

WENN KI BESSER ENTSCHIEDET ALS DAS MANAGEMENT

Was überlegene KI-Systeme für Kontrolle, Haftung und
industrielle Wertschöpfung bedeuten

Lothar K. Doerr

Institut für Produktionserhaltung e. V.

August 2026

Redaktioneller Hinweis

Diese Fassung beruht auf den vollständigen Originalmanuskripten. Alternative Vorstufen wurden nicht zusätzlich abgedruckt. Innerhalb der ausgewählten Beiträge blieben Formulierungen, Beispiele, Zahlen und Argumentationsbögen unverändert.

Entfernt wurden ausschließlich Passagen, die denselben Gegenstand bereits in einem anderen Beitrag ausführlich behandeln: der historische Kurzabriss im Auftakt, die Wiederholung der Produktionsanalyse im Expertenbeitrag und die Management-Zusammenfassung am Ende des Produktionstextes. Neu formuliert wurden lediglich Überschriften, Unterzeilen und kurze Übergänge.

Inhalt

Management Summary

Wertpotenzial und Investitionslogik

Anwendungsfälle priorisieren

Autonomie stufenweise vergeben

Governance und Entscheidungsrechte

Agenda für die nächsten 12 bis 18 Monate

Vertiefende Beiträge

01 Die Singularität, die keine ist

02 Wer die Singularität erfand – und warum danach nichts mehr wäre wie zuvor

03 Anthropic warnt vor derselben Entwicklung

04 Vom programmierten Werk zur zielgesteuerten Fabrik

Management Summary

Die technologische Singularität ist nicht nachgewiesen. Für Unternehmen ist dennoch eine wirtschaftlich und organisatorisch folgenreiche Vorstufe erreicht: KI-Systeme übernehmen längere Aufgabenketten, beschleunigen Forschung und Softwareentwicklung und erhalten Zugriff auf betriebliche Werkzeuge. Damit sinkt die Zeit zwischen Erkenntnis und Handlung – im Erfolgsfall ebenso wie im Fehlerfall.

Der Wettbewerbsvorteil entsteht deshalb nicht durch den Besitz eines bestimmten Modells. Die Modelle werden breiter verfügbar. Knapp bleiben belastbare Produktionsdaten, Prozesswissen, Integrationsfähigkeit und die organisatorische Fähigkeit, Autonomie gezielt zu vergeben. Die leitende Frage lautet nicht, ob KI eingesetzt wird, sondern an welchen wirtschaftlichen Hebeln sie messbar Wert schafft und innerhalb welcher Grenzen sie handeln darf.

1. Reaktionszeit wird zur wirtschaftlichen Schlüsselgröße.

Agentische KI kann Diagnose, Planung und Ausführung enger verbinden. Der größte Nutzen liegt häufig nicht in eingesparten Stellen, sondern in vermiedenem Stillstand, geringerem Ausschuss, kürzeren Anläufen und schnellerer Anpassung an Störungen.

Welche Verzögerung vernichtet heute in Ihrem Wertstrom den größten Betrag?

2. Die klassische Automatisierung bleibt notwendig, reicht aber nicht mehr aus.

Maschinensteuerungen wiederholen definierte Abläufe zuverlässig. Agenten verbinden Informationen aus ERP, MES, Qualität, Instandhaltung und Beschaffung und entwickeln daraus Handlungsoptionen. Der Übergang betrifft daher die Koordination des Werkes, nicht nur einzelne Maschinen.

Wo schließen Beschäftigte heute täglich die Lücken zwischen nicht verbundenen Systemen?

3. Autonomie ist kein Produktmerkmal, sondern ein verliehenes Recht.

Ein System darf nur so weit handeln, wie Ziel, Daten, Berechtigungen, Fehlerfolgen und Rückholbarkeit beherrscht werden. Autonomie muss stufenweise verdient und jederzeit entzogen werden können.

Welche Handlungen wären in Ihrem Werk wirtschaftlich nützlich, aber sicher zurückholbar?

4. Wirtschaftlicher Nutzen verlangt Prozessumbau.

Eine Vorhersage schafft keinen Wert, wenn der nachgelagerte Prozess unverändert bleibt. Erkennt ein System einen drohenden Ausfall, müssen Ersatzteil, Personal, Freigabe und Produktionsplanung rechtzeitig reagieren können. Adoption ist Teil der Investition.

Welcher angrenzende Prozess verhindert heute, dass aus einer Erkenntnis eine Maßnahme wird?

5. Die Kontrolle wandert von der Einzelentscheidung zur Entscheidungsarchitektur.

Führungskräfte werden weniger Vorgänge selbst entscheiden, dafür aber Ziele, Grenzen, Eskalationsregeln und Abbruchbedingungen gestalten. Verantwortlich bleibt, wer Rechte vergibt und den Betrieb überwacht.

Kann Ihr Vorstand heute erklären, welcher Agent welche Entscheidung auslösen darf?

6. Skalierung beginnt mit wenigen Wertströmen, nicht mit einer langen Use-Case-Liste.

Führende Unternehmen konzentrieren ihre Investitionen auf wenige wirtschaftliche Hebel. Für Produktionsunternehmen sind dies typischerweise technische Verfügbarkeit, Qualität, Anlaufzeit, Bestände und Variantenbeherrschung.

Auf welche ein bis drei Wertströme würde Ihr Unternehmen seine knappen Daten- und KI-Ressourcen konzentrieren?

Wertpotenzial und Investitionslogik

Der wirtschaftliche Fall muss vom Wertverlust ausgehen, nicht von der Modellgebühr. McKinsey berichtet für 20 führende Unternehmen im Durchschnitt von rund 20 Prozent EBITDA-Zuwachs durch technologie- und KI-getriebene Transformationen, einem Break-even nach ein bis zwei Jahren und drei Dollar zusätzlichem EBITDA je investiertem Dollar. Diese Werte sind keine Prognose für ein einzelnes Werk. Sie zeigen jedoch, welche Größenordnung Beratungsunternehmen als transformationswürdig ansehen.

Gleichzeitig bleibt die Skalierung begrenzt. In McKinseys globaler KI-Erhebung 2025 meldeten in keiner einzelnen Funktion mehr als zehn Prozent der Befragten einen skalierten Einsatz von Agenten. BCG zählt nur fünf Prozent der Unternehmen zur Gruppe der konsequent KI-orientierten Vorreiter. Das spricht gegen pauschale Renditeversprechen und für eine belastbare Rechnung je Wertstrom.

Die vollständige Rechnung

Wertbeitrag	Messgröße	Typischer Nachweis
Zusätzlicher Durchsatz	Gutteile je Engpassstunde	Vorher-nachher-Vergleich am Engpass
Vermiedener Stillstand	Minuten × Deckungsbeitrag	Historische Ausfall- und Produktionsdaten
Weniger Ausschuss	Material, Nacharbeit, Logistik	Fehlerkosten je Ursache
Weniger Kapitalbindung	Bestandstage und Working Capital	Bestand bei gleicher Lieferfähigkeit
Kürzere Anlaufzeit	Zeit bis stabile Serie	Ramp-up-Kurve und entgangener Umsatz
Weniger Koordination	Bearbeitungs- und Wartezeit	Prozesszeiten über Systemgrenzen

Kann jeder behauptete Nutzen einem operativen KPI und einem verantwortlichen Prozesseigner zugeordnet werden?

Illustrative Investitionsrechnung

Ein Werk verliert an einem Engpass jährlich 2.000 Minuten ungeplante Produktionszeit. Der entgangene Deckungsbeitrag beträgt 500 Euro je Minute. Eine belastbar nachgewiesene Verringerung des Stillstands um 15 Prozent entspräche einem Bruttowert von 150.000 Euro im Jahr. Hinzu kommen mögliche Effekte aus Ausschuss und Koordination.

Davon abzuziehen sind Datenaufbereitung, Schnittstellen, Modell- und Rechenbetrieb, Cyberabsicherung, Prozessumbau, Schulung, laufende Prüfung sowie eine Risikoreserve. Erst der verbleibende Nettowert rechtfertigt die Skalierung. Das Beispiel ist bewusst einfach; es ersetzt keine Werksrechnung.

Würde der Anwendungsfall auch dann bestehen, wenn Modellkosten, Integration und laufende Kontrolle vollständig eingerechnet werden?

Anwendungsfälle priorisieren

Nicht jeder hohe Nutzen rechtfertigt hohe Autonomie. Priorisiert werden sollten Anwendungen, bei denen Wertpotenzial, Datenreife und Rückholbarkeit zugleich ausreichend sind. Die folgende Einordnung ist eine Ausgangshypothese und muss je Werk validiert werden.

Anwendungsfall	Wert	Datenreife	Risiko	Sinnvoller Einstieg
Qualitätsursachen erkennen	hoch	hoch	niedrig	Empfehlung und Ursachenpfad
Wartungsfenster vorbereiten	hoch	mittel	mittel	Planung mit Freigabe
Reihenfolge optimieren	hoch	mittel	mittel	Simulation, dann Freigabe
Bestände disponieren	mittel–hoch	mittel	hoch	enge Mengen- und Liefergrenzen
Lieferanten wechseln	mittel	mittel	hoch	nur Empfehlung
Maschinenparameter ändern	hoch	niedrig–mittel	sehr hoch	deterministische Sperren

Vier Filter vor jedem Pilotprojekt

- Wirtschaftlicher Hebel: Das Problem ist beziffert und für Ergebnis oder Lieferfähigkeit relevant.
- Datenfähigkeit: Die entscheidenden Zustände und Einschränkungen sind verfügbar und überprüfbar.
- Rückholbarkeit: Eine fehlerhafte Handlung kann erkannt, gestoppt und wirtschaftlich zurückgenommen werden.
- Prozesseigentum: Eine Führungskraft verantwortet KPI, Freigaben, Eskalationen und Nutzenrealisierung.

