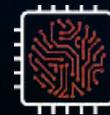


1. Vernetzungstreffen im Rahmen des AI Policy Forums

KI in der
öffentlichen Verwaltung

Mittwoch, 19. Oktober 2022
Festsaal Technisches Museum Wien



AIM AT 2030
Artificial Intelligence Mission Austria



Themensession 3: **Vertrauenswürdigkeit und Ethik**

- **Andreas Windisch**, Know Center Graz - *Trustworthy AI - Konzept und Stand der Forschung*
- **Wolfgang Kabelka**, BRZ - *Vertrauenswürdige KI im BRZ*
- **Ralf Tatto**, BMKÖS & **Peter Biegelbauer**, AIT AI Ethics Lab - *Digitale Verwaltung und Ethik*
- **Björn Lellmann**, BMF - *Akzeptanz und Vertrauen in KI in der Verwaltung*

Trustworthy AI

1. Vernetzungstreffen im Rahmen des AI Policy Forum
October 19th 2022, Vienna, Austria

Andreas Windisch



Outline

(1) Complexity
 $X_{AI} = r X_0 (1 - X_0)$

(2) geopolitical context of AI

(3) Definitions
 $y := 3x + 7$

(4) Human centered AI

(5) Trustworthy AI

(6) Main Take-away

Outline

The chalkboard contains the following content:

- (1) Complexity
 $X_{n+1} = r X_n (1 - X_n)$
- (2) Geopolitical context of AI
[Flags of EU, UK, USA, and Japan]
- (3) Definition
 $y := 3x + 7$
- (4) Human centered AI
[Diagram of a person with arrows pointing to them]
- (5) Take-away
[Shopping cart icon]

A hand in a blue suit sleeve is pointing to the 'Human centered AI' section.

Examples



Defining aspects

- ▲ Nonlinear dynamics
- ▲ Known and unknown unknowns
- ▲ System evolves, cannot be contained or reverted to earlier state
- ▲ System boundaries unclear/hard to define
- ▲ No reliable models, uncertainty

Consequences/Strategies

- ▲ Insufficient knowledge
- ▲ There is no 'truth'
- ▲ Accept uncertainty
- ▲ Monitor evolution closely
- ▲ Adept on the fly

Examples



Defining aspects

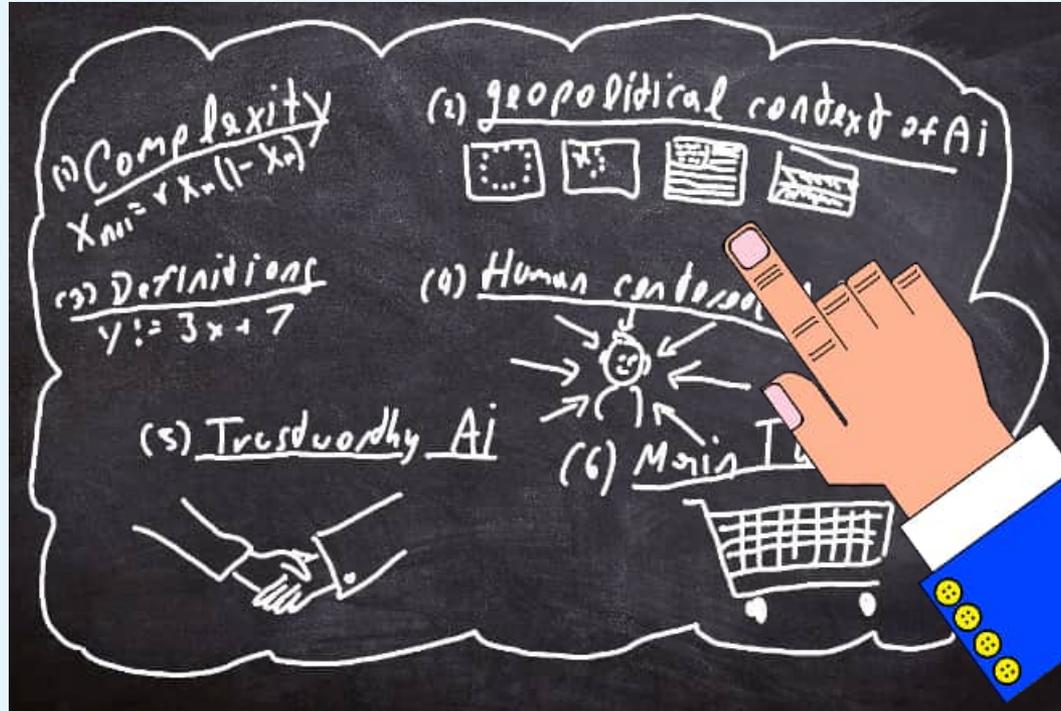
- ▲ Nonlinear dynamics
- ▲ Known and unknown unknowns
- ▲ System evolves, cannot be contained or reverted to earlier state
- ▲ System boundaries unclear/hard to define
- ▲ No reliable models, uncertainty

Consequences/Strategies

- ▲ Insufficient knowledge
- ▲ There is no 'truth'
- ▲ Accept uncertainty
- ▲ Monitor evolution closely
- ▲ Adept on the fly

AI affects/used for all of these systems!

