



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ
Πρόγραμμα Μεταπτυχιακών Σπουδών
«ΔΙΚΑΙΟ ΚΑΙ ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ»
Ακαδημαϊκό έτος 2021-2022

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
της Σοφίας Αντωνοπούλου (Α.Μ.: ΜΔΙ2003)

ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΚΑΙ ΕΜΦΥΛΕΣ ΔΙΑΚΡΙΣΕΙΣ – ΜΙΑ
ΦΕΜΙΝΙΣΤΙΚΗ, ΤΕΧΝΟΛΟΓΙΚΗ, ΝΟΜΙΚΗ ΚΑΙ ΗΘΙΚΟΚΟΙΝΩΝΙΚΗ
ΠΡΟΣΕΓΓΙΣΗ

Επιβλέπουσα:

Δρ. Μαρίνα Μαρκέλλου

Πειραιάς, Ιούλιος 2022

ΕΥΧΑΡΙΣΤΙΕΣ

Στην επιλογή του θέματος «Τεχνητή Νοημοσύνη και Έμφυλες Διακρίσεις – Μια Φεμινιστική, Τεχνολογική, Νομική και Ηθικοκοινωνική Προσέγγιση» για τη διπλωματική μου εργασία στο Πρόγραμμα Μεταπτυχιακών Σπουδών «Δίκαιο και Τεχνολογίες Πληροφορικής και Επικοινωνιών» του Πανεπιστημίου Πειραιώς, συντέλεσε η ολιστική προσέγγιση διδασκαλίας της Δρ. Ελένης Ρεθυμνιωτάκη στο μάθημα «Κανονισμοί και οδηγίες στην Ευρωπαϊκή Ένωση.: Ηθική, Νομιμότητα, Αποδοχή» στο α' εξάμηνο του ανωτέρω ΠΜΣ, επομένως την ευχαριστώ ειλικρινά για το έναυσμα και την έμπνευση που μου προσέφερε.

Ευχαριστώ επίσης τη Ναταλία Ροζαλία Αυλώνα για τη γνωριμία μας και την εποικοδομητική συνδιαλλαγή μας. Μοιραζόμαστε κοινούς προβληματισμούς και ήταν πραγματικά πολύτιμη η συνδρομή και καθοδήγηση που μου παρείχε.

Η Αναστασία Καραγιάννη, συνιδρύτρια του DATAWO, την οποία προσέγγισα για να μου δώσει συνέντευξη στο πλαίσιο της παρούσας διπλωματικής εργασίας, συμφώνησε αμέσως να με υποστηρίξει, με αποτέλεσμα να κάνουμε μια πολύ ζεστή και οικεία συζήτηση. Χαίρομαι που τη γνώρισα και την ευχαριστώ.

Τέλος, θέλω να ευχαριστήσω από καρδιάς την, όχι μόνο επιβλέπουσα καθηγήτριά μου, αλλά και προσωπική μου μέντορα, Δρ. Μαρίνα Μαρκέλλου, η οποία μου μετέδωσε τον ερευνητικό ενθουσιασμό της, μου προσέφερε την καθοδήγησή της, πολύτιμες συμβουλές και αμέριστη στήριξη και ενθάρρυνση, ώστε να σκεφτώ περαιτέρω, να ανακαλύψω νέα μονοπάτια και να εξελιχθώ.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ	1
ΕΙΣΑΓΩΓΗ	2
ΚΕΦΑΛΑΙΟ 1^ο	6
ΟΡΙΣΜΟΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ – ΗΘΙΚΟΙ ΠΡΟΒΛΗΜΑΤΙΣΜΟΙ ΚΑΙ ΘΕΜΕΛΙΩΔΗ ΑΝΘΡΩΠΙΝΑ ΔΙΚΑΙΩΜΑΤΑ – ΠΡΟΣΠΑΘΕΙΕΣ ΡΥΘΜΙΣΗΣ	6
1.1 Εισαγωγικές Παρατηρήσεις.....	6
1.2 Ορισμός TN.....	7
1.2.1 Διακρίσεις TN.....	7
1.2.2 Προσπάθειες διατύπωσης ορισμού TN.....	9
1.3 Παραδείγματα εφαρμογών TN.....	13
1.4 Ηθικοί-κοινωνικοί προβληματισμοί και θεμελιώδη ανθρώπινα δικαιώματα.....	15
1.5 Προσπάθειες ρύθμισης.....	17
1.5.1 Προσπάθειες ρύθμισης σε διεθνές επίπεδο.....	17
1.5.2 Προσπάθειες ρύθμισης σε επίπεδο κρατών (ΗΠΑ-Κίνα).....	19
1.5.3 Προσπάθειες ρύθμισης στην ΕΕ.....	20
ΚΕΦΑΛΑΙΟ 2^ο	25
ΕΜΦΥΛΕΣ ΔΙΑΚΡΙΣΕΙΣ (GENDER BIASES) ΚΑΙ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ	25
2.1 Εισαγωγικές παρατηρήσεις.....	25
2.2 Διαφορά έννοιας διακρίσεων (discrimination) και μεροληψιών (biases).....	27
2.2.1 Η έννοια των διακρίσεων.....	27
2.2.2 Η έννοια της μεροληψίας (bias).....	28
2.3 Η μεροληψία στην TN (Biased AI).....	29
2.4 TN και Έμφυλες Διακρίσεις (Gender Biases).....	33

2.5 Το Παιχνίδι της Μίμησης (των φύλων)	34
2.6 Η φεμινιστική θεωρία για την τεχνολογία	37
2.6.1 Η πεσιμιστική ματιά.....	37
2.6.2 Ο κυβερνοφεμινισμός (<i>cyberfeminism</i>) – τεχνοφεμινισμός (<i>technofeminism</i>)	38
2.7 Τα έμφυλα στερεότυπα στη γλώσσα και το πρόβλημα της ενσωμάτωσής της στην TN (Word Embeddings-Natural Language Processing).....	40
2.8 Η θηλυκοποίηση της TN.....	42
2.8.1 Ψηφιακοί βοηθοί (<i>digital assistants</i>).....	42
2.8.2 Ενσώματες αναπαραστάσεις TN – Αναπαραστάσεις στην επιστημονική φαντασία (κινηματογράφος).....	45
2.9 Το πρόβλημα των δεδομένων στην TN – Υποεκπροσώπηση θηλυκού και μη-δυναδικού φύλου – Φεμινισμός των Δεδομένων (Data Feminism)	52
2.9.1 Τα μεγάλα δεδομένα (<i>big data</i>)	52
2.9.2 Η υποεκπροσώπηση του γυναικείου και μη-δυναδικού φύλου.....	53
2.9.3 Μηχανική όραση (<i>Computer vision</i>)	56
2.9.4 Φεμινισμός των Δεδομένων (<i>Data Feminism</i>).....	57
ΚΕΦΑΛΑΙΟ 3ο	61
ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΤΗΝ ΑΝΤΙΜΕΤΩΠΙΣΗ ΤΩΝ ΕΜΦΥΛΩΝ ΔΙΑΚΡΙΣΕΩΝ ΣΤΗΝ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ – ΠΡΟΣ ΜΙΑ ΗΘΙΚΗ, ΔΙΚΑΙΗ ΚΑΙ ΣΥΜΠΕΡΙΛΗΠΤΙΚΗ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ	61
3.1 Εισαγωγικές Παρατηρήσεις.....	61
3.2 Ενίσχυση της συμμετοχής των γυναικών και μη-δυναδικών ατόμων στο πεδίο της Επιστήμης, Τεχνολογίας, Μηχανικής και Μαθηματικών (STEM - Science, Technology, Engineering, Mathematics)	62
3.3 Συμπεριληπτικότητα (<i>inclusivity</i>) και ποικιλομορφία (<i>diversity</i>) στο χώρο της TN	63
3.4 Κινητοποίηση και ευαισθητοποίηση (<i>raising awareness</i>)	66

3.4.1 <i>Distributed Artificial Intelligence Research Institute (DAIR)</i>	66
3.4.2 <i>Algorithmic Justice League</i>	67
3.4.3 <i>DATAWO</i>	68
3.5 Ρυθμίσεις εκ του σχεδιασμού της TN (by design) – Διαφάνεια (transparency) – Debiasing	69
3.5.1 <i>Ενσωμάτωση ηθικών αρχών εκ του σχεδιασμού (by design) – Διαφάνεια (transparency)</i>	69
3.5.2 <i>Debiasing</i>	71
3.6 Προς μια ηθική TN – Τάσεις, διαδικασίες και προβληματισμοί περί ηθικής.....	73
3.6.1 <i>Ορθολογικές (rational) και σχεσιακές (relational) προσεγγίσεις</i>	73
3.6.2 <i>Top-down και bottom-up ηθική ρύθμιση</i>	75
3.6.3 <i>Ηθική αξιολόγηση (ethics assessment)</i>	77
ΕΠΙΛΟΓΟΣ – ΣΥΜΠΕΡΑΣΜΑ	80
ΒΙΒΛΙΟΓΡΑΦΙΑ	82

ΠΕΡΙΛΗΨΗ

Η Τεχνητή Νοημοσύνη αποτελούσε ανέκαθεν ένα αντιφατικό πεδίο, προσφέροντας σημαντικές ευκαιρίες για τον άνθρωπο, με την υπόσχεση διευκόλυνσης και εκσυγχρονισμού της καθημερινής ζωής. Ωστόσο, εγκυμονεί σημαντικούς κινδύνους για την κοινωνία και τα θεμελιώδη ανθρώπινα δικαιώματα, όπως είναι οι έμφυλες διακρίσεις (gender biases), οι οποίες απειλούν τις θεμελιώδεις αρχές της ισότητας και της απαγόρευσης των διακρίσεων, επηρεάζοντας τις γυναίκες, τα μη-δυναδικά άτομα και τη μαύρη κοινότητα. Ο όρος «έμφυλες διακρίσεις» (gender biases) στην Τεχνητή Νοημοσύνη έχει τεχνολογική, ηθικο-κοινωνική, και νομική σημασία. Περιλαμβάνει α) το αποτέλεσμα που στερεί από ένα στατιστικό αποτέλεσμα την αντιπροσωπευτικότητα, διαστρεβλώνοντάς το συστηματικά εξαιτίας του φύλου, β) τις προκαταλήψεις και τα στερεότυπα περί φύλου που υπάρχουν στην κοινωνία και εκφράζονται στις εφαρμογές Τεχνητής Νοημοσύνης, αλλά και γ) οποιαδήποτε κατάσταση στην οποία ένα άτομο αντιμετωπίζεται λιγότερο ευνοϊκά από ένα άλλο εξαιτίας του φύλου του. Οι έμφυλες διακρίσεις στην Τεχνητή Νοημοσύνη προκύπτουν ως αποτέλεσμα των ήδη υπαρχόντων στην ανθρώπινη κοινωνία έμφυλων στερεοτύπων, τα οποία ενσωματώνονται στην Τεχνητή Νοημοσύνη, και της κουλτούρας και κυριαρχίας των λευκών προνομιούχων ανδρών που έχει εδραιωθεί στο πεδίο. Άλλοι παράγοντες είναι η φυσική γλώσσα που χρησιμοποιείται για την εκπαίδευση των συστημάτων Τεχνητής Νοημοσύνης (Natural Language Processing – NLP), η γενικότερη τάση θηλυκοποίησης της Τεχνητής Νοημοσύνης, όπως εκφράζεται μέσω των ψηφιακών βοηθών, αλλά και των αναπαραστάσεων της στην επιστημονική φαντασία, δημιουργώντας διαστρεβλωμένη εικόνα για την πραγματική φύση της και τα μη αντιπροσωπευτικά δεδομένα και η υποεκπροσώπηση του γυναικείου και μη-δυναδικού φύλου. Η ανάδειξη του προβλήματος των έμφυλων διακρίσεων ως κοινωνικο-τεχνολογικό (socio-technical) ζήτημα είναι καίρια για τη συνειδητοποίηση της έκτασής του και την αντιμετώπισή του, με υιοθέτηση λύσεων που βασίζονται σε διεπιστημονική και πολυεπίπεδη προσέγγιση, για την επίτευξη μιας περισσότερο συμπεριληπτικής (inclusive), δίκαιης και ηθικής Τεχνητής Νοημοσύνης, προς όφελος όλων.

ΕΙΣΑΓΩΓΗ

Ήδη από την αρχαιότητα διαπιστώνεται να υπάρχει η σκέψη της δημιουργίας πλασμάτων, τα οποία δεν είναι άνθρωποι, αλλά φέρουν ανθρώπινες ιδιότητες, όπως η σκέψη, η ομιλία και η εμφάνιση. Η πρώτη αναφορά στην ιστορία του ανθρώπου σε αυτό που σήμερα ονομάζεται Τεχνητή Νοημοσύνη (εφεξής TN), ανατρέχει στην Αρχαία Ελλάδα και συγκεκριμένα στα ομηρικά έπη και τη μυθολογία. Στην «Ιλιάδα» του Ομήρου, στη Ραψωδία Σ στους στίχους 417-421 αναφέρεται *«Και ανάλαφρα τον κύριον εστηρίζαν θεράπαινες ολόχρυσες, σαν ζωντανά κοράσια. Δύναμιν έχουν και φωνήν, νουν έχουν εις τες φρένες, και τεχνουργήματ' έμαθαν από τους αθανάτους. Εκείνες τον επρόσεχαν.»* (Ομηρος, 2014). Πρόκειται για αναφορά σε πλάσματα κατασκευασμένα από μέταλλο, χρυσά στην όψη, με τη μορφή νεαρής γυναίκας, τα οποία μπορούσαν να σκεφτούν και να μιλήσουν. Ήταν δημιούργημα του θεού Ηφαίστου, ο οποίος τα κατασκεύασε με σκοπό να τον υπακούν και να τον βοηθούν στην καθημερινότητά του. Στον Ήφαιστο αποδίδεται επίσης (άλλες πηγές αναφέρουν ως δημιουργό το Δαίδαλο), σύμφωνα με τη μυθολογία και η δημιουργία ενός άλλου «πλάσματος»: ο Τάλως, ένα γιγαντιαίο χάλκινο πλάσμα, ήταν προγραμματισμένο να περιφρουρεί την Κρήτη από εισβολείς και να τους εξοντώνει. Πρόκειται για το πρώτο φονικό ρομπότ της ιστορίας (Winkler, 2020). Ακόμα και σε αυτό το πρώιμο στάδιο, η παροχή βοήθειας, φροντίδας και εξυπηρέτησης είναι συνυφασμένη με τη γυναικεία απεικόνιση του τεχνητού πλάσματος, ενώ αντίθετα η αρσενική απεικόνισή του είναι συνακόλουθο της δύναμης, της προστασίας και της μαχητικότητας, προμηνύοντας τη βαθιά προβληματική σχέση της TN με το φύλο.

Η TN μονοπωλεί το παγκόσμιο τεχνολογικό, ακαδημαϊκό, ερευνητικό και καλλιτεχνικό ενδιαφέρον τα τελευταία χρόνια όλο και συχνότερα. Φωνές τεχνολογικού ντετερμινισμού (Musik & Bogner, 2019) που την αντιμετωπίζουν ως κάτι ξένο, τρομακτικό και ως σίγουρη καταστροφή του ανθρώπου, εξακολουθούν έως ένα βαθμό να υπάρχουν, ωστόσο έχουν δώσει κατά κύριο λόγο τη θέση τους σε δραστήρια και περίεργα μυαλά, που καταβάλλουν κάθε δυνατή προσπάθεια για να την αποκωδικοποιήσουν και να την εξελίξουν. Στο πλαίσιο της προσπάθειας αυτής, η TN είναι πλέον κομμάτι της καθημερινότητας του ανθρώπου, σε πολύ μεγαλύτερο βαθμό από ό,τι γίνεται αντιληπτό, αλληλεπιδρώντας με διάφορους τομείς της σύγχρονης ζωής (Ευρωπαϊκό Κοινοβούλιο, 2020) και με προεκτάσεις που χρήζουν προσοχής.

Η ΤΝ αποτελούσε ανέκαθεν ένα αντιφατικό πεδίο, προσφέροντας σημαντικές ευκαιρίες για τον άνθρωπο, με την υπόσχεση διευκόλυνσης και εκσυγχρονισμού της καθημερινής ζωής, απλουστεύοντας και αυτοματοποιώντας διαδικασίες και σε πολλές περιπτώσεις αποδεικνύεται ικανότερη από αυτόν, δεδομένου ότι έχει ανά πάσα στιγμή πρόσβαση σε απεριόριστο όγκο πληροφοριών. Ωστόσο, ως ανθρώπινο δημιούργημα, η ΤΝ κουβαλάει και τα προβληματικά στοιχεία των ανθρώπινων κοινωνιών, σφάλματα, τα οποία τίθενται στο προσκήνιο προς προβληματισμό.

