Künstliche Intelligenz – Transparenz durch katalogbasierte Plattform für Österreich (KITKA)

White Paper – Jänner 2021

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

FH Oberösterreich
Global Business Management
Driven by Global Sales and Marketing

Center for
Human-Computer Interaction
University of Salzburg

ONTEC

FFG
Forschung wirkt.

Kurzzusammenfassung

Dieses White Paper gibt einen Überblick über die Ziele sowie den Fortschritt des Projekts **„Künstliche Intelligenz – Transparenz durch katalogbasierte Plattform für Österreich“ (KITKA)**. Das einjährige Sondierungsprojekt, welches im Februar 2019 gestartet hat, ist im Zuge des Ideen Labs 4.0 entstanden und wird von der FFG gefördert. Das Konsortium, bestehend aus den drei wissenschaftlichen PartnerInnen *Institut für Höhere Studien*, *Fachhochschule Oberösterreich – Global Sales and Marketing*, *Center for Human-Computer Interaction (Universität Salzburg)* und dem Unternehmen *ONTEC AG*, beschäftigt sich mit der Fragestellung, wie Vertrauen in KI-Systeme unter Berücksichtigung ethischer Grundsätze gestärkt werden kann. Mit einem innovativen, multiperspektivischen Ansatz soll das Projekt dazu beitragen, dass die Transparenz von KI-Systemen erhöht und somit eine Vertrauenssteigung erzielt wird, damit österreichische Unternehmen das Potential von KI erkennen und optimal ausschöpfen.

Das Problem mit dem Vertrauen

KI-Systeme besitzen ein großes Potential für österreichische Unternehmen, wobei dieses von vielen nicht ausgeschöpft wird. Ein Mangel an Vertrauen in und Wissen über KI-Systeme werden als wesentliche Barrieren für deren adäquate Nutzung angesehen.

Nur **18%** der Unternehmen haben KI weitgehend in ihre Angebote und Prozesse aufgenommen[1]

[1] Shattuck, S.: People don't trust AI. We need to change that. Toward Data Science, 2019

Der Ansatz von KITKA – Vertrauenssteigerung durch Transparenz

An dem Projekt beteiligt sind: *Institut für Höhere Studien*, *Fachhochschule Oberösterreich – Global Sales and Marketing*, *Center for Human-Computer Interaction (Universität Salzburg)* und *ONTEC AG*

Genau diesen Barrieren möchte das interdisziplinäre Projektteam von KITKA („Künstliche Intelligenz – Transparenz durch katalogbasierte Plattform für Österreich"), ein von der FFG gefördertes Projekt, entgegenwirken.

Übergeordnetes Ziel von KITKA ist es, **Vertrauen in KI-Systeme durch Transparenz, d.h. durch Zurverfügungstellung umfassender Informationen zu steigern.** Im Projekt wird dabei der Ansatz verfolgt, dass eine maximale Vertrauenssteigerung nur durch eine ganzheitliche Beschreibung der KI-Systeme möglich ist. Dies bedeutet, dass die Beschreibung von KI-Systemen aus den Perspektiven **Ethik, Datenschutz, Management und Human-Computer Interaction (HCI)** als ebenso wichtig wie eine detaillierte **technische Darstellung** angesehen wird.

Für die **ganzheitliche Betrachtung** von KI-Systemen wird im Rahmen des Projekts ein **Kriterienkatalog** entwickelt und evaluiert, welcher Informationen aus den verschiedenen Perspektiven berücksichtigt. Die Definition und Evaluierung der Kriterien erfolgt mit Hilfe eines innovativen Methodenmix, durch den mittels partizipativer Methoden sowohl **KI-EntwicklerInnen**, (potentielle) **KI-AnwenderInnen** sowie **ExpertInnen** aus den oben genannten Bereichen einbezogen werden. Außerdem beschreiben zehn KI-EntwicklerInnen ihr KI-System für einen spezifischen Use Case (= Anwendungsfall) entlang des Kriterienkatalogs.

Um diese Informationen zukünftigen, potentiellen KI-AnwenderInnen zugänglich zu machen, wird eine **open-access Plattform** konzipiert, auf der KI-Systeme für spezifische Use Cases entlang des entwickelten Kriterienkatalogs beschrieben werden sollen. Durch die auf der Plattform dargestellten Informationen sollen österreichische Unternehmen, die von KI-Systemen profitieren könnten, über deren Einsatzmöglichkeiten informiert und das Potential für das eigene Unternehmen sichtbar gemacht werden. Die ganzheitliche Beschreibung der KI-Systeme soll das Vertrauen in diese durch größtmögliche Transparenz und Wissensgenerierung steigern sowie realistische Erwartungen an sie wecken. Im Rahmen des Projekts wird ein **Prototyp der geplanten Plattform** exemplarisch für ein KI-System umgesetzt. Die langfristige Vision des KITKA-Projektteams ist es, Informationen über eine Vielzahl von KI-Systemen und Use Cases aus Österreich auf der Plattform bereitstellen zu können.

