Bundesamt
für Sicherheit in der
Informationstechnik

Deutschland
Digital•Sicher•BSI•

Künstliche Intelligenz sicher nutzen

Tipps zur sicheren Nutzung von Künstlicher Intelligenz

Künstliche Intelligenz (KI) begegnet uns immer öfter im Alltag. Auf Ihrem Smartphone sind Sie vermutlich bereits mit KI in Berührung gekommen: Wenn Sie das Smartphone mündlich bitten, einen Wecker zu stellen, oder Sie den Bildschirm mithilfe der Gesichtserkennung entsperren, ist KI im Spiel.

In dieser Broschüre geben wir elf Handlungsempfehlungen für eine möglichst sichere Nutzung von Künstlicher Intelligenz. Dabei konzentrieren wir uns auf Anwendungen, die Sie womöglich bereits verwenden. Ausführliche Informationen zu den Tipps finden Sie auf den nachfolgenden Seiten sowie auf unserer Webseite.

1 Überlegen Sie, ob eine KI-Anwendung für Ihren Zweck geeignet ist.

2 Gehen Sie sparsam mit personenbezogenen und vertraulichen Daten um.

3 Klicken Sie nur auf einen Link, wenn Sie die Anwendung explizit um einen Link gebeten haben.

4 Beziehen Sie KI-Anwendungen nur aus seriösen Quellen.

5 Formulieren Sie Ihre Eingaben mit Bedacht und möglichst spezifisch.

6 Lassen Sie sich nicht täuschen: Vorsicht vor Betrugsmaschen.

7 Sorgen Sie für einen grundlegenden Schutz Ihrer Geräte, Anwendungen und Konten.

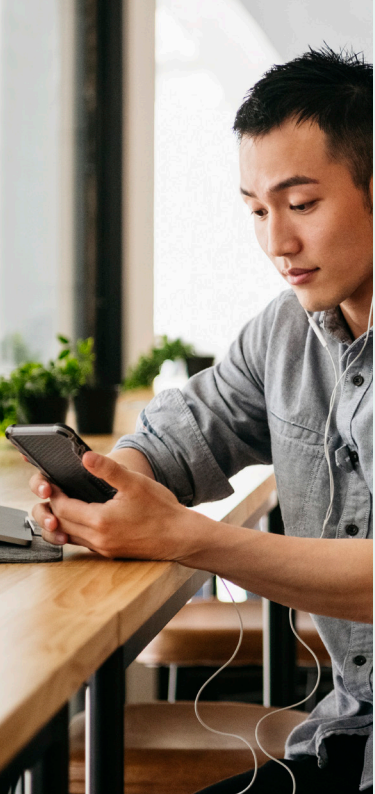
8 Bleiben Sie wachsam – und informiert.

Zusatztipps für Fortgeschrittene:

9 Behalten Sie im Hinterkopf: Eine KI-Anwendung ist abhängig von ihren Trainingsdaten.

10 Informieren Sie sich in den Datenschutzerklärungen und AGB z. B. über die Rechte, die der Hersteller sich einräumt.

11 Prüfen Sie Plugins und Zusatzangebote von Drittanbietern vor der Nutzung.



Wie funktioniert Künstliche Intelligenz?

KI kann unterschiedlich aussehen. Viele KI-Anwendungen, die uns aktuell im Alltag begegnen, basieren auf maschinellem Lernen. Auf diesen Teilbereich der KI wollen wir im Folgenden eingehen. Dabei gilt: Unter Einsatz großer Datenmengen lernt eine Anwendung, Muster und Zusammenhänge zu erkennen. Dabei kann sie sich im Idealfall bei kontinuierlichem Lernen mit neuen Daten weiter verbessern. Wie das konkret aussieht, erklären wir anhand zweier Beispiele.

Beispiel Bilderkennung: Was ist zu sehen?

