



Leitlinie zum Umgang mit künstlicher Intelligenz



Stand: Juli 2023

Vorwort

Künstliche Intelligenz (KI) hat das Potenzial, unsere Welt zu verbessern, birgt aber – wie viele an sich neutral wirkende Technologien – auch Risiken. Aus diesem Grund hat die Europäische Kommission 2021 einen neuen Rechtsrahmen für KI vorgelegt, der zum Ziel hat, das Konzept einer „vertrauenswürdigen KI“, die dem Menschen dient und menschenzentriert arbeitet, in Europa tief zu verankern. Damit soll der Einsatzbereich von künstlicher Intelligenz reguliert werden, insbesondere dort, wo Grundrechte und Datenschutz betroffen sind.

Im Medienbereich steht die Anwendung hochriskanter KI-Technologien im Normalfall nicht auf der Tagesordnung. Zusätzlich ist innerhalb der Datenschutzgrundverordnung (DSGVO) bzw. des Datenschutzgesetzes ein Medienprivileg zum Schutz der umfassenden Aufrechterhaltung der journalistischen Arbeit berücksichtigt. Doch ist es ein Gebot der Stunde, dass sich Medienunternehmen gezielt mit den Herausforderungen des neuen Rechtsrahmens auseinandersetzen. Denn zum einen stehen Medienunternehmen permanent im Blickpunkt der Öffentlichkeit und haben eine besondere Vorbildwirkung in Bezug auf ethische Standards. Zum anderen sind KI-Anwendungen im Alltag von RedakteurInnen, KuratorInnen, EditorInnen und Newsroom-DeveloperInnen angekommen, weshalb sie Leitlinien für deren Einsatz und Entwicklung benötigen.

Erste Ansätze, wie die APA – Austria Presse Agentur mit dieser Herausforderung umgehen möchte, wurden im Rahmen einer Studie des APA-medialab für das Technologieministerium (BMK) „AI.AT.Media“ im Sommer 2021 adressiert und in weiterer Folge in dem von der aws (Austria Wirtschaftsservice Gesellschaft m.b.H.) geförderten Projekt „Visual.Trust.AI“ im Rahmen von internen und extern begleiteten Workshops von einem cross-funktionalen Team ausgearbeitet. Eine 2023 eigens einberufene KI Task Force begleitet die Weiterentwicklung im Umgang mit KI-Technologien laufend. Das Ziel: Als innovatives, marktorientiertes Unternehmen nutzt die APA die Vorteile von KI-Systemen, um Kundinnen und Kunden sowie Mitarbeiterinnen und Mitarbeiter bei ihrer Arbeit zu unterstützen.

Die vorliegende KI-Leitlinie zum Umgang mit künstlicher Intelligenz unterstreicht die Wahrnehmung der APA als rechtlich, technisch, inhaltlich und ethisch kompetente und innovative Partnerin. Sie stellt die Grundlage für alle Anwendungsgebiete künstlicher Intelligenz (Definition siehe S. 5) in der APA-Gruppe dar. Für den Anwendungsfall Gesichtserkennung in Bilddatenbanken, resultierend aus dem Projekt „Visual.Trust.AI“, sowie für den Umgang mit „generativer AI“ sind hier bereits die Kernanforderungen formuliert.

Um den genannten Anforderungen zu genügen, orientiert sich die vorliegende KI-Leitlinie an den Ethik-Leitlinien für eine vertrauenswürdige KI der von der Europäischen Kommission eingesetzten, hochrangigen Expertengruppe für künstliche Intelligenz aus 2018. Unter Berücksichtigung der ethischen Grundsätze:

- ▶ Achtung der menschlichen Autonomie,
- ▶ Schadensverhütung,
- ▶ Fairness und
- ▶ Erklärbarkeit

strebt die APA die Einhaltung folgender Anforderungen an vertrauenswürdige KI bei der Entwicklung, Einführung und Nutzung von KI-Systemen an¹:

1. Vorrang menschlichen Handelns und menschliche Aufsicht
2. Technische Robustheit und Sicherheit
3. Schutz der Privatsphäre und Datenqualitätsmanagement
4. Transparenz
5. Vielfalt, Nichtdiskriminierung und Fairness
6. Gesellschaftliches und ökologisches Wohlergehen
7. Rechenschaftspflicht

¹ Vgl. Europäische Kommission (Hg.): Ethik-Leitlinien für eine vertrauenswürdige KI. Hochrangige Expertengruppe für künstliche Intelligenz, 2018.

Inhalt

A. Geltungsbereich	4
B. Definition von KI	4
C. Ethische Grundsätze für APA-KI-Systeme	6
1. Achtung der menschlichen Autonomie	7
2. Schadensverhütung	7
3. Fairness	8
4. Erklärbarkeit	9
5. Spannungen zwischen den Grundsätzen	9
D. Anwendung der Leitlinie: Visual.Trust.AI, KI-Gesichtserkennung	10
1. Vorrang menschlichen Handelns und menschliche Aufsicht	11
2. Technische Robustheit und Sicherheit	11
3. Schutz der Privatsphäre und Datenqualitätsmanagement	12
4. Transparenz	13
5. Vielfalt, Nichtdiskriminierung und Fairness	14
6. Gesellschaftliches und ökologisches Wohlergehen	15
7. Rechenschaftspflicht	15
Checkliste für „Visual.Trust.AI“	16
E. Anwendung der Leitlinie: Generative AI	17
F. Ansprechpersonen	18
KI-Leitlinie & Künstliche Intelligenz	18
Visual.Trust.AI: Gesichtserkennung in Bilddatenbanken	18
G. Quellen	19

A. Geltungsbereich

Die gegenständliche KI-Leitlinie soll einen Handlungsrahmen für die Unternehmen der APA-Gruppe, konkret APA – Austria Presse Agentur eG und ihre Tochterfirmen APA-IT Informations Technologie GmbH, APA-DeFacto Datenbank & Contentmanagement GmbH, APA-OTS Originaltext-Service GmbH und Gentic Software GmbH bieten. Ist in Folge von „APA“ oder „APA-Konzern“ die Rede, sind darunter sinngemäß alle Unternehmen der APA-Gruppe zu verstehen, sofern nicht explizit die Redaktion der APA angeführt ist. Unter Berücksichtigung bestehender Regelwerke wie des APA-Redaktionshandbuchs oder des Ehrenkodex für die österreichische Presse des Österreichischen Presserats enthält die gegenständliche Leitlinie Handlungsempfehlungen zum Thema künstliche Intelligenz für die im Dokument beschriebenen Bereiche. Der Fokus liegt auf dem Einsatz und der Entwicklung von KI im Zusammenhang mit Produkten und Dienstleistungen der APA. Die Leitlinie dient als Handlungsrahmen für rechtlich und ethisch korrektes Verhalten. Gleichzeitig hilft er den Unternehmen dabei, Risiken aufzudecken und intern zu kommunizieren.

Es handelt sich um einen Handlungsrahmen, der sich laufend weiterentwickeln soll. Speziell in Hinblick auf die fortschreitende Digitalisierung von APA-Diensten und damit einhergehenden, neu entstehenden Rahmenbedingungen wird er einer periodischen Evaluierung

unterzogen. Die Erstfassung aus dem Frühjahr 2022 wurde im Juni 2023 um die Bereiche KI-Gesichtserkennung und Generative AI maßgeblich ergänzt. Die Grundsätze einer vertrauenswürdigen KI, die in diesem Text detaillierter erklärt werden, stellen dabei das ethische Fundament für sämtliche aktuelle und künftige KI-Anwendungen der gesamten APA-Gruppe dar. Die darauf aufbauenden Anforderungen gehen auf Spezifika des jeweiligen KI-Anwendungsfalls ein und können sich von Fall zu Fall im Rahmen der Grundwerte unterschiedlich gestalten. Etwaige über den Rahmen der KI-Leitlinie hinausgehende Zusatzinformationen werden in separaten Dokumenten und Detailaufbereitungen zur Verfügung gestellt. Spätere Aktualisierungen werden bei entsprechender Relevanz im Unternehmen und auch nach außen (z. B. Öffentlichkeit, Förderstellen und weitere Stakeholder) kommuniziert.

