

Comparing Notes on Trustworthy AI 2026

Session 3: Agents, Physical AI, “RentAHuman”: From LLMs to “Meatspace” Workers.

German version below

Host: Liquid Legal Institute

Date: 30 July 2026

Location: TUM Think Tank, Munich

The third event of the 2026 series moved from physical AI's hardware and demos to the software and social impacts, discussing what happens to trust, work and accountability once agentic systems start acting on our behalf. The evening opened with an unfamiliar name for most of the room, "RentAHuman", a platform where people register to be booked for real world tasks, even by AI agents. The panel closed on the same idea from the other direction, depicting a future where AI stops being a tool and starts orchestrating work in the physical world.

The speakers

Eva Simone Lihotzky recently founded Radiant, focused on Responsible AI, Trust and Ethics, after years of specialising in assurance and quality assessment and leading an AI lab she co-founded. She also hosts one of Germany's top three tech podcasts. On the panel she pushed the conversation toward organisational responsibility and the fear employees feel once agents enter the workplace, and argued that companies should upskill people rather than replace them.

Kai Hermsen co-founded DECAID, which brings AI governance to small and medium sized businesses, after working in cybersecurity and tech-human-interaction at Siemens, where he helped launch the “Charter of Trust” initiative. He argued that the real blind spot in AI adoption sits in the output, not the input, and that trust should be the explicit and ultimate outcome.

Prof. Dr. Johann Kranz researches sociotechnical systems, AI in organisations and the circular economy at LMU Munich. He introduced the arc from systems of record to systems of insight to today's systems of action, and returned throughout the evening to the risks and opportunities that come with physical AI and the labor market.

The discussion was moderated by **Dr. Bernhard Walzl**. He is co-founder and co-CEO of the Liquid Legal Institute and Head of Legal Innovation and Operations at BMW Group.

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

From systems of record to systems of action

Organisations have moved through three stages on the way to today's AI systems:

- **Systems of record**, the classic mainframes and IT systems.
- **Systems of insight**, chatbots and early AI layered on top.
- **Systems of action**, today's agentic systems that actually do things rather than just report or suggest.

Johann argued that this shift changes more than individual companies, it reaches into the entire economy once physical AI is added to the mix, raising the prospect of lights-out factories with no human staff on the shop floor at all, a scenario several industry watchers expect within a few decades.

The blind spot in output, not input

The real blind spot in AI adoption isn't in the input, it's in the output. Organisations have become fairly good at managing input risk, this is the GDPR and data privacy questions that companies have spent years learning to handle, and they understand process risk reasonably well. What they're not good at is output risk. AI systems produce results that are non-deterministic and occasionally hallucinated, and unlike broken software that simply fails to run, a flawed AI output often looks entirely convincing, "not wrong, but also not right enough", as Kai put it.

Layer agentic systems on top of that and two further problems appear:

- **Audit trails disappear** once a result passes from agent to agent, so nobody can trace back what went wrong or where.
- There's **no equivalent yet of human style authentication** for individual agents, which means no clean way to know which agent did what, or on whose authority.

What counts as an agent, and the trust equation

The panel spent some time on a working definition of the word "agent", and this wasn't the community's first pass at the question. The 2025 edition of Comparing Notes series was built around AI agents, running from basic definitions and where they create real business value through governance, legal liability, and the calibrated trust needed to keep humans meaningfully in control. This session picked up that thread from the other side, less concerned with how to deploy an agent responsibly and more with what happens once an agent is capable enough to do some of the deploying itself.

An agent is something that acts on someone else's behalf with a degree of autonomy, true of institutions and individuals as much as of software, according to Eva's definition. With AI agents, it becomes an underlying model behaving like someone, carrying out a task you have handed over along with the risk that comes with it. Bernhard's own experience from the legal department tells him that AI started as a research tool, but he found himself running out of arguments for that framing once he saw AI developing strategies and giving recommendations rather than just fetching information.