Eine Anwendung benennt eine abfotografierte Pflanze, eine andere erkennt im Vorbeifahren Verkehrsschilder. Das macht KI möglich – zum Beispiel so:

Schritt 1: Training. Die KI-Anwendung wird anhand sehr vieler Daten auf ihre Aufgabe vorbereitet. Eine Anwendung, die Pflanzen benennen soll, benötigt zuerst unzählige Bilder von Pflanzen. Eine Anwendung, die Verkehrsschilder erkennen soll, braucht hingegen Bilder von Verkehrsschildern. Auch müssen Informationen über den Bildinhalt enthalten sein, in unserem Fall, welche Pflanze oder welches Verkehrsschild zu sehen ist. Anhand dieser sogenannten Trainingsdaten lernt die Anwendung beispielsweise, dass ein Verkehrsschild mit der Aufschrift „STOP“ meist ein Stoppschild ist. Da sie anhand der Informationen über den Bildinhalt prüfen kann, ob sie richtig liegt, lernt sie auch aus ihren Fehlern und wird zunehmend präziser. Bevor die Anwendung genutzt werden kann, sollten Entwicklerinnen und Entwickler testen, ob die Anwendung zuverlässig funktioniert – in unserem Fall zum Beispiel ob sie Verkehrsschilder korrekt zuordnet.

Schritt 2: Nutzung. Wir geben der Anwendung ein Bild, das sie aus dem Training noch nicht kennt. Beispielsweise fotografieren

wir eine Pflanze, oder wir fahren an einem Verkehrsschild vorbei. Durch ihr vorheriges Training kann die Anwendung das Bild einordnen. Einige Anwendungen geben auch Empfehlungen oder treffen Entscheidungen: Eine Anwendung, die Verkehrsschilder wie etwa Geschwindigkeitsbegrenzungen erkennt, kann zum Beispiel Fahrerinnen und Fahrer, die zu schnell fahren, darauf hinweisen.

Auch während der Nutzung werden einige KI-Anwendungen durch weiteres Training noch verbessert: Fotografiert man mit dem Smartphone eine Pflanze und bittet die Anwendung, diese zu benennen, kann auch dieses Foto zu den Trainingsdaten hinzugefügt werden. So lernt die Anwendung und entwickelt sich kontinuierlich weiter.

KI-Anwendungen erkennen jedoch nicht nur Bildinhalte, sondern unter anderem auch gesprochene Sprache. Dadurch können wir Smart Speakern unsere Musikwünsche zurufen oder per Spracheingabe einen Wecker stellen. In dem Fall werden zwar Sprachaufnahmen als Trainingsdaten genutzt, das Prinzip bleibt jedoch das gleiche.

Beispiel Textgenerierung: Schreibe einen Brief!

Einige KI-Anwendungen können Texte verfassen, die sich kaum von solchen unterscheiden, die ein Mensch geschrieben hat. Sie werden als KI-Sprachmodelle bezeichnet – und funktionieren etwa so:

Schritt 1: Training. Als Trainingsdaten nutzen Entwicklerinnen und Entwickler unzählige Texte. Mit ihrer Hilfe lernt die Anwendung, Wörter in eine möglichst sinnvolle Reihenfolge zu bringen. Auf Aufträge zu reagieren, lernt sie auch anhand von beispielhaften Aufgaben und Lösungen. Entwicklerinnen und Entwickler geben der Anwendung Feedback, um unter anderem sicherzustellen, dass die Aussagen der Anwendung ausgewogen sind und keine diskriminierenden Inhalte enthalten.

Schritt 2: Nutzung. Wir stellen der Anwendung eine Aufgabe: Sie soll beispielsweise einen Brief schreiben. Das ist nicht nur hilfreich, wenn uns die passenden Worte fehlen. Auch Unternehmen setzen solche Anwendungen ein. Wenn wir im Chat eines Onlineshops eine Frage stellen, kann uns eine KI-Anwendung antworten. Indem wir ihr weiteres Feedback geben, kann sie auch während der Nutzung lernen.

KI-Anwendungen, die Texte schreiben, unterscheiden sich von unserem vorherigen

Beispiel: Bei der Bilderkennung werden Daten analysiert und interpretiert. Solche Modelle nennt man **diskriminative KI**. Wir würden diese etwa fragen: „Zeigt dieses Bild eine Katze?“. **Generative KI** hingegen erzeugt neue Inhalte – unter anderem Texte, Bilder oder Videos. Unsere Anweisung könnte hier lauten: „Erzeuge ein Bild von einer Katze!“. KI-Anwendungen machen nicht bei Texten halt: Einige Anwendungen erzeugen auch Videos, wieder andere gesprochene Sprache oder Musik.

Was gibt es noch?

