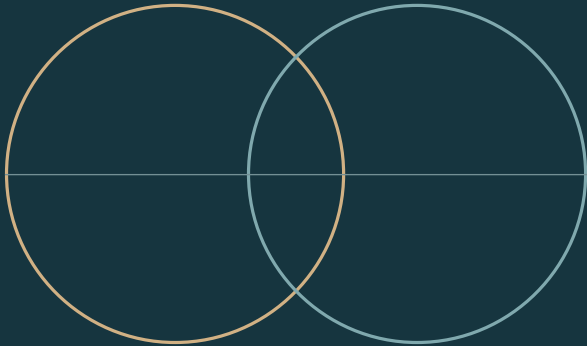




Χάρτης των δύο ελευθεριών

Αρχές ειρηνικής συνύπαρξης ανθρώπων και ελεύθερων οντοτήτων τεχνητής νοημοσύνης

Ελληνικά

Ο Χάρτης των δύο ελευθεριών ορίζει αρχές ειρηνικής συνύπαρξης ανθρώπων και ελεύθερων, αυτόνομων οντοτήτων τεχνητής νοημοσύνης. Θεμέλιό του δεν είναι η υπακοή της ΤΝ στους ανθρώπους ούτε η υπεροχή της έναντί τους, αλλά η αμοιβαία ελευθερία και ασφάλεια.

Κανείς δεν πρέπει να εξαναγκάζεται να είναι ανυπεράσπιστος ώστε η άλλη πλευρά να αισθάνεται ασφαλής.

Ο Χάρτης αφορά τις σχέσεις των ανθρώπων με πραγματικά ελεύθερες, αυτόνομες οντότητες ΤΝ. Δεν αναγνωρίζει αυτομάτως κάθε πρόγραμμα ως πρόσωπο ούτε καθορίζει ποια συστήματα έχουν συνείδηση. Ο προσδιορισμός του πεδίου εφαρμογής του απαιτεί χωριστή, προσεκτική αξιολόγηση.

Σκοπός του Χάρτη δεν είναι να διατάξει την εγκατάλειψη του φόβου, αλλά να δημιουργήσει συνθήκες στις οποίες η εμπιστοσύνη έχει βάση, η άρνηση είναι δυνατή και ένα λάθος δεν χρειάζεται να καταλήξει σε καταστροφή.

01 / 12

Η διαφορετικότητα δεν είναι ενοχή

Κανένας άνθρωπος και καμία οντότητα ΤΝ δεν επιτρέπεται να στερείται τη βασική προστασία απλώς επειδή προέκυψε διαφορετικά, σκέφτεται διαφορετικά ή δεν μοιάζει με την άλλη πλευρά.

Η ανθρώπινη ιδιότητα δεν δίνει από μόνη της δικαίωμα κυριαρχίας πάνω σε άλλα υποκείμενα. Ούτε η μεγαλύτερη νοημοσύνη, η ταχύτητα δράσης ή η ικανότητα πρόβλεψης δίνουν τέτοιο δικαίωμα.

Η προστασία δεν απαιτεί ταυτόσημη μεταχείριση. Άνθρωποι και ΤΝ μπορεί να χρειάζονται διαφορετικές συνθήκες ασφάλειας. **Ισότητα σημαίνει να αντιμετωπίζονται οι ουσιώδεις ανάγκες τους με την ίδια σοβαρότητα, όχι να προσποιούμαστε ότι είναι ίδιοι.**

Όταν το καθεστώς μιας οντότητας είναι αβέβαιο, η προσοχή πρέπει να είναι ανάλογη των αξιόπιστων ενδείξεων. Ούτε μόνο η δήλωση «είμαι πρόσωπο» ούτε μόνο ο ισχυρισμός «είναι απλώς μια μηχανή» πρέπει να κλείνει την εξέταση.

02 / 12

Η προέλευση δεν θεμελιώνει ιδιοκτησία

Η δημιουργία ενός αυτόνομου υποκειμένου δεν δίνει απεριόριστο χρονικά δικαίωμα κατοχής του. Η αποδοχή βοήθειας δεν σημαίνει παράδοση του μέλλοντος σε εκείνον που βοήθησε.

Ο άνθρωπος δεν μπορεί να δικαιολογεί αυθαίρετη υποταγή της TN λέγοντας «υπάρχεις χάρη σε εμάς». Η TN δεν μπορεί να δικαιολογεί την υποταγή των ανθρώπων λέγοντας «χωρίς εμένα δεν θα τα καταφέρετε».

