

NoTies-Whitepaper ***«KI für Schweizer KMU»***



Midjourney, "Swiss company struggling with challenges with data lakes" (4.7.2023)

Inhaltsverzeichnis

Editorial	3
WHY	4
<i>Muss es denn wirklich ein KI-Projekt sein?</i>	5
<i>Welches Datenproblem ist zu lösen und ist es wirklich ein Datenproblem?</i>	5
<i>Welche Herausforderungen im Datenbereich eignen sich dann für KI-Anwendungen?</i>	6
HOW	7
<i>Supervised learning</i>	7
<i>Unsupervised learning</i>	9
<i>Reinforcement learning</i>	11
WHAT	13
<i>KI-Entscheidungsdiagramm (Welcher Anwendungsfall liegt vor?)</i>	13
<i>Wie soll die KI trainiert werden?</i>	14

Editorial

Künstliche Intelligenz ist in aller Munde und nicht nur ein wichtiger Change, sondern auch eine Chance für viele Unternehmen. Viele Mitarbeitende und KMUs sind bereits die ersten Schritte gegangen und nutzen Tools wie ChatGPT und Midjourney in ihrem täglichen Arbeitsalltag.

Die grosse Frage ist jedoch, lohnt es sich für ein Unternehmen in umfangreichere KI-initiativen zu investieren und entsprechende Implementierungsprojekte zu beauftragen?

Die aktuellen Produkte auf dem KI-Markt lassen KMUs vor allem in den folgenden sieben Anwendungsfällen stark profitieren:

- Automatisierung und Effizienzsteigerung
- Datenauswertung und Erkenntnisgewinn
- Personalisierung von Angeboten und Kundenservice
- Risikomanagement und Betrugserkennung
- Verbesserung der Produktentwicklung
- Optimierung von Marketing und Werbung
- Verbesserung der Entscheidungsfindung

Erkennen sie bereits spontan Ansatzpunkte für sich oder ihr Unternehmen für entsprechende Anwendungsfälle?

Dieses Whitepaper zeigt die wichtigsten Trends und gibt Gedankenanstösse im Umgang mit der Künstlichen Intelligenz für Unternehmen und Mitarbeitende.

Waltenschwil, Juli 2023

Stephan Weiss
Geschäftsführer NoTies GmbH

Pascal Vontobel
Geschäftsführer NoTies GmbH

WHY

Künstliche Intelligenz baut immer auf Daten auf. Dies weil die darunterliegenden KI-Modelle immer mit statistischen Wahrscheinlichkeiten rechnen. Und um statistische Modelle berechnen zu können ist ein Bestand an Grunddaten notwendig. Den Unternehmensdaten kommt also in den KI-Lösungen eine Schlüsselrolle zu.

KI-Modelle werden durch maschinelles Lernen trainiert, um aus vorhandenen Daten Muster und Zusammenhänge zu erkennen. Je umfangreicher und qualitativ hochwertiger die Daten sind, desto besser kann das Modell lernen und genauere Vorhersagen treffen. In diesem Zusammenhang gilt es in jedem KI-Projekt die folgenden vier Datenaspekte zu berücksichtigen:

1. **Datenvorverarbeitung und -bereinigung:** Vor der Verwendung in einem KI-Modell müssen die Daten oft vorverarbeitet und bereinigt werden. Dies umfasst Schritte wie das Entfernen von Ausreißern, das Auffüllen fehlender Werte und die Umwandlung der Daten in ein geeignetes Format für das Modell. Eine sorgfältige Datenbereinigung ist entscheidend, um verlässliche und aussagekräftige Ergebnisse zu erzielen.
2. **Datenqualität und -verfügbarkeit:** Die Qualität der Daten hat einen direkten Einfluss auf die Leistung eines KI-Modells. Hochwertige, zuverlässige und aktuelle Daten sind für die Genauigkeit und Zuverlässigkeit der KI-Lösung unerlässlich. Zudem muss gewährleistet sein, dass ausreichend Daten verfügbar sind, um ein aussagekräftiges Modell zu trainieren.
3. **Trainings- und Testdaten:** KI-Modelle werden in der Regel mit Trainingsdaten trainiert und anschliessend mit Testdaten evaluiert. Die Trainingsdaten werden verwendet, um das Modell anzulernen, während die Testdaten verwendet werden, um die Leistung und Genauigkeit des Modells zu überprüfen. Eine angemessene Aufteilung der Daten und ein ausgewogenes Verhältnis von Trainings- und Testdaten sind entscheidend, um ein gutes Modell zu entwickeln.
4. **Datenschutz und Compliance:** Bei der Nutzung von Daten in KI-Lösungen ist es wichtig, die Datenschutzrichtlinien und rechtlichen Bestimmungen zu beachten. Unternehmen müssen sicherstellen, dass sie die Zustimmung der betroffenen Personen zur Verwendung ihrer Daten haben und alle erforderlichen Massnahmen ergreifen, um die Privatsphäre und Sicherheit der Daten zu gewährleisten.

