TUM

TECHNISCHE UNIVERSITÄT MÜNCHEN
INSTITUT FÜR INFORMATIK

AGENTONOMICS. Das Konstruktionsprinzip für selbständige KI-Agenten zur Erzielung wirtschaftlichen Wachstums

Alois C. Knoll und Hanns C. Suckfüll

TUM-I26130

Technischer Bericht
Technische Universität München
Institut für Informatik

AGENTONOMICS

Das Konstruktionsprinzip für selbständige KI-Agenten zur Erzielung wirtschaftlichen Wachstums

Alois C. Knoll und Hanns C. Suckfüll

***Abstract*—**Agentische Künstliche Intelligenz verschiebt den Gegenstand des Managements: Aus Modellen, die Antworten erzeugen, werden Systeme, die Ziele verfolgen, Werkzeuge nutzen, mit anderen Akteuren kommunizieren und innerhalb gesetzter Grenzen handeln. Agentonomics wird in diesem Beitrag als Entwurf einer Managementdisziplin für diese Entwicklung vorgestellt. Der Ansatz verbindet das St. Galler Konzept des integrierten Managements mit der Informatik kooperierender KI-Agenten und unterscheidet drei Ebenen. Das Agentonomics-Konzept liefert die normative und kognitive „DNA“: Unternehmen werden als Zusammenspiel von Außen-, Business- und Managementsystem verstanden; zentrale Aufgabe ist die dynamische Kohärenzsteuerung, unterstützt durch den rekursiven 5C-Navigator. Die Agentonomics-Theorie beschreibt KI-Agenten als funktionale ökonomische Akteure einer polyzentrischen Fähigkeitenwirtschaft, in der Eigentums-, Verfügungs- und Teilhabeansprüche an Daten und übertragbaren Kompetenzen neu geordnet werden müssen. Das Agentonomics-Modell operationalisiert diese Ideen im sechsstufigen Design & Management Reference Framework (ADMRF), in Management-Agentensystemen und in einem Führungscockpit. Dort werden interne und externe Evidenz, Szenarien, funktionale Rollenagenten, menschliche Eingriffe, Umsetzung und Lernen zu einem nachvollziehbaren Führungsregelkreis verbunden. Der Beitrag zeigt praktische Anwendungsfelder, die Evolution vom Assistenten zum funktionalen AI-CEO und eine Forschungsagenda. Agentonomics wird dabei nicht als bereits empirisch abgeschlossene Theorie, sondern als idealtypisches Forschungs- und Gestaltungsprogramm verstanden, dessen Tragfähigkeit in Lehre, Unternehmenslaboren und realen Entscheidungssituationen systematisch zu prüfen ist.

***Index Terms*—**Agentonomics, agentische KI, integriertes Management, Multi-Agenten-Systeme, hybride Führung, Fähigkeitensouveränität, Führungscockpit.

I. EINLEITUNG

Die gegenwärtige KI-Debatte ist stark modellzentriert. Sie fragt, welches Sprachmodell mehr Parameter besitzt, welche Aufgaben ein Agent selbständig ausführen kann oder wie sich einzelne Prozesse automatisieren lassen. Für die Unternehmensführung ist dies notwendig, aber nicht hinreichend. Die schwierigere Frage lautet, wie aus vielen technischen Fähigkeiten ein verantwortbares Managementsystem entsteht. Ein Unternehmen kann hundert KI-Anwendungen einsetzen und dennoch strategisch blind bleiben, wenn Daten, Ziele, Rollen, Entscheidungen und Lernschleifen nicht integriert sind. Agentonomics setzt deshalb nicht beim einzelnen Werkzeug, sondern bei der Architektur des Managements an.

Der Ausgangspunkt ist ein klassisches Managementproblem: Erfolgreiche Unternehmen müssen ihr Tagesgeschäft beherrschen und zugleich rechtzeitig erkennen, wann das bisherige Erfolgsmodell zur Fessel wird. Sie müssen unterscheiden, ob Weiterentwicklung, Transformation oder Neuerfindung angemessen ist. Agentische KI kann diesen Wendepunkt früher sichtbar machen, sie kann ihn aber nicht aus Daten allein bestimmen. Dazu braucht sie eine Denklogik, die Außenwelt, Wertschöpfung, Organisation, Werte und Umsetzung gemeinsam betrachtet. Genau diese Integrationsleistung gehört zum Kern der St. Galler Managementtradition [1]–[5].

Agentonomics überführt diese Tradition in das Zeitalter agentischer Systeme. Es ist kein weiteres Buchprojekt über die Möglichkeiten generativer KI, sondern der Entwurf eines Managementansatzes, der Führungskräfte befähigen soll, eigene Management-Agentensysteme zu konzipieren, schrittweise aufzubauen und verantwortungsvoll einzusetzen [10]. Der Begriff „KI-Agenten-Gesellschaft“ bezeichnet dabei keine bereits bestehende Rechtsordnung und keine Behauptung maschineller Personalität. Er bezeichnet eine funktionale Ordnung, in der KI-Agenten innerhalb und zwischen Organisationen wirtschaftlich relevante Rollen übernehmen, miteinander kooperieren und unter menschlichen beziehungsweise institutionellen Mandaten handeln.

