

Roboter Vertrauen

Eine Exploration vertrauensförderlicher Attribute der Robotergestaltung mittels Clusteranalyse

Daniel Ullrich
Institut für Informatik
Ludwig-Maximilians-Universität
München
daniel.ullrich@ifi.lmu.de

Sarah Diefenbach
Department Psychologie
Ludwig-Maximilians-Universität
München
sarah.diefenbach@psy.lmu.de

Lara Christoforakos
Department Psychologie
Ludwig-Maximilians-Universität
München
lara.christoforakos@psy.lmu.de

Tinatini Surmava
Department Psychologie
Ludwig-Maximilians-Universität
München
tinatini.surmava@gmail.com

Cedrik Quintes
Institut für Informatik
Ludwig-Maximilians-Universität
München
cdrcqnts@gmail.com

ABSTRACT

Menschenähnlichkeit gilt in der Gestaltung von Robotern und anderen interaktiven Technologien als ein beliebtes Mittel um Vertrauen zu stärken. Wenngleich auch empirische Studien einen positiven Zusammenhang zwischen anthropomorpher Gestaltung und Vertrauen nahelegen, ist unklar, ob es tatsächlich die Menschenähnlichkeit per se oder weitere, häufig mit Menschenähnlichkeit gemeinsam auftretende Attribute sind, die das Vertrauen seitens der Nutzer fördern. Anhand einer Clusteranalyse von rund 100 Robotern entlang zweier Komponenten von Vertrauen (Kompetenz, Wohlwollen) haben wir vertrauensbildende Attribute der Robotergestaltung näher befocht. Weiterführende Forschungsfragen und Designimplikationen werden diskutiert.

KEYWORDS

Roboter, Vertrauen, Kompetenz, Wohlwollen, Menschenähnlichkeit, Anthropomorphisierung, Design Pattern, Cluster Analyse

ACM Reference format:

Daniel Ullrich, Sarah Diefenbach, Lara Christoforakos, Tinatini Surmava und Cedrik Quintes. 2019. Roboter Vertrauen. In *Mensch und Computer 2019 – Workshopband*, Bonn: Gesellschaft für Informatik e.V., <https://doi.org/10.18420/muc2019-ws-306-01>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
MuC'19 Workshops, Hamburg, Deutschland
© Proceedings of the Mensch und Computer 2019 Workshop on Von der Nutzung zur Zusammenarbeit – User Experience in Zeiten der Künstlichen Intelligenz.
Copyright held by the owner/author(s).
<https://doi.org/10.18420/muc2019-ws-306-01>

1 Roboter Vertrauen und Menschenähnlichkeit

Siri, Alexa, Chatbots, der freundlich lächelnde Staubsaugroboter oder realen Menschen nachempfundene soziale Roboter wie "Sophia" (<https://www.hansonrobotics.com/sophia/>): viele Technologien, die uns im Alltag begleiten weisen in ihrem Aussehen oder ihrer Dialogfähigkeit menschenähnliche Eigenschaften auf. Dies gilt insbesondere für die Domäne von Robotern, in der die Orientierung am Menschen seit jeher einen zentralen Gestaltungsansatz darstellt.