In Abbildung 1 sind die unterschiedlichen Projektphasen dargestellt.

Erstellung des Kriterienkatalogs

Eine umfangreiche State-of-the-Art **Literaturrecherche** in den Bereichen Technik (Künstliche Intelligenz), Ethik (inklusive DSGVO-Aspekte), Management und Human-Computer Interaction (HCI) wird durchgeführt, um eine wissenschaftlich fundierte Erstfassung des Kriterienkatalogs zur ganzheitlichen Beschreibung von KI-Systemen zu erstellen.

Evaluierung des Kriterienkatalogs

Im Rahmen einer umfangreichen Evaluierung und Weiterentwicklung des Kriterienkatalogs aus mehreren Perspektiven werden verschiedene Stakeholder-Gruppen mittels unterschiedlicher, innovativer, partizipativer Methoden involviert.

Interviews mit KI-EntwicklerInnen

Zehn KI-EntwicklerInnen werden gebeten ihr KI-System bezogen auf einen spezifischen Use Case entlang des Kriterien-katalogs zu beschreiben. Zusätzlich wird die Meinung der EntwicklerInnen hinsichtlich der vorgeschlagenen Kriterien erfasst.

Fokusgruppen mit ExpertInnen

Im Zuge von zwei Fokus-gruppen mit ExpertInnen aus den Bereichen KI, Ethik, Datenschutz, Management und HCI wird der Erstentwurf des Kriterienkatalogs evaluiert und entsprechend weiterentwickelt.

Social Design Lab Workshops

Im Rahmen der eigens dafür entwickelten Methode „Social Design Lab“ (Kombination aus Social Lab und Design Thinking) wird die Angemessenheit der entwickelten Kriterien hinsichtlich Vertrauens-stiftung sowie die Übertragbarkeit der dargestellten KI-Systeme auf den Kontext von potentiellen AnwenderInnen in zwei Workshops überprüft.

Darstellung der Ergebnisse auf einem Prototypen der Plattform

Als Grundlage für den Prototypen der Plattform wird ein strukturelles Plattformkonzept entwickelt und ein dafür angemessenes Design erstellt. Der Prototyp umfasst Informationen zu einem der zehn erhobenen KI-Systeme entlang des in der Projektlaufzeit erarbeiteten Kriterienkatalogs. Des Weiteren wird ein Dynamisierungskonzept erarbeitet, welches Möglichkeiten zur Inbetriebnahme und Erhaltung der Plattform sowie der Skalierung der Use Cases und KI-Systeme aufzeigt.

Abbildung 1. KITKA-Projektphasen

Institut für Höhere Studien Institute for Advanced Studies Vienna
FH Oberösterreich Global Business Management
Center for Human-Computer Interaction University of Salzburg
ONTEC
FFG Forschung wirkt.

Auf dem Weg zu einer ganzheitlichen Beschreibung von KI-Systemen

Basierend auf der Literaturrecherche sowie mehreren Iterationszyklen innerhalb des KITKA-Projektteams ist eine umfangreiche Erstversion des Kriterienkatalogs entstanden. Diese umfasst 161 Fragen zur Beschreibung eines KI-Systems, welche unterschiedliche Stakeholder adressieren (EntwicklerInnen, NutzerInnen oder beide). Die Fragen sind in fünf übergeordnete Kategorien gegliedert: *Technische Aspekte, Wertgenerierung durch KI-Anwendung, Geschäftsabläufe, Ethische Aspekte* und *Beschreibung der Interaktion*.

Die **Interviews** mit den KI-EntwicklerInnen (Cloudfight GmbH, SAS Institute Software GmbH, smec - Smarter Ecommerce GmbH, Xephor Solutions GmbH, Ubitec GmbH, SAIL LABS Technology GmbH, ImageBiopsy Lab GmbH, Ontec AG, Blumatix Consulting GmbH und einem weiteren Unternehmen, das anonym bleiben möchte) resultierten neben der Beschreibung der KI-Systeme für spezifische Use Cases in wichtigen Erkenntnissen hinsichtlich des Kriterienkatalogs und der KITKA-Plattform. Im Folgenden werden zentrale Ergebnisse zusammengefasst.