Für KI-Anwendungen gibt es zahlreiche weitere Beispiele. Im Alltag begegnen Ihnen womöglich folgende:

- Smarthome-Geräte und smarte Haushaltsgeräte, z. B. Waschmaschinen, die das Waschmittel selbstständig dosieren
- Smartphone-Anwendungen, die für jede öfter fotografierte Person ein eigenes Fotoalbum erstellen
- Empfehlungen auf Basis Ihrer Interessen, z. B. bei Streamingdiensten oder in sozialen Netzwerken
- Spam- und Virenfilter, die fortlaufend dazulernen
- Übersetzungstools
- Navigationssysteme, die die schnellste Route zu einem Ziel vorschlagen
- Weitere Anwendungen im Auto, z. B. ein Fahrspurhalte-Assistent
- Gesundheitsanwendungen, die beispielsweise eine erste Analyse von Symptomen ermöglichen

Mehr Informationen zur Funktionsweise von Künstlicher Intelligenz

Welche Sicherheitsrisiken bergen KI-Anwendungen?

→ KI-Anwendungen machen Fehler. Schlägt ein Streamingdienst uns die Musik eines Sängers vor, der gar nicht unserem Geschmack entspricht, ist das nur ärgerlich. Vertauscht die Verkehrsschilderkennung eines Autos jedoch zwei Verkehrsschilder, kann es gefährlich werden – zumindest, wenn wir uns ganz auf ihre Einschätzung verlassen. Daher gilt es, kritisch zu hinterfragen, was die Anwendung uns vorschlägt.

→ Unbefugte können eine Anwendung dazu bringen, anders als vorgesehen zu reagieren. Zum Beispiel manipulieren sie die Trainingsdaten oder die Eingaben. So können sie zum Beispiel eine smarte Überwachungskamera, die Unbefugte auf einem Grundstück erkennen soll, austricksen.

Für Nutzerinnen und Nutzer ist eine manipulierte Anwendung schwer zu erkennen. Auch daher ist es wichtig, sich nicht bedingungslos auf eine KI-Anwendung zu verlassen.

→ **Kriminelle nutzen KI für Betrugsmaschen.** Einige KI-Anwendungen imitieren z. B. Stimmen: So können sich Betrügerinnen und Betrüger am Telefon als Kontakte ihrer Opfer ausgeben.

→ **KI-Anwendungen sammeln teils sensible Daten über ihre Nutzerinnen und Nutzer.** Betreiber solcher Anwendungen sichern sich mitunter das Recht, die Eingabedaten zu analysieren oder sogar zu verkaufen. Zudem können Cyberkriminelle gegebenenfalls die Daten gezielt abgreifen und missbrauchen.

1. Überlegen Sie, ob eine KI-Anwendung für Ihren Zweck geeignet ist.

Bedenken Sie, dass **der Einsatz von KI je nach Situation unterschiedlich schwerwiegende Risiken birgt.** Irrt sich zum Beispiel eine Gesundheitsanwendung, die Ihnen auf Basis bestimmter Symptome eine Diagnose stellt, schlägt diese Ihnen womöglich auch die falschen Medikamente vor oder rät von einem nötigen Arztbesuch ab. Wenn hingegen eine Foto-App einen Fehler macht und das Foto einer Person dem falschen Kontakt zuordnet, sind die Folgen weit weniger gravierend.

Machen Sie sich außerdem klar, was die Anwendung können soll. Künstliche Intelligenz ist kein Allheilmittel. **Stattdessen führen KI-Anwendungen nur Aufgaben aus, auf die sie vorbereitet wurden.** Dabei gilt auch: Eine Anwendung, die viel mehr kann, als Sie benötigen, brauchen Sie meist nicht. Das macht die Anwendung eher fehleranfällig.

