

AI-Governance als strategisches Instrument

Wie nachhaltige AI-Systeme Stakeholder-Interessen und strategische Unternehmensziele integrieren

In ihrer AI-Anwendung müssen Unternehmen ethische Anforderungen wie Fairness, Transparenz und Verantwortlichkeiten mitdenken. Auch Regulierungsbehörden definieren immer strikere Regeln für die Nutzung der wirtschaftlichen und technologischen Vorteile von AI. Das St. Galler Governance Framework skizziert Schritte, um eine AI-Governance mit ethischen und rechtlichen Standards in Einklang zu bringen.

Richard Sentinella, Victoria Honsel und Maël Schnegg

1. Einführung in AI-Governance

Unternehmen stehen heute vor der Herausforderung, die weitreichenden Potenziale künstlicher Intelligenz (in der Folge wird die englische Abkürzung AI = «artificial intelligence» benutzt) zu erschließen, um auch zukünftig wettbewerbsfähig zu bleiben. Jedoch muss dabei sichergestellt werden, dass die Nutzung von AI den Interessen von diversen Stakeholdern (Nutzer, Arbeitgeber, Technologieanbieter etc.) gerecht wird. Die gleichzeitige Berücksichtigung gesellschaftlicher und technologischer Belange aller Interessensgruppen und der wirtschaftliche Druck, AI schnell anzuwenden, bilden ein Spannungsfeld, das es vonseiten des Managements auszubalancieren gilt. Zusätzlich zu diesem Spannungsfeld werden für die Nutzung von AI durch alle rechtlichen Instanzen verstärkte regulatorische Vorgaben gemacht. Diese Vorgaben werden durch bestimmte Charakteristika der AI (wie ihr nicht-deterministischer Charakter oder die oft undurchsichtige Ergebniserzeugung in «Black Boxes») bedingt, die weitergehende ethische Fragen aufwerfen: zum Beispiel bezüglich ihrer Fairness, Transparenz, Verantwortlichkeit, Vertrauenswürdigkeit oder ihres Einflusses auf das menschliche Wohlergehen.

AI-Governance strebt an, diese komplexen Herausforderungen auszubalancieren. Sie bietet Unternehmen ein System mit spezifischen Prozessen, Strukturen und Kontrollen, um den Einsatz von AI so zu steuern, dass Fairness und Verantwortlichkeit gefördert, regulatorische Anforderungen erfüllt

und die AI-Anwendungen auf die strategischen Unternehmensziele abgestimmt werden. In der Regel ist AI-Governance eine Controlling-Herausforderung, bei der die spezifischen Steuerungssysteme so konzipiert und eingesetzt werden müssen, dass die Ressourcen Technik und Know-how effektiv und effizient für die Umsetzung der Unternehmensstrategie genutzt werden können.

In der Folge beleuchten wir aktuelle ethische und regulatorische Herausforderungen von Organisationen und erörtern, wie ein AI-Governance Framework wie das St. Galler Governance Framework für AI (vgl. Sentinella et al., 2023) Organisationen unterstützen kann, ihre AI-Strategien verantwortungsvoll zu gestalten.

2. Regulatorische und Ethische Ansprüche an AI-Nutzung

Die letzten Jahre haben gezeigt, welche vielfältigen wirtschaftlichen und technologischen Vorteile die Nutzung von AI Unternehmen aller Industriebereiche bringen kann. Trotz dieser großen Potenziale können Unternehmen, die neue AI-Technologien und Instrumente unüberlegt einsetzen, vielfältige Risiken eingehen. Die Folgen eines AI-Missbrauchs können dabei variieren: Manchmal ist die einzige Konsequenz ein Projektabbruch mit überschaubaren Kosten und begrenzten Folgen für die Reputation. In gravierenderen Fällen kann die AI-Nutzung auch Rechte oder gesetzliche Bestimmungen bestimmter Stakeholder verletzen. So wird beispielsweise OpenAI von diversen Parteien verklagt, da-

Dr. **Richard Sentinella** ist AI Governance Research Fellow an der International Association of Privacy Professionals und hat an der Universität St.Gallen promoviert. E-Mail: rsentinella@iapp.org

Dr. **Victoria Honsel** ist wissenschaftliche Mitarbeiterin an der Universität St.Gallen, Institut für Accounting, Controlling und Auditing. E-Mail: victoria.honsel@unisg.ch

