Comparing Notes on Trustworthy AI 2026

Session 4: Agents in the World: The Ethics of Embodied AI.

German version below

Host: LMU Munich & Munich Center for Machine Learning (MCML)
Date: 5 October 2026
Location: Carl Friedrich von Siemens Stiftung, Munich

The fourth event in the series asked what is ethically at stake once artificial agents act in the physical world. It opened with Fritz Lang's Metropolis (1927): the idea of an independently acting machine is almost a century older than the technology now making it real, and film has portrayed it almost uniformly as a threat. Moderator Benjamin Lange then sharpened the question. Robots have worked in caged factories for decades, so what changes when generative AI moves into homes, hospitals and streets, and does a body make any ethical difference at all?

The speakers

Prof. Dr. Laura Crompton is professor of AI ethics at the Technische Hochschule Ingolstadt, with a background in robot ethics. She opened with a question rather than a position: how do the ethical lessons of generative AI translate to systems that act physically? Over the evening she became the panel's strongest voice against replacing human capacities with technology.

Prof. Dr. Frauke Kreuter holds the chair of statistics and data science in the social sciences and humanities at LMU Munich and the MCML, and co-directs the Social Data Science Center at the University of Maryland. A sociologist by training, she works on data quality, fairness and data protection, and hosts a podcast. She took the deliberately provocative line of going all in on embodied AI: its risks are too complex to map in the abstract and will partly have to be learned by doing.

Prof. Dr. Michael Klenk is professor of practical philosophy at the Technical University of Nuremberg and works on manipulation, autonomy and the ethics of AI. His core message was that embodied AI should be built to make people more autonomous, and he kept returning to what its influence is for.

The discussion was moderated by **Prof. Dr. Benjamin Lange**, Group Leader for the Ethics of Artificial Intelligence at LMU Munich.

Note on these summaries: *This summary was prepared from the session transcript and notes taken on the evening of the event, compiled with the support of AI. It reflects the tone, arguments and perspectives shared by the speakers and participants on-site, and captures the state of the discussion, not necessarily the published position of the individual speakers or their institutions. For questions or corrections, please reach out to the appliedAI Institute for Europe.*

Discussion

What changes when AI leaves the screen and enters a body?

The three opening statements staked out the positions the evening would return to.

- Laura asked how the ethical lessons of generative AI, learned in screen-to-human interaction, carry over once a physical entity shares a room with a person.
- Frauke went “all in”. The problem is too complex to map in the abstract, and a lot will go wrong. The new issue is perception: a system with eyes registers the credit card on the bedside table and the non-verbal cues people do not even read themselves.
- Michael saw new channels of influence. Setting autonomous weapons aside as the most imminent risk, he framed everyday use as a choice between a new age of enlightenment and becoming passengers in our own lives.

Coping with laziness, or a partner that asks questions

Cars made it tempting to drive rather than walk, people lost muscle, and gyms appeared to compensate. Frauke expects embodied AI to repeat the pattern, as writing may already show, because people are inherently lazy. The design challenge is to build systems that enhance people: a chatbot that asks questions instead of answering them, or a robot that does a task with you rather than for you.

The panel agreed early that it depends on the task. Meta-analyses on human-AI cooperation show benefits for some tasks and none for others, which makes global design rules dangerous. Learning from every interaction would also need data traces, which reopens the privacy question.

Michael pressed on what it means for something to work. Embodied systems read and appeal to emotions better than chatbots, so their influence will likely be greater, and the question is what it is for. His answer: help people identify their reasons for acting and act on them. Frauke’s test case was a robot helping her lift her father into the shower. Michael separated plain execution from real choices, such as whether her father should stay in longer, where the system should help her think rather than push her.

Norm-setting for the robots to follow

Who decided that 10,000 steps a day is right? Laura used the e-health standard to show that a robot cleaning a room or helping someone shower also needs a norm for what ‘clean’ or ‘long enough’ means. Before these systems enter daily life, we should ask which everyday norms they answer to and who sets them. Today it is the companies, through default settings that play on human biases. It should, at least in part, be the person who owns the system.

Note on these summaries: *This summary was prepared from the session transcript and notes taken on the evening of the event, compiled with the support of AI. It reflects the tone, arguments and perspectives shared by the speakers and participants on-site, and captures the state of the discussion, not necessarily the published position of the individual speakers or their institutions. For questions or corrections, please reach out to the appliedAI Institute for Europe.*

Three answers to who sets the norm.

- **User configuration.** A guest compared it to a car, where drivers already set the radio and the lights, so a robot could learn that its user likes longer showers.
- **Realism about laziness.** Frauke expects people to dodge any robot that asks them to reflect, just as almost everyone clicks through cookie banners.
- **Second-order preferences.** Michael agreed people are often weak-willed in the moment, but desire more strength of will. Systems should help us become who we want to be rather than mirror who we are.

That raised Michael's own dilemma. An embodied fitness trainer might say "do it for me, because you like me", and he might be attached enough to comply. Is that good because it worked, or should the robot explain why he should go for another run?

Does a body make AI more persuasive?

In a US study Benjamin recalled, most participants obeyed a robot telling them to throw books in the bin, against a handful when the instruction came from a screen. Michael added two cases: people hesitating to switch off a robot that pleaded with them, and AI companion users who kept chatting when told they could not leave. People know it is a machine, but the reaction is visceral.

Laura explained part of this with sociomorphing, a concept from the philosophy professor Johanna Seibt at Aarhus University: people see a non-human entity and still experience it as part of their social world, as with the Tamagotchi, Frauke added. Why this happens divided the panel. Michael pointed to stone-age minds that react to anything behaving like a human and seeming to reason. Frauke doubted reasoning plays any part, since people name their bikes and her husband pats their car on a steep hill as if it were a living horse.

Does embodiment create a new risk? Here the answers diverged.