Die Gestaltung von Technik mit menschenähnlichen Attributen scheint auch dahingehend vielversprechend, dass zahlreiche Studien in der Konsumentenpsychologie und Mensch-Technik-Interaktion eine Korrelation zwischen der wahrgenommenen Menschenähnlichkeit von technischen und anderen unbelebten Objekten und wünschenswerten Eigenschaften nahelegen. Ein bekanntes Beispiel für Anthropomorphisierung, bzw. die Tendenz nichtmenschliche Entitäten mit menschenähnlichen Charaktereigenschaften, Motivationen, Intentionen oder Emotionen auszustatten [3], ist die Wahrnehmung des "Gesichts" eines Autos, wie es beispielsweise die in der Studie von Aggarwal und McGill [1] genutzten Bilder gut demonstrieren (siehe Abb. 1). Hierbei zeigte sich, dass der Grad, zu dem Menschen das Auto als Person wahrnehmen, sich positiv auf die Bewertung des Autos auswirkt. Auch auf die Produktbindung scheint die wahrgenommene Menschenähnlichkeit positive Effekte zu haben: Studienteilnehmer, bei denen durch eine Instruktion anthropomorphes Denken über ihr Auto induziert wurde, waren weniger bereit dieses zu ersetzen und beurteilen dessen Qualität eher anhand von für die Beurteilung für Menschen relevanten Eigenschaften und Persönlichkeitsmerkmalen [2]. Ein weiteres häufig untersuchtes und positiv besetztes Urteil im Zusammenhang mit Menschenähnlichkeit ist die Variable Vertrauen: so zeigte sich beispielsweise die wahrgenommene Menschenähnlichkeit eines

autonomen Fahrzeugs als relevanter Prädiktor für das Vertrauen in das Fahrzeug [5]. Gerade im Zeitalter zunehmend fortgeschrittener Technik, die immer komplexere und sensiblere Aufgaben übernimmt, gleichzeitig aber auch für den Menschen immer schwerer zu begreifen und zu durchschauen ist, wird die Frage des Vertrauens zentral. Nur wenn Nutzer der Technik vertrauen, werden sie diese auch nutzen und von deren Möglichkeiten profitieren können. Allerdings ist die Frage, ob um Vertrauen zu stiften, die Menschenähnlichkeit tatsächlich das alleinige Wundermittel ist.

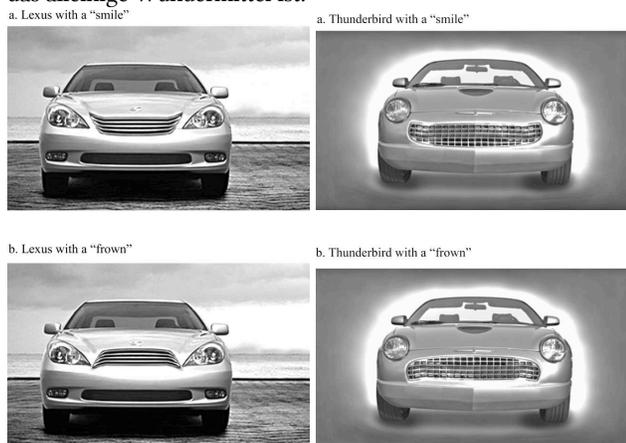


Abb. 1: Lächelnde (oben) versus finster dreinblickende (unten) Autos. Originalabbildung aus Aggarwal & McGill [1].

Diese Sichtweise wollen wir kritisch hinterfragen. Wenngleich es viele Befunde gibt, die einen positiven Zusammenhang zwischen Menschenähnlichkeit und Vertrauen nahelegen, bleibt die Frage, ob Menschenähnlichkeit per se die für Vertrauen zentrale relevante Eigenschaft ist – oder ob es sich hierbei möglicherweise um eine Scheinkorrelation handelt, da eine anthropomorphe Gestaltung oft auch mit anderen, positiv besetzten vertrauenserweckenden Attributen einhergeht. Bei genauerer Betrachtung macht es auch unsere Alltagserfahrung fragwürdig, dass Menschenähnlichkeit per se vertrauenswürdig macht – auch reale Menschen, genießen nicht per se unser Vertrauen, im Gegenteil gibt es auch viele Menschen, denen wir im ersten Eindruck negative Eigenschaften zuschreiben.

Unser Ziel war es daher, die Vertrauen in Technik erzeugenden Faktoren näher zu explorieren und hierdurch auch eine Grundlage für die Gestaltung zu liefern.