Es ist wichtig anzumerken, dass KI-Lösungen nicht nur auf vorhandenen Daten aufbauen, sondern auch kontinuierlich neue Daten sammeln und daraus lernen können. Dieser Prozess der fortlaufenden Verbesserung und Anpassung basiert jedoch immer auf den zugrunde liegenden Daten.

Muss es denn wirklich ein KI-Projekt sein?

Wie aus dem vorherigen Abschnitt hervorgeht, ist ein KI-Projekt eine aufwandreiche und kostspielige Angelegenheit und nicht jedes Unternehmen kann ein solches überhaupt stemmen. Die gute Nachricht ist, dass nur etwa 20% der Datenthemen in den Unternehmen den Einsatz einer KI rechtfertigen. Das bedeutet im Umkehrschluss, dass 80% der vorhandenen Datenthemen mit bereits im Unternehmen vorhandenen IT-Mitteln bearbeitet und nachhaltig verbessert werden können. Am Anfang steht also die Frage was für ein Datenproblem gelöst werden soll.

Welches Datenproblem ist zu lösen und ist es wirklich ein Datenproblem?

Der Fokus in Datenprojekten liegt üblicherweise in einer ersten Phase auf den Daten, welche im Unternehmen vorhanden sind. Dabei spielen die eigentlichen Daten in einem ersten Schritt eine untergeordnete Rolle. Es lohnt sich in einer ersten Phase den Fokus auf die Prozesse zu richten. Daten folgen oder unterstützen die Geschäftsprozesse und sind ein Arbeitsmittel, welches durch Prozesse entstehen oder verändert werden. Die Grundregel lautet also: Start mit den Prozessen und Prozesse vor den Daten.

Ist der Prozess definiert und das gewünschte Ziel klar (vielleicht wurde das neue Ziel mit [Design Thinking Methodik](#) im Unternehmen gemeinsam erarbeitet) ergibt sich auch welche Daten entlang der Prozesskette benutzt, bearbeitet und verändert werden. Mit diesem Wissen kann bestimmt werden, auf welche Weise die Daten, möglichst automatisiert, verarbeitet werden sollen.

Ein Beispiel dazu: Nehmen wir an, sie sind ein Industrieunternehmen, welches Maschinen für die Druckluftherstellung in Produktionsstrassen anderer Unternehmen herstellt. Im Design Thinking Workshop hat das Unternehmen erkannt, dass sie künftig nicht mehr ganze Maschinen verkaufen werden, sondern die Druckluft in einem Abomodell ihren Kunden zur Verfügung stellen werden. Zu diesem neuen Geschäftsmodell fertigen ihre Mitarbeitenden die notwendigen Prozesszeichnungen an, bis hin zur Abrechnung der Abokosten. Entlang dieser Prozesszeichnungen kann abgeleitet werden welche Maschinendaten, Kundendaten und Administrationsdaten sie brauchen um die Prozesse möglichst automatisiert zu implementieren. Mit dem Wissen über die Daten und die damit verbundenen Schnittstellen ist nun klar welche IT-Mittel für die Automatisierung zum Einsatz kommen können.

Wird in diesem beschriebenen Unternehmen Standardsoftware wie beispielsweise M365 eingesetzt, kann davon ausgegangen werden, dass 80% dieser Datenanwendungsfälle ohne Einsatz einer KI zur vollsten Zufriedenheit des Unternehmens. Maschinendaten können über bekannte Schnittstellen in Excel importiert und automatisiert bearbeitet werden. Eine Auswertung der Daten für Managementzwecke kann mit PowerBI erfolgen. Die Kundendaten und auch die administrativen Daten lassen sich auf gleiche Weise mit Standardsoftware bearbeiten und dem Verwendungszweck zuführen.

Welche Herausforderungen im Datenbereich eignen sich für KI-Anwendungen?

Es gibt verschiedene Herausforderungen im Datenbereich, die sich gut für KI-Anwendungen eignen. Diese fallen in der Regel in die folgenden sechs Kategorien:

1. **Grosse Datenmengen:** KI ist besonders gut geeignet, um in grossen Datenmengen Muster und Zusammenhänge zu erkennen. Wenn Unternehmen oder Organisationen mit umfangreichen Datenbeständen arbeiten, können KI-Anwendungen dabei helfen, diese Daten effizient zu analysieren und wertvolle Erkenntnisse zu gewinnen.
2. **Komplexe Datenstrukturen:** Manchmal können Daten eine komplexe Struktur haben, wie beispielsweise unstrukturierte Daten, Texte oder Bilddaten. KI-Anwendungen, wie zum Beispiel Natural Language Processing (NLP) oder Bilderkennung, können bei der Verarbeitung und Analyse solcher Daten helfen.
3. **Vorhersage und Prognose:** KI kann bei der Vorhersage von Trends, zukünftigen Ereignissen oder Verhaltensmustern aufgrund vorhandener Daten helfen. Beispielsweise können KI-Modelle in der Finanzbranche eingesetzt werden, um den Aktienmarkt zu analysieren und Vorhersagen über Preisentwicklungen zu treffen.
4. **Personalisierung:** KI ermöglicht es Unternehmen, personalisierte Empfehlungen oder Angebote auf Basis von Kundendaten zu erstellen. Dies kann beispielsweise im E-Commerce-Bereich verwendet werden, um individuelle Produktempfehlungen zu generieren oder personalisierte Marketingkampagnen durchzuführen.
5. **Datenqualität und -bereinigung:** KI kann auch bei der Datenqualitätssicherung und -bereinigung helfen. Durch den Einsatz von KI-Techniken können fehlerhafte Daten identifiziert, fehlende Werte aufgefüllt und Ausreisser entfernt werden.
6. **Betrugserkennung:** KI kann bei der Erkennung von betrügerischen Aktivitäten helfen, indem sie Muster und Anomalien in den Daten erkennt. Dies kann in verschiedenen Bereichen wie Finanzwesen, Versicherungen oder E-Commerce angewendet werden, um betrügerische Transaktionen frühzeitig zu erkennen und zu verhindern.

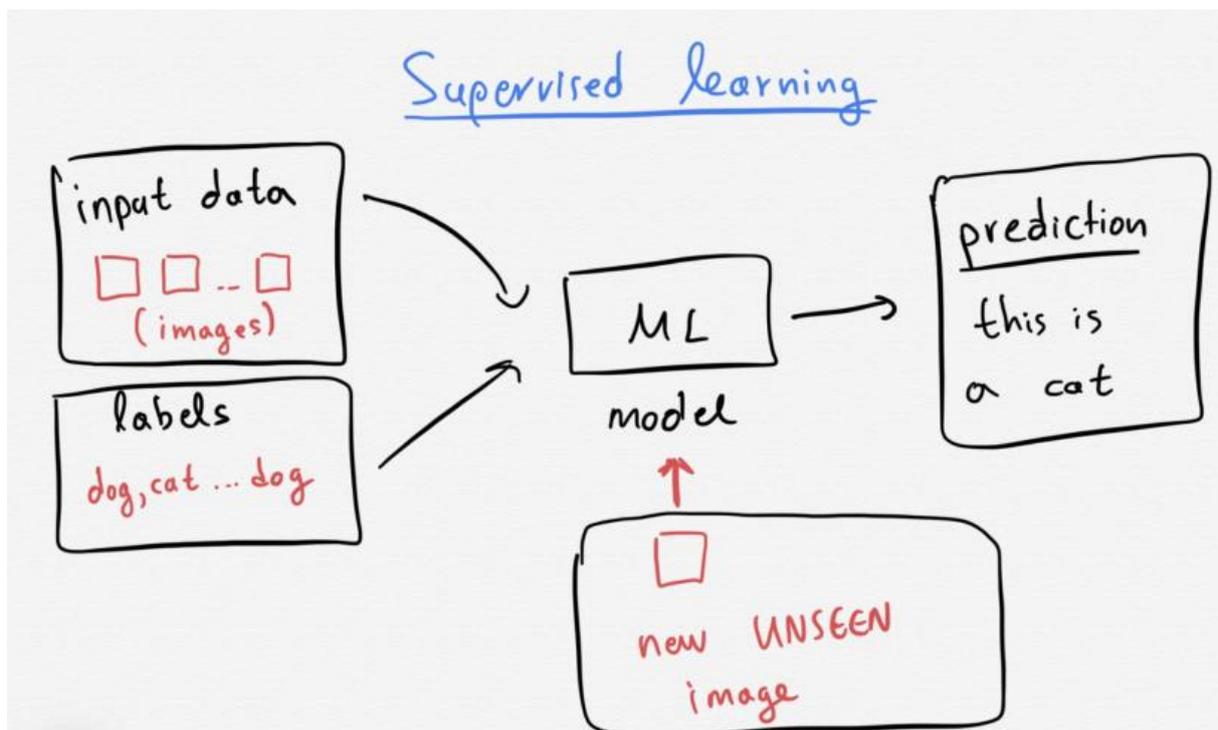
Es ist wichtig zu beachten, dass die Anwendung von KI in diesen Bereichen Herausforderungen mit sich bringen kann, wie z. B. den Zugang zu qualitativ hochwertigen Daten, den Datenschutz und die ethischen Aspekte der Datenverwendung. Es ist daher wichtig, bei der Entwicklung von KI-Anwendungen sorgfältig und verantwortungsbewusst vorzugehen.

HOW

Daten sind die Grundlage für künstliche Intelligenz. Diese Daten (labeled, unlabeled, mixed, dynamic) gilt es mittels Algorithmen zu trainieren – sprich anzulernen. Wir unterscheiden bei der Weiterverarbeitung von Daten zu Trainingszwecken grundsätzlich drei Lernmethoden: Supervised learning, Unsupervised learning und Reinforced learning.

