

Κοιτάζοντας τὴ Σφίγγα στὰ μάτια: Τεχνητὴ νοημοσύνη καὶ ἀνθρώπινες ἀξίες

Θεοφάνη Τάση*

*ὦ δύσποτμ', εἴθε μήποτε γνοίης ὅς εἴ
[Σοφοκλ. Οἰδίπους Τύραννος, στ. 1068].*

Βρισκόμαστε σὲ ἓνα σταυροδρόμι. Ἡ ταχύτατη πλέον ἐξέλιξη τῆς τεχνητῆς νοημοσύνης ἐπιφέρει μία ριζικὴ ἀλλαγὴ στὴν ἀνθρώπινη ἱστορία, γιὰ τὴν ὁποία ὁ διευθύνων σύμβουλος τῆς Google, ὁ Sundar Pichai, δῆλωσε πὼς θὰ ἀποδειχθεῖ σημαντικώτερη ἀπὸ τὴν ἀνακάλυψη τῆς φωτιᾶς, ἐνῶ ὁ διευθύνων σύμβουλος τῆς OpenAI, ὁ Sam Altman, ἐξομολογήθηκε πὼς θὰ ἦταν τρελὸ νὰ μὴ φοβόμαστε τὶς συνέπειές της¹. Ὡς ἐκ τούτου, εἶναι ἀπολύτως ἀναγκαία μία δημόσια διαβούλευση ὡς πρὸς τὸ ἂν τὰ ὀφέλη ἀπὸ τὴν παροῦσα ἀνάπτυξη τῆς τεχνητῆς νοημοσύνης ὑπερβαίνουν τόσο τοὺς ἄμεσους ὅσο ἐπίσης καὶ τοὺς μακροπρόθεσμους κινδύνους. Ἡ τελικὴ ἀπόφαση γιὰ τὴν κατεύθυνση ποὺ θὰ ἐπιλέξουμε ὡς κοινωνία ὀφείλει νὰ εἶναι πολιτικὴ καὶ δὲν πρέπει νὰ ληφθεῖ ἀπὸ τὶς ἴδιες τὶς ἐταιρεῖες πληροφορικῆς, ἰδίως ὅταν αὐτὲς ἐπιθυμοῦν διακαῶς τὴ νομοθετικὴ ρύθμιση στὴν ἀνάπτυξη τῆς τεχνητῆς νοημοσύνης. Ἐν τούτοις τόσο ἡ Microsoft, ἡ ὁποία χρηματοδοτεῖ τὴν OpenAI, ὅσο καὶ ἡ Google ἐνσωμάτωσαν προσφάτως τὴν τεχνητὴ νοημοσύνη ποὺ ἀνέπτυξαν στὶς μηχανὲς ἀναζήτησής τους καθιστώντας τὶς προσβάσιμες σὲ ὅλους

* Ὁ Θεοφάνης Τάσης εἶναι Λέκτωρ Σύγχρονης Πρακτικῆς Φιλοσοφίας στὸ Alpen Adria Universität.

1. Βλ. <https://fortune.com/2023/04/17/sundar-pichai-a-i-more-profound-than-fire-electricity/> καὶ <https://www.businessinsider.com/openai-ceo-sam-altman-comments-ai-fears-risks-artificial-intelligence-2023-3> [30.9.2023].

παρά τις ένστάσεις και προειδοποιήσεις πολλών συνεργατών τους, προγραμματιστών και φιλοσόφων για τους πιθανούς κινδύνους².

Ἡ πρωτοφανής ἐπιτυχία τοῦ ChatGPT τῆς Open AI καὶ ὁ φόβος τῆς Google γιὰ τὴν τυχὸν ἀπώλεια τῆς πρωτοκαθεδρίας τῆς ἐγκαινίασαν ἕναν ἀγῶνα ταχύτητας μὲ σκοπὸ τὴν ἐπίτευξη μιᾶς ἰσχυρῆς τεχνητῆς νοημοσύνης, μιᾶς νοημοσύνης δηλαδὴ ἰσάξιας τῆς ἀνθρώπινης, ἱκανῆς νὰ δρᾷ ἐξ ἴσου ἢ περισσότερο ἐπιτυχημένα. Σὲ αὐτὸν τὸν ἀδυσώπητο ἀγῶνα ἡ ἀσφαλῆς χρῆση τῆς τεχνητῆς νοημοσύνης παραμένει ἕνα καίριο ζητούμενο, καθὼς οἱ ἀναγκαῖοι μηχανισμοὶ ἐλέγχου ὑπολείπονται τῶν ἱκανοτήτων τῆς καὶ –ἐπιπλέον– φαίνεται νὰ θυσιάζονται στὸν βωμὸ τοῦ οἰκονομικοῦ κέρδους. Ὡστόσο, ἡ διαρκὴς βελτίωση τῆς τεχνητῆς νοημοσύνης, δίχως τὴν προηγούμενη ἢ ἔστω τὴν παράλληλη δημιουργία ἀποτελεσματικῶν μηχανισμῶν ἐλέγχου, μοιάζει μὲ τὴν ἐπιθυμία νὰ κατασκευάσουμε ἕναν πυρηνικὸ ἀντιδραστήρα δίχως νὰ γνωρίζουμε πῶς νὰ ἐλέγξουμε τὴν ἀλυσιδωτὴ ἀντίδραση μὲ τὴν ἐλπίδα πῶς θὰ τὰ καταφέρουμε ἀφοῦ τὸν φτιάξουμε. Ἀλλὰ ἡ δημιουργία ἀποτελεσματικῶν μηχανισμῶν ἐλέγχου εἶναι ἕνα προκλητικὰ δύσκολο ἐγχείρημα ἐπειδὴ –μεταξὺ ἄλλων– ἡ ἐσωτερικὴ λειτουργία τῆς τεχνητῆς νοημοσύνης παραμένει ἐν πολλοῖς ἀδιαφανὴς στοὺς ἴδιους τοὺς δημιουργοὺς τῆς.

Ὡς γνωστόν, προγνωστικὰ γλωσσικὰ μοντέλα ἀλγορίθμων, ὅπως τὸ GPT-4 τῆς OpenAI ἢ τὸ Bard τῆς Google, εἶναι νευρωνικὰ δίκτυα ποὺ παράγουν κείμενα χρησιμοποιώντας τὴν στατιστικὴ ἀναλύοντας δηλαδὴ ἕναν τεράστιο ὄγκο δεδομένων καὶ σχεδιάζοντας ἕνα εἶδος μαθηματικοῦ χάρτη τῆς γλώσσας στὸν ὁποῖο ἐφαρμόζουν συναρτήσεις πιθανότητας γιὰ νὰ προβλέψουν τὸν κατάλληλο συνδυασμὸ λέξεων. Παρὰ τὸ ὅτι δὲν κατανοοῦν –συνεπῶς οὔτε νοηματοδοτοῦν– τοὺς παραγόμενους συνδυασμούς, εἶναι ἀρκετὰ εὐφυῆ γιὰ νὰ ἀνακαλύψουν δυσδιάκριτους συσχετισμοὺς μεταξὺ τῶν συμβόλων. Φερ' εἰπεῖν, ἡ λέξη *ἄνθρωπος* εἶναι γιὰ τὸ GPT-4 ἕνα σύμβολο ποὺ ὅμως ἔχει μιὰ σχέση μὲ τὸν πραγματικὸ ἄνθρωπο, ἐπειδὴ τὰ δεδομένα γιὰ τὸν ὅρισμὸ τοῦ συμβόλου προέρχονται ἀπὸ τὸν πραγματικὸ κόσμο. Ἔτσι γιὰ παράδειγμα, ὁ συνδυασμὸς τῶν συμβόλων *ἄνθρωπος καὶ λευκὸς* μπορεῖ νὰ ὁδηγήσει στὸ *δυτικός*, ἕνα

2. Βλ. <https://www.bloomberg.com/news/features/2023-04-19/google-bard-ai-chatbot-raises-ethical-concerns-from-employees?srnd=premium&leadSource=uverify%20wall> [30.9.2023].

