



Making Speech Recognition Work for Children

An Interview with Amelia Kelly

What does it take to build new AI technologies for education? Dr. Amelia Kelly, chief technology officer at SoapBox Labs, shares her experience with us in this interview.

By Amy Adair and Joewie J. Koh

DOI: 10.1145/3589638

Recent progress in natural language processing (NLP) has been monumental in enabling innovative new technologies for facilitating learning. However, many of these NLP advances are limited to text, and bridging the gap between speech and text poses numerous nontrivial challenges—especially when we’re working with children. In this interview, we sat down with Dr. Kelly to chat about her work at SoapBox Labs, a Dublin tech company transforming the landscape of speech-centered educational technology for kids. This interview has been edited for brevity and clarity.

Joewie Koh [JK]: Many of our readers don’t have a computer science [CS] background, but are still very keen to pursue a career in computing. While preparing for this interview, I looked at your resume and noticed you studied physics and astronomy for your undergraduate degree. You’re now Chief Technology Officer [CTO] at SoapBox Labs, working on speech recognition technology. Can you tell us more about how you got to where

you are? What advice would you have for non-CS students hoping to break into the technology industry?

Amelia Kelly [AK]: I was 17 when I went to college and interested in studying many subjects, including English and philosophy, but I chose physics and astronomy. After I finished my undergraduate degree, I traveled for a bit and then decided to go back to college. At this point, I still hadn’t decided on

a career path. I was still interested in English and the arts, but I had also discovered linguistics, which appealed to the more logical side of my brain. So, I went to Trinity College Dublin to complete my master’s in linguistics. In my courses, I was studying everything from morphology to semantics to linguistic typology, as well as speech synthesis and speech recognition. I quickly grasped some concepts, like



signal processing and model fitting, that might have been more challenging if I didn't already have a degree in physics.

At Trinity, I joined a research group that was working on speech synthesis for the Irish language, and I stuck around in the phonetics lab to do my Ph.D. I also completed an internship at a speech technology company, which inspired me to get a job at a Silicon Valley startup once I completed my Ph.D.

I loved my job at the startup. It was so refreshing working in a fast-paced environment. Everybody was just like, "What do we do now? Well, that seems impossible—let's try it!" The attitude was just fantastic. I then worked for IBM Watson, a massive corporation,

but I preferred the startup world. So, when I got the opportunity to work at SoapBox (my current job), I took it. I was one of the first three employees, helping to build the foundations of SoapBox's speech technology. Now, after eight years in the company, I'm the CTO.

In terms of career advice, there is one thing I would advise everybody: Intern at a company you're interested in. You'll be exposed to a lot of the different skill sets and challenges that working in industry brings, which differ from academia. Industry often doesn't have the same financial constraints. You end up working with bigger servers and bigger data sets. Unlike academia, the focus could be putting together two things that al-

ready work to make a new product feature; it might not be a scientific breakthrough, but it could be very useful. If you are handy, or imaginative, or inventive there will always be a place for you in a company.

Mine has been a long and varied career path, but it's one I'd recommend for anybody who isn't absolutely sure of what they want to do.

JK: That's very cool. I spent some time at a startup myself, so I can relate to what you're saying.

Amy Adair [AA]: At SoapBox Labs, your team has been developing speech recognition technology specifically designed for kids. In what ways can this speech recognition technology be used in education, whether that's in a formal or informal setting?

AK: There are a lot of technological challenges for making speech recognition work for children's voices. Children sound so much different from adults—there are physiological and behavioral differences. Also, the applications that you need speech recognition for are different. We're a behind-the-scenes technology provider, which means that we sell to companies and businesses who come up with imaginative products and services that require voice and speech recognition that will work very accurately for kids. Our primary market is in education, where our speech technology powers preK-12 digital screening and assessment tools for literacy, language learning, and other subjects from education publishers like McGraw Hill and Scholastic.

For reading, for example, the aim of our speech technology is to mimic what the teacher has to do for each of the 30 students in their class. To support individual students as they learn to read is very time-consuming for a teacher. They have to sit down next to each student, listen to them read, and provide individualized feedback. Our technology helps automate and scale reading practice and assessments. Let's say that the child is reading a passage from start to finish. Our voice engine will compare what the child read to the prescribed passage and then tell the teacher if they've made any substitution, insertion, or deletion errors. It will also tell them the intonation of the voice—where it went up and down and where it stopped.

