The Voice!

Volume 29 Issue 2
May, 2024

AI

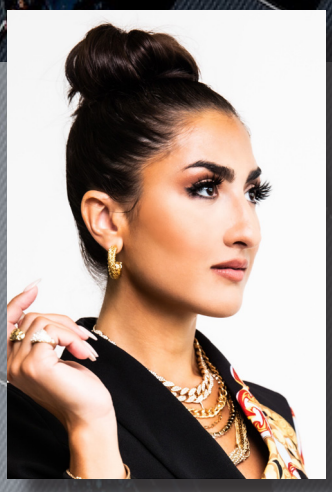
ARTIFICIAL INTELLIGENCE

The Voice of the Editor

Kimberly Steinhauer, Ph.D.
President, Estill Voice International
Pittsburgh, PA

Lately we find it everywhere. It's in journal publications, on stages, in recordings, at the doctor's office for my annual physical. It's even in the last issue of The Voice Foundation Newsletter. No, it's *not* reflux, but it *is* Artificial Intelligence (AI). You may, or may not, be surprised that I actually wrote this introduction without AI assistance. (I was tempted but stuck to my principles.) In this issue, three live authors volunteered their views on AI. Nashville's Vanessa Campagna shares her AI insights from the recording artist/songwriter vantage point. New York City's Sarah Brown contributes her AI conundrums from the Speech-Language Pathologist perspective. And from Erlangen, Germany, Michael Döllinger lends his AI expertise for the scientific scenario. Hopefully, we'll hear from you via real human conversations on artificial intelligence to be continued at the 53rd Annual Symposium: Care of the Professional Voice this May.

In This Issue



The Voice of the Artist

AI AND THE HUMAN EXPERIENCE
Vanessa Campagna
Recording Artist & Songwriter
Nashville, TN

Singing on stage since I was 5 years old, I've seen just how much the music industry has changed over the last 26 years. Now, my typical week is filled by songwriting sessions with other collaborators or sometimes just by myself, recording either at my kitchen table, or in a state-of-the-art studio.

The 2020 pandemic forced us to create either on our own or with others on Zoom sessions. In a way, it was a good thing, because it showed me that you can get much accomplished by working remotely. I collaborate with artists in Europe and Asia, so instead of using the excuse "we'll do it whenever I make it over there," I was able to actually get it done right then and there. That year I had more releases than I've ever had just because I had the technology of Zoom at my fingertips.

Fast forward to 2024 where we are all still taking advantage of what we learned in 2020, but with the addition of Artificial Intelligence. Chat GPT was all the rage at first, and songwriters started using it to push through a writer's block and even to inspire ideas on the spot if their creative well had dried up.

I've never used it, nor have I tried it. Not because I think I'm better than AI, but because to me, that's not being a creative songwriter sharing personal experience. I'm not sure an AI program can create a song that will pull on the heartstrings of millions.

AI in the production world is yet another interesting topic. I work with producers who use AI sites for vocals. My first experience with this practice occurred without my knowledge. I, like I do on all projects, wrote a song, recorded all of my vocals and then sent them to the producer so that he could mix them in and finalize the song for release. When I received the song, I was flabbergasted by what I heard. A deep husky, soulful woman was singing every vocal riff and word that I had recorded. I knew they didn't get another vocalist to sing the song because it matched my inflections impeccably. I called the producer and questioned this. His answer was "oh, I just thought it sounded better with a more soulful voice. I hope you don't mind!" The truth is, I didn't mind. This particular song DID sound better with a more soulful voice. However, I can't put my name on that project. I can say I wrote it, but I can't say that I sang it because even though I did, I really didn't. Robin, Soulful Voice #5 from the AI site did.

"Robin" which isn't her real name, sold her vocal rights to this website, along with thousands of others...something I would

never do. I don't know how much she earns every time someone uses her voice or how much she was paid for her vocal rights, but I would not participate in this real-life situation of "Ariel giving her voice to the sea witch in a seashell."

One pro to AI is that it makes song pitching easier. If I write a song suited for Katy Perry or Rhianna, I can sing the demo, upload the vocal to the AI Generator, and choose the voice of the artist I want to pitch to. Then, when the song gets sent to said artist they can "hear" their voice without ever having to sing a note. Then, it's much easier for them to say YES to recording the song.

