

## Kann man Roboter bestrafen? – Gabriele Peters über Mensch-Maschine-Interaktion

In der Sendung „Neugier genügt“ auf WDR5 wurde Gabriele Peters zu der Frage „Kann man Roboter bestrafen?“ interviewt.

Das Interview führte Thomas Koch am 1. Oktober 2014.

**Thomas Koch** Bei „Neugier genügt“, WDR 5, geht es jetzt um den Zusammenhang zwischen Lernen und Bestrafung. Nun wird der Begriff *Strafe* in der modernen Pädagogik ja gar nicht mehr so gerne gehört aber - keine Sorge – es geht hier auch um die Bestrafung von Robotern, und die kennen ja keinen Schmerz. Aber wie kann man sie dann bestrafen? Denn Strafe ohne Schmerz - wie soll das funktionieren? Und wenn ein Roboter nun über Schmerzerfahrung lernfähig sein sollte, dann hätte er womöglich sogar so etwas wie ein Bewusstsein, würde er empfinden, und dann taucht auch schon die Frage auf, ob man Roboter quälen würde durch Bestrafung, und ob man das überhaupt darf? Und schon sind wir mitten in einer Diskussion, die eine technische, eine philosophische und eine ethische Dimension hat. Wie gut, dass Gabriele Peters, Professorin für Mensch-Computer-Interaktion an der FernUniversität in Hagen bei uns zu Gast ist. Guten Morgen, Frau Peters.

**Gabriele Peters** Guten Morgen, Herr Koch.

**Koch** Fangen wir mal ganz einfach technisch an, weil Sie da auch am besten zu Hause sind, nehme ich mal an. Kann man Roboter bestrafen?

**Peters** Das kann man. Wir entwickeln an unserem Lehrgebiet interaktive Systeme. Das sind Systeme, die möglichst auf eine intuitive Weise mit dem Menschen interagieren sollen. Und eine Forderung, die man an solche Systeme stellt, ist dass sie lernfähig sind. Und zwar auch noch dann, wenn sie schon im Einsatz sind. Sie sollen sich also auf eine sich verändernde Umwelt und den Menschen einstellen können. Wenn man zum Beispiel einen Pflegeroboter nimmt, der die Atmung eines Patienten überwachen und Alarm geben soll, wenn die Atmung sich bedrohlich entwickelt, dann würde man davon ausgehen, dass dieser Roboter sich vorher auf das individuelle Atemmuster dieses Patienten eingestellt hat - es *erlernt* hat.

**Koch** Aber ist denn der Begriff *Strafe* dann überhaupt passend? Wenn wir uns das vorstellen: wenn er dann irgendetwas falsch macht, dass man ihn bestraft - wie könnte das dann aussehen?

**Peters** Man kann Roboter bestrafen, wenn man das erstmal vom technischen Aspekt her sieht. Das Verfahren, das wir anwenden, nennt sich *Reinforcement Learning*. Das kann ich am besten anhand einer Aufgabe erklären, die wir in diesem Semester unseren Studierenden in einem Praktikum stellen. Wir haben einen mobilen Roboter in unserem Labor. Der soll eigenständig einen Pfad vorbei an Hindernissen zu einer Stelle finden - das kann man sich als Futterplatz vorstellen - und zwar ohne, dass die Programmierer die Strategie dafür vorgeben. Es funktioniert so, dass nur Belohnungs- und Bestrafungssignale gegeben werden. Das sind positive und negative Zahlen...

**Koch** Moment. Positive, negative Zahlen sorgen für Belohnung und Bestrafung?

**Peters** Ganz genau.

**Koch** Wie funktioniert das?

**Peters** Das wird so interpretiert. Und zwar läuft der Roboter am Anfang zufällig und erhält einfach diese Signale an den Hindernissen. Zufällig findet er am Anfang dann zum Beispiel auch mal diesen Futterplatz. Dort bekommt er ein positives Signal zurück gemeldet. Und dadurch, dass er viele Versuche durchführt, baut er sich intern eine Wertefunktion auf, die ihm sagt, in welcher Situation welche Aktion gut war - also nach links gehen, nach rechts gehen...

