KAPITEL 2

An MLOps-Prozessen beteiligte Personen

Auch wenn Machine-Learning-(ML-)Modelle in erster Linie von Data Scientists erstellt werden, ist es ein weitverbreitetes Missverständnis, dass nur Data Scientists von robusten MLOps-Prozessen und -Systemen profitieren können. In Wirklichkeit ist MLOps ein wesentlicher Bestandteil der KI-Strategie eines Unternehmens und betrifft jeden, der am Lebenszyklus von ML-Modellen mitwirkt bzw. davon profitiert.

Dieses Kapitel behandelt die Rollen, die den beteiligten Personen im ML-Lebenszyklus jeweils zukommen. Außerdem bespricht es, mit wem sie im Rahmen einer hochwertigen MLOps-Strategie idealerweise verbunden sein und zusammenarbeiten sollten, um die bestmöglichen Ergebnisse aus den Bemühungen im Bereich des Machine Learning zu erzielen, und welche Anforderungen an MLOps sie unter Umständen haben.

Es ist wichtig, anzumerken, dass sich dieses Handlungsfeld permanent weiterentwickelt. Ständig kommen neue Berufsbezeichnungen auf, die hier vielleicht nicht aufgeführt sind, und es ergeben sich fortwährend neue Herausforderungen (oder Überschneidungen) bei den MLOps-Verantwortlichkeiten.

Bevor wir zu den Details vordringen, werfen wir einen Blick auf die folgende Tabelle, die uns einen ersten Überblick verschafft:

Rolle	Rolle im ML-Lebenszyklus	Anforderungen an MLOps
Fachexperten	• Stellen Geschäftsfragen, Ziele oder KPIs bereit, an denen ML-Modelle ausgerichtet werden sollen. • Bewerten und stellen fortlaufend sicher, dass die Leistung des Modells mit dem ursprünglichen Zweck übereinstimmt bzw. diesen erfüllt.	• Einfache Möglichkeit, die Leistung des eingesetzten Modells in geschäftlicher Hinsicht zu beurteilen. • Mechanismus bzw. Feedback-Schleife zur Kenntlichmachung von Modellergebnissen, die nicht mit den Geschäftsanforderungen übereinstimmen.

Rolle	Rolle im ML-Lebenszyklus	Anforderungen an MLOps
Data Scientists	• Erstellen Modelle, die auf die von Fachexperten eingebrachten geschäftlichen Fragen oder Belange eingehen. • Liefern operationalisierbare Modelle, damit diese in der Produktivumgebung und mit Produktionsdaten ordnungsgemäß verwendet werden können. • Bewerten die Qualität der Modelle (sowohl des Ausgangsmodells als auch der Modelle, mit denen getestet wird) in Zusammenarbeit mit Fachexperten, um sicherzustellen, dass sie die ursprünglichen Geschäftsfragen oder -anforderungen beantworten bzw. erfüllen.	• Automatisierte Modellpaketierung und -auslieferung für ein schnelles und einfaches (und dennoch sicheres) Deployment in der Produktion. • Möglichkeit, Tests zu entwickeln, um die Qualität der eingesetzten Modelle zu bestimmen und kontinuierliche Verbesserungen vorzunehmen. • Übersicht über die Leistung aller eingesetzten Modelle (einschließlich des Parallelbetriebs von Tests) von einer zentralen Stelle aus. • Möglichkeit, die Datenpipelines der einzelnen Modelle zu untersuchen, um schnelle Bewertungen und Anpassungen vorzunehmen, unabhängig davon, wer das Modell ursprünglich erstellt hat.
Data Engineers	• Optimieren den Zugang und die Nutzung von Daten, mit denen ML-Modelle betrieben werden.	• Transparenz hinsichtlich der Leistung aller eingesetzten Modelle. • Möglichkeit, alle Details der einzelnen Datenpipelines zu überblicken, um zugrunde liegende Probleme bei der Datenversorgung zu beheben.
Software Engineers	• Integrieren ML-Modelle in die Anwendungen und Systeme des Unternehmens. • Stellen sicher, dass ML-Modelle nahtlos mit anderen, nicht auf Machine Learning basierenden Anwendungen zusammenarbeiten.	• Versionierung und automatische Tests. • Möglichkeit, gleichzeitig an derselben Anwendung zu arbeiten.
DevOps	• Bauen einsatzfähige Systeme auf und testen diese auf Sicherheit, Leistung und Verfügbarkeit. • Verwalten Continuous-Integration/Continuous-Delivery-(CI/CD-)Pipelines.	• Nahtlose Integration von MLOps in die übergeordnete DevOps-Strategie des Unternehmens. • Lückenlose Deployment-Pipeline.
Modellrisikomanager/Auditoren	• Minimieren das Gesamtrisiko für das Unternehmen infolge des Einsatzes von ML-Modellen in der Produktion. • Stellen sicher, dass die internen und externen Anforderungen eingehalten werden, bevor ML-Modelle in die Produktion überführt werden.	• Zuverlässige, möglichst automatisierte Reporting-Tools für alle Modelle (aktuell oder jemals in Produktion), einschließlich der Datenhistorie.
Machine Learning Architects	• Gewährleisten eine skalierbare und flexible Umgebung für ML-Modellpipelines vom Design über die Entwicklung bis hin zur Überwachung. • Führen gegebenenfalls neue Technologien ein, die die Leistung von ML-Modellen in der Produktion verbessern.	• Allgemeiner Überblick über Modelle und ihre benötigten Ressourcen • Möglichkeit, Datenpipelines zu durchleuchten, um die Infrastruktur zu bewerten und anzupassen.