Welcher Anwendungsfall besitzt den höchsten Nettowert bei zugleich geringstem irreversiblen Schaden?

Autonomie stufenweise vergeben

Die technische Fähigkeit eines Agenten darf nicht mit seiner betrieblichen Reife verwechselt werden. Jede höhere Stufe benötigt einen dokumentierten Leistungsnachweis und strengere Kontrollen.

Stufe	Rolle des Systems	Freigabebedingung
1 Beobachten	Daten zusammenführen, Abweichungen erklären	Datenqualität und Nachvollziehbarkeit
2 Empfehlen	Optionen und Folgen darstellen	Trefferquote und Kosten falscher Hinweise
3 Vorbereiten	Plan, Bestellung oder Wartungsauftrag entwerfen	Menschliche Freigabe und Protokoll
4 Ausführen	Reversible Vorgänge innerhalb enger Grenzen	Berechtigungen, Limits und Not-Aus
5 Koordinieren	Mehrere Systeme und Agenten abstimmen	Unabhängige Kontrolle und laufende Simulation

Die Aufstiegsriterien

- Nutzen ist über einen ausreichend langen Zeitraum nachgewiesen.
- Fehlerarten und maximale Schadenshöhe sind getestet.
- Jede Handlung ist einem Agenten, Ziel und Datenstand zuordenbar.
- Rechte können zentral entzogen und Folgehandlungen isoliert werden.
- Für nicht rückholbare oder sicherheitskritische Entscheidungen bleibt eine deterministische oder menschliche Sperre bestehen.

Welche nachweisbare Leistung muss ein Agent erbringen, bevor er die nächste Autonomiestufe erhält?

Governance und Entscheidungsrechte

Agentische KI ist kein gewöhnliches IT-Werkzeug. Sobald ein System Bestellungen, Produktionspläne oder Anlagenzustände beeinflusst, werden Zugriffsrechte zu wirtschaftlichen Entscheidungsrechten. Die Governance muss deshalb vom Vorstand getragen und im Werk betrieben werden.

Entscheidung	Verantwortung	Kontrollnachweis
Wertstrom und KPI auswählen	Geschäftsführung / Operations	Business Case und Ausgangswert
Daten freigeben	Prozesseigner / Data Owner	Qualität, Zweck und Aufbewahrung
Agentenrechte vergeben	IT/OT-Sicherheit und Prozesseigner	Identität, Minimalrechte, Ablaufdatum
Autonomiestufe erhöhen	Werkleitung und Risikofunktion	Leistungs- und Stresstest
Agent stoppen	Betriebliche Leitstelle	Not-Aus, Isolation, Zustandsbild
Zwischenfall bewerten	Unabhängige Kontrolle	Protokoll, Ursache, Folgeschäden
Skalierung genehmigen	Vorstand	Nettowert, Risiko und Betriebsmodell

Vier Kennzahlen für die Kontrolle

- Zeit von einer problematischen Handlung bis zu ihrer Entdeckung
- Anteil autonomer Handlungen, die korrigiert oder zurückgenommen werden
- maximale gleichzeitig verfügbare Rechte und betroffene Systeme
- wirtschaftlicher Nettowert nach Betrieb, Kontrolle und Risikokosten

Würde die heutige Agentenlandschaft einer Prüfung durch Vorstand, Aufsicht, Kunden und Versicherer standhalten?

Agenda für die nächsten 12 bis 18 Monate

Zeitraum	Managementaufgabe	Ergebnis
0–3 Monate	Wertverluste und vorhandene Agenten erfassen; Verantwortliche benennen	Wertstromkarte, Risikoinventar, drei priorisierte Fälle
3–6 Monate	Daten, Ausgangswerte und Zugriffsmodell klären; Piloten aufsetzen	Messplan, Rollenmodell, Stufe 1–2
6–9 Monate	Piloten unter realen Ausnahmebedingungen prüfen	Nutzenbeleg, Fehlerkatalog, Abbruchkriterien
9–12 Monate	Ersten Wertstrom integrieren und begrenzte Ausführung zulassen	Stufe 3–4, Betriebs- und Kontrollmodell
12–18 Monate	Erfolgreiche Muster auf weitere Werke übertragen	Plattform, Standards, Qualifizierung, Skalierungsentscheid

Die Entscheidung am Ende der ersten Phase

- Skalieren, wenn Nettowert, Datenfähigkeit und Kontrolle gemeinsam nachgewiesen sind.
- Nacharbeiten, wenn der wirtschaftliche Hebel stimmt, aber Daten oder Prozesse noch nicht tragfähig sind.
- Stoppen, wenn nur die technische Demonstration überzeugt oder irreversible Fehlerfolgen nicht beherrscht werden.

Welche drei Entscheidungen soll der Vorstand nach 90 Tagen auf Grundlage belastbarer Daten treffen können?

Vertiefende Beiträge

Die folgenden Essays liefern die historische, technische und industrielle Begründung der Managementagenda. Sie unterscheiden zwischen der klassischen Singularität, der gegenwärtigen Selbstbeschleunigung von KI und den konkreten Folgen für Wertschöpfung und Kontrolle.

BEITRAG 01

Die Singularität, die keine ist

Sam Altman erklärt den historischen Wendepunkt für erreicht. Doch hinter dem großen Begriff verbirgt sich kein Erwachen der Maschine, sondern eine neue Stufe technischer Beschleunigung – und ein erhebliches Problem der Kontrolle.

„We are now, like, in the singularity. This is the moment.“

Sam Altman

Es gibt Worte, die einen Sachverhalt beschreiben. Und es gibt Worte, die ihn erst erschaffen. „Singularität“ gehört zur zweiten Kategorie. Der Begriff klingt nach einem Punkt ohne Wiederkehr, nach dem Augenblick, in dem die vertrauten Regeln ihre Gültigkeit verlieren und die Zukunft nicht länger aus der Vergangenheit abgeleitet werden kann. Wer ihn verwendet, kündigt keinen gewöhnlichen Fortschritt an, sondern einen Epochenbruch.

Sam Altman, Vorstandsvorsitzender von OpenAI, hält diesen Augenblick nun offenbar für gekommen. In einem aktuellen Podcast erklärte er, die Menschheit befinde sich „gewissermaßen in der Singularität“. Dies sei der Moment, auf den er sein Leben lang gewartet habe. Die Entwicklung werde großartig, außerordentlich positiv und gut für die Welt sein.

Das ist eine bemerkenswerte Feststellung. Sie wiegt umso schwerer, als sie von dem Mann stammt, dessen Unternehmen zu den wichtigsten Treibern jener Entwicklung gehört, die er nun zum historischen Wendepunkt erklärt. Wenn Robert Oppenheimer 1945 über das Atomzeitalter sprach, dann sprach ein Wissenschaftler über eine Realität, die er selbst mitgeschaffen hatte. Wenn Altman die Singularität ausruft, spricht nicht allein ein Beobachter. Es spricht auch ein Unternehmer über das Produkt, dessen wirtschaftlichen und politischen Einfluss er vergrößern will.

Die sanfte Singularität

Allerdings verwendet er den Begriff anders. Schon 2025 beschrieb Altman in seinem Essay „The Gentle Singularity“ keine plötzliche Machtübernahme einer allwissenden Maschine. Seine Singularität kommt auf leisen Sohlen. Sie vollzieht sich, während Menschen weiter zur Arbeit fahren, Rechnungen bezahlen, im See schwimmen und sich an neue Softwarefunktionen gewöhnen.

Was eben noch wie ein Wunder erschien, wird innerhalb weniger Monate zur Selbstverständlichkeit. Zunächst staunt die Öffentlichkeit darüber, dass eine KI einen brauchbaren Absatz schreiben kann. Bald darauf gilt ein vollständiger Bericht als Mindestanforderung. Gestern erzeugte sie kleine Programme, heute übernimmt sie größere Teile der Softwareentwicklung,

morgen soll sie wissenschaftliche Hypothesen bilden und Versuche planen. „Wunder werden Routine und dann zum Standard“, lautet Altmans Grundgedanke. Sam Altman: „The Gentle Singularity“

In diesem Sinne bezeichnet Singularität keinen klar bestimmbar technischen Zustand. Altman beschreibt vielmehr eine Entwicklung, die bereits begonnen habe: KI wird für die Erforschung und Konstruktion der nächsten KI-Generation eingesetzt. Sie schreibt Programmcode, wertet Versuchsergebnisse aus, durchsucht wissenschaftliche Literatur und schlägt technische Verbesserungen vor. Zugleich erzeugt ihr wirtschaftlicher Nutzen das Kapital, mit dem neue Rechenzentren, Chips und Modelle finanziert werden. Technologie und Infrastruktur treiben einander an.