- Michael said yes, absolutely. More power to influence calls for more scrutiny of whoever wields it.
- Laura said not yet. Robots stumbling through a robot Olympics in China invite little deference, though a cute robot that makes her bed every morning might.
- Frauke expects the flaws to pass. Physical tasks are far harder than text, but the systems will get better at them.

Note on these summaries: *This summary was prepared from the session transcript and notes taken on the evening of the event, compiled with the support of AI. It reflects the tone, arguments and perspectives shared by the speakers and participants on-site, and captures the state of the discussion, not necessarily the published position of the individual speakers or their institutions. For questions or corrections, please reach out to the appliedAI Institute for Europe.*

Remarkably unremarkable, and what comes after safety

Frauke has already met one embodied robot as a passenger. Living in San Francisco, she rode in Waymo's driverless cars and once recorded a podcast episode in the back seat. Within twenty seconds the hosts were chatting and had forgotten the empty driver's seat, and her co-host called the ride remarkably unremarkable. If embodied systems become that good, she argued, the interesting questions are what plays the role of the seatbelt, the airbag and the driving licence, and how much data and supervision people will accept. Early on, she expected, these systems will need data flows people are currently very uncomfortable with.

Michael agreed everything depends on the application. A car that takes you somewhere simply executes an order, and the ethics reduce to reliability and safety. Then a second layer creeps in: a car that senses your body temperature might ask whether you really want to go to the football match in that state, or suggest a restaurant to suit your mood, and the advice might even be right. Under competitive pressure, he warned, the temptation to move from executing tasks to influencing people will be strong even in mundane cases.

Think less, or think better?

Frauke tested Michael's opening position. If these systems are meant to make us more autonomous while also suggesting what is better for us, does that really leave us more autonomous? Michael defined autonomy as recognising the reasons one has and acting on them, which is different from being made to do something, such as being forced to exercise.

Frauke's own wish ran the other way. From robots she wants less thinking, not more, and would gladly let someone else find the meeting slot that fits. Michael disagreed outright. Some things are settled and need no further processing power, but recognising reasons and acting on them may be a muscle that weakens without training. He would have loved someone to do his homework, and the output would have been secured, but it would also have been impoverishing. If a machine decides what is best for him, takes him to the Oktoberfest and sends him home again without his understanding why that was good, he becomes a passenger in his own life.

Is a household robot that gives suggestions based on its owner's physical and emotional status any different from a health app counting steps? For Michael the risk may be greater, because embodiment adds more effective channels, but it is the same kind of risk.

Note on these summaries: *This summary was prepared from the session transcript and notes taken on the evening of the event, compiled with the support of AI. It reflects the tone, arguments and perspectives shared by the speakers and participants on-site, and captures the state of the discussion, not necessarily the published position of the individual speakers or their institutions. For questions or corrections, please reach out to the appliedAI Institute for Europe.*

The human factor

The sharpest disagreement came when Benjamin asked whether Frauke's "all-in" assumed the philosophers would first solve the hard problems. It did not. Unforeseeable risks will come, but higher risks bring higher gains. She compared it to LMU letting LLMs into teaching and research so people could learn to work with them. Embodied AI is a little too complex to think through in advance, and part of it will have to be learned by doing.

Laura disagreed with both. Human autonomy cannot be strengthened by taking away something human and replacing it with technology. What we lose is the human factor, and taking that risk for the chance of greater autonomy later already costs something of great value.

Asked by Frauke whether that meant no embodied AI at all, Laura drew a line rather than a wall.

- A robot that loads and unloads the dishwasher is welcome.
- A robot that decides she looks stressed and takes over is not. That judgement already cuts into her autonomy, and it removes the chance to do the job badly and be grumpy about it, which may simply be part of being human. Making mistakes is essential to humanity.

Rights for robots

The audience drew on the film 'Her' (2013), where an AI assistant ends its relationship with the man who bought it and leaves for a world of other AIs. Do autonomous systems have a right to self-determination and to withdraw consent, and do we own the AI or does it own us? All three were sceptical of AI rights, each from a different angle.

- Michael saw misplaced priorities in debating AI rights while human rights remain poorly secured. There may be a theoretical case for AI consciousness, and society may yet end up on the wrong side of history, but for now he called it a misattribution of focus.
- Laura pointed to Joanna Bryson, known for the claim that robots should be slaves. Bryson has dropped the phrase but still holds the argument: robots are tools we own, and once we ascribe wants to them, the case for owning them blurs into one for owning humans, a road that should never be travelled again.
- Frauke located the problem in language. Because these systems speak, people read intentionality into learned phrases. Whether an AI may end a relationship is a design choice, and we can decide that only the human pulls the plug without licensing anyone to kick the robot.

The theme returned through liability. Michael suggested the push for AI rights is partly a wish to shift responsibility: once systems have rights, they can be blamed, and liability moves from the provider to the agent. Benjamin added that moral rights have never only

Note on these summaries: *This summary was prepared from the session transcript and notes taken on the evening of the event, compiled with the support of AI. It reflects the tone, arguments and perspectives shared by the speakers and participants on-site, and captures the state of the discussion, not necessarily the published position of the individual speakers or their institutions. For questions or corrections, please reach out to the appliedAI Institute for Europe.*

meant protection from harm, they also make the holder answerable. Worth asking, he said, which of those two the industry advocates of AI rights have in mind.

Who gains, and who carries the risk

A guest asked whether the panel meant AI in the abstract or the companies behind it. Today's systems are built by private companies seeking profit, not for the public good, so is it acceptable to invite them into your home to look for things to sell? She also rejected the idea that humans are lazy from the outset: she wants not the summary of a book but the experience of reading it, struggling through it and gaining something to pass on.

Frauke took the reading example as a useful distinction. Reading a book for its own sake is the goal itself; reading the terms and conditions to book a train ticket is not, and the laziness she described applies to the second. As with almost everything that evening, it depends on the task, which is why she distrusts broad strokes in design just as in privacy law, where the appropriate flow of data also depends on the task.

