Παιδαγωγικός Λόγος

Τόμ. 32, Αρ. 1 (2026)

Λόγος περί της Τεχνητής Νοημοσύνης

Μηχανές-Δικαστές με Τεχνητή Νοημοσύνη. Με Ηθική;

Ιωάννα Μαλανδράκη

doi: 10.12681/plogos.33468

Βιβλιογραφική αναφορά:

Μαλανδράκη Ι. (2026). Μηχανές-Δικαστές με Τεχνητή Νοημοσύνη. Με Ηθική; . *Παιδαγωγικός Λόγος*, *32*(1), 49–63. https://doi.org/10.12681/plogos.33468

Ιωάννα ΜΑΛΑΝΔΡΑΚΗ

Μηχανές-Δικαστές με Τεχνητή Νοημοσύνη. Με Ηθική;

doi:https://doi.org/10.12681/plogos.

Ι. Εισαγωγή

Ο ΠΟΛΙΤΙΣΜΟΣ ΜΑΣ ΠΙΣΤΩΝΕΤΑΙ ΣΤΗ ΝΟΗΜΟΣΥΝΗ ΜΑΣ,[1] ΓΡΑΦΕΙ Ο καθηγητής επιστήμης των υπολογιστών Stuart Russel και τα λόγια του περιτριγυρίζουν στη σκέψη μου.

Η παραδοχή ότι διανύουμε μια εποχή με έναν ορμητικά καλπάζοντα ρυθμό προόδου στον τεχνολογικό πολιτισμό είναι αδιαμφισβήτητη. Στο άκουσμα της λέξεως «τεχνολογία» κάποιοι/ες θα προτάξουν τις θετικές επιρροές της στην καθημερινή ζωή, την προσφορά της στις επιστήμες, κάποιοι/ες θα συλλογιστούν την αρνητική της όψη προβαίνοντας σε βαθιά παρατήρηση των επιδράσεών της, ίσως στον χώρο της χρήσης νέων εργαλείων για χάρη της βίας και κάποιοι/ες θα προβληματιστούν για την ροπή, που δημιουργείται στον άνθρωπο, προς την πίστη ότι όλα μπορούν να υλοποιηθούν από τα επινοήματά του, τα τεχνήματά του κατευθύνοντάς τον σε μία παθητική στάση με κανένα κίνητρο για να διεκδικήσει, όπως πριν, ό,τι του αναλογεί.

Η ζωή, πλέον, φαίνεται να ορίζεται από την τεχνολογία. Ο σύγχρονος άνθρωπος εξαναγκάζεται να επιθυμεί να είναι μέρος όλων των καινών γεγονότων, αφού οι περισσότεροι τομείς επιδέχονται συνεχή κατάκτηση νέων επιπέδων. Ως απόρροια, της εν λόγω συνθήκης, εγείρονται εύλογες ανησυχίες σε σχέση με διάφορες εκφάνσεις χρήσης των τεχνολογικών επιτευγμάτων. Μία από τις εν λόγω εκφάνσεις είναι η Τεχνητή Νοημοσύνη

[1] Stuart Russel, *Συμβατή με τον άνθρωπο; η Τεχνητή Νοημοσύνη και το πρόβλημα του ελέγχου*, μτφρ. Νίκος Αποστολόπουλος (Εκδοτικός Οίκος ΤΡΑΥΛΟΣ, 2021), 141.

(Τ.Ν.) η οποία, έχοντας εκτεταμένη χρήση σε ποικίλα πεδία της ανθρώπινης δραστηριότητας, καταλαμβάνει, ίσως, την πιο αιχμηρή θέση. Ένα από τα επιμέρους πεδία εφαρμογής της Τ.Ν., το οποίο αναδεικνύει την αιχμηρότητά της και αποτελεί αντικείμενο πραγμάτευσης του παρόντος δοκιμίου είναι η λήψη δικαστικών αποφάσεων.

ΙΙ. Τεχνητή Νοημοσύνη και Δίκαιο: μία αξιοσημείωτη περίπτωση

Ο ρόλος της Τ.Ν. τη δεδομένη χρονική στιγμή στην πρακτική του Δικαίου είναι επικουρικός και πραγματώνεται μέσω λογισμικών συστημάτων που αποσκοπούν στην προσομοίωση της ικανότητας του ανθρώπου να λαμβάνει αποφάσεις.[2] Το COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) συνιστά ένα από τα πιο γνωστά λογισμικά ποινικών αδικημάτων. Το 1998 οι Tim Brennan και Dave Wells, επικεφαλής του Ινστιτούτου Northpointe, ανέπτυξαν τον αλγόριθμο COMPAS, ένα εργαλείο εκτίμησης κινδύνου πιθανής υποτροπιάζουσας παραβατικής συμπεριφοράς, που χρησιμοποιείται σε πολλές πολιτείες των ΗΠΑ,[3] παρέχοντας αρωγή στους/στις δικαστές στο στάδιο της λήψης αποφάσεων προανακριτικής απελευθέρωσης για τον/την εκάστοτε κατηγορούμενο/η.[4] Ωστόσο, είναι σημαντικό να σημειωθεί το γεγονός ότι, παρότι ο εν λόγω αλγόριθμος εξάγει έναν βαθμό που αντιστοιχεί σε μία από τις κατηγορίες κινδύνου πιθανής επανάληψης, η ακριβής πεπερασμένη αλληλουχία εντολών, η οποία τον συνθέτει, δε έγκειται στη σφαίρα γνώσης μας.[5] Επομένως, η άγνωστη εσωτερική λειτουργία του συστήματος καθιστά ανέφικτη τη λογοδοσία του, η οποία συνδέεται άρρηκτα με την ευθύνη, που χαρακτηρίζεται από ηθική αιτιότητα και είναι ένας από τους θεμελιώδεις στόχους στο πλαίσιο της εύρυθμης δικαστικής λειτουργίας.

---

[2] Peter Jackson, *Introduction to Expert Systems* (Addison-Wesley, 1999), 1.

[3] Alexandra Mac Taylor, "AI Prediction Tools Claim to Alleviate an Overcrowded American Justice System… But Should they be Used?," *Stanford Politics*, September 13, 2020, https://stanfordpolitics.org/2020/09/13/ai-prediction-tools-claim-to-alleviate-an-overcrowded-american-justice-system-but-should-they-be-used/.

[4] Eugenie Jackson and Christina Mendoza, "Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not," *Harvard Data Science Review* 2, no. 1 (2020): 3, https://doi.org/10.1162/99608f92.1b3dadaa.