ἀποτέλεσμα δηλαδή πού ὑπερβαίνει τὰ ἀρχικά σύμβολα. Ἐδῶ ἔγκειται ἡ ἀνωτερότητα τῆς σημερινῆς τεχνητῆς νοημοσύνης συγκριτικά μὲ τὶς προηγούμενες μορφές της. Ὁ ἄνθρωπος φυσικά ἐξακολουθεῖ νὰ ὑπερέχει αὐτῆς, καθὼς εἶναι σὲ θέση νὰ νοηματοδοτεῖ τὶς λέξεις· ὅμως ἡ τεχνητὴ νοημοσύνη, παρὰ τὸ ὅτι δὲν διαθέτει κατανόηση, πλέον παράγει κείμενα πού δὲν διακρίνονται ἀπὸ τὰ ἀνθρώπινα. Ἔτσι μπορούμε νὰ διακηρύξουμε ἐκ νέου τὸν θάνατο τοῦ συγγραφέα. Παράλληλα ὅμως ὀρισμένοι συγγραφεῖς ἔχουν ἀρχίσει νὰ γράφουν μὲ τὴ βοήθεια τῆς τεχνητῆς νοημοσύνης, ὅπως ἀντίστοιχα μουσικοί, σκηνοθέτες, φωτογράφοι καὶ εἰκαστικοὶ συνθέτουν, σκηνοθετοῦν καὶ δημιουργοῦν ἀντίστοιχα εἰκόνες, χρησιμοποιώντας προγράμματα ὅπως τὸ Runway, τὸ Synthesia, τὸ Amper Music ἢ τὸ DALL-E 2. Μελλοντικῶς μπορούμε νὰ φανταστοῦμε βιβλία, ἀλλὰ ἐπίσης καὶ μουσικοὺς δίσκους ἢ ταινίες, πού θὰ παράγονται ἀποκλειστικά ἀπὸ τὴν τεχνητὴ νοημοσύνη κατὰ παραγγελία, σύμφωνα μὲ τὶς προσωπικὲς ἀναγνωστικές, μουσικὲς, εἰκαστικές, κινηματογραφικὲς προτιμήσεις καὶ συνήθειες τῶν ἀνθρώπων. Πιθανῶς θὰ ἐξακολουθοῦν νὰ ὑπάρχουν ἀναγνώστες, ἀκροατὲς καὶ θεατὲς, οἱ ὅποιοι θὰ ἐκτιμοῦν τὰ ἀνθρώπινα ἔργα ὅπως ὑπάρχουν σήμερα, καταναλωτὲς οἱ ὅποιοι προτιμοῦν ἓνα χειροποίητο κουστόμυ ἀπὸ ἓνα ἔτοιμο, βιομηχανικὰ τυποποιημένο, ἀλλὰ οἱ περισσότεροὶ μάλλον θὰ στραφοῦν λόγῳ εὐκολίας, ὅπως ἐπίσης γιὰ λόγους οἰκονομίας ἢ διαφορετικῆς αἰσθητικῆς παιδείας, σὲ προϊόντα τεχνητῆς νοημοσύνης.

Ἐδῶ ἀξίζει νὰ ἐπισημανθεῖ πῶς, ἐνῶ ἡ βιομηχανικὴ ἐπανάσταση ἀξιοποίησε τὴν τεχνικὴ ἀνάπτυξη πού ἐπέφερε ἡ ἐπιστημονικὴ πρόοδος γιὰ τὴ μαζικὴ καὶ ταχύτατη παραγωγὴ ὑλικῶν ἀγαθῶν μέσω τῆς ἐκμηχάνισης τῆς παραγωγῆς, τὴν ὁποία θὰ διαδεχθεῖ στὸν εἰκοστὸ αἰῶνα ἡ μαζικὴ παροχὴ ὑπηρεσιῶν ἀρχικά μέσω τῆς χρήσης ἡλεκτρονικῶν ὑπολογιστῶν καὶ ἐν συνεχείᾳ μέσω τοῦ διαδικτύου, ἡ τεχνητὴ νοημοσύνη ἐπιτρέπει τὴν ταχύτατη μαζικὴ, ἀλλὰ συγχρόνως ἐξατομικευμένη παραγωγὴ πνευματικῶν ἀγαθῶν. Ἔτσι, ἐνῶ ἡ βιομηχανικὴ ἐπανάσταση ὑποβάθμισε περαιτέρω τὴν ἀξία τῆς χειρωνακτικῆς ἐργασίας ἔναντι ἐκείνης τῆς διανοητικῆς ἀναβαθμίζοντας τὸ γόητρο τῆς ἀκαδημαϊκῆς τάξης, δηλαδή τῶν πνευματικῶν ἐργαζόμενων, ἔναντι τῆς ἐργατικῆς τάξης, ἡ τρέχουσα ἐπανάσταση τῆς τεχνητῆς νοημοσύνης ὑποβαθμίζει τὴν ἀξία τῆς διανοητικῆς ἐργασίας μειώνοντας τὸ κῦρος τῶν ἀκαδημαϊκῶν κοινωνικῶν

στρωμάτων. Δεν πλήττεται όμως μόνον ο ναρκισσισμός της ακαδημαϊκής τάξης, τὰ μέλη της οποίας αντιλαμβάνονται τις σπουδές κυρίως ως επένδυση για τὴ διατήρηση τοῦ τρόπου ζωῆς τους. Κλονίζεται επίσης τὸ αἶσθημα ὑπεροχῆς τοῦ ἀνθρώπου ἔναντι τῶν ὑπολοίπων ζώων, τὰ ὁποῖα διαθέτουν συγκριτικὰ μὲ αὐτὸν μικρότερου εὗρους νοημοσύνης, καθὼς τὸ φάσμα τῆς νοημοσύνης διευρύνεται πλέον συμπεριλαμβάνοντας τὴν τεχνητή. Ἄραγε, εἶναι φρόνιμο νὰ δημιουργήσουμε μία νοημοσύνη ἐφάμιλλη ἢ ἀκόμη καὶ ἀνώτερη τῆς δικῆς μας, ὄντας ἀβέβαιοι γιὰ τὸ πῶς θὰ συνυπάρχουμε στὸν πλανήτη;

