I. VORWORT

Geschätzte Unternehmerin, geschätzter Unternehmer, geschätzte Organisationen!

Die vorliegende „Trustworthy AI – by Know Center"- Prüfungsrichtlinie („Richtlinie") enthält die Vergabebestimmungen des Know-Center-Gütesiegels „Trustworthy AI – by Know Center". Das Gütesiegel wurde auf Basis anerkannter Prinzipien vertrauenswürdiger KI entwickelt, um zentrale Aspekte digitaler Souveränität erweitert sowie durch Erkenntnisse aus Forschung und Praxis ergänzt.

Mit der Beantragung des „Trustworthy AI – by Know Center"-Gütesiegels entscheiden Sie sich für ein Qualitätsmanagement mit überdurchschnittlicher Wertigkeit für vertrauenswürdige KI. Diese Leistung wird mit dem „Trustworthy AI – by Know Center"-Gütesiegel kommuniziert.

Die Richtlinie basiert auf folgenden Prinzipien:

- Datenschutz und Sicherheit
- Transparenz, Erklärbarkeit und Fairness
- Menschliche Aufsicht und Entscheidung
- Digitale Souveränität
- Ökonomie
- Nachhaltigkeit
- Digitaler Humanismus

Die Beantragung des „Trustworthy AI – by Know Center"-Gütesiegels ist freiwillig und beinhaltet das Einverständnis, sich den Prüfungskriterien der Richtlinie zu unterziehen.

Das Gütesiegel trägt dazu bei, das Vertrauen in KI, die digitale Souveränität und den Wirtschaftsstandort Österreich/Europa zu stärken.

Mit freundlichen Grüßen

Dr. S. Cathrin von Hesler

Site Manager Wien

©Natascha Unkart

Dr. Kerstin Waxnegger

Senior Legal Counsel & Data Protection Officer

©Opernfoto Hausleitner

II. LEITBILD

Der Schwerpunkt des Qualitätssicherungssystems des Gütesiegels liegt auf den Bereichen vertrauenswürdige KI und digitale Souveränität.

Vertrauenswürdige KI bezeichnet KI, die unter Berücksichtigung ethischer, rechtlicher und technischer Kriterien entwickelt und betrieben wird, sodass sie zuverlässig und nachvollziehbar agiert.

Zentrale Aspekte sind die menschliche Aufsicht, Transparenz, Fairness, Robustheit und der Datenschutz. Vertrauenswürdige KI erfordert einen interdisziplinären Ansatz, der technische Innovationen mit ethischen Leitlinien und regulatorischen Vorgaben verbindet. Nur durch die konsequente Integration dieser Prinzipien kann das Vertrauen der Gesellschaft in KI-Systeme nachhaltig gestärkt werden.

Die Entwicklung und der Einsatz vertrauenswürdiger KI leisten einen wesentlichen Beitrag zur digitalen Souveränität. Digitale Souveränität bezeichnet die Fähigkeit von Organisationen und Staaten, digitale Technologien selbstbestimmt, sicher und rechtskonform zu entwickeln, zu betreiben und zu steuern. Vertrauenswürdige KI stärkt diese Fähigkeit, indem sie Transparenz, Nachvollziehbarkeit, Robustheit, Fairness sowie Datenschutz gewährleistet und damit die kontrollierte Nutzung und unabhängige Bewertung von KI-Systemen und KI-Modellen ermöglicht. Sie ist Teil eines umfassenderen Ansatzes digitaler Souveränität, der darüber hinaus auch technologische Infrastruktur, Datenverfügbarkeit, Interoperabilität und organisatorische Kompetenzen umfasst.

Die konsequente Einhaltung der Prinzipien vertrauenswürdiger KI stärkt das Vertrauen der Gesellschaft und ermöglicht eine kontrollierte, souveräne Steuerung digitaler Infrastrukturen, die die Autonomie in Schlüsselbereichen wie Gesundheitswesen, Mobilität, Energieversorgung und Industrie unterstützt.

Vertrauenswürdige KI und digitale Souveränität bilden die Grundlage für die nachhaltige, selbstbestimmte Gestaltung digitaler Transformationen auf nationaler und europäischer Ebene.

