

## Konzept zur Untersuchung kultureller Einflüsse auf die Akzeptanz und Leistung von automatisierten Systemen

Philip JOISTEN, Marius OBERLE, Ilka ZÖLLER, Bettina ABENDROTH

*Institut für Arbeitswissenschaft, Technische Universität Darmstadt, Otto-Berndt-Straße 2, D-64287 Darmstadt*

**Kurzfassung:** Im Zuge der Digitalisierung und Vernetzung durch die Initiative „Industrie 4.0“ steigt die Automatisierung neuer Systeme und Technologien. Die Akzeptanz von automatisierten Systemen wird u.a. von kulturellen Einflüssen geprägt. Das Ziel dieses Beitrags ist die Evaluation eines Konzepts zur Untersuchung des kulturellen Einflusses auf die Akzeptanz und Leistung der Mensch-Automation Interaktion bei verschiedenen Level der Automation und Kommunikation. Dazu wurde ein Experiment mit  $N = 32$  Probanden mit deutsch- ( $N_1 = 24$ ) und chinesisch-kulturellem ( $N_2 = 8$ ) Hintergrund durchgeführt. Die Ergebnisse zeigen, dass das eingesetzte Versuchskonzept geeignet ist, um kulturelle Einflüsse auf die Akzeptanz von und Leistung mit automatisierten Systemen zu untersuchen. Chinesische Probanden akzeptieren hochautomatisierte Automations- und Kommunikationsstufen eher. Wie zu erwarten steigt mit zunehmendem Automationsgrad die Systemleistung Mensch-Automation, bei gleichzeitiger Abnahme der Gedächtnisleistung des Menschen.

**Schlüsselwörter:** Automation, Akzeptanz, Kultur, Kommunikation, Mensch-Automation Interaktion, Terminplanung

### 1. Einleitung

Die zunehmende Digitalisierung und Vernetzung geht mit einem gleichzeitigen Anstieg der Automatisierung einher (vgl. Apt & Wischmann 2017; Mertens & Barbian 2016). Obwohl Automatisierung ein seit Dekaden etabliertes Forschungsfeld ist, verbleibt das Design und die erfolgreiche Implementierung automatisierter Systeme eine herausfordernde Aufgabe (vgl. Grote et al. 2014). Der Automatisierungsgrad charakterisiert die Interaktion zwischen Mensch und automatisierten System. Ein weiterer Faktor ist die Kommunikation zwischen Mensch und System. Die angemessene Gestaltung von Feedback und Information moderiert bspw. das Systemverständnis (Degani 2006) und kann somit zu einer Verringerung der Unsicherheit führen (Oberle et al. 2017).

Daneben ist die Akzeptanz von automatisierten Systemen von großer Bedeutung (vgl. Lee & See 2004). Auch eine gut gestaltete Maschine ist nutzlos, wenn der Bediener der Maschine nicht vertraut und sie nicht akzeptiert (Moray et al. 2000). Hinsichtlich der Akzeptanz neuer Technologien ist Deutschland eine eher konservative Kultur (Schmidt & Kollmann 2016), was den Prozess der Digitalisierung möglicherweise behindern könnte. Im Gegensatz dazu ist die chinesische Kultur für ihre hohe Akzeptanz und Integration von Technologie in Arbeit und Alltag bekannt (vgl. Schulze & Godehardt 2017).

Da die kulturell geprägte Akzeptanz je nach Automatisierungsgrad in Verbindung mit der Kommunikationsebene variieren kann, müssen beide Faktoren berücksichtigt werden. Schließlich ist die Leistungsfähigkeit des Gesamtsystems für eine vollständige Untersuchung des kulturellen Einflusses zu betrachten.

In diesem Beitrag wird ein Konzept zur Untersuchung der kulturellen Auswirkung auf die Akzeptanz und Leistung der Mensch-Automation Interaktion vorgestellt und evaluiert. Im Vordergrund steht folgende Forschungsfrage: Wie wirkt sich die Kultur auf die Akzeptanz und Leistung der Mensch-Automation Interaktion für verschiedene Automationsgrade und Kommunikationsebenen im Vergleich zwischen deutschen und chinesischen Akteuren aus? Im Folgenden wird das Versuchskonzept näher beschrieben, mit dem der aufgezeigten Forschungsfrage begegnet wird.

