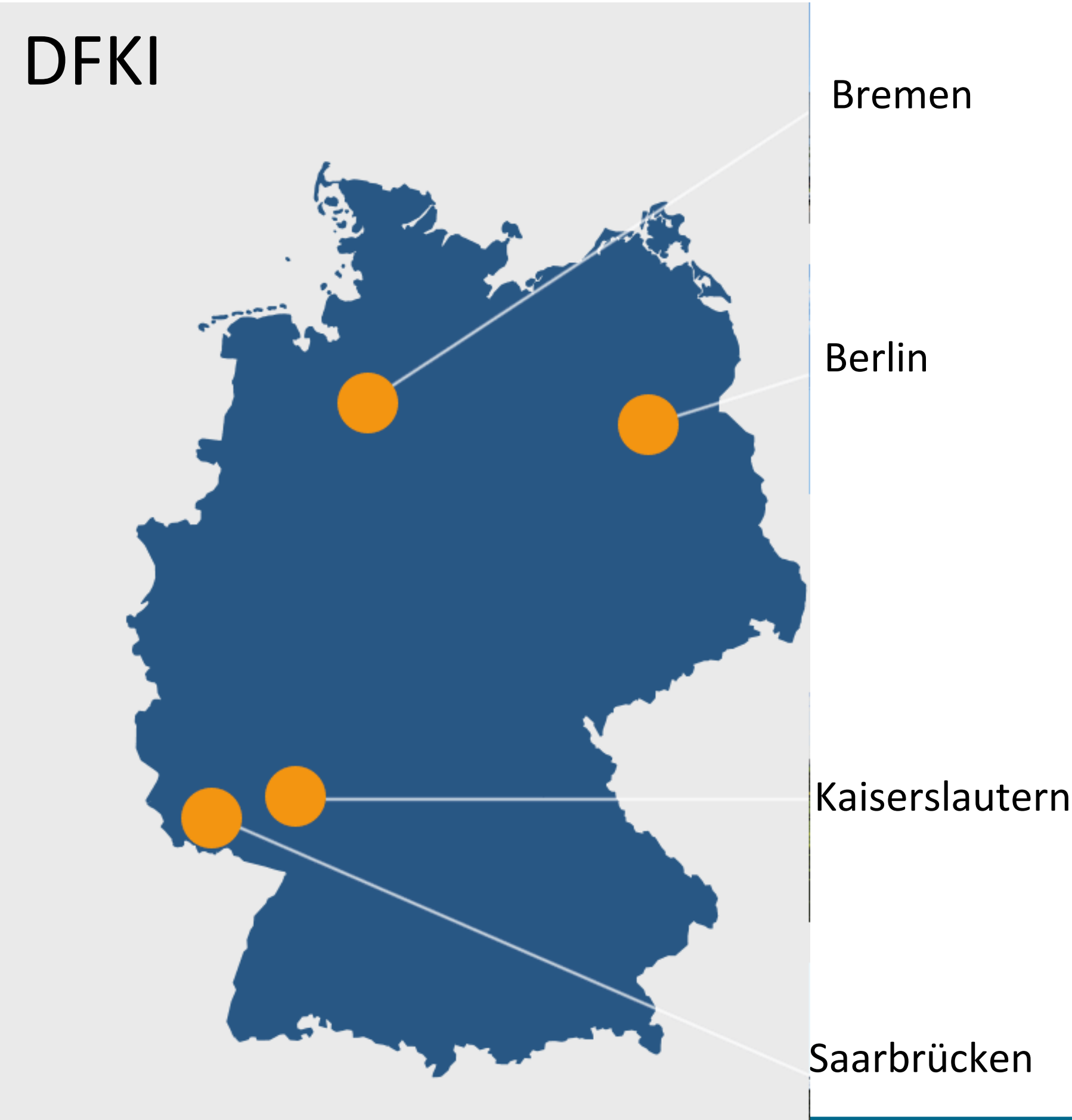


OVERVIEW

- Data Science and Artificial Intelligence
- Machine Learning
- Competence Centers in Germany
- Machine Learning Challenges



