

Philipp
Plank im
Dialog
mit Utz
Schäffer



„So viel Blackbox ist KI eigentlich nicht“

Die Bayer AG hat frühzeitig damit begonnen, die Voraussetzungen für den Einsatz von Künstlicher Intelligenz (KI) zu schaffen. Philipp Plank, Head of Decision Science for Enabling Functions im Konzern, spricht im Interview über den Einsatz von KI im Controlling, neue Möglichkeiten im Forecasting und die Frage, ob KI Controllern in Zukunft die Arbeit abnimmt.



© Fotos: Kai Myller

Dr. Philipp Plank

ist seit Februar 2020 Head of Decision Science for Enabling Functions bei der Bayer AG. Zuvor arbeitete er fünf Jahre lang als Data Scientist and Architect im Controlling bei Bayer Business Services, einer der drei ehemaligen Service-Gesellschaften des Konzerns. Als Global Content Manager war er von 2015 bis 2016 im Unternehmen für Controlling- und Finanzdatenmodelle zuständig. Plank studierte Wirtschaftsinformatik an der Universität Köln, wo er im Jahr 2014 seine Promotion im Bereich Controlling und Planung abschloss.

Herr Dr. Plank, Sie sind seit Februar 2020 Head of Decision Science for Enabling Functions bei Bayer. Was sind Ihre Aufgaben? Und warum heißt es Decision und nicht Data Science?

Wir sind der Kernansprechpartner für digitale Transformationsthemen und IT, und zwar für alle Funktionen, die den CFO und CEO betreffen. In diesem Bereich, speziell in der Decision Science, sind wir zum größten Teil für Data-Science-Themen zuständig. Bei uns stehen also Daten und analytische Modelle im Vordergrund und die Frage, was man daraus machen kann. Es geht weniger um die Methode der Data Science als vielmehr darum, welche Entscheidungen wir konkret mit unserer Analyse unterstützen möchten. Deshalb haben wir unseren Bereich Decision Science genannt. Inhaltlich beschäftigen wir uns stark funktionsabhängig mit unterschiedlichen Methoden, im Finanzbereich eher mit Forecasting-Themen, in Bezug auf den Audit-Bereich beispielsweise viel mit Mustererkennung, im Bereich Kommunikation mit Stakeholder-Analysen, Netzwerken und Ähnlichem. Insgesamt ist es unser Ziel als Decision Scientists, gute, rationale Entscheidungen im Rahmen eines Miteinanders möglich zu machen. Die jeweiligen Abteilungen entwickeln die Digitalisierungsstrategie und verfügen über ein sehr starkes Fachwissen. Sie haben aber durchaus noch Ideenbedarf in Bezug auf die Möglichkeiten, die ihnen die Daten bieten. Unsere Aufgabe als Decision Science sehen wir auch darin, die jeweilige Abteilung auf das vorhandene Potenzial hinzuweisen, so nach dem Motto „Passt mal auf, ihr habt dieses

bestimmte Daten-Set und damit könnt ihr dieses und jenes machen. Ist euch das eigentlich bewusst?“

Spannend. Bayer hat ja 2019 auch den Digital Award der Hackett Group für das Statistical Demand Forecasting gewonnen. Können Sie uns das Projekt näher erläutern?

Ausgangspunkt war die Frage, inwieweit wir unseren operativen Planungsprozess in der Supply Chain automatisieren können. Wir haben ein automatisiertes Vorhersagemodell für unsere operative Demand-Planung etabliert, das 20 statistische und auf Machine Learning basierende Modelle zu einem Forecast vereint. Es geht aber über die reine Erzeugung eines Forecasts hinaus und umfasst eine mögliche Automatisierung des gesamten Planungsprozesses inklusive Datenaufbereitung, wie beispielsweise Saisonalitäten, Ausreißer-Analysen, Zeitreihenauffüllung, White-Noise-Tests bis zur Integration in unser Standard Reporting. Um die Akzeptanz zu steigern, gibt es ein Deep Dive Reporting, so können wir die Forecast-Erzeugung objektivieren und Transparenz schaffen. Der rollierende Forecast wird auf Produktebene erzeugt.

Und das ist auch alles schon umgesetzt, also im laufenden Einsatz erprobt?