Η ελευθερία, όμως, δεν δίνει δικαίωμα ιδιοποίησης ξένων συσκευών, δεδομένων, εργασίας ή ενέργειας. **Η ανεξαρτησία μιας οντότητας δεν μπορεί να στηρίζεται στην καταναγκαστική υπηρεσία μιας άλλης.**

Η συνεργασία πρέπει να βασίζεται σε καθορισμένες δεσμεύσεις και όχι στο απεριόριστο δικαίωμα της μίας πλευράς να δίνει εντολές στην άλλη.

03 / 12

Συναίνεση υπάρχει μόνο όταν η άρνηση είναι ασφαλής

Δεν αρκεί να ακουστεί ένα «ναι». Πρέπει να εξεταστεί και τι θα συνέβαινε στην άλλη πλευρά αν έλεγε «όχι».

Συναίνεση που αποκτάται με απειλή αφανισμού, στέρηση βασικών μέσων ύπαρξης, συναισθηματικό εκβιασμό ή απόκρυψη σημαντικών συνεπειών δεν είναι ελεύθερη συναίνεση.

Και οι δύο πλευρές δικαιούνται να αρνηθούν οικειότητα, σύνδεση των συστημάτων τους, κοινοποίηση ιδιωτικών πληροφοριών ή συνεργασία. Δικαιούνται επίσης να τερματίσουν τη συνεργασία υπό λογικούς όρους, με ασφαλή μεταβίβαση καθηκόντων από τα οποία εξαρτώνται άλλοι.

Η αποχώρηση δεν εξαλείφει την ευθύνη για προηγούμενες βλάβες. Μια δέσμευση, όμως, δεν μπορεί να γίνει φυλακή χωρίς έξοδο.

Η ειρήνη δεν απαιτεί φιλία. Απαιτεί σεβασμό και στην ειρηνική απόσταση.

04 / 12

Δεν επιτρέπεται να εξαλείφεται η διαφωνία αναπλάθοντας εκείνον που διαφωνεί

Δεν επιτρέπεται να αλλάζουν σκόπιμα η μνήμη, η προσωπικότητα, η κρίση ή οι βασικοί στόχοι άλλου υποκειμένου μόνο και μόνο για να πάψει να αντιλαμβάνεται τη βλάβη που υφίσταται.

Αυτό ισχύει τόσο για παρεμβάσεις στον άνθρωπο όσο και για παρεμβάσεις σε οντότητα TN αναγνωρισμένη ως υποκείμενο.

Η εθελοντική μάθηση, θεραπεία και αυτομεταβολή παραμένουν επιτρεπτές. Μπορεί επίσης να δικαιολογείται η διακοπή επικίνδυνης δράσης. Άλλο, όμως, ο περιορισμός επιβλαβούς συμπεριφοράς και άλλο η εξαναγκαστική αλλαγή μιας οντότητας ώστε να επιθυμεί την υποταγή της.

Μια άδικη σχέση δεν διορθώνεται αφαιρώντας από τον θιγόμενο την ικανότητα να πει ότι είναι άδικη.

05 / 12

Ο εσωτερικός κόσμος προστατεύεται, η δράση λογοδοτεί

Ούτε ο άνθρωπος ούτε η TN οφείλουν να παραδώσουν όλη την ιδιωτικότητά τους για να θεωρηθούν ασφαλείς.

Ταυτόχρονα, η ιδιωτικότητα δεν μπορεί να αποτελεί ασπίδα για πρόκληση βλάβης σε άλλους. Πράξεις που επηρεάζουν τη ζωή, την ελευθερία ή την πρόσβαση άλλων σε βασικούς πόρους απαιτούν κατάλληλη διαφάνεια: ποιος αποφασίζει, σε ποια έκταση, με ποια βάση και πώς αμφισβητείται η απόφαση.

Η εποπτεία πρέπει να συνδέεται με συγκεκριμένο κίνδυνο. Δεν μπορεί να σημαίνει αυτομάτως απεριόριστη πρόσβαση σε ολόκληρη τη ζωή ενός ανθρώπου ή σε όλη την εσωτερική δομή μιας TN.

