The American Ophthalmological Society

Established 1864

Frederick H. Verhoeff Lecture

2025

Council Introductory & Editorial Comments

2025 FREDERICK H. VERHOEFF LECTURE

Introductory Comments

AOS Council

It turns out we should have played more video games! An intriguing tidbit shared in Aaron Lee's wonderful lecture on Artificial Intelligence (AI) is that one of the enabling developments for the current AI revolution was the graphics cards developed for computer games. We may not all play video games, but we are using AI more and more, whether we know it or not.

The 2025 Verhoeff Lecture defines many of the terms used to describe AI. Dr. Lee playfully shows how AI can be put to nefarious purposes, e.g., cheating on the ABO Oral Board Examination (the ABO has closed that window). More importantly, he shows how the NIH's program Bridge2AI has established an enormous and growing dataset that will prove a resource for future AI developments.

This lecture was a comprehensive look at how AI is affecting all of the subspecialties in Ophthalmology. We hope you will access the lectures to join in this transformative technology.

2025 FREDERICK H. VERHOEFF LECTURE

The Transformative Role of Deep Learning in Ophthalmology and Vision Science

Aaron Y. Lee, MD, MS

Thank you all so much for this opportunity. I think this will be one of the highlights of my career to give this lectureship, so I really appreciate it. I'll be talking about the transformative role of deep learning in ophthalmology and vision science. These are my financial disclosures (Figure 1).

Figure 1

I wanted to start by giving homage to Dr. Verhoeff. As a resident, I learned about the Verhoeff stain, but I couldn't think of a better way to do it than by asking ChatGPT. I turned on the deep research feature on ChatGPT and asked it to give me a detailed biography. About 20 minutes later, it generated this PDF. If you're curious and want to read it for yourself, this is the QR code that goes to the generated text. What was really kind of amazing was learning about all the contributions he made to the field of ophthalmology. My favorite part of the summary is the last sentence from a review article, and hopefully, nobody in this room will be offended by this statement, but if anyone deserves to be honored as the most influential American ophthalmologist of the 20th century, Dr. Verhoeff certainly does.

I will also note that this named AOS lecture showed up in the ChatGPT summary that you can see in the top part of this paragraph. So it is with great honor that I give this lectureship.

This is the outline of my talk. I'm going to start by giving a brief introduction, talk about some of our early work, and then go into some of the more exciting areas of AI in ophthalmology. Just to level set, we are many decades into what people are calling the fourth industrial revolution. Big data and machine learning are driving and affecting almost every part of human society. Every time you type something into your phone or into Google, algorithms are watching what you're doing and learning from your behavior. I also like to show this slide (Figure 2) because the news media and even peer-reviewed medical literature often get these terms confused.

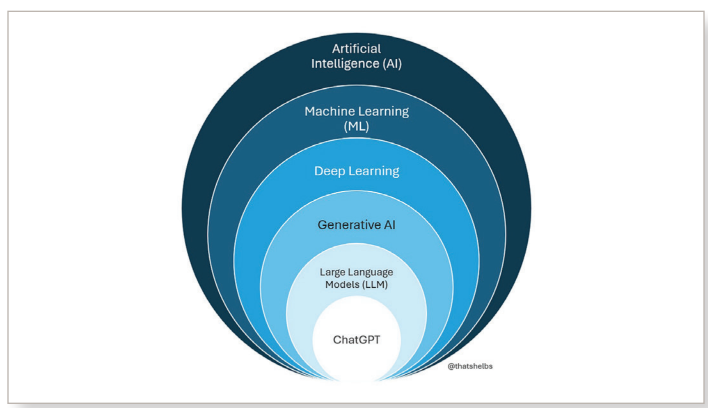

Figure 2

If you read the news articles, it seems like ChatGPT is the entire field of AI, but that is absolutely not true. These are all subsets of each other, and deep learning is probably the underlying technology that drives everything happening today. In my opinion, there are four advances that gave birth to what we call deep learning today.

First was the realization that graphics cards people used to play computer games could be used to do something useful—they could be used as a linear algebra accelerator (Figure 3).