Eva also introduced the evening's central concept, a **trust equation** David Maister had first introduced in 2000 in his book *"The Trusted Advisor"*:

Trust = Credibility + Reliability + Empathy, divided by Self-Orientation

Kai is also a believer in the equation that ties back to the birth moment of his Charter of Trust initiative, and agreed with Eva that AI can imitate all three of the first three qualities. What it cannot do is self-orient the way a person does, and that is precisely why AI itself can never really be described as trustworthy in the full sense. What can be trustworthy, or not, is the company standing behind the model, since a model's incentives ultimately trace back to an organisation with its own interests. The question isn't whether we can trust the output or not, but whether we can trust the source producing it.

Shadow AI and who really wants people's data

Bernhard posed a scenario to the room.

"What if an AI had access to everything a company knows, plus everything publicly available, and simply told you the best decision to make?"

Johann was blunt that this will not happen soon in practice, large regulated companies will not grant that kind of access, and pointed to an example of an OEM whose first instinct with LLMs was to lock employees out entirely rather than explore where the tool might actually help. Meanwhile, many employees already use personal LLM apps for work tasks on their own initiative. This points to a grey area called shadow AI.

Two dimensions of the issue must be addressed, according to Kai.

- On the **technical and governance** side, more data is not automatically better data, which is the same lesson the industry already learned from the big data era. The real work is orchestration around the right data rather than simply pouring in more of it.
- On the **trust** side, he returned to the aforementioned equation with a pointed addition. LLMs themselves may not be self-oriented, but the companies that build them are, to varying degrees, and that is a legitimate reason for an organisation to hesitate before handing over its data.

Agents at work, who is liable?

Companies face a real choice once employees start noticing what agents can do. Some upskill deliberately, Eva cited a CEO who gives every employee room to grow specifically to retain them, while many others have no answer and simply hope the government or the tech industry will solve the problem for them. Once an employee realises an agent can do their job faster, cheaper, without complaint or fatigue, the dominant reaction across the room is fear. Whether responsibility for managing the fear falls to employers, to the tech industry, or to policy, remains an open question.

The conversation then moved to something sharper that explores the possibility of an agent becoming the manager. Despite AI's many advantages in work performance and simulation of human empathy, AI cannot yet take on responsibility in any legal or accountability sense, and that, more than pure orchestration skill, is what justifies what managers are paid for.

“Orchestration itself is not the hard part of a management job, taking responsibility is,” quoted from Kai. He also named a current example of a startup building insurance specifically for AI agents, charging a premium per agent to cover residual risk once the technical governance is in place. Eva added that accountability inside organisations was already diffuse before agents arrived, and agents simply make that diffusion harder to ignore.

“RentAHuman” and the next platform economy

This brought the evening back to its opening hook. RentAHuman lets people register to be booked for real world tasks. The controversy surrounds scenarios where the booking party is an AI agent rather than another human, effectively a credit card handed to an agent that then pays a person to do something physical. The panel drew a direct line to existing platform economies like Uber, which created employment but also brought externalities and precarious conditions.

Johann offered a concrete physical AI example that sat somewhere between funny and uncomfortable. This featured a video footage of a human ironing clothes, recorded specifically to train a robot that will eventually take over the same task. Nobody on the panel, and no political party currently, seems to have a clear answer to what happens next for the displaced workers. The panel's shared conclusion was less a prediction than an admission, reiterating that this kind of arrangement could well be normal in five years.

Jobs, regulation and why choice still matters

On the labor market, Johann noted the research has mixed findings. Some studies point to job losses, others show that companies increase investment in AI, which creates more entry level roles rather than fewer. Cutting entry level jobs would be short-sighted, he argued, given the “**augmentation versus automation paradox.**” Without newcomers growing into a field, there is nobody left to augment, resulting eventually in the expertise disappearing.

On regulation, the panel's general view was that companies cannot be expected to self-regulate toward the public interest purely out of good will. Since that's simply not what they're individually incentivized to do, politics and regulation still have a role to play in setting the frame. The EU AI Act was discussed as a fundamentally different strategy from an unrestrained race, and the room saw a real opportunity to turn a "made in Germany" quality standard into a competitive advantage rather than treating regulation purely as a brake.

