

# Kimberly J. Wilber

Machine learning research infrastructure engineer.

kimmy@kjwilber.org  
<https://kjwilber.org>  
[GitHub](#) · [LinkedIn](#) · [Scholar](#)

## Core expertise

**ML research infrastructure** Data pipelines, evaluation tooling, experiment frameworks, training and inference infrastructure, loss functions and gradient descent optimization, model debugging / interpretability tools

**LLMs** Mechanistic interpretability, embedding systems, approximate nearest neighbor search, linear probes

**On-device ML** quantization, latency optimization, memory optimization, model distillation

**Datasets** Large-scale dataset curation and processing

**Research process** Crowdsourcing and human-in-the-loop ML pipelines, experiments, academic paper writing

## Open-source contributions

- Contributor to various projects including Racket and Node.js.
- [GitHub profile](#): 196 followers.
- Repositories I created have 200 forks and 700 stars.

## Professional Experience

**Technical co-founder, Krnel.ai, New York, NY**

2024 – Present

- Built mechanistic-interpretability tooling for LLM safety enforcement. PoC deployments with 2 hedge funds, 1 hardware vendor, and 1 systems integrator
- Built “Krnel MI Guardrails,” a content moderation safety filter achieving 25 × fewer false alarms than LlamaGuard and 33% lower latency. See [code](#), [blog post](#)
- Designed and implemented [Krnel-graph](#), a content-addressable computation graph for mechanistic-interpretability workflows.
  - Krnel-graph is widely used within all of the company’s research projects
- Led the open-source process for Krnel-graph, including drafting documentation and shepherding research contributions → 22 stars, ≈200 monthly downloads

**Software Engineer, Google AI, New York, NY**

2018 – 2024

- Technical lead for on-device outdoor monocular depth estimation and segmentation models used by [Project Guideline](#) and [ARCore](#). This work launched in ARCore 1.36 and enabled geospatial API and the MediaPipe on-device depth plugin, both demonstrated at Google I/O 2023.
  - Improved segmentation accuracy **0.9296** → **0.9400 mIOU** while reducing latency **28.7ms** → **23ms** on mobile hardware.
  - Feedback from L5 research scientist: “As a direct result of Kimberly’s work, the MediaPipe team was able to use monocular depth as an ‘on-device plugin’ for their flagship image generation task ... *First real-time on-device monocular depth solution open sourced via MediaPipe*”
  - Feedback from L6 research scientist, partner team: “Invaluable contributions in up-leveling monocular depth across multiple product surfaces at Google ... with greatly improved generalization to outdoor scenes as well as latency and memory savings!”
- Co-led development of *Debug Server v2*, an embedding-model inspection platform adopted by 4+ Google teams (Play Ads, RankEmbed, Intent Mining, LocalEmbed) for debugging large-scale retrieval systems.
  - This work consolidated debug infrastructure, reducing resource usage by **≈15 TB RAM** and simplifying deployment to a one-command setup
  - Feedback from L6 peer: “Kimberly’s deep expertise in developing embedding models has positioned her as a *trusted authority and go-to expert, even beyond her own team*”

## Technical stack

**Fluent languages** Python, C, HTML/CSS/JS, POSIX and CLI tools, L<sup>A</sup>T<sub>E</sub>X

**ML frameworks** PyTorch, HuggingFace Transformers, sklearn, inference tooling like llama-cpp, ollama, vLLM

**Data** Parquet, DuckDB/LanceDB, Postgres, SQLite, Pandas, NumPy

**Web** Flask, FastAPI, Node.js, Vue, React

**Sysadmin** Debian/Ubuntu server clusters, Ansible, Docker, AWS EC2, GCP, and self-hosted infrastructure

- Sole maintainer of *VisionKit Model Zoo and Studio*, supporting **120+ internal ML models** and enabling deployment of new model versions across Google Research teams
- Co-led creation of [SANPO](#), a large-scale egocentric scene understanding dataset used for depth estimation research, accessibility products, and ARCore models.
  - Our dataset has 10 citations in research papers,  $\approx 50$  monthly downloads, used by 3+ teams across Google for research and product development
- Co-designed **Cross-Example Softmax**, a novel retrieval loss improving AUC-PR from 0.14  $\rightarrow$  0.20 on Conceptual Captions
  - Built all ML experimentation infrastructure (data pipelines, training tools, hyperparameter search), **improving research throughput 5 – 10  $\times$**  and reducing iteration cycles from days to hours.

**Volunteer publication assistant, Computer vision foundation, Remote** 2013 – Present

- The [CVF Open Access Archive](#) serves most CVPR, ICCV, and WACV papers to the computer vision community since 2013.
- I own the PDF processing pipeline used to build CVPR/ICCV/WACV conference proceedings.
- Our stack (Python, Ghostscript, qPDF) adds page numbers, an attribution banner, and proper PDF metadata with automatic checks for common quality problems (figures overlapping margins, etc).
- We're now able to process 2,700+ papers with a 3-day turnaround time and less than 5 manual corrections per conference thanks to the automation pipeline I built.

**Research Assistant, Cornell Tech, New York, NY** 2014 – 2018

- Conducted computer vision research on perceptual similarity, crowdsourcing, and object recognition.
- Helped establish and maintain the new vision group's presence at Cornell.
- Served as TA for classes including four semesters of "CS5785 Modern Analytics."

**Summer Intern, Google Photos Team, Mountain View, CA** 2017

- Implemented and tested tools to make it easier for ML engineers to prototype UI interactions.
- These tools helped shape the feature roadmap for [Google Photos Sharing Suggestions](#).

