The first-ever AI opera, Chasing Waterfalls, premiered in Dresden.

September 3rd, 2022—curtain up. At the Semperoper in Dresden, a hitherto unknown singing talent faced an audience of opera professionals. After Andrew Lloyd Webber’s Cats in August and The Magic Flute in mid-September wowed opera audiences, the spotlight turned to artificial intelligence (AI) later in the month.

The Semperoper has added an innovative opera project to its program in the form of “Chasing Waterfalls.” The ensemble includes the Staatskapelle Dresden, six singers, and the algorithm “Ego Fluens,” a piece of artificial intelligence which—unsurprisingly—plays itself. Chasing Waterfalls offers up the first-ever live performance of AI as an opera singer.

Is it a classic opera? That’s open to dispute. Multimedia effects and a host of digital elements have left their mark on the opera. The engineered voices of the six characters came together in an incredibly modern composition, with videos, lighting effects, an eight-meter-high waterfall as the central stage element—and of course, AI.

AI meets the opera

Bringing digitalization to the opera and testing the capacity of AI sounds like an interesting digital experiment. For classical artists, the inclusion of the virtual ensemble member created a host of new challenges. Whereas in a classical opera, the actors respond to one another and the conductor guides the orchestra through the performance at the correct tempo, Chasing Waterfalls followed a strict sequence, with conductor Angus Lee facing the challenging task of not only conducting the chamber orchestra, but also coordinating the human actors with the electronic background sounds and the fixed-time appearances of the AI (via click tracker)—all live and in time.
It’s orchestra 2.0, if you will. Angus Lee also composed a large part of the AI-era in collaboration with the Berlin Studio for Sonic Experiences, kling klang klong. "We tried to bridge the gap and connect our very different soundscapes," summarizes the musician, who is originally from Hong Kong. Kling klang klong provided the electronic sounds for Lee’s classical passages and brought everything together in his first full opera score.

But the real mastermind behind the innovative opera project is director and media artist Sven Sören Beyer. He has worked together with Berlin artist collective phase7 performing.arts since 1999, putting together performance productions and installations. He is always sounding out the area of tension between man and machine. In his latest project, he reflects on the influence that AI has had on our lives.

The real world confronts the digital world

The story might remind some people of Tron. In the pioneering Disney film from 1982, programmer Jeff Bridges finds himself a prisoner within a computer network and, with the help of his program Tron (a kind of alter ego) tried to escape—and put a stop to the Master Control Program, the AI alter ego of his adversary David Warner. What was purely science fiction at the time, now, 40 years later, has a real, somewhat bitter taste of dystopia in Chasing Waterfalls.

Norwegian soprano Eir Inderhaug from the cast of the Bavarian State Opera is confronted with not one, but six ‘digital twins’ she needs to grapple with. Back then it was a laser, now it’s a simple login. When the real, physical selves log into the computer, they are faced with digital copies of themselves: the numerous traces that they leave behind in the digital world of the Internet. They encounter their digital twins, Ego Fluentes, which operate in the virtual world as independent personalities, and eventually even team up against the physical selves and attack reality.

Will the physical selves be able to assert themselves as real people against the dissolving boundaries? The boundaries between the virtual and physical world become blurred on stage, with the stage design highlighting this effectively.

But what about the real world? Beyer wants to use his work to contribute to the discussion surrounding how digital our personalities are today, prompting age-old questions such as: What is truth (in an increasingly digitalized world)? What makes us human?

How can AI learn to sing?

Over the course of just 70 minutes, Chasing Waterfalls immersed the audience in a world in which opera and digitalization merged. AI had a key role to play this evening too, before moving on to its next engagement in Hong Kong, the home of composer Angus Lee. Naturally, it didn’t get stage fright and won’t in Hong Kong either—but perhaps it has the mannerisms of an opera diva? How can you make AI actually start singing?

"With all the achievements AI can already boast of a ready-made AI opera singer does not yet exist," smiles Nico Westerbeck, who worked as the technical lead with the AI, training it from a beginner in the girls’ choir to an opera soloist. Westerbeck is a passionate computer and data scientist. He has worked at T-Systems Multimedia Solutions (MMS) in Dresden since 2018. His main focus areas are deep learning for language and text, reinforcement learning, and security. He and the MMS team turned the artists’ innovative ideas into reality and brought the AI to life.

Here, it’s actually not quite correct to say just ‘one’ piece of AI—several were involved, all in all. Librettist Christiane Neudecker worked with GPT 2 and 3 to develop texts, while another learned how to read music, and another how to sing. A team from T-Systems MMS was heavily involved in developing the singing AI.

Will the physical selves be able to assert themselves as real people against the dissolving boundaries? The boundaries between the virtual and physical world become blurred on stage, with the stage design highlighting this effectively.