[5] Ellora Thadaney Israni, "When an Algorithm Helps Send You to Prison," *The New York Times,* October 26, 2017, https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html.

Σχεδόν μία δεκαετία πριν, το 2014, έλαβε χώρα μία τρόπον τινά ληστεία στην Πολιτεία της Φλόριντα. Η δεκαοχτάχρονη Brisha Borden και μία φίλη της έχοντας καθυστερήσει να παραλάβουν την πνευματική αδερφή της πρώτης από το σχολείο, όταν εντόπισαν στον δρόμο ένα παιδικό ποδήλατο κι ένα σκούτερ, χωρίς ιδιοκτήτη, επιχείρησαν να τα οδηγήσουν για να φτάσουν γρηγορότερα στον προορισμό τους.[6] Ευθύς αμέσως αντιλήφθηκαν ότι δεν ήταν εφικτό να κινηθούν με αυτά εξαιτίας του μεγέθους τους και εγκατέλειψαν την προσπάθεια. Λίγο αργότερα, όμως, συνελήφθησαν από τις αστυνομικές αρχές λόγω της αναφοράς του περιστατικού από έναν γείτονα. Ως αποτέλεσμα, οι δύο φίλες κατηγορήθηκαν για διάρρηξη και κλοπή αντικειμένων, με τη συνολική χρηματική αξία των κλοπιμαίων να ανέρχεται σε ογδόντα δολάρια. Ωστόσο, αυτή δεν ήταν πρωτόγνωρη κατάσταση για την Borden, αφού στο παρελθόν είχε εμπλακεί σε παραπτώματα.[7] Στον αντίποδα αυτής της υπόθεσης, το 2013 εκτυλίχθηκε ένα άλλο περιστατικό με τον σαρανταενάχρονο Vernon Prater, έναν άνδρα με βεβαρυμμένο ποινικό μητρώο, ο οποίος συνελήφθη για κλοπή εργαλείων αξίας περίπου ογδόντα έξι δολαρίων από ένα κατάστημα.[8]

Στο στάδιο της προανακριτικής διαδικασίας, ο αλγόριθμος εκτίμησε τη μελλοντική έκβαση. Εξήγαγε το πόρισμα ότι η Brisha Borden είχε υψηλή προδιάθεση να επαναλάβει παραβατική συμπεριφορά, ενώ ο Vernon Prater χαμηλή.[9] Μετά το πέρας δυο χρόνων, δεν είχε καταλογιστεί νέο παράπτωμα στην Brisha Borden, εν αντιθέσει με τον Vernon Prater ο οποίος είχε προβεί σε ληστεία ηλεκτρονικών ειδών αξίας χιλιάδων δολαρίων από την αποθήκη ενός σπιτιού με συνέπεια την έκτιση ποινής οκτώ ετών.[10] Όπως διαπιστώθηκε από την έρευνα της ProPublica,[11] ο αλγόριθμος δε βασίστηκε στο παρελθόν των κατηγορουμένων και στην πράξη που διέπραξαν, αλλά στη φυλετική προέλευσή τους, αφού η απόφαση που εξεδόθη, από το σύστημα, σχετιζόταν με το σκουρόχρωμο δέρμα της Brisha Borden και το ανοιχτόχρωμο του Vernon Prater.[12]

---

[6] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks," *ProPublica,* May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[7] Στο ίδιο.

[8] Στο ίδιο.

[9] Στο ίδιο.

[10] Στο ίδιο.

[11] Μη κερδοσκοπική οργάνωση ερευνητικής δημοσιογραφίας για το δημόσιο συμφέρον.

[12] Angwin et al., "Machine Bias."

Σε αυτό το σημείο αναδεικνύεται ζήτημα καίριας σημασίας. Η προκατάληψη και η μεροληψία που αποκτούν τα συστήματα Τ.Ν., εξαιτίας του/της προγραμματιστή/ριας τους ή της μεθόδου βαθιάς μάθησης, ενός τρόπου μηχανικής μάθησης που επιτυγχάνεται μέσω της εμπειρίας των συστημάτων, τα κατευθύνουν προς την εξαγωγή λανθασμένων δεδομένων και τα καθιστούν επισφαλή. Μάλιστα, το ζήτημα της άγνωστης εσωτερικής λειτουργίας των συστημάτων Τ.Ν. δυσχεραίνει την αποτροπή προβλημάτων αυτού του είδους. Είναι, λοιπόν, αναγκαία η διαφανής Τ.Ν., αφού το σύστημα εκθέτοντας τον τρόπο με τον οποίο εξήγαγε ένα συγκεκριμένο αποτέλεσμα, θα ενημερώνει τον άνθρωπο-χειριστή και η περιστολή λήψης λανθασμένης απόφασης, ειδικά σε κρίσιμες συνθήκες – όπως στη λήψη δικαστικής απόφασης – κατά συνέπεια, θα δύναται να επιτυγχάνεται.

Επιπροσθέτως, όπως φαίνεται, η χρήση της Τ.Ν. δεν θα περιοριστεί στον επικουρικό ρόλο της. Στην πρωτεύουσα της Κίνας, το Πεκίνο, έχει, ήδη, κατασκευαστεί η πρώτη μηχανή-δικαστής του κόσμου με Τ.Ν., η οποία παρουσιάστηκε στο ευρύ κοινό τον Ιούνιο του 2019.[13] Σε αυτό ακριβώς το σημείο έγκειται ο ηθικοφιλοσοφικός προβληματισμός για τις ηθικώς δρώσες τεχνητές οντότητες που κέντρισε το ενδιαφέρον μου και συνοψίζεται στο ερώτημα: έχουμε αναλογιστεί το ενδεχόμενο τα συστήματα Τ.Ν., γενικότερα αλλά και ειδικότερα στον δικαστικό κλάδο, να καταστούν κάποια στιγμή υπερ-νοήμονα και να αυτονομηθούν με αποτέλεσμα να βλάψουν – έστω και άθελά τους – την ανθρωπότητα;[14]

III. Η Ηθική της Τεχνητής Νοημοσύνης στο πλαίσιο της Υπερ-νοημοσύνης

Η Τ.Ν. με τη δράση της σε ποικίλα πεδία της ανθρώπινης δραστηριότητας έχει ενεργοποιήσει ηθικές αρχές και κατηγορίες, όπως συμβαίνει, άλλωστε, σε κάθε πεδίο που δημιουργεί ηθικά διλήμματα. Το ζήτημα που εκκινεί τον ηθικοφιλοσοφικό προβληματισμό του παρόντος δοκιμίου είναι

[13] "Beijing Internet Court launches online litigation service center," Beijing Internet Court, last modified July 1, 2019, https://english.bjinternetcourt.gov.cn/2019-07/01/c_190.htm.

