Testen von KI-Systemen

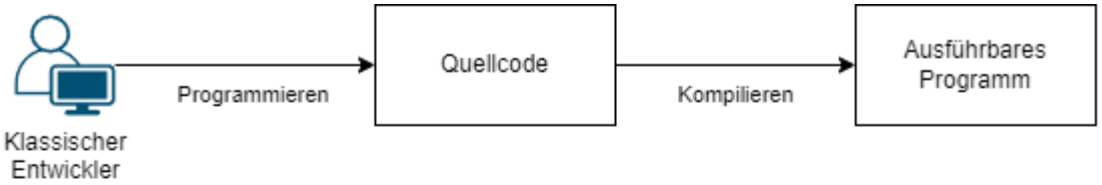
Teil 5 – Tests für KI-Systeme

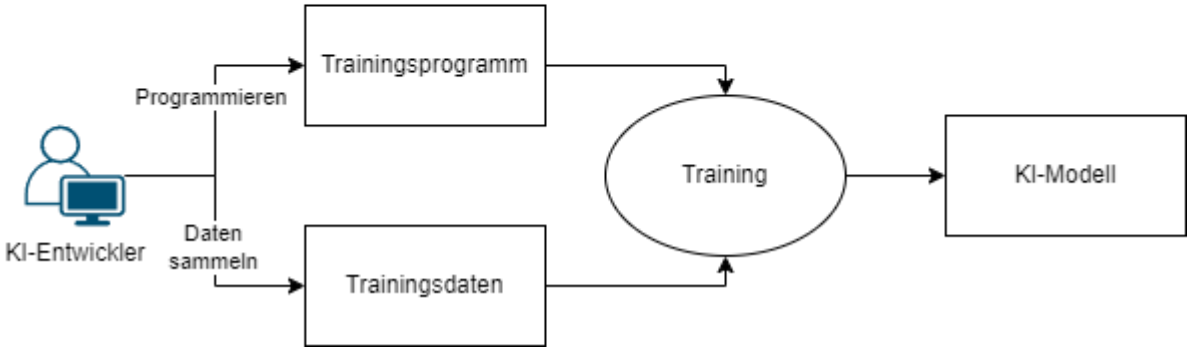
INNOVET
KI B³
Künstliche Intelligenz in die berufliche Bildung bringen

INNOVET
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
bibb Bundesinstitut für Berufsbildung

Gefördert als InnoVET-Projekt aus Mitteln des Bundesministeriums für Bildung und Forschung.

Einführung: ‘Entwicklungsparadigmen’





Qualitätseigenschaften von KI-Systemen

- Flexibilität and Anpassungsfähigkeit
  - Flexibilität und Anpassungsfähigkeit sind eng verwandt.

Qualitätseigenschaften von KI-Systemen

- **Flexibilität** beschreibt die Möglichkeit, das System für Einsatzzwecke zu nutzen, die ursprünglich nicht in den Anforderungen vorgesehen waren.

- **Anpassungsfähigkeit** beschreibt den nötigen Aufwand, um ein System auf neue Einsatzzwecke anzupassen, wie etwa geänderte Hardware oder eine geänderte Betriebsumgebung.

Qualitätsmerkmale für KI-Systeme

- Sowohl Flexibilität als auch Anpassungsfähigkeit sind nützlich, wenn:
  - Die Details aus dem Betriebsumfeld nicht vorab bekannt sind.
  - Das System auch in fremden Betriebsumgebungen zurecht kommen soll.
  - Das System sich an neue Situationen anpassen soll.
  - Das System selbst erkennen soll, wann es sein Verhalten ändern muss.

Qualitätsmerkmale für KI-Systeme

- **Autonomie** ist die Fähigkeit eines Systems, auch über längere Zeit ohne menschliche Aufsicht zu arbeiten.

- Die Autonomie-Anforderungen sind für die Spezifikation und Durchführung geeigneter Tests entscheident.
- Es ist wichtig, herauszufinden, wie lange ein KI-System ohne menschlichen Eingriff funktioniert.
- In welchen Szenarien sollte das autonome KI-System die Kontrolle an einen Menschen zurückgeben?

Qualitätsmerkmale für KI-Systeme

- **Evolution** ist die Fähigkeit eines Systems, sich an veränderte Bedingungen anzupassen.

- In einem Musik-Empfehlungssystem könnte sich der Algorithmus an sich ändernde Charts und sich ändernde Verhaltensweisen von Nutzern anpassen, und so die Relevanz der Musikvorschläge verbessern.

Qualitätsmerkmale für KI-Systeme

- **Voreingenommenheit („Bias“)**
- In KI-Systemen beschreibt die Voreingenommenheit den mathematischen Abstand zwischen den KI-Ergebnissen und „fairen Ergebnissen“, die keine Gruppen bevorzugen.
- Solche Voreingenommenheit kann sich z.B. auf Geschlecht, Rasse, Ethnizität, Sexualität, Einkommen oder Alter beziehen.

Qualitätsmerkmale für KI-Systeme

- **Voreingenommenheit**

- Fälle von Voreingenommenheit wurden schon in vielen Systemen beobachtet. Besonders problematisch äußert sich das bei Jobportalen, Kredit-Scoring und juristischen Systemen

Qualitätsmerkmale für KI-Systeme

- **Ethik** beschreibt einen Moralkodex für Verhaltensweisen.
- Fortgeschrittene KI-Systeme haben viele Vorteile im Leben der Menschen. Mit zunehmender Verbreitung werden auch Bedenken geäußert, ob und wann der Einsatz ethisch vertretbar ist.
- Ethisches Verhalten ändert sich über die Zeit und ist abhängig vom Kulturkreis.
- Bei der Einführung von KI-Systemen an unterschiedlichen Standorten müssen ethische Überlegungen immer neu durchgeführt werden.

