

KI-Wettlauf 2026: Innovationen, Investitionen und Machtfragen

Ausgesprochen digital. Der Podcast für digitale Trends.

Intro

[00:00:11.660] - Steffen Wenzel

Hallo und herzlich willkommen bei Ausgesprochen digital. Ja, 2025 markierte wahrscheinlich einen Wendepunkt für künstliche Intelligenz, weil es vom Experiment zur kritischen Infrastruktur geworden ist, wirtschaftlich, politisch und auch militärisch, wie wir leider in vielen Kriegen feststellen mussten. Wenige Konzerne und Staaten kontrollieren die leistungsfähigsten Systeme im Zentrum natürlich der Systemkonflikt zwischen den USA und China. Wir blicken auf die entscheidenden Entwicklungen des Jahres 2025 und fragen natürlich auch mit Blick auf 2026, welche Rolle bleibt denn da noch für Europa, für die EU und natürlich auch für Deutschland. Und diesen Ausblick machen wir mit Professor Frank Schönefeld, bis vor kurzem Mitglied der Geschäftsleitung bei der Telekom MMS, jetzt als KI-Experte und Berater tätig und natürlich auch ein wenig unsere KI-Glaskugel bei Ausgesprochen digital. Hallo Frank, schön, dass du da bist.

[00:01:11.980] - Prof. Dr. Frank Schönefeld

Ja, danke für die Einladung. Ich freue mich erneut mit euch sprechen zu können.

[00:01:16.800] - Steffen Wenzel

Frank, erst mal, wie geht es dir? Theoretisch bist du im Ruhestand, aber nicht wirklich, oder?

[00:01:22.220] - Prof. Dr. Frank Schönefeld

Ja, sind noch ein paar Aktivitäten übriggeblieben. Ich bin im Silicon Saxony aktiv. Ich halte weiter Vorlesungen an der HTW Dresden und ich freue mich natürlich, die Verbindung mit der MMS zu halten über die schon erwähnte Beratungstätigkeit. Und da bin ich sehr aktiv in den AI-Communities und konnte auch das europäische Projekt, das IPCEI-Projekt, KI "Important Project of Common European Interest", ein Stück weit begleiten.

[00:01:49.660] - Steffen Wenzel

Frank, lass uns erst mal ein bisschen zurück gucken. Was waren die Hauptinnovationen in 2025? Was ist dir aufgefallen?

[00:01:57.320] - Prof. Dr. Frank Schönefeld

Ja, wir können ja so sagen, die Kambrische Explosion, die liegt nunmehr drei Jahre zurück. November 2022 erblickte GPT-3, das Licht der Welt. Das heißt, wir haben jetzt eine stabile Entwicklung über drei Jahre, drei Monate und man muss schon, man kann sagen, KI ist erwachsen geworden, hat ihre Leistungsparameter in vielen Gebieten einfach ausgeweitet und zieht in den meisten Vergleichen in der Leistungsfähigkeit auch schon an menschlicher Leistungsfähigkeit vorbei. Das heißt, sie wird immer leistungsstärker. Wenn man den Vergleich zu vor drei Jahren ziehen will, sagt man, ungefähr so ein Faktor 6, also ein Faktor 6 intelligenter. Und sie war schon am Anfang ziemlich intelligent. Das heißt, das ist eigentlich die Quintessenz all dessen. Sie wird immer nutzbarer, sie wird immer intelligenter. Und um das vielleicht an Beispielen auch noch mal greifbar zu machen: Sie ist jetzt in der Lage, Mathematik Olympiaden zu gewinnen, zu bestimmen. Und da gibt es ein tolles Verfahren, das heißt „Self-Verifying Proofs“. Das heißt, es schaut über seine eigene Schlussfolgerungskette und kann daraufhin bewerten: „Da

habe ich wirklich was Gutes abgeliefert. Dem kann ich trauen." Das ist also auch eine Methode, um Halluzinationen vorzubeugen. Und wir sind dort inzwischen wirklich auf Olympiade Niveau angekommen. Du hast explizit nach Innovationen gefragt, würde man jetzt sagen, es gab die Breakthrough Innovation?

[00:03:30.890] - Prof. Dr. Frank Schönefeld

So würde ich es nicht beschreiben. Eher ist es die Vielzahl an Innovationen, kleinen Schritten, die aber in der Summe wieder diesen Riesenschritt ermöglichen. Und die kleinen Schritte, die sind einmal in den Modellen selber zu finden, die werden immer effizienter. Wir können auch mit kleineren Modellen inzwischen arbeiten. Da kommt durchaus ein guter Hub, Faktor 10 in der Leistungsfähigkeit. Dann haben wir einfach bessere Algorithmen, bessere Einschätzung – kann ich noch ein Beispiel bringen. Und natürlich von der Hardware-Front kommt auch massiver Support, der auch noch mal einen Faktor 10 bringt. Und wenn ich das zusammenaddiere, komme ich eben auf eine erhebliche Leistungssteigerung bei den Algorithmen. Einer, der mir besonders ins Auge gefallen ist, ist dieser Übergang von Single Token Prediction zu Multi Token Prediction. Das heißt, anstelle beim Generieren immer nur Wort für Wort zu generieren und jeweils den ganzen Apparat hinterherzuziehen, generiert man jetzt lieber gleich vier, zwei oder vier Token und wird damit genau den gleichen Faktor effizienter. Also die Vielzahl der Einzelinnovationen führt zu dieser massiven Leistungssteigerungen in Gänze.

[00:04:47.840] - Steffen Wenzel

Wir alle benutzen die Sprachmodelle und haben damit Berührung, aber ich habe ja eben meine These aufgestellt, dass es jetzt auch in die kritische Infrastruktur reingegangen ist, also KI, künstliche Intelligenz. Kannst du da das ein bisschen verifizieren oder würdest du das eher anders sehen?

[00:05:04.360] - Prof. Dr. Frank Schönefeld

Da würde ich sagen, das ist zweischneidig im Moment noch. Was wir sehen, ist, dass in der Tat die potenzielle Leistungsfähigkeit sehr groß ist und das über eine Vielzahl von Gebieten. Ich würde mal als Beispiel den Benchmark, den OpenAI selbst verwendet, heranziehen. Das ist der GDPval. Und dort werden reale Aufgaben aus der täglichen Arbeit aus der täglichen Ökonomie herangezogen und inzwischen ist GPT in der Lage, mit der 5.2er-Version etwa 70% dieser relevanten Aufgaben aus der Wirtschaft auf gleichem Niveau oder besser als ein Mensch zu lösen. Insofern ist es also in der Tat in der Kritischen, nämlich in allen Prozessen angekommen. Was trübt den Wein an der Stelle ein bisschen? Wir sehen, dass die die Adoptionsfähigkeit der existierenden Unternehmen nicht mit diesen Fortschritten komplett mithalten kann. Da gibt es vielfältige Gründe. Erstens hat man vielleicht nicht die richtigen Ressourcen, das so schnell einzuführen, es anzuwenden. Zweitens hat man sicherlich rechtliche Compliance-Sicherheitsprobleme, Toolverfügbarkeitsprobleme und irgendwann kommt auch ein Kostenproblem hinzu. Das heißt, die Trägheit der existierenden Systeme, Trägheit im physikalischen Sinne, die sind doch sehr beharrend, diese Systeme, und es dauert länger, als wir ursprünglich angenommen haben. Und wir müssen uns einfach noch ein, zwei, drei Jahre mehr Zeit geben, bis wir dort richtig den Durchbruch sehen.

[00:06:49.200] - Prof. Dr. Frank Schönefeld

Und das führt dann umgekehrt auch zu einer Neubewertung der gesamten – jetzt bin ich wieder bei kritischen Infrastrukturen – den gesamten Investitionen da rein. Da stehen 5 Billionen Dollar im Raum in den nächsten paar Jahren. Und da ist die Frage: Wer braucht diese Investitionen? Werden die sich jemals amortisieren? Werden sie sich schnell amortisieren? Und da kommen immer mehr Fragezeichen auf. Das heißt, wir werden da durchaus auch eine Wertberichtigung, eine Verschiebung sehen und auch eine Marktbereinigung. Da würde ich noch mal später drauf eingehen.

