

# Nikhil Kapila

☎ +971503564790 | 🌐 nkapila.me | 🐙 github | ✉ nkapila6@gatech.edu

---

## Research Interests

My research interests lie primarily in optimization and inductive biases of deep learning architectures, such as: multimodal image and text models, and generative models.

### Optimization

In optimization, I would like to look into techniques like knowledge distillation, neural architectural search and pruning to create smaller models that outperform their larger counterparts. I recently explored this with my project `CNNtention` [1] wherein we augmented attention layers in a deep ResNet model to see if refined feature maps through attention yield any benefit for an image classification task on MNIST and CIFAR-10. For upcoming work, in my personal downtime, I aim to investigate knowledge distillation techniques, i.e. some form of intermediate feature map alignment using a bigger model.

Apart from architectural changes, compute efficiency can be improved through alternative methods. It could be something simple like decomposing the attention matrix (bilinear attention) by approximating the full  $wx$  to save on compute. Another example is replacing traditional matmuls with Hadamard products as seen in certain GRU implementations [2].

Then, there exist architectures such as WANNs [3] which has some similarity with [2] where weights are fixed but architectural changes are focused on through a genetic algorithm like approach which surprisingly gives decent performance.

### Multimodal and Generative

As I dive more into the world of deep learning, I am able to recognize many similarities across modalities. I realize a lot of the ideas from image could be transferred to other modalities such as speech/text with ease.

One example is the case of using Neural Style Transfer for Images [4] and the case of using specific loss functions to dictate these changes [5]. I see a lot of similarities in the audio/speech modality where one could perform voice cloning by performing a similar style transfer approach in audio [6]. Of course, there are many newer pre-trained few/zero-shot speech models that exist now.

It's fascinating to observe how architectures with similar design principles can operate across different modalities by learning appropriate latent representations, as demonstrated by contrastive learning approaches like CLIP that create aligned embedding spaces. The recent emergence of foundation models has further highlighted how architectural decisions can enable powerful cross-modal capabilities, allowing systems to reason across text, images, audio, and other modalities with increasing coherence and sophistication.

### Biases

Another aspect of modeling is having good knowledge of its inductive biases. One paper that I found really intriguing in this regard is when it was found that CNNs inherently have a texture bias but could easily be converted into shape biases through training [7].

### Tabular modalities

Major focus on NLP and computer vision tasks have led to ignorance of tabular data [8] where ensemble methods are still SOTA. New approaches such as TabFN [9] and CARTE [10] that combine deep learning approaches are now creating new interests in these fields.

# Nikhil Kapila

📞 +971503564790 | 🌐 nkapila.me | 🐙 github | ✉ nkapila6@gatech.edu

---

## Research Trajectory

I am due to complete my MS in Fall of 2025. Currently I am working on a few projects as listed below.

### Coursework

- ICD9 disease prediction from unstructured clinical notes of MIMIC-III. Using the FarSight paper as our base [11].

**Personal Projects:** Whenever I'm free from coursework, I work on the below projects to deepen my understanding in generative modeling.

- **Implementing GANs and VAEs:** Better my understanding of generative modeling, how biases of certain losses change the output.
- **Implementation of models such as GPT2:** To get a better understanding of what goes into pre-training large language models.

### Future Projects:

- I plan to take the Education Technology course in the Fall where we are pushed to do literature review and develop publishable work in the areas of EdTech. I am preparing for it now and looking into systems / approaches that can help “induce” socratic learning.

## Research Impact

Research in model optimization, multimodal learning, and inductive biases has significant practical applications. These approaches can democratize AI by enabling efficient deployment on edge devices, enhancing accessibility technologies, and enabling more natural human-computer interaction (Iron Man-like natural language UX).

# Nikhil Kapila

📞 +971503564790 | 🌐 nkapila.me | 🐙 github | ✉ nkapila6@gatech.edu

---

## References

- [1] Nikhil Kapila, Julian Glattki, and Tejas Rathi, “CNNtention: Can CNNs learn better with Attention?,” Dec. 2024. [Online]. Available: <https://arxiv.org/abs/2412.11657v3>
- [2] Rui-Jie Zhu *et al.*, “Scalable MatMul-free Language Modeling,” Jun. 2024. [Online]. Available: <https://arxiv.org/abs/2406.02528>
- [3] Adam Galer and David Ha, “Weight Agnostic Neural Networks,” Jun. 2019. [Online]. Available: <https://arxiv.org/abs/1906.04358>
- [4] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge, “A Neural Algorithm of Artistic Style,” Aug. 2015. [Online]. Available: <https://arxiv.org/abs/1508.06576v2>
- [5] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” Sep. 2023. [Online]. Available: <https://arxiv.org/abs/1603.08155>
- [6] Rongjie Huang, Yi Ren, Jinglin Liu, Chenye Cui, and Zhou Zhao, “GenerSpeech: Towards Style Transfer for Generalizable Out-Of-Domain Text-to-Speech,” Oct. 2022.
- [7] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel, “ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness,” Nov. 2022. [Online]. Available: <https://arxiv.org/abs/1811.12231>
- [8] Marie Brayer, “Should you care about Tabular AI,” Mar. 06, 2025. [Online]. Available: <https://ontheflyinvesting.substack.com/p/should-you-care-about-tabular-ai>
- [9] Noah Hollmann, Samuel Muller, Katharina Eggensperger, and Frank Hutter, “TABPFN: TRANSFORMER THAT SOLVES SMALL TABULAR CLASSIFICATION PROBLEMS IN A SECOND.” [Online]. Available: <https://arxiv.org/pdf/2207.01848>
- [10] Myung Jun Kim, Leo Grinsztajn, and Gael Varoquaux, “Pretraining and Transfer for Tabular Learning.” [Online]. Available: <https://arxiv.org/pdf/2402.16785>
- [11] Tushaar Gangavarapu, G. Krishnan, S. SowmyaKamath, and Jayakumar Jeganathan, “FarSight: Long-Term Disease Prediction Using Unstructured Clinical Nursing Notes,” Jul. 2021.