Sachbericht zum Verwendungsnachweis (Schlussbericht)

Trainingsansatz zur Vermittlung von Maßnahmen zur Prävention digitaler Desinformationskampagnen – PREVENT

Teilvorhaben: Medienethische und demokratietheoretische Begleitung und Evaluation

Schlussbericht
FKZ: 16KIS1494 (Universität Tübingen)

Ursprüngliche Laufzeit: 01.01.2022–31.12.2024
Laufzeitverlängerung bis 30.04.2025 (Universität Tübingen)

Verbundpartner:
FKZ 16KIS1491K (Universität Potsdam)
FKZ 16KIS1492 (Universität Bamberg)
FKZ 16KIS1493 (Universität zu Köln)
FKZ 16KIS1495 (Virtimo AG)

Prof. Dr. Jessica Heesen, Dr. Wulf Loh
Internationales Zentrum für Ethik in den Wissenschaften
Universität Tübingen
Wilhelmstraße 56
72074 Tübingen
Tel. +49 7071 29-77983
E-Mail: jessica.heesen@uni-tuebingen.de, wulf.loh@uni-tuebingen.de

Verbundkoordination:
Prof. Dr. Stefan Stieglitz (Universität Potsdam)

Projektmitarbeiterin:
Pawelec, Maria

Frühere Projektmitarbeiter:innen:
Dr. Sievi, Luzia
Dr. Vondermaßen, Marcel

Autorinnen des Schlussberichts:
Pawelec, Maria; Dr. Sievi, Luzia

I Kurze Darstellung

1 Ursprüngliche Aufgabenstellung

Das Projekt PREVENT knüpfte an die zunehmende Verbreitung von Falschinformationen in den sozialen Medien an. Diese können den demokratischen Diskurs sowie weitere Rechte und Werte wie Sicherheit bedrohen. Nichtsdestotrotz sind Eingriffe in die öffentliche Kommunikation und die Grundrechte auf freie Meinungsäußerung und Informationsbeschaffung in einem besonderen Maße rechtfertigungsbedürftig, v.a. dann, wenn sie vonseiten staatlicher Akteure erfolgen. PREVENT nahm Behörden und Organisationen mit Sicherheitsaufgaben (BOS) in Deutschland in den Blick, da sie von Falschinformationen potenziell besonders stark betroffen sind, aber gleichzeitig entscheidend dagegen vorgehen könnten. PREVENT untersuchte, vor welchen Herausforderungen BOS in Bezug auf Falschinformationen in den sozialen Medien stehen und wie sie darauf reagieren können und sollten. Hauptziel war dabei, der Entstehung von digitalen Desinformationskampagnen vorzubeugen. Dazu entwickelte das Konsortium einen innovativen Trainingsansatz, der sich an BOS richtet. Dies beinhaltete:

1. die Entwicklung eines **Demonstrators**, der die Entstehung von digitalen Desinformationskampagnen simulierte (Simulationsmodul) und es BOS ermöglichte, wirksame vorbeugende Maßnahmen zu erlernen (Trainingsmodul);
2. die Entwicklung von **wirksamen Präventionsmaßnahmen**, die BOS ergreifen können, um der Entstehung von digitalen Desinformationskampagnen vorzubeugen;
3. die **ethische und rechtliche Bewertung** dieser Maßnahmen, um deren verantwortungsvolle und legale Verwendung in realen Krisen sicherzustellen.

Das IZEW-Teilvorhaben „Ethische Bewertung von Maßnahmen zur Prävention von digitalen Desinformationskampagnen“ untersuchte die Anforderungen an einen von den Grundsätzen der Medienfreiheit und Staatsferne der Medien gedeckten Umgang von BOS mit Desinformationskampagnen. Daneben war es wesentlich beteiligt an der Erstellung medienpädagogisch reflektierter Schulungen und Materialen für BOS. Zentrale Ziele des Teilvorhabens waren:

1. die **forschungsethische Begleitung** der technischen Umsetzung des Demonstrators;
2. die Entwicklung von **ethischen Leitlinien** für die Umsetzung und Evaluation der im Projekt entwickelten Maßnahmen, die BOS ergreifen können, um der Entstehung von digitalen Desinformationskampagnen vorzubeugen;
3. die Entwicklung eines **pädagogischen Konzepts und von Schulungsangeboten** für den im Projekt zu erstellenden Demonstrator.

2 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Für die Bearbeitung der Forschungsfragen wurden vorliegende Studien umfassend zur Kenntnis genommen sowie Debatten auf Tagungen in den Bereichen politische Ethik und Philosophie, politikwissenschaftliche Demokratieforschung, Katastrophenethik, ethische Forschung zu (staatlichen) Maßnahmen gegen Falschinformationen sowie zur Krisenkommunikation und den Social Media-Aktivitäten von BOS kontinuierlich verfolgt. Es zeigte sich, dass die wissenschaftliche (und mediale) Debatte über die negativen Auswirkungen von Falschinformationen u.a. auf die Demokratie und die Sicherheit zwar zunehmend an Bedeutung gewinnt, sich aber nur wenige Beiträge mit konkreten Gegenmaßnahmen auseinandersetzen. Darüber hinaus werden BOS nur sehr selten in den Blick genommen. Es fehlt darüber hinaus an einer ethischen Reflexion von

Gegenmaßnahmen, insbesondere jenen, die BOS ergreifen können (s. Sievi/Pawelec 2025). Dies untermauert die bereits im Projektantrag festgehaltenen eklatanten Forschungslücken. Die Durchführung des Vorhabens berührte keine Schutzrechte.

3 Ablauf des Vorhabens

Die gemäß Antrag durchgeführten Arbeitspakete der Verbundpartner waren:

- AP 1: Stand der Forschung und Analyse des Nutzungskontexts
- AP 2: Entwicklung von Maßnahmen zur Prävention von digitalen Desinformationskampagnen
- AP 3: Technische Implementierung des Trainingsansatzes
- AP 4: Entwicklung eines ethisch-rechtlichen Rahmens für Maßnahmen zur Prävention von digitalen Desinformationskampagnen
- AP 5: Evaluation des Demonstrators, der vorbeugenden Maßnahmen und des ethisch-rechtlichen Rahmens

Die Laufzeitverlängerung wurde durch die Verbundpartner genutzt, um Arbeiten durchzuführen, die aufgrund von Verzögerungen insbesondere durch die Verschiebung von Projektworkshops (v.a. mit dem Ziel, mehr Praktiker:innen zu erreichen), den Umzug von zwei Lehrstühlen und Personalwechsel nicht plangemäß durchgeführt werden konnten.

4 Wesentliche Ergebnisse

Das zentrale Ergebnis des IZEW-Teilprojekts ist eine wesentliche Erweiterung des Wissensstandes zur Rolle von BOS in der Bekämpfung von Falschinformationen und zur ethischen Bewertung möglicher Gegenmaßnahmen. PREVENT zeigte auf der Grundlage empirischer Erhebungen, dass deutsche BOS mit einer Vielzahl verschiedener Falschinformationen konfrontiert sind und dabei als falsch empfundener Kritik an der eigenen Organisation einen besonderen Stellenwert einräumen, was aus demokratietheoretischer Sicht problematisch sein kann. Bisher haben vor allem BOS in Großstädten sowie auf Bundesebene mit organisierten Desinformationskampagnen zu tun, während Gerüchte und Spekulationen (Misinformation) für viele BOS insbesondere in Einsatzlagen eine Herausforderung darstellen. Darüber hinaus konnte PREVENT zahlreiche praktische Herausforderungen im Umgang mit Falschinformationen herausarbeiteten, die konkrete Ansatzpunkte für eine Stärkung der BOS zur Bekämpfung von Falschinformationen fernab politisch umstrittener Regulierunsentscheidungen (etwa zum Umgang mit Social Media-Plattformen) bieten. PREVENT sammelte und systematisierte zudem erstmals eine Vielzahl möglicher individueller, koordinierter und (teil-)automatisierter Gegenmaßnahmen, die BOS selbst bereits ergreifen oder in der Zukunft ergreifen könnten, und unterzog diese einer vertieften ethischen (und rechtlichen) Bewertung.

Die Ergebnisse von PREVENT wurden sowohl mit verschiedenen (inter-)disziplinären wissenschaftlichen Fachcommunities geteilt als auch breiter in die Öffentlichkeit und die BOS-Praxis gestreut. Dazu entwickelte PREVENT didaktisch aufbereitete Lehr- bzw. Lernmaterialien, um die zentralen Forschungsergebnisse sowohl im Rahmen des entwickelten Demonstrators als auch der Open Access-Handreichung mit einer breiten Community an Praktiker:innen zu teilen, so dass sie im Arbeitsalltag von BOS nachhaltig wirksam werden können. Damit stärkt PREVENT eine Gruppe von Akteuren, die in der politisch und gesellschaftlich zunehmend bedeutsamen Debatte über den Umgang mit Falschinformationen in den sozialen Medien häufig vernachlässigt wird, hierbei jedoch sowohl eine große Verantwortung trägt als auch potenziell besonders wirkmächtig ist.

II Eingehende Darstellung

1 Verwendung der Zuwendung und erzielte Ergebnisse

AP 1: Stand der Forschung und Analyse des Nutzungskontexts

(1) Verwendung der Zuwendung

Beschäftigungsentgelte für wissenschaftliche Mitarbeiter:innen, studentische Hilfskräfte, Reisekosten für Projekttreffen und zur Durchführung der Expert:inneninterviews.

(2) Ziele

- AP 1.1: Durchführung einer systematischen Literaturanalyse, deren Ergebnisse in einer gemeinsamen Publikation veröffentlicht werden (UP, UB, UKÖ, IZEW)
- AP 1.2: Analyse von digitalen Desinformationskampagnen (UP, UB)
- AP 1.3: Entwicklung von Erwartungsprofilen, abgeleitet aus einer repräsentativen Umfrage zu den Erwartungen der Bevölkerung, wie BOS mit digitalen Desinformationskampagnen umgehen sollten (UKÖ) sowie Expert:innen-Interviews mit Vertretenden verschiedener BOS zu deren Bedenken, Herausforderungen und Möglichkeiten bezüglich vorbeugender Maßnahmen (IZEW); Publikation der Ergebnisse
- AP 1.4: Integrationskonzept

(3) Arbeitsschritte und erzielte Ergebnisse

In **AP 1.1** führte das IZEW eine systematische Literaturanalyse aktueller und grundlegender philosophischer, politik-, medien- und demokratietheoretischer Literatur über Des- und Misinformation, den Umgang von Behörden damit sowie zur Bedrohung der demokratischen Öffentlichkeit dadurch bzw. durch Gegenmaßnahmen (staatliche Eingriffe) durch. Das IZEW lotete dabei die normativ zugeschriebenen Funktionen (digitaler) Öffentlichkeiten für demokratische Prozesse, die politiktheoretischen Notwendigkeiten und Grenzen staatlicher Eingriffe in politische Bildung und ethisch-politische Selbstverständigungsprozesse, sowie die philosophisch-ethischen Grenzen freier Meinungsäußerung einerseits, sowie staatlicher Zensur andererseits, aus. Auf der Grundlage der im Konsortium durchgeführten Literaturanalysen wurden zunächst begriffliche Fragen (insbesondere die Abgrenzung von Falsch-, Des- und Misinformation) geklärt und eine gemeinsame Definition von Desinformation erarbeitet. Zentrale Erkenntnisse wurden zudem in einem gemeinsam erstellten interdisziplinären Konferenzartikel bei der Internationalen Tagung für Wirtschaftsinformatik 2023 veröffentlicht, der 2025 in einem Sammelband erschien:

> ➢ Stieglitz, Stefan; Fromm, Jennifer; Kocur, Alexander; Rostalski, Frauke; Duda, Michelle; Evans, Alison; Rieskamp, Jonas; Sievi, Luzia; Pawelec, Maria; Heesen, Jessica; Loh, Wulf; Fuchß, Christopher; Eyilmez, Kaan (2025): What Measures Can Government Institutions in Germany Take Against Digital Disinformation? A Systematic Literature Review and Ethical-Legal Discussion, in: Daniel Beverungen, Christiane Lehrer und Matthias Trier (Hrsg.): Transforming the Digitally Sustainable Enterprise. Cham: Springer Nature, S. 319-337 (zuvor Preprint als Konferenzbeitrag: 18. Internationale Tagung Wirtschaftsinformatik 2023)