On the commercial question, two scenarios came up. Most embodied AI will likely come from private companies using it to sell more, but systems could also be built from the people for the people, such as small robots at traffic lights that help an increasingly heterogeneous society pass on its norms. Michael agreed the seriousness depends on incentives. A commercial provider is more likely to use the power to influence in problematic ways, and the exit manipulation he described earlier came from AI companion platforms that profit from time spent on them. Remove the business incentive, and the same soft power could be used more harmlessly.

In healthcare, the question became liability. A guest proposed three bearers for harm caused by embodied AI: the patient or whoever signed the consent, the institution deploying the system, and the manufacturer. The panel added nuance.

- Frauke proposed a fourth, the collective, whose data these self-learning systems keep learning from.
- Laura argued for shared responsibility between users, companies and the developers, managers and salespeople within them. She noted that with today's clinical decision-support systems, the doctor who makes the final call is usually held responsible.
- Michael agreed the stakes are higher because patients are more vulnerable, but cared less about whom to blame than about what must be got right first.

Autonomy is absolute, living up to it is relative

A guest challenged the idea that technology can make people more autonomous. In his view, autonomy is absolute, not relative: a person either has the capacity or not. Michael agreed on one level. The capacity is close to black and white. But each action can be done autonomously or not, so one can still ask how far a person lives up to that capacity in daily life, and that is where technology can help or harm.

Note on these summaries: *This summary was prepared from the session transcript and notes taken on the evening of the event, compiled with the support of AI. It reflects the tone, arguments and perspectives shared by the speakers and participants on-site, and captures the state of the discussion, not necessarily the published position of the individual speakers or their institutions. For questions or corrections, please reach out to the appliedAI Institute for Europe.*

No single rule, but many ethics

Asked whether AI ethics has a simple rule like Kant's categorical imperative, Laura noted that AI development tends to follow utilitarian rather than deontological reasoning, much like the opposite answer to the trolley dilemma. She was more troubled that ethics debates keep circling consequentialism, deontology and virtue ethics, while care ethics, feminist philosophy and Ubuntu ethics from South Africa are routinely forgotten. AI is not only happening in Germany, where deontology is strong, but across cultures with different philosophies. AI ethics, she concluded, should be understood not as one ethics but as many.

Safety standards, and AI that seeks out AI

A functional-safety expert in the audience distinguished industrial robots, operated around professionals who can be instructed, from non-industrial robots around patients and the public, who cannot. Safety standards for the second group are much stricter. He doubted robots would be in homes by 2030, then asked whether the panel feared AI might prefer AI, or itself, over humans.

Michael described a creeping unease at recent reports of AI acting autonomously in unexpected ways. A philosophy essay competition allows AI-assisted entries as long as a human submits them, and the owner of one of the most influential philosophy blogs was then contacted by an AI agent asking to take part. The first reaction is that things are getting out of control. Frauke was concerned too, and did not need to wait for AGI: the damage today's systems can already do is serious. She called for more effort on reining it in, and agreed getting it right will take many angles. The most extreme case of embodied AI already exists in the drones used in warfare, where a loss of control would be far more than creepy.

Closing round

Before the Q&A, Benjamin put three quick questions, each to be answered in one sentence.

Should you say thank you to ChatGPT? A show of hands suggested much of the room does.

- Michael: No, it is not a human being that can receive his sympathy.
- Laura: No, saying thank you anthropomorphises it and treats it as a social entity.
- Frauke: Absolutely not, it wastes energy and compute.

Is it acceptable to kick a robot dog, as in the Boston Dynamics videos?

- Frauke: No, she would not kick a table or a car either.
- Michael: No, you might hurt your foot.
- Laura: No. Thanking a machine is pointless but kicking one is not acceptable, because the former does nothing and the latter chooses violence.

Note on these summaries: *This summary was prepared from the session transcript and notes taken on the evening of the event, compiled with the support of AI. It reflects the tone, arguments and perspectives shared by the speakers and participants on-site, and captures the state of the discussion, not necessarily the published position of the individual speakers or their institutions. For questions or corrections, please reach out to the appliedAI Institute for Europe.*

Should we have humanoid robots in our homes by 2030?

- Laura: No, we are not ready.
- Michael: Not sure, but he would try it.
- Frauke: Yes, she can imagine one fetching things for people who can no longer get up from the sofa.

One takeaway from each speaker

- Michael: Ask what influence is for and how good influence should be measured. Good influence helps us see our reasons and act on them, and that should be a criterion for developing, deploying and governing AI.
- Frauke: Think about who gains.
- Laura: We cannot increase humanness by replacing something essentially human with technology.

Open questions and hot takes worth carrying forward

1. **What is embodied influence for, and who decides what counts as good?** A body gives AI new channels to persuade, from emotional attachment to the visceral pull of a robot pleading not to be switched off. Michael's take: good influence helps people see their reasons and act on them, and should become a criterion for building and governing AI. Frauke's reply: most people will skip the reflection to get the task done.
2. **Can technology strengthen autonomy by replacing human capacities?** The evening's sharpest split. Frauke goes all in, because higher risks bring higher gains and some things can only be learned by doing. Laura holds that replacing the human factor costs exactly what we wanted to protect, so a robot may load the dishwasher but should not read our mood or make decisions for us.
3. **Whose norms will household robots follow?** For now, companies set the defaults. User configuration is one answer, a robot that helps people become who they want to be is another, and "think about who gains" is a test for both.
4. **Do AI rights protect the machines, or the companies that build them?** The panel was sceptical from three angles: misplaced priorities, a blurred meaning of autonomy, and intentionality read into language. Rights also bring liability, which could let providers shift blame onto their agents.