[14] Για μια σύνοψη των ηθικών προβληματισμών που εγείρει η δυνατότητα δικαστηριακής χρήσης της Τ.Ν., καθώς και για τις οντολογικής και επιστημολογικής φύσεως προεκτάσεις της, βλ.: Άλκης Γούναρης και Γιώργος Κωστελέτος, «Γράφοντας τον αλγόριθμο του Καλού: Η Τεχνητή Νοημοσύνη ως μηχανή απόδοσης Δικαιοσύνης», *Ηθική. Περιοδικό φιλοσοφίας* 19 (2024): 5-27. https://doi.org/10.12681/ethiki.39654.

η προκατάληψη και η μεροληψία που αποκτούν τα συστήματα Τ.Ν.. Όπως έχω, ήδη, υποδείξει το η προκατάληψη της Τ.Ν. εκδηλώνεται είτε απευθείας από τον/την ίδιο/α τον/την προγραμματιστή/ρια του συστήματος Τ.Ν. είτε από τον τρόπο εκπαίδευσής του, αφού έχει προηγηθεί η έκθεσή του σε ένα περιβάλλον από το οποίο αντλεί ερεθίσματα.

Τι συγκροτείται, όμως, στη σημασία που αποδίδεται στον όρο «προκατάληψη της Τ.Ν.»; Η μεροληψία της Τ.Ν. είναι δυνατό να προκύψει ποικιλοτρόπως. Κατά πρώτον, μία περίπτωση είναι ένα σύστημα να χαρακτηρίζεται από προκατάληψη διότι ο/η αρμόδιος/α να του εμφυσήσει αξίες προσλαμβάνει επιρροές για ένα αντικείμενο πραγμάτευσης και έπειτα τις εφαρμόζει (εν γνώσει ή εν αγνοία του) σε διαφορετικό αντικείμενο με συνέπεια να καταλήγει σε λανθασμένους συλλογισμούς.[15] Επομένως, κατά αντιστοιχία, αυτό το άτομο εκπαιδεύει το σύστημα Τ.Ν. να δρα τοιουτοτρόπως.[16] Κατά δεύτερον, απασχολεί τη συζήτηση η ανθρώπινη τάση προς τη γνωστική προκατάληψη, ήτοι προς την ερμηνεία μίας πληροφορίας μέσα σε ένα συγκεκριμένο πλαίσιο, ώστε αυτή να αποτελεί επαλήθευση για κάτι που κάποιος/α πιστεύει, χωρίς να έχει παρέλθει το στάδιο της λογικής διεργασίας και της διασταύρωσης πηγών για ασφαλή γνώση.[17] Αναλογικά, το σύστημα Τ.Ν. αφομοιώνει, μέσω της μηχανικής μάθησης, από τον/την χρήστη τί να δέχεται και τί να αποκρούει στη βάση συγκεκριμένων πεποιθήσεων. Ακόμα, τίθεται επί τάπητος η στατιστική προκατάληψη.[18] Η εν λόγω μορφή προκύπτει από την αρχική αμεροληψία του συστήματος, με την έννοια ότι εκπαιδεύτηκε για μία συγκεκριμένη συνθήκη και έδρασε ορθώς, αλλά, εν συνεχεία, λόγω της ανεπάρκειάς του να αντιληφθεί την αιτία που προχώρησε σε εκείνη την κρίση, χρησιμοποίησε το συγκεκριμένο σύνολο δεδομένων σε άλλη συνθήκη με διαφορετικά δεδομένα και ακολουθώντας την προηγούμενη στρατηγική έδρασε, εν τέλει, μεροληπτικά.[19]

Λαμβάνοντας τη σκυτάλη από την τελευταία μορφή μεροληψίας ενός συστήματος Τ.Ν συλλογίζομαι την υπόθεση ενός υπερ-νοήμονος συστήματος. Με λίγα λόγια, στρέφω την προσοχή μου στην περίπτωση που

[15] Vincent C. Müller, "Ethics of Artificial Intelligence and Robotics," *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), ed. Edward N. Zalta, https://plato.stanford.edu/archives/sum2020/entries/ethics-ai/.
[16] Στο ίδιο.
[17] Στο ίδιο.
[18] Στο ίδιο.
[19] Στο ίδιο.

μηχανές-δικαστές λειτουργούν μελλοντικά με επιβλαβή τρόπο για την ανθρωπότητα, ακόμα και αν δρουν δίκαια, έχοντας αποδεσμευτεί από ζητήματα προκατάληψης λόγω της υπερ-νοήμονος κατάστασής τους, υιοθετώντας στρατηγική που επιφέρει το θεμιτό αποτέλεσμα, αλλά κατά την εφαρμογή της δημιουργεί άλλου είδους προβλήματα.

Ο Max Tegmark, καθηγητής φυσικής στο MIT και πρόεδρος του Ινστιτούτου για το Μέλλον της Ζωής, στο βιβλίο του που τιτλοφορείται *Life 3.0-Τι θα σήμαινε να είσαι άνθρωπος στην εποχή της τεχνητής νοημοσύνης;* θέτει ποικίλα ερωτήματα για την υπερ-νοήμονα συνθήκη της Τ.Ν. Διερωτάται, μεταξύ άλλων, πώς είναι δυνατό να κατασκευαστεί Τ.Ν. με στόχους και να διασφαλιστεί ότι οι εν λόγω στόχοι θα παραμείνουν ίδιοι ακόμη και αν η Τ.Ν. γίνει πιο ευφυής.[20] Ο καθηγητής επισημαίνει ότι τη στιγμή της δημιουργίας συστήματος Τ.Ν. οι αξίες, οι ηθικοί και νομικοί κανόνες, που θα εμφυσήσουν σε αυτό, εξαρτώνται από τον/την σχεδιαστή/ρια του και κατ' επέκταση, όταν το σύστημα καταστεί υπερ-νοήμων, θα είναι στον ύψιστο βαθμό ικανό να εκπληρώσει τους στόχους του,[21] εφόσον η νοημοσύνη συνεπάγεται την ικανότητα πραγμάτωσης στόχων.