Das IZEW analysiert hier erstens, inwiefern Desinformation schädlich für die öffentliche Kommunikation und demokratische Deliberation ist und die normativ zugeschriebenen Funktionen der demokratischen Öffentlichkeit stört. Der öffentliche Raum, einschließlich der sozialen Medien, erfüllt in pluralistischen Demokratien wichtige Aufgaben. Dies umfasst, die Bürger:innen zu informieren, politische Meinungsbildung (und dabei auch Kritik an und damit Kontrolle von staatlichen Stellen) zu ermöglichen und Meinungsäußerungen in die politische Sphäre zu übertragen. In einem demokratischen Staat muss der Meinungsbildungsprozess ohne größere Verzerrungen ablaufen und Rede-, Meinungs- und Ausdrucksfreiheit gewahrt werden. Eingriffe in die öffentliche Kommunikation durch Regierungsbehörden bedürfen einer besonderen Rechtfertigung und sind nur legitim, wenn das Funktionieren des öffentlichen

Raums ernsthaft gefährdet ist. Dies kann durch digitale Desinformationskampagnen der Fall sein, da sie die Meinungsbildung verzerren, die Kontrolle von politischen Prozessen erschweren, gesellschaftliche Marginalisierung verschärfen und das Vertrauen in Institutionen wie etablierte Medien oder BOS selbst schwächen. BOS müssen daher die schädlichen Auswirkungen von Desinformation zu einem gewissen Grad eindämmen, ohne demokratische Werte wie Meinungs-, Informations- und Pressefreiheit und das Recht auf Privatsphäre zu untergraben. Zweitens unterziehen wir einzelne Gegenmaßnahmen (Erhöhung der Medienkompetenz, „social inoculation", Debunking, automatisierte Erkennung von Desinformation, social bots) einer ersten ethischen und demokratietheoretischen Analyse und loten Grenzen staatlicher Eingriffe in ethisch-politische Selbstverständigungsprozesse aus. Insbesondere die Förderung von Medienkompetenz steht im Einklang mit dem demokratischen Konzept der freien, gleichberechtigten und politisch aktiven Bürger:innen und ist eine wichtige Aufgabe für Demokratien. Medienkompetenzschulungen sollten jedoch nicht nur kritisches Denken lehren, sondern auch, warum Bürger:innen einzelnen Quellen mehr vertrauen können als anderen. Wir heben darüber hinaus zentrale Kriterien für eine ethisch reflektierte Krisenkommunikation durch Behörden hervor und zeigen, dass die Betonung sozialer Normen als Bevormundung und Benachteiligung von Gruppen wahrgenommen werden sowie zu weiterer Marginalisierung führen kann. Wir diskutieren die Gefahr, dass die Entlarvung von Falschinformationen für Propaganda oder die Unterdrückung unliebsamer Meinungen missbraucht wird. Wenn BOS Algorithmen zur automatischen Identifikation von Desinformation einsetzen, kann dies den Datenschutz und die Privatsphäre gefährden und zu Selbstzensur führen und bedarf einer besonderen Begründung. Transparenz und Widerspruchsmöglichkeiten gegen potenzielle Fehlalarme sind wichtig. Wir zeigen zudem auf, dass die Nutzung von Social Bots durch BOS umstritten ist. Nicht zuletzt identifizieren wir offene ethische und demokratietheoretische Fragen, die BOS bei der Anwendung von Maßnahmen gegen Desinformation mitbedenken sollten. Dazu gehört etwa die Frage, wer darüber entscheidet, was als Desinformation gilt.

AP 1.3 diente der Entwicklung von Erwartungsprofilen, um die Bedenken, Herausforderungen und Möglichkeiten deutscher BOS beim Umgang mit Falschinformationen zu erfassen. Maßgeblich waren Interviews mit Vertretenden deutscher BOS. Das IZEW führte ab April 2022 elf qualitative, leitfadengestützte Interviews mit Expert:innen zum Thema Öffentlichkeitsarbeit und/oder Desinformation von elf deutschen BOS durch: je eine Polizei einer mittelgroßen Stadt und einer Großstadt, die Bundespolizei, ein Desinformationsexperte der Polizei NRW, das BBK, eine weitere, hier anonymisierte Bundessicherheitsbehörde, je eine Feuerwehr einer mittelgroßen Stadt und einer Großstadt, das THW sowie zwei Hilfsorganisationen. Wir interviewten sechs Frauen und neun Männer, von denen sechs Personen Leitungsposten (zumeist als Leiter:innen der Pressestelle oder des Pressereferats) innehatten. Abgesehen von zwei Personen, die von der Institution als in die Pressearbeit neu einzuarbeitende Mitarbeitende zum Interview dazu gebeten wurden, verfügten alle Befragten über mehrjährige Arbeitserfahrung. Wir fragten in den 1-1,5 Stunden langen Interviews nach der allgemeinen Öffentlichkeitsarbeit der Organisation, nach konkreten Erfahrungen und Problemen mit Falschinformation, nach Maßnahmen, die bei deren Bekämpfung eingesetzt werden und entsprechenden Erfahrungen, nach Wünschen und Bedarfen für kommende Schulungen und technische Tools zur Desinformationsbekämpfung sowie nach Einschätzungen zur Verwendung automatisierter Maßnahmen und Social Bots. Die Interviews wurden anonymisiert transkribiert und für die Projektpartner:innen zusammengefasst. Zudem erarbeitete das IZEW zwei interne Paper mit den zentralen Ergebnissen: Erstens zu den bisherigen Erfahrungen der interviewten BOS mit Desinformation sowie ihren Erwartungen an

Schulungen und den Demonstrator, und zweitens zu bereits angewendeten Gegenmaßnahmen sowie den Herausforderungen, denen BOS beim Umgang mit Desinformation begegnen.
Zu diesen Ergebnissen war zunächst eine Veröffentlichung mit der UKÖ geplant, die im AP eine repräsentative Bevölkerungsumfrage durchführte. Aufgrund der Fülle des Materials und einer unterschiedlichen Verwertung der Ergebnisse zeigte sich dann, dass getrennte disziplinäre Publikationen aus wissenschaftlicher Sicht sinnvoller waren. Das IZEW veröffentlichte daher zentrale Erkenntnisse in einem peer-reviewten Open Access-Artikel:

- Pawelec, Maria; Sievi, Luzia (2023): Falschinformationen in den sozialen Medien als Herausforderung für deutsche Sicherheitsbehörden und -organisationen. In: Kriminologie – Das Online-Journal 4(5), S. 316–347.

Der Artikel beruht auf der erfolgten Literaturanalyse (AP 1.1) sowie einer tiefgreifenderen Analyse des Interviewmaterials und dessen Auswertung im Rahmen einer qualitativen Inhaltsanalyse nach Gläser und Laudel. Im Ergebnis zeigte sich, dass BOS-Vertretende einen weiten Begriff von „Des- und Falschinformation“ haben. Viele verbinden damit (als unzutreffend empfundener) Kritik an der eigenen Institution bis hin zum „Shitstorm“ gegen die eigene Organisation. Einige begegnen hauptsächlich Gerüchten und Spekulationen (Misinformation) einer verunsicherten Bevölkerung. Intendierte, politische Desinformation betrifft hauptsächlich BOS in Großstädten und auf Bundesebene. Dazu gehören gegen die Polizei gerichtete sowie rassistische Desinformationskampagnen und ideologisch motivierte Desinformation im Zusammenhang mit Covid-19. In Bezug auf mögliche Gegenmaßnahmen identifizierten wir Schulungen zur Thematik, interne Sprachregelgungen und -prozesse, interne Informationsbeschaffung, -priorisierung und -verifikation, aktive und präventive Kommunikation, Vertrauens- und Communitymanagement, reaktive (Krisen-)Kommunikation, Löschen und Blockieren von Nutzenden und Posts, Kooperation mit anderen BOS sowie bewusste Zurückhaltung. Dabei stehen BOS in einem Spannungsfeld zwischen dem Schutz der demokratischen Öffentlichkeit vor Falschinformationen sowie Meinungsfreiheit und Datenschutz. Alle BOS begegnen zudem Herausforderungen wie geringem behördeninternen Problembewusstsein, Ressourcenmangel, psychischen Belastungen, Abstimmungs- und Zuständigkeitsproblemen. Hier anzusetzen, könnte BOS im Kampf gegen Falschinformationen stärken.
AP 1.4 sicherte die kontinuierliche Zusammenarbeit der Projektpartner durch regelmäßige Verbundtreffen, (Online-)Besprechungen, gemeinsame Workshops und interdisziplinäre Publikationen. Bei den Besprechungen präsentierten die Partner jeweils den aktuellen Arbeitsstand und stellten geplante Schritte und Methoden zur Diskussion. Damit und durch die Teilnahme der assoziierten Partner sowie weiterer BOS an den Workshops konnte die inter- bzw. transdisziplinäre Perspektive gewährleistet und Einigkeit (bzw. transparente Differenzen) über den Forschungsgegenstand hergestellt werden (vgl. Stieglitz et al. 2023, Schewina et al.2024). Die Einrichtung einer Homepage und die Nutzung von Onlinediensten ermöglichte die Bereitstellung und den Austausch von Daten und (Zwischen-)Ergebnissen.
AP 1 schuf die **Grundlagen für das weitere Vorgehen**: Die Literaturanalyse bildete eine Basis, aus der Maßnahmen für BOS entwickelt, sowie im späteren Verlauf ethisch und rechtlich reflektiert wurden. Zudem griffen wir auf die Expert:innen-Interviews zurück, um Maßnahmen zu finden, die BOS bereits anwenden oder sich als mögliche zukünftige Methoden vorstellen können. Zudem konnten wir die bestehenden Werte und Leitbilder sowie die ethischen und rechtlichen Reflexionen zu Problematiken von Desinformation sowie ihrer Bekämpfung durch BOS erheben. Zuletzt gaben uns die Interviews einen Einblick, welche Schulungsangebote bereits für BOS zum Thema Desinformation existieren, welches

Vorwissen wir bei welchen Zielgruppen erwarten können, und welche Vorstellungen und Anforderungen die Interviewten an Schulungsangebote anlegen.

AP 2: Entwicklung von Maßnahmen zur Prävention von digitalen Desinformationskampagnen

(1) Verwendung der Zuwendung

Beschäftigungsentgelte für wissenschaftliche Mitarbeiter:innen, studentische Hilfskräfte, Reisekosten für Projekttreffen und Workshops.

(2) Ziel

Die Verbundpartner setzten gemeinsam das im Projektantrag formulierte Ziel um, individuelle, koordinierte und automatisierte vorbeugende Maßnahmen für BOS zu entwickeln. Das IZEW begleitete und informierte diesen Prozess (forschungs-)ethisch.

(3) Arbeitsschritte und erzielte Ergebnisse

Das IZEW organisierte eine interne Sitzung mit den Konsortialpartnern, um die partizipative Maßnahmenentwicklung gemeinsam mit den assoziierten Projektpartnern und weiteren BOS **forschungsethisch** vorzubereiten und eine wertebasierte Entwicklung (value sensitive design) sowie eine verantwortungsbewusste Forschung (Responsible Research and Innovation) sicherzustellen. Wir diskutierten, welche Art der Forschungspartizipation von BOS PREVENT anstrebt und was für die Projektmitarbeitenden zu beachten ist, damit die Partizipation auch von Seiten der BOS als gelungen erfahren wird. In Anlehnung an Bethmann et al. (2021) stellten wir die verschiedenen Stufen der Partizipation vor. Dazu gehört die Vorstufe (Information, Anhörung und Einbeziehung), die eigentliche Partizipation (Mitbestimmung, teilweise Entscheidungskompetenz und Entscheidungsmacht) sowie die darüber hinausgehende Selbstorganisation. Das IZEW zeigte zudem auf, welche Kosten eine Partizipation für teilnehmende BOS bedeutet: Diese müssen Zeit und Geld investieren, sie erleben möglicherweise Rollenkonflikte und müssen auch bei den Gesprächen jederzeit den Datenschutz im Hinterkopf behalten. Partizipation bedeutet auch Verantwortung, die als Überlast wahrgenommen werden kann. Zudem können die gemeinsam getroffenen Entscheidungen dazu führen, dass die Beziehungen zur eigenen Behörde oder zur Öffentlichkeit beeinträchtigt werden. Das IZEW regte damit eine forschungsethische Reflexion an, welche Form der Forschungspartizipation wir in welchem Format anbieten wollten und wie PREVENT eine gleichberechtigte Teilhabe für unterschiedliche Gruppen ermöglichen konnte. Zudem mahnte das IZEW an, die Beiträge der Teilnehmenden ergebnisoffen zu behandeln und Hyperpartizipation oder „Particitainment“ (Selle 2011) zu vermeiden. Es zeigte sich, dass wir die Teilnehmenden in ihren Wünschen und Ideen anhören sowie sie bei der Entwicklung neuer Maßnahmen einbeziehen können (Stufe 4 und 5 nach Bethmann et al. 2021). Mitbestimmung oder gar Entscheidungskompetenz, welche Maßnahmen von uns ausführlicher beforscht und in den Demonstrator aufgenommen werden, war nicht möglich, um die Unabhängigkeit der eigenen Forschung zu bewahren.