I would not participate in this real-life situation of "Ariel giving her voice to the sea witch in a seashell."

Although tricky, there are pros and cons for AI in background vocals. It's nice to have vocal diversity in a recording, and often the artist doesn't sing their own background vocals (BGVs). AI can save producers time and money since they don't have to hire extra vocalists. Same with recording a choir. Instead of going through the hassle of hiring 20 people, finding a studio large enough, dealing with schedule conflicts etc., you can ask the artist to record the parts (for free) and then change each vocal to make it sound like 20 different people. This process can save the producer and artist days of work.

Here's where it gets tricky. Most vocalists earn their living by singing BGVs, demos, and featuring in choir ensembles. With AI, those people are automatically out of work because it's more efficient for someone to sit at the computer and finish the project quickly. While that process still takes time and money, it won't

take nearly as much as it would to have the real live voices. $25 a month to subscribe to the AI site for unlimited vocal conversions and one day of timing and tuning one vocal can replace 5 days of tracking multiple singers and timing and tuning each individual vocal. The PRO? Unless you're a global star, making money in the music industry is extremely hard. So, I guess it's nice that AI is paying vocalists for their vocal rights.

Finally, live performance. The reason touring makes the most money for artists is because people long for the human experience of seeing and hearing that artist tell their story live. I will always choose the living and breathing artist over the AI hologram. There is an undeniable energy exchange between the audience and the artists and musicians on stage. From the crowd spontaneously screaming the song lyrics to bonding with total strangers during a special musical moment, that experience can't be generated with a hologram.

It all comes down to human connections for me. Robots are cool but they aren't alive. They have no emotion. We live in a HUMAN world where emotion and physical experience is everything. So, if there's a way to find a middle ground with AI where it doesn't take over every single aspect of our life but instead enhances it, then I might change my opinion. AI still freaks me out, but I am open to learning what it has to offer.

Vanessa Campagna's commanding presence and versatile voice shines in collaborations with some of the best in the business and in the biggest music venues across the country. Her diversity and vocal purity earned her the respect of world-renowned director/composer, Marvin Hamlisch & mega-hit maker, Desmond Child, both of whom took Vanessa under their wing and developed her songwriting and artistry. She has written multiple top 10 hits on the UK Pop, Dance Charts and Billboard International charts for other major label artists and has had over 50 major ad placements in the Sync/Tv & Film world as well. 2020 was a breakout year for Vanessa as an artist. She featured alongside multi-platinum Christian/Pop recording artist, Michael W. Smith and garnered her first #1 hit on the Billboard Charts & RIAA Certified Gold Record with the globally loved song "Waymaker." Later, they also went on to win "Worship Song of the Year" at the K-LOVE Fan Awards. Since then, she has had a top 10 hit with the song, "Who I Am" by Wyn Starks, on 4 different Billboard Charts for 20+ weeks. Aside from writing for other artists, Vanessa is currently in the midst of releasing her self-titled solo project "Heart Like Mine," which will consist of several singles all wrapping into an album by late 2024.

