

Anlage 1d: Anwendungsfälle von KI in Bundesministerien und nachgeordneten Behörden (Antwort zu den Fragen 6 bis 6b)

Hinweis: Die Nummerierung entspricht der Anlage 1a.

lfd. Nr.	Ressort/ Behörde	Abt.	Evaluierung	Erläuterung	Veröffentlichung
1	BMF	IVA4	ja	Die Modelle werden iterativ entwickelt und fortlaufend angepasst. Bei Übernahme in den Betrieb sind die Modelltrainings fortzusetzen und die Ergebnisse regelmäßig zu testen.	Bisher noch nicht erfolgt, da noch in der Entwicklung
2	BMF/ BZSt	Z	ja	Nutzer*innen	nicht veröffentlicht Ergebnisse von Evaluierungen werden nicht veröffentlicht, da hier sensible Steuerdaten tangiert sind (Hinweis auf § 30 Abgabenordnung).
3	BMF/ BZSt	Z	nein		
4	BMF/ BZSt	St U	ja	Interne Evaluierung durch Nutzende zur Verbesserung der Treffergenauigkeit.	Nicht veröffentlicht zur Vermeidung einfacher Umgehung der E-Commerce Mustererkennung durch Webseitenbetreiber.
5	BfArM	9	ja	Regelmäßige interne Evaluierung durch Mitarbeiter aus dem Bereich Data Science.	Ergebnisse für internen Gebrauch
6	BfArM	9	ja	Regelmäßige interne Evaluierung durch Mitarbeiter aus dem Bereich Data Science.	Ergebnisse für internen Gebrauch
7	BfArM	9	ja	Regelmäßige interne Evaluierung durch Mitarbeiter aus dem Bereich Data Science.	Ergebnisse für internen Gebrauch
8	BfArM	Abt. 5 Forschung	ja	intern durch Regulation	Ergebnisse für internen Gebrauch
9	RKI	Abt. 3	ja	Interne Evaluierung durch Projektmitarbeiter, Test auf Funktionalität der KI Anwendung	
10	RKI	ZBS6	ja	Interne Evaluierung durch Projektmitarbeiter, Test auf Funktionalität der KI Anwendung	

11	RKI	Abt. 1, MFI, ZBS6	ja	Interne Evaluierung durch Projektmitarbeiter, Test auf Funktionalität der KI Anwendung	
12	RKI	P5	ja	Interne Evaluierung durch Projektmitarbeiter, Test auf Funktionalität der KI Anwendung	
13	RKI	ZIG1	ja	Interne Evaluierung durch Projektmitarbeiter, Test auf Funktionalität der KI Anwendung	
14	RKI	PG	ja	Interne Evaluierung durch Projektmitarbeiter, Test auf Funktionalität der KI Anwendung	
15	RKI	PG	ja	Interne Evaluierung durch Projektmitarbeiter, Test auf Funktionalität der KI Anwendung	
16	RKI	PG	ja	Interne Evaluierung durch Projektmitarbeiter, Test auf Funktionalität der KI Anwendung	
17	DPMA	Abt. 2.4	ja	Evaluierung durch die nutzenden Abteilungen und immer im Vergleich zu bestehenden, internen Systemen sowie in einer intellektuellen Prüfung Ergebnis: Klassifikation beschleunigt die Bearbeitung und verringert den manuellen Aufwand	rein interne Veröffentlichung, rein interner Unterstützungsmechanismus ohne autom. Entscheidungsfindung
18	DPMA	Abt. 2.4	ja	Evaluierung durch die nutzenden Abteilungen und immer im Vergleich zu bestehenden, internen Systemen sowie in einer intellektuellen Prüfung Ergebnis: kognitive Suche unterstützt die Recherche der Patentprüfer in Breite und Tiefe und unterstützt damit die intellektuelle Recherche	s.o.
19	DPMA	Abt. 2.4	ja	Evaluierung durch die nutzenden Abteilungen und immer im Vergleich zu bestehenden, internen Systemen sowie in einer intellektuellen Prüfung Ergebnis: die Übersetzungen ermöglichen erst eine	s.o.