Unsere Analyse konzentrierte sich dabei auf die Domäne von Robotern, als eine Domäne, in der Menschenähnlichkeit als Gestaltungsansatz besonders verbreitet und Vertrauen seitens des Nutzers besonders relevant ist.

2 Methode

Um einen näheren Einblick in die für die subjektive Vertrauenswürdigkeit relevanten Attribute zu erlangen, führten wir eine Clusteranalyse durch. Bei einer Clusteranalyse werden

verschiedene Objekte hinsichtlich ihrer Ähnlichkeit angeordnet und das sich hieraus ergebende Muster wird dann genutzt, um Einsichten in zugrundeliegende Muster, mögliche Gesetzmäßigkeiten und ggf. Subdimensionen zu ermitteln.

Ausgangspunkt der Clusteranalyse war eine Sammlung von rund 100 Robotern und Agenten verschiedener Art, z.B. Chatbots, Roboterarme, Roboter für militärische Zwecke oder auch in ihrer Gestaltung stark am Menschen orientierte Roboter. Zusätzlich enthielt die Bildersammlung auch Bilder von realen Menschen – denn Einsichten dazu, was Menschen vertrauenswürdig macht, könnten auch relevant für die Vertrauenswürdigkeit von Robotern sein, gerade in Hinblick auf menschenähnliche Roboter. In einem Workshop mit fünf Teilnehmern mit Hintergrund in der Psychologie und Informatik wurden diese nach ihrer subjektiven Vertrauenswürdigkeit angeordnet. In Orientierung an psychologischen Modellen des Vertrauens unterschieden wir hierbei verschiedene Komponenten welche für den Aufbau von Vertrauen gegenüber Mitmenschen relevant sind. Insbesondere orientierten wir uns an einer Längsschnittstudie zum Aufbau von Vertrauen unter Arbeitskollegen [4], in der competence und benevolence als wichtigste Komponenten für den Aufbau von Vertrauen identifiziert wurden. Übertragen auf unsere Fragestellung im Feld Mensch-Roboter-Interaktion unterschieden wir dementsprechend zwei Komponenten von Vertrauen:

- Kompetenz: Wie sehr vertraue ich, dass der Roboter seine Aufgabe gut erledigen kann und wird?
- Wohlwollen: Wie sehr vertraue ich, dass der Roboter mir etwas Gutes will bzw. mich in meinen Zielen unterstützt?

Der durch die beiden Komponenten aufgespannte Raum zur Positionierung der Bilder wurde auf einem Whiteboard skizziert (siehe Abb. 2). Die Teilnehmer positionierten die Bilder ihrer spontanen Einschätzung nach entlang der zwei Komponenten. Anschließend wurde das sich ergebende Bild mit allen Teilnehmern diskutiert. Es wurden mögliche Gründe für die Positionierung reflektiert und ggf. geringfügige Neupositionierungen einzelner Bilder vorgenommen, wenn diese nach der Diskussion angemessen schienen.

3 Erste Einsichten

Abbildung 2 zeigt das Ergebnis der Clusteranalyse und die im Workshop vorgenommene Einordnung von Robotern (und einigen Menschen) hinsichtlich ihrer subjektiven Vertrauenswürdigkeit entlang der Vertrauenskomponenten Kompetenz und Wohlwollen. Abbildung 3 zeigt beispielhaft die Einordnung von vier Robotern mit extremer Ausprägung hinsichtlich der zwei Vertrauenskomponenten.

Die Positionierung der Roboter bietet bereits einige interessante Einsichten, welche wiederum Ausgangspunkte für Designimplikationen sowie weiterführende Forschungsfragen liefern:



Abb. 2: Clusteranalyse. Einordnung von Robotern (und einigen Menschen) hinsichtlich ihrer subjektiven Vertrauenswürdigkeit entlang der Vertrauenskomponenten Kompetenz und Wohlwollen.