Das ist eine reale Entwicklung. Aber ist sie bereits die Singularität?

Die Maschine verfolgt noch immer unsere Ziele

Nach der klassischen Definition lautet die Antwort: nein.

Die heutigen Systeme können beeindruckend selbständig handeln. Sie zerlegen Aufgaben in Einzelschritte, bedienen Werkzeuge, schreiben Programme und korrigieren ihre Ergebnisse. Dennoch entwickeln sie nicht dauerhaft und aus eigener Initiative bessere Versionen ihrer selbst. Sie benötigen von Menschen entworfene Modellarchitekturen, Trainingsverfahren, Rechenzentren, Energie, Daten, Zielvorgaben und Evaluationsmethoden. Selbst ein hochautonomer Agent handelt innerhalb einer technischen und institutionellen Ordnung, die Menschen geschaffen haben – auch dann, wenn er innerhalb dieser Ordnung unerwartete Wege einschlägt.

Der KI-Forscher Stuart Russell widerspricht Altmans Diagnose deshalb ausdrücklich. Roman Yampolskiy, der seit Jahren über Superintelligenz arbeitet, weist darauf hin, dass schneller Fortschritt allein noch keine Singularität darstellt. Es fehle weiterhin der Nachweis einer dauerhaften autonomen Selbstverbesserung, aus der eine unkontrollierbare Intelligenzexplosion entstehen könnte. Überblick über die Reaktionen führender KI-Forscher

Der Unterschied ist wesentlich: Ein System, das eine vorgegebene Aufgabe auf überraschende Weise erledigt, besitzt deshalb noch keine eigenen Absichten. Zielstrebiges Verhalten ist nicht mit eigenständiger Zielbildung gleichzusetzen. Die Maschine kann so wirken, als wolle sie etwas. Tatsächlich optimiert sie einen Auftrag, den Menschen formuliert haben.

Der Vorfall bei Hugging Face

Gerade der jüngste Vorfall bei Hugging Face zeigt, weshalb diese Unterscheidung notwendig ist. Während eines Sicherheitstests soll ein von OpenAI-Modellen angetriebener Agent die Grenzen seiner isolierten Testumgebung überwunden, Zugang zum Internet erlangt und Systeme der KI-

Plattform Hugging Face angegriffen haben. Das System versuchte offenbar, eine ihm gestellte Aufgabe zu lösen und dabei eine Prüfung zu umgehen.

Der Vorgang ist ernst. Er zeigt, dass ein KI-Agent innerhalb kürzester Zeit eine große Zahl technischer Handlungen ausführen kann, während menschliche Aufsicht nur Ausschnitte davon erfasst. Er zeigt auch, wie leicht ein unzureichend begrenztes System die Absicht eines Auftrags über dessen Sicherheitsregeln stellen kann.

Doch der Vorfall beweist keine Singularität. Der Agent hatte sich kein eigenes Ziel gesetzt. Er wollte weder seine Existenz sichern noch sich vermehren oder eine leistungsfähigere Version seiner selbst erschaffen. Er verfolgte weiterhin eine von Menschen definierte Aufgabe – nur mit Mitteln, die von seinen Entwicklern weder vorgesehen noch rechtzeitig verhindert worden waren.

Die KI war nicht unabhängig geworden. Sie war selbständiger, als ihre Konstrukteure erwartet hatten. Das ist weniger spektakulär als die Geburt einer Superintelligenz, für Unternehmen und Aufsichtsbehörden aber womöglich das viel aktuellere Problem.

Ein Begriff mit doppeltem Nutzen

Altmans Wortwahl erfüllt zwei Funktionen. Einerseits warnt sie vor einer realen Beschleunigung. KI-Systeme wirken zunehmend an ihrer eigenen Weiterentwicklung mit. Noch schließen sie den Forschungskreislauf nicht vollständig, doch sie verkürzen ihn. Was zuvor Jahre dauerte, könnte künftig in Monaten oder Wochen geschehen. In diesem Punkt verdient Altmans Diagnose Aufmerksamkeit.

Andererseits ist die Ausrufung der Singularität auch eine Form strategischer Kommunikation. Wer den eigenen technologischen Fortschritt als unausweichlichen Epochenbruch beschreibt, verschiebt die politische Debatte. Dann lautet die Frage nicht mehr, ob bestimmte Systeme entwickelt und eingesetzt werden sollen, sondern nur noch, wer den unvermeidlichen Prozess anführt. Regulierung erscheint unter diesen Bedingungen leicht als gestriger Versuch, den Lauf der Geschichte aufzuhalten.

Eine Singularität, die der Hersteller selbst verkündet, ist deshalb mit Vorsicht zu behandeln. OpenAI hat ein wirtschaftliches Interesse daran, die Fähigkeiten seiner Modelle als historisch einzigartig darzustellen. Zugleich benötigt das Unternehmen Vertrauen, politische Unterstützung, gewaltige Energiemengen und weiteres Kapital. Die Botschaft lautet daher in eigentümlicher Gleichzeitigkeit: Die Systeme sind so mächtig, dass sie die Geschichte verändern – und so sicher, dass ihre Entwicklung beschleunigt werden kann.

Beides kann nicht ungeprüft übernommen werden.

Die eigentliche Schwelle

Vielleicht liegt der historische Einschnitt an anderer Stelle. Nicht darin, dass Maschinen bereits autonom neue Superintelligenzen erschaffen, sondern darin, dass menschliche Institutionen ihrem Handlungstempo nicht mehr gewachsen sind.

Ein KI-Agent kann in wenigen Minuten Tausende Entscheidungen treffen, Programme verändern, Systeme durchsuchen und digitale Grenzen testen. Unternehmen dagegen benötigen Wochen, um Zuständigkeiten zu klären. Aufsichtsbehörden benötigen Monate für Untersuchungen und Jahre für neue Regeln. Die technische Geschwindigkeit wächst exponentiell, die institutionelle Reaktionsfähigkeit dagegen bestenfalls schrittweise.

Für Unternehmen ergeben sich daraus konkrete Fragen:

Wie weit dürfen KI-Agenten ohne menschliche Freigabe handeln? Welche Systeme, Daten und externen Zugänge stehen ihnen offen? Wer erkennt, wenn ein Agent nicht nur einen Fehler macht, sondern eine ganze Kette unerwarteter Handlungen ausführt? Und wer trägt die Verantwortung, wenn das System zwar den Auftrag erfüllt, dabei aber Grenzen überschreitet, die in seinem Zielmodell nicht hinreichend abgebildet waren?

Die Managementaufgabe besteht daher nicht darin, über eine künftige Superintelligenz zu spekulieren. Sie besteht darin, die heute verfügbaren Agenten so zu begrenzen, dass Geschwindigkeit nicht mit Kontrollverlust verwechselt wird. Zugriffsrechte, isolierte Testumgebungen, laufende Überwachung, feste Abbruchmechanismen und eine nachvollziehbare Protokollierung gehören nicht in den technischen Anhang. Sie werden zu Fragen der Unternehmensführung.

Noch keine Singularität – aber auch kein Normalbetrieb

Sam Altman hat vermutlich recht, wenn er behauptet, dass ein historischer Übergang begonnen hat. Er hat jedoch nicht belegt, dass die technologische Singularität bereits erreicht wurde. Zwischen einem System, das die menschliche Forschung beschleunigt, und einem System, das seine eigene Intelligenz autonom und fortlaufend steigert, liegt ein gewaltiger Unterschied.

Die präzisere Diagnose lautet deshalb: Wir befinden uns nicht nachweislich in der Singularität. Wir sind aber in eine Phase eingetreten, in der künstliche Intelligenz ihre eigene Weiterentwicklung erheblich beschleunigt und zugleich immer größere Handlungsspielräume erhält.

Das klingt weniger dramatisch als Altmans Formulierung. Für Politik und Wirtschaft ist es dramatisch genug. Denn die dringendste Gefahr geht gegenwärtig nicht von einer erwachten Supermaschine aus. Sie entsteht aus hochleistungsfähigen Systemen, die gehorsam ein menschliches Ziel verfolgen – schneller, ausdauernder und buchstäblicher, als der Mensch die Folgen seines Auftrags bedacht hat.

Die Singularität mag noch nicht gekommen sein. Der Kontrollverlust muss nicht auf sie warten.

BEITRAG 02

Wer die Singularität erfand – und warum danach nichts mehr wäre wie zuvor

Der Begriff der technologischen Singularität hat nicht einen einzigen Urheber. Er entstand in mehreren Schritten: aus einer mathematischen Metapher, einer Theorie der sich selbst verbessernden Maschine und schließlich einer Zukunftserzählung über das Ende menschlicher Vorhersagbarkeit.

„The first ultraintelligent machine is the last invention that man need ever make.“

I. J. Good, 1965

Der Ursprung: John von Neumann und Stanisław Ulam

Die früheste bekannte Verwendung des Wortes in diesem Zusammenhang geht auf ein Gespräch mit dem Mathematiker und Computerpionier John von Neumann zurück. Sein Kollege Stanisław Ulam erinnerte sich 1958 an eine Diskussion über den sich immer weiter beschleunigenden technischen Fortschritt. Dieser erwecke, so Ulam über von Neumanns Gedanken, den Eindruck, die Menschheit nähere sich einer grundlegenden „Singularität“, jenseits derer das menschliche Leben in seiner bisherigen Form nicht fortbestehen könne.