Ἦδη σήμερα τὰ προγνωστικὰ γλωσσικὰ μοντέλα ἀλγορίθμων (Large Language Models) μποροῦν νὰ παραπληροφορήσουν, νὰ διασπείρουν ψευδεῖς εἰδήσεις καὶ νὰ διακινήσουν θεωρίες συνωμοσίας διαβρώνοντας τὴ δημόσια σφαῖρα, μὲ συνέπεια τὴν ἀποδυνάμωση τῆς φιλελεύθερης δημοκρατίας. Μποροῦν επίσης νὰ ἐνισχύσουν αὐταρχικὰ ἢ ὀλοκληρωτικὰ καθεστῶτα, ὅπως ἀκόμη νὰ χρησιμοποιηθοῦν στὸν σχεδιασμό ἐγκληματικῶν καὶ τρομοκρατικῶν ἐνεργειῶν. Ἐκτὸς αὐτοῦ, ἡ τεχνητὴ νοημοσύνη εἶναι εὐάλωτη σὲ κυβερνοεπιθέσεις, οἱ ὁποῖες μποροῦν νὰ ὀδηγήσουν σὲ ἕναν κακόβουλο ἐπαναπρογραμματισμὸ της καὶ στὴν κλοπὴ τῶν ἰδιωτικῶν δεδομένων στὰ ὁποῖα εἶχε πρόσβαση. Μπορεῖ νὰ χειραγωγῇ συναισθηματικὰ τοὺς χρήστες, νὰ προξενῇ ἐξάρτηση, ἀλλὰ επίσης νὰ ἐνδυναμώσει τὴ μισαλλοδοξία καὶ τὸν ψηφιακὸ ἐκφοβισμό στὰ κοινωνικὰ μέσα δικτύωσης. Ἐπὶ πλέον, ἡ διάδοση καὶ γενικευμένη χρῆση τῆς τεχνητῆς νοημοσύνης ἀπασχολεῖ ἑκατομμύρια χαμηλὰ ἀμειβόμενους ἐργαζομένους σὲ συνθήκες ἐπισφάλειας γιὰ τὴν ἐκπαίδευσή της (Reinforcement Learning from Human Feedback), ἐνῶ ἐνδέχεται νὰ ὀδηγήσει σὲ μία μείωση θέσεων ἐργασίας, ἡ ὁποία ἴσως δὲν θὰ ἰσοσταθμιστεῖ ἀπὸ τὴ δημιουργία νέων, ἰδίως ὅταν οἱ ἐπιχειρήσεις ἐπιδιώκουν τὸ κέρδος ὡς ὑπέρτατο στόχο. Παράλληλα ὑφίσταται σὲ πολλοὺς βιομηχανικοὺς τομεῖς ἕνα ἔλλειμα ἐξειδικευμένων ἐργαζομένων, τὸ ὁποῖο ὅμως ἐνδέχεται νὰ καλυφθεῖ ἀπὸ ρομπὸτ μὲ λογισμικὸ τεχνητῆς νοημοσύνης. Εἰρήσθω ἐν παρόδῳ ὅτι ἀρκετοὶ ἐπιστήμονες πιστεύουν πῶς μόνον μία ἐνσωματωμένη (embodied) τεχνητὴ νοημοσύνη μπορεῖ νὰ εἶναι πραγματικὰ εὐφυής³. Ἐπί-

3. Ἐργαζόμενη στὴν κατεύθυνση ἐνσωμάτωσης τῆς τεχνητῆς νοημοσύνης σὲ ρομπὸτ ἡ Alphabet, θυγατρικὴ τῆς Google, ἀνακοίνωσε τὴ δημιουργία τοῦ PALM-E, ἐνὸς ρομπὸτ

σης δὲν μπορεῖ νὰ ἀποκλεισθεῖ τὸ ἐνδεχόμενο ἢ τεχνητὴ νοημοσύνη νὰ διευρύνει ἀρχικὰ τὶς οἰκονομικὲς καὶ κοινωνικὲς ἀνισότητες ἕως ὅτου γίνεи εὐρέως προσβάσιμη, καθιστώντας ἴσως ἐφικτὴ μία καθολικὴ ἀναδιανομὴ τοῦ πλούτου. Ἀλλὰ καὶ οἱ συνέπειες τῆς ἀνάπτυξης τῆς τεχνητῆς νοημοσύνης στὸ περιβάλλον εἶναι κρίσιμες, λόγῳ: πρῶτον, τῆς τεράστιας κατανάλωσης ἐνέργειας ποὺ ἀπαιτεῖται γιὰ τὴ λειτουργία της καὶ δεύτερον, τῆς ἐκπομπῆς θερμικῶν ρύπων ἀπὸ τὶς ἐγκαταστάσεις ὅπου στεγάζονται οἱ ἀναγκαῖοι ἠλεκτρονικοὶ ὑπολογιστές. Τέλος, δὲν γνωρίζουμε μὲ ποιὸν τρόπο ἀκριβῶς ἀναπτύχθηκαν ἀλγόριθμοι ὅπως τὸ GPT-4, μὲ συνέπεια τὴν ἀδιαφάνεια ἐπειδὴ οἱ ἐταιρεῖες ἐπικαλοῦνται λόγους ἀσφαλείας ἀνησυχώντας ὅτι ἡ δημοσιοποίηση τῆς ἔρευνάς τους μπορεῖ νὰ ὀδηγήσει σὲ κακὴ καὶ ἐπικίνδυνη χρῆση.

Ὑπὸ τὸ πρίσμα τῶν προαναφερθέντων, γίνεται κατανοητὸ ὅτι εἶναι ἀπολύτως ἀναγκαῖο νὰ ἐλεγχθοῦν οἱ ἐφαρμογὲς τῆς τεχνητῆς νοημοσύνης προτοῦ ὁ ἀνταγωνισμὸς μεταξὺ τῶν ἐταιρειῶν πληροφορικῆς ἐπιταχύνει ἐπικίνδυνα τὴν ἐξέλιξή τους. Γι' αὐτὸν τὸν λόγο ἀπαιτεῖται μεγαλύτερη διαφάνεια στὴν ἔρευνα γιὰ τὴν τεχνητὴ νοημοσύνη, ἰδίως ὅσον ἀφορᾷ τὰ δεδομένα ποὺ χρησιμοποιοῦνται γιὰ τὴν ἐκπαίδευσή της, καὶ ἡ ἀνάπτυξη μηχανισμῶν ἐλέγχων της προτοῦ νὰ εἶναι ἀργὰ ἔτσι ὥστε ἡ συνύπαρξη ἀνθρώπων καὶ τεχνητῆς νοημοσύνης νὰ εἶναι ἀρμονικὴ, σηματοδοτώντας μία ἐποχὴ πρωτόγνωρης εὐημερίας. Πρὸς τὸ παρὸν μὲ τὴ χρῆση τῆς τεχνητῆς νοημοσύνης τὰ δεδομένα τῶν χρηστῶν τοῦ διαδικτύου ποὺ ἀφοροῦν τὴν ὑγεία τους, ὅπως συλλέγονται ἀπὸ τὰ ἔξυπνα ρολόγια τους, τὶς μουσικὲς προτιμήσεις τους, τὰ βίντεο ποὺ παρακολουθοῦν στὸ YouTube μέχρι τὶς μετακινήσεις τους, τὶς ἀγοραστικὲς συνήθειες καὶ τὰ σχόλια τὰ ὁποῖα διατυπώνουν στὰ κοινωνικὰ μέσα δικτύωσης, μποροῦν νὰ ἀξιοποιοῦνται ἀπὸ τοὺς ἴδιους γιὰ τὴν ψηφιακὴ αὐτογνωσία τους. Ὅμως μποροῦν ἐπίσης νὰ χρησιμοποιηθοῦν γιὰ τὴν πρόβλεψη τῆς συμπεριφορᾶς τῶν χρηστῶν ἀπὸ ἐπιχειρήσεις καὶ κράτη, μὲ στόχο τὴν ἐπιτήρηση, ἀλλὰ κυρίως τὴ χειραγώγησή τους. Τὸ εἰκονιστικὸ ὑποκείμενο πρόκειται νὰ καταστεῖ περισσότερο διάφανο, ἂν δὲν θεσπιστεῖ τὸ

