"That is a 1984 Orwellian future at our doorstep, right?"
Natural Language Processing, Artificial Neural Networks and the Politics of (Democratizing) AI

Andreas Sudmann in conversation with Alexander Waibel, professor for Computer Science at the Karlsruhe Institute of Technology and also professor at the School of Computer Science at Carnegie Mellon University.

Andreas Sudmann: Alex, you are one of the pioneers in the area of Artificial Neural Networks (ANN) and Natural Language Processing (NLP). What was your initial motivation to enter this field of research?

Alexander Waibel: I have been working in academia for around forty years. Born in Germany and having German parents, I went to study at MIT and later to Carnegie Mellon, the leading universities in computer science and AI. And it was at those institutions that I developed the main thrust and inspiration for all my work, which is the question of how we learn and communicate as human beings and how we can build technology to help improve human communication. Back then, in the 1970s, when I went to university, people had already been thinking about AI, given that the first definitions of the field were proposed in the 1950s. In fact, Nobel laureate Herbert Simon, who was part of my thesis committee, had participated in the famous Dartmouth conference on building intelligent machines in 1956, where he and other researchers defined this early vision of AI.

In those days, everybody was attempting to build intelligent machines by search algorithms, rules, and logic formulas. But for me as a student, this seemed a bit like "des Kaisers neue Kleider" (the emperor's new clothes). I was listening to these famous people talking about a problem that to me would never be solvable with a rule-based approach to AI. It was intuitively clear to me that the amount of knowledge and facts that we learn in a lifetime is just so enormous that programming it all into rules would be impossible. And worse, they would have to be changed all the time, because the world around us is changing all the time. In fact, this *is* totally impossible, and so it was an early concern for me to say from the start: We will never achieve such goals unless we develop learning machines that can acquire such knowledge by themselves.

Aside from this fundamental scientific quest, though, another dimension always mattered to me as a scientist: Even though this might now sound like a cliché – it was a continuing desire to make contributions to society and to make the world a better place, as opposed to just following my own personal pursuits or interests. As a researcher, I am not so much simply curiosity-driven, but driven by practical goals. Practical goals provide a way to evaluate progress and can impact society in a positive way, once we are successful. And among them perhaps one of the most consistent goals for my work as a foreigner who grew up with 5 languages, was to build machines that can help us humans to translate between languages, by text or by spoken language. Throughout history, there have been many attempts to use machines for translating texts, which is hard enough in its own right. But when you try to connect people across language barriers, you also have to translate *spoken* language. In the 70's this seemed like a preposterous goal, and indeed, it seemed unsolvable, as speech added a whole new dimension of complexity and complication to the problem due to the fact that turning speech into text (before translating it) was an unsolved AI problem in itself. We did not know how to recognize speech and worse people never speak clean text… they make mistakes when they speak, they stutter, hesitate and correct themselves during speaking. And how would you then combine it with the other hard AI problem of translation? And all of that combined was so hard that it was unthinkable to realize in the early days of AI. But for me the dream was born and with youthful naivety and optimism we went for it. Needless to say, it was and remains a hard problem, a problem that we are still working to this very day. But despite the obstacles, challenges and delays along the way, we were able to see the fruits of our efforts. In retrospect, it is quite a privilege, actually, to be living in the *one* generation of humankind that sees language barriers disappear and to have had the opportunity to be working on the technologies made it possible.

The key to success scientifically was due to progress in machine learning methods combined with the explosive growth in available computing power and data that supports them. But for the vision to become reality also meant that academic progress had to be transferred to societal deployment. To do so, we started several companies that specialized on building wearable speech and language technology and eventually mobile speech translators. One of them, Jibbigo, built and sold the first ever Dialog translator on a phone. It was sold via the App Store and helped Tourists and Healthcare workers to communicate. The company was later acquired by Facebook, and we continued working on even more advanced deployments. For example, we are now developing new interpreting tools that help migrants in Germany to communicate with doctors if they cannot speak the language. In a University setting, we have installed automatic simultaneous interpretations services at KIT, so that foreign students can study in Germany and follow a German lecture by way of simultaneous interpretation during the lecture.