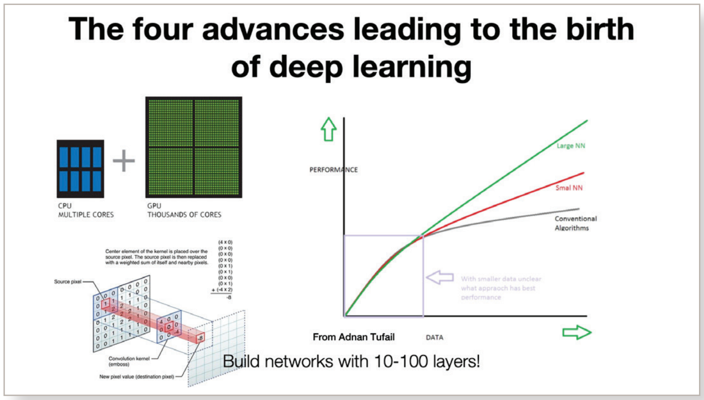

Figure 3

Second was the use of convolutional filters. The first papers describing this came from the 1980s, but they didn't have the computational power to pull them off. The third is the non-linear activation function that people had theorized would be useful. Finally, we entered the era of big data, where our computer hardware had advanced to the point where we could start to collect all the data on the internet. When you put all of these things together, we live in an exciting time. The more data and the larger the computers we have, the more these models seem to exceed human performance in every aspect.

I want to rewind time a little and go back to how I began this journey. It happened right after Russ Van Gelder brought me onto the faculty. I was able to extract all the OCT images from our clinical dataset, and from the EHR, we were able to get all the clinical features attached to these images. The easiest question I could think of was whether deep learning could distinguish between normal OCT scans and AMD ones. From this dataset, I pulled together about 100,000 images and trained a model called VGG 16, which was state-of-the-art at the time. In two days, I constructed this ROC curve (Figure 4). It was really shocking, I couldn't believe how, just by throwing a lot of images at this model, in two days, it achieved that level of performance. But it also made me nervous. It was my first deep learning experiment. What were the chances I had done things correctly? This is a problem we face today: deep learning and AI models are black boxes. We don't really understand how they work.

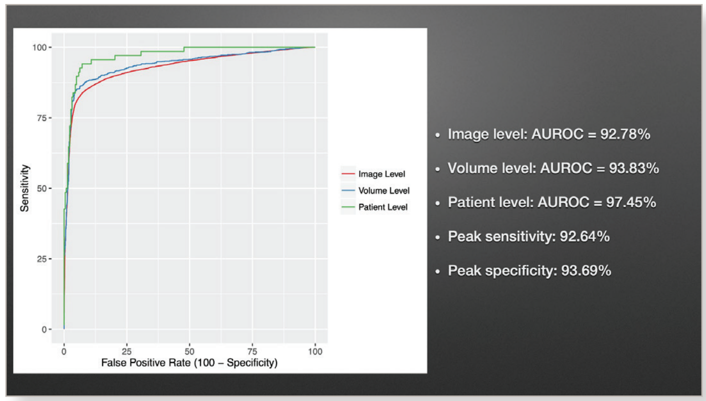

Figure 4

I was pretty unsatisfied with this, and I didn't have a way to check my code. This was a brand-new technology. I didn't feel comfortable publishing the result, so I dug around the computer vision literature and found a technique that I thought was intuitive. The idea is, if I show you this picture [of a dog with a ball in its mouth] and ask if there's a ball in it, you'd say yes with certainty. If I cover up small parts of the picture and keep moving that box around, eventually I will cover up the ball, and you'll say you're not sure anymore. That's what we did—we took OCT scans the model had never seen, systematically occluded every part of the photo, and looked at how model confidence changed. By doing that, we generated heatmaps showing the areas the model relied on most to make its classification. That gave me more confidence that the model was doing something useful.

This led to us publishing many papers showing that deep learning could do X, Y, and Z in ophthalmology. One paper, in particular, was exciting. This was Matthew Zang's idea from the University of Washington. He had been taking pictures of patients undergoing pre- and post-oculoplastic surgery. To my knowledge, this is the first application of deep learning in oculoplastics. He wanted to fully automate clinically relevant measurements like MRD1, MRD2, and brow measurements. We showed in a prospective test set that the model performed as well as human beings. The coolest part wasn't in the paper—it was a video. Julia Owen, the first author, was at home during COVID, turned on the slow-mo feature on her iPhone, and took a video of herself blinking. The deep learning model captured sub-millisecond resolution of upper lid dynamics—previously only possible with special equipment. Now, all you need is a deep learning model and an iPhone.

