Datum: 22.05.2026

INSIDE IT

Online-Ausgabe

inside-it.ch
8400 Winterthur
043/ 243 95 50
https://inside-it.ch/

Medienart: Online
Medientyp: Fachmedien
Page Visits: 220'000

Webansicht

Auftrag: 1093187
Themen-Nr.: 666014

Referenz: 75c93504-68a0-4b0b-bb67-62dd5ed33176
Ausschnitt Seite: 1/4

SATW insights: Wie Schweizer Firmen KI verantwortungsbewusst einsetzen

22.05.2026 Manuel Kugler

Ein Chatbot halluziniert falsche Produktdaten. Ein Bewerbungsalgorithmus benachteiligt Frauen. Was nach lösbaren Problemen klingt, entpuppt sich in der Praxis als komplexes Geflecht aus Ethik, Rechtsfragen und Verhalten.

Wer heute nach Orientierung für den verantwortungsvollen Einsatz von KI sucht, findet reichlich Material. Der "EU AI Act" setzt rechtliche Leitplanken. Die "Assessment List for Trustworthy AI" der EU-Kommission liefert einen umfangreichen Fragenkatalog zu sieben Dimensionen von Human Agency über Technical Robustness bis hin zu Accountability. Dutzende Unternehmen haben eigene Ethical AI Principles veröffentlicht. Das Problem ist aber, dass diese Frameworks zeigen, was sie erreichen sollten. Aber nicht, wie.

Nehmen wir das Prinzip Transparenz. Ein grosses Wort, aber was es für eine IT-Fachperson, die ein neues KI-System einführen soll, genau bedeutet, ist unklar. Muss jede einzelne Modellentscheidung erklärbar sein? Reicht eine allgemeine Beschreibung des Funktionsprinzips? Wem gegenüber muss was transparent sein? Vor, während oder nach der Nutzung?

Im Folgenden geben vier Unternehmen Einblicke, wie sie verantwortungsvolle KI bei sich umsetzen. Sie präsentierten ihre Initiativen an dem Anlass "Responsible AI in der Praxis", die von der Data Innovation Alliance, der Digital Society Initiative der Universität Zürich und der Schweizerische Akademie der Technischen Wissenschaften im Juni vergangenen Jahres organisiert wurde.

Walbusch: Angestellte unterstützen, nicht ersetzen

Das Modeunternehmen Walbusch setzt auf KI-gestützte Automatisierung im Kundenservice. Thorsten Schmelz, Bereichsleiter Kundenservice bei Walbusch, fasst die Unternehmensphilosophie so zusammen: Transparenz schaffen, Ängste entschärfen, niemanden verlieren. Das Unternehmen organisierte Zukunftswerkstätten, in denen Mitarbeitende KI-Tools selbst ausprobieren, Feedback geben und mitgestalten konnten. Das Ziel war von Anfang an klar: Mitarbeitende unterstützen, nicht ersetzen.

Konkret zeigt sich das in der technischen Umsetzung. Wenn das System eine E-Mail automatisch bearbeitet, kommuniziert es das offen: "Um die Anliegen unserer Kundinnen und Kunden aber dennoch möglichst schnell zu klären, lassen wir vereinzelt E-Mails automatisiert bearbeiten. In Ihrem Fall hat unser System erkannt, dass Sie eine Einstellung der Werbezusendungen wünschen. Dieses wurde bereits in Ihrem Kundenkonto vermerkt."

Direkt darunter steht: "Hat sich unser System geirrt oder haben Sie noch weitere Anliegen in Ihrer Nachricht aufgeführt? Das kann vorkommen und wir bitten, dies zu entschuldigen. Bitte antworten Sie in diesem Fall kurz auf diese Mail."

Diese Transparenz hat mehrere Effekte. Sie schafft Vertrauen bei den Kundinnen und Kunden, reduziert die Angst der Mitarbeitenden, für Fehler des Systems verantwortlich gemacht zu werden, und liefert wertvolles Feedback. Jede Korrekturmail zeigt, wo das System noch Schwächen hat.

Bei Walbusch wird die Technologie nicht zu Marketingzwecken eingesetzt. KI übernimmt die repetitiven Standardanliegen, damit sich die Mitarbeitenden auf die komplexen Fälle konzentrieren können. Die Arbeit wird anspruchsvoller, nicht weniger. Dieser positive und transparente Umgang mit KI nützt allen und so bleibt keine Person auf der Strecke.