The Voice of the Speech-Language Pathologist

ARTIFICIAL INTELLIGENCE IN SPEECH PATHOLOGY: EVALUATING THE POTENTIAL AND LIMITATIONS

Sarah K. Brown, MS, CCC-SLP
SKB Voice Studio
New York, NY

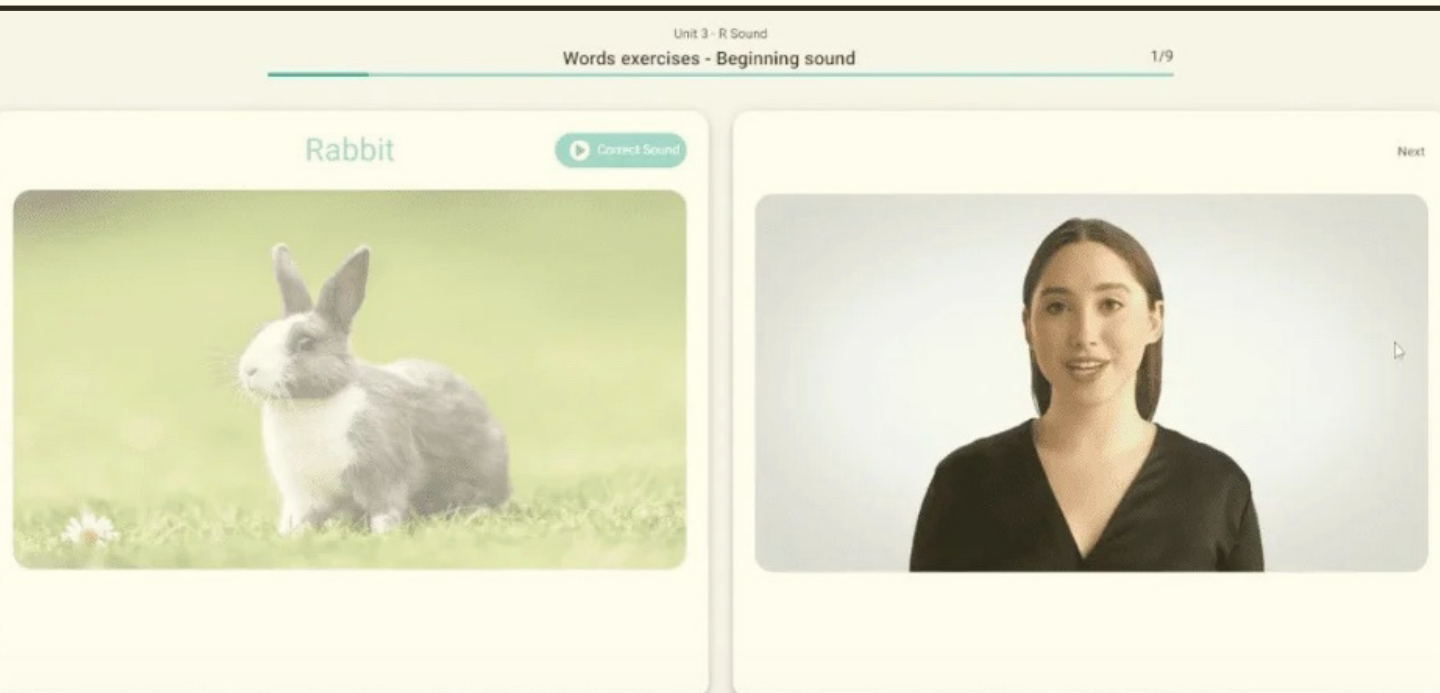
Artificial Intelligence has rapidly gained prominence, manifesting in innovations like Chat GPT, self-driving cars, and AI-assisted health technologies. While AI holds immense potential to enhance various aspects of our world, it also poses potential risks, as evidenced by concerns about job security (the writers' and actors' strikes of 2023 come to mind). In the world of speech pathology, particularly voice therapy, where can AI be a valuable tool and where do its capabilities fall short?

AI has already made significant strides in the field of communication sciences. Perhaps most notably, AI technology played a pivotal role in restoring communication abilities to a paralyzed woman following a brainstem stroke.[1] This AI technology tracks neural activity directly related to attempts to move her facial muscles which is then decoded into speech and enables her to communicate via an avatar. Another form of

AI used in communication disorders is voice banking. This technology is commonly used with progressive neurological conditions such as ALS and Parkinson's disease. Voice banking collects recordings of patients with these conditions while they still have the ability to speak. As they eventually transition to augmentative and alternative communication, their AAC device can produce speech using their pre-recorded voice, helping to make communication feel more personalized and authentic.

In voice therapy, one promising application of AI includes the development of a tool that assesses the perception of femininity/masculinity in voice samples and gives a percentage rating.[2] For example, my voice personally scored at 94% feminine. This tool was designed to aid in gender affirming voice training. Traditionally one of the main measurable targets of gender affirming voice training has been pitch, but there are also many subjective factors that contribute to someone's vocal gender perception: resonance, timbre, registration, and inflection pattern. In our study on this AI tool, we found that some participants were rated more femininely than others despite measuring in a similar pitch range. This AI tool may become revolutionary for assessment as it may be able to rate an individual's voice more holistically beyond pitch. However, approach its use with sensitivity by recognizing that it may evoke gender dysphoria in some individuals.