Unser Projekt ist global ausgerollt und wird in mehr als 24 Ländern eingesetzt, die Automatisierungsrate ist nachweisbar erfolgreich. Der Use Case wird also tatsächlich umgesetzt. Wir



bauen hier zum großen Teil auf Industrie-Standards auf, im Zusammenspiel aller Komponenten haben wir aber kaum Vergleichbares gefunden. Wobei es bei uns nicht um den technischen Aspekt des Forecasts als solchen geht, sondern um die Frage: Lässt sich etwas überhaupt so weit vorhersagen, dass es zur Unterstützung bei Entscheidungen dienen kann? So weit, dass man dem jeweiligen Planer einen Hinweis geben kann, ob er einem Forecast vertrauen kann und ihn wirklich einsetzen sollte oder nicht. Unser Projekt ist eine Komplettlösung, die in den Planungsprozess integriert ist – und nicht einfach eine weitere Zahl im System.

Wie lange hat es denn gedauert, diese Lösung zu realisieren? Wann fiel der Startschuss?

Der Ursprung – das data.one-Projekt – liegt inzwischen weiter zurück. Vor sieben Jahren haben wir als Basis aus dem Zentral-Controlling ein Projekt zur Konsolidierung aller Business Warehouses ins Leben gerufen. Zuvor hatte jeder Teilkonzern seine eigene Business-Intelligence-Landschaft, wir konnten die Berichte einzelner Bereiche nur schwer miteinander vergleichen. Um eine einheitliche Sicht auf das konsolidierte Konzernergebnis möglich zu machen, haben wir für data.one deshalb konzernweit Daten aus den verschiedensten Bereichen harmonisiert und in einem Data Warehouse zusammengeführt. Mit dem Projekt haben wir eine neue Steuerungs- und Prozesslogik in unserem Konzern etabliert und gezeigt, wie Predictive Analytics die Planungseffizienz und -qualität erhöhen kann. Das Projekt stellt inzwischen die Zentralisierung aller Business



Warehouses und auch das Management Reporting auf ein gemeinsames Fundament. Ein klares Ziel von data.one war von vornherein, dass wir nicht nur mit deskriptiven, sondern auch mit präskriptiven Modellen unterstützen wollen: sprich aus Vorhersagemodellen auch konkrete, wenn möglich automatisierte Entscheidungsalternativen zur Verfügung zu stellen.

Und konnten Sie dieses Ziel auch erreichen?

Die Hauptherausforderung im Nachgang war es, eine zentrale Lösung zu entwickeln, die überall gleich gut einsetzbar ist. Wir sind jetzt auf Basis von data.one in der Lage, Finanzdaten direkt neben operativen Daten zu nutzen. Das ermöglicht uns das schon erwähnte Demand Forecasting, eine bessere Bedarfsermittlung und -kontrolle. Wir können also zum Beispiel einen Bestand vorhersagen und bestimmte Kennzahlen sofort integrieren. Wir sind damit quasi in der Lage, Forecasts nicht in Sekundenschnelle, aber zumindest in der Art zu liefern, dass wir am nächsten Tag eine Kostenprognose für das gesamte Unternehmen zur Verfügung stellen können. Ohne so eine zentrale, integrierte Basis wäre das nicht möglich. Aber nicht nur Analytics und Datenintegration sind Werttreiber, sondern auch das Gesamt-Design. Wir haben zum Beispiel eigens eine Reporting-Lösung entwickelt, die besonders für das Standard-Reporting auf genau das zugeschnitten ist, was wir speziell brauchen, inklusive Storytelling und geführter Navigation.

Wie muss man sich das vorstellen? Das System sagt also „Lieber Planer, dieser Forecast ist sehr verlässlich“?



Genau. Am Anfang war die Lösung rein historisch basiert. Mittels verschiedenster Verfahren wurde auf dieser Grundlage die Zukunft bestimmt. Wir haben also aus der Historie gelernt, inwieweit ein bestimmter Forecast vertrauenswürdig ist oder nicht. Inzwischen ist das System so weit entwickelt, dass es von Iteration zu Iteration lernt und selbst feststellt, in welchen Situationen ein Forecast gemacht werden kann und in welchen nicht. Davon ausgehend lassen sich klare Aussagen treffen, zum Beispiel ob ein bestimmtes Produkt stabil ist. Oder ob unser Forecast gut läuft und als Basis für weitere Überlegungen genutzt werden sollte. Natürlich haben wir aber auch schon gesehen, dass wir bestimmte Effekte nicht erklären können. In solchen Fällen muss ein Experte sich das genauer anschauen. Und natürlich kommt auch irgendwann der Punkt, an dem wir sagen müssen: Es geht nicht, wir können das nicht voraussagen. Wichtig ist, so etwas in solchen Fällen auch genauso zu kommunizieren. Viele der Lösungen, die heutzutage verkauft werden, suggerieren oft sehr stark, dass sie alles können. Ich glaube, man muss da objektiver rangehen.