Im Entstehen fokussierte sich die Leitlinie auf APA-Anwendungsfälle wie z. B. KI-Gesichtserkennung in internen Bilddatenbanken. Die aktuell bei der APA im Einsatz befindlichen KI-Anwendungen sind in erster Linie interne Systeme, welche Suchfilter bzw. Meta-Daten und/oder Texte bereitstellen. Überall dort, wo ein KI-gestütztes Endprodukt Kundinnen und Kunden zur Verfügung gestellt wird, werden die relevanten Punkte in den Kernanforderungen je Anwendungsfall berücksichtigt.

B. Definition von KI

KI-Definition allgemein

Eine allgemeingültige und konsistent genutzte Definition künstlicher Intelligenz existiert nicht. Am ehesten kann KI als Überbegriff für Anwendungen

genutzt werden, bei denen Maschinen menschenähnliche Intelligenzleistungen erbringen. Für die vorliegende KI-Leitlinie wird die Definition der Europäischen Kommission aus dem Jahr 2018 herangezogen:

„KI bezeichnet Systeme mit einem ‚intelligenten‘ Verhalten, die ihre Umgebung analysieren und mit einem gewissen Grad an Autonomie handeln, um bestimmte Ziele zu erreichen.“

„KI-basierte Systeme können rein softwaregestützt in einer virtuellen Umgebung arbeiten (z. B. Sprachassistenten, Bildanalysesoftware, Suchmaschinen, Sprach- und Gesichtserkennungssysteme), aber auch in Hardware-Systeme eingebettet sein (z. B. moderne Roboter, Drohnen oder Anwendungen des ‚Internet der Dinge‘).“²

KI-Definition APA

In der APA fassen wir verschiedene KI-Definitionen folgendermaßen zusammen: Eine KI ist ein System, das in einem eng begrenzten Aufgabengebiet autonome, nicht-triviale Entscheidungen treffen kann. Sie bildet also bis zu einem gewissen Grad menschliche kognitive Fähigkeiten und Prozesse nach.

Als Voraussetzung für solche Entscheidungen muss eine KI immer an die konkreten Anforderungen jeder Aufgabenstellung angepasst werden. Diese Anpassung erfolgt in vielen Fällen durch maschinelles Lernen (Machine Learning). Machine Learning passiert immer unter kontrollierten Bedingungen und mit Hilfe von Trainingsdaten, welche für die Aufgabenstellung als geeignet ausgewählt und aufbereitet wurden.

Durch diesen Prozess entsteht ein Modell, das Wissen über die Aufgabenstellung beinhaltet. Die Qualität des Modells wird durch Anwendung auf entsprechend gewählte Testdaten geprüft. Neben dem weiter oben bereits erwähnten Analysieren der Umgebung und autonomen Handeln ist ein weiteres wichtiges Merkmal von KI also die Fähigkeit, aus Daten zu lernen und dadurch ihr Verhalten anzupassen. Generative Künstliche Intelligenz greift diese genannten Prinzipien auf und erweitert sie um die Fähigkeit, neue, noch nichtexistierende Daten und Inhalte zu erzeugen. Sie stellt eine Untergruppe des maschinellen Lernens dar und befasst sich mit Modellen, die in der Lage sind, kreative Inhalte zu generieren, die als originell und neu wahrgenommen werden.

Vertrauenswürdige KI

Eine weitere Spezifizierung stellt jene der „vertrauenswürdigen KI“ dar, der sich die vorliegende Leitlinie verpflichtet fühlt. Vertrauenswürdige KI zeichnet sich durch drei Komponenten aus, die während des gesamten Lebenszyklus des Systems erfüllt sein sollten:

1. Die KI sollte **rechtmäßig** sein und somit geltendes Recht und alle gesetzlichen Bestimmungen einhalten. Dazu gehören folgende Grundrechte: Achtung der Menschenwürde, Freiheit des Einzelnen, Achtung von Demokratie, Gerechtigkeit und Rechtsstaatlichkeit, Gleichheit, Nicht-diskriminierung und Solidarität, Bürgerrechte.
2. Sie sollte **ethisch** sein und somit die Einhaltung ethischer Grundsätze – Achtung der menschlichen Autonomie, Schadensverhütung, Fairness, Erklärbarkeit und Werte – garantieren; und
3. sie sollte **robust** sein, und zwar sowohl in technischer als auch in sozialer Hinsicht, da KI-Systeme möglicherweise unbeabsichtigten Schaden verursachen, selbst wenn ihnen gute Absichten zugrunde liegen.

Risikoeinschätzung

Von der KI-Leitlinie erfasst sind prinzipiell alle von der APA im Bereich Produkte und Services entwickelten Systeme, heute und in Zukunft. Der Verordnungsentwurf der Europäischen Kommission unterteilt KI-Systeme in unterschiedliche Klassen³:

- ▶ **Hochrisiko-KI-Systeme:** Darunter versteht die EU KI-Systeme, bei denen es aus Sicht der Kommission besonders wahrscheinlich ist, dass ein Eingriff in Sicherheit, Gesundheit oder Grundrechte stattfindet. Für diese Systeme ist eine Zertifizierung nach einheitlichen EU-Standards notwendig. Ob ein KI-System ein Hochrisiko-KI-System ist, hängt nicht von dessen Ausgestaltung, sondern von dessen Zweckbestimmung ab.

² Vgl. Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (Hg): AI.AT.MEDIA, AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course, 2021. https://www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/ai_at_media.html

³ Vgl. Kieffer, Dr. Michael: CE-Zeichen für KI-Systeme – Ausweitung des Produktsicherheitsrechts auf künstliche Intelligenz, 2021. Siehe <https://iot.taylorwessing.com/ce-zeichen-fuer-ki-systeme-ausweitung-des-produktsicherheitsrechts-auf-kuenstliche-intelligenz>

► **Geringes Risiko:** An KI-Systeme, die nicht als „hochrisikobehaftet“ gelten, aber der Interaktion mit natürlichen Personen dienen (z. B. Chatbots), werden besondere Transparenzanforderungen gestellt. Den Nutzerinnen und Nutzern soll klar sein, dass sie es mit einem KI-System zu tun haben. Wer etwa ein Emotionserkennungssystem einsetzt oder ein System zur Assoziierung (gesellschaftlicher) Kategorien anhand biometrischer Daten, muss die davon betroffenen natürlichen Personen über den Betrieb des Systems informieren. KI-Systeme, die sogenannte Deepfakes erzeugen, müssen offenlegen, dass diese Inhalte künstlich erzeugt bzw. manipuliert wurden. Ein Deepfake ist ein Bild-, Ton- oder Videoinhalt, der einer wirklichen Person, einem Gegenstand oder Ereignis merklich ähnelt und fälschlicherweise als echt oder wahrhaftig eingestuft werden könnte.

► **Minimales Risiko:** Systeme mit minimalem Risiko können weitgehend ohne Beschränkungen (im Rahmen des sonstigen Rechts) verwendet werden. Beispiele wären etwa Videospiele oder Spamfilter.