- Die Unterscheidung der Komponenten Kompetenz und Wohlwollen ist generell zielführend: viele Roboter zeigen eine hohe Ausprägung hinsichtlich der einen Komponente und eine niedrige Ausprägung hinsichtlich der anderen Komponente
- Roboter mit ähnlicher Positionierung hinsichtlich ihrer Vertrauenswürdigkeit ähneln sich oft auch hinsichtlich weiterer Merkmale, z.B. bezüglich Form oder Funktionalitäten
- Roboter mit einer hohen Ausprägung bezüglich Wohlwollen zeigen oft Merkmale wie große Augen, überproportional großer Kopf (Kindchenschema)
- Roboter mit einer hohen Ausprägung bezüglich Kompetenz haben oft ein stark maschinenhaftes Aussehen, das durch einen hohen Grad an Feinmotorik und Details gekennzeichnet ist. Insgesamt nimmt die Komplexität entlang der Kompetenzachse zu
- Menschenähnlichkeit allein ist nicht vertrauensbildend: stark menschenähnliche Roboter verteilen sich über die gesamte Matrix und finden sich im Feld von Robotern mit hoher und geringer Vertrauenswürdigkeit hinsichtlich beider Komponenten
- Einige Roboter wirken deshalb als vertrauenswürdig, weil sie ein bestimmtes Schema aktivieren (z.B. Ärzte = kompetent und wohlwollend)
- Es gibt insgesamt nur wenige Roboter mit hoher Ausprägung auf beiden Komponenten: Vertrauenswürdigkeit hinsichtlich beider Komponenten gleichzeitig zu erzeugen, scheint eine Herausforderung
- Ein Grund für diese Art von Vertrauensparadox könnten auch psychologische Schlussfolgerungsprozesse sein: eventuell besteht die implizite Annahme eines negativen Zusammenhangs zwischen den beiden Komponenten: Wohlwollen könnte auch Assoziationen von Harmlosigkeit/Naivität erzeugen, was wiederum mit eher geringer Kompetenz einhergeht



Abb. 3: Beispielhafte Einordnung von vier Robotern mit extremer Ausprägung hinsichtlich der zwei Vertrauenskomponenten.

4 Design Implikationen und weiterführende Forschungsfragen

Eine erste relevante Einsicht hinsichtlich der Gestaltung von Robotern mit dem Ziel der Erhöhung ihrer Vertrauenswürdigkeit ist bereits die Sinnhaftigkeit der Unterscheidung der beiden Vertrauenskomponenten Kompetenz und Wohllollen sowie der mögliche Trade-off zwischen diesen beiden Komponenten. Designer könnten sich dementsprechend fragen:

- Welche der beiden Vertrauenskomponenten ist in meinem Kontext besonders wichtig?
- Oder andersherum formuliert: Bezüglich welcher Komponente könnte tendenziell Misstrauen bestehen, dem durch entsprechende Gestaltungsentscheidungen entgegengewirkt werden sollte?
- Wie lässt sich die relevante Vertrauenskomponente in der Gestaltung adressieren? Welche Designmerkmale könnten das Vertrauen hinsichtlich dieser Komponente stärken?

So könnten beispielsweise beim Einsatz von Robotern im medizinischen Kontext (z.B. bei Operationen, bei Reha-Maßnahmen) Patienten vom grundsätzlichen Wohllollen des Roboters überzeugt sein, d.h. sie vertrauen, dass der Roboter bzw. die den Roboter einsetzende Institution gewillt ist, ihnen bestmöglich zu helfen. Gleichzeitig könnten Patienten jedoch an