GERMAN AI MAP

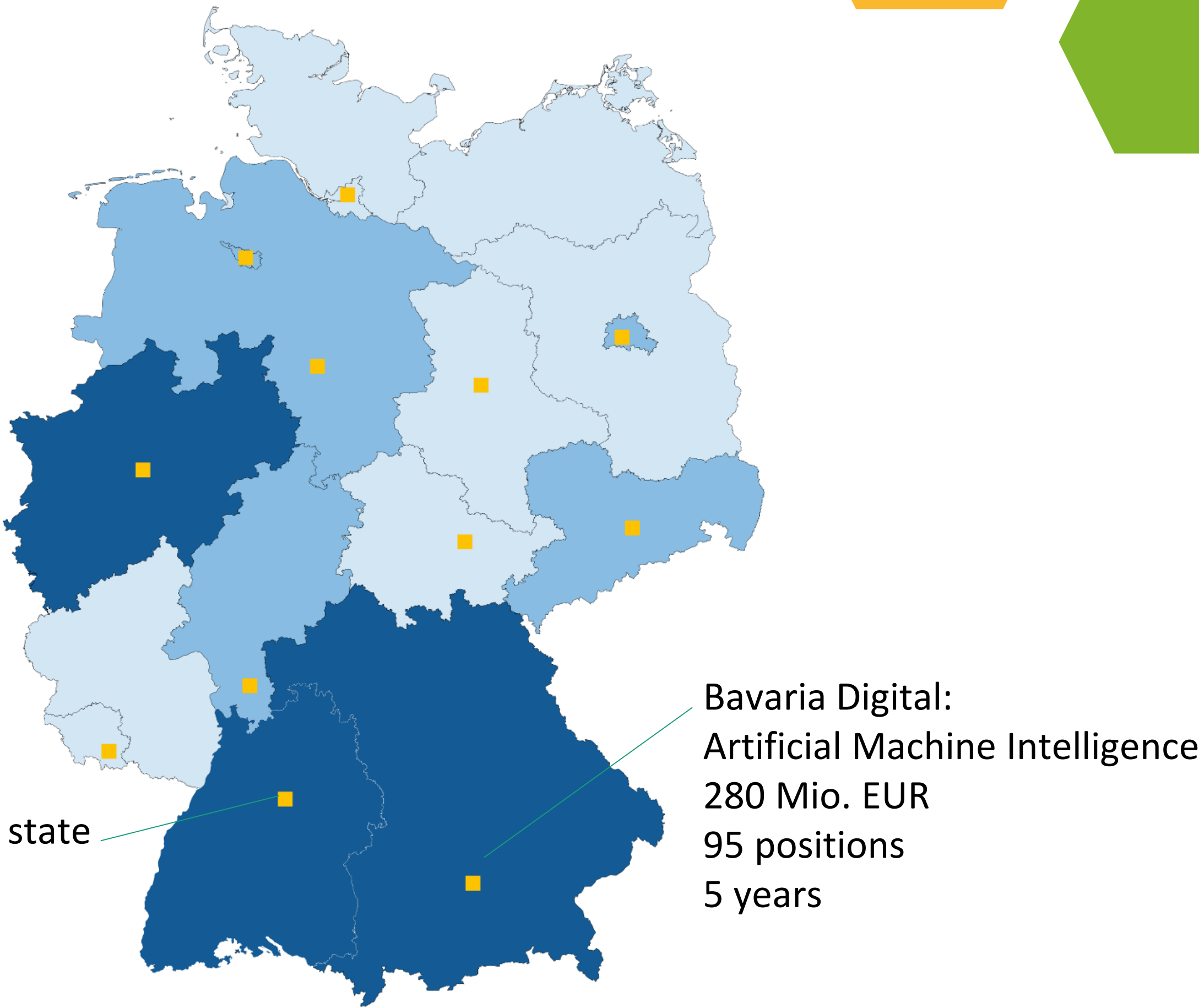


Max Planck
Institutes

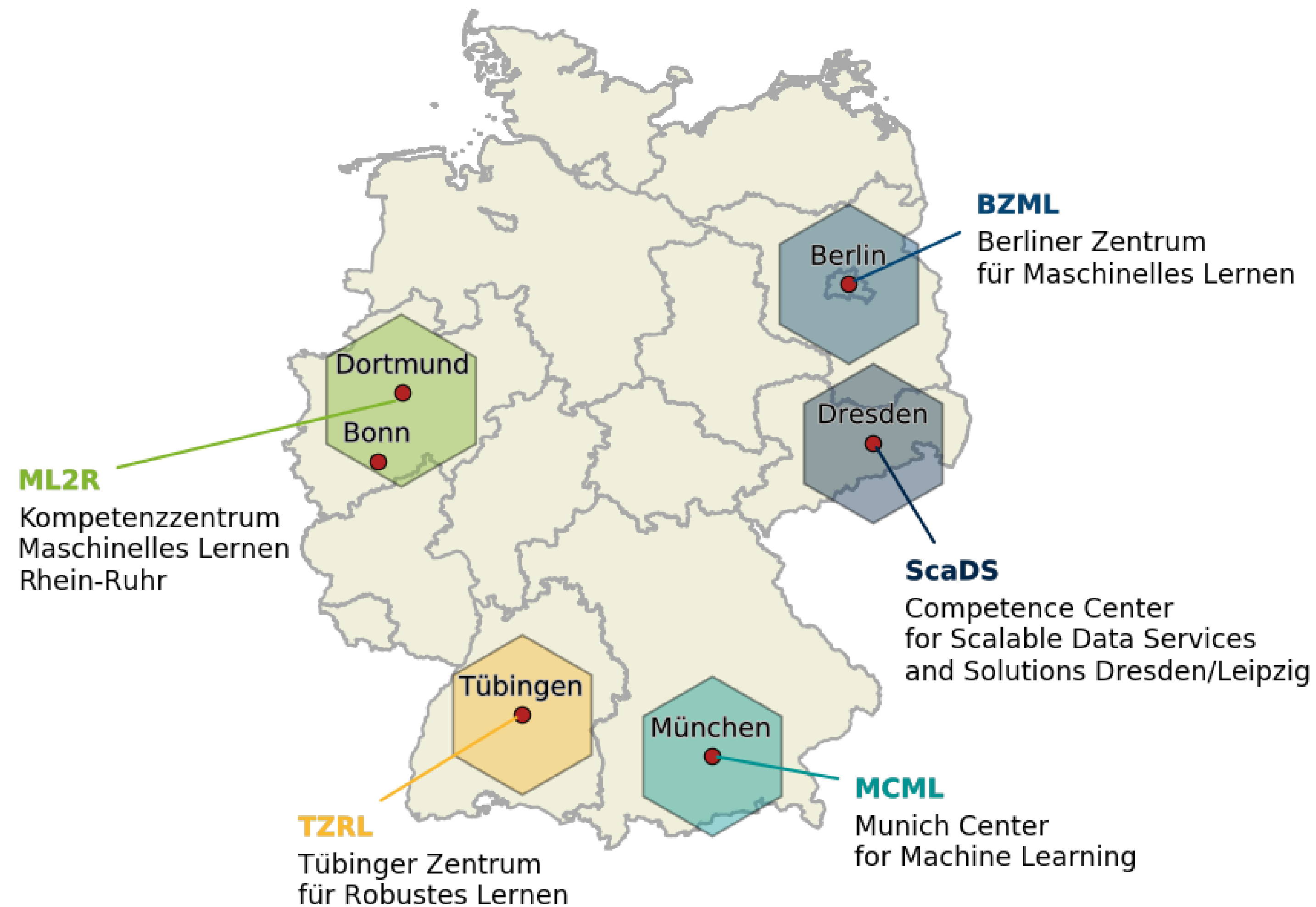
Collaborative
Research
Centers

Infrastructure
for data

CyberValley:
50 Mio. EUR by the state
50 Mio. EUR by
industries
1,25 EUR per year



