

**Beschluss des Kantonsrates
zum Postulat KR-Nr. 323/2022
betreffend Qualitätsanforderungen
beim Einsatz von KI in der Verwaltung**

(vom)

Der Kantonsrat,

nach Einsichtnahme in den Bericht und Antrag des Regierungsrates vom 30. Oktober 2024,

beschliesst:

I. Das Postulat KR-Nr. 323/2022 betreffend Qualitätsanforderungen beim Einsatz von KI in der Verwaltung wird als erledigt abgeschrieben.

II. Mitteilung an den Regierungsrat.

Der Kantonsrat hat dem Regierungsrat am 28. November 2022 folgendes von Kantonsrat Gabriel Mäder, Adliswil, und Mitunterzeichnenden am 12. September 2022 eingereichte Postulat zur Berichterstattung und Antragstellung überwiesen:

Der Regierungsrat wird eingeladen einen Bericht zu verfassen, in dem die qualitativen, rechtlichen und ethischen Anforderungen an die unterschiedlichen Kategorien von KI-Systemen, welche in der Verwaltung des Kantons Zürich zum Einsatz kommen sollen, dargelegt werden. Zudem soll aufgezeigt werden, welchen Gremien der Entscheid zur Zulassung von KI-Systemen sowie der Definierung deren Einsatzgebietes zukommt und wie die langfristige Qualitätskontrolle der KI-Systeme sichergestellt wird.

Bericht des Regierungsrates:

I. Ausgangslage

Der Einsatz von künstlicher Intelligenz (KI) hat in den letzten Jahren nicht nur in Gesellschaft und Wirtschaft, sondern auch bei der öffentlichen Hand sehr rasch an Bedeutung gewonnen. Es ist davon auszugehen, dass die Möglichkeiten und der Einsatz von KI sich auch zukünftig rasch weiterentwickeln und weiter zunehmen werden. Damit einhergehend stellen sich Fragen namentlich zur Zulassung von KI-Systemen und -Anwendungen, zu passenden Einsatzgebieten, zu sachgemässen Rahmenbedingungen für den Einsatz je nach Einsatzgebiet sowie zur Qualitätssicherung und -kontrolle. Dabei sind sowohl eigenständige Applikationen und Systeme als auch eingebettete KI-Systeme und -Anwendungen in Standard-Software oder in Fachapplikationen zu berücksichtigen.

Der Regierungsrat ist sich der Verantwortung und der Wichtigkeit der Qualitätssicherung beim Einsatz von KI bewusst. Bereits im Februar 2021 wurde der – auch im Postulat erwähnte – Bericht «Einsatz Künstlicher Intelligenz in der Verwaltung: rechtliche und ethische Fragen» (Schlussbericht vom 28. Februar 2021, Universität Basel in Zusammenarbeit mit AlgorithmWatch CH im Auftrag der Staatskanzlei; nachfolgend: KI-Studie) veröffentlicht. In der KI-Studie werden die rechtlichen und ethischen Aspekte sowie die Herausforderungen und Chancen von KI behandelt. Unter anderem werden dabei einerseits das Verzeichnis der algorithmischen Entscheidungssysteme (AES) und andererseits die Methode der ethischen Folgenabschätzung vorgestellt. Aus den Ergebnissen dieser Studie zum Aufbau des AES-Verzeichnisses lassen sich auch die nötige Struktur und Zusammensetzung der Gremien zur Sicherstellung der Einhaltung der Vorgaben und zur Qualitätssicherung ableiten.

Der Begriff «AES» wird oft synonym mit «KI» oder «Algorithmen» benutzt, da es keine allgemeingültigen Begriffsdefinitionen gibt. Mit AES werden Informationen erfasst, analysiert und Lösungen für bestimmte Probleme entwickelt oder vorgeschlagen. All dies findet innerhalb eines von Menschen festgelegten Umfelds mit den darin geltenden Regeln statt. Entscheidungen oder Vorschläge durch AES können dieses Umfeld wiederum beeinflussen. Dabei hängt das Ausmass dieses Einflusses auf menschliche Entscheidungen von der Rolle der AES ab (vgl. auch Antrag des Regierungsrates vom 5. Juli 2023 zum Neuerlass des Gesetzes über die Information und den Datenschutz [IDG, LS 170.4], Vorlage 5923, Erläuterungen zu § 13, S. 63 f., unter Bezugnahme auf das Deutsche Institut für Menschenrechte, Algorithmische Entscheidungs-

systeme, Menschenrechtliche Vorgaben und Entwicklungen auf internationaler Ebene [institut-fuer-menschenrechte.de/fileadmin/Redaktion/Publikationen/Information/Information_Algorithmische_Entscheidungssysteme.pdf] sowie die KI-Studie).