III. ALLGEMEINE VERGABEBESTIMMUNGEN

1. Begriffsbestimmungen

Die Begriffe Qualitätssiegel und Gütesiegel werden synonym verwendet.

2. Freiwillige Teilnahme

Die Beantragung des „Trustworthy AI – by Know Center“-Gütesiegels ist freiwillig und beinhaltet das Einverständnis, sich den Prüfungskriterien der Richtlinie zu unterziehen.

3. Lizenzvertrag

Voraussetzung für die Verwendung des „Trustworthy AI – by Know Center“-Gütesiegels im Sinne einer Kennzeichnung, Vermarktung oder Werbung ist der Abschluss eines gültigen Lizenzvertrags mit dem Know Center. Mit Abschluss des Lizenzvertrags wird das Recht zur Führung des „Trustworthy AI – by Know Center“-Gütesiegels erworben.

Lizenznehmer:innen sind alle physischen oder juristischen Personen, die mit dem Know Center einen Vertrag zur Führung des „Trustworthy AI – by Know Center“-Gütesiegels abgeschlossen haben. Mit dem Lizenzvertrag wird das Nutzungsrecht zur Verwendung des geschützten „Trustworthy AI – by Know Center“-Gütesiegels erworben.

IV. SPEZIELLE VERGABEBESTIMMUNGEN

Als spezielle Kriterien des Vergabeverfahrens werden zwei Bestimmungen der Vergabe festgelegt.

Die erste Bestimmung betrifft die Prüfung durch das Know Center in der **Research- und/oder Methoden-/Prototypentwicklungsphase.**

Die zweite Bestimmung betrifft das Eigenprüfungsverfahren des Antragstellers/der Antragstellerin und die Mitteilung des Antragstellers/der Antragstellerin, dass eine entsprechende Eigenprüfung durchgeführt wird.

1. Erste Bestimmung der Vergabe

Erste Voraussetzung der Vergabe ist die Trustworthy-AI-Prüfung in der **Research- und/oder Prototypenphase** bzgl. der Kriterien Datenschutz und Sicherheit, Transparenz, Erklärbarkeit und Fairness, Menschliche Aufsicht und Entscheidung, Digitale Souveränität, Ökonomie, Nachhaltigkeit und Digitaler Humanismus.

Das „Trustworthy AI – by Know Center"-Gütesiegel steht im Bereich KI für folgende Kriterien:

- Datenschutz und Sicherheit

  Technische Sicherheit: Im Vordergrund steht die Betriebssicherheit; es ist darauf zu achten, dass keine Personen zu Schaden kommen.

  Funktionale Sicherheit: Das KI-Tool muss in sicherheitskritischen Systemen vorhersehbar und robust reagieren.

  Datensicherheit und Datenschutz: Der Schutz personenbezogener und anderer schützenswerter Daten beim Training und im Betrieb steht im Fokus.

- Transparenz, Erklärbarkeit und Fairness

  Die Nachvollziehbarkeit der Ergebnisfindung wird durch die anwender:innenspezifische Auswahl technischer Methoden unterstützt.

  Ein fundiertes Verständnis von Bias und Fairness im Daten- und Anwendungskontext legt die Grundlage für gerechte und nichtdiskriminierende Resultate.

- Menschliche Aufsicht und Entscheidung

  Der Erhalt menschlicher Autonomie steht im Vordergrund.

  Zudem wird eine taskadäquate Aufgabenverteilung zwischen Mensch und KI erarbeitet.

- Digitale Souveränität

  Datensouveränität: Die Verarbeitung personenbezogener und anderer geschützter Daten wird entsprechend geschützt.

  Der Forschungsstandort liegt in Europa, die Forschungsdienstleistung wird durch ein europäisches Unternehmen erbracht.

  Weiters wird eine Abhängigkeit von einzelnen Konzernen vermieden.

- Ökonomie

  Eine höhere Autonomie von Prozessen führt zu einer gesteigerten Produktivität.

- Nachhaltigkeit

  Energieverbrauch: Angesichts des hohen Energiebedarfs beim Training großer KI-Modelle ist der Fokus auf energieeffiziente Algorithmen und Hardware zu legen.

