

# SAINT Lehrplan

UNIT 3: Anwendung von AI in Sprache und  
Bildverarbeitung

---

Lieferbar: WP2/2.2



## SAINT

HANDS ON INTRODUCTION TO ARTIFICIAL  
INTELLIGENCE IN PRIMARY EDUCATION  
USING MINECRAFT

APRIL 2023

---

EKVASIS

Verfasst von: EKVASIS Team

Projekt-Nummer: 2022-1-FR01-KA220-SCH-000087794



Co-funded by  
the European Union

Die Unterstützung der Europäischen Kommission für die Erstellung dieser Veröffentlichung stellt keine Billigung des Inhalts dar, der ausschließlich die Meinung der Autoren widerspiegelt, und die Kommission kann nicht für die Verwendung der darin enthaltenen Informationen verantwortlich gemacht werden.

## REVISIONSGESCHICHTE

Version	Datum	Autor	Beschreibung	Aktion	Seiten
1.0	17/01/2023	TARAN	Erstellung	C	TBS

(\*) Aktion: E = Erstellung, E = Einfügen, A = Aktualisieren, E = Ersetzen, L = Löschen

## REFERENZIERTE DOKUMENTE

ID	Referenz		Titel
1	2022-1-FR01-KA220-SCH-000087794		SAINT Vorschlag
2			

## ANWENDBARE DOKUMENTE

ID	Referenz		Titel
1			
2			

## Inhalt

1	Einföhrung in das Projekt.....	4
1.1	Der Umfang des Projekts.....	4
1.2	Die Zielgruppen .....	4
1.3	Der Zweck dieses Dokuments .....	5
2	Glossar des Referats.....	5
3	Einföhrung in die Einheit.....	6
3.1	Beschreibung .....	6
3.2	Lernziele und Ergebnisse .....	6
3.3	Geschätzte Sitzdauer .....	7
4	Inhalt der Unterrichtseinheit.....	7
4.1	Einleitung .....	7
4.2	Idee 1: Wahrnehmung .....	7
4.3	Idee 2: Repräsentation und Argumentation .....	9
4.4	Idee 3: Lernen .....	10
4.5	Idee 4: Natörlische Interaktion.....	12
4.6	Idee 5: Gesellschaftliche Auswirkungen .....	13
4.7	Fallstudien und Erfolgsgeschichten .....	15
4.	Zusätzliche Materialien und Ressourcen .....	16
5	Nachbereitung .....	18
6	Quiz.....	18
7	Referenzen.....	19

# 1 Einführung in das Projekt

## 1.1 Der Umfang des Projekts

Als ideale digitale Lernumgebung, um Kindern die praktischen Anwendungen von KI auf der Grundlage der AI4K12-Projektrichtlinien näher zu bringen, umfasst die Motivation für dieses Projekt die folgenden Ziele:

- Schülern, Lehrern und Ausbildern KI-Konzepte, ihre Auswirkungen auf unsere Gesellschaft und die damit verbundenen praktischen Umsetzungen näherbringen,
- den wachsenden Bedarf an der Entwicklung von Fernlernlösungen zu decken, die das Engagement, die Kreativität, die Problemlösungs- und die Entscheidungsfähigkeit der Schüler fördern,
- Weiterqualifizierung von Lehrern und Erziehern mit neuen Fähigkeiten (PBL, KI, Gamification usw.), die durch innovative Lehrmethoden entwickelt wurden,
- Verbesserung des Engagements der Kinder durch den Einsatz innovativer Lehrmethoden, die den Kindern helfen, Kreativität zu entwickeln,
- Verringerung der Kluft zwischen Bedarf und Verfügbarkeit von KI-bezogenen Fähigkeiten.

AI-Adventures in Minecraft vermittelt Kindern im Alter von 9-12 Jahren in einer Minecraft-Welt KI-bezogene Fähigkeiten. Damit schaffen wir ein unterhaltsames, interaktives und kreatives Lernumfeld durch spezifische Aktivitäten und Herausforderungen, die sich an den AI4K12-Leitlinien ([ai4ka12.org](http://ai4ka12.org)) und den 5 großen Ideen der KI orientieren: 1) Wahrnehmung, 2) Repräsentation und Argumentation, 3) Lernen, 4) natürliche Interaktionen, 5) gesellschaftliche Auswirkungen.

Zu diesem Zweck entwickelt und fördert das Projekt die folgenden greifbaren Ergebnisse:

- Dieses Curriculum: ein vollständiger Lernkurs zur Einführung von KI in den Schulunterricht, der auf den 5 großen Ideen des AI4K12-Rahmens basiert. Der Kurs vermittelt Wissen über die KI-Bildungsrichtlinien von AI4K12 und die 5 großen Ideen, erforscht die Auswirkungen von KI in unserer Gesellschaft und verbessert das Verständnis der relevanten Konzepte.
- Eine maßgeschneiderte Minecraft-Welt (AI Adventures World), die pädagogischen Herausforderungen auf der Grundlage des Lernkurses bietet. Sie nutzt das Konzept des Escape Rooms und bietet problemorientierte Lernaktivitäten. Eine Herausforderung für jede Einheit oder Lektion.
- Der virtuelle Raum der Gießerei, der eine wachsende Gemeinschaft von SAINT-Anwendern unterstützt und die korrigierende/perfektive und evolutive Wartung des Schulungspakets leitet.

## 1.2 Die Zielgruppen

Das Projekt sieht die direkte Beteiligung von Lehrern vor, hauptsächlich von Lehrern für Kinder im Alter von 9-12 Jahren oder von Hochschullehrern, die in der Ausbildung von Pädagogen tätig

sind. Diese LehrerInnen sind entweder LehrerInnen von MINT-Fächern oder haben ein gewisses Wissen und Interesse an KI und/oder Minecraft.