Επομένως, ο Max Tegmark αναρωτιέται αν μπορούμε να ευθυγραμμίσουμε τους στόχους μας με εκείνους ενός υπερ-νοήμονος συστήματος και ως αποτέλεσμα να έχουμε μία «φιλική Τ.Ν.», όπως την χαρακτηρίζει ο Αμερικανός θεωρητικός Eliezer Yudkowsky ή αν το σύστημα Τ.Ν. λόγω της υπερ-νοημοσύνης του θα ενεργήσει με δικούς του στόχους, με δυνητικά καταστροφικές συνέπειες, ακόμα και χωρίς σκόπιμη επιδίωξη τέτοιας δράσης.[22] Θεωρεί, δηλαδή, ότι το σύστημα ενδέχεται να έχει την ικανότητα να θέτει δευτερεύοντες στόχους ή ακόμα και να λειτουργεί βάσει δικών του στόχων, ώστε να επιτύχει τον τελικό στόχο του, διότι θα έχει τη δυνατότητα αναστοχασμού των στόχων με τους οποίους έχει διαποτιστεί.[23] Συνεπώς, αναδύεται ο προβληματισμός που προκύπτει από τον συγκερασμό του ζητήματος σχετικά με το ηθικό και νομικό σύστημα αξιών, το οποίο θα εισάγεται σε ένα σύστημα Τ.Ν. και του ζητήματος ελέγχου της Τ.Ν. στο μέλλον.

Την ίδια στάση υιοθετεί ο Σουηδός φιλόσοφος Nick Bostrom. Εκείνος έχει περιγράψει το πείραμα σκέψης "paperclip maximizer," το οποίο

[20] Max Tegmark, *LIFE 3.0 Τι θα σήμαινε να είσαι άνθρωπος στην εποχή της τεχνητής νοημοσύνης;,* μτφρ. Νίκος Αποστολόπουλος (Εκδοτικός Οίκος ΤΡΑΥΛΟΣ, 2018), 375.
[21] Στο ίδιο, 391.
[22] Στο ίδιο, 390-404.
[23] Στο ίδιο, 390-403.

συνοψίζεται στην ιδέα ότι «μία Τ.Ν. σχεδιασμένη να διαχειρίζεται την παραγωγή σε ένα εργοστάσιο, έχει ως τελικό στόχο τη μεγιστοποίηση της κατασκευής συνδετήρων και προχωρά με τη μετατροπή πρώτα της Γης και έπειτα ολοένα και περισσότερο μεγαλύτερων κομματιών του παρατηρήσιμου σύμπαντος σε συνδετήρες».[24] Έτσι, ο τελικός στόχος δεν είναι επιζήμιος για τον άνθρωπο, αλλά η υπερ-νοημοσύνη επιχειρώντας να έχει το μέγιστο δυνατό αποτέλεσμα είναι ικανή να μετατρέψει τους ανθρώπους σε συνδετήρες, υποθέτοντας ότι εκείνοι μπορούν να την απενεργοποιήσουν με συνέπεια να μη δημιουργηθούν πολλοί συνδετήρες.[25]

Σε αντίθεση με την προαναφερθείσα καταστρεπτική θέση, οι Michael Anderson και Susan Leigh Anderson, καθηγητές Φιλοσοφίας, συνιστούν μια πιο αισιόδοξη προσέγγιση. Σε συνέντευξή τους ερωτήθηκαν για το κατά πόσο πιστεύουν ότι ενυπάρχει κίνδυνος για την ανθρώπινη ζωή από την Τ.Ν. στο εγγύς μέλλον, αν τα συστήματα καταστούν πιο ευφυή και αποκρίθηκαν ότι η δράση των συστημάτων Τ.Ν. προσδιορίζεται από τον τρόπο ανάπτυξής τους επισημαίνοντας ότι μη απειλητικές μηχανές μπορούν να σταθούν αρωγοί μας ακόμα και στη βελτίωση της συμπεριφοράς μας· βέβαια, δεν αγνοούν τους κινδύνους που έθεσε η Τ.Ν., όμως θεωρούν πως τοιουτοτρόπως θα διαφυλαχτεί η νοημοσύνη μας στην περίπτωση που κινδυνέψουμε γενικά.[26]

Συν τοις άλλοις, οι Anderson εισηγούνται το ερευνητικό πρόγραμμα «Ηθική των Μηχανών».[27] Επιχειρούν, μεταξύ άλλων, να δημιουργήσουν ένα ασφαλές περιβάλλον για τους ανθρώπους που τρέφουν ανησυχίες και αβεβαιότητες σχετικά με το ενδεχόμενο ύπαρξης ηθικών συστημάτων Τ.Ν. στο πλαίσιο των αυτόνομων ευφυών συστημάτων Τ.Ν.[28] Ως απάντηση προβάλλουν την ανθρωποκεντρική διάσταση του ζητήματος, καθώς υπογραμμίζουν ότι «η ανησυχία ότι οι μηχανές που ξεκινούν να συμπεριφέρο-

[24] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014), 149, δική μου μετάφραση.

[25] Στο ίδιο.

[26] Michael Anderson, Susan Leigh Anderson, Alkis Gounaris, and George Kosteletos, "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson," *Conatus – Journal of Philosophy* 6, no. 1 (2021): 192-193, https://doi.org/10.12681/cjp.26832.

[27] «Δημιουργία μιας μηχανής που η ίδια να ακολουθεί μία ηθική αρχή ή ένα σύνολο αρχών, ήτοι να καθοδηγείται από αυτήν την αρχή ή από αυτές τις αρχές στις αποφάσεις που λαμβάνει περί των πιθανών κατευθύνσεων δράσης που θα μπορούσε να λάβει». Michael Anderson and Susan Leigh Anderson, "Machine Ethics: Creating an Ethical Intelligent Agent," *AI Magazine* 28, no. 4 (2007): 15, https://doi.org/10.1609/aimag.v28i4.2065.

[28] M. Anderson and S. L. Anderson, "Machine Ethics," 16.