## 2. Versuchskonzept

Die Aufgabe der Teilnehmer ist es, mehrere Terminanfragen unterschiedlicher Priorität über eine eigenständig programmierte Simulationssoftware abzuwickeln, die auf einem Laptop arbeitet. Die Software simuliert verschiedene Grade der Automation sowie Ebenen der Kommunikation. Je nach Unterstützungsgrad müssen die Teilnehmer entweder alle Anfragen selbst bearbeiten (1), werden von der Software durch Empfehlungen zu den Anfragen unterstützt (2) oder haben eine Überwachungsfunktion inne, wobei die Anfragen von der Software selbständig mit (3) oder ohne (4) Benachrichtigung der Teilnehmer über die getroffene Maßnahme bearbeitet werden.

Der Ablauf des Experiments beginnt mit dem Empfang des Probanden. Um Missverständnisse zu vermeiden, wird dem Probanden eine schriftliche Aufgabenstellung vorgelegt. Alle Anweisungen und Fragebögen werden dem Probanden in deutscher Sprache präsentiert. Zusätzlich ist eine englische Übersetzung verfügbar. Nachdem der Proband mit dem Setup, der Software und der Aufgabe vertraut gemacht wurde, wird ein erster Fragebogen zur Erfassung der demografischen Daten (Geschlecht, Alter) ausgehändigt. Wenn es keine weiteren Rückfragen des Probanden gibt, beginnt das eigentliche Experiment. Dem Probanden werden die ersten Anfragen und ein zuvor definierter Unterstützungsgrad präsentiert. Nach der Bearbeitung aller Anfragen wird die Software geschlossen. Als ein Maß für die Leistung wird die Zeit für die Bearbeitung der Anfragen vom Versuchsleiter gestoppt. Als zweites Leistungsmaß wird der Proband gebeten, den resultierenden Zeitplan aus den bearbeiteten Anfragen auf einem Blatt Papier zu reproduzieren. Danach erhält der Proband einen Fragebogen zur Beurteilung der Akzeptanz gegenüber der zuvor verwendeten Software bzw. dem erfahrenen Unterstützungsgrad. Zu diesem Zweck wird eine Adaption des meCue-Fragebogens (Minge et al. 2013) verwendet. Der Fragebogen wurde ursprünglich entwickelt, um die User Experience zu bewerten, lässt aber auch Rückschlüsse auf die Akzeptanz zu. Genutzt werden die Skalen Nützlichkeit (N) und Benutzbarkeit (B) des Moduls I (Wahrnehmung aufgabenbezogener Qualität). Nachdem der Proband den Fragebogen ausgefüllt hat, wird dieser mit einem anderen Modus der Software konfrontiert und der Prozess der Terminplanung, Leistungs- und Akzeptanzmessung wiederholt. Jeder Teilnehmer wird mit zwei verschiedenen Kombinationen von Automatisierungsgrad und Kommunikationsebene konfrontiert (Stufe 1 und Stufe 3 oder Stufe 2 und Stufe 4). Um Sequenzierungseffekte zu vermeiden, wird die Kombination und Reihenfolge der Softwaremodi während der gesamten Studie permutiert. Nach dem zweiten Durchlauf des Experiments endet die Studie für den Probanden.

Zur Auswertung der Daten wurden Kontraste zwischen den kulturellen Gruppen analysiert. Das Signifikanzniveau wurde mit  $\alpha = 0,05$  angenommen. Zur Relevanz der signifikanten Unterschiede wurde die Effektstärke (d) nach Cohen (1988) berechnet. Ein Wert größer oder gleich 0,8 wird als großer Effekt interpretiert (Cohen 1988).

### 3. Evaluation des Versuchskonzepts

Im September 2018 wurde das Versuchskonzept im Rahmen einer Studie mit insgesamt 32 Probanden auf seine allgemeine Anwendbarkeit geprüft.

#### 3.1 Beschreibung des Kollektivs

Im Mittel über alle Probanden ergibt sich ein Alter von 27 Jahren, der Median beträgt 28 Jahre. Der Anteil von Männern (53,1 %) und Frauen (46,9 %) ist nahezu gleichverteilt. Acht Probanden haben einen kulturell chinesisch geprägten Hintergrund, 24 Probanden einen kulturell deutsch geprägten Hintergrund. Die demografischen Merkmale der Probanden sind in Tabelle 1 zusammengefasst.

