dsb
datenschutzbeauftragte
des kantons zürich

Merkblatt

Vorgehen beim Einsatz von KI bei öffentlichen Organen

1 Einleitung

Bei der Nutzung von künstlicher Intelligenz (KI) wird eine Vielzahl von Daten bearbeitet, sei es in Text, Audio oder Bild. Dabei handelt es sich regelmässig auch um Personendaten und besondere Personendaten im Sinne von § 3 Abs. 3 und 4 Gesetz über die Information und den Datenschutz (IDG, LS 170.4).

In folgenden Fällen werden Personendaten durch eine KI-Applikation bearbeitet:

- Personendaten können in den Anweisungen an die KI-Applikation (sogenannte Prompts) vorhanden sein. Weiter können Personendaten in Dokumenten enthalten sein, die in die KI-Applikation zur Bearbeitung gegeben werden, oder in Datenbanken, auf welche die KI-Applikation zugreift.
- Mit KI-Applikationen können Personendaten bearbeitet, beispielsweise verändert, ausgewertet oder ergänzt werden. Somit können Personendaten auch in KI-generierten Inhalten vorkommen.
- Anbieter von KI-Applikationen (KI-Anbieter) können Personendaten wie beispielsweise Fotos dazu benutzen, die KI-Applikation zu verbessern (sogenanntes Training). Werden ganze Dokumente zu Trainingszwecken verwendet, umfasst dies regelmässig ebenfalls Personendaten.
- Anhand der eingegebenen Daten sind Rückschlüsse auf die Person der oder des Nutzenden möglich. Dies kann zudem eine Beurteilung wesentlicher Aspekte der Persönlichkeit der Nutzerin oder des Nutzers erlauben. Datenschutzrechtlich können bei der Verwendung von KI-Applikationen somit auch Persönlichkeitsprofile (§ 3 Abs. 4 lit. b IDG) entstehen. Dies ist selbst dann möglich, wenn die Nutzerin oder der Nutzer sich nicht eingeloggt hat oder ein anonymes Login verwendet.

Wollen öffentliche Organe KI-Applikationen für die gesetzliche Aufgabenerfüllung einsetzen, sind die Vorgaben des IDG zu beachten. Dieses Merkblatt zeigt die Vorgehensweise beim Einsatz von KI-Applikationen in öffentlichen Organen auf.

2 Erarbeitung und Beurteilung der Anwendungsfälle

Vor der Wahl einer KI-Applikation beziehungsweise deren Einsatz durch ein öffentliches Organ ist zu definieren, wie die Nutzung genau ausgestaltet werden soll. Dabei kann es zielführend sein, konkrete Anwendungsfälle zu definieren und folgende Fragen zu klären:

- Welchen Zweck verfolgt der Einsatz der Anwendung?
- Werden für die Erreichung des Zwecks Personendaten bearbeitet?
- Wer soll die Anwendung nutzen (beispielsweise Lehrpersonen, Schülerinnen und Schüler, Mitarbeitende einer bestimmten Abteilung usw.)?
- Wo werden Personendaten benötigt (beispielsweise bei der Anmeldung, bei der Verwendung, im Rahmen von Protokolldaten usw.)?

- Sind besondere Personendaten betroffen oder unterstehen Informationen einem Amts- oder Berufsgeheimnis?
- Werden die eingegebenen Personendaten zur Verbesserung der KI-Applikation verwendet (Trainingsdaten)?
- Wird die KI-Applikation von externen Dienstleistenden (Dritten) angeboten? Werden Trainingsdaten auch für die Verbesserung der KI-Applikation der oder des Dritten eingesetzt?

Weiter ist zu prüfen, ob für die festgestellten Anwendungsfälle hinreichende Rechtsgrundlagen bestehen (§ 8 IDG). Das öffentliche Organ hat für jeden Anwendungsfall zu prüfen, ob der Einsatz der KI-Applikationen zur Erfüllung einer gesetzlich umschriebenen Aufgabe geeignet und erforderlich ist. Werden mit der KI-Applikation besondere Personendaten bearbeitet, ist eine hinreichend bestimmte Regelung in einem formellen Gesetz notwendig. Die vorgesehenen Anwendungsfälle und Abklärungen sind zu dokumentieren.

3 Prüfung Auslagerung und Vertragsverhältnis mit KI-Anbieter

KI-Applikationen werden häufig von externen Dienstleistenden angeboten, oft cloudbasiert. Es handelt sich folglich um ein Bearbeiten im Auftrag (§ 6 IDG). Das öffentliche Organ bleibt für ausgelagerte Datenbearbeitungen verantwortlich. Es muss in der Lage sein, die Pflichten zum Schutz der Informationen (Sach-, Personen- oder besondere Personendaten) wahrzunehmen. Der Auftragnehmer darf die Informationen nur so bearbeiten wie das öffentliche Organ dies tun darf und muss dieselben Sicherheitsanforderungen erfüllen in Bezug auf die Gewährleistung von Vertraulichkeit, Verfügbarkeit und Integrität.

Ein Bearbeiten im Auftrag muss schriftlich vereinbart werden. Der Auftrag muss insbesondere die in § 25 Verordnung über die Information und den Datenschutz (IDV, LS 170.41) genannten Aspekte regeln. Die Abklärungen sind zu dokumentieren.

Weitere Informationen befinden sich im Leitfaden «Bearbeiten im Auftrag».

4 Durchführen einer Datenschutz-Folgenabschätzung (DSFA)

Ist gemäss den Bewertungen nach Ziffer 2 davon auszugehen, dass mit der KI-Anwendung Personendaten bearbeitet werden, ist eine Datenschutz-Folgenabschätzung durchzuführen, um die mit der Bearbeitung einhergehenden Risiken für die Grundrechte der betroffenen Personen einzuschätzen.

Informationen zur Erstellung einer DSFA und ein Formular für die DSFA sind auf der Website der Datenschutzbeauftragten abrufbar.

5 Erstellen eines ISDS-Konzepts

Mit einer Schutzbedarfsanalyse ist festzulegen, mit welchen organisatorischen und technischen Massnahmen die Risiken zu minimieren sind. Ein Informationssicherheits- und Datenschutzkonzept (ISDS-Konzept) fasst die Resultate der Schutzbedarfsanalyse zusammen. Das ISDS-Konzept ist die Grundlage für die Umsetzung des Projekts. Es ist auch nach Inbetriebnahme laufend aktuell zu halten.

Hinweise zur Schutzbedarfsanalyse und zum ISDS-Konzept können der HERMES-Seite des Kantons Zürich entnommen werden, die eine Vorlage zur Verfügung stellt.

6 Einreichen zur Vorabkontrolle

Bei KI-Applikationen handelt es sich um neue Technologien im Sinne von § 24 Abs. 1 lit. c IDV. Das Vorhaben ist folglich der Datenschutzbeauftragten zur Vorabkontrolle vorzulegen (§ 10 Abs. 2 IDG).

Hierzu sind folgende Unterlagen einzureichen:

- Datenschutz-Folgenabschätzung (Ziff. 4),

- ISDS-Konzept (Ziff. 5),
- Rechtsgrundlagenanalyse mit den Abklärungen gemäss Ziffern 2 und 3,
- Entwurf des Vertrags mit dem KI-Anbieter (Ziff. 3).

Weitere Informationen zur Vorabkontrolle sind auf der Website der Datenschutzbeauftragten abrufbar.

V 1.0 / April 2025