Outline



Geopolitical aspects of AI



Players



China vs. USA

- ▲ China 4 times as many engineers
- ▲ China 9am-9pm-6 days a week - work ethics
- ▲ USA / China: two different information spaces
- ▲ Both have large enough markets
- ▲ energy, transportation, e-commerce, quantum computing

Possible measures and strategies

- ▲ Double funding for research
- ▲ High skill immigration
- ▲ Government involvement (free university, 30-day participation)
- ▲ China: busy starting new things
- ▲ US: act now/stay in race
- ▲ EU: Human rights and AI

Outline

The chalkboard contains the following content:

- (1) Complexity
 $X_{n+1} = r X_n (1 - X_n)$
- (2) geopolitical context of AI
[Flags of EU, UK, USA, and Japan]
- (3) Definitions
 $y := 3x + 7$
- (4) Human centered AI
[Diagram of a person with arrows pointing to them]
- (5) Trustworthy AI
- (6) Main Take-away
[Shopping cart icon]

Machine learning

cf. book of Geron 'Hands-On Machine Learning'

Goodfellow Deep Learning

'ML is the field of study that gives computers the ability to learn without being explicitly programmed' GERON

'A computer is said to learn from ^{data} experience E with respect to a task T and some ^{cost function} performance measure P, if the performance on T, as measured by P, improves with experience E.'

Alan Turing 'The imitation game'

paradigm shift

'conventional approach'

rules - program - answers
data

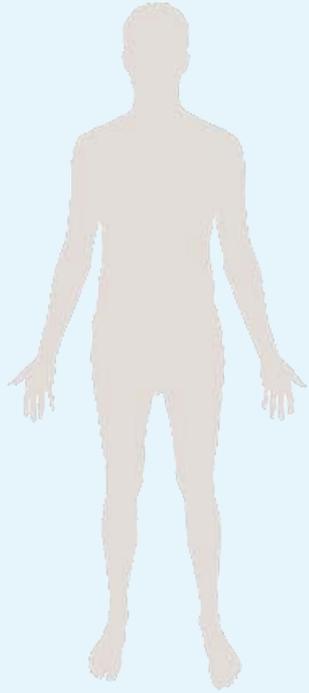
'machine learning'

answers - ML - rules (patterns)
data

Outline

- (1) Complexity
 $X_{n+1} = r X_n (1 - X_n)$
- (2) geopolitical context of AI
[EU Flag] [UK Flag] [USA Flag] [Japan Flag]
- (3) Definitions
 $y := 3x + 7$
- (4) Human centered AI
[Diagram: A person with arrows pointing towards them]
- (5) Trustworthy AI
[Handshake icon]
- (6) Key Take-away

Human centered AI



Who is the 'human' we want to center AI on?

- ▲ 2020: World population approx 7.8 billions
- ▲ More than 70 percent \$ 10 per day
- ▲ Complex systems: system boundaries
- ▲ Hierarchy of scales: Global, intermediate (EU), local (Austria)
- ▲ Different strategies on different scales

Different roles

- ▲ Government
- ▲ Scientists and engineers
- ▲ Entrepreneurs
- ▲ Media
- ▲ Citizens

