Ρόμποκοπ, ρομποηθική, και το μέλλον του πολέμου: Μύθοι και Πραγματικότητα

Είναι 2028. Η Τεχεράνη, πρωτεύουσα του Ιράν, είναι ηλιόλουστη, κυριολεκτικά – αλλά και μεταφορικά επίσης, ίσως? Τεράστιοι δίποδοι ρομποτικοί **φύλακες** φροντίζουν να ανιχνεύουν και να εξουδετερώνουν οποιεσδήποτε πιθανές τρομοκρατικές ενέργειες. Γρήγοροι, ακριβείς, αποτελεσματικοί. Αλάνθαστοι? Όχι κατ'ανάγκη – έτσι και αλλιώς, θα μπορούσες να είσαι **αλάνθαστος**, μόνο εφόσον έχεις συμφωνήσει επακριβώς τι είναι σωστό και τι λάθος, και έχεις στην τσέπη σου τον απόλυτο γνώμονα για την λήψη Ορθών αποφάσεων (ότι και να σημαίνει αυτό!), και την χρυσή ζυγαριά του Δικαίου. Και μια και οι δίποδοι φύλακες της εισαγωγής του Νέου Ρόμποκοπ δεν φαίνεται να έχουνε τσέπες, μέσα στα πρώτα λεπτά της ταινίας έρχεται και το πρώτο «λανθασμένο θετικό» (**False Positive**), όπως λέγεται στη γλώσσα της ασφάλειας: ένα μικρό παιδί που έτρεξε μπροστά στους Φύλακες. Μόνο που το κόστος του λανθασμένου θετικού στην περίπτωση ενός ένοπλου αυτόνομου ρομπότ, είναι ο Θάνατος. Ποιός ευθύνεται όμως για αυτό? Φταίει η ηλεκτρομηχανική φύση των Φυλάκων? Αν υπήρχανε κύτταρα μέσα τους, ή ακόμα και ένα ανθρώπινο μυαλό, θα άλλαζε τίποτε? Αν ήταν απλά άνθρωποι, θα άλλαζε τίποτε? Τελικά το **συνολικό όφελος** των Φυλάκων, είναι θετικό, παρά τα «λανθασμένα θετικά»?

Αυτά και πολλά ακόμα είναι τα ερωτήματα που φτιάχτηκε για να γεννήσει το νέο Ρόμποκοπ στους θεατές του. Και δεν είναι κάταλευκο το πανί που βγήκε από το Χόλυγουντ – αλλά χρωματισμένο, υπέργεια και υπόγεια, με τις δικές του απαντήσεις αλλά και τις υποσυνείδητες αποχρώσεις, που σκοπεύει να εντυπώσει στους νευρώνες των θεατών του. Ποιές όμως είναι τελικά, οι **βασικές ερωτήσεις** που πραγματεύεται το Ρόμποκοπ, και ποιές οι απόψεις που φιλοδοξεί να περάσει? Και κυρίως, πως αυτές σχετίζονται με την σημερινή επιστημονική γνώση και έρευνα, και πως εντάσσονται σε ένα ευρύτερο πλαίσιο φιλοσοφίας και ηθικής? Αυτό θα προσπαθήσω να πραγματευτώ, έστω και εν μεγάλη συντομία, στο κείμενο που ακολουθεί.

Δεν είναι και μικρή η λίστα από τα **θεμελιώδη θέματα** που πραγματεύεται τελικά το Ρόμποκοπ. Ανθρώπινη και Τεχνητή Νοημοσύνη, Λογική και Συναίσθημα, Ελεύθερη Βούληση, Ηθική Συλλογιστική, η σχέση Κυβερνήσεων με ισχυρές Πολυεθνικές, η ευθύνη της Δημόσιας Διανόησης, μέχρι –έστω και πλαγίως- το ερώτημα του αν οι Φύλακες (Αστυνομικοί ή Πλατωνικοί?) είθιστε να έχουνε οικογένεια. Ας τα δούμε όμως ένα-ένα:

Πίσω στο 2028, με μια περιστροφή της υδρογείου, από τη Τεχεράνη στο Ντετρόητ. Ενώ το Ιράν είναι γεμάτο από οπλισμένους ρομποφύλακες, και μάλιστα κατασκευασμένους από μια Αμερικάνικη εταιρεία, οι ΗΠΑ έχουνε μόνο παλιομοδίτικους οπλισμένους αστυνομικούς. Γιατί όμως? Ο "**public intellectual**" Dreyfus έχει πείσει την κοινή γνώμη αλλά και την κυβέρνηση των ΗΠΑ ότι άδεια φόνου (**license to kill**) μπορούν να έχουνε μόνο άνθρωποι και όχι μηχανές. Γιατί όμως? Δεν μπορούνε οι μηχανές να είναι πιο ακριβείς από τους ανθρώπους? Το όνομα του Dreyfus βέβαια αποτελεί έμμεση αναφορά και στον φιλόσοφο Hubert Dreyfus, γνωστού για την κριτική του ενάντια στην παντοδυναμία της τεχνητής νοημοσύνης, και αντίστοιχα του Dennett της

ταινίας στον φιλόσοφο Daniel Dennett, τον κατά μεγάλο μέρος υπέρμαχο του AI. Τι χρειάζεται όμως για να δοθεί σε κάποιον η άδεια φόνου?