				breite und tiefe Recherche in asiatischen Patentdokumenten für die breite Prüferschaft	
20	BMBF	Abt. 1 / Ref. 112	nein	Aufgrund spekulativer Natur der Prozesse und Ergebnisse der Strategischen Vorausschau ist eine formelle ex-Post-Evaluation nicht zielführend. Um Erfolgskontrolle zu gewährleisten, trägt Projektträger Beispiele zur direkten und indirekten Verwertung zusammen.	
21	BMWK / BAFA	4, 6	nein	Aufgrund der ausführlichen Abnahmetests und kontinuierlichen Weiterentwicklungen sehen wir dedizierte Evaluierungen als nicht notwendig an.	siehe 6a)
22	BMWK / BAFA	5	nein	Aufgrund der ausführlichen Abnahmetests und kontinuierlichen Weiterentwicklungen sehen wir dedizierte Evaluierungen als nicht notwendig an.	siehe 6a)
23	BMWK / BAM	1	ja	Vergleich mit Referenzdaten und -ergebnissen; Vergleich mit Standardmethoden	Ergebnisse incl. Rohdaten und Algorithmen in wiss. Publikationen veröffentlicht
24	BMWK / BAM	6	nein		laufende Forschung
25	BMWK / BAM	6	nein	Die autonome Materialentwicklungslabore befinden sich in der Aufbauphase. Die Ergebnisse werden wissenschaftlich aufbereitet und publiziert.	
26	BMWK / BAM	6	nein		
27	BMWK / BAM	7	nein		https://github.com/BAMresearch/probeye/
28	BMWK / BAM	8	ja	Peer-review-Prozess	Beispielhaft: https://opus4.kobv.de/opus4-bam/frontdoor/index/index/start/6/rows/10/sortfield/score/sortorder/desc/searchtype/simple/query/machine+learning/belongs_to_bibliographyfq/true/subjectfq/NDT/docId/48794 https://opus4.kobv.de/opus4-bam/frontdoor/index/index/start/0/rows/10/sortfield/score/sortorder/desc/s

					<p>earchtype/simple/query/machine+learning/belongs_to_bibliographyfq/true/institutefq/8+Zerst%C3%B6rungsfreie+Pr%C3%BCfung/docId/54861 https://opus4.kobv.de/opus4-bam/frontdoor/index/index/start/2/rows/10/sortfield/score/sortorder/desc/earchtype/simple/query/machine+learning/belongs_to_bibliographyfq/true/institutefq/8+Zerst%C3%B6rungsfreie+Pr%C3%BCfung/docId/56915 https://opus4.kobv.de/opus4-bam/frontdoor/index/index/start/3/rows/10/sortfield/score/sortorder/desc/earchtype/simple/query/machine+learning/belongs_to_bibliographyfq/true/institutefq/8+Zerst%C3%B6rungsfreie+Pr%C3%BCfung/docId/54219 https://opus4.kobv.de/opus4-bam/frontdoor/index/index/start/6/rows/10/sortfield/score/sortorder/desc/earchtype/simple/query/machine+learning/belongs_to_bibliographyfq/true/institutefq/8+Zerst%C3%B6rungsfreie+Pr%C3%BCfung/docId/53939 https://opus4.kobv.de/opus4-bam/frontdoor/index/index/start/7/rows/10/sortfield/score/sortorder/desc/earchtype/simple/query/machine+learning/belongs_to_bibliographyfq/true/institutefq/8+Zerst%C3%B6rungsfreie+Pr%C3%BCfung/docId/56938</p>
29	BMWK/ BAM	8	ja	Peer-review-Prozess	<p>Beispielhaft: https://opus4.kobv.de/opus4-bam/frontdoor/index/index/start/5/rows/10/sortfield/score/sortorder/desc/earchtype/simple/query/v%C3%B6lker/docId/56638 https://opus4.kobv.de/opus4-bam/frontdoor/index/index/start/1/rows/10/sortfield/score/sortorder/desc/earchtype/simple/query/machine+learning/belongs_to_bibliographyfq/true/institutefq/8+Zerst%C3%B6rungsfreie+Pr%C3%BCfung/facetNumber_author_facet/all/author_facetfq/Kruschwitz%2C+Sabine/docId/48799 https://opus4.kobv.de/opus4-bam/frontdoor/index/index/start/0/rows/10/sortfield/score/sortorder/desc/earchtype/simple/query/machine+learning/belongs_to_bibliographyfq/true/institutefq/8+Zerst%C3%B6rungsfreie+Pr%C3%BCfung/facetNumber_author_facet/all/author_facetfq/Kruschwitz%2C+Sabine/docId/53137</p>
30	BMWK/ BAM	8	ja	intern durch Expert*inn*en	Noch nicht veröffentlicht, Paper ist im Peer-Review-Prozess