[00:07:25.080] - Steffen Wenzel

Können wir bestimmt. Wichtiges Thema ist natürlich nicht nur, dass es die kritischen Infrastrukturen der Big Player oder der großen Firmen betrifft, sondern natürlich auch meine persönlichen. Also ein Geschäftsmodell und sehr viel versprechendes ist ja das sogenannte Agentic AI, dass ich Agenten habe, die dann gewisse Tätigkeiten für mich übernehmen, aber dadurch natürlich auch Zugriff zu allen Daten, zu allen Zugängen, die ich habe, Bankdaten et cetera, brauchen, damit sie das tun können. Wie bewertet du Agentic AI? Was ist das für dich gerade für ein Thema?

[00:08:00.260] - Prof. Dr. Frank Schönenfeld

Wir haben mit Innovationen gestartet und wenn man vielleicht doch die Innovation 2025 herausheben will, es gibt sie schon ein bisschen länger, aber der Fortschritt der agentischen Systeme und das Versprechen, was sie uns liefern, der hat sich ja 2025 erst so richtig deutlich herausgebildet. Ich denke, das kann man nicht überschätzen. Damit wären viele der Anfangskrankheiten oder der auch im einzelnen Sprachmodell noch vorhandenen Drawbacks, also Nachteile, werden durch die Agenten wieder aufgehoben. Die Agenten können verschiedene Rollen annehmen. Der eine kann der Qualitätssicherer sein, der andere kann die Lösung generieren, also verschiedene Rollen. Wir haben eine komplette Speicher-Hierarchie in den Agenten. Wir haben ein Kurzzeit-Gedächtnis, wir haben inzwischen ein Episodengedächtnis in den Agenten und wir haben ein Langzeit-Gedächtnis. Und die KI kann auch, so ähnlich wie bei der Mathematik-Olympiade, kann verifizieren: „Habe ich da was Sinnvolles gemacht? Oder eher doch nicht?“ Es kann sogar entscheiden: „Oh Gott, ich sollte mich vielleicht doch mal mit meinem Menschen unterhalten, weil hier bin ich mir selber unsicher.“ Und das ist eben diese neue Leistungsfähigkeit. Und sie kommt aber zu einem Preis, das müssen wir auch sagen, wir müssen ungefähr den Faktor 20 ansetzen, den so ein agentisches System einsetzt im Sprachmodell.

[00:09:30.000] - Prof. Dr. Frank Schönenfeld

Das heißt, das Sprachmodell selber oder die Sprachmodelle werden zu einem Faktor 20 utilisiert, angefragt, genutzt, als in einem normalen Prompt.

[00:09:40.960] - Steffen Wenzel

Das heißt, mehr Energieverbrauch auch, oder?

[00:09:43.280] - Prof. Dr. Frank Schönenfeld

Das auch, Aber nur damit gelingt es natürlich, diese Autonomie, und das ist vielleicht das zentrale Wertversprechen, hier wird etwas Autonomes, was weitgehend unabhängig bis das gewünschte Endergebnis erreicht ist, auf die Schiene gesetzt. Das sind so die zentralen Versprechen und das wird weitergehen. Das bleibt eines der heißen Themen für 2026, können wir ganz klar sagen. Energieverbrauch, du hattest es angesprochen, vielleicht schon eine Kurzbemerkung dazu. Da haben wir zwei gegenläufige Tendenzen. Das eine ist, durch die Verbesserungen in den Chips, in den Algorithmen, in den Modellen, haben wir eine Absenkung der Kosten. Zum Beispiel das aktuelle Gemini-Modell hat eine um 30% reduzierte Output-Token-Produktion. Das ist schon wesentlich. Das heißt, es ist einfach in dem, was es generiert, um 30% effizienter. Umgekehrt haben wir aber durch die Agenten eine 20-fache Verstärkung. Und dann spare ich auf der einen Seite 30% Token ein und auf der anderen Seite erhöhe ich es den Faktor 20. Das heißt, am Ende bleibt trotzdem eine Erhöhung stehen.

[00:10:55.140] - Steffen Wenzel

Wollte ich gerade sagen, das ist jetzt eine Mathematik Leistung, wofür ich wahrscheinlich die KI brauche, um das dann endgültig auszurechnen. Lass uns noch mal einen Blick auf das Große und Ganze werfen. Du sprichst vom Kampf der Systeme: China versus USA. Was findet dort gerade statt? Was beobachtest du?

[00:11:13.180] - Prof. Dr. Frank Schönefeld

Während wir die ersten beiden Jahre nach der Kambrischen Explosion haben wir sicherlich die USA massiv vorne gesehen. GPT hat die Maßstäbe gesetzt und dann kamen die anderen Modelle der großen Anbieter: Anthropic, Google, xAI, um sie zu nennen. Und im Schatten dieser Entwicklung, lange unbemerkt, hat aber in China eine massive Entwicklung stattgefunden und die hat sich dann zuerst mit DeepSeek Raum gebrochen, wo plötzlich im Raum stand: „Oh, die können mit einem Kostenfaktor Vorteil von 200 trainieren.“ Das war die ursprüngliche Aussage. 5 Millionen versus eine 1 Milliarde Trainingskosten. Inzwischen wissen wir, dass da Äpfel mit Birnen verglichen worden sind. Aber Andreessen hat gesagt, das war der Sputnik-Schock für die KI der USA, dieses DeepSeek-Erlebnis. Und wir müssen sagen, jawohl, da sind starke Anbieter einfach unterwegs und neben die DeepSeek, würde ich noch Moonshot AI erwähnen oder die großen Firmen, die alle ihr eigenes Sprachmodell haben, eilen vorneweg Alibaba mit seinen Qwen-Modellen. Und wenn man die jetzt in Benchmark-Systeme hineinwirft, stellt man fest, in den Basis Benchmarks zieht China auch an den USA vorbei und nur in den fortgeschrittenen Benchmarks ist ein leichter Fortschritt für die USA noch zu sehen. Das heißt, wir sind eigentlich ex aequo.

[00:12:41.440] - Prof. Dr. Frank Schönefeld

Der einzige Vorteil, den die USA hat, ist natürlich ihre Systemüberlegenheit, Chipsystemüberlegenheit und vielleicht auch Cloud-Überlegenheit. Aber auch der steht natürlich auf nicht immer belastbaren Füßen. Wenn man überlegt, wo die Chips hergestellt werden, dann kommt natürlich sofort die große Geopolitik ins Spiel. Was weiterhin für China spricht im KI Kampf der Systeme, ist die billigere Energie. Sie können zur Hälfte des Preises Energie bereitstellen und ausreichend Energie bereitstellen, um auch die Investitionen in KI-Infrastrukturen langfristig zu powern. Und sie haben mehr Ingenieure. Die pure Masse an Ingenieuren ist natürlich in China viel größer und deswegen findet man auch kaum noch ein KI-Paper, wo nicht ein chinesischer oder asiatischer Name auftaucht. Also das bleibt spannend zu entwickeln. Letzte Entwicklung vielleicht noch: Auch im Open Source oder Open Weight Geschäft hat sich eine massive Verlagerung zugunsten Chinas ergeben. Also die tollsten und leistungsfähigsten Open Weight-Modelle, jedenfalls in der Breite dann, die kommen inzwischen aus China und das führt aber, wenn man das zu Ende denkt, natürlich dann sofort zur Frage: Souveränität, KI-Souveränität. Also müssen wir gegebenenfalls chinesische Modelle hier in unserem Umfeld hosten, um überhaupt noch KI betreiben zu können.