Mit den insgesamt 161 Fragen ist der Kriterienkatalog recht umfangreich, was auch von mehr als der Hälfte der InterviewpartnerInnen angesprochen wurde. Einerseits wurde angemerkt, dass der Kriterienkatalog somit alle wichtigen Bereiche abdeckt, der Umfang könne aber sowohl für die BereitstellerInnen als auch BetrachterInnen der Informationen überwältigend sein. Um diesen Kritikpunkt zu adressieren, wird bei der nächsten Überarbeitung des Kriterienkatalogs besonderer Wert darauf gelegt, Fragen wenn möglich zu bündeln und nach ihrer Wichtigkeit zu priorisieren und Filterfragen zu formulieren. Dadurch soll eine Reduktion des Umfangs des Kriterienkatalogs erreicht werden, ohne auf relevante Informationen zu verzichten. Für die Darstellung der Informationen auf der Plattform wird eine detaillierte Facettenstruktur ausgearbeitet, durch die die Plattform-NutzerInnen strukturiert durch die Fülle an Informationen geführt werden können.

Alle Befragten waren sich einig, dass Informationen über *technische Aspekte* von KI-Systemen notwendig sind, um Vertrauen aufzubauen. Hinsichtlich der anderen Kategorien ergibt sich kein gleichermaßen eindeutiges Bild. Auf spezifische KI-Systeme treffen manche Kategorien (teilweise) nicht zu. Dennoch scheint keine der fünf erfassten Kategorien gänzlich irrelevant zu sein. Auch wurde keine weitere, noch fehlende Kategorie identifiziert. In einem weiteren Schritt muss genauer untersucht werden, wie eine Zuordnung der Kategorien zu bestimmten KI-Systemen und/oder Use Cases erfolgen kann. Eine Hypothese ist, dass ein Zusammenhang zwischen der Relevanz der Kategorie des Kriterienkatalogs und der Branche, für die das KI-System entwickelt wurde, besteht.

Jene Fragen, die auf ihr KI-System zutreffen, konnten von den EntwicklerInnen größtenteils ohne Probleme beantwortet werden. Die meisten Verständnisprobleme traten bei der Kategorie *Ethische Aspekte* auf. Dies kann unter anderem darauf zurückzuführen zu sein, dass erwähnte Konzepte wie „Umwelt“, „menschliche Autonomie“ oder „Werte“ viel Raum für Interpretation lassen und eine konkrete Antwort bzw. Einschätzung deshalb in manchen Fällen schwierig ist. Für die Überarbeitung des Fragenkatalogs bedeutet dies, dass Begrifflichkeiten noch eindeutiger dargestellt und erläutert werden müssen. Zudem muss genauer untersucht werden, wer zuverlässige Auskunft zu diesen Fragen geben kann.

Eine weitere wichtige Erkenntnis aus den Interviews betrifft das Thema soziale Erwünschtheit. Zwei TeilnehmerInnen wiesen explizit darauf hin, dass soziale Erwünschtheit bei der Beantwortung der Fragen eine große Rolle spielt. Da EntwicklerInnen ein Interesse daran haben, ihre Firma und ihre KI-Systeme vorteilhaft zu präsentieren, sei es gemäß den InterviewpartnerInnen bei manchen Fragen unwahrscheinlich, ausschließlich ehrliche Antworten zu erhalten. Dies betreffe nicht zuletzt die ethischen Fragen. So sei es vor allem in kleinen Teams nicht immer möglich, alle Diversitätskriterien einzuhalten. Eine

Herausforderung, die bei der Weiterentwicklung des Kriterienkatalogs und der Plattform demnach gezielt adressiert werden muss, ist die Identifikation von sowie der Umgang mit sozial erwünschten Angaben.
Die meisten InterviewpartnerInnen wären bereit, die durch den Kriterienkatalog erfassten Informationen auf einer open-access Plattform öffentlich zur Verfügung zu stellen, obwohl nur drei dadurch die Gewinnung neuer KundInnen erwarten. Als weitere Motive wurden unter anderem Öffentlichkeitsarbeit für das eigene Unternehmen, aber auch hinsichtlich KI in Österreich im Allgemeinen, MitarbeiterInnenaquise oder das Einnehmen einer Vorbildfunktion durch Offenlegen jeglicher Informationen über das KI-System genannt.

