Let's Talk AI with Martin Wirsing

Martin Wirsing[1] and Barbara Steffen[2]

[1] LMU Munich, Department of Computer Science, Germany,
wirsing@informatik.uni-muenchen.de
[2] METAFrame Technologies GmbH,
barbara.steffen@metaframe.de

"AI is a powerful man-made tool which should be used to improve and support our lives. It should help us achieve the global human goals by promoting health, combating poverty and fighting for our climate."

The Interviewee - Martin Wirsing

My Personal AI Mission:
Explore and foster the use of LLMs for systematic development of safe software systems. In "LLM-assisted Software Engineering" [1] an LLM acts together with other supporting bots and tools to help the human developers in all phases of the software lifecycle.

My Takes on AI

Artificial Intelligence (AI): AI is the ability of a machine to display human-like capabilities such as reasoning, learning, planning and creativity [10].

Trust and Trustworthy AI: Trust refers to the "assured reliance on the character, ability, strength, or truth of someone or something." [11] Trustworthy AI should have the properties of being: (1) lawful - respecting all applicable laws and regulations; (2) ethical - respecting ethical principles and values; (3) robust

- both from a technical perspective while taking into account its social environment [9].

Large Language Model (LLM): A language model is a mathematical model for analyzing and constructing texts in natural language. An LLM is a large-scale, pre-trained, statistical language model based on neural networks [4].

Essential Elements of Human Capabilities: Life; bodily health; bodily integrity; senses, imagination and thought; emotions; practical reason; affiliation; other species; play; and control over one's environment are the 10 "central human capabilities" that are needed for "the dignity of a human being" [5].

Global Human Goals: In 2015 the United Nations defined 17 world development goals called the Sustainable Development Goals with the aim of "peace and prosperity for people and the planet, now and into the future" [12].

The Interview

Barbara *I am delighted that Professor Martin Wirsing is sitting here with me. Could you please briefly introduce yourself and your relationship to artificial intelligence?*

Martin Thank you. I'm delighted to be here. As you mentioned, my name is Martin Wirsing. I'm an emeritus professor at LMU Munich, specializing in computer science, specifically in software development and formal methods. I've been involved with artificial intelligence for at least 30 to 40 years. My primary focus has been on symbolic methods such as logic and constraint reasoning. Sub-symbolic methods like neural networks are relatively new to me. We strive to develop safe systems in artificial intelligence and ensure that AI tools are safe for users, among other applications.

"My greatest fear is that politicians may misuse AI. They might claim that their AI system is an AGI which is superior to humans, and therefore, humans should follow [their] directives."

Barbara *Can you give one or two examples of research questions you are currently addressing in artificial intelligence?*

Martin Let me mention two research questions. The first involves learning algorithms. If I examine the output of a learning algorithm, for instance, in decision support, I want to ensure the decisions are safe and within a well-defined boundary to prevent unsafe actions (see e.g. [7]). The second question relates to my main field of software development. How can we enhance software development with learning tools, specifically with LLMs? Will LLM-Assisted Software Engineering profoundly change the way develop software? [1]

"I'm apprehensive about AGI. I don't believe machines can truly express genuine emotions, though they can simulate them."

Barbara *Is the role of trust in AI adoption related to your first research question?*

Martin Yes, trust is crucial in AI. Trust is a vital component of human relationships, but since LLMs behave like agents, it becomes a top priority. Trust is derived from quality. This encompasses many aspects, for example, that a system should consistently provide accurate answers and prove to be reliable. These quality attributes are essential for software and should also be met by AI systems. Otherwise, they won't be truly useful or widely adopted.

Barbara *Would you suggest that we start using AI in areas where we can control the outcome? Basically, in contexts where we know the correct results or where we have clear expectations that put us in a position of greater control.*

Martin It really depends on the applications. As researchers, we are involved in applications across all fields. We aim to make these systems usable, and they

become usable when quality is built into them. AI systems are less predictable than other systems, so we need to apply special validation and verification techniques.

Barbara *What do you think is the role of ethics in the adoption of AI? Do you have any important measures in mind that you would suggest?*

Martin Indeed, ethical considerations are paramount to me and they should be taken into account right from the requirements stage. When developing a system, we should question the ethical values the system should embody (see e.g. [9]). This way, we can put safeguards and shields in place during development and we can ensure that the AI systems align with the ethical requirements.

Barbara *If you look at AI and its potential future capabilities on a scale of 1 to 10, where 1 represents the artificial intelligence systems like ChatGPT that we see today, which are dedicated to a concrete context and functions, and 10 represents something like an artificial general intelligence (AGI) system that surpasses human capabilities. What do you think will be possible?*

Martin Well, I would say around 6. Software has always, in some ways, surpassed human capabilities. It started with something as simple as a pocket calculator, which performs multiplication – i.e. some form of reasoning - far better than I can by hand. So, software and software systems have always excelled in specific fields compared to humans. With AI, these fields expand. There are many areas where, due to AI, software can now outperform humans. Games like chess and Go [6] are examples; LLMs are able to perform impressive conversations (see also [4]). However, it's certainly not artificial general intelligence. I'm apprehensive about AGI. I don't believe machines can truly express genuine emotions, though they can simulate them. I believe we will need and use AI for many different applications, but not for dictating our actions.