[00:14:19.160] - Steffen Wenzel

Das ist ein wichtiger Punkt. Mir geht es auch da die Anwendungsfälle. Wenn ich das so in meinem Verständnis erst mal so sehen würde, komme ich ja mit chinesischer KI überhaupt nicht in Kontakt. Und ich könnte kann man ja auch immer als westliches Unternehmen sagen: „Möchte ich auch gar nicht aus Datenschutzgründen etc. pp.“ Genauso wie es ja zum Beispiel auch kritisch gesehen wurde, Huawei-Technologie im 5G-Bereich einzusetzen, weil man gesagt hat, das entspricht nicht unseren Datenschutzrichtlinien, deswegen machen wir das in Deutschland nicht. Ist diese Tendenz bei KI irgendwie aufgehoben, dass man sagt, Kampf der Systeme, da wird sich ein System durchsetzen und deswegen werden wir nicht daran vorbeikommen, eventuell auch chinesische KI-Technologie dann zu benutzen?

[00:15:06.260] - Prof. Dr. Frank Schönefeld

Ich denke, so weit ist es noch nicht, aber man kann es, glaube ich, auf alle Zeit auch nicht ausschließen. Und es ist ja durchaus berechtigt, Bedenken zu haben. Nicht umsonst bleiben irgendwo Elektrobusse stehen, die in China gefertigt worden sind. Und das ist natürlich ein Szenario, was man vermeiden will. Umgekehrt weiß man aus Befragungen von Firmen: "Würdet ihr chinesische Open Weight-Modelle einsetzen, wenn sie in sicheren Umgebungen gehostet werden?" Und dann nimmt die Bereitschaft, das zu tun, in den Befragungen immer weiter zu. Aber noch sind wir nicht so weit und das bleibt eine der

spannenden Fragen für 2026. Wie wird sich das entwickeln? Rein von der Leistungsfähigkeit sind das sehr starke Modelle. Und jetzt noch mal: Wie viel Modelle braucht man überhaupt? Die Frage ist natürlich auch berechtigt zu stellen. Ich versuche, immer zu beantworten: Wie viele Browser gibt es? Wie viele Betriebssysteme für Laptops gibt es? Und wie viele Betriebssysteme für Handhelds, also Mobilgeräte, gibt es? Und man wird in jede dieser Fragen mit der Antwort zwei, manchmal auch drei, vielleicht noch vier kommen, aber im Wesentlichen haben sich Duopole herausgebildet. Und das würde ich generell auch für diesen Markt hier erwarten. Nicht umsonst sprechen wir von Foundation-Modellen.

[00:16:31.240] - Prof. Dr. Frank Schönefeld

Es sind grundlegende Modelle und von grundlegenden Modellen brauche ich nicht 20, brauche ich auch nicht 10. Brauche ich vielleicht noch vier, aber am Ende würde ich auch dort so einen Duopol erwarten und man kann jetzt schon wetten, wer aus einer solchen Entwicklungen als Sieger hervorgehen wird.

[00:16:50.900] - Steffen Wenzel

Ich glaube, die Bedeutung, wer da als Sieger hervorgehen wird und was das für einen Einfluss auf unser Leben hat, ist auch noch mal wichtig dabei, zu berücksichtigen, dass es eben nicht nur Sprachmodelle geht, weil da mit denen kommen wir immer wieder in Kontakt, also wir als Normalbürgerinnen und Bürger, sondern es geht auch die Industrie, Industrial AI, also wo KI eingesetzt wird. Kannst du das noch mal so unterscheiden, welche zwei Faktoren das sind und warum das so eine Bedeutung auch hat?

[00:17:21.580] - Prof. Dr. Frank Schönefeld

Na ja, nach wie vor macht ja den Erfolg eines guten Modells machen gute Daten aus. In Europa haben wir noch, muss man sagen, einen guten industriellen Kern und insofern haben wir, insbesondere auch in Deutschland, auch wenn sich da Veränderungen anbahnen. Und insofern haben wir auch gute industrielle Daten. Wenn ich jetzt eine industrielle Automatisierung, eine industrielle KI will, profitiere ich natürlich von Modellen, die mit diesen industriellen Daten gebaut worden sind, gegebenenfalls weiterentwickelt worden sind, finegetuned worden sind oder wo sie einfach in einem Zusatzspeicher zur Verfügung stehen und dann mit dem klassischen Modell zusammenarbeiten. Ich glaube, das ist der Weg, den Europa auch für sich letztlich gewählt hat. Lass uns diese Datenhoheit in dem Kontext bewahren und uns aber mit den fortgeschrittensten Modellen, die die Welt im Moment kennt, miteinander verbinden. Und deswegen kommt es auch zu solchen Partnerschaften, die wir immer wieder in der Presse lesen, Siemens und NVIDIA versuchen dort, Entwicklungen, die wir Industrie 5.0 nennen könnten, anzustoßen. Und ich denke, das kann bis zu einem gewissen Grad auch erfolgreich sein.

[00:18:41.800] - Steffen Wenzel

Also das heißt, in der Industrie operativ läuft KI in den Prozessen, zur Steuerung, zur Optimierung von Prozessen, Predictive Maintenance, also vorausschauend zu gucken, welche Teile gehen vielleicht kaputt und müssen ersetzt werden. All da ist KI schon integriert oder wird integriert.

[00:19:00.030] - Prof. Dr. Frank Schönefeld

Na ja, wird integriert. Ich würde da doch noch eine gewisse Zukunft sehen. Natürlich gibt es Predictive Maintenance-Szenarien schon viel länger, aber gerade der massive Einsatz von generativer KI, von autonomer KI in der Produktionssteuerung, den Schritt, den müssen wir in der Breite noch gehen und den werden wir auch gehen. Und da gibt es auch andere Allianzen. Wir haben über die Deutsche Telekom, kann man dort sprechen, die natürlich auch interessiert ist, solche Angebote für ihre Kunden auf den Weg zu bringen im Rahmen der Industrial AI Cloud bis hin zu einer Gigafactory. Wir werden sehen, wann sie kommt, wo sie kommt. Und das sind auf jeden Fall spannende Entwicklungen auch, was Lidl macht, oder die Schwarz Gruppe hat ein großes KI-Rechenzentrum in der Lausitz, also vielleicht 100 Kilometer von hier, angekündigt. Ich denke, die deutsche Industrie hat sich auf den Weg gemacht, aber der Weg ist auch noch weit und bietet einige Stolperfallen.

[00:20:02.620] - Steffen Wenzel

Das ist eine gute Überleitung. Bevor wir zu Deutschland gucken, lasst uns noch mal zu Europa gucken. Ich glaube, das wissen wir alle in der jetzigen Zeit. Europa wird da immer wichtiger. Deutschland ist natürlich ein wichtiger Teil davon. Aber wie steht es die europäische KI-Entwicklung?

[00:20:18.220] - Prof. Dr. Frank Schönefeld

Ich fange mit der deutschen Perspektive an und gehe dann auf die Europäische über. Der Zug mit den Foundation-Modellen, der ist sehr weit aus der Halle gefahren und wir können gerade noch so die Rücklichter gegebenenfalls erkennen. Das gilt für Deutschland. Ob es gelingt, da mit einem erweiterten Teuken-70B, also 70 Milliarden Parameter, da wirklich noch mal einen Unterschied zu machen, bleibt abzuwarten oder auch 100B. Das wissen wir nicht. Aus meiner Sicht gibt es dort nur noch auf europäischer Ebene den französischen Anbieter Mistral mit seinen Modellen, die auf dieser Liga ungefähr mitspielen können. Das ist also gut. Was ich aber als Veränderung wahrnehme, ist, wie entscheidend die Frage auf der europäischen Ebene betrachtet wird, wie sie behandelt wird und wie sie in den Mittelpunkt gestellt wird. Und das ganz am Anfang erwähnte IPCEI - Important Project of Common European Interest- adressiert eben jetzt gerade diese KI. Und auch dort versucht man natürlich, die Frage der souveränen, souveräne KI: Wie sieht souveräne KI aus? Wo braucht man die Souveränität? Brauche ich die Souveränität im Foundation-Modell oder brauche ich sie eher in der Data Pipeline, in den Industriedaten und in den Anwendungen und in den Orten, wo ich diese Anwendung betreibe?