But what about the real world? Beyer wants to use his work to contribute to the discussion surrounding how digital our personalities are today, prompting age-old questions such as: What is truth (in an increasingly digitalized world)? What makes us human?

How can AI learn to sing?

Over the course of just 70 minutes, Chasing Waterfalls immersed the audience in a world in which opera and digitalization merged. AI had a key role to play this evening too, before moving on to its next engagement in Hong Kong, the home of composer Angus Lee. Naturally, it didn’t get stage fright and won’t in Hong Kong either—but perhaps it has the mannerisms of an opera diva? How can you make AI actually start singing?

With all the achievements AI can already boast of a ready-made AI opera singer does not yet exist," smiles Nico Westerbeck, who worked as the technical lead with the AI, training it from a beginner in the girls’ choir to an opera soloist. Westerbeck is a passionate computer and data scientist. He has worked at T-Systems Multimedia Solutions (MMS) in Dresden since 2018. His main focus areas are deep learning for language and text, reinforcement learning, and security. He and the MMS team turned the artists’ innovative ideas into reality and brought the AI to life.

Here, it’s actually not quite correct to say just ‘one’ piece of AI—several were involved, all in all. Librettist Christiane Neudecker worked with GPT 2 and 3 to develop texts, while another learned how to read music, and another how to sing. A team from T-Systems MMS was heavily involved in developing the singing AI.
Origin of the AI voice: A text-to-speech algorithm

“We didn’t start completely from scratch with our opera singers. We used research results from the area of text-to-speech, primarily the work of Chen et al. (‘HifiSinger’, 2020) and Liu et al. (‘DiffSinger’, 2021), which converted a text-to-speech system into a singing voice synthesis system,” says Westerbeck. As the project progressed, too, the search for up-to-date findings remained a constant companion for the MMS team —after all, singing AIs are new territory. Westerbeck dug through dozens of publications in order to find the key to help make a piece of AI sing.

But there was quite a way to go from ‘Hello, how can I help you?’ to ‘Hell’s vengeance boils in my heart’—and a great deal of code had to be written too. How do you describe a singing voice in precise detail? Language, and moreover singing, are too complex. This is especially true when it comes to mapping them digitally.

A typical music file of 44 kHz contains 44,000 individual sound pressures in one second. How many words can a person speak or sing in this time?

American actor Eddie Murphy sounds like he’s speaking 50 words in Beverly Hills Cop, but in an opera, it’s perhaps five. How do you distribute these 44,000 pulses across five words (and the different notes)? Where does each phoneme—the basic units of sound in language—begin? It’s a real puzzle, and a complex one at that.

Wanted: singing models and suitable data

The MMS team decided to take a pragmatic approach. The AI was to learn from a model. Kling klang klong invited Eir Inderhaug, who would play the role of the real self in the opera performance, to the studio in Berlin. There she initially sang 50 children’s songs, which were digitalized and sent to T-Systems MMS. Why children’s songs? “One of the publications we read recommended children’s songs,” explains the AI specialist. “This was an effective approach too. However, at a later stage of the project, it became clear that we wouldn’t achieve our goal this way. Rudolph the Red-Nosed Reindeer is no operatic aria—even when it’s sung by an opera singer.”

A second visit to the studio was therefore necessary. Inderhaug had to go one better, singing 20 operatic arias, which ultimately provided 10 more minutes of training material.

In the end, 70 songs were used by the AI team as a data source, intended to show the AI how singing functions. “We then had a sufficiently wide spectrum of data to avoid overfitting.”

How the neural network learns to sing

The AI experts at T-Systems MMS then developed a piece of architecture for a neural network, which is able to record notes and texts as input and generate a sound output from them. The team decided to spread out the complexity of the tasks in the neural network and generated a pipeline of multiple neural sub-networks. “In a few years, it’ll perhaps no longer be required, but the complexity we were faced with made this strategy necessary,” admits Westerbeck. These networks were initially as musical as a housefly on the hunt for food.

At least that’s what the first result sounded like. This isn’t surprising, as the first parameters for the neural network were initially created by a random generator.

The AI was trained in so-called epochs. Within an epoch, the AI was shown the complete dataset—which had already been split up into 10,000 snippets. In this way, three million ‘training sessions’ were provided over a total of 300 epochs.

During each process (‘forward pass’) within a training ‘epoch’, the AI was scrutinized at the end. The AI ‘aria’ was compared with the professionally sung version by Eir Inderhaug. This ruthless loss review assessed the performance of the model. The quantified results were then played back automatically into the neural network, which adjusted the originally random parameters.

“At the start of the training, a neural network makes huge errors. That’s not unusual,” explains Westerbeck. “The purpose of the training is to reduce these errors and gradually continue to improve.”

The neural network increasingly started to recognize sounds that were too loud or too high, and learned from this so that it sang better on the next attempt. The second product (after 10 epochs) first sounded like singing. “A bit like a radio that isn’t tuned to an exact station, or secret messages from outer space which have been extracted from cosmic background noise,” smiles Westerbeck.