On the skepticism, whether any profit-driven company's AI product can truly be trusted, two views emerged. One view held that this is desirable but currently idealistic, pointing to social media's drift away from good intentions during the long absence of regulation. The counterpoint was that trust is a spectrum that depends on use cases rather than a binary question. So a risk based approach of choosing the model, the data access, and the degree of human oversight case by case does more good than a blanket judgement. Preserving diversity among LLM providers is also key according to Kai, so nobody is stuck with limited choices between several dominant platforms, the way social media narrowed down the options.

Closing round

An interactive exercise in the end asked the speakers individually to complete the sentence "successful collaboration between humans and AI in 2035 will be defined by..."

- **Johann:** Human acceptance and AI capability.
- **Eva:** Mutual understanding.
- **Kai:** Governance built on the realisation that we're better off together than apart.

Asked for one piece of advice for a CEO or policymaker heading into an agent-driven future, the panel set a shared theme from three angles.

- **Kai:** Try the technology out before forming an opinion on it, and look honestly at the risks alongside the opportunities.
- **Eva:** Understand your own organisation before reaching for new technology.
- **Johann:** Get outside your usual circle to hear different perspectives before deciding on anything. Attending panel discussions like these and getting informed about current topics could also be a start.

Open questions and hot takes worth carrying forward

1. **Where does accountability actually land once a result passes from agent to agent?** The panel agreed this is already diffusing inside organisations without AI, agents just make the gap harder to ignore.
 2. **Can AI itself ever be trustworthy, or only the organisation behind it?** If trust depends on self-orientation, and only companies, not models, have that in any meaningful sense, then every question about trusting an AI product is really a question about trusting a provider.
 3. **Are concepts like "RentAHuman" genuine previews of platform economy 2.0, and who absorbs the externalities?** Existing platform economies already show a pattern worth watching closely, as AI agents become the ones doing the booking.
 4. **Does cutting entry level jobs today cost us the expertise pipeline of tomorrow?** The augmentation versus automation paradox suggests yes, since without newcomers growing into a field, there is nobody left to augment.
-

Comparing Notes on Trustworthy AI 2026

Teil 3: Agenten, physische KI und „RentAHuman“. Von LLMs zu Arbeitskräften im „Meatspace.“

Dieser Bericht ist auf Deutsch und Englisch verfügbar.

Gastgeber: Liquid Legal Institute

Datum: 30. Juli 2026

Ort: TUM Think Tank, München

Die dritte Veranstaltung der Reihe 2026 wechselte von der Hardware und den Demos der physischen KI zu ihren Auswirkungen auf Software und Gesellschaft. Im Mittelpunkt stand die Frage, was aus Vertrauen, Arbeit und Verantwortlichkeit wird, sobald agentische Systeme in unserem Namen handeln. Der Abend begann mit einem für die meisten im Raum unbekannten Namen, „RentAHuman“, einer Plattform, auf der man sich für reale Aufgaben buchen lassen kann, mittlerweile auch von KI-Agenten. Das Panel schloss den Kreis von der anderen Seite und zeichnete das Bild einer Zukunft, in der KI nicht mehr nur Werkzeug ist, sondern Arbeit in der physischen Welt orchestriert.

Die Sprecher:innen

Eva Simone Lihotzky hat vor Kurzem Radiant gegründet, mit Fokus auf Responsible AI, Trust und Ethik. Sie verbrachte Jahre in Assurance und Qualitätssicherung und leitete zuletzt ein KI-Lab, das sie selbst mitgegründet hatte. Sie hostet außerdem einen der top drei Tech-Podcasts Deutschlands. Auf dem Panel lenkte sie das Gespräch auf organisatorische Verantwortung und die Angst, die Mitarbeitende empfinden, sobald Agenten Einzug in den Arbeitsalltag halten. Sie plädierte dafür, dass Unternehmen ihre Belegschaft weiterbilden statt ersetzen sollten.