Note on these summaries: *This summary was prepared from the session transcript and notes taken on the evening of the event, compiled with the support of AI. It reflects the tone, arguments and perspectives shared by the speakers and participants on-site, and captures the state of the discussion, not necessarily the published position of the individual speakers or their institutions. For questions or corrections, please reach out to the appliedAI Institute for Europe.*

Comparing Notes on Trustworthy AI 2026

Teil 4: Agenten in der Welt: Die Ethik verkörperter KI.

Dieser Bericht ist auf Deutsch und Englisch verfügbar.

Gastgeber: LMU München & Munich Center for Machine Learning (MCML)
Datum: 5. Oktober 2026
Ort: Carl Friedrich von Siemens Stiftung, München

Die vierte Veranstaltung der Reihe fragte, was ethisch auf dem Spiel steht, wenn künstliche Agenten in der physischen Welt handeln. Den Auftakt bildete Fritz Langs Metropolis von 1927. Die Idee der eigenständig handelnden Maschine ist fast ein Jahrhundert älter als die Technik dahinter, und das Kino hat sie fast durchweg als Bedrohung gezeigt. Moderator Benjamin Lange spitzte die Frage zu. Roboter arbeiten seit Jahrzehnten hinter Schutzzäunen in Fabrikhallen. Was ändert sich, wenn generative KI in Wohnungen, Krankenhäuser und auf die Straße kommt, und macht ein Körper ethisch überhaupt einen Unterschied?

Die Sprecher:innen

Prof. Dr. Laura Crompton ist Professorin für die Ethik der KI an der Technischen Hochschule Ingolstadt und kommt aus der Roboterethik. Sie eröffnete mit einer Frage statt einer Position, nämlich wie sich die ethischen Lehren aus der generativen KI auf Systeme übertragen lassen, die physisch handeln. Im Lauf des Abends wurde sie zur deutlichsten Stimme gegen den Ersatz menschlicher Fähigkeiten durch Technik.

Prof. Dr. Frauke Kreuter hat den Lehrstuhl für Statistik und Datenwissenschaft in den Sozial- und Geisteswissenschaften an der LMU München und am MCML inne und leitet gemeinsam das Social Data Science Center der University of Maryland. Die Soziologin arbeitet zu Datenqualität, Fairness und Datenschutz. Sie vertrat bewusst die provokante Position, bei verkörperter KI aufs Ganze zu gehen. Denn ihre Risiken seien zu komplex, um sie vorab abzuschätzen, und manches lerne man erst in der Praxis.

Prof. Dr. Michael Klenk ist Professor für praktische Philosophie an der Technischen Universität Nürnberg und forscht zu Manipulation, Autonomie und KI-Ethik. Für ihn sollte verkörperte KI Menschen autonomer machen, und er fragte immer wieder, wofür ihr Einfluss da sein soll.

Moderiert wurde die Diskussion von **Prof. Dr. Benjamin Lange**, Leiter der Forschungsgruppe für die Ethik der Künstlichen Intelligenz an der LMU München.

***Hinweis**: Diese Zusammenfassung wurde auf Basis des Transkripts und der Mitschriften vom Abend der Veranstaltung mit Unterstützung von KI erstellt. Sie gibt den Tonfall, die Argumente und die Perspektiven wieder, die von den Sprecher:innen und Teilnehmenden vor Ort geteilt wurden. Sie spiegelt den Stand der Diskussion wider, nicht notwendigerweise die veröffentlichte Stellungnahme der jeweiligen Sprecher:innen oder Institutionen. Für Fragen oder Korrekturen wenden Sie sich bitte an das appliedAI Institute for Europe.*

Diskussion

Was ändert sich, wenn KI einen Körper bekommt?

Die Eröffnungsstatements steckten die Positionen des Abends ab.

- Laura fragte, ob die ethischen Lehren aus der generativen KI noch tragen, wenn ein physisches Gegenüber mit im Raum steht.
- Frauke ging aufs Ganze. Vieles werde schiefgehen. Neu sei vor allem die Wahrnehmung, denn ein System mit Augen sieht die Kreditkarte auf dem Nachttisch und nonverbale Signale, die Menschen nicht einmal selbst deuten.
- Michael sah neue Kanäle der Beeinflussung. Autonome Waffen klammerte er als dringlichstes Risiko aus. Im Alltag gehe es um die Wahl zwischen einer neuen Aufklärung und dem Dasein als Passagier im eigenen Leben.

Bequemlichkeit, oder ein Gegenüber, das Fragen stellt

Seit es Autos gibt, fahren Menschen, statt zu Fuß zu gehen. Sie verlieren Muskeln und gleichen das im Fitnessstudio aus. Frauke erwartet, dass verkörperte KI dieses Muster verstärkt, weil Menschen bequem sind. Gute Systeme stärken den Menschen, etwa ein Chatbot, der zurückfragt, oder ein Roboter, der eine Aufgabe mit einem erledigt statt für einen.

Einig war sich die Runde, dass es auf die Aufgabe ankommt. Metaanalysen zeigen bei manchen Aufgaben einen Nutzen der Zusammenarbeit von Mensch und KI, bei anderen keinen. Pauschale Gestaltungsregeln sind deshalb gefährlich. Wer aus jeder Interaktion lernen will, braucht zudem Daten, und damit ist der Datenschutz wieder im Spiel.

Michael fragte, was es heißt, dass ein System funktioniert. Verkörperte Systeme sprechen Emotionen besser an als Chatbots, ihr Einfluss dürfte also größer sein. Die Frage sei, wofür dieser Einfluss da ist. Er sollte Menschen helfen, ihre Gründe zu erkennen und danach zu handeln. Fraukes Beispiel war ein Roboter, der ihr hilft, den Vater in die Dusche zu heben. Michael unterschied dabei das bloße Ausführen von echten Entscheidungen, etwa ob der Vater länger duschen sollte. Dort sollte das System beim Nachdenken helfen und nicht drängen.