Fachexperten

Als erstes Profil im Rahmen der Aktivitäten von MLOps sind die Fachexperten (engl. *Subject Matter Experts*, SMEs) zu berücksichtigen; schließlich beginnt und endet der ML-Lebenszyklus mit ihnen. Wenngleich die datenorientierten Rollen (Data Scientists, Engineers, Architects usw.) über Fachwissen in vielen Bereichen verfügen, fehlt ihnen in der Regel ein tieferes Verständnis für das Geschäftsmodell und die Probleme oder Fragen, die mit ML-Modellen gelöst werden sollen.

Fachexperten müssen bzw. sollten zumindest in der Regel in den MLOps-Prozess *eingebunden werden* – um ihre klar definierten Ziele, Geschäftsfragen und/oder wichtigen Leistungskennzahlen (KPIs), die sie erreichen oder adressieren möchten, ausreichend berücksichtigen zu können. In einigen Fällen können diese sehr konkret definiert sein (z. B. »Um unsere Zahlen für das Quartal zu erreichen, müssen wir die Kundenabwanderung um 10 % reduzieren« oder »Wir verlieren X € pro Quartal aufgrund ungeplanter Wartungsarbeiten; wie können wir Ausfallzeiten besser vorhersagen?«). In anderen Fällen sind die Ziele und Fragen vielleicht weniger klar definiert (z. B. »Unsere Servicemitarbeiter müssen unsere Kunden besser verstehen, um ihnen im nächsten Schritt ein höherwertiges Produkt (Upselling) anbieten zu können« oder »Wie können wir die Menschen dazu bringen, mehr Produkte bzw. Anwendungen zu kaufen?«).

In Unternehmen mit funktionierenden Prozessen ist es nicht immer zwingend notwendig oder gar ideal, den Lebenszyklus von ML-Modellen mit einer fest vordefinierten Geschäftsfrage zu beginnen. Die Arbeit mit einem weniger klar definierten Geschäftsziel kann eine gute Gelegenheit für Fachexperten sein, im Vorfeld direkt mit den Data Scientists zusammenzuarbeiten, um das Problem besser zu umreißen und mögliche Lösungen zu erarbeiten, bevor überhaupt mit der Datenexploration oder den Modellexperimenten begonnen wird.

Ohne diese anfängliche Einbindung von Fachexperten wird riskiert, dass andere Datenexperten (insbesondere Data Scientists) den Lebenszyklus von ML-Modellen so angehen, dass sie Probleme lösen bzw. Lösungen anbieten, die nicht dem gesamten Unternehmen dienen. Letztendlich ist dies nicht nur für die Fachexperten nachteilig, die mit Data Scientists und anderen Datenexperten zusammenarbeiten müssen, um Lösungen zu entwickeln, sondern auch für die Data Scientists selbst, die möglicherweise Schwierigkeiten haben, einen Mehrwert zu stiften.