Im Gegensatz zu herkömmlichen Computerprogrammen kann sich bei AES die Entscheidungsgrundlage als soziotechnologisches System laufend ändern. Eine vertiefte Prüfung der Systeme und eine fortlaufende Qualitätssicherung sind bei solchen Systemen deshalb unabdingbar. In der kantonalen Verwaltung sind bereits heute einige wenige AES im Einsatz. Dabei handelt es sich hauptsächlich um Pilotprojekte. Es ist davon auszugehen, dass in Zukunft weitere derartige Systeme zum Einsatz kommen werden.

Das in der KI-Studie beschriebene AES-Verzeichnis ist ein öffentliches Verzeichnis, das einerseits Auskunft über die im Kanton Zürich eingesetzten AES gibt und andererseits darlegt, welche Überlegungen zur Qualitätssicherung dabei gemacht wurden. Es kann einen grossen Mehrwert für die Öffentlichkeit, aber auch für die Mitarbeitenden der kantonalen Verwaltung darstellen. Mit der ethischen Folgenabschätzung wird zudem eine Methode vorgestellt, um die für das Verzeichnis notwendige Dokumentation zu erstellen.

Das vorliegende Postulat knüpft in der Begründung ausdrücklich an die KI-Studie an. Es fordert ein Aufzeigen der Qualitätssicherung beim Einsatz von KI in der Verwaltung und einen Vorschlag dazu, welche Gremien bei der Zulassung von KI-Systemen entscheidend sein könnten. Außerdem ist das vorliegende Postulat inhaltlich eng verknüpft mit dem Postulat KR-Nr. 9/2022 betreffend Transparenz über den Einsatz von künstlicher Intelligenz in der Verwaltung. In diesem Postulat wird der Regierungsrat gebeten, ein AES-Verzeichnis zu erstellen, um Transparenz über den Einsatz von KI in der Verwaltung zu schaffen.

2. Rechtliche Rahmenbedingungen

Am 5. Juli 2023 hat der Regierungsrat dem Kantonsrat die Totalrevision des IDG beantragt (Vorlage 5923). Damit soll u. a. eine gesetzliche Grundlage für ein Verzeichnis der von einem öffentlichen Organ verwendeten AES geschaffen (§ 13 Abs. 3 E-IDG) und das erwähnte Postulat KR-Nr. 9/2022 als erledigt abgeschrieben werden (vgl. Dispositiv IV gemäss Vorlage 5923). Das öffentliche Organ soll demnach ein öffentlich zugängliches Verzeichnis der von ihm verwendeten AES führen, die sich auf die Grundrechte von Personen auswirken können. Der Regierungsrat soll die Einzelheiten für alle öffentlichen Organe in einer Verordnung regeln. Im Einklang mit den Legislaturzielen des Regie-

rungsrates (vgl. RRB Nr. 871/2023, Legislaturziel RRZ 10) soll so einerseits das Vertrauen der Bevölkerung in die Verwaltungstätigkeiten gestärkt und andererseits die Transparenz über Verwaltungstätigkeiten in diesem Bereich verbessert werden.

Im Bericht zur Totalrevision des IDG wird darauf hingewiesen, dass das IDG neben dem Schutz der Grundrechte von Personen, deren Daten bearbeitet werden, auch den Zweck verfolgt, durch Transparenz die freie Meinungsbildung und die Wahrnehmung der demokratischen Rechte zu fördern sowie die Kontrolle staatlichen Handelns zu erleichtern. Entsprechend besteht ein erhebliches Interesse an einer Offenlegung der Verwendung von AES, die einen Einfluss auf die Grundrechte haben können. Dies gilt unabhängig davon, ob es sich um komplexe Systeme, die mit Mustererkennung arbeiten, handelt oder um einfache Systeme, die nach festen Regeln ohne Entscheidungsspielraum handeln. AES können sich u. a. insbesondere dann auf die Grundrechte auswirken, wenn Personendaten oder auch anonymisierte Personendaten mittels AES ausgewertet oder bei der Bearbeitung von Personendaten AES verwendet werden, so z. B. bei der Verwendung von AES für die Auswertung von Bewerbungsdossiers (Vorlage 5923, S. 64).

Sobald die rechtliche Grundlage für ein AES-Verzeichnis vorliegt, wird ein solches rasch geschaffen werden. Das Bestehen eines AES-Verzeichnisses und das Eintragen von AES in das Verzeichnis stellen erste grundlegende Schritte für die Qualitätssicherung beim Einsatz von KI in der Verwaltung dar.