My team in Karlsruhe and I have also performed early experiments at the European Parliament to see if such a technology can be of assistance in this most challenging language environment. So it's really ultimately not only about translation alone but about how we can build technology that can bridge across barriers, that can bring the world together and make people understand each other better. And in order to master this hard problem, you really have to build machines that can learn effectively at multiple different levels.

And you have already worked out the necessary fundamentals in the 1980s and 1990s, especially with your research on so-called TDNN models. Perhaps you can tell us a little about how you came up with this particular approach and explain how it works?
Back while I was writing my PhD thesis at Carnegie Mellon I became fascinated with the idea of building learning algorithms that would mimic more closely the massively parallel, holistic learning that we perform as humans in the brain. I discovered that researchers in the 1950's had already proposed so-called "perceptrons", which did learn, but could only solve very simple classification tasks. Still, the fact that one could actually learn those functions was not only exciting, but seemed to be directly applicable to the fuzzy and ambiguous language and perception problems that I was working on. Again, this was a time when people believed they could solve speech recognition and language translation primarily by rules, a belief that seemed preposterous to me, given the enormity of facts and details that would have to be assembled. Nevertheless, simple perceptrons and similar methods also had severe limitations for speech recognition, because one could only train a single neuron at a time. The whole magic of the brain, by contrast is that it does not train single neurons but it trains entire *networks* of neurons, and that an *ensemble* of neurons can do much more powerful tasks. But how would we train an entire network?

It was just during that time that fortuitously a young assistant professor by the name of Geoffrey Hinton came to Carnegie Mellon, and started working on something called Boltzmann machines. While he was there, we had many wonderful discussions, and he introduced me to something they had been tinkering with at USC San Diego, an algorithm called backpropagation. It was much closer to what I was looking for and I immediately jumped on it. Backpropagation was a simple algorithm, an extension of the simple perceptron – except that this algorithm would now optimize the whole *network* of perceptrons and make sure that it was functioning in an optimal way. If you tell the entire network what it's supposed to do, it can in fact adjust each internal neurons in such a way that each of them will try to contribute to what is best for the whole ensemble of neurons. This seemed like a big step in the right direction, a big improvement toward classifying patterns, but for speech and language this was not enough. Because in most real world problems, recognizing patterns is not the only problem, but finding the

pattern that is to be classified in the first place. This always meant that one would have to segment a signal first to find the interesting patterns (sounds, images) before they could be classified. In speech, one would have to cut speech in such a way that you identify the beginning and the end of a particular phoneme, and then once you have that, you can try to apply a neural network for classifying these sounds and assemble them into a speech sequence. However, this meant compounding multiple separate hard problems. And the hard learn lesson in speech was that this is deadly as each of them makes mistakes. Therefore, it became clear to me that we needed a neural network that was not only a wonderful classifier but that would also recognize patterns independent of position, a property we would call "shift invariance". So what does that mean? It means that you are building a neural network that you do not just apply to a particular pattern, but that you move all neurons over a range of input, and let them essentially scan that input until it lights up whenever it finds a useful, helpful pattern. Networks of such units could thus learn to assemble all useful evident independent of small shifts in the signal. Such shift-invariance is, of course, necessary for speech, because speech flows by and changes all the time, but as it turns out it is also necessary for many other problems in AI, including images, music, games, language, and many more.

In all of these situations, your first challenge is to know *where* the useful patterns are before you can classify them correctly. Hence, classifying things by detecting them in a shift-invariant fashion was the key problem that we needed to solve. With that goal in mind, I then went to Japan as a post-doc, where I had access to some of the most powerful super-computers at that time. And with this computing power, I had the chance to develop a new model which then became known as the time-delay neural network (TDNN). It was still a multilayered ("deep") neural network, but it was now trained specifically for shift invariant classification. And as it turned out it worked fantastically well; it worked better than all other methods that existed back then.

So did this new TDNN model then replace other methods?

Sadly, we still did not have the necessary computing power to build networks that were large enough. Back then in Japan, we used the biggest supercomputers available, and compared them with other statistical or rule-based methods over benchmark data – and we found them to be much, much better. But when we tried to build larger networks and practical speech recognition systems, we still ran up against computational limits and had to make many compromises that hurt performance, and so other researchers could use simpler methods to gradually catch up, and get similar or even slightly better performances then we did. As a consequence – and this was in the late 1990s, early 2000s – people lost interest in neural networks and simply used other statistical methods.