**Tabelle 1:** Demografische Merkmale der Probanden nach Kultur und gesamt (N = 32)

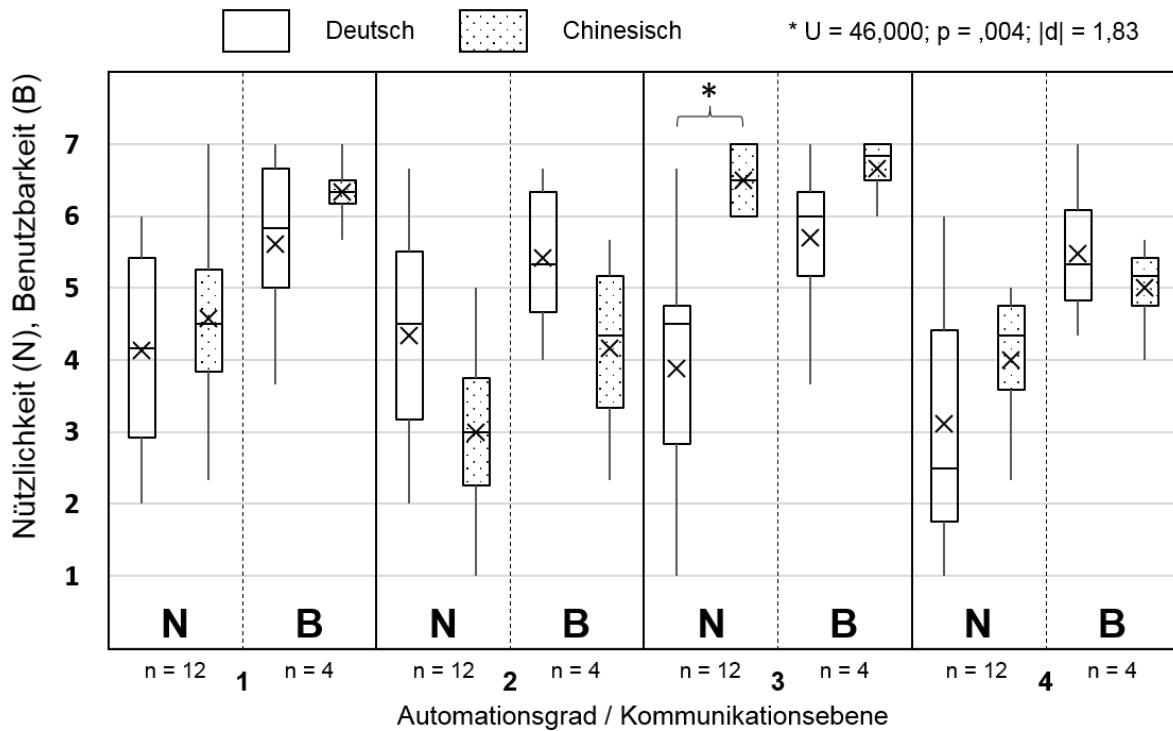
		Kultur		gesamt
		Deutsch	Chinesisch	
<b>Geschlecht</b>	w	11	4	15
	m	13	4	17
<b>Alter [Jahre]</b>	Median	28	26	28
	Mittelwert	27,5	25,5	27
	Standardabw.	2,02	2,62	2,31

#### 3.2 Akzeptanzmaße nach Kultur, Automationsgrad und Kommunikationsebene

Der Vergleich der Akzeptanzmaße zwischen den kulturellen Gruppen zeigt, dass kulturell chinesisch geprägte Probanden die Nützlichkeit des Automationsgrades 3 höher einstufen als kulturell deutsch geprägte Probanden (s. Abbildung 1).

Die Prüfung der Normalverteilung der Daten ist im Einzelfall mit dem Shapiro-Wilk Test durchgeführt worden. Weil eine Normalverteilung nicht für alle Variablen angenommen werden kann, wird für die Auswertung der non-parametrische Mann-Whitney-U Test verwendet. Die Unterschiede hinsichtlich der zentralen Tendenzen in den Gruppen „Kultur“ sind für den Automationsgrad 3 und die Dimension Nützlichkeit signifikant verschieden ( $U = 46,000$ ;  $p = ,004$ ;  $|d| = 1,83$ ). Weitere Unterschiede sind nicht statistisch signifikant.

