Fireside Chat with Kathy Baxter,
Principal Architect of Ethical AI Practice at Salesforce
Video Transcripts

Olaf Groth: Hello everybody, this is Olaf Groth, welcome to our fireside chat session for the Future of Technology. I am very privileged and pleased today to welcome to this fireside chat, my friend and frequent collaborator, Cathy Baxter. Cathy is a principal architect for Ethical AI Practice at Salesforce. She comes to Salesforce with about 20 years of experience in the technology industry, or the technology domain, and has co-authored a number of books on user experience.

Kathy's mandate is to ensure that the AI-based innovation at Salesforce is human-centric and responsible and, as such, she's really the avant garde of software design -- the way it should be practiced, which is with the human user and stakeholder in mind. This is one of the trends that we outline in Future of Technology that is up-and-coming at the moment and that we will see blossom over the next 5 to 10 years if all goes well. Kathy, welcome to our fireside chat, thank you for joining us.

Kathy Baxter: Thank you so much for having me.

Olaf Groth: All right, let's dive right in, can you tell us, what does a principal ethics architect do -- AI ethics architect -- do at salesforce or in general, with whom and for whom?

Kathy Baxter: That is a great question. Principal architect is a fancy word, it's basically the title on the individual contributor ladder, that would be equivalent to VP on the management side of the House, and so. In 2018, I pitched this role of an AI ethicist to our then research scientist Richard Socher, saying that we needed someone to look across all of our Einstein -- that's the name of our AI platform -- all of our Einstein teams in the company, identify ethical risks and considerations and put guidelines in place to help them create our technology more responsibly, but then also putting in guard rails, in-app education, tools to help our customers then use our technology more responsibly.

And so, myself and my co-worker Yoav Schlesinger -- we are part of the larger office of ethical and humane use, and it has a couple of primary areas. We've got ethics by design, and that's working with our product teams to build in ethics into our products, build them responsibly. And that includes our accessibility inclusive design team. So when we think about ethical responsible products, we think about not only concepts of fairness, bias, transparency, explainability, and AI, but it also includes things like inclusive language. Not using words like blacklist, whitelist, master, slave, which is common in computer science terminology, but is not inclusive language. And it also includes being accessible, universally accessible to all. And then the other side of the ethical humanities house is the policy side. So what are the red lines, what are the things that we won't allow customers to use our technology for? Are there certain industries that we won't sell our technology to? And, so one of those red lines is we've never sold or allowed our AI to be used for facial recognition, for example, and so, those are the two main focuses of the team.

Olaf Groth: Got it. So as I always say, you know, ethics and economics are inseparably tied by their hips and the cognitive era, right, where everything is about automating parts of the cognitive processes in our brains. We need to pay more attention to ethics, not less, and you are at the forefront of that. Now, can you maybe just sharpen the edge here a little bit on what's at stake? Like, where could we go wrong? You know, we all have these movie visions of Terminator and Sign It, but that's not really what it's about, right? That's not what your work is about. Can you maybe put a sharper edge on what could go wrong? What do we have to avoid?

Kathy Baxter: So, in some cases, it might seem like creating technology that -- technology is neutral, there's a lot of belief that if we use technologies, that, it's going to be less biased, more fair than humans are, and for processes that are currently very manual, AI is like bacon. Yeah, it makes everything better, and that's not necessarily the case.

In some cases, it makes it worse because AI not only mirrors all of the bias that exists in society already, it actually amplifies it and makes it significantly worse. And so we see that in examples like translation of text online. You go into Google Translate or another tool, and bias in the language that it was trained in can be replicated. So, in Turkish, which is a gender neutral language, Google Translate had a problem with translating “This person is a doctor or medical professional” to “She is a nurse”, “He is a doctor”, and so we end up just replicating stereotype biases.

Amazon tried to create an AI to help identify the best candidates that were applying for their jobs because, not surprisingly, they get thousands and thousands of resumes that they have to go through. But they trained it on past data of who they had hired before and who they had promoted for and after three years, they had to give up because the AI was incapable of recommending women for roles. And so, it can be the mundane of what gets recommended from ads online to much more serious issues, like in the medical, the healthcare domain.

We've seen time and again gender and racial bias that comes through in medical predictions, because the training data isn't representative of everyone. In something even just as simple as the oximeter. So in the time of Covid, I think everyone became really familiar with that little white device that would go on your index finger to measure your oxygen level. It's not as accurate for people with very dark skin, so it can be life impacting in cases where you don't -- you aren't being inclusive and responsible in your design.