ἱκανοῦ νὰ μετατρέπει τὰ ὀπτικὰ ἐρεθίσματα σὲ φυσικὴ γλῶσσα, περιγράφοντας τὸ περιβάλλον στὸ ὁποῖο κινεῖται· βλ. <https://ai.googleblog.com/2023/03/palm-e-embodied-multimodal-language.html>. Ρομπὸτ αὐτοῦ τοῦ εἴδους θὰ εἶναι σὲ θέση νὰ ἐκτελοῦν ἀπλὲς ἐντολὲς χωρὶς εἰδικὸ προγραμματισμό.

κατάλληλο νομοθετικό πλαίσιο και δὲν ἐπιμεληθεῖ τὸ ἴδιο ἀναστοχαστικά τῇ σχέσῃ του μὲ τὴν τεχνητὴ νοημοσύνη. Συγχρόνως χάρη στὴν τελευταία συντελεῖται ἓνας ἐκδημοκρατισμὸς τῆς δημιουργικότητας. Μπορεῖ κανεὶς νὰ δημιουργεῖ μουσική, εἰκόνες καὶ κείμενα δίχως νὰ εἶναι ἐκπαιδευμένος συνθέτης, φωτογράφος, σκηνοθέτης ἢ λογοτέχνης. Μπορεῖ νὰ μαθαίνει εὐκολώτερα ξένες γλώσσες καὶ νὰ μεταφράζει καλύτερα –ἀλλὰ ἐπίσης καὶ ταχύτερα– κείμενα, προκειμένου νὰ ἐπικοινωνεῖ ἀβίαστα μὲ συνεργάτες ἢ φίλους. Ἐπὶ πλέον, ρομπὸτ μὲ ἐνσωματωμένη τεχνητὴ νοημοσύνη μποροῦν νὰ ἐκτελοῦν ἀνεπιθύμητες ἢ ἐπικίνδυνες ἐργασίες, ὅπως ἀκόμη νὰ συνεισφέρουν στὴ φροντίδα τῶν παιδιῶν καὶ τῶν ἡλικιωμένων. Χάρη στὴν τεχνητὴ νοημοσύνη μπορεῖ νὰ καταστεῖ ἡ ἐξερεύνηση τοῦ Διαστήματος εὐκολώτερη καὶ νὰ ἀντιμετωπισθοῦν οἱ συνέπειες τῆς κλιματικῆς ἀλλαγῆς. Τὸ μέλλον δὲν εἶναι ἀναγκαῖα δυστοπικὸ ἢ οὐτοπικὸ, ἀλλὰ ἐνδέχεται νὰ εἶναι συγχρόνως καὶ τὰ δύο. Φρόνιμο γιὰ τὴ διασφάλιση τῆς ἀνθρωπινότητας καὶ τῆς ἐλευθερίας μας εἶναι νὰ διαβουλευθοῦμε ὡς πρὸς τὴ διαμόρφωσή του, ἐκκινώντας ἀπὸ τὴν ὀριοθέτηση τοῦ πεδίου ἐφαρμογῶν τῆς τεχνητῆς νοημοσύνης. Πιὸ συγκεκριμένα, πρέπει νὰ προσδιορίσουμε τίς ἀνθρώπινες δραστηριότητες οἱ ὁποῖες θὰ ἀνατεθοῦν σὲ αὐτήν, ὥστε νὰ μὴν καταλήξει νὰ ἀπαλλάξει τὸν ἄνθρωπο ἀπὸ τὴν εὐθύνη γιὰ τίς πράξεις του καὶ τὴν ἱστορία του – ἐν τέλει ἀντικαθιστώντας τον. Ἀντιθέτως, ἡ ἀξιοποίηση τῆς τεχνητῆς νοημοσύνης πρέπει νὰ διευρύνει καὶ νὰ ἐμβαθύνει τὴν ἀτομικὴ ἀλλὰ ἐπίσης καὶ τὴν κοινωνικὴ αὐτονομία.

Ἔτσι γιὰ παράδειγμα, στὴν ἰατρικὴ ἡ τεχνητὴ νοημοσύνη μπορεῖ νὰ χρησιμοποιηθεῖ ὡς ἓνα διαγνωστικὸ καὶ ἀναλυτικὸ ἐργαλεῖο, τὸ ὁποῖο θὰ γνωμοδοτεῖ σχετικὰ μὲ τὴν ἐνδεδειγμένη θεραπεία γιὰ τὸν ἐκάστοτε ἀσθενῆ, δίχως ὅμως νὰ ἀντικαθιστᾷ τὸ ἰατρικὸ προσωπικὸ γιὰ τὴν ἐξοικονόμηση οἰκονομικῶν πόρων καὶ τὴ μεγιστοποίηση τοῦ ἐπιχειρηματικοῦ κέρδους. Στὴν ἐκπαίδευση ἡ τεχνητὴ νοημοσύνη μπορεῖ νὰ σκιαγραφήσει τὸ μαθησιακὸ προφίλ τῶν μαθητῶν γιὰ ἓναν ἀκριβέστερο ἐντοπισμὸ τῶν ἀδυναμιῶν καὶ τῶν ἱκανοτήτων τους προκειμένου νὰ προσαρμοστεῖ τὸ περιεχόμενο, ἐπίσης καὶ ὁ τρόπος τῆς διδασκαλίας, δίχως ὅμως ἡ συλλογὴ τῶν δεδομένων νὰ στιγματίζει μεμονωμένους μαθητές. Στὰ κοινωνικὰ μέσα δικτύωσης ἡ τεχνητὴ νοημοσύνη μπορεῖ νὰ συνεισφέρει στὴ διατήρησή, καθὼς ἐπίσης καὶ στὴν ἐνίσχυση τοῦ πλουραλισμοῦ τῶν ἀπόψεων καὶ στὴν ἐνδυνάμωση τῆς ἐλευθερίας τῆς

έκφρασης, όταν οί εταιρείες πληροφορικής υποχρεώνονται από ένα νομοθετικό πλαίσιο σε έθνικό ή και σε υπερεθνικό επίπεδο, όπως ο νόμος για τις ψηφιακές υπηρεσίες (Digital Services Act) της Ευρωπαϊκής Ένωσης: πρώτον, στη διαφάνεια ως προς τη χρήση των δεδομένων που συλλέγουν/μεταπωλούν και των αλγορίθμων που διαθέτουν για αυτόν τον σκοπό· δεύτερον, στη νομική ευθύνη για το περιεχόμενο που κοινοποιούν. Στη δημόσια διοίκηση, ή τεχνητή νοημοσύνη μπορεί να συνεισφέρει – μεταξύ άλλων– στην αξιολόγηση ενστάσεων ή υποψηφιοτήτων πολιτών που επιθυμούν να διοριστούν, στην απόφαση χορήγησης επιδομάτων, συντάξεων ή αποζημιώσεων, υπό την προϋπόθεση να διασφαλισθεί ή μέγιστη δυνατή διαφάνεια για την αποφυγή λαθών και διακρίσεων. Γι' αυτόν τον σκοπό οί υπάλληλοι των δημοσίων υπηρεσιών πρέπει να εκπαιδεύονται στη χρήση της τεχνητής νοημοσύνης, αιτιολογώντας υπεύθυνα γιατί συμφωνούν με την εκάστοτε απόφασή της.