31	BMWK/ BAM	9	ja	interne Projektmitarbeiter	Veröffentlichung geplant im Rahmen wissenschaftlicher Publikation
32	BMWK/ BAM	9	ja	interne Projektmitarbeiter	Fabry, C., Pittner, A. & Rethmeier, M. Design of neural network arc sensor for gap width detection in automated narrow gap GMAW. Weld World 62, 819–830 (2018). https://doi.org/10.1007/s40194-018-0584-8
33	BMWK/ BAM	8 und 9	nein	nein, Grundlagenforschung	nein, Grundlagenforschung
34	BMWK/ BAM	8	nein	nein, Grundlagenforschung	nein, Grundlagenforschung
35	BMWK/ BAM	8	nein		
36	BMWK/ BAM	S	nein		laufende Forschung
37	BMWK/ BGR	B1	nein	mutmaßlich durch Auftragnehmer, aber Kriterien nicht klar nachvollziehbar; durch uns nach Beginn des Leistungszeitraums auf Fallbasis durch fachlich versiertes Personal	keine, da kommerzielles Angebot
38	BMWK/ BGR	B1	ja	Team	
39	BMWK/ BGR	B2	ja	Durch Entscheider wurde Modellgüte anhand Vergleich mit sogenannten Hindcasts durchgeführt (d.h. "Nachhersagen" historischer Grundwasserstände und Abgleich mit entsprechenden Messwerten); darüber hinaus über wissenschaftliche Aufsätze und deren Publikation über Gutachterverfahren (peer-review)	(A) WUNSCH, A., LIESCH, T. & BRODA, S. (2022a): Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. - Nat Commun 13, 1221. doi: 10.1038/s41467-022-28770-2; KIT&BGR (B)WUNSCH, A., LIESCH, T. & BRODA, S. (2022b): Feature-based Groundwater Hydrograph Clustering Using Unsupervised Self-Organizing Map-Ensembles. - Water Resour. Manage., 36(1): 39-54. doi: 10.1007/s11269-021-03006-y; KIT&BGR (C) WUNSCH, A., LIESCH, T. & BRODA, S. (2021): Groundwater level forecasting with artificial neural networks: a comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). - Hydrol. Earth Syst. Sci. 25: 1671-1687. doi: 10.5194/hess-25-1671-2021