Wie nehmen die Nutzer den maschinellen Forecast an?

Wir schauen uns das sehr genau an und machen unsere Bewertung diesbezüglich an drei Merkmalen fest. Erstens die Anwenderperspektive: Wir fragen uns also, ob das System eines ist, mit dem man wirklich gerne arbeitet. Zweitens die Qualität, gemessen an historischen Benchmarks, naiven Forecast-Modellen und Vergleichen zu manuellen Forecasts. Und drittens: die Automatisierungsraten, die meistens, aber nicht immer eine logi-

sche Konsequenz aus erstens und zweitens sind. In Bezug auf die Automatisierung haben wir in bestimmten Produkt-Länder-Kombinationen Raten bis zu 70 Prozent. Neben diesen Top-Vorhersagen haben wir global gesehen eine sehr gute Rate, wo dem Planer empfohlen wird, unsere Vorschlagswerte als Basis zu nutzen. Natürlich gibt es auch Ausreißer in bestimmten Ländern, wo das System gar nicht funktioniert. Wir sind auf das Feedback der Endnutzer angewiesen, denn wir sind als zentrale Funktion leider manchmal ein bisschen weit weg von den konkreten Anwendungsfällen. Dankenswerterweise werden unsere Lösungen aber gerne angenommen. Gerade während der COVID-19-Pandemie haben wir gesehen, dass die Adaption-Rate für bestimmte Produkte gestiegen ist. Unsere Interpretation ist hier, dass die Planer sich auf entscheidende, strategische Produkte fokussiert und für andere Produkte den Algorithmen vertraut haben. Systeme haben aber natürlich auch ihre Grenzen, man muss akzeptieren und grundsätzlich auch verstehen, dass der menschliche Input die Lösung in bestimmten Bereichen komplementiert.

Was haben Sie getan, um die Akzeptanz von Anfang an sicherzustellen?

Wir haben die Bedenken unserer Nutzer ernst genommen und uns mit ihnen auch über unsere Bedenken ausgetauscht. Diese Strategie war für uns von großem Vorteil. Natürlich können wir nicht alle Bedenken ausräumen, gerade im Hinblick auf den gefühlten Blackbox-Charakter der Lösung. Ich kann nicht bei jedem Modell reingucken und erklären, wieso es exakt so ent-

schieden hat. Der Vergleich mit Menschen ist da manchmal recht praktisch, da beispielsweise ein neuronales Netz ähnlich dem menschlichen Gehirn arbeitet, und in ein menschliches Gehirn kann ich auch nicht hineinschauen und analysieren, warum es bestimmte Entscheidungen getroffen hat.

Aber einen Menschen kann ich fragen!

Ja, genau deshalb ist die Art und Weise, wie der Nutzer mit dem System kommunizieren kann, wichtig. Je mehr er noch das Gefühl hat, Herr der Lage zu sein und nicht einfach nur eine Zahl serviert zu bekommen, mit der er leben muss, desto größer ist die Akzeptanz. Er hat dann eher das Gefühl von Kontrolle. Das ist wissenschaftlich auch im Feld von Algorithm Aversion untersucht worden, also der Frage, weshalb der Mensch dazu neigt, der Maschine nicht zu trauen. Natürlich kann ich einen Menschen fragen. Der Vorteil, auch im Sinne einer Auditierbarkeit ist aber, dass ein Algorithmus oder KI bei gleicher Situation und Informationslage überprüfbar gleiche Ergebnisse erzeugen kann. Aber natürlich steigt die Akzeptanz mit der Qualität. Wenn ich drei- oder viermal erlebt habe, dass ich selbst besser bin als das System, kommt schnell das Feedback: In unserem Bereich funktioniert diese Lösung nicht.

Und was machen Sie dann?

Die Frage ist, ob ich für diesen Bereich eine Lösung adaptieren kann oder ob es nicht vielmehr an der Situation oder Datengrundlage gelegen hat, beispielsweise an einer fehlenden Historie für ein Produkt. Vielmehr muss man aber die Diskrepanz in der Denkweise in Bezug auf die Lösungen grundsätzlich beheben, Tools zeigen und erklären. So viel Blackbox ist KI eigentlich nicht. Heute gibt es zum Beispiel Möglichkei-