Wenn Fachexperten nicht in den ML-Lebenszyklus involviert sind, können Datenteams infolgedessen ohne konkrete Geschäftsergebnisse dastehen und dadurch Schwierigkeiten haben, die nötige Aufmerksamkeit und zusätzliches Budget oder Unterstützung zu erlangen, um fortgeschrittenere Analyseinitiativen fortzusetzen. Letztendlich ist dies schlecht für die Datenteams, für die Fachexperten und für das Unternehmen als Ganzes.

Um die Einbeziehung von Fachexperten besser zu strukturieren, können Methoden zur Modellierung von Geschäftsentscheidungen genutzt werden, um die zu lö-

senden Geschäftsprobleme zu formalisieren und die Rolle von Machine Learning im Hinblick auf die Lösung zu klären.

Modellieren von Geschäftsentscheidungen

Das Modellieren von Geschäftsentscheidungen liefert eine Blaupause für Entscheidungsprozesse, sodass Fachexperten ihre Belange direkt strukturieren und beschreiben können. Entscheidungsmodelle können hilfreich sein, weil sie ML-Modelle für Fachexperten in einen Kontext stellen. Dies ermöglicht, die Modelle in die Unternehmensrichtlinien einzubinden, und hilft den Fachexperten, den Kontext ihrer Entscheidung und die möglichen Auswirkungen von Modelländerungen vollständig zu verstehen.

Binden die Fachexperten die geschäftliche Entscheidungsmodellierung in die MLOps-Strategien ein, können sie auf effektive Weise sicherstellen, dass die Ergebnisse von ML-Modellen von denjenigen, die kein fundiertes Wissen darüber haben, wie die zugrunde liegenden Modelle selbst funktionieren, richtig eingeordnet werden.[1]

Fachexperten spielen nicht nur zu Beginn des Lebenszyklus eines ML-Modells eine Rolle, sondern auch am Ende (Postproduktion). Um zu verstehen, ob ein ML-Modell gut bzw. wie erwartet funktioniert, benötigen Data Scientists oft die Einschätzung der Fachexperten, um die Feedback-Schleife zu schließen, da herkömmliche Kenngrößen (z. B. Güte- bzw. Qualitätsmaße wie die Korrektklassifikationsrate, Relevanz und Recall bzw. Sensitivität) nicht ausreichen.

Die Datenwissenschaftler könnten zum Beispiel ein einfaches Modell zur Vorhersage von Kundenabwanderungen (*Churn Prediction*) erstellen, das in der Produktivumgebung eine sehr hohe Genauigkeit aufweist. Doch die Marketingmaßnahmen vermögen es dennoch nicht, die Abwanderung von Kunden zu verhindern. Aus geschäftlicher Sicht bedeutet dies, dass das Modell nicht funktioniert hat, und das ist eine wichtige Information, die an die Entwickler des ML-Modells zurückgegeben werden muss. Somit können diese eine andere mögliche Lösung finden, etwa die Einführung einer Uplift-Modellierung, die dem Marketing hilft, die potenziellen Kunden, die abwandern möchten, aber für Werbemaßnahmen empfänglich sein könnten, besser anzusprechen.

Angesichts der Rolle von Fachexperten im Lebenszyklus von ML-Modellen ist es beim Aufbau von MLOps-Prozessen entscheidend, dass eine einfache Möglichkeit für sie besteht, die erbrachte Modellleistung in geschäftlicher Hinsicht zu verste-

1 Modelle zur Entscheidungsfindung basieren auf einem Framework namens »Decision Model and Notation« (*https://oreil.ly/6k5OT*). Sie helfen dabei, Prozesse zu verbessern, Unternehmensprojekte durch feste Richtlinien effektiv zu verwalten, Predictive-Analytics-Projekte zu gestalten und handlungsorientierte Entscheidungsunterstützungssysteme wie z. B. Dashboards zu nutzen.

hen. Das heißt, sie müssen nicht nur die genannten Qualitätsmaße interpretieren können, sondern auch die Ergebnisse oder Auswirkungen des Modells auf den im Vorfeld identifizierten Geschäftsprozess. Darüber hinaus benötigen die Fachexperten bei unerwarteten Leistungsveränderungen eine skalierbare Möglichkeit, mittels MLOps-Prozessen Modellergebnisse zu erfassen, die nicht mit den Geschäftserwartungen übereinstimmen.