Outline

(1) Complexity
 $X_{n+1} = r X_n (1 - X_n)$

(2) geopolitical context of AI

(3) Definitions
 $y := 3x + 7$

(4) Human centered AI

(5) Trustworthy AI

(6) Main Take-away

Trustworthy AI



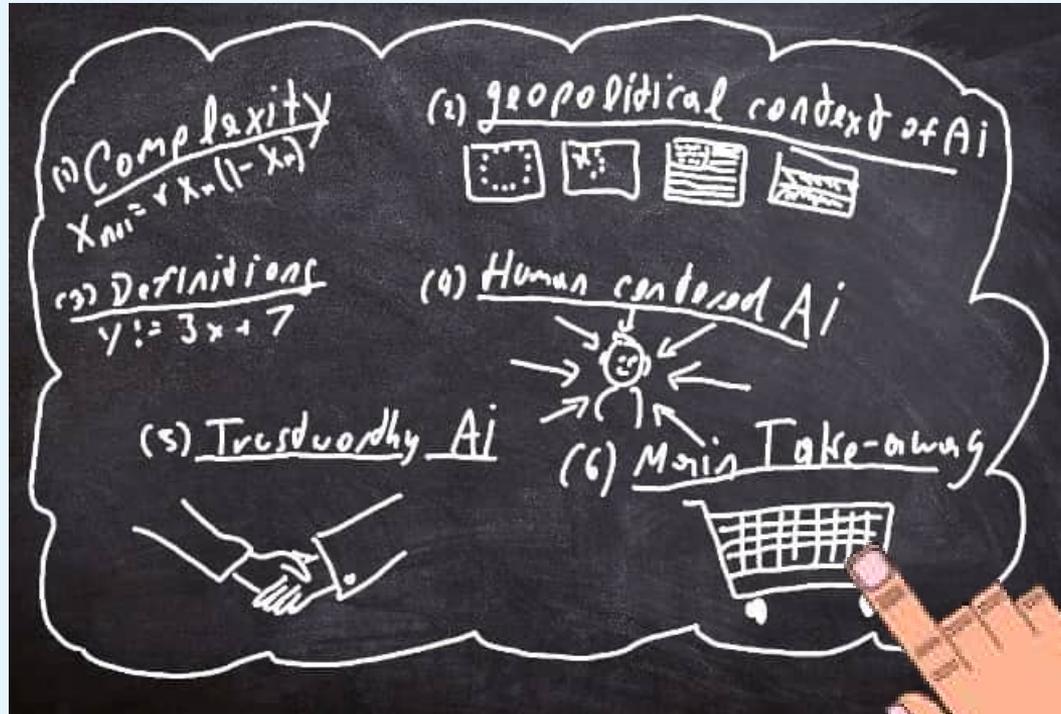
EU guidelines on trustworthy AI

- ▲ lawful
- ▲ ethical
- ▲ robust (technically and socially)

Seven key requirements

- ▲ Human agency and oversight
- ▲ Technical robustness and safety
- ▲ Privacy and data governance
- ▲ Transparency
- ▲ Diversity, non-discrimination and fairness
- ▲ Societal and environmental wellbeing
- ▲ Accountability

Outline



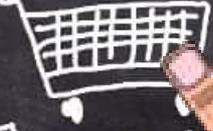
(1) Complexity
 $X_{n+1} = r X_n (1 - X_n)$

(2) geopolitical context of AI


(3) Definitions
 $y := 3x + 7$

(4) Human centered AI


(5) Trustworthy AI


(6) Main Take-away




Summary

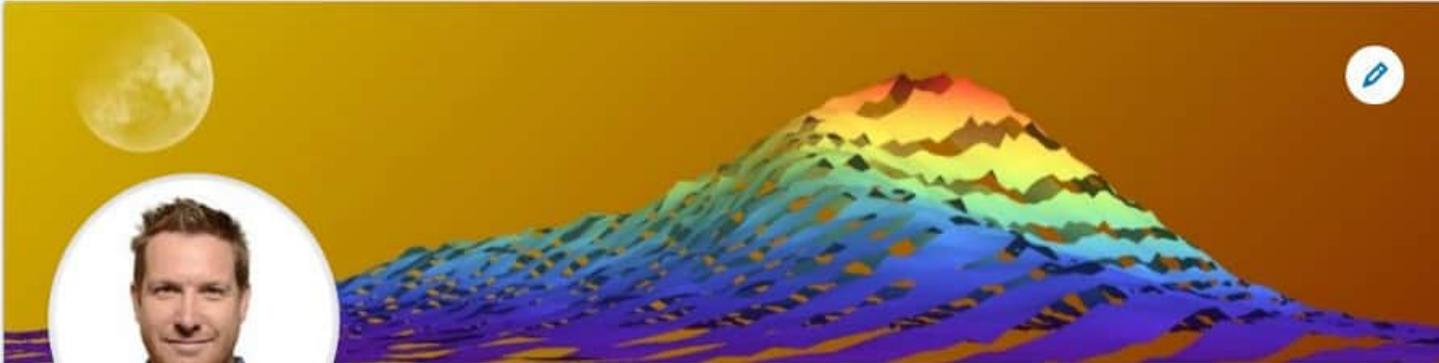
complexity \Rightarrow there is no 'truth'
geopolitics + AI \Rightarrow fierce competition
definition AI \Rightarrow machine learning
human centered AI \Rightarrow different scales
trustworthy AI \Rightarrow lawful, ethical, robust

\Downarrow

talk to each other, discuss,
educate, support, engage



Connect with me on LinkedIn and start a discussion!



Andreas Windisch

Theoretical Physicist, AI specialist, high performance computing enthusiast and passionate private pilot

Graz, Styria, Austria · [500+ connections](#) · [Contact info](#)



Know-Center GmbH

Vertrauenswürdige AI im BRZ

1. Vernetzungstreffen des AI-Policy Forums zu KI in der Verwaltung

Artificial Intelligence, Bundesrechenzentrum

Dipl.-Ing.ⁱⁿ Dr.ⁱⁿ Katharina Birner

Mag. Wolfgang Kabelka

Festsaal Technisches Museum Wien

19. Oktober 2022

Die Verwaltung muss Maßstäbe beim Einsatz von KI setzen **BRZ**

Fünf Gründe für ein Engagement der Verwaltung

1.

Öffentliche Diskussion anregen und Maßstäbe setzen

2.

Wirksame Kontrolle für den Einsatz von künstlicher Intelligenz sicherstellen

3.

Unabhängige Prüfkompetenz etablieren

4.

Algorithmenkompetenz auf allen Ebenen der Verwaltung sicherstellen

5.

