



## Τεχνητή νοημοσύνη για το κοινωνικό όφελος: αποφεύγοντας την παγίδα της μονομέρειας

ΕΥΡΩΠΑΪΚΟΙ ΘΕΣΜΟΙ & ΠΟΛΙΤΙΚΕΣ

**Ανδρέας ΤΣΑΜΑΔΟΣ**

*Διδακτορικός Ερευνητής στο Oxford Internet Institute,  
Πανεπιστήμιο Οξφόρδης*



# Τεχνητή νοημοσύνη για το κοινωνικό όφελος: αποφεύγοντας την παγίδα της μονομέρειας

Copyright © 2021 | All Rights Reserved

HELLENIC FOUNDATION FOR EUROPEAN & FOREIGN POLICY (ELIAMEP)

49, Vasilissis Sofias Ave., 10676, Athens, Greece

Tel.: +30 210 7257 110 | Fax: +30 210 7257 114 | [www.eliamep.gr](http://www.eliamep.gr) | [eliamep@eliamep.gr](mailto:eliamep@eliamep.gr)

Το ΕΛΙΑΜΕΠ δεν υιοθετεί ως ίδρυμα πολιτικές θέσεις. Καταβάλλει μάλιστα προσπάθεια να παρουσιάζονται στα πλαίσια των δημοσιεύσεών του και στο μέτρο του δυνατού όλες οι υπάρχουσες απόψεις. Υπό το πρίσμα αυτό, οι αναλύσεις και οι γνώμες που δημοσιεύονται στις σειρές του θα πρέπει να αποδίδονται αποκλειστικά στους συγγραφείς και να μην θεωρείται ότι αντιπροσωπεύουν απαραίτητα τις απόψεις του Ιδρύματος.

## Ανδρέας ΤΣΑΜΑΔΟΣ

Διδακτορικός Ερευνητής στο *Oxford Internet Institute* του Πανεπιστημίου της Οξφόρδης

### Περίληψη

- Η Τεχνητή Νοημοσύνη μπορεί να παίζει ουσιαστικό ρόλο στην επίλυση παγκόσμιων, ολοένα πιο διασυνδεδεμένων και περίπλοκων προβλημάτων.
- Τα διακριτά πλεονεκτήματα προβλέψεων, κατηγοριοποιήσεων και αποφάσεων που προσφέρουν τα συστήματα Τεχνητής Νοημοσύνης δύνανται να δημιουργήσουν μοναδικές προκλήσεις, αλλά και να επιδεινώσουν ήδη υπάρχοντα κοινωνικά προβλήματα.
- Η κατανόηση της Τεχνητής Νοημοσύνης για το κοινωνικό όφελος ως παγκόσμιο φαινόμενο παραμένει ανεπαρκής, ενώ απουσιάζει και ένα συγκροτημένο πλαίσιο αξιολόγησης τέτοιων προσπαθειών.
- Η προσπάθεια εκμετάλλευσης της δυναμικής της Τεχνητής Νοημοσύνης από τη διεθνή κοινότητα για την επίλυση ορισμένων από τα πιο επείγοντα ζητήματα δίχως ένα αξιόπιστο πλαίσιο διακυβέρνησης μπορεί να αλλοιώσει την αξιολόγηση ρεαλιστικών και αξιόλογων προσφερόμενων ευκαιριών.
- Το εύρος και η σχεδόν καθολική αποδοχή των Στόχων Βιώσιμης Ανάπτυξης του ΟΗΕ δίνουν τη δυνατότητα σε έργα που επιδιώκουν να αξιοποιήσουν την τεχνητή νοημοσύνη με τρόπο κοινωνικά επωφελή, αλλά και στους ερευνητές που τα παρακολουθούν, να χρησιμοποιήσουν τους Στόχους ως αφετηρία και σημείο αναφοράς ως προς την αξιολόγησή τους.
- Το υφιστάμενο σώμα έρευνας σχετικά με τους Στόχους Βιώσιμης Ανάπτυξης περιλαμβάνει μελέτες και μεθόδους μέτρησης σχετικά με τον υπολογισμό της προόδου επίτευξης εκάστου εκ των 17 Στόχων Βιώσιμης Ανάπτυξης και των 169 συνδεδεμένων στόχων, όπως αυτοί ορίστηκαν στην Ατζέντα 2030 για τη Βιώσιμη Ανάπτυξη. Αυτές οι μέθοδοι μπορούν να χρησιμοποιηθούν για την αξιολόγηση της επίδρασης συνδυαστικών έργων Τεχνητής Νοημοσύνης και Στόχων Βιώσιμης Ανάπτυξης.
- Η ανάλυση καταδεικνύει πως ο συνδυασμός Τεχνητής Νοημοσύνης και Στόχων Βιώσιμης Ανάπτυξης αποτελεί ένα παγκόσμιο φαινόμενο, με έργα να εκτελούνται σε πέντε ηπείρους, αλλά και πως το φαινόμενο αυτό δεν αφορά εξίσου όλους τους Στόχους Βιώσιμης Ανάπτυξης. Ο Στόχος 3 («Καλή υγεία και ευζωία») βρίσκεται στην πρώτη θέση, με τους στόχους 5 («Ισότητα των φύλων»), 16 («Ειρήνη, Δικαιοσύνη και Ισχυροί Θεσμοί») και 17 («Συνέργειες για την επίτευξη των Στόχων») να εντάσσονται σε λιγότερα από πέντε έργα.

## Εισαγωγή

“...η Τεχνητή Νοημοσύνη μπορεί να διαδραματίσει σημαίνοντα ρόλο στην επίλυση παγκόσμιων προβλημάτων, τα οποία παρουσιάζουν ολοένα πιο αυξημένο βαθμό αλληλεπίδρασης και περιπλοκότητας.”

Η Τεχνητή Νοημοσύνη (AI) εγκαθίσταται σιγά-σιγά στις κοινωνίες μας, που ξεπερνούν τον αρχικό αιφνιδιασμό, κάτι συνηθισμένο για τις περισσότερες επαναστατικές τεχνολογίες. Κατά την τελευταία δεκαετία, η Τεχνητή Νοημοσύνη συχνά απεικονίστηκε είτε ως μια δραματική τεχνολογική μεταβολή δυστοπικού χαρακτήρα, είτε ως ένα μαγικό ραβδί για την επίλυση των περισσότερων, αν όχι όλων, των πιθανών προβλημάτων, από την αντιμετώπιση της κλιματικής αλλαγής έως την πρόληψη του εγκλήματος παγκοσμίως. Αυτό το άρθρο συμβάλλει σε μια πιο σύγχρονη και ρεαλιστική αξιολόγηση που αρχίζει να βλέπει το φως: Η Τεχνητή Νοημοσύνη είναι μια ισχυρή τεχνολογία που μπορεί να μας βοηθήσει να αντιμετωπίσουμε ένα ευρύ φάσμα επαρκώς προσδιορισμένων προβλημάτων, υπό την προϋπόθεση του κατάλληλου σχεδιασμού, της ανάπτυξης, της ολοκλήρωσης και της παρακολούθησης.