Another promising application of AI in speech pathology is exemplified by AI Therapist Jessica,[3] which is an exciting



prospect for making SLP services more accessible to underserved areas. Jessica utilizes advanced speech recognition and natural language processing to assess speech patterns and deliver targeted interventions. While its potential benefits for articulation and language therapy are evident, it is unclear to me its potential in serving voice patients, particularly in addressing the nuanced aspects of vocal efficiency. Articulation therapy for example is more objective – an articulatory sound is either intelligible or not. Training vocal efficiency in voice therapy does not look nearly as black and white.

This did make me wonder, what would an AI voice therapist look like? I believe the most practical and realistic use is its potential role in evaluation and outcome measurement. For instance, an AI assessment tool could gauge speaking voice ratings, providing a new standardized means to track progress in therapy. However, defining what constitutes a “normal” voice poses a challenge, as there is a wide range of what “normal” or functional voice is. This highlights the importance of meticulous data curation and consideration of diverse vocal norms.

Assuming such a tool was developed, it could enhance the current acoustic and perceptual measures used in voice evaluation. Using the gender AI voice tool as a model and rating “normal” voice on a percentage may be a practical way to see percentage improvement across the course of voice therapy.

It is possible that machine learning could better capture an overall rating of the abstract pieces of vocal functionality that we can't determine with our current instrumental measures. It also may have the potential to be a more standardized perceptual evaluation than our currently human performed CAPE-V and GRBAS. Such advancements in evaluation could extend to aiding patients with home practice and assessing their progress outside of therapy sessions.

AI assessment tool could gauge speaking voice ratings, providing a new standardized means to track progress in therapy

Despite the potential of AI improving the evaluation stage of voice therapy, I cannot overstate the indispensable role of human voice therapists in delivering personalized care in the treatment stage. Voice therapy necessitates a blend of art and science, demanding human qualities like compassion, empathy, and understanding. Therapists tailor treatment protocols based on individual needs and preferences, incorporating elements of counseling to motivate and support patients effectively—a level of human interaction that AI struggles to replicate.

Nevertheless, AI can complement human therapists by facilitating specific tasks, such as generating practice sentences tailored to therapeutic goals. For example, asking Chat GPT to compose a list of sentences using as many nasal consonants as possible for resonant voice therapy or using all H starting sentences for easy onset work. Yet, caution is advised as AI systems may occasionally produce inappropriate content, underscoring the need for human oversight.

While AI holds tremendous potential to augment voice evaluation and improve access to services, its current capabilities do not negate the need for human intervention, particularly in the application of voice therapy treatment. Ethical considerations and regulations are paramount as we continue to integrate AI into healthcare practices. As technology advances, it is essential to maintain a balance between leveraging AI's benefits and preserving the irreplaceable human touch in therapy.

Sarah K. Brown is a singing teacher and licensed speech pathologist based in New York City. Before launching her own voice studio and going into full-time private work, Sarah maintained a clinical caseload at the Voice Center of Mount Sinai for 5 years as well as an adjunct faculty position at Pace University teaching the graduate level voice disorders course. Her research has covered a wide variety of topics in voice, including her collaboration on a study regarding an AI voice evaluation tool. Sarah also regularly teaches singing masterclasses, most recently with Yale University School of Drama. Her specialized combination of singing background, performance experience, and clinical education and expertise have enabled her to serve Broadway performers, Metropolitan Opera singers, and TV/film actors. She holds a BM in Vocal Performance from Chapman University and an MS in Speech-Language Pathology from Northwestern University.