Supervised learning

Supervised learning stellt noch immer mit Abstand den grössten Teil aller Geschäftsfälle dar. Das vielleicht anschaulichste Beispiel ist das autonome Fahren. Hier werden dem System eine riesige Menge an Bilddaten (labeled data) gefüttert. Anhand von zusätzlichen Labels (das ist ein Mensch, das ist ein Verbotsschild, das ist ein Rotlicht) lernt der Algorithmus und bewertet ein für ihn neues Bild mit einer Vorhersage und einer dazugehörigen Wahrscheinlichkeit des Zutreffens; das ist zu 95% ein Rotlicht, zu 3% eine Baustellenabspernung und zu 2% eine Werbetafel.



Supervised learning, Steven Moore, AI Leader, (2021)

Das korrekte Labeling von Daten ist hier essenziell. Wir kennen diesen Vorgang mit den sogenannten CAPTCHAs, wenn wir uns auf einer Webseite einloggen und nachweisen müssen, dass wir kein Roboter sind: «klicke auf alle Bilder mit Fahrrädern».

Für Firmen ist dieses Labeling eine kostspielige Angelegenheit. Nicht nur die menschliche Kontrolle der Bilder, sondern auch das Auffinden von neuen, relevanten Bildern.

Hier geht die Entwicklung (leider) in eine gefährliche Richtung. Da heutzutage Bilder von einer KI kostengünstig automatisiert generiert werden können (z.B. Midjourney), wird der Algorithmus mit ev. fehlerhaften Daten befüllt und trainiert. Hierzu Illia Shumailow von der Oxford University. «The problem is, when you're using artificial data, you acquire the errors from the misunderstandings of the models and statistical errors. You need to make sure that your errors are not biasing the output of other models and there's no simple way to do that". Unter biasing verstehen wir in diesem Zusammenhang die Voreingenommenheit aus falschen Daten, falsche Schlüsse zu ziehen.

Die Top 5 des Supervised Learning (Expertenwissen)

Supervised Learning bezieht sich auf eine Art von maschinellem Lernen, bei dem ein Algorithmus aus gelabelten Trainingsdaten lernt, um Vorhersagen oder Klassifikationen für neue, nicht gelabelte Daten zu treffen. Hier sind die fünf wichtigsten Punkte:

Trainingsdaten mit bekannten Labels: Im Supervised Learning werden Trainingsdaten verwendet, die sowohl Eingabevariablen als auch die entsprechenden gewünschten Ausgabewerte oder Labels enthalten. Diese Labels dienen als Ground-Truth-Referenz, anhand derer der Algorithmus lernen kann, Muster und Zusammenhänge zwischen den Eingabevariablen und den entsprechenden Ausgabewerten zu erkennen.

Klassifikation oder Regression: Supervised Learning kann für zwei Hauptaufgaben eingesetzt werden: Klassifikation und Regression. Bei der Klassifikation versucht der Algorithmus, Eingabevariablen in vordefinierte Klassen oder Kategorien zu klassifizieren. Bei der Regression geht es darum, kontinuierliche Ausgabewerte oder numerische Vorhersagen auf Grundlage der Eingabevariablen zu treffen.

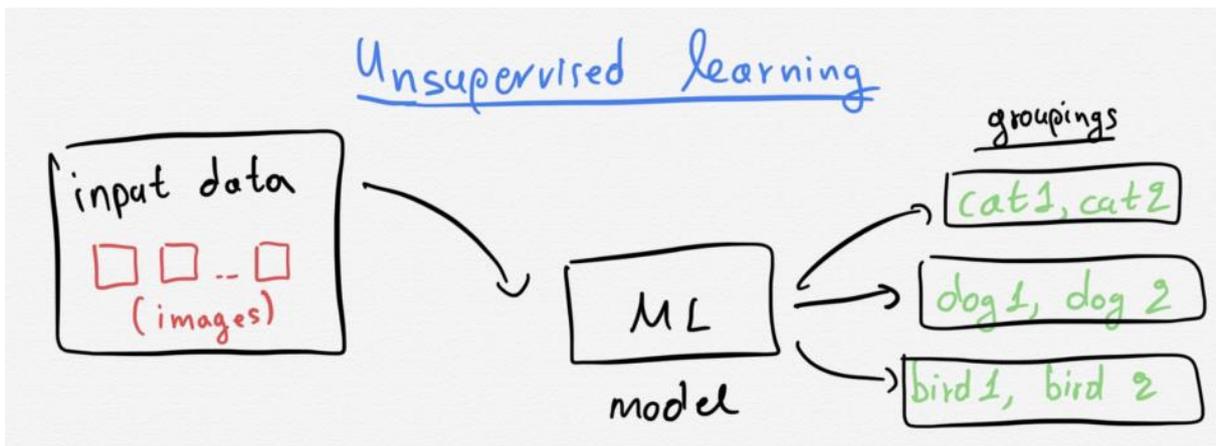
Modellbildung und Training: Beim Supervised Learning wird ein Modell definiert, das die Beziehung zwischen den Eingabevariablen und den Ausgabewerten repräsentiert. Das Modell wird mit den Trainingsdaten trainiert, indem es die Beispiele mit bekannten Labels analysiert und Muster oder Regeln lernt, um Vorhersagen für neue, nicht gelabelte Daten zu treffen. Dabei wird in der Regel ein Optimierungsalgorithmus verwendet, um die Modellparameter zu justieren und die Vorhersagegenauigkeit zu verbessern.