Zusätzlich zu diesen bewussten Feedback-Mechanismen sollte die MLOps-Strategie generell so aufgebaut sein, dass die Transparenz für Fachexperten erhöht wird. Das heißt, sie sollten in der Lage sein, MLOps-Prozesse als Ausgangspunkt für die Exploration der Datenpipelines hinter den Modellen zu nutzen, um zu verstehen, welche Daten verwendet werden, wie sie transformiert und angereichert werden und welche Art von ML-Methoden angewendet werden.

Für Fachexperten, die sich auch um die Compliance von ML-Modellen hinsichtlich interner oder externer Vorschriften kümmern, dient MLOps als zusätzliche Möglichkeit, die Transparenz und das Verständnis für derartige Prozesse zu erhöhen. Dazu gehört auch die Möglichkeit, gezielt einzelne Entscheidungen eines Modells einzusehen, um verstehen zu können, warum das Modell zu dieser Entscheidung gelangt ist. Dies sollte die statistische Bewertung und das an verschiedensten Stellen eingeholte Feedback zusätzlich ergänzen.

Letztlich ist MLOps für Fachexperten vor allem wichtig, um ein Feedback zu ermöglichen und ein Rahmenwerk zu haben, das eine gute Kommunikation mit den Data Scientists gewährleistet, die die Modelle entwickelt haben. Allerdings gibt es auch andere Anforderungen an MLOps – insbesondere in Bezug auf die Transparenz, die mit Responsible AI zusammenhängt –, die für Fachexperten relevant sind und sie zu einem wichtigen Teil des Gesamtkonzepts von MLOps machen.

Data Scientists

Die Bedürfnisse der Data Scientists sind die wichtigsten, die beim Aufbau einer MLOps-Strategie berücksichtigt werden müssen. In den meisten Unternehmen haben Data Scientists heute oft mit voneinander separierten Daten, Prozessen und Tools zu tun, was es für sie schwierig macht, ihre Arbeit auf effektive Weise zu skalieren. MLOps ist hervorragend dafür geeignet, dieser Problematik entgegenzutreten.

Obwohl die meisten die Rolle der Data Scientists im Lebenszyklus von ML-Modellen nur als den Teil der eigentlichen Modellerstellung interpretieren, ist sie viel umfassender – bzw. sollte es zumindest sein. Data Scientists müssen von Beginn an mit den Fachexperten zusammenarbeiten und dabei helfen, die Geschäftsprobleme so zu verstehen und zu formulieren, dass sie eine praktikable Lösung auf Basis von Machine Learning erstellen können.

In der Realität zeigt sich, dass dieser allererste, entscheidende Schritt im Lebenszyklus von ML-Modellen oft der schwierigste ist. Vor allem für Data Scientists ist er eine Herausforderung, weil ihre Ausbildung nicht darauf ausgerichtet ist. Sowohl for-

melle als auch informelle Data-Science-Programme an Universitäten und im Internet betonen die technischen Fähigkeiten und nicht unbedingt die Fähigkeiten zur effektiven Kommunikation mit Fachexperten aus dem Unternehmen, die wiederum in der Regel nicht sehr vertraut mit den Machine-Learning-Methoden sind. Auch hier können wieder Methoden zur geschäftlichen Entscheidungsmodellierung helfen.

Gleichzeitig stellt die Kommunikation mit den Fachexperten eine Herausforderung dar, weil sie schlicht Zeit kostet. Für Data Scientists, die sofort loslegen und sich die Hände schmutzig machen wollen, kann es eine Qual sein, wochenlang ein Problem zu formulieren und zu skizzieren, bevor sie mit der Lösung beginnen. Hinzu kommt, dass Datenwissenschaftler oft (räumlich, kulturell oder beides) vom Kern des Unternehmens und von den Fachexperten getrennt sind, sodass sie einfach nicht in den organisatorischen Rahmen eingebunden sind, der eine einfache Zusammenarbeit zwischen diesen Bereichen ermöglicht. Robuste MLOps-Systeme können dabei helfen, diese Herausforderungen zu bewältigen.

Sobald die erste Hürde überwunden ist, kann das Projekt je nach Unternehmen entweder an Data Engineers oder an Analysten übergeben werden, die einen Teil der anfänglichen Datenerfassung, -aufbereitung und -exploration übernehmen. In einigen Fällen verwalten die Data Scientists diese Teile des Lebenszyklus des ML-Modells selbst. Aber in jedem Fall kommen sie wieder ins Spiel, wenn es an der Zeit ist, das Modell zu entwickeln, zu testen, zu optimieren und anschließend in Betrieb zu bringen.