REFERENCES

1. Marks, R., & Kurtzman, L. (2024, March 18). How artificial intelligence gave a paralyzed woman her voice back. How Artificial Intelligence Gave a Paralyzed Woman Her Voice Back | UC San Francisco. https://www.ucsf.edu/news/2023/08/425986/how-artificial-intelligence-gave-paralyzed-woman-her-voice-back

2. Bensoussan, Y., Park, C., Johns, M., Brown, S., Pinto, J., Chang, J., & Courey, M. (2021). A comparison of an artificial intelligence tool to fundamental frequency as an outcome measure in people seeking a more feminine voice. The Laryngoscope, 131(11), 2567–2571. https://doi.org/10.1002/lary.29605

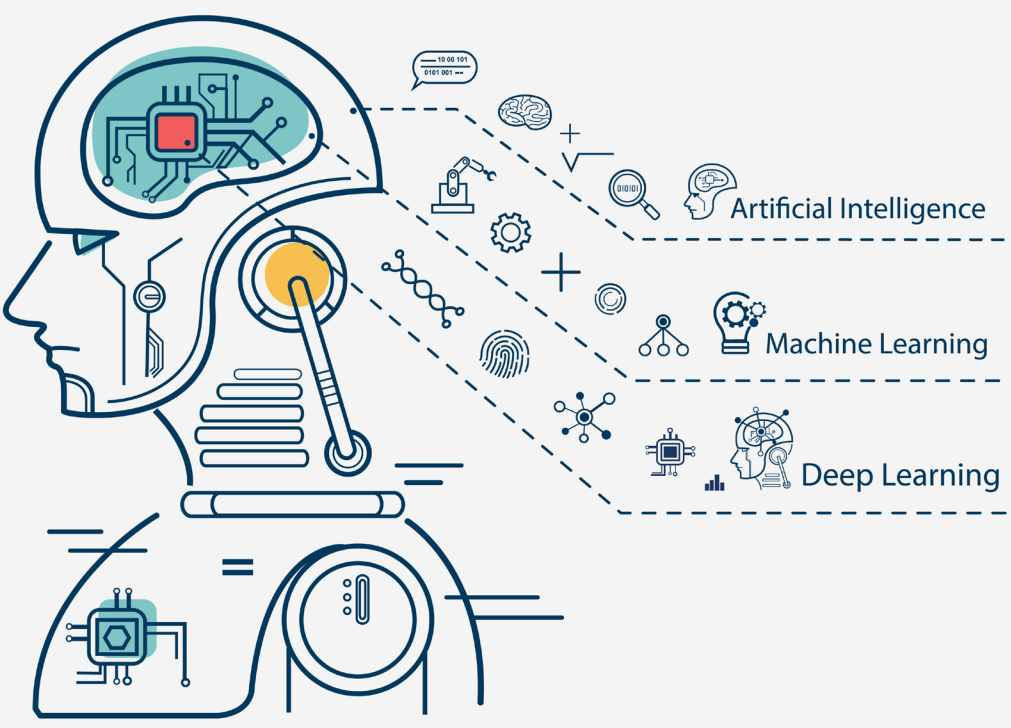
3. Better Speech. (2023, August 7). Meet jessica - the first AI speech therapist. https://www.betterspeech.com/post/meet-jessica-the-first-ai-speech-therapist

Join The Voice Foundation Today!