Βασιζόμενο σε προηγούμενη έρευνα που παρουσιάστηκε στα έργα (Cowls et al., 2021a) και (Cowls et al., 2021b), το παρόν άρθρο αναλύει το περίπλοκο ζήτημα της χρήσης Τεχνητής Νοημοσύνης για την επίτευξη κοινωνικά επωφελών αποτελεσμάτων. Οι δύο πρώτες ενότητες περιλαμβάνουν μία επισκόπηση των διάφορων περιορισμών των συστημάτων Τεχνητής Νοημοσύνης, καθώς και υφιστάμενων έργων που επιδιώκουν να αξιοποιήσουν την Τεχνητή Νοημοσύνη για το ευρέως εννοούμενο «κοινωνικό όφελος». Με τη σειρά μου, υποστηρίζω πως παρά τις κοινωνικοτεχνικές προκλήσεις η Τεχνητή Νοημοσύνη μπορεί να διαδραματίσει σημαίνοντα ρόλο στην επίλυση παγκόσμιων προβλημάτων, τα οποία παρουσιάζουν ολοένα πιο αυξημένο βαθμό αλληλεπίδρασης και περιπλοκότητας. Η τελευταία ενότητα του άρθρου περιλαμβάνει την παρουσίαση των αποτελεσμάτων μίας διετούς έρευνας (Cowls et al., 2021a) σχετικά με την Τεχνητή Νοημοσύνη, η οποία υποστηρίζει τους Στόχους Βιώσιμης Ανάπτυξης των Ηνωμένων Εθνών και τη χρήση τους ως σημείο αναφοράς για την αξιολόγηση της συνεισφοράς της Τεχνητής Νοημοσύνης στο κοινωνικό όφελος.

## Ούτε φορέας αισθητήριας αντίληψης, ούτε τέλειο δημιουργήμα

Πριν ξεκινήσουμε να αναλογιζόμαστε τη δυναμική αυτής της ισχυρής νέας τεχνολογίας είναι σημαντικό να συμφωνήσουμε ως προς το τι εννοούμε με τον όρο «Τεχνητή Νοημοσύνη». Εν προκειμένω επέλεξα να υιοθετήσω τον ορισμό της Ευρωπαϊκής Επιτροπής, η οποία από το 2016 κατάφερε να ηγηθεί των διεθνών προσπαθειών σε ό,τι αφορά στρατηγικές και κανονισμούς για την Τεχνητή Νοημοσύνη (Roberts et al., 2021). Η Ομάδα Εμπειρογνομόνων σε ζητήματα Τεχνητής Νοημοσύνης της Ευρωπαϊκής Ένωσης (AI HLEG) παρέχει έναν ολοκληρωμένο ορισμό της Τεχνητής Νοημοσύνης (AI HLEG, 2019), ο οποίος συνίσταται από τα ακόλουθα:

“Η Τεχνητή Νοημοσύνη δεν αποτελεί κάποιο φορέα αισθητήριας αντίληψης, ικανό να καταλάβει τον πλανήτη αν «ξεφύγει από τον ανθρώπινο έλεγχο» όπως αναφέρεται συχνά στις ειδήσεις.”

- i) συστήματα λογισμικού (αλλά και υλικό)
- ii) σχεδιασμένα από ανθρώπους, τα οποία
- iii) δεδομένου ενός σύνθετου στόχου
- iv) ενεργούν σε φυσική ή ψηφιακή διάσταση αντιλαμβανόμενα το περιβάλλον τους μέσω καταγραφής δεδομένων
- v) ερμηνεύοντας τα συλλεχθέντα δομημένα ή αδόμητα δεδομένα
- vi) καταλήγοντας σε συμπεράσματα ή επεξεργαζόμενα τις πληροφορίες που προκύπτουν

από τα δεδομένα

vii) και αποφασίζοντας για τις βέλτιστες ενέργειες προς την επίτευξη του δεδομένου στόχου.

Η Τεχνητή Νοημοσύνη δεν αποτελεί κάποιο φορέα αισθητήριας αντίληψης ικανό να καταλάβει τον πλανήτη αν «ξεφύγει από τον ανθρώπινο έλεγχο» όπως αναφέρεται συχνά στις ειδήσεις (Friend, 2018). Ένας τρόπος κατανόησης των γενικών περιορισμών της Τεχνητής Νοημοσύνης είναι η αντίληψη της τεχνολογίας ως ικανότητα χωρίς κατανόηση.

*“Τα διακριτά πλεονεκτήματα προβλέψεων, κατηγοριοποιήσεων και αποφάσεων που προσφέρουν τα συστήματα Τεχνητής Νοημοσύνης δύνανται να δημιουργήσουν μοναδικές προκλήσεις, αλλά και να επιδεινώσουν υπάρχοντα κοινωνικά προβλήματα. Η αναγνώριση και αντιμετώπιση ορισμένων εξ αυτών αποτελεί προϋπόθεση για οποιαδήποτε έργα ή προσπάθειες που επιδιώκουν να αξιοποιήσουν την Τεχνητή Νοημοσύνη για κοινωνικά ευεργετικά αποτελέσματα.”*

Ένα σύστημα Τεχνητής Νοημοσύνης μπορεί να δίνει την εντύπωση πως ζωγραφίζει με τη χρήση της δικής του φαντασίας ή πως διαμορφώνει τις δικές του ιδέες και επιχειρήματα, τα χαρακτηριστικά αυτά εντούτοις δεν γίνονται κατανοητά από το ίδιο το σύστημα. Αντιθέτως, το σύστημα απλώς παράγει αποτελέσματα που εκφράζονται με πιθανολογικούς όρους (Tsamados et al., 2021). Για παράδειγμα, κατά τη διάρκεια της έρευνάς μου σχετικά με τη μαζική παραγωγή πολιτικής προπαγάνδας, προγραμματίσα έναν μοντέλο γλωσσικού μετασχηματισμού (μια μορφή τεχνητής νοημοσύνης) ώστε να γράφει ποιήματα στο ύφος του Κωνσταντίνου Καβάφη. Το σύστημα κατάφερε να δημιουργήσει παρόμοια ποιήματα με εκείνα του Καβάφη όχι επειδή κατόρθωσε πώς να αποδίδει ζωντανές εικόνες και συναισθήματα όπως ο περίφημος ποιητής, αλλά επειδή κάθε φορά που εισήγαγα μία λέξη στο σύστημα (π.χ. «εκκλησία») αυτό απλώς αναζητούσε όλα τα γραπτά του Καβάφη και υπολόγιζε ποια λέξη είχε τη μεγαλύτερη πιθανότητα να ακολουθήσει μετά την «εκκλησία» στο έργο του Καβάφη.

Καθώς περισσότερα συστήματα Τεχνητής Νοημοσύνης ενσωματώνονται σε κρίσιμες υποδομές και σχεδόν όλες οι πτυχές της ανθρώπινης δραστηριότητας, από τα δίκτυα ηλεκτρικού ρεύματος ως τα νοσοκομεία, την επιλογή φίλων και μουσικής στο διαδίκτυο, οι περιπτώσεις κατάρρευσης των συστημάτων αυτών ή/και πρόκλησης βλάβης σε ανθρώπους αυξάνονται με ανησυχητικό ρυθμό (Tsamados et al., 2021). Τα διακριτά πλεονεκτήματα προβλέψεων, κατηγοριοποιήσεων και αποφάσεων που προσφέρουν τα συστήματα Τεχνητής Νοημοσύνης δύνανται να δημιουργήσουν μοναδικές προκλήσεις, αλλά και να επιδεινώσουν ήδη υπάρχοντα κοινωνικά προβλήματα. Η αναγνώριση και αντιμετώπιση ορισμένων εξ αυτών (ανάλογα με το πλαίσιο) αποτελεί προϋπόθεση για οποιαδήποτε έργα ή προσπάθειες που επιδιώκουν να αξιοποιήσουν την Τεχνητή Νοημοσύνη για κοινωνικά ευεργετικά αποτελέσματα. Για τους σκοπούς του παρόντος άρθρου, δεν θα απარიθμήσω όλους τους πιθανούς ηθικούς κινδύνους αναφορικά με την εξέλιξη και την ενσωμάτωση συστημάτων Τεχνητής Νοημοσύνης (βλ. Tsamados et al., 2021 για περισσότερα), αλλά θα παρουσιάσω εκείνους που θεωρώ πιο σχετικούς με τη συζήτηση για την Τεχνητή Νοημοσύνη και το κοινωνικό όφελος.