Wettbewerb im Einsatz von künstlicher Intelligenz stärken (d. h. keine Monopole)

Das sind aktuelle Herausforderungen

Stakeholder Workshop – Adopt-AI Studie – EU DG CNECT

- Ressourcen-Engpässe
- Mangelndes Wissen zur Formulierung von Anforderungen und Projektumsetzung
- Unklare Kost/Nutzen-Erwartung im Hinblick auf AI-Einsatz
- Mangel an klaren Grundsätzen welche Art von Anwendungsfällen umgesetzt werden können

Beschaffung

Daten

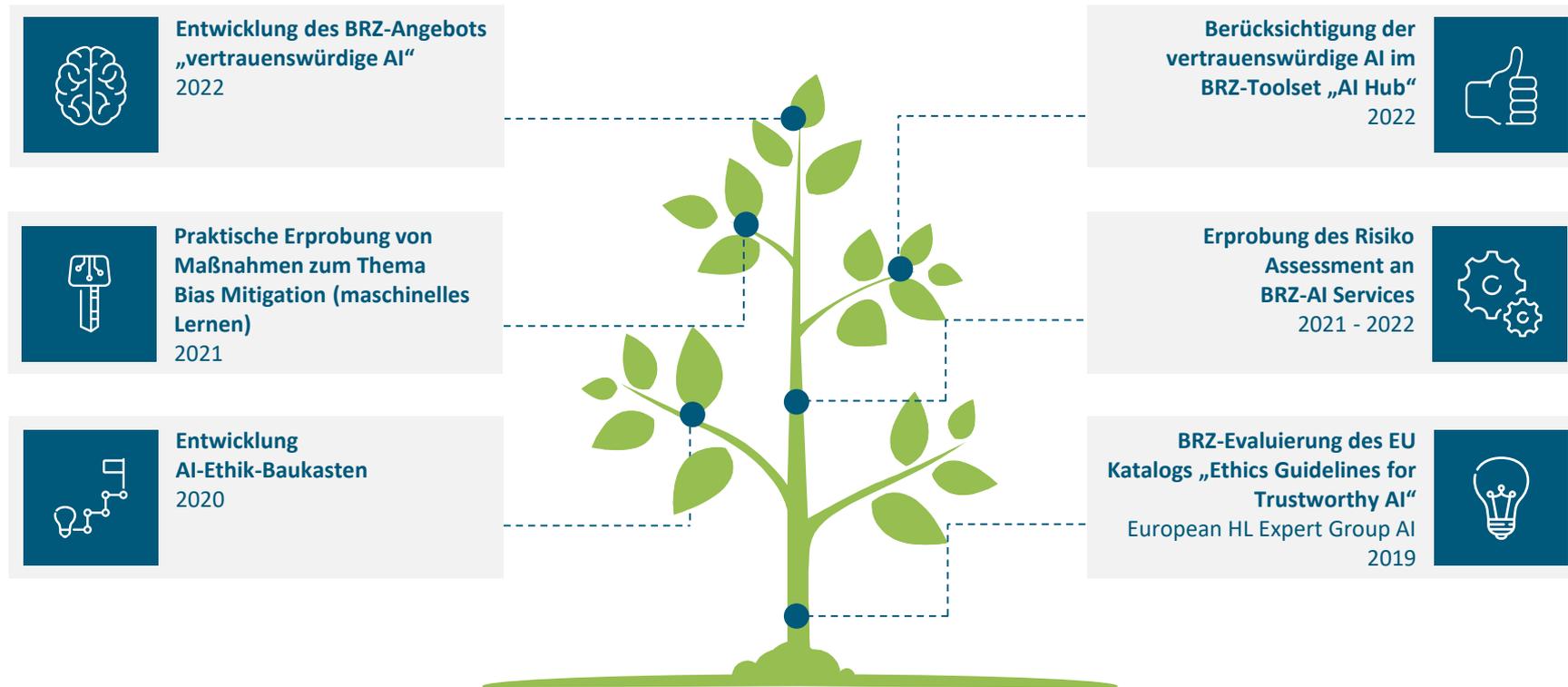
- Datenschutz hindert Erstellung von Trainingsdaten
- Unklares Datennutzungsrecht
- Fehlende Datenstrategie für öffentlichen Sektor
- Fehlendes Wissen zum Thema Datenqualität

- Mangelndes Vertrauen in Ergebnisse
- fehlende Zertifizierung
- AI-Effekte sind schwer abschätzbar
- Verantwortlichkeit einer Bundesverwaltung zählt höher als die der Privatwirtschaft

