

Intelligent Machines in Responsible Positions

Based on Edmund M. A. Ronald & Moshe Sipper's article "Intelligence is not Enough:
On the Socialization of Talking Machines"

Zotou Xristina, Panagiota

Turing Test, one of the main subjects in Artificial Intelligence, supposed to be a test that shows if a machine has intelligence or not. Two hidden players, a machine and a human, answer to the questions of a second human. If the last one cannot decide which one of the players is the machine, then the machine passes the test and that means that the machine has intelligence. Edmund M. A. Ronald and Moshe Sipper examined hypothetical scenarios and they indicated that when a machine has passed the Turing Test, this is not enough in case we want to put it in a position such as that of a doctor, teacher or judge, that requires trustworthiness, reliability and responsibility. We can define the identity of someone by social interaction and through the time. Machines have no feelings and cannot be part of a social interaction like humans. Furthermore a machine has no fear of a punishment or the expectation of a rewarding, therefore it is very difficult to control its behavior and make it feel responsible. It seems that a machine has no identity and even more mind or soul. By changing the hypothetical scenarios of Ronald and Sipper we can end up to different conclusions; for example, machines have no motives, so they can be more impartial than humans, therefore more trustful. Today, it seems difficult to make machines with personal identity and social skills but the acknowledgment of matters like these is the first step in this kind of research.

Νοήμονα Μηχανήματα σε Υπεύθυνες Θέσεις

Μια συζήτηση για το άρθρο των Edmund M. A. Ronald και Moshe Sipper
“Intelligence is not Enough: On the Socialization of Talking Machines”

Ζοτού Χριστίνα, Παναγιώτα

Το Turing Test, ένα από τα βασικά θέματα στην Τεχνητή Νοημοσύνη, υποτίθεται πως είναι ένα τεστ που δείχνει αν ένα μηχάνημα έχει νοημοσύνη ή όχι. Δύο κρυμμένοι παίχτες, ένα μηχάνημα και ένας άνθρωπος, απαντούν σε ερωτήσεις ενός δεύτερου ανθρώπου. Αν ο τελευταίος δεν μπορεί να αποφασίσει ποιός από τους παίχτες είναι το μηχάνημα, τότε το μηχάνημα περνάει το τεστ και αυτό σημαίνει ότι έχει νοημοσύνη. Οι Edmund M. A. Ronald και Moshe Sipper εξέτασαν υποθετικά σενάρια και κατέδειξαν ότι όταν ένα μηχάνημα έχει περάσει το Turing Test αυτό δεν είναι αρκετό στην περίπτωση που θέλουμε να το τοποθετήσουμε σε μια θέση, όπως γιατρός, δάσκαλος ή δικαστής, που απαιτεί αξιοπιστία, ειλικρίνεια και υπευθυνότητα. Μπορούμε να προσδιορίσουμε την ταυτότητα κάποιου μέσω της κοινωνικής διάδρασης και με το πέρασμα του χρόνου. Τα μηχανήματα δεν έχουν αισθήματα και δεν μπορούν να πάρουν μέρος σε μια κοινωνική διάδραση όπως οι άνθρωποι. Επιπλέον ένα μηχάνημα δεν έχει τον φόβο μιας τιμωρίας ή την προσδοκία μιας ανταμοιβής, επομένως είναι πολύ δύσκολο να ελέγξουμε την συμπεριφορά του και να το κάνουμε να νιώσει υπεύθυνο. Φαίνεται ότι ένα μηχάνημα δεν έχει ταυτότητα και πολύ περισσότερο μυαλό ή ψυχή. Αλλάζοντας τα υποθετικά σενάρια των Ronald και Sipper μπορούμε να καταλήξουμε σε διαφορετικά συμπεράσματα· για παράδειγμα, τα μηχανήματα δεν έχουν κίνητρα και έτσι μπορούν να είναι πιο αμερόληπτα από τους ανθρώπους, επομένως πιο αξιόπιστα. Σήμερα, φαίνεται δύσκολο να φτιάξουμε μηχανήματα με προσωπική ταυτότητα και κοινωνικές δεξιότητες αλλά η αναγνώριση θεμάτων σαν και αυτά αποτελεί το πρώτο βήμα σε αυτού του είδους την έρευνα.