MACHINE LEARNING COMPETENCE CENTERS



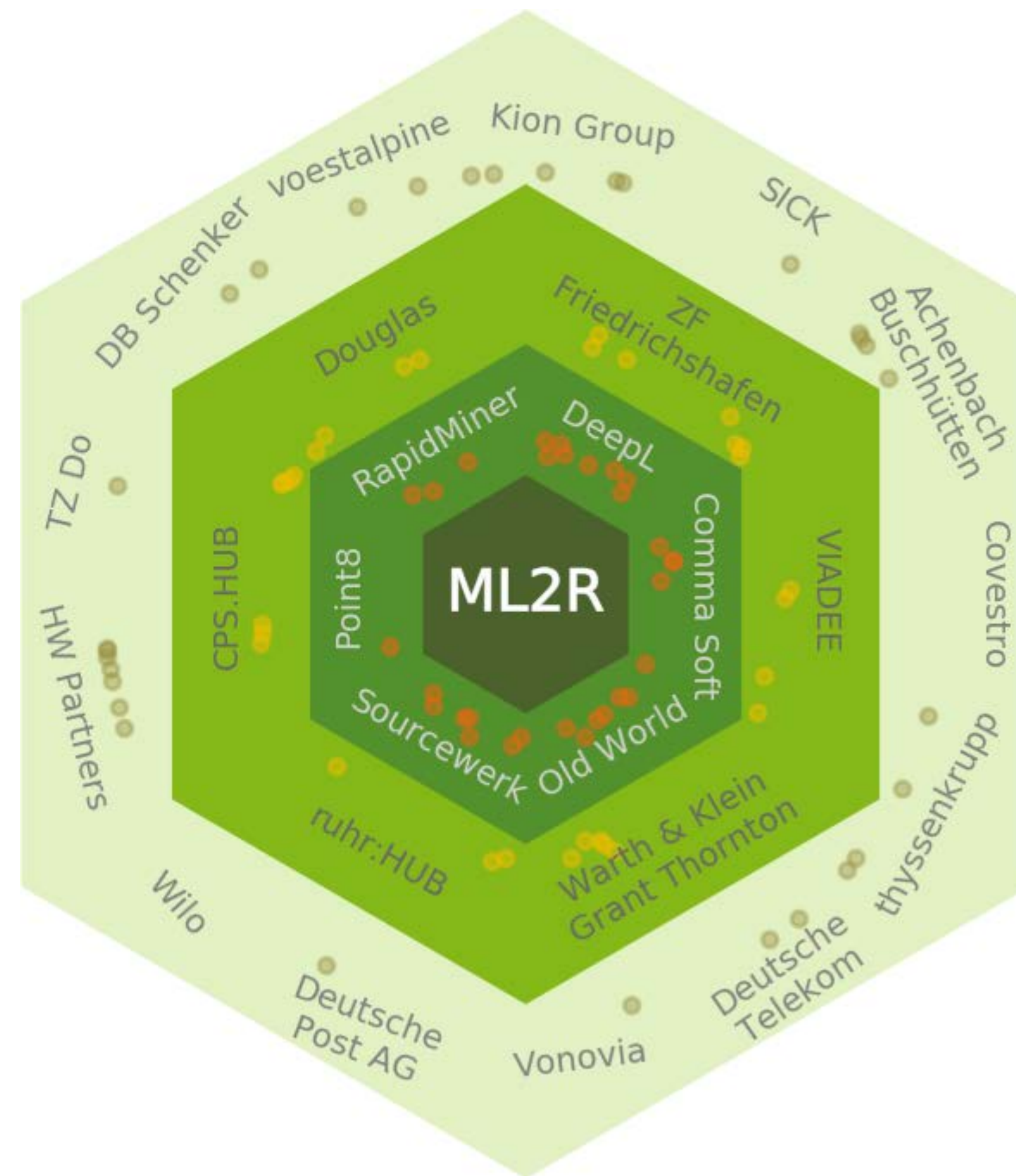
Research

Networking

Summer Schools

Transfer to Applications

TRANSFER TO MARKET: LOCAL TRANSFER

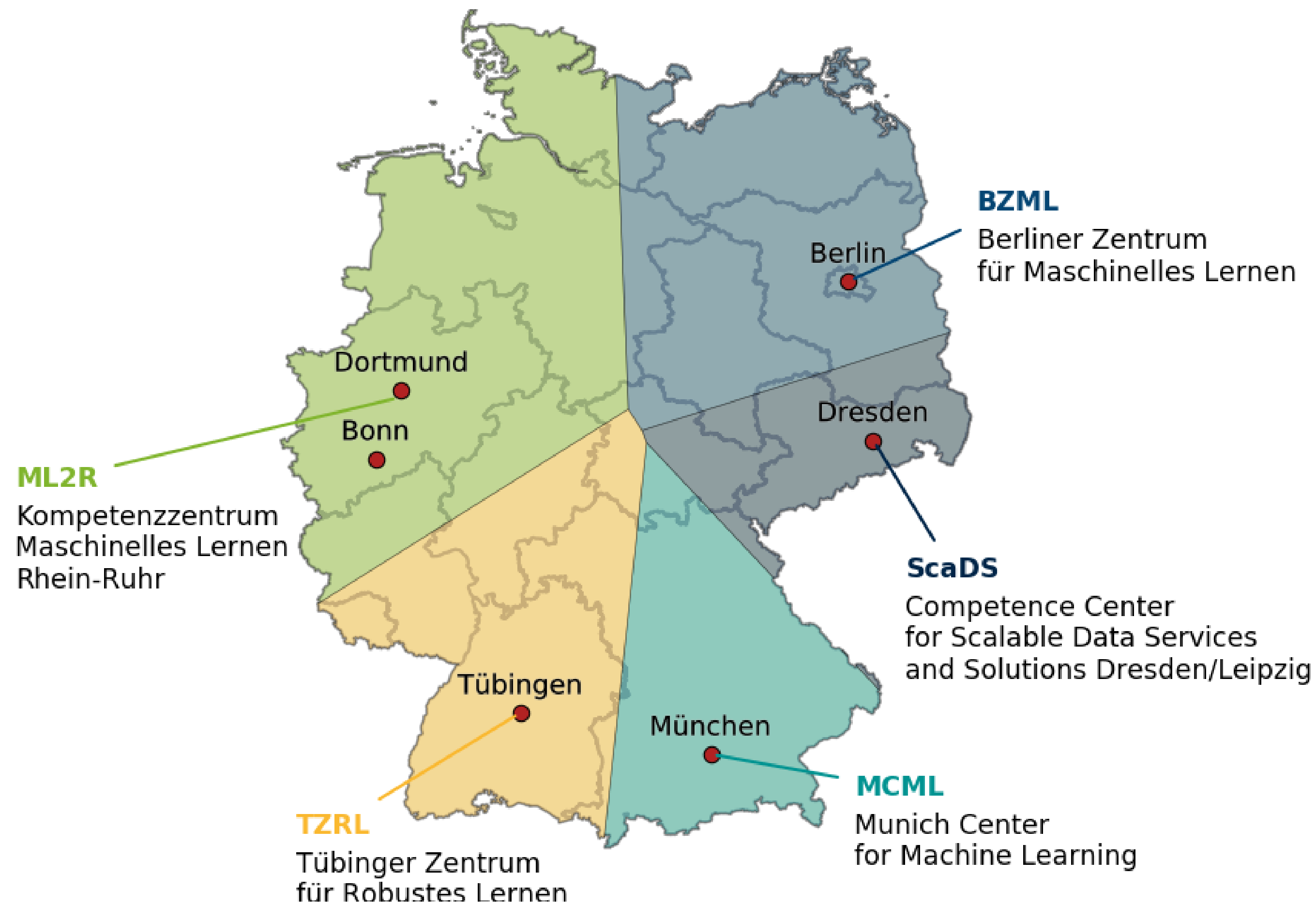


Research

Start-ups specialized
on machine learning

Companies applying
machine learning

NETWORK OF CENTERS AND THEIR LOCAL TRANSFER



OVERVIEW

- Data Science and Artificial Intelligence
- Machine Learning
- Competence Centers in Germany
- Machine Learning Challenges



MACHINE LEARNING CHALLENGES

Robust Learning

- Concept drift
- Transfer learning

Learning in Complex Data Structures

- Graphs, knowledge as input
- Graphical, spatiotemporal models

Learning and Hardware

- Learning on (for) restricted hardware
- Hardware for machine learning

Human-Oriented Machine Learning



MACHINE LEARNING WITH COMPLEX KNOWLEDGE

- Extracting Knowledge from Texts
- Exploiting User Feedback
 - Interactive modeling
 - Active learning
- Using Simulations
 - Simulations guide machine learning
 - Learned models enhance simulations
- Tailoring Learning based on Knowledge
 - Regularization of the optimization problem
 - Restricting the parameter space of learning



