

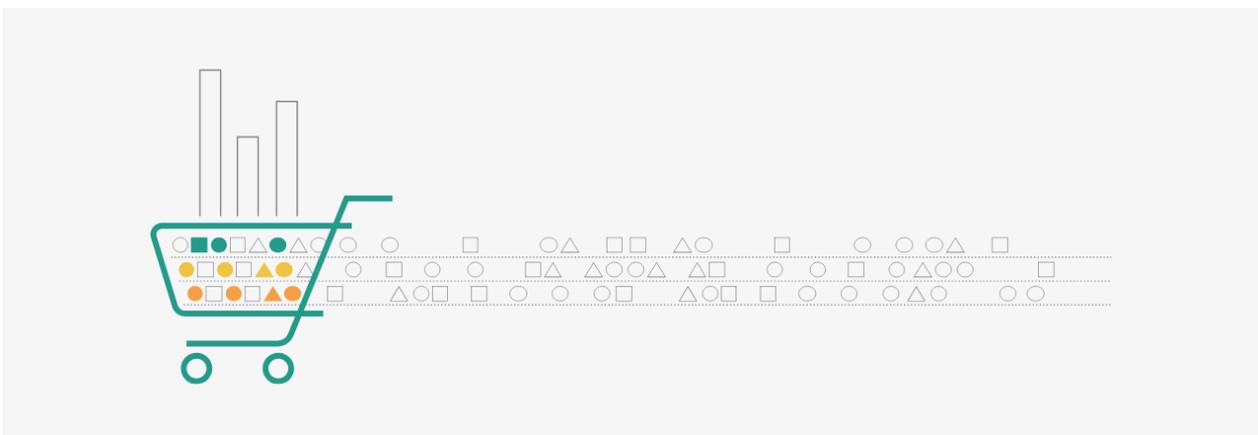
Grenzen und Potenziale von KI bei der Testdatenbereitstellung

Agentenbasierte KI bietet großes Potenzial für die Testautomatisierung – aber die Grundlage für jede Automatisierung sind reproduzierbare, konsistente, realistische und aktuelle Testdaten. Kann KI Testern dabei helfen, in Ihren Testumgebungen die dringend benötigte Datengrundlage zu schaffen?

Künstliche Intelligenz verändert kontinuierlich die Testlandschaft – sie verspricht, die Komplexität drastisch zu senken und Automatisierung zum Kinderspiel zu machen. Da KI bereits jetzt die Softwareentwicklung beschleunigt, benötigen Teams schnellen und zuverlässigen Zugang zu Testdaten, um effektiv Qualitätssicherung zu betreiben. Die kurzfristige Bereitstellung der richtigen Daten an die richtigen Personen wird zunehmend zum Engpass. Dies ist nicht nur eine Produktivitätsbremse, sondern führt im schlimmsten Fall zu verminderter Qualität und Sicherheit, da Tester gezwungen sind, sich auf veraltete Daten oder Workarounds zu verlassen, um ihre Ziele zu erreichen.

Agentenbasierte KI bietet großes Potenzial für die Testautomatisierung - aber die Grundlage für jede Automatisierung sind reproduzierbare, konsistente, realistische und aktuelle Testdaten. Kann KI Testern dabei helfen, in Ihren Testumgebungen die dringend benötigte Datengrundlage zu schaffen?

Datenverfügbarkeit – ein unterschätzter Schmerzpunkt, der Geld kostet



Datenbankadministratoren und Betreiber („Operators“) sehen sich einer immer komplexer und schneller werdenden Nachfrage gegenüber, was sich am Ende auf Produktivität, Qualität,

Test Data mit KI

Lösung XDM

Resilienz sowie Compliance auswirkt. Eine Gartner-Umfrage belegt, dass schlechte Datenqualität Organisationen inzwischen durchschnittlich 12,9 Millionen Dollar jährlich kostet (<https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality>).

Datenverfügbarkeit herzustellen ist alles andere als trivial, denn es gilt, gleich drei Herausforderungen zu überwinden:

1. Technische Hürden:

Es muss sichergestellt werden, dass referentiell intakte, realistische Testdaten aus verschiedenen Plattformen oder Datenbanken zur Verfügung gestellt werden.