**Koch** Aber wie kann er das unterscheiden? Wie kann ein Roboter positive und negative - letztendlich elektronische Signale unterscheiden?

**Peters** Die kann er erst einmal nur auf einer technischen Ebene mit einer mathematischen Funktion unterscheiden. Man kann sich vielleicht zunächst die Frage stellen, welchen Zweck eine Bestrafung bei einem Menschen oder einem Tier hat. Der Hauptsinn einer Bestrafung ist es ja, eine Verhaltensveränderung herbei zu führen. Und wenn man diese Frage erst einmal auf einem niedrigen, auf einem technischen Niveau versteht, kann man schon sagen, dass man auch bei einem Roboter eine Verhaltensänderung dadurch erzielen kann, dass man ihn mit einer numerischen Bestrafung versieht. Das geht schon. Aber natürlich kann man diese Frage auch umgangssprachlicher verstehen. Dann würde man natürlich vermuten, dass man eine gewisse Art von Einsicht voraussetzen muss, um wirkungsvoll bestraft werden zu können.

**Koch** Oder eine Emotionalität...

**Peters** Genau.

**Koch** ... die wir ja auch sehr stark damit verbinden.

**Peters** Also, das ist heute noch nicht der Fall bei Robotern.

**Koch** Also die Roboter leiden nicht.

**Peters** Sie leiden *noch* nicht.

**Koch** *Noch* nicht?

**Peters** Genau. An der Stelle ist man an einem interessanten Punkt. Denn da stellt sich die Frage nach der Natur des Bewusstseins, und ob es möglich ist, dass es vielleicht irgendwann einmal künstliche Systeme gibt, die tatsächlich auch Bewusstsein haben.

**Koch** Das wäre dann aber auch eine Programmierungsleistung, das letztendlich herzustellen, oder?

**Peters** Wenn Sie mich jetzt rein nach der Möglichkeit fragen, ob es solche Systeme irgendwann geben wird, dann sage ich erst einmal: zunächst gibt es für mich keinen Grund, das auszuschließen. Aber vielleicht sind das dann eher Systeme, die Mischwesen sind, die technische Bestandteile und biologische Bestandteile haben.

**Koch** Aber das ist kein Tabu in der Forschung? Das sind Dinge, die schon jetzt nicht so ganz die Super-Science-Fiction-Musik sind, oder?

**Peters** Ich würde sagen, dass erst einmal grundsätzlich gar nichts in der Forschung tabu ist. Aber das ist schon in sehr weiter Ferne. Was aber der Grund ist, warum die Roboter heute noch kein Bewusstsein haben, ist, dass man das *menschliche* Bewusstsein noch nicht verstanden hat. Es gibt keine allgemeingültige Definition von

Bewusstsein. Die Philosophen, Informatiker, Neurowissenschaftler sind sich nur einig darüber, dass bestimmte Aspekte dazugehören, wie zum Beispiel Aufmerksamkeit, Gedächtnis, Lernfähigkeit. Und einige dieser Aspekte sind eben einem rechnerbasierten Ansatz zugänglicher, wie zum Beispiel das Lernen, andere sind schwieriger nachzubilden, wie zum Beispiel Motivation oder Selbsterkenntnis.

**Koch** Was sind das für Fragen mit denen Sie sich dann – weil Sie gerade auf Ihre Studenten angesprochen haben zu Beginn, als Sie das Beispiel vorgestellt haben – was sind das dann für konkrete Fragen, um die es bei der Arbeit bei Ihnen geht?