Πρώτον, όσα περιγράψαμε παραπάνω (v) ως ερμηνεία δεδομένων μπορούν να δημιουργήσουν ζητήματα ανεπιθύμητης μεροληψίας και διάκρισης, καθώς τα δεδομένα μπορεί να μην είναι ακριβή, επίκαιρα ή κατάλληλα για ένα συγκεκριμένο πλαίσιο. Τέτοια ζητήματα μπορούν επίσης να οφείλονται σε προβληματικές υποθέσεις πριν τη συλλογή και την επεξεργασία των δεδομένων ή στην εξαγωγή δεδομένων από ένα ήδη άδικο και προκατειλημμένο σύστημα (Tsamados et al., 2021). Τα αποτελέσματα που θα παράγει ένα σύστημα με τα παραπάνω ελαττώματα δεν μπορούν να διορθωθούν όσο και αν αυξηθεί η ποσότητα των δεδομένων. Άνθρωποι έχουν αντιμετωπιστεί διαφορετικά σε νοσοκομεία ή τους έχει επιβληθεί μεγαλύτερη ποινή φυλάκισης εξαιτίας του χρώματος του δέρματός τους και της διαιώνισης από τα συστήματα Τεχνητής Νοημοσύνης των δομικών ανισοτήτων στα δεδομένα προγραμματισμού τους (Benjamin, 2019; Obermeyer et al., 2019).



Ένα δεύτερο και σχετικό ζήτημα αφορά στον κίνδυνο παραβίασης της ανθρώπινης αυτονομίας από την Τεχνητή Νοημοσύνη (Cowls et al., 2021b). Η αυξανόμενη εξάρτησή μας από τις συστάσεις των συστημάτων Τεχνητής Νοημοσύνης, όπως η πρόσληψη πολιτικών πληροφοριών ή ακόμη και η κατανομή προγραμμάτων κρατικής πρόνοιας, μπορεί να έχουν σοβαρές επιπτώσεις στην ανθρώπινη αυτενέργεια. Η ψευδαίσθηση της μηχανιστικής αντικειμενικότητας με την οποία έχουμε συνδυάσει τις «έξυπνες μηχανές» το μόνο που καταφέρνει είναι να διαιωνίζει αυτήν την απώλεια αυτονομίας.

Τρίτον, τη βιβλιογραφία σχετικά με την Τεχνητή Νοημοσύνη απασχολεί εδώ και καιρό το ζήτημα της διαφάνειας (Ananny and Crawford, 2018). Η αδιαφάνεια πολλών συστημάτων Τεχνητής Νοημοσύνης μπορεί να οφείλεται σε διάφορους λόγους, όπως η συσκοτίση των δεδομένων προγραμματισμού και η έλλειψη εργαλείων για την οπτικοποίηση και παρακολούθηση του συνόλου δεδομένων ή εξελισσόμενων παραμέτρων του συστήματος. Με τη σειρά του, αυτό μπορεί να οδηγήσει σε έλλειψη λογοδοσίας και διάβρωση της εμπιστοσύνης σε αυτά τα συστήματα ή τους προγραμματιστές τους.

Τέταρτον, η συλλογή προσωπικών δεδομένων για τον προγραμματισμό των συστημάτων Τεχνητής Νοημοσύνης εγείρει σημαντικά ερωτήματα για την ιδιωτικότητα ατόμων και ομάδων (Taylor, Floridi and van der Sloot, υπό έκδοση). Ακόμη και αν αυτή αφορά γενικά παραδεκτούς καλούς σκοπούς, όπως η παρακολούθηση της εξάπλωσης του κορωνοϊού, η συλλογή και η χρήση προσωπικών δεδομένων αποτελεί ευαίσθητο ζήτημα με επιπτώσεις στα θεμελιώδη ανθρώπινα δικαιώματα, οι οποίες καλύπτονται εκτενώς από τον Γενικό Κανονισμό Προστασίας Δεδομένων της ΕΕ.

Πέμπτον, η τρέχουσα ανάπτυξη της Τεχνητής Νοημοσύνης απαιτεί σημαντικές ποσότητες ενέργειας (Taddeo et al., 2021), ιδιαίτερα κατά τη φάση προγραμματισμού των συστημάτων Τεχνητής Νοημοσύνης καθώς οι σημαντικότερες ανακαλύψεις στον τομέα της Τεχνητής Νοημοσύνης από το 2012 απαιτήσαν ολοένα και μεγαλύτερη υπολογιστική ισχύ. Μεταξύ 2012 και 2018 η υπολογιστική ισχύς για τον προγραμματισμό μεγάλων συστημάτων Τεχνητής Νοημοσύνης διπλασιάζεται κάθε 3,4 μήνες (Amodei and Hernandez, 2018). Αυτό σημαίνει πως αν η ενεργειακή αποδοτικότητα αυτών των συστημάτων και της σχετικής υποδομής (π.χ. κέντρα δεδομένων και ηλεκτρικά δίκτυα) δεν συμβαδίσουν με την αύξηση της ζήτησης για υπολογιστική ισχύ, το αποτύπωμα άνθρακα για την έρευνα και ανάπτυξη της Τεχνητής Νοημοσύνης θα αυξηθεί ραγδαία. Ένα σχετικό ζήτημα που αξίζει να αναφερθεί είναι πως το μεγαλύτερο ποσοστό έρευνας και ανάπτυξης για την Τεχνητή Νοημοσύνη σήμερα λαμβάνει χώρα σε δυτικές βιομηχανικές χώρες, διαιωνίζοντας τοιούτοτρόπως την περιβαλλοντική ανισορροπία που ήδη επέφερε η αποικιακή εκμετάλλευση και η μαζική μόλυνση αυτών των κρατών (Cho, 2020; Taddeo et al., 2021).

Ο έκτος και τελευταίος κίνδυνος αφορά την καλή διακυβέρνηση ή μάλλον την έλλειψή της, σε ζητήματα Τεχνητής Νοημοσύνης και ειδικότερα σε ζητήματα Τεχνητής Νοημοσύνης για κοινωνικό όφελος. Αυτός θα μπορούσε να θεωρηθεί ως ένας δευτερεύων κίνδυνος, με πιθανότητα μεγέθυνσης άλλων κινδύνων και προκλήσεων. Ο μεγάλος αριθμός έργων Τεχνητής Νοημοσύνης για το κοινωνικό όφελος σε όλο τον κόσμο και ο πειραματισμός με ταχείς ρυθμούς με νέες εφαρμογές της Τεχνητής Νοημοσύνης δημιουργεί τον κίνδυνο επιβλαβείς πρακτικές να διαφύγουν της προσοχής, με δυσμενείς συνέπειες σε διάφορους πληθυσμούς αλλά και στο περιβάλλον (Cowls et al., 2021b). Η παρακάτω ενότητα πραγματεύεται πτυχές του πολύπλευρου ζητήματος της καλής διακυβέρνησης της Τεχνητής Νοημοσύνης και οδηγεί στην πρόταση του άρθρου για ένα συνεκτικό πλαίσιο βελτίωσης των προσπαθειών αξιοποίησης της Τεχνητής Νοημοσύνης για κοινωνικά επωφελή αποτελέσματα.