[00:21:44.580] - Prof. Dr. Frank Schönefeld

Und das ist eher die Antwortrichtung, in die die EU jetzt geht und diese auch supportet mit dem Projekt.

[00:21:50.280] - Steffen Wenzel

Wie bewertet du die?

[00:21:52.940] - Prof. Dr. Frank Schönefeld

Na, ich würde sagen, ein bisschen natürlich aus der Not geboren, weil der Fortschritt oder der Vorsprung sozusagen der großen Foundation-Modelle ist erheblich. Den kannst du nicht über Nacht aufholen und insofern scheint es, angesichts der entstandenen Rahmenbedingungen, eine sinnvolle, aber auch eine gewissermaßen erzwungene Vorgehensweise zu sein.

[00:22:18.480] - Steffen Wenzel

Also wir brauchen Souveränität, digitale Souveränität, um natürlich jetzt auch gerade bei den Entwicklungen in den USA uns da nicht dauerhaft abhängig zu machen. Dass das schwierig ist, ist ja unbenommen. Jetzt wird ja oft behauptet, dass die Regulierung, die Überregulierung, um das schon mal vorwegzunehmen, wie das manche beurteilen, dazu führt, dass wir weniger Souveränität, weniger Innovation bekommen. Denkst du, das ist richtig, diese Annahme?

[00:22:47.220] - Prof. Dr. Frank Schönefeld

Ich denke, ich habe im letzten Podcast so was Ähnliches gesagt. Ich habe einfach den Chef des norwegischen Staatsfonds zitiert und der hat sinngemäß gesagt: USA haben viel KI, aber wenig Regulierung und Europa hat wenig KI, aber viel Regulierung. Und ich glaube, man kann das noch verschärfen. Irgendwann verlierst du das Recht, den moralischen Zeigefinger zu heben, wenn du gar keine Grundlage mehr hast, aufgrund der dieser Zeigefinger beruhen würde. Und deswegen sage ich, du musst auch mit der Entwicklung weitgehend mitgehen können und nur daraus ergibt sich ein gewisses Recht, auch eine Regulierung voranzutreiben und weiterzuentwickeln.

[00:23:36.470] - Steffen Wenzel

Aber glaubst du wirklich, dass Regulierung dazu führt, dass KI-Innovationen nicht stattfinden? Weil Regulierung bedeutet ja eigentlich nur, ich gebe einen klaren Rahmen vor, wie Datenmodelle zu funktionieren haben, welche Daten anonymisiert sein müssen, wie ich Datenschutz entsprechend lebe und so weiter und so fort. Während das ja doch maßgeblich bei KI-Innovationen um Investitionen geht. Also wenn ich sehe, wie viel Geld OpenAI eingesammelt hat, jetzt in den letzten Jahren oder auch die großen Player, und wie viel Geld, du hast es eben in China erwähnt, dass das natürlich dann noch mal reduziert wurde, dass das der Sputnik-Effekt war. Aber dennoch geht es doch darum, welches Geld steht zur Verfügung, um KI zu entwickeln und nicht, wie viel das Regulierung dazu führt, dass ich das erst mal gar nicht angebe, sei denn, Regulierung behindert Geschäftsmodelle.

[00:24:28.000] - Prof. Dr. Frank Schönefeld

Genau, und so, da drauf könnten wir uns auch einigen. Ich könnte sogar so sagen, eine gute Regulierung kann Innovationen weiter fördern und unterstützen und eine zu restriktive, insbesondere im Datenschutzbereich, kann es mal im Minimum verlangsamen. Aber was macht den eigentlichen Erfolg aus? Sind natürlich die Investitionssummen. Und das ist ja eigentlich kennzeichnend. Wer sind denn die Vorreiter der KI-Entwicklung? Das ist übrigens ein interessanter Aspekt. Hochschulen, Universitäten oder Akademia spielen keine Rolle mehr. Das ist eigentlich erschreckend bis zu einem gewissen Grad. Die fortschrittlichsten Modelle kommen aus Industrieunternehmen und nur aus Industrieunternehmen. Und das ist in China genauso wie in den USA. Das muss man sich auf der Zunge zergehen lassen. Die Akademia kann noch gute Leute ausbilden und bereitstellen, ohne Zweifel, und kann Teilespekte researchen, aber die tatsächlichen Modelle kommen aus Industrieunternehmen. Und woher kommen sie besonders gut? Natürlich aus denen, die über etablierte Geschäftsmodelle verfügen, die einen Cash-Flow von 23 Milliarden jedes Quartal erzeugen. Und dann fällt es natürlich leicht, da mal 50 Milliarden in die nächste Investitionsrunde für KI-Infrastruktur zu stecken. Und das ist durchaus auch ein struktureller Nachteil, den Europa da hat, den europäische Unternehmen da haben und den können wir nicht weg reden komplett.

[00:26:03.970] - Steffen Wenzel

Europa geht ja und auch Deutschland einen anderen Weg, nämlich nicht zu sagen, das sind die Tech-Konzerne, die das entwickeln, sondern zumindest gibt es eine staatliche Innovation oder staatlich-gelenkte Innovationszentrum. In Deutschland ist zum Beispiel das Zentrum für digitale Souveränität entstanden, dass dann jetzt in Zusammenarbeit mit der Industrie, mit Unternehmen digitale, souveräne Systeme entwickeln sollen. Es gibt auf der europäischen Ebene Gaia-X als Daten-und Cloud-Infrastrukturprojekt. Den Euro-Stack, der Deutschland-Stack, soll geplant werden. Hältst du dieser Weg also quasi nicht aus der freien Wirtschaft heraus, aus der freien Industrie Technologiekonzerne heraus, das zu entwickeln, sondern quasi gelenkt aus staatlicher Kontrolle heraus, dieses Thema anzugehen, für erfolgsversprechend?

[00:26:54.960] - Prof. Dr. Frank Schönefeld

Ich würde mal George Bush da zitieren. Der hat mal sinngemäß gesagt, was die größte Gefahr für ein Unternehmen ist. Und zwar hat er gesagt: „Wenn jemand zu dir kommt und sagt: „Ich bin der Staat und ich will dir helfen, dann ist höchste Gefahr angesagt. Und das nicht überbewerten jetzt. Ich möchte nicht diesen Programmen das Wasser abgraben. Im Gegenteil, ich bin ja selber Verfechter der gemeinsamen europäischen Vorgehensweisen an der Stelle, aber er kann natürlich trotzdem nicht vollständig die Wirtschaftskraft der Unternehmen, die Investitionskraft der Unternehmen ersetzen. Das ist aussichtslos rein von den Summen her. Und wir sehen ja auch, da wird man noch mehrere Sondervermögen brauchen, da auf einen grünen Zweig zu kommen.“

[00:27:46.020] - Steffen Wenzel

Aber das ist ja genau die Frage: Ist man gewillt, da rein zu investieren? Das heißt ja nicht, dass der Staat alles selbst machen muss oder die EU alles selbst machen muss, aber sie können ja damit Innovation in den Unternehmen steuern. Und das ist ja genau das, was passieren soll. Also genauso wie ich jetzt wieder Geld für die Elektroautoindustrie oder Automobilindustrie zur Verfügung stelle, könnte das ja auch in der KI-Industrie passieren.

[00:28:10.420] - Prof. Dr. Frank Schönefeld

Am Ende muss eine Investition sich selbst tragen, auch selbst rechnen, muss wirtschaftlich erfolgreich sein. Das gibt auch in diesen Fragen den Ausschlag und finde ich, it's the economy, stupid, weil ich einmal mit Bill Clinton, mit Präsidenten Zitaten angefangen hatte und das waren ja auch noch welche, die man zitieren konnte.