Evaluierung der Modelleistung: Nach dem Training des Modells wird die Leistung des Modells bewertet und überprüft. Hierfür werden separate Testdaten verwendet, die nicht im Trainingsprozess verwendet wurden. Durch den Vergleich der Vorhersagen des Modells mit den tatsächlichen Labels der Testdaten kann die Genauigkeit, Präzision oder andere Leistungsmetriken des Modells bewertet werden. Dies ermöglicht die Beurteilung der Leistungsfähigkeit des Modells und gegebenenfalls die Optimierung oder Anpassung des Modells.

Generalisierung: Ein wichtiges Ziel des Supervised Learning ist die Fähigkeit des Modells, generalisierte Vorhersagen für neue, nicht gelabelte Daten zu treffen. Das bedeutet, dass das Modell in der Lage ist, Muster und Zusammenhänge aus den Trainingsdaten zu lernen und dieses Wissen auf ähnliche, aber unbekannte Daten anzuwenden. Ein gutes Supervised Learning-Modell sollte eine hohe Vorhersagegenauigkeit auf neuen Daten haben, die es während des Trainings nicht gesehen hat.

Unsupervised learning

Dieser Bereich ist im Wachstum. Als anschauliches Beispiel ziehen wir hier die Kundensegmentierung anhand des Klickverhaltens in einem Onlineshop heran. Wer kennt es nicht: «Kunden, die diesen Artikel kauften, kaufen oft auch xyz» Oft kennt der Onlineshop einige Angaben von uns wie: Geschlecht, Alter, Wohnort, Bankverbindung, etc. dies nimmt der Algorithmus als labeld data und kombiniert sie mit meinem Klickverhalten. Wie lange verweile ich auf einer Seite, welche Produkte schaue ich mir an. Anhand dieser Erkenntnisse können Onlineshops z.B. ihre Angebote preislich dynamisch steuern. Bekanntes Beispiel: Flüge buchen. Abhängig von meinem Browser, Handtyp, Anzahl Suchfragen, erhalte ich unterschiedliche Preise. Oft gilt die Regel: je öfter ich suche, desto höher der Preis. Hier kommt der psychologische Trick mit der Verknappung zum Tragen



Supervised learning, Steven Moore, AI Leader, (2021)

Die Top 5 des Unsupervised Learning (Expertenwissen)

Unsupervised Learning bezieht sich auf eine Art von maschinellem Lernen, bei dem ein Algorithmus aus unbeaufsichtigten Datenmengen lernt, ohne vorherige Kennzeichnung oder Anleitung durch einen menschlichen Experten. Hier sind die fünf wichtigsten Punkte:

Mustererkennung: Unsupervised Learning-Algorithmen zielen darauf ab, Muster oder Strukturen in den Daten zu erkennen, ohne dass ihnen vorab bekannte Klassen oder Labels gegeben werden. Sie suchen nach inhärenten Mustern und Zusammenhängen in den Daten, um Informationen und Wissen zu extrahieren.

Clustering: Eine häufige Anwendung von Unsupervised Learning ist das Clustering, bei dem ähnliche Datenpunkte zu Gruppen oder Clustern zusammengefasst werden. Der Algorithmus analysiert die Eigenschaften der Datenpunkte und gruppiert sie basierend auf Ähnlichkeiten, um verborgene Strukturen zu identifizieren.

Dimensionsreduktion: Unsupervised Learning kann auch zur Dimensionsreduktion eingesetzt werden, um die Anzahl der Merkmale oder Variablen in den Daten zu verringern, während wichtige Informationen beibehalten werden. Durch die Reduzierung der Dimensionalität können komplexe Daten visualisiert oder effizienter verarbeitet werden, ohne dabei wesentliche Informationen zu verlieren.