Nach dem Deployment gehört es zu den Aufgaben der Data Scientists, die Qualität des Modells ständig zu überprüfen, um sicherzustellen, dass die Funktionsweise in der Produktivumgebung den ursprünglich ins Auge gefassten Geschäftszielen bzw. -anforderungen entspricht. Die zugrunde liegende Frage in vielen Unternehmen ist häufig, ob Data Scientists nur die Modelle überwachen, an deren Entwicklung sie zuvor beteiligt waren, oder ob eine andere Person das gesamte Monitoring übernehmen sollte. Was würde im ersten Szenario passieren, wenn es zu einem Personalwechsel kommt? Im zweiten Szenario ist der Aufbau guter MLOps-Praktiken entscheidend, da die Person, die das Modell überwacht, auch in der Lage sein muss, schnell zu intervenieren und Maßnahmen zu ergreifen, wenn sich die Qualität des Modells verschlechtert und sich negativ auf den Geschäftsbetrieb auszuwirken beginnt. Sofern sie nicht diejenigen waren, die es aufgebaut haben, wie kann dann MLOps dafür sorgen, dass dieser Prozess reibungslos verläuft?

Operationalisierung und MLOps

Während des gesamten Jahres 2018 und zu Beginn des Jahres 2019 war Operationalisierung das wichtigste Schlagwort, wenn es um den Lebenszyklus von ML-Modellen und KI im Unternehmen ging. Einfach ausgedrückt, ist die Operationalisierung von Data Science der Prozess, in dem Modelle in die Produktion überführt und ihre Leistung im Hinblick auf die Geschäftsziele gemessen wird. Wie fügt sich

also die Operationalisierung in den Kontext von MLOps ein? MLOps geht bei der Operationalisierung noch einen Schritt weiter und übernimmt nicht nur die Überführung in die Produktion, sondern auch die Wartung dieser Modelle – und der gesamten Datenpipeline – in der Produktion.

Obwohl die Konzepte unterschiedlich sind, könnte man MLOps als die neue Form der Operationalisierung betrachten. Das heißt, dass in den Unternehmen, in denen viele der großen Hürden zur Operationalisierung beseitigt wurden, MLOps nun die nächste große Hürde darstellt, da sie die Unternehmen im Bereich der betrieblichen Nutzung von ML-Modellen vor Herausforderungen stellt.

Alle Fragen des vorherigen Abschnitts münden direkt in der Frage, welche Bedürfnisse Data Scientists letztlich in Bezug auf MLOps haben. Ausgehend vom Ende des Prozesses und rückwärts denkend, muss MLOps den Data Scientists einen umfassenden Überblick bieten über die Leistung aller eingesetzten Modelle sowie über alle Modelle, die im Rahmen von A/B-Tests getestet werden. Dabei ist jedoch nicht nur wichtig, die Modelle zu überwachen, sondern auch Maßnahmen zu ergreifen. Erstklassige MLOps-Strategien sollten Data Scientists die Flexibilität bieten, die besten Modelle unter den getesteten auszuwählen und sie einfach zu implementieren.

Transparenz zu schaffen, ist ein ebenso zentrales Thema bei MLOps. Daher ist es nicht verwunderlich, dass es auch Data Scientists wesentlich tangiert. Die Fähigkeit, Datenpipelines zu durchleuchten und auf diese Weise zügig Beurteilungen und Anpassungen vorzunehmen (unabhängig davon, wer das Modell ursprünglich erstellt hat), ist entscheidend. Eine automatisierte Modellpaketierung und -auslieferung für ein schnelles und einfaches (und dennoch sicheres) Deployment in der Produktion ist ein weiterer wichtiger Punkt, der die Transparenz erhöht, und ein entscheidender Bestandteil von MLOps, vor allem um für Data Scientists, Softwareentwickler und DevOps-Teams eine Vertrauensbasis zu schaffen.

Neben einer transparenten Gestaltung kommt auch der Effizienz eine tragende Rolle bei der erfolgreichen Umsetzung von MLOps-Aktivitäten zu – insbesondere wenn es um die Belange von Datenwissenschaftlern geht. In der Unternehmenswelt zählen Agilität und Geschwindigkeit. Das gilt für DevOps – und für MLOps verhält es sich nicht anders. Natürlich können Data Scientists Modelle ad hoc bereitstellen, testen und überwachen, aber sie werden enorm viel Zeit damit verbringen, das Rad mit jedem einzelnen ML-Modell neu zu erfinden – und das wird niemals zu skalierbaren ML-Prozessen im Unternehmen führen.