Die **Entwicklung von Maßnahmen** zur Prävention digitaler Desinformationskampagnen erfolgte dann zunächst basierend auf unserer Literaturanalyse und den durchgeführten Expert:inneninterviews (AP 1). Dabei sammelten wir verschiedene Maßnahmen, die BOS bereits durchführen oder anwenden könnten. Es folgte laufend eine enge Abstimmung mit unseren assoziierten Projektpartnern sowie weiteren eingeladenen BOS in drei Workshops in Paderborn, Köln und Potsdam, bei denen wir auch gemeinsam weitere Maßnahmen entwickelten. Das IZEW, die UB und die UKÖ strukturierten daraufhin in mehreren Iterationen die so entstandene Sammlung an bestehenden sowie möglichen zukünftigen

Gegenmaßnahmen. Die folgende Abbildung zeigt das Ergebnis dieser Strukturierung der identifizierten individuellen, koordinierten und automatisierten Maßnahmen von BOS zur Vorbeugung und Bekämpfung von Falschinformationen:

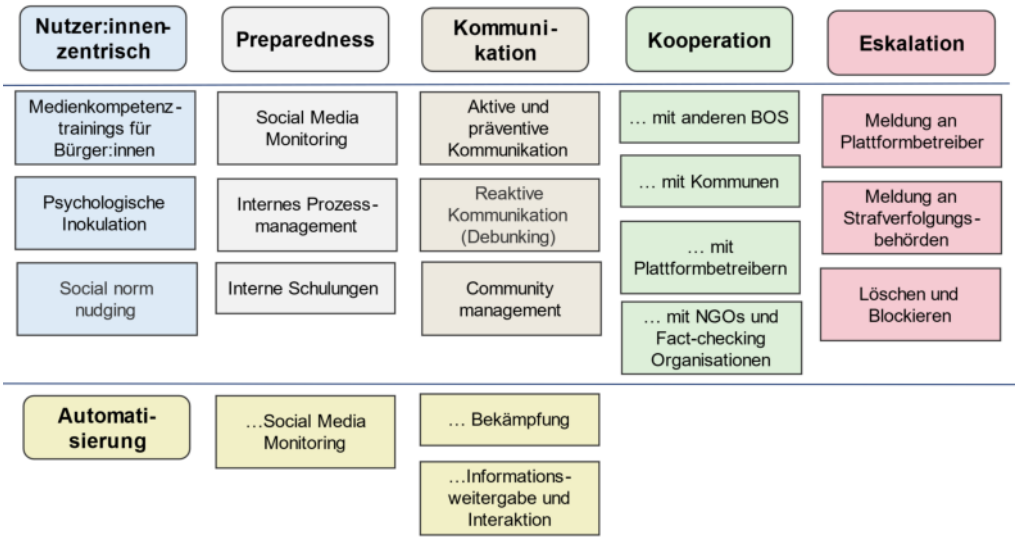

Abbildung 1: (Mögliche) Maßnahmen von BOS gegen Falschinformationen (eigene Darstellung)

Nutzer:innenzentrische Maßnahmen konzentrieren sich auf die Nutzenden sozialer Medien. Diese sollen nicht zur Verbreitung falscher Informationen beitragen oder diese nicht ohne weitere Prüfung glauben. Diese Maßnahmengruppe wird vor allem in der Fachliteratur vorgeschlagen. Sie bezieht sich größtenteils auf Regierungen (Medienkompetenztrainings, die v.a. in Schulen und in der politischen Bildung angeboten werden sollen) und Plattformen (die mit Nudging die Nutzenden dazu anregen sollen, Falschinformationen weniger zu verbreiten). In unseren Interviews und Workshops zeigte sich, dass die meisten BOS keine eigenen Medienkompetenztrainings für eine breite Bevölkerung durchführten und, dass die Maßnahmen Nudging und psychologische Inokulation den teilnehmenden BOS nicht bekannt waren und auch nicht umgesetzt wurden. Eine zweite Gruppe individueller Maßnahmen trägt zur Vorbereitung ("Preparedness") der BOS bei: Social Media Monitoring, die Schaffung interner Prozesse für den Umgang mit Falschinformationen sowie interne Schulungen. Unsere Interviews zeigten, dass Prepardness eine wichtige Rolle spielt. Oftmals sind Hierarchieebenen in den Behörden ein Hemmnis für eine schnelle und zeitnahe Bearbeitung von Desinformationen, weil eine Reaktion der Social Media Abteilung mit vielen Ebenen abgestimmt werden muss. Auch ein geringes Bewusstsein für die Problematik von Falschinformationen bei den Vorgesetzten und Kolleg:innen erschwert laut den Interviews die Bearbeitung von Falschinformationen. Drittens spielt die Kommunikation der BOS sowohl präventiv vor dem Auftreten von Desinformation als auch reaktiv eine wichtige Rolle. BOS können zudem eine aktive, loyale Gemeinschaft von Follower:innen ihrer Kanäle aufbauen. Das Community- und Vertrauensmanagement wurde insbesondere in den Interviews als wichtige Maßnahme für BOS hervorgehoben. Schließlich können BOS auf besonders eingriffsintensive individuelle Maßnahmen zurückgreifen ("Eskalation"), etwa, dass sie unwahre Inhalte an Plattformbetreiber oder an die Strafverfolgungsbehörden melden, oder dass sie – allerdings nur auf ihren eigenen Social-Media-Seiten – Nutzende oder bestimmte Wörter sperren und sogar Beiträge löschen.

Im Rahmen der koordinierten Maßnahmen ist es zudem denkbar, dass BOS mit anderen Organisationen und Institutionen kooperieren, um falsche Informationen zu bekämpfen. Unsere Interviews zeigen, dass BOS dies v.a. in komplexen und unvorhergesehenen Lagen

tun und dabei in Krisenreaktionszentren etwa mit der Kommune und weiteren BOS zusammenarbeiten, um Situationen wie Naturkatastrophen oder Gewalttaten zu bewältigen. Die komplexe rechtliche Lage in Bezug auf die Zuständigkeiten einzelner BOS wurde durch die UKÖ für den Demonstrator und die Handreichung aufgearbeitet.
Darüber hinaus können BOS im Sinne der automatisierten Maßnahmen technische Tools nutzen, um bestimmte Vorgänge wie das Social Media Monitoring oder die Weitergabe von Informationen an die Bevölkerung zu erleichtern oder sogar zu automatisieren. Unsere Interviews und Workshops zeigen, dass BOS dies zu einem gewissen Grad bereits tun und beispielsweise auf Programme zurückgreifen, die eine so genannte „Sentiment-Analyse" von Social Media-Posts erlauben, also KI-basiert kategorisieren, ob es sich um einen eher positiven, neutralen oder negativen Post handelt. Spezielle Tools, um Falschinformationen aufzudecken oder gar automatisiert zu bekämpfen, scheinen BOS jedoch nicht vorzuliegen.

AP 3: Technische Implementierung des Trainingsansatzes

(1) Verwendung der Zuwendung

Beschäftigungsentgelte für wissenschaftliche Mitarbeiterinnen, studentische Hilfskräfte, Reisekosten für Projekttreffen und Konferenzen.

(2) Ziele

- AP 3.1: Entwicklung eines Technologiekonzepts, das technische und funktionale Anforderungen an den Demonstrator formuliert.
- AP 3.2: Implementierung eines Basismoduls
- AP 3.3: Implementierung des Simulationsmoduls einschließlich Modellierung geeigneter Parameter für realistische Simulation der Entstehung digitaler Desinformationskampagnen.
- AP 3.4: Implementierung des Trainingsmoduls, einschließlich (vonseiten des IZEW) Entwicklung eines entsprechenden pädagogischen Konzepts und der Erstellung von Lernmaterialien, die zur ethischen Reflektion der Maßnahmen anregen.

(3) Arbeitsschritte und erzielte Ergebnisse

In Zusammenarbeit mit den Praxispartner:innen entwickelte das Konsortium in **AP 3.1** technische und funktionale Anforderungen an den Demonstrator. Diese wurden im von der Virtimo AG organisierten Technologie-Workshop im Oktober 2022 diskutiert und priorisiert. Das IZEW brachte in das Anforderungsdokument sowie in den Workshop die Erwartungsprofile der BOS aus den Expert:inneninterviews, die Ergebnisse der Literaturrecherche sowie forschungsethische Überlegungen ein. Beispielsweise sollte Nutzenden erläutert werden, auf welchen Einsichten, Forschungsergebnissen und Vorannahmen die voreingestellte Simulation sowie nicht anpassbare Parameter beruhen. Das Simulationsmodul sollte zudem auf ethische Bedenken hinweisen, wenn Nutzende bestimmte Maßnahmen auswählen, denn es ist denkbar, dass bestimmte Maßnahmen zwar wirkungsvoll sind (wie das Löschen oder Blockieren von Nutzenden oder Posts) aber ethisch (und rechtlich) umstritten. Der Erfolg von Maßnahmen zur Bekämpfung oder Prävention von Desinformation könnte an mehr als einem Parameter gemessen werden; ein zweiter Parameter könnte einen demokratietheoretisch relevanten Faktor wie das Vertrauen der Menschen in Institutionen und die Demokratie abdecken. Konkret sollte der Demonstrator auch auf inklusives Design achten, z.B. dass die Farben rot und grün nicht zur inhaltlichen Differenzierung genutzt werden, um Farbenblinde auszuschließen.
Auf Basis der technischen und funktionalen Anforderungen entwickelte die Virtimo AG dann ein Technologiekonzept, um digitale Desinformationskampagnen realistisch zu simulieren.

Das IZEW unterzog dieses Konzept einer ethischen Analyse, ergänzte weitere ethische Anforderungen und brachte diese in den Prozess der Implementierung des Simulationsmoduls (**AP 3.3**) ein. So machen wir im Technologiekonzept darauf aufmerksam, dass bei der Speicherung von Daten ein besonderes Augenmerk auf Privatheit und die Sicherheit persönlicher Daten zu legen ist (privacy by design, Datenschutzstandards). Beim Simulationsmodul weisen wir darauf hin, welche Problematiken für die Nutzenden durch Simulationen als vereinfachte Darstellungen der „Wirklichkeit" entstehen und dass diese daher transparent gestaltet werden müssen. Nutzende sollten nicht das Gefühl bekommen, dass die Simulation „die Realität" abbildet bzw. abbilden kann, und ihre Handlungsentscheidungen in der Praxis unverhältnismäßig stark von Ergebnissen der Simulation abhängig machen. Wir problematisieren die Nutzung bestimmter Metaphern und Begriffe sowie den möglichen Bias der Forschenden in Bezug auf ihr Menschenbild und die Wahrnehmung verschiedener Gruppen.
Gemeinsam mit der UKÖ entwickelte das IZEW im Rahmen von Online-Treffen in **AP 3.4** ein pädagogisches Konzept zur Gestaltung der Trainingsmaterialien. Das Ziel ist, dass die BOS-Mitarbeitenden im Demonstrator-Trainingsmodul Wissen zu Falsch- und Desinformation und ihrer Bekämpfung optimal lernen und einüben können sowie zu ethischen und rechtlichen Reflexionen darüber angeregt werden. Grundsätzlich fassen wir Lernen als „konstruktiven Prozess" auf, „in dem sich Lernende neues Wissen aneignen und in individuell vorhandene kognitive Strukturen integrieren. Dabei bauen sie ihren Lernprozess auf bestehenden Wissensbestände[n] und Kompetenzen auf." (Schumacher 22.04.2008; vgl. Dubs 2010). Wichtig war für uns daher, zunächst die Zielgruppen des Trainings festzulegen, damit wir deren Wissensbestände adressieren können. Basierend auf den Expert:inneninterviews sowie den Projekt-Workshops identifizierten wir drei Zielgruppen: Erstens BOS-Mitarbeitende, die schon länger im Bereich Social Media arbeiten. Bei diesen gehen wir davon aus, dass sie vielfältige Erfahrungen in der Social-Media-Kommunikation haben und auch in der Praxis bereits mit Falschinformationen umgehen mussten. Vermutlich verfügen viele von ihnen bereits über grundlegendes Fachwissen zu Falschinformationen – es kann aber auch sein, dass gerade in kleineren Städten oder ländlichen Gegenden absichtliche Desinformation in der Arbeit dieser Mitarbeitenden noch keine Rolle spielte und daher kein spezifisches Wissen angeeignet wurde. Die zweite Zielgruppe umfassen jene Mitarbeitenden, die neu in den Bereich Social Media einsteigen oder nur sporadisch (zu Nacht- oder Wochenendzeiten) aushelfen. Uns wurde berichtet, dass gerade am Wochenende oder in starken Stressphasen bei Feuerwehren oder Polizeien oft Kolleg:innen die Social Media Accounts mitbetreuen, die ansonsten andere Aufgaben innehaben. Es ist davon auszugehen, dass diese Zielgruppe über wenig Praxiserfahrung und Fachwissen zu Falschinformationen verfügt. Zuletzt machten uns BOS in den Interviews und Workshops uns darauf aufmerksam, dass ihre Arbeit erschwert, dass ihre Vorgesetzten und Kolleg:innen oft unwissend und in Bezug auf das Thema Falschinformationen und Social-Media-Kommunikation nicht sensibilisiert sind. Einige Interviewte wünschten sich explizit, dass das Training Sensibilisierungskomponenten für Vorgesetzte und Kolleg:innen enthält.
Um die Zielgruppen ansprechend zu adressieren und sie bei ihren Wissensbeständen abzuholen, wählten wir eine modulare Herangehensweise für die Implementierung des Trainingsmoduls: Neben einführenden Texten werden weiterführende Elemente angeboten. Einzelne Themenbereiche stehen für sich, so dass es möglich ist, entsprechende Trainings zeitlich und inhaltlich flexibel zu gestalten sowie die Trainingsmaterialien in Offline- sowie Online-Trainings einzubauen. Dies beruht auf den erarbeiteten Erwartungsprofilen (AP 1.3): BOS hatten hier viele verschiedene Bedürfnisse in Bezug auf entsprechende Trainings