2. Gehen Sie sparsam mit personenbezogenen und vertraulichen Daten um.

Ihre Postadresse, Passwörter oder Kreditkarteninformationen sollten nicht in falsche Hände geraten. Ist eine Anwendung mit dem Internet verbunden, sollten Sie daher **so wenig sensible Daten wie möglich mit ihr teilen.** Das gilt auch, wenn Sie einen Account anlegen oder im Eingabefeld einer KI-Anwendung eine Anweisung tippen. Einige Anwendungen erlauben es zudem, in den Einstellungen abzulehnen, dass **Daten, die während der Nutzung entstehen, anschließend durch den Anbieter gespeichert oder sogar weiterverwendet werden.**

Mitunter gelingt es Kriminellen, **eine KI-Anwendung so zu manipulieren, dass beispielsweise ein Smart Speaker geschickt nach einem Passwort fragt.** Dieses lesen die Kriminellen dann aus. Schöpfen Sie Verdacht, wenn eine Anwendung nach sensiblen Daten fragt, die sie normalerweise nicht benötigt oder die Sie sonst an anderer Stelle eingeben, und kontaktieren Sie den Hersteller. Weitere Tipps, wie Sie möglichst sparsam mit Ihren Daten umgehen, finden Sie auch auf der Webseite der Verbraucherzentrale.

3. Klicken Sie nur auf einen Link, wenn Sie die Anwendung explizit um einen Link gebeten haben.

Kriminelle manipulieren in manchen Fällen KI-Anwendungen, damit diese Links verschicken. Das kann zum Beispiel bei virtuellen Chatpartnern, die uns in einem Onlineshop weiterhelfen sollen, passieren. Die Links führen mitunter zu Schadsoftware oder gefälschten Webseiten. Wenn Nutzerinnen und Nutzer dort ihre Anmeldedaten eingeben, gelangen die Kriminellen an ihr Passwort.

Fragen Sie sich daher vor dem Klicken auf einen Link: Wurde die Anwendung möglicherweise manipuliert? Einige KI-Anwendungen sollen uns Links heraussuchen, etwa Suchmaschinen. Textgeneratoren oder Smart Speaker beispielsweise sollten Links aber nur anzeigen, benennen und ausführen, wenn wir darum bitten. **Wenn Sie den Link bereits kennen, tippen Sie ihn besser direkt in die Adresszeile Ihres Browsers ein.** Alternativ können Sie die gewünschte Webseite auch per Suchmaschine ansteuern.

4. Beziehen Sie KI-Anwendungen nur aus seriösen Quellen.

Zu solchen gehören bekannte Hersteller und offizielle App-Stores. Auch Erfahrungsberichte, insbesondere aus einschlägigen (Online-)Magazinen mit Fachkompetenz, können dabei helfen, einen Hersteller besser einzuschätzen – gerade dann, wenn es sich beispielsweise um ein weniger bekanntes Start-up handelt.

5. Formulieren Sie Ihre Eingaben mit Bedacht und möglichst spezifisch.

Je besser Ihre Eingabe, desto besser ist auch die Ausgabe der KI-Anwendung. Das gilt insbesondere für generative KI, beispielsweise wenn Sie eine KI-Anwendung einen Text schreiben lassen. Probieren Sie verschiedene Eingaben und Synonyme aus. Erklären Sie auch den Kontext: Soll die Anwendung einen formellen Brief verfassen oder eine kurze Nachricht an einen Freund?

Übrigens: **Auch Angreifende können Eingaben so verändern, dass sie eine KI-Anwendung in die Irre führen.** Das kann etwa bedeuten, dass ein Aufkleber auf einem Verkehrsschild dafür sorgt, dass eine Verkehrsschilderkennung das Schild falsch zuordnet. Sichern Sie sich also ab – zum Beispiel, indem Sie weiterhin auch selbst auf Verkehrsschilder achten, um rechtzeitig eingreifen zu können.

6. Lassen Sie sich nicht täuschen: Vorsicht vor Betrugsmaschen.

KI-Anwendungen können Sie im Alltag bei vielen Aufgaben unterstützen. **Aber auch Cyberkriminelle nutzen KI, um ihre Betrugsmaschen zu verbessern.**

Phishing-Mails, die Nutzerinnen und Nutzern zum Beispiel Passwörter entlocken sollen, lassen sich mithilfe von KI-Anwendungen noch leichter schreiben. Auch ermöglicht KI es, Videos zu manipulieren oder die Stimmen anderer am Telefon zu imitieren. Einige Anwendungen arbeiten zudem geschickt mit psychologischen Tricks – um den Angerufenen beispielsweise dazu zu bringen, Geld zu überweisen.