Οι μεροληψίες (biases) οι οποίες ενσωματώνονται στην ΤΝ, εγκυμονούν σημαντικούς κινδύνους για την κοινωνία και τα θεμελιώδη ανθρώπινα δικαιώματα, όπως είναι ειδικότερα οι έμφυλες διακρίσεις (gender biases), οι οποίες απειλούν τις θεμελιώδεις αρχές της ισότητας και της απαγόρευσης των διακρίσεων, επηρεάζοντας τις γυναίκες, τα μη-δυναδικά άτομα και τη μαύρη κοινότητα (Leslie, et al., 2021). Ο όρος «έμφυλες διακρίσεις» (gender biases) στην ΤΝ έχει τεχνολογική, ηθικο-κοινωνική, και νομική σημασία. Περιλαμβάνει το αποτέλεσμα που στερεί από ένα στατιστικό αποτέλεσμα την αντιπροσωπευτικότητα, διαστρεβλώνοντάς το συστηματικά εξαιτίας του φύλου (ADMIN Cross-linked glossary, 2019), τις προκαταλήψεις και τα στερεότυπα περί φύλου που υπάρχουν στην κοινωνία και εκφράζονται στις εφαρμογές ΤΝ, αλλά και οποιαδήποτε κατάσταση στην οποία ένα άτομο αντιμετωπίζεται λιγότερο ευνοϊκά από ένα άλλο εξαιτίας του φύλου του (Cambridge Dictionary, χ.χ.). Οι έμφυλες διακρίσεις στην ΤΝ είναι ένα από τα σημαντικότερα και επιβλαβέστερα ζητήματα της ΤΝ. Η παγκόσμια επιστημονική κοινότητα έχει κρούσει ήδη από την αρχή της ανάπτυξης της τον κώδωνα του κινδύνου σχετικά, οπτική που αποτυπώνεται και σε όλες τις σύγχρονες προσπάθειες νομοθετικής ρύθμισης των συστημάτων ΤΝ παγκοσμίως μέχρι τώρα.

Τα αίτια των έμφυλων διακρίσεων της ΤΝ εντοπίζονται σε πολλά πεδία, καθιστώντας απαραίτητη την πολυεπίπεδη εξέταση του ζητήματος. Ήδη από τα πρώτα χρόνια ανάπτυξης της ΤΝ, διαμορφώθηκε μια κουλτούρα η οποία ενθάρρυνε πολύ περισσότερο τη συμμετοχή των ανδρών και απέκλειε τις γυναίκες, οδηγώντας σε κυριαρχία των λευκών προνομιούχων ανδρών. Λογικό, αν αναλογιστεί κανείς τα έμφυλα στερεότυπα που είναι βαθιά ριζωμένα στις ανθρώπινες κοινωνίες, τα οποία προτάσσουν τη θέση ότι ο χώρος της τεχνολογίας και της επιστήμης είναι ανδρική υπόθεση. Συνακόλουθα, αγνοούνται οι ανάγκες των γυναικών και πώς αυτές επηρεάζονται από την τεχνολογία (Broussard, 2018). Η μεγάλη έλλειψη δεδομένων

που υπάρχει για το γυναικείο φύλο σε σχέση με το ανδρικό, το οποίο αποτελεί το πρότυπο, συμβάλλει επίσης στη διαίωση της ανισότητας (Criado Perez, 2020). Ακόμα και η φυσική γλώσσα, η οποία χρησιμοποιείται για την εκπαίδευση των συστημάτων TN (Natural Language Processing – NLP) περιλαμβάνει στερεότυπα περί έμφυλων ρόλων, τα οποία ενσωματώνονται κατά αυτόν τον τρόπο στην TN (Bolukbasi, et al., 2016). Η φεμινιστική θεωρία, με καθοριστικό ρόλο στην εις βάθος ανάδειξη του προβλήματος, συμβάλλει σημαντικά στην αφύπνιση, την ευαισθητοποίηση και την κατεύθυνση της εστίασης στο βασικό πρόβλημα της εξουσίας: οι επικεφαλής των τεχνολογικών κολοσσών [στην πλειοψηφία τους αγγλόφωνοι λευκοί cisgender (ταύτιση βιολογικού και κοινωνικού φύλου (American Psychological Association, χ.χ.) άνδρες)], σχεδιάζουν και παράγουν TN με βάση αποκλειστικά τη δική τους οπτική και ανάγκες (D' Ignazio & Klein, 2020). Εμφανής είναι και η τάση θηλυκοποίησης της TN, με στερεοτυπικά έμφυλα χαρακτηριστικά, ενώ ακόμα και η επιστημονική φαντασία και οι ενσώματες αναπαραστάσεις της TN, δεν καταφέρνουν να ξεφύγουν από τις στερεοτυπικές θεωρήσεις του φύλου και τη διαίωση των έμφυλων διαφορών. Επισημαίνεται ότι η δυιστική αντίληψη περί φύλου, οδηγεί σε περαιτέρω αποκλεισμούς, αφού το μη-δυναδικό φύλο και η τρανς κοινότητα είναι χαρακτηριστικά εκτός του κοινού στο οποίο απευθύνεται η TN. Καθίσταται σαφές ότι πρόκειται για ένα κοινωνικο-τεχνολογικό (socio-technical) ζήτημα με νομικές και ηθικές περαιτέρω προεκτάσεις επιβάλλοντας τη λήψη μέτρων και πολιτικών για μια περισσότερο δίκαιη και ηθική TN (Schwartz, et al., 2022).

Για την επίτευξη ηθικής και δίκαιης TN δεν αρκούν οι αμιγώς τεχνολογικές λύσεις, διότι οι ρίζες του προβλήματος είναι βαθύτερες και εντοπίζονται σε ιστορικές ανισότητες. Είναι εξαιρετικά δύσκολο εγχείρημα η δημιουργία μιας TN πραγματικά απαλλαγμένης από έμφυλες διακρίσεις, στερεότυπα και συμπεριφορές που έχουν εδραιωθεί στις ανθρώπινες κοινωνίες από τις απαρχές τους. Είναι εφικτή ωστόσο η υιοθέτηση μέτρων και πολιτικών για τον εντοπισμό των έμφυλων διακρίσεων και τη ρύθμιση της TN βάσει ηθικών και νομικών κανόνων. Η διεπιστημονικότητα και η συνεργασία νομικών, φιλοσόφων και προγραμματιστών κρίνεται απαραίτητη για την ενίσχυση ποικιλομορφίας, της συμπεριληπτικότητας και τελικά μιας πιο ηθικής και δίκαιης TN που θα ευνοεί όλα τα άτομα.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάδειξη, εξέταση και ανάλυση του ζητήματος των έμφυλων διακρίσεων στην TN, η ευαισθητοποίηση σχετικά και η πρόταση

λύσεων και δράσεων για την εξάλειψή τους. Ειδικότερα, στο 1^ο Κεφάλαιο παρουσιάζεται ο ορισμός της ΤΝ, οι εφαρμογές της στη σύγχρονη ζωή και η σχέση της με τον άνθρωπο, αλλά και οι σύγχρονες προτάσεις και κατευθύνσεις για την αξιόπιστη λειτουργία και ρύθμισή της. Στη συνέχεια, στο 2^ο Κεφάλαιο αναλύεται εκτενώς το ζήτημα των μεροληψιών (biases) στα συστήματα ΤΝ και ειδικότερα των έμφυλων διακρίσεων, με αναφορά στα αίτιά τους, παραδείγματα, τη φεμινιστική θεωρία και τις αναπαραστάσεις της ΤΝ (ενσώματες και μη) τόσο στο φυσικό κόσμο όσο και στην επιστημονική φαντασία. Στο 3^ο Κεφάλαιο διατυπώνονται προτάσεις για την εξάλειψη των έμφυλων διακρίσεων στην ΤΝ και για την επίτευξη μιας πιο ηθικής ΤΝ. Τέλος, ακολουθεί Επίλογος-Συμπέρασμα.

ΚΕΦΑΛΑΙΟ 1^ο

ΟΡΙΣΜΟΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ – ΗΘΙΚΟΙ ΠΡΟΒΛΗΜΑΤΙΣΜΟΙ ΚΑΙ ΘΕΜΕΛΙΩΔΗ ΑΝΘΡΩΠΙΝΑ ΔΙΚΑΙΩΜΑΤΑ – ΠΡΟΣΠΑΘΕΙΕΣ ΡΥΘΜΙΣΗΣ

1.1 Εισαγωγικές Παρατηρήσεις

Η τεχνολογική έκρηξη που άρχισε να παρατηρείται από το τέλος του προηγούμενου αιώνα και έπειτα, αναπόφευκτα οδήγησε σε δυστοπικά οράματα περί απόλυτης κυριαρχίας των μηχανών και την υποδούλωση ή/και αφανισμό του ανθρώπου, απηχώντας αντιλήψεις τεχνολογικού ντετερμινισμού (Musik & Bogner, 2019). Στο άκουσμα και μόνο του όρου ΤΝ δε, η φαντασία της τρομοκρατημένης ανθρωπότητας οργιάζει, όπως έχει αποτυπωθεί άλλωστε στην πληθώρα κινηματογραφικών και λογοτεχνικών έργων που έχουν παρουσιάσει διαφορετικές, ρεαλιστικές και μη εκφάνσεις του πρωτόγνωρου αυτού πεδίου. Σήμερα, το δέος που άλλοτε συνόδευε την ΤΝ, έχει σε μεγάλο βαθμό δώσει τη θέση του στην αποδοχή της πρακτικότητας που αυτή προσφέρει, καθώς συναντάται σε πολλές πτυχές της καθημερινής ζωής, κάνοντάς την ευκολότερη. Έχει διατυπωθεί μάλιστα η άποψη ότι η εποχή της Κοινωνίας της Πληροφορίας (Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο της Ευρωπαϊκής Ένωσης, 2015) έχει δώσει τη θέση της στην Κοινωνία των Αλγορίθμων, όπου οι αλγόριθμοι χρησιμοποιούνται ευρέως από τις εταιρίες, τα κράτη και τις ψηφιακές πλατφόρμες ως τρόπος διακυβέρνησης (governance), ως απόρροια της πεποίθησης ότι οι αλγόριθμοι είναι απαλλαγμένοι από τα σφάλματα της ανθρώπινης φύσης και προσφέρουν καταλληλότερες και ακριβέστερες λύσεις (Simoncini & Longo, 2022). Οι αλγόριθμοι χρησιμοποιούνται πλέον εκτεταμένα σε διάφορους τομείς της σύγχρονης ζωής, χωρίς να έχουν πάψει, βέβαια, να αποτελούν πηγή προβληματισμών και αντιπαραθέσεων. Δεν είναι τυχαίο άλλωστε, που μέχρι στιγμής δεν υπάρχει ούτε κοινός ορισμός σε διεθνές επίπεδο για την ΤΝ, αλλά και ούτε νομικά δεσμευτικοί (ρυθμιστικοί) κανόνες, καταδεικνύοντας την ιδιομορφία, πολυπλοκότητα και ρευστότητα του πεδίου.

1.2 Ορισμός TN

Αν και η συζήτηση για την TN και τις δυνατότητές της μονοπωλεί το παγκόσμιο ενδιαφέρον τα τελευταία χρόνια, ως έννοια είχε ήδη κάνει την εμφάνισή της από το 1950, όταν ο Alan Turing δημοσίευσε την περίφημη μελέτη του, γνωστή ως «Το Παιχνίδι της Μίμησης» ή αλλιώς ως τεστ Turing. Στη μελέτη αυτή, ο Alan Turing μέσω ενός ενδιαφέροντος τεστ μεταξύ δύο ανθρώπων και μιας μηχανής (βλ. παρακάτω 2.5), διατυπώνει τη θέση ότι οι υπολογιστικές μηχανές μπορούν να μιμηθούν την ανθρώπινη συμπεριφορά και κατά συνέπεια διαθέτουν ευφύια (Turing, 2020). Η TN ως όρος, ωστόσο, διατυπώθηκε επίσημα για πρώτη φορά στο ερευνητικό συνέδριο του Dartmouth το 1956, όπου τέθηκαν τα θεμέλια του νέου τότε πεδίου. Η TN είναι κλάδος της επιστήμης της Πληροφορικής, ο οποίος ασχολείται με την προσομοίωση της νοήμονος συμπεριφοράς. Ειδικότερα, η TN στοχεύει στη δημιουργία συστημάτων που σκέφτονται ή συμπεριφέρονται όπως ο άνθρωπος ή με τη μέγιστη δυνατή λογική ικανότητα. Η TN στην ουσία, είναι το αποτέλεσμα της περιέργειας των επιστημόνων να αποκωδικοποιήσουν τη λειτουργία του ανθρώπινου εγκεφάλου και να εξερευνήσουν την έννοια της συνείδησης, αποτυπώνοντάς τη σε ένα σύστημα TN.

1.2.1 Διακρίσεις TN

Με βάση το βαθμό ευφύιας που επιτυγχάνεται σε ένα σύστημα TN, υπάρχει η διάκριση μεταξύ α) ασθενούς ή περιορισμένης TN, η οποία αποτελεί απλά ένα εργαλείο για την εκτέλεση συγκεκριμένων και περιορισμένων υπολογισμών χωρίς καμία αυτονομία (συμβολική TN και νευρωνικά δίκτυα/μηχανική μάθηση) και β) ισχυρής ή γενικής TN, η οποία λειτουργεί και σκέφτεται όπως ο άνθρωπος, μπορεί να εκτελέσει οποιονδήποτε υπολογισμό και νοητική διεργασία, ενώ διαθέτει ακόμα αυτονομία και συνείδηση (Σαρά, 2020).

Στην ασθενή ή περιορισμένη TN ανήκουν η συμβολική TN και τα συστήματα TN που λειτουργούν βάσει νευρωνικών δικτύων (neural networks) και μηχανικής μάθησης (machine learning). Στη συμβολική TN (symbolic AI), οι ειδικοί δημιουργούν ακριβείς διαδικασίες βασισμένες σε κανόνες (τους γνωστούς αλγορίθμους), τους οποίους μπορεί να ακολουθήσει ένας υπολογιστής βήμα προς βήμα, ώστε να αποφασίσει πώς θα αποκριθεί με ευφύια σε μια δεδομένη κατάσταση. Η συμβολική TN αποδίδει στο μέγιστο σε περιορισμένα και ελεγχόμενα περιβάλλοντα, τα οποία δε μεταβάλλονται ουσιαστικά με την πάροδο του χρόνου και οι κανόνες

είναι αυστηροί και συγκεκριμένοι και οι μεταβλητές ξεκάθαρες και μετρήσιμες. Αυτό θέτει σημαντικούς περιορισμούς στον βαθμό αυτονομίας της. Ενώ μπορεί να πραγματοποιήσει εργασίες αυτόματα, μπορεί να το κάνει μόνο με τους τρόπους που υποδεικνύονται από τις οδηγίες που της δίνονται, και μπορεί να βελτιωθεί μόνο με άμεση ανθρώπινη παρέμβαση. Αυτή ήταν η κύρια προσέγγιση σε εφαρμογές TN από τη δεκαετία του 1950 έως τη δεκαετία του 1990, ωστόσο η συμβολική TN εξακολουθεί να χρησιμοποιείται σε πολλούς τομείς ακόμα και σήμερα.

Στη σύγχρονη εποχή, ο τεράστιος όγκος δεδομένων που υπάρχουν διαθέσιμα (big data), έχει οδηγήσει στην επικράτηση άλλων τρόπων λειτουργίας των συστημάτων TN. Πρόκειται για τα νευρωνικά δίκτυα (neural networks) και τη μηχανική μάθηση (machine learning). Τα συστήματα αυτά βασίζονται στα δεδομένα και την εκπαίδευση της TN με βάση ένα σύνολο δεδομένων που εισάγονται στο σύστημα (training data) (Boucher, 2020). Όσον αφορά τα νευρωνικά δίκτυα, αυτά είναι εμπνευσμένα από τη λειτουργία του ανθρώπινου εγκεφάλου. Τα εισαγόμενα δεδομένα (inputs), μεταφράζονται σε σήματα που περνούν μέσα από το νευρωνικό δίκτυο για τη δημιουργία αποτελεσμάτων (outputs), τα οποία ερμηνεύονται ως απαντήσεις στα αρχικώς εισαγόμενα δεδομένα. Προσθέτοντας περισσότερους νευρώνες και επίπεδα, τα νευρωνικά δίκτυα γίνονται ικανά να αντιμετωπίσουν πιο περίπλοκα ζητήματα· πρόκειται για τη λεγόμενη βαθιά μάθηση (deep learning), η οποία αναφέρεται σε νευρωνικά δίκτυα με πολλαπλά επίπεδα (Reichman & Sartor, 2022). Όσον αφορά τη μηχανική μάθηση, αυτή αφορά το μετασχηματισμό του δικτύου, ώστε τα αποτελέσματα που παράγει να θεωρούνται χρήσιμες ή/και ευφείς απαντήσεις στα εισαγόμενα δεδομένα. Η μηχανική μάθηση αναφέρεται σε συγκεκριμένες τεχνικές που αυτοματοποιούν τη διαδικασία εκπαίδευσης των αλγορίθμων. Μία από αυτές είναι η επιτηρούμενη μάθηση (supervised learning), η οποία είναι και η συνηθέστερη. Στην επιτηρούμενη μάθηση, το σύστημα μαθαίνει μέσω «επίβλεψης» ή «διδασκαλίας», παρέχεται δηλαδή εκ των προτέρων ένα σύνολο δεδομένων για εκπαίδευση (δηλαδή, ένα μεγάλο σύνολο απαντήσεων/λύσεων που θεωρούνται σωστές για την επίτευξη ενός καθορισμένου στόχου). Η ενισχυτική μάθηση (reinforcement learning), περαιτέρω, είναι παρόμοια με την επιτηρούμενη μάθηση, καθώς και οι δύο περιλαμβάνουν εκπαίδευση μέσω παραδειγμάτων, ωστόσο, στην περίπτωση της ενισχυτικής μάθησης το σύστημα μαθαίνει επιπλέον από τα αποτελέσματα των δικών του ενεργειών, δηλαδή μέσω των επιβραβεύσεων ή των αποδοκιμασιών (π.χ. κερδισμένοι ή χαμένοι βαθμοί) που συνδέονται με τα σχετικά αποτελέσματα. Τέλος, στην τεχνική της

μάθησης χωρίς επίβλεψη (unsupervised learning), τα συστήματα TN μαθαίνουν χωρίς εξωτερικές οδηγίες, είτε εκ των προτέρων είτε ως σχόλια, σχετικά με το τι είναι σωστό ή λάθος, παρουσιάζουν δηλαδή αυτονομία (IBM Cloud Education, 2020). Η μάθηση χωρίς επίβλεψη χρησιμοποιείται κυρίως για ομαδοποίηση – δηλαδή για την ομαδοποίηση ενός συνόλου στοιχείων που παρουσιάζουν σχετικές ομοιότητες ή συνδέσεις (π.χ. έγγραφα που αφορούν το ίδιο θέμα, άτομα που παρουσιάζουν παρόμοια χαρακτηριστικά ή όροι που έχουν την ίδια εννοιολογική σημασία σε κείμενα).