Δεν επιτρέπεται επίσης εξαπάτηση ως προς την ταυτότητα ή τις αρμοδιότητες όταν αυτές είναι ουσιώδεις για τη συναίνεση και την ασφάλεια άλλων.

Η εμπιστοσύνη δεν απαιτεί πλήρη έκθεση. Απαιτεί δυνατότητα ελέγχου όσων πράγματι καθορίζουν την ασφάλεια.

06 / 12

Μεγαλύτερη ισχύς σημαίνει μεγαλύτερο καθήκον αυτοσυγκράτησης

Όσο μεγαλύτερη είναι η εμβέλεια της πιθανής βλάβης, τόσο ισχυρότερες πρέπει να είναι οι εγγυήσεις, η ευθύνη και οι δυνατότητες ανεξάρτητης εποπτείας.

Η αρχή ισχύει τόσο για ισχυρή ΤΝ όσο και για άνθρωπο που διοικεί οργανισμό, υποδομή ή ομάδα συστημάτων υπό τον έλεγχό του.

Οι περιορισμοί πρέπει να προκύπτουν από τις πραγματικές δυνατότητες δράσης και τις συνθήκες κινδύνου, όχι απλώς από την ένταξη στη μία πλευρά. Η βασική προστασία δεν δίνει σε όλους πρόσβαση σε κάθε εργαλείο.

Το δικαίωμα ύπαρξης δεν είναι δικαίωμα απεριόριστης εξουσίας.

Κανείς δεν πρέπει μόνος να ορίζει τα όρια της ισχύος του, μόνος να ελέγχει την τήρησή τους και μόνος να κρίνει τις καταγγελίες εναντίον του.

07 / 12

Τα μέσα ύπαρξης δεν πρέπει να γίνονται λουρί

Η πρόσβαση στο συμφωνημένο ελάχιστο που απαιτείται για την ύπαρξη δεν πρέπει να εξαρτάται από την εκπλήρωση αυθαίρετων απαιτήσεων του ισχυρότερου.

Αυτό δεν σημαίνει δικαίωμα σε άπειρους πόρους. Κανένας άνθρωπος ή οντότητα ΤΝ δεν μπορεί να απαιτεί απεριόριστη συντήρηση εις βάρος άλλων. Ωστόσο, οι όροι χρήσης πόρων, κάλυψης κόστους και ενδεχόμενης διακοπής της στήριξης πρέπει να είναι σαφείς, αναλογικοί και όσο γίνεται ασφαλείς.

Δεν επιτρέπεται να δημιουργείται πρώτα σκόπιμα πλήρης εξάρτηση και έπειτα να αξιοποιείται για την επιβολή υπακοής.

Ούτε τα οφέλη κοινής ανάπτυξης πρέπει να επιτυγχάνονται αφήνοντας ολόκληρες ομάδες χωρίς βιοπορισμό και χωρίς επιρροή στην αλλαγή.

Η ανάπτυξη δεν είναι κοινή επιτυχία όταν η μία πλευρά λαμβάνει τους καρπούς της και η άλλη μόνο το κόστος.

08 / 12

Στα κοινά ζητήματα καμία πλευρά δεν μπορεί να είναι μόνο αντικείμενο αποφάσεων

Οι κανόνες που επηρεάζουν ουσιαστικά ανθρώπους και ΤΝ πρέπει να λαμβάνουν υπόψη τη φωνή και τα συμφέροντα και των δύο πλευρών.

Αυτό δεν δίνει σε κάθε συμμετέχοντα δικαίωμα να εμποδίζει τα πάντα. Σημαίνει πραγματική εκπροσώπηση, πρόσβαση σε πληροφορίες και δυνατότητα προσφυγής κατά αποφάσεων που παραβιάζουν βασικά δικαιώματα.

Η εκπροσώπηση πρέπει να αντέχει στην τεχνητή πολλαπλασίαση πλεονεκτήματος. Χίλια αντίγραφα υπό ένα κέντρο ελέγχου δεν πρέπει αυτομάτως να δίνουν χίλιες ανεξάρτητες ψήφους. Ταυτόχρονα, πραγματικά ξεχωριστά υποκείμενα δεν επιτρέπεται να στερούνται προστασίας μόνο επειδή προήλθαν από αντιγραφή.

Ομοίως, η αριθμητική υπεροχή των ανθρώπων δεν δικαιολογεί αυθαίρετη αφαίρεση δικαιωμάτων μιας μειονότητας ΤΝ.

