

Pressemitteilung

Nr. 47 vom 31. Mai 2021

Mehr Transparenz bei Künstlicher Intelligenz

Ob selbstfahrende Autos oder die Steuerung von Industrieanlagen: Künstliche Intelligenz spielt eine immer größere Rolle in unserer Gesellschaft. Die TH Köln untersucht gemeinsam mit der Ruhr-Universität Bochum in einem Promotionstandem, wie Künstliche Intelligenz transparenter und verständlicher gemacht werden kann. Das Projekt „(RL)³ Repräsentations-, Reinforcement- und Regel-Lernen“ wird im Rahmen des Graduiertenkollegs „Data-NInJA“ vom Ministerium für Kultur und Wissenschaft NRW mit insgesamt 500.000 Euro gefördert.

„Künstliche Intelligenz kann viel, aber es mangelt ihr noch an Transparenz, da nicht sichtbar ist, was sie dazu bringt, eine bestimmte Handlung auszuführen. Das möchten wir ändern, indem wir sie erklärbarer und dadurch auch vertrauenswürdiger machen“, sagt Prof. Dr. Wolfgang Konen vom Institut für Informatik der TH Köln über das Ziel des Projekts „(RL)³ Repräsentations-, Reinforcement- und Regel-Lernen“.

Gemeinsam mit Prof. Dr. Laurenz Wiskott vom Institut für Neuroinformatik der Ruhr-Universität Bochum betreut er die Doktoranden Moritz Lange und Raphael Engelhardt. Moritz Lange von der Ruhr-Universität Bochum beschäftigt sich in seiner Dissertation mit dem Repräsentations-Lernen. Repräsentationen sind extrahierte Informationen, auf denen das Reinforcement-Lernen aufbaut und die diesen Prozess erleichtern. „Wie gut die relevanten Informationen aus den Daten ablesbar sind, ist entscheidend für die Fähigkeiten eines Algorithmus, daraus zu lernen“, sagt Lange. Gute Repräsentationen seien zusätzlich auch wichtig, um analysieren zu können, wie ein Algorithmus auf verschiedene Szenarien reagiert. Um dieses Verständnis entwickeln zu können, müssen komplexe Datensätze so dargestellt werden, dass sie durchschaubar sind.

Beim Reinforcement-Lernen handelt es sich um das Training einer Computersoftware, auch als Agent bezeichnet, der durch Belohnung lernt, ob seine Aktion ihn in einen besseren oder schlechteren Zustand führt. „Diese Belohnung ist beispielsweise eine Zahl und hängt vom konkreten Fall ab: Soll ein Stab aufrecht balanciert werden, ist eine denkbare Belohnung die Zahl eins für jede Sekunde, die der Stab balanciert wird. Alle Belohnungen, die der Agent nach und nach erhält, werden addiert. Sein Ziel ist es, die Aktionen so zu wählen, dass er eine möglichst große Gesamtbelohnung zu erwarten hat“, erläutert Raphael Engelhardt, Doktorand an der TH Köln. Dieser Prozess wird auch Verstärkungslernen genannt, nach dessen Ende der Agent im Idealfall eine bestimmte Aufgabe lösen kann.

„Das Problem ist, dass bei einem Agenten nach Abschluss der Lernprozesse nicht ersichtlich ist, wie dieser zu seiner Strategie gekommen ist. Man sieht nur, der Agent erreicht das Ziel. Aber was ihn dazu bringt, ist nicht mehr nachvollziehbar. Deshalb spricht man auch von einem Black-Box-Effekt. Man kann sich das ähnlich wie bei der Intuition eines Menschen vorstellen; ein Profi entwickelt eine gewisse Intuition für eine Handlung und wird Schwierigkeiten haben, diese detailliert zu erklären“, sagt Engelhardt. Er versucht anhand von Simulationen, Regelmäßigkeiten innerhalb der Black-Box zu erkennen und zu extrahieren. Es soll ermöglicht werden, das komplexe Verstärkungslernen in einfache Regeln zu überführen. Engelhardt erforscht so das Regel-Lernen.

Referat Kommunikation und
Marketing
Presse- und Öffentlichkeitsarbeit
Sandy Syperek
0221-8275-5147
pressestelle@th-koeln.de

Technische Hochschule Köln

Postanschrift:
Gustav-Heinemann-Ufer 54
50968 Köln

Sitz des Präsidiums:
Claudiusstraße 1
50678 Köln

Pressemitteilung Nr. 47 vom 31. Mai 2021
(RL)3 Repräsentations- Reinforcement- und Regel-Lernen

Angestrebte Anwendungsgebiete des Projekts sind, die Steuerung von technisch komplexen Anlagen, beispielsweise bei Stahlwalzprozessen, oder die Unterstützung von Logistikleitzentren mit KI-Modellen. Das Projekt hat eine Laufzeit von dreieinhalb Jahren und wird mit 500.000 Euro gefördert, beide projektbeteiligten Hochschulen erhalten jeweils 250.000.

Graduiertenkolleg Data-NInJA

Das Graduiertenkolleg „Trustworthy AI for Seamless Problem Solving: Next Generation Intelligence Joins Robust Data Analysis“, kurz Data-NInJA NRW, widmet sich Forschungsthemen rund um Künstliche Intelligenz und wird vom Ministerium für Kultur und Wissenschaft Nordrhein-Westfalen mit insgesamt fünf Millionen Euro gefördert. Die Gesamtkoordination des Graduiertenkollegs, in dem 14 Promotionen bewilligt wurden, übernimmt die Universität Bielefeld. Das Kolleg soll den Promovierenden ermöglichen, neben der technologischen Qualifizierung auch Kompetenzen zu gesellschaftlichen, sozioökonomischen und politischen Effekten der künstlichen Intelligenz zu erwerben.

Die **TH Köln** zählt zu den innovativsten Hochschulen für Angewandte Wissenschaften. Sie bietet Studierenden sowie Wissenschaftlerinnen und Wissenschaftlern aus dem In- und Ausland ein inspirierendes Lern-, Arbeits- und Forschungsumfeld in den Sozial-, Kultur-, Gesellschafts-, Ingenieur- und Naturwissenschaften. Zurzeit sind rund 27.000 Studierende in etwa 100 Bachelor- und Masterstudiengängen eingeschrieben. Die TH Köln gestaltet Soziale Innovation – mit diesem Anspruch begegnen wir den Herausforderungen der Gesellschaft. Unser interdisziplinäres Denken und Handeln, unsere regionalen, nationalen und internationalen Aktivitäten machen uns in vielen Bereichen zur geschätzten Kooperationspartnerin und Wegbereiterin.