



Wenn KI Code schreibt, wird die Anforderung zum Engpass

Als ein österreichischer Versicherer eingehende Schadenmeldungen mit KI erfassen lassen wollte, begann das Projekt mit einem Workshop bei den Sachbearbeiter:innen. Auf ihren Haftnotizen standen Sätze wie »Eintippen nervt« und »Alle Daten sammeln nervt«.

Bild: zVg

Im Gespräch kamen Regeln zum Vorschein, die in keiner Anforderung auftauchten. So lässt sich die Sparte nach der Erfassung nicht mehr ändern, das Schadendatum nur innerhalb desselben Jahres, und für 40 bis 50 Prozent der Fälle braucht die Deckungsprüfung noch ein Altsystem. Erfahrene Sachbearbeiter:innen kennen solche Regeln, ein KI-Agent kennt sie nicht.

Für Banken, Versicherungen und Behörden ist das die eigentliche Hürde auf dem Weg zur KI-nativen Softwareentwicklung, in der Menschen und KI-Agenten gemeinsam größere Teile der Entwicklung übernehmen. In klassischen Projekten beschreibt ein Ticket meist nur das Was, und die Lücken füllen Menschen mit implizitem Wissen, mit Rückfragen beim Fachbereich oder in Gesprächen auf dem Flur. Ein Agent merkt dagegen oft nicht, dass ihm Wissen fehlt, etwa wie eine Ausnahmeregel bei der Kreditvergabe gemeint ist, und trifft eine plausible Annahme, die sich später als falsch herausstellt. Als fleißiger Umsetzer baut er auch eine unvollständige Anforderung in hohem Tempo. In einer Bank wird aus einer solchen Annahme im Code schnell ein Befund der Innenrevision. »Wo der Plan Lücken hat, füllt sie der Agent mit Annahmen«, sagt Tobias Kilga, Head of BU Product, Data & AI bei SQUER, einem Technologie- und Transformationspartner mit Fokus auf angewandte KI im DACH-Raum. Je schneller die Umsetzung wird, desto mehr hängt das Ergebnis davon ab, wie gut die fachlichen Regeln beschrieben sind. »Wer KI nur in der Entwicklung einsetzt, verschiebt den Engpass bloß. Den ganzen Nutzen realisiert nur, wer KI entlang der gesamten Wertschöpfungskette einsetzt, vom Fachbereich bis in den Betrieb. Alles andere bleibt ein lokales Optimum«, sagt Manuel Klein, Mitgründer von SQUER.

SQUER setzt deshalb auf Intent Engineering, bei dem die Absicht hinter einem Vorhaben ausdrücklich beschrieben wird. Die Idee ist mit Spec-driven Development verwandt. Der Unterschied liegt nach Kilgas Darstellung darin, dass das Warum versioniert mit dem Code verbunden bleibt, lesbar für die Agenten und nachvollziehbar für alle, die später verstehen wollen, warum etwas so gebaut wurde.

In regulierten Branchen müssen Sicherheit und Nachvollziehbarkeit von Beginn an Teil der Entwicklung sein, und viele Organisationen setzen dafür bisher auf lange Freigabeketten am Projektende.

Fragt die Revision in zwei Jahren, warum Glasbruch anders behandelt wird als ein Sturm Schaden, soll die Antwort im Intent neben dem Code stehen. Dafür hat SQUER eine eigene Rolle eingeführt, den Intent Engineer, der zwischen Fachbereich, Produkt und Technik arbeitet und den fachlichen Kontext gemeinsam mit den Fachexpert:innen schärft. Harness Engineers bauen den technischen Harness, also die Arbeitsumgebung mit Regeln, Werkzeugen und Prüfungen, in der die Agenten arbeiten. Die Verantwortung für die Absicht und für die Prüfung der Ergebnisse bleibt bei Menschen.

Intents sind in Ebenen gegliedert. Übertragen auf den Versicherer beschreibt ein Strategic Intent das Ziel, Schäden schneller zu bearbeiten. Ein Domain Intent legt ein messbares Ergebnis fest, etwa dass bei Glasbruch die Pflichtfelder einer Meldung automatisch erkannt werden. Ein Tactical Intent beschreibt ein Arbeitspaket, das ein Team überblicken kann, zum Beispiel das Auslesen dieser Felder aus einer E-Mail. Intents entstehen dort, wo die Arbeit stattfindet, durch Beobachtung und Gespräche wie im Workshop beim Versicherer. Sie beschreiben das Was und das Warum; Technologieentscheidungen und Code-Strukturen legen sie bewusst nicht fest.