geäußert. Ergänzt wird die modulare Herangehensweise durch interaktive Elemente wie Verweise auf die Simulation (Trial&Error, exploratives Lernen), Reflexionsfragen für Einzel- und Gruppenarbeiten, juristische und praktische Fallbeispiele (Fallstudiendidaktik), sowie Multiple-Choice-Quizfragen.
Konkret teilten wir das Training für die verschiedenen Zielgruppen in zwei Stufen auf: Texte und Übungen der ersten Stufe dienen der Einführung und Sensibilisierung, um die Zielgruppen zwei und drei anzusprechen. Diese sind, passend für Personen mit geringem Vorwissen, didaktisch so aufbereitet, dass sie eine übersichtliche Struktur bieten (z.B. Übersichten, Merksätze und eine reduzierte, prägnante Darstellung, vgl. Kerres 2021). In der zweiten Stufe wird für Zielgruppe eins Expert:innenwissen zu den spezifischen Themen vermittelt, das mehr Quellen und Verlinkungen zu weiterführenden Texten bietet und mit Praxisbeispielen und Simulationen mehr Spielraum für die Auseinandersetzung mit den Texten bietet (vgl. ebd.). Weiterhin wurden die Trainingsmaterialien mit Hilfe der Lernzieltaxonomie nach Bloom (Bloom et al. 1956) gestaltet, um verschiedene Stufen des Lernens anzusprechen und das Wissen nachhaltiger zu vermitteln.

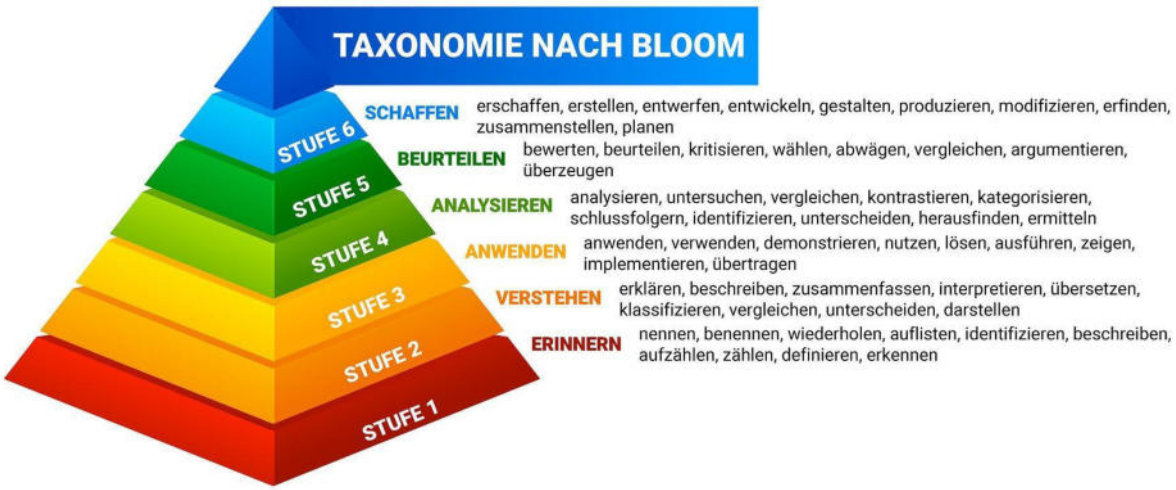

Abbildung 2: Taxonomie nach Bloom (Quelle: Goecke 2024)

Wir erweiterten die Lernzieltaxonomie um zwei Schritte. Erstens ist es notwendig, Akzeptanz zu schaffen: Wir verdeutlichen in einführenden Lerneinheiten, warum das Thema Falschinformationen wichtig ist und warum sich die Lernenden damit auseinandersetzen sollten, selbst wenn sie nicht täglich mit Falschinformationen konfrontiert sind. Zweitens muss den Lernenden überhaupt erst grundlegendes Wissen zur Verfügung gestellt werden. Dieses wird mit einführenden oder weiterführenden Texten und Bildern vermittelt.
Damit die Lernenden das vorgestellte Wissen besser memorieren, bieten wir Quizze an, die wichtige Punkte abfragen und erklären, warum vorgegebene Antworten richtig oder falsch sind. Die Quizze testen auch das Verständnis des Gelernten. In Simulationen können die Lernenden somit das Gelernte übertragen. Dadurch wird ebenfalls das Verstehen, Anwenden und Analysieren der Teilnehmenden erhöht. Daneben enthält das Trainingsmodul auch Fallbeispiele, um das Gelernte in praktischer Ebene angewendet zu sehen und es besser analysieren zu können. Ein wichtiger Teil des Trainings sind ethische Abwägungen und Bewertungen, die teilweise in den Texten erfolgen. Wir bieten aber auch Reflexionsübungen an, um Abwägungen zu üben, sich eigene Werte bewusst zu machen und sie zu priorisieren. Zuletzt können mit dem Simulator auch eigene Szenarien entwickelt werden. Für genauere Informationen zur Lernzieltaxonomie nach Bloom und deren Anwendung auf die

Trainingsmaterialien verweisen wir auf den im Projekt entwickelten Schulungsleitfaden (vgl. AP 5.2).
In Bezug auf die konkreten Trainingshalte definierten wir in Abstimmung mit der UKÖ zentrale Themenbereiche. Diese sind: Einführung in die Problematik und Relevanz des Themas, rechtliche Fragen der behördlichen Zuständigkeit, datenschutzrechtliche Grundlagen, sowie Maßnahmen gegen Falschinformationen und ihre ethisch-rechtliche Bewertung, konkret a) nutzer:innenzentrierte Maßnahmen wie eine Steigerung der Medienkompetenz, psychologische Inokulation [eine Art „Impfung“ gegen Falschinformationen], sowie Nudging [das subtile Lenken des Verhaltens in eine bestimmte Richtung], b) die interne Informationsbeschaffung, -verifikation und -priorisierung (einschließlich Social Media Monitoring), c) präventive und reaktive Krisenkommunikation (einschließlich Debunking, also Richtigstellungen), d) Vertrauens- und Communitymanagement, e) Eskalation, f) automatisierte Maßnahmen.

AP 4: Entwicklung eines ethisch-rechtlichen Rahmens für Maßnahmen zur Prävention von digitalen Desinformationskampagnen

(1) Verwendung der Zuwendung

Beschäftigungsentgelte für wissenschaftliche Mitarbeiterinnen, studentische Hilfskräfte, Reisekosten für Projekttreffen und Konferenzen, Verwaltungsausgaben (v.a. Catering für den Workshop in Stuttgart am 12.10.2023).

(2) Ziele

- AP 4.1: Ethisch-soziale Bewertung der entwickelten Präventionsmaßnahmen
- AP 4.2: Rechtliche Bewertung und Überlegungen de lege ferenda
- AP 4.3 Entwicklung eines ethisch-rechtlichen Rahmens und konkreter Empfehlungen für BOS
- Gemeinsame Publikation zu den entwickelten Maßnahmen (AP 2) und ihrer ethisch-rechtlichen Bewertung

(3) Arbeitsschritte und erzielte Ergebnisse

Die ethisch-demokratietheoretische Bewertung der Maßnahmen in **AP 4.1** erfolgte in drei Iterationen. Zunächst reflektierten wir die in der Literaturanalyse erhobenen Maßnahmen (vgl. Stieglitz et al. 2023) und ergänzten dies später um die von BOS bereits verwendeten Maßnahmen (vgl. Vortrag von Sievi am 27.10.2023, Luiss University Rome). Als zweite Iteration entwickelten wir ein Konzept für eine weitergehende ethisch-soziale Bewertung. Dabei legten wir vier Rollenverständnisse der BOS (als rechtsstaatliche sowie vertrauenswürdige Institution, als Institution mit Schutzaufgaben sowie als solche mit demokratischem Menschenbild) zugrunde, um zentrale Werte der BOS zu identifizieren. Dazu zählen Gerechtigkeit, Gleichheit, Transparenz, Freiheit, Autonomie, Partizipation, Pluralismus, Sicherheit, Zuverlässigkeit, Wahrhaftigkeit, Überparteilichkeit/Neutralität sowie informationelle Selbstbestimmung. Dann arbeiteten wir heraus, wie diese Werte für die Falschinformationsbekämpfung umgesetzt werden und welche Wertekonflikte dabei entstehen. Beispielsweise sollten BOS, wenn sie Sicherheit herstellen wollen, darauf achten, dass sie richtige bzw. gut geprüfte, vollständige und unzweideutige, sowie schnelle bzw. rechtzeitige Informationen veröffentlichen, die aber keinen Schaden anrichten sollen. Hierbei kann jedoch ein Konflikt entstehen, ob BOS Informationen lieber schnell, aber dafür weniger gut geprüft oder besser geprüft, aber dafür langsamer herausgeben sollten. Unser Konzept für eine weitergehende ethisch-soziale Maßnahmenbewertung stellten wir beim

Konsortialtreffen am 21.06.2023 vor und präsentierten es einer breiteren Fachöffentlichkeit im Rahmen eines Vortrags beim Graduierten-Netzwerk Zivile Sicherheit am 28.06.2023.
2023 führten wir die ethisch-demokratietheoretische Bewertung von Gegenmaßnahmen auf Grundlage der Forschungsliteratur und des Inputs der BOS fort. V.a. bewerteten wir die entwickelten Maßnahmen gemeinsam mit Vertretenden verschiedener BOS. Hierzu führten wir am 12.10.2023 in Stuttgart den „Workshop mit Wissenschaft und Praxis zu sicherheitsrelevanten Falschinformationen“ durch. 16 Vertretende verschiedener BOS aus Baden-Württemberg sowie dem weiteren Bundesgebiet nahmen teil. Wir stellten den Teilnehmenden die verschiedenen Maßnahmen vor und diskutierten diese gemeinsam in Gruppenarbeiten. Neben der Wirksamkeit der Maßnahmen reflektierten wir vor allem ethische Problematiken der Bekämpfung von Falschinformationen sowie der verschiedenen Maßnahmen. Darüber hinaus teilten die Teilnehmenden die Leitbilder und -werte ihrer Organisation sowie ihre Erfahrungen mit Falschinformationen und reflektierten dies ethisch.
Die Ergebnisse dieser ethischen Bewertung haben wir in einem peer-reviewten englischsprachigen Fachjournal Open Access veröffentlicht:

- Sievi, Luzia; Pawelec, Maria (2025): (How) Should security authorities counter false information on social media in crises? A democracy-theoretical and ethical reflection. In: International Journal of Disaster Risk Reduction 116, S. 1-24.