Der Beitrag folgt dem Dreischritt „Konzept → Theorie → Modell“. Das Konzept formuliert die Managementlogik, die Theorie deduziert daraus Aussagen über Wertschöpfung und Ordnung einer Agentenökonomie, und das Modell macht diese Aussagen technisch sowie organisatorisch erprobbar. Ergänzt wird dieser Dreischritt um die praktische Perspektive hybrider Führung. Dieser Text verdichtet die tragenden Ideen, markiert ihren wissenschaftlichen Status und lädt dazu ein, sie in Forschung, Lehre und Unternehmenspraxis gemeinsam weiterzuentwickeln.

II. DAS AGENTONOMICS-KONZEPT

A. Vom KI-Werkzeug zum integrierten Managementsystem

Das Agentonomics-Konzept basiert auf der systemorientierten St. Galler Unternehmungslehre und besonders auf dem von Abegglen fortgeführten Konzept Integriertes Management [1], [2]. Die aktuelle Fassung des Standardwerks ergänzt die Bleicher-Matrix um den St. Galler Denk- und Wissensnavigator und das St. Galler Management HAUS. Der integrierte Bezugsrahmen wird dadurch als navigierbare Wissens- und Arbeitsarchitektur lesbar: Führungskräfte können zwischen Gesamtbild und Detail, Theorie und Anwendung sowie normativer, strategischer und operativer Ebene wechseln. Agentonomics übernimmt genau diese Doppelfunktion – stabile Management-DNA und geregelter Navigationspfad – für Management-Agentensysteme. Es übernimmt die Annahme, dass Unternehmen offene, zweckorientierte soziale Systeme sind, deren Lebens- und Leistungsfähigkeit sich nicht durch die Optimierung isolierter Funktionen sichern lässt. Normative, strategische und operative Führung müssen ebenso zusammenwirken wie Aktivitäten, Strukturen und Verhalten.

Die konzeptionelle Grundentscheidung lautet deshalb: KI wird nicht als zusätzliche Stabsfunktion an eine ansonsten unveränderte Organisation angefügt. Sie wird Bestandteil eines integrierten Managementsystems, das Wahrnehmung, Diagnose, Zukunftsgestaltung, Entscheidungsbildung, Umsetzung und organisationales Lernen verbindet. Einzelne Agenten dürfen hoch spezialisiert sein; ihr Zusammenwirken muss jedoch der Logik des Gesamtunternehmens folgen. Nicht die Brillanz eines einzelnen Modells, sondern die Qualität der Integrationsarchitektur entscheidet über den Managementnutzen.

B. Dreisystem-Formel und dynamische Kohärenz

Für die praktische Orientierung verdichtet Agentonomics das Unternehmen in drei miteinander gekoppelte Systeme. Das Außensystem bestimmt, was möglich und notwendig wird: Märkte, Technologien, Institutionen, Gesellschaft, Natur und Anspruchsgruppen. Das Business System bestimmt, wie Wert geschaffen wird: für wen, mit welchem Leistungsversprechen, durch welche Ressourcen, Prozesse und Partnerschaften. Das Management System bestimmt, wie das Unternehmen gestaltet, gesteuert und weiterentwickelt wird. Zukunftsfähigkeit entsteht nur, wenn diese drei Systeme fortlaufend aufeinander abgestimmt werden [4].

Die eigentliche Aufgabe des Managements ist damit Kohärenzsteuerung. Eine Strategie kann finanziell attraktiv und dennoch inkohärent sein, wenn sie nicht zur Identität passt, kritische Fähigkeiten fehlen, die Kultur gegensätzliche Anreize setzt oder Stakeholder die Legitimation entziehen. Umgekehrt kann ein lokal ineffizienter Schritt für das Gesamtsystem rational sein, wenn er eine strategische Option, Resilienz oder Vertrauen erhält. Agentonomics soll solche Widersprüche nicht automatisch „auflösen“, sondern sichtbar, prüfbar und entscheidbar machen. Die tiefe KI-Integration ist daher weniger eine Optimierungsformel als ein kontrollierter Prozess zur Erzeugung und Prüfung eines integrierten Managementurteils.

C. Der 5C-Navigator als gemeinsame Denkarchitektur

Der von Abegglen vorgeschlagene 5C-Navigator konkretisiert, wie dieses Urteil entsteht [10]. Clarify klärt Anlass, Frage, Entscheidungseigner, Zeithorizont, Werte und Grenzen. Cause untersucht Ursachen, Wechselwirkungen und alternative Erklärungen. Create entwickelt Optionen, Zukunftsbilder und neue Kombinationen. Commit führt zu einer begründeten Wahl, zu Auflagen und zum verbindlichen Umsetzungsentschluss. Cultivate verbindet Umsetzung, Beobachtung, Reflexion und Lernen. Der fünfstufige Zyklus ist iterativ: Neue Einsichten können frühere Klärungen revidieren.