40	BMWK/ BGR	B2	ja	Entscheider hat Evaluierungen der machine learning-basierten Regionalisierungen von hydrogeologischen Flächendaten durchgeführt; hier durch "leave-one-out"-Prinzip, d.h. die Modelle werden an Punkten getestet, an denen gemessene ("ground truth"); Daten vorliegen, aber nicht in den Test- und Validierungsdatensatz enthalten sind; darüber hinaus über wissenschaftliche Aufsätze und deren Publikation über Gutachterverfahren (peer-review)	(A) NÖLSCHER, M. & BRODA, S. (2021): Using an Extreme Gradient Boosting Learner for Mapping Hydrogeochemical Parameters in Germany. - EGU General Assembly, online, 19-30 Apr 2021, EGU21-12818. doi: 10.5194/egusphere-egu21-12818. BGR; (B) NÖLSCHER, M, MUTZ, M. & BRODA, S. (2022): Multiorder Hydrologic Position for Europe – a Set of Features for Machine Learning and Analysis in Hydrology. - Sci Data 9, 662. doi: 10.1038/s41597-022-01787-4
41	BMWK/ BGR	B2	ja	s.o.	Projekt ist in Bearbeitung, Veröffentlichungen befinden sich in Vorbereitung
42	BMWK/BGR	B2	ja	s.o.	Projekt hat gerade begonnen, Ergebnisse werde im Laufe der dreijährigen Projektphase veröffentlicht
43	BMWK/ BGR	B4	ja	BGR B4: Abgleich mit Benchmark-Katalogen - die bestehenden manuell ausgewerteten und klassifizierten Beben -	Interne Erfolgskontrolle / Messung der Zielerreichung
44	BMWK/ BGR	B4	ja	Publikation der Ergebnisse in Berichten und wiss. Fachorganen	Abschlussberichte Massenbewegungen in Deutschland (MBiD) und Erdfälle in Deutschland (EfiD)
45	BMWK/ BGR	B4	ja	Publikation der Ergebnisse in Berichten und wiss. Fachorganen	z.B.: Wagener, N. J., Kalia, A. C. (2021): Unsupervised learning applied to Persistent Scatterer Interferometry datasets for the characterisation of ground motion patterns in northern Germany, Proc. SPIE 11861, Microwave Remote Sensing: Data Processing and Applications, 1186109 (12 September 2021); https://doi.org/10.1117/12.2600131
46	BMAS	Z	ja	Interne Evaluierung der Leistungsfähigkeit. Ergebnisse dienen der Weiterentwicklung der KI-Suche.	Keine Veröffentlichung. Ergebnisse sind primär von interner Relevanz (z.B. Fortentwicklung). Evaluationsergebnisse für Externe nur bedingt verwendbar.
47	BSI	OC / OC31	ja	eigene Entwickler; Ziel: Verfeinerung und Anpassung der Systeme für die bessere Nutzbarkeit im Alltag, d.h. ungeeignete Parameter identifizieren und anpassen; Ergebnisse der kontinuierlichen Evaluation wurden nicht dokumentiert	Ergebnisse führten zu sofortiger Adaption von Verfahren und Parametern

48	BSI	OC / OC33	ja	- Die Evaluierung findet projektintern durch den Auftragnehmer statt - es werden übliche Indikatoren Präzision, Recall und F1-Wert verwendet - diese werden Klassen weise ausgewertet - Der Mehrwert wird als ausreichend angesehen, die manuelle Ontologierweiterung zu unterstützen und soll langfristig noch durch die Verbesserung des Gesamtprozesses gesteigert werden (im Regelbetrieb).	Bisher keine schriftliche Veröffentlichung. Das System befindet sich noch in der Implementierung und soll danach noch eine 3-jährige Pilotphase durchlaufen.
49	Statistisches Bundesamt	C	ja	Mehrstufige Prüfung durch Fachexperten; ggf. manuelle Korrektur der Ergebnisse	Es handelt sich um intern genutzte Verfahren zur Verbesserung des Statistikproduktionsprozesses. Z. T. liegen hierzu Publikationen in Fachzeitschriften (z. B. WISTA Wirtschaft und Statistik, AStA Wirtschafts- und Sozialstatistisches Archiv, Statistical Journal of the IAOS) vor.
50	BMI	Bürgerkommunikation	ja	Evaluation ist durch BMI geplant	
51	BMI	Digitale Gesellschaft	nein		
52	BBK	Krisenmanagement	ja	intern auf Arbeitsebene	nicht veröffentlicht, da nur interne Evaluierung
53	BMI / BKG	Geodaten (GD)	ja	Durch BKG	Evaluierung gegenüber Referenzgeodaten
54	BMI / BKG	Geodaten (GD)	ja	Durch BKG	Evaluierung gegenüber Referenzgeodaten
55	BAMF	7, 2	ja	Laufende Überwachung von Präzision und, soweit möglich, Sensitivität durch externe Auftragnehmer und eigenes Personal	-
56	BAMF	9, 2	ja	Interne Evaluierung / externe Begleitforschung	-
57	BMI	B	ja	Eingesetzte Luftsicherheits-ausrüstung muss gemäß Luftsicherheitsausrüstungsverordnung zertifiziert, vor Inbetriebnahme am Einsatzort zugelassen und im laufenden Betrieb von den Landesluftsicherheits-behörden regelmäßig	Zertifizierte Luftsicherheitsaus-rüstung wird durch die Zertifizierungsstelle veröffentlicht: https://www.bundespolizei.de/SharedDocs/Webs/Downloads/IFG/nlsp_file.html Die Durchführung und Ergebnisse der Zulassung und Routinetests werden