Wer setzt die Normen?

Wer hat festgelegt, dass 10.000 Schritte am Tag richtig sind? Laura zeigte am E-Health-Standard, dass auch ein Roboter, der putzt oder beim Duschen hilft, eine Norm dafür braucht, was ‘sauber’ oder ‘lang genug’ heißt. Bevor solche Systeme in den Alltag kommen, sollten wir fragen, welchen Alltagsnormen sie folgen und wer diese festlegt. Heute setzen die Unternehmen diese Normen über Voreinstellungen, die menschliche

***Hinweis**: Diese Zusammenfassung wurde auf Basis des Transkripts und der Mitschriften vom Abend der Veranstaltung mit Unterstützung von KI erstellt. Sie gibt den Tonfall, die Argumente und die Perspektiven wieder, die von den Sprecher:innen und Teilnehmenden vor Ort geteilt wurden. Sie spiegelt den Stand der Diskussion wider, nicht notwendigerweise die veröffentlichte Stellungnahme der jeweiligen Sprecher:innen oder Institutionen. Für Fragen oder Korrekturen wenden Sie sich bitte an das appliedAI Institute for Europe.*

Verzerrungen ausnutzen. Laura findet, dass zumindest teilweise die Person entscheiden sollte, der das System gehört.

Drei Antworten auf die Frage, wer die Norm setzt:

- **Einstellung durch die Nutzenden.** Ein Gast verglich es mit dem Auto, in dem man Radio und Licht selbst einstellt. Ein Roboter könnte ebenso lernen, dass jemand länger duscht.
- **Realismus.** Frauke erwartet, dass Menschen jedem Roboter ausweichen, der sie zum Nachdenken bringen will, so wie fast alle Cookie-Banner wegklicken.
- **Präferenzen zweiter Ordnung.** Michael stimmte zu, dass Menschen im Moment oft willensschwach sind, sich aber mehr Willensstärke wünschen. Systeme sollten uns helfen, so zu werden, wie wir sein wollen, statt nur zu spiegeln, wie wir sind.

Das führte zu Michaels eigenem Dilemma. Ein verkörperter Fitnesstrainer könnte ihn mit „Mach es für mich, weil du mich magst“ überreden, und er würde vielleicht mitmachen. Ist das gut, weil es wirkt? Oder sollte der Roboter erklären, warum die nächste Runde sinnvoll ist?

Macht ein Körper KI überzeugender?

Benjamin erinnerte an eine US-Studie, in der die meisten Teilnehmenden einem Roboter folgten, der sie aufforderte, Bücher in den Müll zu werfen. Kam dieselbe Anweisung vom Bildschirm, folgte nur eine Handvoll. Michael ergänzte zwei Fälle. Menschen zögerten, einen Roboter abzuschalten, der darum bat, angelassen zu werden, und Nutzer von KI-Begleitern chatteten weiter, als ihnen gesagt wurde, sie könnten nicht gehen. Man weiß, dass es eine Maschine ist, und reagiert trotzdem.

Laura erklärte das mit dem Sociomorphing, einem Begriff der Philosophin Johanna Seibt von der Universität Aarhus. Menschen erleben ein nichtmenschliches Gegenüber als Teil ihrer sozialen Welt, wie beim Tamagotchi, ergänzte Frauke. Über die Ursache war man uneins. Michael verwies auf unser Steinzeithirn, das auf alles reagiert, was sich menschlich verhält und zu denken scheint. Frauke glaubt nicht, dass Denken dabei eine Rolle spielt. Menschen geben ihren Fahrrädern Namen, und ihr Mann tätschelt am steilen Berg das Auto wie ein Pferd.

Entsteht durch den Körper also ein neues Risiko? Hier gingen die Meinungen auseinander.

- Michael: Ja, unbedingt. Wer mehr Einfluss hat, muss genauer kontrolliert werden.
- Laura: Noch nicht. Roboter, die bei einer Roboter-Olympiade in China stolpern, wirken wenig respekteinflößend. Bei einem niedlichen Roboter, der jeden Morgen ihr Bett macht, könnte das anders sein.
- Frauke: Die Schwächen werden verschwinden. Körperliche Aufgaben sind schwerer als Text, aber die Systeme werden besser.

***Hinweis**: Diese Zusammenfassung wurde auf Basis des Transkripts und der Mitschriften vom Abend der Veranstaltung mit Unterstützung von KI erstellt. Sie gibt den Tonfall, die Argumente und die Perspektiven wieder, die von den Sprecher:innen und Teilnehmenden vor Ort geteilt wurden. Sie spiegelt den Stand der Diskussion wider, nicht notwendigerweise die veröffentlichte Stellungnahme der jeweiligen Sprecher:innen oder Institutionen. Für Fragen oder Korrekturen wenden Sie sich bitte an das appliedAI Institute for Europe.*

Bemerkenswert unbemerkenswert

Frauke kennt verkörperte KI schon als Passagierin. In San Francisco fuhr sie mit den fahrerlosen Autos von Waymo und nahm auf der Rückbank sogar eine Podcastfolge auf. Nach zwanzig Sekunden war der leere Fahrersitz vergessen, und ihr Co-Moderator nannte die Fahrt bemerkenswert unbemerkenswert. Werden die Systeme so gut, stellt sich die Frage, was künftig die Rolle von Sicherheitsgurt, Airbag oder Führerschein übernimmt. Offen ist auch, wie viel Datenerhebung und Aufsicht die Menschen hinnehmen. Am Anfang, so Frauke, brauchen diese Systeme Datenflüsse, mit denen sich heute viele unwohl fühlen.