MACHINE LEARNING UNDER RESOURCE CONSTRAINTS

Small Devices Produce Massive Data Streams in the Internet of Things

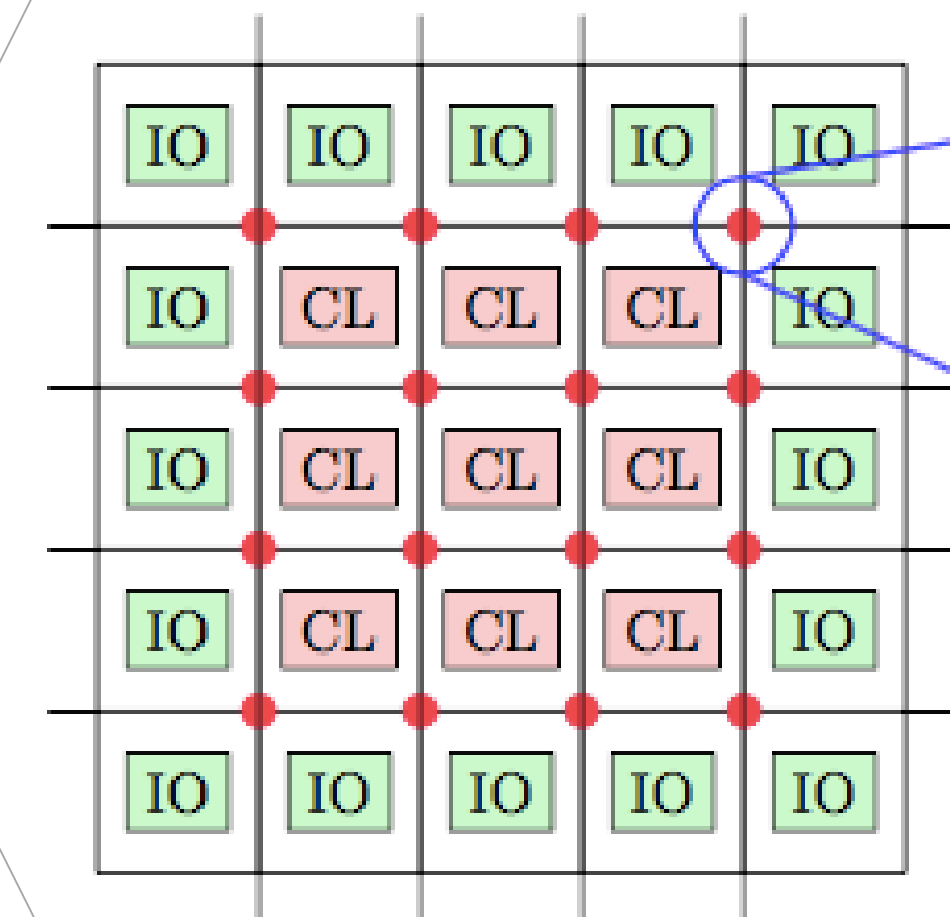
- Data summaries
- Distributed, federated learning

New Machine Learning Models Require Less

- Memory
- Communication
- Computing capacity

New Hardware Dedicated to Machine Learning

- Tensor Processing Unit (Google), Lake Crest (Intel)
- Quantum computing



Field Programmable Gate Array
FPGA

SUSTAINABILITY: ENERGY

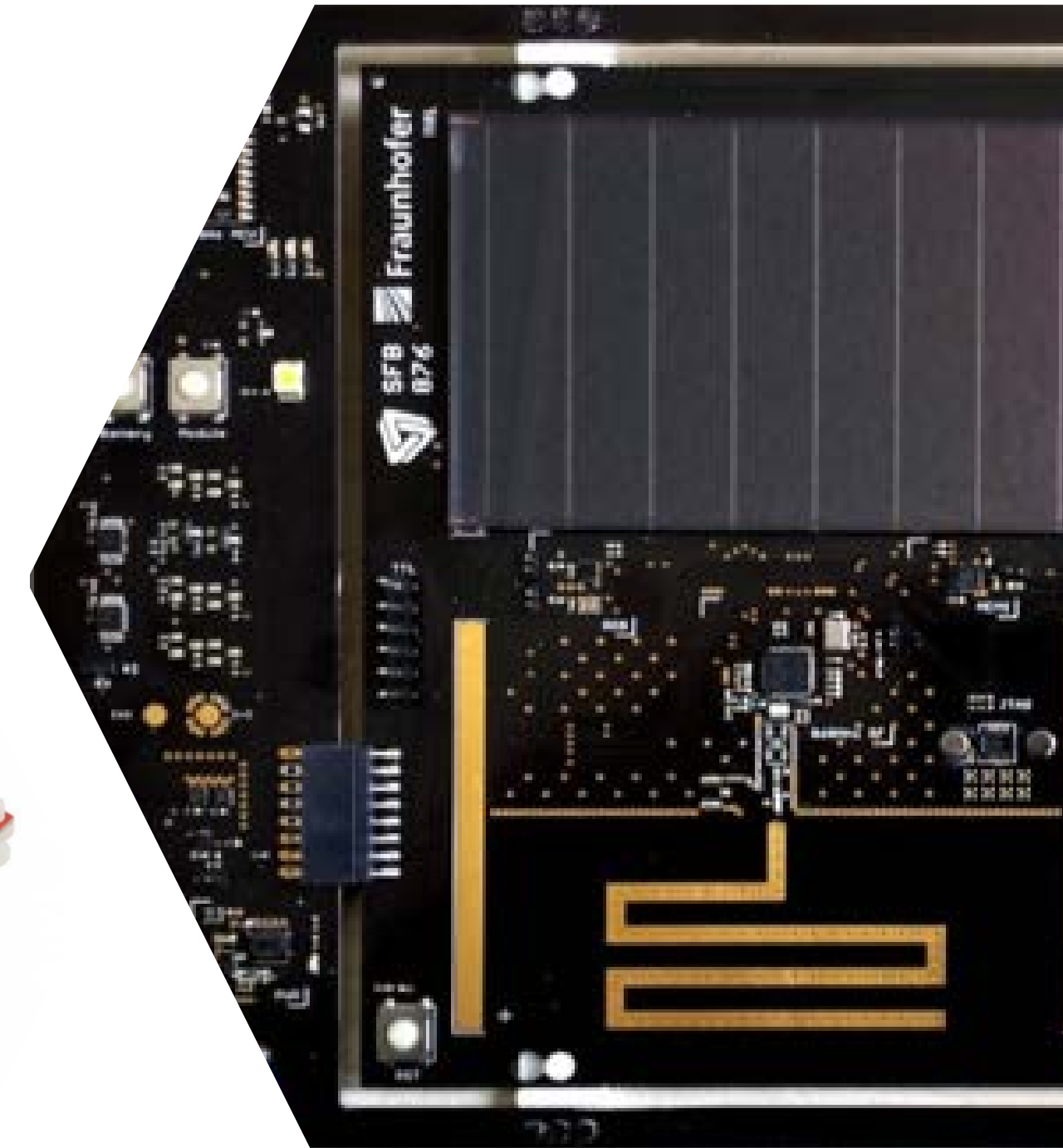
- Google's most important application of machine learning was in their computing centers.
- Decreasing the cost of cooling by 40 %!
Vinton Cerf interview 25.3.2017
- Google's total yearly energy consumption is 2 terawatt hours (2024 watt hours).
European Network of Excellence in Internet Science, report in Ubiquity June, 2015



 Machine learning reduces energy!

MACHINE LEARNING USING LESS ENERGY

- Ultra-low power microcontrollers
 - only 0.0048 watt
 - only 64 kb memory
 - no floating point unit
- Develop machine learning models that demand less resources!
 - Graphical models using only integer numbers and operations!
 - Implementing learning algorithms on FPGA, e.g. decision trees, deep learning