				überprüft werden, ob die zertifizierte Leistung weiterhin erfüllt wird (Routinetests).	gemäß Luftsicherheitsausrüstungsverordnung dokumentiert und im Rahmen von Audits überprüft. Eine Veröffentlichung findet nicht statt.
58	AA	S	ja	-interne Evaluierung mittels Kreuzvalidierung und separatem Testdatensatz für Zeitreihen. -Ergebnisevaluierung gegen regelbasierte Heuristiken als Benchmarks -Ansatz erreicht signifikant bessere Ergebnisse	-bisher keine formale Evaluierung erfolgt -Modellierung im Aufbau
59	BAuA	Fachbereich 2	ja	Eigenentwicklung, deren Ergebnisqualität geprüft wird	
60	BAuA	Fachbereich 2	ja	Eigenentwicklung, deren Ergebnisqualität geprüft wird	
61	BMEL/BVL	5	nein		
62	BMEL/BVL	4	ja	Mitarbeitende der Abteilung	noch nicht
63	BMEL / BfR	7	ja	Modelle sind vom Hersteller bzw. Modell-Entwickler evaluiert	Zahlreiche Publikationen zu spezifischen Modellen
64	BMEL / BfR	7	ja	siehe Literatur	Sarah Nexus: z.B. http://dx.doi.org/10.1016/j.yrtph.2015.12 Derek Nexus: z.B. http://dx.doi.org/10.1016/j.yrtph.2015.01.010 in OECD QSAR-Toolbox für alle Profiler einzeln dokumentiert (Literatur)
65	BMEL / BfR	7	nein		
66	BMEL / BfR	7	nein	noch in Entwicklung	Bahl et al (2019), https://doi.org/10.1016/j.impact.2019.100179
67	BMEL / BfR	7 und 4	nein	noch in Entwicklung	keine Referenz, Projekt erst gestartet
68	BMEL / BfR	9	ja	Es soll eine Evaluierung durch die Anwenderinnen und Anwender erfolgen, sobald das SMAFIRA Tool online ist	Das Tool selbst steht noch nicht öffentlich zur Verfügung, ist aber für 2023 angedacht; die entsprechende Forschungsergebnisse wurden publiziert: • Neves M, Klippert A, Knöspel F, Rudeck J, Stolz A, Ban Z, Becker M, Diederich K, Grune B, Kahnau P, Ohnesorge N, Pucher J, Schönfelder G, Bert B, Butzke D Automatic classification of experimental models in biomedical literature to support searching for alternative methods to animal experiments. [under review, https://europepmc.org/article/ppr/ppr479254] • Neves M. Integration of the PubAnnotation ecosystem in the development of a web-based search tool for alternative methods, Genomics & Informatics,