Öffentlicher Sektor

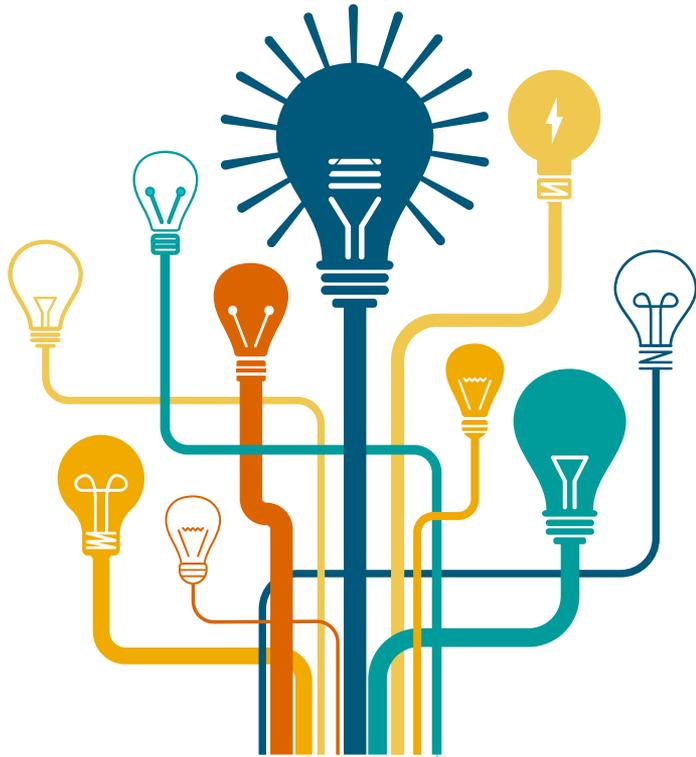
So kann das BRZ AI-Team unterstützen

Die Aktivitäten der letzten 1.111 Tage



Haben sie schon bedacht, dass ...

Beispiele für das kleine 1x1 der vertrauenswürdige AI



exemplarische Fragestellungen

- > Offenlegung der Eigenschaften des verwendeten Algorithmus/Modells. War das Modell schon im Einsatz; wurde es getestet?
- > Ist die Funktionsweise verständlich und anschaulich dargestellt?
- > Wurde eine Mediationsstelle zur Bereinigung und Klärung von Zugänglichkeitskonflikten eingerichtet?
- > Ist die Verantwortung in Bezug auf Externe (betroffene Personen oder Nutzer:innen) geklärt?
- > Wird die Wechselwirkung eines Systems mit lernenden Komponenten mit seiner Umwelt überwacht?
- > Wurden Anforderungen, Ziele und Aufgabenbeschreibungen auf implizite und explizite diskriminierende Faktoren untersucht?

Noch ist Zeit für Maßnahmen (AI Act) ...

Weitergabe von BRZ-Wissen

Geplante Kundenvorhaben werden vor Umsetzung auf ethische Compliance evaluiert

Compliance-Evaluierung der 5 Bereiche:

- Transparenz
- Verantwortung
- Datenschutz
- Zuverlässigkeit
- Gerechtigkeit

Aufzeigen von potentiellen Risiken bei der Verarbeitung von Daten und Informationen



Ergebnis: Compliance-Evaluierungsergebnis inklusive Vorschläge für Maßnahmen zur Mitigation identifizierter Risiken

Nutzen



Erkennen kritischer Aspekte in Kundenvorhaben



Wissen zur vertrauenswürdigen Datenverarbeitung



Proaktiv Maßnahmen ergreifen



Beitrag zur Compliance des Bundes



> **Dipl.-Ingⁱⁿ Drⁱⁿ Katharina Birner**

Technical Consultant

katharina.birner@brz.gv.at

> **Mag. Wolfgang Kabelka**

Digital Advisor

wolfgang.kabelka@brz.gv.at



@brz_gmbh



@Bundesrechenzentrum



@Bundesrechenzentrum



@Bundesrechenzentrum



@Bundesrechenzentrum

Digitale Verwaltung und Ethik

Ein Projekt mit dem Fokus auf die öffentliche Verwaltung

Ralf M. Tatto
BMKÖS

Peter Biegelbauer
AIT AI Ethics Lab

Digitalisierung und Verwaltung

- Erwartungen in der und an die Verwaltung zu Digitalisierung und KI Systemen:
 - Rationalisierung von Arbeitsprozessen
 - Leistungsverbesserungen
 - Vereinfachung für Bürger:innen
- Einsatz von KI Systemen erfordert u.a.:
 - Schutz von Persönlichkeitsrechten bzw. Grundrechten
 - Schutz vor Verzerrung durch Bias und Diskriminierung
 - Transparenz und Nachvollziehbarkeit von Entscheidungen
 - Allgemeiner: Ethik im Einsatz von KI

Handlungsnotwendigkeiten in der Verwaltung

- Viele Aktivitäten in Ö zum Thema – diese Netzwerkveranstaltung ist dem gewidmet
- Reichlich Anwendungsbeispiele
 - Non-Stop-Verfahren (AVG, E-GovG)
 - Automatisierte Erledigungen (AVG, VwGVG, VStG)
 - Verpflichtender elektronischer Verkehr zw. Behörden (E-GovG)
 - Chatbot-Antrag und –Manuduktion (AVG)
 - Gebührenfreie Registerabfragen (AVG, GebG, GGG)
 - Ersetzendes Scannen (E-GovG)

AI findet sich zur Bewältigung komplexer Abläufe im Anwendungs-Dreieck von Recht – Nutzbarkeit – Sicherheit.

