

# Abram Yorde

Englewood, OH abram.yorde@outlook.com (937) 507-1958 Abram Yorde AbramYorde

## Objective

---

Enable internal and external customer success via data and analytics solutions and models.

Ensure these customer's success and guide them from reactionary to proactive and guided decision making.

## Education

---

<b>MS</b>	<b>Georgia Institute of Technology</b> , Analytics	Atlanta, GA
	• GPA: 3.66 / 4.00	Jan 2019 – Dec 2021
<b>BS</b>	<b>Wright State University</b> , Mechanical Engineering	Dayton, OH
	• GPA: 3.71 / 4.00	Jan 2015 – Dec 2017

## Experience

---

<b>Copeland</b> , Senior Data Engineer	Remote
• Lorem Epsum	Apr 2024 – present 1 year 9 months
<b>Copeland</b> , Senior Data Scientist	Sidney, OH
• Lorem Epsum	Jan 2023 – Aug 2024 1 year 8 months
<b>Emerson Climate Technologies</b> , Data Scientist	Sidney, OH
• Lorem Epsum	Jan 2020 – Jan 2023 3 years 1 month
<b>Emerson Climate Technologies</b> , Refrigeration Project Engineer - Data Science	Sidney, OH
• Created on-device neural network compression pipeline deployed across 50M+ devices	Jan 2018 – Jan 2020
• Filed 2 patents on efficient model quantization techniques for edge inference	2 years 1 month
<b>Emerson Climate Technologies</b> , Refrigeration Internships (Multiple Departments)	Sidney, OH
• Implemented novel self-supervised learning framework for low-resource language modeling	Aug 2015 – Jan 2018
• Research integrated into Azure Cognitive Services, reducing training data requirements by 60%	2 years 6 months

## Projects

---

<b>FlashInfer</b>	Jan 2023 – present
Open-source library for high-performance LLM inference kernels	
• Achieved 2.8x speedup over baseline attention implementations on A100 GPUs	
• Adopted by 3 major AI labs, 8,500+ GitHub stars, 200+ contributors	
<b>NeuralPrune</b>	Jan 2021
Automated neural network pruning toolkit with differentiable masks	
• Reduced model size by 90% with less than 1% accuracy degradation on ImageNet	
• Featured in PyTorch ecosystem tools, 4,200+ GitHub stars	

## Selected Honors

---

- MIT Technology Review 35 Under 35 Innovators (2024)
- Forbes 30 Under 30 in Enterprise Technology (2024)
- ACM Doctoral Dissertation Award Honorable Mention (2023)
- Google PhD Fellowship in Machine Learning (2020 – 2023)
- Fulbright Scholarship for Graduate Studies (2018)

## Skills

---

**Languages:** Python, C++, CUDA, Rust, Julia

**ML Frameworks:** PyTorch, JAX, TensorFlow, Triton, ONNX

**Infrastructure:** Kubernetes, Ray, distributed training, AWS, GCP

**Research Areas:** Neural architecture search, model compression, efficient inference, multi-agent RL

## Patents

---

1. Adaptive Quantization for Neural Network Inference on Edge Devices (US Patent 11,234,567)
2. Dynamic Sparsity Patterns for Efficient Transformer Attention (US Patent 11,345,678)
3. Hardware-Aware Neural Architecture Search Method (US Patent 11,456,789)

## Invited Talks

---

4. Scaling Laws for Efficient Inference — Stanford HAI Symposium (2024)
3. Building AI Infrastructure for the Next Decade — TechCrunch Disrupt (2024)
2. From Research to Production: Lessons in ML Systems — NeurIPS Workshop (2023)
1. Efficient Deep Learning: A Practitioner’s Perspective — Google Tech Talk (2022)

## Any Section Title

---

You can use any section title you want.

You can choose any entry type for the section: TextEntry, ExperienceEntry, EducationEntry, PublicationEntry, BulletEntry, NumberedEntry, or ReversedNumberedEntry.

Markdown syntax is supported everywhere.

The design field in YAML gives you control over almost any aspect of your CV design.

See the [documentation](#) for more details.