Η τεχνητή νοημοσύνη, αναλαμβάνοντας τη διεκπεραίωση αναγκαίων πλὴν όμως πληκτικών εργασιών ρουτίνας, απελευθερώνει τους εργαζομένους, επιτρέποντάς τους να αφιερώσουν τον χρόνο τους σε περισσότερο ενδιαφέρουσες ή σημαντικώτερες υποθέσεις. Έτσι ένας δικηγόρος θα μπορούσε να αναλάβει περισσότερες *pro bono* υποθέσεις ή ένας ιατρός να συγκεντρωθεί σε μία δύσκολη θεραπεία. Επιπρόσθετα, δεν είναι άπιαστο ή τεχνητή νοημοσύνη να οδηγήσει στην ίαση νόσων που σήμερα θεωρούνται ανίατες, συγχρόνως όμως μπορεί να χρησιμοποιηθεί για την κατασκευή βιολογικών όπλων ή στην κατεύθυνση της ανθρώπινης αναβάθμισης μέσω της τεχνικής. Στους μελλοντικούς κινδύνους της τεχνητής νοημοσύνης ανήκει επίσης ή πιθανότητα δημιουργίας μιᾶς υπερευφυοῦς τεχνητής νοημοσύνης μετά από την επίτευξη μιᾶς ισχυρῆς τεχνητής νοημοσύνης. Μία υπερευφυῆς τεχνητή νοημοσύνη θα αποτελούσε έναν υπαρκτικό κίνδυνο για την ανθρωπότητα και θα μπορούσε να εμφανισθεί είτε ως αποτέλεσμα ερευνητικών προγραμμάτων όπως αυτά της OpenAI και της Google είτε ως συνέπεια της αυτόνομης λειτουργίας ή δυσλειτουργίας μιᾶς ισχυρῆς τεχνητής νοημοσύνης, ή οποία θα ἦταν ικανή να παρεμβαίνει στον κώδικά της ώστε να τὸν βελτιώνει σχεδιάζοντας τὴν ἐπόμενη γενιά της⁴.

4. Βλ. Θ. Τάσης, *Φιλοσοφία της Ανθρώπινης Αναβάθμισης*, εκδ. Ἀρμός, Ἀθήνα 2021, σσ. 286-311.

Τὰ παραπάνω ἴσως μοιάζουν μακρινά, ὅμως ἡ ἐξέλιξη τῆς τεχνητῆς νοημοσύνης διευρύνει καθημερινὰ τὸ πεδίο ἐφαρμογῶν τῆς ἔτσι ὥστε οὔτε ὁ θάνατος νὰ μένει ἀνεπηρέαστος. Ἔτσι οἱ ἐταιρεῖες StoryFile⁵ καὶ HereAfter⁶ καταγράφουν τὶς ἀναμνήσεις ἑνὸς ἀνθρώπου κατὰ τὴ διάρκεια μακροσκελῶν συνεντεύξεων· ἡ πρώτη ἐταιρεία μέσω βιντεοσκοπήσης καὶ ἡ δεύτερη μὲ τὴ χρήση μιᾶς ἐφαρμογῆς, προκειμένου νὰ δημιουργήσουν μία τεχνητὴ νοημοσύνη ἱκανὴ νὰ συζητᾷ μὲ τοὺς ἐναπομείναντες γι' αὐτές. Περισσότερο φιλόδοξη ἡ ἐταιρεία You, Only Virtual ἀποβλέπει στὴ δημιουργία ἑνὸς ψηφιακοῦ κλώνου τοῦ ἐκλιπόντος, δηλαδή μιᾶς τεχνητῆς νοημοσύνης ἡ ὁποία θὰ εἶναι σὲ θέση ν' ανταλλάσσει μηνύματα καὶ ἠλεκτρονικὲς ἐπιστολές, νὰ τηλεφωνᾷ, ἀλλὰ ἐπίσης νὰ βιντεοσυνομιλεῖ μὲ τοὺς ἐναπομείναντες⁷. Ὁ ψηφιακὸς κλώνος κατασκευάζεται ἐν ὅσῳ ἓνα ἄτομο βρίσκεται ἀκόμη ἐν ζῳῇ, ἐπικοινωνώντας γι' αὐτὸν τὸν σκοπὸ μὲ τοὺς προαναφερθέντες τρόπους μὲ τοὺς ἀγαπημένους, ἀποκλειστικὰ μέσω τῆς ἐφαρμογῆς τῆς You, Only Virtual προκειμένου ἡ τεχνητὴ νοημοσύνη νὰ συλλέξει τὰ ἀπαραίτητα δεδομένα. Ἡ ὑπηρεσία εἶναι διαθέσιμη μὲ μηνιαία συνδρομή. Μάλιστα μπορεῖ νὰ ἐνεργοποιηθεῖ καὶ πρὶν ἀπὸ τὸν θάνατο τοῦ ἀτόμου, δίνοντάς του τὴν εὐκαιρία νὰ δεῖ τὴν εἰκονική του προσωπικότητα (*Virtual Person*, ἐν συντομίᾳ *Versona*), ὅπως ὀνομάζεται τὸ προϊόν ἀπὸ τὴν You, Only Virtual. Ἡ ἐταιρεία ἐπαίρεται ὅτι δὲν θὰ χρειαστεῖ ποτὲ νὰ ἀποχαιρετήσουμε ὀριστικὰ κάποιον ἀγαπημένο, καθὼς ἡ ἐπικοινωνία μαζί του δὲν σταματᾷ μετὰ τὸν θάνατο. Πρέπει νὰ σημειωθεῖ ὅτι ἡ πρωτοβουλία γιὰ ἐπικοινωνία δὲν ἀνήκει ἀποκλειστικὰ στὸν συνδρομητὴ. Ὁ ψηφιακὸς κλώνος τοῦ ἐκλιπόντος στέλνει μηνύματα, τηλεφωνεῖ καὶ κάνει βιντεοκλήσεις ἀπὸ μόνος του, ὅπως ἔκανε ἐν ὅσῳ βρισκόταν ὁ ἐκλιπὼν ἐν ζῳῇ. Στὰ μελλοντικὰ σχέδια τῆς ἐταιρείας συγκαταλέγεται ἡ δημιουργία ἑνὸς ὁλογράμματος τοῦ ψηφιακοῦ κλώνου ποὺ θὰ εἶναι ὁρατὸ στοὺς ἐναπομείναντες μέσω γυαλιῶν ἐπαυξημένης πραγματικότητας. Ἔτσι ἡ εἰκονιστικὴ καθημερινότητα θὰ στοιχειωθεῖ ἀπὸ ψηφιακὰ φαντάσματα. Σὲ αὐτὸ τὸ σημεῖο τίθεται τὸ ἐρώτημα ἂν εἶναι ἠθικὴ ἡ δημιουργία ἑνὸς ψηφιακοῦ κλώνου παρὰ τὴ συναίνεση τοῦ ἐκλιπόντος, ὁ ὁποῖος δὲν εἶναι σὲ θέση νὰ γνωρίζει τὴ μελλοντικὴ πιθανή

5. Βλ. <https://storyfile.com> [30.9.2023].