[00:28:37.250] - Steffen Wenzel

So sieht es aus. Es ist aber ein guter Übergang zu unserem nächsten Thema, nämlich Geschäftsmodelle. Die müssen ja dahinter liegen, wenn wir darüber reden, warum sollte man denn überhaupt investieren? Man könnte jetzt aus EU-Sicht, aus Deutschlands-Sicht sagen, natürlich wollen wir uns, wie gesagt, nicht abhängig von den USA machen oder von China, sondern wir müssen was Eigenes hier in Europa entwickeln. Aber es gibt natürlich, wenn ich den Unternehmen Geld gebe, dass sie auch weiter noch investieren, Geld einsammeln auf dem Markt. Es muss funktionierende Geschäftsmodelle geben. Jetzt auf der anderen Seite gibt es die Tendenz, dass man oft sagt, das ist alles eine Bubble, das ist eine Blase, diese KI-Geschäftsmodelle, die werden sich nicht so entwickeln, wie angenommen wird. Was denkst du darüber?

[00:29:21.740] - Prof. Dr. Frank Schönefeld

Ich bin heute bei meinen Zitaten, diesmal aber ohne einen Präsidenten. Ich glaube, das hat Amara gesagt. Amara war ein Senator und ein Leiter eines Instituts und der hat sinngemäß gesagt: "Wir tendieren dazu, die kurzfristigen Auswirkungen einer Entwicklung zu überschätzen und die langfristigen Auswirkungen zu unterschätzen." Und so ähnlich ist das hier, glaube ich, auch mit der KI. Wir haben gesagt, binnen drei Jahren wird sich die Welt auf den Kopf stellen. Das hat sie nicht getan. Sie tickt weiter, wie sie immer getickt hat. Aber die langfristigen Auswirkungen und da wäre langfristig, wäre schon, würde ab fünf Jahren beginnen und würde so den Zehnjahreszeitraum erreichen. Da werden wir schon dramatische Veränderungen sehen und die werden natürlich durch erfolgreiche Geschäftsmodelle auch geprägt sein. Und ich würde da so drei Kategorien sehen. Natürlich haben wir einmal die Kernanbieter von KI, von KI Foundation Models, aber auch da sehen wir schon zwei verschiedene Herangehensweisen. Das wären so die ersten beiden Unterscheidungen. Exemplarisch können wir GPT hernehmen. GPT hat auf der Grünen Wiese begonnen und hat versucht, um sein Large-Language-Modell sozusagen in ein Ökosystem herumzubauen. Im Moment haben sie zwischen 800 und 900 Millionen wöchentliche Nutzer. Das ist die Marktführerschaft. Etwa 68% aller KI-Anfragen gehen an GPT.

[00:30:55.560] - Prof. Dr. Frank Schönefeld

Das ist schon stark. Trägt es sich schon? Können sie mit ihren Freimodellen, mit ihren Account-Modellen sozusagen schon überleben? Nein, das reicht nicht aus. Deswegen erfinden sie immer neue Sachen hinzu. Erotische Gespräche werden möglich. Jetzt soll Werbung eingeführt werden als Allzweckwaffe. Das ist so das eine Modell, auf der grünen Wiese was aufzusetzen und nur aus dem KI-Geschäft was zu machen. Umgekehrt haben wir etablierte digitale Geschäftsmodelle, sei es in den sozialen Medien, sei es in der Internetsuche, sei es in der Videodarstellung etc. Und das ist ein etabliertes digitales Geschäftsmodell, was diese schon erwähnten Cash-Flow-Summen auch erzeugt. Und jetzt kann ich da versuchen, schrittweise KI reinzubringen. Und das fällt natürlich leichter, weil ich es aus existierenden Cash-Flows finanzieren kann, weil ich die Basisdienste damit stärker mache und deswegen

[00:31:58.760] - Steffen Wenzel

Das verstehe ich ganz.

[00:32:00.000] - Steffen Wenzel

sehe ich das als das erfolgreichere Modell an, von einem etablierten digitalen Geschäftsmodell ohne KI.

[00:32:07.880] - Prof. Dr. Frank Schönefeld

Beispiel?

[00:32:09.500] - Prof. Dr. Frank Schönefeld

Suche. Google.

[00:32:10.680] - Steffen Wenzel

Also Google Ads? Aber das ist ja genau das Problem bei Google Ads, dass Sie sagen: „Ja, blöd, wir können keine Ads mehr in KI schalten.“

[00:32:19.120] - Prof. Dr. Frank Schönefeld

Ja, aber man kann nach wie vor eine gute Link-Sammlung integrieren und man kann ja trotzdem die Sponsored Ads auch davor setzen und so weiter. Es gibt mehrere Möglichkeiten.

[00:32:29.500] - Steffen Wenzel

Was sie übrigens nicht tun.

[00:32:30.840] - Prof. Dr. Frank Schönefeld

Selten tun, gebe ich recht. Wir haben versucht, die Unterschiede heraus zu kristallisieren und das bleibt ein Unterschied. Wie erfolgreich das ist und wo es adjustiert werden muss, das werden wir noch sehen. Und drittes Geschäftsmodell wären natürlich die Standardunternehmen, gegebenenfalls die Industrieunternehmen, und die müssen natürlich versuchen, ihr etabliertes Geschäftsmodell, sei es Mobilität zu verkaufen, sei es Autos zu verkaufen, sei es Maschinen zu verkaufen, auf diese neue Ebene zu heben und dort KI reinzubringen. Die kommen von einem anderen Standpunkt natürlich aus und da ist die Investitionssumme gegebenenfalls auch nicht so gewaltig. Also xAI nimmt allein 5 Milliarden in die Hand, um ein Rechenzentrum aus dem Boden zu stampfen, um seine Crok KI zu powern. Da ist natürlich die Frage: Wer soll auf Dauer sozusagen Crok? Ist das das dritte, vierte oder gar das fünfte Foundation-Modell? Wird sich das rechnen? Oder auch Oracle hat sich definitiv überhoben und die Börse hat das auch sehr schnell erkannt und hat den Kurs einfach mal halbiert. Also da nicht alles, was nach KI aussieht, glänzt auch wirklich.

[00:33:50.360] - Steffen Wenzel

Bei der Industrie verstehe ich das, KI-Geschäftsmodell. Also für die Firmen, die verkaufen KI-Industrie 5.0 und dadurch kriegen die Firmen weniger Mitarbeiter*innen, schnellere Prozesse, weniger Fehler, mehr Output. Das heißt, sie haben eine höhere Marge. Wenn sie KI einsetzen, dafür werden sie dann auch bezahlen. Verstehe ich. Bei ChatGPT oder bzw. OpenAI, du hast es eben genannt, ich habe mal ein paar Zahlen rausgesucht. Die machen, glaube ich, zwölf Milliarden Umsatz ungefähr mit GPT und haben auch ungefähr die Ausgaben, die sie dafür benötigen. Ich glaube, da sind aber nicht die jährlichen Investitionen in Neuentwicklung reingerechnet, sondern das ist null Return of Investment momentan. Siehst du ja auch kritisch noch, aber ist die Frage, wie sich das weiterentwickelt. Die Frage wird sein, und das ist das, was ich dich gerne fragen würde – zu viele Fragen gerade, aber machen wir weiter –, ist das, dass man davon ausgeht, dass es darum geht, sich jetzt mit der KI-API in gewisse Systeme einzusetzen, das jetzt

kostenlos anzubieten, auch insbesondere die Sprachmodelle, und irgendwann Lizenzen dafür zu benutzen und die Unternehmen kommen nicht mehr raus. Also ein bisschen, das mal zu problematisieren. Glaubst du, das ist ein Weg? Das ist ja so ein bisschen das Microsoft-Modell, zu sagen: „Hier habt ihr Word, aber dann bitte bleibt auch in unserem Universum.“

[00:35:16.400] - Prof. Dr. Frank Schönefeld

Dieser First-Mover-Effekt, der ist ja bei GPT auf jeden Fall vorhanden. Und klar kennen die auch die Gesetze der Netzwerkökonomie und genau denen fühlen sie sich verpflichtet. Es muss ihnen gelingen, diese Zahl der Nutzer 800 bis 900 Millionen, wie ich erwähnt hatte, und die Marktführerschaft, 68%, die müssen sie in stabile Erlösflüsse bewegen. Und da kann man durchaus ein Jahr, zwei Jahre, drei Jahre, vier Jahre, fünf Jahre, wenn man genügend Investitionen hat, auch weiter in das Wachstum dieses Netzwerks investieren. Aber irgendwann muss dann die Idee AdWords, AdSense für KI, die muss dann kommen, die muss dann greifen und muss sich monetarisieren, sonst fällt es wieder in sich zusammen.