Kai Hermesen hat DECAID mitgegründet, das KI-Governance für kleine und mittlere Unternehmen anbietet, nach Jahren in der Cybersicherheit und Mensch-Technik-Interaktion bei Siemens, wo er die „Charter of Trust“ Initiative mitbegründete. Er vertrat die These, dass der eigentliche blinde Fleck bei der KI-Einführung nicht im Input, sondern im Output liegt, und dass Vertrauen das explizite und oberste Ziel sein muss.

Prof. Dr. Johann Kranz forscht an der LMU München zu sozio-technischen Systemen, KI in Organisationen und der Kreislaufwirtschaft. Er führte die Entwicklungslinie von Systems of Record über Systems of Insight zu den heutigen Systems of Action ein und kam im Laufe des Abends immer wieder auf die Chancen und Risiken physischer KI und ihre Folgen für den Arbeitsmarkt zurück.

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.

Die Diskussion moderierte **Dr. Bernhard Walzl**. Er ist Mitgründer und Co-CEO des Liquid Legal Institute und Head of Legal Innovation and Operations bei der BMW Group.

Diskussion

Von Systems of Record zu Systems of Action

Organisationen haben auf dem Weg zu den heutigen KI-Systemen drei Phasen durchlaufen:

- **Systems of Record**, die klassischen Mainframes und IT-Systeme.
- **Systems of Insight**, Chatbots und frühe KI, die darauf aufgesetzt wurden.
- **Systems of Action**, die heutigen agentischen Systeme, die tatsächlich handeln, statt nur zu berichten oder Vorschläge zu machen.

Johann argumentierte, dass dieser Wandel weit über einzelne Unternehmen hinausreicht. Sobald physische KI hinzukommt, erfasst er die gesamte Wirtschaft. Er verwies auf die Aussicht auf Lights-out-Fabriken ohne jegliches menschliches Personal auf dem Shopfloor, das viele innerhalb der nächsten Jahrzehnte erwarten.

Der blinde Fleck liegt im Output, nicht im Input

Der Schwachpunkt bei der KI-Einführung liegt nicht im Input, sondern im Output. Organisationen sind inzwischen recht gut darin, Input-Risiken zu vermindern. Das sind Fragen zu DSGVO und Datenschutz, mit denen sich Unternehmen seit Jahren beschäftigen. Auch Prozessrisiken verstehen sie mittlerweile gut. Worin sie schlecht sind, ist der Umgang mit Output-Risiken. KI-Systeme liefern Ergebnisse, die nicht deterministisch und teilweise halluziniert sind. Anders als fehlerhafte Software, die einfach nicht läuft, wirkt ein fehlerhafter KI-Output oft völlig überzeugend. „Nicht falsch, aber auch nicht richtig genug“, wie Kai es formulierte.

Kommen agentische Systeme hinzu, entstehen zwei weitere Probleme:

- **Audit-Trails verschwinden**, sobald ein Ergebnis von einem Agenten zum nächsten weitergegeben wird. Niemand kann dann zurückverfolgen, was schiefgelaufen ist oder wo.
- **Es gibt noch kein Äquivalent zur menschlichen Authentifizierung** für einzelne Agenten. Das bedeutet, es gibt keine klare Möglichkeit zu wissen, welcher Agent was getan hat oder in wessen Auftrag.

Was einen Agenten ausmacht, und die Trust-Gleichung

Die Runde nahm sich Zeit für eine Arbeitsdefinition des Begriffs „Agent“, und das war nicht der erste Anlauf der Community zu dieser Frage. Comparing Notes drehte sich 2025 komplett um KI-Agenten, von grundlegenden Definitionen und der Frage, wo sie echten Geschäftswert schaffen, über Governance und rechtliche Haftung bis hin zu dem kalibrierten Vertrauen, das nötig ist, damit Menschen die Kontrolle behalten. Diese Session griff den Faden von der anderen Seite auf. Es ging weniger darum, wie man einen Agenten verantwortungsvoll einsetzt, sondern darum, was passiert, sobald ein Agent fähig genug ist, selbst einen Teil des Einsetzens zu übernehmen.