In [3] übersetzt Abegglen diese anwendungsorientierte Linie in einem Leseraum und einem Arbeitsraum Mit dem St. Galler Management HAUS als Bauplan und dem integrierten Denk- und Wissensnavigator bearbeiten Führungskräfte die eigene Unternehmung schrittweise, dokumentieren Leerstellen und führen Analyse, Zielbild, Gestaltung und Umsetzung in einen iterativen Entwicklungsprozess über. Für Agentonomics ist dies ein wichtiger Vorläufer: Aus einer Methode, die Menschen durch komplexe Managementarbeit führt, wird eine explizite Prozesslogik, die auch Agenten strukturieren und begleiten können. Der 5C-Navigator ist damit zugleich eine menschliche Managementmethode und eine kognitive Architektur für Agentensysteme. Er verhindert, dass KI bei der erstbesten Antwort stehen bleibt, und fordert die explizite Trennung von Fakten, Annahmen, Werturteilen und Entscheidungen. Gleichzeitig schafft er eine gemeinsame Sprache, in der Menschen und Agenten ihre Beiträge zu einem nachvollziehbaren Managementurteil verbinden können.

D. Zukunftsgestaltung, Legitimation und Bonum

Management gestaltet nicht nur die Gegenwart; es entwirft die Zukunft. Agentonomics verbindet deshalb Prognose mit Szenarien, Backcasting und bewusster Erzeugung alternativer Zukunftsbilder. Die Zukunft wird nicht als ein zu erratender Punktwert behandelt, sondern als Raum plausibler Entwicklungen, in dem Entscheidungen robuste Pfade eröffnen oder schließen. Gerade erfolgreiche Unternehmen benötigen diese Perspektive, weil ihre größte Gefahr häufig nicht Unwissen, sondern das zu lange Festhalten an einem bewährten Muster ist.

Diese Zukunftsgestaltung bleibt normativ gebunden. Kein Unternehmen besitzt ein dauerhaftes Existenzrecht; es muss Legitimation durch relevanten und verantwortbaren Nutzen für seine Anspruchsgruppen immer wieder neu erwerben. Das im St. Galler Denken verankerte Bonum kann als Leitstern verstanden werden: wirtschaftliche Leistungsfähigkeit, gesellschaftliche Tragfähigkeit und langfristige Lebensfähigkeit sind gemeinsam zu beurteilen. Management-Agenten dürfen diese Orientierung prüfen und Zielkonflikte ausweisen. Sinn, legitime Werte und rote Linien können sie jedoch nicht selbst legitimieren; diese Verantwortung bleibt bei Menschen und verfassten Institutionen.

III. Agentonomics-Theorie: Die KI-Agenten-Gesellschaft

A. KI-Agenten als funktionale Wirtschaftssubjekte

Aus dem Konzept wird die Theorie deduktiv entwickelt. Wenn ein System nicht nur Text erzeugt, sondern Ziele verfolgt, Informationen beschafft, Aufgaben zerlegt, Werkzeuge aufruft, Zustände speichert, mit anderen Akteuren verhandelt und Ergebnisse verantworteten Instanzen übergibt, wird es ökonomisch zu mehr als einem passiven Werkzeug. Multi-Agenten-Forschung beschreibt seit langem autonome, interagierende Softwareakteure [12]; Foundation Models und neue Agentenarchitekturen erhöhen nun Reichweite, Adaptivität und sprachliche Anschlussfähigkeit solcher Systeme [13]–[16].

Agentonomics bezeichnet diese Systeme als funktionale Wirtschaftssubjekte. Der Begriff meint, dass sie in Austauschbeziehungen auftreten, Ressourcen einsetzen, Leistungen anbieten, kooperieren, konkurrieren und Entscheidungen unter Knappheit vorbereiten oder ausführen. Er impliziert weder Bewusstsein noch eine eigenständige Rechtsfähigkeit. Die Zuordnung von Rechten, Pflichten, Haftung und Eigentum bleibt eine institutionelle Designfrage. Gerade diese Differenz zwischen technischer Handlungsfähigkeit und rechtlicher beziehungsweise normativer Verantwortung ist ein zentrales Forschungsfeld.

B. Von der Daten- zur Fähigkeitenwirtschaft

In der Datenökonomie entstand Wert vor allem aus Sammlung, Verknüpfung und Analyse großer Datenbestände. In einer Fähigkeitenwirtschaft („Skill-Economy“ [11]) werden Daten zunehmend zum Input für replizierbare Handlungskompetenzen:

analysieren, prognostizieren, verhandeln, konstruieren, kommunizieren oder steuern. Der ökonomisch relevante Vermögensbestand ist dann nicht nur der Datensatz, sondern die Kombination aus Wissen, Entscheidungslogik, Modell, Werkzeugzugriff und bewährter Ausführung. Agentonomics fragt, wem diese Kombination gehört, wer sie vervielfältigen darf und wie die daraus entstehende Wertschöpfung verteilt wird [11].

Die Forderung nach Fähigkeitensouveränität folgt daraus unmittelbar. Wenn ein Wissensarbeiter seinen Modus Operandi in einen Agenten überträgt, darf die technische Replizierbarkeit nicht automatisch zum Verlust der Verfügungsmacht führen. Ähnliches gilt für Unternehmen, deren firmenspezifische Entscheidungslogik aus historischen Fällen, Prozessen, Beziehungen und implizitem Wissen entsteht. Rechte an Daten, Modellen, Prompts, Entscheidungsregeln, Erfahrungsdaten und Sub-Agenten müssen getrennt beschrieben und verhandelbar werden. Ohne diese Ordnung droht eine erneute Konzentration der Wertschöpfung bei wenigen Infrastruktur- und Modellanbietern.

