Initiative „vertrauenswürdige KI“ im Rahmen von aws KIplus
in Kooperation zwischen aws und Plattform Industrie 4.0

Die Idee ist, österreichischen Unternehmen die Teilnahme an Gremien im Bereich Normung & Standardisierung zu ermöglichen. Durch die aktive Teilnahme in Arbeitsgruppen haben die Unternehmen die Chance auf standardisierte Vorgehensweisen bei der Softwareprogrammierung und damit bei Produkten, Prozessen und Dienstleistungen, die auf KI basieren, wodurch Wettbewerbsvorteile generiert werden.
Mit Hilfe der Plattform soll ein Konzept erarbeitet werden, wie man österreichische Unternehmen bei der Teilnahme an Gremien unterstützen kann. Dabei soll nicht nur Trustworthiness, sondern auch Security und Data Science adressiert werden.
Man startet als Mitglied in einem nationalen Gremium und kann sich als Entsandter von Österreich delegieren lassen.

Ausgewählte Gremien:

Austrian Standards AG 001 42 - Artificial Intelligence (Spiegelgremium zu ISO/IEC JTC 001/SC 42, -SC 22, -SC 38)

AG2 Data Science[1]
AG3 Trustworthiness
AG4 Use Cases

Die Digitalisierung der Industrie hat den Bereich der IT Standardisierung verändert:

- Nicht-Technologische Anforderungen wie ethische und gesellschaftliche Aspekte und die Möglichkeit zur Schaffung von vertrauenswürdigen Systemen sind wesentliche Eckpunkte.
- Die Vielfalt der Stakeholder hat deutlich zugenommen (z.B. Regulierungsbehörden, Sozialwissenschaftler etc.)
- Frühzeitiges Engagement der verschiedenen Interessengruppen ist zur üblich geworden.
- Die Anzahl der IT Use Cases hat stark zugenommen.
- Das Verstehen von Anwendungen, der Nachweis von Business Cases und die Entwicklung von Standards erfolgen nun zeitgleich.
- Das "Daten-Ökosystem" ist genauso wichtig wie Hardware, Software und Betriebstechnologien.

Alle 2-3 Wochen findet für die Mitglieder eine Arbeitssitzung statt, wo aktive Teilnahme und Ergebnisse verlangt werden.

Kontakt: Dr. Karl Grün, Austrian Standards, k.gruen@austrian-standards.at

[1] Nicht technische Aspekte, sondern die Entwicklung von KI

ISO/IEC JTC1 / SC42 Artificial intelligence

WG 2 Big Data

WG 3 Trustworthiness

WG 4 Use Cases

-Betrachtung eines breiten Spektrums von Themen im Zusammenhang mit Vertrauenswürdigkeit, Sicherheit und Datenschutz im Kontext von KI.
-Wichtige Stakeholder sehen dies als einen notwendigen Bereich für den Erfolg und die breite Marktakzeptanz von KI.
-Häufig im Kontext von KI-Anwendungsbereichen diskutiert. Internationale Standards könnten enorm helfen. Standards können Fragen der KI-Ethik und gesellschaftliche Bedenken entschärfen und ermöglichen eine breite und schnellere Einführung.
-Aus einer breiteren Branchenperspektive liegt das Interesse an KI in den aktuellen und wachsenden Anwendungsbereichen (Use Cases)
-Use cases sind die "Währung" zwischen SDO-Gremien.
-Seit der Freigabe wurden über 130 Use Cases gesammelt.
-Die ethischen und gesellschaftlichen Belange im Kontext der Use Cases werden berücksichtigt.
-WG2 Big Data: Referenzarchitektur (Rahmen und Anwendungsprozess), AI (Prozessmanagementrahmen für Big Data Analysen)
-WG3 Trustworthiness: AI (Bias in AI-Systemen und AI unterstütztes decision making, Bewertung der Robustheit neuronaler Netze, Risk Management, Überblick über ethische und soziale Angelegenheiten)
-WG4 Use Cases: AI Use Cases

ISO/IEC AWI TR 24368 Overview of ethical and societal concerns[2]

JWG 1 - ISO/IEC JTC1/SC 40: Governance implications of AI[3]

Was sich in einem Unternehmen durch den Einsatz von KI verändert.
Die Notwendigkeit, sich mit den Governance-Implikationen für den Einsatz von KI in Organisationen auseinanderzusetzen ist von größter Bedeutung geworden.
Die Motivation ist, den Vorständen und Führungskräften von Unternehmen zu helfen, wichtige Fragen zu KI-Technologien zu stellen und zu beantworten.
Durch die Kombination der Expertise des SC 42, der sich mit dem gesamten KI-Ökosystem befasst, mit der des SC 40, der sich mit IT-Governance beschäftigt, wurde eine gemeinsame Arbeitsgruppe gebildet um einen ISO/IEC-Standard zu den Governance-Implikationen von KI zu entwickeln.