Ten years passed and few people continued to work on neural networks, until around 2008, when – rather by coincidence – various people in the US actually tried these old neural networks again, but with the help of much more computing power and with much larger amounts of learning data that is now available over the internet. And, as it turned out, these neural network methods that had already been developed in the 1980s suddenly worked amazingly better than any other approach in the field of AI. And they did not just work a *little* better, but in fact they worked like 30 percent better. In our area, you know, entire PhD theses are written when progress of half a percent is made; so doing something that is 30 percent better is simply revolutionary. As a result, the entire community switched to neural approaches within two years. The other thing that came as a surprise is what happens when you add more layers in the network. In the 1980s, we had one or two so-called hidden layers and that was all we could compute. But now, with all this new computing power, we can do three, four, five, and more layers. No one expected that this would continue to improve performance, but it did. Today, we have networks used for speech recognition in our laboratory with 40, 80 or even hundreds of layers. And the exciting neural models that worked so well 20 years ago work even better today, too. TDNN's went on and got applied to image processing, games, speech, and other problems and became known by the more generic name: "Convolutional Neural Nets". They can now be found at the heart of most modern AI engines.

In terms of having access to powerful hardware and large amounts of data, it was certainly helpful that you worked for Facebook for some time.

Right, I was with Facebook for two years as a director, but of course I also have many friends working at Google, Amazon and Microsoft. Many of our students are now with Google, Amazon, Microsoft and so on. And many of them graduated from our labs. The massive amounts of data that these companies control is of course a treasure trove for learning programs. In those large Internet companies, they train huge neural networks over huge amounts of data using huge amounts of computing power, and the performance gains still grow. And that's surprising and impressive. But if you ask me what's the new breakthrough in AI today as opposed to 20 years ago, I would have to tell you: not that much. They are again very much the same network techniques and training algorithms as we were working on in the 1980s, except that we now use orders of magnitude more data and more computing power, and they actually work much, much better than we ever imagined.

So far, we have mainly talked about the technological aspects of speech recognition, machine translation, and ANN. Perhaps we can now talk about the political dimension of these models and applications of AI. What would you consider to be the most relevant political aspects concerning the field of natural language processing in general and speech recognition and machine translation in particular?

There is of course much to say about this. One important aspect would be the politics of research funding that affects us directly in terms of how we are doing science. Scientific support depends on political factors, and sometimes they work, sometimes they don't. And there are really fascinating differences between the US and Germany, or between countries in Europe and Asia, because each of these countries or cultures approaches scientific support differently and therefore has specific strengths and weaknesses. So I am politically active at that level to introduce and improve better mechanisms for research support in Europe. The other political dimension, though, is what we do with our research. As I have mentioned before, in my view of the world, I like to do projects with which I try to improve some aspect of society. And as I said before, research always meant for me to make people understand each other better. If with our research we can build machines that allow us to communicate better, then this means having fewer misunderstandings.

Throughout my career, I have founded several companies, and one of them was for building a handheld speech translator on a phone. It was the first mobile speech translation system on a phone ever. We launched that in 2009. The start-up company was called "Mobile Technologies" and the product was called "Jibbigo". You could speak into the phone and then the system translated the input into another language. It was a huge success. Apple, for example, ran commercials with it. It was used everywhere and people came back to us, saying: "I can finally understand my in-laws!", and "I can really understand other people!" And we also started doing humanitarian projects, for example, we built systems, say, in Thai and Khmer, so that American, European, or Japanese doctors could help rural people get healthcare, and we deployed similar things in South America.

Due to broad interest in this type of technology, the company was then acquired by Facebook (making the world "open and connected") in 2013 and for two years I led a team of scientists to build translation technology there. At Facebook, the use of the technology for translation of posts and other company use cases, however, turned out to be of higher priority than the interactive communication aspect of our speech translators I was keen on advancing, and so I returned to the University to continue our work on the educational and humanitarian aspects of this technology.

