

Khoa D. Doan

🏠 khoadoan.me

✉ khoadoan@vt.edu

ABOUT

I am interested in developing computational frameworks that enable existing complex/deep models to be more suitable for practical uses. I focus on improving the following aspects of existing models: (i) training, (ii) inference, (iii) realistic assumptions, and (iv) security understanding. Most of my ML/AI solutions center around generative-based approaches that have low computational complexity and require less human effort. Read more about my interests [HERE](#).

EDUCATION

Ph.D in Computer Science

Virginia Tech (VT)

Virginia, USA

August 2021

MS in Computer Science

University of Maryland, College Park (UMCP)

College Park, Maryland, USA

August 2010-May 2015

BS in Computer Science with a Minor in Mathematics

Webster University

Saint Louis, Missouri, USA

August 2003-May 2006

INDUSTRY EXPERIENCE

Research Scientist – Baidu Research USA, Seattle, WA

2020-

- Perform fundamental research in Generative Modeling, especially Generative Adversarial Models (GANs) and Deep Energy-based Generative Models, Deep Hashing Models (including Billion-scale Search) and AI Security.

Researcher – Criteo AI Lab, Palo Alto, CA

2019-2020

- Perform research in Generative Modeling, especially Generative Adversarial Models (GANs). Besides investigating their theoretical aspects, I studied the applications of generative models into several problem domains, including computational look-alike modeling in advertising, learning to hash for text and images, and regression.

Senior Data Scientist – Verve Mobile/Carlsbad, CA

2015-2019

- Be the principle data scientist behind Verve's Momentum-product (<https://www.verve.com/products/momentum/>) which build highly-scalable statistical models of measuring the effectiveness of advertising campaigns. This is a fully-automatic platform that aggregates user-interaction data from multiple, large-scale data sources, build counterfactual models to estimate the *incremental KPIs* of the selected advertising campaigns, and exposes the analytical results to Verve's customers.
- Be a part of the Audience Modeling Team which is responsible for building production-ready predictive models such as Age, Gender, Home Associations, etc... Responsible for modeling the core algorithms behind Verve's Audience System.
- Be a part of the team that builds the core, patented Home Association algorithm at Verve (see below).
- Be a part of the Data Infrastructure Team that design the underlying architecture of the company's data-warehouse/data-analytic platform on Amazon AWS, which is later migrated to Google GCP.

Faculty Research Associate – University of Maryland, College Park, MD

2012-2015

- Participate in high performance, distributed system research on NASA large-scale, remote-sensing data.
- Involve in Automated Event Service (AES) and Data Environment For Rapid Exploration And Characterization Of Hydrometeorological Organized Systems (DERECHOS) research projects at NASA.
- Responsible for development of large scale data (statistical) analysis algorithms in distributed shared-nothing environments using standard tools such as Hadoop, Spark, Cassandra, and SciDB (a multidimensional in-memory array-based scientific database).

Senior Developer – Aquilent/Laurel, MD (acq. by Booz Allen Hamilton) 2008-2012

- Lead application development and maintenance activities for enterprise applications at US Department of Energy and Department of Veteran Affairs. This effort involves the integration of Open Source API's into existing system architecture.
- At the Department of Veteran Affairs, I am the main developer behind the Enrollment System Redesign (ESR) Project. ESR is the new implementation of the current standalone Enrollment System at VA. The effort is to establish a fully integrated environment which will replace not only the current Enrollment System, but also all the standalone systems at VA. The development environment at VA is: BEA Weblogic 8.1 and 10.0, Oracle 10g, J2EE, Spring, Hibernate, Struts, BizFlow BPM, ILog Rule Engine, Jasper Report.