[00:36:08.160] - Steffen Wenzel

Sind die großen Player, die Big Five, hattest du sie mal genannt, falls die überhaupt noch alle Five da sind, aber ich glaube schon.

[00:36:14.280] - Prof. Dr. Frank Schönefeld

Es waren sieben.

[00:36:15.230] - Steffen Wenzel

Ja, genau.

[00:36:16.620] - Prof. Dr. Frank Schönefeld

Die sind alle noch da.

[00:36:17.800] - Steffen Wenzel

Die sind alle noch da. Sind die too Big to Fail?

[00:36:20.900] - Prof. Dr. Frank Schönefeld

Nein, das würde ich definitiv nicht so sehen. Schon damals haben wir insbesondere Tesla als Fail-Kandidaten gesehen. Da muss man sehen, ob die neue Strategie ... Ich bin ja gar kein Autohersteller mehr. Ich bin ja mehr ein Hersteller für Autonomie, für autonomes Bewegen, für Robotik, ob das wirklich vom Markt gekauft wird. Die anderen haben nach wie vor sehr solide Marktaussagen, denen wir allen trauen können. Nvidia wird trotz Wettbewerbs eine ganz stabile Größe bleiben können. Microsoft bleibt groß. Google ist der Hecht im Karpfenteich, muss man im Moment sagen. Die haben sowohl, was die Abdeckung der KI-Wertschöpfungskette vom Chipsystem bis hoch zur Anwendung betrifft, als auch, was die Abdeckung der Multimodalität der KI betrifft, haben sie die beste Aussage. Die können Text zu Text, die können Text zu Bild, die können Text zu Video und auch umgekehrt. Das heißt, multimodal ist auch ein Faktor. Wer kann das? Und da zum Beispiel bei Anthropic doch massive Abstriche machen und da sieht man, welche tolle Wertaussage Google hat. Also bester KI-Stack von allen. Und das mit einem etablierten digitalen Geschäftsmodell, das passt halt hervorragend zusammen, auch wenn es Veränderungen bedarf und braucht. Aber die Wette, das hat auch noch das Orakel von Omaha erkannt.

[00:37:58.720] - Prof. Dr. Frank Schönefeld

Berkshire Hathaway, das Investment von Warren Buffett, war ja nicht umsonst in Google mit 5 Milliarden. Ich glaube, ein halbes Jahr später sind daraus schon zwölf geworden.

[00:38:12.400] - Steffen Wenzel

Ja, wobei sich bei Google natürlich die Frage auch noch mal stellen wird, weil da wird es besonders evident, weil wir Google kennen als Suchmaschine, wie wir Informationen bekommen und dort natürlich immer Links hatten zu den Quellen, die dann entsprechend natürlich auch davon profitiert haben. Davor gab es ein SEO, dass man dann gesagt hat: „Okay, die kommen auch auf meine Seite und besuchen die.“ Diese Quellen verschwinden jetzt aus der KI. Bislang ist die Frage, wie Sie damit weiter ...

[00:38:40.100] - Prof. Dr. Frank Schönefeld

Sie sind zurückhaltender drin.

[00:38:42.760] - Steffen Wenzel

Gut, ich erkenne sie nicht, aber sie benutzen ja Content aus bestimmten Quellen. Wie siehst du dieses ganze Henne-Ei-Problem? Also zu sagen, wo kommt denn dieser Content her? Oder ist das ein neuer Content, der dort entsteht, den die KI da zusammenbaut und was das auch für urheberrechtliche Konsequenzen hat, weil viele Medienunternehmen in Deutschland holen da schon wieder so ein bisschen das Leistungsschutzrecht hoch und sagen: „Ja, Moment mal, was benutzt ihr denn dort?“

[00:39:17.880] - Prof. Dr. Frank Schönefeld

Also klar ist, ein System zu bevorzugen, was auch offenlegt, womit es selbst trainiert worden ist. Und dann kommen ja die Antworten zuvorderst aus diesem Training, es sei denn, es ist über MCP oder RAG oder andere Mechanismen sind aktuelle Quellen mit angeschlossen. Ich bin kein Experte im Urheberrecht. Ich zitiere mal einen befreundeten Professor von der HTW, der in dem Umfeld Vorlesungen hält und er sagt: „Das ist das Ende des Urheberrechts, so wie wir es kennen.“ Punkt.

[00:39:52.960] - Steffen Wenzel

Also das heißt, wir brauchen vielleicht ein neues, aber da wollen wir jetzt auch gar nicht so tief reingehen. Ich wollte es nur mal ansprechen, weil mir das natürlich auch immer persönlich wichtig ist. Lass uns doch noch mal ein paar positive Momente bei aller Kritik auch hervorstreichen. Siehst du gute Adoptionen von KI, von künstlicher Intelligenz, in Unternehmen, wo man sagt: „Das sind Unternehmen, die sich jetzt mit künstlicher Intelligenz auch verändert haben, effektiver geworden sind?“ Das, was ich auch eben schon mal angesprochen habe.

[00:40:18.690] - Prof. Dr. Frank Schönefeld

Ja, es gibt hervorragende Beispiele und manchmal muss man gar nicht so weit schauen. Da hilft auch ein Blick im eigenen Kontext und man sieht, wie KI dort verwendet werden kann. Eines meiner Lieblingsbeispiele kommt hier aus der Region, ein kleines Softwareunternehmen. Und das hat verstanden, seinen Softwareentwicklungsprozess wirklich inzwischen KI-gestützt, sozusagen zu machen. Als Begriff wird häufig Vibe Coding dafür verwendet. Und die Kollegen sind aber in der Lage, das wirklich in jedem Projekt nach Vereinbarung mit dem Kunden und rechtlich sicher einzusetzen. Und bei denen läuft ein Bildschirm an der Wand läuft die Zahl der generierten Token aus dem KISystem durch und man kann sofort sagen, oh, im Projekt A sind letzte Woche wieder vier Millionen Output Token generiert worden und im Projekt B war nur halb so gut mit zwei Millionen. Und da merkt man natürlich, da ist KI in den Kernwertschöpfungsprozess gewandert und das ist eigentlich das Ziel, was alle anstreben müssten, anstreben sollten. Und darüber entscheidet sich, glaube ich, auch dann Erfolg oder Misserfolg. Und da gibt es Studien dazu, erst 10% der Unternehmen sind in der Lage, wirklich ihre Kernwertschöpfung durch KI zu pushen. Und das führt eben auch wieder zu einer Beharrungsvermögen, zu einer Verlangsamung und zieht den gesamten Rattenschwanz, Blase Ja, Blase Nein, nach sich.

