«RESPONSIBLE AI» KÜNSTLICHE INTELLIGENZ

Text: **Peter Kasahara, Christian Westermann, Jörg Gerigk, alle PwC**

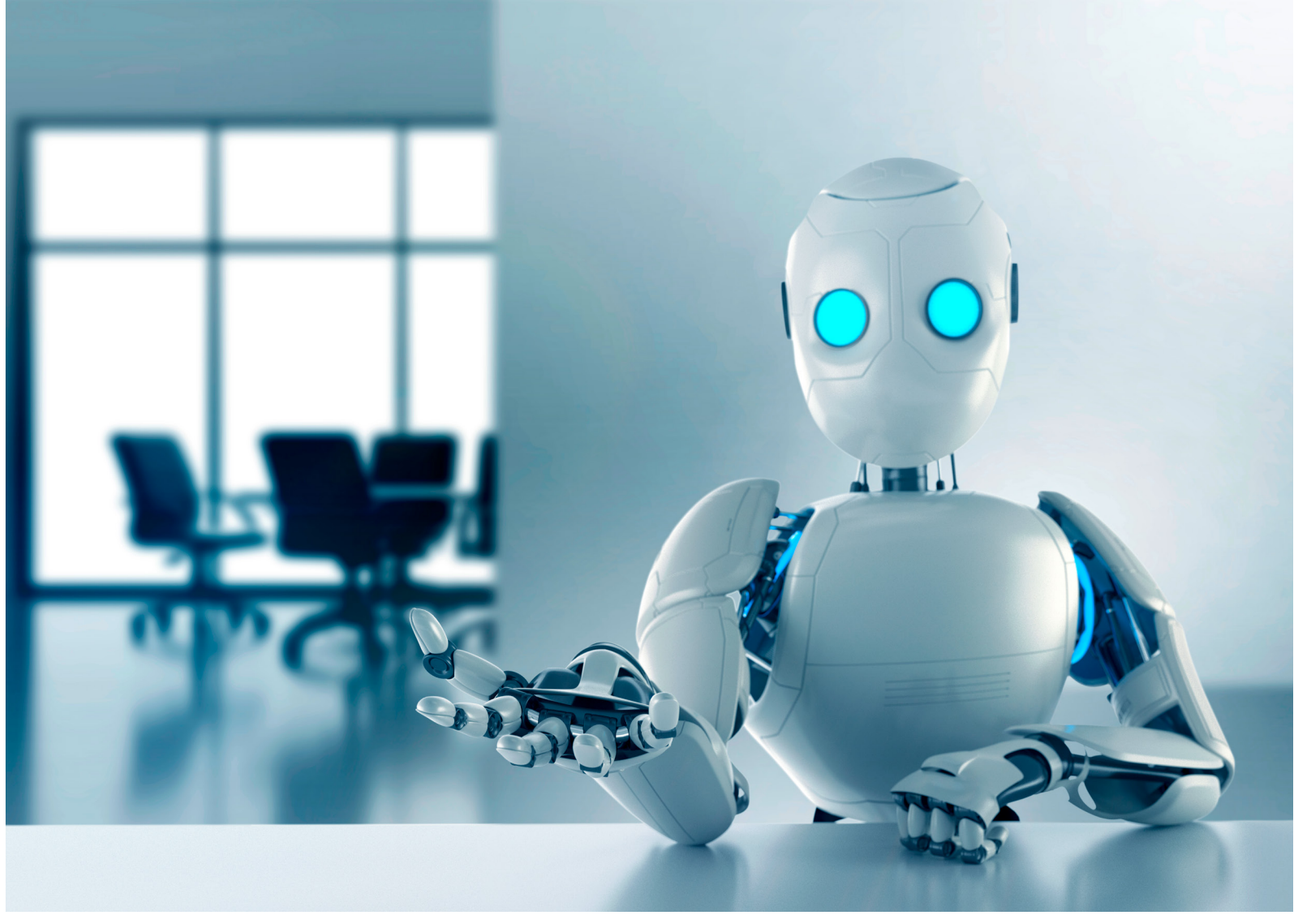
Die Kunst, die natürliche Intelligenz der Menschen zu nutzen, um die Risiken künstlicher Intelligenz zu zähmen. Oder wie wir Effizienzgewinne durch den Einsatz künstlicher Intelligenz gewissenhaft nutzen können und was dafür notwendig ist.