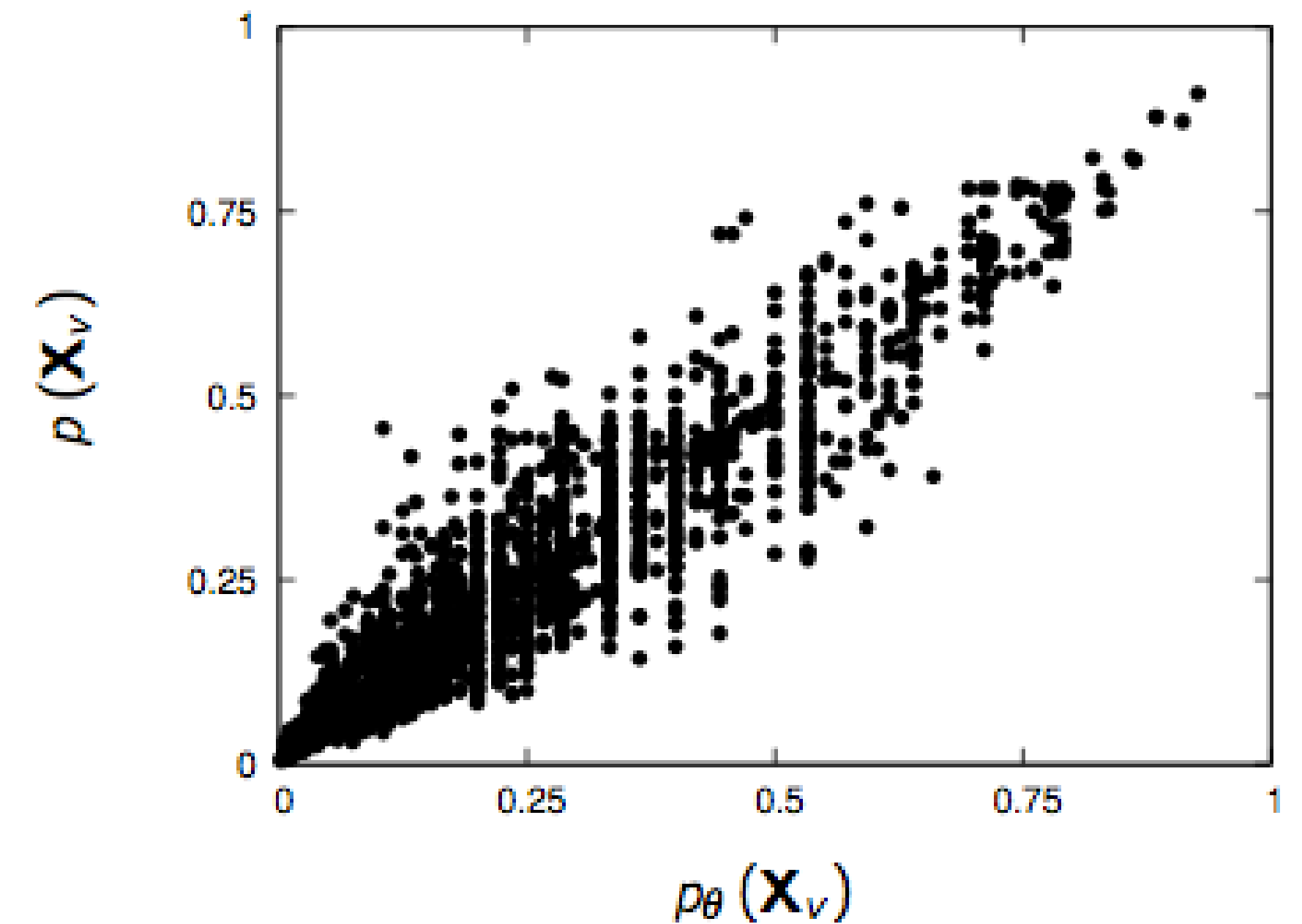
INVESTIGATING RESOURCE DEMANDS OF LEARNING

Energy saving

- Restrict the parameter space, e.g. to integer numbers.
Bound the approximation error.

Model compression for less memory

- Exploit redundancy in the parameters of a model.
Bound on distance between true parameters and the compressed version.



UNIVERSAL REPARAMETRIZATION COMPRESSES THE MODEL

- Reparametrize model

$$\Delta_t \approx \theta_{t+1} - \theta_t$$

Δ regularized by L1, L2 norm

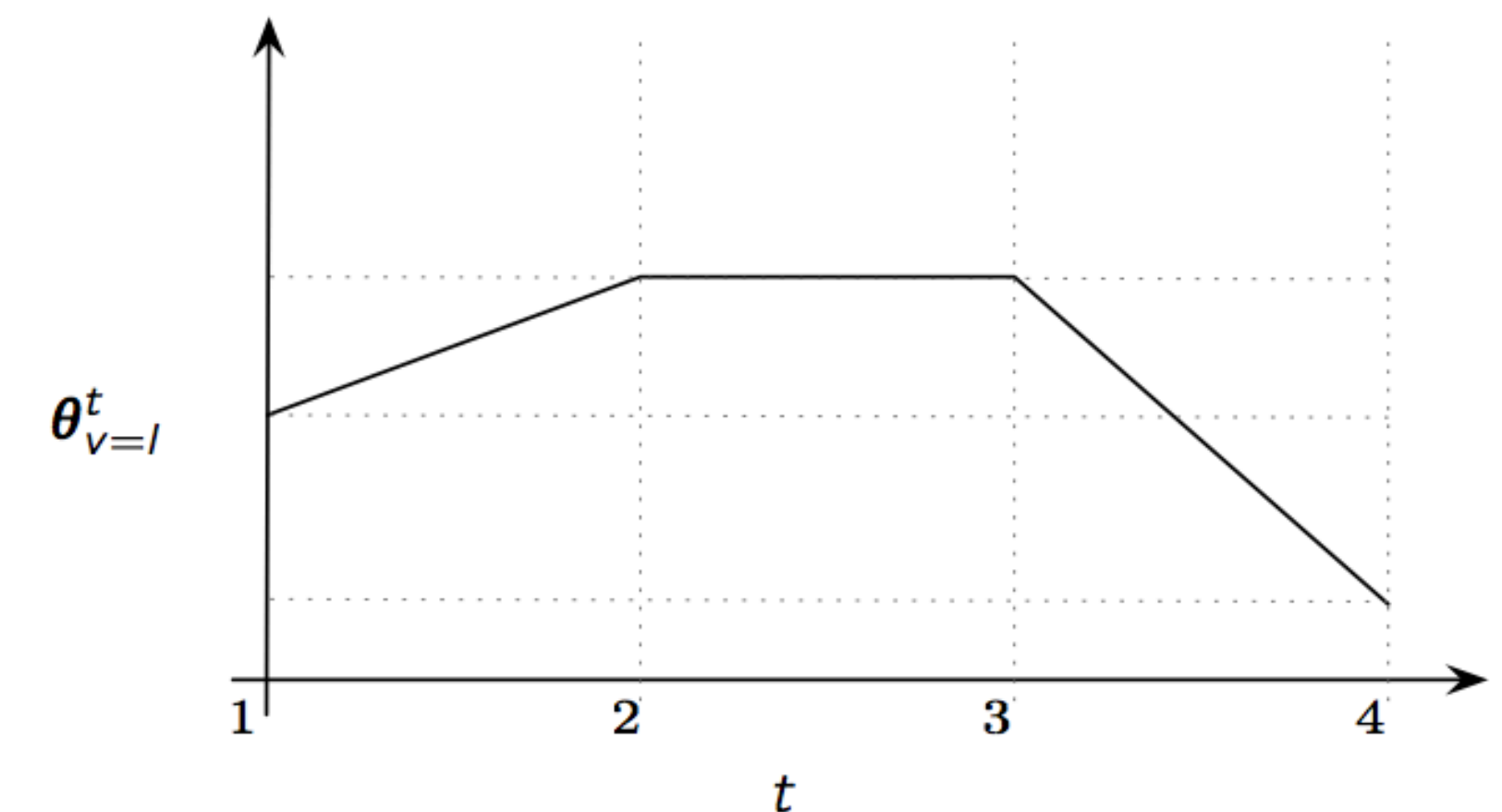
- There are not many changes over time.

- Bound on distance between true θ and $v(\Delta)$;
Sparsity in estimate implies redundancy in the true parameter.
Proof Piatkowski (2018)

- Learning is faster.

- Quality is not at all less than MRF, 4NN.