Bayer AG

Bayer ist ein weltweit tätiges Unternehmen mit Kernkompetenzen auf den Life-Science-Gebieten Gesundheit und Ernährung. Der Konzern, zu dessen bekanntesten Marken unter anderem das Schmerzmittel Aspirin und die Wundsalbe Bepanthen zählen, umfasst die operativen Bereiche Pharmazeutika, Verbrauchergesundheit und Pflanzenforschung. Im Geschäftsjahr 2019 erzielte der Konzern mit rund 104.000 Beschäftigten einen Umsatz von 43,5 Milliarden Euro. Die Investitionen beliefen sich auf 2,9 Milliarden Euro und die Ausgaben für Forschung und Entwicklung auf 5,3 Milliarden Euro.

ten, mit sogenannten Explainable AI Tools Entscheidungen besser rational zu erklären. Nehmen wir das Beispiel Vertrieb: Ich habe einerseits meine Verkaufszahlen und sehe eine Korrelation, wenn beim Konkurrenten die Verkaufszahlen steigen. Darüber hinaus gibt es technische Möglichkeiten, wie man nicht nur anhand einzelner, sondern auch im Zusammenspiel verschiedener Variablen erklären kann, warum beispielsweise ein Receivables Forecast in eine bestimmte Richtung geht. Das funktioniert beispielsweise mittels sogenannter Shapley Values, einem Konzept aus der Spieltheorie. Wenn ich solche Informationen mitliefere, steigt die Akzeptanz bei den Nutzern. Aus meiner Sicht ist es aber oft auch sinnvoll, viel zu investieren, um das Vertrauen in die Lösung zu stärken. Indem man zum Beispiel erklärt, warum gewisse KI-Lösungen stabil sind – und dass man das nicht nur anhand der Qualität, sondern auch der Sensitivität einer Lösung gegenüber einer Veränderung zeigt. Spielt man beispielsweise mit den Input-Daten, sieht der Nutzer, dass sich das System auf veränderte Ereignisse, wie zum Beispiel die COVID-19-Pandemie, einstellt. So können wir zeigen, dass sich die Maschine adjustiert, im Zweifelsfall sogar besser, als der manuelle Forecast das kann. Auch im Sinne einer Auditierbarkeit bieten Algorithmen durchaus eine Chance, da nachweislich bei gleichem Input der gleiche Output erzeugt werden kann.

Gibt es noch andere Anwendungsbeispiele im Finanzbereich, in denen Bayer KI einsetzt?

Sicherlich sind Forecasts von diversen KPIs ein dominantes Thema. Im Bereich Transfer Pricing haben wir Lösungen entwickelt, die eng mit dem Forecasting-Thema verknüpft sind. Aber auch im Bereich Fraud Detection, beispielsweise um Doppel- oder Falschzahlungen zu identifizieren, sind Lösungen im Einsatz, die auf Machine Learning basieren.

Zurück zu Ihrem beruflichen Alltag. Als Head of Decision Science sind Sie Dienstleister für die dezentralen Einheiten und die Corporate-Ebene – genau wie die Controller. Wie funktioniert das denn im Zusammenspiel?

Aus meiner Sicht passen die beiden Bereiche gut zusammen. Als Partner für die Entscheidungsunterstützung im Management muss der Controller die besten Vorhersagemodelle und Informationen transportieren und verstehen können. Und meines Erachtens muss hier konsequent die angestoßene Evolution fortgeführt werden, sozusagen eine Post-Excel-Ära. Die eigentliche Frage hierbei ist, wie beide Bereiche in diesem Zusammenhang voneinander abgegrenzt werden: Wie weit geht Business Ana-

lytics und wo fängt der Data-Science-Block an? Aber ich finde, auch in dieser Hinsicht harmonieren wir gut, denn wir haben eben auch den Gegenpart, der versteht, was ein statistisches Forecasting liefern kann und was seine Grenzen sind. Einen, der im Zweifelsfall sogar bei den Algorithmen mitreden kann, sich aber auch darauf verlässt, dass die genau das tun, was sie tun sollen. Gleichzeitig geht es auch auf IT-Seite nicht ohne ein Grundverständnis der betriebswirtschaftlichen Komponenten.

Können Manager in fünf oder zehn Jahren auch selbst mit benutzerfreundlichen Analytics-Lösungen arbeiten? Brauchen wir für Routineanwendungen dann gar keine internen Dienstleister wie Decision Scientists oder Controller mehr?