"AI systems are less predictable than other systems, so we need to apply special validation and verification techniques."

Barbara *So it's more about how people adopt AI, rather than that AI will or should become autonomous?*

Martin Yes, autonomous systems are a crucial field. I've worked on this for several years (see e.g. [7, 8]). We need autonomous machines that support us in our daily live and in our jobs. However, my greatest fear is that politicians may misuse AI. They might claim that their AI system is an AGI which is superior to humans, and therefore, humans should follow the directives of the AGI (see e.g. [2]). So politicians could abdicate responsibility, treating the machine as an infallible entity. But ultimately, it's their responsibility to guide their country, and they shouldn't treat machines as infallible. This horrifying scenario could be exploited by authoritarian politicians, and I'm firmly against it.

Barbara *That brings me to my next question. Many theoretically possible future scenarios are being discussed today. They range from dystopian to utopian. Where would you position yourself?*

Martin I believe AI, like any form of software, should enhance and support our lives. AI should help us achieve the global human goals [12]. We need to promote health, combat poverty, and fight for our climate. AI could be an invaluable tool for these issues, as it enables us to solve more complex problems than we were able to before.

Barbara *Because AI addresses these challenges in a more holistic way, or why would AI be helpful in these cases?*

Martin Yes, it is a holistic approach which encompasses many different fields. AI is software, and this software should be used together with engineering solutions and social and political measures. Computer scientists don't create the applications; they come from other fields. Software engineers always need to collaborate with other fields. To solve these huge problems, computer scientists need to cooperate with scientists from many other disciplines.

"When developing a system, we should also question the ethical values the system should embody."

Barbara *That leads me to my next question. Reflecting on the last few days, what are some insights that were particular surprising or interesting to you?*

Martin The discussion about ethical AI is very important. We also need to understand and discuss the EU regulations. I believe we should discuss these issues with our students. In their education, students need to learn about ethical aspects and how to incorporate them into systems.

Barbara *Is there a specific research question or topic that you would like to see addressed by multiple disciplines? If so, which disciplines should be involved?*

Martin As a software engineer, I've worked on systems for predicting various outcomes. For instance, for predicting how climate and land use will change in a region like Bavaria, you need to collaborate with many disciplines, including meteorologists, plant scientists, economists, and tourism experts [3]. For all major and important applications, we need cooperation between different disciplines.

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

Martin As I mentioned earlier, AI should be used to enhance our lives and help to solve the world's global problems.

Barbara *Do you have anything to add?*

Martin No, thank you. I enjoyed our discussion. I wish you all the best and success in your career, whatever you choose to do.

Barbara *Thank you, Martin, for your time and your perspective on AI. Enjoy the rest of the conference!*

Martin You're welcome.

References

1. L. Belzner, T. Gabor, M. Wirsing: Large Language Model Assisted Software Engineering: Prospects, Challenges, and a Case Study. AISoLA 2023, 355-374.
2. W. Hesse: Das Zerstörungspotenzial von Big Data und Künstlicher Intelligenz für die Demokratie. Inform. Spektrum 43(5), 2020, 339-346.
3. W. Mauser, M. Prasch (eds.): Regional Assessment of Global Change Impacts. Springer, 2016.
4. S. Minaee, T. Mikolov, N. Nikzad, M. Chenaghlu et al.: Large Language Models: A Survey. CoRR abs/2402.06196, 2024.
5. M. Nussbaum: Frontiers of Justice: Disability, Nationality, Species Membership, Cambridge, MA: Harvard University Press, 2006.
6. J. Schrittwieser, I. Antonoglou, T. Hubert, M. Chenaglu et al.: Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. Nat. 588 (7839), 2020, 604-609.
7. M. Wirsing, L. Belzner: Towards Systematically Engineering Autonomous Systems Using Reinforcement Learning and Planning. Analysis, Verification and Transformation for Declarative Programming and Intelligent Systems 2023, 281-306.
8. M. Wirsing, M. Hölzl, N. Koch, P. Mayer (eds.): Software Engineering for Collective Autonomic Systems - The ASCENS Approach. LNCS 8998, Springer, 2015.
9. European Commission: Ethics guidelines for trustworthy AI. Published 08/04/2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, accessed 26/02/2024.
10. European Commission: What is artificial intelligence and how is it used? Published 04/09/2020, https://www.europarl.europa.eu/topics/en/article/20200827STO85804/, accessed 26/02/2024.
11. Merriam-Webster Dictionary: Trust, https://www.merriam-webster.com/dictionary, accessed 26/02/2024.
12. United Nations: The 17 Goals. https://sdgs.un.org/goals, accessed 28/02/2024.