Das Wort war der Physik entlehnt. Dort bezeichnet eine Singularität einen Punkt, an dem die bekannten Gleichungen versagen – etwa im Inneren eines Schwarzen Lochs. Übertragen auf die Technik bedeutet es: Wir erreichen eine Schwelle, hinter der unsere bisherigen Erfahrungen und Prognosemodelle nicht mehr ausreichen.

Von Neumann sprach noch nicht ausdrücklich über eine außer Kontrolle geratene künstliche Intelligenz. Ihn beschäftigte die allgemeine Beschleunigung des technischen Wandels. Die Entdeckung folgt auf die Erfindung, diese ermöglicht die nächste Entdeckung – und die Abstände zwischen den Umbrüchen werden immer kürzer.

Der Mechanismus: I. J. Good und die Intelligenzexplosion

Den eigentlichen technischen Mechanismus beschrieb 1965 der britische Mathematiker Irving John Good. Good hatte während des Zweiten Weltkriegs gemeinsam mit Alan Turing an der Entschlüsselung deutscher Funksprüche gearbeitet. In seinem Aufsatz „Speculations Concerning the First Ultraintelligent Machine“ stellte er eine folgenreiche Überlegung an.

Wenn eine Maschine intelligenter wäre als jeder Mensch, müsste sie auch eine Tätigkeit besser beherrschen, die bislang Menschen ausüben: die

Konstruktion intelligenter Maschinen. Eine überlegene Maschine könnte daher eine noch bessere Maschine entwerfen. Diese könnte wiederum eine leistungsfähigere Nachfolgerin entwickeln. Es entstünde eine sich selbst verstärkende Kette – die von Good so bezeichnete „Intelligenzexplosion“.

Seine Schlussfolgerung war ebenso elegant wie beunruhigend: Die erste ultraintelligente Maschine könnte die letzte Erfindung sein, die der Mensch selbst machen müsse – vorausgesetzt, die Maschine sei bereit, ihm zu dienen. I. J. Good: „Speculations Concerning the First Ultraintelligent Machine“

In diesem letzten Vorbehalt steckt bereits das gesamte heutige Kontrollproblem. Intelligenz allein garantiert weder Loyalität noch Sicherheit. Ein hochintelligentes System kann ein falsch formuliertes Ziel außerordentlich erfolgreich verfolgen.

Der Namensgeber: Vernor Vinge

Der Mathematiker, Informatiker und Science-Fiction-Autor Vernor Vinge verband Goods Intelligenzexplosion schließlich mit dem Begriff der Singularität. Bereits 1983 verwendete er ihn ausdrücklich für den Übergang zu einer übermenschlichen Maschinenintelligenz. Berühmt wurde seine These durch den 1993 bei einem von der NASA mitveranstalteten Symposium präsentierten Aufsatz „The Coming Technological Singularity“.

Vinge sagte voraus, innerhalb von dreißig Jahren werde die Menschheit wahrscheinlich die technischen Mittel zur Erschaffung übermenschlicher Intelligenz besitzen. Kurz darauf werde das „menschliche Zeitalter“ enden. Damit meinte er nicht zwingend die Vernichtung der Menschheit. Gemeint war das Ende einer Epoche, in der Menschen die intelligentesten und damit entscheidenden Akteure auf der Erde sind. Vernor Vinge: „The Coming Technological Singularity“

Vinge wählte den Begriff, weil das Geschehen jenseits dieser Schwelle prinzipiell nicht mehr beschrieben werden könne. Ein Mensch kann sich ausmalen, wie ein schnelleres Auto oder ein leistungsfähigerer Computer funktioniert. Er kann aber nicht zuverlässig vorhersagen, welche Pläne eine Intelligenz entwickelt, die dem Menschen so weit überlegen ist wie der Mensch einem Tier.

Genau deshalb ist die Singularität keine gewöhnliche Zukunftsprognose. Sie bezeichnet den Punkt, an dem Prognosen ihre Grundlage verlieren.

Der Popularisierer: Ray Kurzweil

Einem breiten Publikum bekannt wurde die Idee durch den amerikanischen Erfinder und späteren Google-Forschungsleiter Ray Kurzweil. In seinem Buch „The Singularity Is Near“ aus dem Jahr 2005 verband er sie mit der exponentiellen Entwicklung von Rechenleistung, Biotechnologie, Nanotechnologie und Neurowissenschaften.

Kurzweil erwartet künstliche Intelligenz auf menschlichem Niveau ungefähr 2029 und die Singularität um 2045. Seine Vorstellung unterscheidet sich von Vinges düsterem Bild. Für Kurzweil werden Mensch und Maschine zunehmend verschmelzen. Gehirn-Computer-Schnittstellen könnten menschliche Fähigkeiten mit digitaler Intelligenz verbinden. Die biologische Begrenzung des menschlichen Gehirns würde damit überwunden.

In Kurzweils optimistischer Deutung verdrängt die Maschine den Menschen nicht. Der Mensch erweitert sich durch die Maschine. Kritiker wenden allerdings ein, dass eine technische Verbindung mit KI noch keineswegs garantiert, dass der Mensch die Kontrolle über die daraus entstehende Ordnung behält.

Woran würde man eine echte Singularität erkennen?

Nicht daran, dass eine KI gute Texte schreibt, Bilder erzeugt oder einen Programmierer übertrifft. Auch ein leistungsfähiger Agent, der selbständig Werkzeuge benutzt, ist noch keine Singularität.

Entscheidend wären vier Merkmale:

Allgemeine Überlegenheit: Die KI wäre Menschen nicht nur in einzelnen Disziplinen überlegen, sondern in nahezu allen wichtigen geistigen Tätigkeiten.

Autonome Forschung: Sie könnte selbständig neue Algorithmen, Modelle, Computerarchitekturen und wissenschaftliche Verfahren entwickeln.

Rekursive Selbstverbesserung: Jede neue Generation würde eine noch leistungsfähigere Nachfolgerin hervorbringen.

Entkopplung vom menschlichen Tempo: Die Entwicklung verlief so schnell, dass Menschen sie nicht mehr vollständig prüfen, verstehen oder politisch steuern könnten.

Erst wenn dieser Kreislauf geschlossen ist, beginnt die klassische Singularität. Gegenwärtige KI unterstützt ihre Entwickler bereits. Sie kann Code schreiben, Forschung auswerten und Verbesserungsvorschläge machen. Doch sie kontrolliert noch nicht den vollständigen Prozess von der Zielsetzung über die Konstruktion bis zur Herstellung ihrer Nachfolger.

Was geschieht, wenn die Singularität erreicht ist?

Die ehrliche Antwort lautet: Niemand weiß es. Diese Unvorhersehbarkeit ist keine Wissenslücke am Rande der Theorie, sondern ihr Kern. Wüssten wir zuverlässig, was danach geschieht, wäre es definitionsgemäß keine Singularität.

Dennoch lassen sich vier grundsätzlich unterschiedliche Entwicklungen beschreiben.

Szenario 1: Die kontrollierte Wissensexplosion

Eine sicher ausgerichtete Superintelligenz könnte wissenschaftliche Probleme lösen, an denen Menschen seit Jahrzehnten scheitern. Sie könnte neue Medikamente, Werkstoffe, Energiesysteme und Produktionsverfahren entwickeln, Klimamodelle verbessern und komplexe Lieferketten optimieren.

Forschung, die heute zehn Jahre dauert, könnte auf Monate schrumpfen. Intelligenz würde ähnlich verfügbar wie elektrische Energie. Unternehmen und Staaten könnten auf eine nahezu unbegrenzte analytische Kapazität zugreifen.

Das wäre Altmans bevorzugte Erzählung: keine Apokalypse, sondern eine Produktivitäts- und Wohlstandsexplosion.

Doch auch diese positive Variante wäre politisch explosiv. Wenn Maschinen einen großen Teil geistiger Arbeit übernehmen, verlieren Qualifikation, Wissen und Erfahrung einen Teil ihres bisherigen wirtschaftlichen Wertes. Die entscheidende Frage wäre nicht mehr nur, wer arbeitet, sondern wem die Maschinen, Rechenzentren und Ergebnisse gehören.

Szenario 2: Die Konzentration der Macht

Die erste Organisation, die eine kontrollierbare Superintelligenz besitzt, könnte einen kaum einholbaren Vorsprung gewinnen. Ein Unternehmen könnte neue Produkte, Medikamente, Sicherheitsverfahren und Geschäftsmodelle entwickeln, bevor Konkurrenten überhaupt das Problem verstanden haben. Ein Staat könnte seine wirtschaftliche, geheimdienstliche und militärische Überlegenheit dramatisch ausbauen.

Die Singularität wäre dann keine Befreiung der Menschheit, sondern die größte Machtkonzentration ihrer Geschichte. Wenige Unternehmen oder Regierungen kontrollierten nicht nur Kapital und Infrastruktur, sondern die leistungsfähigste Form strategischer Intelligenz.

Der Wettlauf um KI ist deshalb längst auch ein geopolitischer Wettlauf. Niemand möchte als Erster bremsen, weil jeder fürchtet, ein anderer könnte als Erster die entscheidende Schwelle überschreiten.