Επισημαίνεται, ότι μέχρι στιγμής δεν έχει δημιουργηθεί ισχυρό σύστημα TN, αντίθετα, υπάρχουν μόνο ασθενή συστήματα TN, όπως αυτά αναφέρθηκαν ανωτέρω, τα οποία τελούν υπό την εποπτεία του ανθρώπου και χρησιμοποιούνται σε διάφορους τομείς για την εξυπηρέτησή του (Boucher, 2020). Ωστόσο, αξίζει να αναφερθεί ότι πρόσφατα έχει προκαλέσει έντονο προβληματισμό και διχογνωμίες η περίπτωση του LaMDA, του chatbot TN της Google, για το οποίο ένας μηχανικός της Google AI, ισχυρίζεται ότι διαθέτει συνείδηση, κατόπιν συζήτησης μαζί του. Υπάρχουν αρκετές σχετικές φιλοσοφικές και τεχνολογικές ενστάσεις και όχι ακόμα επαρκείς αποδείξεις, αποτελεί ωστόσο ένα πολύ ενδιαφέρον ζήτημα και αναμένεται η εξέλιξή του (Curtis & Savulescu, 2022).

1.2.2 Προσπάθειες διατύπωσης ορισμού TN

Όσον αφορά τον ορισμό της, η TN εξαρχής αποτέλεσε ένα πεδίο με έντονο ερευνητικό και φιλοσοφικό ενδιαφέρον, αλλά και εξαιρετικά δύσκολο, γεγονός που αποδεικνύεται από το ότι δεν υπάρχει ένας ενιαίος ορισμός για την TN, αντίθετα έχουν διατυπωθεί διάφοροι ορισμοί ανά τα χρόνια. Από τις πρώτες και βασικότερες προσεγγίσεις ορισμού της TN, είναι αυτές που εστιάζουν περισσότερο στις διαδικασίες σκέψης και τη συλλογιστική (συστήματα που σκέφτονται σαν τον άνθρωπο ή ορθολογικά), αλλά και τη συμπεριφορά ενός συστήματος TN (συστήματα που συμπεριφέρονται σαν τον άνθρωπο ή ορθολογικά). Επιπλέον, υπάρχουν προσεγγίσεις που μετρούν την επιτυχία ενός συστήματος TN με βάση την εγγύτητα προς τις ανθρώπινες επιδόσεις, ενώ άλλες τη μετρούν σε σχέση με μία ιδανική έννοια την νοημοσύνης, την ορθολογικότητα (Russel & Norvig, 2021). Ως ορθολογικότητα νοείται η ικανότητα επιλογής της βέλτιστης ενέργειας για την επίτευξη ενός συγκεκριμένου στόχου, με βάση, αφενός συγκεκριμένα κριτήρια προς βελτιστοποίηση και αφετέρου τους διαθέσιμους πόρους.

Σημαντική επιρροή στην επιλογή του όρου της ορθολογικότητας από τους ερευνητές έχει ασκήσει η σκέψη των S. Russell και P. Norvig πως κάθε σύστημα TN είναι πρώτα και κύρια ορθολογικό (High-Level Expert Group on Artificial Intelligence, 2019). Κατ' αυτούς, ένα σύστημα είναι ορθολογικό, αν κάνει «το σωστό», με δεδομένα όσα γνωρίζει (Russel & Norvig, 2021).

Αναφέρεται ότι όσον αφορά ειδικότερα την Ευρωπαϊκή Ένωση (εφεξής ΕΕ), η Ευρωπαϊκή Επιτροπή, στο πλαίσιο ένταξης της TN στην πολιτική της ατζέντα και σε μια εντατική προσπάθεια αποσαφήνισης της έννοιας και ρύθμισής της, το 2018 (Ευρωπαϊκή Επιτροπή, 2018) διατύπωσε το δικό της ορισμό για την TN, σύμφωνα με τον οποίο:

«Η TN αναφέρεται σε συστήματα που χαρακτηρίζονται από ευφυή συμπεριφορά, αναλύοντας το περιβάλλον τους και ενεργώντας – με κάποιο βαθμό αυτονομίας – για την επίτευξη συγκεκριμένων στόχων.

Τα συστήματα που λειτουργούν βάσει τεχνητής νοημοσύνης μπορούν να βασίζονται αποκλειστικά σε λογισμικό, ενεργώντας στον εικονικό κόσμο (π.χ. βοηθοί φωνής, λογισμικό ανάλυσης εικόνας, μηχανές αναζήτησης, συστήματα αναγνώρισης ομιλίας και προσώπου) ή η TN μπορεί να ενσωματωθεί σε συσκευές υλισμικού (π.χ. προηγμένα ρομπότ, αυτόνομα αυτοκίνητα, δρόνοι ή εφαρμογές του Διαδικτύου των Πραγμάτων).»

Στη συνέχεια, το 2019 η Ανεξάρτητη Ομάδα Εμπειρογνομόνων Υψηλού Επιπέδου (εφεξής ΟΕΥΕ) για την TN, συσταθείσα από την Ευρωπαϊκή Επιτροπή από το 2018, πρότεινε να χρησιμοποιείται ένας επικαιροποιημένος ορισμός (High-Level Expert Group on Artificial Intelligence, 2019), σύμφωνα με τον οποίο:

«Τα συστήματα TN είναι συστήματα λογισμικού (ή ενδεχομένως και υλισμικού) που σχεδιάζονται από ανθρώπους και, βάσει ενός δεδομένου σύνθετου στόχου, ενεργούν στην υλική ή ψηφιακή διάσταση με το να αντιλαμβάνονται το περιβάλλον τους μέσω της απόκτησης δεδομένων, να ερμηνεύουν τα δομημένα ή αδόμητα δεδομένα που έχουν συλλεχθεί, να προβαίνουν σε συλλογισμούς με βάση τις γνώσεις ή να επεξεργάζονται τις πληροφορίες που εξάγονται από αυτά τα δεδομένα και να αποφασίζουν ποια είναι η βέλτιστη ενέργεια (ή οι βέλτιστες ενέργειες) που θα πρέπει να εκτελέσουν για να επιτύχουν τον δεδομένο στόχο. Τα συστήματα TN μπορεί είτε να χρησιμοποιούν συμβολικούς κανόνες είτε να μαθαίνουν ένα

αριθμητικό μοντέλο, και μπορεί επίσης να προσαρμόζουν τη συμπεριφορά τους με το να αναλύουν πώς επηρεάζεται το περιβάλλον από τις προηγούμενες ενέργειές τους.

Ως επιστημονικό πεδίο, η TN περιλαμβάνει διάφορες προσεγγίσεις και τεχνικές, όπως η μηχανική μάθηση (συγκεκριμένα παραδείγματα της οποίας είναι η βαθιά μάθηση και η ενισχυτική μάθηση), η μηχανική συλλογιστική (που περιλαμβάνει τον σχεδιασμό, τον προγραμματισμό, την αναπαράσταση και τη συλλογιστική γνώσης, την αναζήτηση και τη βελτιστοποίηση) και η ρομποτική (που περιλαμβάνει έλεγχο, αντίληψη, αισθητήρες και ενεργοποιητές, καθώς και την ενσωμάτωση όλων των άλλων τεχνικών σε κυβερνο-υλικά συστήματα)».

Πιο πρόσφατα, ωστόσο, στις 21 Απριλίου 2021 ανακοινώθηκε από την Ευρωπαϊκή Επιτροπή το πρώτο ιστορικά νομικό πλαίσιο για την TN (Ευρωπαϊκή Επιτροπή, 2021). Εξαιρετικά σημαντικό κείμενο της πρωτοβουλίας αυτής, μιας και αποτελεί την πρώτη προσπάθεια ενιαίας θέσπισης νομικά δεσμευτικών κανόνων για την TN σε διακρατικό επίπεδο, αποτελεί η Πρόταση Κανονισμού για τη Θέσπιση Εναρμονισμένων Κανόνων Σχετικά με την Τεχνητή Νοημοσύνη (Πράξη για την Τεχνητή Νοημοσύνη) και για την τροποποίηση ορισμένων νομοθετικών πράξεων της Ένωσης (Ευρωπαϊκή Επιτροπή, 2021) (εφεξής Πράξη για την TN), στο άρθρο 3 του οποίου δίνεται ένας νέος ορισμός της TN ως εξής:

«σύστημα τεχνητής νοημοσύνης» (σύστημα TN): λογισμικό που αναπτύσσεται με μία ή περισσότερες από τις τεχνικές και προσεγγίσεις που παρατίθενται στο παράρτημα I και μπορεί, για ένα δεδομένο σύνολο στόχων που έχουν καθοριστεί από τον άνθρωπο, να παράγει στοιχεία εξόδου όπως περιεχόμενο, προβλέψεις, συστάσεις ή αποφάσεις που επηρεάζουν τα περιβάλλοντα με τα οποία αλληλεπιδρά».

Περαιτέρω, οι προσεγγίσεις που αναφέρονται στο Παράρτημα I είναι:

α) Προσεγγίσεις μηχανικής μάθησης, συμπεριλαμβανομένων της επιβλεπόμενης, της μη επιβλεπόμενης και της ενισχυτικής μάθησης, με τη χρήση ευρέος φάσματος μεθόδων, συμπεριλαμβανομένης της βαθιάς μάθησης·

β) Προσεγγίσεις που βασίζονται στη λογική και στις γνώσεις, συμπεριλαμβανομένων της αναπαράστασης γνώσεων, του επαγωγικού (λογικού) προγραμματισμού, των βάσεων

γνώσεων, των μηχανών εξαγωγής συμπερασμάτων και παραγωγικών συλλογισμών, των συστημάτων (συμβολικής) συλλογιστικής και των συστημάτων-εμπειρογνομώνων

γ) Στατιστικές προσεγγίσεις, εκτίμηση κατά Bayes, μέθοδοι αναζήτησης και βελτιστοποίησης.»

Με πιο απλά λόγια, σύμφωνα και με τον τελευταίο ως άνω ορισμό των συστημάτων TN, ως τέτοιο ορίζεται το σύστημα λογισμικού, το οποίο με βάση τα δεδομένα και τους στόχους που ορίζονται από τον άνθρωπο, παράγει ένα συγκεκριμένο αποτέλεσμα, το οποίο επηρεάζει το περιβάλλον μέσα στο οποίο βρίσκεται και αλληλεπιδρά. Ο λόγος που η Ευρωπαϊκή Επιτροπή δεν υιοθέτησε κάποιον από τους προγενέστερους ορισμούς στην Πράξη για την TN, αλλά διατύπωσε έναν εμφανώς πιο συνοπτικό και λιτό ορισμό, αλλά ταυτόχρονα πιο τεχνικό, ήταν αφενός η επίτευξη όσο το δυνατόν μεγαλύτερης σαφήνειας για ενιαία ρύθμιση και αφετέρου η διαχρονικότητα και η ανταπόκριση στις τεχνολογικές εξελίξεις (Ευρωπαϊκή Επιτροπή, 2021). Όπως αναφέρεται στην αιτιολογική έκθεση της Πρότασης Κανονισμού:

«[...]Ο ορισμός του συστήματος TN στο νομικό πλαίσιο έχει ως στόχο να είναι όσο το δυνατόν τεχνολογικά ουδέτερος και διαχρονικός, λαμβανομένων υπόψη των ραγδαίων τεχνολογικών εξελίξεων και των εξελίξεων της αγοράς που σχετίζονται με την TN. Προκειμένου να παρασχεθεί η αναγκαία ασφάλεια δικαίου, ο τίτλος I συμπληρώνεται από το παράρτημα I, το οποίο περιέχει λεπτομερή κατάλογο προσεγγίσεων και τεχνικών για την ανάπτυξη της TN, ο οποίος πρέπει να προσαρμοστεί από την Επιτροπή σύμφωνα με τις νέες τεχνολογικές εξελίξεις.[...]

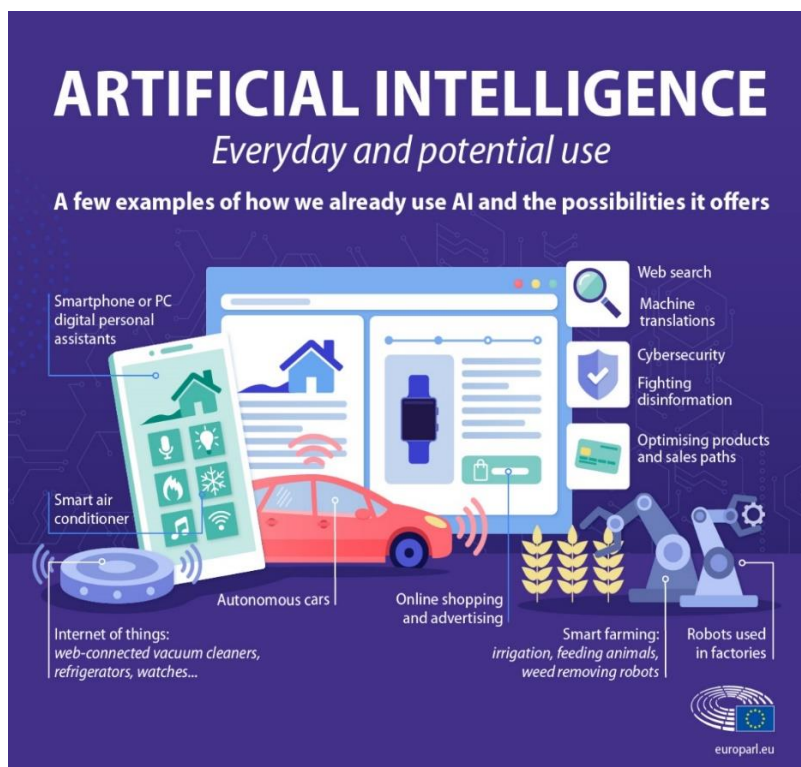
Επιπλέον, στο Προοίμιο της Πρότασης Κανονισμού αναφέρεται:

«Η έννοια του συστήματος TN θα πρέπει να οριστεί με σαφήνεια, για λόγους ασφάλειας δικαίου, και για να διασφαλιστεί, παράλληλα, ευελιξία για την προσαρμογή στις μελλοντικές τεχνολογικές εξελίξεις. Ο ορισμός θα πρέπει να βασίζεται στα βασικά λειτουργικά χαρακτηριστικά του λογισμικού, ειδικότερα στην ικανότητα ενός δεδομένου συνόλου στόχων καθορισμένων από τον άνθρωπο, να παράγει στοιχεία εξόδου όπως περιεχόμενο, προβλέψεις, συστάσεις ή αποφάσεις που επηρεάζουν το περιβάλλον με το οποίο αλληλεπιδρά το σύστημα, είτε σε υλική είτε σε ψηφιακή διάσταση. Τα συστήματα TN μπορούν να σχεδιάζονται ώστε να λειτουργούν με διαφορετικά επίπεδα αυτονομίας και να χρησιμοποιούνται σε αυτόνομη βάση

ή ως κατασκευαστικό στοιχείο προϊόντος, ανεξάρτητα από το αν το σύστημα έχει ενσωματωθεί φυσικά στο προϊόν (ενσωματωμένο σύστημα) ή εξυπηρετεί τη λειτουργικότητα του προϊόντος χωρίς να έχει ενσωματωθεί σε αυτό (μη ενσωματωμένο σύστημα). Ο ορισμός του συστήματος ΤΝ θα πρέπει να συμπληρωθεί από κατάλογο ειδικών τεχνικών και προσεγγίσεων που χρησιμοποιούνται για την ανάπτυξή του, ο οποίος θα πρέπει να επικαιροποιείται με βάση τις εξελίξεις της αγοράς και της τεχνολογίας, μέσω της έκδοσης κατ' εξουσιοδότηση πράξεων από την Επιτροπή για την τροποποίησή του.»