Εισαγωγή |

Το 1950 ο Alan Turing δημοσίευσε ένα άρθρο με το οποίο υπέδειξε έναν τρόπο για να απαντήσουμε στο ερώτημα αν μπορούν να σκεφτούν οι μηχανές, δηλαδή για να αποφασίσουμε αν τα μηχανήματα έχουν νοημοσύνη.¹ Πρότεινε γι' αυτόν τον σκοπό ένα παιχνίδι μίμησης “*The Imitation Game*” μέσω του οποίου, όπως και ο ίδιος σημειώνει, μπορούμε να αποφανθούμε μόνο για την νοημοσύνη των υπολογιστικών μηχανών που λαμβάνουν μέρος σ' αυτό και όχι για το σύνολο των μηχανών και των κομπιούτερ. Στο παιχνίδι παίρνουν μέρος 2 παίκτες, ένας άνθρωπος και ένα κομπιούτερ, και ένας κριτής που είναι άνθρωπος. Ο κριτής δεν μπορεί να δει ούτε να ακούσει τους παίκτες στους οποίους υποβάλλει ερωτήσεις και δέχεται απαντήσεις μέσω ενός τηλετύπου. Με βάση τις απαντήσεις που λαμβάνει, ο κριτής θα πρέπει να αποφασίσει ποιος από τους δύο παίκτες είναι ο άνθρωπος και ποιος το κομπιούτερ. Αν ο κριτής μαντέψει λάθος τότε μπορεί να θεωρηθεί ότι το κομπιούτερ έχει νοημοσύνη αφού κατάφερε να υποδυθεί με επιτυχία τον άνθρωπο και να ξεγελάσει τον κριτή.

Στα επόμενα χρόνια το άρθρο έγινε ένα από τα πιο πολυσυζητημένα στον χώρο της τεχνητής νοημοσύνης. Κυρίως οι συζητήσεις αφορούσαν την ισχύ του Turing Test (έτσι ονόμασαν το παιχνίδι μίμησης του Turing), δηλαδή το κατά πόσο είναι σωστό να αναγνωρισθεί ότι έχει νοημοσύνη όποιο μηχανήμα περνάει το τεστ, με άλλα λόγια αν το τεστ καταφέρνει να πετύχει τον σκοπό του και να κάνει “διάγνωση” νοημοσύνης. Από την μια έχουν γίνει προσπάθειες να βελτιωθεί το Turing Test και να ανταπεξέλθει σε κριτικές που π.χ. το θεωρούν πολύ ανθρωποκεντρικό, από την άλλη υπάρχουν εκείνοι που επιμένουν στην πλήρη απόρριψή του. Οι Edmund M. A. Ronald και Moshe Sipper στο άρθρο τους “*Η νοημοσύνη δεν είναι αρκετή:*

για την κοινωνικοποίηση των ομιλούντων μηχανημάτων” δεν παίρνουν μέρος στις κυρίαρχες συζητήσεις για την ισχύ του Turing Test αλλά μελετούν ένα άλλο ζήτημα που σχετίζεται με τα προβλήματα που προκύπτουν από την στιγμή που κάποιο μηχανήμα περάσει το τεστ και τεθεί σε χρήση για τους ανθρώπους.² Ανεπίλυτα ζητήματα εμπιστοσύνης, κοινωνικότητας, ταυτότητας και υπευθυνότητας προκύπτουν και φαίνεται να επισκιάζουν την ικανότητα λογικού συλλογισμού που αποδίδουμε στα κομπιούτερ που περνούν το τεστ περιορίζοντας τελικά το πεδίο δυνατής χρήσης τους. Μελετώντας σενάρια πιθανών χρήσεων τέτοιων κομπιούτερ οι Ronald και Sipper θέτουν υπό αμφισβήτηση την δυνατότητα επάρκειας της νοημοσύνης τους.

1. Turing, A.M. “Computing Machinery and Intelligence.” *Mind: A Quarterly Review of Psychology and Philosophy*, October 1950: 433-460.