Τρία από τα βασικά συστατικά απαντήσεων σε αυτό το ερώτημα είναι συνήθως τα εξής: Πρώτον, δυνατότητα λήψης **ορθής απόφασης** για το αν και πότε θα πατήσει την σκανδάλη. Δεύτερον, δυνατότητα ακριβούς και γρήγορης **στόχευσης**, και τρίτον, **νομική** ευθύνη.

Πληρούν όμως αυτές τις τρείς προϋποθέσεις τα **ρομπότ**, ή οι **άνθρωποι**? Και πέρα από τα δύο αυτά άκρα του βιολογικού-τεχνητού φάσματος, θα τις πληρούσε ένα **βιολογικό-τεχνητό υβρίδιο**, ένα «μυαλό στη γυάλα» (brain-in-a-vat) στη φιλοσοφική αργκό, όπως είναι ο Ρόμποκοπ 'Αλεξ στην ταινία, που αποτελείται από ένα ανθρώπινο μυαλό, που αφού έζησε πολλά χρόνια στο σώμα του, αποσπάστηκε και τοποθετήθηκε σε ένα μηχανικό σώμα – και μάλιστα επαυξήθηκε με ολοκληρωμένα κυκλώματα (τσιπάκια) που φυτεύτηκαν στο μυαλό του. Ας δούμε τις προϋποθέσεις μια-μια:

Πρώτον, η δυνατότητα λήψης ορθής απόφασης για εκτέλεση. Μια παραδοσιακή σχολή σκέψης υποστηρίζει ότι μόνο η λογική, η απαλλαγμένη από κάθε είδους **συναισθηματική στρέβλωση**, μπορεί να οδηγήσει σε ορθές αποφάσεις. Συμφωνεί όμως αυτό με τα νευροεπιστημονικά δεδομένα? Και αρκεί ένας ορθός μηχανισμός λήψης αποφάσεων για ορθές αποφάσεις, ή μήπως χρειαζόμαστε τουλάχιστον ακόμα δύο συστατικά: πρώτον, πληροφορίες και δεύτερον, ένα ξεκάθαρο – και υπολογίσιμο – ηθικό σύστημα?

Είναι όμως το **συναίσθημα πάντα αντίπαλος της λογικής?** Το 1848, μια μεγάλη έκρηξη, έσπρωξε μια μεταλλική ράβδο μεγάλου μήκους με τόση ορμή, ώστε διαπέρασε το κρανίο ενός νεαρού εργάτη σιδηροδρόμου: του Φινέα Γκέητζ, το όνομα του οποίου έμεινε ιστορικό στις νευροεπιστήμες. Πάρα πολλοί άλλοι άνθρωποι με συγκεκριμένες **εγκεφαλικές κακώσεις**, έχουν αποτελέσει σημαντικές πηγές για την μελέτη του ανθρωπίνου εγκεφάλου και της σχέσεως του με την νόηση και την συμπεριφορά. Το ενδιαφέρον για μας στοιχείο από τέτοιες μελέτες είναι ότι άνθρωποι με κακώσεις σε περιοχές που σχετίζονται κυρίως με την συναισθηματική (και όχι την λογική) λειτουργία, τελικά καταλήγουν να έχουνε σημαντικό πρόβλημα και σε διαδικασίες που εκ πρώτης όψεως φαίνονται καθαρά «λογικές», όπως η επίλυση μαθηματικών προβλημάτων (Damasio, το σφάλμα του Καρτεσίου).

Επίσης, αναλύοντας την διαδικασία σκέψης της επίλυσης ενός μαθηματικού προβλήματος, φαίνεται ότι πέρα από τα εύκολα βήματα μιας απόδειξης, όπου υπάρχει συνειδητή επίγνωση και δοκιμή διαφόρων πιθανών επομένων βημάτων, υπάρχουνε και βήματα που απαιτούνε ένα «νοητικό άλμα», στα οποία παύει να υπάρχει επίγνωση μιας αλληλουχίας σκέψεων που οδηγεί στη λύση – και η λύση φαίνεται απλά να «μας έρχεται». Κατά τη διάρκεια αυτών ακριβώς των νοητικών αλμάτων, που αποτελούνε μέρος της επίλυσης προβλημάτων που φαίνονται **καθαρά λογικά, φαίνεται να υπεισέρχονται εγκεφαλικές λειτουργίες και περιοχές που σχετίζονται με συναισθήματα**: μια αποτίμηση της γεύσης χιλιάδων δυνατών επομένων βημάτων συμβαίνει αστραπιαία, και μερικά από αυτά εξέρχονται ως λύση της διαδικασίας. Άρα,

αυτό που φαινομενικά ονομάζουμε «λογική νοητική διαδικασία» αλληλοσυμπληρώνεται με κάποιες μορφές συναισθηματικά επιχρωματισμένων διαδικασιών ταχείας ολιστικής αποτίμησης - οπότε **δεν είναι πάντα το συναίσθημα αντίπαλος της λογικής** – αλλά σε κάποιες περιπτώσεις φαίνεται να είναι και απαραίτητος σύμμαχος της, ώστε να ληφθούν βέλτιστες αποφάσεις!