Assumption: Smoothness over time



Idea:

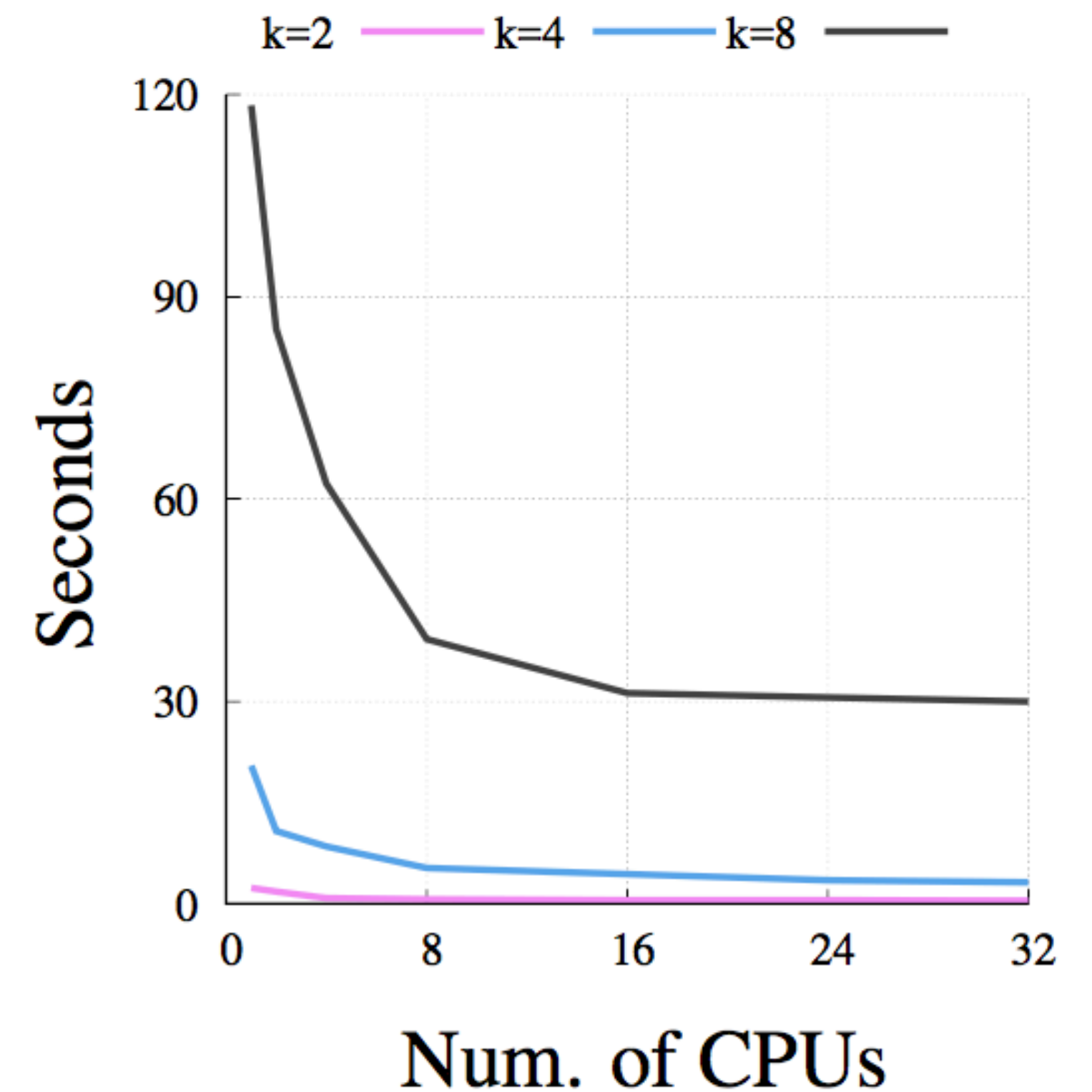
- Remove the near zero slopes, while retaining the performance

INVESTIGATING RESOURCE DEMANDS OF LEARNING

Scalability

- Separate hard-to-compute parts from the easy parts; precompute the hard parts; implement parallel algorithms.
- Approximate integral based on Chebyshev polynomials; numerical approximation with bounded error.

Algorithm	Complexity	Quality
JT (Lauritzen & Spiegelhalter, 1988)	$\mathcal{O}(L^w)$	Exact
MF (Weiss, 2001)	$\mathcal{O}(InL\Delta)$	Lower bound
LBP (Heskes, 2002; Yedidia et al., 2003)	$\mathcal{O}(ImL^2\Delta)$	Local minimum of Bethe free energy
TRW (Wainwright et al., 2005)	$\mathcal{O}(ImL^2\Delta + m \log n)$	Upper bound
WISH (Ermon et al., 2013)	$\mathcal{O}(n \ln(n/\zeta)) \times \text{Time(MAP)}$	$(16, \zeta)$ -approx
DCCQ (Alg. 1)	$\mathcal{O}(k_\varepsilon^2 d^{2k_\varepsilon})$	ε -approx (Theorem 5)
SDCCQ (Alg. 2)	$\mathcal{O}(k_\varepsilon^2 d^{2k_\varepsilon}) + \mathcal{O}(k_\varepsilon^2 m_\zeta)$	(ε, ζ) -approx (Theorem 7)



HUMAN-ORIENTED MACHINE LEARNING

Understanding Learned Models

- Explainable (explain afterwards)
- Interpretable (easily understandable)
- Inspectable (investigate examples)