Due to Cambridge Analytica and other scandals, Facebook has increasingly been confronted with massive criticism, which is why the tech giant is all the more under pressure to meet their idealistic agenda. At the same time, companies like Microsoft or OpenAI are demanding the democratization of AI. What is your general opinion on this concept?
Again, you know, the world is much more complex than a simple slogan suggests. Facebook is a good case in point. I am sure their initial goal was to democratize news. If anyone can post news, how wonderful that can be! If anyone can provide facts on Wikipedia, how wonderful would that be! No more experts dictating their opinions, right? But if you really think this through – if anybody can publish trash about anybody and reach a worldwide audience – the benefit is not necessarily that you are making people heard that were unheard before, but you also open up a worldwide potential for abuse and manipulation. And that is exactly what we are realizing now. So democratization is fine. But the potential for massive manipulation and abuse is equally there in the same process, and therefore one has to be really careful.

In other words: You are more or less skeptical about this concept?
Once again, we should be careful and keep on thinking about what we are doing because tools like the Internet are so powerful. Sometimes you create things that have unintended consequences. And one must reevaluate the technology and strive to move it in a good direction. While the internet lead to democratization of information, we now see again massive concentration of information and power as well. Would you rather have a world in which only Google, Microsoft, Facebook, Amazon or Apple can have intelligent systems and everybody else is at their mercy with regard to using this technology? Would you like to have a world in which only one of the big tech giants can recognize anybody's face by a machine and nobody else is able to? These technologies effectively encourage monopolies, that are hording incredible amount of data and generate a lot of knowledge, but – despite the best intentions – at the same time also provide a lot of potential for manipulation. We have recently seen that this is the case with the Cambridge Analytica scandal. So, again, the question is: Would you like to see all data and AI to be concentrated with only three or four companies in the world?

These concerns also play out on a geo-political stage. While the internet was designed to be a great global unifying force, it now also threatens to break into major regional spheres with different moral and societal attitudes that compete for supremacy. In China, where there are fewer laws or restrictions to data collection and handling, we see that AI feeds the emergence of an automated mass surveillance state that is overseen by the government. Will this – by way of competition – undermine Western values of privacy, freedom, and independence? That is a 1984 Orwellian future at our doorstep, right? Indeed, democratizing it at least distributes the technology to a broader set of players and that is why antitrust

efforts are so important, domestically. But, if we talk about global balances – for example – Europe versus America: Europe does not have a large internet company and this creates asymmetries, where one continent is critically dependent on AI systems from another for its information and data management. This still works, because relations between America and Europe are amicable and supportive because both are Western democracies. But what about China, Russia, India? China is making dramatic progress in AI right know, and is spending billions of dollars on AI. So it is only a matter of time until China and others will be on par, if not more advanced then the US. And, without its own clear technology base and vision, Europe, could be at the mercy of what other players are doing and be much more vulnerable to external meddling and manipulation. For me, these are worrisome developments.

Nevertheless, it still makes a difference whether we talk about American or Chinese tech monopolies.

Right, the so-called GAFA [Google, Apple, Facebook, Amazon] have to be considered differently in some sense. In China, companies like Baidu, Alibaba or Tencent have very strong government connections. In the US, the playing field looks different because there is a public that watches these companies, and whenever one of them abuses their data power, it becomes a scandal and is immediately all over the news, which is very bad for the company. While I was with Facebook, I do have to say that – despite all the scandals – I was impressed by how much the company actually attempted to deal with their data in a responsible way. And the fact that they still produce scandals simply shows how hard it is to do that and how sensitive an issue it is. But I think that companies in the US also embrace the idea of democratizing AI because it is part of their business model. Of course, companies ultimately are very selfish and try to do what is best for them. But in the American context, protecting data is good for business, since scandals are terrible. Hence, Microsoft and Google are into democratizing AI because it supports their business strategies. Take Amazon, for example: One of its largest businesses is Amazon Web Services (AWS) that, among other things, includes renting nodes, so that a small company can do its computation on Amazon's servers. Microsoft or Google provide similar computational resources, and if they can provide AI services on top of that, then it is obviously also good for their business.

So which would you consider to be the most important political challenges of AI in the near future, from your perspective as a computer scientist?
I think that, first of all, we as scientists have a responsibility to be vigilant. But it gives rise to optimism that there are actually a lot of idealistic people working inside those big companies. Thus, the fact that there are scandals is good news, because it means that you cannot keep such things secret, that it forces society to keep thinking these things through, which is good. And regarding how politics should respond: Well, if you look at some of these senatorial debates, you realize that politicians cannot be deeply involved in every aspect of every technology, and hence may lack intuition about where it may go and how to respond. For this reason, I think it is very important for AI scientists to be vocal and active in a public dialog, so we (science, public, and governments) can ensure that we build these technologies to serve humanity, as opposed to serving our own political or financial interest.