Βέβαια, αυτό δεν σημαίνει ότι δεν υπάρχουνε συναισθηματικές και άλλες συστηματικές στρεβλώσεις (**cognitive biases**) που μας κάνουνε να αποκλίνουμε από τις ορθές αποφάσεις – και πολλές από αυτές έχουνε μελετηθεί σε βάθος, π.χ. σε σχέση με την λήψη οικονομικών αποφάσεων – όπως στο Nobel οικονομικών των Kahneman και Tverksy. Αλλά ακόμα και κάποιες από αυτές τις φαινομενικές αποκλίσεις, μπορούνε να δικαιολογηθούν σε κάποιες περιπτώσεις αν τις δούμε μέσα σε ένα ευρύτερο πλαίσιο, π.χ. της νοητικής οικονομίας και του bounded rationality. Δηλαδή, τα πράγματα σε σχέση με την εμπλοκή συναισθηματικών διαδικασιών στην λήψη ορθών αποφάσεων, **απλά, δεν είναι τόσο απλά**! Και εξαρτώνται πολύ από το πως κανεί οριοθετεί την τεχνητή διχοτόμηση μεταξύ λογικής και συναισθήματος – που κατά άλλους δεν αποτελούνε διακριτές κατηγορίες, αλλά συμμετέχουνε σε ένα πιο ολιστικό συνεχή χάρτη νοητικών ποικιλιών.

Οπότε, δεν υπερτερούνε κατ'ανάγκη και πάντα οι μηχανές από τους ανθρώπους ως προς τη λήψη ορθών αποφάσεων, τουλάχιστον όχι λόγω της «έλλειψης συναισθηματικών στρεβλώσεων». Όμως, αν είναι έτσι, και αν τελικά απαιτείται μερικές φορές και τεχνητή συναισθηματική νοημοσύνη από τις μηχανές, **θα μπορέσουνε ποτέ οι μηχανές να «έχουνε συναισθήματα»**? Και εδώ εισέρχεται ένας ακόμα δημοφιλής μύθος: ότι τα συναισθήματα αποτελούνε μόνο ανθρώπινη ιδιότητα. Παραταύτα, και πάλι βέβαια αφού αποσαφηνίσει κανείς εμπειρικά ελέγξιμες ενδείξεις που να επιβεβαιώνουνε την πρόταση «η οντότητα Χ έχει συναισθήματα» (όπου Χ είμαι εγώ, ο Γιώργος, ο Γάτος μου, ή το Ρομπότ μου) για να ξέρουμε για τι μιλάμε, φαίνεται ότι ακόμα και οι σημερινές μηχανές (πόσο μάλλον οι μελλοντικές) έχουνε κάποιες δυνατότητες συναισθηματικής νοημοσύνης! Π.χ. υπάρχουνε προγράμματα που διαβάζουνε ανθρώπινες εκφράσεις στο πρόσωπο ή τον τόνο της φωνής και τον κατηγοριοποιούνε ώς «φοβισμένο» ή «ευτυχισμένο» κλπ. Και υπάρχουνε και εικονικοί χαρακτήρες ή ρομπότ που χαμογελάνε ή κατσουφιάζουνε ανάλογα με την διάδρασή τους – όπως και το πολύ δημοφιλές στην Ιαπωνία παιχνίδι **Tamagochi** – ένα εικονικό ανθρωπάκι τσέπης που θέλει φροντίδα για να παραμένει ευτυχισμένο και να αναπτύσσεται! Σε κάποιες περιπτώσεις μάλιστα, αυτά τα προγράμματα μπορεί να είναι πιο ακριβή από τους ανθρώπους στην αναγνώριση λεπτών ενδείξεων συναισθημάτων. Και το μέλλον του "**affective computing**" ως τομέα διαγράφεται λαμπρό.

Όμως, το να μπορεί ένα μηχάνημα να **αναγνωρίζει συναισθήματα**, ή να δίνει την **αίσθηση ότι βρίσκεται σε συναισθηματικές καταστάσεις**, αυτό σημαίνει ότι **ΕΧΕΙ ΟΝΤΩΣ συναισθήματα**? Από μια πιο οντολογική προσέγγιση, αυτό είναι ενδεχομένως αμφιλεγόμενο. Με μια πιο φαινομενολογική όμως, όχι και τόσο. Σκεφτείτε π.χ. το τι σας κάνει να πιστεύετε ότι οποιοσδήποτε άλλος άνθρωπος, πέρα από το εαυτό σας, έχει συναισθήματα: παρατηρείτε απλά τις εκφράσεις του και την συμπεριφορά του, και έτσι συνάγετε ότι «Για να φαίνεται