Ein Agent ist etwas, das im Auftrag eines anderen mit einem gewissen Grad an Autonomie handelt, das gilt für Institutionen und Einzelpersonen genauso wie für Software, so Evas Definition. Bei KI-Agenten wird daraus ein zugrunde liegendes Modell, das sich wie jemand verhält und eine übergebene Aufgabe übernimmt, samt dem Risiko, das damit einhergeht. Bernhards eigene Erfahrung aus der Rechtsabteilung zeigt ihm, dass KI als Recherche-Tool begann. Doch er merkte, dass ihm die Argumente dafür ausgingen, sobald er sah, wie KI Strategien entwickelt und Empfehlungen gibt, statt nur Informationen zu beschaffen.

Eva brachte außerdem das zentrale Konzept des Abends ein, die Trust-Gleichung, die David Maister erstmals 2000 in seinem Buch „*The Trusted Advisor*“ vorgestellt hatte:

Trust = Glaubwürdigkeit + Zuverlässigkeit + Empathie, geteilt durch Selbstorientierung

Kai ist ebenfalls ein Anhänger dieser Gleichung, die auf die Gründungsphase seiner Charter of Trust-Initiative zurückgeht, und stimmte Eva zu, dass KI alle drei der ersten Eigenschaften imitieren kann. Was sie nicht kann, ist sich selbst zu orientieren wie ein Mensch, und genau deshalb kann KI selbst im eigentlichen Sinne niemals wirklich als vertrauenswürdig bezeichnet werden. Vertrauenswürdig oder nicht ist stattdessen das Unternehmen, das hinter dem Modell steht, da die Anreize eines Modells letztlich auf eine Organisation mit eigenen Interessen zurückgehen. Die Frage ist nicht, ob wir dem Output vertrauen können, sondern ob wir der Quelle vertrauen können, die ihn produziert.

Schatten-KI, und wer eigentlich an die Daten möchte

Bernhard stellte der Runde ein Szenario vor.

„Was, wenn eine KI Zugriff auf alles hätte, was ein Unternehmen weiß, plus alles, was öffentlich verfügbar ist, und einem einfach sagen würde, welche Entscheidung die beste wäre?“

Johann war unmissverständlich, dass dies in der Praxis so schnell nicht passieren wird. Große, regulierte Unternehmen werden diese Art von Zugriff nicht gewähren. Er nannte das Beispiel eines OEM, dessen erster Instinkt bei LLMs war, Mitarbeitende komplett

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.

auszuschließen, statt zu erkunden, wo das Werkzeug tatsächlich helfen könnte. Gleichzeitig nutzen viele Mitarbeitende bereits aus eigener Initiative private LLM-Apps für Arbeitsaufgaben. Das verweist auf eine Grauzone namens Schatten-KI.

Laut Kai müssen zwei Dimensionen dieses Problems betrachtet werden.

- Auf der **Governanceebene** ist mehr Daten nicht automatisch besser. Das ist dieselbe Lektion, die die Industrien schon aus der Big-Data-Ära gelernt haben. Tatsächlich liegt die Arbeit in der Orchestrierung rund um die richtigen Daten, nicht darin, einfach mehr davon hineinzuschütten.
- Auf der **Vertrauensebene** griff er die bereits erwähnte Gleichung wieder auf, mit einer pointierten Ergänzung. LLMs selbst mögen nicht selbstorientiert sein, aber die Unternehmen, die sie bauen, sind es in unterschiedlichem Ausmaß. Das ist ein legitimer Grund für eine Organisation, zu zögern, bevor sie ihre Daten weitergibt.

Agenten am Arbeitsplatz, wer haftet?

