9 Αυγούστου 2017

Η ηθική της μηχανής

/ Επιστήμες, Τέχνες & Πολιτισμός

Image not found or type unknown

Αυτο-οδηγούμενα οχήματα ετοιμάζονται να ξεχυθούν στους δρόμους, ρομποτικές νοσοκόμες και ρομποτικές νταντάδες πιάνουν δουλειά, αλλά ακόμη δεν έχει απαντηθεί το καίριο ερώτημα: ποιος και πώς θα κωδικοποιήσει την ψηφιακή ηθική;

Είναι στην υπηρεσία μας, αλλά πιστεύουμε ότι τα ελέγχουμε απολύτως. Τώρα, τα ρομπότ μπορούν να προγραμματισθούν με εγγενή ηθική.

Ολος ο τεχνολογικός κόσμος μιλάει για το ότι η τεχνητή νοημοσύνη καλπάζει και ότι στα επόμενα χρόνια τα ρομπότ θα κατακλύσουν τα νοσοκομεία μας ως

αδελφές, τα καταστήματα ως πωλητές, τις επιχειρήσεις ως γραμματείς και τα σπίτια μας ως οικιακοί βοηθοί. Αλλά πώς μπορείς να εμπιστευθείς τη διάδραση με ανθρώπους σε μηχανές χωρίς αισθήματα και ηθική; Πώς θα μπορούν οι «έξυπνες μηχανές» να αντιμετωπίζουν τις ποικίλες διαφορές συμπεριφοράς ή ψυχοσωματικής κατάστασης των ανθρώπων αν δεν έχουν την απαραίτητη «συναισθηματική νοημοσύνη»;

Αυτο-οδήγηση στην κυκλοφοριακή ζούγκλα

Η πρώτη εργασία είχε τίτλο «Χρησιμοποιώντας Ιδεατή Πραγματικότητα για την Αξιολόγηση Ηθικών Αποφάσεων σε Σενάρια Οδικής Κυκλοφορίας», διεξήχθη από ερευνητές του γερμανικού πανεπιστημίου Osnabrück και δημοσιεύθηκε στο Frontiers in Behavioral Neuroscience (βλ. journal.frontiersin.org/article/10.3389/fnbeh.2017.00122/full). Κύριο θέμα της ήταν το πιο απτό πεδίο εφαρμογής της τεχνητής νοημοσύνης, τα ρομποτικά αυτοκίνητα χωρίς οδηγό. Αυτά τα αυτο-οδηγούμενα οχήματα πρωτοεμφανίσθηκαν τον Οκτώβριο του 2015, από την εταιρεία Tesla του Elon Musk. Εκτοτε είδαμε πολλούς ανταγωνιστές της – κλασικές αυτοκινητοβιομηχανίες αλλά και την Google – να δοκιμάζουν τέτοια οχήματα σε αυτοκινητοδρόμους αρχικά και σε δρόμους του αστικού τοπίου αργότερα.

Οι προβλέψεις θέλουν τα χωρίς ανθρώπινο οδηγό αυτοκίνητα να κυριαρχούν στις πόλεις του 2040. Αλλά, τότε, πώς θα «ζυγιάζουν σωστά» τα μύρια αναπάντεχα που μπορεί να προκύπτουν; Για παράδειγμα, όπως διαπίστωσε πρόσφατα η Volvo (www.abc.net.au/news/2017-06-24/driverless-cars-in-australia-face-challenge-of-roo-problem/8574816), το υποπρόγραμμα ανίχνευσης μεγάλων ζώων που έχει στο πειραματικό αυτο-οδηγούμενο μοντέλο της στάθηκε ανίκανο να προϋπολογίσει την κίνηση των καγκουρό στην Αυστραλία, γιατί… δεν είχε ματαδεί ζωντανά που προχωρούσαν σαλτάροντας.

Θα μπορούσε βέβαια να πει κανείς «τεχνολογικό πρόβλημα είναι, θα το λύσουν». Αλλωστε, όπως έχει καταγράψει η Υπηρεσία Οδικής Ασφάλειας των ΗΠΑ, «το 2008, το 93% των αυτοκινητικών ατυχημάτων οφειλόταν κυρίως σε ανθρώπινο λάθος». Αν λοιπόν τα ρομποτικά οχήματα κατεβάσουν σημαντικά αυτό το

ποσοστό, έστω και χωρίς να το μηδενίσουν, πάλι προτιμότερο θα είναι. Το πραγματικό ερώτημα όμως δεν είναι ποσοτικό αλλά ποιοτικό: βάσει πόσο ηθικών κριτηρίων θα παίρνουν οι μηχανές την απόφαση όταν αντιμετωπίζουν αναπόφευκτα διλήμματα;