*“Μεταξύ 2012 και 2018 η υπολογιστική ισχύς για τον προγραμματισμό μεγάλων συστημάτων Τεχνητής Νοημοσύνης διπλασιάζεται κάθε 3,4 μήνες. Αυτό σημαίνει πως αν η ενεργειακή αποδοτικότητα αυτών των συστημάτων και της σχετικής υποδομής δεν συμβαδίσουν με την αύξηση της ζήτησης για υπολογιστική ισχύ, το αποτύπωμα άνθρακα για την έρευνα και ανάπτυξη της Τεχνητής Νοημοσύνης θα αυξηθεί ραγδαία.”*

## Κενά διακυβέρνησης και η παγίδα της μονομέρειας

Από την άφιξη της βαθιάς μάθησης, μίας μορφής τεχνητής νοημοσύνης, το 2012 η έρευνα και η ανάπτυξη της Τεχνητής Νοημοσύνης γνώρισαν μία νέα άνθηση τόσο σε σχέση με την ποσότητα των εφαρμογών αυτής της τεχνολογίας, όσο και ως προς το ενδιαφέρον που συγκέντρωσε από πλευράς κυβερνήσεων, ακαδημαϊκών και μελών της κοινωνίας των πολιτών (Perrault et al., 2019). Εμπορικά και μη έργα που στοχεύουν στην αξιοποίηση της Τεχνητής Νοημοσύνης για την παραγωγή κοινωνικά επωφελών αποτελεσμάτων αυξήθηκαν επίσης με ταχείς ρυθμούς (McKinsey, 2018; Vinuesa et al., 2020). Πρόκειται για έργα που χρησιμοποιούν την υφιστάμενη τεχνολογία για την παραγωγή κοινωνικά επωφελών αποτελεσμάτων, τα οποία ήταν προηγουμένως «αδύνατα, υπερβολικά ακριβά ή απλώς λιγότερο εφικτά από άποψης αποδοτικότητας και αποτελεσματικότητας» (Cowls et al., 2021a).

Πράγματι η Τεχνητή Νοημοσύνη έχει χρησιμοποιηθεί για τη μείωση της κατανάλωσης ενέργειας σε κέντρα δεδομένων, για την παραγωγή αξιόπιστων προβλέψεων σχετικά με την εξάπλωση πυρκαγιών ή την τήξη των πάγων της Αρκτικής με τη χρήση δορυφορικών εικόνων, ακόμη και την παροχή πρόσβασης σε υπηρεσίες υγείας και στοιχειώδεις διαγνώσεις μέσω κινητού τηλεφώνου για συγκεκριμένες περιοχές του κόσμου (Cowls et al., 2021a, 2021b). Εντούτοις, η κατανόηση της Τεχνητής Νοημοσύνης για το κοινωνικό όφελος ως παγκόσμιο φαινόμενο παραμένει ανεπαρκής, ενώ απουσιάζει και ένα συγκροτημένο πλαίσιο αξιολόγησης τέτοιων προσπαθειών. Αυτό αποτελεί ζήτημα διακυβέρνησης το οποίο απαιτεί επαρκή κατανόηση των προαναφερθέντων περιορισμών και κινδύνων της Τεχνητής Νοημοσύνης, αλλά και του τύπου της σκέψης που συχνά καθοδηγεί λανθασμένα πολλά έργα αξιοποίησης της Τεχνητής Νοημοσύνης. Προϋποθέτει επίσης διεθνή συντονισμό στη βάση επαρκώς ορισμένων και ρεαλιστικών στόχων.

Η Τεχνητή Νοημοσύνη για κοινωνικό όφελος έχει μέχρι στιγμής αναπτυχθεί σε μεμονωμένες περιπτώσεις και έχουν αναλυθεί μόνο συγκεκριμένοι τομείς εφαρμογής της (Vinuesa et al., 2020). Όπως εξηγήσαμε στο Cowls et al., 2021a:

*«Η προσέγγιση αυτή μπορεί να καταδεικνύει την παρουσία ενός φαινομένου, δεν μπορεί όμως να εξηγήσει τι καθιστά την Τεχνητή Νοημοσύνη κοινωνικά επωφελή, ούτε μπορεί να καταδείξει πώς [η κοινωνικά επωφελής Τεχνητή Νοημοσύνη] μπορεί και θα πρέπει να σχεδιαστεί και να εφαρμοστεί για την πλήρη αξιοποίηση αυτής της τεχνολογίας».*

Με τη σειρά τους, τα ζητήματα αυτά αναδεικνύουν κινδύνους σχετικά με την αναπάντεχη κατάρρευση έργων, την παράβλεψη ευκαιριών, αδικαιολόγητες παρεμβάσεις και εκπτώσεις ηθικής (ethical dumping), με τις τελευταίες να αναφέρονται σε τάσεις εκμετάλλευσης μη ηθικών ερευνητικών πρακτικών σε χώρες με λιγότερους κανονισμούς ερευνητικής προστασίας ή μικρότερες δυνατότητες παρακολούθησης των παραβιάσεων και επιβολής των κανονισμών σε σύγκριση με πλουσιότερες χώρες (Floridi, 2019).

Η προσπάθεια εκμετάλλευσης του δυναμικού της Τεχνητής Νοημοσύνης από τη διεθνή κοινότητα για την επίλυση ορισμένων από τα πιο επείγοντα ζητήματα δίχως ένα αξιόπιστο πλαίσιο διακυβέρνησης μπορεί να αλλοιώσει την αξιολόγηση ρεαλιστικών και αξιόλογων προσφερόμενων ευκαιριών. Οι προαναφερθέντες κίνδυνοι συχνά συνοδεύονται από τη λεγόμενη «τεχνολογική μονομέρεια» ή «τη μονομέρεια της Τεχνητής Νοημοσύνης». Για να παραφράσουμε τον Evgeny Morozov (Morozov, 2013) ο οποίος καθιέρωσε την ιδέα της

*“Η προσπάθεια εκμετάλλευσης του δυναμικού της Τεχνητής Νοημοσύνης από τη διεθνή κοινότητα για την επίλυση ορισμένων από τα πιο επείγοντα ζητήματα δίχως ένα αξιόπιστο πλαίσιο διακυβέρνησης μπορεί να αλλοιώσει την αξιολόγηση ρεαλιστικών και αξιόλογων προσφερόμενων ευκαιριών.”*

«τεχνολογική μονομέρειας», η μονομέρεια μπορεί να κατανοηθεί ως μία «διανοητική παθολογία» η οποία παρουσιάζει

«όλες τις περίπλοκες κοινωνικές καταστάσεις είτε ως επαρκώς ορισμένα προβλήματα με συγκεκριμένες υπολογιστικές λύσεις, είτε ως διαφανείς και αυτοπροσδιοριζόμενες διαδικασίες που μπορούν εύκολα να βελτιστοποιηθούν αν υπάρχουν οι κατάλληλοι αλγόριθμοι!»