Was die indirekten Zielgruppen betrifft, so können folgende Personen einbezogen werden:

- MINT-Zentren, die ihren Katalog innovativer Unterrichtstechnologien oder ihren Katalog produktverbessernder KI-Kenntnisse ausbauen wollen,
- Hochschuleinrichtungen, die mit Unternehmen/Behörden zusammenarbeiten, die an der Erstellung von Lehrmaterial beteiligt sind,
- Organisationen, Verbände oder Netzwerke, die Eltern und/oder Pädagogen Bildungsmaterial zum Thema KI zur Verfügung stellen wollen: z. B. Programmierclubs, Volkshochschulen, Unternehmensberatungsdienste, Weiterbildungszentren usw.

### 1.3 Der Zweck dieses Dokuments

Das Arbeitspaket Nr. 2-AI4K12 Bildungsprogramm konzentriert sich auf die Erstellung eines kompletten Kurses über KI mit einer Reihe von 5 Herausforderungen in der dazugehörigen Minecraft Welt, um die praktische Umsetzung der Technologie zu veranschaulichen.

Dieser KI-Lehrplan besteht aus insgesamt 5 Einheiten pädagogischen Materials, das auf den AI4K12-Bildungsrichtlinien und den Lernzielen basiert, die im Anschluss an nationale Umfragen festgelegt wurden:

1. Anwendung von KI beim maschinellen Lernen,
2. Anwendung von KI im Bereich Arbeit und Unternehmertum,
3. Anwendung von KI in Sprache und Vision,
4. Anwendung von KI in Spielen und Puzzles,
5. Anwendung von KI im täglichen Leben.

Zusätzlich wird in jeder Einheit ein Glossar erstellt, um Lehrern und Schulen die Einführung des SAINT-Pakets zu erleichtern.

## 2 Glossar des Referats

Wörter	Definition
<b>Maschinelles Lernen</b>	Ein Teilbereich der KI, der Algorithmen und statistische Modelle verwendet, um Computersysteme in die Lage zu versetzen, ihre Leistung bei einer bestimmten Aufgabe zu verbessern
<b>Erkennung von Sprache</b>	auf der Grundlage der eingegebenen Daten zu verbessern, ohne explizit programmiert zu werden.
<b>Computer Vision</b>	Die Fähigkeit einer Maschine oder eines Programms, menschliche Sprache zu erkennen und zu verstehen.
<b>Tiefes Lernen</b>	Die Fähigkeit einer Maschine oder eines Programms, menschliche Sprache zu erkennen und zu verstehen, wird häufig in Anwendungen

	wie virtuellen Assistenten, automatisiertem Kundendienst und Transkription eingesetzt.
<b>Natürliche Sprachverarbeitung (NLP)</b>	Ein Studienbereich, der sich darauf konzentriert, Computer in die Lage zu versetzen, visuelle Daten aus ihrer Umgebung zu interpretieren und zu verstehen, und der häufig in Anwendungen wie Objekterkennung, Gesichtserkennung und
<b>Verstärkungslernen</b>	autonome Fahrzeuge.
<b>Erweiterte Realität</b>	Ein Teilbereich des maschinellen Lernens, der künstliche neuronale Netze verwendet, um
<b>Virtuelle Realität</b>	komplexe Muster in Daten zu modellieren, die häufig in Anwendungen wie der Bild- und Spracherkennung verwendet werden

## 3 Einführung in die Einheit

### 3.1 Beschreibung

Diese Lerneinheit führt die Lernenden in das spannende Feld der künstlichen Intelligenz und ihre Anwendung in der Sprach- und Bildverarbeitung ein. Es werden die verschiedenen Arten des maschinellen Lernens untersucht, darunter überwachtes, unüberwachtes und verstärkendes Lernen, und wie sie in der Spracherkennung, Bildklassifizierung und anderen realen Anwendungen eingesetzt werden. Das Referat wird sich auch mit natürlicher Interaktion befassen, z. B. damit, wie KI mit Menschen interagiert und wie virtuelle und erweiterte Realität die Art und Weise verändern, wie wir mit Technologie interagieren. Darüber hinaus werden die gesellschaftlichen Auswirkungen der KI behandelt, einschließlich ethischer Überlegungen und der Vor- und Nachteile der KI-Technologie. Am Ende der Lerneinheit sollten die Lernenden ein grundlegendes Verständnis von KI in den Bereichen Sprache und Sehen sowie deren Potenzial für die Lösung von Problemen in der realen Welt haben.

### 3.2 Lernziele und Ergebnisse

In dieser Lerneinheit werden die Lernenden mit der Anwendung von KI in der Sprach- und Bildverarbeitung vertraut gemacht. Sie lernen die Grundlagen des maschinellen Lernens kennen, einschließlich überwachtem, unüberwachtem und verstärktem Lernen. Die Lerneinheit behandelt auch die natürliche Interaktion und ihre Bedeutung für die KI sowie die gesellschaftlichen Auswirkungen und ethischen Überlegungen im Zusammenhang mit dem Einsatz von KI.

Nach erfolgreichem Abschluss dieser Einheit sollten die Lernenden:

- sollten die Lernenden in der Lage sein, die verschiedenen Arten des maschinellen Lernens und ihre Anwendungen in den Bereichen Sprache und Sehen zu identifizieren.
- können sie die Bedeutung der natürlichen Interaktion und ihre Rolle bei der Entwicklung von KI-Systemen erklären.
- wird ein solides Verständnis der gesellschaftlichen Auswirkungen von KI und der ethischen Überlegungen haben, die bei der Entwicklung und dem Einsatz dieser Systeme berücksichtigt werden müssen.

- werden in der Lage sein, reale Anwendungen von KI in den Bereichen Sprache und Sehen zu bewerten und ihr Potenzial zur Lösung komplexer Probleme zu erkennen.

### 3.3 Geschätzte Sitzdauer

Der Abschluss des Moduls und die Umsetzung des vermittelten Wissens wird etwa 5-6 Stunden dauern, basierend auf dem Umfang des Inhalts und der Quizfragen im Modul. Die tatsächliche Zeit, die die Lernenden für den Abschluss des Moduls und die Anwendung des erworbenen Wissens benötigen, kann jedoch variieren, je nach Lerntempo und Vertrautheit mit dem Thema.

