Ein ABC aktueller Herausforderungen für sichere interaktive Systeme

Tilo Mentler

Institut für Multimediale und Interaktive Systeme (IMIS), Universität zu Lübeck

mentler@imis.uni-luebeck.de

Zusammenfassung

Die Potenziale und Grenzen maschineller Lernverfahren (unter dem Schlagwort „Artificial Intelligence“), großer verfügbarer Datenmengen („Big Data“) sowie vernetzter softwaretechnischer und mechanischer Komponenten („Cyber-Physical Systems“) werden derzeit hinsichtlich verschiedener sicherheitskritischer Domänen diskutiert. Unabhängig von konkreten Anwendungen (z.B. Umgang mit Fake News, „smarte“ Energieverteilnetze oder E-Health) gilt es, Herausforderungen für die Gestaltung entsprechender Computersysteme systematisch zu erfassen. Sie müssen dann insbesondere hinsichtlich der Mensch-Maschine-Schnittstelle bewertet werden. In diesem Beitrag wird auf Grundlage der Forschungsarbeiten im Projekt „Artificial Intelligence and the Automated Ordering of Digital Communication“ das zuvor benannte ABC aktueller Herausforderungen für die Gestaltung sicherer interaktiver Systeme diskutiert und Forschungsbedarf im Bereich Mensch-Technik-Interaktion identifiziert.

1 Einleitung

Im Rahmen der Erforschung und Entwicklung von sicherheitskritischen Systemen, d.h. Systemen mit *„hohem Risikopotenzial“* (Herczeg, 2014), müssen die Potenziale und Grenzen technischer Innovationen sorgfältig und kritisch bewertet werden. Es gilt *„Hype-Zyklen“* (vgl. Kreutzer, 2015) zu begegnen, d.h. Entwicklungen und Auswirkungen kurzfristig nüchtern und aufgeregt zu begleiten, ohne jedoch Dynamiken sowie mittel- bis langfristige Potenziale zu unterschätzen. Amara formulierte hierzu prägnant: *„We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run”* (Ratcliffe, 2016).

In diesem Zusammenhang wird nachfolgend auf die Forschungs- und Entwicklungsarbeiten zu Themen eingegangen, die derzeit sowohl in verschiedenen Wissenschaftsdisziplinen als auch in der breiten Öffentlichkeit intensiv diskutiert werden (siehe z.B. Abdul et al., 2018; Dreyer & Schulz, 2018; O’Neil, 2016):

Veröffentlicht durch die Gesellschaft für Informatik e. V. 2018 in
R. Dachselt, G. Weber (Hrsg.):
Mensch und Computer 2018 – Workshopband, 02.–05. September 2018, Dresden.
 https://doi.org/10.18420/muc2018-ws12-0457

- maschinelle Lernverfahren unter dem Schlagwort *Artifizielle/Künstliche Intelligenz*,
- die Verfügbarkeit und Auswertbarkeit großer Datenmengen unter dem Stichwort *Big Data*,
- die Vernetzung softwaretechnischer, mechanischer und elektronischer Komponenten unter dem Oberbegriff *Cyber-Physical Systems*.

Dabei ist neben der Betrachtung der einzelnen Konzepte und Technologien auch und gerade ihr Zusammenwirken zu berücksichtigen. So wird beispielsweise im Rahmen des Projektes „Artificial Intelligence and the Automated Ordering of Digital Communication“ von Kommunikations- und Medienwissenschaftlern, Juristen sowie Informatikern untersucht, wie Betreiber von sozialen Medien mit Mengen an nutzergenerierten Inhalten umgehen und kritische Aufgabenstellungen, z.B. Umgang mit rechtswidrigem Material, mithilfe von KI-Verfahren bewältigen können. In diesem Beitrag werden grundsätzliche Herausforderungen für die Gestaltung entsprechender sicherheitskritischer Systeme aus der Perspektive der Mensch-Computer-Interaktion systematisiert und weiterer Forschungsbedarf identifiziert.

2 Hintergrund

In diesem Abschnitt werden grundlegende und aktuelle Forschungsarbeiten zu den Themen Artificial Intelligence, Big Data und Cyber-Physical Systems, insbesondere im Zusammenhang mit sicherheitskritischen Anwendungsdomänen, zusammengefasst.