Αυτό οδηγεί τους ανθρώπους στην αναζήτηση γνωστών λύσεων (στην περίπτωση της Τεχνητής Νοημοσύνης) για την αντιμετώπιση οποιουδήποτε προβλήματος ανεξαρτήτως συνθηκών. Μία ακόμη τάση των ανθρώπων και των ομάδων που υιοθετούν αυτή τη νοοτροπία είναι «η αποφυγή ενασχόλησης με μη επιλύσιμα προβλήματα» ή η λανθασμένη κατηγοριοποίησή τους ως επιλύσιμα και η κατά συνέπεια «πρόκληση κυμάτων ακούσιων συνεπειών» (Morozov, 2013).

*“Παγκόσμιες  
υπαρξιακές  
απειλές, όπως η  
κλιματική αλλαγή  
και οι πανδημίες,  
απαιτούν  
φιλόδοξες, ακόμη  
και ριζοσπαστικές  
απαντήσεις από τη  
διεθνή κοινότητα.”*

Οι προκλήσεις αυτές δεν συνιστούν απώλειες ζώων και δεν προκαλούν απαραίτητα βλάβες. Εντούτοις, είναι σημαντικό να τις αναδείξουμε στις συζητήσεις για την Τεχνητή Νοημοσύνη για το κοινωνικό όφελος και την καλή διακυβέρνηση της Τεχνητής Νοημοσύνης, έστω και αυτό περιορίζεται στην ενημέρωση όσο το δυνατόν περισσότερων ατόμων (ειδικότερα διαμορφωτών πολιτικής) σχετικά με τη δυσκολία επίτευξης της κατάλληλης ισορροπίας. Αξίζει να σημειωθεί πως μελέτες έδειξαν πως στα περισσότερα κορυφαία ερευνητικά κείμενα σχετικά με την Τεχνητή Νοημοσύνη «οι εκτιμήσεις περί αρνητικών συνεπειών είναι εξαιρετικά σπάνιες» (Birhane et al., 2021).

Παγκόσμιες υπαρξιακές απειλές, όπως η κλιματική αλλαγή και οι πανδημίες, απαιτούν φιλόδοξες, ακόμη και ριζοσπαστικές απαντήσεις από τη διεθνή κοινότητα. Εντούτοις, ούτε η κλίμακα αυτών των απειλών, ούτε η αισιόδοξη, αν όχι αφελής, οπτική του δυναμικού της Τεχνητής Νοημοσύνης δεν θα πρέπει να καθιστά αόρατη τη σημασία των υφιστάμενων και υπό εξέλιξη κανόνων και των κοινωνικών προσδοκιών σχετικά με τη χρήση τεχνολογιών πιθανού υψηλού κινδύνου (Cowls et al., 2021b).

## Κοινό και εύφορο έδαφος: Στόχοι Βιώσιμης Ανάπτυξης και Τεχνητή Νοημοσύνη

Παρά τους προαναφερθέντες περιορισμούς και κινδύνους, η απόρριψη του δυναμικού της Τεχνητής Νοημοσύνης για κοινωνικά επωφελή αποτελέσματα και ως τεχνολογία που μπορεί να βοηθήσει στον συντονισμό των διεθνών προσπαθειών επίλυσης παγκόσμιων προβλημάτων θα ήταν εξίσου αντιπαραγωγική. Τα ευρήματα της διετούς μας έρευνας σχετικά με την Τεχνητή Νοημοσύνη για το κοινωνικό όφελος (Cowls et al., 2021a) ξεκινούν με την πρόταση για έναν αξιόπιστο ορισμό του τι μπορεί να θεωρηθεί Τεχνητή Νοημοσύνη για το κοινωνικό όφελος. Η ενότητα υποστηρίζει επίσης πως οι 17 Στόχοι Βιώσιμης Ανάπτυξης του ΟΗΕ προσφέρουν ένα ισχυρό σημείο αναφοράς για την αξιολόγηση και την παρακολούθηση από τη διεθνή κοινότητα έργων με έμφαση στην Τεχνητή Νοημοσύνη και την παραγωγή κοινωνικά επωφελών αποτελεσμάτων.

Όπως υποστηρίξαμε με τους συναδέλφους μου (Cowls et al., 2021a), η Τεχνητή Νοημοσύνη για το κοινωνικό καλό μπορεί να αναλυθεί επί τη βάση των αποτελεσμάτων της. Ένα έργο μπορεί να κριθεί επιτυχημένο στον βαθμό που καταφέρνει να βοηθήσει στην άμβλυνση ή ακόμη και στην επίλυση ενός συγκεκριμένου ζητήματος δίχως να

“Το επόμενο βήμα προς την εγκαθίδρυση ενός σταθερού πλαισίου και σημείου αναφοράς για την αξιολόγηση της Τεχνητής Νοημοσύνης για το κοινωνικό όφελος αποτελείται από «την αποτελεσματική αναγνώριση των προβλημάτων που θεωρείται ότι επηρεάζουν αρνητικά την ανθρώπινη ζωή ή το φυσικό περιβάλλον».”

δημιουργεί νέα ή να επιδεινώνει υφιστάμενα. Κατά συνέπεια, η Τεχνητή Νοημοσύνη για το κοινωνικό καλό ορίζεται επίσημα (Cowls et al., 2021a) ως «ο σχεδιασμός, η ανάπτυξη και η εφαρμογή συστημάτων Τεχνητής Νοημοσύνης με τρόπους που να επιτρέπουν:

- i) την πρόληψη, άμβλυνση ή/και επίλυση προβλημάτων που επηρεάζουν αρνητικά την ανθρώπινη ζωή ή/και το φυσικό περιβάλλον ή/και
- ii) κοινωνικά προτιμητέες ή περιβαλλοντικά βιώσιμες λύσεις,
- iii) χωρίς την πρόκληση νέων βλαβών ή την επιδείνωση υφιστάμενων ανισοτήτων.

Το επόμενο βήμα προς την εγκαθίδρυση ενός σταθερού πλαισίου και σημείου αναφοράς για την αξιολόγηση της Τεχνητής Νοημοσύνης για το κοινωνικό όφελος αποτελείται από «την αποτελεσματική αναγνώριση των προβλημάτων που θεωρείται ότι επηρεάζουν αρνητικά την ανθρώπινη ζωή ή το φυσικό περιβάλλον» (Cowls et al., 2021a). Για διάφορους λόγους οι 17 Στόχοι Βιώσιμης Ανάπτυξης του ΟΗΕ παρέχουν εύφορο έδαφος για κάτι τέτοιο. Ίσως ο σημαντικότερος παράγοντας είναι ότι οι στόχοι αυτοί υιοθετούνται από 193 χώρες και απολαμβάνουν σχεδόν καθολική αναγνώριση.



Γράφημα 1. Οι 17 Στόχοι Βιώσιμης Ανάπτυξης του ΟΗΕ.

Οι Στόχοι Βιώσιμης Ανάπτυξης υπογράφηκαν από τη διεθνή κοινότητα με σκοπό να παρέχουν κοινούς στόχους που θα τέμνουν οριζοντίως την οικονομική, κοινωνική και περιβαλλοντική διάσταση της βιώσιμης ανάπτυξης. Λόγω τους εύρους και της αποδοχής αυτών των στόχων, έργα που αποσκοπούν στην αξιοποίηση της τεχνητής νοημοσύνης για κοινωνικά επωφελή αποτελέσματα, όπως και οι ερευνητές που τα παρακολουθούν, μπορούν να χρησιμοποιήσουν τους Στόχους Βιώσιμης Ανάπτυξης ως αφετηρία και σημεία αναφοράς ως προς την αξιολόγηση.