έτσι, πρέπει να αισθάνεται Χ». Έιχατε όμως ποτέ άμεση, **πρωτογενή πρόσβαση**, στην εσωτερική κατάσταστη οποιουδήποτε άλλου ανθρώπου πλην του εαυτού σας? Ή τελικά, ακόμα και αυτό είναι μια **δευτερογενής πεποίθηση** (πιστεύω ότι πιστεύει Χ, πιστεύω οτι αισθάνεται Χ κλπ.)? Και μια και από αυτά που βλέπουμε πιστεύουμε ότι ο Γιώργος αισθάνεται χαρούμενος, και όχι μόνο ο Γιώργος, αλλά και ο σκύλος μας όταν κουνάει την ουρά του, για ποιό λόγο να είναι κάτι το εντελώς διαφορετικό όταν ένα ρομπότ χαμογελά? Έτσι και αλλιώς, ποτέ δεν μπήκαμε στο μυαλό του Γιώργου ή του σκύλου μας – όπως και του ρομπότ, για να αισθανθούμε άμεσα ότι αισθάνεται. Και δεν αναφέρομαι σε όλα αυτά για να σας οδηγήσω σε ένα σκεπτικιστικό σολιπσισμό (στην αργκό της φιλοσοφίας), απλά για να σας θυμήσω ότι μόνο στα δικά σας συναισθήματα έχετε πρωτογενή (άμεση) πρόσβαση (και πάλι, μάλιστα, μερική!), και ότι τα συναισθήματα οποιασδήποτε άλλης οντότητας (ανθρώπινης, ζωικής, ή ρομποτικής) απλά μπορείτε να τα υποθέσετε δευτερογενώς, σε σχέση με αυτά που βλέπετε. Άρα: **ναι, και οι μηχανές μπορούνε να έχουνε όχι μόνο συναισθηματική νοημοσύνη αλλά και συναισθήματα**, σύμφωνα με την παραπάνω θεώρηση.

Άρα, σε σχέση με την πρώτη προϋπόθεση της δυνατότητας έκδοσης «άδειας φόνου», αυτήν της **λήψης ορθών αποφάσεων**, δεν φαίνεται να υπάρχει καθαρό προβάδισμα ανθρώπων ή μηχανών, τουλάχιστον καθολικό ή εκ προοιμίου, και σίγουρα όχι λόγω της «συναισθηματικής μόλυνσης» των ανθρώπων – κάθε άλλο μάλιστα. Έτσι και αλλιώς, η έλλειψη καθαρού προβαδίσματος ισχύει και σε πολλά ακόμα πεδία σύγκρισης μεταξύ ανθρώπων και μηχανών: π.χ. το Deep Blue της IBM νικάει τον παγκόσμιο πρωταθλητή σκακιού Kasparov – ενώ σε πολλά προβλήματα αναγνώρισης προτύπων οι άνθρωποι υπερτερούνε κατά πολύ των μηχανών – ακόμα τουλάχιστον.

Βέβαια, δεν είδαμε ακόμα δύο άλλες πτυχές της ορθής απόφασης: Πρώτον, την ύπαρξη **επαρκούς πληροφορίας, και σχετικής με την απόφαση προς λήψη**: δεν αρκεί μόνο ο νοητικός μηχανισμός- πρέπει να τροφοδοτηθεί και με τα σωστά δεδομένα για μια σωστή απόφαση! Εκεί, υπεισέρχεται μια πολύ ενδιαφέρουσα πτυχή της υπόθεσης: ένα μεγάλο μέρος αυτής της πληροφορίας, ενδέχεται να μην περιέχεται μόνο στις αισθήσεις ή τη μνήμη του ρομπότ ή ανθρώπου: αλλά να είναι **εξωτερική πληροφορία** – π.χ. στο διαδίκτυο, σε μια ειδική βάση δεδομένων υπόπτων, ή σε εικόνες από ένα δίκτυο καμερών που καλύπτουνε μέρη μιας πόλης. Έτσι, και στην νέα ταινία του ρόμποκοπ, τα ρομπότ ή ο υβριδικός Άλεξ έχουνε πρόσβαση στις κάμερες και στις αποθηκευμένες βιντεοσκοπήσεις και τις βάσεις δεδομένων, που βρίσκονται στο διαδίκτυο, έξω από το σώμα τους. Εκεί οι μηχανές φαίνεται να έχουνε ένα προβάδισμα σε σχέση με τους ανθρώπους: έχουνε άμεση πρόσβαση στην πληροφορία αυτή, ενώ οι άνθρωποι χρειάζονται κάποιο ειδικό ενδιάμεσο (**human-machine interface**) για την παρουσίαση της πληροφορίας ώστε να διέρθει μέσω των αισθήσεων στο μυαλό τους. Από την άλλη μεριά όμως, ένα μεγάλο μέρος των πληροφοριών που υπάρχουνε, τουλάχιστον αυτή τη στιγμή, στο διαδίκτυο ή οι εικόνες που παράγονται, δεν είναι σε μορφή άμεσα κατανοήσιμη από μηχανές (**machine understandable format**); οπότε και πάλι, τουλάχιστον σήμερα, τα πράγματα δεν φαίνεται να κλίνουν πολύ καθαρά προς την μεριά των μηχανών ή των ανθρώπων, ακόμα και σε σχέση με αυτή τη

δεύτερη πτυχή της ορθής λήψης αποφάσεων, δηλαδή της πρόσβασης σε επαρκή και σχετική πληροφορία.