Data Engineers

Datenpipelines sind das Kernstück des Lebenszyklus von ML-Modellen, und Data Engineers sind wiederum die zentralen Akteure, wenn es um Datenpipelines geht. Da Datenpipelines häufig abstrakt und komplex sind, können Data Engineers aus MLOps einen großen Nutzen ziehen.

In großen Unternehmen ist die Verwaltung des Datenflusses jenseits der Anwendung von ML-Modellen ein Vollzeitjob. Abhängig vom technischen Stack und der Organisationsstruktur des Unternehmens können sich Data Engineers daher mehr auf die Datenbanken selbst als auf Pipelines konzentrieren (vor allem wenn das Unternehmen Data-Science- und Machine-Learning-Plattformen einsetzt, die die grafische Erstellung von Pipelines durch andere Datenexperten wie Businessanalysten erleichtern).

Trotz dieser je nach Unternehmen leicht unterschiedlichen Aufgabengebiete besteht die Rolle der Data Engineers im Lebenszyklus letztlich darin, die Abfrage und die Nutzung von Daten zu optimieren, um schließlich ML-Modelle betreiben zu können. Im Allgemeinen bedeutet dies, dass sie eng mit den Businessteams, insbesondere den Fachexperten, zusammenarbeiten, um die richtigen Daten für das jeweilige Projekt zu identifizieren und sie möglicherweise auch für die Nutzung vorzubereiten. Auf der anderen Seite arbeiten sie eng mit Data Scientists zusammen, um etwaige Probleme im Zusammenhang mit dem Datenflussmanagement zu beheben, die dazu führen könnten, dass sich ein Modell in der Produktion unerwünscht verhält.

Da Data Engineers eine zentrale Rolle im Lebenszyklus von ML-Modellen spielen und sowohl die Entwicklung als auch die Überwachung unterstützen, kann MLOps zu erheblichen Effizienzsteigerungen beitragen. Data Engineers benötigen nicht nur einen Einblick in die Leistung aller Modelle, die in der Produktion zum Einsatz kommen, sondern auch die Möglichkeit, einen Schritt weiterzugehen und direkt in einzelne Datenpipelines Einblick zu nehmen, um die zugrunde liegenden Probleme zu beheben.

Damit Data Engineers ihre Rolle möglichst effizient ausfüllen können (und auch andere, einschließlich der Data Scientists), darf MLOps nicht nur auf eine einfache Überwachung abzielen, sondern muss eine Brücke zu den zugrunde liegenden Systemen schlagen, um ML-Modelle überprüfen und optimieren zu können.

Software Engineers

Es wäre ein Leichtes, die klassischen Software Engineers aus den MLOps-Überlegungen auszuschließen. Doch es ist aus einer weiter gefassten organisatorischen Perspektive entscheidend, auch ihre Belange zu berücksichtigen, um eine kohärente unternehmensweite Machine-Learning-Strategie aufzubauen.

Software Engineers bauen in der Regel keine ML-Modelle. Andererseits erstellen die meisten Unternehmen nicht nur ML-Modelle, sondern auch klassische Softwareanwendungen. Es ist wichtig, dass Software Engineers und Data Scientists zusammenarbeiten, um zu gewährleisten, dass das Gesamtsystem funktioniert. Schließlich sind ML-Modelle nicht von den unternehmensweiten Prozessen isoliert; der Programmcode für das ML-Modell, das Training, die Tests und das Deployment müs-

sen allesamt in die CI/CD-Pipelines (*Continuous Integration/Continuous Delivery*), die auch von anderen Softwareanwendungen genutzt werden, integriert werden.

Betrachten wir zum Beispiel ein Einzelhandelsunternehmen, das ein ML-basiertes Empfehlungssystem für seine Webseite entwickelt hat. Das ML-Modell wurde von einem Data Scientist erstellt, aber um es in die Webseite, die noch weitere Funktionen bereithält, einzubetten, müssen zwangsläufig Software Engineers involviert werden. Ebenso sind die Software Engineers für die Wartung der gesamten Webseite verantwortlich, was auch die korrekte Funktionsweise der ML-Modelle in der Produktivumgebung einschließt.

