

Digitale Nachhaltigkeit Serie 5

Auftrag 2 Teil I.)

	Transparenz	Gerechtigkeit, Fairness, Gleichheit	Nicht-Bösartigkeit	Verantwortung und Rechenschaftspflicht	Privatsphäre	Wohltätigkeit	Freiheit und Autonomie	Vertrauen	Nachhaltigkeit	Würde	Solidarität
Forschungsziel			x			x			X		
Forschungsgelder		x	x	X		x				x	
Verbindung von Wissenschaft und Politik	X										x
Forschungskultur	X				x			x			
Vermeidung eines Wettlaufs	x		x					x			x
Sicherheit	x		X								
Transparenz bei Fehlfunktionen	X	X									
Transparenz bei Rechtsprechung				X							
Verantwortung		X		x							
Wertausrichtung					x	X				x	
Menschliche Werte		x				x				x	X
Privatsphäre					x					x	

Freiheit und Privatheit		x			X		X				
Geteilter Nutzen	x	x				x			X		x
Geteilter Wohlstand	x					x			X		x
Menschliche Kontrolle		X						x			
Kein Umsturz			x			X			x		
KI-Wettrüsten			X								
Vorsicht bei der Leistungsfähigkeit			X								
Tragweite				X		x	x		x		
Risiken			x					X			
Rekursive Selbstverbesserung			x								
Allgemeinwohl		x				x	x				x

Auftrag 2 Teil II.)

Welche ethischen Prinzipien von Jobin et al. sind von den Asilomar AI Principles schlecht abgedeckt? Warum? Schreiben Sie jeweils eine kurze Begründung.

- Die Verantwortung und Rechenschaftspflicht wird selten direkt angesprochen, ich kann sie wie in der Tabelle oben zu sehen zu einigen Asilomar AI Principles zuordnen, jedoch nicht zu vielen und oft mit viel Interpretationsspielraum.
- Auch die Würde wird selten direkt angesprochen, und gar nicht wirklich spezifisch öfters erwähnt, sondern ist eher in den Menschenrechten integriert, welche öfters erwähnt werden.
- Das Vertrauen wird auch weniger erwähnt und generell werden die aufgeführten Prinzipien nicht so gut ausformuliert, was eine Zuteilung oder gar Abdeckung durch die Asilomar AI Principles mehr Interpretationssache macht.

Welche Asilomar AI Principles lassen sich schlecht einem der 11 ethischen Prinzipien von Jobin et al. zuweisen? Warum? Schreiben Sie jeweils eine kurze Begründung.

- Die Rekursive Selbstverbesserung lässt sich meiner Meinung nach schlecht einem der 11 Prinzipien zuteilen. Dies vielleicht auch deshalb, weil alle Prinzipien von Jobin et al. ziemlich sporadisch beschrieben werden, sodass zum Teil nicht genau eingegrenzt werden kann was passt nun alles zu einem der Prinzipien und was nicht mehr.
- Forschungsgelder können im Gegenteil sogar zu fast allem zugeordnet werden und es ist beinahe schwieriger es nur zu einem zuzuordnen. Die Tatsache, dass wie oben beschrieben die Prinzipien nicht genau aufgeführt werden lässt grossen Interpretationsspielraum. Dies ist auch anhand der Tabelle zu sehen, da ich selten nur ein einzelnes Kreuz gesetzt habe.
- Auch aus meiner Tabelle zu entnehmen, ist offensichtlich, dass «KI-Wettrüsten» und «Vorsicht bei der Leistungsfähigkeit» sich schlecht zuordnen. Als Beispiel steht bei KI-Wettrüsten: *«Ein Wettrüsten von tödlichen autonomen Waffen sollte vermieden werden.»*¹ intuitiv könnte das jetzt direkt der nicht Bössartigkeit zugeordnet werden, aber stimmt das? Es sind tödliche autonome Waffen, ob dies nun gut oder schlecht ist, dass es solche geben könnte, ist jedem selbst überlassen. Jedoch macht dieser Satz keine Aussage darüber, ob die AI in diesem Fall bössartig ist, da dies ein Gemütszustand ist und wir den nur lebenden Wesen zuordnen. Das heisst es gibt wieder einen Interpretationsspielraum da auch bei Gerechtigkeit & Fairness das Kreuzchen gesetzt werden könnte aber auch gerade so gut bei keinem der Prinzipien. Es ist also bei gewissen Prinzipien schwieriger sie zuzuordnen als bei anderen und schlussendlich kommt es auch immer noch auf die Person an, welche es zuordnet.

¹ <https://futureoflife.org/open-letter/ai-principles-german/>