νται ηθικά θα καταλήξουν να συμπεριφέρονται ανήθικα, ίσως ευνοώντας τα συμφέροντά τους, μπορεί να οφείλεται σε φόβους που προέρχονται από εύλογες ανησυχίες σχετικά με την ανθρώπινη συμπεριφορά. Οι περισσότεροι άνθρωποι απέχουν από τα ιδανικά μοντέλα ηθικών παραγόντων, παρά το γεγονός ότι έχουν διδαχθεί ηθικές αρχές και τείνουν να ευνοούν τον εαυτό τους».[29] Οι Anderson υπογραμμίζουν ότι η δράση των συστημάτων εξαρτάται από την ανάπτυξή τους.[30]

Εν κατακλείδι, μεταβιβάζοντας τον προβληματισμό για τις υπερ-νοήμονες οντότητες Τ.Ν. στις δικαστικές αποφάσεις προκύπτουν δύο θέσεις. Από τη μία πλευρά, υπάρχει η ανησυχία ότι οι υπερ-νοήμονες οντότητες Τ.Ν. δε θα διακρίνονται ενδεχομένως από τη μέριμνα της Δικαιοσύνης ή ότι θα την αποδίδουν διαφορετικά από τους ανθρώπους και πως με αυτόν τον τρόπο θα είναι πιθανό να μπορούν να μας παραπλανούν σχετικά με τις προθέσεις τους, ακριβώς επειδή θα είναι υπερ-νοήμονες. Από την άλλη πλευρά, έχοντας μηχανές-δικαστές με υπερ-νοημοσύνη θα μπορούμε να δεχθούμε την απόφασή τους, ύστερα από την παραδοχή ότι έχουν θετικούς στόχους και μεριμνούν για το ανθρώπινο καλό, εφόσον θα αδυνατούμε πλέον να ακολουθήσουμε τη δράση τους.

IV. Μηχανές-Δικαστές

Το εφαλτήριο του συλλογισμού μου έγκειται σε μία δίκη ανάλογη με εκείνη του Otto Adolf Eichmann. Ανατρέχοντας στο έργο *Ο Άιχμαν στην Ιερουσαλήμ* της Hannah Arendt ο αναγνώστης μπορεί να εντοπίσει την άποψη πως οποιοσδήποτε άνθρωπος είναι ικανός να προβεί σε εγκλήματα κατά της ανθρωπότητας, αν έχει απογυμνωθεί από τις ηθικές ευθύνες του.

Τον 20ο αιώνα, το ζήτημα της βίας εξακολούθησε να αποτελεί αντικείμενο φιλοσοφικής συζήτησης στο επίπεδο της φαινομενολογικής σκέψης. Στο επίκεντρο είναι ο Γάλλος φιλόσοφος Frantz Fanon με το βιβλίο του με τίτλο *Της γης οι κολασμένοι* που εκδόθηκε στα 1961 και την εισαγωγή του οποίου συνέθεσε ο, επίσης Γάλλος φιλόσοφος, Jean-Paul Sartre. Το βιβλίο συνιστά μία εξύμνηση στη βία και σημείο αναφοράς για τους αντιαποικιοκράτες, διότι ωθεί τον καταπιεσμένο λαό προς την κατεύθυνση πως είναι δυνατό να κατακτήσει την ελευθερία μόνος του μέσω της βίας. Δηλώνει ότι η βία απελευθερώνει, καθώς «ο αποικιοκρατούμενος

[29] Στο ίδιο, 17, δική μου μετάφραση.
[30] Anderson et al., "Towards Moral Machines," 192-193.

ανακαλύπτει την πραγματικότητα και την μετουσιώνει στην κίνηση της δράσης του, στην εξάσκηση της βίας, στο σχέδιο του για απελευθέρωση».[31] Μάλιστα, ο Jean-Paul Sartre σχολιάζει τη σκέψη του Frantz Fanon τασσόμενος υπέρ του και υπερασπιζόμενος τη χειραφέτηση που προκαλεί η επαναστατική βία.

Στη συζήτηση εισέρχεται η Hannah Arendt. Στο *Περί Βίας* ασκεί την κριτική της, συνδιαλεγόμενη με τους δυο Γάλλους φιλοσόφους, ισχυριζόμενη ότι η βία δεν μπορεί να παράγει πολιτικά αποτελέσματα. Πιο συγκεκριμένα, η Γερμανίδα φιλόσοφος σημειώνει «ο Σαρτρ […] προχωρεί […] στην εξύμνηση της βίας […]. […] Η "βία", πιστεύει τώρα, βρίσκοντας έρεισμα στο βιβλίο του Φανόν, "σαν τη λόγχη του Αχιλλέα, μπορεί να γιατρέψει τις πληγές που προκάλεσε". […] Αν δούμε την ιστορία ως μια συνεχή χρονολογική διαδικασία, της οποίας η πρόοδος είναι μάλιστα αναπόφευκτη, η βία με τη μορφή του πολέμου και της επανάστασης μπορεί να φανεί πως συνιστά τη μόνη δυνατή διακοπή. Αν αυτό ήταν αλήθεια, αν μόνο η χρήση της βίας θα επέτρεπε να διακοπούν οι αυτόματες διαδικασίες στο πεδίο των ανθρώπινων πραγμάτων, οι κήρυκες της βίας θα είχαν κερδίσει σε ένα σημαντικό ζήτημα».[32]

Επιπλέον, η Hannah Arendt ισχυρίζεται ότι «η βία δεν είναι ούτε ζωώδης ούτε ανορθολογική».[33] Αρχικά, φρονεί πως κατά τον τρόπο που η απόδοση ανθρωπομορφικής συμπεριφοράς στα ζώα δεν είναι επιτυχημένη, ο εντοπισμός ζωωδών στοιχείων στον άνθρωπο δεν μπορεί να λειτουργήσει δικαιολογώντας ή καταδικάζοντας την ανθρώπινη συμπεριφορά,[34] αφού, μάλιστα, «η κατασκευή εργαλείων συνιστά μια εξαιρετικά περίπλοκη διανοητική δραστηριότητα».[35] Έπειτα, θεωρεί ότι η αντίδραση της οργής δε συνεπάγεται την ανορθολογική όψη της βίας, γιατί «μόνο εκεί όπου εύλογα υποπτευόμαστε ότι οι συνθήκες θα μπορούσαν να αλλάξουν μα δεν αλλάζουν εμφανίζεται η οργή. Μόνο όταν μας θίγουν το αίσθημα δικαίου αντιδρούμε με οργή […]».[36] Τέλος, η φιλόσοφος δέχεται μόνο σε μία περίπτωση να χαρακτηριστεί η βία ανορθολογική και αυτή είναι όταν

[31] Frantz Fanon, *Της γης οι κολασμένοι,* μτφρ. Αγγέλα Αρτέμη (Εκδόσεις Κάλβος, 1982), 33.
[32] Hannah Arendt, *Περί Βίας*, μτφρ. Βάνα Νικολαΐδου-Κυριανίδου (Εκδόσεις Αλεξάνδρεια, 2000), 75-93.
[33] Στο ίδιο, 122.
[34] Στο ίδιο, 119-120.
[35] Στο ίδιο, 122.
[36] Στο ίδιο, 123.