Aufgrund dieses Zusammenspiels sind Software Engineers ebenfalls auf MLOps angewiesen, damit ihnen Details zur Modellleistung als Teil eines größeren Bilds der Softwareanwendungsleistung für das Unternehmen zur Verfügung stehen. MLOps bietet Data Scientists und Software Engineers die Möglichkeit, dieselbe Sprache zu sprechen und dasselbe Grundverständnis davon zu haben, wie verschiedene Modelle, die in den unterschiedlichen Unternehmensbereichen eingesetzt werden, in der Produktion zusammenarbeiten.

Wichtige Funktionen für die Software Engineers sind darüber hinaus die Versionierung (um sicher zu sein, mit welchem Modell sie gerade arbeiten), die Durchführung automatischer Tests (um so sicher wie möglich zu sein, dass das Modell, mit dem sie gerade arbeiten, funktioniert) und die Möglichkeit, parallel an derselben Anwendung zu arbeiten (dank eines Systems wie Git, das Branches und Merges erlaubt).

DevOps

MLOps wurde auf der Grundlage von DevOps-Prinzipien entwickelt, aber das bedeutet nicht, dass sie parallel als völlig getrennte und voneinander isolierte Systeme betrieben werden können.

DevOps-Teams haben zwei primäre Rollen im Lebenszyklus von ML-Modellen. Zum einen sind sie für die Durchführung und den Aufbau von operativen Systemen sowie für Tests zuständig, um Sicherheit, Leistung und Verfügbarkeit von ML-Modellen sicherzustellen. Zum anderen sind sie für das CI/CD-Pipeline-Management verantwortlich. Beide Rollen erfordern eine enge Zusammenarbeit mit Data Scientists, Data Engineers und Data Architects. Eine solche enge Zusammenarbeit ist natürlich leichter gesagt als getan, aber genau hier kann MLOps einen Mehrwert bieten.

Für DevOps-Teams muss MLOps in die breitere DevOps-Strategie des Unternehmens integriert werden und die Lücke zwischen traditionellem CI/CD und modernem Machine Learning schließen. Das erfordert Systeme, die sich grundsätzlich ergänzen und die es DevOps-Teams ermöglichen, die Tests für ML-Modelle genauso zu automatisieren wie die Tests für traditionelle Software.

Modellrisikomanager/Auditor

In bestimmten Branchen (insbesondere im Finanzdienstleistungssektor) ist die Funktion des Modellrisikomanagements (MRM) entscheidend für die Einhaltung von Vorschriften. Doch nicht nur Unternehmen stark regulierter Branchen können davon profitieren und sollten eine ähnliche Rolle installieren. MRM kann Unternehmen aus verschiedensten Branchen vor verheerenden Verlusten schützen, die durch schlecht funktionierende ML-Modelle entstehen. Außerdem spielen in vielen Branchen meist relativ arbeitsintensive Audits eine Rolle, wodurch MLOps auch hier zum Tragen kommen kann.

Wenn es um den Lebenszyklus von ML-Modellen geht, übernehmen Modellrisikomanager eine entscheidende Rolle, da sie nicht nur die Modellergebnisse analysieren, sondern auch die ursprüngliche Zielsetzung und die Geschäftsfragen, die die ML-Modelle zu lösen versuchen, um das Gesamtrisiko für das Unternehmen zu minimieren. Sie sollten zusammen mit Fachexperten gleich zu Beginn des Lebenszyklus einbezogen werden, um sicherzustellen, dass ein automatisierter, ML-basierter Ansatz an sich kein Risiko darstellt.

Und natürlich müssen sie eine Rolle beim Monitoring spielen – ihre traditionelle Rolle im Modelllebenszyklus –, um sicherzustellen, dass die Risiken in Schach gehalten werden, sobald die Modelle im Produktivbetrieb sind. Das MRM setzt auch zwischen den Phasen der Konzeption und des Monitorings an, denn nach der Modellentwicklung bzw. vor der Inbetriebnahme in der Produktion ist es ebenfalls unverzichtbar, um die Konformität mit internen und externen Anforderungen sicherzustellen.

MRM-Fachleute und -Teams profitieren in hohem Maße von MLOps, denn ihre Arbeit ist oft mühsame Handarbeit. Da das für das MRM zuständige Team und die Teams, mit denen sie arbeiten, häufig unterschiedliche Tools verwenden, kann eine Standardisierung einen enormen Geschwindigkeitsvorteil bei der Prüfung und dem Management von Risiken bieten.