Olaf Groth: That makes sense, and you know, it's been proven lots now that, not only is this, of course, the right thing to do, we want to be responsible. We want to be the good people with good moral compass. That's the first rule, of course. But then, it's also true and it's been proven that, you know, responsible conduct, diversity -- well managed diversity --, ESG -- environmental, social, governance -- are actually top line and bottom line enhancing from an economic standpoint. So it makes sense, both from an interpersonal, societal, moral, but also from an economic standpoint to heed these things. Why should an executive, who is not in the tech industry per se, who is not designing AI, maybe, for whom AI is not at the center of the product offering -- why should they care about this? What's at stake for the broader economy?

Kathy Baxter: You know, Marc Benioff has often said, “You can do well and do good,” and we very much believe that and have demonstrated that we have recognized as a society that companies must be accountable. That they must be responsible for what they bring into the world -- you can't put the toothpaste back into the tube, as they say. And so, when we think about, how do we make decisions as companies, for who do we give loans to? Who is eligible for social benefits, like food, housing, medical care? When we are thinking about -- who are our customers? Who do we want to make sure that we are serving? And who are we not serving?

Every single one of those are ethical considerations that have much broader impact for who gets economic opportunities, who gets financial opportunities. Questions of privacy and surveillance of all kinds come into this and, so, every single corporation, as well as every single employee in those corporations have a responsibility to ask, is this the right thing to do? Should this even exist in the first place?

And Salesforce you know, whenever new employees are hired, as part of their bootcamp, we have a 30-minute ethics module to communicate why this is so important, why Salesforce values this so much, and we talked about it in the same ways that we do security. We have a massive security team, but every single employee is responsible for ensuring that they have a strong password, that they don't jump on janky WiFi, free WiFi, or that they're not clicking on dodgy links in emails, and so, it's similar with ethics -- it takes every single employee to think about and ask those hard questions.

Olaf Groth: Because ethics is a systemic problem, right? You can't be ethical in one area of the business and unethical in another without it being impacting the entire company and your entire stakeholders.

Kathy Baxter: Absolutely.

Olaf Groth: Yeah, and then, the other element, I guess, is even if you are a non-tech company with technologies that were promulgating out there that are permeating all products, I'm thinking about, you know, smart-enabled toys now, right? The management or the staff in the toy company may not be sufficiently aware, right, of the ethics pitfalls of modern AI and data science, and yet it will have an impact on how the toys are perceived, the impact of toys have, and the potential backlash that will ensue.

Kathy Baxter: Yeah, absolutely. I'm part of the World Economic Forum's working group on a project called Generation AI. And our goal is to come up with a series of guidelines for engineers, product managers, the people creating technology, AI-centered technology for kids and teens, as well as CEOs and their boards that should be overseeing this work. And then, the guardians -- parents, teachers using this technology -- give them the tools to understand, what does responsible technology look like for children, particularly when AI is involved?

Recognizing that when you develop anything for kids, it's not just -- you take an existing version for adults and take some features out or slap some cartoon characters on it. This is, you have to think very differently about it and understanding neuro-diverse children. Different contexts and being inclusive and equitable and how they get access to the information and then controlling it, being mindful of their privacy and aware of manipulative design patterns, on the very impressionable children. And what do we expect as a society, when we give these tools to kids?

Olaf Groth: As we design these ethical solutions, what are some of the common pitfalls, you know? Assume somebody who is new to this whole area is attempting to design ethical governance frameworks and ethical design frameworks, what are some of the common pitfalls you have experienced, either in-house yourself or with customers?

Kathy Baxter: I think one of the first things I often hear for people that are just stepping into this area, they immediately asked the question of “Whose ethics?” And that it can feel very subjective, and it is important to ask whose ethics are we focused on? There are nearly 200 different sets of AI principles alone that have been created by individual companies, like Salesforce, but then other organizations, governments, and the vast majority are based on Western philosophy and religious beliefs, where the individual's rights and autonomy are supreme over broader societal benefits.

And so we do have to think about whose values are we talking about? Who is being included in these discussions, and who is not? But we bring our conversations back to the UN guiding principles on human rights. So, first and foremost, our number one principle is protecting and respecting human rights and those have been agreed upon by nearly every country in the globe. And then working down from there to think about harms modeling.

So what exactly is the harm that this particular tool could unintentionally cause, and how do we, how do we resolve it? So when you start being -- when you are concrete about human rights and harms, and you ensure that you are inclusive of everyone that's going to be potentially impacted, it feels much less subjective in that particular case.