Hier erörtern wir, ob und wie BOS auf Falschinformationen reagieren sollten und mit welchen Wertkonflikten sie hierbei konfrontiert werden. Der Artikel kombiniert deskriptive und normative Ethik und konzentriert sich auf Werte wie Freiheit, Autonomie, Neutralität, Privatsphäre, Nichtdiskriminierung und Sicherheit. Wir bewerten vier Maßnahmen für BOS zur Bekämpfung von Falschinformationen und kommen zu dem Ergebnis, dass Wertkonflikte bei der Falschinformationsbekämpfung nicht pauschal beantwortet werden können. Daher schlagen wir Fragen und Überlegungen für eine ethische Reflexion vor, die sich die BOS kontextbezogen stellen sollten, wenn sie konkreten Fällen von Falschinformationen begegnen. Dazu gehört: Wer entscheidet, ob eine Information falsch oder verlässlich ist? Sind Entscheidungsträger:innen geschult und reflektieren sie ihre eigenen Vorurteile kritisch? Wird bei der Wahl der Maßnahmen darauf geachtet, ob die Intervention in einem angemessenen Verhältnis zu dem behandelten Problem steht? Welche Wertekonflikte sind damit verbunden?
Unsere Analyse ergibt zudem konkreter, dass die BOS Medienkompetenzschulungen anbieten könnten, um Bürger:innen zu lehren, wie sie falsche Informationen in sozialen Medien erkennen und damit umgehen können. Solche Schulungen befähigen Bürger:innen und stehen im Einklang mit demokratischen Werten. Neben praktischen und rechtlichen Erwägungen müssen BOS jedoch sicherstellen, dass die Schulungen unvoreingenommen sind. Darüber hinaus können Medienkompetenztrainings bestimmte Teile der Bevölkerung, die den „Mainstream“-Medien und dem Staat skeptisch gegenüberstehen, nicht erreichen oder können sogar kontraproduktiv sein, wenn sie Bürger:innen nur beibringen, bestimmten Quellen zu misstrauen, nicht aber, wie und warum sie anderen vertrauen sollen.
Die Literatur diskutiert auch die psychologische Inokulation, also eine "Impfung" gegen Falschinformationen, als potenziell wirksame Strategie, allerdings bislang nicht im Zusammenhang mit BOS. BOS könnten sie trotzdem einsetzen. Sie müssen jedoch darauf achten, dass sie dabei keine Falschinformationen verbreiten oder von den Bürger:innen als aggressiv und paternalistisch wahrgenommen werden. Das Einverständnis der Teilnehmenden, ein geringer zeitlicher Abstand zwischen Vorwarnung und tatsächlicher „Impfung" sowie die Verwendung unpolitischer fiktiver Beispiele können diesem Ziel dienen.
BOS könnten auch soziale Normen nutzen, um Nutzende subtil dazu zu bringen, weniger Falschinformationen zu glauben und zu teilen („nudging“) – eine weitere Maßnahme, die laut

Literatur und unseren Interviews für BOS neu ist. Um sich an demokratische Normen zu halten, dürfen BOS jedoch (neben praktischen Einschränkungen) nur solche Normen hervorheben, die sich die demokratische Gemeinschaft selbst in Form ihrer demokratischen Verfassung oder Gesetze gegeben hat. Sie dürfen nicht für einen bestimmten Lebensstil oder eine politische Agenda werben oder bestimmte soziale Gruppen benachteiligen.
Beim Social Media Monitoring müssen BOS über den Umfang entscheiden: Ein Fokus auf die eigene Organisation und Arbeit ist ressourcensparend. Er trägt dazu bei, den Erhalt und die Legitimität der Institution zu sichern und Verantwortlichkeits- und Datenschutzfragen zu vermeiden. Möglicherweise wird er der Sicherheitsbedrohung durch Falschinformationen in Krisen jedoch nicht gerecht. BOS können daher umfassender überwachen, sollten aber möglichst Voreingenommenheiten (bias) bei der Definition von Suchbegriffen vermeiden. Außerdem mus die Überwachung sozialer Medien verhältnismäßig und notwendig sein. BOS müssen ein präventives (öffentliches) oder eigenes Interesse daran haben, persönliche Daten zu sammeln und zu verarbeiten. Zum Schutz der Privatsphäre sollten sie technische Instrumente zur Aggregation und/oder Anonymisierung von Beiträgen in Betracht ziehen. Darüber hinaus müssen die BOS bei der Bewertung aufgedeckter Falschinformationen die sich daraus ergebenden Bedrohungen gegen die Meinungsfreiheit der Bürger:innen abwägen. Solche Bewertungen sind zwangsläufig politisch, da BOS-Mitarbeitende von ihrem eigenen Hintergrund und politischen Einstellung und der Organisationskultur der BOS beeinflusst werden. Daher müssen sie ihre eigenen Voreingenommenheiten und „blinden Flecken“ sowie die sozialen Gruppen, die sie mit ihren Urteilen benachteiligen könnten, genau untersuchen und prüfen, ob ein Eingreifen in die öffentliche Meinungsbildung wirklich gerechtfertigt ist. Strukturell könnten die BOS auch die entsprechenden Bewertungskriterien formalisieren und die Öffentlichkeit und die Wissenschaft einladen, diese zu überprüfen.
Präventive Kommunikation kann entscheidend sein, um Gerüchten und Fehlinformationen in Krisen vorzubeugen und verunsicherte und schlecht informierte Bürger:innen zu beruhigen. Darüber hinaus ist die reaktive Kommunikation, einschließlich Debunking, entscheidend, um falsche Informationen, die bereits in den sozialen Medien kursieren, zu zerstreuen. Unsere Analyse ergab, dass die deutschen BOS schnell und präzise auf eindeutige Falschmeldungen reagieren, aber von Fall zu Fall entscheiden, wenn es um Falschinformationen geht, bei denen sie unsicher sind. Darüber hinaus üben einige BOS Zurückhaltung bei politisch und ideologisch aufgeladenen Falschinformationen, z.B. im Zusammenhang mit Maßnahmen gegen Covid-19. Kritik an der eigenen Institution und Arbeit steht auch im Vordergrund, wenn deutsche BOS über Falschinformationen im Internet berichten. Sie sind jedoch geteilter Meinung darüber, ob sie die Kritik einfach hinnehmen, ohne zu reagieren, oder ob sie ihr entgegentreten.
Sowohl deontologische als auch utilitaristische philosophische Positionen unterstreichen die Pflicht der BOS, die Öffentlichkeit über Risiken in Krisen zu informieren, um Gefahren abzuwenden und die Bürger:innen zu stärken. Gleichzeitig bestehen aber Wertkonflikte in der BOS-Krisenkommunikation: Wahrhaftigkeit, Sicherheit, Privatsphäre, Gerechtigkeit, Nichtdiskriminierung und Meinungsfreiheit sind nicht konfliktfrei gemeinsam umzusetzen. BOS-Krisenkommunikation muss daher ethisch reflektiert sein. Dazu müssen BOS ihre knappen Finanz-, Personal- und Aufmerksamkeitsressourcen gerecht verteilen. Eventuell müssen sie über verschiedene Kanäle, Stile und sogar Sprachen kommunizieren, um Randgruppen zu erreichen. Gleichzeitig dürfen sie ihre allgemeinen Informations- und Schutzaufgaben nicht vernachlässigen. Darüber hinaus besteht ein Spannungsverhältnis zwischen schneller und präziser BOS-Kommunikation. Wenn eine langsame Kommunikation zur Gewährleistung der Genauigkeit aufgrund unmittelbarer Gefahr nicht mehr zu

rechtfertigen ist, sollten BOS offen und ehrlich kommunizieren und den unsicheren Status veröffentlichter Informationen klären. Auch sollte ihre Kommunikation befähigend sein und Informationen über angemessene Reaktionen der Bürger:innen enthalten. Ein weiteres Spannungsverhältnis besteht, wenn BOS z.B. aus taktischen oder rechtlichen Gründen (z.B. Datenschutz) Informationen nicht veröffentlichen dürfen, die zum Debunken falscher Informationen und zur Information der Öffentlichkeit erforderlich wären. Aus ethischer Sicht müssen BOS von Fall zu Fall abwägen, ob und inwieweit eine größere Transparenz die Bürger:innen stärken und die Sicherheit erhöhen würden und ob demgegenüber Gründe für die Zurückhaltung von Informationen überwiegen. Außerdem kann Debunking durch BOS schwerwiegende Folgen für Einzelpersonen haben. Daher sollten formelle Widerspruchsmöglichkeiten gegen solche Richtigstellungen möglich sein. BOS sollten außerdem eigene Fehler in Bezug auf die Veröffentlichung falscher Informationen oder die fälschliche Behauptung, dass andere dies getan haben, transparent machen. Schließlich müssen BOS bei Kritik an ihrer eigenen Institution darauf achten, diese nicht vorschnell oder unangemessen als falsch zu bezeichnen und unverhältnismäßig zu reagieren, denn Kritik an Regierung und Behörden ist ein Kernelement liberaler Demokratien. BOS müssen sorgfältig abwägen, ob die Kritik gerechtfertigt oder eine legitime Meinungsäußerung ist und ob die gewählte Reaktion angemessen ist.

Vertrauens- und Communitymanagement bekämpft ebenfalls Falschinformationen: BOS bauen eine ihnen wohlgesonnene Gruppe an Follower:innen in den sozialen Medien auf, die im Krisenfall ihre Informationen verbreiten und ggf. selbst Falschinformationen berichtigen. Die Veröffentlichung informativer und sachlicher Inhalte zu diesem Zweck steht im Einklang mit dem gesetzlichen Auftrag der BOS und informiert und befähigt Bürger:innen. BOS sollten jedoch Stilmittel und Inhalte vermeiden, die ihrem Ruf schaden und Vertrauen mindern könnten und der Jagd nach Likes dienen. Zudem dürfen BOS ihre Follower:innen niemals als bloßes Mittel zum Zweck (d.h. zur Bekämpfung von Falschinformationen) betrachten und müssen sich stets der Kosten bewusst sein, die Bürger:innen durch ihre Mitarbeit entstehen. BOS sollten auch bedenken, ob sie bestimmte gesellschaftliche Gruppen durch ihren Sprachgebrauch oder Stil ausschließen.

In Ap 4.1 trug das IZEW zudem zu einer **interdisziplinären Publikation** gemeinsam mit den Konsortialpartnern bei. Darin stellte das IZEW die erarbeitete Systematik möglicher Maßnahmen von BOS gegen Falschinformationen ebenso vor wie grundsätzliche ethische Überlegungen, die BOS anstellen sollten, bevor sie Maßnahmen ergreifen. Die Publikation bereitet wesentliche Projektergebnisse praxisnah für den deutschsprachigen Raum auf:

- Schewina, Kai, Pawelec, Maria, Sievi, Luzia, Rieskamp, Jonas, Duda, Michelle, Hochstrate, Eric (2024). Maßnahmen zur Bekämpfung digitaler Desinformation: Interdisziplinäre Perspektiven für Sicherheitsbehörden. SIAK Journal für Polizeiwissenschaft und polizeiliche Praxis.

Eine abschließende dritte Iteration der ethisch-sozialen Bewertung erfolgte im Rahmen der weiteren Entwicklung des ethisch-rechtlichen Rahmens. Weitere Maßnahmengruppen wurden einer ethisch-sozialen Bewertung unterzogen, darunter Maßnahmen der Eskalation wie das Löschen und Blockieren von User:innen und Accounts sowie automatisierte Maßnahmen wie das KI-unterstützte Social Media Monitoring oder die Nutzung von Chatbots durch BOS zur Interaktion mit Bürger:innen. Zentrale Erkenntnisse dieser Bewertung sind, dass Löschen und Blockieren äußerst eingriffsintensive Maßnahmen sind, die einer besonderen Rechtfertigung bedürfen und sowohl die öffentliche Meinungsbildung verzerren als auch gravierende Auswirkungen auf die Betroffenen haben können. Sie sollten daher nur vorsichtig eingesetzt werden. Darüber hinaus unterliegen BOS praktischen Einschränkungen: Die meisten eskalierenden Maßnahmen können nur die Betreiber der Social Media-

Plattformen durchsetzen. BOS können höchstens auf ihren eigenen Kanälen löschen und blockieren. Diese praktischen Einschränkungen und ethischen (sowie rechtlichen) Bedenken verweisen auf die geringe Wirksamkeit und auch Wünschbarkeit derart eskalierender Eingriffe. BOS können darüber hinaus KI nutzen, um auf die wachsende Flut an Falschinformationen zu reagieren. Insbesondere das Social Media Monitoring kann durch KI unterstützt werden. Die automatisierte Erkennung von Falschinformationen ist jedoch niemals perfekt; es gibt stets falsch-positive und falsch-negative Meldungen. Die zugrundeliegende KI kann zudem verzerrt sein („bias“) und beispielsweise blinde Flecken gegenüber bestimmten politischen Meinungen aufweisen. KI kann auch nicht zwischen Falschinformationen und Meinungsäußerungen unterscheiden und erkennt Satire nicht. Darüber hinaus kann KI keine Werteabwägungen beispielsweise zwischen der Meinungsfreiheit und Sicherheit treffen. Die Folgen einer Kennzeichnung von Posts als falsch durch BOS und etwaiger Reaktionen der Behörden können für einzelne Betroffene enorm sein. Zentral ist daher die menschliche Letztentscheidung: KI kann nur ein unterstützendes und vorab filterndes Instrument sein. Die letztendliche Entscheidung und Verantwortung liegen beim Menschen. Chatbots zur Interaktion mit Bürger:innen und Social Bots zur Verbreitung und Personalisierung von BOS-Botschaften müssen darüber hinaus klar gekennzeichnet werden (Transparenz). Doch auch gekennzeichnete Bots können BOS schaden, wenn Bürger:innen das Gefühl haben, von KI „abgewimmelt werden“ und nicht mehr zu echten Menschen in den Behörden vordringen. Das kann dem Vertrauen in BOS schaden und verunsichert Menschen v.a. in Krisenlagen zusätzlich. Darüber hinaus können Large Language Models „halluzinieren“; es besteht also die Gefahr, dass BOS-Bots Falschinformationen verbreiten. Zudem kann der Arbeitsaufwand für BOS sogar erhöht werden, wenn Bots eine menschliche Reaktion der Behörden versprechen. Nicht zuletzt ist die Robustheit und Cybersicherheit der Systeme entscheidend: Würde beispielsweise ein BOS-Bot gehackt, könnten über als vertrauenswürdig eingeschätzte Kanäle Falschinformationen oder andere schädigende Inhalte verbreitet werden. Dies würde dem Vertrauen in BOS, der öffentlichen Sicherheit und der Demokratie massiv schaden.