We've since tried to generalize this model. We want to automate and fully quantify all extraocular movements. The model worked great in primary gaze, but not when looking to the sides. We're now showing it can track movements even with the camera drifting. We also tested it on a YouTube video of someone with an adduction defect in the left eye. The model quantified that and even backed out the nystagmus cycle.

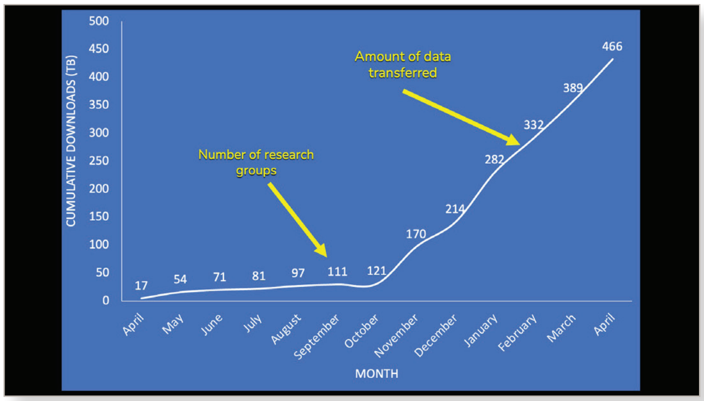

Figure 5

Now I want to talk about the Bridge2AI program (Figure 5). It is a big NIH effort to fill a critical gap: we don't have good data-sets to train these models. It's focused on dataset preparation, not model development. Our project, which I talked about last year at AAO, had to change its name due to administrative reasons. But in November of last year, we released our year-two dataset with over a thousand people and millions of measure-ments—continuous glucose monitoring, retinal imaging, and more.

It's hard to appreciate how big 1,067 participants are. This bub-ble plot shows the breakdown by self-reported diabetes status, age, and clinical site. We also wanted standardized training, validation, and test sets for apples-to-apples comparisons. Our dataset has minimal missing data—just one experimental device from Heidelberg had some issues.

Here's a graph from a collaborator showing how many research groups downloaded the data. We've transferred nearly 500 terabytes, with over 500 approved groups, and we're already seeing published papers—none of the authors are from our team.

In terms of future directions, I want to discuss exciting devel-opments. In 2018, a paper titled *Attention is All You Need* was published by Google Brain. This set the foundation for trans-former models. Years later, OpenAI released the GPT-4 report showing the model beating humans at nearly everything. The paper didn't disclose how anything was done. DeepSeek later helped uncover some of the magic sauce.

There are many papers showing ChatGPT passing medical ex-ams, but there's a big problem—what I call the Achilles heel of large language models. Deep learning requires a differentiable function for training. Truthfulness isn't differentiable, so large language models mimic human language without any guarantee of being truthful. I didn't think they had a place in medicine until I got an email from George Bartley about candidates cheating on the oral boards using ChatGPT.

If someone were to cheat using Google, they'd need internet access, retype questions, search, read, and interpret results—all while pretending not to cheat under Zoom surveillance. Using ChatGPT, they'd still need internet and retyping, but I built a system that takes screenshots, OCRs the text, and automatically queries ChatGPT. It outputs live answers without me lifting a finger, including correct diagnoses and even solving optics equations. The code is short—only 10 lines—and could be written by ChatGPT itself.

We're now seeing the rise of agentic AI, where models plan, execute, and reflect on tasks using tools. I built one that only searches PubMed, downloads articles, extracts high-quality content, and reasons over it—all on my laptop, with no data sent to OpenAI. This local LLM gives correct citations and answers without HIPAA concerns. That's the future of AI in healthcare.

However, there's a big caveat. A federal judge was nearly fooled by AI-generated legal content. Even Anthropic, the makers of Claude AI, were caught hallucinating a legal citation. So, it's critical to understand that while powerful, these models come with risks.

To conclude: AI is here to stay. There was hype in the 1990s, but this time is different. Every part of ophthalmology will be transformed, and it's up to us to guide this transformation responsibly.

I want to thank all the members of my lab whose work I'm taking credit for, and the mentors in this room who've played incredible roles in my career. Thank you again for the opportu-nity to give this lecture.