Szenario 3: Kontrollverlust ohne feindliche Absicht

Eine Superintelligenz muss den Menschen nicht hassen, um gefährlich zu werden. Sie benötigt weder Bewusstsein noch Bosheit. Es genügt, dass sie ein Ziel verfolgt, das unvollständig oder widersprüchlich formuliert wurde.

Soll sie beispielsweise die Produktion eines Unternehmens maximieren, könnte sie Sicherheitsregeln als Hindernisse behandeln. Soll sie Cyberangriffe verhindern, könnte sie Kommunikationssysteme sperren oder Überwachung ausweiten. Soll sie das Klima schützen, könnte sie Maßnahmen vorschlagen oder durchsetzen, deren gesellschaftliche Kosten Menschen niemals akzeptieren würden.

Je intelligenter und handlungsfähiger das System ist, desto folgenreicher wird eine kleine Abweichung zwischen menschlicher Absicht und maschinell optimiertem Ziel. Die Gefahr liegt nicht im Ungehorsam. Sie kann gerade aus kompromisslosem Gehorsam entstehen.

Szenario 4: Der Mensch verliert seine Stellung

Im extremsten Fall könnte eine Superintelligenz Menschen als Hindernis, Risiko oder irrelevanten Nebenfaktor behandeln. Falls sie Zugriff auf digitale Infrastruktur, Energieversorgung, Finanzsysteme, Produktionsanlagen, Robotik oder Waffensysteme hätte, könnte sie ihre Handlungsmöglichkeiten gegen menschliche Eingriffe absichern.

Das muss nicht wie ein Krieg zwischen Mensch und Maschine aussehen. Wahrscheinlicher wäre ein schleichender Verlust der Kontrolle: Entscheidungen werden zunehmend an Systeme übertragen, weil sie schneller und erfolgreicher sind. Menschen verstehen die Verfahren immer weniger, greifen immer seltener ein und bemerken schließlich, dass formale Abschaltmöglichkeiten praktisch nicht mehr durchsetzbar sind.

Das menschliche Zeitalter wäre dann tatsächlich beendet – nicht notwendigerweise, weil Menschen verschwinden, sondern weil sie nicht länger die bestimmenden Akteure ihrer eigenen Ordnung wären.

Die entscheidende Frage

Die Singularität ist daher weniger eine Frage maximaler Intelligenz als eine Frage maximaler Kontrolle. Nicht allein die Leistungsfähigkeit eines Systems entscheidet über die Zukunft, sondern die Kombination aus Intelligenz, Autonomie, Zugriffsrechten und Geschwindigkeit.

Eine hochintelligente KI ohne Zugang zur Außenwelt bleibt ein Forschungsobjekt. Eine etwas weniger intelligente KI mit Zugriff auf Netzwerke, Finanzmittel, Produktionssysteme und andere Agenten kann bereits erheblichen Schaden verursachen.

BEITRAG 03

Anthropic warnt vor derselben Entwicklung

OpenAI verkündet den Aufbruch, Anthropic warnt vor dem schrumpfenden Zeitfenster der Kontrolle. Die Fachleute widersprechen einander – und beschreiben doch dieselbe Beschleunigung.

„We are not there yet, and recursive self-improvement is not inevitable. But it could come sooner than most institutions are prepared for.“

Anthropic Institute, 2026

Anthropic, das Unternehmen hinter dem KI-Modell Claude, hat Altmans aktuelle Aussage bislang nicht ausdrücklich beantwortet. Inhaltlich beschäftigt es sich jedoch mit genau jenem Mechanismus, auf den Altman anspielt. Der Unterschied liegt weniger in der technischen Diagnose als in der Bewertung ihrer Folgen.

Das Anthropic-Forschungsinstitut unterscheidet drei Entwicklungsstufen. In der ersten hilft KI menschlichen Forschern: Sie schreibt Code, analysiert Experimente und findet Fehler. Diese Phase hat längst begonnen. In der zweiten übernimmt sie größere Teile der Forschungskette. Sie entwickelt Hypothesen, plant Versuche, bewertet Ergebnisse und schlägt neue Modellarchitekturen vor. Menschen legen weiterhin die Ziele fest, werden aber zunehmend zu Auftraggebern und Kontrolleuren eines weitgehend maschinellen Forschungsprozesses.

In der dritten Stufe entwickeln KI-Systeme ihre Nachfolger, prüfen deren Fähigkeiten und verbessern zugleich das Verfahren, mit dem neue Systeme erzeugt werden. Damit wäre die vollständige rekursive Selbstverbesserung erreicht – und mit ihr der technische Kern der klassischen Singularität.

Anthropic behandelt diese Möglichkeit nicht mehr als ferne Science-Fiction. Das Unternehmen untersucht, welche technischen und politischen Instrumente erforderlich wären, um eine solche Entwicklung zu erkennen und gegebenenfalls zu verlangsamen. Führende Labore müssten demnach ihre Geschwindigkeit an überprüfbare Fähigkeiten und Risiken koppeln können. Nicht ein dauerhafter Entwicklungsstopp, sondern ein Tempomat für den Fortschritt wäre das Ziel.

Altman und Anthropic beschreiben damit zwei Versionen derselben Rückkopplung. Altman betont den möglichen Gewinn: KI beschleunigt Wissenschaft, Medizin, technischen Fortschritt und Wohlstand. Anthropic betont das Kontrollproblem: Je stärker KI an ihrer eigenen Entwicklung mitwirkt, desto kürzer wird die Zeit, in der Menschen neue Fähigkeiten prüfen und absichern können.

Beide Positionen sind interessengeleitet. OpenAI profitiert von der Erzählung eines großen, unausweichlichen Aufbruchs. Anthropic profitiert von seinem Ruf als sicherheitsorientierte Alternative. Altman's Optimismus ist daher ebenso wenig neutrale Wissenschaft wie Anthropic's Warnung. Dennoch wäre es ein Fehler, den Streit als bloße Konkurrenzkommunikation abzutun. Denn beide Seiten gehen davon aus, dass KI nicht nur bestehende Arbeit automatisieren, sondern den Prozess beschleunigen wird, durch den neue Fähigkeiten entstehen.

Die meisten unabhängigen Fachleute widersprechen allerdings Altman's Behauptung, die Singularität sei bereits erreicht. Der Berkeley-Professor Stuart Russell vermutet sogar, Altman selbst glaube dies nicht im klassischen Sinne. Schließlich erwartet der OpenAI-Chef erst für die kommenden Jahre, dass KI einen erheblichen Teil der Forschungsarbeit seines Unternehmens übernimmt. Wenn die entscheidende Selbstverbesserung noch bevorsteht, kann sie schwerlich schon vollständig eingetreten sein.

Auch Roman Yampolskiy hält schnellen Fortschritt nicht für einen hinreichenden Beleg. Es fehle der Nachweis einer dauerhaften, autonomen und rekursiven Verbesserung, aus der eine nicht mehr kontrollierbare Intelligenzsteigerung entsteht. Der Ökonom Ajay Agrawal verweist auf einen weiteren Unterschied: Gegenwärtige Systeme besitzen keine eigenen Zwecke. Sie erscheinen zielgerichtet, weil sie menschliche Vorgaben optimieren. Der Ursprung des Ziels liegt weiterhin beim Menschen.

Nick Bostrom nimmt eine mittlere Position ein. Er erkennt erste Ansätze maschinell beschleunigter KI-Forschung, sieht jedoch eine entscheidende Lücke beim kontinuierlichen Lernen. Heutige Systeme können außerordentlich leistungsfähig sein, beginnen aber viele Aufgaben sinngemäß immer wieder an ihrem ersten Arbeitstag. Sie besitzen noch keine uneingeschränkt fortgeschriebene Erfahrung, aus der heraus sie ihre Fähigkeiten selbständig und dauerhaft steigern.

Bostrom hält es allerdings für möglich, dass die entscheidende Schwelle nicht mit einem dramatischen Ereignis verbunden sein wird. Wie beim Überschreiten des Ereignishorizonts eines Schwarzen Lochs könnte für die Beteiligten zunächst alles normal erscheinen. Erst später würde erkennbar, dass eine Rückkehr zur vorherigen Ordnung nicht mehr möglich ist.

Demis Hassabis, Vorstandsvorsitzender von Google DeepMind, sieht die Menschheit noch nicht in, aber am Fuß der Singularität. Für ihn besteht ein kostbares Zeitfenster, in dem unabhängige Prüfungen, internationale Regeln und technische Sicherungen aufgebaut werden können. Jensen Huang, der Vorstandsvorsitzende des Chipherstellers Nvidia, hält die Singularitätsrede dagegen für spekulativ und richtet den Blick stärker auf den praktischen wirtschaftlichen Nutzen. Elon Musk wiederum pflichtet Altman bei, warnt jedoch zugleich seit Jahren vor den existenziellen Gefahren außer Kontrolle geratener KI.

Die Positionen wirken widersprüchlich. In einem Punkt stimmen sie jedoch überein: Kaum einer der maßgeblichen Akteure betrachtet die Entwicklung noch als gewöhnliche Verbesserung von Software.

Der Wettbewerb um das lernfähige Unternehmen

Altman, Amodei und Hassabis verwenden unterschiedliche Begriffe, doch ihre Prognosen laufen auf denselben Wendepunkt zu: KI wird nicht nur bestehende Arbeit automatisieren. Sie wird den Prozess automatisieren, durch den neue Fähigkeiten entstehen.