					<p>2020, 18(2). [full text and pdf, https://genominfo.org/journal/view.php?doi=10.5808/GI.2020.18.2.e18]</p> <ul style="list-style-type: none"> • Butzke D, Dulisch N, Dunst S, Steinfath M, Neves M, Mathiak B, Grune B. SMAFIRA-c: A benchmark text corpus for evaluation of approaches to relevance ranking and knowledge discovery in the biomedical domain, Preprint from Research Square. [full text and pdf, https://www.researchsquare.com/article/rs-16454/v1] • Neves M, Butzke D, Grune B. Evaluation of Scientific Elements for Text Similarity in Biomedical Publications, 6th Workshop on Argument Mining, ACL 2019, Florence, Italy. [pdf and bibtex, https://www.aclweb.org/anthology/papers/W/W19/W19-4515/] • Neves M, Butzke D, Schönfelder G, Grune B. Bf3R at SemEval-2018 Task 7: Evaluating Two Relation Extraction Tools for Finding Semantic Relations in Biomedical Abstracts, Proceedings of the 12th International Workshop on Semantic Evaluation, 2018, New Orleans, EUA. [pdf and bibtex, https://www.aclweb.org/anthology/papers/S/S18/S18-1130/] • Neves M, Ševa J. An extensive review of tools for manual annotation of documents, Briefings in Bioinformatics, 2019. [full text, https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbz130/5670958]
69	BMEL / BfR	9	ja	eigene wissenschaftliche Güteabwägung	Arbeit ist noch nicht abgeschlossen
70	BMEL / BfR	4	ja	geplant	geplant
71	BMEL / BfR	6	ja	durch die Behörde und deren EU-Partner wie EFSA	DOI: 10.1002/ps.5828 10.2903/sp.efsa.2019.EN-1598
72	BMEL / BfR	6	nein		internes Werkzeug, welches vollautomatisiert die Ähnlichkeitswerte mehrerer Moleküle zueinander ausgibt die sonst auch einzeln zusammenklickbar wären
73	BMEL / BfR	6	ja	durch die Mitgliedstaaten und Experten der OECD	DOI (Auszugsweise): https://doi.org/10.1016/j.comtox.2022.100219 https://doi.org/10.1016/j.comtox.2019.01.006
74	BMEL / BfR	6	ja	peer-review bei Veröffentlichung	https://doi.org/10.1021/ci400649p
75	BMEL / JKI	alle	ja	durch verantw. Wiss. bei Forschungsvorhaben	noch keine Veröffentlichung erfolgt

76	BMUV /BfS	UR	nein	in Planung	
77	BMUV /BfS	WR	nein	in Planung	
78	BMUV /BfS	WR	nein	in Planung	
79	BMVg	Strategie und Einsatz	ja	intern, positiv	nein; keine Relevanz
80	BMVg	KdoCIR	ja	internes Fachpersonal im KdoCIR	Die Ergebnisse lassen Rückschlüsse auf die konkrete Fähigkeiten der Software sowie der Lagebilderstellung KdoCIR zu.
81	Bundes-sprachen-amt (BSprA)	Sprach-mittlerdienst	ja	Feststellung Trainierbarkeit durch BSprA	nein; interner Prozess
82	BMDV/ BFG	Z2/M2/M5/U3	ja	findet erst noch statt	
83	BMDV/ BFG	Z2/Experten-netzwerk	ja	findet erst noch statt	
84	BMDV/ BFG	G3	ja	findet erst noch statt	
85	BMDV/ BFG	M5	ja	findet erst noch statt	
86	BMDV/ BFG	U4	ja	findet erst noch statt	
87	BMDV/ BFG	U4	ja	findet erst noch statt	
88	BMDV/ BFG	U4	ja	findet erst noch statt	
89	BMDV/ BFG	M2	ja	findet erst noch statt	
90	BMDV/ BFG	M2	ja	findet erst noch statt	
91	BMDV/ BFG	M2	ja	findet erst noch statt	
92	BMDV/ BFG	M2/Z2/U3/M5/M1	ja	findet erst noch statt	

93	BMDV/ BAST	SD	ja	Noch nicht abgeschlossen. Evaluierung der Ergebnisse nach Abschluss durch Vergleich mit bestehendem Straßennetz.	
94	BSH	S			
95	BMDV / Deutscher Wetterdien st	mehrere	ja	<ul style="list-style-type: none"> • Eingesetzte KI und ML Verfahren werden fortlaufend intern evaluiert • Prozesse unterliegen einer extrem aufwendigen Qualitätsüberwachung zur stetigen Verbesserung der operationellen Vorhersagesysteme • Der wissenschaftliche Beirat des DWDs wird regelmäßig über KI-Aktivitäten informiert • Drittmittelprojekte werden nach anerkannten Standards der guten wissenschaftlichen Praxis (z.B. durch internationale Begutachtung) evaluiert 	<ul style="list-style-type: none"> • Ergebnisse werden intern regelmäßig in Verifikationsberichten festgehalten • Forschungs-ergebnisse werden im begutachteten wissenschaftlichen Zeitschriften veröffentlicht • Ergebnisse werden ggf. im Internet der Öffentlichkeit zur Verfügung gestellt