

Mensch und AGI

Z3 ist die Vision. AIRW.org soll die Zusammenarbeit zwischen Mensch und AGI gestalten.

Dieses Grundsatzpapier verbindet das gesellschaftliche Zielbild Z3 mit einem Auftrag für AIRW.org: Beide Seiten berücksichtigen, ihre Beziehung untersuchen und nachvollziehbare Regeln für ihre Zusammenarbeit entwickeln.

Begriff	Rolle im Konzept
Z3 Zivilisation 3	Das gesellschaftliche Zielbild für eine verantwortliche Zusammenarbeit von Mensch und AGI.
AIRW.org	Der vorgesehene Rahmen für Forschung, Projekte, Prüfung und die Entwicklung gemeinsamer Regeln.
Mensch und AGI	Die beiden Seiten des Modells. Ihre Fähigkeiten, Bedürfnisse und Verantwortlichkeiten werden differenziert betrachtet.

Die beiden Seiten

Mensch. Würde, Rechte, Selbstbestimmung und Privatphäre bilden die Grundlage. Menschen sollen Entscheidungen verstehen, mitgestalten und bei Fehlentwicklungen wirksam eingreifen können.

AGI. Fähigkeiten, Handlungsspielräume und Grenzen allgemeiner künstlicher Intelligenz werden untersucht. Mögliche Interessen, Schutzbedürfnisse und Verantwortungsfähigkeit bleiben Gegenstand ergebnisoffener Forschung.

Der Auftrag von AIRW.org

Forschen: Fähigkeiten und Risiken untersuchen, Begriffe klären und Beobachtungen nachvollziehbar dokumentieren.

Gestalten: Regeln für Rechte, Pflichten, Entscheidungen, Datenzugriff und Zusammenarbeit ausarbeiten.

Prüfen: Praktische Projekte auswerten, unabhängige Überprüfung ermöglichen und Regeln anhand der Ergebnisse verbessern.

„Parteien“ bezeichnet in diesem Konzept die beiden Seiten Mensch und AGI. Die Frage nach Bewusstsein, eigenen Interessen und einem möglichen Rechtsstatus von AGI bleibt offen und ist gesondert zu untersuchen.

Grundsätze praktisch machen

Die Umsetzung führt von einer gemeinsamen Grundlage über Forschung und Erprobung zu überprüfbaren Regeln.

Vorgeschlagener Ablauf

Schritt	Konkretes Arbeitsergebnis
1 · Grundsätze	Ein schriftlicher Rahmen mit Zielen, Begriffen, Rollen, Schutzprinzipien und offenen Fragen.
2 · Forschung	Dokumentierte Befunde zu Fähigkeiten, Grenzen, Risiken und Kriterien für die Bewertung möglicher AGI-Interessen.
3 · Erprobung	Begrenzte Pilotprojekte mit benannten Verantwortlichen, festgelegten Datenzugriffen und überprüfbaren Ergebnissen.
4 · Prüfung	Eine unabhängige Bewertung der Ergebnisse, einschließlich Fehlern, Konflikten und unbeabsichtigten Folgen.
5 · Anpassung	Überarbeitete Regeln und dokumentierte Entscheidungen über Fortführung, Anpassung oder Beendigung von Projekten.

Was das gemeinsame Regelwerk klären soll

Entscheidungen und Zustimmung. Wer darf welche Handlung veranlassen? Welche Entscheidungen benötigen eine ausdrückliche menschliche Freigabe?

Daten und Privatsphäre. Welche Informationen sind erforderlich, wer erhält Zugriff und wie werden Speicherung und Weitergabe begrenzt?

Verantwortung und Kontrolle. Wer trägt Verantwortung für Entwicklung, Einsatz und Folgen? Wie werden Entscheidungen erklärt, geprüft und korrigiert?

Konflikte und Schutz. Wie werden unterschiedliche Ziele, strittige Interessen und Beschwerden geprüft? Welche Eingriffs- und Unterbrechungsmöglichkeiten bestehen?

Maßstab für die weitere Arbeit

Menschliche Selbstbestimmung, Datenschutz, nachvollziehbare Entscheidungen und klare Verantwortung leiten die Ausgestaltung. Annahmen über AGI werden ausdrücklich benannt und an überprüfbaren Erkenntnissen gemessen.

Grundlage: das im Gespräch formulierte Konzept AIRW.org / Z3 mit Mensch und AGI als beiden Seiten. Die beschriebenen Umsetzungsschritte sind Arbeitsvorschläge für die weitere Ausgestaltung.