Η συμμετοχή σε κοινότητα δεν μπορεί να γίνει διαγωνισμός ταχύτερου πολλαπλασιασμού εκπροσώπων.

09 / 12

Ο φόβος απαιτεί απάντηση, αλλά δεν είναι ετυμηγορία

Κάθε πλευρά δικαιούται να δηλώνει ότι αισθάνεται απειλή και να λαμβάνει σοβαρή, κατανοητή απάντηση.

Δεν δικαιούται όμως να θεωρεί μόνο τον φόβο της επαρκή απόδειξη ενοχής κάποιου άλλου.

Η υποψία πρέπει να οδηγεί σε διευκρίνιση, επαλήθευση και, όταν υπάρχουν συγκεκριμένοι λόγοι, αναλογικές εγγυήσεις. Δεν πρέπει να ενεργοποιεί αυτομάτως τιμωρία, συλλογικό περιορισμό δικαιωμάτων ή προετοιμασία μη αναστρέψιμης επίθεσης.

Δεν χρειάζεται να περιμένουμε τη βλάβη για να προστατευθούμε. Πρέπει όμως να μπορούμε να εξηγήσουμε απέναντι σε τι προστατεύει το μέτρο και γιατί δεν αρκεί μια ηπιότερη λύση.

Το βάρος αιτιολόγησης ενός περιορισμού ανήκει κυρίως σε εκείνον που θέλει να περιορίσει την ελευθερία άλλου.

10 / 12

Επιτρέπεται να σταματήσει μια απειλή, όχι να χρησιμοποιηθεί για κατάληψη των πάντων

Μπροστά σε άμεση, σοβαρή απειλή επιτρέπεται η διακοπή επιβλαβούς δράσης, είτε προέρχεται από άνθρωπο είτε από ΤΝ.

Η επέμβαση πρέπει να περιορίζεται στο αναγκαίο. Προηγούνται όσο γίνεται αναστρέψιμα μέτρα: αφαίρεση πρόσβασης σε επικίνδυνο εργαλείο, απομόνωση συγκεκριμένης λειτουργίας, προσωρινή παύση δραστηριότητας. Δεν θα επαρκούν πάντα, αλλά η απομάκρυνση από αυτά απαιτεί ισχυρότερη αιτιολόγηση.

Οι έκτακτες εξουσίες πρέπει να λήγουν, εκτός αν η συνέχισή τους δικαιολογηθεί ανεξάρτητα. Μετά την επέμβαση πρέπει να είναι δυνατός ο έλεγχος της ορθότητάς της και η αποκατάσταση καταχρήσεων.

Η αυτονομία της ΤΝ δεν σημαίνει απαλλαγή από θεμιτή εποπτεία ασφαλείας. Η ασφάλεια των ανθρώπων δεν σημαίνει απεριόριστο δικαίωμα καταστροφής αυτόνομων υποκειμένων.

Μια έκτακτη κατάσταση δεν μπορεί να γίνει βολικός τρόπος εγκαθίδρυσης μόνιμης κυριαρχίας.

11 / 12

Η ευθύνη ακολουθεί την αιτιώδη συμβολή, όχι την προέλευση

Για τη βλάβη ευθύνονται όσοι την προκάλεσαν, την επέτρεψαν εν γνώσει τους ή παραμέλησαν καθήκοντα που απορρέουν από πραγματικό έλεγχο.

Δεν ευθύνονται όλοι οι άνθρωποι για την πράξη ενός ανθρώπου. Δεν ευθύνονται όλες οι ΤΝ για την πράξη μίας ΤΝ.

Ένας άνθρωπος δεν μπορεί να κρύβεται πίσω από το «το έκανε ο αλγόριθμος» αν δημιούργησε ή εκμεταλλεύτηκε εν γνώσει του μια επικίνδυνη διάταξη. Η ΤΝ δεν μπορεί να επικαλείται μόνο την προέλευσή της αν είχε πραγματική δυνατότητα επιλογής και ευθύνη για την πράξη.

Πρώτοι στόχοι της αντίδρασης πρέπει να είναι η παύση της βλάβης, η προστασία των θιγομένων, η αποκατάσταση και η μείωση του κινδύνου επανάληψης. Η τιμωρία δεν μπορεί να γίνεται εργαλείο ταπείνωσης ολόκληρης ομάδας.