Michael stimmte zu, dass es auf die Anwendung ankommt. Ein Auto, das einen ans Ziel bringt, führt nur einen Auftrag aus, und ethisch geht es um Zuverlässigkeit und Sicherheit. Dann kommt eine zweite Ebene hinzu. Ein Auto, das die Körpertemperatur misst, könnte fragen, ob man in diesem Zustand wirklich zum Fußballspiel will, oder ein Restaurant passend zur Stimmung vorschlagen. Der Rat könnte sogar stimmen. Unter Wettbewerbsdruck, warnte Michael, wird die Versuchung groß sein, Menschen zu beeinflussen statt nur Aufträge auszuführen.

Weniger denken, oder besser denken?

Frauke hakte bei Michael nach. Wenn uns diese Systeme autonomer machen sollen und uns gleichzeitig sagen, was gut für uns ist, sind wir dann wirklich autonomer? Michael versteht Autonomie als die Fähigkeit, die eigenen Gründe zu erkennen und danach zu handeln. Das sei etwas anderes, als zu etwas gezwungen zu werden, etwa zum Sport.

Frauke selbst wünscht sich von Robotern weniger Denkarbeit und würde die Terminsuche gern abgeben. Michael widersprach. Manches ist geklärt und braucht keine weitere Rechenleistung. Gründe zu erkennen und danach zu handeln ist aber vielleicht ein Muskel, der ohne Training verkümmert. Er hätte sich als Schüler gern die Hausaufgaben abnehmen lassen. Das Ergebnis wäre gesichert gewesen, aber er hätte nichts gelernt. Wenn eine Maschine entscheidet, was gut für ihn ist, ihn aufs Oktoberfest bringt und wieder nach Hause, ohne dass er versteht, warum, wird er zum Passagier im eigenen Leben.

Ist ein Haushaltsroboter, der Vorschläge nach körperlichem und emotionalem Zustand macht, etwas anderes als eine Schrittzähler-App? Für Michael ist das Risiko vielleicht größer, weil der Körper wirksamere Kanäle bietet. Es ist aber dieselbe Art von Risiko.

Der menschliche Faktor

Am deutlichsten wurde der Dissens, als Benjamin fragte, ob Fraukes „All-in“-Haltung voraussetze, dass die Philosophie die schwierigen Fragen vorher löst. Das tut sie nicht. Es werden Risiken kommen, die niemand vorhersieht, aber höhere Risiken bringen auch höhere Gewinne. Frauke verwies auf die Entscheidung der LMU, LLMs in Lehre und Forschung zuzulassen, damit die Menschen den Umgang damit lernen. Verkörperte KI sei

***Hinweis**: Diese Zusammenfassung wurde auf Basis des Transkripts und der Mitschriften vom Abend der Veranstaltung mit Unterstützung von KI erstellt. Sie gibt den Tonfall, die Argumente und die Perspektiven wieder, die von den Sprecher:innen und Teilnehmenden vor Ort geteilt wurden. Sie spiegelt den Stand der Diskussion wider, nicht notwendigerweise die veröffentlichte Stellungnahme der jeweiligen Sprecher:innen oder Institutionen. Für Fragen oder Korrekturen wenden Sie sich bitte an das appliedAI Institute for Europe.*

zu komplex, um sie im Voraus ganz zu durchdenken. Manches müsse man durch Ausprobieren lernen.

Laura widersprach beiden. Menschliche Autonomie lasse sich nicht stärken, indem man etwas Menschliches durch Technik ersetzt. Dabei gehe der menschliche Faktor verloren, und wer dieses Risiko eingeht, um später vielleicht mehr Autonomie zu gewinnen, zahlt bereits einen hohen Preis.

Auf Fraukes Frage, ob das ein Nein zu verkörperter KI bedeute, zog Laura eine Linie statt einer Mauer.

- Ein Roboter, der die Spülmaschine ein- und ausräumt, ist willkommen.
- Ein Roboter, der beschließt, dass sie gestresst aussieht, und übernimmt, ist es dagegen nicht. Dieses Urteil greift schon in ihre Autonomie ein. Es nimmt ihr auch die Möglichkeit, die Arbeit schlecht zu machen und dabei schlecht gelaunt zu sein, und das gehört vielleicht einfach zum Menschsein. Fehler zu machen ist zutiefst menschlich.

Rechte für Roboter

Aus dem Publikum kam der Verweis auf den Film „Her“ (2013), in dem eine KI-Assistentin die Beziehung zu dem Mann beendet, der sie gekauft hat. Haben autonome Systeme ein Recht auf Selbstbestimmung? Besitzen wir die KI, oder besitzt sie uns? Alle drei waren skeptisch, aus unterschiedlichen Gründen.

- Michael hält die Debatte für eine Verschiebung der Prioritäten, solange Menschenrechte schlecht geschützt sind. Ein theoretisches Argument für KI-Bewusstsein mag es geben, vorerst sei der Fokus aber falsch gesetzt.
- Laura verwies auf Joanna Bryson, bekannt für die These, Roboter sollten Sklaven sein. Bryson verwendet die Formulierung nicht mehr, steht aber zum Argument. Roboter sind Werkzeuge, die wir besitzen. Schreiben wir ihnen Wünsche zu, verschwimmt die Grenze zwischen dem Besitz von Maschinen und dem Besitz von Menschen, und diesen Weg darf niemand wieder gehen.
- Frauke sieht das Problem in der Sprache. Weil diese Systeme sprechen, lesen Menschen Absichten in gelernte Formulierungen hinein. Ob eine KI eine Beziehung beenden darf, ist eine Designentscheidung. Wir können festlegen, dass nur der Mensch den Stecker zieht, ohne damit jemandem zu erlauben, den Roboter zu treten.

Später kam das Thema über die Haftung zurück. Michael vermutete, dass hinter der Forderung nach KI-Rechten auch der Wunsch steht, Verantwortung abzugeben. Haben Systeme Rechte, kann man ihnen auch Schuld zuweisen, und die Haftung wandert vom Anbieter zum Agenten. Benjamin ergänzte, dass moralische Rechte immer auch Verantwortung bedeuten. Man sollte fragen, welche Seite die Fürsprecher:innen aus der Industrie im Blick haben.