An den **Fokusgruppen** nahmen insgesamt 12 ExpertInnen aus den Bereichen Künstliche Intelligenz, Ethik und Diversität, Management, Datenschutz und Human-Computer Interaction teil. Die Fokusgruppen, die im Oktober 2020 online stattgefunden haben, hatten als zentrales Thema Vertrauen in KI-Systeme. Entsprechend wurde diskutiert, was für die teilnehmenden ExpertInnen vertrauensstiftend im Kontext Künstliche Intelligenz bedeutet bzw. welche Informationen für die Vertrauensbildung in KI-Systeme aus ihrer Sicht notwendig sind.
Viele Bereiche, die im Kriterienkatalog bereits abgedeckt sind, wurden auch von den ExpertInnen als relevant eingestuft. So basiere Vertrauen laut den ExpertInnen maßgeblich auf einer klaren Darstellung des Ziels, Nutzens und der Kosten des KI-Systems, der Beschreibung der technischen sowie datenschutzrelevanten Aspekte und Informationen zur Konformität des KI-Systems mit Gesetzen und Regularien. Auch ethischen Aspekten wurde eine besondere Bedeutung für die Vertrauensbildung zugeschrieben. Darunter werden sowohl die Berücksichtigung der Menschenrechte und darüber hinaus existierender gemeinsamer gesellschaftlicher Werte als auch eine Abschätzung der Auswirkungen des KI-Systems auf den Menschen und die Umwelt verstanden. Eine besondere Herausforderung stelle die Erfassung dieser Informationen durch konkrete Indikatoren dar. Im Hinblick auf die Folgenabschätzung der KI-Systeme wurde betont, dass eine Momentaufnahme nicht ausreichend ist. Vielmehr müssen die potentiellen Folgen des KI-Systems kontinuierlich neu evaluiert und veröffentlicht werden.
In all diesen Bereichen sei es wichtig, auch die Grenzen des KI-Systems aufzuzeigen. Nur so können sich die ExpertInnen vorstellen, dass realistische Erwartungen geschaffen werden, welche Voraussetzung für ein langfristiges Vertrauen sind.
Weiters stellte sich heraus, dass neben Informationen über das jeweilige KI-System auch andere Faktoren Einfluss auf das Vertrauen in das System haben können. Dazu zählen die Reputation der EntwicklerInnenfirma sowie die vorherrschende KI-Kultur im Unternehmen potentieller AnwenderInnen. Beide Faktoren entwickeln sich über einen längeren Zeitraum und besonders die KI-Kultur muss erlebt anstatt lediglich durch Informationen erläutert werden. Aus diesem Grund betonte ein/e FokusgruppenteilnehmerIn „weiche Faktoren“ in diesem Kontext.

Weiteres Vorgehen

Sowohl die Erkenntnisse, die aus den Interviews gewonnen wurden, als auch jene aus den Fokusgruppen mit ExpertInnen, sind äußerst hilfreich und wertvoll für die Weiterentwicklung des Kriterienkatalogs und der KITKA-Plattform. Deshalb bedankt sich das KITKA-Projektteam nochmals herzlich bei allen TeilnehmerInnen.
Um die Ergebnisse exemplarisch darstellen zu können, wird in einem nächsten Schritt der Prototyp der Plattform erstellt.

Fragen, Anregungen, Ideen, Kooperationen…

… das KITKA Projektteam freut sich über jede Kontaktaufnahme.

Projektpartnerin	Ansprechpersonen
Institut für Höhere Studien (IHS) Josefstädterstraße 39 1080 Wien	Mag. Dr. Elisabeth Frankus frankus@ihs.ac.at Julia Schmid, BA schmid@ihs.ac.at Mag. Milena Wuketich wuketich@ihs.ac.at
Universität Salzburg Center for Human-Computer Interaction Jakob-Haringer-Straße 8 / Techno 5 5020 Salzburg	Hanna Braun, M.Sc. hanna.braun@sbg.ac.at Mag.a Dr.in Alina Krischkowsky alina.krischkowsky@sbg.ac.at
FH OÖ Forschungs & Entwicklungs GmbH Studiengang Global Sales and Marketing Wehrgrabengasse 1-3 4400 Steyr	FH-Prof. DI Dr. Margarethe Überwimmer margarethe.ueberwimmer@fh-steyr.at Alexandra Fratric, BA MA alexandra.fratric@fh-steyr.at
ONTEC AG Ernst-Melchior-Gasse 24/DG 1020 Wien	Tobias Eljasik-Swoboda, M.Sc. tobias.eljasik-swoboda@ontec.at Christian Rathgeber christian.rathgeber@ontec.at