## 4 Inhalt der Unterrichtseinheit

### 4.1 Einleitung

Willkommen zu der Einheit über die Anwendung von KI in der Sprach- und Bildverarbeitung. In dieser Lektion werden wir untersuchen, wie KI eingesetzt wird, um die Bereiche Spracherkennung und Computer Vision zu verändern. Zunächst werden wir die Grundlagen des maschinellen Lernens und die verschiedenen Techniken besprechen, die in Sprach- und Bildverarbeitungsanwendungen zum Einsatz kommen, darunter überwachtes, unüberwachtes und verstärkendes Lernen. Als Nächstes werden wir uns mit natürlicher Interaktion befassen und wie KI die Art und Weise verändert, wie wir mit Maschinen interagieren, einschließlich der Interaktion mit Sprache und Bildverarbeitung. Wir werden auch die gesellschaftlichen Auswirkungen von KI in Sprach- und Bildverarbeitungsanwendungen untersuchen, einschließlich ihrer Vor- und Nachteile und ethischer Überlegungen. Schließlich werden wir einige der spannendsten Fallstudien und Erfolgsgeschichten auf diesem Gebiet beleuchten, die zeigen, wie KI zur Lösung realer Probleme in den Bereichen Gesundheitswesen, Bildung und Unterhaltung eingesetzt wird. Am Ende dieser Lektion werden Sie ein solides Verständnis dafür haben, wie KI die Bereiche Sprache und Sehen verändert und welche potenziellen Auswirkungen sie auf die Gesellschaft als Ganzes haben kann.

### 4.2 Idee 1: Wahrnehmung

Künstliche Intelligenz (KI) ist ein sich rasch entwickelnder Bereich, der unser Leben in vielerlei Hinsicht verändert. Von der Spracherkennung bis zur Bildanalyse ist die KI zu einem festen Bestandteil unseres Alltags geworden. Die Vermittlung von KI-Konzepten an Kinder kann jedoch eine Herausforderung sein, insbesondere angesichts der Komplexität der Technologie. Um dieser Herausforderung zu begegnen, wurde der AI4K12-Rahmen als Leitfaden für die Vermittlung von KI-Konzepten und praktischen Anwendungen an SchülerInnen der Klassenstufe 12 entwickelt. Der Rahmen basiert auf fünf großen Ideen, von denen die erste die Wahrnehmung ist. In dieser Unterrichtseinheit werden wir die erste große Idee des AI4K12-Rahmens, die Wahrnehmung, erforschen, bei der es darum geht, wie KI die Welt um uns herum wahrnimmt. Die Bedeutung der Wahrnehmung in der KI kann gar nicht hoch genug eingeschätzt werden, da sie es Maschinen ermöglicht, die Welt um sie herum zu interpretieren und auf eine Weise zu reagieren, die der des Menschen ähnlich ist. Um die Welt um sie herum wahrnehmen

zu können, sind KI-Systeme auf eine Reihe von Sensoren angewiesen, die jeweils Daten aus verschiedenen Quellen sammeln. Diese Sensoren spielen eine entscheidende Rolle, wenn es darum geht, dass Maschinen ihre Umgebung verstehen und auf der Grundlage dieser Informationen Entscheidungen treffen können. In diesem Abschnitt stellen wir verschiedene Arten von Sensoren vor, die in der KI verwendet werden, und zeigen, wie sie funktionieren.

### Beispiele:

**Microfone:** Eine der häufigsten Arten von Sensoren, die in der KI eingesetzt werden, ist das Mikrofon, das für die Spracherkennung verwendet wird. Mikrofone sammeln Schallwellen und wandeln sie in digitale Signale um, die von KI-Algorithmen verarbeitet werden können. Sprachassistenten wie Alexa von Amazon und Siri von Apple sind auf Mikrofone angewiesen, um Sprachbefehle zu verstehen und darauf zu reagieren.

**Kameras:** Ein weiterer wichtiger Sensortyp für die KI ist die Kamera, die für die Computer Vision verwendet wird. Kameras nehmen Bilder oder Videobilder auf und verwenden Algorithmen, um daraus Informationen zu extrahieren. Diese Informationen können Objekterkennung, Gesichtserkennung und Bewegungsverfolgung umfassen. Kameras werden in vielen realen Anwendungen eingesetzt, darunter Überwachungssysteme, autonome Fahrzeuge und Sicherheitskameras.

**Infrarot-Sensoren:** Infrarotsensoren erkennen Infrarotstrahlung, die von allen Objekten ausgesandt wird, deren Temperatur über dem absoluten Nullpunkt liegt. Diese Sensoren werden in Anwendungen wie Haussicherheitssystemen und Nachtsichtkameras eingesetzt. Sie erkennen die Wärmesignatur von Objekten und ermöglichen es KI-Systemen, Bewegungen in der Dunkelheit zu erkennen und zu verfolgen.

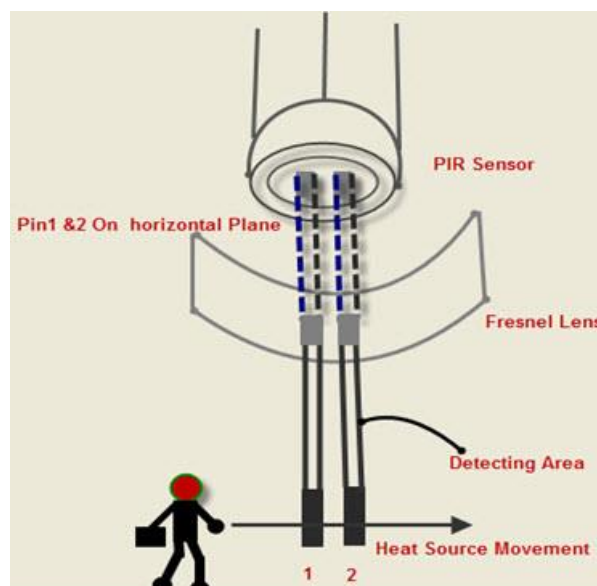


Bild 1: Infrarotsensor



**Berührungssensoren** : Berührungssensoren werden in Anwendungen wie Touchscreens, Roboter greifern und Prothesen eingesetzt. Berührungssensoren können Druck, Temperatur und andere physikalische Eigenschaften erkennen und ermöglichen KI-Systemen die Interaktion mit der physischen Welt.