BSI: Code of Conduct AI

ARGUS DATA INSIGHTS® WISSEN ZUM ERFOLG
ARGUS DATA INSIGHTS® Schweiz AG | Rüdigerstrasse 15, Postfach, 8027 Zürich
T +41 44 388 82 00 | E mail@argusdatainsights.ch | www.argusdatainsights.ch

BSI Software entwickelt IT-Lösungen für Banken, Versicherungen und Retailer. Das Schweizer Unternehmen hat einen eigenen Code of Conduct AI erarbeitet, der bei der Umsetzung von jedem KI-System zum Einsatz kommt, das BSI für den Produktiveinsatz entwickelt. Bei dessen Entwicklung stützte sich BSI auf vorhandene Grundlagen: die Swico Ethik Charta, das Data Fairness Label von Swiss Insights sowie den EU AI Act.

Der Code of Conduct umfasst sechs Grundsätze: Schadensvermeidung, Fairness, Selbstbestimmung, Transparenz, Verantwortung und ethischer Diskurs. Dieses Framework spannt sich über sämtliche Projektphasen – vom Projekteingang, über die Umsetzung und den ethischen Diskurs bis hin zur Übergabe an die Kundin.

Der Prozess beginnt mit einer systematischen Stakeholderanalyse. Für ihren Proof-of-Concept zu KI-assistierter E-Mail-Beantwortung identifizierte das BSI-Team Customer-Service-Agents mit Toolzugriff, Agents ohne Zugriff, Kundinnen, Administratoren der Wissensdatenbank, das Projektteam selbst und potenziell auch Medien, die über das Tool berichten könnten. Jede Gruppe hat andere Interessen und Risiken.

Dann folgte die Risikoanalyse und das Team fragte sich, welche "Dark Patterns" auftreten könnten: Ist es möglich, dass das System Mitarbeitende emotional manipuliert, eine Antwort zu versenden, die sie sonst geprüft hätten? Gibt es Hürden für die Nutzenden, Fehler zu melden? Sind Default-Einstellungen so gewählt, dass sie bestimmte Entscheidungen begünstigen?

Das Team definierte einen kontinuierlichen Validierungsprozess. Das System wurde dann so entwickelt, dass KI-generierte Textpassagen visuell hervorgehoben werden. Mitarbeitende sehen auf den ersten Blick, was von der Maschine stammt. Das erhöht die Wahrscheinlichkeit, dass sie genau diese Stellen kritisch prüfen.

Swisscom: Risk-based Governance

Der Telekommunikationskonzern Swisscom entwickelt seit Jahren KI-Lösungen für den Kundenservice, Netzwerkoptimierung und interne Prozesse. Mit der Swiss AI Platform verfolgt das Unternehmen einen explizit verantwortungsvollen Ansatz, der auf Trust, Transparency, Reliability und Neutrality basiert.

Bei der Entwicklung von internen KI-Anwendungen kommt ein umfassendes AI Governance Framework nach dem Risk-based Approach der EU-KI-Verordnung zur Anwendung. Dieses hilft dabei, Anwendungsfälle sicher und verantwortungsvoll sowie unter Berücksichtigung von Datenschutzgesetzen und dem EU AI Act umzusetzen.

Jedes KI-System durchläuft zunächst eine Risikoklassifizierung: Prohibited (verboten), High Risk (hohes Risiko), Low Risk (geringes Risiko) oder Minimal Risk (minimales Risiko). Die Klassifizierung erfolgt anhand eines einfachen Fragebogens mit Kriterien des EU AI Acts plus Swisscom-spezifischer Anforderungen. Je nach Kategorie, in die eine Anwendung fällt, müssen Entwicklungsteams unterschiedliche Massnahmen ergreifen.

Ein Beispiel ist die Inspektion von Glasfaserkabelschächten. Swisscom entwickelte ein KI-System, das Anomalien in der Infrastruktur erkennen kann. Doch das Trainieren einer solchen KI ist nicht unkritisch. Denn obwohl darauf keine Personen zu erkennen sind, handelt es sich bei den Bildaufnahmen um sensitive Daten: Schächte mit Glasfaserkabeln zählen zur kritischen Infrastruktur. Hier halfen synthetische Trainingsdaten dabei, die KI-Lösung verantwortungsvoll und sicher zu entwickeln.

Daniel Dobos, Forschungsdirektor bei Swisscom, betont die Bedeutung der Systematik: "Wir können nicht bei jedem Use Case von vorne anfangen. Das Framework gibt uns Orientierung, ohne Innovation zu bremsen." Ein zentraler Punkt ist auch die Dokumentation: Swisscom protokolliert alle Governance-Entscheidungen, um später gegenüber Auditorinnen und Auditoren oder im Rahmen des EU AI Act nachweisen zu können, dass man sorgfältig gearbeitet hat.

