Berkcan Kapusuzoglu

+1 571-505-5914 berkcankapusuzoglu@gmail.com /berkcankapusuzoglu /kapusuzoglu
Falls Church, VA 22043

Research Interests

LLM, Sparse Training, PINN: My research focuses on advancing large language models (LLMs) through novel training methodologies, particularly in continual learning and catastrophic forgetting mitigation. I combine theoretical approaches (random matrix theory) with practical implementations (model merging, sparse training) to develop production-ready artifical intelligence (AI) systems, while leveraging my background in physics-informed neural network (PINN) to bridge domain expertise with modern deep learning.

Education

Vanderbilt University **Aug. 2017 – Mar. 2022**
Ph.D. in Civil Engineering, (GPA: 3.88) *Nashville, TN*
Thesis title: Physics-informed machine learning for uncertainty quantification and optimization

- Explored the integration of physics-based models and experimental data to craft diverse physics-informed machine learning models. These models, designed for optimization under uncertainty, achieved a remarkable 300x reduction in computational effort while preserving accuracy.
- Spearheaded the development of adaptive sampling and multi-level Bayesian calibration strategies. These advancements were applied to a complex multivariate time-dependent multi-component system, resulting in a noteworthy enhancement of computational efficiency and accuracy by a factor of 120 and 16%, respectively.

Delft University of Technology **July 2015 – Nov. 2016**
Master of Science in Applied Mathematics *Delft, The Netherlands*

University of Erlangen-Nuremberg **Aug. 2014 – May 2015**
Master of Science in Computational Engineering *Erlangen, Germany*

Bilkent University **Sep. 2010 – June 2014**
Bachelor of Science in Mechanical Engineering *Ankara, Turkey*

Experience

Capital One - AI Foundations Team **Dec. 2023 – Present**
Principal Associate, Data Scientist *USA*

- Led development of novel continual pre-training framework for LLMs, reducing catastrophic forgetting by 46% while improving the learning of new internal tasks.
- Implemented multi-node distributed training pipeline via Kubernetes, scaling Llama 3.1 models (8B-70B parameters) efficiently across multiple GPUs and nodes.
- Pioneered data rehearsal techniques for continual pre-training and supervised fine-tuning using internal datasets, developing custom data curation and synthesis pipelines to ensure high-quality training data while maintaining data privacy.
- Automated evaluation pipeline for NLP tasks, enabling rapid iteration and validation.

Capital One **Nov. 2022 – Nov. 2023**
Principal Associate, Data Scientist *USA*

- Collaborated closely with a cross-functional team of data scientists, software engineers, and product managers to deliver accurate forecasts for Comprehensive Capital Analysis and Review (CCAR) and routine business operations. Executed pivotal adjustments to the model forecast, resulting in a significant total allowance impact of $600M.
- Applied a diverse technology stack, including Python, AWS, and Spark, to enhance the precision of loss forecasts. Conducted comprehensive sensitivity analysis on account-level loss models across different economic scenarios, leading to the identification and resolution of a data issue with a consequential impact of $700M in provision.
- Engineered a custom Continuous Integration/Continuous Deployment (CICD) pipeline to create a summarization and visualization tool. This tool enables the broader team to efficiently summarize model outputs and visualize them on Tableau, fostering streamlined communication and decision-making processes.

InspiRD, Inc. **Jan. 2022 – Oct. 2022**
Machine Learning, Modeling and Simulation Engineer *USA*

- Engineered machine learning models for both regression and classification tasks, seamlessly integrating advanced adaptive sampling techniques to enhance efficiency, while concurrently devising numerical statistical techniques for in-depth similarity analysis of models and parameter distributions.
- Led the development of machine learning models tailored for complex multi-fidelity simulations, achieving a remarkable 500x acceleration in computational speed. Additionally, implemented multi-fidelity transfer learning and physics-informed machine learning (PIML), resulting in a notable 65x speed increase and a 55% enhancement in accuracy.
- Innovated a dimension reduction technique utilizing Autoencoders, and rSVD, building upon variational autoencoder models. This approach accelerated similarity search using in-house BERT models, resulting in a quadratic speed-up and a minimum 12x decrease in memory usage.
- Contributed to the development of Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) proposals, showcasing a commitment to advancing innovative research initiatives.

Neue Materialien Fürth GmbH **May 2015 – Aug. 2015**

Research Engineer (Intern) *Fürth, Germany*

- Developed a user-friendly graphical user interface (GUI) utilizing MATLAB to facilitate multi-objective optimization with a focus on statistical design of experiments, streamlining the process and enhancing accessibility.
- Applied advanced data mining techniques to elevate the efficiency of Kriging models by an impressive 23%. This optimization specifically targeted electron beam melting processes for metallic materials, contributing to increased precision and effectiveness in manufacturing.