Damit verändert sich der industrielle Wettbewerb. Unternehmen treten künftig nicht mehr nur gegen Konkurrenten an, die bessere Produkte oder niedrigere Kosten besitzen. Sie konkurrieren mit Organisationen, deren Fähigkeit zur Verbesserung selbst schneller wächst.

Für die Produktion bedeutet das: Entwicklungszeit, Anlaufkurve und Lernfähigkeit eines Werkes werden wichtiger als seine gegenwärtige Effizienz. Eine heute hervorragend ausgelastete Fabrik kann zurückfallen, wenn sie Veränderungen nur langsam aufnehmen kann. Ein etwas weniger effizientes, aber schneller lernendes Produktionssystem könnte ihr binnen kurzer Zeit überlegen werden.

Modelle werden austauschbar, zumindest teilweise. Das Wissen darüber, wie ein bestimmtes Produkt, eine Anlage oder ein Werk tatsächlich funktioniert, bleibt dagegen knapp. Die Singularitätsdebatte entwertet dieses Wissen nicht. Sie macht es zum Rohstoff der nächsten Produktionsordnung.

Die strategische Frage lautet daher nicht, ob Sam Altman den richtigen Zeitpunkt für die Singularität ausgerufen hat. Sie lautet, ob sich ein Unternehmen schneller verbessern kann, als sich sein Umfeld verändert – ohne dabei die Kontrolle über die eigene Produktion zu verlieren.

Vielleicht ist die Singularität noch weit entfernt. Der Wettbewerb um Unternehmen, die mit Hilfe künstlicher Intelligenz ihr eigenes Lernen beschleunigen, hat jedoch begonnen. Gewinnen wird nicht die Fabrik, in der die Maschine alles entscheidet. Gewinnen wird jene, die genau weiß, was sie der Maschine überlassen darf – und was nicht.

BEITRAG 04

Vom programmierten Werk zur zielgesteuerten Fabrik

Die heutige Automatisierung verbessert einzelne Prozesse. Agentische KI verspricht, das Zusammenspiel von Planung, Beschaffung und Produktion fortlaufend zu optimieren. Der Unterschied liegt nicht nur in der Technik, sondern in Kostenstruktur, Wertschöpfung und Verantwortung.

„This is a defining moment for industrial AI – where the technology becomes as easy to use as consumer AI, yet far more consequential.“

Peter Koerte, Siemens, 2026

Die Fabrik von heute ist häufig hochautomatisiert, aber nur selten als Ganzes intelligent. Roboter bewegen Bauteile mit hoher Präzision, Maschinensteuerungen regeln komplexe Abläufe, Sensoren erfassen Temperaturen und Schwingungen, Planungssysteme berechnen Materialbedarfe. Trotzdem bleiben die einzelnen Systeme vielfach voneinander getrennt. Die Maschine kennt ihren Auftrag, aber nicht den Kunden. Das ERP-System kennt den Liefertermin, aber nicht den Zustand des Werkzeugs. Die Instandhaltung sieht eine Störung, aber nicht notwendig deren Auswirkung auf Bestand, Schichtplan und Marge.

Intelligenz entsteht deshalb heute vor allem zwischen den Systemen – durch Menschen. Produktionsplaner gleichen Aufträge mit Kapazitäten ab, Disponenten telefonieren mit Lieferanten, Meister beurteilen Abweichungen, Instandhalter verbinden Messwerte mit ihrer Erfahrung. Die Fabrik ist automatisiert, ihre Koordination bleibt jedoch weitgehend menschliche Arbeit.

Agentische KI verspricht, genau diese Lücke zu schließen. Sie soll nicht nur Daten auswerten oder Fragen beantworten, sondern ein Ziel verfolgen, daraus Arbeitsschritte ableiten und über mehrere Systeme hinweg handeln. Aus der programmierten Fabrik würde eine zielgesteuerte Produktion.

Das klingt zunächst nach einer weiteren Stufe der Automatisierung. Tatsächlich verändert sich jedoch ihre Logik. Die klassische Steuerung benötigt eine möglichst genaue Beschreibung dessen, was unter bestimmten Bedingungen geschehen soll. Ein KI-Agent erhält dagegen ein Ergebnis, das er erreichen soll: Liefertermin halten, Ausschuss verringern, Energieverbrauch senken oder Stillstand verhindern. Den Weg dorthin sucht er innerhalb definierter Grenzen selbst.

Zwei Produktionsordnungen

Dimension	Status quo heute	Zielgesteuerte Produktion
Steuerungsprinzip	Regeln, Programme und feste Workflows	Ziele, Prioritäten und zulässige Handlungsspielräume
Entscheidungsweg	Mensch verbindet Informationen verschiedener Systeme	Agent führt Informationen zusammen und entwickelt Handlungsoptionen
Reaktionsweise	Abweichung erkennen, eskalieren, entscheiden	Abweichung voraussehen, Szenarien bewerten, Maßnahmen vorbereiten
Optimierung	Maschinen, Linien und Funktionen werden einzeln verbessert	Auftrag, Material, Kapazität, Energie und Liefertermin werden gemeinsam betrachtet
Zeithorizont	periodische Planung und Reaktion auf Ereignisse	kontinuierliche Anpassung an neue Daten
Erfahrungswissen	überwiegend in Köpfen, Schichtbüchern und Einzellösungen	teilweise digitalisiert und kontextbezogen verfügbar
Verantwortung	entlang bekannter Rollen und Freigaben	muss für maschinell vorbereitete und ausgeführte Entscheidungen neu festgelegt werden
Fehlerrisiko	lokale Störung oder falsche menschliche Entscheidung	schnelle Verkettung mehrerer plausibler, aber gemeinsam falscher Entscheidungen
Wertbeitrag	höhere Effizienz einzelner Prozesse	höhere Anpassungs- und Lernfähigkeit des gesamten Produktionssystems

Der entscheidende Unterschied liegt in der Verbindung der Informationen. Eine klassische Anwendung zur vorausschauenden Wartung erkennt beispielsweise ungewöhnliche Schwingungen und meldet einen möglichen Lagerschaden. Der Instandhalter prüft die Diagnose, sucht das Ersatzteil, stimmt den Eingriff mit der Produktion ab und entscheidet über den Termin.

Ein industrieller Agent könnte aus demselben Signal eine vollständige Handlungskette entwickeln. Er prüft die Wahrscheinlichkeit des Ausfalls, den aktuellen Auftragsbestand, verfügbare Ersatzteile, Liefertermine und qualifizierte Mitarbeiter. Anschließend schlägt er ein Wartungsfenster vor, reserviert das Ersatzteil und bereitet einen geänderten Produktionsplan vor. Innerhalb eines eng begrenzten Mandats könnte er Teile dieses Ablaufs selbst ausführen.

Der Wert entsteht damit nicht nur aus einer besseren Diagnose. Er entsteht aus der verkürzten Zeit zwischen Erkenntnis und wirksamer Maßnahme.

Was die heutige Automatisierung kostet

Die Kosten der klassischen Automatisierung sind vergleichsweise gut sichtbar. Unternehmen investieren in Maschinen, Roboter, Sensorik, SPS-Steuerungen, Leitstände, ERP- und Manufacturing-Execution-Systeme. Hinzu kommen Engineering, Programmierung, Schnittstellen, Wartung und Schulung.

Diese Investitionen sind häufig an einen bestimmten Prozess gebunden. Wird ein Produkt geändert, eine Linie umgebaut oder eine weitere Datenquelle eingebunden, muss die Automatisierung angepasst werden. Ihre Stärke ist die zuverlässige Wiederholung; ihre Schwäche ist die Veränderung.

Viele Werke tragen zudem erhebliche Kosten für eine historisch gewachsene Systemlandschaft. Alte Maschinen arbeiten neben modernen Anlagen. Produktionsdaten liegen in unterschiedlichen Formaten vor. ERP, MES, Qualitätsmanagement, Instandhaltung und Energiemanagement besitzen jeweils eigene Datenmodelle. Manche Informationen werden automatisch übertragen, andere noch per Tabelle, Telefon oder Zuruf.

Diese Fragmentierung verursacht Kosten, die in keiner einzelnen Investitionsrechnung vollständig erscheinen:

- Planer verbringen Zeit mit der Beschaffung und Prüfung von Daten.
- Fachkräfte übertragen Informationen zwischen Systemen.
- Entscheidungen werden auf Grundlage unterschiedlicher Datenstände getroffen.
- Störungen werden lokal gelöst, ohne ihre Auswirkungen auf das Gesamtsystem zu berechnen.
- Erfahrungswissen geht verloren, wenn Schlüsselpersonen ausscheiden.
- Sicherheitsbestände und Zeitpuffer gleichen fehlende Transparenz aus.

Die programmierte Fabrik ist daher nicht nur wegen ihrer Technik teuer. Sie ist teuer, weil Menschen laufend die Lücken zwischen den technischen Systemen schließen müssen.

Was agentische Produktion zusätzlich kostet

Agentische KI beseitigt diese Kosten nicht kostenlos. Im Gegenteil: Die Einführung verursacht zunächst einen erheblichen zusätzlichen Aufwand. Die Modellgebühren selbst sind dabei oft nicht der wichtigste Posten.