2.1 Artificial Intelligence

Eine der ersten Konzeptualisierungen des Begriffes Künstlicher Intelligenz (KI) wurde von McCarthy et al. (1955, zitiert nach McCarthy et al., 2006) Mitte der 1950er-Jahre vorgenommen: *„…how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves“*. Die Forschung zur Theorie und Praxis von KI in den darauffolgenden Jahrzehnten lässt sich nach Haun (2014) in folgende Epochen einteilen (siehe Tabelle 1).

Epoche	**Zeitraum**	**Fragestellungen und Ansätze**
Klassische Epoche	1955-1965	Lösung beliebiger Probleme, General Problem Solver (GPS), Formelmanipulation, Mustervergleiche
Romantische Epoche	1965-1975	Repräsentation von Wissen, Suchen in Bäumen, LISP/PROLOG als Programmiersprachen
Moderne Epoche	1975-1994	Problemspezifisches Wissen, wissensbasierte Systeme, Produktionsregelsysteme, Expertensysteme
Postmoderne Epoche	1995-2000	Kommerzielle Nutzung von KI-Werkzeugen und Expertensystemen, Integration logischer und funktionaler Programmierung

Epoche der Renaissance und des Cognitive Computing	seit 2000	Zuwendung zu Alltagsproblemen, Integration von Bild-, Wissens- & Sprachverarbeitung, Zusammenwirken von Philosophie, Psychologie, Informatik, Linguistik, Mathematik & Kognitionswissenschaften

Tabelle 1: Historie der KI in Epochen nach Haun (2014)

Ohne im Detail auf die mathematischen und informationstheoretischen Grundlagen eingehen zu wollen – hierzu seien die Lesenden auf Rich (1983) oder Ertel (2016) verwiesen – kann festgestellt werden, dass die praktische Relevanz von KI-Verfahren derzeit insbesondere im Zusammenhang mit Deep-Learning-Verfahren wahrgenommen wird. Deren Grundidee lässt sich wie folgt zusammenfassen: *„Deep learning enables the computer to build complex concepts out of simpler concepts"* (Goodfellow et al., 2016). Voraussetzung dafür sind jedoch umfangreiche Mengen an Trainingsdaten (vgl. Wick, 2017).

KI- und insbesondere Deep-Learning-Verfahren werden in sicherheitskritischen Domänen u.a. zur Diagnose von Krankheiten (Danaee et al., 2017; Kourou et al., 2015) oder zur Entscheidungsunterstützung in Krisensituationen (Lauras, & Comes, 2015) diskutiert und erprobt. Open-Source-Plattformen wie „AIDR – Artificial Intelligence"[1] nutzen dabei Massen an nutzergenerierten Daten aus sozialen Medien.

2.2 Big Data

Große Datenmengen, *„die in ihrer Größe klassische Datenhaltung, Verarbeitung und Analyse […] auf konventioneller Hardware übersteigen"* (Merv, 2011, zitiert nach Fasel & Meyer, 2016), werden unter dem Begriff Big Data zusammengefasst. Dabei ist zu beachten, dass keine einheitliche Definition vorliegt und beispielsweise auch die von Laney (2001) vorgeschlagenen Dimensionen als den Diskurs prägend genannt werden müssen:

- *Volumen*: Menge der Daten;
- *Velocity*: Geschwindigkeit des Datenaufkommens;
- *Variety*: Vielzahl und Heterogenität der Daten.

Darüber hinaus wurden und werden weitere "V" definiert (Furht & Villanustre, 2016; Godfrey et al., 2016; Olshannikova, 2015):

- *Veracity*: Vertrauenswürdigkeit der Daten;
- *Value*: (Unternehmerischer) Mehrwert der Daten.

Unter Vernachlässigung einer exakten Definition können Forschungs- und Entwicklungsarbeiten zur Nutzbarmachung großer Datenmengen in zahlreichen sicherheitskritischen Domänen ausgemacht werden, u.a. Krisenmanagement (Reuter et al., 2016), Verkehrswesen

[1] http://aidr.qcri.org/

(Toyoda, 2015; Zheng et al., 2016), Flugsicherheit (Becker, 2016) oder Energieverwaltung (Holzinger et al., 2015).

2.3 Cyber-Physical Systems

Cyber-Physical Systems sind *„gekennzeichnet durch eine Verknüpfung von realen (physischen) Objekten und Prozessen mit informationsverarbeitenden (virtuellen) Objekten und Prozessen über offene, teilweise globale und jederzeit miteinander verbundene Informationsnetze“* (acatech, 2012).