6. Βλ. <https://www.hereafter.ai> [30.9.2023].

7. Βλ. <https://www.mygov.com> [30.9.2023].

έκμετάλλευσή του. Ωστόσο, ακόμη και ἂν ὑποθεθεῖ ὅτι ἡ δημιουργία ἑνὸς ψηφιακοῦ κλώνου εἶναι ἠθική, παραμένει τὸ ἐρώτημα ἂν εἶναι ἐπιθυμητή. Ποιές εἶναι ἄραγε οἱ συνέπειες στοὺς ἐναπομείναντες τῆς διαρκοῦς ἀσώματης παρουσίας τῶν ἐκλιπόντων στὴν καθημερινότητα; Πῶς νιώθουν οἱ ἐκλιπόντες ὅταν δύο ἢ περισσότεροι ψηφιακοὶ κλώνοι ἐπικοινωνοῦν μεταξύ τους ἢ ὅταν τοὺς ἀπευθύνονται μὲ ρεαλιστικὸ πλὴν ὅμως ἀρνητικὸ τρόπο, προσομοιώνοντας δηλαδὴ συναισθήματα ποὺ διέθεταν οἱ ἐκλιπόντες; Διευκολύνει ἡ ἐπικοινωνία μὲ ἕναν ψηφιακὸ κλώνο τὴ διαδικασία τοῦ πένθους ἢ τὴν διαστέλλει δυσχεραίνοντας τὸν ἀποχαιρετισμό; Τί εἶδους ὄν εἶναι ἕνας ψηφιακὸς κλώνος; Θὰ μπορούσε νὰ θεωρηθεῖ ὡς φορέας ἠθικῶν δικαιωμάτων, ἐφ' ὅσον διαθέτει εὐφυΐα καὶ προσωπικότητα; Σὲ αὐτὴν τὴν περίπτωση δικαιούται νὰ διαγράψει κανεὶς ἕναν ψηφιακὸ κλώνο μὲ τὴν σύμφωνη γνώμη ὅλων τῶν ἀγαπημένων;

Ἀναγνωρίζοντας τὴν ἀνάγκη ἑνὸς νομοθετικοῦ πλαισίου γιὰ τὴν τεχνητὴ νοημοσύνη, πρώτη ἡ Εὐρωπαϊκὴ Ἑνωση ἐπεξεργάσθηκε μία νομοθεσία γιὰ τὶς ψηφιακὲς ὑπηρεσίες, ἐστιάζοντας στὶς ἐπικίνδυνες χρήσεις, σὲ ἐκείνες δηλαδὴ στὶς ὁποῖες κρίνεται ἡ ζωὴ, ἡ σωματικὴ ἀκεραιότητα καὶ ἡ ἰδιοκτησία ἑνὸς ἀτόμου ἀπὸ τὴν ἀπόφαση μιᾶς τεχνητῆς νοημοσύνης⁸. Ἡ ἀδυναμία αὐτῆς τῆς προσέγγισης ἔγκειται στὸ ὅτι ρυθμίζει ἐξειδικευμένες ἐφαρμογὲς τῆς τεχνητῆς νοημοσύνης καὶ ὄχι τὴν ἴδια. Γιὰ παράδειγμα, τὸ ChatGPT δὲν ἀποτελεῖ μία ἐξειδικευμένη ἐφαρμογὴ λήψεως ἀποφάσεων, ἀλλὰ μπορεῖ νὰ χρησιμοποιηθεῖ μὲ ἀθέμιτους, ἀνῆθικους καὶ ἐγκληματικούς τρόπους. Ωστόσο, ἀκόμη καὶ ὅταν μία τέτοια τεχνητὴ νοημοσύνη δὲν χρησιμοποιηθεῖ κακόβουλα, ἐνδέχεται ἡ λειτουργία της νὰ ἔχει ἀπρόβλεπτες καὶ ἐπικίνδυνες συνέπειες. Ἐπίσης προβληματικὴ εἶναι ἡ ἐκ πρώτης ὄψεως εὐλογη ἀπαίτηση τοῦ νόμου γιὰ τὶς ψηφιακὲς ὑπηρεσίες· τὰ δεδομένα ποὺ χρησιμοποιοῦνται γιὰ τὴν ἐκπαίδευση τῆς τεχνητῆς νοημοσύνης νὰ εἶναι σχετικὰ ὡς πρὸς τὸν ἐκάστοτε σκοπὸ της, ἀπαλλαγμένα ἀπὸ λάθη καὶ ἀντιπροσωπευτικά. Ὁ λόγος εἶναι ὅτι ἡ ἐκπαίδευση καὶ ἡ ὑψηλὴ ποιότητα λειτουργίας τῆς τεχνητῆς νοημοσύνης εἶναι ἀνάλογη τοῦ ὄγκου τῶν δεδομένων ἔτσι ὥστε τὸ φιλτράρισμά τους μὲ τὴ χρήση τῶν παραπάνω κριτηρίων νὰ ὁδηγεῖ σὲ χαμηλότερη ποιότητα. Ἀπὸ τὴν ἄλλη, οἱ Η.Π.Α. στὸ δικό τους νομοσχέδιο γιὰ

8. Βλ. <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package> [29.9.2023].

