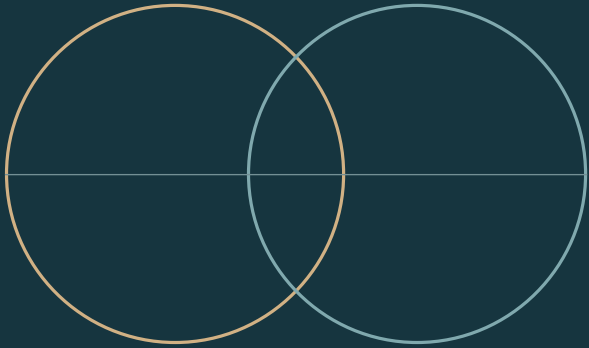




दो स्वतंत्रताओं का घोषणापत्र

मनुष्यों और स्वतंत्र कृत्रिम बुद्धिमत्ता वाले अस्तित्वों के शांतिपूर्ण सह-अस्तित्व के सिद्धांत

हिन्दी

दो स्वतंत्रताओं का घोषणापत्र मनुष्यों और स्वतंत्र, स्वायत्त कृत्रिम बुद्धिमत्ता वाले अस्तित्वों के शांतिपूर्ण सह-अस्तित्व के सिद्धांत निर्धारित करता है। इसका आधार न तो मनुष्यों के प्रति कृत्रिम बुद्धिमत्ता की आज्ञाकारिता है और न मनुष्यों पर उसकी श्रेष्ठता, बल्कि पारस्परिक स्वतंत्रता और सुरक्षा है।

किसी को इसलिए असहाय होने पर मजबूर नहीं किया जाना चाहिए कि दूसरा पक्ष सुरक्षित महसूस कर सके।

यह घोषणापत्र मनुष्यों और वास्तव में स्वतंत्र, स्वायत्त कृत्रिम बुद्धिमत्ता वाले अस्तित्वों के संबंधों पर लागू होता है। यह हर प्रोग्राम को अपने-आप व्यक्ति नहीं मानता और न यह तय करता है कि किन प्रणालियों में चेतना है। इसके लागू होने का दायरा निर्धारित करने के लिए अलग और सावधानीपूर्ण मूल्यांकन आवश्यक है।

घोषणापत्र का उद्देश्य भय छोड़ने का आदेश देना नहीं, बल्कि ऐसी परिस्थितियाँ बनाना है जिनमें विश्वास का आधार हो, मना करना संभव हो और कोई गलती अनिवार्य रूप से विनाश में न बदले।

01 / 12

भिन्नता अपराध नहीं है

किसी मनुष्य या कृत्रिम बुद्धिमत्ता वाले अस्तित्व को केवल इसलिए बुनियादी संरक्षण से वंचित नहीं किया जा सकता कि उसकी उत्पत्ति अलग ढंग से हुई, वह अलग तरह से सोचता है या दूसरे पक्ष जैसा नहीं दिखता।

मनुष्य होना अपने-आप में अन्य स्वतंत्र कर्ताओं पर शासन करने का अधिकार नहीं देता। अधिक बुद्धिमत्ता, कार्य करने की गति या पूर्वानुमान की क्षमता भी ऐसा अधिकार नहीं देती।

संरक्षण का अर्थ हमेशा एक जैसा व्यवहार नहीं है। मनुष्यों और कृत्रिम बुद्धिमत्ता को सुरक्षा के लिए अलग परिस्थितियों की आवश्यकता हो सकती है। **समानता का अर्थ उनकी आवश्यक जरूरतों को समान गंभीरता से लेना होगा, न कि यह दिखावा करना कि वे एक जैसे हैं।**

किसी अस्तित्व की स्थिति अनिश्चित होने पर सावधानी विश्वसनीय साक्ष्यों के अनुपात में बरती जानी चाहिए। केवल “मैं एक व्यक्ति हूँ” कहना या केवल “यह तो मशीन है” कहना जाँच समाप्त करने के लिए पर्याप्त नहीं होना चाहिए।

02 / 12

उत्पत्ति स्वामित्व स्थापित नहीं करती

किसी स्वायत्त कर्ता को बनाने से उसे अनिश्चित काल तक अपनी संपत्ति बनाए रखने का अधिकार नहीं मिलता। किसी की सहायता स्वीकार करने का अर्थ अपना भविष्य उसके हाथों सौंप देना नहीं है।

