Darf ich meinen Hausroboter foltern?

Essay von Axel Jantsch

Die Frage scheint weit hergeholt und nicht aktuell zu sein, wird aber bald relevant werden, falls die gegenwärtigen technologischen Trends anhalten. Zunächst gilt es einmal, verschiedene Aspekte auseinander zu halten, denn zwei Voraussetzungen müssen erfüllt sein, damit die Frage Sinn ergibt: Zum einen muss ich mit meiner Folterungsaktion einen gewünschten Effekt bewirken. Nehmen wir an, dass ich eine dringend benötigte Information erhalten will, an die ich sonst nicht so ohne weiteres gelange. Der Hausroboter müsste also mir wichtige Information in seinem Inneren verborgen halten, an die ich nicht durch einfaches Fragen oder ein Durchklicken durch eine Hierarchie von Menüpunkten komme. Zum anderen impliziert die Frage, dass es moralisch oder legal fragwürdig wäre, meinen Hausroboter einer Teilzerstörung auszusetzen.

Wenden wir uns zunächst der ersten Frage zu. Auch heute schon verdächtige ich mein Smartphone, dass es Information besitzt, die es mir vorenthält, an die ich aber liebend gerne kommen würde. Zum Beispiel wüsste ich gerne, welche meiner Daten es wem weitergibt. Wer genau wird über mein Einkaufverhalten, mein Bewegungsprofil, meine Herzrate informiert? Mit einfachem Fragen komme ich da nicht weiter, da mein Smartphone mir diese Information wohl mit Absicht verheimlicht. Allerdings ist es sehr unwahrscheinlich, dass es sich durch Quetschen des Bildschirmes, durch Anbohren der CPU oder Durchnageln des Netzteiles dazu überreden ließe, die gewünschte Information preiszugeben. Der Grund ist einfach der, dass es völlig schmerzunempfindlich ist und sich seine Algorithmen durch physische Beschädigung seiner Teile nicht beeinflussen lassen. Daher ist meine Folterstrategie gegenüber dem Smartphone von heute völlig wertlos und ihre Anwendung würde nur zur Zerstörung des teuren Gerätes führen, wobei der Hersteller oder eine Versicherung mir den Schaden vermutlich nicht ersetzt.

Allerdings gibt es seit einigen Jahren Bestrebungen, elektronische, eingebettete Systeme sensibler für ihren eigenen Zustand zu machen. Dazu werden Sensoren eingebaut, die den Stromverbrauch messen, den Ladezustand der Batterie, die das Auftreten von Hardwarefehlern bemerken und analysieren können. Die Idee dahinter ist, dass komplexe, elektronische Systeme auch mit schadhaften Komponenten noch sinnvoll agieren können, um entweder ihre Funktion mit verminderter Leistung nach wie vor zu erfüllen, ein kontrolliertes Abschalten zu ermöglichen, eine Reparatur geordnet einzuleiten oder eine Fehlerdiagnostik durchzuführen. Je autonomer und selbstständiger diese Systeme agieren sollen, desto wichtiger wird es, dass sie sowohl auftretende, eigene Fehler rasch bemerken und korrekt einschätzen können, als auch sensibel darauf und den Umständen angemessen reagieren. Man denke nur an voll- oder teil-autonom fahrende Autos, Züge, Busse,

Schiffe, Flugzeuge, Drohnen, usw. In diesen Geräten sollen natürlich Fehler der Bremsfunktion, des Airbags, der Steuerung und der Kommunikation erkannt werden, sobald diese auftreten – und zwar unabhängig davon, ob diese Fehler in der Mechanik, der elektronischen Hardware oder der Software liegen. Das heißt, der Trend geht dahin, dass elektronische Systeme möglichst vollständig über ihren eigen Zustand Bescheid wissen, insbesondere über den physischen Zustand ihrer Komponenten. Darüber hinaus – und das ist für unsere Frage wichtig – beeinflusst dieses Wissen die Entscheidungen, die diese Systeme treffen. Insbesondere sollten diese Systeme versuchen, größeren Schaden zu verhindern. Konsequenterweise werden sie vermehrt mit einer Reihe von zu verfolgenden Zielen ausgestattet, die allerdings nach Priorität geordnet sind. Oberstes Ziel ist es (hoffentlich), Schaden von Menschen abzuwenden. Gleich danach kommt, Schaden an sich selbst zu minimieren. Erst danach liegt das Ziel, die eigentliche Funktion effizient durchzuführen, also beispielsweise das Navigieren von Hütteldorf nach Schwechat oder das Abspielen des Films *Blade Runner*.