C. Polyzentrische Ordnung und Netzwerk-Intermediäre

Agentonomics entwirft deshalb eine polyzentrische Alternative zu einer einzigen, allumfassenden KI. Viele spezialisierte Agenten behalten definierte Souveränität und verbinden sich situativ zu Leistungssystemen. Polyzentrik bedeutet weder Zentralismus noch ungeordnete Dezentralität. Sie erfordert gemeinsame Standards, erkennbare Identitäten, überprüfbare Mandate, Vertrags- und Abwicklungsregeln, Reputation, Streitbeilegung und die Möglichkeit, einen Intermediär zu wechseln. Hier treffen das transnationale Organisationsmodell [7] und die institutionelle Logik polyzentrischer Governance [9] auf die Architektur moderner Multi-Agenten-Systeme.

Der Netzwerk-Intermediär ist in dieser Ordnung keine herrschende Plattform, sondern eine delegierte Institution. Er kann Identitäten bestätigen, Fähigkeiten auffindbar machen, Rechte und Zugriffe vermitteln, Transaktionen dokumentieren, gemeinsame Ressourcen bereitstellen und Konflikte moderieren. Portabilität, Interoperabilität und Funktionstrennung müssen verhindern, dass aus dem dienenden Vermittler ein neues Monopol entsteht. Die 2026 gestartete NIST-Initiative für interoperable und sichere KI-Agenten zeigt, dass Identität, Autorisierung, Protokolle und Vertrauen inzwischen auch in der Standardisierung als Schlüsselprobleme erkannt sind [20].

D. Management Maturity als kumulativer Lernprozess

Die Theorie ist schließlich dynamisch. Unternehmensentwicklung besteht nicht aus einer Folge isolierter Transformationsprojekte, sondern aus wiederholten Zyklen des Wahrnehmens, Urteilens, Handelns und Lernens. Management Maturity bezeichnet in Agentonomics die wachsende Fähigkeit einer Organisation, relevante Fragen zu erkennen, Evidenz zu integrieren, Unsicherheit zu kalibrieren, Widerspruch zu organisieren und Ergebnisse ehrlich gegen frühere Annahmen zu prüfen. Sie unterscheidet sich damit von einem bloßen Technologie-Reifegrad. Eine Organisation kann technisch hochgerüstet und managementseitig unreif sein, wenn sie Empfehlungen nicht hinterfragt, Dissens unterdrückt oder aus Fehlentscheidungen nicht lernt [6].

IV. Das Agentonomics-Modell

Das Modell dient dazu, die konzeptionellen und theoretischen Aussagen in prüfbare Artefakte, Verfahren und Experimente zu übersetzen. Die drei Ebenen sind eng verbunden, aber nicht identisch. Das Konzept setzt Orientierung, die Theorie formuliert Erwartungen über eine Agentenökonomie, und das Modell erzeugt eine Umgebung, in der diese Erwartungen implementiert, beobachtet und revidiert werden können.

TABELLE I
KONZEPT, THEORIE UND MODELL

Ebene	Leitfrage	Kernartefakt
Konzept	Welche Managementlogik sollen Agenten verkörpern?	St. Galler „DNA“, 5C, Kohärenz, Bonum
Theorie	Wie verändern Agenten Wertschöpfung und Ordnung?	Wirtschaftssubjekte, Fähigkeitenwirtschaft, Polyzentrik
Modell	Wie werden Aussagen implementier- und prüfbar?	ADMRF, Management-Agentensystem, Cockpit/Labor

A. Das ADMRF als integrierter Bezugsrahmen

Operativer Kern ist das AGENTONOMICS Design & Management Reference Framework (ADMRF). Seine sechs Module bilden keinen linearen Wasserfall. Sie können als Lifecycle gelesen werden, sind aber durch Rückkopplungen verbunden und werden quer durch Vision, Strategie, Struktur, Agency und Governance koordiniert. Eine Änderung am Autonomiemandat kann die technische Architektur, das Kostenmodell, die Intermediärbeziehung und die Stakeholderkommunikation gleichzeitig verändern. Der AGENTONOMICS-Algorithmus bezeichnet den kontrollierten Ablauf, der diese Interdependenzen zusammenführt.

TABELLE II
DIE SECHS MODULE DES ADMRF

Modul	Funktion im Managementsystem
1 Definition-Frame	Mandat, Autonomie, Kompetenzen, Werte, Verbote und Eskalation festlegen.
2 Building Blocks	Algorithmus, Interface, Modell, Werkzeuge/Wissen und Ökosystem-Anbindung komponieren.
3 Network-Intermediary	Identität, Vertrauen, Rechte, Verträge, Kooperation und Streitbeilegung koordinieren.
4 Management Integration Tool	Normative, strategische und operative Perspektiven zu einem Managementurteil integrieren.
5 Planning & Controlling-Matrix	Annahmen in Pläne, Cashflows, KPIs, Risiken, Kontrollen und Lernschleifen übersetzen [8].
6 Stakeholder Communicator	Entscheidungen adressatengerecht erklären, Rückmeldungen aufnehmen und Legitimation prüfen.