Και υπολείπεται και η τρίτη προς εξέτασιν πτυχή: η υπάρξη ενός ξεκάθαρου – και υπολογίσιμου – ηθικού συστήματος. Ορθή απόφαση, αλλά με ποιά θεώρηση του σωστού ή λάθους, καλού ή κακού? Πολλές οι σχετικές έννοιες σε μια τέτοια συζήτηση. Υπάρχει «απόλυτο κακό» ή «απόλυτο καλό», ή έστω μια εύκολη σχέση διάταξη του τύπου «καλύτερο» ή «χειρότερο»? Αν ναι, πως τοποθετούμε τα όρια της μελέτης των επιπτώσεων μιας πράξης? Καλό για ποιόν? Για μας? Για το παιδί μας? Για την Ελλάδα? Για την Ανθρωπότητα? Για την μητέρα Γη, μαζί με την βιόσφαιρά της? Και σύμφωνα με το τι η κάθε ομάδα θεωρεί καλό? Σε αυτά που δεν θα θέλαμε να συμβούνε σε ένα άνθρωπο – υπάρχει αρκετή πανανθρώπινη συμφωνία – και έτσι και έχουνε προκύψει θεμέλια όπως τα **Δικαιώματα του Ανθρώπου**. Σε αυτά όμως τα θετικά που θα ήθελε ο κάθε άνθρωπος? Γνώση? Ισχύ? Χρήμα? Αυτοπραγμάτωση? Αγάπη? Είχε ο Επίκουρος δίκαιο, ο Ιησούς, ο Πλάτωνας, ή ο Εκκλησιαστής της Παλαιάς Διαθήκης? Έστω, μπορεί να πει κανείς, ο φύλακας – ρομπότ δεν προσπαθεί να μεγιστοποιήσει διαφορετικές όψεις του «Κοινού Καλού» (δεν είναι το ηλεκτρονικό ισοδύναμο των Κυβερνητων του Πλάτωνα – αλλά ένας αστυνομικός, που θα άνηκε στην δεύτερη μάχιμη τάξη της Πολιτείας), απλά προσπαθεί να αποφύγει τα πασιφανώς και ομόφωνα άσχημα – π.χ. την δολοφονία ενός πολίτη από έναν εγκληματία. Αλλά και πάλι, μπορεί να υπολογίσει τις επιπτώσεις της πράξης του, και πως θα σταθμίσει το βάρος τους? Και κυρίως, ποιό σύστημα ηθικής συλλογιστικής (**ethical reasoning**) θα χρησιμοποιήσει?

Οι τρεις νόμοι της ρομποτικής του Asimov – αποτελούνε απλά ένα χρησιμότατο αλλά πολύ χονδροειδές περίβλημα – που και πάλι έχει μεγάλες υπολογιστικές δυσκολίες, όταν προσπαθεί κανείς να το υλοποιήσει στην πράξη. Και πέρα από popular science θεωρήσεις, όπως του Asimov, τα **θεμελιώδη ερωτήματα της ηθικής συλλογιστικής** είναι βαριά, και αποτελούνε και σημείο διχασμού της κοινής γνώμης. «Ο σκοπός αγιάζει τα μέσα», «το κόστος της ανθρώπινης ζωής», και άλλες τέτοιες εκφράσεις βρίσκονται στο κέντρο αυτών των συζητήσεων. Μια πτήση, ακυβέρνητη, με έναν μη ειδικευμένο πιλότο ημιαναίσθητο, ζητάει άδεια να προσγειωθεί στο Ελευθέριος Βενιζέλος, Αύγουστο μήνα, γεμάτο με κόσμο. Θα την λάβει? Θα την καταρρίψει ο Ρόμποκοπ με πυραύλους, για να μειώσει την πιθανότητα ενός πολύ πιο πολύνεκρου ατυχήματος? Το κλασσικότατο «**πρόβλημα του τρόλλευ**» της Φιλίπα Φουτ, κεντρικό στην ηθική συλλογιστική, είναι ιδιαίτερα παρόμοιο με το παραπάνω.

Και εκτός αυτού, υπάρχει και μια άλλη μεγάλη διάσταση της υπολογιστικής υλοποίησης της ηθικής με την οποία θα θελήσουμε να εφοδιάσουμε ένα μηχάνημα: αυτή της **πρακτικής υλοποιησιμότητας**. Αρκετά ηθικά συστήματα, κυρίως αυτά που βασίζονται στην μεγιστοποίηση της στατιστικά μέσης τιμής κάποιας συνάρτησης χρησιμότητας που προκύπτει από τις πιθανές εκβάσεις των συνεπειών της κάθε πράξης προς επιλογή, στηρίζονται στην υπολογιστική προβλεψιμότητα όχι μόνο των πρωτογενών, αλλά και των δευτερογενών συνεπειών των πράξεων μας, και μάλιστα αυτών που διαμεσολαβούνται από άλλους ανθρώπους. Τι εννοώ όμως με αυτό? Απλά, ότι αυτές οι θεωρίες θα μπορούσαν απλοϊκά να συνοψιστούν στο: «Διάλεξε εκείνη την πράξη που θα