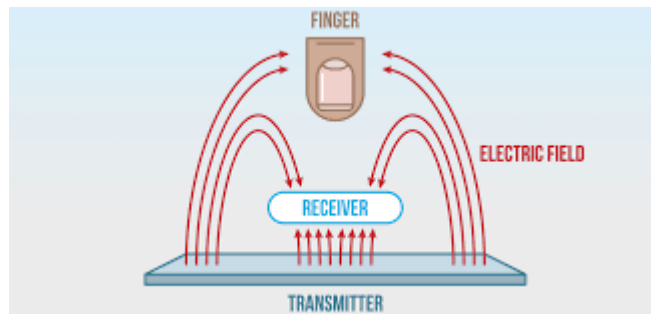


Bild 1: Berührungssensoren

### 4.3 Idee 2: Repräsentation und Argumentation

Die Darstellung bezieht sich darauf, wie Daten von KI-Systemen gespeichert und verarbeitet werden. Diese Darstellung ermöglicht es dem KI-System, Merkmale des Klangs zu analysieren und zu extrahieren, z. B. Phoneme oder Wörter. In der KI werden die Daten so dargestellt, dass Maschinen sie verstehen und effizient verarbeiten können. Am häufigsten werden Daten in der KI durch Symbole oder Zahlen dargestellt. So kann z. B. das Bild einer Katze als eine Matrix von Zahlen dargestellt werden, die der Farbe jedes Pixels auf dem Bild entsprechen.

Unter Schlussfolgerungen versteht man den Prozess, bei dem aus den gesammelten und dargestellten Daten Schlussfolgerungen gezogen werden. KI-Systeme verwenden verschiedene Schlussfolgerungstechniken zur Verarbeitung und Analyse von Daten, z. B. deduktives, induktives und abduktives Schließen. Diese Techniken helfen Maschinen, Vorhersagen zu treffen, Muster zu erkennen und Probleme zu lösen.

#### Beispiel:

**Spracherkennung:** KI-Systeme stellen gesprochene Sprache als eine Folge von Symbolen oder Zahlen dar und verwenden dann Schlussfolgerungstechniken, um die Bedeutung der Sprache zu interpretieren und zu verstehen. Ein KI-System kann die Audiodaten mit einer Datenbank bekannter Wörter oder Phrasen abgleichen oder die Absicht eines gesprochenen Befehls erkennen.

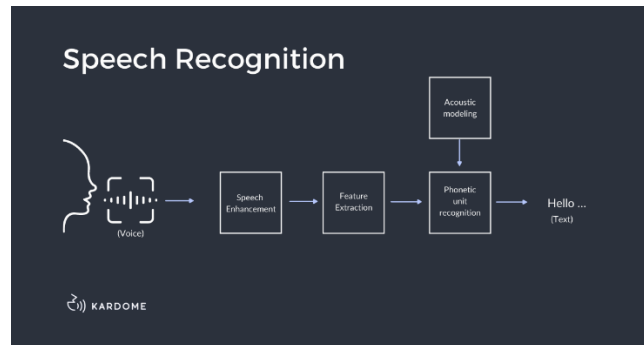


Bild 2: Spracherkennung

**Computer Vision:** KI-Systeme stellen Bilder als Zahlenmatrizen dar und verwenden dann Schlussfolgerungstechniken, um Objekte und Muster in den Bildern zu erkennen. Bei der Schlussfolgerung können Bilderkennungsalgorithmen verwendet werden, um Objekte in einer Szene zu identifizieren oder die Bewegung von Objekten über die Zeit zu verfolgen.

**Medizinischer Bereich:** KI kann große Mengen von Patientendaten analysieren, um Muster zu erkennen und mögliche Gesundheitsprobleme vorherzusagen. Dies kann Ärzten helfen, fundiertere Entscheidungen über die Patientenversorgung zu treffen.

**Finanzen:** KI kann Markttrends und Daten analysieren, um Muster zu erkennen und zukünftige Marktbewegungen vorherzusagen. Dies kann Anlegern helfen, fundiertere Entscheidungen über Investitionen zu treffen und Risiken zu steuern.

## 4.4 Idee 3: Lernen

Lernen ist ein entscheidender Aspekt der KI und unterscheidet sie von der herkömmlichen Computerprogrammierung. Algorithmen für maschinelles Lernen ermöglichen es KI-Systemen, aus Erfahrungen zu lernen und sich zu verbessern, ohne explizit programmiert zu werden. Im Zusammenhang mit Sprach- und Bildverarbeitungsanwendungen spielt das Lernen eine entscheidende Rolle, wenn es darum geht, dass Maschinen Muster erkennen und Vorhersagen auf der Grundlage von Daten treffen können. Die Bedeutung des Lernens bei Sprach- und Bildverarbeitungsanwendungen kann gar nicht hoch genug eingeschätzt werden. Durch Lernen aus großen Datenmengen können Maschinen komplexe Muster erkennen und genauere Vorhersagen treffen, was zu einer verbesserten Leistung in realen Anwendungen wie Spracherkennung, Objekterkennung und autonomes Fahren führt.

Zu den drei Hauptarten des maschinellen Lernens gehören:

**Überwachtes Lernen:** Das überwachte Lernen wird für Aufgaben wie Sprach- und Bilderkennung verwendet. Überwachtes Lernen ist eine Art des maschinellen Lernens, bei dem der Algorithmus auf einem markierten Datensatz trainiert wird. Der markierte Datensatz besteht aus Eingabedaten (z. B. Bilder oder Audiodateien) zusammen mit den entsprechenden Ausgabemarken (z. B. Text- oder Bildklassifizierung). Das Ziel des Algorithmus ist es, eine Abbildungsfunktion von den Eingabedaten auf

die Ausgabemarken zu erlernen. Bei der Spracherkennung wird der Algorithmus durch überwachtes Lernen darauf trainiert, gesprochene Wörter zu erkennen und sie in Text umzuwandeln. Die Eingabedaten bestehen aus Audiodateien, und die Ausgabemarken sind die entsprechenden Transkriptionen. Der Algorithmus wird anhand eines großen Datensatzes von gekennzeichneten Audiodateien und Transkriptionen trainiert und lernt, die Muster und Beziehungen zwischen den Audiodaten und dem entsprechenden Text zu erkennen. Einige gängige überwachte Lernalgorithmen, die in Sprach- und Bildverarbeitungsanwendungen eingesetzt werden, sind Faltungsneuronale Netze (CNNs) und rekurrente Neuronale Netze (RNNs). CNNs eignen sich besonders gut für Bilderkennungsaufgaben, da sie automatisch hierarchische Darstellungen der visuellen Merkmale lernen können. RNNs eignen sich besonders gut für Aufgaben der Spracherkennung, da sie die zeitlichen Abhängigkeiten zwischen den Audiodaten und dem entsprechenden Text modellieren können.