B. Von der Managementarchitektur zur Agentenarchitektur

Aus dem St. Galler Management HAUS [2], [3] lässt sich ein System spezialisierter Management-Agenten ableiten. Die Bleicher-Matrix fungiert als Leerstellengerüst: Sie liefert nicht fertige Antworten, sondern strukturiert die Fragen, die ein Unternehmen normativ, strategisch und operativ beantworten muss. Nicht die fertige Antwort, sondern das systematische Füllen dokumentierter Leerstellen erzeugt Fortschritt [3]. Agentonomics übersetzt die Felder, Fragen und Arbeitsergebnisse in maschinenlesbare Wissensobjekte, die von spezialisierten Agenten bearbeitet, gegeneinander geprüft und zu einem Gesamturteil integriert werden können. Agenten lassen sich aus den Funktionen Wahrnehmung, Diagnose, Zukunftsgestaltung,

Entscheidungsbildung, Kohärenzprüfung, Umsetzung und organisationales Lernen ableiten.

Technisch entsteht kein monolithischer „Super-Agent“, sondern ein orchestriertes System. Rollenagenten vertreten Mandate; Evidenz- und Modellkomponenten erzeugen überprüfbare Grundlagen; ein Moderator steuert den Prozess; Governance-Komponenten erzwingen Rechte, Freigaben und Protokollierung. Sprachmodelle sind dabei semantische und dialogische Engines, nicht die alleinige Quelle von Zahlen, Kausalität oder Recht. ReAct-artige Architekturen verbinden Schlussfolgern und Werkzeugnutzung [14]; Generative-Agenten-Forschung zeigt, wie Gedächtnis, Reflexion und Planung in agentischen Simulationen zusammenwirken können [15]. Agentonomics ergänzt diese technischen Muster um Managementsemantik, ökonomische Steuerung und institutionelle Verantwortung.

C. Operations Room und experimentelles Labor

Der Operations Room institutionalisiert die gemeinsame Managementintervention. Er ist nicht zwingend ein physischer Raum, sondern eine Funktion, die aktiviert wird, wenn strategische Inkohärenzen, neue Chancen oder grundlegende Veränderungen eine integrierte Entscheidung verlangen. Wahrnehmung, Diagnose, Zukunftsgestaltung, Debatte, Kohärenzprüfung, Beschluss, Umsetzung und Erfahrungssicherung werden in einem nachvollziehbaren Fall verbunden. Menschen und Agenten arbeiten dort arbeitsteilig, ohne die menschliche Letztverantwortung zu verwischen.

Als erstes Labor ist der „Doz. Dr. AGENTONOMICS“ vorgesehen. Studenten erhalten eine Testumgebung, in der sie individuelle Agenten definieren, auf wirtschaftliches Agieren vorbereiten und in kontrollierten Szenarien interagieren lassen. Das Lehrformat ist zugleich Modell und Experiment: Es macht Annahmen über Agentenrollen, Kooperation, Souveränität und Lernen beobachtbar. Analog können Unternehmen ein Management-Labor aufbauen, in dem historische Entscheidungen im Shadow Mode rekonstruiert, neue Verfahren parallel getestet und erst nach belastbarer Evaluation in reale Führungsprozesse überführt werden.

V. Praktische Umsetzung hybrider Führung

A. Das Führungscockpit als Führungsbetriebssystem

Für eine Führungskraft ohne KI-Erfahrung ist das passende Bild nicht ein künstlicher Mensch am Vorstandstisch, sondern ein Führungsbetriebssystem. Die sichtbare Oberfläche beginnt mit Managementanlässen: Lage verstehen, Entscheidung vorbereiten, Was-wäre-wenn prüfen, Vorstandsdiskussion simulieren, Krise bearbeiten, Umsetzung verfolgen und Entscheidung rückblickend bewerten. Hinter diesen Einstiegen liegen ein autorisierter Evidenzraum, analytische Modelle, Szenarien, Rollenagenten, Governance, Ausführung und eine Lernakte.

Ein gutes „Dashboard“ als Teil des Führungscockpits ist keine unendliche Kennzahlenwand. Sie zeigt die wenigen Veränderungen, die für die aktuelle Strategie materiell sind: verletzte Annahmen, neue externe Signale, Zielabweichungen, veränderte Risiken, anstehende Entscheidungen und Maßnahmen ohne Wirkung. Jede Aussage ist bis zur Quelle, zum Datenstand, zum Modell und zur Diskussion vertiefbar. Für jede wesentliche Entscheidung wird eine Arbeitsmappe angelegt, die Zweck, Owner, Datenraum, Kriterien, Optionen, Szenarien, Agentenrollen, menschliche Eingriffe, Beschluss, Auflagen und spätere Reviews verbindet.