[2] TR – Technical Report
[3] JWG - Joint Working Group

IEEE C/S2ESC - Software & Systems Engineering Standards Committee

EMELC-WG - Engineering Methodologies for Ethical Life-Cycle Concerns Working Group

IEEE P7000 - IEEE Draft Model Process for Addressing Ethical Concerns During System Design[4]

IEEE ECPAIS Zertifizierung - Ethics Certification Program for Autonomous and Intelligent Systems, Transparency, Accountability, Algorithmic Bias, Privacy, environmental sustainability[5]

Gütesiegel für ethische Aspekte:

The ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS), für die Entwicklung von AI ethischen Zertifizierungskriterien für verantwortungsbewusste Innovation und autonome intelligente Systeme. Die Kriterien betreffen Transparenz, Verantwortlichkeit, algorithmische Bias und Privacy. Die Zertifizierung kann für sowohl private, als auch öffentliche Institutionen angewendet werden, um die Vertrauenswürdigkeit und Vorteile der Nutzung von intelligenten Systemen gegenüber ihren Kunden, MitarbeiterInnen und der Öffentlichkeit darzustellen und Vertrauen aufzubauen. Die Regierung kann die Zertifizierung zu Informationszwecken für die Politik und Öffentlichkeit nutzen und damit zeigen, wie sie autonome intelligente Systeme nutzen.
Zu den Aspekten wurde noch Interesse an einer Gründung einer Projektgruppe zu Environmental Sustainability ausgedrückt.

IEEE AIS Financial Services Playbook konzentriert sich darauf, abstrakte Konzepte von Ethik, Vertrauen und Fairness in praktische Werkzeuge zu überführen, um eine faire, verantwortliche und transparente Nutzung von KI-Systemen zu fördern.
Link: The IEEE Trusted Data & Artificial Intelligence Systems (AIS) Playbook for Finance Initiative

IEEE Ethically Aligned Design v2 schuf die globale Perspektive, welche viele KI-Ethik Systemprinzipien weltweit beeinflusst hat. Dieses Projekt wurde initiiert durch die IEEE Global Initiative für KI-Ethik Systeme. Die Teilnahme ist kostenlos.
Link: Ethically Aligned Design v2

IEEE P2863 Interne Governance Struktur von Unternehmen, bei KI-Implementierungen

Hierbei geht es um Governance-Kriterien wie Sicherheit, Transparenz, Rechenschaftspflicht, Verantwortung und Minimierung von Bias sowie Prozessschritte für eine effektive Implementierung, Leistungsprüfung, Schulung und Compliance bei der Entwicklung oder Nutzung von künstlicher Intelligenz in Organisationen.

IEEE 7010-2020:

Genehmigter Entwurf einer empfohlenen Praxis zur Bewertung der Auswirkungen von autonomen und intelligenten Systemen auf das menschliche Wohlbefinden. Die empfohlene Vorgehensweise ist verankert in wissenschaftlich validen Indizes zum Wohlbefinden und basiert auf einem Prozess der Einbeziehung von Interessengruppen. Der Zweck der empfohlenen Praxis ist, die Produktentwicklung zu leiten, Bereiche für Verbesserungen zu identifizieren, Risiken zu managen, die Leistung zu bewerten und beabsichtigte und unbeabsichtigte Benutzer, Verwendungen und Auswirkungen auf das menschliche Wohlbefinden durch autonome und intelligente Systeme zu identifizieren.

[4] Ist bereits kurz vorm Abschluss, den Draft bekommen wir.

[5] Die Gruppe ist in Vorbereitung, neu dabei ist nun der Aspekt der Umweltverträglichkeit, wo mit uns ein neues Projekt gestartet werden könnte, zudem wäre es auch möglich zu einem der ersten 4 Aspekte eine Manufactury Gruppe zu gründen;

IEEE P7005 - Transparente Employer Data Governance (Datenschutz für Beschäftigte)

Der Standard definiert spezifische Methoden, um Arbeitgebern bei der Zertifizierung zu helfen, wie sie den Zugriff auf das Sammeln, Speichern, Nutzen, Teilen und Vernichten von Mitarbeiterdaten angehen.