## Ανάλυση

Οι Στόχοι Βιώσιμης Ανάπτυξης αποτελούν επαρκώς εμπειρικά και μη αμφιλεγόμενα σημεία αναφοράς για την αξιολόγηση της θετικής κοινωνικής επίδρασης της Τεχνητής Νοημοσύνης για το κοινωνικό όφελος παγκοσμίως. Η κίνηση χαρακτηρισμού της «Τεχνητής Νοημοσύνης για κοινωνικό όφελος» ως Τεχνητής Νοημοσύνης που υποστηρίζει τους Στόχους Βιώσιμης Ανάπτυξης (AIxSDGs) μπορεί να φαίνεται περιοριστική δεδομένου πως υπάρχουν αναμφίβολα πολλαπλά παραδείγματα κοινωνικά επωφελών χρήσεων της Τεχνητής Νοημοσύνης εκτός του πλαισίου των Στόχων Βιώσιμης Ανάπτυξης. Εντούτοις, η προσέγγιση αυτή συνοδεύεται από σημαντικά πλεονεκτήματα, μεταξύ των οποίων



ξεχωρίζουν πέντε.

Πρώτον, οι Στόχοι Βιώσιμης Ανάπτυξης παρέχουν ξεκάθαρα, επαρκώς ορισμένα και κοινοποιήσιμα όρια ως προς τη θετική αναγνώριση του τι συνιστά κοινωνικά επωφελή Τεχνητή Νοημοσύνη (τι θα πρέπει να γίνει σε αντίθεση με το τι θα πρέπει να αποφευχθεί), αν και δεν θα πρέπει να θεωρηθεί πως καταδεικνύουν τι δεν συνιστά κοινωνικά επωφελή Τεχνητή Νοημοσύνη.

Δεύτερον, οι Στόχοι Βιώσιμης Ανάπτυξης αποτελούν διεθνείς συμφωνημένους στόχους για την ανάπτυξη και καθοδηγούν τον σχεδιασμό σχετικών πολιτικών παγκοσμίως, εγείρουν κατά συνέπεια λιγότερα ερωτήματα σχετικά με τη σχετικότητα και πολιτιστική εξάρτηση των αξιών. Παρότι διαθέτουν προφανώς περιθώρια βελτίωσης, αποτελούν ό,τι κοντινότερο διαθέτουμε σε μία πανανθρώπινη συναίνεση για το τι θα πρέπει να γίνει για την προώθηση της θετικής κοινωνικής αλλαγής και την προστασία του φυσικού μας περιβάλλοντος.

*“...οι Στόχοι Βιώσιμης Ανάπτυξης παρέχουν ξεκάθαρα, επαρκώς ορισμένα και κοινοποιήσιμα όρια ως προς τη θετική αναγνώριση του τι συνιστά κοινωνικά επωφελή Τεχνητή Νοημοσύνη.”*

Τρίτον, το υφιστάμενο σώμα έρευνας σχετικά με τους Στόχους Βιώσιμης Ανάπτυξης περιλαμβάνει ήδη μελέτες και μεθόδους μέτρησης σχετικά με τον υπολογισμό της προόδου επίτευξης εκάστου εκ των 17 Στόχων Βιώσιμης Ανάπτυξης και των 169 συνδεδεμένων στόχων, όπως αυτοί ορίστηκαν στην Ατζέντα 2030 για τη Βιώσιμη Ανάπτυξη. Αυτές οι μέθοδοι μπορούν να χρησιμοποιηθούν για την αξιολόγηση της επίδρασης συνδυαστικών έργων Τεχνητής Νοημοσύνης και Στόχων Βιώσιμης Ανάπτυξης (Vinuesa et al., 2020).

Τέταρτον, η επικέντρωση στην επίδραση έργων Τεχνητής Νοημοσύνης σε διαφορετικούς Στόχους Παγκόσμιας Ανάπτυξης μπορεί να βελτιώσει της υφιστάμενες και να δημιουργήσει νέες συνέργειες μεταξύ έργων που ασχολούνται με διαφορετικούς Στόχους, αξιοποιώντας ακόμα περισσότερο την Τεχνητή Νοημοσύνη για την απόκτηση πληροφοριών από μεγάλα και ποικίλα σύνολα δεδομένων, και μπορεί να οδηγήσει σε πιο φιλόδοξα συνεργατικά έργα.

Πέμπτον, η κατανόηση της Τεχνητής Νοημοσύνης για κοινωνικό όφελος με όρους συνδυασμού Τεχνητής Νοημοσύνης και Στόχων Βιώσιμης Ανάπτυξης επιτρέπει καλύτερο σχεδιασμό και κατανομή πόρων από τη στιγμή που θα γίνει ξεκάθαρο ποιοι Στόχοι χρήζουν περισσότερης προσοχής και γιατί.

Ενόψει των πλεονεκτημάτων της χρήσης των Στόχων Βιώσιμης Ανάπτυξης του ΟΗΕ ως σημείου αναφοράς για την αξιολόγηση της Τεχνητής Νοημοσύνης για κοινωνικό όφελος, διεξαγάγαμε μία διεθνή έρευνα συνδυαστικών έργων Τεχνητής Νοημοσύνης και Στόχων Βιώσιμης Ανάπτυξης. Η έρευνα διεξήχθη μεταξύ Ιουλίου 2018 και Νοεμβρίου 2020 και αφορούσε στη συλλογή δεδομένων για συνδυαστικά έργα Τεχνητής Νοημοσύνης και Στόχων Βιώσιμης Ανάπτυξης που πληρούσαν τα παρακάτω πέντε κριτήρια:

i) Μόνο έργα που όντως ασχολούνταν με τουλάχιστον ένα από τους 17 Στόχους Βιώσιμης Ανάπτυξης.

ii) Μόνο εφαρμοσμένα και συγκεκριμένα έργα που βασίζονταν σε κάποια μορφή Τεχνητής Νοημοσύνης (συμβολική Τεχνητή Νοημοσύνη, νευρικά δίκτυα, μηχανική μάθηση, έξυπνα ρομπότ, επεξεργασία φυσικής γλώσσας κ.ο.κ.) και όχι απλά με αναφορά στην Τεχνητή Νοημοσύνη (ένα γενικό πρόβλημα μεταξύ νεοφυών επιχειρήσεων τεχνητής νοημοσύνης).

iii) Μόνο έργα που δημιουργήθηκαν και χρησιμοποιήθηκαν στην πράξη για τουλάχιστον έξι μήνες, σε αντίθεση με θεωρητικά έργα ή ερευνητικά έργα που είναι ακόμα υπό εξέλιξη

(για παράδειγμα ευρεσιτεχνίες ή επιχορηγούμενα προγράμματα).