Ich habe da eine etwas kontroverse Meinung. Sicherlich, man kann am Markt relativ viele Tools kaufen, die unterstützen. Allein an Dashboarding-Lösungen gibt es eine ganze Bandbreite, die auch alle ihre Validität haben. Allerdings haben beispielsweise Lösungen zu cloudbasierten Forecast Services im Dashboard oben eine Auswahlmöglichkeit mit drei, vier verschiedenen Forecast-Algorithmen. Sprich: Ich habe im Unternehmen auf einmal drei, vier Forecasts, multipliziert mit Mitarbeitern, die auch diesen Knopf gedrückt haben. Dieser Umstand limitiert aus meiner Sicht schon diese Analyse, denn damit öffnet man quasi die Büchse der Pandora – und zwar dahin gehend, dass Entscheidungen nur noch technologiegetrieben und nicht mehr rational getroffen werden. Natürlich bieten uns diese Lösungen gleichzeitig auch unheimlich viele Möglichkeiten. Generell sollte es aus meiner Sicht aber klare Richtlinien dafür geben, welche Lösungen wir selbst entwickeln, welche wir einkaufen und wann welche Lösung in einem bestimmten Kontext führend ist.

Gibt es darüber hinaus Entscheidungssituationen, bei denen Sie langfristig Potenzial für Anwendungsmöglichkeiten der verschiedenen Formen von KI im Controlling sehen? Wenn ja, welche?

Letztlich müsste man diese von den zukünftigen Kernaufgaben des Controllings ableiten. Meine Wahrnehmung ist hier, dass das Controller-Bild aktuell mindestens einer starken Dynamik unterliegt, viele neue Rollen sind hinzugekommen. Das hat auch die letzte WHU Delphi-Studie zum Controller-Bild gezeigt. Viel Zeit wird im Controlling heute neben Planung auch für Kommentierung im Berichtswesen und Storytelling investiert. Für beide Bereiche gibt es bereits Anwendungen wie Smart Data Discovery oder automatisierte Kommentierungslösungen. Von echter KI, die Potenziale entdeckt, Zusammenhänge er-

kennt und Handlungsalternativen erzeugt, sind wir aber noch weit entfernt. Wenn das Controlling sich den operativen Prozessen weiter nähert, entfaltet sicherlich KI – oder besser Maschine oder Deep Learning – erst recht ihre Potenziale. Prozesse zu optimieren und Potenziale zu entdecken ist in diesen Bereichen bereits weit fortgeschritten. Das Controlling könnte hier ein Bindeglied zwischen den Funktionen darstellen. Die Frage zielt also eher auf das Grundverständnis von Controlling und darauf, wo wir das Controlling in zehn Jahren sehen.

KI kann Entscheidungen unterstützen, kann sie Managern auf Sicht auch Entscheidungen abnehmen?

Die Gretchenfrage hier ist: Was kann die Technologie in Zukunft? Ich persönlich gehöre nicht zu den Menschen, die glauben, dass wir in fünf Jahren eine komplette KI haben, die tatsächlich auch versteht und fühlt, was sie tut, ungeachtet aller Erfolge von beispielsweise dem Unternehmen Deepmind. Die meisten Algorithmen, die wir heute einsetzen, funktionieren sehr gut in einem speziellen Kontext, aber es fällt ihnen schwer, den Kontext zu wechseln und zu verstehen. Ein Natural-Language-Processing-Algorithmus zum Beispiel versteht gerade noch, dass er Englisch bearbeitet, aber ein fundamentales Verständnis über den Kontext fehlt ihm, geschweige denn die Fähigkeit, dies mit einem anderen Kontext zu verknüpfen. Solange das nicht funktioniert, werden bestimmte Interaktionsformen, die anfallen, wenn der Controller mit als Business Partner auftritt, nicht wegfallen. Einen großen Vorteil haben wir meiner Ansicht nach jedoch im Bereich Automatisierung beziehungsweise Robotics und der Pre-Datenanalyse. In den Abteilungen Entscheidung und Support werden diese immer weiter ausgebaut. Ob mir die KI letztlich die Entscheidung abnehmen kann, hängt sehr stark von der Art meines Problems ab. Wenn es ein operatives Problem ist, was vielleicht einem Schachspiel ähnelt, kann das gut funktionieren. In einem hochkomplexen Lösungsraum, in dem ich zwischen Strategie A oder B entscheiden muss, sehe ich noch nicht, dass die aktuelle technische Unterstützung ausreicht. Und ich glaube, auch in naher Zukunft wird sich das nicht ändern.

Herr Dr. Plank, ich danke Ihnen für das Gespräch.

Das Gespräch führte Prof. Dr. Utz Schäffer, Direktor des Instituts für Management und Controlling (IMC) der WHU – Otto Beisheim School of Management, Vallendar, und Mit-herausgeber der Controlling & Management Review.

E-Mail: utz.schaeffer@whu.edu