Moving Targets: die Regulierung von KI

- Regulierung von KI in vielen Ländern (z.B. US, China, UK, NL)
- Hohes Aktivitätsniveau auf EU-Ebene:
 - Digital Services Act, Digital Market Act, Data Act, AI Act etc.
 - EU hat internationale Vorreiterfunktion seit DSGVO
- Unmittelbare Auswirkungen auf österreichische Verwaltung, bspw. verpflichtendes Risikomanagement (i.e. Folgenabschätzung) von KI Tools
- Besondere Rolle staatlicher Verwaltung:
 - Wahrung des Vertrauens der Bürger:innen in den Staat
 - Einhalten neuer Regulierung sich rasch entwickelnder Technologien

Ziele des Projektes

- Ethische Standards für den Umgang mit digitalen Zukunftstechnologien in der Bundesverwaltung
- Reflexionsrahmen für den Einsatz von digitalen Technologien in Österreich
- Rahmenbedingungen für Folgenabschätzungen und dazugehörige Prozess- und Strukturnotwendigkeiten
- Standard für die Aus- und Weiterbildung von Verwaltungsbediensteten schaffen



https://commons.wikimedia.org/wiki/File:Target_Flat_Icon.svg

Projekt-Output

- Praxisleitfaden für den ethischen Umgang mit Digitalisierung und KI im Verwaltungskontext
- Curriculum zur Aus- und Weiterbildung mit dem Fokus auf KI-Ethik im Verwaltungseinsatz
- Beitrag zur Vernetzung unterschiedlicher verwaltungsbezogener Initiativen



Quelle: GettyImages

Workshops

- **Kick-off Workshop Verwaltung**
 - Themen vorstellen, Diskussionsrahmen herstellen
- **1. Workshop Wissenschaft**
 - Voraussetzungen für verantwortungsvolle Einführung von KI in der Verwaltung
- **2. Workshop Verwaltung**
 - Vertrauen, Ethische Prinzipien, Standards und rechtliche Rahmenbedingungen
- **3. Workshop NGOs**
 - Voraussetzungen für verantwortungsvolle Einführung von KI in der Verwaltung
- **4. Workshop Verwaltung**
 - Kosten-Nutzen Abwägung, Folgenabschätzungen, Risiko Assessment von KI

Einige Ergebnisse Kick-off Workshop

- Ziele: Ermittlung der zentralen Themen, Fragen nach Rahmenbedingungen & Kompetenzen
- Transparenz und Nachvollziehbarkeit im Hinblick auf Verwaltungsentscheidungen, insbes. bei Maschinenlernen Anwendungen (Stichworte Rechenschaftspflicht und Fairness)
- Diskriminierung bei KI Systemen durch Bias in Grundannahmen, Modellen und Datensätzen sowie Forderung nach „human in the loop“ und Entscheidungs**unterstützung**
- KI-Expert:innen sollten in jeder Sektion vorhanden sein, insb. zur Nutzung von Domänenwissen bei Einführung und (Fort)-Entwicklung von KI-Anwendungen

Danke für Ihre Aufmerksamkeit!

Kontakt:

ALⁱⁿ Mag.^a Ursula Rosenbichler

BMKÖS III/C/9

ursula.rosenbichler@bmkoes.gv.at

Kontakt:

PD Dr. Peter Biegelbauer

AIT AI Ethics Lab

peter.biegelbauer@ait.ac.at



Akzeptanz und Vertrauen in KI in der Verwaltung

Der „Rules as Code“-Ansatz und das Projekt „Grants4Companies“

Björn Lellmann, BMF, Abt.V/B/5
AI Policy Forum, Wien
19.10. 2022

Künstliche Intelligenz in eGovernment-Verfahren

Die öffentliche Verwaltung braucht gesicherte und nachvollziehbare Entscheidungen!

B-VG Art. 18 (1) „Die gesamte staatliche Verwaltung darf nur auf Grund der Gesetze ausgeübt werden.“!

Für die Akzeptanz der verwendeten Methoden bei den Betroffenen ist insbesondere notwendig:

- **Transparenz:** Die für die Entscheidung verwendeten Methoden und Grundlagen müssen offen einsehbar sein
- **Nachvollziehbarkeit:** Der Prozess der Entscheidungsfindung muss verständlich und erklärbar sein.

Künstliche Intelligenz - Ansätze

Subsymbolische KI

Stärken insb.:

Lernen von Regeln aus Daten,
Bilderkennung, Spracherkennung,
Big Data, Deep learning,
etc

Problem hier: schwer nachvollziehbar,
basiert auf historischen (potentiell
voreingenommen) Daten
Ergebnis mit Unsicherheitsfaktor

Symbolische KI

Stärken insb.:

Schlussfolgern mit gegebenen Regeln,
Logische Systeme,
Ontologien,
Intelligent Automation, etc

Essentiell hier: Gesicherte,
nachvollziehbare, transparente
Entscheidungen

Dieser
Vortrag

Zukunftsperspektive: Rules as Code

Vision:

Wo möglich sind Regeln/Gesetze

- öffentlich einsehbar
- in einer formalisierten, logikbasierten, ausführbaren Regeldarstellung
- mit einer klar definierten Semantik gegeben.

Regeln/Gesetze und deren ausführbare kodierte Darstellung haben denselben Stellenwert.