Another interesting application of our technology is rapid automatized naming (RAN), a way of detecting dyslexia early. We have some clients who are working on products that would ask a child to read out words in order, for example: "house, dog, dog, house, dog, pig, cat, house." They time how long it takes the child to read the words, how long the child pauses in between words, and how many mistakes the child makes. Then, they show the child the same concepts except in pictures instead of words and compare the time it takes the child to do the same task. If there's a decided difference between the reading and

the picture identification, they can flag for dyslexia in certain cases.

AA: When developing speech recognition technology for kids, what are some specific challenges that your team needed to consider? And how do you address these challenges?

AK: There were two big challenges we needed to consider. The first was background noise. Children use our technology in everyday environments, from noisy kitchens to loud classrooms, so it was imperative that we built our technology to work for kids wherever they are.

In the past, children's speech datasets (or any voice datasets at all) were always recorded in very clean, quiet conditions. Picture a bunch of children in soundproof booths recording audio with a headset mic. If you build a speech system with this clean dataset, the result would be a speech system that only worked for children in those quiet environments. At SoapBox, we used a different approach, ensuring that our models were built on real-world noisy data, and therefore would work in the kinds of environments where we knew children would be using voice-enabled products—such as the classroom, noisy cars, and kitchens—with all kinds of different atmospheric background noise. We began building our system around the time when modern advances were gaining ground over more traditional system architectures, and we've leveraged this perfect storm of increased computational power, big data, and sophisticated neural network architecture to build the SoapBox engine that we have today.

If you are building an AI system that is going to be used blindly in any way, you need to take responsibility for the fact that you're putting this out into the world.

The second big challenge was ensuring our technology works equally well for all accents and dialects. When you're making speech technology, it needs to work equally well for every child, regardless of race, background, age, or ethnicity—not just for the children who happen to be in your training set. At SoapBox, we make a real concerted effort to test our system with diverse accents and dialects. Last year, we became the first and only company to be awarded the Digital Promise certification for Prioritizing Racial Equity in AI Design, which we're extremely proud of.

JK: You've been talking about some of the cool things that y'all are doing at SoapBox Labs. What kind of team does it take to work on this kind of system? What types of expertise do you draw upon to build this product?

AK: I lead the Speech Technology team at SoapBox, which is made up of speech engineers, scientists, and computational linguists, who work in partnership with the Engineering team led by Robert O'Regan, our VP of engineering.

A very cool part of our company is called SoapBox Studio, our voice consulting services, born out of the realization that it takes a lot of imagination and innovation to bring clients along the thought and design process of voice-enabling their products. They may know that they want to incorporate voice into their products, but they don't know where to start. That's where SoapBox Studio—which consists of our voice UX, product, and pedagogy experts—helps. For example, Studio works with clients who are adding voice to an existing product and will help them solve problems like: How does the child know when to start speaking? How do we keep them engaged throughout the entire voice experience? How do we reward them and let them know it's time to move onto the next activity?

JK: There are fears that increasing use of technology in education will exacerbate inequalities, due to systematic bias in the technology or through inequality of access. What are some ways you think these impacts could be mitigated?

AK: I think it's a hugely concerning problem. If you're building an AI

system, basically what you're building is a system made up of programming rules, which you feed a lot of data and information and tell it to make up its own rules about the future based on what it has seen in the past.

This type of activity has led to some real blunders. One example is a CV sorter that discarded all the female applicants for engineering jobs because, historically, most engineers the companies had hired were male. Another thing that happened was people started realizing that they could copy and paste the job ad and put it at the bottom of their resumes in white, so it was unreadable, but the system would still be able to parse it. The text would match the job description so exactly that these applicants went straight to the top of the pile. These are examples of why it can be very risky to blindly trust an AI system that hasn't been trained with diversity and inclusion and equity in mind.

If you are building an AI system that is going to be used blindly in any way, you need to take responsibility for the fact that you're putting this out into the world. For me, that means ensuring that SoapBox's dataset is diverse, ensuring that it's labeled properly, ensuring that the system is tested adequately, and ensuring that we're always on the lookout for potential bias.