In regulierten Branchen müssen Sicherheit und Nachvollziehbarkeit von Beginn an Teil der Entwicklung sein, und viele Organisationen setzen dafür bisher auf lange Freigabeketten am Projektende. In der KI-nativen

Softwareentwicklung werden fachliche Regeln und regulatorische Vorgaben dagegen versioniert und laufend geprüft. Der Harness legt fest, auf welche Modelle, Werkzeuge und Daten Agenten zugreifen dürfen, und verankert automatisierte Prüfungen in der CI/CD-Pipeline. Weil Modellanbieter nach DORA als IKT-Drittdienstleister gelten, regelt er auch, welche Anbieter zum Einsatz kommen und wo die Daten verarbeitet werden. An festgelegten Stellen prüfen Menschen oder geben frei. Liegen Vorgaben etwa aus dem EU AI Act oder von BaFin und FMA als strukturierte Intents vor, werden sie von Anfang an mitgeprüft, und die Compliance-Abteilung ist früh eingebunden. Eine Studie des Stanford Digital Economy Lab zu erfolgreichen KI-Einführungen zeigt, dass Sicherheitsanforderungen, die anfangs bremsen, später den Umgang mit sensiblen Daten erst möglich machen.

»Governance gehört in das System, nicht ans Ende einer Freigabekette. Nur so können Teams mit KI schneller arbeiten, ohne die Kontrolle abzugeben«, sagt Tobias Kilga. Moralisch handeln kann eine KI nicht, erklärt Kilga. Die Verantwortung dafür bleibt bei Menschen. Schwieriger ist das Vier-Augen-Prinzip. Wenn Agenten in diesem Tempo Code schreiben, kann niemand mehr jede Zeile lesen. Das zweite Augenpaar bleibt – es schaut nur woanders hin. Automatisierte Prüfungen lesen jede Zeile, Menschen prüfen Absicht, Architektur und alles, was das System als riskant markiert. Welche Stellen als riskant

gelten, legen Menschen vorher fest, und die Freigabe bleibt menschlich. Dass sich jede Änderung lückenlos auf eine freigegebene Absicht zurückführen lässt, ist das Ziel.

Beim Versicherer zeigte sich schon in der Analyse, dass die historischen Daten zahlreiche Unstimmigkeiten enthielten. Ein Großteil der Arbeit floss deshalb in Datenqualität und Sicherheit, vom geprüften Referenzdatensatz über die Pseudonymisierung bis zum Konzept für CISO und Datenschutz. Das Team begann mit Glasbruch, Marderschäden und Verstopfungen, drei Schadenarten, die selten Rückfragen erzeugen, und legte die Abnahmekriterien vor der Entwicklung fest. Heute arbeiten die Sachbearbeiter:innen mit echten Schadenmeldungen im System. Was sie früher abgetippt haben, schlägt ihnen die KI vor, und sie prüfen und korrigieren es.

Der erste Schritt zur KI-nativen Softwareentwicklung ist deshalb aus Kilgas Sicht selten ein weiterer Coding-Assistent. Er rät Unternehmen, vorher zu klären, welches Wissen fehlt, was die neue Lösung bewirken soll und woran sich erkennen lässt, ob sie funktioniert. Wie wenig das Gefühl dabei taugt, zeigt eine Studie des Forschungsinstituts METR aus dem Jahr 2025. Erfahrene Open-Source-Entwickler:innen schätzten darin, dass KI sie um rund 20 Prozent beschleunigt habe, gemessen waren sie 19 Prozent langsamer. Die Werkzeuge sind seither besser geworden, die Studie zeigt aber, wie weit gefühlte und gemessene Produktivität auseinanderliegen können. Den durchgängigen Weg von der Spezifikation bis zur freigegebenen Software erprobt SQUER derzeit in einem ersten Projekt unter realen Bedingungen. Agenten bauen, prüfen gegen die Spezifikation und testen in Schleifen, bis ein Mensch das Ergebnis freigibt. Gemessen werden Kosten und Durchlaufzeit, aber auch Qualität und Prüfbarkeit, jeweils gegen eine klassische Schätzung desselben Umfangs, die beide Seiten vor dem Start festlegen. Die Rechenkosten der Agenten zählen mit. »Erst wenn die Zahlen stimmen, lohnt es sich, weitere Vorhaben auf diesem Weg umzusetzen«, sagt KI-Experte Tobias Kilga.