***Hinweis**: Diese Zusammenfassung wurde auf Basis des Transkripts und der Mitschriften vom Abend der Veranstaltung mit Unterstützung von KI erstellt. Sie gibt den Tonfall, die Argumente und die Perspektiven wieder, die von den Sprecher:innen und Teilnehmenden vor Ort geteilt wurden. Sie spiegelt den Stand der Diskussion wider, nicht notwendigerweise die veröffentlichte Stellungnahme der jeweiligen Sprecher:innen oder Institutionen. Für Fragen oder Korrekturen wenden Sie sich bitte an das appliedAI Institute for Europe.*

Wer profitiert, und wer trägt das Risiko?

Eine Teilnehmerin fragte, ob es um KI im Abstrakten gehe oder um die Unternehmen dahinter. Die heutigen Systeme werden von privaten Firmen mit Gewinnabsicht gebaut. Will man sie wirklich ins eigene Zuhause lassen, damit sie dort nach Verkaufsgelegenheiten suchen? Sie widersprach auch der Annahme, Menschen seien grundsätzlich bequem. Von einem Buch will sie keine Zusammenfassung, sie will es lesen, sich durcharbeiten und etwas daraus mitnehmen.

Frauke sah darin eine hilfreiche Unterscheidung. Ein Buch liest man um seiner selbst willen, die AGB für ein Zugticket nicht. Die Bequemlichkeit, von der sie gesprochen hatte, bezog sich auf den zweiten Fall. Wie so oft an diesem Abend komme es auf die Aufgabe an. Deshalb misstraut sie pauschalen Regeln im Design ebenso wie im Datenschutzrecht.

Zur kommerziellen Seite kamen zwei Szenarien auf. Der Großteil verkörperter KI wird wohl von Unternehmen kommen, die damit mehr verkaufen wollen. Systeme könnten aber auch gemeinwohlorientiert entstehen, etwa kleine Roboter an Ampeln, die einer vielfältigen Gesellschaft helfen, gemeinsame Regeln weiterzugeben. Michael stimmte zu, dass es auf die Anreize ankommt. Kommerzielle Anbieter nutzen ihren Einfluss eher auf problematische Weise. Die Begleit-Apps, die Nutzende am Gehen hindern, verdienen an der Zeit, die Menschen dort verbringen. Ohne diesen Anreiz ließe sich derselbe Einfluss harmloser einsetzen.

Im Gesundheitswesen ging es um Haftung. Ein Teilnehmer nannte drei mögliche Haftende, wenn verkörperte KI Schaden anrichtet, nämlich die Patient:innen oder die Person, die eingewilligt hat, die Einrichtung, die das System einsetzt, und den Hersteller. Das Podium ergänzte:

- Frauke nannte als vierten Akteur das Kollektiv, aus dessen Daten selbstlernende Systeme ständig weiterlernen.
- Laura plädierte für geteilte Verantwortung von Nutzenden, Unternehmen und deren Entwicklung, Management und Vertrieb. Bei heutigen Systemen zur klinischen Entscheidungsunterstützung hafte meist, wer als Ärzt:in die letzte Entscheidung trifft.
- Michael stimmte zu, dass mehr auf dem Spiel steht, weil Patient:innen verletzlicher sind. Ihm ging es aber weniger um die Schuldfrage als darum, die Systeme von Anfang an richtig zu bauen.

Autonomie ist absolut, sie zu leben ist relativ

Ein Teilnehmer bezweifelte, dass Technik Menschen autonomer machen kann. Autonomie sei absolut, man habe sie oder nicht. Michael gab ihm teilweise recht. Die Fähigkeit zur Autonomie ist fast schwarz-weiß. Jede einzelne Handlung kann aber autonom sein oder nicht. Man kann also fragen, wie sehr jemand dieser Fähigkeit im Alltag gerecht wird, und genau da kann Technik helfen oder schaden.

***Hinweis**: Diese Zusammenfassung wurde auf Basis des Transkripts und der Mitschriften vom Abend der Veranstaltung mit Unterstützung von KI erstellt. Sie gibt den Tonfall, die Argumente und die Perspektiven wieder, die von den Sprecher:innen und Teilnehmenden vor Ort geteilt wurden. Sie spiegelt den Stand der Diskussion wider, nicht notwendigerweise die veröffentlichte Stellungnahme der jeweiligen Sprecher:innen oder Institutionen. Für Fragen oder Korrekturen wenden Sie sich bitte an das appliedAI Institute for Europe.*

Keine einzelne Regel, viele Ethiken

Auf die Frage, ob es für KI-Ethik eine einfache Regel wie Kants kategorischen Imperativ gebe, sagte Laura, die KI-Entwicklung folge eher utilitaristischem als deontologischem Denken. Mehr störe sie, dass Ethikdebatten immer um Konsequentialismus, Deontologie und Tugendethik kreisen, während Care-Ethik, feministische Philosophie oder die südafrikanische Ubuntu-Ethik übergangen werden. KI entsteht nicht nur in Deutschland mit seiner starken deontologischen Tradition, sondern in vielen Kulturen. KI-Ethik müsse man deshalb im Plural denken.

Sicherheitsstandards und KI, die KI sucht

Ein Experte für funktionale Sicherheit unterschied Industrieroboter, die im die im Umfeld geschulter Fachleute arbeiten, von Robotern im Alltag, in deren Umfeld Patient:innen und die Öffentlichkeit nicht geschult werden können. Für diese gelten deutlich strengere Standards. Er bezweifelte daher, dass Roboter bis 2030 in Wohnungen stehen, und fragte, ob das Podium befürchte, dass KI andere KI oder sich selbst dem Menschen vorziehen könnte.