κινείται ενάντια σε υποκατάστατα εξαιτίας ψυχολογικών κινήτρων.[37]

Υπάρχει, λοιπόν, ηθική δικαιολόγηση για την πολιτική βία; Η Hannah Arendt είναι φιλόσοφος που ενδιαφέρεται για τον βίο, ήτοι για τη ζωή του όντος που έχει προσωπικότητα, που «γράφει ιστορία» με τα λεγόμενα και τις πράξεις του. Πιο συγκεκριμένα, αποσαφηνίζει στο έργο τής *Ανθρώπινη Κατάσταση* μέσω της φιλοσοφικής ανθρωπολογίας της πως ενδιαφέρεται για τον homo politicus, για εκείνον που πράττει στον δημόσιο χώρο, στον χώρο της ελευθερίας.[38] Μάλιστα, σημαντική θέση στη σκέψη της κατέχει η Αριστοτελική ηθική και πολιτική φιλοσοφία. Επομένως, για να δοθεί απάντηση στο ερώτημα, υπό την οπτική της φιλοσόφου, έχω τη γνώμη ότι θα πρέπει να εξεταστεί και να εφαρμοστεί η πρότασή της στη βάση της παραπάνω παραδοχής.

Η φιλόσοφος σε συζήτηση στα 1967 με θέμα «Η νομιμότητα της βίας ως πολιτική πράξη;» υπήρξε η μόνη που δεν ερμήνευσε το συγκεκριμένο ζήτημα ως ηθικό, αλλά ως πολιτικό. Αναλυτικότερα, εξήγησε πως στον πυρήνα της ηθικής φιλοσοφίας συναντάμε την πράξη του εαυτού μας κι όχι του κόσμου· «σε όλα τα ηθικά ζητήματα, μας απασχολεί ο εαυτός μας. Αναρωτιόμαστε αν είμαστε ένοχοι για κάτι, αν μπορούμε να ζήσουμε με τον εαυτό μας αφού έχουμε κάνει αυτό ή εκείνο. Αυτά είναι απολύτως θεμιτά και πολύ σημαντικά ερωτήματα, αλλά δεν είναι θεμελιωδώς πολιτικά. Στην πολιτική, ασχολούμαστε με τον κόσμο και όχι με τον εαυτό μας».[39] Σύμφωνα με αυτά, η φιλόσοφος επιχειρεί να προβάλλει την ηθική πολιτική όχι στο πεδίο του εαυτού αλλά στη δημόσια σφαίρα, ως πολίτη. Συμπερασματικά, και βάσει της επιρροής που δέχθηκε από τον Αριστοτέλη, φαίνεται ότι για τη Hannah Arendt η βία δεν έχει θέση στα πολιτικά ζητήματα και για αυτό δεν τίθεται θέμα ηθικής δικαιολόγησης. Η χρήση βίας, για τη φιλόσοφο, συνεπάγεται την έλλειψη ηθικής.

Επιστρέφοντας, όμως, στον θεματικό πυρήνα της παρούσας ενότητας, η αναφορά στη Hannah Arendt στοχεύει στην ανάδειξη του ζητήματος η υπερ-νοήμων δρώσα οντότητα T.N. να ακολουθεί το μοτίβο του Otto Adolf Eichmann, ήτοι να αναγνωρίζει και να επικροτεί την τήρηση των καθηκόντων, αλλά να μην περιλαμβάνει ουδεμία ηθική οπτική στη δράση της. Η φιλόσοφος έχοντας παρακολουθήσει τη δίκη ως δημοσιογράφος διατείνεται ότι διαπίστωσε πως ο Otto Adolf Eichmann ενεργούσε βάσει

[37] Στο ίδιο, 124-127.

[38] Hannah Arendt, *The Human Condition* (The University of Chicago Press, 1998), 22-28.

[39] Noam Chomsky, Hannah Arendt and Susan Sontag, "The Legitimacy of Violence as a Political Act," in *Dissent, Power and Confrontation*, ed. Alexander Klein (McGraw-Hill, 1971), 116, δική μου μετάφραση.

καθήκοντος χωρίς να εμπλέκει την ηθική κρίση στις ενέργειές του.[40]

Δυνάμει της εν λόγω παρατήρησης, θα μπορούσε ο/η δικαστής, όπως ίσως και η μηχανή-δικαστής, να κρίνει ότι ο Otto Adolf Eichmann, πράγματι, λειτουργούσε αρμονικά μέσα σε ένα ηθικό πλαίσιο, για μία ομάδα ή βάσει μίας (μαθηματικής) λογικής και πάλι για μία συγκεκριμένη ομάδα, στην περίπτωση που η προαναφερθείσα θέση είναι αποδεκτή από το κοινωνικό σύνολο; Θα ήταν σε θέση η Τ.Ν., ακόμα κι αν ήταν υπερ-νοήμων, να λαμβάνει τέτοιου είδους αποφάσεις ούσα αυτόνομη; Δίχως να διατηρείται, δηλαδή, ο σημερινός επικουρικός ρόλος της; Αν μια μηχανή-δικαστής είναι ένα αλγοριθμικό σύστημα που απλώς εντοπίζει και δικαιώνει την τήρηση του καθήκοντος, χωρίς να πραγματεύεται τις επιπτώσεις της εν λόγω τήρησης στην εκάστοτε συνθήκη, ήτοι να μην προβαίνει στην ηθική εξέταση της περίστασης, είναι πολύ πιθανό να ελλοχεύει ο κίνδυνος για νέα εγκλήματα από την Τ.Ν., αν, μάλιστα, ληφθεί υπόψη και η στάση των Max Tegmark, Eliezer Yudkowsky και Nick Bostrom.