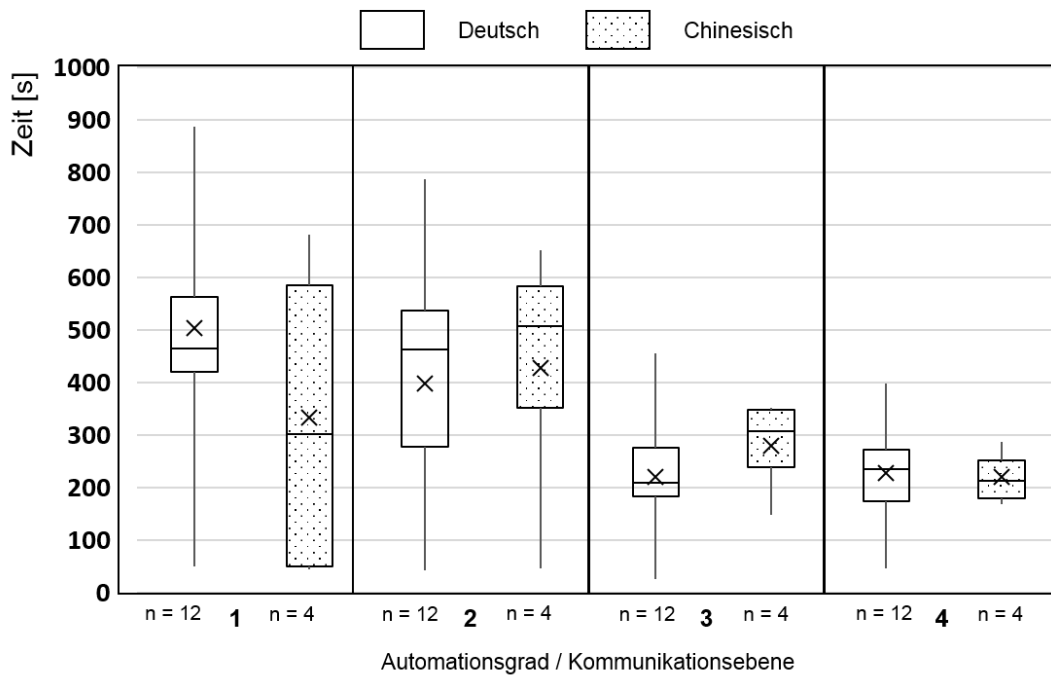
Bei der Betrachtung der Akzeptanzmaße von Unterstützungsgrad 3 (Software führt selbständig Operationen aus, mit Benachrichtigung des Probanden) und 4 (ohne Benachrichtigung des Probanden) wird der Einfluss der Kommunikation sichtbar. Hochautomatisierte Systeme, die nicht über die getroffenen Maßnahmen informieren, werden tendenziell weniger akzeptiert, als Systeme, die den Bediener durch Benachrichtigungen über ihre Handlungen informieren.



**Abbildung 1:** Akzeptanzmaße (Nützlichkeits- „N“, Benutzbarkeits- „B“) nach Kultur, Automationsgrad und Kommunikationsebene; 1 (lehne völlig ab) – 7 (stimme völlig zu)

### 3.3 Leistungsmaße nach Kultur, Automationsgrad und Kommunikationsebene

Mit zunehmenden Automationsgrad sinkt tendenziell die Zeit, die die Probanden für die Erfüllung der Aufgabe benötigen (s. Abbildung 2).