2. Einhaltung von Datenschutzvorgaben:

Datenschutzvorschriften sind strikt zu befolgen, der Zugriff auf Produktionsdaten muss streng kontrolliert sowie lückenlos dokumentiert werden.

3. Unterschiedliche Anforderungen der Datenkonsumenten:

Die konkreten Anforderungen an die Testdaten variieren erheblich: Ein manueller Tester oder Entwickler benötigt spezifische Daten, um einzelne Funktionen zu testen oder Produktionsfehler zu reproduzieren, während z. B. Automatisierungingenieure reproduzierbare Massendaten für ihre nächtlichen Regressionstests benötigen.

All diese Anforderungen unterscheiden sich je nach Team und Abteilung, was eine erhebliche organisatorische Herausforderung mit sich bringt – vor allem in großen Unternehmen oder in stark regulierten Branchen.

KI allein kann das Problem der Testdatenverfügbarkeit nicht lösen



Es erscheint verlockend, das Problem „mit KI zu bekämpfen“ und Daten bei Bedarf von Grund auf neu zu generieren, anstatt auf maskierte Produktionsdaten zurückzugreifen. Wir sehen

Test Data mit KI

Lösung XDM

diese Herangehensweise – bestenfalls – als Teillösung:

Generative KI bringt inhärente Probleme mit sich, etwa Halluzinationen, mangelnde Reproduzierbarkeit und begrenzte Genauigkeit. Diese Herausforderungen sind für bestimmte Anwendungsfälle mitigierbar oder sogar hinnehmbar: zum Beispiel ist das automatisierte Ausführen von 80 % Ihrer Tests mithilfe von KI ein großer Produktivitätsschub, auch wenn die restlichen 20 % manuell erfolgen müssen. Doch was bedeutet das für die Generierung von Testdaten? Im Bereich der generativen KI wird eine Genauigkeit von 80 % häufig als zufriedenstellend betrachtet. Würden Sie riskieren, Ihre Falsch-Positiv-Rate um 20 % zu erhöhen, nur weil Ihre Testdaten inkorrekt oder inkonsistent sind?

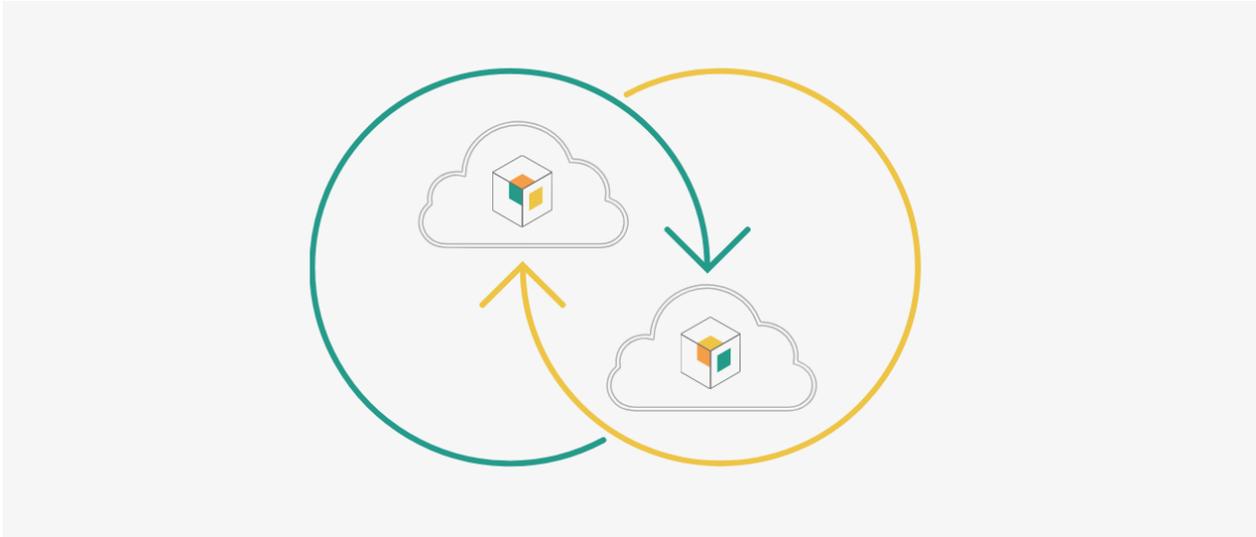
- **Selbst bei vollständig automatisierten Tests** muss nach jedem Durchlauf ein Mensch die fehlgeschlagenen Tests evaluieren – was das ursprüngliche Problem wieder aufwirft: Wie kann sichergestellt werden, dass nicht die Testdaten schuld an den Fehlern sind? Wie erhält der Prüfer JETZT einen referentiell intakten, realistischen Datensatz, um dies zu validieren?
- **Mangelnde Reproduzierbarkeit** von Testdaten führt dazu, dass zum Beispiel in jedem Testlauf andere Szenarien fehlschlagen. Das kann hier nicht der Anspruch sein: was in anderen Bereichen ggf. tolerabel ist, ist im Kontext Testdaten offensichtlich unverhandelbar.