Reproducibility

- Sampling, data generation
- Counterfactual modeling
- Meta-Learning

Certificates of Data and Models

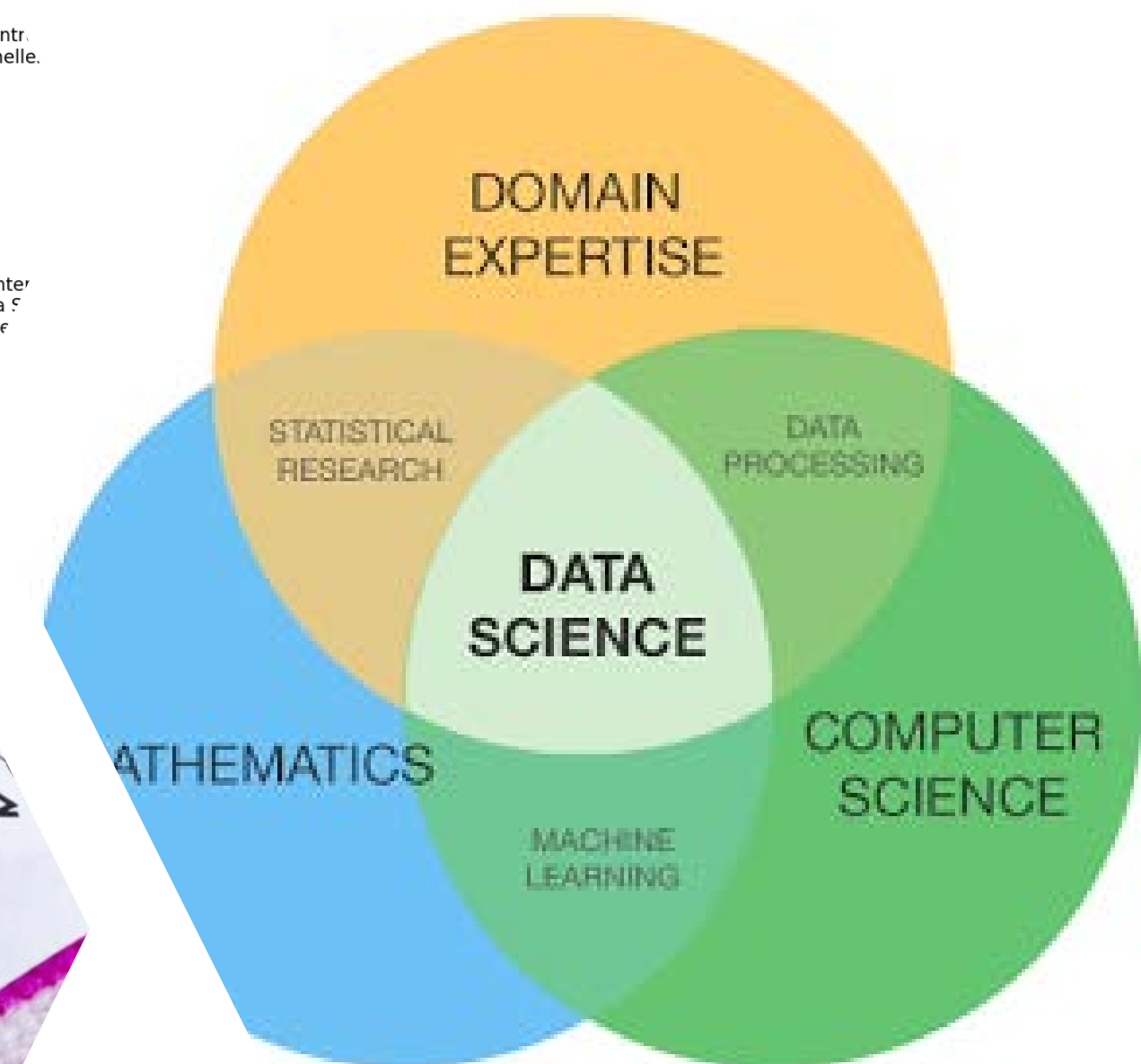
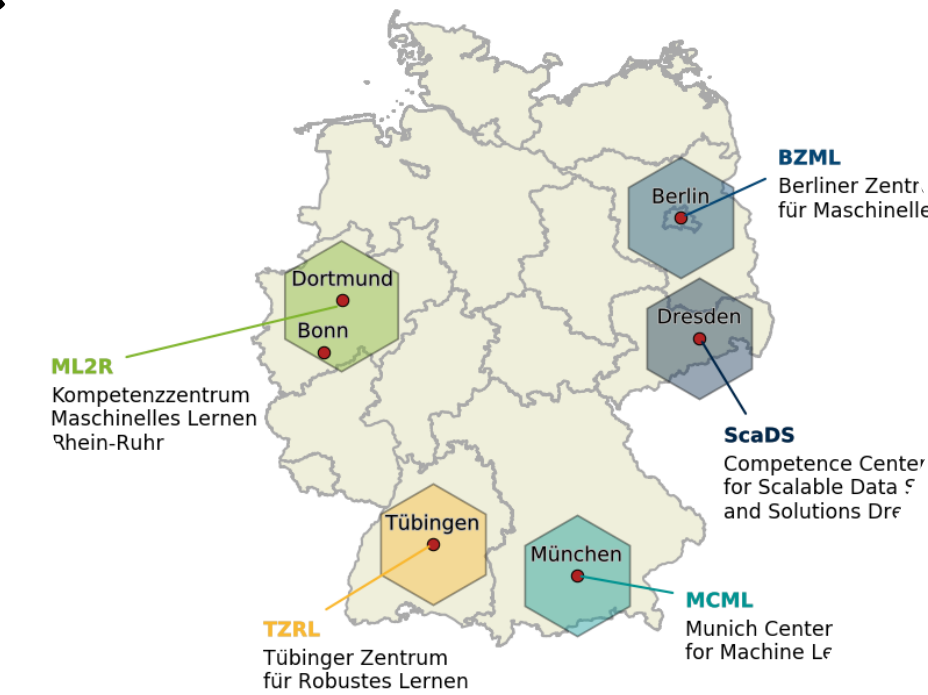
- Translating theoretical proofs into labels



SUMMARY

- Data Science and Artificial Intelligence
- Machine Learning
- Competence Centers in Germany
- Machine Learning fuels
 - Good labor
 - Scientific insight
 - Sustainability
- Machine Learning must
 - Give guarantees for learned models
 - Express guarantees and risks easy to read

Further research is urgently needed!



Source: Palmer, Shelly. Data Science for the C-Suite. New York: Digital Living Press, 2015. Print.

LITERATURE

- Nico Piatkowski, Sangkyun Lee, Katharina Morik (2013) Spatio-Temporal Random Fields: Compressible Representation and Distributed Estimation, Machine Learning Journal, Vol. 93, No. 1, S. 115 – 139
- Nico Piatkowski, Sangkyun Lee, Katharina Morik (2016) Integer undirected graphical models for resource-constrained systems, Neurocomputing, Vol. 173, No. 1, 9 – 23
- Nico Piatkowski, Katharina Morik (2018) Fast Stochastic Quadrature for Approximate Maximum Likelihood Estimation, Procs. 34th Uncertainty in AI
- Nico Piatkowski (2018) Exponential Families on Resource-Constrained Systems, Ph D thesis TU Dortmund
- Sebastian Buschjäger, Katharina Morik (2018) Decision Tree and Random Forest Implementations for Fast Filtering of Sensor Data, IEEE Transactions on Circuits and Systems I: Regular Papers, Vol. PP, No. 99, p. 1 – 14
- Sebastian Buschjäger, Kuan-Hsun Chen, Jian-Jia Chen, Katharina Morik (2018) Realization of Random Forest for Real-Time Evaluation through Tree Framing, Procs. ICDM 2018