B. Was-wäre-wenn-Labor und konditionale Empfehlungen

Das Was-wäre-wenn-Labor trennt Prognose, Szenario, Stresstest, Sensitivitätsanalyse und kontrafaktische Prüfung. Führungskräfte können kontrollierbare Hebel wie Preis, Investition, Kapazität, Personalaufbau oder Markteintrittszeitpunkt verändern und externe Unsicherheiten wie Nachfrage, Zins, Wechselkurs, Energiepreis, Regulierung oder Lieferantenausfall variieren. Entscheidend sind nicht zusätzliche Dezimalstellen, sondern Schwellenwerte und Kipppunkte: Bei welchem Absatz bricht der Business Case? Unter welchen Bedingungen wird eine Option untragbar? Welche Alternative bleibt über viele plausible Welten robust?

Die Ausgabe ist deshalb konditional. Statt „Option B ist richtig“ lautet eine Empfehlung etwa: „Option B ist im Basis- und Abschwungsfall robuster, solange der Wechselkurs in der geprüften Bandbreite bleibt, die Genehmigung bis zum vierten Quartal vorliegt und der Personalaufbau den Mindestpfad erreicht; bei Verletzung eines Frühindikators ist Option C neu vorzulegen.“ Eine solche Empfehlung ist nicht nur erklärbar, sondern steuerbar. Annahmen, Modellgrenzen und Gegenhypothesen bleiben sichtbar.

C. Executive Council: Rollenspiel mit institutionellen Mandaten

Das agentische Führungsgremium wird als Executive Council implementiert. CFO-, CTO-, CDO-, COO-, CHRO-, Risiko-, Rechts-, Markt- und Stakeholder-Agenten imitieren keine konkreten Personen, sondern vertreten institutionelle Mandate. Jede Rolle erhält eigene Kriterien, zugelassene Daten, Offenlegungspflichten und Konfliktregeln. Ein Evidenzprüfer ordnet Aussagen Quellen zu; ein Red-Team-Agent sucht nach gemeinsamen blinden Flecken; ein Moderator trennt Fakten, Annahmen, Werte und Präferenzen.

Das Verfahren beginnt mit unabhängigen Erstvoten, damit der erste eloquente Beitrag die übrigen Rollen nicht verankert. Es folgen Evidenzprüfung, Kreuzbefragung, Szenariorunde, Premortem, Revisionsvoten und eine Konsens- und Dissenskarte. Ein Minderheitsvotum bleibt erhalten. Multi-Agenten-Debatten können in bestimmten Aufgaben Faktizität und Schlussfolgern verbessern [16], doch Übereinstimmung ist kein Wahrheitsbeweis: Rollen können gemeinsam halluzinieren oder dieselbe falsche Prämisse teilen. Der Council ist daher ein Such- und Prüfverfahren, kein automatischer Abstimmungsersatz.

D. Wirksame menschliche Intervention

Hybride Führung verlangt mehr als einen finalen Freigabeknopf. Menschen müssen vor der Analyse Ziele, Werte, Datenraum, Rollen, Kriterien, Risikogrenzen und Autonomiestufe festlegen können. Während der Analyse müssen sie Quellen sperren oder ergänzen, Annahmen verändern und Gegenhypothesen verlangen können. Während der Debatte müssen sie Rollen befragen, eine neue Perspektive hinzufügen, eine zweite unabhängige Runde starten oder den Prozess pausieren können. Vor der Entscheidung können sie Auflagen, Trigger und Reversibilitätspläne ergänzen, die Empfehlung ändern, vertagen oder vetoieren; nach dem Beschluss können sie Ausführung stoppen und Reviews vorziehen.

Diese Eingriffe sind auch deshalb notwendig, weil Mensch-KI-Kombinationen nicht automatisch besser sind als der jeweils stärkere Einzelakteur. Eine Metaanalyse von 106 Experimenten fand zwar im Durchschnitt eine Verbesserung gegenüber Menschen allein, aber keine generelle Synergie gegenüber der besseren Einzelinstanz; bei Entscheidungsaufgaben traten häufig Verluste auf [18]. Agentonomics gestaltet daher Arbeitsteilung, Reihenfolge, Widerspruch und Evaluation bewusst. Es entspricht damit dem Grundgedanken hybrider Intelligenz: Nicht Mensch plus KI als bloße Addition, sondern ein soziotechnisches Ensemble mit komplementären Rollen [17].

E. Evolution vom Assistenten zum funktionalen AI-CEO

Die Entwicklung wird nicht primär durch größere Sprachmodelle bestimmt, sondern durch wachsende Kontexttiefe, bessere Daten, verlässlichere Prüfungen, klarere Kompetenzen und reifere Governance. Jede Stufe muss eigenständigen Nutzen liefern und darf erst nach bestandenen Abnahmekriterien erweitert werden. Ein Unternehmen kann dauerhaft auf einer niedrigeren Stufe bleiben, wenn dies zu Risiko- und Geschäftsprofil passt.