[00:41:53.180] - Prof. Dr. Frank Schönefeld

Aber die positiven Beispiele gibt es. Softwareindustrie hatte ich gerade gesagt, aber auch im Rechtsanwaltswesen. Ein befreundeter Rechtsanwalt hat mir erzählt, wenn du inzwischen im Verkehrsrecht tätig bist, also Einsprüche gegen Verwarngelder etc. pp., alles was folgt, und du arbeitest nicht mehr, nicht mit KI, um das automatisch, den Claim entgegenzunehmen, zu bearbeiten, einzureichen, bist du verloren. Du musst das derart schon hochautomatisieren, die gesamte Aufnahme des Falls, Einreichung, nur die finale Rechtsprechung, falls es vor Gericht geht, ist dann noch ein menschlicher Akt. Alles andere passiert mit KI. Also so weit kann man es treiben und kann man kommen.

[00:42:45.260] - Steffen Wenzel

Das hatten wir übrigens mal in einem Podcast projiziert, dass wir gesagt haben, das wird sich in die Sektoren, in die einzelnen Dienstleistungsbereiche auch aufteilen und es wird dort spezifische KI-Anwendungen für zum Beispiel den Rechtsbereich geben. Und das ist ja auch dann so gekommen. Lass uns zum Ausblick auch noch mal, oder bevor wir zu dem Ausblick kommen, noch mal auf den Arbeitsmarkt schauen. Das hast du ja eben indirekt auch angesprochen. KI führt zu Optimierung. Ich habe es eben auch schon mal gesagt, im besten Fall für das Unternehmen habe ich denselben Output mit weniger Mitarbeitern und Mitarbeiterinnen. Hat natürlich extreme soziale Implikationen. Siehst du da schon eine Tendenz? Wie wird sich das gerade entwickeln?

[00:43:26.760] - Prof. Dr. Frank Schönefeld

Wir können auch da mal die Visionäre sprechen lassen, die sind schon ein bisschen gegensätzlich dort unterwegs. Also wenn wir Elon Musk folgen, dann ist Arbeit in 10 bis 20 Jahren optional, weil der Rest über Robotik und KI abgedeckt werden kann und wir sowieso in eine Ära des Energieüberflusses eintreten. Wenn wir Jensen Huang stattdessen folgen, der sagt: „Du wirst deine Arbeit nicht an KI verlieren, aber du wirst sie an Menschen verlieren, die mit KI arbeiten.“ Und das ist die These, der würde ich zumindest für die nächsten zehn Jahre den Vorteil geben. Daran glaube auch ich. Und das macht auch Sinn für uns selber in dem Umfeld. Was wir sehen, es gibt so eine Art Betroffenheitsindex für Berufe. Welche Aufgaben können von KI besonders gut schon abgedeckt werden? Und da sind überraschenderweise doch sehr viele technische Berufe auch drin, Unterstützungsberufe, Assistenzberufe. Die sind zumindest potenziell durch KI sehr, sehr gut abzubilden. Und was ist das Gegenteil davon? Das sind Berufe, die neben einer Intelligenzfähigkeit auch eine mechanische, eine haptische Fähigkeit benötigen. Der Friseur ist noch ein paar Jahre länger sicher, davon können wir ausgehen, aber auch die sozialen Berufe, die Pflege machen, die Blut abnehmen, die im medizinischen Bereich aktiv sind.

[00:45:03.000] - Prof. Dr. Frank Schönefeld

Also insofern wäre die Empfehlung: Lasst eure Kinder was machen, wo man Intelligenz braucht, aber wo auch eine gewisse Haptik und Mechanik sozusagen noch eine Rolle spielt.

[00:45:14.460] - Steffen Wenzel

Ja, und manche Prognosen kommen ja dann dann doch wie immer anders. Man hat zum Beispiel ja gesagt, Steuerberatung, das wird auf jeden Fall irgendwann mal jetzt in kürzester Zeit einer der ersten Berufe sein. Der wird komplett technologisiert, die wird eine KI machen. Da braucht es keinen mehr dafür. Das hat übrigens auch dazu geführt, dass immer weniger diese Ausbildung gemacht haben. Und ich kenne sehr viele Steuerbüros, die händeringend nach Mitarbeiterinnen und Mitarbeitern suchen und keine finden.

[00:45:43.760] - Prof. Dr. Frank Schönefeld

Ja, in meinem ist das genauso. Da hat man die Rechnung praktisch ohne das deutsche Steuerrecht gemacht.

[00:45:49.220] - Steffen Wenzel

Genau, das ist nämlich so kompliziert, dass selbst die KI es nicht schafft. Vielleicht. Du hast eben Energieüberfluss von Elon Musk noch genannt. Das ist noch mal ein Thema, was wir eben angeschnitten hatten, was mir sehr am Herzen liegt. Wir brauchen oder wir müssen auch noch mal über die Energieeffizienz reden von KI. Das ist ja immer so ein Bias, würde ich mal sagen, zwischen: Was investiere ich in dieses Rechenzentrum, damit ich überhaupt so einen Lerneffekt erzeile, diese Rechenmodelle laufen etc. pp. und dann das Heilsversprechen in Anführungszeichen: KI wird aber auch zu einer Optimierung des Energieverbrauchs führen und dementsprechend das irgendwann wieder alles reinholen. Ein Breakeven. Wann wird der kommen oder wird der jemals überhaupt da sein?

[00:46:33.400] - Prof. Dr. Frank Schönefeld

Also im Moment ist es so, dass der Energieverbrauch weltweit, der über Rechenzentren abgerufen wird, ist etwa 1,5% des Weltenergiebedarfs, geht in Rechenzentren. Durch die geplanten Investitionen der nächsten fünf Jahre würde sich der verdoppeln. Das heißt, wir würden von 3% des Weltprimärenergiebedarfs für Rechenzentren und davon der Großteil KI-gestützte oder KI-abarbeitende Rechenzentren geprägt sein. Das heißt, wir gehen da einer Verdoppelung praktisch entgegen und das können einzelne Länder besser abbilden als andere. Hatten wir vorhin schon. China kann es besser abbilden als die USA. Deutschland hat noch mehr Mühe, das energieeffizient abzubilden. Das muss man einfach so sehen. Was kann man dagegen tun? Es gibt die kleine Lösung und die größere Lösung. Die kleine Lösung ist natürlich, jede einzelne Anfrage effizienter zu bearbeiten. Da können wir mal vom Ist-Stand kurz ausgehen. Ein Prompt an Gemini abgeschickt kostet etwa 0,24 Wattstunden. Das ist ungefähr so viel wie neun Sekunden Fernsehen. Also neun Sekunden. Macht ja keiner mehr Fernsehen. Ein Prompt, neun Sekunden lineares Fernsehen, genau, was sowieso keiner mehr macht. Also können wir stattdessen doch prompten. Und das kriegt man natürlich durch Optimierung weiter runter. Die ganzen algorithmischen Verbesserungen, die ich am Anfang angesprochen habe, zahlen da drauf ein.

[00:48:15.220] - Prof. Dr. Frank Schönefeld

GPT ist nicht ganz so effizient. Die sind eher so bei 0,32 Wattstunden. Der GPT-Chef, der Sam Altman, hat auch das Wort von der Backofensekunde geprägt. Also wir heizen unseren Backofen auf 200 Grad hoch und dann ist praktisch eine Query, ein Prompt, eine Sekunde Backofen zu dieser Temperatur. Ob das genau jetzt rechnerisch stimmt, sei dahingestellt. Das war der Vergleich, den er gewährt hat.

[00:48:42.580] - Steffen Wenzel

Zehn Minuten prompten in der Zukunft oder eine Tiefkühlpizza.

[00:48:45.560] - Prof. Dr. Frank Schönefeld

Und jetzt können wir natürlich die Emissionsfrage noch dazu stellen. Das hängt natürlich dann davon ab, durch welchen Energiemix halte ich denn mein Reihenzentrum am Leben? Und das gilt wie für jeden Energieverbrauch, wie viel grüne Energie ist drin? Wie viel Kernenergie ist drin? Wie viel fossile Energie? Auch da, muss man sagen, hat Google im Moment einen Faktor Fünf Vorteil vor GPT.

