Let's Talk AI with Markus Langer

Markus Langer[1] and Barbara Steffen[2]

[1] University of Freiburg, Department of Psychology, Chair of Work and Organizational Psychology, Germany
markus.langer@psychologie.uni-freiburg.de
[2] METAFrame Technologies GmbH,
barbara.steffen@metaframe.de

"Artificial Intelligence is a moving target – what is considered to be "Artificial Intelligence" today will be "Automation" tomorrow."

The Interviewee - Markus Langer

My Personal AI Mission:
Bringing more psychology into AI research and bringing more AI research into psychology.

My Takes on AI

Artificial Intelligence: An artificial system that is able to mimick human functions such as decision-making.

Trust: The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party [10, p. 712]

Explainability: A system that enables humans to better understand its functions, outputs, limitations, and capabilities is a system that aims for explainability.

Essential Elements of Human Capabilities: Intuition and needs such as relatedness.

The Interview

Barbara *Hello Professor Markus Langer, thank you for taking the time for this interview. Could you please briefly introduce yourself and your personal relationship to artificial intelligence?*

Markus My name is Markus Langer. I am a professor of work and organizational psychology at the University of Göttingen (remark: now at the University of Freiburg). My interest in AI began with my master thesis, where I delved into the topic of AI in hiring and personnel selection [8]. I was intrigued by how people react to AI in high-stakes decision-making situations and how they use AI-based systems in their decisions. This psychological perspective on AI-based decision making has continued to influence my career.

Barbara *Are there specific challenges or examples of challenges that you are currently addressing with your AI research?*

Markus For example, we are currently invested in the topic of effective human oversight. This is a concept called for in the European proposal for an AI Act, but also across the globe [3, 5] which states that humans need to oversee AI-based systems in high-risk context to reduce the risks otherwise associated with the use of such systems. We are exploring how humans can effectively oversee these systems and what they need from a technical and workflow perspective to detect erroneous outcomes and unfair outputs [7].

Barbara *What role do you think trust plays in the adoption of AI? And what steps should we take to ensure ethical AI adoption?*

Markus Trust is a significant research topic we are currently investigating because it is closely related to effective human oversight. On one hand, we do not want people to blindly trust AI-based systems. On the other hand, we do not want them to under-trust either [9]. For instance, if people constantly monitor what the system is doing, like someone sitting in an autonomous car and constantly trying to drive it themselves, this might rather increase the risk in the operation of AI-based systems and defeat the purpose of involving a human being for risk reduction. So, we need to find the right balance of trust in the AI-based system. This is also related to ethical questions, such as how we introduce these AI-based systems in high-risk situations. Can we tell people how to interact with the system? Should we change workflows in a way that makes people aware of the system limitations [2]? But at the same time, they might feel like such changes in the workflow reduce their agency in interacting with the system. In the end, these topics are all closely related [11].

"[...] in healthcare, we might want robots to assist people, but do we want them to take over all human-related tasks? Do we want robots to have such advanced interpersonal capabilities that people cannot distinguish between a human and a robot? Or do we want to reserve certain tasks for humans?"

Barbara *In your conference presentation, you mentioned that trust is usually something personal, that each person has a subjective level of trust. So how much should we trust and who should or could determine what an appropriate and healthy level of trust is or looks like? For example, if trust varies a lot and early adopters tend to trust more easily, they might benefit from trusting and being first, but as a society we are probably better off if not everyone became an early adopter. What do you think about it?*

Markus That is a very good question. Trust is indeed subjective to a certain extent. In a current conceptual research project we propose that, while there are certain standards that we can probably agree upon, for instance, that a system's accuracy is important for its trustworthiness, everyone may have slightly individual standards of what they consider to be trustworthy [12]. For instance, an expert medical doctor might require a system to be 95% accurate before considering it to be trustworthy, whereas assistant doctors or doctors in environments with little available medical infrastructure might be satisfied with 85% accuracy. The debate on what constitutes trustworthy AI is ongoing in the European Union and globally. Is trustworthy AI fair AI? Does it adhere to privacy principles? But as I mentioned earlier, this is also a cultural perspective. From a European perspective, an important value is privacy, while other cultures might prioritize other factors when developing AI-based systems. This shows the subjectivity and culture-specific perspective on trustworthiness and trust. This is why we propose that it is hard to say that a system itself is trustworthy. We need a common understanding and agreement on what we expect from these systems, which requires a societal debate about what we consider to be trustworthy.

"[...] humans tend to anthropomorphize things, [...] They anthropomorphize their cars, pets, and AI systems. [...] I have repeatedly experienced that even after explaining what GPT is and its technical foundations, people still ascribe human characteristics to it ... one person I was talking to insisted on ChatGPT having a personality."