Liip: Quantifizierbare Qualitätsmetriken

Die Schweizer Digitalagentur Liip entwickelte mit "Züricity GPT" einen Chatbot für die Stadt Zürich, der Fragen zur

Stadtverwaltung beantwortet. Dabei stand das Team vor einer grundsätzlichen Frage: Wie stellt man sicher, dass ein KI-System im öffentlichen Sektor die Anforderungen an Würde, Persönlichkeit und Eigenverantwortung erfüllt, wie es die Kantonsverfassung vorschreibt?

Gemäss dem Projektverantwortlichen Max Reichen wandte Liip dafür den hippokratischen Eid der Medizin auf KI an: Das System soll den Nutzenden dienen und keinen Schaden anrichten. Konkret hat Liip ein Evaluationssystem entwickelt, das unter anderem zwei zentrale Metriken kontinuierlich misst:

Answer Correctness: Wie faktisch korrekt ist die Antwort? Das System gleicht die Ausgaben mit offiziellen Informationen der Stadt Zürich ab und bewertet jede Antwort. Bei einem Test lag die Korrektheit bei 73%. Ein klares Signal, wo noch Verbesserungsbedarf besteht.

Faithfulness: Bleibt das System bei den Fakten oder halluziniert es? Diese Metrik prüft, ob die KI nur Informationen aus der Wissensbasis verwendet oder eigene "Erfindungen" hinzufügt. Im Test erreichte das System 84% Faithfulness.

Die Metriken werden nicht nur einmalig, sondern kontinuierlich während des Betriebs erhoben. Jede Antwort wird automatisch bewertet und dokumentiert. Das Team kann so gezielt nachsteuern, wenn bestimmte Themenbereiche problematisch sind. Bei einer Frage zu Asthma lieferte das System beispielsweise eine fachlich inkorrekte Antwort. Das Evaluationssystem erkannte dies sofort und markierte die Antwort als fehlerhaft.

Max Reichen beschreibt den Entwicklungsprozess als kontinuierlichen Zyklus: Experimentieren – Probieren – Verbessern. Im Zentrum stehen dabei immer die Bedürfnisse der Nutzenden. "Wir können nicht alles vorhersehen", sagt Reichen. "Aber wir können systematisch messen, transparent kommunizieren und kontinuierlich besser werden."

Die Nutzungsdaten zeigen: 66% der Anfragen kommen von städtischen Mitarbeitenden, 32% von externen Nutzenden. Nur 2% entfallen auf Liip selbst. Das System ist also tatsächlich im produktiven Einsatz und nicht nur ein Vorzeigeprojekt. Auf "Züricity GPT" folgten denn auch weitere Chatbots wie beispielsweise für den Kanton Basel-Stadt und das Genfer Universitätsspital HUG.

Fünf Empfehlungen für die Praxis

Die vier Beispiele zeigen, dass der verantwortungsbewusste Einsatz von KI kein theoretisches Konzept, sondern praktisch umsetzbar ist. Allerdings erfordert dieser mehr als gute Absichten. Fünf Empfehlungen dienen zur Orientierung:

Stakeholder systematisch einbinden. Nicht nur informieren, sondern aktiv in den Entwicklungsprozess integrieren. Wer Bedenken äussern kann, trägt zum Erfolg bei.

Ethische Bewertung kontinuierlich durchführen. Nicht bis zum Go-live warten. Die kritischen Weichenstellungen erfolgen bei der Datenauswahl und beim Modelldesign. Bereits da müssen die ethischen Fragen gestellt werden.

Von Prinzipien zu Metriken übersetzen. "Das System soll fair sein" ist keine prüfbare Anforderung; eine bessere Formulierung ist: "Die Falsch-Positiv-Rate darf zwischen Gruppe A und B um maximal fünf Prozent abweichen".

Transparenz technisch umsetzen. KI-generierte Inhalte klar kennzeichnen, einfache Korrekturmöglichkeiten bieten und relevante Nutzungsmuster für Qualitätssicherung protokollieren.

Dokumentation von Anfang an. Der EU AI Act verlangt nachvollziehbare Risikomanagement-Systeme. Wer Entscheidungen, Tradeoffs und Validierungen von Beginn an dokumentiert, spart später enormen Aufwand.

Die Tools und Frameworks existieren bereits. Was fehlt, ist oft der Mut, sie konsequent anzuwenden. Denn die Einführung eines verantwortungsvollen KI-Systems kostet zunächst Zeit und Geld. Die Frage ist: Was kostet es, sie nicht zu machen?

Über den Autor

Über die Kolumne

Manuel Kugler. Foto: zVg