Qualitätsmerkmale für KI-Systeme

- **Erklärbare KI**

- Technisch sind Entscheidungen der KI nicht leicht nachzuvollziehen. Oft ist dies jedoch nötig, und muss bei der Entwicklung des Systems berücksichtigt werden. Das stärkt auch das Vertrauen der Nutzer in das KI-System.

Qualitätsmerkmale für KI-Systeme

- **Erklärbare KI** lässt sich in 3 Eigenschaften aufteilen:

- **Transparenz**: Wie leicht kann man herausfinden, mit welchem Algorithmus und welchen Trainingsdaten das Modell erstellt wurde?
- **Interpretierbarkeit**: Wie gut verstehen die Stakeholder die verwendete Technologie?
- **Erklärbarkeit**: Wie leicht kann der Nutzer feststellen, wie die KI zu einem bestimmten Ergebnis kommt?

Qualitätsmerkmale für KI-Systeme

- **Sicherheit** beschreibt die Erwartung, das Systeme keine Schäden an Menschen, Gegenständen oder der Umwelt verursachen.
- KI-Systeme, die in Gesundheitswesen, Produktion, Militär und Straßenverkehr benutzt werden, können diese Sicherheit verletzen.

- Die Herausforderung, all diese Merkmale zu testen, wird in den anderen Abschnitten behandelt.

Arbeitsablauf

1. Modellentwicklung
    1. Framework & Algorithmus wählen
    2. Testdaten vorbereiten
    3. Modellerzeugung (n.B. wiederholen)
        1. Modell Trainieren
        2. Modell evaluieren
        3. Modell anpassen
    4. Modell Testen
2. Modell Ausrollen
3. Modell wird benutzt
4. Überwachung und Anpassung des Modells

Arbeitsablauf

In diesem Arbeitsablauf werden insbesondere vier Elemente immer wieder wichtig:

- Algorithmen, die verschiedene Stärken und Schwächen haben
- Daten, die erzeugt, geprüft und beschriftet werden müssen
- Interaktionen mit anderen KI-Systemen
- Tests unter Realbedingungen

Arbeitsablauf

- **Trainings-, Validierungs- und Testdatensätze**
  - Übung: Datensatz aufteilen
    - Teile einen Datensatz für ein Empfehlungssystem in Trainings-, Validierungs- und Testdaten auf.
    - Trainiere damit ein Empfehlungsmodell mit *Supervised Learning*.
    - Vergleiche die Leistung des Modells auf Validierungs- und Testdaten!

Arbeitsablauf

Datenqualität

- Daten können falsch, unvollständig, unzureichend, irrelevant, veraltet, unausgewogen, unfair sein; sie können Duplikate enthalten, falsch oder gar nicht klassifiziert sein und die Privatsphäre oder Sicherheit verletzen.

Arbeitsablauf

Auswirkungen der Datenqualität auf das KI-Modell

- Die Qualität der KI-Modells kann nur so hoch sein, wie die Datenqualität des Trainingsdatensatzes.
- Schlechte Datenqualität führt zu verminderter Genauigkeit, Voreingenommenheit, kompromittierten Modellen und vielen weiteren Problemen

Arbeitsablauf

Auswirkungen der Datenqualität auf das KI-Modell

- **Verminderte Genauigkeit:**
  - Wird meist von falschen, unvollständigen, falsch oder nicht klassifizierten, unzureichenden, veralteten und irrelevanten Daten verursacht.

  - Wenn das Recommender-System nur mit Daten über Liebesfilmen trainiert wird, sind die Vorschläge für Actionfilme oder Komödien vermutlich sehr ungenau.

Arbeitsablauf

Auswirkungen der Datenqualität auf das KI-Modell

- **Voreingenommenheit**:
  - Wird meist von unvollständigen, unausgewogenen, unfairen Daten verursacht.
  - Wird das Recommender-System nur auf Daten von 10-20 Jahre alten Mädchen trainiert, empfiehlt das System vermutlich keine Filme, die andere Zielgruppen ansprechen.

Arbeitsablauf

Auswirkungen der Datenqualität auf das KI-Modell

- **Kompromittiertes Modell:**
  - Ein kompromittiertes Modell entsteht, wenn die Trainingsdaten die Privatsphäre oder Sicherheit verletzen.

  - Wenn ein KI-System zur Autovervollständigung von Text mit privaten Adressen trainiert wird, kann es möglicherweise private Adressdaten automatisch vervollständigen.

INNOVET
KI B³
Künstliche Intelligenz in die berufliche Bildung bringen

Tobias Eisenreich

Universität Stuttgart
Institut für Software Engineering
Empirisches Software Engineering

Umm-e-Habiba

Universität Stuttgart
Institut für Software Engineering
Empirisches Software Engineering

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
Institut für Software Engineering

IHK Industrie- und Handelskammer Reutlingen
Reutlingen | Tübingen | Zollernalb

IHK Region Stuttgart

IHK Industrie- und Handelskammer Karlsruhe

LMU Ludwig-Maximilians-Universität München

GEFÖRDERT VOM
INNOVET
Bundesministerium für Bildung und Forschung
bibb Bundesinstitut für Berufsbildung

Gefördert als InnoVET-Projekt aus Mitteln des Bundesministeriums für Bildung und Forschung.

Lizenzbestimmungen

„Tests für KI-Systeme“ von Umm-e-Habiba und Tobias Eisenreich, KIB3 / Uni Stuttgart

Quellenverzeichnis

https://unsplash.com/de/fotos/vpOeXr5wmR4