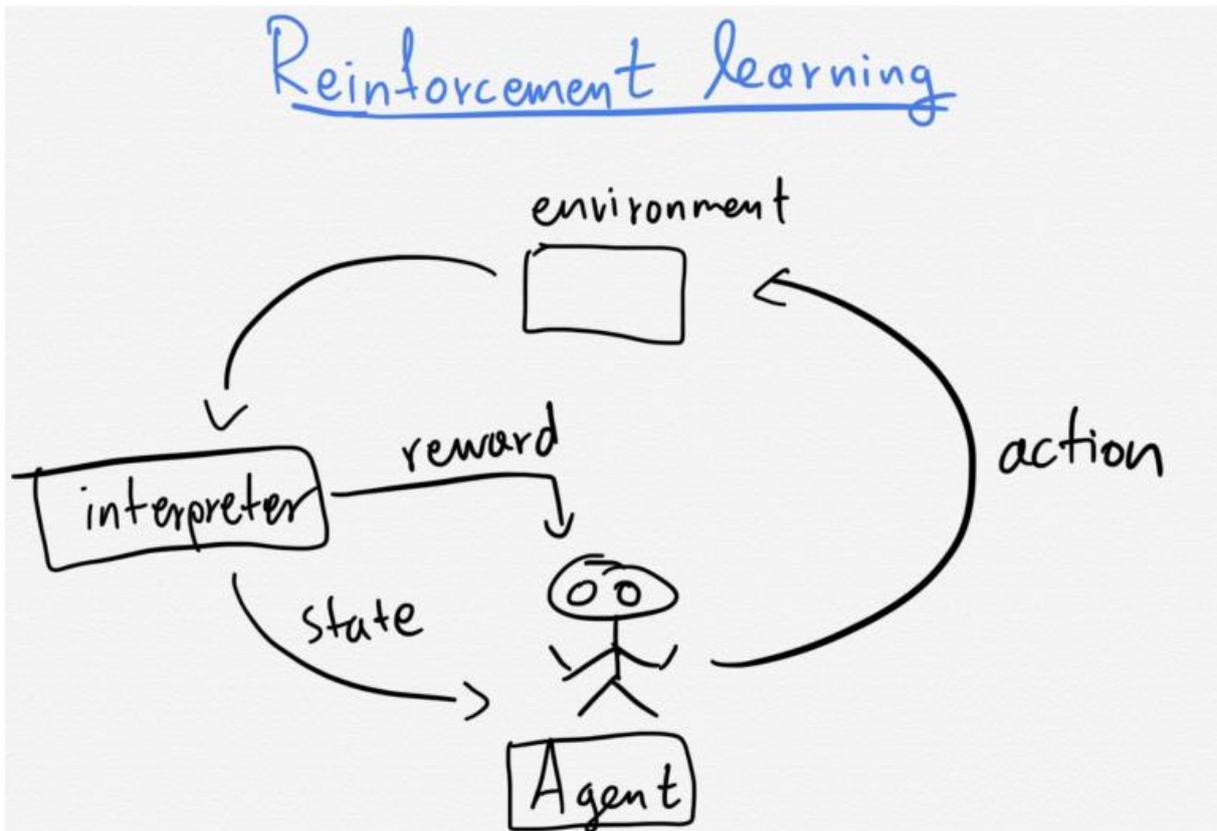
Anomalieerkennung: Eine weitere wichtige Anwendung von Unsupervised Learning ist die Erkennung von Anomalien oder Ausreißern in den Daten. Der Algorithmus lernt die normalen Muster der Daten und kann dann Abweichungen oder ungewöhnliche Beobachtungen identifizieren, die von den normalen Mustern abweichen.

Generative Modelle: Unsupervised Learning kann auch verwendet werden, um generative Modelle zu erstellen, die in der Lage sind, neue Daten zu generieren, die ähnlich zu den vorhandenen Daten sind. Diese Modelle erlernen die Wahrscheinlichkeitsverteilung der Daten und können dann neue Beispiele erzeugen, die den gelernten Mustern entsprechen.

Reinforcement learning

Das ist die Königsklasse und somit auch das komplexeste Umfeld. Es macht in der Schweiz noch einen geringen Teil der Geschäftsfälle aus. Obschon wenn wir von KI sprechen, auch sehr oft diesen Lerntyp meinen.

Als anschauliches Beispiel wird hier der Industrieroboter genommen, welcher für ihn unbekannte Objekte erkennt, diese selbstständig sortiert und seine Fehler lernt und diese korrigiert – easy? Nicht wirklich. Denn bei diesem Typ wird dem Roboter ein Optimierungsziel gegeben mit einem Belohnungssystem. Diese Belohnung gilt es zu Maximieren.



Supervised learning, Steven Moore, AI Leader, (2021)

Die Top 5 des Reinforcement Learning (Expertenwissen)

Reinforcement Learning ist ein Bereich des maschinellen Lernens, bei dem ein Agent durch Interaktion mit einer Umgebung lernen soll, wie er Aktionen ausführt, um maximale Belohnungen zu erhalten. Hier sind die fünf wichtigsten Punkte:

Agent-Umgebung-Interaktion: Im Reinforcement Learning interagiert ein Agent mit einer Umgebung. Der Agent führt Aktionen aus, die die Umgebung beeinflussen, und erhält als Antwort Beobachtungen und Belohnungen. Der Agent lernt durch Erfahrung, indem er die Auswirkungen seiner Aktionen beobachtet und die besten Entscheidungen trifft, um seine Belohnung zu maximieren.