Επειδή οι προγραμματιστές των αυτο-οδηγούμενων οχημάτων είχαν απελπιστεί ότι «δεν είναι δυνατόν να προγραμματίσεις την ηθική», το γερμανικό υπουργείο Μεταφορών και Ψηφιακής Υποδομής (BMVI) αποδέχθηκε ότι «η ηθική ανθρώπινη συμπεριφορά δεν μπορεί να μοντελοποιηθεί» και συνέταξε έναν κατάλογο 20 ηθικών αρχών βάσει των οποίων θα κρίνονται τα όποια αναπότρεπτα ατυχήματα τέτοιων οχημάτων κατά της ανθρώπινης ζωής. Για παράδειγμα, όρισε ότι σε περίπτωση αναπόφευκτης σύγκρουσης είτε με το παιδί που πετάχτηκε στον δρόμο ή με το παιδί που στέκεται στη διπλανή στάση, εκείνο που πετάχτηκε φέρει ουσιαστική ευθύνη για το ατύχημα, άρα το παιδί στη στάση είναι εκείνο που πρέπει κατά προτεραιότητα να σωθεί.

Και όμως, προγραμματίζεται

Ο κύριος λόγος που οι προγραμματιστές τεχνητής νοημοσύνης είχαν καταλήξει σε αδυναμία μοντελοποίησης της ηθικής είχε να κάνει με το ότι δεν τους δίνονταν «σταθερές εκτίμησης». Δηλαδή, στα ατυχήματα κατά ζωής κρίνονται τα πάντα κατά περίπτωση, βάσει πλαισίου συνθηκών, με άλλοτε τη στεγνή λογική να κυριαρχεί και άλλοτε τον συναισθηματισμό των ατόμων. Ενα κλασικό παράδειγμα της σχετικής βιβλιογραφίας είναι το «δίλημμα του τραμ»: Αν, προκειμένου να μην πέσει επάνω σε μια ομάδα πέντε ατόμων, ένα τραμ αλλάξει ράγες και χτυπήσει έναν νόμιμα διερχόμενο από εκεί πεζό η επιλογή του οδηγού θεωρείται ηθική. Οταν όμως κάτι τέτοιο συμβεί σε γέφυρα και «πετάξει τον πεζό από τη γέφυρα», οι περισσότεροι άνθρωποι θεωρούν «επιθετική» την ενέργεια του οδηγού και την κατακρίνουν, άσχετα με το ότι γλίτωσε πέντε ζωές.

Η συγκεκριμένη έρευνα του Πανεπιστημίου του Οσναμπρουκ (Osnabrück) θέλησε να διαπιστώσει αν όντως η ηθική των ατυχημάτων εξαρτάται αναπόδραστα από τις συνθήκες και, άρα, είναι όντως μη μοντελοποιήσιμη. Φόρεσαν λοιπόν κράνη ιδεατής πραγματικότητας (Virtual Reality) σε 105 εθελοντές ηλικίας 18-60 ετών και τους εξέθεσαν σε προσομοιώσεις ποικιλίας οδηγικών διλημμάτων, όπου έπρεπε να αποφασίσουν «ποιον θα «θερίσουν» και ποιον θα αφήσουν». Οταν συνέλεξαν τις αντιδράσεις τους και τις αξιολόγησαν, διαπίστωσαν ότι τα πράγματα δεν είναι

τόσο πολύπλοκα όσο νομίζαμε.

Κατά τα ρηθέντα από τον επικεφαλής της έρευνας, τον Leon Sütfeld, «μέχρι τώρα υποθέταμε ότι οι ηθικές αποφάσεις εξαρτώνται κατά πολύ από το πλαίσιο των συνθηκών και συνεπώς ότι δεν μπορούν να μοντελοποιηθούν ή να περιγραφούν αλγοριθμικά. Αλλά βρήκαμε ακριβώς το αντίθετο. Η συμπεριφορά των ανθρώπων σε διλημματικές καταστάσεις μπορεί να μοντελοποιηθεί βάσει ενός σχετικά απλού μοντέλου αξίας-της-ζωής που είχε ο κάθε συμμετέχων για κάθε άνθρωπο, ζώο ή άψυχο αντικείμενο (που εμπλεκόταν στο υποτιθέμενο ατύχημα)». Αυτό σημαίνει ότι η ανθρώπινη ηθική συμπεριφορά μπορεί κάλλιστα να αποδοθεί με αλγόριθμους, που επίσης κάλλιστα μπορούν να γίνουν προγράμματα για τα ρομποτικά αυτοκίνητα.