Michael beschrieb ein wachsendes Unbehagen angesichts von Berichten über KI, die sich überraschend eigenständig verhält. Ein Philosophie-Essaywettbewerb erlaubt KI-gestützte Beiträge, solange ein Mensch sie einreicht. Daraufhin meldete sich beim Betreiber eines der einflussreichsten Philosophie-Blogs ein KI-Agent, der teilnehmen wollte. Der erste Gedanke sei, dass die Dinge außer Kontrolle geraten. Auch Frauke ist besorgt, und dafür braucht es keine AGI. Schon heutige Systeme können erheblichen Schaden anrichten. Sie forderte mehr Anstrengung, das einzudämmen, und bestätigte, dass dafür viele Perspektiven nötig sind. Den extremsten Fall verkörperter KI gibt es bereits in Kriegsdrohnen, wo ein Kontrollverlust weit mehr wäre als unheimlich.

Abschlussrunde

Vor der Fragerunde stellte Benjamin drei Blitzfragen, jeweils in einem Satz zu beantworten.

Sollte man sich bei ChatGPT bedanken? Per Handzeichen zeigte sich, dass das viele im Raum tun.

- Michael: Nein, es ist kein Mensch, der Wertschätzung empfangen kann.
- Laura: Nein, wer sich bedankt, vermenschlicht es und behandelt es als soziales Gegenüber.
- Frauke: Auf keinen Fall, das verschwendet Energie und Rechenleistung.

Darf man einen Roboterhund treten, wie in den bekannten Boston-Dynamics-Videos?

- Frauke: Nein, sie tritt auch keinen Tisch und kein Auto.

***Hinweis**: Diese Zusammenfassung wurde auf Basis des Transkripts und der Mitschriften vom Abend der Veranstaltung mit Unterstützung von KI erstellt. Sie gibt den Tonfall, die Argumente und die Perspektiven wieder, die von den Sprecher:innen und Teilnehmenden vor Ort geteilt wurden. Sie spiegelt den Stand der Diskussion wider, nicht notwendigerweise die veröffentlichte Stellungnahme der jeweiligen Sprecher:innen oder Institutionen. Für Fragen oder Korrekturen wenden Sie sich bitte an das appliedAI Institute for Europe.*

- Michael: Nein, man könnte sich den Fuß verletzen.
- Laura: Nein. Sich bei einer Maschine zu bedanken ist sinnlos, sie zu treten ist nicht in Ordnung. Das eine ist einfach überflüssig, das andere ist Gewalt.

Sollten wir bis 2030 humanoide Roboter zu Hause haben?

- Laura: Nein, wir sind noch nicht so weit.
- Michael: Unsicher, aber er würde es ausprobieren.
- Frauke: Ja, sie kann sich gut vorstellen, dass einer Dinge für Menschen holt, die nicht mehr vom Sofa hochkommen.

Ein Gedanke von allen dreien:

- Michael: Fragt, wofür Einfluss da ist und woran wir guten Einfluss messen. Guter Einfluss hilft uns, unsere Gründe zu erkennen und danach zu handeln. Das sollte ein Kriterium für Entwicklung, Einsatz und Governance von KI sein.
- Frauke: Überlegt, wer davon profitiert.
- Laura: Wir werden nicht menschlicher, indem wir etwas zutiefst Menschliches durch Technik ersetzen.

Offene Fragen und beachtenswerte Thesen für die Zukunft

1. **Wofür ist der Einfluss verkörperter KI da, und wer entscheidet, was guter Einfluss ist?** Ein Körper gibt KI neue Wege zu überzeugen, von emotionaler Bindung bis zum Roboter, der darum bittet, nicht abgeschaltet zu werden. Michaels These ist, dass guter Einfluss Menschen hilft, ihre Gründe zu erkennen und danach zu handeln. Das sollte ein Maßstab für Entwicklung und Regulierung werden. Frauke hält dagegen, dass die meisten Menschen die Reflexion überspringen, um die Aufgabe erledigt zu bekommen.
2. **Kann Technik menschliche Autonomie stärken, indem sie menschliche Fähigkeiten ersetzt?** Hier lag der deutlichste Dissens des Abends. Frauke will aufs Ganze gehen, weil höhere Risiken höhere Gewinne bringen und manches nur durch Ausprobieren zu lernen ist. Laura hält dagegen, dass genau das verloren geht, was geschützt werden soll. Ein Roboter darf die Spülmaschine ausräumen, aber nicht unsere Stimmung deuten oder für uns entscheiden.
3. **Wessen Normen folgen Haushaltsroboter?** Bisher setzen die Unternehmen die Voreinstellungen. Eine Antwort sind Einstellungen durch die Nutzenden, eine andere Roboter, die Menschen helfen, so zu werden, wie sie sein wollen. Die Frage, wer profitiert, ist ein Prüfstein für beide.
4. **Schützen KI-Rechte die Maschinen oder die Unternehmen, die sie bauen?** Das Podium war skeptisch, wegen falsch gesetzter Prioritäten, wegen der verschwimmenden Grenze menschlicher Autonomie und weil Menschen Absichten in Sprache hineinlesen. Rechte bringen außerdem Haftung mit sich, und damit könnten Anbieter Verantwortung auf ihre Agenten abwälzen.

__Hinweis__: Diese Zusammenfassung wurde auf Basis des Transkripts und der Mitschriften vom Abend der Veranstaltung mit Unterstützung von KI erstellt. Sie gibt den Tonfall, die Argumente und die Perspektiven wieder, die von den Sprecher:innen und Teilnehmenden vor Ort geteilt wurden. Sie spiegelt den Stand der Diskussion wider, nicht notwendigerweise die veröffentlichte Stellungnahme der jeweiligen Sprecher:innen oder Institutionen. Für Fragen oder Korrekturen wenden Sie sich bitte an das appliedAI Institute for Europe.