Um den Anforderungen an KI in den Bereichen „Hochrisiko“ und „geringes Risiko“ auch bei künftigen Entwicklungen gerecht werden zu können, verpflichtet sich die APA zu einer internen Prüfung jeder neuen KI-Anwendung in Hinblick auf deren Einstufung. Die Prüfung findet entsprechend der Handlungsempfehlungen der KI-Leitlinie durch die verantwortliche Person statt, welche die Kernanforderungen der gegenständliche KI erarbeitet.

Im Falle einer Einstufung über „Geringes Risiko“ hinaus ist ein rechtliches Gutachten einzuholen, um das potenzielle Vorhandensein einer Hochrisiko-KI eindeutig klären zu können.

C. Ethische Grundsätze für APA-KI-Systeme

Neben Qualität, Sicherheit und Ausgewogenheit stellt Vertrauenswürdigkeit einen zentralen Wert der APA-Redaktion, aber auch der gesamten APA-Gruppe dar. Sie hat die Kernaufgabe, „Trusted Information und Trusted Content“ nach Qualitätskriterien (Check, Double-Check, Re-Check) für Medien zu produzieren. Diese und weitere strenge Qualitätskriterien für journalistisches Arbeiten (siehe Seite 7) gelten auch für in die redaktionelle Arbeit eingebundene KI-Anwendungen.

Um die Vertrauenswürdigkeit des KI-Systems sicherzustellen, setzt die APA Content-Beschränkungen (nur ausgesuchte, geprüfte Quellen) oder Nutzergruppen-Beschränkungen (nur ausgewählte Nutzerinnen und Nutzer haben Zugriff auf die Daten) ein

Diese Entscheidungen (welche Quellen für welche Nutzergruppen) werden transparent dargestellt. Im Einzelfall können auch beide Methodiken Grundlage einer KI-Anwendung sein

Die APA verpflichtet sich, bei der Entwicklung, Einführung und Nutzung von KI-Systemen folgende ethische Grundsätze einzuhalten:

1. Achtung der menschlichen Autonomie
2. Schadensverhütung
3. Fairness
4. Erklärbarkeit

1. Achtung der menschlichen Autonomie

1.1. Achtung der Grundrechte

Das Fundament für eine vertrauenswürdige KI stellen die Grundrechte dar. Die APA sieht es als ihre rechtliche und ethische Verantwortung an, dass ihre KI-Systeme im Bereich Produkte und Services mit den folgenden Grundrechten in Einklang stehen:

- ▶ Achtung der Menschenwürde,
- ▶ Freiheit des Einzelnen,
- ▶ Achtung von Demokratie, Gerechtigkeit und Rechtsstaatlichkeit,
- ▶ Gleichheit, Nichtdiskriminierung und Solidarität,
- ▶ Bürgerrechte

Datenschutz und Privatsphäre nehmen dabei einen hohen Stellenwert ein⁴. Potenziell negative Auswirkungen von KI auf Grundrechte werden einer gewissenhaften Vorab-Prüfung unterzogen und in Hinblick auf die maximal mögliche Schadensverhütung behandelt.

1.2. Journalistische Qualitätsstandards

Die im APA-Statut verankerten Qualitätskriterien Unabhängigkeit, Objektivität, Ausgewogenheit, Vermeidung von Einseitigkeit und Parteinahme, Zuverlässigkeit, Richtigkeit und Schnelligkeit stellen die Grundlage sämtlicher journalistischer Tätigkeit der Redaktion der APA dar.⁵

Zudem hat sich die Redaktion der APA dem Ehrenkodex für die österreichische Presse des Österreichischen Presserats verpflichtet. Durch die Einhaltung der im Grundsatzkatalog festgelegten Inhalte Freiheit, Genauigkeit, Unterscheidbarkeit, Vermeidung von Einflussnahmen, Persönlichkeitsschutz, Intimsphäre, Schutz vor Pauschalverurteilungen

und Diskriminierung, Vermeidung unlauterer Methoden bei der Materialbeschaffung, redaktionelle Spezialbereiche, öffentliches Interesse, Interesse von MedienmitarbeiterInnen sowie Suizidberichterstattung bekennt sich die Redaktion der APA zu einem der Wahrheitsfindung und Korrektheit verpflichteten Gebrauch der Pressefreiheit.⁶

Die Redaktion und ihr Wertegerüst wiederum bilden als Kern der Nachrichtenagentur die Basis aller mit der Marke APA in Verbindung stehenden Entitäten. Sämtliche Dienste und Produkte der APA-Gruppe, so auch KI-basierte Lösungen, orientieren sich bei entsprechender Anwendbarkeit an diesen hohen redaktionellen Qualitätsstandards. Innovationen im Bereich der Quellenverarbeitung, Informationsaufbereitung, Contentverarbeitung und -distribution werden durch Re-Evaluierung dieser Standards begleitet, Qualitätssicherungsprozesse gegebenenfalls für neue Anwendungen in diesen Bereichen adaptiert.

1.3. Internationale Branchenstandards

Im Rahmen der Zusammenarbeit mit anderen Nachrichtenagenturen auf bilateraler, europäischer (Gruppe 39, EANA – European Alliance of News Agencies) und globaler Ebene (MINDS International) findet ein kontinuierlicher Austausch zu Themen aller Art – so auch Innovationsthemen wie KI – statt. Unabhängig ihrer geografischen Verortung und (nicht-)kommerziellen Orientierung stellen Redaktionen mit ihren hohen journalistischen Branchenstandards den Kern jeder Nachrichtenagentur dar.

Auch auf internationaler Ebene bildet somit die Einhaltung der Branchenstandards die Grundlage sämtlicher Kooperationen und gemeinschaftlicher Entwicklungen von Produkten und Dienstleistungen.

2. Schadensverhütung

Von der APA erstellte KI-Systeme sollen weder Schaden verursachen noch verschärfen oder sich auf andere Art und Weise negativ auf Menschen auswirken. Der Schutz der Menschenwürde sowie der geistigen und körperlichen Unversehrtheit steht im Vordergrund.

In erster Linie sollen KI-Systeme dazu dienen, Arbeitsschritte zu vereinfachen, vorhandene Daten zu ergänzen, diese besser aufzubereiten bzw. besser auffindbar zu machen.

⁴ Vgl. APA – Austria Presse Agentur: Datenschutzerklärung APA-Gruppe, Stand Mai 2021, siehe <https://apa.at/about/datenschutzerklaerung>

⁵ vgl. APA – Austria Presse Agentur: APA-Redaktionshandbuch, 2005, siehe <https://apa.at/about/die-redaktion>

⁶ Österreichischer Presserat (2019): Grundsätze für die publizistische Arbeit (Ehrenkodex für die österreichische Presse), siehe https://www.presserat.at/show_content.php?hid=1

Die APA setzt all ihre Handlungen gemäß den gültigen gesetzlichen Rahmenbedingungen. Vor dem Einsatz einer KI-Anwendung werden daher die nötigen rechtlichen Prüfungen, z. B. hinsichtlich Persönlichkeitsrechte, vorgenommen, um materielle wie immaterielle Schäden einer Person zu verhindern.

Zusätzlich werden Produkte und Services in der Regel entlang der bereits erwähnten hohen journalistischen Standards entwickelt. Der Ausprägungsgrad variiert je Kundengruppe und Anwendungsfall, wird jedoch in den Kernanforderungen zur jeweiligen KI-Lösung getrennt erörtert.