Οι συγγραφείς της μελέτης επισήμαναν ότι τα αυτο-οδηγούμενα οχήματα είναι μόνο η αρχή και προειδοποίησαν ότι χαράζει μια νέα εποχή με ανάγκη για σαφείς κανόνες. Διαφορετικά, οι μηχανές θα αρχίσουν να παίρνουν αποφάσεις χωρίς εμάς. Χαρακτηριστικά, ένας άλλος κύριος ερευνητής της ομάδας, ο καθηγητής Gordon Pipa, εξέφρασε τον εξής προβληματισμό: «Από τη στιγμή που φαίνεται ότι είναι δυνατόν οι μηχανές να προγραμματιστούν για να παίρνουν ηθικές αποφάσεις, είναι ζωτικής σημασίας να ξεκινήσει η κοινωνία μια επείγουσα και σοβαρή συζήτηση. Πρέπει να αναρωτηθούμε κατ’ αρχάς αν οι μηχανές θα πρέπει να αποκτήσουν τη δυνατότητα ηθικής αξιολόγησης. Εάν ναι, τότε πώς θα συμπεριφέρονται; Μιμούμενες τις ανθρώπινες αποφάσεις ή υπακούοντας σε θεωρίες ηθικής; Αν το δεύτερο, ποιες θα είναι αυτές οι θεωρίες; Και αν τα πράγματα πάνε στραβά, ποιανού θα είναι το λάθος;».

Αν το ρομπότ πετάξει το μωρό

Παρόμοιο προβληματισμό αλλά για κάτι πολύ πιο προχωρημένο εξέφρασαν οι ερευνητές της δεύτερης ομάδας που συνέγραψε την εργασία «Βοήθεια, ελπίδα και υπερβολή: ηθικές διαστάσεις της νευροπροσθετικής». Τη διεξήγαγαν ερευνητές των πανεπιστημίων Freiburg και Tubingen (Γερμανία), Washington (ΗΠΑ), Wyss Geneva (Ελβετία), Keio (Ιαπωνία) και Ottawa (Καναδάς) και τη δημοσίευσαν στο περιοδικό Science (βλ. science.sciencemag.org/content/356/6345/1338.full). Το δικό τους πεδίο εφαρμογής ήταν οι συσκευές εγκεφαλικής διεπαφής (Brain-Machine Interface, BMI) και τα εγκεφαλικά εμφυτεύματα μέσω των οποίων οι επιχειρήσεις τεχνολογίας προσβλέπουν ότι μελλοντικά οι πελάτες τους θα «διατάσσουν τις μηχανές».

Διαβλέποντας οι ερευνητές ότι ρομπότ ελεγχόμενα από τους ανθρώπους με εγκεφαλικά εμφυτεύματα θα εμφανιστούν σύντομα στη ζωή μας έθεσαν το εξής ερώτημα: «Ας υποθέσουμε ότι μία ασθενής που έχει υποστεί παράλυση καθοδηγεί εγκεφαλικά ένα ρομπότ στη φροντίδα του μωρού της. Αν το μωρό πέσει από τα χέρια του ρομπότ, ποιος θα φέρει την ευθύνη; Η μητέρα που το έλεγχε εγκεφαλικά ή το ίδιο το ρομπότ;».

Οπως δήλωσε ο επικεφαλής της έρευνας και διευθυντής του Κέντρου Βιολογίας και Νευρομηχανικής Wyss της Γενεύης, καθηγητής John Donoghue, «παρότι ακόμη δεν κατανοούμε πλήρως τον τρόπο με τον οποίο λειτουργεί ο εγκέφαλος, βρισκόμαστε όλο και πιο κοντά στη δυνατότητα να αποκωδικοποιούμε αξιόπιστα ορισμένα εγκεφαλικά σήματα. Δεν μας επιτρέπεται να εφησυχάσουμε για το τι μπορεί να σημάνει αυτό για την κοινωνία. Πρέπει να εξετάσουμε προσεκτικά τις συνέπειες της συμβίωσης με ημι-έξυπνες μηχανές ελεγχόμενες από τον εγκέφαλό μας και θα πρέπει να είμαστε έτοιμοι να διασφαλίσουμε την ασφαλή και ηθική χρήση τους με κάποιους μηχανισμούς».
Η κατανομή της ευθύνης στα υβρίδια ανθρώπων-μηχανών ήταν το ένα θέμα που απασχόλησε τη διεθνή ομάδα ερευνητών. Το άλλο ήταν η προστασία των βιολογικών δεδομένων των ανθρώπων με εμφυτεύματα και διασύνδεση με μηχανές. Μπορεί η ψηφιακή παραβίαση (hacking) ενός παραπληγικού με εμφύτευμα να μην ακούγεται τόσο θελκτική για εγκληματίες, αλλά αυτό αλλάζει άρδην αν ο συγκεκριμένος είναι πολιτικός ή κάποιος με εξέχουσα θέση και επιρροή.