Das IZEW erarbeitete unter Federführung der UKÖ zudem einen ethisch-rechtlichen Rahmen (**AP 4.3**), der bei der Entwicklung und Evaluation der BOS-Gegenmaßnahmen sowie der Demonstrators maßgeblich war. Als erste Grundlage dieses ethisch-rechtlichen Rahmens diente die Erhebung der verschiedenen Rollenverständnisse und damit einhergehenden Werte der BOS im Rahmen der Erwartungsprofile sowie des Workshops in Stuttgart im Oktober 2023. Die Arbeit am ethisch-rechtlichen Rahmen wurde 2024-2025 dann gemeinsam im Rahmen der gemeinsamen Arbeit mit der UKÖ an Lehr- und Lernmaterialien und dem damit einhergehenden Austausch über ethisch wünschenswerte und rechtlich zulässige Maßnahmen von BOS gegen Falschinformationen fortgeführt.

Um die entsprechenden Erkenntnisse über die Fachöffentlichkeiten hinaus auch einem breiteren, praxisnahen Publikum zugänglich zu machen, beschlossen das IZEW und die UKÖ, die Ergebnisse der ethisch-rechtlichen Bewertung über den Demonstrator und bestehende wissenschaftliche Publikationen hinausgehend in Form von Lehr-/Lernmaterialien deutschen BOS und weiteren Stakeholdern auch in einer didaktisch aufbereiteten Handreichung zur Verfügung zu stellen. Dies erhöht den praktischen Impact der Forschungsarbeit und stellt die Projektergebnisse in einer dauerhaft zugänglichen Form interessierten BOS und weiteren Stakeholdern zur Verfügung. Die Kerninhalte und -elemente der Handreichung entsprechen dem vom IZEW und der UKÖ entwickelten pädagogischen Konzept und den Schulungsinhalten des Demonstrators (vgl. auch AP 3.4). Diese wurden mit weiteren Quellen und weiterführender Literatur angereichert und grafisch ansprechend aufbereitet. Es folgte

eine Open Access Publikation im Rahmen der IZEW-Reihe „Materialien zur Ethik in den Wissenschaften“, um die Auffindbarkeit der Publikation zu erhöhen. Sie wurde in den Social Media-Kanälen der Projektpartner sowie auf der PREVENT-Webseite beworben und im Anschluss an die Abschlussveranstaltung des Projekts am 8. April 2025 an zahlreiche BOS-Mitarbeitende disseminiert. Über Kontakte wurde sie zudem im Intranet der Polizei NRW beworben sowie über die Bund-Länder-Arbeitsgruppe Hybride Bedrohungen den Innenressorts aller Länder, dem Bundesministerium der Verteidigung und den Nachrichtendiensten des Bundes zugesandt.

- ➢ Pawelec, Maria; Duda, Michelle und Sievi, Luzia (2025): Rechtssicher und ethisch reflektiert auf Falschinformationen reagieren. Eine Handreichung für Behörden und Organisationen mit Sicherheitsaufgaben. Tübingen: IZEW, Materialien zur Ethik in den Wissenschaften, Band 26. ISBN: 978-3-935933-23-0.

AP 5: Evaluation des Demonstrators, der vorbeugenden Maßnahmen und des ethisch-rechtlichen Rahmens

(1) Verwendung der Zuwendung

Beschäftigungsentgelte für wissenschaftliche Mitarbeiterinnen, Studentische Hilfskräfte, Reisekosten für Projekttreffen und Konferenzen, Verwaltungsausgaben (v.a. Catering für den Workshop in Stuttgart am 18.07.2024).

(2) Ziele

- AP 5.1: Evaluation des Trainingsansatzes
- AP 5.2: Evaluation des ethisch-rechtlichen Rahmens, Erarbeitung eines Konzepts zur Verstetigung der Projektergebnisse und eines Schulungsleitfadens

(3) Arbeitsschritte und erzielte Ergebnisse

Zur Evaluation des Trainingsansatzes (**AP 5.1**) nutzten wir in der ersten Iteration die Reflexionen und Diskussionen der baden-württembergischen BOS in unserem Projektworkshop in Stuttgart im Oktober 2023, um zu überprüfen, inwiefern die von uns im Projekt entwickelten und vermittelten Inhalte dem Wissensstand und -bedarf der anwesenden BOS entsprachen. Zudem formulierte das IZEW für den Simulator und den Trainingsansatz, wie ihn die VAG bis 2023 umgesetzt hatte, schriftliches Feedback. Hierbei achteten wir insbesondere auf die Nutzerfreundlichkeit und den Aufbau des Simulators.

Die im Projekt erarbeiteten, zu vermittelnden ethisch-rechtlichen Inhalte sowie technischen und inhaltlichen Funktionalitäten des Demonstrators wurden im Rahmen eines Workshops in Stuttgart am 18.07.2024 mit BOS dann in einer zweiten Iteration evaluiert. Teilnehmende waren u.a. Mitarbeitende von Polizeibehörden, Feuerwehren, der Bundesanstalt Technisches Hilfswerk, der Malteser, der Johanniter Unfall-Hilfe, des Kompetenzzentrum gegen Rechtsextremismus in Baden-Württemberg (konex) sowie des EU-Forschungsprojekts VIGILANT. Inhalte des unter der Leitung des IZEW interdisziplinär gestalteten Evaluationsworkshops waren eine juristisch-ethische Arbeitseinheit zur Reaktion auf Falschinformation und Abgrenzung zur Meinungsäußerung (einschließlich der Arbeit an juristischen Fallbeispielen), eine Arbeitseinheit zu möglichen Gegenmaßnahmen und ihrer ethischen Betrachtung, eine ethisch-juristische Arbeitseinheit zum zentralen Thema Krisenkommunikation sowie eine Vorstellung und Erprobung des Simulationstools. Eine zentrale Frage beim Workshop war auch die Umsetzbarkeit des ethisch-rechtlichen Rahmens in Bezug auf die Lernzieltaxonomie nach Bloom (s. AP 5.2).

In einer dritten Iteration der Evaluation des Trainingsansatzes wurden die Akzeptanz und das Bürger:innenvertrauen in die entwickelten vorbeugenden Maßnahmen Bestandteil einer Umfrage, die von der UB durchgeführt und vom IZEW begleitet wurde. Die Analyse zeigt, dass Befragte die Social-Media-Reaktionen von BOS als sachlich, klar, professionell und korrekt

wahrnehmen. Dennoch bestehen deutliche Potenziale zur Optimierung: Viele Befragte empfinden den Kommunikationsstil als zu formell, distanziert oder wenig empathisch – insbesondere im Umgang mit emotionalisierten oder verschwörungsgläubigen Zielgruppen. Kritisiert werden zudem mitunter eine fehlende Tiefe und eine geringe strategische Wirkung der BOS-Kommunikation. Eine stärker bürgerorientierte, emotional intelligentere und medienwirksamere Ansprache scheint somit erforderlich, um den Herausforderungen durch Desinformation wirkungsvoll zu begegnen.
Das IZEW unterstützte darüber hinaus beratend mit ethischer Expertise eine experimentelle Evaluation der UP des im Projekt entwickelten Chatbots sowie zur Effektivität von BOS-Debunking. Bei einem Online-Experiment mussten die Teilnehmenden entscheiden, ob sie ihnen präsentierte Informationen für wahr oder falsch hielten. Eine Gruppe sollte mit dem Chatbot interagieren, eine andere nicht. Zudem waren für weitere Gruppen die falschen Informationen mit Kontext (Debunking) durch BOS bzw. Community Notes (also Kommentaren der Community) versehen. Die Ergebnisse deuten auf die Wirksamkeit des Chatbots sowie von Debunking durch BOS sowie Community Notes hin. Das IZEW ergänzte hierbei eine ethische Befragung, die die Akzeptanz von Debunking durch BOS sowie die Community erfragt und zu dem Schluss kommt, dass beides überwiegend akzeptiert wird, jedoch auch hier die Community höhere Akzeptanz erfährt. Weitere Informationen können dem Abschlussbericht der Universität Potsdam entnommen werden.
Eine erste Evaluation (der Umsetzbarkeit) des ethisch-rechtlichen Rahmens (**AP 5.2**) erfolgte beim Workshop im Oktober 2023, um die Akzeptanz und das Vertrauen der entwickelten Maßnahmen zu überprüfen. Hier wurden die von uns aus der Literatur und den Interviews heraus entwickelten Rollenverständnisse und zentralen Werte von BOS mit den anwesenden BOS diskutiert, reflektiert und ergänzt und somit mit der Praxis abgeglichen.
Die zweite Iteration der Evaluation des ethisch-rechtlichen Rahmens einschließlich der Umsetzbarkeit der Empfehlungen erfolgte bei einem Online-Workshop bei der Fachtagung Katastrophenvorsorge am 23.04.2024 mit Nichtregierungsorganisationen und BOS. Diesen veranstaltete das IZEW zusätzlich zu den geplanten Projektworkshops mit dem Ziel einer noch größeren und bundesweiten Fundierung der Evaluation des ethisch-rechtlichen Rahmens. Dabei gaben die Teilnehmenden Feedback aus der Praxis zu den erarbeiteten Maßnahmen und unserer ethischen Bewertung. Auf einem interaktiven Online-Whiteboard gaben sie u.a. an, inwiefern ihre Organisation die Maßnahmen bereits nutzt und ob ihre eigene Einschätzung der Maßnahme von unserer abweicht. Die Impulse aus der Praxis flossen wiederum in die Demonstratorinhalte (AP 3.4) sowie die ethisch-soziale Bewertung und den ethisch-rechtlichen Rahmen (AP 4.1, 4.3) ein. Dazu gehört etwa die Überlegung, wie ältere Menschen durch BOS-Medienkompetenztrainings erreicht werden könnten, oder zur Unterscheidung zwischen konstantem und anlassbezogenem Social Media-Monitoring.
Darüber hinaus holte das IZEW im Anschluss an den Workshop im Juli 2024 in Stuttgart Feedback der anwesenden BOS mit Blick auf die Lernzieltaxonomie nach Bloom ein. Kernfragen betrafen die Verständlichkeit der Schulungsinhalte (Stufe 2 der Taxonomie), die Bedeutung des Simulationsmoduls und weiterer interaktiver Elemente der Schulung (Stufen 3-5) und eine mögliche Verstetigung der Projektergebnisse. Die Auswertung dieses Feedbacks stellt die dritte Iteration der Evaluation des ethisch-rechtlichen Rahmens dar. Sie floss wiederum in die Ausgestaltung der Schulung, der Lehr-/Lerninhalte (siehe AP 3.4) und des ethisch-rechtlichen Rahmens (AP 4.3) ein. Das Feedback zeigte, dass fast alle Teilnehmenden die Schulungsinhalte sowie den Aufbau der Schulung verständlich fanden. 90% waren der Meinung, dass die geplante Schulung alle für sie relevanten Inhalte in Bezug auf Falschinformationen abdeckt. 70% der Befragten sahen einen „großen Mehrwert“ im