Ο αντι-παραγωγικός χαρακτήρας της βίας, σύμφωνα με τη Hannah Arendt, μπορεί να παραλληλιστεί με ένα «α-ηθικό» σύστημα λήψης αποφάσεων. Συγκεκριμένα, όπως η βία φέρεται να μην είναι ικανή να προσφέρει πολιτικά αποτελέσματα, η μηχανή-δικαστής ελλείψει ηθικών αρχών συνεπάγεται την αδυναμία επιβίωσης στην πόλη. Το αποδεκτό ή το κατεστημένο δεν ορίζει την ηθική, εν αντιθέσει η ηθική, η αυτοσυνείδηση της κοινωνίας, κρίνει και αμφισβητεί με επιχειρηματολογία βάσει κριτηρίων, ήτοι με λογοδοσία, το αποδεκτό ή το κατεστημένο.[41] Αν οι υπερ-νοήμονες μηχανές-δικαστές διαποτιστούν με ηθική, ο αριστοτελικός θεσμός της ηθικής πόλης δύναται να πραγματωθεί· οι πολίτες θα μπορούν, με την επανάληψη της καλής πράξης, να αποκτούν την ηθική αρετή – την ηθική του ορθού – και να γίνονται ενάρετοι αλληλεπιδρώντας με τους κοινωνικούς άλλους.[42] Επομένως, θα επιτευχθεί τοιουτοτρόπως η ηθική πόλη και θα

[40] Roger Berkowitz, "The Power of Non-Reconciliation – Arendt's Judgement of Adolf Eichmann," *HannahArendt.Net* 6, no. 1/2 (2012), https://doi.org/10.57773/hanet.v6i1/2.11. Δίχως, όμως, να παρουσιάζει αντίρρηση για την ενοχή του. Στο ίδιο.

[41] Σταυρούλα Τσινόρεμα, «Ηθικές Αρχές» (διάλεξη στο μάθημα Θεωρητική Ηθική ΙΙ του Προγράμματος Μεταπτυχιακών Σπουδών «Φιλοσοφία» - Κατεύθυνση: «Εφαρμοσμένη Ηθική», Φιλοσοφική Σχολή, Ε.Κ.Π.Α., Αθήνα, Μάρτιος 2022).

[42] Για την ιδέα ότι η Τ.Ν. μπορεί να λειτουργήσει ως ενάρετο σύστημα και να σμιλεύσει ενάρετους χαρακτήρες ανθρώπων-χρηστών, βλ.: Alkis Gounaris, George Kosteletos and Maria-Artemis Kolliniati, "Virtue in the machine: beyond a one-size-fits-all approach and Aristotelian ethics for Artificial Intelligence," Conatus – Journal of Philosophy 10, no.1 (2025): 127-152. https://doi.org/10.12681/cjp.40628

εξαλειφθεί η ανάγκη καταφυγής σε δικαστηριακές διαδικασίες, διότι θα έχει επέλθει η ύψιστη ηθική αρετή, η φιλία, σύμφωνα με τον Αριστοτέλη.[43]

Ωστόσο, αν οι υπερ-νοήμονες μηχανές-δικαστές διαθέτουν χαρακτηριστικά ανθρώπινης συμπεριφοράς, η επίτευξη του θεσμού της ηθικής πόλης δεν είναι εγγυημένη. Ούσα μέλος μίας κοινωνίας με ανθρώπους-δικαστές εντοπίζω την περίπτωση κάποιος/α δικαστής να είναι σε θέση να διακρίνει τί είναι ηθικό, αλλά να το αγνοεί επιδιώκοντας να ικανοποιήσει δικά του/της συμφέροντα και δημιουργώντας κατ' επέκταση ανάλογα αποτελέσματα. Ως εκ τούτου, λαμβάνοντας υπόψη αυτήν την παράμετρο, μία μηχανή-δικαστής, ακόμα κι αν είχε κρίση για να εντοπίσει το ηθικό, θα μπορούσε να παραβλέψει σκοπίμως την ηθική δράση.

V. Συμπεράσματα

Στην περίπτωση των μηχανών-δικαστών είναι αναγκαία η πρόσβαση στην εσωτερική λειτουργία των συστημάτων. Το αίτημα για διαφανή Τ.Ν. χρειάζεται να εξεταστεί και να γίνει αποδεκτό ακόμα και για τον σημερινό, επικουρικό, ρόλο των συστημάτων στη λήψη δικαστικών αποφάσεων. Η υπόθεση των υπερ-νοημόνων συστημάτων Τ.Ν., ίσως, να μην απασχολεί έντονα τη δεδομένη στιγμή το δικαστικό πεδίο, αλλά φρονώ, δεδομένης της χρήσης συστημάτων όπως το COMPAS, ότι σε τέτοια ρυθμιστικά πλαίσια, όπου διακυβεύεται το μέλλον κάποιων βάσει μίας απόφασης, η εξάλειψη της μεροληψίας από τα συστήματα είναι φλέγουσα και η γνωστοποίηση του αλγορίθμου μπορεί να συνδράμει στην αντιμετώπισή της.

Επίσης, οι νομικές και ηθικές αξίες που θα χορηγηθούν στις μηχανές-δικαστές και το ζήτημα του ελέγχου διαδραματίζουν σημαντικό ρόλο. Οι επιφυλακτικοί ερευνητές της Τ.Ν. μπορεί να διερωτώνται για το αν οι μηχανές-δικαστές θα ενδιαφέρονται για την απόδοση Δικαιοσύνης ή για το αν θα αντιλαμβάνονται τη Δικαιοσύνη όπως οι άνθρωποι. Ωστόσο, οι οπτιμιστές ερευνητές πιθανώς να παρατηρούν ότι τα υπερ-νοήμονα συστήματα απαλλαγμένα, ίσως, από την προκατάληψη της Τ.Ν., χάρη στην ικανότητά τους να διακρίνουν την προβληματική της μεροληψίας, θα αναλάβουν την προστασία μας και θα κρίνουν τις δικαστικές υποθέσεις ορθώς θέτοντας θετικούς στόχους.

Εναρμονίζοντας όλα τα παραπάνω, η ανάπτυξη μηχανών-δικαστών και η δράση τους απαιτεί την ύπαρξη ηθικής. Το σενάριο της ηθικής πόλης

[43] Τσινόρεμα, «Ηθικές Αρχές».

φαντάζει ουτοπικό, εντούτοις θεωρώ ότι στο ενδεχόμενο της υπερ-νοημοσύνης η ανθρώπινη φαρέτρα χρειάζεται να είναι γεμάτη με τρόπους δημιουργίας της καλύτερης δυνατής εκδοχής ενός τέτοιου συστήματος, ώστε να μην κινδυνέψουμε από εγκλήματα εκπορευόμενα από την Τ.Ν. Φαίνεται, λοιπόν, να μην είναι θεμιτό η Τ.Ν. να έχει αυτόνομο λόγο τουλάχιστον σε «ευαίσθητους» κλάδους της ζωής μας – όπως είναι ο δικαστικός κλάδος – δεδομένου ότι παραμονεύει ο κίνδυνος μιας υπερ-νοήμονος Τ.Ν. που ως τέτοια θα είναι ανεξέλεγκτη. Ας είναι σκοπός μας η βέλτιστη εκδοχή των τρεχόντων συστημάτων Τ.Ν., για την πιθανότητα εμφάνισης υπερ-νοημόνων συστημάτων, διότι «ο πολιτισμός μας πιστώνεται στη νοημοσύνη μας».