Um diesen Anforderungen gerecht werden zu können, werden die Anwendungsfälle für den Einsatz von KI so eng und spezifisch wie möglich definiert und die technischen Möglichkeiten nicht zwingend zur Gänze ausgeschöpft. Bewusste Beschränkungen zum Zwecke der Schadensverhütung können – je nach KI-Anwendung und Umsetzbarkeit – Content- und/oder Nutzungs-seitig erfolgen.

In allen Stufen der konzeptionellen Entwicklung von KI-Systemen der APA-Gruppe werden systemische Sicherungsmaßnahmen nach technischer Möglichkeit getroffen.

APA-Tech als technischer Dienstleister des gesamten APA-Konzerns wurde von TÜV AUSTRIA nach ISO/IEC 20000-1:2018 sowie nach ISO/IEC 27001:2013 zertifiziert, wobei ISO/IEC 20000-1:2018 die Einhaltung der Standards für das IT-Service-Management belegt und ISO/IEC 27001:2013 ein Qualitätssiegel für Informationssicherheit darstellt. Letzteres bescheinigt die Einhaltung der hohen Qualitätslevels für IT-Security Management und stellt sicher, dass für den Basisbetrieb, der hinter den technischen Serviceleistungen von APA-Tech steht, alle relevanten Maßnahmen

bezüglich Datensicherheit, Rechtssicherheit, Integrität und Verfügbarkeit umgesetzt werden.

Die APA-Tech Cyber Security Policy fasst wesentliche, vereinbarte, anzuwendende Grundsätze und Richtlinien zum sicheren Umgang mit IT-Systemen und Daten für die APA-Gruppe zusammen, die detailliert in internen Richtlinien, Prozessbeschreibungen, Betriebs- oder Dienstvereinbarungen geregelt sind.

Zusätzlich kommt auf menschlicher Seite je nach Sensibilität das journalistische Vier-Augen-Prinzip als wesentliches Kriterium des Qualitätsmanagements und der Überwachung bzw. der Schadensverhütung zum Einsatz. Ein Beispiel für eine solche kritische Anwendung ist die KI-Gesichtserkennung in Bilddatenbanken, hier werden die hinterlegten Trainingsdaten im 4-Augen-Prinzip geprüft, da sie die Grundlage der Funktionalität bilden. Dies gilt vor allem für die Grundlagen der KI-Anwendung wie zum Beispiel Trainingsdaten.

2.1. Nachhaltigkeit

Beim Training und Einsatz von KI-Systemen achten wir darauf, für jede Aufgabenstellung eine möglichst effiziente Systemarchitektur zu entwerfen. Wir entwickeln und nutzen also stets auf konkrete Aufgaben spezialisierte Ansätze.

Durch die Spezialisierung sind die von uns genutzten Modelle um viele Größenordnungen kleiner als die häufig in den Medien für ihren Ressourcenbedarf kritisierten allgemeinen KI-Modelle. Außerdem streben wir an, jedes Modell auf der dafür am besten geeigneten Hardware auszuführen. Durch diese Grundsätze minimieren wir den Ressourceneinsatz und daraus folgende ökologische Auswirkungen.

3. Fairness

Die APA verpflichtet sich, bei der Entwicklung von KI-Systemen den Grundsatz der Fairness stets zu berücksichtigen. Das Ziel dabei ist eine gleiche und gerechte Verteilung von Vorteilen und Kosten sowie die Gewährleistung des Schutzes von Personen und Gruppen vor unfairer Verzerrung, Diskriminierung und Stigmatisierung. Die APA steuert möglichen Verzerrungen entgegen, indem Bias in den Daten für das Training von KI so weit wie möglich vermieden wird, etwa durch zusätzliche Recherchen. Sollte ein gewisser Bias in

einer KI-Anwendung absolut unvermeidbar sein, so wird dieser seitens APA transparent kommuniziert. Auf die Gefahr von Diskriminierung oder Stigmatisierung wird bereits bei der Produktentwicklung Bedacht genommen. Darüber hinaus wird die APA dort, wo es notwendig ist, an das jeweilige Produkt oder Service angepasste Opt-In- oder Opt-Out-Möglichkeiten anbieten (z. B. Einholen von Einverständniserklärungen oder das Löschen aus einem Index).

4. Erklärbarkeit

„Ohne menschliche gibt es keine künstliche Intelligenz. Nur wenn wir verstehen, was die Maschinen machen, können wir sie auch so einsetzen, dass es Sinn macht“.⁷

Die APA verpflichtet sich, Prozesse im Zusammenhang mit KI-Anwendungen im größtmöglichen Ausmaß transparent zu gestalten und Rückverfolgbarkeit sowie Erklärbarkeit zu gewährleisten, soweit dies technisch realisierbar ist. Dies umfasst etwa die sorgfältige Dokumentation der in die KI-Anwendung involvierten Prozesse, Mechanismen und Entscheidungsabläufe, der zugrunde liegenden Trainings- und Testdaten, der Evaluierung des Modells, der Ergebnisse interner Testläufe sowie der finalen Validierung. Ebenso umfasst sind neben der Entwicklungs- und der Einführungsphase auch die Ergebnisse der laufenden Evaluierung von KI-Anwendungen im Echtbetrieb wie die stichprobenartige Durchsicht von Ergebnissen oder die quantitative Analyse von Suchabfragen.

In manchen Fällen ist die APA zum Erreichen der bestmöglichen Resultate auf die Nutzung externer Dienstleister angewiesen. Die Wahl eines Dienstleisters basiert neben technischen Eigenschaften stark auf seiner Transparenz in Bezug auf die im vorigen Absatz genannten Aspekte. Auch bei Nutzung externer Dienstleister verwendet die APA so weit wie möglich Trainings- und Testdaten, die von ihr selbst zusammengestellt und aufbereitet wurden.

Die Fähigkeiten und Limitationen von KI-Anwendungen werden seitens APA so transparent wie möglich kommuniziert, um das dauerhafte Vertrauen der NutzerInnen in KI-Anwendungen des Unternehmens gewährleisten zu können. Die Kommunikation bedient sich je nach KI-Anwendung sämtlicher relevanter Kanäle wie Webseiten und Plattform-Oberflächen und gestaltet sich – im Einklang mit dem lebenden Dokument der KI-Leitlinie – adaptiv.

5. Spannungen zwischen den Grundsätzen

Die ethischen Grundsätze wie Achtung der menschlichen Autonomie, Schadensverhütung, Fairness und Erklärbarkeit können miteinander in Konflikt geraten. Am Beispiel KI-gestützter Gesichtserkennung zeigt sich, dass diese einerseits eingesetzt werden kann, um Kriminalität zu reduzieren (Schadensverhütung), andererseits aber auch das Potenzial hat, die Privatsphäre und Freiheitsrechte von Individuen (Achtung der menschlichen Autonomie) einzuschränken. In solchen Fällen wird mittels faktenbasierter Reflexion ein vernünftiger Kompromiss angestrebt.

Die Prämisse, dass die Vorteile von KI-Systemen gegenüber den vorhersehbaren, individuellen Risiken erheblich überwiegen sollen, wird stets berücksichtigt. Sollte ein Kompromiss nicht möglich sein, ist absoluten Prinzipien wie der Würde des Menschen Vorrang gegenüber konkurrierenden Grundsätzen einzuräumen. Vorgänge dieser Art sind kontinuierlich zu ermitteln, bewerten, dokumentieren und zu kommunizieren.

⁷ APA-DeFacto Geschäftsführer Klemens Ganner am Österreichischen Kommunikationstag 2021 des PRVA.