Prof. Dr. **Maël Schnegg** ist Assistenzprofessor an der Universität St.Gallen, Institut für Accounting, Controlling und Auditing. E-Mail: mael.schnegg@unisg.ch

runter die New York Times. Der Vorwurf lautet, dass Zeitungsartikel ohne Genehmigung für das Training der Sprachmodelle genutzt worden seien (vgl. Grynbaum/Mac, 2023). Zudem haben französische Regulierungsbehörden Google eine Geldstrafe in Höhe von 250 Millionen Euro auferlegt, weil das Unternehmen sein Modell mit Daten von Verlagen und Nachrichtenagenturen ohne entsprechende Genehmigung trainiert hat (vgl. Orru, 2024). Solche Urheberrechtsverletzungen sind nur einige der zahlreichen juristischen Herausforderungen in der AI-Nutzung; Hinzu kommen auch Datenschutz- und Antidiskriminierungsrechte (insbesondere wegen der autonomen und nicht-deterministischen Natur der AI). Dabei sind nicht nur Unternehmen von diesen Herausforderungen betroffen: Auch die niederländische Regierung musste 2021 eine Geldbuße zahlen, nachdem festgestellt worden war, dass ein automatisiertes System zur Betrugsbekämpfung im Kindergeldsystem Minderheiten diskriminierte (vgl. Dachwitz, 2021).

Angesichts der fortschreitenden Entwicklung von AI passen Gesetzgeber die rechtlichen Rahmenbedingungen fortlaufend an. Die Europäische Union hat mit dem AI-Gesetz wegweisende Regelungen definiert, die die Nutzung von AI strenger normieren und klare Richtlinien für Transparenz und Governance vorschreiben. Die EU AI Verordnung (vgl. European Commission, 2024) wurde kürzlich vom Europäischen Parlament angenommen und wird in den kommenden Monaten schrittweise in Kraft treten. Sie unterscheidet sich von anderen Verordnungen durch ihren breiten Anwendungsbereich, denn sie gilt sehr allgemein für AI und nicht für spezifische Kontrollen. Und was noch wichtiger ist: Durch ihre Durchsetzungskraft, im Gegensatz zu Empfehlungen der EU, können Verstöße gegen diese Verordnung mit erheblichen Strafen geahndet werden. Ergänzend wurden in den letzten Jahren umfassende Verordnungen wie die «Datenschutz-Grundverordnung» und der «Digital Markets Act» eingeführt, um den Einsatz von AI noch gezielter zu steuern.

Neben diesen gesetzlichen Vorgaben sehen sich Unternehmen auch mit kritischen Stimmen von Stakeholdern konfrontiert, die eine faire, rechenschaftspflichtige und transparente Nutzung von AI fordern. Die nicht-deterministische Natur von AI kann selbst bei identischen Eingaben zu unterschiedlichen Ergebnissen führen, was den zuvor genannten «Black Box» Eindruck nur weiter verstärkt. Ein Mangel an Vertrauen entsteht demnach, wenn die Stakeholder nicht nachvollziehen können, wie und warum AI-Entscheidungen getroffen werden. Dies wirft bei diversen Stakeholdern auch Fragen zur Fairness auf, wenn zum Beispiel Trainingsdaten für die Programme voreingenommen ausgewählt werden oder unvollständig sind. Beispielsweise würde eine HR-AI Software, die ausschließlich mit Daten männlicher Bewerber trai-

niert wird, unbewusst systematisch weibliche Bewerbungen ausschließen. Die fehlende Transparenz erschwert es dabei, die Grundlagen von AI-Entscheidungen zu verstehen, was wiederum auch ihre Rechenschaftspflicht beeinträchtigt. Angesichts dieser Eigenschaften und der möglichen Verzerrungen von Ergebnissen zögern viele Akteure, Verantwortung für den Einsatz von AI zu übernehmen.

AI-Nutzung birgt somit noch grundlegende ethische Fragen, besonders in technologischen Grenzbereichen und wenn sie ohne menschliche Aufsicht stattfindet. Unternehmen müssen sich diesen Herausforderungen stellen, um sicherzustellen, dass ihre AI-Anwendungen verantwortungsvoll gestaltet sind und menschliches Wohlergehen nicht beeinträchtigen. Diese Verantwortung erfordert eine sorgfältige Auseinandersetzung mit den ethischen Implikationen von AI auf organisatorischer Ebene.