entwickelten Demonstrator. 88% begrüßten dessen Verstetigung. Allerdings sahen nur 37,5% ihre Institution in der Lage dies organisatorisch zu unterstützen und nur 12,5% finanziell. Alle Befragten sahen zudem einen Mehrwert in einer Schulung zum Thema Falschinformationen für ihre Organisation, und 90% in der geplanten Handreichung. Bezüglich der konkreten Darbietung einer entsprechenden Schulung gingen die Meinungen auseinander: 72% begrüßten einen Self-Study-Onlinekurs (ja/eher ja), während 28% dies eher ablehnten. 90% wünschten sich eine Vor-Ort-Schulung zum Thema.
In AP 5.2 entwickelte das IZEW zudem ein Konzept zur Verstetigung der Projektergebnisse von PREVENT nach Projektende. Es eruiert Optionen, den entwickelten Demonstrator weiter zu betreiben und Schulungen für BOS anzubieten. Der Betrieb des Demonstrators könnte durch Partnerschaften mit Ministerien, BOS oder Universitäten sowie durch Open-Source-Veröffentlichungen finanziert werden. Schulungen sollten sowohl als Präsenz-Workshops als auch online verfügbar sein, und ein Gebührenmodell oder Crowdfunding könnten deren Finanzierung sichern. Die Ergebnisse wurden mittels Open Access-Publikationen, praxisnahen Handreichungen und Social-Media-Kampagnen verbreitet und gelangen so in die Praxis. Regelmäßige Evaluierungen und Aktualisierungen der Inhalte wären aufgrund der schnellen Entwicklungen im Bereich der Desinformation erforderlich, lassen sich jedoch ohne weitergehende Projektfinanzierung nicht umsetzen. Insgesamt basiert die Verstetigungsstrategie auf einer vielfältigen Mischung von Maßnahmen, die wissenschaftliche Publikationen, praktische Anwendungen und Trainingsmaterialien kombinieren.
Des Weiteren entwickelte das IZEW in AP 5.2 einen Schulungsleitfaden, der erläutert, wie die im Projekt entwickelten Lehr-/Lernmaterialien, die sowohl im Demonstrator implementiert als auch als Handreichung verschriftlicht wurden, in verschiedenen Formen didaktisch aufbereitet werden und in Schulungen einfließen können (vgl. AP 3.4).

2 Wichtigste Positionen des zahlenmäßigen Nachweises

Mit Verweis auf den detaillierten zahlenmäßigen Nachweis, der dem Projektträger vorliegt, werden im Folgenden ausgewählte Positionen dargestellt.

Position 0812 (Beschäftigte E12-E15)

Aus Position 0812 wurden die Beschäftigungsentgelte für die wissenschaftlichen Angestellten gemäß TV-L E13 gezahlt. Die genaue Aufstellung der Posten ist im zahlenmäßigen Nachweis aufgelistet. Projektmitarbeitende waren: Vondermaßen, Dr. Marcel (01.01.2022-01.04.2022), Sievi, Dr. Luzia (01.01.2022–31.10.2024), Pawelec, Maria (30.06.2022- 30.04.2025).

Position 0822 (sonstige Beschäftigungsentgelte)

Aus Position 0822 wurden die Beschäftigungsentgelte für studentische und wissenschaftliche Hilfskräfte gezahlt, welche die Arbeiten innerhalb des Projekts unterstützten. Dies waren in wechselnder Besetzung und Stundenanzahl: Maria Staecker (stud. Hilfskraft), Amelie Seifert (stud. Hilfskraft), Jan-David Bühler (stud. Hilfskraft), Deborah Methner (stud. Hilfskraft), Kristina Janackova (BA), Martin Bux (BA) und Kim Lemke (BA).

Position 0840 (Ausgaben für den Kauf von Literatur)

Es wurde auf die Anschaffung von Literatur verzichtet, um Catering für die Workshops in Stuttgart vom 12.10.2023 und 18.07.2024 zu ermöglichen. Hierzu wurden die Mittel entsprechend umgewidmet.

Position 0844 und 0845 (Dienstreisen)

Aus Position 0846 wurden die Kosten für Reisen zu Interviews, Projekttreffen, Konferenzen und Workshops gezahlt:

- Projektworkshop in Paderborn, 08.11.2022
- Interviews in Paderborn und Müllheim an der Ruhr
- 08.09.2022 Vortrag von Dr. Luzia Sievi auf der Mancept Konferenz: „The Relationship of Right-Wing Discourse to Science Is Not Antagonistic, It Is Hybrid.", Manchester, Großbritannien.
- Projektworkshop in Köln, 02.03.2023
- Projektworkshop in Potsdam, 28.06.2023
- Projektworkshop in Stuttgart, 12.10.2023
- 27.10.2023: Vortrag Luzia Sievi auf der Konferenz "The democratic containment of fake news and bad beliefs": “How Security Organisations and Authorities Combat False Information on Social Media and What This Means for Democracy – A Case Study of Germany”, Luiss University, Rom, Italien.
- Workshop in Köln, 14.02.2024
- Workshop in Stuttgart, 18.07.2024
- Abschlussveranstaltung in Potsdam, 08.04.2025

3 Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die Bedeutung von Falschinformationen in den sozialen Medien ist enorm und steigt aktuell weiter durch technische Entwicklungen v.a. im Bereich generativer KI, welche die Erstellung gezielter Desinformation weiter erleichtern und neue technische Möglichkeiten etwa der Fälschung audio-visueller Medien (Deepfakes) bieten. BOS stehen daher vor enormen Herausforderungen, die in der Forschung bisher unzureichend untersucht wurden. Es fehlt an einer wissenschaftlichen Auseinandersetzung mit der besonderen Rolle und Verantwortung von BOS bei der Bekämpfung von Falschinformationen, an ethisch-demokratietheorischen Analysen möglicher Gegenmaßnahmen und an praxisnahen Empfehlungen für BOS, insbesondere in Deutschland. PREVENT setzte an diesen Forschungslücken an und trug durch die interdisziplinäre Zusammenarbeit entschieden zur Erweiterung des wissenschaftlichen Kenntnisstands bei. Veröffentlichungen für verschiedene Zielgruppen und Vorträge in verschiedenen Kontexten gaben der Forschungslandschaft wichtige Impulse. Das IZEW war zudem in der Wissenschaftskommunikation aktiv und gab zahlreiche Medieninterviews. Impulse für die Praxis gab PREVENT zudem praxisnahe wissenschaftliche Veröffentlichungen, die Entwicklung des Demonstrators, die Erstellung umfassender rechtlich und ethisch fundierter Schulungsmaterialien und deren Verbreitung in Form einer Handreichung. Dadurch stärkte PREVENT deutsche BOS bei der alltäglichen Social Media-Arbeit ebenso wie in besonderen Einsatzlagen. Insbesondere die zwei in Stuttgart vom IZEW veranstalteten Workshops zeigten, dass die Projektarbeit auf großes Interesse und Resonanz von Seiten der BOS stieß. Wir konnten beim ersten Workshop über 20 Teilnehmende aus mehr als zehn verschiedenen BOS aus ganz Baden-Württemberg gewinnen. Die Resonanz war sehr positiv und das Interesse so groß, dass der THW Landesverband seine Räumlichkeiten für weitere Workshops anbot. Auf dieses Angebot kamen wir beim zweiten Workshop im Juli 2024 dankend zurück. Hieran nahmen 16 Teilnehmende aus verschiedenen BOS bundesweit teil. Hier eingeholtes Feedback verdeutlichte das große Interesse der BOS am Thema Falschinformationen, an entsprechenden Schulungen sowie an den im Projekt entwickelten Materialien (vgl. AP 5.2). PREVENT entspricht einem zivilgesellschaftlichen Interesse an der Stärkung der Demokratie und der Sicherheit in Deutschland. Daher war eine Förderung, die die Unabhängigkeit der Forschung gewährleistet, unumgänglich.

4 Voraussichtlicher Nutzen und Verwertbarkeit des Ergebnisses

Die Projektergebnisse sind durch Veröffentlichungen und sonstige Formen der Dissemination dauerhaft verfügbar und stellen wichtige Wissensbestände für zukünftige Forschungen und die zivilgesellschaftliche Debatte dar. Die Beiträge in Fachzeitschriften, auf Konferenzen und in Interviews zeigen auf, wieso Falschinformationen die öffentliche Meinungsbildung gefährden und insbesondere in Krisensituationen die Sicherheit bedrohen. Die praxisnahe Handreichung enthält ethisch und rechtlich fundierte Bewertungen einer Vielzahl an möglichen Gegenmaßnahmen von BOS gegen Falschinformationen sowie didaktisch

aufbereitete Lehr-/Lernmaterialien und kann BOS als wichtige Orientierung beim zukünftigen Umgang mit Falschinformationen dienen. Es sind zudem verschiedene Optionen der weiteren Verstetigung der Projektergebnisse denkbar. Diesbezüglich wird auf das Konzept zur Verstetigung der Projektergebnisse verwiesen (vgl. AP 5.2).

5 Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Im Laufe des Projekts entstanden Kontakte zu PREVENCY, einem Unternehmen, das Simulations- und Schulungssoftware zur Krisenbewältigung und -verhinderung entwickelt. Mitarbeitende werden in Schulungen dabei z.B. einem „Shitstorm“ gegen die eigene Organisation ausgesetzt und lernen, angemessen zu reagieren. Es bestehen Überschneidungen zum PREVENT-Demonstrator; PREVENCY konzentriert sich allerdings nicht auf Falschinformationen und nicht ausschließlich auf BOS; die Simulation ist nicht empirisch und wissenschaftlich fundiert. PREVENCY wurde im Verlauf des Projekts als assoziierter Partner eingebunden.

Darüber hinaus wurde ein Austausch mit dem EU-geförderten Projekt VIGILANT („Vital Intelligence to Investigate Illegal Disinformation“) etabliert. Der Austausch erfolgte im Rahmen eines Online-Arbeitstreffens sowie der Teilnahme eines Mitarbeitenden der Deutschen Hochschule der Polizei (DHPol) an dem Workshop mit BOS im Juli 2024 in Stuttgart sowie an der Abschlussveranstaltung im April 2025 in Potsdam. Die in PREVENT zu erstellende Handreichung wurde VIGILANT darüber hinaus zur Verfügung gestellt. VIGILANT unterstützt Polizeibehörden dabei, Desinformation im Internet besser zu identifizieren, nachzuverfolgen und entsprechend zu ermitteln. Dazu wird eine auf KI-Methoden basierte Software entwickelt, die eine Identifizierung und Analyse von Desinformationen erlaubt. Die DHPol konzipiert und führt Schulungen für europaweite Polizeibehörden in der Verwendung der Plattform durch. Es bestehen somit inhaltliche Überschneidungen zum Projekt PREVENT. Der Fokus von VIGILANT ist jedoch thematisch enger (Fokus auf Polizeibehörden und keine weiteren BOS) und geografisch weiter (europaweit) als der von PREVENT. VIGILANT konzentriert sich darüber hinaus auf eine technische Gegenmaßnahme zu Falschinformationen und nicht auf das breite Spektrum möglicher Gegenmaßnahmen und ihre vertiefte ethische (und rechtliche) Bewertung. Darüber hinaus sind keine einschlägigen Publikationen oder Projekte bekannt, die sich aktuell mit der Bekämpfung von Falschinformationen durch deutsche BOS befassen.

6 Erfolgte und geplante Veröffentlichungen der Ergebnisse

Wissenschaftliche Veröffentlichungen

Stieglitz, Stefan; Fromm, Jennifer; Kocur, Alexander; Rostalski, Frauke; Duda, Michelle; Evans, Alison; Rieskamp, Jonas; Sievi, Luzia; Pawelec, Maria; Heesen, Jessica; Loh, Wulf; Fuchß, Christopher; Eyilmez, Kaan (2025): What Measures Can Government Institutions in Germany Take Against Digital Disinformation? A Systematic Literature Review and Ethical-Legal Discussion, in: Daniel Beverungen, Christiane Lehrer und Matthias Trier (Hrsg.): Transforming the Digitally Sustainable Enterprise. Cham: Springer Nature, S. 319-337 (zuvor Preprint als Konferenzbeitrag: 18. Internationale Tagung Wirtschaftsinformatik 2023)

Pawelec, Maria, & Sievi, Luzia (2023): “Falschinformationen in den sozialen Medien als Herausforderung für deutsche Sicherheitsbehörden und -organisationen“. Kriminologie - Das Online-Journal | Criminology - The Online Journal, 4(5), 316–347. https://doi.org/10.18716/ojs/krimoj/2023.4.7

Schewina, Kai, Pawelec, Maria, Sievi, Luzia, Rieskamp, Jonas, Duda, Michelle, Hochstrate, Eric (2024). Maßnahmen zur Bekämpfung digitaler Desinformation: Interdisziplinäre Perspektiven für Sicherheitsbehörden. SIAK Journal für Polizeiwissenschaft und polizeiliche Praxis (4), S. 15-29.