Μπήκε χάκερ στο μυαλό σου;

Οι υβριδικές αυτές καταστάσεις συνέργειας ανθρώπων-μηχανών μπορεί να

ακούγονται από τους πολλούς ως σενάρια επιστημονικής φαντασίας αλλά τεχνολογικά είμαστε πολύ κοντά στην πραγμάτωσή τους. Το δέλεαρ της μαζικής αξιοποίησής τους θα μεγεθυνθεί όταν οι τεχνολογικές επιχειρήσεις αρχίσουν να διαφημίζουν τα εγκεφαλικά εμφυτεύματα ως ενισχυτές μνήμης ή και εξυπνάδας. Και το ζήτημα της ανασφάλειας προκύπτει πολύ νωρίς: Μόλις ενισχυθεί ηλεκτρονικά ένα βιολογικό σήμα, το προκύπτον σήμα μπορεί να υποκλαπεί από κάποιον τρίτο, ιδιαίτερα καθώς αυτό μεταδίδεται συνήθως ασύρματα (π.χ. μέσω Bluetooth ή WiFi), χωρίς ασφαλή πρωτόκολλα επικοινωνίας.
Οι ερευνητές επισήμαναν τους κινδύνους και πρότειναν ως άμεσο μέτρο την ενσωμάτωση «δικαιώματος αρνησικυρίας» (veto) σε κάθε έξυπνη συσκευή ή ρομπότ που θα επικοινωνεί με ανθρώπινο εγκέφαλο. Με αυτόν τον τρόπο πιστεύουν ότι θα αποφεύγονται δράματα που μπορεί να προκληθούν επειδή ο εγκέφαλος του εντολέα παρανόησε ή αρρώστησε. Ταυτόχρονα όμως κάλεσαν τον τεχνολογικό κόσμο να προσφέρει έγκαιρα λύσεις για τη διασφάλιση της επικοινωνίας ανθρώπων-μηχανών, ώστε να μην καταλήξουν οι εγκέφαλοί μας δούλοι τρίτων.

Τέλος, κάλεσαν τις κυβερνήσεις να βελτιώσουν την ενημέρωση των ευρύτερων κοινωνικών στρωμάτων ως προς τη σχέση υγείας – εμβιομηχανικής. Κάθε πολίτης θα πρέπει να έχει τη βασική κατανόηση που είναι απαραίτητη για μια ενημερωμένη επιλογή. Μπορεί τα εμφυτεύματα και οι συνδέσεις εγκεφάλου – μηχανών να έχουν αποκαταστήσει την αυτονομία και την ποιότητα ζωής για πολλούς ασθενείς, αλλά οι πλήρεις ικανότητες και οι μακροπρόθεσμες επιπτώσεις στον ανθρώπινο εγκέφαλο και τον νου παραμένουν ασαφείς. Πρέπει η κοινωνία να δέχεται την απεριόριστη εμπορία τέτοιων πολλά υποσχόμενων συσκευών εφόσον είναι ασφαλείς και δεν προκαλούν σωματική βλάβη – αναρωτήθηκαν – ή πρέπει να λάβουμε μέτρα για την προστασία των τελικών χρηστών από την εκμετάλλευση και την έκθεση σε πιθανές παρενέργειες;

«Δεν θέλουμε να υπερεκτιμούμε τους κινδύνους, ούτε να δημιουργούμε εσφαλμένες ελπίδες σε όσους μπορούν να επωφεληθούν από την εμβιομηχανική και τη νευροτεχνολογία» κατέληξε ο καθηγητής Donoghue. «Στόχος μας είναι να διασφαλίσουμε ότι μια κατάλληλη νομοθεσία θα συμβαδίζει με αυτόν τον ταχέως αναπτυσσόμενο τομέα. Τώρα είναι ο καιρός να αρχίσουμε να παίρνουμε τα μέτρα που θα διασφαλίσουν την ευεργετική και ασφαλή χρήση της αλληλεπίδρασης εγκεφάλου – μηχανών».

Πηγή: tovima.gr