Die Deutsche Akademie für Technikwissenschaften bezeichnete Cyber-Physical Systems als *„Innovationsmotor für Mobilität, Gesundheit, Energie und Produktion“* (acatech, 2011) und somit für sicherheitskritische Mensch-Maschine-Systeme in vielen Bereichen.

3 Herausforderungen

Werden die drei im vorherigen Abschnitt vorgestellten Konzepte zusammengeführt, lässt sich das folgende abstrakte Szenario für computerbasierte Lösungen auf Basis von Artificial Intelligence, Big Data und Cyber-Physical Systems (im Folgenden: ABC-Systeme) formulieren:

Große Mengen an Daten, die von einer Vielzahl und den Nutzenden nur teilweise bekannten informationsverarbeitenden und physischen Komponenten produziert werden, werden mithilfe von selbstlernenden Systemen maschinell weiterverarbeitet. Entscheidungen, die nicht nur einzelne Nutzende, sondern ganze Gesellschaften betreffen, können somit innerhalb von wenigen Sekunden und datengestützt getroffen werden. Weder Nutzende noch Entwickelnde können direkten Einfluss auf das Computersystem nehmen.

Während diese kurze Beschreibung positiv im Sinne einer Utopie der Vollautomatisierung und Objektivierung nicht nur von Produktions-, sondern Entscheidungsprozessen generell gedeutet werden kann, muss aus Sicht der Mensch-Technik-Interaktion die Bewertung anders ausfallen. Dabei müssen nicht nur, aber auch und gerade, die Ironien der Automatisierung (Bainbridge, 1983) Warnung und Mahnung zugleich sein. Nachfolgend wird auf einige der zu bewältigenden Herausforderungen eingegangen.

3.1 Menschzentrierte Entwicklungsprozesse für ABC-Systeme

Menschzentrierte und partizipative Vorgehensmodelle der Systementwicklung sollen die Gebrauchstauglichkeit und Akzeptanz computerbasierter Lösungen gewährleisten. Der in DIN EN ISO 9241-210:2011 beschriebene iterative menschzentrierte Entwicklungsprozess hat sich in vielen Anwendungsbereichen bewährt. Die dem Prozess zugrundeliegenden Prinzipien haben sich jedoch seit Einführung der Vorgänger-Norm ISO 13407 nicht wesentlich geändert (vgl. DIN EN ISO 924-210:2011). Daraus lässt sich nicht per se Änderungsbedarf

ableiten. Jedoch sollten die etablierten Prozesse angesichts folgender Veränderungen kritisch beleuchtet werden:

- Die *„allgegenwärtige Mensch-Computer-Interaktion“* (Koch & Alt, 2017) verändert die Gewohnheiten und Erwartungen der Nutzenden auch in sicherheitskritischen Kontexten. Mobile und am Körper tragbare Endgeräte werden teilweise kontextübergreifend genutzt. Die umfassende Analyse von Nutzungskontexten, einer der grundlegenden Phasen menschzentrierter Entwicklung, wird dadurch schwieriger.
- Das mit Big-Data- und KI-Anwendungen verbundene *Programmierparadigma der datengetriebenen Entwicklung* (vgl. Falcini & Lami, 2017) verändert Rollen und Verantwortlichkeiten auf Seiten der Entwickelnden. Während die mathematischen Grundlagen von Deep-Learning-Verfahren ggf. noch nachvollzogen werden können, lässt sich das jeweils konkrete, von den Trainingsdaten abhängende Wissen- und Weltmodell des Anwendungssystems zunächst einmal nur sehr beschränkt nachvollziehen.
- Computerbasierte Lösungen in sicherheitskritischen Kontexten betreffen teilweise nicht mehr nur die Nutzenden oder andere mittelbar Betroffene (z.B. mögliche Opfer von Schäden), sondern *gesellschaftliche Gruppen oder ganze Gesellschaften* (z.B. verpflichtende E-Health-Lösungen oder die Gefahr von Wahlmanipulationen über Soziale Netzwerke).

Alles in allem muss daher das von Rahwan (2018) skizzierte Konzept der *Society-in-the-Loop* operationalisiert und methodisch fundiert werden. Dazu müssen u.a. rechtliche und sozialwissenschaftliche Perspektiven in noch stärkerem Maße in die Entwicklung integriert werden. Liegt der Fokus bei formativen und summativen Evaluationen derzeit i.d.R. noch auf einzelnen Nutzenden oder Arbeitsgruppen, wird es zukünftig auch darum gehen müssen, gesellschaftliche Auswirkungen von Systemkonzepten und Gestaltungslösungen zu bewerten.