Text Christian Kolb

Der Engpass wandert

Banken, Versicherungen und Behörden gelten als langsam, wenn es um neue Technologien geht. Dabei können gerade sie viel gewinnen, wenn KI-Agenten Teile der Softwareentwicklung übernehmen. Dafür müssen Regeln und Verantwortung von Anfang an im System verankert sein.

Tobias Kilga, Head of BU Product, Data & AI bei SQUER, über den Konflikt zwischen Tempo und Kontrolle.



Tobias Kilga
Head of BU Product, Data & AI

Vor welchen besonderen Herausforderungen stehen regulierte Unternehmen?

Sie müssen jede Änderung begründen und belegen können, gegenüber Revision, Aufsicht und oft auch gegenüber Kund:innen. Ich habe vor SQUER jahrelang die Konzern-IT eines stark regulierten Medizintechnikherstellers geleitet und kenne das aus Sicht des Kundenunternehmens. Dazu kommen gewachsene Systemlandschaften, in denen fachliches Wissen nur noch im Code oder in den Köpfen einzelner Leute steckt. Und der Druck wächst, schneller zu werden, weil Fachkräfte fehlen und Kund:innen digitale Angebote erwarten. Viele lösen den Konflikt zwischen Tempo und Kontrolle, indem sie Freigaben verlängern. Das kostet Zeit, und Fehler fallen trotzdem erst spät auf. Mit KI wird der Konflikt schärfer, weil die Umsetzung plötzlich sehr schnell geht und die Prüfung nicht mithält.

Ist die digitale Transformation in erster Linie ein Hardware-, Software- oder Mindset-Thema?

Am ehesten ist sie eine Frage der Organisation. Hardware und Software lassen sich kaufen oder bauen, und auch ein offenes Mindset hilft wenig, wenn niemand entscheiden darf. Mit KI-Agenten wird diese Frage sehr konkret. Welche Entscheidungen darf ein Agent selbst treffen, wo braucht er eine Freigabe, und wer trägt die Verantwortung? Wer das offen lässt, hat am Ende schnelle Agenten und langsame Freigaben.

Warum ist der Einsatz von KI auch in regulierten Branchen so wichtig?

Weil bei ihnen viel IT-Kapazität in Wartung und Modernisierung alter Systeme steckt. Agenten können dort viel übernehmen, etwa Wissen aus altem Code herausarbeiten, Tests ergänzen und Migrationen vorbereiten. Eine Studie des Stanford Digital Economy Lab beschreibt einen Versicherer, der ein Altsystem neu geschrieben hat. Veranschlagt waren 5.000 Stunden mit sieben Leuten, gebraucht wurden 600 Stunden mit drei. Ein gut aufgesetztes agentisches System dokumentiert seine Arbeit vollständiger als die meisten manuellen Prozesse, was bei jeder Prüfung hilft.

Wie verändert sich der Entwicklungsprozess, wenn KI-Agenten mitarbeiten?

Früher war das Schreiben von Code der teure Teil. Heute ist es das Verstehen: Was soll das System leisten, welche Regeln gelten, woran erkennen wir, dass es richtig funktioniert? Ein Agent setzt eine Anforderung in Minuten um, egal ob sie gut oder schlecht ist. Wir stecken deshalb mehr Zeit in Intents und Akzeptanzkriterien vor der Umsetzung und in automatisierte Prüfungen danach.

Warum ist Intent Engineering die Voraussetzung dafür, KI in der Entwicklung zu skalieren?

Einzelne Entwickler:innen mit einem Coding-Assistenten kommen mit unscharfen Vorgaben zurecht, weil sie nachfragen und mitdenken. Mit zehn Agenten parallel geht das nicht mehr. Jede Unschärfe in der Anforderung vervielfacht sich. Intent Engineering sorgt dafür, dass die Absicht hinter einer Änderung strukturiert und überprüfbar vorliegt, bevor gebaut wird. Dann arbeiten Agenten zuverlässiger, und Menschen können die Ergebnisse gegen ein festgelegtes Ziel prüfen. Laut Stanford AI Index 2026 sind die Gewinne dort am größten, wo Aufgaben klar abgegrenzt sind und Qualität messbar ist.

Wie messen Sie den wirtschaftlichen Wert (ROI) von KI-Agenten im Entwicklungsalltag?

Gegen eine klassische Schätzung, die wir vor dem Start gemeinsam mit den Kund:innen festlegen, messen wir Kosten und Durchlaufzeit bis zur Produktion, einschließlich der Rechenkosten der Agenten. Dazu kommen Fehler, Nacharbeit und die fachliche Abnahme, weil schnell gebauter Code ohne Abnahme nichts wert ist. Langfristig zeigt sich der Wert für mich daran, ob das Kundenunternehmen danach selbst schneller weiterbauen kann.