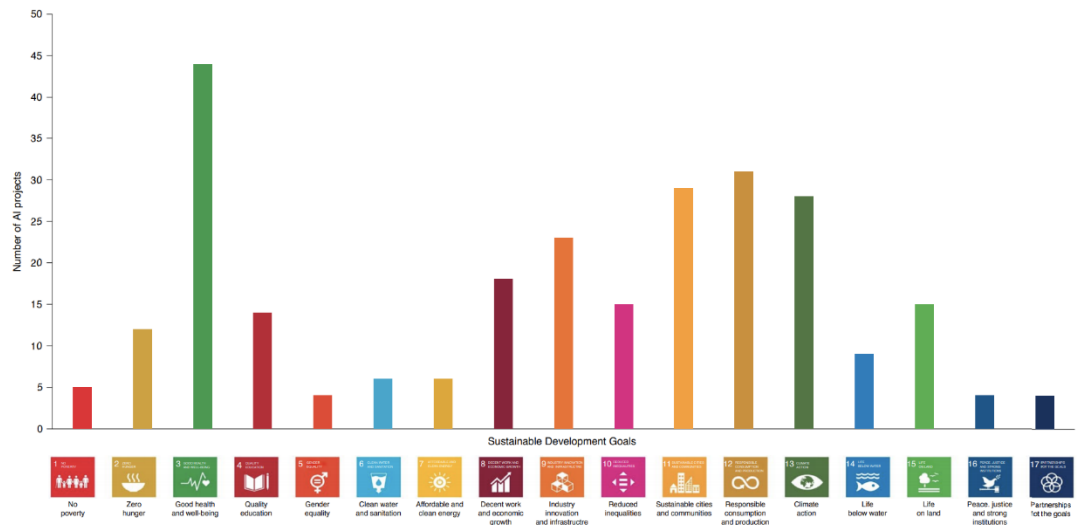
iv) Μόνο έργα με καταγεγραμμένη θετική επίδραση, για παράδειγμα μέσω ιστοσελίδων, άρθρων στον Τύπο, επιστημονικών άρθρων, εκθέσεων ΜΚΟ κ.ο.κ.

v) Μόνο έργα με καμία ή ελάχιστες ενδείξεις αρνητικών επιδράσεων.

Η έρευνα κατέληξε σε 108 έργα στα Αγγλικά, τα Ισπανικά και τα Γαλλικά που να πληρούν τα παραπάνω κριτήρια. Τα δεδομένα σχετικά με συνδυαστικά έργα Τεχνητής Νοημοσύνης και Στόχων Βιώσιμης Ανάπτυξης που συλλέχθηκαν από αυτήν την έρευνα είναι διαθέσιμα στη σχετική βάση δεδομένων (<https://www.aiforsdgs.org/allprojects>), η οποία ανήκει στο Oxford Research Initiative on Sustainable Development Goals and Artificial Intelligence. Μια προκαταρκτική εκδοχή της ανάλυσής μας παρουσιάστηκε τον Σεπτέμβριο 2019 σε παράπλευρη εκδήλωση κατά τη διάρκεια της ετήσιας Γενικής Συνέλευσης των Ηνωμένων Εθνών.

Η ανάλυσή μας (βλ. παρακάτω) καταδεικνύει πως κάθε Στόχος Βιώσιμης Ανάπτυξης αντιστοιχεί σε τουλάχιστον ένα έργο Τεχνητής Νοημοσύνης. Η ανάλυση καταδεικνύει επίσης πως ο συνδυασμός Τεχνητής Νοημοσύνης και Στόχων Βιώσιμης Ανάπτυξης αποτελεί ένα αυξανόμενο παγκόσμιο φαινόμενο, με έργα να εκτελούνται σε πέντε ηπείρους, αλλά και πως το φαινόμενο αυτό δεν αφορά εξίσου όλους τους Στόχους Βιώσιμης Ανάπτυξης (Σχ. 2). Ο Στόχος 3 («Καλή υγεία και ευζωία») βρίσκεται στην πρώτη θέση, με τους στόχους 5 («Ισότητα των φύλων»), 16 («Ειρήνη, Δικαιοσύνη και Ισχυροί Θεσμοί») και 17 («Συνέργειες για την επίτευξη των Στόχων») να εντάσσονται σε λιγότερα από πέντε έργα (Σχ. 2). Είναι σημαντικό να σημειωθεί πως η χρήση Τεχνητής Νοημοσύνης για την επίτευξη τουλάχιστον ενός από τους Στόχους Βιώσιμης Ανάπτυξης δεν συνεπάγεται απαραίτητα επιτυχές αποτέλεσμα. Παρατηρούμε επίσης πως ένα έργο μπορεί να καλύπτει ταυτόχρονα πολλαπλούς Στόχους Βιώσιμης Ανάπτυξης ή σε διαφορετικές χρονικές περιόδους και με διαφορετικούς τρόπους. Επιπλέον, ακόμη και η πλήρης επιτυχία ενός έργου δύσκολα να συνεπαγόταν την εξάλειψη όλων των σχετικών με τον εκάστοτε Στόχο Βιώσιμης Ανάπτυξης προκλήσεων. Αυτό συμβαίνει κυρίως λόγω του γεγονότος πως κάθε Στόχος Βιώσιμης Ανάπτυξης έχει να κάνει με εδραιωμένες προκλήσεις ευρύτερης και δομικής φύσεως. Η διαπίστωση αυτή αντικατοπτρίζεται στον τρόπο οργάνωσης των Στόχων Βιώσιμης Ανάπτυξης: και οι 17 Στόχοι περιλαμβάνουν πολλαπλούς δευτερεύοντες στόχους, οι οποίοι με τη σειρά τους διαθέτουν περισσότερους από έναν δείκτες επιτυχίας. Η έρευνα καταδεικνύει για ποιους Στόχους Βιώσιμης Ανάπτυξης γίνεται χρήση Τεχνητής Νοημοσύνης σε υψηλό επίπεδο, απαιτείται όμως λεπτομερέστερη ανάλυση για την αξιολόγηση του εύρους της θετικής επίδρασης της χρήσης Τεχνητής Νοημοσύνης σε σχέση με συγκεκριμένους δείκτες Στόχων Βιώσιμης Ανάπτυξης, καθώς και πιθανές ακούσιες συνέπειες.

“...η κατανόηση της Τεχνητής Νοημοσύνης για κοινωνικό όφελος με όρους συνδυασμού Τεχνητής Νοημοσύνης και Στόχων Βιώσιμης Ανάπτυξης επιτρέπει καλύτερο σχεδιασμό και κατανομή πόρων από τη στιγμή που θα γίνει ξεκάθαρο ποιοι Στόχοι χρήζουν περισσότερης προσοχής και γιατί.”



Γράφημα 2 Έργα που αφορούν τους ΣΒΑ. Δείγμα έρευνας σε 108 έργα τεχνητής νοημοσύνης παγκοσμίως

Με βάση την ανάλυση, τα 108 εγχειρήματα που πληρούν τα κριτήρια βρίσκουν αντιστοιχία σε επτά βασικούς παράγοντες για Τεχνητή Νοημοσύνη για το κοινωνικό όφελος, όπως έχουν εντοπιστεί σε πρόσφατες έρευνες (Floridi et al., 2020): διαψευσιμότητα και σταδιακή ανάπτυξη, διασφάλιση έναντι ενδεχόμενου χειρισμού στις μεταβλητές πρόβλεψης, παρεμβατικότητα και εξήγηση ενημερωμένη από μεριάς των δεκτών, διαφάνεια όσον αφορά τους σκοπούς, προστασία προσωπικών δεδομένων και συγκατάθεση των ατόμων που προσφέρουν δεδομένα, δίκαιη διαχείριση της εκάστοτε κατάστασης και σημασιοποίηση υπό όρους φιλικούς προς τον άνθρωπο. Κάθε παράγοντας αντιστοιχεί τουλάχιστον σε μια από τις ηθικές γραμμές της Τεχνητής Νοημοσύνης – δράση με ευεργετική πρόθεση, δράση με μη κακόβουλη πρόθεση, δικαιοσύνη, αυτονομία, εξηγησιμότητα – όπως προκύπτουν από την συγκριτική ανάλυση των Floridi και Cowls(2019). Η συνέπεια μεταξύ παραγόντων και ηθικών γραμμών είναι καθοριστικής σημασίας: Τεχνητή Νοημοσύνη που στηρίζει Στόχους Βιώσιμης Ανάπτυξης δεν μπορεί να μην συνάδει με τα ηθικά πλαίσια που οδηγούν τη δημιουργία και αξιολόγηση κάθε είδους Τεχνητής Νοημοσύνης.

