

# CUONG (ERIK) DAO

Stockholm, Sweden | +46 (0) 76 766 6457 | [cuong@neuraltalks.io](mailto:cuong@neuraltalks.io) | [LinkedIn](#) | [GitHub](#) | [Google Scholar](#) | [Homepage](#)

## EXPERIENCE

---

### Machine Learning Engineer Intern

Spotify

Jun 2022 – Present

Stockholm, Sweden

- Owned the ranking model distillation project inside the search ranking team; implement and experiment different approaches to create ranking models that are lighter but maintain on-par performance with production models
- Implemented a package to measure the inference speed of ranking models in Java and Scala Scio
- Assisted team in upgrading codebase and solving technical debt tickets

### Research Engineer, Division of Robotics, Perception and Learning

KTH Royal Institute of Technology

Jul 2021 – Mar 2022

Stockholm, Sweden

- Devised a pipeline to leverage vision transformer features for weakly-supervised semantic segmentation; investigated and pinpoint the weakness of current transformer-based models for segmentation
- Established and executed a research plan on semi-supervised semantic segmentation; designed transformer-based networks to find optimal settings for semi-supervised models to achieve on-par performance with fully supervised ones devised and implemented novel deep network architecture in PyTorch, trained on a large nVidia A100s GPU cluster
- Mentored an undergraduate intern with little programming experience to work efficiently in building segmentation pipeline in PyTorch, on GPU clusters within two months

### Lead Software Engineer, Insight Data Science

Series-B, Y-Combinator Startup

Apr 2018 – Feb 2021

San Francisco, USA

- Initiated and led a team of 5 engineers to build a professional network, a messaging platform (with Python Django, AWS, PostgreSQL), serving 4000+ data science professionals with 4 million messages quarterly
- Established best engineering practices (e.g., 8-hour code review, automated testing, continuous integration); communication and development practices (e.g., async communication, agile) and reinforced the culture across team
- Designed and built analytic pipelines, dashboards for all metrics of company's products using PostgreSQL, Redis, Apache Superset; leveraged data to accelerate bringing job opportunities to company's fellows amid COVID-19

### Research Intern, Image and Video Understanding Lab

King Abdullah University of Science and Technology

Aug 2017 – Jan 2018

Jeddah, Saudi Arabia

- Co-authored a paper on a novel Lasso-based text similarities metric and a large-scale dataset for measuring the robustness of VQA models; published at AAI 2019 (Oral presentation, acceptance rate: 16% of 7095 submissions) [\[Link\]](#)
- Implemented a spatiotemporal pooling operation for video frames in TensorFlow; devised deep network architecture for weakly-supervised localization of video action localization that was on-par with supervised methods [\[Link\]](#)
- Designed a tool to crowd-source annotation data for videos on Amazon Mechanical Turk and implemented a deep architecture for fish detection in underwater video footage from Australian coasts with 93% accuracy

## TECHNICAL SKILLS

---

- **Machine Learning:** PyTorch, PyTorch Lightning, Scikit-Learn, Numpy, Pandas, Apache Spark, Kubeflow, TensorFlow, CNN, RNN, Vision Transformer, Object Detection, Semantic Segmentation, Distributed training
- **Programming Languages:** Python, SQL, Java, Typescript
- **Cloud Computing:** Amazon AWS (EC2, S3, ElastiCache, Elasticsearch), Google Cloud (GCS, Dataflow)
- **Web Development:** Python Django, ReactJS, Redux, Git, PostgreSQL, Redis, Java Spring, Docker

## EDUCATION

---

### KTH Royal Institute of Technology

Master's Programme in Machine Learning

Aug 2020 – Present

Stockholm, Sweden

*Thesis: Incorporating Sparse Attention Mechanisms to Transformer for Object Detection in Images*

### Hanoi University of Science and Technology

Engineering Degree in Information and Communication Technology

Aug 2011 – Mar 2017

Hanoi, Vietnam