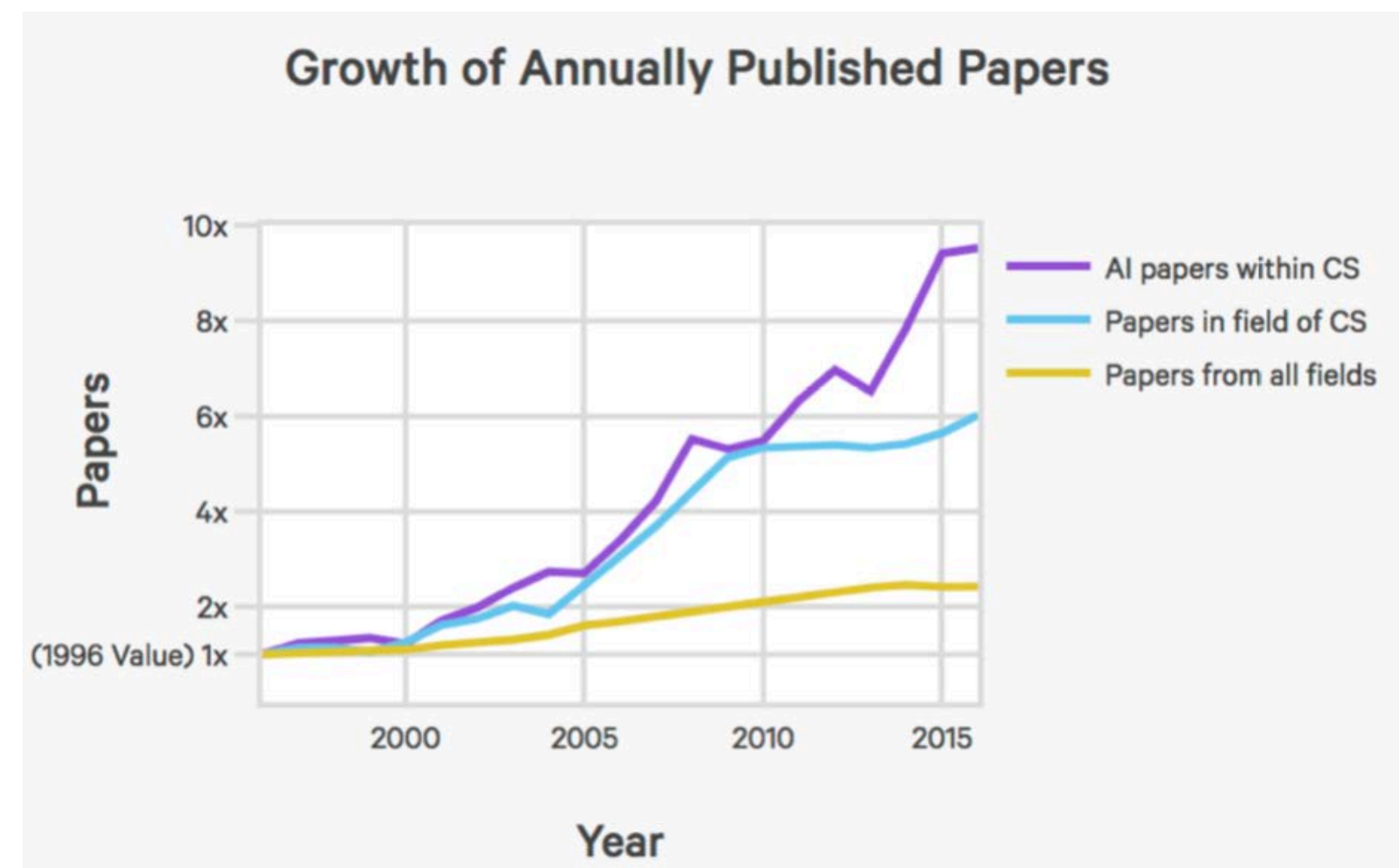
STUDIES ON DEMANDS AND OPPORTUNITIES OF AI IN GERMANY



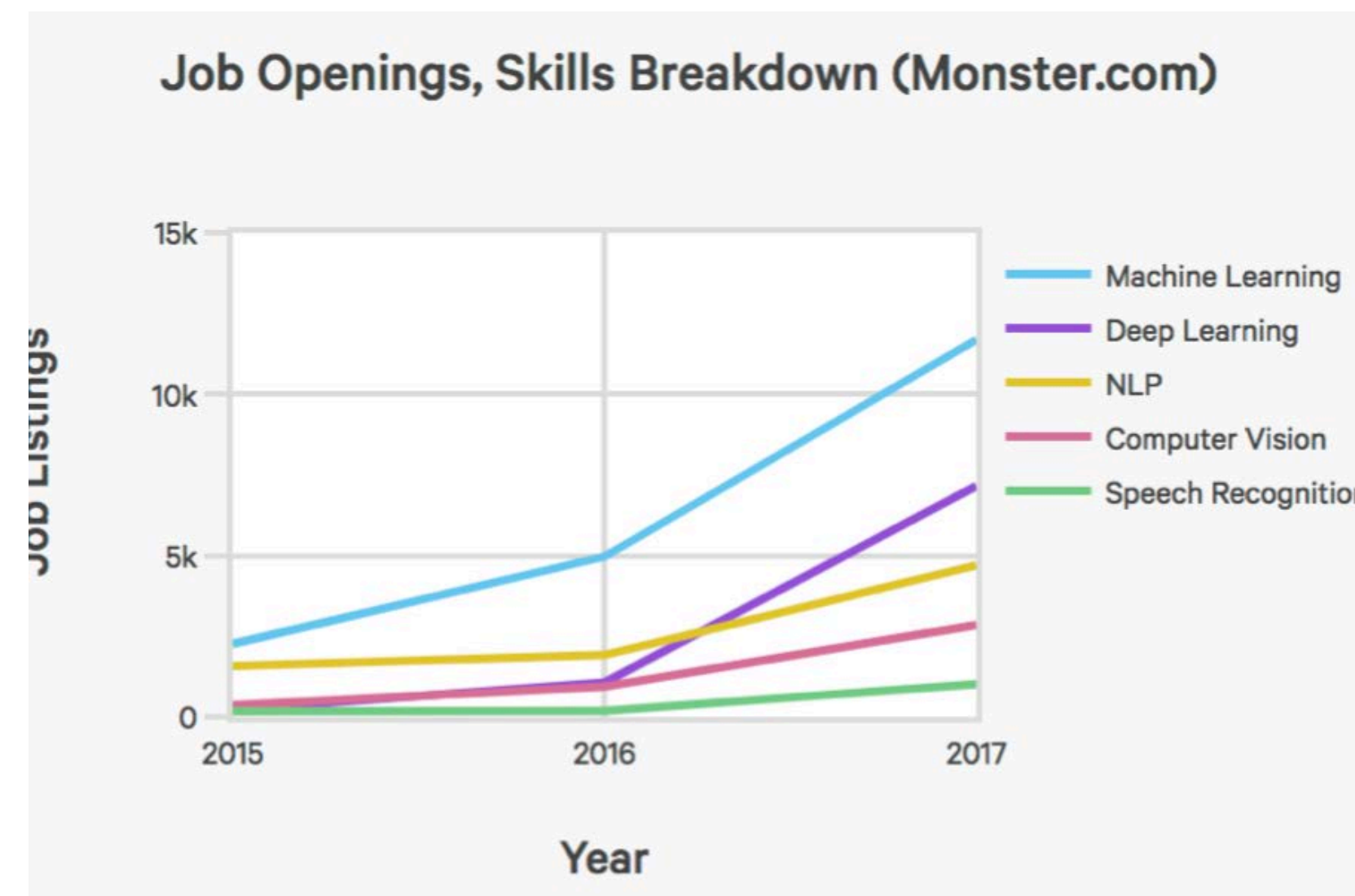
- **Fraunhofer-Allianz Big Data (Hg.) (2017): Zukunftsmarkt Künstliche Intelligenz – Potenziale und Anwendungen,** https://www.bigdata.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/KI-Potenzialanalyse_2017.pdf
- **Fraunhofer-Allianz Big Data (Hg.) (2018): Maschinelles Lernen – Eine Analyse zu Kompetenzen, Forschung und Anwendung,** https://www.bigdata.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer-Studie_ML_2018_WEB.PDF
- **EFI – Expertenkommission Forschung und Innovation (2018): Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands 2018,** https://www.e-fi.de/fileadmin/Gutachten_2018/EFI_Gutachten_2018.pdf
- **Begleitforschung PAiCE, iit-Institut für Innovation und Technik in der VDI / VDE Innovation + Technik GmbH (Hg.) (2018): Potenzial der Künstlichen Intelligenz im produzierenden Gewerbe in Deutschland,** <https://www.bmwi.de/Redaktion/DE/Publikationen/Studien/potenziale-kuenstlichen-intelligenz-im-produzierenden-gewerbe-in-deutschland.html>
-

Growth of the Field, Demand of Skills

- 457 Open positions in Germany Maschinelles Lernen
- 470 Open positions in Germany Machine Learning
- 154 Deep Learning
- 118 Natural Language Processing incl. Speech
- 200 Image, Vision



<https://aiindex.org>



(Monster.de am 23.9.2018)

MODULAR MACHINE LEARNING