Transparenz durch

- Öffentlich einsehbare Regelbasis
- Klar definierte Semantik

Nachvollziehbarkeit durch regelbasierte Darstellung und Erklärungen der Entscheidungen

Parallele Erstellung der Regeln/Gesetze erlaubt:

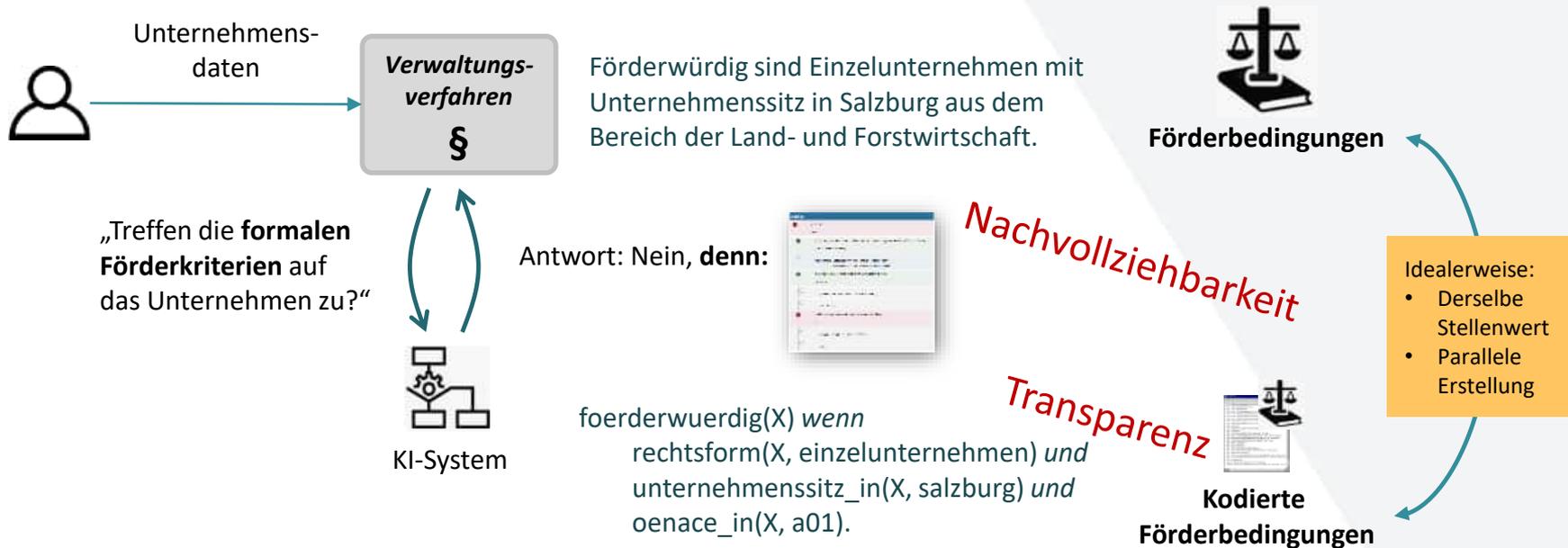
- Klarere Regeln
- Fehlervermeidung
- Zeitersparnis

Schnelle Reaktion auf Krisen durch flexibel anpassbare Regelbasis

Anwendungsmöglichkeiten



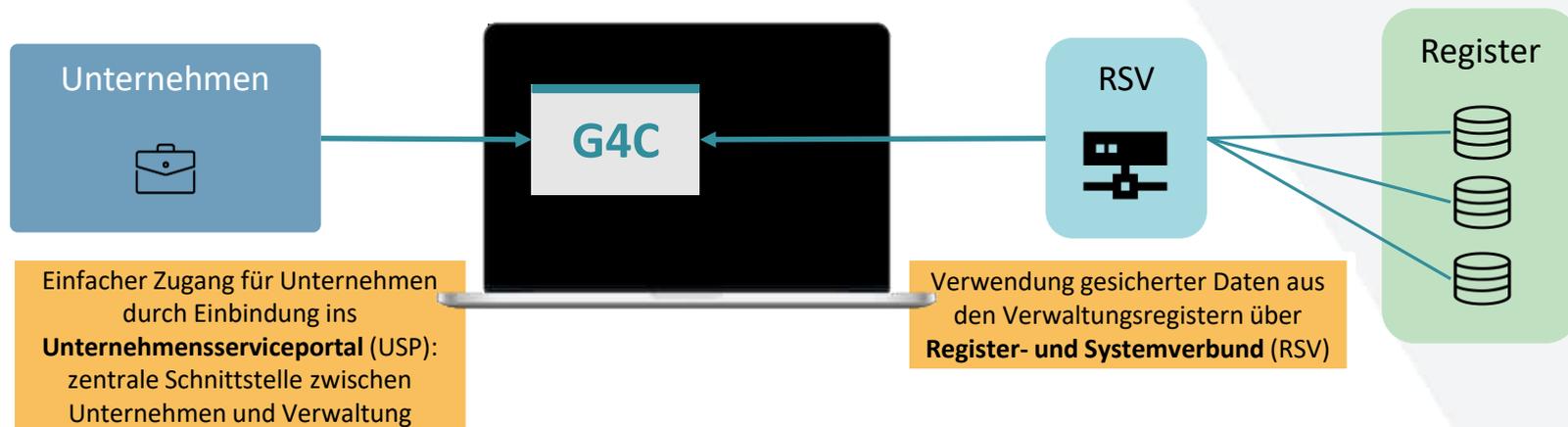
Rules as Code am Beispiel Unternehmensförderungen