## Συμπεράσματα

Σκοπός του παρόντος άρθρου είναι να αντικρούσει εξίσου το κινδυνολογικό αφήγημα για την ανάπτυξη της Τεχνητής Νοημοσύνης ως δυστοπίας αλλά και την παρουσίασή της ως πανάκειας για όλα τα παγκόσμια προβλήματα. Αμφότερα τα παραπάνω αφηγήματα ανακυκλώνονται στους τίτλους ειδήσεων ή προωθούνται από φορείς με ιδιοτελή ατζέντα και βλάπτουν πολύ περισσότερο από ό,τι ωφελούν.

Εν κατακλείδι, η Τεχνητή Νοημοσύνη δεν αποτελεί λύση από μόνη της, η τεχνολογική μονομέρεια εξυπηρετεί μόνο όσους αναπτύσσουν τέτοια συστήματα. Πράγματι, τα συστήματα Τεχνητής Νοημοσύνης μπορούν να βοηθήσουν σημαντικά τις προσπάθειες επίλυσης διασυνδεδεμένων και περίπλοκων παγκόσμιων ζητημάτων. Επίσης τα συστήματα Τεχνητής Νοημοσύνης θέτουν πράγματι ζητήματα ηθικής, η σοβαρότητα των οποίων εξαρτάται από το πλαίσιο εφαρμογής τους. Τα παραπάνω όμως δεν σημαίνουν πως πρέπει να διακόψουμε κάθε μορφή έρευνας και ανάπτυξης γύρω από την Τεχνητή Νοημοσύνη. Οι Στόχοι Βιώσιμης Ανάπτυξης του ΟΗΕ παρέχουν ένα σταθερό πλαίσιο για κάθε έργο που αποσκοπεί στην προώθηση του κοινωνικού οφέλους μέσω της αξιοποίησης της Τεχνητής Νοημοσύνης. Το πλαίσιο αυτό απολαμβάνει σχεδόν καθολική

αποδοχή χάρη στη στήριξη που παρείχαν 193 κράτη, και λειτουργεί καταλυτικά για την περαιτέρω συνεργασία και μεγιστοποίηση της θετικής επίδρασης έργων Τεχνητής Νοημοσύνης.

## Βιβλιογραφία

AI HLEG (2019) *A Definition of AI: Main Capabilities and Disciplines*. Available at:

[https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=56341](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341).

Amodei, D. and Hernandez, D. (2018) *AI and Compute, OpenAI*. Available at: <https://openai.com/blog/ai-and-compute/> (Accessed: 31 August 2020).

Ananny, M. and Crawford, K. (2018) 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability', *New Media & Society*, 20(3), pp. 973–989. doi: 10.1177/1461444816676645.

Benjamin, R. (2019) *Race after technology: abolitionist tools for the new Jim code*. Medford, MA: Polity.

Birhane, A. et al. (2021) 'The Values Encoded in Machine Learning Research', *arXiv:2106.15590 [cs]*. Available at: <http://arxiv.org/abs/2106.15590> (Accessed: 8 July 2021).

Cho, R. (2020) 'Why Climate Change is an Environmental Justice Issue', *State of the Planet*, 22 September. Available at: <https://blogs.ei.columbia.edu/2020/09/22/climate-change-environmental-justice/> (Accessed: 14 April 2021).

Cowls, J. et al. (2021a) 'A definition, benchmark and database of AI for social good initiatives', *Nature Machine Intelligence*, 3(2), pp. 111–115. doi: 10.1038/s42256-021-00296-0.

Cowls, J. et al. (2021b) 'The AI Gambit — Leveraging Artificial Intelligence to Combat Climate Change: Opportunities, Challenges, and Recommendations', *SSRN Electronic Journal*. doi: 10.2139/ssrn.3804983.

Eubanks, V. (2017) *Automating inequality: how high-tech tools profile, police, and punish the poor*. First Edition. New York, NY: St. Martin's Press.

Fink, K. (2018) 'Opening the government's black boxes: freedom of information and algorithmic accountability', *Information, Communication & Society*, 21(10), pp. 1453–1471. doi: 10.1080/1369118X.2017.1330418.

Floridi, L. (2019) 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical', *Philosophy & Technology*, 32(2), pp. 185–193. doi: 10.1007/s13347-019-00354-x.

Floridi, L. et al. (2020) 'How to Design AI for Social Good: Seven Essential Factors', *Science and Engineering Ethics*, 26(3), pp. 1771–1796. doi: 10.1007/s11948-020-00213-5.



Floridi, L. and Cowls, J. (2019) 'A Unified Framework of Five Principles for AI in Society', *Harvard Data Science Review*. doi: 10.1162/99608f92.8cd550d1.

Friend, T. (2018) *How Frightened Should We Be of A.I.? | The New Yorker*. Available at: <https://www.newyorker.com/magazine/2018/05/14/how-frightened-should-we-be-of-ai> (Accessed: 8 July 2021).

McKinsey (2018) *Marrying artificial intelligence and the sustainable development goals: The global economic impact of AI | McKinsey*. Available at: <https://www.mckinsey.com/mgi/overview/in-the-news/marrying-artificial-intelligence-and-the-sustainable> (Accessed: 8 July 2021).

Morozov, E. (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*.

Obermeyer, Z. *et al.* (2019) 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science*, 366(6464), pp. 447–453. doi: 10.1126/science.aax2342.

Perrault, R. *et al.* (2019) *Artificial Intelligence Index Report 2019*.

Ram, A. (2019) 'Europe's AI start-ups often do not use AI, study finds', 5 March. Available at: <https://www.ft.com/content/21b19010-3e9f-11e9-b896-fe36ec32aece>.

Roberts, H. *et al.* (2021) *Achieving a 'Good AI Society': comparing the aims and progress of the EU and the US*. SSRN Scholarly Paper ID 3851523. Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.3851523.

Taddeo, M. *et al.* (2021) 'Artificial intelligence and the climate emergency: Opportunities, challenges, and recommendations', *One Earth*, 4(6), pp. 776–779. doi: 10.1016/j.oneear.2021.05.018.

Taylor, L., Floridi, L. and van der Sloot, B. (eds) (Forthcoming) *Group Privacy: New Challenges of Data Technologies*. Hildenberg: Philosophical Studies, Book Series, Springer.

Tsamados, A. *et al.* (2021) 'The ethics of algorithms: key problems and solutions', *AI & SOCIETY*. doi: 10.1007/s00146-021-01154-8.

Vinuesa, R. *et al.* (2020) 'The role of artificial intelligence in achieving the Sustainable Development Goals', *Nature Communications*, 11(1), p. 233. doi: 10.1038/s41467-019-14108-y.