Am Ende kommt man an Produktionsdaten bzw. einer geplanten Datenstrategie nicht vorbei, insbesondere wenn es um Randfälle und die Behebung von Produktionsfehlern geht. KI allein kann die nuancierte Realität von Produktionsdaten nicht ersetzen – insbesondere bei seltenen Konstellationen. KI-gestützte Datengenerierung hat ihren Platz in der Datenaugmentation, also im Füllen von Lücken und beim Bereitstellen von (großen) Datenmengen, wo in Produktion unzureichende Daten vorhanden sind. Hierbei sind Expertenwissen als auch praktische Erfahrung essenziell – sowohl auf Seiten der Betreiber als auch unter den Datenkonsumenten.

Vom reaktiven Prozess zur proaktiven Datenstrategie

Test Data mit KI

Lösung XDM



Viele Organisationen entwickeln immer komplexere, hausgemachte Lösungen, die selbst intern gewartet und qualitätsgesichert werden müssen. Dies bindet hochqualifizierte Administratoren („Operators“), deren Aufgabe es vor allem ist, die Entwicklungs-Pipeline am Laufen zu halten und die Produktivität der Teams sicherzustellen. Um wiederkehrende operative Aufgaben zu reduzieren, entscheiden sich immer mehr Unternehmen dafür, gebrauchsfertige Data Orchestration Software zu einzukaufen.

Neben der Aufwandsersparnis und erweiterten Funktionen verspricht der Wechsel von einem reaktiven Ansatz zu einer proaktiven Datenstrategie die Leistung sämtlicher Teams über Abteilungen hinweg erheblich zu verbessern. Häufig beobachten wir, dass nach der Pilotierung in einem Bereich andere Abteilungen schnell aufspringen und sich starke Synergieeffekte ergeben.

Wir empfehlen ein hybrides Modell, bei dem realistische Produktionsdaten bedarfsgerecht durch KI ergänzt werden und gepaart mit zielgerichteter Automatisierung steigende Produktivität unterstützen. Für kürzere Lead Times und eine bessere Developer-Experience muss eine ideale Datenorchestrierungsplattform:

- **Hochqualifizierte Operators entlasten:** Sie können sich auf strategisch wichtigere Aufgaben konzentrieren.
- **Daten-on-Demand ermöglichen:** Datenkonsumenten haben jederzeit Zugriff auf die benötigten Testdaten-Objekte – über Datenbank-Grenzen hinweg.
- **Compliance und Governance erzwingen:** Rollen-basierte Zugriffskontrolle und regulatorische Vorgaben sind innerhalb der Plattform konfigurierbar und jede Verteilung wird revisionssicher dokumentiert.
- **Flexibel anpassbar:** Das Tool ist auf die spezifischen Bedürfnisse Ihrer Organisation hin

Test Data mit KI

Lösung XDM

konfigurierbar und integriert sich nahtlos in vorhandene CI/CD-Prozesse.

- **KI gezielt nutzen:** Dort, wo Produktionsdaten fehlen oder nicht verfügbar sind, wird KI zur Datenaugmentation eingesetzt.
- **Konfiguration durch Generative KI:** Bei der Erstellung, Weiterentwicklung und Wartung der Konfiguration ist KI eine gute und oft gebrauchte Hilfe

Im folgenden Abschnitt beschreiben wir anhand eines Beispiels aus der Praxis, wie ein erfolgreicher Einsatz eines Data Orchestration Tools sowohl technische, regulatorische als auch organisatorische Herausforderungen nachhaltig löst.