Grants4Companies

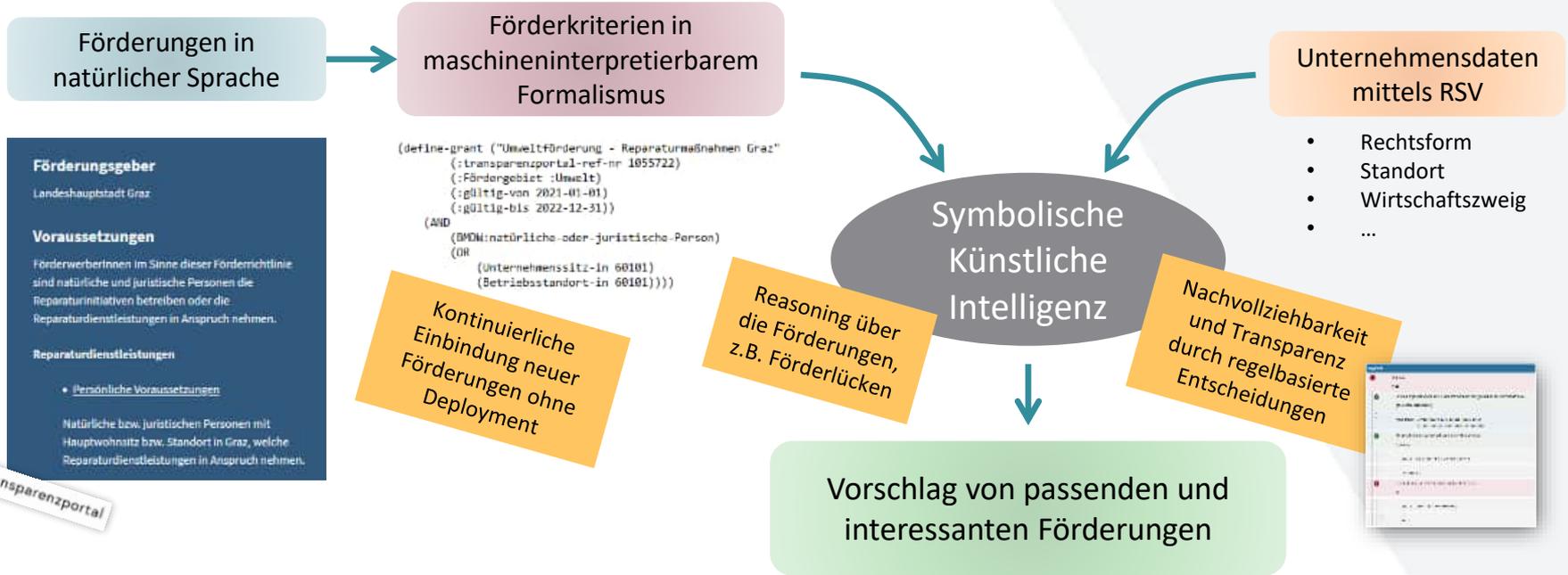
Unternehmensförderungen: Synergien von USP und RSV

Unternehmen sehen sich **Herausforderungen bei der Suche nach Förderungen** gegenüber:
Eine Vielzahl an Fördergebern und komplexe Förderbedingungen erschweren die Suche.



Grants4Companies nutzt symbolische Künstliche Intelligenz, um passende Förderungen vorzuschlagen.

Grants4Companies: Übersicht



Grants4Companies



Maschineninterpretierbare Darstellung von bereits 26
Unternehmensförderungen. Kontinuierliche Erweiterung geplant.

Bewertung der formalen Förderkriterien und Anzeige der Förderungen nach:
„treffen zu“ / „treffen nicht zu“ / „treffen potenziell zu“

Unternehmensdaten: über RSV derzeit aus Unternehmensregister und
Firmenbuch. Erweiterung auf weitere Register nach Verfügbarkeit geplant

G4C hat reinen **Informationscharakter**: Die Entscheidung über Gewährung
der Förderungen liegt rein bei den Förderstellen

Proof of Concept wurde mit dem **eGovernment Award 2021** ausgezeichnet

GoLive im Unternehmensserviceportal: Q4 2022

Kernpunkte

- **Akzeptanz und Vertrauen** in KI in der Verwaltung durch **transparente und nachvollziehbare** Entscheidungen
- Die Methoden der **symbolischen KI** können derartige Entscheidungen liefern
- Die Anwendung **Grants4Companies** illustriert die Vorteile dieser Methoden und bringt Mehrwert für Unternehmen
- Vision des „**Rules as Code**“-Ansatzes: Regeln/Gesetze und deren formalisierte und ausführbare Form haben denselben Stellenwert und werden parallel erstellt

Kontaktaufnahme bezüglich Anmerkungen, Anwendungsmöglichkeiten,
oder Erfahrungsberichten ist jederzeit willkommen!

Vielen Dank!

Dr. Björn Lellmann
Bundesministerium für Finanzen
Abteilung V/B/5
Bjoern.Lellmann@bmf.gv.at