But of course this still wasn’t enough for the operatic artists. Eule sums up the development of the AI—and you can tell from his words the slight disdain from the

The songs sung served as a data source for the AI.

Overfitting

Overfitting describes an unpleasant feature of AI, which can jeopardize entire AI projects. If the AI is trained with too little data, it tends to focus on this input and not on acquiring general knowledge. It would be like a student learning by heart which answer boxes to cross in a multiple-choice quiz (without knowing their meaning) rather than readily knowing the correct answers. In this case, the AI would be able to sing a few songs correctly, but would fail when it came to different, unknown notes and texts. It wouldn’t learn to sight-read the music.
classic operatic artists at the beginning of
the project: “When we heard the first
singing samples, we were unsure as to
whether we really wanted to venture into
this experiment,” says dramaturge Eule.

Algorithm Ego Fluens passes the
improvisation test

Indeed, if you think that the AI would
now only ever deliver the same version of
the song, you’d be wrong. The neural
network is dynamic and is always slightly
changing its playback. Neither listeners
nor vocal coaches are able to say what
they will hear exactly, just that an ‘audi-
ble’ result will be created that suitably
reproduces the played notes and the
text. The neural network will provide
‘just’ an optimized result, which is always
slightly different from the original—and
rightly so. Whether a singer rolls an R or
hisses an S is entirely up to them. The
same is true for the AI. It is perhaps a bit
more human in this regard than one
might think.

But the makers of the opera wanted to
go one better. Not only was algorithm
Ego Fluens to sing songs, for which it re-
ceived the (previously unknown) notes
and text from the score, a 4-minute pas-
sage was also provided in which it was
supposed to improvise. An extemporiza-
tion. For this passage, the AI received no
texts or notes from humans. Thanks to
the work of the team at kling klang klong,
it was possible to combine the AI lan-
guage model GPT-3 with a note compe-
tition model, providing the synthetic vo-
cal model (the singing AI) with different
texts and notes for each performance.

Algorithm Ego Fluens passes the
improvisation test

After the AI received these during the
performance—those responsible checked
first that the texts did not contain racist or
sexist content You never know with AI...
The audience of experts shouldn’t have
noticed any differences in the passages
in which text and notes were fixed be-
forehand and the AI followed the score
and the libretto. “Here, the result within
certain parameters was quite clear—and
we were quite relaxed,” remembers
Westerbeck. And while the AI didn’t
deliver world-class soprano-level singing,
the singing AI received these during the
performance—those responsible checked
first that the texts did not contain racist or
sexist content You never know with AI...

The audience of experts shouldn’t have
noticed any differences in the passages
in which text and notes were fixed be-
forehand and the AI followed the score
and the libretto. “Here, the result within
certain parameters was quite clear—and
we were quite relaxed,” remembers
Westerbeck. And while the AI didn’t
deliver world-class soprano-level singing,

Things start to sound more and more melodic: version one after 150 epochs,
version four after circa 200 epochs, and version ten with the final AI song.

For those in the know: the pipeline parts initially consisted of a transformer-
based acoustic model, a diffusion decoder, and a GAN-based vocoder.

The end result was a neural network which not only made few mistakes,
but also had a general understanding of how singing works. “In a best-
case scenario, it’s therefore able to sing any popular song for people—with
the voice of Eir Inderhaug.” ‘A generalization’ like this makes any data
scientist happy.
the freshly generated text and the unknown notes in real time into a live aria.

The experiment was a success; the opera AI had proven itself. Even the libretto impressed certain critics. Would you like to hear a sample? And if the AI assures you that "I am so much more than a machine... My heart is just a cold hard drive," it really does seem like it is worried about its own existence.

Will the new Maria Callas come from a computer?

What’s next? Whoever is able to create a new AI Maria Callas from a computer could land on the idea to invent the digital opera too. A new business model—one without prima donnas saying things like ‘I am not in tune today’ and conduct lengthy negotiations over pay. Despite the successful pilot, Sven Sören Beyer doesn’t think this will happen. "I don’t believe that it will replace human creativity, but rather act as a catalyst. However, I do believe that we are on the brink of something huge. Optimization will become more and more creative and give us tools that we can’t even fathom yet."

Cultural creatives, therefore, have the all-clear—music will remain in the domain of humans, and AI will not be listed with a talent agency, but will find its feet as a support for creatives. AI expert Pavol Bauer also shares this view in the previous issue. Once it returns from Hong Kong this November, algorithm Ego Fluens will take its first holiday to the MMS data archives. We’ve not ruled out an appearance at the Christmas party. ✪

Watch trailer

A four-minute extempore by the AI was part of the performance.

FuturePractice  |  The opera diva from the computer