**Peters** Bei uns findet das Ganze eher auf einer wirklich technischen Ebene statt. Wir haben nicht das Ziel, künstliches Bewusstsein zu erzeugen in dem Praktikum in diesem Wintersemester. Was mir aber wichtig ist, wenn wir jetzt über künstliches Bewusstsein sprechen, ist der folgende Gedanke: Wenn es wirklich einmal Systeme geben sollte, die auf Bestrafung mit Einsicht reagieren können, dann würde das bedeuten, dass die dann auch ein Ich-Gefühl und eine Erste-Person-Perspektive haben. Sie müssten ja die Bestrafung auf *sich selbst* beziehen können. Das würde aber auch heißen, dass sie dann leidensfähig sind. Dann ist natürlich die Frage, ob man so etwas überhaupt entwickeln will. Und spätestens dann müsste man sich auch darüber Gedanken machen, ob man solchen Wesen dann auch Rechte einräumen müsste. Das heißt, man braucht dann tatsächlich auch eine Ethik *für* Roboter. - Das ist aber noch in weiterer Ferne.

**Koch** Aber ist das für Sie und Ihre Studenten und Studentinnen, wenn sie sich intensiv damit beschäftigen - es fängt ja im Grunde damit an, dass wir den Maschinen eine Identität geben, wenn wir dem Navi sagen „Wo willst du hin?“ oder unserem Computer sagen, wenn wir über ihn reden, „Was macht er denn wieder da?“ oder „Er hat mich nicht reingelassen in dieses System.“ Ist das so bei Ihnen schon auf einer anderen Ebene, wenn es um die Kommunikation mit künstlicher Intelligenz geht?

**Peters** Was Sie jetzt angesprochen haben, ist tatsächlich schon eine Ebene weiter, als das, was wir bei uns mit den Studierenden machen. Ich bin von Haus aus Mathematikerin / Informatikerin und keine Philosophin. Ich interessiere mich nur persönlich selbst zunehmend für diese philosophischen und ethischen Fragen. Konkret mit den Studierenden entwickeln wir solche Systeme, die auf dieser technischen Ebene belohnt und bestraft werden. Was Sie angesprochen haben - dass diese Systeme tatsächlich etwas empfinden - ist wirklich noch in weiterer Ferne.

**Koch** Gibt es denn schon Situationen, wo Sie spüren, dass wir diese Diskussion als Gesellschaft vielleicht bald führen müssten?

**Peters** Unbedingt! Was ich gerade angesprochen habe - Ethik für Roboter - ist mir tatsächlich nicht so wichtig. Anderes ist dringender: dass wir uns nämlich zuerst mit ethischen Fragen befassen, die die *Menschheit* betreffen, und die sich aus den Problemen ergeben, die die *aktuell existierenden* autonomen Systeme schon produzieren. Ich denke da zum Beispiel an autonome Fahrerassistenzsysteme. Google testet gerade Fahrzeuge, die komplett autonom navigieren - ohne Fahrer. Und da kann man sich vorstellen, dass diese Systeme, wenn sie in der Realität eingesetzt werden, früher oder später in schwierige Verkehrssituationen kommen, wo dann das System entscheiden muss, was es tut. Also wenn es jetzt ausweichen muss: fährt es nach links, wo es vielleicht einen Mopedfahrer ohne Sturzhelm treffen würde, oder fährt es nach rechts, wo es vielleicht ein Radfahrer mit Helm treffen würde. Diese Entscheidungen treffen dann die Systeme. Und ein anderes Beispiel sind Kampfroboter, die autonom entscheiden, ob das, was sie da vor sich erkennen, ein feindliches Ziel ist, das sie angreifen sollen, oder nicht.

**Koch** Also eine Abschätzung vornehmen.

**Peters** Eine Abschätzung. Und es ist möglich, dass dann auch Ziele angegriffen werden, deren Angriff niemand befohlen hat, und was auch gar nicht beabsichtigt war. Was diesen Beispielen gemeinsam ist, ist, dass quasi ein Outsourcing von Entscheidungen und Verantwortung an künstliche Systeme stattfindet. Und es ist nicht klar, wer dann die Verantwortung trägt, wenn Schaden angerichtet wird. Das ist unklar.