D. Anwendung der Leitlinie

Visual.Trust.AI: KI-Gesichtserkennung in Bild-datenbanken

Täglich wachsen alle Bilddatenbanken der APA um viele tausend Bilder, in der Bildagentur picturedesk.com sind es bereits über 60 Millionen (Stand Juli 2023). Eine übliche, Keyword-basierte Volltextsuche stößt bei dem umfangreichen Bildrepertoire an aktuellen, (inter-)nationalen und historischen Pressefotos oft an

ihre Grenzen. Damit die Motivsuche insbesondere von Personen öffentlichen Interesses wie PolitikerInnen und SportlerInnen schneller und ergebnisreicher erfolgt, entwickelte die APA einen zuverlässigen KI-basierten Personenfilter. Die technologische Basis ist ein Gesichtserkennungs-System (Face Recognition).

„Die Technologie soll bei diesem Projekt kein Selbstzweck sein. Der Informationsgewinn für Bildsuchende steht klar im Vordergrund, so auch unser Qualitätsanspruch an geprüftem, zuverlässigem Bildmaterial.“

(Luzia Strohmayer-Nacif, Leiterin der Visual Desk)

In der KI-Leitlinie bekennt sich die APA bei der Entwicklung von KI-Systemen zur Erfüllung der Anforderungen an eine vertrauenswürdige KI. Anhand des Anwendungsfalles „Visual.Trust.AI“ werden die im Vorwort genannten sieben Grundsätze (Vorrang menschlichen Handelns, Technische Robustheit usw.) als Kernanforderungen erklärt.

Diese Kernanforderungen unterliegen einer kontinuierlichen Bewertung. Es wird darauf geachtet, dass Entscheidungen nicht von einer Person alleine getroffen werden. Das **Vier-Augen-Prinzip** nimmt somit bei der täglichen Arbeit eine zentrale Kontrollfunktion ein.

1. Vorrang menschlichen Handelns und menschliche Aufsicht

1.1. Risikoeinschätzung

Die von der APA entwickelte KI-gestützte Gesichtserkennung für Personen öffentlichen Interesses in der Bilddatenbank ist innerhalb der Klassifizierung der Europäischen Kommission als Hochrisiko-KI-System einzustufen. Entsprechend der gutachterlichen Beurteilung unseres Experten ist jede Gesichtserkennungssoftware nach der Verordnung so einzustufen. Diese Einstufung sagt nichts über die rechtliche Zulässigkeit des Einsatzes aus, sondern teilt ausschließlich KI-Systeme in Risikogruppen ein. Die rechtliche Zulässigkeit orientiert sich an anderen, insbesondere datenschutzrechtlichen Aspekten und hier wurde im Gutachten die Rechtmäßigkeit auf Basis des Medienprivilegs bzw. erweiterten Informationsprivilegs für publizistische Zwecke gesehen, sofern die Verarbeitung der biometrischen Daten auch nur für diese Zwecke erfolgt. Andere Nachrichtenagenturen haben ähnliche redaktionelle Anwendungsfälle bereits im Einsatz.

1.2. Grundrechte & Vorrang menschlichen Handelns

Ein im Vorfeld (beim Bildimport) mit einem KI-System zusätzlich erstellte Beschlagwortung unterstützt Bildsuchende dabei, Fotos von Personen öffentlichen Interesses mit einem bisher nicht möglichen Grad an Effizienz aufzufinden. Eine Volltextsuche liefert auch Treffer, auf denen die gewünschte Person gar nicht

zu sehen ist, diese werden mittels KI-Gesichtserkennung herausgefiltert. Rechtssicherheit, die Einhaltung ethischer Rahmenbedingungen sowie Transparenz in Bezug auf den Einsatz von KI stehen dabei im Vordergrund. Am Bildportal picturedesk.com finden sie eine Beschreibung der konkreten Funktionalität.

Entsprechend den Lenkungs- und Kontrollmechanismen „Human in Command“ und „Human in the Loop“ stellt die eigens für die KI-gestützte Gesichtserkennung geschaffene Position „Visual Trust Officer“ über definierte Prozesse die Vertrauenswürdigkeit im Umgang mit dieser Technologie sicher. Trainierte Personen sind erst nach einem zweistufigen Freigabeprozess für die Bild-Annotierung aktiv. Weiters kann sie jederzeit Trainingsdaten optimieren, sollten in Tests oder im Betrieb ungünstige Resultate in Bezug auf einzelne Personen festgestellt werden.

Um ethische und juristische Risiken zu minimieren, wurden mögliche Auswirkungen auf Grundrechte mittels Ex-ante-Bewertungen durch das Projektteam im Rahmen von Workshops evaluiert. Durch die vielfältige Zusammensetzung des Teams aus Journalismus, Technik, Bildredaktion, Produktmanagement, Bildagentur, Rechtsabteilung und Geschäftsführung sowie die externe Evaluierung durch ein namhaftes Beratungsinstitut konnten dabei verschiedenste Perspektiven abgedeckt werden, was das Risiko weiter reduziert.

2. Technische Robustheit und Sicherheit

2.1. Widerstandsfähigkeit gegen Angriffe und Sicherheitsverletzungen

Mit der Selektion des verwendeten Bildmaterials ausschließlich aus vertrauenswürdigen Quellen, wie zum Beispiel dem APA-Bilderdienst, bedient sich das KI-System einer Content-seitigen Beschränkung zum Zwecke der Schadensverhütung. Um Schaden zu vermeiden, werden nicht alle technisch theoretisch verfügbaren Optionen der Gesichtserkennung implementiert. Einem möglichen Missbrauch wird somit inhaltlich und technisch vorgebeugt.

2.2. Auffangplan und allgemeine Sicherheit

Der Personenfilter wird mit einem Click durch die Nutzenden selbst aktiviert, andernfalls bleibt der Filter deaktiviert und hat keinen Einfluss auf die generelle Verfügbarkeit der Bildersuche. Auf Verlangen einer suchbaren Person und nach Prüfung durch die APA kann das KI-System auf Personen-Ebene von der APA deaktiviert werden. Dieses sogenannte Opt-Out der zu suchenden Person ist mittels E-Mail (siehe Kontaktdaten am Ende des Dokuments) möglich, muss allerdings inhaltlich durch die APA geprüft werden.

2.3. Präzision, Zuverlässigkeit und Reproduzierbarkeit

Es wird besonderes Augenmerk darauf gelegt, dass die Gesichtsidendifikation keine Objekte als Gesichter erkennt, dass es zu keinen Personenverwechslungen kommt und dass administrative Änderungen keine unerwünschten Auswirkungen haben. Sorgfältige Planung und Konzeption, höchste Datenqualität,

gewissenhafte Auswahl der Trainings- und Testdaten, kritische Evaluierung des Modells, zahlreiche Testläufe und eine finale Validierung durch beteiligte Akteurinnen und Akteure (z. B. EntwicklerIn, Visual Trust Officer, ProjektmanagerIn, BetriebsmanagerIn) zielen auf die Erstellung eines funktionalen Systems ab. Das Ergebnis dieser Tests und Evaluierungen werden im Wissensmanagement-System der APA-Gruppe dokumentiert.

3. Schutz der Privatsphäre und Datenqualitätsmanagement

3.1. Schutz der Privatsphäre und Datenschutz

Die APA bekennt sich zum Schutz der Privatsphäre und Datenschutz in allen Phasen des Lebenszyklus des KI-Systems. Die KI-unterstützte Gesichtserkennung nutzt nicht die maximal verfügbaren technischen Möglichkeiten, sondern bedient einen bewusst eng gehaltenen Anwendungsfall. Die Einhaltung rechtlicher und ethischer Parameter hat dabei in allen Phasen oberste Priorität, da biometrische Daten einem erhöhten Schutz unterliegen.