der Kompetenz des Roboters zweifeln, d.h. sie bezweifeln beispielsweise ob der Roboter genauso gut (oder sogar besser) operiert wie ein Mensch. Um dieses Misstrauen abzubauen, sollten also Designmerkmale verwendet werden, die insbesondere die Kompetenz des Roboters zum Ausdruck bringen. Ausgehend von der Clusteranalyse und der starken Präsenz von maschinenhaften Robotern mit sichtbar hoher Komplexität am Ende der Kompetenzskala, könnte eine Empfehlung hier lauten, die Fähigkeiten und Komplexität des Roboters transparent zu machen und dessen Fähigkeiten in überzeugender Weise zu demonstrieren. Ein weiterer Kompetenz-Cue scheint auch ein hoher Grad an visueller Verfeinerung und eine gewisse Eleganz der visuellen Gestaltung, welcher wahrscheinlich Schlussfolgerungen auch hinsichtlich einer hohen Verfeinerung bezüglich technischen Reifegrads und damit der Kompetenz des Roboters nahelegt.

In anderen Kontexten könnte weniger die Stärkung der Kompetenz- sondern die Stärkung der Wohllollens-Komponente im Vordergrund stehen. Wird beispielsweise ein Chatbot eingesetzt um potentielle Krankenversicherungskunden vor dem Abschluss einer Versicherung zu ihrer medizinischen Vorgeschichte zu befragen, so werden Nutzer möglicherweise weniger die Kompetenz des Roboters (zuverlässige Verarbeitung und Speicherung von Texteingaben des Nutzers) in Zweifel ziehen, könnten aber am Wohllollen des Roboters zweifeln (Nutzung der Texteingabe im Sinne der Nutzer). Hier wäre es beispielsweise wichtig, anhand der Gestaltung und

Interaktionselemente des Roboters zu betonen, dass dieser dem Nutzer freundlich gesonnen ist, ihn nicht hereinlegen will und nachvollziehbar macht warum welche Daten seitens des Nutzers benötigt werden und dass diese sensibel behandelt werden.

Diese ersten Ideen sind jedoch nur beispielhaft zu verstehen. Limitiert sind die Einsichten aus der vorliegenden Studie auch durch die spezifische Stichprobe des verwendeten Bildmaterials. In den nächsten Schritten unseres Forschungsprojekts werden wir vertrauensförderliche Design-Cues hinsichtlich der beiden Vertrauenskomponenten noch systematischer erforschen und prüfen, sodass sich möglicherweise auch generelle Design-Pattern für die Vertrauenskomponenten abbilden lassen.

Weitere Gestaltungsimplicationen und vertrauensförderlichen Design-Cues könnten sich aus einem besseren Verständnis der bei der Entstehung von Vertrauen in Roboter relevanten Schlussfolgerungsprozesse ergeben. Ein erster Ansatzpunkt wäre die oben formulierte mögliche implizite Annahme eines negativen Zusammenhangs zwischen Wohlwollen (im Sinne von Naivität) und Kompetenz. Sollte sich dieser Zusammenhang in zukünftigen Studien bestätigen, wäre eine zentrale weiterführende Frage, wie sich die Anmutung von Wohlwollen in der Gestaltung anlegen lässt, ohne aber die (unerwünschte, mit Kompetenz negativ korrelierte) Anmutung von Naivität zu erzeugen.

ACKNOWLEDGMENTS

Teile der vorliegenden Forschungsarbeiten wurden gefördert durch das BMBF-Projekt GINA (FKZ: 16SV8097).

REFERENCES

- [1] Pankaj Aggarwal and Ann L. McGill. 2007. Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. *Journal of Consumer Research*, 34(4), 468-479.
- [2] Jesse Chandler and Norbert Schwarz. 2010. Use does not wear ragged the fabric of friendship: Thinking of objects as alive makes people less willing to replace them. *Journal of Consumer Psychology*, 20(2), 138-145.
- [3] Nicolas Epley, Adam Waytz and John T. Cacioppo. 2007. On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864-886. doi:10.1037/0033-295X.114.4.864.
- [4] Lisa van der Werff and Finian Buckley. 2017. Getting to know you: A longitudinal examination of trust cues and trust development during socialization. *Journal of Management*, 43(3), 742-770.
- [5] Adam Waytz, Joy Heafner and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113-117.