3. Integration von Stakeholder-Interessen und Strategiezielen

AI-Governance bietet Unternehmen Werkzeuge und Prozesse, um den Einsatz von AI zu managen, die Einhaltung gesetzlicher Vorgaben zu überwachen und den Schutz von betroffenen Stakeholdern zu gewährleisten. AI-Governance sollte dabei sowohl als ein übergreifendes System angesehen werden, das den Einsatz von AI in der Organisation strategisch angeht, als auch als ein flexibler Werkzeugkasten, der für spezielle Anforderungen genutzt werden kann. Das St. Galler Governance Framework für AI (vgl. Sentinella et al., 2023) fungiert als ein solcher Werkzeugkasten: Es beinhaltet zahlreiche Werkzeuge, um die Nutzung von AI zu steuern, und dient damit als Modell zur Umsetzung der Unternehmensstrategie und -operationen. Mehrere dieser Werkzeuge, die im St. Galler Governance Framework für AI (vgl. Sentinella et al., 2023) vorgestellt werden, sind Teil eines breiteren Steuerungssystems, das zur Umsetzung der Unternehmensstrategie und -operationen dienen kann (vgl. Abb. 1).

Der erste Schritt des Frameworks besteht darin, die verschiedenen Stakeholder-Anforderungen, die Unternehmensstrategien und -ziele sowie die rechtlichen Vorgaben für den Einsatz von AI zu erfassen. Dazu sind verschiedene Aktivitäten notwendig: Erstens die Definition der Strategie und der operationellen Ziele, die mit dem Einsatz von AI verfolgt werden. Zweitens die Identifikation des Stakeholder-Netzwerks der Strategie und der von der Implementierung und Nutzung von AI betroffenen Parteien. Drittens die Identifikation und Formalisierung von möglichen ethischen Bedenken. Und viertens die frühzeitige Festlegung der Kommunikation der AI-Strategie und der operativen Maßnahmen, wobei diese angesichts der dynami-