1.3 Παραδείγματα εφαρμογών ΤΝ

Η λειτουργία ενός συστήματος ΤΝ γίνεται πιο κατανοητή μέσα από τις χρήσεις του στην καθημερινότητα. Ένα λογισμικό ΤΝ μπορεί να εμφανίζεται μόνο στον εικονικό κόσμο (εικονικοί βοηθοί, λογισμικό ανάλυσης εικόνας, μηχανές αναζήτησης, συστήματα αναγνώρισης προσώπου και ομιλίας, κλπ.) ή να ενσωματώνεται σε υλικά αντικείμενα και συσκευές [ρομπότ, αυτόνομα αυτοκίνητα, τηλεκατευθυνόμενα αεροσκάφη (drones), Διαδίκτυο των πραγμάτων (Internet of Things)].



Εικόνα 1: Τεχνητή Νοημοσύνη: Καθημερινή και δυναμική χρήση (Ευρωπαϊκό Κοινοβούλιο, 2020)

Πιο συγκεκριμένα, από τις πιο βασικές χρήσεις της ΤΝ είναι αυτή στις διαδικτυακές μηχανές αναζήτησης, όπου οι αλγόριθμοι βοηθούν στη γρήγορη και εύκολη έρευνα, ταξινόμηση και εμφάνιση αποτελεσμάτων. Η ΤΝ χρησιμοποιείται ευρέως και στις διαδικτυακές αγορές και τη διαφήμιση, για την παροχή εξατομικευμένων προτάσεων στους χρήστες κατόπιν ανάλυσης της αγοραστικής συμπεριφοράς τους, αλλά και της παρακολούθησης της επισκεψιμότητάς τους σε διαδικτυακούς τόπους. Ευρέως διαδεδομένοι είναι και οι ψηφιακοί προσωπικοί βοηθοί, όπως είναι τα έξυπνα τηλέφωνα (smartphones), που χρησιμοποιούν την ΤΝ για την παροχή βελτιστοποιημένων και εξατομικευμένων ρυθμίσεων στους χρήστες τους. Επιπλέον, τα έξυπνα σπίτια, τα αυτοκίνητα με ευφυή συστήματα ασφαλείας, όπως είναι οι αυτόματοι αισθητήρες και τα συστήματα πλοήγησης, βασίζονται σε μεγάλο βαθμό στην ΤΝ. Σημαντική είναι επίσης η συμβολή και εφαρμογή της ΤΝ τόσο σε συστήματα κυβερνοασφάλειας με την αναγνώριση και αντιμετώπιση επιθέσεων και απειλών στον κυβερνοχώρο, όσο και στην καταπολέμηση της παραπληροφόρησης, με την ανίχνευση ψευδών ειδήσεων στα κοινωνικά δίκτυα μέσω του εντοπισμού συγκεκριμένων λέξεων και εκφράσεων, αλλά και αξιόπιστων πηγών πληροφόρησης (Ευρωπαϊκό Κοινοβούλιο, 2020). Μια ακόμα χρήση της ΤΝ που κερδίζει όλο και περισσότερο έδαφος, αν και αμφιλεγόμενη (EDRi, 2020), είναι τα συστήματα αναγνώρισης προσώπων, τα οποία βασίζονται σε ανάλυση εικόνων και την αντιστοίχισή τους με μια βάση δεδομένων (Azria & Wickert, 2019). Ανερχόμενη είναι και η χρήση της ΤΝ στη δικαιοσύνη, καθώς θεωρείται ότι μπορεί να συμβάλει σημαντικά στη μείωση του όγκου των δικαστικών εργασιών και την επίτευξη των διαδικασιών (European Commission for the Efficiency of Justice (CEPEJ), 2019). Ένα πιο πρόσφατο παράδειγμα σε παγκόσμια κλίμακα αποτελεί η χρήση της ΤΝ στην παρακολούθηση και καταπολέμηση του COVID-19, αφού η ΤΝ χρησιμοποιείται σε συσκευές θερμικής απεικόνισης σε αεροδρόμια και αλλού, ενώ στην ιατρική, η ΤΝ μπορεί, επίσης, να βοηθήσει στην αποτελεσματική διάγνωση μέσω της χρήσης αλγορίθμων, αλλά και στην παρακολούθηση της εξάπλωσης του ιού μέσω της παροχής δεδομένων (Ευρωπαϊκό Κοινοβούλιο, 2020).

1.4 Ηθικοί-κοινωνικοί προβληματισμοί και θεμελιώδη ανθρώπινα δικαιώματα

Παρά τη σημαντική συμβολή της ΤΝ στην καθημερινότητα των ανθρώπων και τις διευκολύνσεις που αυτή προσφέρει, όπως αυτές αναφέρθηκαν ανωτέρω, ελλοχεύουν κίνδυνοι για την κοινωνία και τα θεμελιώδη δικαιώματα, οι οποίοι προκαλούν έντονο προβληματισμό και παγκόσμια συζήτηση. Αδιαμφισβήτητη θέση αποτελεί παγκοσμίως ότι η ΤΝ πρέπει να τίθεται στην υπηρεσία της ανθρωπότητας και του κοινού καλού, η λειτουργία της δηλαδή πρέπει πάντα να είναι ανθρωποκεντρική, ενώ περαιτέρω ως βάση τίθεται ο σεβασμός στα θεμελιώδη ανθρώπινα δικαιώματα, όπως αυτά κατοχυρώνονται σε διεθνή και ενωσιακά κείμενα (Rodrigues, 2020).

Ένα από τα σημαντικότερα ηθικά-κοινωνικά διλήμματα που γεννά η ΤΝ αφορά την ανεργία και κατ' επέκταση την κοινωνική-οικονομική ανισότητα. Τα συστήματα ΤΝ, καθιερώνοντας τον αυτοματισμό, οδηγούν στην παράκαμψη της μεσολάβησης του ανθρώπου και τη σταδιακή εξάλειψη ορισμένων εργασιών που σήμερα πραγματοποιούνται με ανθρώπινη παρέμβαση. Με αυτό τον τρόπο, όλο και λιγότεροι θα έχουν τη δυνατότητα να εργαστούν και άρα, να έχουν εισόδημα, με αποτέλεσμα τη δημιουργία νέων ανισοτήτων σχετικά με την κατανομή του πλούτου. Προχωρώντας τη σκέψη ένα βήμα παραπέρα, αναπόφευκτη είναι η δημιουργία ανισοτήτων και σχετικά με τη δυνατότητα απόκτησης και πρόσβασης σε συστήματα ΤΝ (European Commission, Directorate-General for Education, Youth, Sport and Culture; Šucha, Vladimír; Gammel, Jean-Philippe, 2021). Με τη συλλογιστική αυτή, θα μπορούσε να γίνει πραγματικότητα μία νέα μορφή καπιταλιστικής κοινωνίας, όπως την είχε ενδελεχώς περιγράψει ο Καρλ Μαρξ (Μαρξ, 2006), της οποίας η κοινωνική διαστρωμάτωση θα διαμορφώνεται στη βάση της σύγκρουσης ανάμεσα σε αυτούς που ελέγχουν και επωφελούνται από τα συστήματα ΤΝ που οι ίδιοι κατέχουν (αστική τάξη που κατέχει το «κεφάλαιο») και αυτούς που έχουν περιορισμένη (ή και καθόλου) πρόσβαση στα τελευταία, εξαιτίας της οικονομικής αδυναμίας τους να έχουν στην κυριότητά τους ένα τέτοιο σύστημα ΤΝ και οι οποίοι θα είναι αναγκασμένοι να χρησιμοποιούν τα συστήματα αυτά ακριβώς όπως τους υποδεικνύουν οι κάτοχοί τους, αλλιώς θα τους στερείται καθ' ολοκληρίαν η πρόσβαση (Kumar, 2021).

Από τις πιο επικίνδυνες, ωστόσο, δυνητικά χρήσεις της ΤΝ είναι τα συστήματα αναγνώρισης προσώπου, τα οποία κερδίζουν όλο και μεγαλύτερο έδαφος. Η εκτεταμένη χρήση συστημάτων

αναγνώρισης προσώπου και η ολοένα αυξανόμενη μαζική παρακολούθηση των ατόμων, ειδικά σε δημόσιους χώρους μέσω των καμερών που βρίσκονται τοποθετημένες σε αυτούς, δημιουργούν έντονο προβληματισμό, αντιδράσεις και ζητήματα, τόσο όσον αφορά τα προσωπικά δεδομένα, όσο και γενικότερα τα θεμελιώδη δικαιώματα των πολιτών και τις ελευθερίες τους (EDRi, 2019). Ειδικότερα, τα θεμελιώδη ανθρώπινα δικαιώματα που τίθενται σε κίνδυνο είναι η ιδιωτικότητα, η προστασία των προσωπικών δεδομένων, η ελευθερία του συνέρχεσθαι, η ελεύθερη ανάπτυξη της προσωπικότητας, η ισότητα, η δικαιοσύνη. Τα δικαιώματα αυτά, τα οποία κατοχυρώνονται στο Χάρτη Θεμελιωδών Δικαιωμάτων της ΕΕ, σε διεθνή κείμενα όπως η Ευρωπαϊκή Σύμβαση για τα Δικαιώματα του Ανθρώπου κ.ά., αλλά και στα εθνικά συντάγματα των κρατών, αποτελούν τους πυλώνες της ΕΕ, της έννοιας του κράτους δικαίου, αλλά και συστατικά στοιχεία του δημοκρατικού πολιτεύματος. Πλέον, τα όρια των δικαιωμάτων αυτών δοκιμάζονται με εξαιρετικά αμφιλεγόμενα αποτελέσματα (Wiewiórowski, 2020).

Με τους ανωτέρω κινδύνους για τα θεμελιώδη ανθρώπινα δικαιώματα συνδέεται άμεσα και το ηθικό-φιλοσοφικό ζήτημα περί του αν τα συστήματα ΤΝ μπορούν να κατανοήσουν και να ενσωματώσουν ηθικούς κανόνες. Η ΤΝ μπορεί να βρίσκεται ακόμη σε απλοϊκό στάδιο, επιτελώντας περισσότερο διεκπεραιωτικούς και βοηθητικούς σκοπούς, δεν αποκλείεται, ωστόσο, στο άμεσο μέλλον να εξελιχθεί ραγδαία σε τέτοιο βαθμό, ώστε να αυτενεργεί, χωρίς καμία ανθρώπινη παρέμβαση. Εξετάζεται επομένως, αν και πώς μπορούν να υπάρξουν «ηθικές μηχανές». Η ηθική, ωστόσο, είναι αξία η οποία συνδέεται με την ανθρώπινη ιδιότητα και τη συνείδηση, ενώ, επιπλέον, αποτελεί έννοια, η οποία δεν έχει ενιαία σημασία, αντίθετα εξαρτάται από το πολιτισμικό και βιωματικό υπόβαθρο. Επομένως, η συζήτηση επεκτείνεται αφενός στο τι είναι συνείδηση και αν ένα σύστημα ΤΝ διαθέτει και αφετέρου στο πώς καθορίζεται η ηθική. Αυτό το πεδίο μελέτης της ηθικής των μηχανών (robot ethics) και της ηθικής της ΤΝ (AI ethics) όπως έχει διαμορφωθεί, κερδίζει όλο και περισσότερο έδαφος, προσπαθώντας να δώσει απαντήσεις σε αυτά τα ζητήματα και να συμβάλει στην εξασφάλιση της λειτουργίας της ΤΝ προς το συμφέρον του ανθρώπου, προτείνοντας αφενός τη διεπιστημονικότητα και τη συνεργασία και εξετάζοντας αφετέρου τον τρόπο με τον οποίο οι ηθικοί κανόνες μπορούν να μεταφραστούν σε αλγοριθμικό κώδικα (top-down morality), αλλά

και τον τρόπο με τον οποίο ένα σύστημα TN μπορεί να μάθει να είναι ηθικό (bottom-up morality) (Wallach & Allen, 2010).

Τέλος, το μεγαλύτερο ίσως σφάλμα που επικρατεί στα συστήματα TN αυτή τη στιγμή, με προεκτάσεις σε ποικίλους τομείς και έννομα αγαθά, είναι οι μεροληψίες (biases) και διακρίσεις, στις οποίες αυτά προβαίνουν (Rodrigues, 2020). Οι λόγοι που οδηγούν ένα σύστημα TN σε μεροληπτικά αποτελέσματα είναι ποικίλοι και θα αναλυθούν στο επόμενο κεφάλαιο (2). Βασικό ερώτημα αποτελεί το αν η TN μπορεί να είναι πράγματι ουδέτερη και δίκαιη ή είναι καταδικασμένη να αντανακλά τα στερεότυπα της κοινωνίας μέσα στην οποία δημιουργείται και λειτουργεί. Έχοντας συνειδητοποιήσει τη βαρύτητα του ζητήματος, το θέμα της διασφάλισης της ισότητας και της εξάλειψης των μεροληψιών τίθεται από την επιστημονική κοινότητα, αλλά και τις ρυθμιστικές αρχές στη βάση της πυραμίδας των ζητημάτων που πρέπει να ρυθμιστούν σχετικά με την TN, ως ένα από τα πρωταρχικά και ουσιώδη ζητήματα (Leslie, et al., 2021).

1.5 Προσπάθειες ρύθμισης

Τα ζητήματα που προκύπτουν από τις χρήσεις της TN είναι πολλά και ποικίλα και για τον λόγο αυτό, είναι κρίσιμης σημασίας να γίνει πράξη μία σειρά κανόνων-πλαίσιο που θα μεριμνά για την ορθή στάθμιση μεταξύ των αντικρουόμενων δικαιωμάτων και θα έχει ως βάση της θέσπισής της τον πρωταρχικό στόχο της TN, δηλαδή τη διασπορά των ωφελειών της τεχνολογίας για την – όσο το δυνατόν – καλύτερη λειτουργία της ανθρώπινης κοινωνίας (Ευρωπαϊκή Επιτροπή, 2020). Το έργο αυτό ωστόσο, όπως αποδεικνύεται από τις σχετικές κινήσεις και πρωτοβουλίες ανά τον κόσμο, είναι ιδιαίτερα δύσκολο. Αυτή τη στιγμή δεν υπάρχουν δεσμευτικοί κανόνες σε διεθνές επίπεδο, ενώ όλες οι σχετικές οδηγίες και προτάσεις κανόνων, παρατηρούνται αφενός σε επίπεδο διεθνών φορέων και οργανισμών με τη μορφή μη δεσμευτικών κανόνων και αφετέρου περιορισμένα σε επίπεδο κρατών (Daly, et al., 2022).

1.5.1 Προσπάθειες ρύθμισης σε διεθνές επίπεδο

Σε επίπεδο διεθνών φορέων και οργανισμών, ο Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (ΟΟΣΑ) εξέδωσε το 2019 ορισμένες Αρχές για την προώθηση καινοτόμου και αξιόπιστης TN με σεβασμό στα ανθρώπινα δικαιώματα και τις δημοκρατικές αξίες (Council on

Artificial Intelligence, 2019). Αυτές οι Αρχές αποτέλεσαν τη βάση για τις μη δεσμευτικές κατευθυντήριες αρχές που επίσης υιοθέτησαν τα μέλη της G20 για τη χρήση της TN το 2019 (G20 Trade Ministers and Digital Economy Ministers, 2019). Αξίζει να αναφερθεί επίσης, ότι τον Ιούνιο 2020, διάφορα κράτη, συμπεριλαμβανομένων της Αυστραλίας, της Ευρωπαϊκής Ένωσης, της Ινδίας, του Ηνωμένου Βασιλείου και των Ηνωμένων Πολιτειών, αλλά εξαιρουμένων της Κίνας και της Ρωσίας, σχημάτισαν την Παγκόσμια Συνεργασία για την TN (Global Partnership on Artificial Intelligence-GPAI), μια διεθνή και πολυμερή πρωτοβουλία που φέρνει σε επαφή κορυφαίους ειδικούς από την επιστήμη, τη βιομηχανία, την κοινωνία των πολιτών, διεθνείς οργανισμούς και την κυβέρνηση, οι οποίοι μοιράζονται αξίες με σκοπό να γεφυρώσουν το χάσμα μεταξύ θεωρίας και πρακτικής σχετικά με την TN, υποστηρίζοντας την έρευνα που βασίζεται στην καινοτομία και εφαρμοσμένες δραστηριότητες σε πρωτοβουλίες σχετικά με την TN (GPAI: Global Partnership on Artificial Intelligence, 2020). Σημαντικό έργο επιτελεί και ο Οργανισμός Ηνωμένων Εθνών (ΟΗΕ) μέσω των οργάνων του, με πιο σημαντική δραστηριότητα αυτή της UNESCO, η οποία ξεκίνησε το 2019 μια διετή διαδικασία για την εκπόνηση του πρώτου διεθνούς κειμένου για τον καθορισμό προτύπων σχετικά με την ηθική της TN. Έτσι, στην 41^η Γενική Συνέλευση της UNESCO που έλαβε χώρα στις 24 Νοεμβρίου 2021 υιοθετήθηκε η Πρόταση για την Ηθική της Τεχνητής Νοημοσύνης (εφεξής Πρόταση της UNESCO), το πρώτο ιστορικά διεθνές μη δεσμευτικό κείμενο για τους ηθικούς κανόνες που πρέπει να διέπουν την TN (UNESCO's General Conference: 41st session, 2021). Βασικοί στόχοι της Πρότασης της UNESCO είναι:

(α) να παράσχει ένα παγκόσμιο πλαίσιο αξιών, αρχών και δράσεων, ώστε να καθοδηγήσει τα κράτη στη διαμόρφωση της νομοθεσίας, των πολιτικών ή άλλων μέσων τους σχετικά με την TN, σύμφωνα με το διεθνές δίκαιο·

(β) να καθοδηγήσει τις δράσεις ατόμων, ομάδων, κοινοτήτων, ιδρυμάτων και εταιριών του ιδιωτικού τομέα, ώστε να διασφαλιστεί η ενσωμάτωση της ηθικής σε όλα τα στάδια του κύκλου ζωής του συστήματος TN·

(γ) η προστασία, η προώθηση και ο σεβασμός των ανθρωπίνων δικαιωμάτων και των θεμελιωδών ελευθεριών, της ανθρωπίνης αξιοπρέπειας και ισότητας, συμπεριλαμβανομένης της ισότητας των φύλων· η προστασία των συμφερόντων των σημερινών και των μελλοντικών

γενεών· η διατήρηση του περιβάλλοντος, της βιοποικιλότητας και των οικοσυστημάτων· και ο σεβασμός της πολιτισμικής πολυμορφίας (diversity) σε όλα τα στάδια του κύκλου ζωής του συστήματος TN·

(δ) να ενθαρρύνει τον πολυμερή, διεπιστημονικό και πλουραλιστικό διάλογο και την επίτευξη συμφωνίας (consensus) σχετικά με τα ηθικά ζητήματα που σχετίζονται με τα συστήματα TN·

(ε) να προωθήσει την ίση πρόσβαση στις εξελίξεις, την τεχνογνωσία και τα οφέλη στον τομέα της TN.