What worries me in this context, however, is the fact that large companies are voraciously hiring scientists, and that universities have difficulties retaining talented people. And we should remember that universities are (or should be) spaces for open discussion and debate so that we are not manipulated by economic or political interests. So my point is that we should maintain a strong academic environment in all major areas in which AI is used. And this is a particular challenge in Europe: Without major internet companies, it naturally suffers from a continuing loss of talent. With a reference to my own AI laboratory in Germany, I can tell you that many of the best scientists, as soon as they are done with their pitches or degrees, get offers that are like seven times higher (or even more) than the ones we can afford at a university. And when young people are being offered those amounts of money plus a chance at building something with a major company like Amazon, Apple, or Facebook, they jump at the opportunity. In other words, the brain drain is enormous, not just from academia to industry, but between countries. Therefore, in an age of AI, Europe must move much more aggressively to provide for its future.

What can Europe do to change that? And how do politicians, people, or the public know which experts they can trust?
Well, as to your last question, I think by being less risk averse and doing more to encourage technology disruption: Europe has outstanding scientists and engineers. There is also outstanding support and freedom in Europe to carry out innovative and fundamental science. Europe has very bright, well-educated, and idealistic scientists. I could even argue that many of the top scientists in America were trained or started their career in Europe. That is not just an empty phrase, I could name many famous examples. But the area in which we are doing badly is letting the scientific advances challenge the status-quo in society. What is needed

is fast, practical, and disruptive moonshot projects. DARPA in the US, for example, has done that very successfully for the government. And so do companies like Google, Tesla, Amazon, all of which did not exist 30 years ago.

The other thing that should be improved is the technology transfer into industrial exploitation. In Europe, we actually have many entrepreneurs who start companies. The risk takers are there, the young people are there, the bright ideas are there, and the excitement and the eagerness to do this are there. What's missing is capital to support such activities and also more willingness and speed of mergers and acquisitions. For example, in the US, small companies are bought up very quickly. Some of the companies exist only twelve or eighteen months before they are being absorbed by a larger corporation. This is a healthy process as it speeds the transition from idea to concept to product to industry. But in Europe, that is very rare. Here, it is very difficult for companies to be bought. It takes a long time to go public, to enter the stock market and so on. The transition from a small successful start-up to a large business has so much friction in Europe that it misses many opportunities; speed is of the essence in this kind of game. And this ultimately drives many small companies and their young entrepreneurs to the US and China.

Another political-ethical concern that many people talk about these days is the problem of algorithmic biases. How are these problems related to your research in natural language processing and the translation of spoken languages?

I am glad that you are bringing us back to this topic. So far, we talked a lot about how AI affects society. Another important political dimension is to discuss how we pick projects that contribute to a society that we want to live in. And for me that means speech translation, because I think this is one of the big problems in Europe. Europeans speak 23 different languages, and these are only the official ones. In fact, there are many more languages in Europe. And this situation generates separation, misunderstandings, and also a considerable loss in business opportunities. One big reason why e-commerce is more challenging in Europe is because it is so fragmented. Each country in Europe has a different legal system, a different delivery system, and much of that is of course fossilized in language, because if everything has to be done in multiple languages, it complicates transnational business exchanges. But saying that everyone should learn English, Esperanto, or something like that would be ridiculous and also not desirable. In fact, having the variety and diversity of languages is something Europeans are rightfully proud of.

Against this background, technology must not be regarded as an obstacle, but as a tremendous problem solver if we want to develop a technology capable of text and speech translation that bridges these language barriers on all fronts, so that we actually have a language-transparent world. Imagine you are going to China, Russia, or Spain, and like to operate in these countries as if you are at home, with-

out any language disadvantages. But if you think that through, what such a scenario would mean if you are in all these countries without knowing the respective language, what all kind of assistance it would require so that you do not notice the language barrier anymore. And indeed to achieve this is the very vision we are working on.