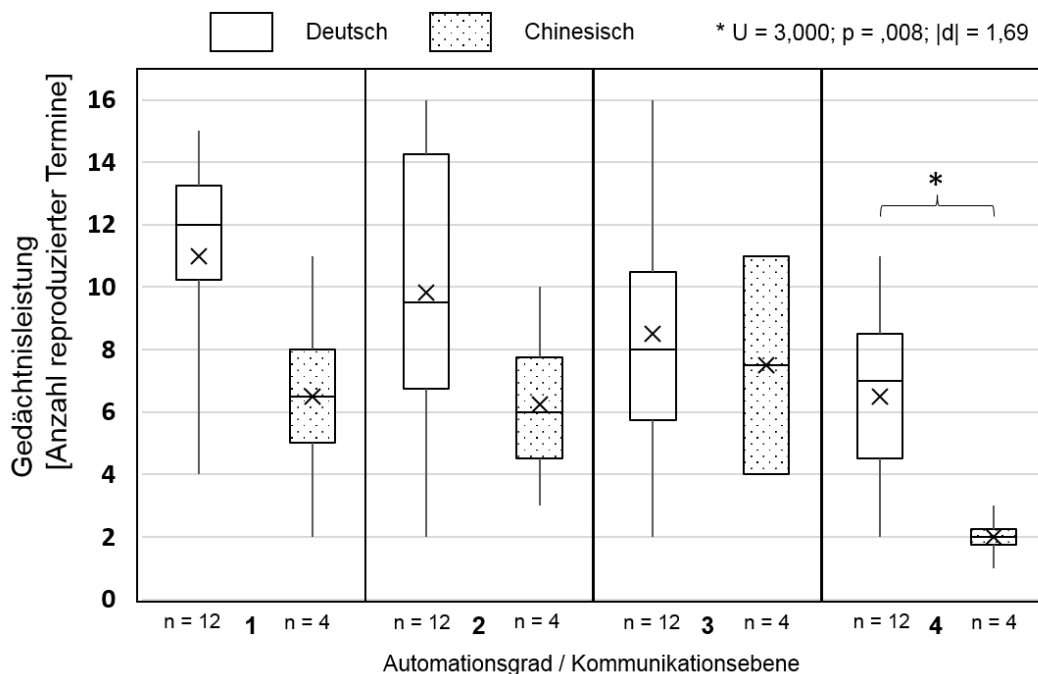


**Abbildung 2:** Benötigte Zeit zur Erfüllung der Aufgabe nach Kultur, Automationsgrad und Kommunikationsebene

Der Shapiro-Wilk-Test auf Normalverteilung zeigt, dass die Daten im Einzelfall normalverteilt sind. Der Test auf Varianzhomogenität (Levene-Test) ist nicht signifikant, es liegt Varianzhomogenität vor. Die Analyse der Kontraste (Vergleichsgruppe „Kultur“) wurde deshalb mit dem T-Test für unabhängige Stichproben durchgeführt. Der T-Test legt keine statistisch signifikanten Unterschiede der benötigten Zeit zur Erfüllung der Aufgabe in den Gruppen „Kultur“ für die verschiedenen Automationsgrade offen.

Mit zunehmenden Automationsgrad sinkt tendenziell die Anzahl der korrekt reproduzierten Termine nach Erfüllung der Aufgabe (Gedächtnisleistung, s. Abbildung 3).

Weil eine Normalverteilung nicht für alle Variablen angenommen werden kann, wird für die Auswertung der non-parametrische Mann-Whitney-U Test verwendet. Die Unterschiede hinsichtlich der zentralen Tendenzen in den Gruppen „Kultur“ sind für den Automationsgrad 4 signifikant verschieden ( $U = 3,000$ ;  $p = ,008$ ;  $|d| = 1,69$ ). Weitere Unterschiede sind nicht statistisch signifikant.



**Abbildung 3:** Gedächtnisleistung nach Erfüllung der Aufgabe nach Kultur, Automationsgrad und Kommunikationsebene

#### 4. Diskussion

Die Evaluation des Konzepts zeigt, dass es sich zur Erhebung von kulturellen Einflüssen auf die Akzeptanz von und Leistung mit automatisierten Systemen eignet. Problematisch zeigte sich die Bedienbarkeit des Interfaces der programmierten Simulationssoftware, wodurch ein Einfluss auf die Messung der Akzeptanz nicht auszuschließen ist. Eine Optimierung des Interfaces ist für weitere Studien zu empfehlen.

Das in der Studie verwendete automatisierte System simuliert den Prozess eines digitalen Terminplaners. Die Aufgabe stellt eine gängige und leicht zugängliche Situation dar. Alle Probanden sollten mit vergleichbaren Aufgabe vertraut sein, was unbeabsichtigte Einflüsse, wie Trainingseffekte, reduziert.

Basierend auf dem Modell zur Beschreibung der Mensch-Maschine Interaktion von Oberle (2017), von dem die Mensch-Automation Interaktion ein Sonderfall darstellt,

sind 67 menschliche Einflussfaktoren zu berücksichtigen. Um die Anzahl der Faktoren zu reduzieren und den Fokus auf kulturelle Unterschiede zu erhöhen, wurde in dieser Studie eine möglichst homogene Population gewählt. Die Probanden sind junge Menschen zwischen 22 und 30 Jahren. In nachfolgenden Studien wären weitere Einflussfaktoren zu berücksichtigen.