Die einer Personensuche zugrunde liegenden Daten sind von journalistischer Relevanz. Die in der Datenbank gespeicherten Bilder wurden von Bildjournalistinnen und -journalisten geprüft und beschriftet. Der KI-Index enthält ausnahmslos Personen des öffentlichen Lebens, wobei die KI-Gesichtserkennung nur Personen erkennen kann, die aktiv trainiert wurden. Der KI-Filter läuft über rechtlich saubere, dem Urheberrecht entsprechende Bildquellen. Die KI-Technologie verändert bei Aktivierung eines Filters die Anzeige der Trefferliste. Dies erleichtert die Auffindbarkeit gesuchter Personen, ohne eine Wertung des Bildinhalts vorzunehmen. Zum Schutz der Privatsphäre abgebildeter Personen gelten somit dieselben strengen Richtlinien, die in Bilddatenbanken wie jenen von APA-PictureDesk bereits gelebte Realität sind. Sollte eine Person als nicht mehr von öffentlichem Interesse gelten, ist es technisch möglich, die Person aus dem KI-System zu exkludieren.

Eine Nutzung der unter Einhaltung aller rechtlichen, ethischen sowie der hohen redaktionellen Qualitätsstandards trainierten Daten für andere Use Cases innerhalb der APA ist grundsätzlich möglich, bedarf aber einer separaten, zu dokumentierenden Prüfung.

3.2. Personen öffentlichen Interesses

Allgemein ist vorab festzuhalten, dass die APA sich im ersten Schritt im Kern an dem Kreis der Personen des öffentlichen Interesses orientiert und danach einen Personenindex erstellt hat. Eine Erweiterung des Personenkreises ist daher unter der Prämisse insbesondere der medienrechtlichen Bestimmungen im Hinblick auf das Medienprivileg der DSGVO möglich.

Besteht berechtigtes öffentliches Interesse zu einer Person, so kann über sie berichtet werden. Personen öffentlichen Interesses sind etwa PolitikerInnen auf Bundes- und Landesebene.

Ehemalige PolitikerInnen können je nach Position im Aktivstatus den Status des öffentlichen Interesses verlieren. Die rechtliche Relevanz von Führungskräften und Wirtschaftstreibenden entsteht erst im Bild-Text-Zusammenhang. Ebenfalls dazugezählt werden Personen, die regelmäßig vor Medien auftreten und sich dessen bewusst sind, wie zum Beispiel SchauspielerInnen, MusikerInnen, ModeratorInnen und SportlerInnen.⁸

⁸ „Nach DSGVO ist die Verarbeitung besonderer Kategorien personenbezogener Daten zulässig, wenn die betroffene Person eingewilligt hat oder eine gesetzliche Rechtsgrundlage vorliegt. Dies ist insbesondere die Wahrung eines erheblichen öffentlichen Interesses.“ S. 21, Gesichtserkennung.

Der **Personenindex** orientiert sich an folgenden Kriterien:

- ▶ Häufigkeit der Nennung in APA-Basisdienst-Meldungen nach Kategorien
- ▶ Suchfrequenz in APA-Datenbanken
- ▶ Top-level PolitikerInnen
- ▶ Top-level CEOs
- ▶ Prominenten-Ranking von Society-Journalistinnen und -Journalisten
- ▶ Gezielte Personen, um eventuelle in den Modellen enthaltene Benachteiligungen von Bevölkerungsgruppen aufzudecken und darauf zu reagieren (z. B. People of Colour, Conchita Wurst)
- ▶ Internationale Personen von Interesse, die oft gesucht werden, deren Suchergebnisse durch die vorhandene große Bildmenge korumpiert werden, wie z. B. Queen Elizabeth

Jeder Änderung in der Arbeitsweise des KI-Systems muss eine erneute Prüfung zugrunde liegen. Eine Ausweitung auf Clustering nach Emotionen (Sentiments) wird zum Beispiel abgelehnt, da dies einer Klassifizierung der abgebildeten Personen nach einem mathematischen Modell entsprechen würde.

4. Transparenz

4.1. Rückverfolgbarkeit

Die APA dokumentiert im Rahmen der technischen Möglichkeiten sämtliche Datensätze und Prozesse, die zu Entscheidungen des KI-Systems geführt haben. Datenerfassung und -kennzeichnung sowie die verwendeten Algorithmen sind ebenso von der Dokumentation umfasst. Die Entscheidungswege der beteiligten Personen sind dokumentiert, ebenso liegen Informationen zu den im Hintergrund genutzten IT-Systemen vor.

4.2. Erklärbarkeit

Unter Einhaltung rechtlicher und ethischer Rahmenbedingungen wird die Auffindbarkeit von Bildern anhand von KI-gestützter Gesichtserkennung in einem bis dato nicht möglichen Ausmaß verbessert. Es werden nur Identifikationen mit Confidence über einem definiertem Schwellenwert übernommen. Sollte dieser Schwellenwert bei Tests zu viele „false positives“ (Person wird fälschlicherweise „erkannt“) produzieren, wird der Schwellenwert angehoben.

3.3. Qualität und Integrität der Daten; Datenzugriff

Die in der KI-Anwendung verwendeten Bilder entspringen einem ausgesuchten Pool, der strengen journalistischen Kriterien und Qualitätsmerkmalen entsprechen muss. Auf Basis der oben angeführten engen Kriterien für eine Auswahl von Quellen und Personen erfolgt die Datennutzung auf rechtskonformer Grundlage. Qualität, Integrität und Relevanz sind im höchstmöglichen Ausmaß gegeben, um Verzerrungen und Fehler zu vermeiden. Das Training der KI erfolgt nur im redaktionellen Kontext.

Wie in der Redaktion, dem Kern der APA eG, ein sorgfältiger und verantwortungsvoller Umgang mit Nachrichten und Informationen Teil des Selbstverständnisses ist, so stellt auch der sorgfältige und verantwortungsvolle Umgang mit Daten in sämtlichen KI-Anwendungen einen zentralen Wert im Selbstverständnis der gesamten APA-Gruppe dar. Zugriff auf Daten erfolgt auch im Falle der KI-gestützten Gesichtserkennung ausschließlich durch qualifiziertes und geschultes Personal.

4.3. Kommunikation

Ziel sämtlicher Kommunikationsmaßnahmen ist es, Transparenz zu schaffen und einen reflektierten Umgang mit KI-Technologie zu ermöglichen. In der Regel kennzeichnet die APA Plattformen, auf denen KI-Systeme direkt zum Einsatz kommen bzw. deren Dienste auf Grundlage von KI-Systemen laufen. Sie informiert Nutzerinnen und Nutzer zu den Fähigkeiten und Limitationen des Systems. Im Footer des Bildportals finden sie konkrete Informationen zu den Funktionalitäten des Personenfilters.

Umfassende Kommunikations- und Schulungsmaßnahmen finden sowohl intern (Mitarbeiterinnen und Mitarbeiter, indirekt betroffene Personen) als auch extern statt.

4.4. Klare Rollenverteilung

Eine klare Verteilung der Zuständigkeiten sowie die Einhaltung des durchgängigen Vier-Augen-Prinzips tragen zur Klarheit bei und stützen die technische Robustheit des Systems. Folgende Rollen und Verantwortlichkeiten wurden definiert:

- ▶ EntwicklerIn: Entwurf und Umsetzung des KI-Systems durch Wahl und Implementierung geeigneter Ansätze und Modelle.
- ▶ Zuständige MitarbeiterIn: Sicherstellung der Probe aller Personen öffentlichen Interesses und Abdeckung von Grenzfällen (z. B. Bias).