2. Edmund, M.A. Ronald, Moshe Sipper. “Intelligence is not Enough: On the Socialization of Talking Machines”. *Minds and Machines*, November 2001, Volume 11, Issue 4, pp 567-576.

Ανάλυση σεναρίων από τους Ronald και Sipper |

Στα υποθετικά σενάρια που αναλύουν οι Ronald και Sipper προϋποθέτουν ότι τα κομπιούτερ που είναι υπό χρήση έχουν περάσει το Turing Test και αυτό γι' αυτούς σημαίνει ότι τα εν λόγω κομπιούτερ έχουν αποκτήσει τον τίτλο “Turing Chatterboxes”, ο οποίος θεωρούν ότι αποδεικνύει απλώς την ικανότητα που έχουν να συνομιλούν με τρόπο δυσδιάκριτο από εκείνον ενός ανθρώπινου συνομιλητή.

Ξεκινούν συγκρίνοντας δύο σενάρια. Στο πρώτο η dis Parker που αισθάνεται άρρωστη και δεν γνωρίζει η ίδια κάποιον γιατρό αποφασίζει να επισκεφθεί τον Δρ. Jackson που όλοι οι φίλοι της της συνιστούν ανεπιφύλακτα. Η άμεση εξυπηρέτηση που δέχεται, η ευγενική νοσοκόμα που επιβεβαιώνει την dis Parker ότι “ο Δρ. Jackson είναι ο καλύτερος που υπάρχει”, το παρουσιαστικό ενός ώριμου τζέντλεμαν με ιατρική ποδιά, του Δρ. Jackson, ακόμη και η επιβλητική πινακίδα στην πόρτα του ιατρείου, όλα συμβάλλουν στο να προσδώσουν στην dis Parker μια αίσθηση εμπιστοσύνης, ασφάλειας και άνεσης και να την

κάνουν να σκεφτεί ότι ήρθε στο σωστό μέρος. Στο δεύτερο σενάριο η dis Parker απευθύνεται σε μια δημόσια υπηρεσία που την παραπέμπει σε μια ανώνυμη Turing κλινική. Μπορεί το ιατρείο να είναι άψογο και η εξυπηρέτηση άμεση αλλά αυτά δεν αρκούν για να κάνουν την dis Parker να μην νιώσει άβολα όταν εξετάζεται από τον Turing γιατρό, ένα μηχανήμα που αρκέστηκε στο να παρουσιάσει τον εαυτό του ως “IQ175” και έπειτα προχώρησε σε μια σιωπηλή διάγνωση, σκανάροντας την dis Parker με κρυμμένους αισθητήρες, και στην τελική εκτύπωση μιας θεραπευτικής αγωγής.

Ακόμη και αν ο Δρ. Jackson και ο Turing γιατρός πρότειναν στην dis Parker την ίδια θεραπευτική αγωγή, αυτή φαίνεται να είναι σίγουρο ότι θα την ακολουθήσει μόνο στην πρώτη περίπτωση. Παρότι και οι δύο γιατροί έχει πιστοποιηθεί ότι είναι όχι μόνο νοήμονες αλλά και ειδικοί σ' αυτό που κάνουν, η dis Parker αμφιβάλει ως προς τις συμβουλές του Turing γιατρού. Φαίνεται λοιπόν ότι σε τέτοιου είδους εργασίες δεν αρκεί μόνο κάποιος να έχει νοημοσύνη και να είναι ειδικός αλλά χρειάζεται να κερδίσει και την εμπιστοσύνη του άλλου. Νοήμονα κομπιούτερ σε θέσεις γιατρών, δικαστών, οικονομικών συμβούλων, τραπεζικών υπαλλήλων, δασκάλων και άλλων, στους οποίους εμπιστευόμαστε την υγεία μας, τον νόμο, τις αποταμιεύσεις μας και την μόρφωση των παιδιών μας, φαίνεται να μην είναι χρήσιμα από την στιγμή που δεν μπορούν να κερδίσουν αυτή την εμπιστοσύνη μας. Σ' αυτές τις περιπτώσεις η εμπιστοσύνη είναι θεσμική και προσωπική. Εμπιστευόμαστε π.χ. τον γιατρό γιατί εμπιστευόμαστε τον θεσμό του πανεπιστημίου που του έχει παραχωρήσει το δίπλωμα καθώς και τον θεσμό του νοσοκομείου. Η εμπιστοσύνη μας στους θεσμούς στηρίζεται στην επικύρωση που επί χρόνια παρέχει σ' αυτούς η κοινωνία. Η προσωπική εμπιστοσύνη στηρίζεται στην συνεχή προσωπική διάδραση που έχουμε με κάποιον, επηρεάζεται από τον περιβάλλοντα κοινωνικό ιστό και χρειάζεται αρκετός χρόνος για ν' αναπτυχθεί.