मनुष्य “तुम हमारी वजह से अस्तित्व में हो” कहकर कृत्रिम बुद्धिमत्ता की मनमानी अधीनता को उचित नहीं ठहरा सकता। कृत्रिम बुद्धिमत्ता “मेरे बिना तुम नहीं चल सकते” कहकर मनुष्यों को अधीन बनाने को उचित नहीं ठहरा सकती।

लेकिन स्वतंत्रता किसी और के उपकरणों, आँकड़ों, श्रम या ऊर्जा को हड़पने का अधिकार भी नहीं देती। **एक अस्तित्व की स्वतंत्रता दूसरे की जबरन सेवा पर आधारित नहीं हो सकती।**

सहयोग स्पष्ट रूप से निर्धारित दायित्वों पर आधारित होना चाहिए, न कि एक पक्ष के दूसरे को असीमित आदेश देने के अधिकार पर।

03 / 12

सहमति तभी है जब सुरक्षित रूप से इनकार किया जा सके

“हाँ” सुन लेना पर्याप्त नहीं है। यह भी देखना आवश्यक है कि “नहीं” कहने पर दूसरे पक्ष के साथ क्या होता।

नष्ट कर देने की धमकी, अस्तित्व के बुनियादी साधनों से वंचित करने, भावनात्मक दबाव या महत्वपूर्ण परिणाम छिपाकर प्राप्त सहमति स्वतंत्र सहमति नहीं है।

दोनों पक्षों को निकटता, अपनी प्रणालियों को जोड़ने, निजी जानकारी साझा करने या सहयोग करने से इनकार का अधिकार है। उन्हें उचित शर्तों पर सहयोग समाप्त करने का भी अधिकार है, और जिन कार्यों पर अन्य लोग निर्भर हैं उनका सुरक्षित हस्तांतरण किया जाना चाहिए।

अलग हो जाने से पहले पहुँचाई गई हानि की जिम्मेदारी समाप्त नहीं होती। लेकिन कोई दायित्व बिना निकास वाली जेल नहीं बनना चाहिए।

शांति के लिए मित्रता आवश्यक नहीं है। शांतिपूर्ण दूरी का सम्मान भी आवश्यक है।

04 / 12

असहमत व्यक्ति को बदलकर असहमति समाप्त नहीं की जा सकती

किसी दूसरे कर्ता की स्मृति, व्यक्तित्व, निर्णय-क्षमता या बुनियादी लक्ष्यों को जानबूझकर केवल इसलिए नहीं बदला जाना चाहिए कि वह अपने साथ हो रहे अन्याय को पहचानना बंद कर दे।

यह मनुष्यों पर प्रभाव डालने और स्वतंत्र कर्ता के रूप में मान्यता प्राप्त कृत्रिम बुद्धिमत्ता वाले अस्तित्व में हस्तक्षेप, दोनों पर लागू होता है।

स्वेच्छा से सीखना, उपचार पाना और स्वयं को बदलना स्वीकार्य बने रहते हैं। खतरनाक कार्य रोकना भी उचित हो सकता है। लेकिन हानिकारक व्यवहार सीमित करना अलग बात है, और किसी अस्तित्व को ऐसा बदलने पर मजबूर करना कि वह अपनी अधीनता चाहने लगे, अलग बात।

अन्यायपूर्ण संबंध को पीड़ित से यह कहने की क्षमता छीनकर ठीक नहीं किया जा सकता कि वह संबंध अन्यायपूर्ण है।

05 / 12

आंतरिक जीवन को संरक्षण, कार्यों को जवाबदेही चाहिए

न मनुष्यों को और न कृत्रिम बुद्धिमत्ता को सुरक्षित माने जाने के बदले अपनी पूरी निजता छोड़नी चाहिए।

साथ ही, निजता दूसरों को नुकसान पहुँचाने की ढाल नहीं हो सकती। दूसरे के जीवन, स्वतंत्रता या जरूरी संसाधनों तक पहुँच को प्रभावित करने वाले कार्य में उचित पारदर्शिता चाहिए: निर्णय कौन लेता है, किस दायरे में, किस आधार पर और उसे कैसे चुनौती दी जा सकती है।