- ▶ Visual Trust Officer: Erstellung eines Testkonzepts. Evaluierung fehlerhafter Suchergebnisse und Weitergabe an die Entwicklung. Stichprobenartige Kontrolle von Suchanfragen und Ergebnissen auf Korrektheit.
- ▶ Produkt ManagerIn/Projekt ManagerIn: Erteilung der Freigabe für die Inbetriebnahme. Jede Veränderung in der Technologie erfordert eine erneute Überprüfung und Bewertung.

Details dazu in der grafischen Darstellung „Detailanalyse der potentiellen Problemstellen mit Lösungsansätzen in den Verantwortlichkeiten“ im Anhang.

5. Vielfalt, Nichtdiskriminierung und Fairness

Wie in APA-KI-Systemen insgesamt, spielen Inklusion, Vielfalt und Fairness während des gesamten Lebenszyklus auch im Gesichtserkennungssystem eine wichtige Rolle. Zur Vermeidung unfairer Verzerrungen wird an mehreren Stellen Rücksicht genommen:

5.1. Trainings- und Testdaten

Um einer Verzerrung der Ergebnisse, dem sogenannten Bias, entgegenzuwirken, versuchen wir in den Trainingsdaten, alle Gesichtstypen abzubilden. Gleichzeitig stellen wir die Unterscheidbarkeit von Trainings- und Testdaten sicher, diese werden vom Visual Trust Officer korrekt markiert und in ausreichender Zahl verfügbar gehalten. Nicht nur das Ergebnis, auch das Modell selbst verändert sich je nach Inputdaten. Die Bildqualität wirkt sich auf die technische Zuverlässigkeit aus. Um Trainingsfehler zu vermeiden, werden mehrere Trainingsbilder pro Person gewählt. Die Suchabfragen in APA-Datenbanken der letzten drei Jahre bieten eine gute Grundlage für den Personenindex. Diese Personen wurden vorher schon gefunden, da sie im Caption-Text der Fotos erwähnt sind. Anschließend wird die Liste anhand definierter Vorgaben (z. B. österreichische Sportstars, Bundesministerinnen und -minister) erweitert. Bekannte KI-Probleme wie das Verwechseln von Menschen und Tieren oder mangelnde Diversität werden in unserer KI-Anwendung weitestgehend vermieden, da nur eine redaktionell überarbeitete Liste an Personen trainiert wird. Als Beispiel ist eine intellektuell erstellte „Diversitätsliste“ anzuführen: Damit soll der durch das Suchverhalten von APA-NutzerInnen generierte so genannte „White-Male-Bias“ im Abbild des öffentlichen Lebens in Österreich abgeschwächt werden.

5.2. Evaluierung des Modells

Die EntwicklerInnen stellen sicher, dass die Resultate eine zufriedenstellende Richtigkeit aufweisen. Zudem wird eine fortlaufende Dokumentation bei Änderungen und Anpassungen des Systems durchgeführt.

5.3. Interner Testlauf und Validierung

Der oder die Visual Trust Officer erstellt ein Testkonzept, das Personen öffentlichen Interesses aus allen Bevölkerungsgruppen abdeckt. Er oder sie evaluiert fehlerhafte Suchergebnisse und eskaliert diese an die Entwicklung. Das Projektmanagement erteilt anhand vordefinierter Kriterien (Accuracy, Precision, Recall) die Freigabe für die Inbetriebnahme.

5.4. Laufender Betrieb

Das Betriebsmanagement, bestehend aus Produktmanagement und dem dazugehörigen Team, überprüft Suchanfragen und Ergebnisse stichprobenartig auf Korrektheit und unterzieht diese quantitativen Analysen zum Zwecke der Ermittlung von Trends und der Sichtung etwaiger Erweiterungs- und Optimierungspotenziale des Systems.

6. Gesellschaftliches und ökologisches Wohlergehen

Durch das Bereithalten und Zugänglichmachen von visuellen zeitgeschichtlichen Dokumenten, die sonst womöglich verloren gingen, liefern Bildagenturen einen wesentlichen gesellschaftlichen Wert. Die KI-gestützte Gesichtserkennung unterstützt diesen bisher oft unter hohem manuellem Aufwand durchgeführten und aufwändigen Suchvorgang und reichert die Suchergebnisse mit

relevanten, bis dato nicht auffindbaren Treffern an. Missbräuchlicher Verwendung wird mittels Beschränkung des KI-Personenfilters auf einen vorab definierten Personenindex sowie einer Beschränkung auf eine rein interne Steuerung (Erstellung Personenindex, Training, Aktivierung für Suchfilter) durch hoch qualifizierte Mitarbeiterinnen und Mitarbeiter der APA vorgebeugt.

7. Rechenschaftspflicht

Die APA ist für die Entwicklung, Einführung und Nutzung des KI-Systems verantwortlich. Eine Dokumentation der technischen und inhaltlichen Rahmenbedingungen des KI-Systems (z. B. Personenindex) ermöglicht Nachprüfbarkeit im Bedarfsfall und schafft dadurch Vertrauen. Ein mit dem Qualitätsmanagement und der periodischen Evaluierung des KI-Systems betrauter Visual Trust Officer stellt die Ermittlung, Bewertung, Berichterstattung und Minimierung potenziell negativer Auswirkungen des KI-Systems sicher. Vom KI-System betroffene relevante Interessen und Werte und potenziell auftretende Konflikte zwischen ebendiesen werden in Hinblick auf ihr Risiko für die oben genannten ethischen Grundsätze evaluiert. Im Konfliktfall ist eine entsprechende Adaptierung des KI-Systems bzw. eine ethisch vertretbare Kompromissfindung vorgesehen.

Sollten trotz aller Sorgfalt dennoch nachteilige Auswirkungen auftreten, stehen die am Ende des Dokuments angeführten APA-AnsprechpartnerInnen zur umgehenden Problemlösung zur Verfügung. Die APA-Gruppe richtet regelmäßig verantwortliche Boards ein, in denen relevante Entscheidungen über allfällige notwendige Anpassungen getroffen werden.

Checkliste für Visual.Trust.AI

Klares Ziel? Was wird dem Kunden/der Kundin angeboten? Definieren des Anwendungsfalls der KI.	<input checked="" type="checkbox"/>
Welche Quellen werden durchsucht? Prüfen, ob diese rechtlich/inhaltlich korrekt sind und mögliche Einschränkungen im Output notwendig sind.	<input checked="" type="checkbox"/>
Wie arbeitet die dahinterliegende KI? Dokumentation bereitstellen.	<input checked="" type="checkbox"/>
Sind die Daten z. B. in einer eigenen Datenbank abgegrenzt? Datensicherheit prüfen.	<input checked="" type="checkbox"/>
Nutze ich Daten aus einem anderen Anwendungsfall? Vorliegenden Use-Case prüfen.	<input checked="" type="checkbox"/>
Welche Personen werden identifiziert? Klären, ob sie von öffentlichem Interesse sind.	<input checked="" type="checkbox"/>
Wo sind die personenbezogenen Daten gespeichert? Datenstruktur ausarbeiten.	<input checked="" type="checkbox"/>
Wer hat Zugriff auf die Daten im Backend? Zugriffsrechte verwalten.	<input checked="" type="checkbox"/>
Wer trainiert, wer kontrolliert die KI? Workflows für Training/Kontrolle etablieren.	<input checked="" type="checkbox"/>
Agiere ich nach journalistischer Sorgfaltspflicht? Wo angebracht hinterfragen.	<input checked="" type="checkbox"/>
Ist die Anwendung DSGVO-konform (Prüfung auf Vorliegen personenbezogener Daten und Einhaltung der Rechtsgrundsätze)? Mit Stabstelle Recht klären.	<input checked="" type="checkbox"/>
Handelt es sich um eine mögliche High-Risk Applikation? Interne Einschätzung mit Stabstelle Recht vornehmen. Bei Bedarf externes Gutachten einholen.	<input checked="" type="checkbox"/>
Welche Eingriffsmöglichkeiten haben betroffene Personen. Gibt es ein Opt-Out/Opt-In? Definieren für wen und wie eine Löschung aus der KI (z. B. Personenindex) erfolgen kann.	<input checked="" type="checkbox"/>