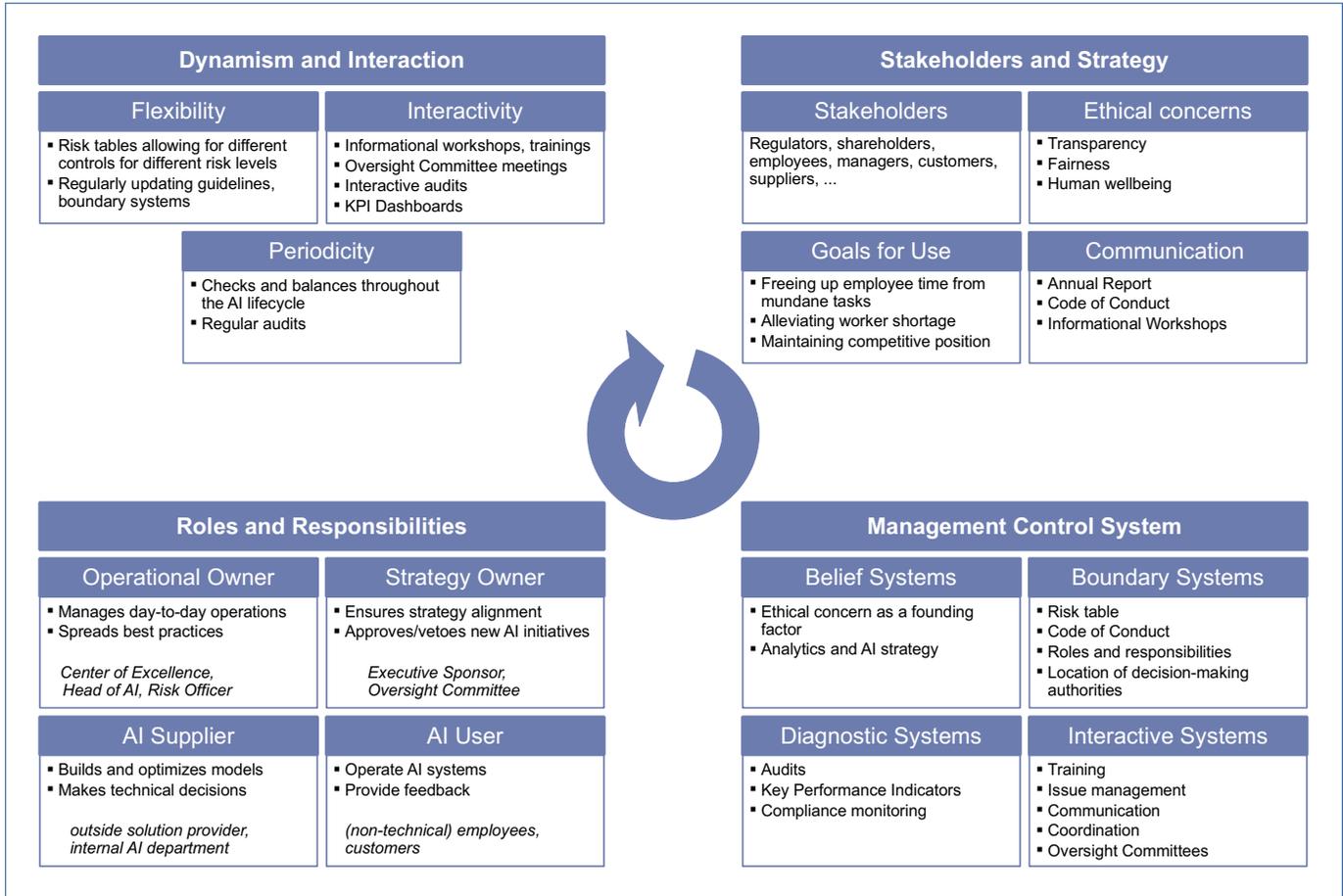


Abb. 1: St. Gallen Governance Framework for AI (sg-GFAI)

schon Entwicklung der AI regelmäßig überprüft werden müssen. Es geht somit nicht darum, jede dieser Komponenten genau im Vorhinein festzulegen. Dennoch ist es notwendig, sie zu formalisieren. Die Art der Formalisierung ist dabei variabel und reicht von «Strategy Maps» bis hin zu Verhaltenskodizes, die sich im Laufe der Zeit ebenfalls weiterentwickeln können. Entscheidend ist, dass diese Herausforderungen zunächst gemeinsam diskutiert und in Folge formalisiert werden.

Die initiale Klärung dieser Aspekte erleichtert die Implementierung eines passenden Steuerungssystems, wie zum Beispiel dem bekannten «Levers of Control» System von Robert Simons (1995). Dieses System bietet ein Rahmenwerk, das es Führungskräften ermöglicht, sowohl kreative Innovationen zu fördern als auch Risiken durch klare Steuerungsmechanismen zu managen. Im Kontext von AI-Governance ist es besonders geeignet, da es sowohl die strategische Ausrichtung durch «Belief Systems» unterstützt als auch durch «Boundary Systems» klare Grenzen für den Einsatz von AI setzt. Zum Beispiel könnte es genutzt werden, um Verhaltenskodizes zu definieren, die unerwünschte Anwendungen von AI identifizieren und eingrenzen, während gleichzeitig Überwachungsmechanismen installiert werden, um die Einhaltung dieser Vorgaben kontinuierlich zu überprüfen. Alternativ könnte auch das COSO-Framework heran-

gezogen werden, das sich auf das Risikomanagement konzentriert und in ähnlicher Weise die Integration von Kontrollmechanismen und die Sicherstellung der Compliance ermöglicht.

Die festgelegte Richtung und Strategie münden in Werkzeuge, die genau definieren, welche Verhaltensweisen und Aktivitäten im akzeptierten Bereich des Unternehmens liegen. Verhaltenskodizes könnten beispielsweise klare Richtlinien zur Transparenz und Fairness von AI-Entscheidungen festlegen, indem sie vorschreiben, dass AI-Modelle regelmäßig auf Bias überprüft werden müssen. Risikotabellen könnten spezifische Risiken wie Datenschutzverletzungen oder Fehlentscheidungen durch AI auflisten und Maßnahmen zur Risikominimierung definieren, wie etwa regelmäßige Audits der Algorithmen. Sobald der Rahmen abgesteckt ist, ermöglichen diagnostische Werkzeuge die regelmäßige Überprüfung von KPIs und die Überwachung der AI-Nutzung. Als ein konkretes diagnostisches Werkzeug könnte so zum Beispiel ein AI-Dashboard fungieren, welches KPIs wie die Genauigkeit der Vorhersagen und die Häufigkeit von Fehlauflösungen überwacht, um damit die AI-Nutzung kontinuierlich zu evaluieren und zu optimieren. Die Adaption von AI bietet dabei die Möglichkeit, bestehende Strategien zu optimieren oder sogar grundlegend zu erneuern. Mechanismen zur Interaktion mit den Stakeholdern sind außerdem un-

erlässlich, um neue Chancen zu nutzen und potenzielle Bedrohungen zu vermeiden. Dies kann durch Schulungen, intern gesteuerte Kommunikationswege oder regelmäßige persönliche Treffen erfolgen.