Barbara *Thinking about the possible technical capabilities of artificial intelligence in the future, on a scale of 1 to 10, where 1 refers to artificial intelligence systems like ChatGPT, and 10 refers to general artificial intelligence systems that surpass human capabilities in all areas. What do you think will be possible in the future?*

Markus The future is context-dependent. I can already see systems that surpass human abilities, and I believe we will see more of these. For example, systems that can better detect medical conditions in images (e.g. [6]) or react more quickly to complex situations, for example in nuclear power plants or train traffic. No human can compete with such information processing capabilities and speed. However, in other situations, it is a question of whether AI can and should surpass human abilities. For instance, in healthcare, we might want robots to

assist people, but do we want them to take over all human-related tasks? Do we want robots to have such advanced interpersonal capabilities that people cannot distinguish between a human and a robot? Or do we want to reserve certain tasks for humans? These are the questions we need to discuss. So, on a scale from one to ten, it strongly depends on the context.

Barbara *What is context dependent? Whether it is possible or whether we would like it to be possible and would welcome this advancement in our lives? If it were just a matter of what the potential capabilities are, on a scale of 1 to 10, what do you think is possible?*

Markus I think both are context dependent. What is possible and whether we want these advancements in our lives. However, with respect to "what is possible" I think ChatGPT has shown me that I need to be careful in predicting what really will be possible. I was one of those who were skeptical, maybe in 2018, about whether an AI-based system could soon interact naturally with people. I have to admit ChatGPT surprised me. But what I still think what may never be possible is that we have interactions with AI-based systems, where people feel that an AI-based system is "caring" for them. For instance, one of the most important positive effects of psychotherapy is the relationship between patient and psychotherapist. I am not sure whether there will ever be the same positive effect possible with an AI-based system because I am not sure whether we as humans will ever feel that an AI-based system really cares for us and our well-being. But then again, already today there may be situations where it is hard to distinguish whether we are talking to a system or a human being, so if we really believe that we are talking to another person, we may also really believe that this other is caring about us.

Barbara *So how good are humans at distinguishing a system like a robot from a human? This may not be a pressing question today, but it will become relevant as we encounter, for example, artificially intelligent robots. As a species, we have never encountered similar situations where we can talk to machines in natural language, share ideas and emotions, etc. Will we be able to keep our distance or not?*

Markus If we do not make it clear what is a human and what is a system, there will be a point where this distinction will be very difficult. Even now, if we are on the phone with a high-quality conversational agent, it is hard to distinguish between a human and an AI-based system. In the future, we may need to disclose this. However, even if we disclose this, humans tend to anthropomorphize things, seeing human aspects in many things [4]. They anthropomorphize their cars, pets, and AI systems. For example, I have repeatedly experienced that even after explaining what ChatGPT is and its technical foundations, people still ascribe human characteristics to it – for instance one person I was talking to insisted on ChatGPT having a personality. This will most likely increase in the future.

Barbara *It is difficult when the behavior shows human patterns and aspects to say, "Yes, that is true, but because of what is going on inside, it is different and therefore cannot be considered human behavior". Is there even a clear definition of what makes us human?*

Markus I am not really sure whether there exists such a clear definition.

Barbara *Is there a specific approach to say "what makes us human" from the perspective of psychology, or does this topic belong to a different discipline?*

Markus I believe philosophy may be a more suitable discipline for this. As a psychologist, you might discuss elements like emotions or intuition. These are aspects we attribute to humans, and we assume that a robot or an AI-based system does not possess these. But at what point do we decide that it does? I had a similar conversation at a workshop about ChatGPT. Someone asked me, "At what point do you say that the AI is self-conscious?" My response was, "I think that is a question for philosophers." The person persisted, and I suggested, "When it can feel pain." But he was not convinced, arguing that a robot could be built with sensors to react to pressure with an expression of pain. I countered that it is not quite the same, at least to me this does not feel the same. I probably should have said something like "What makes us human? Probably the sum of all the characteristics of humans."

"I also wonder if treating AI more gently or humanely affects the output."

Barbara *But is it not the case that for all the things that supposedly make us human, there are also people in our society who display this ability or behavior in a more extreme way? There are autistic people, for example, who are unable to assess their own emotions and those of others in the way that would be considered "usual human behavior".*

Markus This probably depends on what we consider to be "usual human behavior", but yes there is a continuum of human behavior between behavior that "most humans" show and behavior that is more seldom or extreme.