IEEE P7006 Standard for Personal Data Artificial Intelligence (AI) Agent

Dieser Standard beschreibt die technischen Elemente, die erforderlich sind, um eine personalisierte Künstliche Intelligenz (KI) zu erstellen und ihr Zugang zu gewähren, der Input, Lernen, Ethik, Regeln und Werte umfasst, die von Einzelpersonen kontrolliert werden.

IEEE P7008 Standard for Ethically Driven Nudging for Robotic

Intelligent and Autonomous: "Nudges", wie sie von robotischen, intelligenten oder autonomen Systemen gezeigt werden, sind definiert als offene oder versteckte Vorschläge oder Manipulationen, die das Verhalten oder die Emotionen eines Benutzers beeinflussen sollen. Dieser Standard legt eine Abgrenzung typischer Nudges fest (die derzeit verwendet werden oder erstellt werden könnten). Sie enthält Konzepte, Funktionen und Vorteile, die notwendig sind, um ethisch begründete Methoden für das Design von robotischen, intelligenten und autonomen Systemen, die sie einbeziehen, zu etablieren und sicherzustellen.

IEEE P7009 Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems

Dieser Standard legt eine praktische, technische Basis von spezifischen Methoden und Werkzeugen für die Entwicklung, Implementierung und Verwendung von effektiven Fail-Safe-Mechanismen in autonomen und teilautonomen Systemen fest. Dieser Standard beinhaltet (aber nicht nur): klare Verfahren zum Messen, Testen und Zertifizieren der Fähigkeit eines Systems, auf einer Skala von schwach bis stark, sowie Anweisungen zur Verbesserung im Falle einer nicht zufriedenstellenden Leistung.

IEEE P7011 Standard for the Process of Identifying and Rating the Trustworthiness of News Sources

Dieser Standard stellt halbautonome Prozesse zur Verfügung, die Standards verwenden, um Bewertungen von Nachrichtenanbietern zum Zwecke der öffentlichen Wahrnehmung zu erstellen und zu pflegen. Er standardisiert Prozesse zur Identifizierung und Bewertung der sachlichen Richtigkeit von Nachrichten, um eine Bewertung von Online-Nachrichtenanbietern und dem Online-Teil von Multimedia-Nachrichtenanbietern zu erstellen. Dieser Prozess wird verwendet, um Vertrauens-Scorecards durch facettenreiche und mit mehreren Quellen versehene Ansätze zu erstellen. Der Standard definiert einen Algorithmus, der Open-Source-Software und ein Scorecard-Bewertungssystem als Methodik für die Bewertung der Vertrauenswürdigkeit verwendet, um Vertrauen und Akzeptanz zu schaffen.

IEEE P7012 Standard for Machine Readable Personal Privacy Terms

Der Standard identifiziert/adressiert die Art und Weise, wie persönliche Datenschutzbestimmungen angeboten werden und wie sie von Maschinen gelesen und akzeptiert werden können.

IEEE P7014 Standard for Ethical considerations in Emulated Empathy in Autonomous and Intelligent Systems

Dieser Standard definiert ein Modell für ethische Überlegungen und Praktiken zur Erstellung und zum Einsatz von empathischer Technologie, die Systeme mit Fähigkeiten umfasst, die affektive Zustände (zB Emotionen, kognitive Zustände) identifizieren, zählen, darauf reagieren oder simulieren können. Dies beinhaltet die Abdeckung von 'affective computing', 'emotion Artificial Intelligence' und verwandten Gebieten.

IEEE P2671 - Allgemeine Anforderungen an die Online-Erkennung auf Basis von Machine Vision (Bildverarbeitung) in der intelligenten Fertigung

Dieser Standard spezifiziert durch die allgemeinen Anforderungen der Online-Erkennung auf Basis der Bildverarbeitung, einschließlich der Anforderungen an das Datenformat, die Datenübertragungsprozesse, die Definition von Anwendungsszenarien und die Leistungsmetriken zur Bewertung der Wirkung des Einsatzes der Online-Erkennung.

IEEE P2672 – Leitfaden für die allgemeinen Anforderungen der flexibilisierten Massenproduktion

Dieser Leitfaden enthält die Definitionen, Terminologie, Betriebsverfahren, Systemarchitekturen, wichtigsten technologischen Anforderungen, Datenanforderungen und Anwendungen von und im Zusammenhang mit der benutzerorientierten Massenproduktion. Dieser Leitfaden bietet Referenzinformationen, die von Produktionsunternehmen für die Gestaltung und Implementierung von Geschäftsmodellen der flexibilisierten Massenproduktion verwendet werden können.

Stand: November 2021

Nikolina Grgic, MSc
Senior Referentin, Plattform Industrie 4.0
+43 (0) 664 883 48 169
nikolina.grgic@plattformindustrie40.at