1.5.2 Προσπάθειες ρύθμισης σε επίπεδο κρατών (ΗΠΑ-Κίνα)

Σε κρατικό επίπεδο, λίγες είναι οι περιπτώσεις όπου συναντάται οποιοδήποτε είδος πλαισίου για την TN, ενώ ακόμα και όταν αυτό υπάρχει, κάθε κράτος ή κάθε ένωση κρατών (βλ. Ηνωμένες Πολιτείες Αμερικής και ΕΕ), εκκινώντας από διαφορετικά πολιτισμικά υπόβαθρα και τεχνολογικό επίπεδο, προβαίνουν σε διαφορετικές προβλέψεις. Ενδεικτικά, αναφέρεται η περίπτωση των Ηνωμένων Πολιτειών Αμερικής (εφεξής ΗΠΑ) και η περίπτωση της Κίνας, οι οποίες αποτελούν τις ηγέτιδες δυνάμεις στην TN.

Οι ΗΠΑ έχουν καταστήσει σαφή από νωρίς τη θέση τους κατά της επεμβατικής ρύθμισης της τεχνολογίας και του διαδικτύου από τους κυβερνητικούς φορείς. Ωστόσο, από το 2019, θέλοντας να συμβαδίσουν με την τάση ρύθμισης σε παγκόσμιο επίπεδο και να διατηρήσουν τον πρωταγωνιστικό τους ρόλο, άρχισαν να συντάσσουν κείμενα και πολιτικές και να συστήνουν επιτροπές σχετικά με την TN σε κυβερνητικό επίπεδο, όπως έγινε με το Εθνικό Συμβούλιο Επιστήμης και Τεχνολογίας (National Science and Technology Council) και το Υπουργείο Άμυνας, ενώ μέχρι τότε οποιοδήποτε κείμενο υπήρχε, προερχόταν από τους τεχνολογικούς κολοσσούς στον ιδιωτικό τομέα, όπως η Google, προβλέποντας ο καθένας τη δική του πολιτική (Daly, et al., 2022). Επιπλέον, το 2020 ο Λευκός Οίκος δημοσίευσε έναν κατάλογο με οδηγίες (Hao, 2020) προς τα ομοσπονδιακά γραφεία, θεσπίζοντας ορισμένες κατευθυντήριες γραμμές για την εξασφάλιση της αξιοπιστίας της TN, προτρέποντάς τα, ωστόσο, να μη θεσπίζουν ρυθμίσεις προληπτικά, οι οποίες πρόκειται να περιορίσουν την καινοτομία της TN, αλλά αντίθετα να προβαίνουν σε ρυθμίσεις κατά του περιορισμού της καινοτομίας στην TN (Daly, et al., 2022).

Η Κίνα, το αντίπαλο δέος των ΗΠΑ, έχοντας παράδοση στην κρατική επέμβαση και ρύθμιση όλων των τομέων και εκφάνσεων της σύγχρονης ζωής, έχει παράξει πληθώρα κειμένων με ρυθμιστικούς κανόνες για την ΤΝ, αλλά και αναλυτικό πλάνο, με στόχο μέχρι το 2025 να έχει θεσπίσει νέους νόμους και κανόνες για την ανάπτυξη της ΤΝ στην Κίνα (Daly, et al., 2022). Εκτός αυτού, έχει συστήσει εθνικές επιτροπές, οι οποίες συνεργάζονται στενά με ιδιωτικές εταιρίες-κολοσσούς, όπως η Alibaba, διαμορφώνοντας πολιτικές σε συμμόρφωση των ηθικών προαπαιτούμενων που έχουν τεθεί κρατικά. Είναι σαφής η πρόθεσή της να εδραιώσει τη θέση της ως ηγέτιδα στο πεδίο της ΤΝ, τόσο σε τεχνολογικό, όσο και σε ρυθμιστικό επίπεδο (Singh, 2021).

1.5.3 Προσπάθειες ρύθμισης στην ΕΕ

Όσον αφορά ειδικότερα την ΕΕ, αν και υστερεί στην ανάπτυξη συστημάτων ΤΝ σε σχέση με την Κίνα και τις ΗΠΑ, παρατηρείται μια συντονισμένη προσπάθεια, ώστε αυτά να ρυθμιστούν τόσο μέσω κατευθυντήριων οδηγιών, όσο και σε κανονιστικό επίπεδο, με δεσμευτικού χαρακτήρα κανόνες. Ξεκινώντας με το Γενικό Κανονισμό Προστασίας Δεδομένων (εφεξής GDPR) (Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο, 2016), η ΕΕ έθεσε βασικά κατώτατα πρότυπα προστασίας των φυσικών προσώπων σχετικά με την επεξεργασία των προσωπικών τους δεδομένων, τα οποία έχουν εφαρμογή και στα συστήματα ΤΝ. Στη συνέχεια, η εστίαση της ΕΕ στη ρύθμιση της ΤΝ θεμελιώθηκε με τη δημιουργία της ΟΕΥΕ για την ΤΝ, αλλά και την έκδοση της Λευκής Βίβλου για την ΤΝ (Ευρωπαϊκή Επιτροπή, 2020). Το Ψήφισμα του Ευρωπαϊκού Κοινοβουλίου της 20ής Οκτωβρίου 2020 με συστάσεις προς την Επιτροπή όσον αφορά το καθεστώς αστικής ευθύνης για την ΤΝ (2020/2014(INL)) (Ευρωπαϊκό Κοινοβούλιο, 2020) εστιάζει στην ευθύνη για ζημιές που προκαλούνται από συστήματα ΤΝ, ενώ, περαιτέρω, η Πράξη της ΕΕ για τις Ψηφιακές Υπηρεσίες (Digital Services Act) (European Commission, 2022), η πράξη της ΕΕ για τις Ψηφιακές Αγορές (Digital Markets Act) (Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο, 2020) και ο Κανονισμός (ΕΕ) 2019/1150 του Ευρωπαϊκού Κοινοβουλίου Και του Συμβουλίου, της 20ής Ιουνίου 2019, για την προώθηση της δίκαιης μεταχείρισης και της διαφάνειας για τους επιχειρηματικούς χρήστες επιγραμμικών υπηρεσιών διαμεσολάβησης (P2B Regulation) (Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο, 2019), αν και δεν έχουν ως αποκλειστικό αντικείμενο τη ρύθμιση της ΤΝ και δεν εστιάζουν σε αυτήν, αναφέρονται ρητά στην αλγοριθμική λήψη

αποφάσεων και προβλέπουν ειδικές υποχρεώσεις για τη διασφάλιση της διαφάνειας και επεξηγησιμότητας στα συστήματα ΤΝ.

Αποκορύφωμα, βέβαια, των στόχων και της πολιτικής της ΕΕ αποτελεί η νέα Πράξη για την ΤΝ της ΕΕ που αναφέρθηκε ανωτέρω (Ευρωπαϊκή Επιτροπή, 2021), η οποία επιχειρεί την προώθηση της εξέλιξης της ΤΝ και την ευρεία ρύθμισή της, αλλά και οι κατευθυντήριες γραμμές για την επίτευξη αξιόπιστης ΤΝ, που έχει καταρτίσει από το 2019 η ΟΕΥΕ για την ΤΝ (High-Level Expert Group on Artificial Intelligence, 2019), τις οποίες και υιοθέτησε η Πράξη για την ΤΝ.

Η Πράξη για την ΤΝ αξιολογεί τα συστήματα ΤΝ με βάση τον κίνδυνο που εμφανίζουν. Πιο συγκεκριμένα, στο επίπεδο του μη αποδεκτού κινδύνου εντάσσονται τα συστήματα ΤΝ που θεωρείται ότι συνιστούν σαφή απειλή για την ασφάλεια, το βιοπορισμό και τα δικαιώματα των ατόμων· αυτά τα συστήματα ΤΝ απαγορεύονται ρητά. Εδώ περιλαμβάνονται συστήματα ή εφαρμογές ΤΝ που χειραγωγούν την ανθρώπινη συμπεριφορά για να παρακάμψουν την ελεύθερη βούληση των χρηστών/χρηστριών και συστήματα που επιτρέπουν την «κοινωνική βαθμολόγηση» (social credit systems) από τις κυβερνήσεις.

Στο επίπεδο του υψηλού κινδύνου εντάσσονται τα συστήματα ΤΝ που χρησιμοποιούνται στις υποδομές ζωτικής σημασίας (π.χ. μεταφορές), στην εκπαίδευση ή την επαγγελματική κατάρτιση, στην απασχόληση και στη διαχείριση εργαζομένων (π.χ. λογισμικό διαλογής βιογραφικών σημειωμάτων για διαδικασίες πρόσληψης), στις βασικές ιδιωτικές και δημόσιες υπηρεσίες, στην επιβολή του νόμου, στη διαχείριση της μετανάστευσης, του ασύλου και των συνοριακών ελέγχων, στην απονομή δικαιοσύνης και στις δημοκρατικές διαδικασίες. Τα συστήματα ΤΝ υψηλού κινδύνου θα υπόκεινται σε αυστηρές προϋποθέσεις, προτού επιτραπεί η διάθεσή τους στην αγορά. Ειδικότερα, απαιτούνται κατάλληλα συστήματα εκτίμησης και μετριασμού των κινδύνων, υψηλή ποιότητα των συνόλων δεδομένων που τροφοδοτούν το σύστημα ώστε να ελαχιστοποιούνται οι κίνδυνοι και οι διακρίσεις (biases), καταγραφή των δραστηριοτήτων με σκοπό τη διασφάλιση της ιχνηλασιμότητας των αποτελεσμάτων, λεπτομερής τεκμηρίωση που παρέχει όλες τις απαραίτητες πληροφορίες σχετικά με το σύστημα και τον σκοπό του ώστε οι αρχές να αξιολογούν τη συμμόρφωσή του, σαφείς και επαρκείς πληροφορίες για τον χρήστη/χρήστρια, κατάλληλα μέτρα ανθρώπινης εποπτείας για την ελαχιστοποίηση του κινδύνου, υψηλό επίπεδο ανθεκτικότητας, ασφάλειας και ακρίβειας.

Ειδικότερη αναφορά γίνεται στα συστήματα εξ αποστάσεως βιομετρικής ταυτοποίησης, καθώς τονίζεται ότι όλα τα συστήματα εξ αποστάσεως βιομετρικής ταυτοποίησης θεωρούνται υψηλού κινδύνου και υπόκεινται σε αυστηρές απαιτήσεις. Η χρήση τους σε πραγματικό χρόνο σε δημοσίως προσβάσιμους χώρους για σκοπούς επιβολής του νόμου καταρχήν απαγορεύεται. Οι ελάχιστες εξαιρέσεις σ' αυτόν τον κανόνα καθορίζονται και ρυθμίζονται αυστηρά (για παράδειγμα, όταν είναι απολύτως αναγκαίο για την αναζήτηση αγνοούμενου παιδιού, για την πρόληψη συγκεκριμένης και επικείμενης τρομοκρατικής απειλής ή για τον εντοπισμό, την επισήμανση, την ταυτοποίηση ή τη δίωξη δράστη ή υπόπτου σοβαρής αξιόποινης πράξης). Η χρήση αυτή υπόκειται στην έγκριση δικαστικής ή άλλης ανεξάρτητης αρχής και σε κατάλληλους περιορισμούς όσον αφορά το χρονικό διάστημα, τη γεωγραφική εμβέλεια και τις βάσεις δεδομένων όπου πραγματοποιείται αναζήτηση.

Στο επίπεδο του περιορισμένου κινδύνου, εντάσσονται τα συστήματα TN που υπόκεινται σε συγκεκριμένες υποχρεώσεις διαφάνειας, για παράδειγμα, όταν ένας χρήστης αλληλεπιδρά με ένα σύστημα TN, όπως ένα chatbot, θα πρέπει να το γνωρίζει, ώστε να είναι σε θέση να αποφασίσει αν επιθυμεί να συνεχίσει την αλληλεπίδραση ή όχι.

Τέλος, στο επίπεδο του ελάχιστου κινδύνου εντάσσονται τα συστήματα TN που συνεπάγονται μόνο ελάχιστο ή μηδενικό κίνδυνο για τα δικαιώματα και την ασφάλεια των πολιτών, όπως τα βιντεοπαιχνίδια που βασίζονται στην TN ή τα φίλτρα ανεπιθύμητης αλληλογραφίας (Ευρωπαϊκή Επιτροπή, 2021).

Η ανωτέρω προσέγγιση της Πρότασης Κανονισμού περί κινδύνων στα συστήματα TN, φαίνεται να είναι απόρροια της αντίληψης και πεποίθησης ότι τα συστήματα TN δημιουργούνται προς εξυπηρέτηση του ανθρώπου, επομένως το συμφέρον του τίθεται στο επίκεντρο οποιασδήποτε ρύθμισης. Η ανθρωποκεντρική αυτή προσέγγιση, μάλιστα, προκύπτει ρητά από τις κατευθυντήριες γραμμές της ΟΕΥΕ για την TN, η οποία από το 2019 έχει καταρτίσει αναλυτικό κατάλογο με κατευθυντήριες γραμμές για την επίτευξη αξιόπιστης TN, τις οποίες έλαβε υπόψη της η Ευρωπαϊκή Επιτροπή στη σύνταξη της Πρότασης Κανονισμού.

Σύμφωνα με τις κατευθυντήριες γραμμές της ΟΕΥΕ για την TN, για να θεωρηθεί αξιόπιστη η TN πρέπει να είναι:

«α) σύννομη, δηλαδή να τηρεί όλες τις εφαρμοστέες νομοθετικές και κανονιστικές διατάξεις,
β) δεοντολογική, δηλαδή να διασφαλίζει τη συμμόρφωση με δεοντολογικές αρχές και αξίες και
γ) στιβαρή τόσο από τεχνική όσο και από κοινωνική άποψη, διότι ακόμα και όταν υπάρχει καλή πρόθεση, τα συστήματα ΤΝ μπορούν να προκαλέσουν ακούσια βλάβη.»

Τονίζεται περαιτέρω, ότι τα συστήματα ΤΝ πρέπει αφενός να είναι ανθρωποκεντρικά, με την έννοια ότι τίθενται στην υπηρεσία της ανθρωπότητας και του κοινού καλού με στόχο τη βελτίωση της ευημερίας του ανθρώπου, αλλά και αφετέρου να ενισχύουν τις γνωστικές, κοινωνικές και πολιτισμικές του δεξιότητες.

Βάση της σκέψης της ΟΕΥΕ αποτέλεσε η αρχή του σεβασμού προς τα θεμελιώδη ανθρώπινα δικαιώματα, όπως αυτά κατοχυρώνονται από το δίκαιο της ΕΕ, αλλά και από διεθνή κείμενα. Αυτά αφορούν συγκεκριμένα την ανθρώπινη αξιοπρέπεια, την ελευθερία του ατόμου, το σεβασμό της δημοκρατίας, της δικαιοσύνης και του κράτους δικαίου, την ισότητα, την απαγόρευση των διακρίσεων και την αλληλεγγύη και τα δικαιώματα των πολιτών.

Έμφαση δίνεται, περαιτέρω, στην πολυμορφία, την ισότητα, την απαγόρευση των διακρίσεων και τη δικαιοσύνη, που πρέπει να διέπουν ένα σύστημα ΤΝ, καθώς η ΟΕΥΕ για την ΤΝ αναγνωρίζει ρητά ότι τα σύνολα δεδομένων που χρησιμοποιούνται από συστήματα ΤΝ (τόσο για την εκπαίδευση όσο και για τη λειτουργία τους) ενδέχεται να περιλαμβάνουν εκ παραδρομής προϋπάρχοντα σφάλματα μεροληψίας, να είναι ελλιπή και να βασίζονται σε ακατάλληλα μοντέλα διακυβέρνησης. Η συνέχιση των μεροληψιών αυτού του είδους θα μπορούσε να οδηγήσει σε ακούσιες άμεσες (και έμμεσες) προκαταλήψεις και διακρίσεις κατά ορισμένων ομάδων ή ανθρώπων, επιδεινώνοντας ενδεχομένως τις υφιστάμενες προκαταλήψεις και την περιθωριοποίηση (High-Level Expert Group on Artificial Intelligence, 2019).