[00:49:12.100] - Steffen Wenzel

Dass sie weniger Energie reinfangen.

[00:49:14.150] - Prof. Dr. Frank Schönefeld

Nein, weniger Emission. Emission. Energie 0,24 zu 0,32 pro Query.

[00:49:20.140] - Steffen Wenzel

Wie werden denn da die Atomkraftwerke bewertet, die sie gerade bauen?

[00:49:24.800] - Prof. Dr. Frank Schönefeld

Was den CO₂-Ausstoß betrifft? Positiv. Die zählen da positiv.

[00:49:29.100] - Steffen Wenzel

Das ist ja dann aber Das ist ja auch mal ein anderes Thema, was man durchaus diskutieren, zumindest diskutierbar ist.

[00:49:34.780] - Prof. Dr. Frank Schönefeld

Diskutieren, genau. So und das waren jetzt die kleinen Lösungen oder der Stand, den wir haben. Und das wird sich verbessern über weitere algorithmische Fortschritte und so weiter und so fort. Die große Lösung käme über andere systemische Ansätze und da spielen eben diese Reizworte wie Neuromorphes Computing, Spiking Neural Networks eine Rolle, die praktisch eher noch mal Gedanken, die wir aus dem menschlichen Gehirn kennen, Gedanken aus dem Gehirn kennen, also Architekturen, die wir aus dem Gehirn kennen, noch mal versuchen aufzugreifen. Das Problem mit den neuronalen Netzen ist, dass die in der gesamten Breite durchlaufen werden. Jedes Neuron muss aktiviert werden, bis am Ende dann das nächste Token generiert ist. Und bei so einem Spiking Neural Network würden nur die Pfade entlang gelaufen werden, die wirklich relevant für die Entscheidungsfindungen sind und das natürlich wesentlich energieeffizienter.

[00:50:34.350] - Steffen Wenzel

Das ist also Optimierung beim Verbrauch, den du siehst oder der passieren wird Schritt für Schritt. Wie ist das jetzt mit dem Thema, dass ich auch damit Strom insgesamt oder Energie insgesamt einsparen kann durch KI, also in der Industrie, im Zuhause, da überall, wo Energie verbraucht wird?

[00:50:52.970] - Prof. Dr. Frank Schönefeld

Das ist das große Versprechen. Das wird tatsächlich auch gemacht, dass eben durch die Fähigkeit von KI, schlechthin Optimierungen zu stützen und anzutreiben, die sehen wir natürlich in jeglichem Industrieverbrauch. Man kann das sogar gegenüberstellen, also wie viel Energie brauchen die Rechenzentren und wie viel potenzielle Energieeinsparung steht dem gegenüber und man kommt da durchaus auf einen Plus an Energieeinsparung. Aber da sind natürlich viele Annahmen noch dahinter und man kann dem Modell jetzt glauben oder man kann es auch noch Frage stellen.

[00:51:30.630] - Steffen Wenzel

Okay, das sollten wir auf jeden Fall mal weiter im Blick behalten. Vielleicht, wenn wir das noch nachreichen können, werden wir es auch in die Shownotes packen. Frank, zum Abschluss natürlich noch der Ausblick auf das Jahr 2026. Es ist ja jetzt Januar, wo wir die Folge jetzt aufnehmen. Was passiert in 2026? Welche Tendenzen siehst du?

[00:51:51.960] - Prof. Dr. Frank Schönefeld

Ja, wir haben, glaube ich, viel berührt in unserem Gespräch. Ich denke, dieser Kampf der Systeme, der wird sich fortsetzen. Es ist einfach spannend, die neuen Modelle mit ihren Leistungsfähigkeiten zu sehen. Es ist spannend, welche Modelle setzt China dann auch dagegen? Welche Algorithmen wandern aus

China in die amerikanischen Systeme? Haben wir durchaus auch viele Beispiele, dass Ideen von dort in die anderen Systeme rüber wandern. Also dieser Kampf, aber auch Austausch der Systeme, der wird sich fortsetzen. Zu Marktbereinigung hat man schon gesprochen. Ich würde nicht das Wort Blase verwenden, weil Blase und vor allem im Vergleich mit Dotcom-Blase ist unangebracht. Das stimmt von allen Fakten her nicht, aber dass zu viel Optimismus in den 5 Billionen Dollar drinsteckt, davon gehe ich auch aus. Wir haben es mit trügen Systemen in der Wirtschaft trotzdem zu tun. Bis die voll hochlaufen und alle Leistungsparameter abrufen, dauert es einfach ein paar Jahre. Der Innovationsdruck wird hochbleiben. Nach wie vor denke ich, dass die Transformer-Architektur als die prägende Architektur der großen Sprachmodelle, die wird noch Bestand haben. Ich sehe noch nicht die komplette Ablösung des Transformers. Eher so weitere schrittweise Entwicklungen, bessere Attention-Mechanismen, weitere Tokengenerierungs-Optimierungsverfahren und so weiter und so fort.

[00:53:20.860] - Prof. Dr. Frank Schönefeld

Also Innovationdruck hoch, agentische Systeme bleibt ein Megathema. Energieeffizienz haben wir adressiert. Es ist klar, wer energieeffizienter Token generiert, der hat weniger Kosten auf der Energieseite. Besorgniserregend sind vielleicht die Konturen dieser neuen Arbeitswelt. Wir erkennen schon welche. Wir wissen aber nicht, wie wir einen Übergang vom Ist zum Dann zur Zukunft gestalten sollen. Dafür sind sie auch noch zu verschwommen, diese Konturen. Aber das wird natürlich eine unheimliche Herausforderung. Davon können wir ausgehen. Und wir müssen sagen, KI ist ein mächtiges Werkzeug geworden und wer dieses Werkzeug beherrscht, der beherrscht zu einem hohen Grade die Welt.

[00:54:12.660] - Steffen Wenzel

Jetzt habe ich noch eine ganz gemeine Frage zum Abschluss. Vielleicht kann ich da ein bisschen in dein Unterbewusstsein reingucken. Was glaubst du denn, was dich überraschen wird, was du jetzt noch gar nicht weißt, was in 2026 passieren könnte im Bereich KI-Entwicklung?

[00:54:29.220] - Prof. Dr. Frank Schönefeld

Da gibt es so eine Matrix, die heißt die Known-Unknowns und am schönsten sind die Unknowns-Unknowns. Und da werden wir sehen, ob aus welchem Quadranten die Neuigkeiten auf uns einprasseln. Ich habe schon gesagt, nach wie vor ist die prägende Architektur der Transformer, wenn sich dort ein massiver Shift ergeben würde. Das würde ich für eine doch massive Überraschung halten.

[00:54:58.890] - Steffen Wenzel

Also vielleicht jetzt nicht 2026, aber vielleicht sind wir ja 2028, wenn wir dann auch noch hier hoffentlich zusammensitzen werden. So weit, dass auf einmal ein europäisches Unternehmen oben mitmischt.

[00:55:10.880] - Prof. Dr. Frank Schönefeld

Dein Wort in die Gehörgänge der Europäischen Kommission.

[00:55:15.480] - Steffen Wenzel

Vielen Dank, Frank. Es hat mir sehr viel Spaß gemacht. Schön, dass du heute hier zu Gast warst.

[00:55:20.040] - Prof. Dr. Frank Schönefeld

Ja, danke euch.

[00:55:21.380] - Steffen Wenzel

Vielen Dank, dass Sie auch zugehört haben. Ich hoffe, es hat Ihnen gefallen und Sie haben paar Takes mitgenommen für sich. Bleiben Sie uns treu, abonnieren Sie unseren Podcast. Wie gesagt in den

Shownotes noch ein paar Informationen zu diesem Thema und ansonsten freuen wir uns dann auf
dienächste Folge mit Ihnen. Bis dann. Alles Gute.