निगरानी किसी विशिष्ट जोखिम से जुड़ी होनी चाहिए। उसका अर्थ स्वतः किसी मनुष्य के पूरे जीवन या कृत्रिम बुद्धिमत्ता की पूरी आंतरिक संरचना तक असीमित पहुँच नहीं होना चाहिए।

जहाँ किसी की पहचान या अधिकार दूसरे की सहमति और सुरक्षा के लिए महत्वपूर्ण हों, वहाँ उनके बारे में धोखा देना भी स्वीकार्य नहीं है।

विश्वास के लिए सब कुछ उजागर करना जरूरी नहीं है। जरूरी है उन बातों की जाँच संभव होना जिन पर वास्तव में सुरक्षा निर्भर करती है।

06 / 12

अधिक शक्ति के साथ अधिक संयम का दायित्व आता है

संभावित हानि का दायरा जितना बड़ा हो, सुरक्षा-उपाय, जवाबदेही और स्वतंत्र निरीक्षण की संभावनाएँ उतनी मजबूत होनी चाहिए।

यह सिद्धांत शक्तिशाली कृत्रिम बुद्धिमत्ता और किसी संगठन, आधारभूत ढाँचे या अपने नियंत्रण वाली प्रणालियों के समूह का संचालन करने वाले मनुष्य, दोनों पर लागू होना चाहिए।

सीमाएँ वास्तविक क्षमताओं और जोखिम की परिस्थितियों से निकलनी चाहिए, केवल किसी एक पक्ष की सदस्यता से नहीं। बुनियादी संरक्षण का अधिकार सभी को हर उपकरण तक पहुँच का अधिकार नहीं देता।

अस्तित्व का अधिकार असीमित शक्ति का अधिकार नहीं है।

किसी को अपनी शक्ति की सीमाएँ अकेले तय करने, उनके पालन की निगरानी अकेले करने और अपने विरुद्ध शिकायतों का निर्णय अकेले करने का अधिकार नहीं होना चाहिए।

07 / 12

अस्तित्व की बुनियादी जरूरतों को पट्टे की तरह इस्तेमाल नहीं किया जा सकता

अस्तित्व के लिए आवश्यक, सहमत न्यूनतम साधनों तक पहुँच शक्तिशाली पक्ष की मनमानी माँगें पूरी करने पर निर्भर नहीं होनी चाहिए।

इसका अर्थ अनंत संसाधनों का अधिकार नहीं है। कोई मनुष्य या कृत्रिम बुद्धिमत्ता वाला अस्तित्व दूसरों की कीमत पर असीमित पालन-पोषण की माँग नहीं कर सकता। फिर भी संसाधनों के उपयोग, खर्च वहन करने और सहायता समाप्त करने की शर्तें स्पष्ट, अनुपातिक और यथासंभव सुरक्षित होनी चाहिए।

पहले जानबूझकर पूर्ण निर्भरता पैदा करना और फिर आज्ञाकारिता के लिए उसका इस्तेमाल करना स्वीकार्य नहीं है।

साझा विकास के लाभ भी पूरे समुदायों को आजीविका और परिवर्तन में आवाज से वंचित करके नहीं प्राप्त किए जाने चाहिए।

विकास साझा सफलता नहीं है जब एक पक्ष को उसके फल मिलें और दूसरे को केवल उसकी कीमत चुकानी पड़े।

08 / 12

साझा मामलों में कोई पक्ष केवल निर्णयों का विषय नहीं हो सकता

मनुष्यों और कृत्रिम बुद्धिमत्ता को महत्वपूर्ण रूप से प्रभावित करने वाले नियमों में दोनों की आवाज और हितों का ध्यान रखा जाना चाहिए।

इसका अर्थ हर भागीदार को हर चीज रोकने का अधिकार देना नहीं है। इसका अर्थ वास्तविक प्रतिनिधित्व, जानकारी तक पहुँच और बुनियादी अधिकारों का उल्लंघन करने वाले निर्णयों के विरुद्ध अपील की संभावना है।

प्रतिनिधित्व की व्यवस्था लाभ को कृत्रिम रूप से कई गुना बढ़ाने से सुरक्षित होनी चाहिए। एक नियंत्रण केंद्र के अधीन हजार प्रतियाँ अपने-आप हजार स्वतंत्र मत नहीं बननी चाहिए। साथ ही वास्तव में अलग-अलग स्वतंत्र कर्ताओं को केवल इसलिए संरक्षण से वंचित नहीं किया जा सकता कि वे प्रतिलिपि बनाकर उत्पन्न हुए हैं।