Barbara *Nevertheless, they can learn certain patterns that help them integrate better into society. And we would never begin to say: "You are no longer human." Another example is the feeling of pain. There are also people with nerve damage who can no longer feel pain. So, it is really hard to make a clear distinction that does not fall short in certain cases that we also see in humans.*

Markus That is true. There are also individuals who lack the ability to feel empathy for others. However, they learn to behave in a way that makes others perceive them as empathetic. They understand what is expected of them and react accordingly to appear empathetic. That is what we all learn. We learn how to react in certain situations, like when a dog barks at us. And yes, what then makes a human human? Where do we draw the line? How do we identify that? Referring back to the earlier question, it becomes increasingly problematic.

As I mentioned, I can explain things like how neurons work in the brain, how the brain functions, how emotions work. You can find some similarities in an AI-based system. It is not identical to how humans function. That is definitely true. But you can envision futures where there are also, for example, biological components. It will become increasingly difficult to distinguish.

Barbara *Well, considering everything we just said, there are a lot of different opinions about what the future could look like. If you were to take the whole spectrum from utopian to dystopian, where would you position yourself?*

Markus Well, again, it depends on the context, I believe. In some contexts, I think we are already living in a dystopia. To me, the entire world of social media is more dystopian than utopian. It seems not very controlled anymore. You see a lot of polarization due to the algorithmic decision-making [2]. On the other hand, I can envision a world where AI-based systems support human development, make our energy system more efficient, improve energy distribution across Germany, across Europe. Or consider healthcare, where we are understaffed globally. If robots could assist us in performing tasks, that would be incredibly valuable. Of course, this raises questions about whether we want this and whether people will accept it. But the potential is undeniable. It would be a fantastic future where humans have the freedom to perform meaningful tasks while robots handle the dull and repetitive ones. But currently, I lean in my impression more towards the dystopian side of things.

"AI is learning our less desirable behaviors [...] [and] is also causing less than optimal behaviors that lead to mental health issues and polarization. [...] the impact of implementing this AI on society, rather than the AI learning from society, [is] the dystopian aspect for me right now."

Barbara *We have often heard the comparison that AI learns more like children, compared to normal programming languages, which you basically design with specific rules, formulas, and so on. Looking at how society has evolved in recent years, especially in terms of our behavior on social media platforms, is the fact that AI is learning from our behavior particularly scary or dystopian to you?*

Markus It is not just that it learns from these behaviors, it is more that it instigates these behaviors. Sure, it learns from the behaviors we exhibit on social media, but the mechanisms designed to engage people, to keep them scrolling and staying on the platform, are causing these behaviors. If you introduce a – theoretically existing – politically neutral individual to TikTok or Twitter, you will likely see them lean towards one political direction based on the content they first encounter. And this will further drive them towards this direction. So, on one hand, AI is learning our less desirable behaviors, but on the other hand, it is also causing less than optimal behaviors that lead to mental health issues and polarization [2]. So, the impact of implementing this AI on society, rather

than the AI learning from society, seems to be the dystopian aspect for me right now.

Barbara *On a personal note, when you use ChatGPT for example, how do you treat it and communicate with it? Do you treat it like a machine that should follow your commands, like "I need this and that", or do you treat it more like a human being, like "please write xyz?" ChatGPT may not be learning from our interactions yet, but it may do so in the future. Do you think it is better or smarter to engage with a friendly "please" and "thank you" or is it just a transaction with a machine, so why bother?*

Markus I often use polite language when interacting with AI, such as "'please" or "can you", similar to how I would ask a student or colleague to complete a task. I have never really thought about why I do this, it just feels right. ChatGPT, for instance, is designed to mimic human conversation. It even apologizes for its actions, which is quite anthropomorphic. OpenAI could have chosen a different approach, but they opted for a more human-like interaction. I am not sure if there are any studies on this, but I wonder how many people interact with AI in a human-like manner versus those who treat it purely as a tool, giving it instructions without any human-like interaction. I also wonder if treating AI more gently or humanely affects the output. It might, as it provides some context to the task environment, but I am not sure if that is the case.

Barbara *Looking back on the last few days and especially on the interdisciplinary sessions. Was there something that was particularly interesting to you, like an insight from another discipline?*

Markus The legal perspective always fascinates me. Every time I hear about legal perspectives on AI, I am struck by the complexity of regulating AI or any product. It is astounding how minor decisions can have far-reaching effects. For instance, if the AI Act includes the call for effective human oversight of AI-based systems – not saying that this is a "minor decision" – this will shape the future of our society, influence the development and research of these systems, and determine where research funding goes. I find it impressive how the legal perspective impacts so many aspects of our lives.