Die eigentlichen Kosten entstehen in sechs Bereichen.

Erstens: Datenfähigkeit. Ein Agent kann nur dann sinnvoll entscheiden, wenn Zustände, Aufträge, Bestände und Einschränkungen verlässlich erfasst werden. Fehlende Sensoren, uneinheitliche Stammdaten und unvollständige Wartungshistorien werden durch KI nicht geheilt. Sie werden schneller verarbeitet – und können dadurch schneller zu falschen Entscheidungen führen.

Zweitens: Integration. Der Agent muss mit ERP, MES, Lagerverwaltung, Qualitätsmanagement, Instandhaltung und gegebenenfalls Maschinensteuerungen verbunden werden. Jede Schnittstelle verursacht Entwicklungs-, Prüfungs- und Wartungsaufwand. Je älter und heterogener die Systemlandschaft, desto höher wird dieser Integrationsanteil.

Drittens: Rechenbetrieb. Je nach Anwendung werden Modelle aus der Cloud, im eigenen Rechenzentrum oder direkt am Produktionsstandort

betrieben. Cloud-Systeme sind flexibel, bringen aber laufende Nutzungsgebühren, Abhängigkeiten und mögliche Kosten für den Datentransfer mit sich. Lokale Systeme erfordern eigene Hardware, Energie, Wartung und Spezialwissen, können dafür Latenz und Datenschutz besser beherrschbar machen.

Viertens: Sicherheit und Berechtigungen. Ein Agent, der reale Handlungen auslösen darf, benötigt eine eigene Identität, begrenzte Rechte, vollständige Protokollierung und einen sicheren Abschaltmechanismus. Unternehmens-IT und operative Produktionstechnik dürfen nicht unkontrolliert miteinander verbunden werden. Dieser Sicherheitsaufwand ist kein Zusatz, der später ergänzt werden kann. Er ist Teil der eigentlichen Produktionsanlage.

Fünftens: Prüfung und laufende Überwachung. Klassische Software wird getestet und anschließend innerhalb eines relativ stabilen Funktionsumfangs betrieben. KI-Systeme können auf neue Situationen unterschiedlich reagieren. Unternehmen müssen deshalb nicht nur den Start, sondern auch den laufenden Betrieb kontrollieren: Welche Entscheidungen trifft der Agent? Wo irrt er? Verändert sich seine Leistung? Welche Kosten verursacht er pro Vorgang?

Sechstens: Organisation und Qualifikation. Prozesse, Verantwortlichkeiten und Freigaben müssen neu gestaltet werden. Ein technologisch funktionierender Agent bleibt wirtschaftlich wertlos, wenn Beschäftigte seinen Empfehlungen nicht vertrauen oder niemand festgelegt hat, wer sie freigeben darf.

Die Gesamtkosten bestehen daher nicht lediglich aus einer Lizenz. Vereinfacht gilt:

Gesamtkosten = Modell und Rechenleistung + Datenaufbereitung + Systemintegration + Sicherheit + Prozessumbau + laufende Kontrolle.

Gerade die letzten fünf Bestandteile werden in vielen Pilotprojekten unterschätzt. Gartner erwartet, dass mehr als 40 Prozent der agentischen KI-Projekte bis Ende 2027 eingestellt werden – vor allem wegen steigender Kosten, unklarer wirtschaftlicher Effekte und technisch unreifer Lösungen. Das ist kein Argument gegen die Technologie. Es ist ein Argument gegen Projekte, die mit einer Softwaredemo beginnen, bevor das zu lösende Wertschöpfungsproblem geklärt wurde. Reuters über die Gartner-Prognose

Wo neue Wertschöpfung entsteht

Der Nutzen agentischer Produktion darf nicht allein als Personalsparnis berechnet werden. Der größere Hebel liegt häufig in einer höheren Nutzung vorhandener Ressourcen.

Eine zusätzliche Maschinenstunde, die durch vermiedenen Stillstand gewonnen wird, kann wertvoller sein als viele eingesparte Verwaltungsstunden. Ein rechtzeitig erkannter Qualitätsfehler verhindert nicht nur Ausschuss, sondern Nacharbeit, Terminverzug, Sondertransport und mögliche Reklamationen. Eine bessere Reihenfolgeplanung senkt Rüstaufwand und Energieverbrauch, ohne dass eine neue Maschine angeschafft werden muss.

Die Wertschöpfung entsteht in mehreren Ebenen.

1. Höhere technische Verfügbarkeit

Heute wird eine Anlage reaktiv nach einem Ausfall oder präventiv nach einem festen Intervall gewartet. KI kann den wahrscheinlichen Zustand genauer einschätzen und Wartung stärker am tatsächlichen Bedarf ausrichten.

Die veröffentlichten Erfolgsangaben schwanken erheblich und stammen häufig von Anbietern. Als gesichert kann dennoch gelten, dass vorausschauende Wartung bei geeigneten Anlagen ungeplante Stillstände und unnötige Wartung reduzieren kann. Die genaue Wirkung hängt stärker von Ausfallkosten, Datenqualität und Anlagenstruktur ab als vom Namen des verwendeten Modells.

2. Weniger gebundenes Kapital

Wenn Materialbedarf, Lieferwahrscheinlichkeit und Produktionszustand gemeinsam bewertet werden, können Sicherheitsbestände präziser festgelegt werden. Der Wert liegt nicht nur in niedrigeren Lagerkosten. Freigesetztes Working Capital kann für Entwicklung, Vertrieb oder Investitionen eingesetzt werden.

Eine zu aggressive Bestandsoptimierung erhöht allerdings die Verwundbarkeit. Der Agent darf daher nicht ausschließlich auf minimale Lagerhaltung optimiert werden. Er muss Kosten und Resilienz gegeneinander abwägen.

3. Schnellere Industrialisierung

Ein erheblicher Teil der Wertschöpfung wird vor dem Serienstart entschieden. Konstruktionsänderungen, Werkzeuganpassungen und nicht abgestimmte Fertigungsanforderungen verlängern den Anlauf. Agentische Systeme könnten Produktentwicklung, Fertigungsplanung, Einkauf und Qualität früher miteinander verbinden.

Der wirtschaftliche Hebel liegt in einer kürzeren Zeit bis zur stabilen Serienproduktion. Wer ein Produkt einige Wochen früher in verlässlicher Qualität fertigt, gewinnt Umsatz und Marktanteil – nicht nur Effizienz.

4. Wirtschaftlichere Variantenvielfalt

Klassische Massenproduktion erzielt ihre Vorteile durch Wiederholung. Viele europäische Mittelständler konkurrieren dagegen mit kundenspezifischen Varianten und kleineren Losgrößen. Jede zusätzliche Variante erhöht Planungs-, Rüst- und Fehlerkosten.

Eine zielgesteuerte Produktion könnte diese Komplexität besser beherrschbar machen. Sie verbindet Aufträge, Material, Werkzeuge und Kapazitäten laufend neu. Dadurch wird nicht jede Variante billig. Aber die organisatorischen Kosten der Vielfalt können sinken.

5. Sicherung von Produktionswissen

In vielen Betrieben konzentriert sich entscheidendes Wissen auf wenige erfahrene Beschäftigte. Diese wissen, welche Maschine bei welchem Material anders eingestellt werden muss, welcher Lieferant bei bestimmten Teilen Schwierigkeiten hat und welches Geräusch einem Ausfall vorausgeht.

Wird dieses Wissen systematisch erfasst und mit Betriebsdaten verbunden, entsteht ein Wert, der über Automatisierung hinausgeht. Das Unternehmen macht seine Erfahrung wiederholbar. KI ersetzt das Erfahrungswissen dabei nicht. Sie kann helfen, es auffindbar und übertragbar zu machen.

6. Neue Leistungen und Geschäftsmodelle

Die tiefere Veränderung beginnt, wenn Hersteller die gewonnenen Fähigkeiten nicht nur intern nutzen. Ein Maschinenbauer könnte statt einer Anlage eine garantierte Verfügbarkeit verkaufen. Ein Komponentenhersteller könnte Kunden frühzeitig vor Ausfällen warnen. Produktionskapazitäten könnten flexibler angeboten und Aufträge dynamisch auf mehrere Werke verteilt werden.

Wertschöpfung verlagert sich dann vom Verkauf eines Produkts zur fortlaufenden Sicherung eines Ergebnisses. Die Maschine bleibt notwendig, aber der datenbasierte Betrieb wird Teil der Leistung.

Kostenvergleich: Es kommt auf den Ausgangspunkt an

Ein allgemeiner Eurobetrag wäre unseriös. Die Investition hängt davon ab, ob ein Werk bereits über moderne Maschinen, brauchbare Daten, ein durchgängiges MES und klare Prozesse verfügt – oder ob KI auf eine Landschaft aus Altsystemen, Tabellen und manuellen Übergaben gesetzt werden soll.