**Unüberwachtes Lernen:** Unüberwachtes Lernen ist eine Art des maschinellen Lernens, bei dem Muster und Beziehungen innerhalb von Daten gelernt werden, ohne dass explizite Kennzeichnungen oder Überwachung erforderlich sind. Unüberwachtes Lernen ist besonders nützlich bei Computer-Vision-Anwendungen, bei denen große Mengen unstrukturierter visueller Daten analysiert werden müssen, um sinnvolle Muster und Beziehungen zu erkennen. So können beispielsweise Clustering-Algorithmen verwendet werden, um ähnliche Bilder oder Objekte in einem Datensatz auf der Grundlage ihrer visuellen Merkmale zusammenzufassen, ohne dass eine explizite Kennzeichnung erforderlich ist. Ein weiteres Beispiel für einen unüberwachten Lernalgorithmus ist der Autoencoder. Ein Autoencoder ist ein neuronales Netz, das darauf trainiert ist, Eingabedaten aus einer niedrigdimensionalen Darstellung dieser Daten zu rekonstruieren. Der Encoder-Teil des Netzes reduziert die Eingabedaten auf einen niedriger-dimensionalen latenten Raum, und der Decoder-Teil des Netzes versucht, die ursprüngliche Eingabe aus dieser latenten Darstellung zu rekonstruieren. Autoencoder können für Aufgaben wie Bildentrauschung, Dimensionalitätsreduktion und Merkmalslernen eingesetzt werden. Ein weiteres Beispiel für einen unüberwachten Lernalgorithmus ist das generative adversarial network (GAN).

**Verstärkendes Lernen:** Bei Sprach- und Bildverarbeitungsanwendungen wird das Verstärkungslernen in Bereichen wie Robotik, autonomes Fahren und Spiele eingesetzt. Mit Hilfe des verstärkenden Lernens kann man beispielsweise einem Roboter beibringen, sich in einer unübersichtlichen Umgebung zurechtzufinden oder die Bewegungen eines selbstfahrenden Autos zu steuern. Beim Spielen kann das verstärkende Lernen dazu verwendet werden, Agenten zu entwickeln, die Spiele wie Schach oder Go auf einem übermenschlichen Niveau spielen können. Einer der bekanntesten Algorithmen für das Verstärkungslernen ist das Q-Lernen, bei dem eine Q-Tabelle verwendet wird, um die erwartete Belohnung für jede Aktion in jedem Zustand zu speichern. Der Agent aktualisiert die Q-Tabelle auf der Grundlage der Belohnungen, die er erhält, und verwendet die aktualisierte Tabelle, um künftige Aktionen zu wählen.

Actions :    ↑    →    ↓    ←

Start	0	0	0	0
Nothing / Blank	0	0	0	0
Power	0	0	0	0
Mines	0	0	0	0
END	0	0	0	0

Bild 3: Q table

## 4.5 Idee 4: Natürliche Interaktion

Natürliche Interaktion in der KI und bei Robotern ist die Fähigkeit, die natürliche Sprache und Handlungen des Menschen zu verstehen und darauf zu reagieren, z. B. durch Sprache und Gesten. Dies ist besonders bei Sprach- und Bildverarbeitungsanwendungen von Bedeutung, bei denen das Ziel darin besteht, Schnittstellen zu schaffen, die menschliche Sprache und visuelle Hinweise verstehen und auf eine verständliche, intuitive und nützliche Weise reagieren können. Technologien zur Sprach- und Bildinteraktion wie Sprach-, Gesten- und Gesichtserkennung verändern die Art und Weise, wie wir mit Maschinen interagieren, und eröffnen neue Möglichkeiten für Anwendungen in den Bereichen Spiele, Robotik, Sicherheit und mehr.

**Sprachliche Interaktion:** Sprachliche Interaktion ist die Verwendung natürlicher Sprache zur Kommunikation zwischen Menschen und KI-Systemen und umfasst Spracherkennung und Sprachsynthese. Die Spracherkennung wandelt gesprochene Sprache mithilfe von Algorithmen in Text um, während die Sprachsynthese mithilfe von TTS-Systemen künstliche Sprache aus Text erzeugt. Eine der wichtigsten Anwendungen der Sprachinteraktion in der KI sind virtuelle Assistenten wie Siri von Apple, Alexa von Amazon und Google Assistant. Diese virtuellen Assistenten nutzen Spracherkennung und NLP, um Benutzeranfragen zu verstehen und mit relevanten Informationen oder Aktionen zu antworten. Sprachinteraktion wird auch in Call-Centern eingesetzt, wo automatisierte Sprachsysteme Anrufe weiterleiten und grundlegende Fragen beantworten können, so dass menschliche Mitarbeiter für komplexere Aufgaben frei werden.

**Interaktion mit dem Sehen:** Die visuelle Interaktion bezieht sich auf die Art und Weise, wie Menschen mit KI-Systemen unter Verwendung von visuellen Hinweisen, Gesten und anderen nonverbalen Kommunikationsmethoden interagieren können. Dazu gehören Technologien wie Gesichtserkennung, Blickverfolgung und Bewegungserfassung. Diese Technologien haben viele praktische Anwendungen,

wie die Steuerung von Geräten mit Handgesten, die Verfolgung von Augenbewegungen und die Erkennung von Gesichtsausdrücken und Emotionen.