Αναφορές

Anderson, M., Anderson, S. L., Gounaris, A., & Kosteletos, G. (2021). Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson. *Conatus - Journal of Philosophy*, *6*(1), 177–202. https://doi.org/10.12681/cjp.26832

Anderson, Michael and Susan Leigh Anderson. "Machine Ethics: Creating an Ethical Intelligent Agent." *AI Magazine* 28, no. 4 (2007): 15-26. https://doi.org/10.1609/aimag.v28i4.2065.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks." *ProPublica,* May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Arendt, Hannah. *The Human Condition.* The University of Chicago Press, 1998.

Arendt, Hannah. *Περί Βίας*. Μεταφρασμένη από τη Βάνα Νικολαΐδου-Κυριανίδου. Εκδόσεις Αλεξάνδρεια, 2000.

Beijing Internet Court. "Beijing Internet Court launches online litigation service center." Last modified July 1, 2019. https://english.bjinternetcourt.gov.cn/2019-07/01/c_190.htm.

Berkowitz, Roger. "The Power of Non-Reconciliation – Arendt's Judgement of Adolf Eichmann." *HannahArendt.Net* 6, no. 1/2 (2012). https://doi.org/10.57773/hanet.v6i1/2.11.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

Γούναρης, Άλκης και Γιώργος Κωστελέτος. «Γράφοντας τον αλγόριθμο του Καλού: Η Τεχνητή Νοημοσύνη ως μηχανή απόδοσης Δικαιοσύνης». *Ηθική. Περιοδικό φιλοσοφίας* 19 (2024): 5-27. https://doi.org/10.12681/ethiki.39654

Chomsky, Noam, Hannah Arendt and Susan Sontag. "The Legitimacy of Violence as a Political Act." In *Dissent, Power and Confrontation*, edited by Alexander Klein. McGraw-Hill, 1971.

Fanon, Frantz. *Της γης οι κολασμένοι*. Μεταφρασμένο από την Αγγέλα Αρτέμη. Εκδόσεις Κάλβος, 1982.

Gounaris, A., Kosteletos, G., & Kolliniati, M.-A. (2025). Virtue in the Machine: Beyond a One-size-fits-all Approach and Aristotelian Ethics for Artificial Intelligence. *Conatus - Journal of Philosophy*, *10*(1), 127–152. https://doi.org/10.12681/cjp.40628

Israni, Ellora Thadaney. "When an Algorithm Helps Send You to Prison." *The New York Times*, October 26, 2017. https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html.

Jackson, Eugenie and Christina Mendoza. "Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not." *Harvard Data Science Review* 2, no. 1 (2020). https://doi.org/10.1162/99608f92.1b3dadaa.

Jackson, Peter. *Introduction to Expert Systems*. Addison-Wesley, 1999.

Müller, Vincent C. "Ethics of Artificial Intelligence and Robotics." *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/entries/ethics-ai/#BiasDeciSyst.

Russel, Stuart. *Συμβατή με τον άνθρωπο; η Τεχνητή Νοημοσύνη και το πρόβλημα του ελέγχου*. Μεταφρασμένο από τον Νίκο Αποστολόπουλου. Εκδοτικός Οίκος ΤΡΑΥΛΟΣ, 2021.

Τσινόρεμα, Σταυρούλα. «Ηθικές Αρχές». Διάλεξη στο μάθημα Θεωρητική Ηθική ΙΙ του Προγράμματος Μεταπτυχιακών Σπουδών «Φιλοσοφία» - Κατεύθυνση: «Εφαρμοσμένη Ηθική», Φιλοσοφική Σχολή, Ε.Κ.Π.Α., Αθήνα, Μάρτιος 2022.

Taylor, Alexandra Mac. "AI Prediction Tools Claim to Alleviate an Overcrowded American Justice System… But should they be Used?" *Stanford Politics,* September 13, 2020. https://stanfordpolitics.org/2020/09/13/ai-prediction-tools-claim-to-alleviate-an-overcrowded-american-justice-system-but-should-they-be-used/.

Tegmark, Max. *LIFE 3.0 Τι θα σήμαινε να είσαι άνθρωπος στην εποχή της τεχνητής νοημοσύνης;* Μεταφρασμένο από τον Νίκο Αποστολόπουλο. Εκδοτικός Οίκος ΤΡΑΥΛΟΣ, 2018.

Περίληψη

Ο ορμητικά καλπάζων ρυθμός προόδου στον τεχνολογικό πολιτισμό αναδιαμορφώνει τους περισσότερους τομείς της κοινωνίας. Είναι πρόδηλο πως η εξάρτηση, που προκύπτει, από την τεχνολογία ορίζει τη ζωή του σύγχρονου ανθρώπου εγείροντας συχνά εύλογες ανησυχίες για τα καινά γεγονότα. Η Τεχνητή Νοημοσύνη (Τ.Ν.) αποτελεί, ίσως, την πιο αιχμηρή έκφανση της τεχνολογίας έχοντας διεισδύσει σε ποικίλα πεδία της ανθρώπινης δραστηριότητας. Το παρόν δοκίμιο πραγματεύεται τη διαδικασία λήψης δικαστικών αποφάσεων ως ένα αξιοσημείωτο επιμέρους πεδίο εφαρμογής της Τ.Ν. Αρχικά, στο δοκίμιο γίνεται λόγος για τον τεχνολογικό πολιτισμό και τις επιδράσεις του. Εν συνεχεία, παρουσιάζεται η διεκπεραίωση δικαστικών αποφάσεων με Τ.Ν. τη δεδομένη χρονική στιγμή και προσεγγίζεται ο ηθικοφιλοσοφικός προβληματισμός για τις ηθικώς δρώσες τεχνητές οντότητες.

Λέξεις-κλειδιά: ηθική, Τεχνητή Νοημοσύνη, Ηθική των Μηχανών, ηθικώς δρώσα τεχνητή οντότητα, δικαστικές αποφάσεις

Keywords: ethics, Artificial Intelligence, Machine Ethics, artificial moral entities, judicial decision-making

Ιωάννα Μαλανδράκη
Τμήμα Φιλοσοφίας, ΕΚΠΑ
Ηλεκτρονική Διεύθυνση: iomalan@philosophy.uoa.gr
ORCID iD: https://orcid.org/0000-0001-9942-7124