3. Umsetzung

Um das AES-Verzeichnis umzusetzen, hat die Staatskanzlei ein entsprechendes Projekt eingeleitet. Beabsichtigt ist der Aufbau einer Datenbank und einer webbasierten Oberfläche zur öffentlichen Abfrage und Einsicht in die Datenbank. Die KI-Studie stellt dabei ein wichtiges Grundlagenpapier mit wertvollen Überlegungen für den Aufbau des AES-Verzeichnisses dar. Auch in der Literatur (vgl. Nadja Braun Binder / Liliane Obrecht, White Paper: Transparenz durch öffentliches Verzeichnis, Center for Information Technology, Society und Law, Universität Zürich, Juni 2024 [nachfolgend: White Paper]) wird die Wichtigkeit eines gesetzlich verankerten AES-Verzeichnisses festgehalten und damit die Aktualität der KI-Studie von 2021 bestätigt. Im White Paper werden zudem Empfehlungen abgegeben, die als sachdienliche Grundlage dienen können bei der Festlegung, welche Informationen in einer Dokumentation im AES-Verzeichnis aus rechtlicher Sicht auf Verordnungsebene präzisiert und im Verzeichnis zur Verfügung gestellt werden könnten bzw. sollten. Es sind dies u. a. Basisinformationen zum verantwortlichen Organ,

der Einsatzbereich des Systems einschliesslich der Rechtsgrundlage dafür, der Zweck einer allfälligen Bearbeitung von Personendaten, die Funktionsweise der Datenbearbeitung, Informationen über den Algorithmus und ob dieser regelbasiert oder datenbasiert ist, die Auswirkungen auf die Bevölkerung in Form einer Folgenabschätzung zu den Risiken, zur Ausgestaltung der menschlichen Aufsicht und Informationen zu den Metadaten über das System wie z. B. das Datum der Inbetriebnahme oder das Datum von Überarbeitungen der Informationen (siehe White Paper, S. 5 f.). Diese Informationen decken sich weitgehend mit denen, die mit der in der KI-Studie vorgestellten Methode der ethischen Folgenabschätzung erfasst werden (vgl. KI-Studie, S. 74 ff.). Die ethische Folgenabschätzung zielt insbesondere darauf ab, den Grundrechtsschutz zu gewährleisten und die dafür relevanten Information zu dokumentieren. Auch in der im Rahmen der Totalrevision des IDG vorgesehenen Regelung von § 13 Abs. 3 (gemäss Vorlage 5923) ist der Grundrechtsbezug das Kriterium für die Aufnahme eines Systems ins AES-Verzeichnis (siehe die Ausführungen vorstehend, 2. Abschnitt).

Die ethische Folgenabschätzung, wie sie in der KI-Studie vorgestellt wird, ist so aufgebaut, dass auch Personen ohne Fachkenntnisse in Ethik, KI oder Recht mit der beschriebenen zweistufigen Methode eine Dokumentation zu den einzelnen Systemen erstellen können, die eine ethische Einschätzung der Systeme erlaubt. Dabei stehen die ethischen Prinzipien Schadensvermeidung, Gerechtigkeit/Fairness, Autonomie und Benefizienz als zu erreichende Ziele im Zentrum. Diese sollen mithilfe der instrumentellen Prinzipien Kontrolle, Transparenz und Rechenschaftspflicht erreicht werden (siehe KI-Studie, S. 66). Alle diese Prinzipien sind breit anerkannt, haben sich bewährt und kommen in dieser Form seit Jahren auch in der biomedizinischen Ethik erfolgreich zur Anwendung. Zudem sind dies diejenigen ethischen Prinzipien, die – zu grösseren oder kleineren Teilen – auch vom Grundrechtsschutz erfasst sind. Daher bietet sich die ethische Folgenabschätzung als Methode an, um ethische Fragen für grundrechtsrelevante AES zu beurteilen. Zudem werden mit der ethischen Folgenabschätzung nicht nur Probleme identifiziert und aufgezeigt, sondern es wird auch dokumentiert, wie mit ihnen umgegangen wurde.

Konkret handelt es sich bei der ethischen Folgenabschätzung um ein zweistufiges Verfahren (KI-Studie, S. 66). Mithilfe von zwei Fragelisten (in der KI-Studie als «Checklisten» bezeichnet, siehe S. 74 ff.) werden Risiken erkannt und der Umgang mit ihnen dokumentiert. Dabei werden zunächst die «Triage-Fragen» beantwortet. Dies sind einfach zu beantwortende Fragen, die nur mit «JA» oder «NEIN» beantwortet werden. Im Zweifelsfall, das heisst, wenn eine Frage nicht eindeutig mit «NEIN» beantwortet werden kann, ist als Antwort «JA» auszuwählen. Für jedes