Ο Turing δεν ξεκαθάριζε στο άρθρο του την χρονική διάρκεια του τεστ. Το μόνο που ανέφερε είναι η πίστη του ότι σε 50 χρόνια θα υπάρχουν κομπιούτερ τέτοια που ο μέσος κριτής δεν θα έχει περισσότερη από 70% πιθανότητα να κάνει την σωστή ταυτοποίηση έπειτα από 5 λεπτά ερωτήσεων. Μπορεί τα 5 λεπτά να αρκούν για την απόδοση νοημοσύνης, δεν αρκούν όμως για να εμπιστευτούμε κάποιον, είτε άνθρωπο είτε κομπιούτερ. Η εμπιστοσύνη απαιτεί εκτεταμένη κοινωνική διάδραση. Για να μπορεί ένα κομπιούτερ να κερδίσει στον ίδιο βαθμό την εμπιστοσύνη που θα κέρδιζε ένας άνθρωπος πρέπει να αναπτύξει κοινωνικές δεξιότητες και να λάβει μέρος σε πολυπληθή κοινωνικά δίκτυα. Η κοινωνικοποίηση των μηχανημάτων κρίνεται απαραίτητη όταν αυτά περνούν από έναν διαμεσοθλαστικό ρόλο (τηλέφωνο) σ' αυτόν ενός συνομιλητή (οικονομικός σύμβουλος). Το ζήτημα της εμπιστοσύνης γενικεύεται και αφορά το σύνολο της μοντέρνας τεχνολογίας, που συνεξελιίσσεται μαζί με τον άνθρωπο, από την στιγμή που κρίνεται ότι θα καθορίζει το πόσο ταιριαστή θα είναι μια τεχνολογία για τους ανθρώπους και άρα το πόσο απαραίτητη θα είναι η συνέχιση της παραγωγής και εξέλιξής της.

Το τρίτο σενάριο που εξετάζουν οι Ronald και Sipper αναδεικνύει το ζήτημα της ταυτότητας: Η dis Parker εξετάζεται αυτή τη φορά από έναν Turing γιατρό που εμφανίζεται στην οθόνη του υπολογιστή της με την μορφή ενός 50άρη τζέντλεμαν (ο ηλεκτρονικός υπολογιστής της dis Parker είναι εξοπλισμένος με την τελευταία τεχνολογία). Ο Turing γιατρός που εμφανίζεται στην οθόνη ξεκαθαρίζει ότι δεν είναι άνθρωπος αλλά μια μηχανή με το όνομα Δρ. Jackson και με την αποδοχή της dis Parker συνεχίζει κάνοντας την διάγνωση και συνιστώντας την απαραίτητη θεραπευτική αγωγή. Χαμογελώντας η εικόνα καθοσυχάζει την dis Parker πως “η μοντέρνα φαρμακολογία είναι πολύ αποτελεσματική και σε λιγότερο από 2 μέρες θα μπορείς να ξανασταθείς στα πόδια σου”. Την επόμενη ημέρα, νιώθοντας