Olaf Groth: You know, that's a good -- there's some really key takeaways here, right, which is that, of course, it's always interesting to watch these conversations on different ethics principles across cultures, and there are some valid differences there. However, you know, it's very easy to get hung up on that and it's much better to start, as you say -- modeling prototyping, starting to talk about the second, third order effects, the impact of these solutions, and then talk about, does this impact -- does negative consequence matter to the stakeholders around the table? And, of course, more often than not, the answer is yes, and you see eye-to-eye on that, because nobody, as one colleague of ours at the World Economic Forum, used to say, “Whether you live in China or in Africa somewhere, or in Europe or the US or wherever, nobody likes to lose an arm in some kind of automated process” okay.

And so you know, it also, I think, behooves us to remember that if you step away from the term ethics, but maybe look at values, that value sets are often very similar around the world, but the hierarchies in which they are regarded are slightly different, right? And so there is much more common ground and difference, but we get so hung up on the differences, right, and some are important and some, I think, fade away in prototyping. So that's what gets me to the next question, which is you know, at Salesforce, how do you design ethically across all these different geographic boundaries from Mongolia to Moldova from Norway to Namibia right from Chile to Canada? I mean, you're a global company, how do you navigate them?

Kathy Baxter: Well, even more complex, we are a platform, and so we often say that we are the largest global company that you've never heard of. We're used by 95% of the Fortune 500 companies, and so, going across all these different industries, as well as countries and use cases, and as a platform, our customers can customize what our product can do, and so thinking about, how do we put in meaningful defaults? How do we put in in-app guidance to help the users, that are leveraging our products in these different use cases, make informed decisions?

We can't predict every possible application or understand their exact use case, but giving the tools to those that are using it to make an informed decision on what is the right thing to do in that particular use case. In other cases where we have to make policy decisions, then we are going to bring in region experts to really ensure that we have a deep understanding of that culture or that region's context of use, value systems, potential legal implications, and that we are not making a decision based on our US-centric values, laws, and understanding. So again, it's just so important that you are bringing in individuals from the groups that are going to be impacted by your decision making.

Olaf Groth: And you know, I suppose that would help you avoid sort of driving for ethics colonization, right, making sure that Western values don't sort of steamroll over other cultures. Thinking about, you know, different types of products that could touch on very sensitive cultural issues like reproductive health, or, you know, distribution of roles around society and things of that nature.

Kathy Baxter: Inclusive language is another example, so that is a big project that we are working on. But right now, we are very clearly acknowledging that this work is applicable to English, US-only. Words that are sensitive here may not be sensitive in other regions, and there may be words that are sensitive in other regions that mean nothing to us here. So being very specific when we are designing solutions, who they are applicable for, and who they are not applicable for.

Olaf Groth: Yeah, so really getting down into the nuances, right, and into norms, not just the high-level values, but norms that are nuanced and granular now on the use case by use case basis. Really important, yeah. So, do you give stakeholders real veto power or co-creation empowerment in your solution, or do you just -- is this like a focus group, where you go out, you get feedback and you incorporate that yourself?

Kathy Baxter: So we work with our user research and insights team to understand our customers. We will work with our solutions engineers that partner with customers in applying our solutions to their particular use case, so we are seeking information from all different sources to understand those different use cases.

Olaf Groth: I understand, okay. If you -- I want you to look out now, further, if you had access to an oracle that could give you answers to any sort of forward-leaning futures focus question on ethics, values, human-centric design, right, what would you ask the oracle?

Kathy Baxter: Oh, goodness. I think the thing that I am very much looking into the future about is the development of standards. We don't have standards right now, for, how do you clearly say this is ethical, this is not ethical. Or in the case of models, you can never say that a model is 100% bias-free. You can only say -- this is the type of bias we looked for, this was how we looked for it, and this is the amount of bias, and, just like in the pharmaceutical industry, no medication or treatment is 100% risk free.

It's about risks versus benefits, the whole Hippocratic oath of “Do no harm” -- it's not entirely accurate. The medical industry is never “do zero harm,” it's about the benefits outweighing the risks and understanding what those benefits and risks are for different populations, and so, I think really deeply seeing into the future, what are some of those standards? So that we can say for this type of model, this is where the line should be drawn to say, this is the amount of bias that is acceptable that we feel comfortable moving forward, because if the goal is perfection, it can't go out the door.

Olaf Groth: Yeah, I guess what you're saying is almost like an acid product movement toward the ideal state -- you know you can never really get there.

Kathy Baxter: Yes.

Olaf Groth: But that doesn't mean you shouldn't approximate it and have a continuous process of getting ever closer, right?

Kathy Baxter: Absolutely.

Olaf Groth: Kathy, this has been a great conversation -- I would love to talk to you for another hour about all of this. But again, as I always say, this is the beginning and not the end of our conversations, and I thank you for your time and effort today, and I look forward to the next chance to talk to you.

Kathy Baxter: Wonderful, thank you so very much for having me -- I really enjoyed talking to you.