3.2 Visualisierungs- und Interaktionskonzepte für ABC-Systeme

Damit Nutzende in der Lage bleiben bzw. in diese versetzen werden, Systemzustände bewerten und Entscheidungen treffen zu können, werden Visualisierungs- und Interaktionskonzepte benötigt, die einerseits große Datenmengen nutzer- und aufgabenspezifisch aufbereiten[2] und andererseits algorithmische Entscheidungen und Lernprozesse nachvollziehbar und transparent machen. Dabei gilt es einerseits etablierte Prinzipien, wie z.B. das *Visual Information Seeking Mantra* (*“Overview first, zoom and filter, then details-on-demand“*) von Shneiderman (1996), oder epochenüberdauernden Problemstellungen, wie z.B. die Erklärbarkeit von Künstlicher Intelligenz, mit neuen Konzepten zu begegnen.

[2] „We are preparing for the day soon when visualization becomes the sixth V of big data“ (Godfrey et al., 2016).

In diesem Zusammenhang ist beispielsweise das Forschungsgebiet *„Explainable AI“* (Gunning, 2016; Holzinger, 2018) zu nennen. Seine Vertreterinnen und Vertreter setzen sich u.a. das Ziel, das sich Nutzende zukünftig folgende Fragen beantworten können bzw. vom Anwendungssystem beantwortet bekommen:

- Warum hast du [das System] das gemacht und nicht etwas Anderes?
- Wie bewertest du [das System] Erfolg und Misserfolg?
- Wie sehr kann ich dir [dem System] vertrauen?
- Wie kann ich Fehler korrigieren?

All diese Fragen adressieren den Umstand, dass es auch und gerade in sicherheitskritischen Systemen in entscheidendem Maße um zwei V geht:

- Vertrauen in die Korrektheit der Vorschläge und Entscheidungen von Computern;
- Verantwortung im Sinne der *„A-priori-Zuschreibung einer Pflicht, Fehler zu vermeiden bzw. für auftretende Schäden zuständig zu sein“* (Herczeg, 2014).

4 Zusammenfassung und Ausblick

Artificial Intelligence, Big Data und Cyber-Physical Systems können in ihren Wechselwirkungen als grundlegendes ABC der Herausforderungen bei der Gestaltung interaktiver Systeme für sicherheitskritische Kontexte angesehen werden. Damit soll nicht angedeutet werden, dass diese drei Konzepte und Technologien den Problemraum sicherheitskritischer Mensch-Maschine-Systeme ganzheitlich abdecken. Sie stellen jedoch wichtige Aspekte dar, deren Potenziale und Einflüsse systematisch und grundlegend untersucht werden müssen, um Forschung und Entwicklung sicherheitskritischer Mensch-Maschine-Systeme voranzutreiben. Dies gilt sowohl für grundlegende Fragen zu den Prozessen und Prinzipien menschzentrierter Entwicklung als auch für Gestaltungsrichtlinien, z.B. Visualisierungs- und Interaktionskonzepte für Big-Data-Anwendungen. Dabei kann teilweise auf bewährten Erkenntnissen, u.a. zu Fragen der Automatisierung und Assistenz, aufgebaut werden. Andere Fragen, beispielsweise zur Thematik Society-in-the-Loop, lassen jedoch grundsätzlichen Forschungsbedarf erkennen.

Danksagung

Das Projekt „Artificial Intelligence and the Automated Ordering of Digital Communication” wird von der VolkswagenStiftung mit einem Planning Grant gefördert.

Literaturverzeichnis

Abdul, A., Vermeulen, J., Wang, D., Lim, B., & Kankanhalli, M. (2018). Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). ACM.

acatech. (2011). *Acatech Position. Cyber-physical systems: Innovationsmotor für Mobilität, Gesundheit, Energie und Produktion*. Berlin, Heidelberg: Springer.

acatech. (2012). Integrierte Forschungsagenda Cyber-Physical Systems. Verfügbar unter: http://www.acatech.de/?id=1405

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779.

Becker, T. (2016). Big Data Usage. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New Horizons for a Data-Driven Economy* (pp. 143–165). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-21569-3_8

Danaee, P., Ghaeini, R., & Hendrix, D. A. (2017). A Deep Learning Approach for Cancer Detection and Relevant Gene Identification. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, *22*, 219–229.