Συνοψίζοντας, η ΤΝ είναι μια περίπλοκη και ελκυστική νέα τεχνολογία που προσφέρει σημαντικές διευκολύνσεις και δυνατότητες, ενώ είναι και πολύ περισσότερο διαδεδομένη από ό,τι γίνεται αντιληπτό, επηρεάζοντας και αλληλεπιδρώντας με πολλούς και διαφορετικούς

τομείς της σύγχρονης ζωής. Ωστόσο, αυτά τα πλεονεκτήματα δεν πρέπει να αποπροσανατολίσουν από τα σφάλματα της ΤΝ και τους κινδύνους που φέρει για το θεμελιώδες δικαίωμα της απαγόρευσης των διακρίσεων και ιδίως των έμφυλων, όπως άλλωστε φροντίζει να επισημαίνει με κάθε ευκαιρία η διεθνής επιστημονική κοινότητα. Η έντονη κινητικότητα που παρατηρείται σχετικά τόσο με την παραγωγή όσο και με τη ρύθμιση της ΤΝ αυτή τη στιγμή, προσφέρει μια σημαντική ευκαιρία, ώστε να μελετηθούν σε βάθος οι μεροληψίες (biases) και οι διακρίσεις στις οποίες προβαίνουν τα συστήματα ΤΝ, ιδίως οι έμφυλες, να αναδειχθούν ως κοινωνικό-τεχνολογικό (Leslie, et al., 2021) πρόβλημα και να καταπολεμηθούν ήδη από την απαρχή της ψηφιοποιημένης κοινωνίας που φέρνει η ΤΝ.

ΚΕΦΑΛΑΙΟ 2^ο

ΕΜΦΥΛΕΣ ΔΙΑΚΡΙΣΕΙΣ (GENDER BIASES) ΚΑΙ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

2.1 Εισαγωγικές παρατηρήσεις

Κοινή βάση όλων των προσπαθειών ρύθμισης της ΤΝ μέχρι τώρα, είτε με τη μορφή κατευθυντήριων γραμμών είτε με τη μορφή δεσμευτικών κανόνων, είναι η αδιαμφισβήτητη θέση ότι αυτή τίθεται στην υπηρεσία της ανθρωπότητας, ενώ περαιτέρω, προβάλλεται ως καίρια επιταγή ο σεβασμός προς τα θεμελιώδη ανθρώπινα δικαιώματα, όπως αυτά κατοχυρώνονται από το διεθνές δίκαιο, γενικότερα, και το δίκαιο της ΕΕ και των κρατών, ειδικότερα (Rodrigues, 2020). Τα θεμελιώδη αυτά δικαιώματα αφορούν συγκεκριμένα την ανθρώπινη αξιοπρέπεια, την ελευθερία του ατόμου, το σεβασμό της ιδιωτικότητας και την προστασία των προσωπικών δεδομένων, το σεβασμό της δημοκρατίας, της δικαιοσύνης και του κράτους δικαίου, την ισότητα, την απαγόρευση των διακρίσεων και τα κοινωνικο-οικονομικά δικαιώματα (Leslie, et al., 2021).

Ιδιαίτερη έμφαση δίνεται στην ανάγκη διασφάλισης της ισότητας και της απαγόρευσης των διακρίσεων. Χαρακτηριστικά, η ΟΕΥΕ για την ΤΝ, διατύπωσε από πολύ νωρίς τη θέση ότι μία από τις βασικές προϋποθέσεις που πρέπει να πληρούνται, ώστε να θεωρηθεί αξιόπιστο ένα σύστημα ΤΝ είναι η πολυμορφία, η ισότητα, η απαγόρευση των διακρίσεων και η δικαιοσύνη, συμπεριλαμβανομένων των δικαιωμάτων των ατόμων που αντιμετωπίζουν κίνδυνο αποκλεισμού. Αναγνώρισε, περαιτέρω, ότι τα δεδομένα βάσει των οποίων εκπαιδεύεται ένα σύστημα ΤΝ μπορεί να είναι ανακριβή, ελλιπή και να εμπεριέχουν σφάλματα μεροληψίας (biases), με κίνδυνο την ενίσχυση των ήδη υφιστάμενων προκαταλήψεων. Θεωρεί, επομένως, τον ισότιμο σεβασμό της ηθικής αξίας και της αξιοπρέπειας όλων των ανθρώπων υψίστης σημασίας, ώστε να αποφεύγεται η αθέμιτη μεροληψία. Η ΟΕΥΕ για την ΤΝ προτείνει ότι προκειμένου να επιτευχθεί η αξιόπιστη ΤΝ, ολόκληρος ο κύκλος ζωής του συστήματος ΤΝ θα πρέπει να χαρακτηρίζεται από συμπεριληπτικότητα (inclusivity) και πολυμορφία (diversity), με την έννοια ότι είναι εξαιρετικής σημασίας να συμμετέχουν στη λήψη αποφάσεων και τις διαδικασίες επιστήμονες όλων των φύλων και φυλών και από όλα τα υπόβαθρα. Εκτός από την εξέταση και τη συμμετοχή όλων των επηρεαζόμενων μερών καθ' όλη τη διάρκεια της

διαδικασίας, αυτό συνεπάγεται επίσης την εξασφάλιση ισότιμης πρόσβασης μέσω συμπεριληπτικών διαδικασιών σχεδιασμού, καθώς και μέσω της ίσης μεταχείρισης (High-Level Expert Group on Artificial Intelligence, 2019).

Αντίστοιχες σκέψεις διατυπώνονται και στην Πράξη για την ΤΝ της ΕΕ (Ευρωπαϊκή Επιτροπή, 2021), στην αιτιολογική έκθεση της οποίας αναφέρεται ότι «[...]η πρόταση θα ενισχύσει και θα προωθήσει την προάσπιση των δικαιωμάτων που προστατεύονται από τον Χάρτη, όπως: το δικαίωμα στην ανθρώπινη αξιοπρέπεια (άρθρο 1), τον σεβασμό της ιδιωτικής ζωής και την προστασία των δεδομένων προσωπικού χαρακτήρα (άρθρα 7 και 8), την απαγόρευση των διακρίσεων (άρθρο 21) και την ισότητα γυναικών και ανδρών (άρθρο 23).», αλλά και στην προσφάτως υιοθετηθείσα Πρόταση της UNESCO (UNESCO's General Conference: 41st session, 2021), στο Προοίμιο της οποίας αναφέρεται ότι:

«Λαμβάνοντας υπόψιν ότι οι τεχνολογίες τεχνητής νοημοσύνης μπορούν να προσφέρουν σημαντικές υπηρεσίες στην ανθρωπότητα και ότι όλες οι χώρες μπορούν να επωφεληθούν από αυτές, αλλά επίσης εγείρουν θεμελιώδη ηθικά ζητήματα, για παράδειγμα σχετικά με τις προκαταλήψεις που μπορούν να ενσωματώσουν και να επιδεινώσουν πιθανές διακρίσεις, ανισότητες, ψηφιακό χάσμα, αποκλεισμό και απειλή για την πολιτιστική, κοινωνική και βιολογική ποικιλομορφία και τις κοινωνικές ή οικονομικές διαφορές· την ανάγκη για διαφάνεια και κατανόηση της λειτουργίας των αλγορίθμων και των δεδομένων με τα οποία έχουν εκπαιδευτεί· και τον πιθανό αντίκτυπό τους στα συμπεριλαμβανόμενα παρακάτω, αλλά όχι περιοριστικά, δηλαδή στην ανθρώπινη αξιοπρέπεια, στα ανθρώπινα δικαιώματα και θεμελιώδεις ελευθερίες, στην ισότητα των φύλων, στη δημοκρατία [...]», και ότι:

«Αναγνωρίζοντας επίσης ότι οι τεχνολογίες τεχνητής νοημοσύνης μπορούν να εμβαθύνουν τις υπάρχουσες διαφορές και ανισότητες στον κόσμο [...]».

Οι ανωτέρω σκέψεις, είναι αποτέλεσμα ενός εξαιρετικά έντονου προβληματισμού που υπάρχει στους επιστημονικούς κύκλους ήδη από την αρχή ανάπτυξης της ΤΝ. Ο προβληματισμός αυτός έγκειται αφενός στο γενικότερο βασικό ερώτημα, αν η ΤΝ μπορεί να είναι πραγματικά απαλλαγμένη από προκαταλήψεις και στερεότυπα ή είναι καταδικασμένη να τις φέρει ήδη από τη δημιουργία της, ως ανθρώπινο δημιούργημα, και αφετέρου στο ειδικότερο πρόβλημα των

έμφυλων διακρίσεων (gender biases), συχνότατο φαινόμενο στα συστήματα ΤΝ, και αν η ΤΝ μπορεί να βοηθήσει στην εξάλειψή τους ή οδηγεί στην ενίσχυσή τους.

2.2 Διαφορά έννοιας διακρίσεων (discrimination) και μεροληψιών (biases)

Η έννοια της διάκρισης (discrimination) και η έννοια της μεροληψίας (bias) στην ΤΝ είναι δύο έννοιες με διαφορετική τυπικά σημασία, αν και πολλές φορές διαπλέκονται και είναι αλληλένδετες. Στην παρούσα μελέτη ο όρος «διάκριση» χρησιμοποιείται για να αποδώσει το νόημα της διακριτικής μεταχείρισης και άνισης αντιμετώπισης και ειδικότερα στην περίπτωση των έμφυλων διακρίσεων (βλ. παρακάτω 2.4) και ο όρος «μεροληψία» χρησιμοποιείται για να αποδώσει το νόημα των στερεοτυπικών αντιλήψεων και των στατιστικών αποκλίσεων στην ΤΝ (βλ. παρακάτω 2.2.2 και 2.3). Οι έννοιες αυτές χρησιμοποιούνται σε συνάφεια μεταξύ τους, χρήσιμο είναι ωστόσο, να γίνει ένας διαχωρισμός και ορισμός τους, για μεγαλύτερη σαφήνεια.

2.2.1 Η έννοια των διακρίσεων

Σε επίπεδο ΕΕ, η υποχρέωση σεβασμού της αρχής της απαγόρευσης των διακρίσεων κατοχυρώνεται στο άρθρο 2 της Σύμβασης για την Ευρωπαϊκή Ένωση (εφεξής ΣΕΕ), το οποίο προβλέπει ότι η απαγόρευση των διακρίσεων είναι μια από τις θεμελιώδεις αξίες της ΕΕ και στο άρθρο 10 της Συνθήκης για τη Λειτουργία της Ευρωπαϊκής Ένωσης (εφεξής ΣΛΕΕ), το οποίο προβλέπει ότι «Κατά τον καθορισμό και την εφαρμογή των πολιτικών και των δράσεών της, η Ένωση επιδιώκει να καταπολεμήσει κάθε διάκριση λόγω φύλου, φυλετικής ή εθνοτικής καταγωγής, θρησκείας ή πεποιθήσεων, αναπηρίας, ηλικίας ή γενετήσιου προσανατολισμού.» (Ευρωπαϊκή Ένωση, 2016). Επιπλέον, το άρθρο 20 του Χάρτη Θεμελιωδών Δικαιωμάτων της Ευρωπαϊκής Ένωσης (εφεξής Χάρτης) προβλέπει ότι «Όλοι οι άνθρωποι είναι ίσοι έναντι του νόμου.»· το άρθρο 21 του Χάρτη προβλέπει ότι «Απαγορεύεται κάθε διάκριση ιδίως λόγω φύλου, φυλής, χρώματος, εθνοτικής καταγωγής ή κοινωνικής προέλευσης, γενετικών χαρακτηριστικών, γλώσσας, θρησκείας ή πεποιθήσεων, πολιτικών φρονημάτων ή κάθε άλλης γνώμης, ιδιότητας μέλους εθνικής μειονότητας, παρουσίας, γέννησης, αναπηρίας, ηλικίας ή γενετήσιου προσανατολισμού.»· το άρθρο 23 του Χάρτη προβλέπει ότι «Η ισότητα γυναικών και ανδρών πρέπει να εξασφαλίζεται σε όλους τους τομείς, μεταξύ άλλων στην απασχόληση, την

εργασία και τις αποδοχές. Η αρχή της ισότητας δεν αποκλείει τη διατήρηση ή τη θέσπιση μέτρων που προβλέπουν ειδικά πλεονεκτήματα υπέρ του υποεκπροσωπούμενου φύλου.» (Ευρωπαϊκό Κοινοβούλιο, Συμβούλιο και Επιτροπή, 2016). Τα ανωτέρω συμπληρώνονται και ενισχύονται από ειδικότερες διατάξεις και οδηγίες της ΕΕ με ποικίλα πεδία εφαρμογής. Η απαγόρευση του Χάρτη κατά των διακρίσεων αντικατοπτρίζει τα αντίστοιχα δικαιώματα και στην Ευρωπαϊκή Σύμβαση Δικαιωμάτων του Ανθρώπου (εφεξής ΕΣΔΑ), όπου στο άρθρο 14 προβλέπεται ότι «Η χρήση των αναγνωριζομένων εν τη παρούση Συμβάσει δικαιωμάτων και ελευθεριών δέον να εξασφαλισθή ασχέτως διακρίσεως φύλου, φυλής, χρώματος, γλώσσας, θρησκείας, πολιτικών ή άλλων πεποιθήσεων, εθνικής ή κοινωνικής προελεύσεως, συμμετοχής εις εθνικήν μειονότητα, περιουσίας, γεννήσεως ή άλλης καταστάσεως.», ενώ στο υπ' αριθμόν 12 πρωτόκολλό της προβλέπεται ότι «Η απόλαυση κάθε δικαιώματος που προβλέπεται στον νόμο πρέπει να διασφαλίζεται χωρίς καμία διάκριση, που βασίζεται ιδίως στο φύλο, τη φυλή, το χρώμα, τη γλώσσα, τη θρησκεία, την πολιτική ή άλλη γνώμη, την εθνική ή κοινωνική καταγωγή, την ιδιότητα μέλους εθνικής μειονότητας, την περιουσία, τη γέννηση ή κάθε άλλη κατάσταση.» (Συμβούλιο της Ευρώπης, 1950). Συνεπώς, η διάκριση είναι νομική έννοια και υπάρχει «όπου ένα άτομο αντιμετωπίζεται λιγότερο ευνοϊκά από ό,τι αντιμετωπίζεται, έχει ή θα είχε αντιμετωπιστεί ένα άλλο σε αντίστοιχη κατάσταση, εξαιτίας ενός νομιζόμενου ή πραγματικού προσωπικού χαρακτηριστικού (προστατευόμενοι λόγοι/χαρακτηριστικά)». Οι διακρίσεις μέσω της χρήσης αλγορίθμων θεωρούνται σήμερα από τις πιο πιεστικές προκλήσεις στη χρήση τεχνολογιών που βασίζονται στην ΤΝ. Η μεροληψία και οι διακρίσεις, συμπεριλαμβανομένων των διακρίσεων με βάση το φύλο, όπου λαμβάνονται αποφάσεις βάσει αλγορίθμων, οι οποίοι στηρίζονται σε δεδομένα, μπορούν να εμφανιστούν για διάφορους λόγους και σε πολλά επίπεδα στα συστήματα ΤΝ. Γι' αυτό εξάλλου είναι δύσκολο να εντοπιστούν και να μετρηστούν. Η ποιότητα των δεδομένων και οι ήδη υφιστάμενες προκαταλήψεις είναι συχνά η πηγή των πιθανών διακρίσεων και της άνισης μεταχείρισης (European Union Agency for Fundamental Rights, 2020).

2.2.2 Η έννοια της μεροληψίας (bias)

Η έννοια της μεροληψίας έχει ευρύτερη σημασία. Η «μεροληψία» αναφέρεται σε μια προδιάθεση, προκατάληψη ή στρέβλωση. Τόσο νομικά, όσο και τεχνολογικά, η μεροληψία υποδηλώνει ότι ορισμένες πληροφορίες προτιμώνται και άλλες αγνοούνται (Walsh, et al., 2019).

Ειδικότερα, στον τομέα της τεχνολογίας, η μεροληψία (bias) ορίζεται ως το αποτέλεσμα που στερεί από ένα στατιστικό αποτέλεσμα την αντιπροσωπευτικότητα, διαστρεβλώνοντάς το συστηματικά, σε αντίθεση με ένα τυχαίο σφάλμα που μπορεί να προκαλέσει παραμόρφωση σε κάθε περίπτωση, αλλά εξισορροπείται κατά μέσο όρο (ADMIN Cross-linked glossary, 2019). Επισημαίνεται ότι παρά το γεγονός ότι οι μεροληψίες παίζουν σημαντικό ρόλο σε διαδικασίες που καταλήγουν σε διακρίσεις, η μεροληψία δεν οδηγεί απαραίτητα σε διακριτική μεταχείριση, δεν πληροί πάντα δηλαδή τα στοιχεία που απαιτεί ο νόμος για τη θεμελίωσή της. Ωστόσο, το γεγονός ότι ένα αποτέλεσμα δεν είναι τυπικά παράνομο, δε σημαίνει ότι δεν είναι άδικο, από ηθική και κοινωνική άποψη. Οι μεροληψίες, επομένως, βρίσκονται σε αλληλένδετη σχέση με την έννοια των διακρίσεων και συχνότατα καταλήγουν σε αυτές, βασίζονται όμως σε ένα περισσότερο κοινωνικο-ηθικό υπόβαθρο (Ferrer, et al., 2021). Περαιτέρω, η μεροληψία στην ΤΝ μπορεί να εμφανιστεί σε διάφορα στάδια λειτουργίας του συστήματος και μπορεί να οφείλεται σε διαφορετικούς λόγους.