Sievi, Luzia; Pawelec, Maria (2025): (How) Should security authorities counter false information on social media in crises? A democracy-theoretical and ethical reflection. In: International Journal of Disaster Risk Reduction 116, S. 1-24.

Pawelec, Maria; Sievi, Luzia (im Erscheinen): Zum Umgang mit Falschinformationen in den sozialen Medien: Ethische und demokratietheoretische Reflexionen möglicher Gegenmaßnahmen von Sicherheitsbehörden. In: SIAK-Journal - Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis (2/2025).

Beiträge in Publikumsmedien und Formaten der Zivilgesellschaft

- Vondermaßen, Marcel: Artikel im Schwäbischen Tagblatt „Schutz ohne Bevormundung“, 19.04.2022.
- Pawelec, Maria (2022): What are deepfakes and why do they matter? Hintergrundartikel für das European-Israeli Forum for Technology and Society, eine Zusammenarbeit zwischen dem Israel Public Policy Institute (IPPI) und der Heinrich Böll Stiftung.
- Pawelec, Maria (2022): Was sind Deepfakes? Newsletterbeitrag. In: in.media.res: die Kreativregion Stuttgart.
- Pawelec, Maria: Deepfakes: Wo liegen Risiken, aber auch Chancen für unsere Gesellschaft? Beitrag von Maximilian Brose für die Sendung "Systemfragen", 1. Juni 2023 (mehr Informationen zum Beitrag hier).
- Pawelec, Maria: Du als Klon! Betrug mit deiner Stimme, Beitrag von Tasnim Rödder im Auftrag des SWR für die Sendung "VOLLBILD - Recherchen, die mehr zeigen", 18.07.2023.
- Pawelec, Maria: Deepfake Music? Wie die Technologie Einfluss auf die Musik nehmen kann. 4-teilige Mini-Serie von Kathrin Schunn, Hochschule für Musik Karlsruhe, veröffentlicht auf Junger Kulturkanal, 03.09.2023.
- Pawelec, Maria: Deepfakes – Mediale Glaubwürdigkeit in Gefahr. Interview mit Sebastian Wellendorf in der Sendung "mediasres", 21.09.2023.
- Pawelec, Maria: Gefahren von Deepfakes für die mediale Glaubwürdigkeit. Interview in Radio Bremen Zwei, 22.09.2023.
- Pawelec, Maria: Mal angenommen, künstliche Intelligenz ersetzt uns. Was dann? tagesschau Zukunfts-Podcast "mal angenommen" von Kristin Becker und Gábor Halász, 05.10.2023.
- Pawelec, Maria: Pizza beim Parteitag der Grünen: So bewerten Experten den KI-Fake vom Grünen-Parteitag. Beitrag von Sascha Maier in den Stuttgarter Nachrichten, 30.11.2023.
- Pawelec, Maria (2024): Deepfakes: Auf einen Blick, Anwendungsbereiche von Deepfakes, Politische Manipulation und Desinformation, „Softfakes“ in Wahlkämpfen – Ein Ausblick sowie Chancen für die Demokratie. Kapitel in Dossier der Bundeszentrale für politische Bildung (bpb) “Wenn der Schein trügt – Deepfakes und die politische Realität”.
- Pawelec, Maria: Wir generieren uns eine eigene Realität. Beitrag von Olaf Pollaske im CHIP Magazin 01/2024, 05.01.2024.
- Pawelec, Maria: Erste Parteien machen in Sachsen Politik mit künstlicher Intelligenz - Experten warnen. Beitrag von Jim Kerzig, Freie Presse Sachsen, 22.01.2024.
- Pawelec, Maria: «Superwahljahr 2024»: Open AI sagt, Chat-GPT könne nicht für politische Kampagnen eingesetzt werden. Diese Recherche zeigt das Gegenteil. Beitrag von Gioia da Silva, NZZ, 14.02.2024.
- Pawelec, Maria: „Unmöglich, sich zu schützen“: Erschreckend viele Frauen werden Opfer von Deepfake-Pornografie. Beitrag von Anika Zuschke, Frankfurter Rundschau, 02.04.2024.
- Pawelec, Maria (2025): Hintergrund Wenn der Schein trügt: Deepfakes und Wahlen. Hintergrundtext zu Unterrichtsmaterialien der Bundeszentrale für politische Bildung zum Thema Deepfakes und Wahlen (bpb).
- Pawelec, Maria:“Über Geschichte aufklären können: Deepfakes als Chance für die politische-historische Bildung”, Beitrag von Sam Howe in “Campus und Karriere”, Deutschlandfunk, 27.01.2025.
- Pawelec, Maria:“Satire mit KI-Fakes: Zwischen Humor und Falschinformation”, Beitrag von Sophie Rohrmeier, BR24 Faktenfuchs, 15.02.2025.

Vorträge

- 08.09.2022: Vortrag Luzia Sievi: "The Relationship of Right-Wing Discourse to Science Is Not Antagonistic, It Is Hybrid" im Panel "Misinformation, Expertise and Challenges to Democracy", Mancept Konferenz Manchester.
- 17.11.2022: Geladene Keynote Maria Pawelec (mit Céline Gressel) „Ethical implications of immersive technologies and deepfakes in art and culture“ beim Workshop „Immersive Technologien in Kunst und Kultur“, Goethe Institut in Taschkent/Usbekistan (virtuelle Zuschaltung).
- 23.04.2023: Workshop von Maria Pawelec und Luzia Sievi: „(Wie) Sollten Behörden und Organisationen mit Sicherheitsaufgaben (BOS) gegen Falschinformationen in den sozialen Medien umgehen? Ethische und demokratietheoretische Perspektiven“, Fachtagung Katastrophenvorsorge 2024 (virtuelle Zuschaltung).
- 15.05.2023: Diskussion mit Maria Pawelec zum Thema Deepfakes im Seminar „Technology and Media Ethics“, Prof. Christoph Bieber, Hochschule der Medien Stuttgart.
- 24.05.2023: Vortrag Maria Pawelec: „Ethical deepfakes? The values held by deepfake developers and service providers, and their governance potential“, CEPE 2023 (International Conference on Computer Ethics: Philosophical Enquiry).
- 28.06.2023: Vortrag Luzia Sievi und Maria Pawelec beim Graduierten-Netzwerk Zivile Sicherheit: „Maßnahmen von BOS gegen sicherheitsrelevante Falschinformationen und ihre ethische Bewertung“.
- 06.07.2023: Paneldiskussion mit Maria Pawelec: „KI im Gespräch: Deepfakes – Risks and Opportunities“, StudiNight-Diskussionsveranstaltung im KI-Makerspace in Tübingen.
- 27.10.2023: Vortrag Luzia Sievi auf der Konferenz "The democratic containment of fake news and bad beliefs" an der Luiss University in Rom: “How Security Organisations and Authorities Combat False Information on Social Media and What This Means for Democracy – A Case Study of Germany”.
- 08.11.2023: Fish-Bowl-Diskussion mit Luzia Sievi “Wie resilient ist unsere Gesellschaft? Rechtsextremismus als Gefahr für unsere Demokratie”, Auftaktveranstaltung Science Innovation Days der Universität Tübingen.
- Kurzinterview mit Maria Pawelec auf Instagram: "Was ist das wichtigste Sicherheitsproblem in Ihrem Bereich? Was tun?", April 2024.

- 20.06.2024: Vortrag Luzia Sievi und Maria Pawelec: „Should security authorities counter false information on social media in ermergencies, disasters, and catastrophes?“ (beim International Congress on Disaster Ethics, Oviedo, Spanien (virtuelle Zuschaltung).
- 08.10.2024: Geladener Vortrag Maria Pawelec: „‘Olympics has fallen‘? Chancen und Gefahren von Deepfakes für Politik, Gesellschaft und Sport“, Wissenschaftsforum des Württembergischen Landessportbund e.V. (WLSB), Stuttgart.
- 29.11.2024: Geladener Vortrag Maria Pawelec: „Vertrauen in Zeiten der Desinformation? Demokratische Narrative angesichts von KI und technischem Fortschritt“, Akademie der Diözese Rottenburg-Stuttgart.
- 12.03.2025: Geladene Keynote Maria Pawelec: „Gefahren und Potenziale audio-visueller generativer KI für Politik und Gesellschaft“ bei den 47. Stuttgarter Tagen der Medienpädagogik, Stuttgart.
- 14.03.2025: Geladener Vortrag Maria Pawelec: „Chancen, Risiken und Gefahren für die Demokratie und das Gemeinwesen“, Symposium “Deepfakes und das Recht” des Institut für Urheber- und Medienrecht, München.

Weitere Disseminationsaktivitäten

- Kontinuierlich: LinkedIn-Beiträge zum Projekt von Maria Pawelec und dem IZEW
- 26.01.2022 Pressemitteilung zum Projektstart der Universität Tübingen: „Desinformation erkennen und bekämpfen: Projekt unter Beteiligung von Tübinger Medienethikerin untersucht Fake News“
- 07.02.2022 Pressemitteilung zum Projektstart am IZEW: „Desinformation erkennen und bekämpfen: Projekt untersucht Fake News“
- Pawelec, Maria (2023): Teilnahme am Wissenschaftskommunikationsproject "I'm a Scientist" zur Interaktion mit Schüler:innen, organisiert von "Wissenschaft im Dialog", Themenrunde "KI im Film", Januar 2023.
- Pawelec, Maria (2024): Interviewbeiträge zu interaktiver Plattform zu Deepfakes, erstellt von Masterstudierenden der Medienwissenschaften an der Universität Tübingen, 2024.
- Pawelec, Maria (2025): Workshop “Deepfakes: Chancen und Gefahren synthetischer audio-visueller Medien und wie wir als Gesellschaft mit ihnen umgehen sollten” mit Teilnehmenden der Christlich Akademischen Vereinigung (CAV), Internationales Forum Burg Liebenzell, 4. Januar 2025.
- Pawelec, Maria (2025): Stellungnahme zum Antrag der Fraktion der FDP: „Entschlossen gegen digitale Gewalt: Deepfakes und Pornfakes stoppen!" (Stellungnahme 18/2258) zur Anhörung am 16. Januar 2025, Landtag Nordrhein-Westfalen.
- Pawelec, Maria (2025): Teilnahme als geladene Sachverständige an Anhörung des Ausschusses für Gleichstellung und Frauen des Landtags Nordrhein-Westfalen zum Antrag der Fraktion der FDP „Entschlossen gegen digitale Gewalt: Deepfakes und Pornfakes stoppen!“, Landtag Nordrhein-Westfalen (Drucksache 18/10528), 16. Januar 2025 (online).
- Pawelec, Maria (2025): Teilnahme am Wissenschaftskommunikationsproject "I'm a Scientist" mit Schüler:innen, organisiert von "Wissenschaft im Dialog", Themenrunde "KI und Politik", Februar 2025.
- Pawelec, Maria; Duda, Michelle und Sievi, Luzia (2025): Rechtssicher und ethisch reflektiert auf Falschinformationen reagieren. Eine Handreichung für Behörden und Organisationen mit Sicherheitsaufgaben. Tübingen: IZEW, Materialien zur Ethik in den Wissenschaften, Band 26. ISBN: 978-3-935933-23-0.

7 Zitierte Literatur

Bethmann, Andreas/Hilgenböcker, Elke/Wright, Michael (2021). Partizipative Qualitätsentwicklung in der Prävention und Gesundheitsförderung. In: Michael Tiemann/Melvin Mohokum (Hg.). Prävention und Gesundheitsförderung. Mit 169 Abbildungen und 117 Tabellen. Berlin, Springer, 1083–1095.

Bloom, Benjamin S./Engelhardt, Max D./Furst, Edward J./Hill, Walker H./Krathwohl, David R. (1956). Taxonomy of Educational Objectives. The Classification of Educational Goals. Longmans.

Dubs, Rolf (2010). Bildungspolitik und Schule - wohin? Altstätten, Tobler.

Goecke, Tobias (2024). Die Taxonomie von Bloom. URL: https://supratix.com/blog/upskill-manager/die-taxonomie-von-bloom/ (abgerufen am 18.03.2025).

Kerres, Michael (2021). Didaktik. Lernangebote gestalten. Münster/New York, Waxmann.

Schumacher, Eva-Maria (2008). Didaktisches Leitbild für kompetenz- und lernzentriertes Lehren und Lernen an Hochschulen. URL: https://ilias-hdw.fh-bielefeld.de/goto.php?target=file_2345_download&client_id=IHDW (abgerufen am 18.03.2025).

Selle, Klaus (2011). »Particitainment« oder: Beteiligen wir uns zu Tode? PNDonline, 1–19. URL: https://publications.rwth-aachen.de/record/140376/files/2011_selle_particitainment.pdf (abgerufen am 17.03.2025).