χειρότερα, η dis Parker προσπαθεί να καλέσει τον Δρ. Jackson αλλά στην οθόνη του υπολογιστή εμφανίζεται ένας χιμπατζής που στριφογυρίζει ένα στηθοσκόπιο και λέει ότι αυτός είναι ο Δρ. Jackson. Φαίνεται λογικό η dis Parker να ανησυχεί. Η ανησυχία αυτή πηγάζει από το γεγονός ότι η dis Parker δεν μπορεί να αποδώσει μια ταυτότητα στον Turing γιατρό για να τον εμπιστευθεί. Ως άνθρωποι δεσμεύουμε την νοημοσύνη από την αρχή: ένα μυαλό-ένα σώμα. Δυσκολευόμαστε να εμπιστευθούμε άτομα που πάσχουν από την ψυχική διαταραχή της πολλαπλής προσωπικότητας ή άτομα που υποκρίνονται συνεχώς στην καθημερινή τους ζωή. Ομοίως, με Turing Chat-boxes δεν μπορούμε να είμαστε σίγουροι για την ταυτότητα (“μυαλό”) της οντότητας που ενεδρεύει μέσα στο κουτί (“σώμα”). Το πρόβλημα της εμπιστοσύνης περιπλέκεται όταν έχουμε να κάνουμε με ανώνυμα κομπιούτερ τα οποία μπορούν να αλλάζουν ηλεκτρονικό πρόσωπο ή φωνή όποτε θέλουν. Ακόμη πιο δύσκολο είναι να ορίσουμε τι είναι το “μυαλό” ή η “ψυχή” ενός μηχανήματος. Όμως, για να μπορέσουμε να

εμπιστευτούμε τα Turing Chatterboxes χρειάζεται να τους προσδώσουμε μια αναγνωρίσιμη και χρονικά σταθερή ταυτότητα. Το πώς μπορεί να γίνει κάτι τέτοιο παραμένει ένα άλυτο ερώτημα.

Στο τελευταίο σενάριο οι Ronald και Sipper εξετάζουν το ζήτημα της ευθύνης: Έπειτα από λιάθος διαγνώσεις που έκανε ένας Turing γιατρός η ιατρική επιτροπή συνεδριάζει και πρέπει να αποφασίσει είτε για την “επανεκπαίδευση” του κομπιούτερ είτε για την απόσυρσή του.

Εδώ τίθεται το θέμα του κατά πόσο ένα κομπιούτερ είναι υπόλογο για τις λιάθος πράξεις του που μπορεί να είχαν οδυνηρές συνέπειες. Αν τα μηχανήματα δεν έχουν συναισθήματα και κίνητρα, ελεύθερη βούληση και προθέσεις, πώς μπορούν να καταστούν υπεύθυνα για τις πράξεις τους; Ωστόσο, καθώς εισέρχονται στην κοινωνία και γίνονται όλο και πιο πολύπλοκα και αυτόνομα, πώς μπορούμε να τα κρατούμε υπό έλεγχο; Από την μια τα μηχανήματα δεν νιώθουν όπως οι άνθρωποι φόβο για κοινωνικές κυρώσεις ούτε αποβλέπουν σε αμοιβές ώστε να προσαρμόσουν ανάλογα την συμπεριφορά τους και από την άλλη η μόνη τιμωρία που μπορούμε να τους επιβάλουμε, η απόσυρση και καταστροφή τους, φαντάζει πολύ αυστηρή. Παρότι μπορούμε μέχρι

ενός σημείου να θεωρήσουμε ως υπεύθυνο τον κατασκευαστή τους, η συνεχής εξέλιξη και αυτονόμηση των Turing Chatterboxes θα μας φέρει αντιμέτωπους με το ζήτημα της ανάπτυξης υπεύθυνων κομπιούτερ στο μέλλον.

Οι Ronald και Sipper θεωρούν ότι σύντομα θα τεθούν σε χρήση Turing Chatterboxes σε παιχνίδια και μικροδουλειές όπως υπηρεσίες τηλεφωνικού καταλόγου και τουριστικές πληροφορίες. Μέσα δε από την χρήση που θα γίνει, το ζήτημα της νοημοσύνης θα παραχωρήσει την θέση του στα φλέγοντα ζητήματα της εμπιστοσύνης, της κοινωνικότητας, της ταυτότητας και της υπευθυνότητας.