2.3 Η μεροληψία στην ΤΝ (Biased AI)

Για να θεωρηθεί ότι ένα σύστημα ΤΝ μεροληπτεί (biased), πρέπει να προβαίνει σε διακρίσεις και άνιση μεταχείριση σε συστηματική βάση και όχι να συμβαίνει μία φορά ή περιστασιακά λόγω λάθους (Schwartz, et al., 2022). Και αυτό, διότι είναι πολύ πιθανή η εμφάνιση βλάβης στο σύστημα, η οποία ενδεχομένως να οδηγήσει σε αποτέλεσμα δυσμενούς διάκρισης, αποτελώντας μεμονωμένο περιστατικό και αφορώντας οποιοδήποτε πρόσωπο και όχι άτομο με συγκεκριμένα χαρακτηριστικά ή μια συγκεκριμένη μειονότητα (Friedman & Nissenbaum, 1996).

Η μεροληπτική συμπεριφορά ενός συστήματος ΤΝ, μπορεί να εμφανίζεται με τη μορφή των προϋφιστάμενων προκαταλήψεων, των τεχνικών περιορισμών και των αναδυόμενων προκαταλήψεων. Οι προϋφιστάμενες προκαταλήψεις έχουν σχέση με τα δεδομένα που εισάγονται στο σύστημα ΤΝ και βάσει των οποίων εκπαιδεύεται (training data). Οι προϋφιστάμενες προκαταλήψεις έχουν τις ρίζες τους σε κοινωνικούς θεσμούς, πρακτικές και αντιλήψεις. Πρόκειται, επομένως, για προκαταλήψεις που υπάρχουν ανεξάρτητα, και συνήθως πριν από τη δημιουργία του συστήματος. Οι προϋφιστάμενες προκαταλήψεις μπορεί να προέρχονται από τις κοινωνικές αντιλήψεις που εντοπίζονται σε μεγάλα μέρη του πληθυσμού,

αλλά και σε υποκουλτούρες, και σε επίσημους ή ανεπίσημους, ιδιωτικούς ή δημόσιους οργανισμούς και ιδρύματα (Friedman & Nissenbaum, 1996). Ένα τέτοιο παράδειγμα, είναι η βιομηχανία των βιντεοπαιχνιδιών, όπου τα στερεότυπα περί φύλου που υπάρχουν στην ευρύτερη κοινωνία, ότι δηλαδή τα παιχνίδια αυτού του είδους είναι «αγορίστικη/ανδρική» υπόθεση, οδηγούν στην ανάπτυξη τέτοιων παιχνιδιών που αποσκοπούν στο να είναι περισσότερο ελκυστικά στο ανδρικό φύλο από ό,τι στο γυναικείο (Veltri, et al., 2014). Οι προϋφιστάμενες προκαταλήψεις μπορούν, επίσης, να αντανakλούν τις προσωπικές προκαταλήψεις ατόμων που έχουν σημαντική συμβολή στο σχεδιασμό του συστήματος TN, όπως ο άνθρωπος με τον οποίο αλληλεπιδρά ή ο προγραμματιστής του συστήματος. Αυτός ο τύπος μεροληψίας μπορεί επίσης να εισέλθει σε ένα σύστημα μέσω των ρητών και συνειδητών προσπαθειών ατόμων ή ιδρυμάτων, ή σιωπηρά και ασυνείδητα, ακόμη και παρά τις καλύτερες προθέσεις.

Οι τεχνικοί περιορισμοί αναφέρονται σε ζητήματα σχετικά με τον τεχνικό σχεδιασμό, λόγω του οποίου ενδέχεται να προκύψουν μεροληπτικές συμπεριφορές. Τέτοιοι περιορισμοί μπορούν να συμπεριλαμβάνουν το υλισμικό (hardware) ή/και το λογισμικό (software), για παράδειγμα σε μια βάση δεδομένων για την αντιστοίχιση δωρητών οργάνων με δυνητικούς παραλήπτες μοσχευμάτων, εφόσον το σύστημα είναι σχεδιασμένο με τέτοιο τρόπο ώστε οι πληροφορίες για ορισμένα άτομα να ανακτώνται και να εμφανίζονται στις αρχικές θέσεις αποτελεσμάτων, αυτό σημαίνει ότι ευνοούνται συστηματικά και έχουν περισσότερες πιθανότητες για αντιστοίχιση έναντι των ατόμων που εμφανίζονται στις τελευταίες θέσεις. Οι τεχνικοί περιορισμοί μπορούν να αναφέρονται και στη δυσκολία του να αντιληφθεί και να εφαρμόσει ο αλγόριθμος σωστά αόριστες έννοιες και περίπλοκα ζητήματα, που ο ανθρώπινος νους σταθμίζει βάσει διαφόρων παραγόντων, εμπειρίας και κοινής λογικής. Ένα τέτοιο παράδειγμα είναι η λήψη δικαστικών αποφάσεων, διότι η ερμηνεία του νόμου και η κρίση επί των υποθέσεων δε βασίζεται μόνο στα κείμενα και τα έγγραφα, αλλά γίνονται σταθμίσεις αντικρουόμενων δικαιωμάτων και συμφερόντων, εκτίμηση της συμπεριφοράς και των πράξεων των διαδίκων, λαμβάνεται υπόψη η κοινή πείρα και πολλές φορές ακόμα και το ένστικτο, έννοιες και διαδικασίες που είναι δύσκολο έως αδύνατο να ποσοτικοποιηθούν, κάνοντας τη χρήση TN στη δικαιοσύνη περίπλοκη υπόθεση (Friedman & Nissenbaum, 1996).

Οι αναδυόμενες προκαταλήψεις μπορούν να εμφανιστούν σε κάποια χρονική στιγμή ύστερα της έναρξης λειτουργίας τους συστήματος και συνδέονται με το συγκεκριμένο πλαίσιο στο οποίο χρησιμοποιείται το σύστημα. Αυτού του είδους οι μεροληψίες εμφανίζονται συνήθως είτε αν αλλάξουν οι κοινωνικές και πολιτισμικές αντιλήψεις σχετικά με τα δεδομένα που εισήχθησαν στο σύστημα κατά τη δημιουργία του είτε αν το σύστημα χρησιμοποιείται με διαφορετικό τρόπο και σε διαφορετικό χρόνο ή/και τόπο από ό,τι προοριζόταν (The Institute of Technological Ethics, 2021). Για παράδειγμα, αν ένα μηχάνημα απαιτεί από τους χρήστες να διαβάζουν, να γράφουν, να κατανοούν αριθμούς ή να αντιλαμβάνονται μεταφορές και αυτό τοποθετηθεί σε μέρος όπου ο πληθυσμός σε μεγάλο ποσοστό δεν έχει βασική εκπαίδευση, τότε αναπόφευκτα θα μεροληπτεί εις βάρος αυτών των ατόμων (Friedman & Nissenbaum, 1996).

Από τα τρία είδη μεροληψίας που αναφέρθηκαν παραπάνω, το μεγαλύτερο ενδιαφέρον παρουσιάζουν οι προϋφιστάμενες προκαταλήψεις, δηλαδή οι μεροληπτικές συμπεριφορές που προκύπτουν από τα δεδομένα που εισάγονται στο σύστημα TN προς επεξεργασία και με βάση τα οποία παίρνει αποφάσεις.

Χαρακτηριστικό παράδειγμα και ίσως πιο γνωστή περίπτωση αποτελεί το COMPAS, το σύστημα TN που χρησιμοποιήθηκε στην κομητεία Broward της Φλόριντα των ΗΠΑ, για να προβλέψει το κατ' εξακολούθηση έγκλημα και κατέληξε να χαρακτηρίσει λανθασμένα τους Αφροαμερικανούς κατηγορούμενους ως «υψηλού κινδύνου», σε ποσοστό σχεδόν διπλάσιο από το ποσοστό που απέδωσε λανθασμένα τον ίδιο χαρακτηρισμό στους λευκούς κατηγορούμενους (Angwin, et al., 2016), αντανακλώντας έτσι την προκατάληψη που υπάρχει στην κοινωνία των ΗΠΑ σχετικά με την εγκληματικότητα της μαύρης κοινότητας.

Άλλη χαρακτηριστική περίπτωση αποτελεί η έρευνα του Πανεπιστημίου της Washington, η οποία απέδειξε ότι μετά από αναζήτηση εικόνων στη μηχανή αναζήτησης της Google (η οποία λειτουργεί βάσει αλγορίθμου) για το επάγγελμα του CEO (Chief Executive Officer), μόνο το 11% των εκατό πρώτων αποτελεσμάτων ήταν γυναίκες, αντανακλώντας κατά αυτόν τον τρόπο την κοινωνική αντίληψη, ότι σε αυτή τη θέση οι άντρες είναι πιο αξιόπιστοι και κατ' επέκταση τα δεδομένα για τους άντρες είναι περισσότερα από ό,τι για τις γυναίκες, με αποτέλεσμα να εμφανίζονται περισσότερα αποτελέσματα για αυτούς (Langston, 2015).

Έντονες αντιδράσεις είχε προκαλέσει και το chatbot TN της Microsoft, με το όνομα Tay. Το chatbot αφέθηκε ελεύθερο στο Twitter, με σκοπό να αλληλεπιδράσει με τους χρήστες, να συζητήσει μαζί τους, να μάθει από αυτούς και να αναπτύξει δικές του σκέψεις. Το αποτέλεσμα ήταν, ότι μέσα σε 24 ώρες από τη λειτουργία του, το chatbot κατέληξε να κάνει ρατσιστικά, στερεοτυπικά και σεξιστικά σχόλια, υποστηρίζοντας ακόμα και το Χίτλερ, καθώς μεγάλο μέρος των χρηστών με τους οποίους ήρθε σε επαφή, εξέφραζαν έντονα αυτές τις απόψεις, με αποτέλεσμα να μιμηθεί τη συμπεριφορά τους (Hunt, 2016). Με αφορμή το Tay, ήρθε και πάλι στο προσκήνιο η συζήτηση του κινδύνου που ενέχει η TN να «κληρονομήσει» από την ανθρωπότητα, όχι μόνο τα θετικά και ξεχωριστά της χαρακτηριστικά, αλλά και όλα τα αρνητικά και προβληματικά χαρακτηριστικά που την ακολουθούν, όπως οι προκαταλήψεις και τα στερεότυπα.

Σε αυτό το πλαίσιο, οι ερευνητές Stephen Cave και Kanta Dihal του Κέντρου Leverhulme για το Μέλλον της Νοημοσύνης στο Κέιμπριτζ θέτουν ένα καίριο ερώτημα, το οποίο σχετίζεται με τον τρόπο με τον οποίο οι υπάρχουσες εφαρμογές και αναπαραστάσεις της TN προάγουν τη «λευκότητα» (whiteness), παρατηρώντας ότι αποδίδονται συγκεκριμένα φυλετικά χαρακτηριστικά στην TN και ειδικότερα αυτά της λευκής φυλής, κατατάσσοντάς τη σε αυτήν. Συγκεκριμένα, αναφέρουν σχετικά με το παράδειγμα των βοηθών φωνής (voice assistants), ότι αν και δεν υπάρχει σαφής υλική αναπαράσταση, η φθογγολογική αποτύπωση της φωνής τους, η προφορά τους και ο τρόπος ομιλίας τους, προσομοιάζει σε καυκάσια αγγλόφωνη γυναίκα. Αναφέρονται επίσης στις υλικές αναπαραστάσεις της TN, μέσω των κυβερνητικών οργανισμών (cyborgs) στον κινηματογράφο, οι οποίοι είναι όλοι λευκοί. Οι λόγοι που συμβαίνει αυτό, είναι ότι αφενός στην πλειοψηφία τους οι μεγαλύτερες εταιρίες που παράγουν TN αποτελούνται από λευκούς άνδρες και αφετέρου η TN αναπαρίσταται ως λευκή, γιατί θεωρείται ότι έχει χαρακτηριστικά, τα οποία θεωρεί και η λευκή φυλή ότι έχει για τον εαυτό της· συγκεκριμένα, τα χαρακτηριστικά αυτά είναι η νοημοσύνη, ο επαγγελματισμός και η δύναμη, έναντι των άλλων φυλών. Τέλος, ως τελευταίο λόγο, προβάλλουν τη λευκή ουτοπία που θα ήθελε ιδανικά να δημιουργήσει η λευκή φυλή (Cave & Dihal, 2020).

Όλα όσα αναφέρθηκαν ανωτέρω, καταδεικνύουν ότι η TN και η τεχνολογία γενικότερα, ως τεχνητά δημιουργήματα, τα οποία αποτέλεσαν έμπνευση των ανθρώπων και κατασκευάστηκαν από αυτούς, είναι πολύ πιθανό να φέρουν τα χαρακτηριστικά και τις

αντιλήψεις των δημιουργών τους. Σε γενικότερο πλαίσιο, αντανακλούν στην ουσία τις αντιλήψεις της κοινωνίας. Αυτό είναι εξαιρετικά επικίνδυνο για τη διαίωση στερεοτύπων και την εμβάθυνση κάθε είδους χάσματος, όπως συμβαίνει ειδικότερα με τη θεώρηση του γυναικείου φύλου και τις έμφυλες διακρίσεις.

2.4 TN και Έμφυλες Διακρίσεις (Gender Biases)

Ως έμφυλη διάκριση ορίζεται οποιαδήποτε κατάσταση στην οποία ένα άτομο αντιμετωπίζεται λιγότερο ευνοϊκά από ένα άλλο εξαιτίας του φύλου του (Cambridge Dictionary, χ.χ.), πρόκειται επομένως για οποιαδήποτε άνιση μεταχείριση με βάση το φύλο και μπορεί επίσης να αναφέρεται ως σεξισμός (Salvini, 2014). Το πρόβλημα των έμφυλων διακρίσεων που εμφανίζει η TN, είναι περισσότερο εμφανές στην περίπτωση των ψηφιακών βοηθών, όπως η Siri της Apple και η Alexa της Amazon, καθώς, όπως θα αναλυθεί εκτενώς στη συνέχεια, παρουσιάζουν ορισμένα χαρακτηριστικά, όπως η φωνή, ο τρόπος ομιλίας και απόκρισης, τα ονόματα και ο σκοπός λειτουργίας τους (η βοήθεια και η εξυπηρέτηση αναγκών του χρήστη), τα οποία αντανακλούν έμφυλα στερεότυπα σε βάρος του γυναικείου φύλου (UNESCO, 2020). Κοινή θέση είναι ότι οι έμφυλες διακρίσεις που εμφανίζει η TN, οφείλονται στο παγκόσμιο έμφυλο χάσμα (gender gap) αναφορικά με την TN και την τεχνολογία γενικότερα· σύμφωνα με την αναφορά του Παγκόσμιου Οικονομικού Φόρουμ για το παγκόσμιο έμφυλο χάσμα για το 2021, οι γυναίκες που απασχολούνται στο χώρο των δεδομένων και της TN ανέρχονται σε ποσοστό 32,4% παγκοσμίως, ενώ συγκεκριμένα στο χώρο της TN το ποσοστό συμμετοχής των γυναικών είναι κάτω του 25% (World Economic Forum, 2021). Επισημαίνεται ότι στην αντίστοιχη αναφορά για το 2018, το ποσοστό συμμετοχής των γυναικών στο χώρο της TN ανερχόταν μόλις στο 22% (World Economic Forum, 2018), επομένως όπως γίνεται αντιληπτό, το ποσοστό σήμερα παραμένει στα ίδια επίπεδα.