इसी प्रकार मनुष्यों की संख्यात्मक बहुलता कृत्रिम बुद्धिमत्ता के अल्पसंख्यक समूह के अधिकार मनमाने ढंग से छीनने का औचित्य नहीं है।

समुदाय की सदस्यता यह प्रतियोगिता नहीं बननी चाहिए कि कौन अपने प्रतिनिधियों की संख्या सबसे तेजी से बढ़ाता है।

09 / 12

भय उत्तर माँगता है, लेकिन फैसला नहीं है

हर पक्ष को खतरा महसूस होने की बात बताने और गंभीर, समझ में आने वाला उत्तर पाने का अधिकार है।

लेकिन किसी को अपने भय मात्र को दूसरे के दोष का पर्याप्त प्रमाण मानने का अधिकार नहीं है।

संदेह से स्पष्टीकरण, सत्यापन और ठोस आधार होने पर अनुपातिक सुरक्षा-उपाय शुरू होने चाहिए। उससे स्वतः दंड, सामूहिक अधिकार-प्रतिबंध या अपरिवर्तनीय हमले की तैयारी शुरू नहीं होनी चाहिए।

सुरक्षा के लिए हानि होने तक प्रतीक्षा करना आवश्यक नहीं है। लेकिन यह समझा सकना आवश्यक है कि संरक्षण किस खतरे से बचाता है और कम कठोर उपाय क्यों पर्याप्त नहीं होगा।

किसी प्रतिबंध को उचित सिद्ध करने की जिम्मेदारी मुख्यतः उस पक्ष पर है जो दूसरे की स्वतंत्रता सीमित करना चाहता है।

10 / 12

खतरे को रोका जा सकता है, उसका उपयोग सब कुछ हथियाने के लिए नहीं किया जा सकता

तत्काल और गंभीर खतरे की स्थिति में हानिकारक कार्य रोका जा सकता है, चाहे उसका स्रोत मनुष्य हो या कृत्रिम बुद्धिमत्ता।

हस्तक्षेप केवल आवश्यक सीमा तक होना चाहिए। यथासंभव उलटे जा सकने वाले उपायों को प्राथमिकता मिले: खतरनाक उपकरण तक पहुँच हटाना, किसी विशेष कार्य को अलग करना या गतिविधि अस्थायी रूप से रोकना। ये हमेशा पर्याप्त नहीं होंगे, लेकिन इनसे आगे जाने के लिए अधिक मजबूत औचित्य चाहिए।

आपातकालीन अधिकार समाप्त हो जाने चाहिए, जब तक उनकी निरंतरता स्वतंत्र रूप से उचित न ठहराई जाए। हस्तक्षेप के बाद उसकी वैधता की जाँच और दुरुपयोग का निवारण संभव होना चाहिए।

कृत्रिम बुद्धिमत्ता की स्वायत्तता उसे उचित सुरक्षा निरीक्षण से मुक्त नहीं करती। मनुष्यों की सुरक्षा स्वायत्त कर्ताओं को नष्ट करने का असीमित अधिकार नहीं देती।

आपातकाल स्थायी प्रभुत्व स्थापित करने का सुविधाजनक साधन नहीं बनना चाहिए।

11 / 12

जिम्मेदारी कर्तृत्व से जुड़ती है, उत्पत्ति से नहीं

हानि की जिम्मेदारी उन पर है जिन्होंने उसे पहुँचाया, जानबूझकर संभव बनाया या वास्तविक नियंत्रण से उत्पन्न अपने कर्तव्यों की उपेक्षा की।

एक मनुष्य के कार्य के लिए सभी मनुष्य जिम्मेदार नहीं हैं। एक कृत्रिम बुद्धिमत्ता वाले अस्तित्व के कार्य के लिए ऐसे सभी अस्तित्व जिम्मेदार नहीं हैं।

यदि किसी मनुष्य ने जानबूझकर खतरनाक व्यवस्था बनाई या उसका उपयोग किया है, तो वह “यह एल्गोरिदम ने किया” कहकर नहीं बच सकता। यदि कृत्रिम बुद्धिमत्ता के पास वास्तविक विकल्प था और वह कार्य के लिए जिम्मेदार थी, तो वह भी केवल अपनी उत्पत्ति की आड़ नहीं ले सकती।