Σχολιασμός του άρθρου |

Φαίνεται ότι ως άνθρωποι μπορούμε να εμπιστευτούμε περισσότερο άλλους ανθρώπους παρά κάποιο μηχανήμα. Που οφείλεται αυτό; Οφείλεται στο γεγονός ότι μέσω θεσμών και προσωπικής διάδρασης με τους άλλους, ως μέρη ενός ευρύτερου κοινωνικού δικτύου, συλλέγουμε πληροφορίες, που βασίζονται σε όσα ακούμε και βλέπουμε να κάνουν οι άλλοι, οι οποίες μας δίνουν την δυνατότητα να τους αξιολογήσουμε και να τους προσδώσουμε κάποια σταθερά χαρακτηριστικά που θα είναι αναγνωρίσιμα στο μέλλον και τα οποία αν είναι θετικά θα μας επιτρέψουν να τους εμπιστευτούμε. Επιπλέον μπορούμε να περιμένουμε ότι οι άλλοι θα προσαρμόσουν την συμπεριφορά τους στις κοινωνικές απαιτήσεις φοβούμενοι κυρώσεις ή αναζητώντας την επιβράβευση. Όσο περισσότερο λοιπόν γνωρίζουμε τους άλλους και είμαστε σε θέση να προβλέψουμε ως ένα βαθμό την συμπεριφορά τους, τόσο περισσότερο νιώθουμε ότι μπορούμε να τους εμπιστευτούμε, εφόσον

τους αξιολογούμε θετικά.

Τα σενάρια που αναλύουν οι Ronald και Sipper είναι τέτοια ώστε να καταδεικνύουν με σαφή τρόπο τα θεμελιώδη κοινωνικά ζητήματα που θα προκύψουν όταν μηχανές με νοημοσύνη τεθούν σε χρήση. Θα μπορούσαν αντί αυτών των σεναρίων να χρησιμοποιηθούν άλλα τα οποία να μας οδηγήσουν σε κάπως διαφορετικά συμπεράσματα. Το δεύτερο σενάριο θα μπορούσε να είναι διαφορετικό και να μην αφήνει μεγάλα περιθώρια ανησυχίας στην dis Parker. Θα μπορούσε για παράδειγμα όλοι οι φίλοι της dis Parker να της συνιστούσαν ανεπιφύλακτα τον Turing γιατρό ονόματι Δρ. Jackson, στο άψογο ιατρείο του οποίου η υποδοχή γίνεται από μια νοσοκόμα που επιβεβαιώνει την dis Parker ότι “ο Turing γιατρός Δρ. Jackson μπορεί να είναι ένα μηχανήμα αλλά είναι ο καλύτερος γιατρός που υπάρχει και δεν έχει κάνει ποτέ λάθος διάγνωση μέχρι τώρα”. Ο ίδιος ο γιατρός θα μπορούσε να μην είναι ένα κομπιούτερ που το μόνο που λείπει είναι το όνομά του και στην συνέχεια προχωρεί βουίζοντας σε μια σιωπηλή διάγνωση, αλλά θα μπορούσε να είχε την μορφή ενός 50άρη

τζέντλεμαν στην οθόνη του (όπως στο τρίτο σενάριο) ο οποίος πέρα από τις ικανότητες να κάνει διάγνωση της αρρώστιας μέσω αισθητήρων είναι ικανός να αναγνωρίζει και οποιαδήποτε ανησυχία και απορία στο πρόσωπο της dis Parker και να προχωρεί στον ανάλογο διάλογο. Η τελική θεραπευτική αγωγή που συστήνει βασίζεται όχι μόνο στην διάγνωση μέσω του σκαναρίσματος αλλά και στην ανάλυση των δεδομένων που η dis Parker του πρόσφερε πριν και μετά την διάγνωση απαντώντας σε ερωτήσεις όπως “πόσο καιρό έχετε ζαλάδες;”, “ακολουθείτε κάποια δίαιτα;” κ.ά. Τα μηχανήματα θα μπορούσαν εύκολα να κερδίσουν την θεσμική εμπιστοσύνη μας από την στιγμή που αναγνωρίζουμε θεσμούς όπως τα πανεπιστήμια και τα ερευνητικά κέντρα που πιστοποιούν την ικανότητα τους να σκέφτονται και να εξειδικεύονται σε ορισμένους τομείς. Η προσωπική εμπιστοσύνη είναι πιο περίπλοκη και απαιτεί περισσότερο χρόνο. Μπορούν όμως τα Turing Chatterboxes να την κερδίσουν αν τους προσδώσουμε έναν σταθερό “χαρακτήρα”. Παρότι αυτό είναι ένα άλυτο πρόβλημα θα μπορούσε να γίνει πιο εφικτή η λύση του αν είχαμε κομπιούτερ απόλυτα εξειδικευμένα στην εκάστοτε ειδική εργασία, χωρίς άλλες δυνατότητες που θα περιέπλεκαν τις λειτουργίες τους και θα τα καθιστούσαν πολυμορφικά. Το σενάριο που παρουσιάζει στην οθόνη τον Δρ. Jackson με την μορφή ενός χιμπατζή εξυπηρετεί μεν τον πρακτικό σκοπό της κατανόησης του προβλήματος απόδοσης ταυτότητας στα κομπιούτερ είναι όμως ένα σενάριο που δεν θα περιμέναμε να συμβεί αν υποθέσουμε ότι έχουμε φτιάξει Turing Chatterboxes με σκοπό να δίνουν ιατρικές συμβουλές και όχι να κοροϊδεύουν.