Publications

Journal Publications

1. **Kapusuzoglu, B.**, Mahadevan, S., Matsumoto, S., Miyagi, Y., Taba, S., Watanabe, D. Adaptive Surrogate Modeling for High-Dimensional Spatio-Temporal Output. Structural and Multidisciplinary Optimization 65.10, 2022: 300. doi: 10.1007/s00158-022-03402-x.
2. **Kapusuzoglu, B.**, Mahadevan, S., Matsumoto, S., Miyagi, Y., Taba, S., Watanabe, D. Multi-Level Bayesian Calibration of a Multi-Component Dynamic System Model. ASME JCISE, 1-15, 2022. doi: 10.1115/1.4055315.
3. **Kapusuzoglu, B.**, Nath, P., Sato, M., Mahadevan, S., Witherell, P. Multi-Objective Optimization Under Uncertainty of Part Quality in Fused Filament Fabrication. *ASME J. Risk Uncertainty Part B*; 8(1): 011112, 2022. doi: 10.1115/1.4053181.
4. **Kapusuzoglu, B.**, Mahadevan, S. Information fusion and machine learning for sensitivity analysis using physics knowledge and experimental data. *Reliability Engineering and System Safety, Special Issue: Sensitivity Analysis of Model Outputs*, 2020. doi: 10.1016/j.ress.2021.107712.
5. **Kapusuzoglu, B.**, Mahadevan, S. Physics-Informed and Hybrid Machine Learning in Additive Manufacturing: Application to Fused Filament Fabrication. *JOM*, 1-11, 2020. doi: 10.1007/s11837-020-04438-4.
6. **Kapusuzoglu, B.**, Sato, M., Mahadevan, S., Witherell, P. Process optimization under uncertainty for improving the bond quality of polymer filaments in Fused Filament Fabrication. *Journal of Manufacturing Science and Engineering*, 1-46, 2020. doi: 10.1115/1.4048073
7. Karve, P. M., Guo, Y., **Kapusuzoglu, B.**, Mahadevan, S., Haile, M. A. Digital twin approach for damage-tolerant mission planning under uncertainty. *Engineering Fracture Mechanics*, 106766, 2020. doi: 10.1016/j.engfracmech.2019.106766
8. Oskay, C., Su, Z., **Kapusuzoglu, B.**. Discrete eigenseparation-based reduced order homogenization method for failure modeling of composite materials. *Computer Methods in Applied Mechanics and Engineering*, 112656, 2019. doi: 10.1016/j.cma.2019.112656

Conference Publications

1. Sarwar, Z., Panda, A., Thérien, B. Rawls, S., Das, A., Balasubramaniam, K., **Kapusuzoglu, B.**, Zhang, S., Sahu, S., Naphade, M., Chakraborty, S. StructMoE: Structured Mixture of Experts Using Low Rank Experts. *NeurIPS Efficient Natural Language and Speech Processing Workshop. PMLR* (p. 182-193), 2024. https://proceedings.mlr.press/v262/sarwar24a.html.
2. **Kapusuzoglu, B.**, Guo, Y., Mahadevan, S., Matsumoto, S., Miyagi, Y., Taba, S., Watanabe, D. Dimension Reduction for Efficient Surrogate Modeling in High-Dimensional Applications. *AIAA SCITECH 2022 Forum* (p. 1440), 2022. doi: 10.2514/6.2022-1440.
3. Karve, P. M., Guo, Y., **Kapusuzoglu, B.**, Mahadevan, S., Haile, M. A. Fatigue Crack Growth Diagnosis and Prognosis for Damage-Adaptive Operation of Mechanical Systems. *Model Validation and Uncertainty Quantification*, Volume 3 (pp. 233-236). Springer, Cham., 2020.
4. Karve, P. M., Guo, Y., **Kapusuzoglu, B.**, Mahadevan, S., Haile, M. A. Resilience-enhancing operations of aerostructural systems under uncertainty: a digital twin approach. *AIAA SCITECH 2020 Forum*, 2020.

Skills

Large Language Models & NLP: Pre-training & Fine-tuning (Instruction Tuning, RLHF), Parameter-Efficient Training (LoRA, QLoRA, Adapter), Model Optimization (Model Merging, Sparse Training), Distributed Training (Model/Data Parallelism)
Machine Learning: PyTorch, TensorFlow, HuggingFace Transformers, Megatron, Keras, scikit-learn, pandas, numpy
MLOps & Infrastructure: Distributed Training (Kubernetes, Docker), Cloud Computing (AWS), CI/CD Pipelines, Model Monitoring
Programming & Development: Python, SQL, Git, VSCode, Jupyter, LaTeX, C++
Languages: Turkish (native), English (fluent), German (intermediate)

Awards & Certificates

- Neural Networks and Deep Learning, *Coursera*
- Structuring Machine Learning Projects, *Coursera*
- Improving Deep Neural Networks, *Coursera*
- Pao Chung Chen Fellowship
- Erasmus Mundus Scholarship

Leadership / Extracurricular

Civil and Environmental Engineering Graduate Student Council **Aug. 2019 – Aug. 2020**
Career Development Representative *Vanderbilt University*

- Cultivated valuable career-related work experiences, elevated student engagement, and imparted effective job searching techniques to foster professional development and success.

Model United Nations (MUN) **Sep. 2012 – Apr. 2013**
Delegate & Press Member *Ankara, Turkey*

- Designed brochures using Adobe InDesign, Photoshop and Illustrator, and attended several MUN conferences.