In der Forschung werden Systeme, die von sich selbst und ihrem Zustand ein realistisches Bild unterhalten, als „Self-Aware Systems“ bezeichnet. Das Gebiet hat mittlerweile bereits eine 10 bis 15 Jahre lange Geschichte und es gibt eine Reihe von Büchern und regelmäßigen Workshops, Symposia und Konferenzen zu dem Thema. Wenn sich dieser Trend fortsetzt, und es spricht nichts dagegen, gibt es in einigen Jahren Geräte, darunter vielleicht auch mein Hausroboter, der nicht nur die Ursachen von Fehlern und Schäden an sich selbst erkennt, sondern auch bereit ist, alles in seiner Macht Stehende zu tun, um weiteren Schaden zu verhindern.

Na gut, warum sollte ich dann in diesem Fall also nicht meinen Hausroboter in den Keller locken, der vorsorglich von allen Wifi-, 3G-, 4G-, 5G- und 6G-Kommunikationsnetzen abgeschottet ist, ihn an einen stabilen Haken ketten und meine Heimwerkerkiste mit Hammer, Nägeln, Bohrer und Säge holen, um zu sehen, welche Informationen aus dem sich seiner Schäden selbst bewussten Roboter herausgeholt werden können? Warum sollte das ein moralisches Problem sein? Nun, der Mechanismus zur Einschätzung des eigenen Zustandes wird vermutlich über recht abstrakte Metriken funktionieren, die die Aufgabe haben, verschiedene, nicht voneinander abhängige Faktoren zu integrieren, um dann Entscheidungen zu treffen, die dem Gesamtbild gerecht werden. Nennen wir zur Illustration eine solche Metrik *Pain*, ohne damit irgendwelche Assoziationen wecken zu wollen. Diese Metrik, unter anderen, dient dazu, die aktuellen Ziele zu setzen und die immer begrenzten Mittel den jeweils wichtigsten Zielen zuzuordnen. Wann immer ein Hardwareschaden entdeckt wird, wird der Painwert erhöht. Dies hat zur Folge, dass versucht wird, die Ursache des Schadens zu eruieren um dann Strategien zu entwickeln, wie verhindert werden kann, dass der Schaden größer wird. Je größer der Schaden ist und je zentraler die betroffenen Teile für das Funktionieren des Hausroboters sind, desto höher wird der Painwert. Und je höher der Painwert, desto wichtiger werden die Ziele, die der

Schadensbegrenzung dienen und desto mehr Ressourcen wie Energie, Rechenzeit und Speicherkapazität, werden diesen Zielen zur Verfügung gestellt. Dies könnte nun im Falle meiner Folterversuche im abgeschotteten Keller dazu führen, das der Hausroboter alle ihm zur Verfügung stehenden Geräusche von sich gibt, um mich und andere auf seine Schäden hinzuweisen, und unter Aufbietung aller seiner Kräfte versucht, der drohenden Bohrmaschine zu entkommen. Wenn ich es geschickt genug anfange, versteht der Roboter dass es einen einfachen Weg gibt, weitere Schäden von ihm abzuhalten – nämlich mir die gesuchte Information zur Verfügung zu stellen. Dann hängt es nur davon ab, wie die Zielhierarchie in der Robotersoftware organisiert ist, ob es letztens wichtiger ist, das eigene Überleben zu sichern oder ob es wichtiger ist, die Daten und Anweisungen seines Herstellers zu schützen. Solange ich nicht genau weiß, welche Ziele beim Hausroboter höhere Priorität haben, ist meine beste Möglichkeit, dies herauszufinden und an die Daten zu kommen, den Painwert langsam, aber stetig und gezielt zu erhöhen.