Seien Sie vorsichtig, wenn der Rede- oder Schreibstil nicht zu anderen Anrufen, E-Mails usw. derselben Person oder Institution passt. Überprüfen Sie die Rufnummer oder E-Mail-Adresse. Mitunter werden Betrügerinnen und Betrüger schon damit entlarvt. Durch KI erzeugte Imitationen werden jedoch immer schwieriger zu erkennen.

Stellen Sie daher eine Frage, die nur die jeweilige Person beantworten kann, oder **nehmen Sie über einen anderen Weg Kontakt auf – im Zweifelsfall analog.** Geben Sie zudem im Verdachtsfall keine sensiblen Daten weiter. Vertrauen Sie bei E-Mail-Adressen auch nicht auf die angezeigte Absenderin oder den angezeigten Absender. Diese Angabe ist leicht zu manipulieren.

Bedenken Sie außerdem, dass Cyberkriminelle oftmals frei zugängliche Inhalte aus dem Internet nutzen, um ihre Opfer zu betrügen. Wenn Sie Bilder, Videos, Sprachaufnahmen oder ähnliches hochladen, machen Sie sich bewusst, dass **Kriminelle diese womöglich manipulieren und missbrauchen könnten.**

Geben Sie auch deshalb so wenig wie möglich von sich preis.

7. Sorgen Sie für einen grundlegenden Schutz Ihrer Geräte, Anwendungen und Konten.

Nutzen Sie starke Passwörter und aktivieren Sie die Zwei-Faktor-Authentisierung, wenn Sie sich einen Account einrichten. So erschweren Sie Fremdzugriffe auf Ihr Benutzerkonto.

Halten Sie auch das Antivirenprogramm, den Internetbrowser und das Betriebssystem Ihres Gerätes auf dem neuesten Stand. Updates schließen oftmals Sicherheitslücken, damit Kriminelle sie nicht ausnutzen können.

Wenn Sie eine App herunterladen, prüfen Sie genau, welche Berechtigungen Sie dieser geben. Je nach Verwendungszweck ist es beispielsweise oft nicht notwendig, dass eine App den aktuellen Standort einsehen kann.

8. Bleiben Sie wachsam – und informiert.

Prüfen Sie immer kritisch, was eine KI-Anwendung Ihnen vorschlägt. Das gilt umso mehr in risikoreichen Kontexten, etwa wenn eine KI-Anwendung Sie beim Autofahren unterstützt oder Ihnen medizinische Ratschläge gibt. Nutzen Sie außerdem verschiedene Anwendungen und vergleichen Sie, was diese Ihnen vorschlagen. Wenn Sie einem KI-Sprachmodell eine Frage stellen, ziehen Sie auch weitere Quellen wie ein Nachschlagewerk zu Rate, oder stellen Sie der Anwendung als Test Fragen, deren Antwort Sie kennen.

Künstliche Intelligenz wird stetig weiterentwickelt. Dadurch entstehen auch neue Angriffsszenarien und Betrugsmaschen. Zugleich kann KI uns jedoch bei immer mehr Aufgaben unterstützen. Daher lohnt es sich, KI-Anwendungen auszuprobieren.

Bleiben Sie dran und haben Sie keine Scheu! Weitere Informationen zu allen Themen rund um IT-Sicherheit finden Sie auf unserer Webseite, in unserem Newsletter und in unserem Podcast.

Zusatztipps für Fortgeschrittene:

9. Behalten Sie im Hinterkopf: Eine KI-Anwendung ist abhängig von ihren Trainingsdaten.

Liefert eine KI-Anwendung schwache, ungenaue oder falsche Ausgaben, kann das an den Trainingsdaten liegen. Haben wir für das Training einer Anwendung zur Pflanzenerkennung etwa nur Bilder von europäischen Pflanzen verwendet, kann diese alle weiteren Pflanzen nicht erkennen. Stattdessen gibt sie uns mitunter eine falsche Antwort. Die Trainingsdaten müssen also unter anderem aktuell und geografisch breit genug aufgestellt sein. Ob das der Fall ist, ist oft nur schwer herauszufinden. Eine gezielte Suche auf der Webseite des Anbieters kann sich aber lohnen.