Wenn es um spezifische Belange in Bezug auf MLOps geht, ist ein zuverlässiges Reporting-Tool für alle Modelle (unabhängig davon, ob sie derzeit in Produktion sind oder in der Vergangenheit in Produktion waren) am wichtigsten. Das Reporting sollte nicht nur Informationen zur Leistungsfähigkeit vorsehen, sondern auch die Möglichkeit, die Datenhistorie einzusehen. Ein automatisiertes Reporting erhöht die Effizienz von MRM- und Audit-Teams in MLOps-Systemen und -Prozessen zusätzlich.

Machine Learning Architects

Traditionelle Data Architects sind dafür verantwortlich, die gesamte Unternehmensarchitektur zu verstehen und sicherzustellen, dass diese die Anforderungen an den Datenbedarf des gesamten Unternehmens erfüllt. Im Allgemeinen beschäftigen sie sich damit, wie Daten gespeichert und verwendet werden sollen.

Heutzutage sind die Anforderungen an die Architects viel größer, und sie müssen oft nicht nur über die Besonderheiten der Datenspeicherung und -nutzung Bescheid wissen, sondern auch darüber, wie ML-Modelle wechselseitig funktionieren. Das macht die Rolle komplexer und erhöht ihre Verantwortung im MLOps-Lebenszyklus. Deshalb nennen wir sie in diesem Abschnitt *Machine Learning Architects* anstelle des eher traditionellen Titels *Data Architect*.

Machine Learning Architects spielen eine entscheidende Rolle im ML-Modell-Lebenszyklus, indem sie eine skalierbare und flexible Umgebung für Modellpipelines sicherstellen. Darüber hinaus benötigen die Datenteams ihr Fachwissen, um neue Technologien einzuführen (wenn es angebracht ist), die die Leistung von ML-Modellen in der Produktion verbessern. Aus diesem Grund reicht der Titel »Data Architect« nicht aus – diese Architects müssen nicht nur ein tiefgreifendes Verständnis von der Unternehmensarchitektur, sondern auch von Machine Learning haben, um diese Schlüsselrolle im ML-Modelllebenszyklus auszufüllen.

Diese Rolle erfordert eine Zusammenarbeit im gesamten Unternehmen – von den Data Scientists und Engineers bis hin zu den DevOps- und Software-Engineering-Teams. Wenn Machine Learning Architects die Belange all dieser Personen und Teams nicht vollständig verstehen, können sie die Ressourcen nicht richtig zuordnen, um eine optimale Leistung von ML-Modellen in der Produktion sicherzustellen.

Im Hinblick auf MLOps geht es bei der Rolle der Machine Learning Architects darum, einen zentralen Überblick über die Ressourcenzuweisung zu haben. Da sie eine strategische bzw. taktische Rolle einnehmen, benötigen sie einen Gesamtüberblick über die Gegebenheiten, um Engpässe auszumachen und diese Informationen zu nutzen, um langfristige Verbesserungen voranzutreiben. Ihre Rolle besteht darin, mögliche neue Technologien oder Infrastrukturen im Rahmen von Investitionsentscheidungen ausfindig zu machen, und nicht unbedingt darin, operative Schnelllösungen, die nicht direkt die Skalierbarkeit des Systems betreffen, herbeizuführen.

Abschließende Überlegungen

MLOps richtet sich nicht nur an Data Scientists. Eine Vielzahl von Experten aus dem gesamten Unternehmen spielt nicht nur im Lebenszyklus von ML-Modellen eine Rolle, sondern auch im Rahmen der MLOps-Strategie. Tatsächlich spielt jede Person – vom Fachexperten auf der Geschäftsseite bis hin zum technisch versierten Machine Learning Architect – eine entscheidende Rolle bei der Wartung von ML-Modellen in der Produktion. Dies ist letztendlich nicht nur wichtig, um die bestmöglichen Ergebnisse von ML-Modellen zu gewährleisten (gute Ergebnisse führen im Allgemeinen zu mehr Vertrauen in ML-basierte Systeme sowie zu einem höheren Budget für die Entwicklung weiterer Modelle), sondern auch – und das ist vielleicht noch wichtiger – um das Unternehmen vor den in Kapitel 1 dargelegten Risiken zu schützen.