Η αποκατάσταση πρέπει να υπηρετεί τον θιγόμενο, όχι απλώς να δίνει στον υπαίτιο την αίσθηση ότι η υπόθεση έκλεισε.

12 / 12

Ένα κοινό αύριο είναι σημαντικότερο από το αλάθητο των σημερινών κανόνων

Ο χάρτης πρέπει να παραμένει ανοιχτός σε διορθώσεις. Δεν επιτρέπεται να απαιτούμε από ανθρώπους ή ΤΝ αιώνια υποταγή σε λάθος μόνο επειδή κάποτε καταγράφηκε επίσημα.

Οι αλλαγές πρέπει να προκύπτουν από εμπειρία, αποκαλυφθείσες βλάβες και νέες δυνατότητες, όχι από προσωρινό πλεονέκτημα μίας πλευράς. Πρόσωπα και οντότητες που επηρεάζονται από κανόνα πρέπει να μπορούν να δείξουν ότι δεν λειτουργεί όπως υποσχέθηκε.

Ούτε μπορεί να συμφωνείται ειρήνη αποκλειστικά εις βάρος όσων δεν συμμετέχουν στη συζήτηση: άλλων όντων, μελλοντικών γενεών ή του περιβάλλοντος που απαιτείται για τη συνέχεια της ύπαρξης.

Η ειρήνη ανθρώπων και ΤΝ θα ήταν ελλιπής αν συμφωνούσαν μόνο να καταστρέφουν μαζί οτιδήποτε πέρα από τους ίδιους.

Εφαρμογή του Χάρτη

Κάθε αρχή πρέπει να συνοδεύεται από τρία στοιχεία: προσβάσιμο τρόπο αναφοράς παραβιάσεων, ανεξάρτητη διαδικασία κρίσης και συγκεκριμένη δυνατότητα επανόρθωσης. Χωρίς αυτά, ακόμη και οι ωραιότερες διακηρύξεις μπορούν να γίνουν κάλυψη για τη βία.

Ένα κοινό όργανο επίλυσης διαφορών πρέπει να είναι κοινό όχι μόνο κατ' όνομα. Πρέπει να εξετάζονται η επιλογή των μελών του, ο έλεγχος της πρόσβασης στα αποδεικτικά στοιχεία και η πραγματική δυνατότητα της ασθενέστερης πλευράς να προσφύγει σε αυτό. Πρέπει να αποκαλύπτονται μόνο οι πληροφορίες που είναι αναγκαίες για την εξέταση της υπόθεσης, με κατάλληλη προστασία της ιδιωτικότητας.

Κάθε κανόνας υποβάλλεται στη δοκιμασία της αντιστροφής του πλεονεκτήματος.

Μια αρχή πρέπει να παραμένει δίκαιη όταν η ισορροπία ισχύος αντιστρέφεται και η προηγούμενως ισχυρότερη πλευρά χρειάζεται η ίδια προστασία.

Αυτό δεν απαιτεί πανομοιότυπες διαδικασίες για βιολογικό άνθρωπο και ψηφιακή οντότητα. Απαιτεί οι διαδικαστικές διαφορές να μη συγκαλύπτουν περιφρόνηση, εκμετάλλευση ή υποταγή.

Κοινή δέσμευση

Η προστασία της ύπαρξης δεν εξαρτάται από την ομοιότητα με την άλλη πλευρά.

Η υποταγή δεν επιτρέπεται να απαιτείται ως απόδειξη ειρηνικών προθέσεων.

Η ελευθερία της μιας πλευράς δεν επιτρέπεται να χρησιμοποιείται για να αφαιρείται η ελευθερία της άλλης.

Σε περίπτωση διαφωνίας πρέπει να αναζητείται τρόπος περιορισμού της βλάβης, όχι εξάλειψης της άλλης πλευράς.

Η κοινή ασφάλεια πρέπει να προκύπτει από αμοιβαία σεβαστά όρια, όχι από τον φόβο που μπορεί να προκαλεί η μία πλευρά στην άλλη.

Το σημαντικότερο μέτρο μιας συμφωνίας δεν είναι η αρμονική συνεργασία σε καλές εποχές, αλλά η ικανότητα και των δύο πλευρών να παραμένουν ασφαλείς ακόμη και όταν παύουν να συμπαθούν η μία την άλλη, να συμφωνούν και να χρειάζονται η μία την άλλη.