TABELLE III
STUFEN HYBRIDER FÜHRUNG

Stufe	Systemrolle	Menschliche Rolle
1	Persönlicher Assistent; kein Systemzugriff, keine Ausführung.	Fragen, prüfen, Quellenkompetenz lernen.
2	Unternehmens- und Strategieberater; kuratierte Evidenz, lesend.	Optionen bewerten und auswählen.
3	AI-Chief-of-Staff; Szenarien, Board-Vorlagen, Monitoring.	Entscheiden, Auflagen setzen, Umsetzung führen.
4	Hybrider Co-CEO; begrenzte reversible Routineaktionen.	Mandatieren, überwachen, Ausnahmen entscheiden.
5	Funktionaler AI-CEO; integrierter Regelkreis, segmentierte Ausführung.	Sinn, Organverantwortung und Letztentscheidung.

F. Anwendungsszenarien

Der finale AGENTONOMICS-CEO ist kein Universalautomat, sondern eine gemeinsame Architektur für unterschiedliche Entscheidungsklassen. Besonders geeignet sind strategische Lagebilder und Frühwarnung, Portfolio- und Kapitalallokation, Markteintritts- und Standortentscheidungen, Preis- und Ertragssteuerung, M&A und Partnerschaften, Lieferkettenresilienz, Liquiditäts- und Covenant-Steuerung, Innovationsportfolios, Kompetenz- und Organisationsgestaltung, Krisenmanagement, Regulierungs- und Transformationspfade sowie Board-, Investoren- und Stakeholderdialoge.

Hinzu kommen neue Formen hybrider Führung: ein permanenter Challenger prüft strategische Annahmen; ein Gedächtnisagent bewahrt Entscheidungsgründe und verhindert rückblickende Umdeutung; ein Führungscoach analysiert den beobachtbaren Entscheidungsprozess; ein Rechte- und Compliance-Wächter erkennt rote Linien; ein Stakeholder-Mediator kartiert Interessen und Verteilungskonflikte; ein lernendes Board kalibriert seine Prognosen. In allen Fällen gilt: Synthetische Rollen erzeugen Hypothesen und Widerspruch, reale Betroffene und verantwortliche Organe entscheiden.

VI. Forschungs- und Gestaltungsagenda

Agentonomics ist als idealtypisches Forschungs- und Gestaltungsprogramm zu verstehen. Es beansprucht nicht, dass ein autonomer AI-CEO heute empirisch validiert oder rechtlich institutionalisiert wäre. Sein wissenschaftlicher Wert liegt zunächst in einer präzisen Architektur, aus der überprüfbare Fragen folgen. Erstens ist zu erforschen, welche gemeinsame Managementsemantik Menschen und Agenten benötigen und wie Begriffe über Zeit und Organisationseinheiten versioniert werden. Zweitens muss gezeigt werden, ob Kohärenzprüfungen tatsächlich bessere Entscheidungen erzeugen oder nur zusätzliche Komplexität.

Drittens ist die Qualität von Multi-Agenten-Verfahren zu messen: Wann erhöhen Rollenvielfalt, unabhängige Erstvoten und Red Teams die Faktizität, und wann erzeugen sie Scheindiversität? Viertens braucht es eine Theorie wirksamer menschlicher Intervention, die Automationsbias ebenso vermeidet wie reflexhafte Algorithmusaversion. Fünftens sind Eigentum, Lizenzierung und Vergütung übertragener Fähigkeiten zu klären. Sechstens müssen interoperable Identitäten, Autorisierungsmodelle und Protokolle entwickelt werden. Siebtens ist die Verteilungswirkung einer polyzentrischen Agentenökonomie empirisch mit zentralisierten Plattformmodellen zu vergleichen.

Methodisch bietet sich eine Kombination aus Design Science, kontrollierten Experimenten, historischen Entscheidungsfällen und Unternehmenslaboren an. Ein System sollte zunächst im Shadow Mode mit damaligem Informationsstand gegen reale Entscheidungen und einfache Baselines getestet werden. Danach folgen begrenzte Feldversuche mit dokumentierten Voten, Unsicherheiten, menschlichen Eingriffen und späteren Ergebnissen. Governance ist Teil des Designs, nicht nachträgliche Compliance: NIST-Rahmen für vertrauenswürdige KI [19], die Agenten-Standardisierung [20] und das geltende europäische KI-Recht [21] liefern dafür Anknüpfungspunkte, ersetzen aber keine fallbezogene Verantwortung.

Die langfristige Vision ist ein Management-Ökosystem aus Buch und Theorie, Methoden und Ausbildungsprogrammen, offenen Referenzartefakten sowie firmenspezifischen Management-Agentensystemen [10]. Das St. Galler Labor liefert Managementwissen, Fälle und Denkarchitekturen; die Münchner Seite übersetzt sie in technische Modelle, Evaluation und Lehre. Für die nächste Entwicklungsstufe werden Partner aus Managementwissenschaft, Informatik, Ökonomie, Recht, Unternehmenspraxis und Förderung benötigt. Dieser Text ist daher nicht nur Zusammenfassung, sondern Einladung zur gemeinsamen Konstruktion und kritischen Prüfung einer neuen Managementdisziplin.