**Summer Intern, Adobe Research, San Jose, CA** 2016

- Curated [BAM](#), one of the first large-scale collections of professional commercial artwork, intended for ML object classification and emotion understanding.
- Built a data loader system in Python and Redis to quickly analyze millions of images for ML training and inference workloads, speeding up training by 5 $\times$ .

**Summer Intern, Dropbox Photos Team, San Francisco, CA** 2014

- Conducted product-focused computer vision research.
- Introduced our team to more efficient tools and technologies.
- Maintained a computer vision evaluation and experimentation pipeline for crowdsourced data collection of face images.

**Research Assistant, University of California, San Diego, San Diego, CA** 2013

**Software Engineer, Securics, Inc., Colorado Springs, CO** 2012 – 2013

**Assistant Researcher, Vision and Security Technology (VAST) Laboratory at UCCS, Colorado Springs, CO** 2009 – 2013

**Summer Researcher, NSF REU Program, University of Colorado Colorado Springs, Colorado Springs, CO** 2011

**NSF RAHSS High School Intern, Securics, Inc., Colorado Springs, CO** 2009 – 2010

## Education

- 2014 – 2018 **Ph.D. in Computer Science, Cornell Tech**  
 Advised by Dr. Serge Belongie. Thesis topic: combining computer vision and crowdsourcing techniques. Supported by the National Science Foundation Graduate Research Fellowship (NSF GRFP)
- 2009 – 2013 **Bachelor of Innovation in Computer Science, University of Colorado Colorado Springs**  
 Supported by the Kane Family Foundation Scholarship

## Conferences, Awards, Scholarships

- 2011 – Present **Student Volunteer at several conferences:**  
 CVPR (IEEE Conference on Computer Vision and Pattern Recognition) 2011, 2012, 2013, 2014, 2015  
 WACV (Winter Conference on Applications of Computer Vision) 2012, 2013, 2014
- 2013 – 2016 **National Science Foundation GRFP Awardee, UCSD/Cornell**
- 2010 – 2013 **Dean's List, UCCS**
- 2010 – 2013 **Kane Family Foundation Scholarship Recipient, Full tuition and books, UCCS**
- 2010 – 2013 **Braxton Scholarship Recipient, UCCS**

## Selected Publications

Note that some work before 2018 is published under a previous name. See [here](#) for the unabridged list.

- 2024 [PolyMaX: General Dense Prediction with Mask Transformer](#)  
 Xuan Yang; Liangzhe Yuan; **Kimberly Wilber**; Astuti Sharma; Xiuye Gu; Siyuan Qiao; Stephanie Debats; Huisheng Wang; Hartwig Adam; Mikhail Sirotenko; Liang-Chieh Chen. *Winter Conference on Applications of Computer Vision (WACV 2024)*
- 2023 [SANPO: A Scene Understanding, Accessibility, Navigation, Pathfinding, Obstacle Avoidance Dataset](#)  
 Sagar M. Waghmare; **Kimberly Wilber**; Dave Hawkey; Xuan Yang; Matthew Wilson; Stephanie Debats; Cattalyya Nuengsigkapan; Astuti Sharma; Lars Pandikow; Huisheng Wang; Hartwig Adam; Mikhail Sirotenko. *ArXiv*
- 2019 [Understanding Image Quality and Trust in Peer-to-Peer Marketplaces](#)  
 Xiao Ma; Lina Mezghani; **Kimberly Wilber**; Hui Hong; Robinson Piramuthu; Mor Naaman; Serge Belongie. *Winter Conference on Applications of Computer Vision (WACV 2019)*
- 2017 [BAM! The Behance Artistic Media Dataset for Recognition Beyond Photography](#)  
**M. Wilber**; Chen Fang; Hailin Jin; Aaron Hertzmann; John Collomosse; Serge Belongie. *International Conference on Computer Vision (ICCV 2017)*
- 2016 [Residual Networks Behave Like Ensembles of Relatively Shallow Networks](#)  
 Andreas Veit; **M. Wilber**; Serge Belongie. *Neural information processing systems (NIPS 2016)*
- 2016 [Can we still avoid automatic face detection?](#)  
**M. Wilber**; Vitaly Shmatikov; Serge Belongie. *Winter Conference on Applications of Computer Vision (WACV 2016)*
- 2015 [Learning Concept Embeddings with Combined Human-Machine Expertise](#)  
**M. Wilber**; Iljung Sam Kwak; Serge Belongie. *International Conference on Computer Vision (ICCV 2015)*
- 2014 [Cost-Effective HITs for Relative Similarity Comparisons](#)  
**M. Wilber**; Iljung Sam Kwak; Serge Belongie. *AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2014)*
- 2014 [Good Recognition is Non-Metric](#)  
 Walter J. Scheirer; **M. Wilber**; Michael Eckmann; Terry Boulton. *E. Pattern Recognition 47 (8), 2014*
- 2013 **Best paper award: [Animal Recognition in the Mojave Desert: Vision Tools for Field Biologists](#)**  
**M. Wilber**; Walter J. Scheirer; Phil Leitner; et. al.. *Workshop on Applications of Computer Vision (WACV 2013)*
- 2012 [PRIVV: Private Remote Iris Authentication with Vaulted Verification](#)  
**M. Wilber**; Walter J. Scheirer; Terry Boulton. *Conference on Computer Vision and Pattern Recognition Biometrics Workshop (CVPR 2012)*