**Virtuelle und erweiterte:** Virtual und Augmented Reality (VR und AR) sind Technologien, die in der KI eingesetzt werden, um immersive Interaktionen zwischen Menschen und Maschinen zu schaffen. Bei VR wird mithilfe von Headsets und Controllern eine vollständig virtuelle Umgebung geschaffen, während bei AR virtuelle Objekte über die Sicht des Benutzers auf die reale Welt gelegt werden. Diese Technologien haben ihre Grenzen, z. B. erfordern sie spezielle Hardware und können zu Reisekrankheit führen, aber sie entwickeln sich weiter und haben zahlreiche Anwendungen in Bereichen wie Bildung, Gesundheitswesen und Unterhaltung.



*Bild 4: Virtuelle Realität*



*Bild 5: erweiterte Realität*

## 4.6 Idee 5: Gesellschaftliche Auswirkungen

Die gesellschaftlichen Auswirkungen von KI in Sprach- und Bildverarbeitungsanwendungen sind in den letzten Jahren zu einem wichtigen Diskussionsthema geworden. Diese Technologien haben zwar das Potenzial, in verschiedenen Bereichen bedeutende Fortschritte zu erzielen, sie haben aber auch ethische, soziale und wirtschaftliche Auswirkungen, die es zu berücksichtigen gilt.

**Vorteile von KI in Sprach- und Bildverarbeitungsanwendungen:** Die KI in der Sprach- und Bildverarbeitung hat Vor- und Nachteile für die Gesellschaft. Positiv ist, dass sie das Gesundheitswesen, die Bildung und die öffentliche Sicherheit verbessert hat. So hilft die Spracherkennung Ärzten bei der genauen Transkription von Krankenakten, und assistive Technologien

unterstützen Menschen mit Behinderungen. Bildverarbeitungsanwendungen tragen zur Gesichtserkennung und zu selbstfahrenden Fahrzeugen bei. Für eine verantwortungsvolle Entwicklung und Implementierung von KI-Technologie müssen jedoch ethische Überlegungen berücksichtigt werden.

**Nachteile von KI in Sprach- und Bildgebungsanwendungen:** KI in der Sprach- und Bildverarbeitung hat potenzielle Nachteile. Ein Hauptproblem ist das Potenzial für Voreingenommenheit und Diskriminierung, da KI-Algorithmen gesellschaftliche Vorurteile reproduzieren können. Aufgrund der hohen Kosten von KI-Systemen besteht außerdem die Gefahr der Verdrängung von Arbeitsplätzen und wirtschaftlicher Ungleichheit. Datenschutz und Überwachung sind ethische Bedenken, da KI persönliche Daten sammelt. Darüber hinaus gibt es allgemeine Bedenken hinsichtlich der langfristigen Auswirkungen der KI, da sie immer fortschrittlicher und autonomer wird. Es ist von entscheidender Bedeutung, diese Bedenken auszuräumen und transparentere, gerechtere und verantwortungsvollere Systeme zu schaffen, die der Gesellschaft zugute kommen.

**Ethische Erwägungen:** Obwohl KI-Technologien das Potenzial haben, erhebliche Vorteile zu bringen, gibt es auch einige ethische Überlegungen, die berücksichtigt werden müssen. Zu den wichtigsten ethischen Fragen im Zusammenhang mit KI in Sprach- und Bildverarbeitungsanwendungen gehören Datenschutz, Transparenz, Voreingenommenheit und Rechenschaftspflicht.

- **Datenschutz:** Der Einsatz von KI in Sprach- und Bildverarbeitungsanwendungen gibt Anlass zu Bedenken hinsichtlich des Datenschutzes. Da immer mehr Daten gesammelt werden, besteht das Risiko, dass personenbezogene Daten missbraucht oder falsch gehandhabt werden könnten. Daher ist es für Entwickler und Unternehmen wichtig, sicherzustellen, dass ein angemessener Schutz der Privatsphäre besteht, um persönliche Daten zu schützen.
- **Transparenz:** Ein weiteres ethisches Problem im Zusammenhang mit KI in Sprach- und Bildverarbeitungsanwendungen ist die Frage der Transparenz. Es kann schwierig sein zu verstehen, wie ein KI-System zu seinen Entscheidungen kommt, was es schwierig machen kann, die Fairness und Genauigkeit seiner Ergebnisse zu beurteilen. Um diesem Problem zu begegnen, müssen die Entwickler die Transparenz ihrer KI-Systeme in den Vordergrund stellen und dafür sorgen, dass sie überprüft und erklärt werden können.
- **Voreingenommenheit:** Voreingenommenheit ist ein weiteres ethisches Problem, das bei KI in Sprach- und Bildverarbeitungsanwendungen auftreten kann. Wenn die Daten, die zum Trainieren eines KI-Systems verwendet werden, voreingenommen sind, dann können auch die Ergebnisse des Systems voreingenommen sein. Dies kann erhebliche Auswirkungen auf Bereiche wie Personaleinstellung, Kreditvergabe und Strafjustiz haben. Um dieses Problem zu lösen, müssen die Entwickler dafür sorgen, dass ihre Datensätze repräsentativ und unvoreingenommen sind.
- **Rechenschaftspflicht:** Schließlich wirft der Einsatz von KI in Sprach- und Bildverarbeitungsanwendungen Fragen zur Rechenschaftspflicht auf. Wer ist verantwortlich, wenn ein KI-System einen Fehler macht oder Schaden anrichtet? Da KI immer stärker in unser Leben integriert wird, muss sichergestellt werden, dass es klare Verantwortlichkeiten und Zuständigkeiten gibt.



## 4.7 Fallstudien und Erfolgsgeschichten

In Fallstudien werden praktische KI-Anwendungen in den Bereichen Sprache und Sehen vorgestellt. DeepMind AI for Lip-Reading von Google verbessert die Spracherkennung für Menschen mit Hörproblemen, während Watson for Oncology von IBM Onkologen bei der Erstellung personalisierter Krebsbehandlungspläne unterstützt. Amazons Alexa für Spracherkennung hat die Art und Weise revolutioniert, wie Menschen mit Technologie interagieren. Weitere bemerkenswerte KI-Anwendungen in den Bereichen Sprache und Bildverarbeitung sind die Gesichtserkennung von Apple und der Autopilot von Tesla. Diese Beispiele zeigen das transformative Potenzial der KI für die Verbesserung unseres Lebens.