Before Christmas, I had a great conversation with the co-founders of The EdTech Equity Project, Nidhi Hebbur and Madison Jacobs. They created the AI in Education Toolkit for Racial Equity, which is available online for free [1]. If you are a builder of AI, especially in education, this toolkit has a practical list of things you can do to make sure that you're not building a system that is going to amplify and perpetuate already existing biases in AI.

As a leader of an AI team, I ensure all team members know that building for equity is a priority. We meet as a team once per month to discuss and challenge our assumptions and how we label our dataset. I also put a lot of thought into who we hire, keeping diversity and inclusion top of mind, so that we don't end up with a homogeneous group of people who all think the same way. It's very important to hire diverse skill sets, which make

Intern at a company you're interested in. You'll be exposed to a lot of the different skill sets and challenges that working in industry brings, which differ from academia.

a stronger team that's better able to build a system that will be robust to product pitfalls.

AA: That's a wonderful answer. I guess I have another big question for you. In your opinion, what's next for the field of AI in education? And what are the major challenges that are yet to be solved? What opportunities are out there?

AK: There's still a lot of low-hanging fruit when it comes to the education industry adopting AI. There's a little bit of pushback with automating learning tools with AI because people fear that computers are going to replace teachers. I think that this shouldn't be the case. Nor do I think it will be the case. Nor do I think it can be the case. I think the teacher is the most fundamental unit of society. Teachers have a hugely important and emotionally challenging job with a lot of responsibility. AI and technology can help them by offering them tools that are useful for them, save them time, and support them in ways that they deem most beneficial.

Let's take speech therapy as an example. In Ireland, the waiting lists for children's speech therapy are very long. Children could be waiting two years to get screened to see if they need speech therapy. After their consultation, they could be waiting another year and a half to actually see a therapist. This process—from screening to exercises issued by a speech therapist—could be automated with AI technology in a way that doesn't impinge negatively on a therapist's time or energy.

There's a lot of hype around AI,

especially with the release of ChatGPT. New startups will come out with products that are based on ChatGPT outputs. I think entrepreneurs and founders developing these products are going to make some really exciting new technologies. I hope that some of it can be adopted to make people's lives easier, to make children's lives easier, to make the learning journey easier. That said, I think the challenge remains to explain AI in a meaningful way, to keep it safe, and to ensure it's adopted properly and responsibly.

JK: This is an excellent point. For the last question: If you were to solve child speech recognition tomorrow, what would you want to work on next, and why?

AK: That's a tough question! I really like working on technology that impacts education because, not only is there a challenge, it's a scientific challenge. I love the work I do because I'm using my brain, solving problems daily, and helping people. And as my colleague Brian said very inspiringly the other day: "We're all here because we want to help little children. Isn't that the best kind of thing you can say about your job when you get up in the morning?"

References

- [1] Hebbur, N., Jacobs, M., Volk, M., and Allendorf, E. AI in Education Toolkit for Racial Equity. The Edtech Equity Project. Aspen Tech Policy Hub, 2020; <https://www.edtechequity.org/work/ai-in-education-toolkit>

Biographies

Amy Adair is a National Science Foundation (NSF) Graduate Research Fellow and doctoral candidate in the Ph.D. in education program at Rutgers University. Advised by Dr. Janice Gobert, Amy works at the intersection of the learning sciences and AI in education fields to design and develop evidence-based tools for STEM teaching, learning, and assessment. Prior to graduate school, she received a B.S. in mathematics from Louisiana State University and taught high school math and robotics classes in Louisiana. Drawing from her former experience as a mathematics teacher, her dissertation work focuses on automatically assessing and supporting the ways in which students develop and write about mathematical models in virtual science labs within the Inquiry Intelligent Tutoring System (Inq-ITS; inqits.com).

Joewie J. Koh is a doctoral student in the department of computer science at the University of Colorado Boulder. His research is primarily centered on reinforcement learning and spans multiple domains, including robotics, multi-agent systems, and human-AI interaction. Before starting his doctoral studies, Joewie played a pivotal role as the first employee of a Denver-based startup, where he successfully applied state-of-the-art techniques from natural language processing to tackle novel challenges in cybersecurity and legal informatics. He holds bachelor's degrees in mathematics and economics, as well as master's degrees in computer science and electrical engineering.

Copyright is held by owner/author.
1528-4972/23/03