**Koch** Geht das dann zurück auf einen Programmierer / eine Programmiererin, wo man sagt: Du hast das ja so angelegt, dass er jetzt doch nicht in die Kindergruppe gefahren ist, sondern in die Seniorengruppe, weil das aus gesellschaftlichen Gründen vielleicht attraktiver ist?

**Peters** Das ist genau das Problem, das nicht geklärt ist. Und die Gesellschaft muss sich mit diesen Fragen auseinandersetzen. Meiner Meinung nach liegt die Verantwortung der Informatik oder allgemeiner gesprochen der Wissenschaft darin, zunächst erst einmal Aufklärung zu betreiben: also was sie in Ihren Labors entwickelt und wie das technisch funktioniert, damit die Leute es verstehen - und auch auf Gefahren und Risiken aufmerksam zu machen. Ich finde, da hat die Wissenschaft eine Bringschuld gegenüber der Öffentlichkeit – zumal wenn sie öffentlich finanziert wird. Aber die Gesellschaft als Ganze sollte diese Fragen dann diskutieren. Sie sollte diskutieren, ob sie so etwas haben will, und welches Risiko sie bereit ist, dafür einzugehen. Die Gesellschaft als Ganze trägt meiner Meinung nach da auch Verantwortung. Zum Beispiel schon durch ihr Konsumverhalten werden ja bestimmte neue Technologien entwickelt.

**Koch** Ich habe den Eindruck, dass es eher in eine andere Richtung geht. Dass die Gesellschaft oder die Einzelnen (*die Gesellschaft* ist ja immer so bequem, wenn man das sagt): letztendlich nutzt man diese Systeme auch ein bisschen, um Verantwortung abzugeben. Dass man nachher genau eben sagt: „Was hast du denn da wieder gemacht?“, „Was hast du denn da ausgedruckt, das wollte ich doch gar nicht?“, „Was hast du denn da wieder runtergeladen?“.

**Peters** Genau. Das ist natürlich eine bequeme Möglichkeit sich aus der Verantwortung zu stehlen.

**Koch** Und wie kriegen Sie das mit, was da für eine Brisanz drin liegt? Und wo vielleicht auch die Stellen sind, an denen Sie merken: nee, die Leute wollen sich eigentlich nicht darauf einlassen, diese Diskussion zu führen?

**Peters** Also, ich habe persönlich noch nicht festgestellt, dass die Leute sich nicht darauf einlassen wollen, diese Diskussion zu führen. Ich habe eher das Gefühl, dass diese Themen zu wenig diskutiert werden. Ich muss sagen, die Wissenschaft an sich hat auch eine Verantwortung, die Leute erst einmal aufzuklären. Und die Wissenschaft sollte sich vielleicht auch ein wenig intensiver an dieser ethischen Debatte beteiligen. Wobei man natürlich auch sagen muss, was dabei nicht passieren darf: die Wissenschaftsfreiheit sollte dabei so wenig wie möglich eingeschränkt werden, denn sie ist auch ein hohes Gut für die Allgemeinheit. Die Wissenschaftsfreiheit garantiert ja die freie Entwicklung der Gesellschaft unabhängig von ideologischen und kommerziellen Einflussnahmen.

**Koch** Das haben Sie uns wunderbar zum Schluss erklärt, welchen Konflikt wir da haben, dem wir uns in irgendeiner Art und Weise stellen müssen, und dass das nicht einfach und durchaus ein widersprüchliches Thema ist. Dankeschön an Gabriele Peters, Professorin für Mensch-Computer-Interaktion an der FernUniversität in Hagen. Sie war zu Gast bei WDR 5 „Neugier genügt“.

**Peters** Danke für die Einladung.