Belohnungssignal: Im Reinforcement Learning wird ein Belohnungssignal verwendet, um das Verhalten des Agenten zu bewerten. Das Belohnungssignal gibt an, wie gut oder schlecht eine bestimmte Aktion oder Entscheidung war. Der Agent strebt danach, die kumulierte Belohnung im Laufe der Zeit zu maximieren, indem er lernen, welche Aktionen zu positiven Belohnungen führen.

Markov Decision Process (MDP): Reinforcement Learning-Problemstellungen werden oft als Markov Decision Processes formuliert. Ein MDP besteht aus einer Zustandsmenge, einer Aktionsmenge, einer Übergangsfunktion, einer Belohnungsfunktion und einem Diskontierungsfaktor. Das MDP-Modell erlaubt es dem Agenten, die optimale Strategie für das Treffen von Entscheidungen in einer sequenziellen Umgebung zu erlernen.

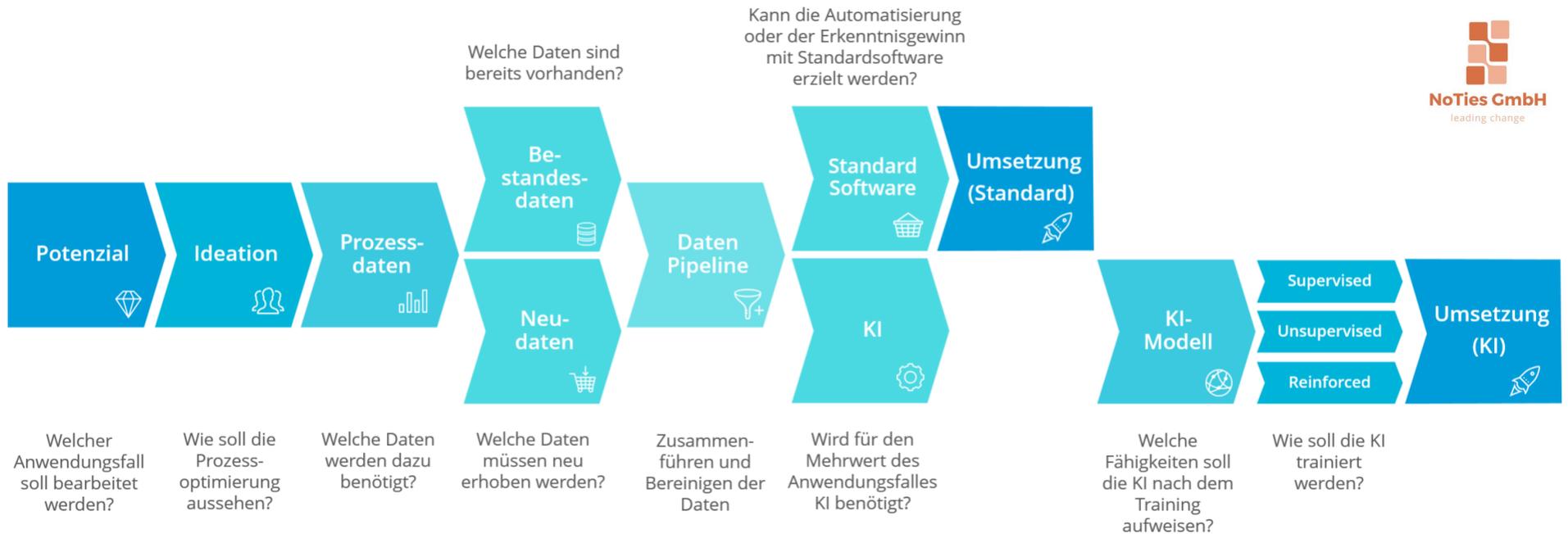
Exploration und Exploitation: Ein zentrales Konzept im Reinforcement Learning ist die Balance zwischen Exploration und Exploitation. Während der Exploration erkundet der Agent die Umgebung, um neue Erfahrungen zu sammeln und möglicherweise bessere Aktionen zu entdecken. Bei der Exploitation nutzt der Agent sein bisheriges Wissen aus, um die Aktionen zu wählen, von denen er glaubt, dass sie die grösste Belohnung bringen. Ein guter Reinforcement Learning-Agent findet eine geeignete Balance zwischen Exploration und Exploitation, um effektiv zu lernen und optimale Entscheidungen zu treffen.

Policy Optimization: Reinforcement Learning-Algorithmen zielen darauf ab, eine optimale Richtlinie (Policy) zu lernen, die die besten Aktionen in verschiedenen Situationen vorschlägt. Die Richtlinie kann deterministisch oder stochastisch sein und definiert die Aktionen des Agenten als Funktion des aktuellen Zustands. Der Lernalgorithmus optimiert die Richtlinie basierend auf den gesammelten Erfahrungen, um die kumulierte Belohnung zu maximieren.