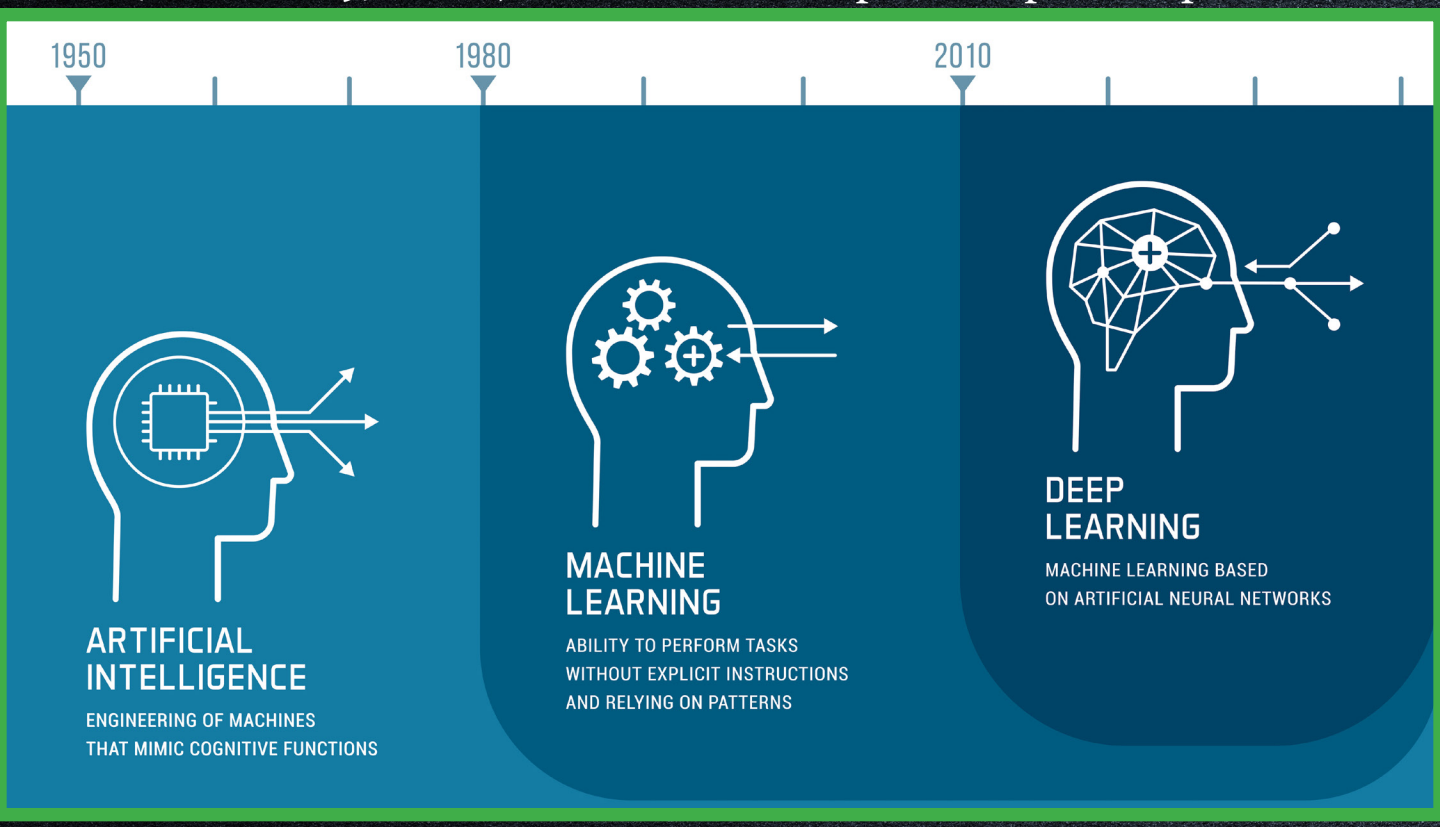
The Voice of the Scientist

DEEP LEARNING: GUESS WHO'S BACK?
Prof. Dr.-Ing. Michael Döllinger
University Hospital Erlangen, Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology Head & Neck Surgery
Department Computer Science, Friedrich-Alexander-University Erlangen-Nuernberg
Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-University Erlangen-Nuernberg

Since ChatGPT was launched in November 2022, Artificial Intelligence (AI) seems to have become officially available for the broader public and has been a permanent topic in the public media. Suddenly, there is a widespread belief that AI will now revolutionize our lives within our (first world) societies. However, certain subareas of AI, like Robotics, Pattern Recognition, and Computer Vision, have been around in industry and science for decades. For example, I received my habilitation degree in Medical Pattern Recognition in 2006 and contributed to my first Machine Learning paper in 2010 on classification of voice disorders applying Support-Vector-Machines[1]. For years, AI methods have been found practically everywhere around us like in our cars, TVs, cell phones, or when we book our holidays online.