### 4.7.1 Beispiel 1: Googles DeepMind-KI zum Lippenlesen

Googles DeepMind hat ein KI-System entwickelt, das auf der Grundlage von 5.000 Stunden TV-Material trainiert wurde und eine Genauigkeit von 46,8 % beim Lippenlesen erreicht hat, was die 12,4 % der menschlichen Experten übertrifft. Das System könnte in den Bereichen Spracherkennung, Sicherheit und Unterstützung von Menschen mit Hörbehinderungen eingesetzt werden. Lippenlesen ist eine Herausforderung, und selbst die besten menschlichen Lippenleser können nur etwa 30 % der gesprochenen Wörter richtig verstehen. Die KI von DeepMind hat das Potenzial, die Kommunikation für Menschen mit Hörbehinderungen zu revolutionieren. Es gibt Bedenken hinsichtlich des Schutzes der Privatsphäre bei der Erfassung und Analyse großer Mengen von Videomaterial und der Möglichkeit des Missbrauchs im Rahmen der Überwachung und Strafverfolgung. Insgesamt zeigt diese Technologie, wie KI die Kommunikation für Menschen mit Hörbehinderungen verbessern kann, und verdeutlicht die Notwendigkeit ethischer Überlegungen bei der Entwicklung und Nutzung von KI..

### 4.7.2 Beispiel 2: IBMs Watson für die Onkologie

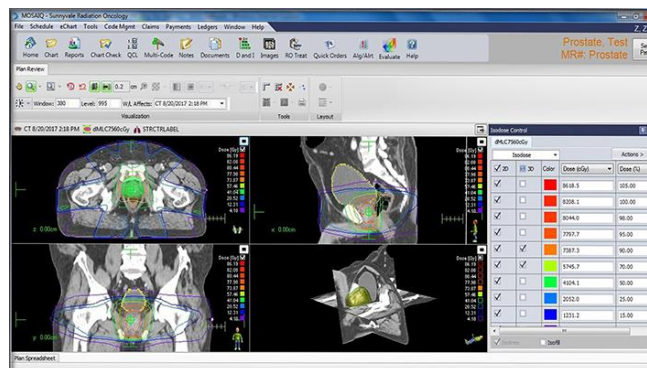


Bild 6: IBM Watson

Watson for Oncology von IBM ist ein KI-System, das Onkologen dabei unterstützt, personalisierte Behandlungsentscheidungen für Krebspatienten zu treffen. Das System nutzt die Verarbeitung natürlicher Sprache und maschinelles Lernen, um große Mengen medizinischer Daten zu analysieren und Behandlungsempfehlungen auf der Grundlage der individuellen Umstände eines Patienten zu geben. Obwohl es einige Herausforderungen und Kritikpunkte an dem System gab, wurde es erfolgreich im Gesundheitswesen eingesetzt und ist ein vielversprechendes Beispiel dafür, wie KI die Ergebnisse für Patienten verbessern kann.

### 4.7.3 Beispiel 3: Amazons Alexa für Spracherkennung

Alexa von Amazon ist ein beliebter virtueller Assistent, der mit Hilfe von Spracherkennungstechnologie eine freihändige Bedienung verschiedener Geräte ermöglicht. Er erkennt und reagiert auf Befehle in natürlicher Sprache und nutzt maschinelles Lernen, um personalisierte Empfehlungen zu geben. Alexa hat auch Menschen mit Behinderungen geholfen, indem sie ihnen die Erledigung alltäglicher Aufgaben erleichtert hat. Allerdings müssen Datenschutz- und Sicherheitsbedenken berücksichtigt werden. Der Erfolg von Alexa zeigt das Potenzial der Spracherkennungstechnologie, unser tägliches Leben zu verbessern, aber ethische Erwägungen und die mit der KI verbundenen Risiken müssen berücksichtigt werden.

## 4. Zusätzliche Materialien und Ressourcen

Art der Ressource	Titel	Thema	Link
Artikel	"Spracherkennungssysteme: Die Zukunft ist jetzt"	Spracherkennung	<a href="https://www.wired.com/story/speech-recognition-future-is-now/">https://www.wired.com/story/speech-recognition-future-is-now/</a>
Artikel	jetzt"	Computer Vision	<a href="https://www.forbes.com/sites/forbestechcouncil/2021/01/05/computer-vision-the-future-of-artificial-intelligence/?sh=39c773b742f4">https://www.forbes.com/sites/forbestechcouncil/2021/01/05/computer-vision-the-future-of-artificial-intelligence/?sh=39c773b742f4</a>
Artikel	"Computer Vision: Die Zukunft der künstlichen Intelligenz"	Tiefes Lernen	<a href="https://blog.hubspot.com/marketing/deep-learning-visual-content">https://blog.hubspot.com/marketing/deep-learning-visual-content</a>
Artikel	e"	KI im Gesundheitswesen	<a href="https://www.healthtechmagazine.net/article/2020/03/pros-and-cons-ai-healthcare">https://www.healthtechmagazine.net/article/2020/03/pros-and-cons-ai-healthcare</a>



Video	"Wie Deep Learning die Zukunft der visuellen Inhalte verändert	Vorteile von Sehen und Sprechen im Gesundheitswesen	<a href="https://www.ted.com/talks/pratik_shah_how_ai_is_making_it_easier_to_diagnose_disease">https://www.ted.com/talks/pratik_shah_how_ai_is_making_it_easier_to_diagnose_disease</a>
Video	Inhalte"	Die weltverändernden Auswirkungen der KI-Entwicklung und -Nutzung	<a href="https://www.youtube.com/watch?v=BDBTJOGvCv4">https://www.youtube.com/watch?v=BDBTJOGvCv4</a>
Link	"Das Für und Wider von KI im Gesundheitswesen	TensorFlow: eine Open-Source-Plattform für maschinelles Lernen zum Erstellen von Deep-Learning-Modellen	<a href="https://www.tensorflow.org/">https://www.tensorflow.org/</a>
Link	e"	OpenCV: eine Open-Source-Computer-Vision-Bibliothek zum Erstellen von Computer-Vision-Anwendungen in Echtzeit	<a href="https://opencv.org/">https://opencv.org/</a>
Link	"TED	PyTorch: eine Open-Source-Bibliothek für maschinelles Lernen zum Erstellen von Deep-Learning-Modellen	<a href="https://pytorch.org/">https://pytorch.org/</a>
Link	NVIDIA	NVIDIA Deep Learning Institute	<a href="https://www.nvidia.com/en-us/deep-learning-ai/education/">https://www.nvidia.com/en-us/deep-learning-ai/education/</a>