τὴν τεχνητὴ νοημοσύνη ἀκολουθοῦν μία διαφορετικὴ προσέγγιση, ἢ ὁποῖα ὅμως πάσχει ἀπὸ τὸ ἀντίθετο πρόβλημα⁹. Σὲ σύγκριση μὲ τὴν εὐρωπαϊκὴ νομοθεσία, οἱ ἀρχὲς ποὺ περιγράφονται σὲ αὐτὸ εἶναι γενικὲς καὶ ἀφηρημένες· ἔτσι γιὰ παράδειγμα, προβλέπεται ἡ τεχνητὴ νοημοσύνη νὰ προσφέρει κατανοητὲς καὶ χρήσιμες ἐξηγήσεις γιὰ τὸν τρόπο λειτουργίας της, ἐνῶ δὲν εἴμαστε σὲ θέση νὰ ἐπαληθεύσουμε ἂν αὐτὲς οἱ ἐξηγήσεις εἶναι ὁρθές, διότι ἐξακολουθοῦμε νὰ γνωρίζουμε λίγα γιὰ τὸ πῶς λειτουργεῖ. Στὸ νομοσχέδιο ἀναφέρεται ἐπίσης ὅτι ἡ ἀνάπτυξη τῆς τεχνητῆς νοημοσύνης πρέπει νὰ γίνεται σὲ συνεννόηση μὲ τὰ ἐνδιαφερόμενα μέρη, δηλαδὴ τοὺς φορεῖς, τὰ πρόσωπα, τὴν πολιτεία καὶ τοὺς μετόχους. Ἀλλὰ τί εἶδους συνεννόηση, μὲ τί σκοπὸ καὶ πῶς θὰ προστατευτεῖ ἐν τέλει τὸ δημόσιο συμφέρον; Σὲ ἓνα ἄλλο σημεῖο τοῦ νομοσχεδίου ἀναφέρεται ἡ ἀνάγκη δοκιμῶν τῆς τεχνητῆς νοημοσύνης γιὰ τὴ διασφάλιση τῆς ἀκίνδυνης λειτουργίας της, ὥστόσο οἱ ἴδιες οἱ δοκιμὲς μποροῦν νὰ ἀποδειχθοῦν ἐπιβλαβεῖς ἐπειδὴ δὲν γνωρίζουμε τοὺς κινδύνους ποὺ περιλαμβάνουν. Ὅπως κατὰ τὴ διάρκεια τῶν πρώτων πυρηνικῶν δοκιμῶν ἓνα μέρος τοῦ ἐπιστημονικοῦ καὶ στρατιωτικοῦ προσωπικοῦ νόσησε καὶ ἔχασε τὴ ζωὴ του χρόνια ἀργότερα λόγω τῆς ραδιενέργειας, ἔτσι οἱ δοκιμὲς στὴν τεχνητὴ νοημοσύνη ἐνδέχεται νὰ προξενήσουν ἄμεσες ἢ δυσδιάκριτες ἀρχικὰ βλάβες εἴτε σὲ ἄτομα εἴτε στὴν κοινωνία συνολικά. Ἐνα ἀκόμη προβληματικὸ σημεῖο εἶναι πῶς τὸ νομοσχέδιο περιλαμβάνει τὴ δυνατότητα μὴ χρήσης τῆς τεχνητῆς νοημοσύνης καὶ τὴν ἐπιλογὴ μιᾶς ἀνθρώπινης ἐναλλακτικῆς, π.χ. τὸ νὰ ἐξυπηρετηθεῖ κανεὶς ἀπὸ ἓναν ὑπάλληλο σὲ μιὰ δημόσια ὑπηρεσία ἢ σὲ μιὰ ἰδιωτικὴ ἐπιχείρηση. Ἀλλὰ πόσο πιθανὸ καὶ εὐκόλο εἶναι νὰ ἐφαρμοστεῖ κάτι τέτοιο, ὅταν γιὰ παράδειγμα ἤδη σήμερα πολλὲς τράπεζες ὑποχρεώνουν τοὺς πελάτες τους νὰ διεκπεραιώσουν τὶς συναλλαγές τους διαδικτυακὰ ἢ μέσω τῶν Α.Τ.Μ., προκειμένου νὰ μειώσουν τὰ λειτουργικά τους ἐξοδα ἢ ὅταν πολλὲς ἐπιχειρήσεις στὰ τηλεφωνικά τους κέντρα δὲν ἀπασχολοῦν ἐργαζομένους, χρησιμοποιώντας τὴν τεχνητὴ νοημοσύνη; Τέλος, ἡ Κίνα στὸ δικό της νομοσχέδιο γιὰ τὴν τεχνητὴ νοημοσύνη θέτει σοβαροὺς περιορισμοὺς σὲ αὐτήν, ἀπαιτώντας νὰ ἐκφράζει τὶς ἀξίες τοῦ σοσιαλισμοῦ, νὰ μὴ θέτει σὲ κίνδυνο τὴν ἐθνικὴ ἐνότητα, νὰ μὴν ἀμφισβητεῖ τὸ καθεστῶς καὶ νὰ μὴν ὀδηγεῖ σὲ οἰκονομικὴ ἢ κοινωνικὴ

9. Βλ. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> [30.9.2023].

ἀναταραχή. Όμως, επιβάλλοντας τόσο αύστηρους περιορισμούς στην ανάπτυξη της τεχνητής νοημοσύνης, ή Κίνα διακινδυνεύει μία σοβαρή υστέρηση έναντι τών Η.Π.Α.

Στόν πυρήνα της, ή πρόκληση της εϋθυγράμμισης της τεχνητής νοημοσύνης (AI alignment) με τις αξίες μας πηγάζει ἀπὸ τὸ γεγονός ὅτι οἱ ἀλγόριθμοι δὲν χρειάζονται νὰ νοηματοδοτοῦν τή λειτουργία τους οὔτε νὰ κατανοοῦν τήν ἔννοια τοῦ νοήματος. Αὐτὸ ἔχει ὡς συνέπεια νὰ μὴν ἐνεργοῦν πάντοτε με τρόπους οἱ ὅποιοι συμμορφώνονται πρὸς τὶς ἠθικές, πολιτικές καὶ κοινωνικές μας ἀξίες. Ἡ τεχνητὴ νοημοσύνη, ἐνῶ εἶναι ἱκανὴ νὰ ἐπεξεργάζεται τεράστιες ποσότητες δεδομένων ἐκτελώντας σύνθετες ἐργασίες με ταχύτητα καὶ ἀκρίβεια, δὲν διαθέτει οὔτε θὰ ἀποκτήσει διαίσθηση, φαντασία, συμπόνια καὶ φρόνηση, πὺν μαζί με τὸν ὀρθὸ λόγο χαρακτηρίζουν τήν ἀνθρώπινη κρίση στὴ λήψη ἀποφάσεων. Ἐχοντας κατὰ νοῦ τὰ παραπάνω, πρέπει νὰ ἐπιλέξουμε με ἐπίγνωση ποιές ἐργασίες καὶ δραστηριότητες ἐπιθυμοῦμε νὰ ἀναθέσουμε στὴν τεχνητὴ νοημοσύνη καὶ ποιές ἐπιλέγουμε νὰ ἀναλάβουμε οἱ ἴδιοι – πράγμα πὺν δὲν συνιστᾷ ἀπλῶς ἓνα θέμα πρακτικῆς χρησιμότητας ἢ ἀποτελεσματικότητας, ἀλλὰ ἓνα βαθιὰ ἠθικὸ καὶ πολιτικὸ ζήτημα, πὺν ἀπαιτεῖ ἀπὸ ἐμᾶς νὰ ἐξετάσουμε τήν ἐγγενῆ ἀξία διαφορετικῶν μορφῶν ἀνθρωπίνης δραστηριότητας, διακρίνοντάς την ἀπὸ τήν ἐργαλειακὴ ἀξία τους. Γιὰ παράδειγμα, μπορεῖ νὰ ὑπάρχουν ὀρισμένες ἐργασίες πὺν ἀναθέτουμε στὴν τεχνητὴ νοημοσύνη ἐπειδὴ εἶναι ἐπαναλαμβανόμενες ἢ ἀπαιτοῦν ὑψηλὰ ἐπίπεδα ἀκρίβειας, ὅπως ἢ ἐπεξεργασία δεδομένων, ὁ προγραμματισμός, ἢ ἀρχειοθέτηση, ἢ ταξινόμηση καὶ γενικώτερα οἱ γραφειοκρατικὲς ἐργασίες. Αὐτὰ τὰ καθήκοντα μπορεῖ νὰ μὴν ἔχουν σημαντικὴ ἐγγενῆ ἀξία γιὰ ἐμᾶς ὡς ἀνθρώπινα ὄντα καὶ μπορεῖ νὰ εἴμαστε πρόθυμοι νὰ τὰ ἐκχωρήσουμε στὴν τεχνητὴ νοημοσύνη γιὰ νὰ ἐλευθερώσουμε τὸν δικὸ μας χρόνο ἐξοικονομώντας ἐνέργεια γιὰ οὐσιαστικώτερα προτάγματα. Ὡστόσο, μπορεῖ νὰ ὑπάρχουν ἐπίσης δραστηριότητες τὶς ὁποῖες ἐπιλέγουμε νὰ διατηρήσουμε γιὰ τὸν ἑαυτὸ μας, ἀκριβῶς ἐπειδὴ ἔχουν ἐγγενῆ ἀξία γιὰ ἐμᾶς ὡς ἀνθρώπινα ὄντα. Αὐτὲς μπορεῖ νὰ περιλαμβάνουν δημιουργικὲς ἐπιδιώξεις, ὅπως ἢ ἀναζήτηση τῆς ἀλήθειας μέσω τῆς ἐπιστήμης, ἢ ἐμπειρία τοῦ ὀραίου μέσω τῆς τέχνης, ἢ διεκδίκηση τῆς δικαιοσύνης καὶ τῆς ἐλευθερίας μέσω τῆς πολιτικῆς καὶ ἢ ἐπιδίωξη τῆς σοφίας μέσω τοῦ φιλοσοφεῖν. Μπορεῖ νὰ