Unternehmen stehen vor einer echten Entscheidung, sobald Mitarbeitende merken, was Agenten können. Manche setzen bewusst auf Weiterbildung, Eva nannte das Beispiel eines CEO, der jedem Mitarbeitenden gezielt Raum zum Wachsen gibt, um sie zu halten. Viele andere haben dagegen gar keine Antwort und hoffen einfach, dass die Politik oder die Tech-Branche das Problem für sie löst. Sobald Mitarbeitende merken, dass ein Agent ihre Aufgabe schneller, günstiger und ohne Beschwerden oder Erschöpfung erledigen kann, macht sich im Raum vor allem Angst breit. Ob Arbeitgeber, die Tech-Branche oder die Politik die Verantwortung trägt, diese Angst aufzufangen, bleibt eine offene Frage.

Das Gespräch wandte sich dann einem schärferen Punkt zu, der Möglichkeit, dass ein Agent selbst zur Führungskraft wird. Trotz vieler Vorteile der KI bei der Arbeitsleistung und der Simulation menschlicher Empathie kann sie noch keine Verantwortung im rechtlichen oder haftungsrelevanten Sinne übernehmen. Genau das, mehr als reine Orchestrierungsfähigkeit, rechtfertigt laut Kai, was Führungskräfte verdienen.

„Orchestrierung selbst ist oft nicht der schwierige Teil eines Führungsjobs, Verantwortung zu übernehmen ist es“, so Kai. Er nannte außerdem ein aktuelles Beispiel, ein Startup, das gezielt Versicherungen für KI-Agenten aufbaut und pro Agent eine Prämie erhebt, um Restrisiken abzudecken, sobald die technische Governance steht. Eva ergänzte, dass Verantwortlichkeit innerhalb von Organisationen schon vor den Agenten diffus war, und Agenten diese Unschärfe nur schwerer zu ignorieren machen.

„RentAHuman“ und die nächste Plattformökonomie

Damit kehrte der Abend zu seinem Ausgangspunkt zurück. Auf RentAHuman können sich Menschen registrieren, um für reale Aufgaben gebucht zu werden. Der kontroverse Punkt sind Szenarien, in denen die buchende Partei ein KI-Agent ist und nicht ein anderer Mensch. Im Prinzip bekommt der Agent eine Kreditkarte und bezahlt damit einen Menschen dafür, etwas Physisches zu tun. Das Panel zog eine direkte Linie zu

bestehenden Plattformunternehmen wie Uber, die zwar Beschäftigung schufen, aber auch Externalitäten und unsichere Arbeitsbedingungen mit sich brachten.

Johann brachte ein konkretes Beispiel für physische KI ein, das irgendwo zwischen komisch und unangenehm lag. Es handelte sich um Videoaufnahmen eines Menschen beim Bügeln, die gezielt aufgenommen wurden, um einen Roboter zu trainieren, der diese Aufgabe irgendwann übernehmen soll. Weder das Panel noch derzeit irgendeine politische Partei scheinen eine klare Antwort darauf zu haben, was für die verdrängten Arbeitskräfte als Nächstes passiert. Das Panel traute sich keine Prognose zu, sondern gestand offen ein, dass diese Art von Arrangement in fünf Jahren durchaus normal sein könnte.

Jobs, Regulierung, und warum Wahlfreiheit weiterhin zählt

Zum Arbeitsmarkt merkte Johann an, dass die Forschung uneinheitliche Ergebnisse liefert. Manche Studien zeigen Jobverluste, andere zeigen, dass Unternehmen mit steigenden KI-Investitionen mehr Einstiegsjobs schaffen statt weniger. Einstiegspositionen pauschal zu streichen sei kurzsichtig, so Johann, angesichts des sogenannten **Augmentation-versus-Automation-Paradoxons**. Ohne Nachwuchs, der in ein Feld hineinwächst, gibt es niemanden mehr, den man weiterentwickeln könnte, und das Fachwissen verschwindet irgendwann.