«JA» muss eine bestimmte Auswahl aus den «Transparenz-Fragen» detailliert beantwortet werden. In der KI-Studie wird ein Flussdiagramm vorgestellt (siehe S. 90), das durch diesen Prozess führt. Die Beantwortung der Fragen aus beiden Listen findet in verschiedenen Phasen eines Projekts statt: beim Projektstart, während der Umsetzung des Projekts und im laufenden Betrieb. Für die Beantwortung der «Transparenz-Fragen» wird Expertise aus verschiedenen Perspektiven benötigt, sodass diese nicht alle von einer einzigen Person bzw. Rolle beantwortet werden. Auf diese Weise wird gewährleistet, dass AES nicht als rein technische Systeme, sondern als soziotechnologische Systeme erkannt und beurteilt werden. Aus den gesammelten Antworten auf die «Transparenz-Fragen» wird eine ausführliche Dokumentation erstellt, welche die Grundlage für den öffentlichen Eintrag zum System im AES-Verzeichnis bildet.

Für die Aufnahme von KI-System und -Anwendungen in das AES-Verzeichnis ist das Kriterium des Grundrechtbezugs demnach nicht nur durch den Regelungsvorschlag von § 13 Abs. 3 IDG (gemäß Vorlage 5923) vorgesehen, sondern auch gestützt auf die vorstehenden Ausführungen zur ethischen Folgenabschätzung als sinnvoll und sachgemäß einzustufen. Dabei ist der Grundrechtsschutz weit, d. h. über das Recht auf informationelle Selbstbestimmung und auf Zugang zu Informationen gemäß Art. 17 der Kantonsverfassung (LS 101) hinausgehend, zu interpretieren (vgl. Vorlage 5923, S. 65, wo darauf hingewiesen wird, dass neben dem Diskriminierungsschutz auch an die Meinungsfreiheit, die politischen Rechte oder die Verfahrensgrundrechte zu denken ist und bereits die Möglichkeit eines AES, die Grundrechte zu verletzen, ausreichen soll). Bei der Beurteilung des Grundrechtsbezugs im Einzelfall ist das infrage stehende KI-System im Zweifelsfall grundsätzlich in das Verzeichnis aufzunehmen.

Neben einer Überarbeitung der ethischen Folgenabschätzung, um diese an die in der Verwaltung bereits bestehenden Prozesse und Qualitätskontrollen (wie namentlich die im Rahmen von Projekten zu erstellende Rechtsgrundlagenanalyse und Schutzbedarfsanalyse sowie die datenschutzrechtlich vorgesehene Datenschutzfolgenabschätzung und Vorabkontrolle) anzupassen, ist im Rahmen des Aufbaus des AES-Verzeichnisses zu prüfen, welche Organisationseinheiten bzw. Anlaufstellen bei Fragen zur Zulassung von KI-Systemen und -Anwendungen und ihrer Einsatzgebiete zuständig sein sollen. Es ist dabei davon auszugehen, dass für solche Fragen interdisziplinäre fachliche Expertise aus verschiedenen Direktionen notwendig sein wird. In diese Prüfung einzubeziehen ist zudem die Frage, in welchen Fällen über die Zulassung von KI-Systemen und über die mit dem Einsatz von KI-Systemen verbundenen grundlegenden Fragen durch bestehende Gremien zu entscheiden ist. Weitere wichtige Elemente, die im Rahmen des Projekts zu prüfen sind, sind die

Sicherstellung der notwendigen Kontaktstellen für Rückfragen zu den einzelnen AES aus der Öffentlichkeit, das kontinuierliche Monitoring des Erfassungsprozesses und die damit verbundene Qualitätssicherung sowie die Befähigung der Mitarbeitenden für eine adäquate und qualitativ zufriedenstellende Erfassung der relevanten AES im Verzeichnis.

Mit der vorgesehenen Regelung von § 13 Abs. 3 E-IDG wird, wie dargelegt, die gesetzliche Grundlage für das AES-Verzeichnis und damit für den Umgang mit qualitativen, ethischen und rechtlichen Anforderungen an KI-Systeme geschaffen werden. Zusammen mit der Methode der ethischen Folgenabschätzung und den skizzierten Fragen, die im Rahmen des Projekts AES-Verzeichnis näher zu klären sind, kann die Qualitäts sicherung und -kontrolle in geeigneter Weise auch im Hinblick auf künftige Entwicklungen sichergestellt werden.

Gestützt auf diesen Bericht beantragt der Regierungsrat dem Kantonsrat, das Postulat KR-Nr. 323/2022 als erledigt abzuschreiben.

Im Namen des Regierungsrates

Die Präsidentin: Die Staatsschreiberin:
Natalie Rickli Kathrin Arioli