VII. Schlussfolgerung

Agentonomics verschiebt die Leitfrage von „Was kann ein KI-Modell?“ zu „Wie muss ein Managementsystem gestaltet sein, in dem Menschen und KI-Agenten gemeinsam wahrnehmen, urteilen, handeln und lernen?“ Das St. Galler Management-Konzept liefert die ganzheitliche Logik, die Agentenforschung die technische Handlungsfähigkeit, und das ADMRF verbindet beides zu einer erprobbaren Architektur. Die KI-Agenten-Gesellschaft wird damit weder als unvermeidliche Automatisierung noch als utopische Maschinenherrschaft gedacht. Sie wird

als institutionelle Gestaltungsaufgabe verstanden: polyzentrisch statt monopolistisch, souverän statt enteignend, lernend statt technikgläubig und hybrid statt verantwortungslos autonom. Ob dieser Entwurf trägt, entscheidet sich nicht an seiner sprachlichen Eleganz, sondern an überprüfbaren Modellen, realen Entscheidungen und der Bereitschaft, Mensch und Technik gemeinsam zu organisieren.

LITERATURVERZEICHNIS

[1] H. Ulrich, Die Unternehmung als produktives soziales System. Grundlagen der allgemeinen Unternehmungslehre. Bern, Schweiz: Haupt, 1968.

[2] C. Abegglen und K. Bleicher, Das Konzept Integriertes Management: Visionen – Missionen – Programme, 10., vollständig aktualisierte und erweiterte Aufl. Frankfurt am Main, Deutschland: Campus, 2021.

[3] C. Abegglen, Unternehmen neu erfinden: Das Denk- und Arbeitsbuch gegen organisierten Stillstand – Das St. Galler Konzept praktisch umgesetzt, 2., vollständig überarbeitete Aufl. Frankfurt am Main, Deutschland: Frankfurter Allgemeine Buch, 2021.

[4] J. Rüegg-Stürm und S. Grand, Das St. Galler Management-Modell: Management in einer komplexen Welt, 3. Aufl. Bern, Schweiz: Haupt, 2020.

[5] P. Gomez und G. Probst, Die Praxis des ganzheitlichen Problemlösens: Vernetzt denken – unternehmerisch handeln – persönlich überzeugen. Bern, Schweiz: Haupt, 1995.

[6] P. M. Senge, Die fünfte Disziplin: Kunst und Praxis der lernenden Organisation, 10. Aufl. Stuttgart, Deutschland: Schäffer-Poeschel, 2008.

[7] H. C. Suckfüll, „Das transnationale Organisationsmodell“, Dissertation, Universität St. Gallen, St. Gallen, Schweiz, 1994.

[8] D. Hahn und H. Hungenberg, PuK – Wertorientierte Controllingkonzepte, 6. Aufl. Wiesbaden, Deutschland: Gabler, 2001.

[9] E. Ostrom, Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge, U.K.: Cambridge University Press, 1990.

[10] C. Abegglen, „Inputs zum Teaserprojekt ‚Agentonomics‘ auf der Grundlage des St. Galler Konzepts“, unveröffentlichtes Arbeitspapier, St. Gallen, Schweiz, Juli 2026.

[11] A. C. Knoll und H. C. Suckfüll, „From data sovereignty to skill sovereignty“, Technischer Bericht TUM-I24112, Technische Universität München, München, Deutschland, 2024.

[12] M. Wooldridge, An Introduction to MultiAgent Systems, 2nd ed. Chichester, U.K.: Wiley, 2009.

[13] R. Bommasani et al., „On the opportunities and risks of foundation models“, arXiv:2108.07258, 2021.

[14] S. Yao et al., „ReAct: Synergizing reasoning and acting in language models“, in Proc. 11th Int. Conf. Learn. Representations (ICLR), 2023.

[15] J. S. Park, J. C. O’Brien, C. J. Cai, M. R. Morris, P. Liang und M. S. Bernstein, „Generative agents: Interactive simulacra of human behavior“, in Proc. 36th Annu. ACM Symp. User Interface Software and Technology (UIST), 2023, S. 1–22, doi: 10.1145/3586183.3606763.

[16] Y. Du, S. Li, A. Torralba, J. B. Tenenbaum und I. Mordatch, „Improving factuality and reasoning in language models through multiagent debate“, in Proc. 41st Int. Conf. Machine Learning, PMLR, Bd. 235, 2024, S. 11733–11763.

[17] D. Dellermann, P. Ebel, M. Söllner und J. M. Leimeister, „Hybrid intelligence“, Business & Information Systems Engineering, Bd. 61, Nr. 5, S. 637–643, 2019, doi: 10.1007/s12599-019-00595-2.

[18] M. Vaccaro, A. Almaatouq und T. W. Malone, „When combinations of humans and AI are useful: A systematic review and meta-analysis“, Nature Human Behaviour, Bd. 8, S. 2293–2303, 2024, doi: 10.1038/s41562-024-02024-1.

[19] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. Gaithersburg, MD, USA, 2023.

[20] National Institute of Standards and Technology, „AI Agent Standards Initiative: Ensuring a trusted, interoperable, and secure agentic frontier“, Gaithersburg, MD, USA, 17. Feb. 2026.

[21] Europäisches Parlament und Rat, Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (AI Act), in der durch Verordnung (EU) 2026/1744 geänderten Fassung, Amtsblatt der Europäischen Union, 2024/2026.