κάνει στατιστικά το περισσότερο καλό, όχι μόνο διαμέσω των άμεσων αποτελεσμάτων της, αλλά και μέσω των έμμεσων – δηλαδή, διαμέσω των μελλοντικών πράξεων άλλων ανθρώπων που θα προκύψουν ως αποτέλεσμα του πως επηρέασε η πράξη σου άλλους ανθρώπους». Αυτό μπορεί να ακούγεται όμορφο, όμως συχνά υπολογιστικά είναι αδύνατο να προβλεφθούν δευτερογενείς συνέπειες, είτε λόγω έλλειψης επαρκών πληροφοριών είτε λόγω εγγενούς υπολογιστικής πολυπλοκότητας. Οπότε, σε αυτή την περίπτωση, απλοϊκά και άκαμπτα ηθικά συστήματα (όπως π.χ. ένα **σύνολο «εντολών» για το τι να κάνει η τι να μην κάνει κανείς** σε εύκολα ελέγξιμες συνθήκες) μπορεί να καταλήγουν να είναι πολύ πιο αποδοτικά στην πράξη τους από τα παραπάνω.

Έτσι λοιπόν καλύψαμε και τις τρεις πτυχές της πρώτης προϋπόθεσης ως προς την απονομή «άδειας φόνου» σε έναν άνθρωπο ή μια μηχανή ή ακόμα και ένα υβρίδιο ανθρώπου-μηχανής, που θα δράσει π.χ. ως αυτόνομος αστυνομικός φύλακας. Οι άλλες δυο είναι αρκετά απλούστερες: η **δυνατότητα ακριβούς και γρήγορης στόχευσης**, και η νομική ευθύνη. Το πρώτο είναι θέμα μηχανικής κατασκευής και αισθητηριο-κινητικών μηχανισμών – και εκεί φαίνεται οι μηχανές να προπορεύονται ήδη των ανθρώπων, και έτσι ακόμα και ο υβριδικός Άλεξ του Ρόμποκοπ, που έχει εφοδιαστεί με ηλεκτρονικά συστήματα στόχευσης και πυροδότησης-σε-δράση, υπερτερεί του προγενέστερου ανθρώπινου Άλεξ.

Το δεύτερο όμως είναι αρκετά πιο σύνθετο, όπως υποδεικνύει και η δυσκολία **νομικής ευθύνης** υπογραφής μηχανικού για έργα πληροφορικής: ένα πληροφοριακό σύστημα που αλληλεπιδρά μέσω αισθητήρων και κινητήρων με το περιβάλλον, είναι αδύνατο να ελεγχθεί διεξοδικά σε σχέση με την ορθότητα της λειτουργίας του, μια και απλούστατα υπάρχουνε χιλιάδες διαφορετικές συνθήκες περιβάλλοντος και λειτουργίας που μπορεί να προκύψουν. Έτσι, ακόμα και αν χρησιμοποιηθούν οι ελάχιστες (και ιδιαίτερα δύσχρηστες) μέθοδοι που υπάρχουν για τον εγγυημένα ορθό σχεδιασμό λογισμικού (**formal design methods**), υπάρχουνε πολλές περιπτώσεις για τις οποίες ένα σφάλμα λειτουργίας (ενδεχομένως προκαλώντας αναίτιους θανάτους στην περίπτωση του Ρομποφρουρού) δεν μπορεί να αποδοθεί στην ευθύνη ενός προγραμματιστή που σχεδίασε το σύστημα. Ακόμα περισσότερο, για ένα πολύ σύνθετο έργο όπως ένα μελλοντικό Ρόμποκοπ, **υπάρχουνε πολλοί που μπορεί να φταίνε για ένα ατύχημα**: ο αγοραστής-χρήστης, η εταιρεία κατασκευής, κάποιοι σχεδιαστές, ή ακόμα και μια λάθος πληροφορία που έφτασε στο σύστημα μέσω του διαδικτύου ή μέσω μιας κάμερας. Έτσι, όχι μόνο δεν υπάρχει αυτή τη στιγμή νομικό πλαίσιο για απόδοση ευθυνών ενός αυτόνομου μηχανήματος με άδεια φόνου, αλλά και η απόδοση ευθυνών για μια δυσλειτουργία μάλλον θα απαιτεί μακροπρόθεσμες έρευνες παρόμοιες με αυτές που συμβαίνουν μετά από αεροπορικά ατυχήματα.

Είναι αξιοσημείωτο ότι στις ένοπλες δυνάμεις των ΗΠΑ, όπου έχουνε **αρχίσει να χρησιμοποιούνται πειραματικά τα ένοπλα ρομπότ**, και όπου τηλεχειριζόμενα αεροπλάνα χωρίς πιλότο επιχειρούν και φονεύουν καθημερινά στο Αφγανιστάν, ακόμα η άδεια φόνου δίνεται μόνο σε ανθρώπους – και δεν επιτρέπεται σε αυτόνομες μηχανές να αποφασίσουν μόνες τους πότε θα τραβήξουν την σκανδάλη, αλλά πάντα κάποιος άνθρωπος δίνει την εντολή, ώστε