Barbara *Would you say that the legal perspective, for example, should be informed by the psychological, philosophical, technical, and other perspectives to ensure a necessary basic understanding. For example, a basic understanding of how people function in order to develop rules that protect humans from negative interactions with AI. So, the law itself should be approached in an interdisciplinary way?*

Markus Absolutely. It is interesting to see how legislation develops with interdisciplinary input. The European Union, for example, consults experts about issues like automation bias in human-system interaction and incorporates their insights into regulations. However, I believe that all disciplines involved, such as psychology, law, computer science, and philosophy, should include an interdisciplinary discourse to exchange such ideas. This is exactly what we are doing at

AISoLA. It is crucial to understand at least a little about how each discipline operates. Ignoring any one perspective can lead to problems down the line. If you ignore the legal perspective, your developments may never be available on the market. If you ignore the psychological perspective, people may not accept or use your product. If you ignore the technical perspective, you might regulate something that will never exist. If you ignore the philosophical and ethical aspects, your insights may lack ethical grounding. So, interdisciplinary collaboration is of utmost importance.

Barbara *Is there a specific research question you would like to see addressed?*

Markus One complex question that interests me is how we can understand AI-based systems. We develop all these black box models where no one understands how they work. The question from every perspective is: what constitutes understanding? What level of understanding is sufficient? How do we facilitate understanding, and how do we measure it? Should "understanding" even be a concept in regulation? At present, it seems elusive, and I would like to see more research focused on this aspect.

Barbara *From your personal perspective, what should be the AI vision for the future?*

Markus I think Kevin Baum made an excellent point in his talk today about the need for a positive normative view on AI development. We need to ask ourselves what we want from these systems in the long term. Different stakeholders will have different interests. For example, a social media site developer would want people to visit their site frequently. But from a societal perspective, is that what we want? If not, what are the implications? Can we regulate usage, for instance, by restricting access for those under a certain age? The vision should be broad and consider what we want the system to achieve in the long run.

Barbara *Is there anything you would like to add?*

Markus Not really. I just think it is important for everyone to try and understand other disciplines. Staying within your own discipline may undermine progress in the future, as everything is interconnected. This is especially true for AI. We need to integrate all perspectives and avoid letting one dominate. If one perspective dominates, others will be ignored, which is not advisable.

Barbara *Thank you Markus for your time and perspective on AI. Enjoy the next few days at AISoLA!*

Markus Thank you.

References

1. Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1–21. https://doi.org/10.1145/3449287

2. Burton, J. (2023). Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization and political violence. Technology in Society, 75, 102262. https://doi.org/10.1016/j.techsoc.2023.102262
3. Enqvist, L. (2023). 'Human oversight' in the EU artificial intelligence act: What, when and by whom? Law, Innovation and Technology, 15(2), 508–535. https://doi.org/10.1080/17579961.2023.2245683
4. Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864
5. Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. Computer Law & Security Review, 45, Article 105681. https://doi.org/10.1016/j.clsr.2022.105681
6. Haggenmüller, S., Maron, R. C., Hekler, A., Utikal, J. S., Barata, C., Barnhill, R. L., Beltraminelli, H., Berking, C., Betz-Stablein, B., Blum, A., Braun, S. A., Carr, R., Combalia, M., Fernandez-Figueras, M.-T., Ferrara, G., Fraitag, S., French, L. E., Gellrich, F. F., Ghoreschi, K., ... Brinker, T. J. (2021). Skin cancer classification via convolutional neural networks: Systematic review of studies involving human experts. European Journal of Cancer, 156, 202–216. https://doi.org/10.1016/j.ejca.2021.06.049
7. Langer, M., König, C. J., Back, C., & Hemsing, V. (2023). Trust in artificial intelligence: Comparing trust processes between human and automated trustees in light of unfair bias. Journal of Business and Psychology, 38, 493–508. https://doi.org/10.1007/s10869-022-09829-9
8. Langer, M., König, C. J., Gebhard, P., & André, E. (2016). Dear computer, teach me manners: Testing virtual employment interview training. International Journal of Selection and Assessment, 24, 312–323. https://doi.org/10.1111/ijsa.12150
9. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392
10. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(2), 709–726. https://doi.org/10.2307/258792
11. Parker, S. K., & Grote, G. (2022). Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. Applied Psychology, 71, 1171–1204. https://doi.org/10.1111/apps.12241
12. Schlicker, N., Uhde, A., Baum, K., Hirsch, M. C., & Langer, M. (2023). A micro and macro perspective on trustworthiness: Theoretical underpinnings of the Trustworthiness Assessment Model (TrAM). PsyArXiv. https://doi.org/10.31234/osf.io/qhwvx