  Nachhaltige Lösungen werden durch die Einbindung relevanter Stakeholder erzielt, die mittels KI eine strukturelle Problemlösung ermöglichen.

- Digitaler Humanismus

  Der Mensch muss im Mittelpunkt stehen; KI soll dem Menschen dienen – nicht der Technologie.

  Weiters hat die Anwendung dem Anwendungskontext entsprechend an menschlicher Werteorientierung ausgerichtet zu sein.

  Missbrauchsprävention: Einer schädlichen Nutzung des KI-Tools wird entgegengewirkt.

2. Zweite Bestimmung der Vergabe

Nach der abgeschlossenen Erstprüfung gemäß V.1. durch das Know Center wird als zweite Voraussetzung der Vergabe eine in einem festgelegten Rhythmus zu erfolgende Eigenüberwachung durch den Antragsteller/die Antragstellerin verlangt. Er/Sie hat dem Know Center als qualifizierende Stelle zu bestätigen, dass eine Eigenüberwachung durchgeführt wird.

Der Antragsteller/Die Antragstellerin hat zur Eigenüberwachung ein Konzept nach den unter V.1. genannten Kriterien zu erstellen. Weiter muss der Antragsteller/die Antragstellerin den Rhythmus der Eigenüberwachung festlegen. Zusätzlich muss er/sie in seinem/ihrem Konzept ausführen, wie die KI-Kompetenz gesichert ist. Zur KI-Kompetenz sind keine Ausführungen erforderlich, wenn der Antragsteller/die Antragstellerin ein entsprechendes Kompetenzevaluierungsergebnis für seine/ihre Mitarbeitenden nach DigComp-Standard (Digital Competence Framework EU, AT o.Ä.) belegen kann. Der Antragsteller/Die Antragstellerin hat das Konzept beim Know Center als qualifizierende Stelle einzureichen.

Die Ergebnisse der Eigenüberwachung sind vom Antragsteller/von der Antragstellerin zu protokollieren. Die Dokumentation der Eigenprüfung durch den Antragsteller/die Antragstellerin ist 5 Jahre aufzubewahren.

V. VERLEIHUNG

Das Know Center verleiht bei Erfüllung der Vergabebestimmungen das Gütesiegel und gibt dem Unternehmen die verliehene Gütesiegel-ID schriftlich bekannt. Weiters erhält der Antragsteller/die Antragsstellerin eine Urkunde über die Verleihung des Rechts zur Führung des Gütesiegels sowie das entsprechende Gütesiegel zur weiteren

Verwendung. Der Antragsteller/Die Antragstellerin hat nur mit seiner IP-Adresse Zugriff auf das Siegel (Privacy by design).

Als qualifizierende Institution steht dem Know Center ein Prüfungsermessen zu. Es besteht kein Rechtsanspruch auf Verleihung des Gütesiegels und der Gütesiegel-ID.

VI. GÜLTIGKEIT

Die Gültigkeit des Gütesiegels beträgt 24 Monate. Eine Verlängerung ist möglich, soweit die Standards weiterhin erfüllt sind. Ein Verlängerungsantrag kann vom Know Center abgelehnt werden, wenn geänderte Vergabebestimmungen, bspw. aufgrund gesetzlicher Änderungen, vorliegen.

VII. WIDERRUF

Das Gütesiegel kann widerrufen werden, wenn die Standards im Sinne dieser Richtlinie nicht mehr erfüllt sind. Dem Know Center steht ein Prüfungsermessen zu. Die Entscheidung ist schriftlich mitzuteilen. Mit dem Widerruf erlischt das Nutzungsrecht zur Führung des „Trustworthy AI – by Know Center“-Gütesiegels.

Eine Erstattung der Gebühren findet nicht statt.

GELTUNG

Diese Richtlinie gilt ab 01. April 2026

IMPRESSUM

Medieninhaber und Hersteller: Know Center Research GmbH

A-8010 Graz, Sandgasse 34, Tel: +43 316 873 30801, info@know-center.at

Gestaltung: corporate identity prihoda gmbH