Use Case: Versicherung optimiert und skaliert die Datenbereitstellung für mehrere QA-Teams mit vielfältigen Datenbanktechnologien



Eine Versicherung agiert in einem hart umkämpften und streng regulierten Markt. Mit zahlreichen Anwendungstestteams, die auf verschiedenen Plattformen wie Db2 z/OS, Db2 LUW, Oracle und MS SQL Server arbeiten, steht die Organisation vor der Herausforderung, robuste und qualitativ hochwertige Testdaten bereitzustellen, um ihre ständig weiterentwickelten Anwendungen zu unterstützen.

Bisher wurden Testdaten nur zweimal jährlich an die Akzeptanzumgebung geliefert. Diese unregelmäßige Bereitstellung brachte mehrere Probleme mit sich:

Test Data mit KI

Lösung XDM

- **Unzureichende Datenqualität:** Die halbjährliche Lieferung führte zu veralteten oder unvollständigen Datensätzen, die für umfassende Tests nicht ausreichten – viele potenzielle Fehler blieben dadurch unentdeckt.
- **Uneinigkeit der Testteams:** Die Abstimmung darüber, welche Testdatensätze benötigt werden und wie sie strukturiert sein sollten, gestaltete sich zwischen den beteiligten Testteams schwierig. Diese Abstimmungsprobleme führten zu inkonsistenten Testpraktiken und beeinträchtigten die Zuverlässigkeit der Testergebnisse.

Um diese Herausforderungen zu meistern, entschied sich die Versicherung für die Investition in ein fortschrittliches Datenmanagement-Tool. Nach der Evaluierung mehrerer Optionen im Rahmen eines rigorosen Proof-of-Concept (POC) kristallisierte sich UBS Hainers XDM als das einzige Tool heraus, das alle Anforderungen erfüllte. Ausschlaggebend waren:

- **Skalierbare Datenverarbeitung:** XDM konnte 59 Tabellen mit bis zu 100 Millionen Zeilen aus einer DB2-Umgebung kopieren und bewies damit seine Effizienz im Umgang mit großen Datenmengen.
- **Konsistente Subset-Extraktion:** Das Tool extrahierte zuverlässig ein konsistentes Subset dieser Tabellen – eine entscheidende Fähigkeit, um sicherzustellen, dass die Testumgebung die Produktionsumgebung realitätsgetreu abbildet.
- **Datenanonymisierung:** XDM bietet robuste, zentral integrierte Anonymisierungsfunktionen, die gewährleisten, dass sensible Produktionsdaten gemäß den Datenschutzbestimmungen transformiert werden.

Weitere Entscheidungsfaktoren waren der zentrale Steuerungspunkt von XDM, der den gesamten Prozess durch Automatisierung und Wiederholbarkeit optimiert, sowie der hervorragende Support von UBS Hainer, der sich während der Implementierung als entscheidend erwies.

Die Einführung des Orchestrierungs-Tools hat die Organisation grundlegend transformiert:

- **Tägliche Datenbereitstellung:** Testdaten werden nun täglich aus der Produktion geliefert, sodass Teams stets mit frischen, relevanten und aktuellen Informationen arbeiten können. Dieser dramatische Anstieg der Bereitstellungsfrequenz hat den Testzyklus erheblich beschleunigt und die Zeit bis zur Fehlererkennung verkürzt.
- **Maßgeschneiderte Datensätze:** Die Lösung unterstützt die individuellen Bedürfnisse der verschiedenen Testteams, sodass jedes Team einen speziell auf die zu testenden Funktionen oder Features ausgerichteten Datensatz erhält.
- **KI für Data Augmentation:** In Fällen, bei denen formale Korrektheit eine untergeordnete