να υπάρχει και υπαιτιότητα. Παραταύτα, σε περιπτώσεις συμπλοκών που ο **χρόνος απόκρισης** είναι σημαντικότατος, αυτό δίνει ένα συγκριτικό μειονέκτημα σε σχέση με την άδεια φόνου σε μηχανές. Βέβαια, υπάρχει μεγάλη ελαστικότητα στο τι σημαίνει «να δίνεις την εντολή»: μπορεί απλά να σημαίνει «να επιτρέπεις στο σύστημα να αποφασίζει αυτόνομα για το αν θα τραβήξει την σκανδάλη ή όχι από εδώ και πέρα». Εκτός αυτού, αν το σκεφτεί κανείς καλύτερα, η «άδεια φόνου» σε μηχανές με μεγάλο χρονικό ορίζοντα και χωρίς άμεση ανθρώπινη εποπτεία δεν είναι κάτι το καινούργιο στην στρατιωτική ιστορία. Ένα από τα πιο απάνθρωπα όπλα, οι **νάρκες**, που στοχεύουν στην ακινητοποίηση προσωπικού και εκμεταλλεύονται τον εγγενή πόνο για το συνάνθρωπο για να δημιουργήσουνε περισσότερους νεκρούς, με τα εκατομμύρια άμαχα θύματά τους με απώλειες μελών, αποτελούνε λαμπρά μελανό παράδειγμα αυτής της κτηνωδίας.

Πέρα από τα καθαρά τεχνητά αυτόνομα ρομπότ, και τους απλούς ανθρώπους – που τοποθετείται ο **υβριδικός Άλεξ** της ταινίας, με μηχανικό σώμα και βιολογικό μυαλό? Ενώ φαίνεται πως αυτό το πρωτότυπο βιονικό κατασκεύασμα αποτελεί εξεχόντος ιδιαίτερη κατηγορία, ίσως τελικά και να μην είναι. Πρώτα από όλα, ένα σημαντικό ερώτημα σε τέτοια υβρίδια είναι το **που τελειώνει ο άνθρωπος και που αρχίζει η μηχανή**? Παραδείγματος χάριν, σήμερα υπάρχει ερευνητικά τεχνολογία για εξωσκελετικά **ρομποτικά πρόσθετα** (**exoskeletons**), που επιτρέπουν αύξηση σωματικών δυνατοτήτων σε στρατιώτες, χωρίς να χάνουνε κανένα βιολογικό σωματικό μέρος τους. Εκτός αυτού, υπάρχουνε ερευνητικά και **νοητικά πρόσθετα** για ανθρώπους: παραδείγματος χάριν μια φορητή έξυπνη κάμερα με αποθήκευση που αναγνωρίζει πρόσωπα και θυμάται προηγούμενες συναντήσεις και διαλόγους – ένα νοητικό πρόσθετο μνήμης δηλαδή. Επίσης, κάτι πολύ απλούστερο που μπορεί να θεωρηθεί ως ένα χαλαρά συζευγμένο υβρίδιο ανθρώπου-μηχανής είναι ένα τανκ με έναν ανθρώπινο χειριστή. Και ένα άλλο παράδειγμα υβριδίου, με μερική απώλεια βιολογικού σώματος (ακούσια βέβαια), αποτελεί ένα προσθετικό ρομποτικό χέρι για αναπήρους.

Τι συμβαίνει όμως όταν έχουμε ένα θωρακισμένο τανκ με δύο ανθρώπινους χειριστές, και μάλιστα ένα που να υποστηρίζεται μέσω ασυρμάτου από εξωτερικούς παρατηρητές? Τότε σιγά-σιγά ξεφεύγουμε από τα όρια της ατομικής τεχνητής ή βιολογικής νοημοσύνης, και των υβριδίων ανθρώπων-μηχανών, χαλαρά ή στενά συζευγμένων, και εισερχόμαστε στην ενσώματη υβριδική συλλογική νοημοσύνη (**collective intelligence**) με ένα μίγμα ανθρώπων και μηχανών σε σύζευξη και διάδραση. Αυτό το μίγμα αρχίζει και δρα σαν μια συλλογική ευφυής οντότητα με κατανεμημένο σώμα, τα όρια της νοημοσύνης και δυνατοτήτων της οποίας ενδέχεται να ξεπερνούνε πολύ όχι μόνο τους μεμονωμένους ανθρώπους ή ρομπότ, αλλά και τις μη μικτές ομάδες ανθρώπων η ρομπότ. Και κάπου εκεί βρίσκεται το μέλλον της αστυνόμευσης, του στρατού, αλλά και της νοημοσύνης γενικότερα: στην δυνατότητα δημιουργίας ισχυρά αποδοτικών και αρμονικών ομάδων ανθρώπων και μηχανών, με την νοημοσύνη τους να αντλείται και να διανείμεται μέσω δικτύων. Και περίπου έτσι είναι τα πράγματα και στον Ρόμποκοπ: η πραγματική αύξηση δυνατοτήτων έρχεται όχι μόνο από το μηχανικό σώμα, αλλά από το δίκτυο καμερών, τα άμεσα προσβάσιμα αποθηκευμένα αρχεία της αστυνομίας με

πληροφοριακό και οπτικοακουστικό υλικό, τις αυτόματες εκτιμήσεις επικινδυνότητας, και όλα αυτά τα στοιχεία **κατανεμημένης νοημοσύνης**.