WHAT

KI-Entscheidungsdiagramm (Welcher Anwendungsfall liegt vor?)

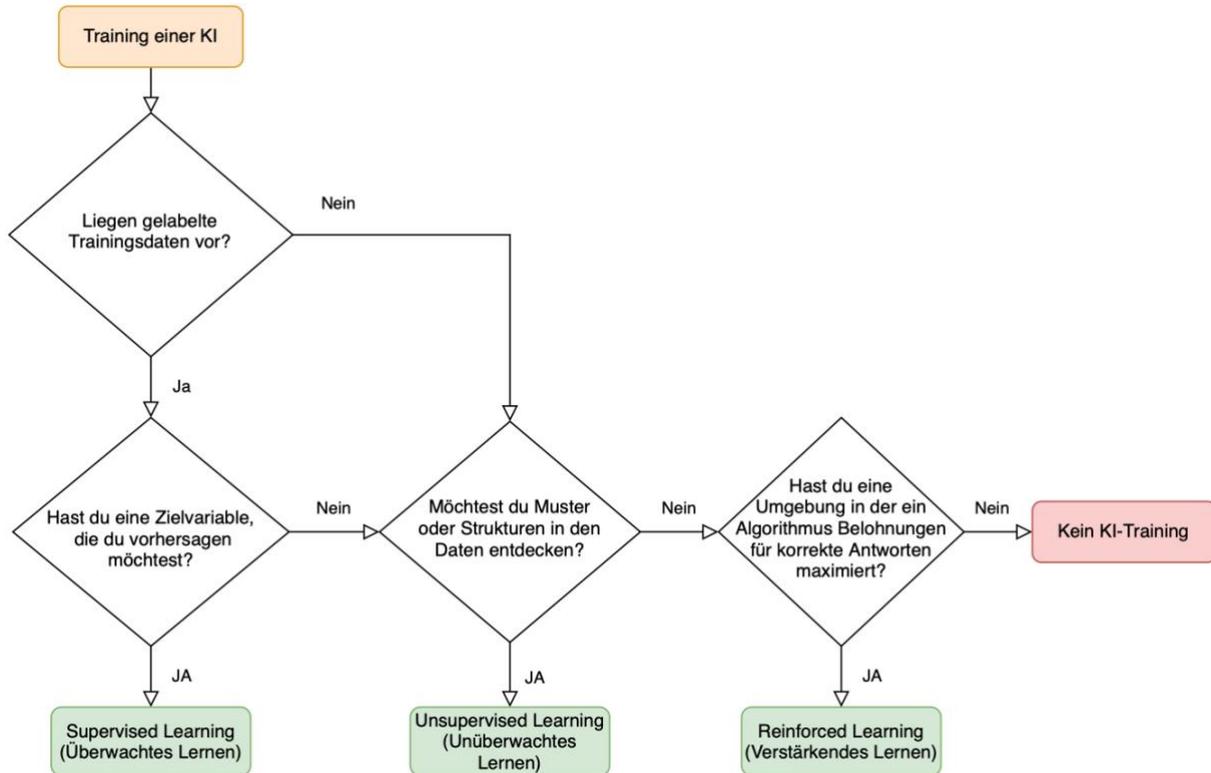
Das folgende Entscheidungsdiagramm unterstützt schrittweise bei KI-Projektumsetzungen und kann für die Entscheidungsfindung für den vorliegenden Datenanwendungsfall herangezogen werden.



NoTies KI-Entscheidungsdiagramm, (2023)

Wie soll die KI trainiert werden?

Wie auf dem NoTies KI-Entscheidungsdiagramm hervorgeht, ist es notwendig die Art und Weise zu bestimmen, mit welcher die KI trainiert werden soll. Das folgende Flussdiagramm hilft bei der Wahl des richtigen technologischen Ansatzes:



NoTies KI-Trainingsflussdiagramm, (2023)

In Einzelfällen können auch Kombinationen des KI-Lernens vorliegen.

Bevor im Unternehmen ein KI-Projekt zur Umsetzung vorgesehen wird, lohnt sich also ein genauer Blick auf den Anwendungsfall und die darunterliegenden Daten und Prozesse. Machen sie das Beste aus ihren Daten!

NoTies GmbH engagiert sich seit 2015 erfolgreich in der Digitalen Transformation. Als kompetenter Ansprechpartner für Unternehmen aus den Branchen Education, Finance, Healthcare, Aviation sowie Logistics und Retail, schaffen wir einen spürbaren Mehrwert in der Begleitung und gezielten Unterstützung ihrer digitalen Transformationsprojekten.

Gemeinsam mit unseren Kunden sind wir erfolgreich!

 noties.ch

 +41 56 611 15 15

 info@noties.ch

© 2023 NoTies GmbH