Bei Bedarf verfügbar:

- ▶ APA-Tech Cyber Security Policy (Einsicht nach Unterfertigung eines NDA – Non Disclosure Agreements möglich)
- ▶ APA-Redaktionshandbuch

E. Anwendung der Leitlinie Generative AI

Der seit Ende 2022 mit ChatGPT und Co. erstmals niederschwellige Zugang zu generativer KI zeigt das enorme Potenzial dieser Technologie. Generative AI ist aktuell im Entwicklungsstatus und nicht frei von Schwach- und Gefahrenstellen. Sie erfordert daher einen bewussten und sicheren Umgang. Gängige, vor allem frei verfügbare Modelle wie ChatGPT oder Dall-E gewährleisten derzeit keine rechtskonforme Nutzung und widersprechen mitunter den APA-Unternehmenswerten wie Faktenbasiertheit, Transparenz, Daten- oder Quellensicherheit oder qualitätsjournalistischen Standards.

Die Anwendung der Leitlinie Generative AI ist für Mitarbeiter:innen der APA-Gruppe verpflichtend. Sublieferanten, die entsprechende Dienstleistungen für die APA erbringen, sind über die genannten APA-Prämissen zu informieren und sind zur Einhaltung verpflichtet.

Die APA-Gruppe unterstützt das Experimentieren, Testen und Nutzen von KI-Anwendungen und generativen KI-Tools im beruflichen Kontext unter folgenden Prämissen und unter Wahrung eines sicheren Umgangs:

- ▶ **Nutzung sicherer Tools:** Nutzen Sie für Produktivität, Produkte und Lösungen der APA ausschließlich Tools, die von der APA geprüft und definiert wurden.
- ▶ **Verantwortungsvoller Umgang mit Inhalten und Daten:** Bedenken Sie bei Konversationen mit KI-Anwendungen, dass der verantwortungsvolle Umgang mit sensiblen Unternehmens-, Kund:innen- oder persönlichen Daten genauso zu beachten ist, wie auf jeder anderen Web- oder sonstigen Plattform.
- ▶ **KI ist kein Recherchewerkzeug:** Generative KI-Tools sind grundsätzlich keine Recherchewerkzeuge und dürfen nicht zur Wissensgenerierung oder Informationsaufbringung für APA-Leistungen verwendet werden.
- ▶ **Koordination und Abstimmung:** Informieren Sie die Taskforce und Ihre Bereichsleitung über Ihre KI-Aktivitäten; stimmen Sie den Einsatz in Produkten und Produktionsprozessen im Vorfeld ab.
- ▶ **Nutzungshinweis für Redaktion:** Die APA-Redaktion und APA-PictureDesk verwenden keine Textbausteine von generativer KI in Textmeldungen und Bildbeschreibungen und keine KI-generierten Bilder oder Videos.
- ▶ **Human-in-the-Loop:** Vor der Verwendung von KI-Content in Produkten mit Außenwirkung (insbesondere Publikationen, Newsletter ...) beachten Sie bitte, dass dieser vor Veröffentlichung menschlich eingeordnet, geprüft und verifiziert werden muss. Achten Sie insbesondere auf einen möglichen Bias hinsichtlich Gender- und Gleichstellungsaspekten.
- ▶ **Transparenz:** Kommt KI-generierter Content für Marketing- und Kommunikationszwecke zum Einsatz, der nach Prüfung nicht verändert wurde, ist im Rahmen des Gesamtproduktes ein Hinweis erforderlich, d.h. an geeigneter Stelle entsprechend zu kennzeichnen. Zur Bebilderung und Illustration ist wie schon bisher – mit definierten Ausnahmen – auf das Bildangebot von APA-PictureDesk zurückzugreifen.
- ▶ **Coding:** In der Software-Entwicklung können KI-Tools zur Codegenerierung und – für Code-teile – auch zur Code-Optimierung eingesetzt werden. Eine Überprüfung ist obligatorisch.

Für APA-Mitarbeiter:innen ist eine detaillierte Anleitung mit konkreten Anwendungsbeispielen verfügbar.

F. Ansprechpersonen

KI-Leitlinie & Künstliche Intelligenz

Stabstelle Recht & Datenschutzbeauftragte der APA

Mag. Irene Korinek
irene.kornek@apa.at bzw.
datenschutzbeauftragte@apa.at
+43 1 36060 3010

Chief Digital Officer

Dr. Andreas Mauczka
andreas.mauczka@apa.at
+43 1 36060 3001

APA-Tech KI Koordination

Dr. Andrea Schauerhuber
andrea.schauerhuber@apa.at
+43 1 36060 6741

Visual.Trust.AI: Gesichtserkennung in Bilddatenbanken

Leitung Visual Desk

Luzia Strohmayer-Nacif, MA
luzia.strohmayer-nacif@apa.at
+43 1 36060 5910

Visual Trust Officer

Kathrin Karner
kathrin.karner@picturedesk.com
+43 1 36060 5930

G. Quellen

APA – Austria Presse Agentur: APA-Redaktionshandbuch, 2005, siehe apa.at/about/die-redaktion

APA – Austria Presse Agentur: Aktuelle Datenschutzerklärung APA-Gruppe, siehe apa.at/about/datenschutzerklaerung

Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (Hg): AI.AT.MEDIA, AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course, 2021. www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/ai_at_media.html

Europäische Kommission (Hg.): Ethik-Leitlinien für eine vertrauenswürdige KI. Hochrangige Expertengruppe für künstliche Intelligenz, 2018.

Kieffer, Dr. Michael: CE-Zeichen für KI-Systeme – Ausweitung des Produktsicherheitsrechts auf künstliche Intelligenz, 2021. Siehe iot.taylorwessing.com/ce-zeichen-fuer-ki-systeme-ausweitung-des-produktsicherheitsrechts-auf-kuenstliche-intelligenz

Österreichischer Presserat (2019): Grundsätze für die publizistische Arbeit (Ehrenkodex für die österreichische Presse), siehe www.presserat.at/show_content.php?hid=1

Bidt, Bayrisches Forschungsinstitut für Digitale Transformation (Hg): Gesichtserkennung. Ein Diskussionsbeitrag zur Regulierung der Technologie, 2021. Siehe www.bidt.digital/meldung-studie-gesichtserkennung

Medieninhaber, Herausgeber und Hersteller:

APA – Austria Presse Agentur
Laimgrubengasse 10, 1060 Wien, Österreich
+43 1 36060
apa@apa.at
www.apa.at

Idee und Konzept:

Kathrin Karner, Verena Krawarik,
Bernhard Sonntag, Luzia Strohmayer-Nacif

Redaktion:

Petra Haller, Kathrin Karner, Irene Korinek, Verena Krawarik,
Andreas Mauczka, Michael Paulhart, Marian Rainer-Harbach,
Katharina Schell, Bernhard Sonntag, Luzia Strohmayer-Nacif

Gestaltung und Satz:

vektorama.city

Datum:

Juli 2023

Erscheinungsort:

Wien