Καλύψαμε λοιπόν ένα μεγάλο αριθμό από **θεμελιώδη ερωτήματα** που προκύπτουν από την νέα ταινία, όπως: ποιά είναι η σχέση ανθρώπινης και τεχνητής νοημοσύνης? Υπερτερούν τα ρομπότ των ανθρώπων, ή τα υβρίδια όπως ο Άλεξ? Είναι το συναίσθημα αντίπαλος της λογικής? Γίνεται οι μηχανές να έχουν συναισθήματα? Πότε θα έπρεπε να δίνεται η άδεια φόνου σε έναν άνθρωπο η μια μηχανή? Τι είναι η **ρομπο-ηθική**, και ποιά η σχέση της με την φιλοσοφική ηθική, το δίκαιο, και την υπολογιστική υλοποίηση ηθικής συλλογιστικής? Για να φτάσουμε έτσι σιγά-σιγά στα νοητικά και μηχανικά πρόσθετα, και στην κατανεμημένη και συλλογική νοημοσύνη, και να εξετάσουμε αν όντως τα ρομπότ σήμερα έχουνε την άδεια να σκοτώνουνε.

Πολλά από τα άλλα βασικά ερωτήματα του ρόμποκοπ, όπως η ελεύθερη βούληση, η θέση των Δημοσίων Διανοουμένων και η σχέση κυβερνήσεων και μεγάλων ιδιωτικών συμφερόντων, αλλά και το θέμα της οικογένειας για τους φύλακες, δεν καλύφθηκαν, αλλά θα ασχοληθούμε μαζί τους σε επόμενο άρθρο.

Παραταύτα, αξίζει να κλείσουμε με μια άλλη επίκαιρη σχετική παρατήρηση. Το 1981, εμφανίσθηκε ο πρώτος «**προσωπικός υπολογιστής**», το περίφημο IBM PC. Πέρα όμως από την ονομασία του, κατάφερε να διαδοθεί τόσο και να γίνει πραγματικά ο πρώτος «προσωπικός» υπολογιστής, χάριν σε δυο εφαρμογές που τρέχανε σε αυτόν: τον επεξεργαστή κειμένου και το λογιστικό φύλλο Lotus 123. Αυτά λοιπόν ήταν τα «**killer apps**» που υποστήριξαν την μεγάλη διάδοση των υπολογιστών. Παρόμοιες εφαρμογές που να επιτρέψουν την ραγδαία διάδοση των ρομπότ, δεν έχουνε προκύψει ακόμα. Ίσως να είναι η οικιακή βοήθεια; ίσως κάποια άλλη υποστηρικτική εφαρμογή. Και υπάρχει ένα ευρύ φάσμα κοινωνικά ωφέλιμων τομέων που άρχισαν να καλύπτονται: από βοηθητικά ρομπότ για ηλικιωμένους και άτομα με ειδικές ανάγκες, μέχρι ρομπότ διάσωσης, ιατρικά ρομπότ, και πολλά άλλα.

Παραμένει όμως το ερώτημα: **ποια θα είναι η «killer app» που θα καθιερώσει την ευρύτατη χρήση ρομπότ** έξω από την βιομηχανία? Ελπίζω δυστηχώς να μην είναι έτσι, αλλά πολύ φοβάμαι, ότι η απάντηση είναι ότι αυτή η «killer app» δεν θα είναι μόνο μεταφορικά "killer", αλλά δυστυχώς και κυριολεκτικά. Ο **επόμενος μεγάλος πόλεμος, πιθανότατα να διεξαχθεί με ρομπότ**. Και δεν θα είναι μόνο ρομπότ εναντίον ρομπότ, αλλά και ρομπότ εναντίον ανθρώπων. Τι διαφορά έχει αυτό από ένα τανκ που στοχεύει σε ένα ακάλυπτο στρατιώτη, θα μου πείτε? Έχει, και δεν έχει. Και ένα μεγάλο μέρος όλων των θεμελιωδών ερωτημάτων που αγγίξαμε εδώ, έχουνε άμεση σχέση με αυτή τη διαφορά. Ας ελπίσουμε ότι τουλάχιστον η μη στρεβλωμένη επίγνωσή τους, και οι συζητήσεις που θα προκύψουνε, οι νέοι δημόσιοι διανοούμενοι, τα ΜΜΕ και η κοινή γνώμη θα βοηθήσουνε ώστε η ανθρωπότητα να μην ανακαλύψει και πάλι το **εύρος της διττής της υπόστασης,** που εκτείνεται από την μέγιστη ωφέλιμη δημιουργία ως την μέγιστη επιζήμια κτηνωδία, μετά από ένα ακόμα μεγάλο ανθρωπιστικό δράμα, όπως προέκυψε μετά την εισαγωγή των χημικών και των πυρηνικών όπλων. Πολύ πιθανόν όμως και εμείς να το ζήσουμε, και τα παιδιά μας. Σκεφτείτε λοιπόν καλά, και κινητοποιηθείτε!