

## Ethische KI

Software Plattform für die Entwicklung von ethischer KI

<b>Programm / Ausschreibung</b>	IWI, IWI, Basisprogramm Ausschreibung 2023	<b>Status</b>	abgeschlossen
<b>Projektstart</b>	01.02.2023	<b>Projektende</b>	31.01.2024
<b>Zeitraum</b>	2023 - 2024	<b>Projektlaufzeit</b>	12 Monate
<b>Keywords</b>			

### Projektbeschreibung

Ethische KI

Künstliche Intelligenz hat enormes Potential. Einen kleinen Teil dessen, was mit KI möglich ist, kann man heute bereits in Anwendungen wie Sprach-Assistenten, Gesichtserkennung oder autonomem Fahren erleben. KI wird die Arbeitswelt grundlegend verändern. Trotz der rasanten technischen Entwicklung von Verfahren zum Maschinellen Lernen wurden ethische Perspektiven bisher kaum berücksichtigt. Selbst großen Unternehmen kamen immer wieder in Situationen in denen KI Modelle von Vorurteilen geprägt waren oder in denen es möglich war, aus Datenbeständen vertrauliche Informationen abzuleiten.

Um zukünftigen ethischen Regularien zu entsprechen, muss die Entwicklung von KI-Anwendungen und Modellen des Maschinellen Lernens völlig neu und unter grundsätzlicher Beachtung ethischer Grundsätze betrachtet werden. Ethische Kriterien müssen von Anfang an in der Entwicklung berücksichtigt werden. Dies stellt eine gesellschaftliche Herausforderung und eine große wirtschaftliche Chance dar.

Das Ziel von Gradient Zero ist es daher, gemeinsam mit universitären Partnern Software-Entwicklungswerkzeuge zu entwickeln, die die Programmierung von KI Anwendungen ermöglichen, welche ethischen Prinzipien folgen. Mit einer solchen KI Anwendung können Unternehmen, die über Datenbestände verfügen, diese Datenbestände unter Einhaltung aller ethischer Kriterien Dritten zur Verfügung stellen.

Im Verlauf des dreijährigen Projektes gilt es, erst in Kooperation mit der Universität Wien, Abteilung für Philosophie, die theoretischen Grundlagen Ethischer KI auszuarbeiten und darauf aufbauend die genannte Entwicklungsplattform sowie Prüfmechanismen für Kriterien Ethischer AI wie Erklärbarkeit, Verhinderung von Vorurteilen, Datenschutz, Transparenz und Nachvollziehbarkeit auszuarbeiten und in Pilotprojekten zu evaluieren.

Das Team verfügt bereits über Erfahrung im Projektbereich und hat in der Vergangenheit bereits einige erfolgreiche ethische KI-Projekte in den Bereichen Gesundheitswesen und Verwaltung umgesetzt.

### Endberichtkurzfassung

Künstliche Intelligenz (KI) birgt wie jede fortschrittliche Technologie Potenziale, aber auch vielseitige Gefahren wie z. B. in den Bereichen Bias und Diskriminierung, Datenschutz und Überwachung sowie hinsichtlich Ethik und moralischer Fragestellungen. Um diesen Gefahren zu begegnen, ist es wichtig, dass die Entwicklung und Anwendung von KI

verantwortungsvoll gestaltet wird.

In diesem Vorhaben gelang in Kooperation mit dem Philosophischen Institut der Universität Wien sowie dem Österreichischen Forschungsinstitut für Künstliche Intelligenz die interdisziplinäre Erforschung und Entwicklung von konkreten Ethik-von-Design-Prinzipien sowie technischen Werkzeugen, die Entwicklern von Anwendungen der Künstlichen Intelligenz dabei unterstützen ihre Entwicklungsprojekte nach ethischen Kriterien auszurichten und diese konkret hinsichtlich der wichtigen Kriterien Privacy, Transparency, Fairness und Impact zu prüfen und zu validieren.

Gemeinsam mit sehr renommierten Pilotkunden und in unterschiedlichen Anwendungsfällen gelang im abschließenden Forschungsjahr eine Validation der durchgeführten F&E. Die wissenschaftlichen Erkenntnisse führen zur Veröffentlichung mehrerer wissenschaftlicher Paper.

Die Entwicklungsergebnisse der Gradient Zero GmbH können nach einer zukünftigen Produktifizierung durch Entwickler von KI-Anwendungen direkt in deren Entwicklungsprozess integriert werden, um das eigene Entwicklungsvorhaben ethischer auszurichten.

Weitere Details können auch auf der Webseite des Forschungsvorhabens unter <https://ekip.ai/> nachgelesen werden.

### **Projektkoordinator**

- Gradient Zero SoftwareentwicklungsgmbH

### **Projektpartner**

- Österreichische Studiengesellschaft für Kybernetik, abgekürzt "ÖSGK"