Deep Learning (DL), a subarea of Machine Learning (ML) using large data sets to train Neuronal Networks (NN), as done in ChatGPT, lately came into public view. The underlying ideas and mathematical fundamentals of NNs were already introduced in the 1950 and 1960 by scientists like J. McCarthy, F. Rosenblatt, M. Minsky. Then, NN quickly disappeared when it was clear that, although these methods may be of interest, they are too complex to be integrated in existing computer systems. This lack of sufficient computational power continued until the noughties (2000-2009). Then, thanks to the computer chip developers



following Moore's law from 1965, computers finally were able to process the huge amount of data that was needed to apply NNs and similar DL methods. After graphics processing units (GPUs) were found to be an excellent choice for DL approaches by significantly speeding up the training and learning processes of DL algorithms, DL was suddenly applicable not only for high-performance-computational (HPC) users, but basically for everybody in possession of a gaming computer with a state-of-the-art graphics card (GPU). In summary, the current DL push is rather based on improved and faster computer hardware than on new mathematical inventions or theories. However, now that DL based methods are widely applied in science and industry, the underlying mathematical formulations can be better explored, refined,

and enhanced.

It is safe to say that Artificial Intelligence, or more specifically Machine/Deep Learning, will play a major role in clinical diagnostics and treatment. In voice science many new approaches have been suggested lately to not only improve diagnostics and treatment[2] but also to enhance basic research on the phonation process.[3] As can be seen in the cited review papers above, applying DL in the clinical environment will surely enhance and improve patient healthcare. However, DL in its current state should be viewed as a decision support system for, instead of a potential replacement of physicians and therapists. Also, as outlined in Alter et al. 2024,[4] introducing DL in the healthcare system may increase new challenges and ethical questions.

So before applying DL methods, one should always ask: What is my goal? Is DL the right method to reach my goal? Does my data reflect the problem to be solved? Do I have a sufficient amount of data for my DL model?

Currently, it seems that every other scientific publication, also in our field, includes some sort of AI or DL methods. At the moment, it is almost expected to use DL methods for data analysis in scientific publications and grant proposals. Hence, sometimes it appears that DL is used only for the sake of methodology. However, DL methods are just mathematical formulas that do exactly what they are told without asking if the requested task makes sense or not. So before applying DL methods, one should always ask: What is my goal? Is DL the right method to reach my goal? Does my data reflect the problem to be solved? Do I have a sufficient amount of data for my DL model?

To make a long story short, Deep Learning or Artificial Intelligence is a highly useful asset when correctly applied. However, it is not and will not become a magic wand solving all our scientific questions and clinical challenges. Successful science still needs human curiosity and enthusiasm. Patient treatment still needs human care and human empathy.

References
1. Voigt D, Döllinger M, Braunschweig T, Yang A, Eysholdt U, Lohscheller J. Classification of functional voice disorders based on Phonovibrograms. Artif Intell Med, 49(1):51-9;2010.
2. Barlow J, Sragi Z, Rivera-Rivera G, Al-Awady A, Daşdöğen Ü, Courey MS, Kirke DN. The Use of Deep Learning Software in the Detection of Voice Disorders: A Systematic Review. Otolaryngol Head Neck Surg. 2024 Jan 3. doi: 10.1002/ohn.636. Epub ahead of print. PMID: 38168017.
3. Döllinger M, Zhang Z, Schoder S, Sidlof P, Tur B, Kniesburges S. Overview on state-of-the-art numerical modeling of the phonation process. Acta Acoustica 2023,7,5. https://doi.org/10.1051/aacus/2023014
4. Alter IL, Chan K, Lechien J, Rameau A. An introduction to machine learning and generative artificial intelligence for otolaryngologists-head and neck surgeons: a narrative review. Eur Arch Otorhinolaryngol. 2024 Feb 23. doi: 10.1007/s00405-024-08512-4. Epub ahead of print. PMID: 38393353.

Michael Döllinger studied in mathematics at Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany.

He received the Diploma (M.Sc.) degree in February 2000, and the Ph.D. degree from FAU, in November 2002. From 2003 to 2005, he was a Postdoctoral Fellow with the University of California, Los Angeles. Then, he returned to Erlangen, Germany, and he received his habilitation degree in Medical Pattern Recognition at FAU in 2006. In 2008 he became Professor and the Head of Research. He was a Scientific Head of the DFG funded Research Group FOR894 “Fundamentals on Flow Dynamics in Voice Production”, from 2008 to 2013. Since 2008, he has been an Adjunct Professor with Louisiana State University, Baton Rouge. He is co-chair of the Division of Phoniatrics and Pediatric Audiology. His scientific contributions have been published in 180 peer reviewed journal papers and 380 conference contributions.