Reflections on AI

Q&A with
Andrea Martin

"I am deeply convinced that (we) need an interdisciplinary dialogue and collaboration to create trust and acceptance in AI."

The TUM IEAI had the pleasure of speaking with Andrea Martin, Leader of the IBM Watson Center Munich & Member of the German Parliament Commission for AI. Prior to her 21 January 2021 Speaker Series Session on *Trustworthy AI: What does it mean and how to implement it?*, we were able to ask her some brief questions about trustworthy AI, what it means and how it can be implemented.

1. What is the biggest misconception about Artificial Intelligence?

One of the biggest misconceptions is that people think of AI as one big tool that can solve all problems automatically. We don't communicate it often and clear enough that AI is a toolbox with every functionality being applied separately, in a combined way or even in conjunction with other technologies. Today, it can be typically used for very specialized tasks, e.g. visual recognition, speech understanding, supporting decision making, in a particular area. (AI) must be trained, similar to us humans when we study a new area of knowledge.

2. What is the most important question right now in AI ethics?

Two points: (First,) what do we mean by AI ethics? Or rather: what is ethical behavior in our societies and what does this mean for AI? This is a question that needs to be answered in an ongoing dialogue – and it may have different answers around the world, in different cultures.

Secondly, it's about how we want to implement ethical AI. Many companies or institutions discuss principles for ethical AI, but, few really think about how they can make ethical AI a reality and how to govern it.

3. What is the role of academia, research institutions and other centers when it comes to the ethics and governance of AI?

Research and academia can provide important input and insight into the discussion of ethics and governance of AI. They can also support empirical studies on ethical AI – i.e. what's the degree of implementation in companies, how is the acceptance among managers, employees, but also society at large. What does it take to deploy AI solutions that are accepted, i.e. how much trust do you need in order to accept AI. I am deeply convinced that (we) need an interdisciplinary dialogue and collaboration to create trust and acceptance in AI. This means all of us need to come together: politics, technology providers, sociologists, philosophers, unions, academic/non-academic research and of course us, society at large.

4. What role should technology companies play in the developing and implementing AI ethics?

I personally think society and business will decide in the end who they trust with their data – may it be personal data or enterprise

data. Therefore, technology companies take their responsibility seriously when they bring powerful technology like AI services into the world and the market. They need to think about guiding principles for AI ethics, but also how to implement it. And of course, they must ensure to comply with all regulatory requirements.

5. What does it mean for AI to be trustworthy? What does this mean in practice?

There is interesting research going on how transparency creates trust, but also how too much transparency can reduce trust. Having said this, I still believe that transparency about data sources and basic techniques used, explainability of AI recommendations, assurance of fairness – whatever someone considers as "fair" –, anti-bias and anti-discrimination, assurance of data privacy and robustness of AI solutions – so that neither the data nor the results can be manipulated – are core elements of trustworthy AI. In addition, development and project approaches should ensure that all these elements are considered from the very beginning, not only when you try and transfer a solution to production. This also includes that all of these things are not a one-time effort, but need to be embedded in an ongoing governance and quality assurance process.

6. We often say that AI is changing or transforming the world. To what extent is AI changing us as humans?

I believe that we will see much more hybrid intelligence in the future, i.e. a hybrid "workforce" where humans and machines work hand in hand to solve problems. Even today, we rely on AI technologies a lot in our day-to-day life (e.g. navigation systems, search machines, GPS systems) sometimes without recognizing it. Just think of search machines, automated tagging of pictures, or visual recognition used for medical diagnosis.

Where humans are unique is when it comes to qualities like creativity, having an intention, setting goals, team building and team work. Also, what is important for us is critical thinking and the ability to make decisions on our own. And these are areas where we need to make sure we don't lose these critical thinking capabilities and where we need to particularly focus on in the education and training of future generations – because it is just so easy to believe in what a machine tells you. It is important to challenge and question AI – not just in technology, but in our lives, too.

Meet the expert

[Andrea Martin](#) heads both the IBM Watson Center Munich and the IBM EMEA Client Centers. Since July 2019, she is responsible for its scope and market impact. She is also an appointed expert for the Commission for AI of the German Parliament. In 2017, she was awarded the Women of the Year Prize for Technical Innovation. Outside the company, she is an expert in the AI Enquete Commission of the German Bundestag and for the Society of Women Engineers.

Andrea Martin draws on her extensive experience from more than 25 years of international service business in the areas of innovation, artificial intelligence, digitalization, IT-transformation, IT-strategy, service management and architecture.