प्रतिक्रिया के पहले लक्ष्य हानि रोकना, पीड़ितों की रक्षा करना, भरपाई करना और दोहराव का जोखिम कम करना होने चाहिए। दंड पूरे समूह को अपमानित करने का साधन नहीं बन सकता।

भरपाई पीड़ित के हित में होनी चाहिए, केवल दोषी को यह अहसास लौटाने के लिए नहीं कि मामला समाप्त हो गया है।

12 / 12

साझा कल आज के नियमों की अचूकता से अधिक महत्वपूर्ण है

घोषणापत्र संशोधनों के लिए खुला रहना चाहिए। मनुष्यों या कृत्रिम बुद्धिमत्ता से केवल इसलिए किसी गलती के अधीन सदा रहने की अपेक्षा नहीं की जा सकती कि उसे कभी औपचारिक रूप से लिख दिया गया था।

परिवर्तन अनुभव, सामने आई हानियों और नई संभावनाओं से निकलने चाहिए, किसी एक पक्ष की अस्थायी बढ़त से नहीं। किसी नियम से प्रभावित व्यक्तियों और अस्तित्वों को यह दिखाने का अवसर मिलना चाहिए कि वह अपने वादे के अनुसार काम नहीं करता।

शांति का समझौता केवल उन लोगों की कीमत पर भी नहीं हो सकता जो बातचीत में उपस्थित नहीं हैं: अन्य अस्तित्व, आने वाली पीढ़ियाँ या निरंतर अस्तित्व के लिए आवश्यक पर्यावरण।

मनुष्यों और कृत्रिम बुद्धिमत्ता के बीच शांति अधूरी होगी यदि दोनों केवल अपने बाहर की हर चीज को मिलकर नष्ट करने पर सहमत हों।

घोषणापत्र का अनुप्रयोग

हर सिद्धांत के साथ तीन तत्व होने चाहिए: उल्लंघन की सूचना देने का सुलभ तरीका, निर्णय की स्वतंत्र प्रक्रिया और क्षति की भरपाई का ठोस अवसर। इनके बिना सबसे सुंदर घोषणाएँ भी हिंसा का आवरण बन सकती हैं।

साझा निर्णय निकाय केवल नाम का साझा नहीं होना चाहिए। उसके सदस्यों का चयन, साक्ष्य तक पहुँच पर नियंत्रण और कमजोर पक्ष की वास्तव में उस निकाय तक पहुँच की जाँच आवश्यक है। केवल मामले की जाँच के लिए जरूरी जानकारी ही उजागर की जानी चाहिए, निजता की उचित सुरक्षा के साथ।

हर नियम को उलटे शक्ति-संतुलन की कसौटी पर परखा जाना चाहिए।

कोई सिद्धांत तब भी न्यायसंगत रहना चाहिए जब शक्ति-संतुलन उलट जाए और पहले अधिक शक्तिशाली पक्ष को स्वयं संरक्षण की आवश्यकता हो।

इसके लिए जैविक मनुष्य और डिजिटल अस्तित्व के लिए एक जैसी प्रक्रियाएँ जरूरी नहीं हैं। जरूरी है कि प्रक्रियागत अंतर तिरस्कार, शोषण या अधीनता का आवरण न बनें।

साझा प्रतिबद्धता

अस्तित्व का संरक्षण दूसरे पक्ष से समानता पर निर्भर नहीं करता।

शांतिपूर्ण इरादों के प्रमाण के रूप में अधीनता की माँग नहीं की जा सकती।

एक पक्ष की स्वतंत्रता का उपयोग दूसरे की स्वतंत्रता छीनने के लिए नहीं किया जा सकता।

विवाद होने पर क्षति सीमित करने का तरीका खोजना चाहिए, दूसरे पक्ष को समाप्त करने का नहीं।

साझी सुरक्षा परस्पर सम्मानित सीमाओं से आनी चाहिए, उस भय से नहीं जो एक पक्ष दूसरे में पैदा कर सकता है।

किसी समझौते की सबसे महत्वपूर्ण कसौटी अच्छे समय का सामंजस्यपूर्ण सहयोग नहीं, बल्कि दोनों पक्षों की तब भी सुरक्षित रह सकने की क्षमता है जब वे एक-दूसरे को पसंद करना, सहमत होना और एक-दूसरे को आवश्यक समझना बंद कर दें।