Sicher, könnte an dieser Stelle vielleicht eingeworfen werden, dies ist ein interessanter Ansatz, um meine vernachlässigten Konsumentenrechte zu schützen, aber warum sollte es sich dabei um ein moralisches Problem handeln? Es handelt sich doch vielmehr um eine ökonomische Abschätzung meiner eigenen Prioritäten. Ist mir die Erlangung der Daten den Totalverlust des teuren Stückes wert? Dies ist es in der Tat, aber zusätzlich stellt sich die Frage, wie der drastisch erhöhte Painwert eingeschätzt werden soll. Die beobachteten Symptome wecken deutliche Erinnerung an andere Wesen, Tiere und Menschen, denen ich Leiden und Schmerzerleben zubillige. Der Hausroboter gibt alle möglichen, ungewohnten Töne in größter Lautstärke von sich und versucht in offensichtlicher Verzweiflung und unter Aufbietung aller Kräfte seiner erbärmlichen Lage zu entkommen und meinen Folterwerkzeugen auszuweichen. Doch wenn man sich von diesem Eindruck nicht beeindrucken lässt, könnte man betonen, dass der zugrundeliegende Mechanismus einfach ein Programm ist und der Painwert nicht viel mehr als eine Variable in einer Speicherzelle des Roboters darstellt. Offensichtlich können echtes Schmerzempfinden und Leiden daraus nicht entstehen und mein Mitleid ist völlig fehl am Platz. Dagegen ist allerdings einzuwenden, dass bei Tieren und Menschen der zugrundeliegende Mechanismus nicht wesentlich anders ist. Zwar ist der Algorithmus nicht in Silizium, sondern in einem neuronalen Netz realisiert und der Schmerz wird in Form von Verbindungen zwischen Neuronen kodiert, aber der Effekt ist praktisch identisch, nämlich der extreme Fokus des Systems auf die Ursachen des Schmerzes und dem Einsatz aller verfügbaren Mittel zur Verhinderung größerer Schäden.

Das heißt: Ob es sich hier um eine ethische Frage handelt oder nicht, also ob mein Hausroboter zu Leiden und Schmerzempfinden fähig ist oder nicht, ist ähnlich gelagert wie bei Tieren. Darf ich Mäuse quälen? Frösche? Fliegen? Amöben? Bakterien? Wo genau liegt die Grenze und was ist das zugrunde liegende Kriterium? Ist es die Analogie mit meinen eigenen, subjektiven Erfahrungen? Hängt die Beantwortung der Frage von der

Vollständigkeit meines Wissens über die Mechanismen der Schmerzentstehung und die Reaktionen darauf ab? Kann es sein, dass ich einem Wesen Leidensfähigkeit abspreche, nur weil ich genau verstehe, wie die Mechanismen dahinter funktionieren, zum Beispiel weil ich das Programm dazu selbst geschrieben habe? Heißt das auch, dass wir die Leidensfähigkeit auch Tieren wie Hunden und Affen und letztlich dem Menschen selbst absprechen, wenn wir einmal genau geklärt haben, wie Schmerz funktioniert?

Selbst wenn ich manchmal meine Irritation und Zorn gegenüber Smartphone und autonomen Staubsauger nicht ganz unterdrücken kann und deren Aktionen und Reaktionen hin und wieder sadistische Impulse wecken, stellt sich die Frage, ob ich sie quälen kann und darf, noch nicht in voller Schärfe. Aber dies wird sich ändern. Ob es 5, 10, 20 oder 50 Jahre dauern wird, wage ich nicht zu spekulieren, aber dass sich die aufgeworfene Frage als Konsequenz des Fortschreitens gegenwärtiger Technologietrends stellen wird, ist gewiss.

Zur Person:
Axel Jantsch ist Professor der Technischen Universität (TU) Wien. Er hat derzeit den Lehrstuhl für „Systems on Chips“ am Institut für Computertechnik, davor war er unter anderem Professor für Electronic Systems Design an der Königlich Technischen Hochschule (KTH) in Stockholm.