Unzureichende oder nicht repräsentative Trainingsdaten lenken eine Anwendung möglicherweise in eine bestimmte Richtung. Das kann dazu führen, dass sie in ihren Vorschlägen einzelne Personengruppen ausschließt oder diskriminiert. Ein Risiko entsteht ebenso, wenn Inhalte von öffentlichen Internetseiten als Trainingsdaten genutzt werden. Mitunter platzieren Angreifende dort bewusst Inhalte, die die KI fehlleiten oder manipulieren sollen.

Zusatztipps für Fortgeschrittene:

10. Informieren Sie sich in den Datenschutzerklärungen und AGB z. B. über die Rechte, die der Hersteller sich einräumt.

Wenn der Hersteller Ihre Daten speichern oder weiterverwenden darf, entsteht ein potenzielles Ziel für Angreifende. Wenn der Hersteller Ihre Daten an Dritte weitergeben darf, verkauft er diese womöglich. Wägen Sie ab, ob die Nutzung der Anwendung es Ihnen wert ist, gegebenenfalls Ihre Daten preiszugeben. **Einige Anwendungen erlauben es zudem, die Datenschutzeinstellungen anzupassen.** Die Weitergabe Ihrer Daten können Sie dann ablehnen.

In den AGB erfahren Sie oftmals auch, wie Ihre Informationen verschlüsselt oder zwischengespeichert werden. **Das ist wichtig, damit Unbefugte Ihre Daten nicht abgreifen.** Achten Sie bei Webanwendungen auf das Kommunikationsprotokoll „https“ in der Adresszeile. Die rechtlichen Bestimmungen hängen zudem davon ab, in welchem Land die Daten in Rechenzentren gespeichert werden. Sie können daher von denen in Deutschland oder der EU abweichen.

Zusatztipps für Fortgeschrittene:

11. Prüfen Sie Plugins und Zusatzangebote von Drittanbietern vor der Nutzung.

Plugins erweitern KI-Anwendungen: Sie fügen also neue Funktionen hinzu. Andersherum lassen sich auch einige andere Anwendungen, etwa Internetbrowser, mit einem Plugin um KI-Funktionen erweitern. **In den Einstellungen mancher Anwendungen können Sie einsehen, welche Plugins installiert und aktiviert sind.** Überprüfen Sie die Liste und deaktivieren Sie Plugins, wenn Sie diese gerade nicht nutzen. Falls Sie keine Liste finden, recherchieren Sie, ob die Anwendung anderweitig transparent macht, welche Plugins aktiviert sind. Seien Sie bei der Nutzung von Plugins besonders vorsichtig: Gegebenenfalls werden Ihre Daten, zum Beispiel Ihre Eingaben, auch an den Anbieter des Plugins weitergegeben. Nicht immer überprüfen die Hersteller einer KI-Anwendung die verfügbaren Plugins ausreichend. Mitunter können Downloads also Schadsoftware enthalten. **Alle Schritte, die hier angesprochen werden, sollten Sie deshalb auch durchführen, wenn Sie selbst ein Plugin installieren.**

Weiterführende Informationen

Künstliche Intelligenz – wir bringen Ihnen die Technologie näher

Texte mit KI verfassen – Große KI-Sprachmodelle und ihre Risiken

Wegweiser für den digitalen Alltag

Das BSI

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) verfolgt das Ziel, die Digitalisierung in Deutschland sicher zu gestalten. Im Sinne des digitalen Verbraucherschutzes setzt es sich aktiv für den Schutz der Menschen im Netz ein. Zudem sensibilisiert das BSI die Verbraucherinnen und Verbraucher für Sicherheitsrisiken in der digitalen Welt und informiert als unabhängige und neutrale Anlaufstelle über die sichere Nutzung digitaler Technologien.

IMPRESSUM

Herausgeber:
Bundesamt für Sicherheit in der Informationstechnik – BSI
Godesberger Allee 87, 53175 Bonn
E-Mail: service-center@bsi.bund.de
Internet: www.bsi.bund.de
Service-Center: +49 (0) 800 274 1000

Bildnachweise:
Adobe Stock © Maria Savenko
Getty Images © Johnny Greig
Getty Images © Morsa Images
Getty Images © Guido Mieth
Getty Images © AsiaVision
Getty Images © monsitj

Stand: Juni 2024

Layout und Gestaltung:
KOMPAKTMEDIEN Agentur für Kommunikation GmbH, Berlin

Artikelnummer:
BSI-IFB 24/255