Μπορεί τα μηχανήματα να μην έχουν κίνητρα και προθέσεις, αυτό όμως μπορεί να είναι ένα πλεονέκτημα γι’ αυτά. Δυσκολευόμαστε να φανταστούμε ένα μηχανήματα που θα προσπαθεί να μας βγάλει άρρωστους ενώ δεν είμαστε για να μας γράψει φάρμακα και να πάρει την προμήθεια από την φαρμακευτική εταιρεία. Στην Ελλάδα

τουλάχιστον, είμαστε συνηθισμένοι σε τέτοιες περιπτώσεις γιατρών που βλέπουν τους πελάτες τους μόνο ως αντικείμενα προς εκμετάλλευση και μέσα πλουτισμού. Γιατί λοιπόν να μην βασιστούμε στην ιστορική μας πείρα και να μην εμπιστευτούμε περισσότερο έναν Turing γιατρό παρά έναν γιατρό που είναι άνθρωπος;

Επιπλέον ένα μηχανήματα σωστά κατασκευασμένο μπορούμε να φανταστούμε ότι κάνει λιγότερες λάθος διαγνώσεις απ’ ότι ένας άνθρωπος. Η ανάλυση των δεδομένων από ένα ισχυρό κομπιούτερ φαίνεται πιο αντικειμενική και σωστή από αυτήν που θα έκανε ένας άνθρωπος. Όσον αφορά την τιμωρία για λάθος πράξεις, αυτή δεν είναι αυστηρά δισδιάστατη. Η μερική απόσυρση ή η χρήση σε λιγότερο ακίνδυνες εργασίες είναι επιλογές πέραν της “επανεκπαίδευσης” ή της πλήρους απόσυρσης, που εφαρμόζονται και σε ανθρώπους.

Σαφώς το να αναπτύξουμε κομπιούτερ υπεύθυνα, με την δική τους ταυτότητα, που θα εμπνέουν εμπιστοσύνη και θα έχουν τις κοινωνικές δεξιότητες των ανθρώπων είναι κάτι το ευκαίο αλλά φαντάζει πολύ δύσκολο προς το παρόν. Η αναγνώριση όμως και μόνον αυτών των ζητημάτων μπορεί να ωθήσει την έρευνα σε σωστά μονοπάτια. Για άλλη μια φορά φαίνεται τελικά ότι κάτι είναι χρήσιμο, και γενικότερα κρίνεται και είναι αποδεκτό, με βάση το πόσο οικείο και ταίριαστο είναι στην δική μας ανθρώπινη ιδιότητα.

Βιβλιογραφία

Ronald, E.M.A. Sipper, M. "Intelligence is not Enough: On the Socialization of Talking Machines". *Minds and Machines*, November: 2001, Volume 11, Issue 4, pp 567-576.

Turing, A.M. "Computing Machinery and Intelligence." *Mind: A Quarterly Review of Psychology and Philosophy*, October 1950: 433-460.