Test Data mit KI

Lösung XDM

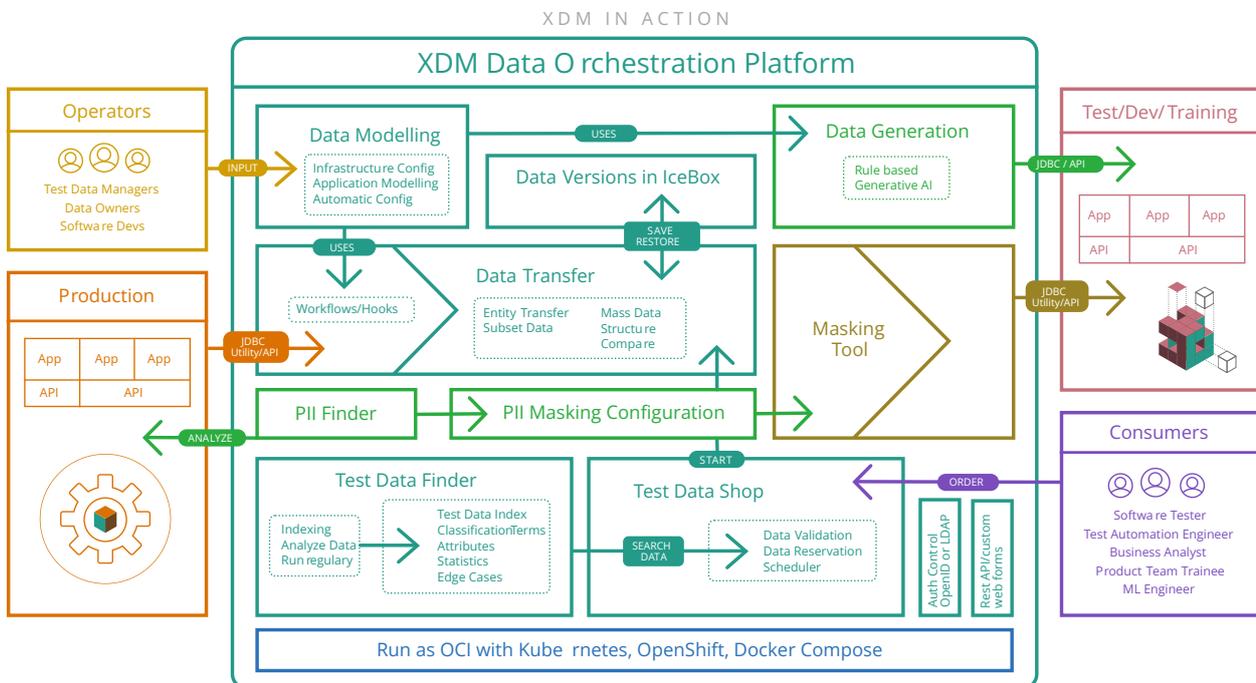


Rolle spielt (z.B. bestimmte funktionale Tests, Load Testing) oder wo Daten aus der Produktivumgebung in unzureichender Menge verfügbar sind (neue Features) werden Daten KI-gestützt generiert. Beispielsweise werden zu bestehenden Verträgen fiktive „Angebote“ generiert.

- **Verbesserte operative Effizienz:** Dank Automatisierung und zentraler Verwaltung wurde der manuelle Aufwand für Operators deutlich reduziert. Dies vermindert Fehlerquellen und ermöglicht es hochqualifizierten Mitarbeitern, sich auf strategisch wichtigere Aufgaben zu konzentrieren.
- **Gesteigerte Qualität:** Der Übergang zu einer kontinuierlichen Datenbereitstellung hat die Gesamtqualität der Tests, die operative Resilienz und die Einhaltung von Compliance-Vorgaben spürbar verbessert.
- **Pull-Effekt:** mehr und mehr Abteilungen planen, auf die neue Technologie umzusteigen.



Fazit: Zukunftssicherheit durch ganzheitliche Lösung



Durch den Einkauf einer fertigen Software-Lösung incl. Experten-Betreuung bei der Integration hat die Versicherung ihren Datenbereitstellungsprozess von einem langsamen, fragmentierten Ansatz in ein robustes, agiles und hochautomatisiertes System verwandelt, das über Abteilungsgrenzen hinweg neue Standards setzt. Diese ermöglicht es dem Unternehmen, strenge Qualitätsstandards aufrechtzuerhalten und gleichzeitig den

Test Data mit KI

Lösung XDM

Anforderungen eines sich rasant wandelnden regulatorischen und technologischen Umfelds gerecht zu werden. Die Kombination aus zentraler Steuerung, effizientem Umgang mit großen Datenmengen, konsistenter Extraktion und effektiver Anonymisierung erfüllt nicht nur die aktuellen Anforderungen, sondern positioniert die Organisation auch optimal für zukünftiges Wachstum und Innovation.