DIN Deutsches Institut für Normung e. V. (2011). *Ergonomie der Mensch-System-Interaktion – Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme*. (DIN EN ISO, 9241-210). Berlin: Beuth.

Dreyer, S., & Schulz, W. (2018). General Data Protection Regulation Falls Short of Algorithms and Artificial Intelligence. Report for the Bertelsmann Foundation. Verfügbar unter: https://doi.org/10.11586/2018011.

Ertel, W. (2016). *Grundkurs Künstliche Intelligenz*. Wiesbaden: Springer Fachmedien Wiesbaden.

Falcini, F., & Lami, G. (2017). Deep Learning in Automotive: Challenges and Opportunities. In A. Mas, A. Mesquida, R. V. O'Connor, T. Rout, & A. Dorling (Eds.), Communications in Computer and Information Science. Software Process Improvement and Capability Determination (Vol. 770, pp. 279–288). Cham: Springer International Publishing.

Fasel, D., & Meier, A. (2016). *Big Data*. Wiesbaden: Springer Fachmedien Wiesbaden.

Furht, B., & Villanustre, F. (2016). Big Data Technologies and Applications. Cham: Springer International Publishing.

Godfrey, P., Gryz, J., Lasek, P., & Razavi, N. (2016). Interactive Visualization of Big Data. In S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, & D. Kostrzewa (Eds.), Communications in Computer and Information Science: Vol. 613. Beyond databases: Advanced technologies for data mining and knowledge discovery (Vol. 613, pp. 3–22). Cham: Springer.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*: MIT Press.

Gunning, D. (2016). Explainable Artificial Intelligence (XAI). Verfügbar unter: https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf

Haun, M. (2014). *Cognitive Computing*. Berlin, Heidelberg: Springer.

Herczeg, M. (2014). *Prozessführungssysteme: Sicherheitskritische Mensch-Maschine-Systeme und interaktive Medien zur Überwachung und Steuerung von Prozessen in Echtzeit*. München: de Gruyter Oldenbourg.

Holzinger, A. (2018). Explainable AI (ex-AI). *Informatik-Spektrum, 41*(2), 138–143.

Koch, M., & Alt, F. (2017). Allgegenwärtige Mensch-Computer-Interaktion. *Informatik-Spektrum, 40*(2), 147–152.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety, Application Delivery Strategies. Verfügbar unter: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8–17.

Kreutzer, R. T. (2015). Der Gartner Hype Cycle als prognostischer Hintergrund. In R. T. Kreutzer (Hrsg.), *essentials. Digitale Revolution* (pp. 3–6). Wiesbaden: Springer Fachmedien Wiesbaden.

Lauras, M., & Comes, T. (2015). Special Issue on Innovative Artificial Intelligence Solutions for Crisis Management. Engineering Applications of Artificial Intelligence, 46, 287–288.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence: August 31, 1955. *AI Magazin*, 27(4), 12–14.

Mentler, T. (2018). Usability Engineering und User Experience Design sicherheitskritischer Systeme. In C. Reuter (Ed.), *Sicherheitskritische Mensch-Computer-Interaktion* (S. 41–60). Wiesbaden: Springer Fachmedien.

O'Neil, C. (2016): Weapons of math destruction: how big data increases inequality and threatens democracy. New York: Crown Publishers.

Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing Big Data with augmented and virtual reality: challenges and research agenda. *Journal of Big Data*, *2*(1), 22.

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. Ethics and Information Technology, 20(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8

Ratcliffe, S. (Ed.). (2016). *Oxford essential quotations* (4 ed.). [Oxford]: Oxford University Press.

Reuter, C., Ludwig, T., Kotthaus, C., Kaufhold, M.-A., Radziewski, E. von, & Pipek, V. (2016). Big Data in a Crisis? Creating Social Media Datasets for Crisis Management Research. *i-com*, *15*(3).

Rich, E. (1983). *Artificial Intelligence*: McGraw-Hill.

Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: *VL '96, Proceedings of the 1996 IEEE Symposium on Visual Languages*. Washington, DC, USA: IEEE Computer Society.

Toyoda, M. (2015). Big Data Analytics, 9498, 108–120.

Wick, C. (2017). Deep Learning. *Informatik-Spektrum, 40*(1), 103–107.

Zheng, Y., Wu, W., Chen, Y., Qu, H., & Ni, L. M. (2016). Visual Analytics in Urban Computing: An Overview. IEEE Transactions on Big Data, 2(3), 276–296.