Wenn Agenten in Echtzeit Code generieren und integrieren: Wie verändert sich die Fehlerkultur?



Governance gehört in das System, nicht ans Ende einer Freigabekette. Nur so können Teams mit KI schneller arbeiten, ohne die Kontrolle abzugeben.

Ein Agent kann denselben Fehler in kurzer Zeit an vielen Stellen wiederholen. Je später man ihn findet, desto teurer wird er. Wir prüfen deshalb laufend und automatisiert, bei jeder Änderung. Wenn ein Agent etwas falsch umsetzt, fragen wir zuerst, welcher Kontext gefehlt hat, und verbessern dann den Intent oder die Leitplanken. Ein Fehler wird damit zum Hinweis auf fehlenden Kontext, über den das Team offen spricht. So lernt das System mit, und derselbe Fehler wird mit jeder Runde unwahrscheinlicher.

Was ist die künftige Rolle des Menschen in diesem System?

Menschen legen fest, was gebaut wird und warum. Sie gestalten die Leitplanken und tragen die Verantwortung für das Ergebnis. Fachexpert:innen werden wichtiger, weil ihr Wissen direkt in die Intents einfließt. In unseren Projekten ist ihre Zeit heute oft der eigentliche Engpass. Entwickler:innen arbeiten stärker als Architekt:innen und Prüfer:innen, Routinearbeit nimmt ab. Unternehmen sollten deshalb in Weiterbildung investieren und Rollen wie den Intent Engineer schaffen.

Was können regulierte Unternehmen konkret tun, um jetzt zu starten?

Mit einem kleinen Vorhaben anfangen, das trotzdem zählt, etwa einem Altsystem, das ohnehin modernisiert werden muss, oder einem Prozess mit viel manueller Prüfung in der Dokumentenverarbeitung. Vorher festlegen, woran man Erfolg misst. Dann Fachbereich, Entwicklung und Compliance an einen Tisch holen und die Intents gemeinsam erarbeiten. Parallel braucht es eine technische Umgebung mit festen Regeln für Daten, Modelle und Freigaben, in der Agenten sicher arbeiten können. Erst nach den ersten Iterationen ausweiten.

Was tut SQUER ganz konkret für die Kund:innen?

Meistens beginnen wir mit einem Workshop von ein bis zwei Tagen bei den Kund:innen, in dem wir mit Fachbereich, IT und

Compliance klären, welches Problem zuerst gelöst wird und welche Vorgaben gelten. Danach arbeiten unsere Intent Engineers und Harness Engineers im Tandem mit Leuten der Kund:innen. Den Einstieg bildet ein klar abgegrenzter Pilot zum Festpreis mit vorher vereinbarten Kriterien. Verfehlen wir sie, tragen wir die Mehrkosten. Am Ende gehört die Software Factory den Kund:innen, und ihre eigenen Leute können damit Software bauen und ändern. Wissen, das vorher in wenigen Köpfen oder im alten Code steckte, liegt dann als Intent vor, den andere prüfen, übernehmen und weiterentwickeln können.

Wie hat KI die Unternehmenskultur bei Ihnen verändert?

Am deutlichsten beim Lernen. Mit einem eigenen Werkzeug messen wir, wie sicher unsere Leute mit KI arbeiten, und mit einem Bootcamp für KI-native Entwicklung schließen wir die Lücken, die wir dabei sehen. Auch die Teams außerhalb der Technik arbeiten inzwischen mit KI. Verändert hat sich außerdem, wie wir über Fehler sprechen. Mit Agenten wird schneller sichtbar, was nicht funktioniert, und dann muss es auch früh auf den Tisch. Ich halte diplomatische Unschärfe für aufgeschobene Kosten.

Wie wird der konsequente Einsatz von KI diese Branchen verändern?

Die Zeit von der Idee bis zum produktiven System wird kürzer. Wenn Agenten einen großen Teil der Umsetzung übernehmen, verliert das Wie an Gewicht, und die Frage, warum wir etwas bauen, wird wichtiger. Banken, Versicherungen und Behörden werden Software deshalb in kleineren, überprüfbaren Schritten planen, und wer seine fachliche Absicht aufschreiben kann, reagiert schneller auf neue Vorschriften und Wünsche der Kund:innen. Wie schnell das geht, weiß ich nicht. Dass es kommt, sehe ich in jedem Projekt.

Weitere Informationen unter:
squer.io