Zur Regulierung war die allgemeine Ansicht des Panels, dass Unternehmen nicht allein aus gutem Willen im öffentlichen Interesse selbstregulieren werden, denn individuell gibt es dafür keinen Anreiz. Politik und Regulierung haben also weiterhin eine Rolle bei der Gestaltung des Rahmens. Der EU AI Act wurde als andere Strategie im Vergleich zu einem Wettlauf ohne Leitplanken diskutiert, und im Raum sah man eine echte Chance, den „Made in Germany“-Qualitätsstandard in einen Wettbewerbsvorteil zu verwandeln, statt Regulierung nur als Bremse zu behandeln.

Zur Skepsis, ob man dem KI-Produkt eines gewinnmaximierenden Unternehmens überhaupt vertrauen kann, standen sich zwei Positionen gegenüber. Die eine hielt dies für wünschenswert, aber derzeit idealistisch, und verwies auf die Abkehr von Social Media von ihren guten Absichten während der langen Abwesenheit von Regulierung. Die Gegenposition war, dass Vertrauen ein Spektrum ist, das vom Anwendungsfall abhängt, statt eine binäre Frage zu sein. Deshalb bringt ein risikobasierter Ansatz, bei dem Modell, Datenzugriff und Grad der menschlichen Aufsicht von Fall zu Fall gewählt werden, mehr als ein pauschales Urteil. Laut Kai ist der Erhalt von Vielfalt unter den LLM-Anbietern der entscheidende Faktor für Vertrauen, damit niemand auf eine kleine Zahl dominanter Plattformen beschränkt bleibt, so wie es bei Social Media hinauslief.

Abschlussrunde

Eine interaktive Übung bat die Sprecher:innen zum Abschluss, den Satz zu vervollständigen: „Erfolgreiche Zusammenarbeit zwischen Mensch und KI wird 2035 geprägt sein durch...“

- **Johann:** Menschliche Akzeptanz und KI-Fähigkeiten.
- **Eva:** Gegenseitiges Verständnis.
- **Kai:** Governance, die auf der Erkenntnis basiert, dass wir gemeinsam besser dastehen.

Nach einem Ratschlag für eine Führungskraft oder Politiker:in mit Blick auf eine agentengetriebene Zukunft gefragt, setzte das Panel aus drei Richtungen ein gemeinsames Thema.

- **Kai:** Probiert die Technologie selbst aus, bevor ihr euch eine Meinung bildet, und schaut euch die Risiken genauso ehrlich an wie die Chancen.
- **Eva:** Versteht erst eure eigene Organisation, bevor ihr nach neuer Technologie greift.
- **Johann:** Verlasst euren gewohnten Kreis, um andere Perspektiven zu hören, bevor ihr etwas entscheidet. Auch der Besuch von Diskussionen wie dieser und das Einholen aktueller Informationen kann ein guter Anfang sein.

Offene Fragen und beachtenswerte Thesen für die Zukunft

1. **Wo landet Verantwortlichkeit, sobald ein Ergebnis von Agent zu Agent weitergereicht wird?** Das Panel war sich einig, dass sich das schon innerhalb von Organisationen ohne KI verflüchtigt, Agenten machen diese Lücke nur schwerer zu ignorieren.
2. **Kann KI selbst je vertrauenswürdig sein, oder nur die Organisation dahinter?** Wenn Vertrauen von Selbstorientierung abhängt, und nur Unternehmen, nicht Modelle, diese in einem bedeutsamen Sinne besitzen, dann ist jede Frage nach dem Vertrauen in ein KI-Produkt eigentlich eine Frage nach dem Vertrauen in einen Anbieter.
3. **Sind Konzepte wie „RentAHuman“ echte Vorboten der Plattformökonomie 2.0, und wer trägt die Externalitäten?** Bestehende Plattformunternehmen zeigen bereits ein Muster, das es genau zu beobachten gilt, sobald KI-Agenten selbst zu den Buchenden werden.
4. **Kostet uns das Streichen von Einstiegsjobs heute die Fachkräfte-Pipeline von morgen?** Das **Augmentation-versus-Automation-Paradoxon** legt das nahe, denn ohne Nachwuchs, der in ein Feld hineinwächst, gibt es niemanden mehr, den man weiterentwickeln könnte.