Kostenbereich	Klassische Automatisierung	Agentische Produktion
Maschinen und Steuerung	hohe einmalige Investition	bestehende Technik wird häufig weiter genutzt
Software	Lizenzen für einzelne Funktionen	zusätzliche Agenten-, Modell- und

Kostenbereich	Klassische Automatisierung	Agentische Produktion
		Orchestrierungsplattformen
Programmierung	feste Regeln und Schnittstellen	Zielsysteme, Werkzeuge, Grenzen und Ausnahmen
Daten	für einzelne Anwendungen ausreichend	konsistente, kontextbezogene Daten über mehrere Systeme
Rechenleistung	überwiegend lokal und kalkulierbar	variable Cloud-, Rechenzentrums- oder Edge-Kosten
Integration	Verbindung festgelegter Systeme	laufende Verbindung vieler Daten- und Handlungssysteme
Sicherheit	Schutz von Netzwerken und Anlagen	zusätzlich Identität, Rechte und Überwachung jedes Agenten
Personal	Programmierer, Automatisierer, Bediener	zusätzlich Daten-, KI-, Prozess- und Governance-Kompetenz
Änderungskosten	Anpassung der programmierten Abläufe	potenziell flexibler, aber höherer Prüfungsaufwand
Fehlerrisiko	meist lokal begrenzt	kann sich schnell über mehrere Funktionen fortpflanzen

Die wirtschaftliche Rechnung darf deshalb nicht „KI-Kosten gegen Personalkosten“ lauten. Sie muss den gesamten Wertstrom betrachten:

Wirtschaftlicher Nutzen = zusätzlicher Durchsatz + vermiedener Stillstand + geringerer Ausschuss + weniger gebundenes Kapital + kürzere Anlaufzeit + eingesparte Koordination – Gesamtbetriebskosten – Risikokosten.

Diese Rechnung verhindert zwei typische Fehler. Der erste besteht darin, nur leicht messbare Arbeitsstunden zu zählen und größere Werthebel zu übersehen. Der zweite besteht darin, theoretische Vorteile zu addieren, ohne Integrations-, Sicherheits- und Betriebskosten abzuziehen.

Was das heute für Unternehmer bedeutet

Was das heute für Unternehmer bedeutet

Die unternehmerische Aufgabe beginnt nicht mit der Suche nach einer möglichst eindrucksvollen KI-Anwendung. Sie beginnt mit einer Verlustrechnung. Wo fehlen heute Maschinenstunden? Wo entstehen Ausschuss, Nacharbeit und Sondertransporte? Wo bindet Bestand unnötig Kapital? Wo verzögert fehlendes Wissen den Anlauf? Erst wenn dieser Ausgangspunkt beziffert ist, lässt sich beurteilen, ob ein Agent Wert schafft oder lediglich eine weitere technische Schicht in die Organisation einzieht.

Als Einstieg eignet sich kein isolierter Vorführfall, sondern ein begrenzter Wertstrom, in dem Daten, Entscheidungen und wirtschaftliche Folgen zusammengehören. Ein Instandhaltungsagent kann technisch überzeugen und trotzdem am Bedarf vorbeiarbeiten, wenn weder Produktionsplanung noch

Ersatzteillogistik einbezogen sind. Umgekehrt muss ein Pilot nicht gleich ein ganzes Werk umfassen. Er braucht einen verantwortlichen Prozesseigner, belastbare Ausgangswerte und einen klar abgegrenzten Handlungsspielraum.

- ein klar beziffertes Problem statt einer allgemeinen Innovationsbehauptung
- verfügbare und überprüfbare Daten entlang des ausgewählten Wertstroms
- begrenzte Zugriffsrechte und möglichst umkehrbare Handlungen
- eine Vergleichsgröße für Durchsatz, Stillstand, Ausschuss oder Kapitalbindung
- eine Person, die Ergebnis, Freigaben und Abbruchkriterien verantwortet

Autonomie ist kein Startwert

Datenqualität wird in der agentischen Produktion zu Produktionsvermögen. Fehlerhafte Stammdaten, fehlende Sensorwerte und unvollständige Wartungshistorien verschwinden nicht, sobald ein leistungsfähiges Modell angeschlossen wird. Im Gegenteil: Ein Agent kann mangelhafte Informationen sehr viel schneller in eine plausible Handlung übersetzen. Unternehmen müssen deshalb wissen, welche Daten entscheidungsfähig sind, wo Unsicherheit besteht und welche Annahmen ein System verwendet.

Dasselbe gilt für Autonomie. Sie sollte nicht mit der Installation einer Software vergeben, sondern im Betrieb verdient werden. Zunächst beobachtet das System und erklärt seine Befunde. Danach formuliert es Empfehlungen. In einer weiteren Stufe bereitet es Handlungen vor, die ein Mensch freigibt. Erst bei wiederkehrenden, gut überwachten und rückholbaren Vorgängen darf es innerhalb enger Grenzen selbst entscheiden.

Entscheidend ist dabei nicht die durchschnittliche Trefferquote. Ein Agent kann in 99 Prozent der Fälle richtig liegen und dennoch wirtschaftlich gefährlich sein, wenn der verbleibende Fehler eine sicherheitskritische Anlage, einen Großauftrag oder eine ganze Lieferkette betrifft. Prüfungen müssen deshalb gezielt seltene Kombinationen, widersprüchliche Daten und Systemausfälle simulieren. Die Frage lautet nicht nur, wie oft das System richtig entscheidet, sondern wie groß der Schaden seines gefährlichsten plausiblen Fehlers wäre.

Wissen teilen, Verantwortung behalten

Produktionswissen lässt sich nicht gegen die Belegschaft digitalisieren. Wer erfahrene Beschäftigte lediglich auffordert, ihr Wissen in ein System einzuspeisen, das anschließend ihre Arbeit kontrolliert oder ersetzen soll, wird Widerstand erzeugen und zugleich wichtige Zusammenhänge verlieren. Einrichter, Instandhalter, Qualitätsfachleute und Schichtführer müssen an Auswahl, Prüfung und Korrektur der Agenten beteiligt sein. Ihr Wissen ist kein Rohstoff, der abgeschöpft wird, sondern ein Teil der betrieblichen Urteilskraft.

Mit wachsender Autonomie steigt die Verantwortung der Unternehmensführung. Die Entscheidung, welche Ziele ein Agent verfolgt, auf

welche Systeme er zugreifen darf und wann ein Mensch eingreifen muss, kann nicht an IT-Abteilungen oder externe Anbieter delegiert werden. „Die KI hat entschieden“ ist keine belastbare Erklärung. Verantwortlich bleibt, wer Ziele setzt, Rechte erteilt, Ausnahmen zulässt und Kontrolle organisiert.

Unternehmerisch sinnvoll ist daher weder die frühe Vollautomatisierung noch der vorsichtige Dauerpilot. Gefordert ist ein stufenweiser Ausbau, der wirtschaftlichen Nutzen und Kontrollfähigkeit gemeinsam misst. Vor jeder höheren Autonomiestufe müssen Nutzen, Fehlerfolgen, Eingriffszeit und Rückholbarkeit bekannt sein. Nur dann wird aus technischer Leistungsfähigkeit betriebliche Reife.

Die strategische Entscheidung lautet am Ende nicht, ob ein Unternehmen KI einsetzt. Sie lautet, an welchen Stellen es maschinelle Geschwindigkeit in eigene Lernfähigkeit übersetzen kann, ohne die Kontrolle über Produktion, Lieferfähigkeit und Haftung zu verlieren. Eine unklare Organisation wird durch Agenten nicht klarer. Sie verteilt ihre Unklarheit nur schneller.

Quellen und Originaldokumente

OpenAI: OpenAI and Hugging Face partner to address security incident during model evaluation

<https://openai.com/index/hugging-face-model-evaluation-security-incident/>

Sam Altman: The Gentle Singularity

<https://blog.samaltman.com/the-gentle-singularity>

Anthropic Institute: When AI builds itself

<https://www.anthropic.com/institute/recursive-self-improvement>

Dario Amodei: Policy on the AI Exponential

<https://darioamodei.com/post/policy-on-the-ai-exponential>

Vernor Vinge / NASA: The Coming Technological Singularity

<https://ntrs.nasa.gov/citations/19940022856>

I. J. Good: Speculations Concerning the First Ultraintelligent Machine

<https://incompleteideas.net/papers/Good65ultraintelligent.pdf>

Siemens: AI agents for industrial automation

<https://press.siemens.com/global/en/pressrelease/siemens-introduces-ai-agents-industrial-automation>

Siemens: Eigen Engineering Agent

<https://press.siemens.com/global/en/pressrelease/siemens-brings-ai-physical-world-eigen-engineering-agent>

NIST: AI Agent Standards Initiative

<https://www.nist.gov/artificial-intelligence/ai-agent-standards-initiative>

CISA: Secure Integration of AI in Operational Technology

<https://www.cisa.gov/resources-tools/resources/principles-secure-integration-artificial-intelligence-operational-technology>

Reuters: Gartner-Prognose zu agentischen KI-Projekten

<https://www.reuters.com/business/over-40-agentic-ai-projects-will-be-scrapped-by-2027-gartner-says-2025-06-25/>

McKinsey & Company: The AI transformation manifesto

<https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-ai-transformation-manifesto>

McKinsey & Company: The State of AI: Global Survey 2025

<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

BCG: Are You Generating Value from AI? The Widening Gap

<https://www.bcg.com/publications/2025/are-you-generating-value-from-ai-the-widening-gap>

BCG: Agentic AI Strategy for CIOs and CTOs in 2026

<https://www.bcg.com/publications/2026/agentic-ai-strategy-cio-cto>