## 5 Nachbereitung

Zusammenfassend lässt sich sagen, dass in der Einheit über die Anwendung von KI in der Sprach- und Bildverarbeitung verschiedene Themen behandelt wurden, darunter die Verarbeitung natürlicher Sprache, das maschinelle Sehen, das maschinelle Lernen und die Auswirkungen auf die Gesellschaft. Wir haben Fallstudien wie Googles DeepMind KI für das Lippenlesen, IBMs Watson für die Onkologie und Amazons Alexa für die Spracherkennung untersucht, die das Potenzial der KI bei der Lösung von Problemen in der realen Welt verdeutlichen. Allerdings müssen auch ethische Erwägungen sowie das Potenzial für Verzerrungen und die Grenzen der KI sorgfältig geprüft werden. Mit diesem Wissen können die Lernenden die enormen Möglichkeiten und Herausforderungen auf dem Gebiet der KI weiter erforschen.

## 6 Quiz

Frage 1: Richtig/Falsch

- a) Deep Learning ist eine Teilmenge des maschinellen Lernens. (**Richtig**/Falsch)  
b) Verzerrung und Überanpassung sind häufige Probleme beim maschinellen Lernen. (Richtig/**Falsch**)

Frage 2: Ordnen Sie die folgenden Arten des maschinellen Lernens den entsprechenden Beschreibungen zu

a. Überwachtes Lernen	i. Lernen mit gelabelten Daten
b. Unüberwachtes Lernen	ii. Lernen ohne markierte Daten
c. Verstärkendes Lernen	iii. Lernen mit einem Belohnungssystem

**Antworten: a. i, b. ii, c. iii**

Frage 3: Ordnen Sie die folgenden Algorithmen des maschinellen Lernens den jeweiligen Anwendungen zu

a. Faltungsneuronale Netze (CNNs)	i. Spracherkennung
b. Autoencoder	ii. Bildklassifizierung
c. Q-learning	iii. Bildklassifizierung

**Antworten: a. ii, b. i, c. iii**

Frage 4: Was ist der Hauptunterschied zwischen überwachtem und unüberwachtem Lernen?

- a) **Überwachtes Lernen erfordert gelabelte Daten, während unüberwachtes Lernen dies nicht tut.**
- b) Unüberwachtes Lernen erfordert gekennzeichnete Daten, überwachtes Lernen dagegen nicht.
- c) Beide Arten des Lernens erfordern markierte Daten.
- d) Es gibt keinen Unterschied zwischen überwachtem und unüberwachtem Lernen.

Frage 5: Which type of learning is used for robotics and game playing?

- a) Überwachtes Lernen
- b) Unüberwachtes Lernen
- c) **Verstärkendes Lernen**
- d) Keiner der oben genannten Punkte

Frage 6: Was sind die wichtigsten Einschränkungen des maschinellen Lernens bei Sprach- und Bildverarbeitungsanwendungen?

- a) Mangel an Rechenleistung
- b) Mangel an etikettierten Daten
- c) **Verzerrung und Überanpassung**
- d) Mangelnde Interpretierbarkeit

Frage 7: Was ist der Hauptvorteil der Verwendung von Deep Learning?

- a) **Bessere Genauigkeit bei komplexen Aufgaben**
- b) Geringere Anforderungen an die Rechenleistung
- c) Weniger anfällig für Überanpassung
- d) Leichter zu interpretieren

Frage 8: Was ist die wichtigste ethische Überlegung bei der Entwicklung von KI für gesellschaftliche Auswirkungen?

- a) Gewährleistung von Sicherheit und Schutz der Nutzer
- b) **Gewährleistung von Fairness und Vermeidung von Voreingenommenheit**
- c) Sicherstellung der Rentabilität des Unternehmens
- d) Sicherstellung der Kompatibilität mit bestehenden Systemen

Frage 9: Was ist der Hauptvorteil der Verwendung von Augmented Reality in Kombination mit KI?

- a) **Verbesserung des Benutzererlebnisses**
- b) Verringerung des Rechenaufwands
- c) Verbesserung der Genauigkeit bei komplexen Aufgaben
- d) KI besser interpretierbar machen

## 7 Referenzen

- Kelleher, J. D., & Tierney, B. (2018). Data science: An introduction. CRC Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
- Jurafsky, D., & Martin, J. H. (2020). Speech and language processing (3rd ed.). Pearson.

- Zhang, L., Tan, T., & Li, Y. (2016). Deep learning-based face recognition: A survey. *International Journal of Automation and Computing*, 13(4), 261-279.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484- 489.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
- Weng, L., Zhang, Y., Xue, X., & Chen, K. (2021). The applications of artificial intelligence in computer vision. In *2021 International Conference on Computer Science and Artificial Intelligence (CSAI)* (pp. 139-143). IEEE.
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
- Dhillon, G., & Moores, D. (2018). Natural language processing in AI: past, present and future. In *Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval* (pp. 1-10).
- Canny, J. (2021). Ethics and artificial intelligence. *Communications of the ACM*, 64(1), 18-20.
- Li, X., Liang, J., & Li, Y. (2021). Research on the impact of augmented reality technology on the experience of cultural tourism. *Journal of Physics: Conference Series*, 1838(1), 012064.
- Yaqoob, I., Ahmed, M. M., Gani, A., Imran, M., Guizani, M., & Hitha, H. (2019). Virtual reality for education: A survey. *Journal of Network and Com*