Es zeigt sich, dass Probanden mit chinesisch-kulturellem Hintergrund eine höhere Akzeptanz für hochautomatisierte Automations- und Kommunikationsstufen aufweisen. Erwartungsgemäß steigt mit zunehmendem Automationsgrad die Systemleistung Mensch-Automation, bei gleichzeitiger Abnahme der Gedächtnisleistung des Menschen. Während beim Leistungsmaß „benötigte Zeit zur Aufgabenerfüllung“ keine statistisch signifikanten kulturellen Einflüsse nachgewiesen werden konnten, zeigen sich statistisch signifikante Unterschiede hinsichtlich des Leistungsmaßes „Gedächtnisleistung“ bei Automationsstufe 4. Dies könnte aus dem geringen Stichprobenumfang der kulturell-chinesisch geprägten Probanden resultieren.

Weiter zeigt sich, dass bei der Gestaltung hochautomatisierter Systeme geeignete Kommunikationsstrategien zwischen System und Mensch zur Erhöhung der Akzeptanz des Nutzers berücksichtigt werden sollten.

Für nachfolgende Untersuchungen ist die Erweiterung des Leistungsmaßes um ein Fehlermaß, die Erweiterung der Stichprobengröße sowie die Erweiterung auf weitere Kulturkreise geplant.

## 5. Literatur

- Apt W, Wischmann S (2017) Neue Gestaltungsmöglichkeiten für die Arbeitswelt. In: Wittpahl V (Hrsg.) Digitalisierung. Bildung / Technik / Innovation. Berlin / Heidelberg: Springer Vieweg, 109-117.
- Cohen J (1988) Statistical Power Analysis for Behavioral Sciences. New York: Academic Press.
- Degani A (2006) Taming HAL: Designing Interfaces Beyond 2001. New York: Palgrave Macmillan.
- Grote G, Weyer J Stanton NA (2014) Beyond human-centered automation – concepts for human-machine interaction in multi-layered networks. *Ergonomics* 57:289-294.
- Lee JD, See KA (2004) Trust in automation: Designing for appropriate reliance. *Human Factors* 46:50-80.
- Mertens P, Barbian D (2016) Digitalisierung und Industrie 4.0 – Trend mit modischer Überhöhung? *Informatik Spektrum* 39:301-309.
- Minge M, Riedel L, Thuring M (2013) Modulare Evaluation interaktiver Technik. Entwicklung und Validierung des meCUE Fragebogens zur Messung der User-Experience. In: Brandenburg E (Hrsg.) Grundlagen und Anwendungen der Mensch-Maschine-Interaktion. 10. Berliner Werkstatt Mensch-Maschine-Systeme, 10.-12. Oktober 2013. Berlin: Univ.-Verl. der TU Berlin, 28-36.
- Moray N, Inagaki T, Itoh M (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of experimental psychology, Applied*, 6:44-58.
- Oberle M (2017) Development of a Method for Characterization, Assessment and Control of Human Induced Uncertainty During Usage. Dissertation, Darmstadt: Technische Universität Darmstadt.
- Oberle M, Sommer E, König C (2017) Reduction of Human Induced Uncertainty Through Appropriate Feedback Design. In: Nunes LI (Hrsg.) Advances in Human Factors and Systems Interactions: Proceedings of the AHFE 2016 Conference on Human Factors and System Interactions, July 27-31, 2016, Walt Disney World, Florida USA. Cham: Springer International Publishing, 43-55.
- Schmidt H, Kollmann T (2016). Deutschland 4.0: Wie die Digitale Transformation gelingt. Wiesbaden: Springer Fachmedien.
- Schulze D, Godehardt N (2017) China 4.0: Reaktionen in Partei und Gesellschaft auf die digitale Transformation. Berlin: SWP - Deutsches Institut für Internationale Politik und Sicherheit.

**Danksagung:** Wir bedanken uns bei Simon Dechant, Toni Dokoza, Tim Graf, Susanne Koch, Victor Listmann, Mirko Peraica und Julia Wenzel für die Mithilfe bei der Konzeptionierung, Programmierung und Evaluation des Versuchskonzeptes.