Computer sind seit Dekaden in Unternehmen im Einsatz. Obwohl etablierte Verfahren die Qualität der Programmerstellung sicherstellen, sind allfällige Probleme aus fehlerhafter Software jedem Anwender bekannt. Kombinationen von Eingabewerten und Systemzuständen sind nicht richtig bedacht oder implementiert worden und bei Tests unerkannt geblieben. Ein Programmierer analysiert und behebt den Fehler.

Was verändert sich in diesem Szenario beim Einsatz von softwarebasierter künstlicher Intelligenz? Kürzlich hat ein autonom fahrendes Auto in Arizona einen Fussgänger nicht beachtet und dieser ist bei der Kollision verstorben. Bei Fehlersuche und -behebung zeigt sich der fundamentale Unterschied zwischen maschinellem Lernen und menschlicher Programmierung: Es gibt keinen Programmierer.

Als Microsoft seinen Chatbot «Tay» ins Netz gestellt hat, konnten spielfreudige menschliche Nutzer schnell erkennen, dass die künstliche Intelligenz ihr Sprachverstehen autonom weiterentwickelt. Was früher ein explizierter, für den menschlichen Experten nachvollziehbarer und verständlicher Programmcode war, ist nun ein für uns unverständlicher, vieldimensionaler Entscheidungsraum geworden, in dem aus einer fast unzähligen Zahl an Verbindungen zwischen den Neuronen des künstlichen Gehirns ein Entscheid gefällt wurde. Wie im menschlichen Hirn können wir Fehler beispielsweise aus falsch memorierter Information, auf neuronaler Ebene nicht beheben. Hier zeigt sich das zentrale Manko der bisher bestehenden künstlichen Intelligenz: Menschen können sich miteinander auf symbolischer Ebene verstehen und Fehler beheben. Das neuronale Netz ist in seinen Trainingsdaten gefangen und kann nur über Repetition neuer Muster sein antrainiertes Verhalten ändern. Die Maschine versteht weder ihr Handeln, noch kann sie dieses erklären. Microsoft musste das Programm zurückziehen.

Zwei Komponenten sind notwendig, um Effizienzgewinne gewissenhaft nutzen zu können, die der Einsatz künstlicher Intelligenz mit sich bringt:



Komponente 1:
Der Einsatz zertifizierter «vertrauensvoller» Computer, die Systeme kontrollieren und verifizieren. Ein sehr aktives, wissenschaftliches Forschungsfeld arbeitet daran, wie mathematisch sichergestellt werden kann, ob ein Computersystem nach menschlichem Ermessen und Wertvorstellungen «richtig» entscheidet. Der Qualität der verwendeten Daten ist dabei grösste Aufmerksamkeit zu widmen: Da maschinelles Lernen nur aus diesem lernt, besteht u. a. das Risiko, dass nicht alle möglichen Entscheidungssituationen ausreichend abgebildet sind. Dies mag im einfachen Fall peinlich sein – eine Emotionserkennung, die nur mit dunkelhäutigen Gesichtern trainiert worden ist, wird ein hellhäutiges Gesicht nicht als Gesicht erkennen. Für Unternehmen entstehen jedoch schnell haftungsrechtliche Fragen, z. B. wenn bei maschinellen Entscheiden gewisse Personengruppen – Geschlecht, Alter, Wohnsitz – ungerechtfertigt benachteiligt werden.

Komponente 2:
Die Schaffung verbindlicher Kontrollrahmen, die sicherstellen, dass bei der Erstellung alle professionellen Standards eingehalten werden. Ähnlich den Vorgaben, die bei klassischer Software oder mathematischen Berechnungsmodellen bestehen. Wir müssen sicherstellen, dass die eingesetzten Modelle in ihrer Arbeitsweise nachvollziehbar und Dritten gegenüber erklärbar sind. Zukünftig werden wir zentrale Anwendungen durch eine vertrauenswürdige Instanz testen müssen und insbesondere die Einhaltung der gesellschaftlich akzeptierten Wertevorstellungen prüfen.

Konsequente Prozesskontrollen hätten das eingangs erwähnte Unglück verhindern können: Aufgrund von häufigen Fehlalarmen durch vom Wind aufgewirbelte Plastiktüten war der Schwellwert zum Abbremsen des Fahrzeugs hochgesetzt worden. Die Person auf der Fahrbahn war ausreichend früh erkannt worden: Die Steuerungsreaktion wurde jedoch unterdrückt. Dieser Fehler konnte durch den menschlichen Eingriff wieder rückgängig gemacht werden. Der Einsatz der Maschine enthebt den Menschen nicht von seiner Verantwortung – sicherlich nehmen wir diese als natürlich intelligente Wesen war.