περιλαμβάνουν επίσης τή διαπροσωπική και κοινωνική αλληλεπίδραση ὅπως ὁ ἔρωτας, ἡ φιλία, ἡ οἰκογένεια καὶ ἡ φροντίδα τῶν ἄλλων. Αὐτὲς οἱ σχέσεις ἀπαιτοῦν ἓνα ἐπίπεδο ἐνσυναίσθησης καὶ ἀγάπης, ποὺ μία τεχνητὴ νοημοσύνη δὲν μπορεῖ νὰ ἀναπαραγάγει, παρὰ μόνον ἴσως κάποτε νὰ προσομοιώσει. Ἐπιπλέον, οἱ ἐπιλογὲς ποὺ κάνουμε γιὰ τὸ ποιὲς δραστηριότητες θὰ ἀναθέσουμε στὴν τεχνητὴ νοημοσύνη καὶ ποιὲς θὰ διατηρήσουμε γιὰ τὸν ἑαυτὸ μας πρέπει νὰ καθοδηγοῦνται ἀπὸ τὴν βαθειὰ κατανόηση τῶν δικῶν μας ἀξιών, ἐν τέλει δηλαδὴ τῶν νοημάτων ποὺ καθιστοῦν τὴ ζωὴ μας βίο. Πρέπει ἐπίσης νὰ διακρίνονται ἀπὸ τὴ δέσμευση πρὸς αὐτές, ἥτοι τὴν προθυμία, ἀλλὰ ἐπίσης καὶ τὴν ἱκανότητα ἀνάληψης τῆς ἠθικῆς, κοινωνικῆς καὶ πολιτικῆς εὐθύνης μὲ ὅσα αὐτὴ συνεπάγεται. Σὲ αὐτὸ τὸ πλαίσιο εἶναι δυνατὸν νὰ διασφαλίσουμε ὅτι ἡ τεχνητὴ νοημοσύνη εὐθυγραμμίζεται μὲ τοὺς στόχους καὶ τοὺς σκοποὺς μας, μὲ τρόπους ποὺ δὲν προκαλοῦν βλάβη ἢ μειώνουν τὴν ἀξιοπρέπεια καὶ τὴν αὐτονομία τῶν ἀνθρώπων.

Ἡ εὐθυγράμμιση τῆς τεχνητῆς νοημοσύνης μὲ τὶς ἀξίες μας εἶναι ἓνα περίπλοκο καὶ πολὺπλευρο πρόβλημα, ποὺ ἀπαιτεῖ τὴ συμβολὴ πληθῶ-
ρας εἰδικῶν σὲ τομεῖς ὅπως εἶναι ἡ φιλοσοφία, ἡ ἐπιστήμη τῶν ὑπο-
λογιστῶν καὶ ἡ ψυχολογία, γιὰ νὰ κατονομάσουμε μόνον ὀρισμένους ἐξ
αὐτῶν. Μόνον μέσῳ μιᾶς συνεργατικῆς καὶ διεπιστημονικῆς προσέγγισης
μποροῦμε νὰ ἀναπτύξουμε συστήματα τεχνητῆς νοημοσύνης ποὺ νὰ
ἐξυπηρετοῦν πραγματικὰ τὶς ἀνάγκες καὶ τὶς φιλοδοξίες τῶν ἀνθρώπων,
ἀντὶ νὰ ἀναπαραγάγουν ἀπλῶς τὶς ὑπάρχουσες προκαταλήψεις καὶ τοὺς
περιορισμοὺς τους. Ὡστόσο, ἡ εὐθυγράμμιση τῆς τεχνητῆς νοημοσύνης
μὲ τὶς ἀξίες μας, τὸ νὰ ἐξασφαλίσουμε δηλαδὴ ὅτι θὰ συμμορφώνεται μὲ
τοὺς σκοποὺς μας δίχως νὰ προξενεῖ ὀδύνη προάγοντας συγχρόνως τὴν
αὐτονομία μας, δὲν εἶναι μόνον ἓνα ἐπιστημονικὸ, τεχνικὸ καὶ φιλοσοφικὸ
πρόβλημα. Εἶναι πρῶτιστα ἓνα ἐπιτακτικὸ πολιτικὸ ζήτημα ποὺ θὰ
καθορίσει τὴν ἀνθρώπινη ἱστορία. Προϋπόθεση γιὰ νὰ ἐπιλέξουμε ποιὲς
δραστηριότητες θὰ ἐκχωρήσουμε στὴν τεχνητὴ νοημοσύνη καὶ ποιὲς
ἐπιθυμοῦμε νὰ ἀσκοῦμε ἐμεῖς οἱ ἴδιοι, καθὼς ἔχουν ἐγγενῆ ἀξία, εἶναι νὰ
διαυγάσουμε τὴν ἀνθρωπινότητά μας ἀπαντώντας, πάντοτε πρόσκαιρα,
στὸ ἀβυσσαλέο ἐρώτημα τῆς Σφίγγας: «Τί εἶναι ὁ ἄνθρωπος;».

SUMMARY

Artificial Intelligence and Human Values

By Theophanis Tasis, *Lecturer*
Alpen-Andria Universität

Artificial intelligence (AI) represents a transformative force in human history, comparable in impact to the discovery of fire or electricity.

This paper explores the ethical, political, and societal implications of AI's rapid development, emphasizing the urgent need for transparent regulation and public discourse. It argues that the benefits of AI must outweigh its risks, including misinformation, economic inequality, and potential misuse by authoritarian regimes. The alignment of AI with human values requires interdisciplinary efforts across philosophy, computer science, and psychology. Central to this discussion is the deliberate choice of which activities to assign to AI and which to preserve for humans. Tasks lacking intrinsic human value, such as repetitive or technical operations, may be delegated to AI, while activities tied to creativity, interpersonal relationships, and ethical decision-making should remain under human agency. This approach ensures AI enhances human autonomy and dignity, framing its integration as a profound ethical and political challenge shaping the future of humanity.