Software Developer – Insight Distribution Software/Portland, OR 2007-2008

- Design and develop Tomra Data Interchange Package to interact with the vendor's system and our legacy application.
- Design and develop the customizable database model

Software Engineer – Conde Nast Publication/New York, NY 2006-2007

HONORS AND SCHOLARSHIPS

Criteo Research Award – Virginia Tech 2018

NSF Urban Computing Fellowship – Virginia Tech 2016-2017

Graduation Honor, Summer Cum Laude – Webster University 2006

Regional ACM Collegiate Computing Contest Honor – Webster University 2004

University Scholarship – Webster University 2003

PUBLICATIONS

1. **K. D. Doan**, Y. Lao, & P. Li, "Backdoor Attacks by the Most Flexible Adversary ". *Under Submission*.
2. **K. D. Doan**, Y. Peng, & P. Li, "One Loss for Quantization: Deep Hashing with Discrete Wasserstein Distributional Matching". *2022 Conference on Computer Vision and Pattern Recognition (CVPR)*.
3. **K. D. Doan**, Y. Lao, Y. Peng, & P. Li, "On the Vulnerability of Vision Transformer against Backdoor Attacks". *Under Submission*.
4. **K. D. Doan**, J. Xie, W. Zhao, & P. Li, "Generative Cooperative Hashing Network". *Under Submission*.
5. Y. Peng, Y. Lao, **K. D. Doan**, & P. Li, "On the Robustness of Vision Transformers to Black-Box Adversarial Attacks". *Under Submission (ECCV)*.

6. **K. D. Doan**, S. Tan, W. Zhao, & P. Li, "Asymmetric Hashing for Fast Ranking via Neural Network Measures". *Under Submission (SIGIR)*.
7. **K. D. Doan**, Y. Lao, & P. Li, "Backdoor Attack with Imperceptible Input and Latent Modification". *Thirty-fifth Conference on Neural Information Processing Systems 2021 (NeurIPS)*.
8. **K. D. Doan**, Y. Lao, W. Zhao, & P. Li, "LIRA: Learnable, Imperceptible and Robust Backdoor Attacks". *2021 IEEE International Conference on Computer Vision (ICCV)*.
9. **K. D. Doan**, S. Manchanda, S. Mahapatra, & CK. Reddy, "Interpretable Graph Similarity Computation via Differentiable Optimal Alignment of Node Embeddings", *In Proceedings of International ACM SIGIR conference on research and development in Information Retrieval 2021 (SIGIR)*
10. **K. D. Doan**, F. Wang, S. Manchanda, S. Selvaraj, A. Bhowmik & CK. Reddy, "Image Generation via Minimizing Fréchet Distance in Discriminator Feature Space. *Under Submission (TPAMI Journal)*.
11. **K. D. Doan**, S. Badirli, & CK. Reddy, "Generative Hashing Network". *Under Submission (Cybernetics Journal)*.
12. **K. D. Doan**, S. Manchanda, S. Badirli, & C. K. Reddy. "Image Hashing by Minimizing Discrete Component-wise Wasserstein Distance". *Under Submission (Cybernetics Journal)*.
13. S. Badirli, X. Liu, **K. D. Doan**, Z. Xing, A. Bhowmik & SS. Keerthi. Gradient Boosting Neural Networks: GrowNet. <https://arxiv.org/abs/2002.07971>.
14. S. Manchanda, **K. Doan** & SS Keerthi. Regression via Implicit Models and Optimal Transport Cost Minimization. <https://arxiv.org/abs/2003.01296>.
15. **K. Doan** & C. K. Reddy. Efficient Implicit Unsupervised Text Hashing using Adversarial Autoencoder. *In Proceedings of The Web Conference, 2020 (WWW)*.
16. S. Manchanda, P. Yadav, **K. Doan**, & K. Sathiya. Targeted display advertising: the case of preferential attachment. *In Proceedings of International Conference on Big Data, 2019*.
17. **K. Doan**, P. Yadav & C. K. Reddy. Adversarial Factorization Autoencoder for Look-alike Modeling. *In Proceedings of ACM International Conference on Information and Knowledge Management, 2019 (CIKM)*.
18. **K. Doan**, G. Yang & C. K. Reddy. An Attentive Spatio-Temporal Neural Model for Successive Point Of Interest Recommendation. *In Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2019*.
19. **K. Doan**, A. O. Oloso, K. S. Kuo, T. L. Clune, H. Yu, B. Nelson, & J. Zhang. Evaluating the impact of data placement to Spark and SciDB with an earth science use case. *In Proceedings of IEEE International Conference on Big Data, 2016*.
20. **K. Doan**, A. O. Oloso, K. S. Kuo & T. L. Clune (2014, December). Performance comparison of big-data technologies in locating intersections in satellite ground tracks. *In Proceedings of ASE BigData Conference, 2014*.