Die erfolgreiche Umsetzung und Nutzung von AI beruht in diesem Sinne stark auf der Kompetenz der verantwortlichen Stakeholder. Organisationen müssen klare Verantwortlichkeiten zuweisen, zum Beispiel durch die Benennung eines Betriebs- oder Strategieverantwortlichen für die AI-Nutzung. Diese Entscheidungsträger beeinflussen sowohl die Beziehung zum AI-Anbieter als auch die Erfahrungen der Nutzer. Die strategischen Verantwortlichen gewährleisten zudem, dass die Bedürfnisse der Organisation, die rechtlichen Anforderungen und die Erwartungen der Stakeholder berücksichtigt werden. Operative Verantwortliche managen die tägliche Implementierung und Anwendung von AI, sorgen für die Einhaltung von Richtlinien und fördern den Austausch von «Best Practices». Abschließend ist es für Unternehmen unerlässlich, dass diese Governance-Maßnahmen kontinuierlich aktualisiert werden, um mit der extrem schnellen Entwicklung der AI-Technologie Schritt zu halten und proaktiv auf neue Herausforderungen und Chancen zu reagieren.

4. Zukunft der AI-Governance

Angesichts der rasanten Entwicklung von AI mag die Auseinandersetzung mit regulatorischen und ethischen Herausforderungen zunächst entmutigend wirken. Unternehmen, die sich jedoch proaktiv dieser Herausforderung stellen, können ihre AI-Nutzung effizienter gestalten und besser auf neue Chancen und Interessenkonflikte reagieren.

Durch die gezielte Anwendung eines ganzheitlichen Steuerungssystems für den Einsatz von AI im Unternehmenskontext lassen sich Bedenken von Stakeholdern direkt adressieren, rechtliche Anforderungen erfüllen und strategische Ziele in

die Anwendung von AI integrieren. Ein solches System umfasst nicht nur umfassende Steuerungsmechanismen und Regelwerke, sondern bietet Organisationen auch die Möglichkeit, mit minimalen Investitionen zu beginnen. Bereits existierende Instrumente, Strukturen und Prozesse können kreativ angepasst werden, um den spezifischen Herausforderungen der AI-Nutzung gerecht zu werden. Unternehmen, die keine Maßnahmen ergreifen, setzen sich dem Risiko aus, dass Mitarbeiter AI unkontrolliert einsetzen oder dass sie, wenn AI nicht eingesetzt wird, möglicherweise im Wettbewerbsvergleich zurückfallen.

Eine proaktive Implementierung von Governance-Strukturen ermöglicht es Organisationen nicht nur, den Fortschritt technologischer Innovationen gezielt zu nutzen, sondern stärkt dabei auch das Vertrauen aller Beteiligten – eine unabdingbare Voraussetzung für den nachhaltigen Einsatz von AI in geschäftskritischen Bereichen.

Literatur

- Dachwitz, I., Kindergeldaffäre: Niederlande zahlen Millionenstrafe wegen Datendiskriminierung, in: Netzpolitik.Org, 29.12.2021.
- Orru, M., Google Fined Roughly \$270 Million in France Over Dispute With News Publishers, in: The Wall Street Journal, 20.03.2024.
- Grynbaum, M. G./Mac, R., The Times sues OpenAI and Microsoft over A.I. use of copyrighted work, in: The New York Times., 27.12.2023.
- Sentinella, R./Schnegg, M./Möller, K., A Management Control oriented Governance Framework for Artificial Intelligence, in: Die Unternehmung, 77. Jg. (2023), H 2, S. 162–184.
- Simons, R., Levers of control: how managers use innovative control systems to drive strategic renewal. Harvard Business Press 1995.
- European Commission, EU AI Act, 2012, artificialintelligenceact.eu, Stand: 12.08.2024.