ACADEMIC SERVICE

Program Committee (Invited)

- International Conference on Learning Representations (**ICLR**): 2021-2022
- Conference on Neural Information Processing Systems (**NeurIPS**): 2020-2022
- International Conference on Machine Learning (**ICML**): 2020-2022
- Conference on Computer Vision and Pattern Recognition (**CVPR**): 2020-2022
- International Conference on Computer Vision (**ICCV**): 2021
- European Conference on Computer Vision (**ECCV**): 2020, 2022
- AAAI Conference on Artificial Intelligence (**AAAI**): 2021-2022
- IEEE International Conference on Big Data (**BigData**): 2020, 2022
- 1st International Workshop on Industrial Recommendation Systems (**IRS**): 2020-2021

Journal Reviewer

- ACM Transactions on Internet Technology (**TOIT**): 2020
- ACM Transactions on Knowledge Discovery from Data (**TKDD**): 2018, 2019, 2020, 2021

Conference Reviewer

- ACM SIGKDD International Conference on Knowledge discovery and data mining (**KDD**): 2017-2019
- ACM International Conference on Information and Knowledge Management (**CIKM**): 2017-2019
- ACM International Conference on Web Search and Data Mining (**WSDM**): 2017-2019
- The Web Conference (WWW): 2017-2019
- International Joint Conference on Artificial Intelligence (**IJCAI**): 2017-2019

TALKS/PRESENTATIONS

- **K. Doan**. Toward Practical Machine Learning Applications with Generative Models: Data Generation and Beyond (Seminar). Singapore Management University (SMU), Singapore. 2022.
- **K. Doan**. Information Retrieval with Deep Hashing (*Seminar*). VinAI, Vietnam, 2022.
- **K. Doan**. Generative models meet similarity search (*Seminar*). Baidu Cognitive Computing Lab, USA, 2020.
- **K. Doan**, B. Cave & C. K. Reddy (*Poster*). CrimeLab: A data-driven approach. Virginia Tech Urban Computing Day, 2017.
- H. Avik, G. Takahara & **D. Khoa** (*Talk*). Social Media Analytics using Bayesian Multistate Intensity Models. 43rd Annual Meeting of the Statistical Society of Canada. <https://bit.ly/3extrE>

PATENTS

B. E. Crook, **K. Doan**, G. K. Ng, C. G. Nicotra, M. J. Wrona. Systems, methods, and apparatus for reverse geocoding. US Patent US20160330592A1.

REFERENCES

Chandan K. Reddy, *Professor*

Department of Computer Science, Virginia Tech

Website: <https://people.cs.vt.edu/~reddy/>

Email: reddy@cs.vt.edu

Ping Li, *Professor & Head of Research*

Cognitive Computing Laboratory, Baidu Research

Website: <http://research.baidu.com/People/index-view?id=111>

Email: liping11@baidu.com

Sathiya Keerthi Selvaraj, *Principal Staff Scientist*

AI Group, LinkedIn

Website: <http://www.keerthis.com/>

Email: keselvaraj@linkedin.com

Gary Ng, *VP of Data Science & Engineering*

Verve Mobile

Website: <https://www.linkedin.com/in/gkcng/>