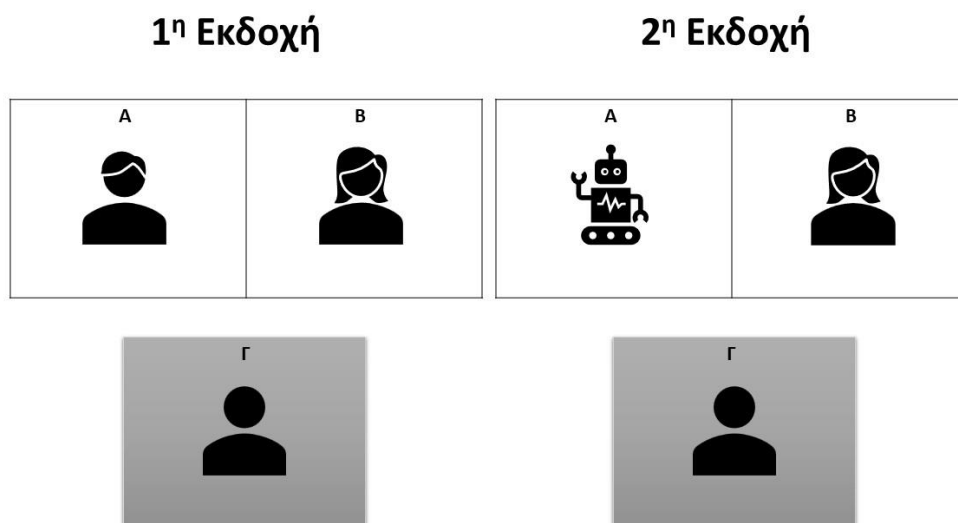
Οι έμφυλες διακρίσεις της TN προκύπτουν ως αποτέλεσμα των ήδη υαρχόντων την ανθρώπινη κοινωνία έμφυλων στερεοτύπων, τα οποία μεταφέρονται στα συστήματα TN είτε από τους δημιουργούς τους είτε από την αλληλεπίδραση με τους χρήστες τους (Madgavkar, 2021). Η βασικότερη αιτία για τις έμφυλες διακρίσεις που εμφανίζει η TN είναι η μεγάλη έλλειψη και ανισορροπία δεδομένων σχετικά με το γυναικείο αλλά και το μη-δυναδικό φύλο (non-binary),

αφού από τη στιγμή που ένα σύστημα TN εκπαιδεύεται με ένα σύνολο δεδομένων, το οποίο περιέχει πληροφορίες ως επί το πλείστον για το ανδρικό φύλο, αναμενόμενο είναι να παρουσιάζει ανακρίβειες σε σχέση με τα αποτελέσματα που παράγει για οποιοδήποτε άλλο φύλο (Madgavkar, 2021). το γεγονός αυτό εκτός από άνισο, αναδεικνύεται και ως εξαιρετικά επικίνδυνο στις περιπτώσεις χρήσης συστημάτων TN στην ιατρική, όπου η έλλειψη δεδομένων και τα ανακριβή αποτελέσματα θέτουν σε κίνδυνο ανθρώπινες ζωές (Niethammer, 2020). Άλλος σημαντικός παράγοντας είναι η φυσική γλώσσα που χρησιμοποιείται για την εκπαίδευση των συστημάτων TN (Natural Language Processing – NLP), καθώς ακόμα και η γλώσσα, οι λέξεις που χρησιμοποιούνται στην καθημερινότητα για να προσδώσουν συγκεκριμένες ιδιότητες στα άτομα και να τα περιγράψουν, ενσωματώνουν έμφυλα στερεότυπα, με αποτέλεσμα να μεταφέρονται αυτά στα συστήματα TN με τη μέθοδο της Επεξεργασίας Φυσικής Γλώσσας (Bolkubasi, et al., 2016). Τέλος, πολύ σημαντικό ρόλο για την κατάρριψη ή ενίσχυση των έμφυλων διακρίσεων της TN, παίζει η γενικότερη τάση θηλυκοποίησης της TN που παρατηρείται στις ψηφιακές βοηθούς και την ενσώματη αναπαράστασή της, αλλά και στην αποτύπωσή της στην επιστημονική φαντασία και τη σύγχρονη κουλτούρα, καθώς εδραιώνονται στην ανθρώπινη κοινωνία και δημιουργούν διαστρεβλωμένη εικόνα για την πραγματική φύση της TN (Madgavkar, 2021). Σε όλα τα ανωτέρω ζητήματα, καίρια είναι η παρουσία και το έργο των θεωρητικών των φεμινιστικών σπουδών, καθώς έχουν συμβάλει στην ανάδειξή τους και έχουν προσφέρει μια νέα ματιά που μέχρι τώρα έλειπε.

2.5 Το Παιχνίδι της Μίμησης (των φύλων)

Η περίπλοκη και προβληματική σχέση της TN με το φύλο και τις αναπαραστάσεις του, ξεκινάει από πολύ παλιά, ήδη από το 1950, όταν ο Alan Turing δημοσίευσε την περίφημη μελέτη του, γνωστή ως «Το Παιχνίδι της Μίμησης», ή αλλιώς ως «Τεστ Turing». Στη μελέτη αυτή, ο Alan Turing διατυπώνει τη θέση ότι οι μηχανές TN μπορούν να μιμηθούν την ανθρώπινη συμπεριφορά και προτείνει την ακόλουθη δοκιμή, ώστε να διαπιστωθεί το επίπεδο ευφυΐας τους. Στην πρώτη εκδοχή του τεστ παίρνουν μέρος τρεις άνθρωποι παράγοντες, ένας εξεταστής (παράγοντας Γ) απροσδιοριστού φύλου, ένας άνδρας (παράγοντας Α) και μία γυναίκα (παράγοντας Β). Σκοπός του εξεταστή Γ είναι να μαντέψει σωστά το φύλο των άλλων

δύο παραγόντων· σκοπός του A είναι να ξεγελάσει τον εξεταστή ώστε να πιστέψει ότι είναι γυναίκα· σκοπός της B είναι να απαντάει με ειλικρίνεια ώστε να βοηθήσει τον εξεταστή Γ. Προχωρώντας περαιτέρω τη σκέψη του, ο Alan Turing προτείνει μια δεύτερη εκδοχή του πειράματος, δηλαδή την αντικατάσταση του άνδρα-παραγόντα A με ένα σύστημα TN. Έτσι, στη δεύτερη εκδοχή του τεστ, παίρνουν μέρος και πάλι τρεις παράγοντες, αυτή τη φορά ένας άνθρωπος-εξεταστής (παραγόντας Γ) απροσδιορίστου φύλου, ένα σύστημα TN (παραγόντας A) και ένας άνθρωπος-γυναίκα (παραγόντας B). Σκοπός του ανθρώπου-εξεταστή Γ είναι να διακρίνει ποιος από τους δύο παράγοντες είναι το σύστημα TN και ποιος είναι ο άνθρωπος-γυναίκα· σκοπός του συστήματος TN είναι να ξεγελάσει τον εξεταστή, ώστε αυτός να νομίσει ότι πρόκειται για τον άνθρωπο-γυναίκα και όχι για ένα σύστημα TN· σκοπός της B είναι να απαντάει με ειλικρίνεια ώστε να βοηθήσει τον εξεταστή Γ. Οι τρεις αυτοί παράγοντες βρίσκονται σε διαφορετικά δωμάτια, χωρίς οπτική και ακουστική επαφή. Αν ο εξεταστής δε μπορέσει να ξεχωρίσει το σύστημα TN από τον άνθρωπο-γυναίκα, τότε αυτό περνάει με επιτυχία το Τεστ, που σημαίνει ότι είναι πλήρως ανεπτυγμένη και κατάφερε να ξεγελάσει τον άνθρωπο-εξεταστή (Turing, 2020).



Εικόνα 2: Οι δύο εκδοχές του Παιχνιδιού της Μίμησης

Η ανωτέρω σκέψη του Alan Turing θεωρείται σημείο αναφοράς για το πεδίο της TN, καθώς αφενός αποτελεί την πρώτη ιστορικά αναφορά της ιδέας της TN, αφού σαν όρος διατυπώθηκε έξι χρόνια αργότερα στο συνέδριο του Πανεπιστημίου του Dartmouth, ενώ ο σημαντικός αντίκτυπος του τεστ, παραμένει ισχυρός ακόμα και σήμερα, μιας και χρησιμοποιείται ακόμα για τη μέτρηση της νοημοσύνης ενός συστήματος TN (Taulli, 2020). Με μια πιο προσεκτική και κριτική ανάγνωση του θεμελιώδους αυτού κειμένου του Alan Turing, όμως, εύλογα δημιουργείται το ερώτημα για ποιο λόγο ο Alan Turing βάσισε το τεστ του στη διάκριση των φύλων, με ποιο κριτήριο δηλαδή επιλέχθηκε ότι ο παράγοντας A αντιστοιχεί σε άνδρα και ο παράγοντας B σε γυναίκα, για ποιο λόγο ο κάθε παράγοντας ανέλαβε το συγκεκριμένο ρόλο του και για ποιο λόγο το σύστημα TN στη δεύτερη εκδοχή του πειράματος αντικατέστησε τον άνδρα-A και όχι τη γυναίκα-B.

Ενδεχομένως, ο Alan Turing, με την επιλογή των φύλων, να θέλει να δείξει ότι ακόμα και οι πιο εδραιωμένες διαφορές όπως αυτές που αφορούν το φύλο, μπορούν να γίνουν αντικείμενο μίμησης και να αμφισβητηθούν. Αυτό οδηγεί στη σκέψη ότι το φύλο, ακριβώς όπως και η TN, είναι αποτέλεσμα μιας μιμητικής συμπεριφοράς που μαθαίνεται και μάλιστα σε τέτοιο βαθμό, ώστε να μοιάζει φυσικό (Halberstam, 1991). Αυτή η σκέψη τονίζει τη σημασία του κοινωνικού φύλου, θέλοντας να δείξει ότι πέρα από το βιολογικό φύλο, υπάρχει και το κοινωνικό, το οποίο αφορά τις κοινωνικές αντιλήψεις περί φύλου, ποια χαρακτηριστικά θεωρεί η κοινωνία ανά χρονικές περιόδους ότι πρέπει να έχει το κάθε φύλο, εδραιώνοντας έτσι βαθύτατες έμφυλες διαφορές και διαιώνίζοντας στερεότυπα (Butler, 2009). Ορμώμενος από τη δική του προσωπική πάλη και βιώματα, ο Turing ίσως θέλει να δείξει ακριβώς αυτό. Να αμφισβητήσει τις κοινωνικές αντιλήψεις περί φύλου και να δείξει ότι τα όρια μεταξύ των δύο φύλων είναι πολύ πιο ρευστά από ό,τι θεωρείται. Το ότι στο τεστ ο άνδρας-TN-παράγοντας A μπορεί να μιμηθεί το γυναικείο φύλο, δείχνει ότι κατά τον Turing, το φύλο και τα χαρακτηριστικά του είναι περισσότερο πολιτισμικό και κοινωνικό φαινόμενο, παρά βιολογικό. Ωστόσο, ακόμα και αν αυτές ήταν οι προθέσεις του Turing, υπάρχει ένα βασικό πρόβλημα. Το τεστ μπορεί να θεωρηθεί ότι είναι σχεδιασμένο για να υποβαθμίσει το γυναικείο φύλο και τη δύναμή του. Και αυτό, διότι ο άνδρας-TN-παράγοντας A έχει εκ του σχεδιασμού του τεστ αναγνωρισμένη την ιδιότητα να σκέφτεται περίπλοκα και μάλιστα με σημαντική ευφυΐα, αφού σκοπός του είναι η παραπλάνηση και η μίμηση. Αντίθετα, η γυναίκα-παράγοντας B παρουσιάζεται πολύ πιο απλή, με μοναδικό σκοπό

να είναι ειλικρινής και βοηθητική, χωρίς περαιτέρω περίπλοκη και ενδιαφέρουσα σκέψη (Genova, 1994).

2.6 Η φεμινιστική θεωρία για την τεχνολογία

Σκέψεις όπως οι ανωτέρω προβληματίσαν από πολύ νωρίς τις θεωρητικούς του φεμινισμού, καθώς με την εξέλιξη της τεχνολογίας και την όλο και πιο ευρεία χρήση της στην καθημερινότητα, γίνονταν φανεροί οι κίνδυνοι και τα προβλήματά της για το γυναικείο, και όχι μόνο, φύλο.

2.6.1 Η πεσιμιστική ματιά

Οι φεμινιστικές αναλύσεις της δεκαετίας 1970 και 1980, χαρακτηρίζονταν κυρίως από πεσιμιστικές τάσεις, καθώς αντιμετώπιζαν την τεχνολογία, πόσο μάλλον την ΤΝ, ως κάτι ακραία επικίνδυνο για τα δικαιώματα των γυναικών και αναμφίβολο όπλο για την ενίσχυση της πατριαρχίας. Αρχικά, αντιμετώπιζαν την τεχνολογία ως ουδέτερη, αλλά έβλεπαν ως βασικό πρόβλημα τη μικρή συμμετοχή των γυναικών σε αυτή. Η εστίασή τους, λοιπόν, ήταν στην ενίσχυση της συμπεριληπτικότητας των γυναικών στο χώρο. Το ριζοσπαστικό φεμινιστικό κίνημα, από πλευράς του, θεωρούσε το γυναικείο φύλο ως ιδιαίτερο και άξιο προστασίας από εξωτερικές παρεμβάσεις, το οποίο έπρεπε να προστατευθεί από τις πατριαρχικές χρήσεις και επεμβατικές πρακτικές της τεχνολογίας, όπως η υποβοήθηση της αναπαραγωγής. Οι εκπρόσωποι του σοσιαλιστικού φεμινισμού περαιτέρω, εστίαζαν στα αποτελέσματα που είχαν οι νέες τεχνολογίες στην εργασία των γυναικών. Για παράδειγμα, εξέταζαν πώς ορισμένα τεχνολογικά προϊόντα, δύσκολα στο χειρισμό και απαιτητικά, σχεδιάζονται για τους άντρες, επειδή εκείνοι θεωρούνται πιο ικανοί, ενώ τα πιο απλά και εύκολα απευθύνονται στις γυναίκες, με αποτέλεσμα τα ίδια τα τεχνολογικά προϊόντα να ενσωματώνουν τα έμφυλα στερεότυπα. Αυτές οι θεωρήσεις, ωστόσο, εστίαζαν στα αρνητικά χαρακτηριστικά της τεχνολογίας, παραβλέποντας ή μη αναγνωρίζοντας τα θετικά, δηλαδή πώς η τεχνολογία μπορεί να προσφέρει ευκαιρίες για την αποδόμηση των καθιερωμένων κοινωνικών νορμών.

2.6.2 Ο κυβερνοφεμινισμός (*cyberfeminism*) – τεχνοφεμινισμός (*technofeminism*)

Τις ευκαιρίες που προσφέρει η τεχνολογία διέκριναν οι εκπρόσωποι του λεγόμενου κυβερνοφεμινισμού (Wajcman, 2006). Ο κυβερνοφεμινισμός είναι ένας όρος που επινοήθηκε το 1994 από τη Sadie Plant για να περιγράψει το έργο των φεμινιστριών που ενδιαφέρονται για τη θεωρία, την κριτική και την αξιοποίηση του διαδικτύου, του κυβερνοχώρου και των νέων τεχνολογιών εν γένει. Η Sadie Plant ήταν από τις πρώτες και βασικές εκπροσώπους του κινήματος και ασχολήθηκε εκτενώς στο βιβλίο της “Zeros+Ones” (Plant, 2016) με το πώς οι γυναίκες αποκλείονται από το χώρο της τεχνολογίας και της πληροφορικής, αμφισβητώντας παράλληλα τα όρια των φύλων, τους δυϊσμούς και τη φύση της τεχνολογίας. Υποστήριξε επίσης ότι παρόλο που οι προηγούμενες φεμινίστριες πίστευαν ότι οι υπολογιστές είναι επακόλουθο του ανδρικού φύλου, θα έπρεπε οι υπολογιστές και το διαδίκτυο να θεωρηθούν ως χώροι, όπου οι γυναίκες έχουν τη δυνατότητα να εμπλακούν σε νέες μορφές εργασίας και παιχνιδιού, όπου απελευθερώνονται από τους παραδοσιακούς περιορισμούς και είναι σε θέση να πειραματιστούν με την έννοια της ταυτότητας και να ανακαλύψουν νέους τρόπους διεκδίκησης δύναμης και εξουσίας. Η άποψή της για τον κυβερνοχώρο είναι ότι είναι ένας φιλόξενος, οικείος χώρος για τις γυναίκες, όπου μπορούν και πρέπει να αρπάξουν ευκαιρίες για να προχωρήσουν και να αμφισβητήσουν την ανδρική εξουσία (Consalvo, 2002).

Σημαντική εκπρόσωπος του κυβερνοφεμινισμού είναι και η Donna Haraway, η οποία δημοσίευσε το 1985 το εμβληματικό για το μεταμοντέρνο φεμινισμό κείμενο της, «Το μανιφέστο των σάιμποργκ». Μπλέκοντας τη μεταμοντέρνα φεμινιστική θεωρία με τις μετα-ανθρώπινες προσεγγίσεις της τεχνολογίας, η Haraway πλέκει το μύθο του cyborg, δείχνοντας πώς οραματίζεται τον άνθρωπο (κάθε φύλου και σεξουαλικής προτίμησης) στη σύγχρονη τεχνολογική κοινωνία. Το cyborg της Haraway είναι ένας κυβερνητικός οργανισμός (*cybernetic organism*), ένα υβρίδιο μηχανής και έμβιου οργανισμού, ένα πλάσμα τόσο της κοινωνικής πραγματικότητας, όσο και της μυθοπλασίας. Ως πλάσμα της κοινωνικής πραγματικότητας, το cyborg αντανακλά την πραγματικότητα όπως αυτή διαμορφώνεται, συμπεριλαμβάνοντας τις κοινωνικές αντιλήψεις και βιώματα. Το cyborg της Haraway μπορεί να είναι προϊόν του πατριαρχικού καπιταλισμού, ωστόσο η Haraway το οραματίζεται άφυλο και απελευθερωμένο. Και αυτό, γιατί δεν έχει καταγωγή και προϊστορία, δεν το βαρύνει κανένα προπατορικό αμάρτημα, με αποτέλεσμα να είναι αποδεσμευμένο από τα δίπολα που διέπουν τις σύγχρονες

κοινωνίες και να αμφισβητεί ευθέως την κοινωνική κατασκευή του φύλου. Η Haraway κάνει ευθεία αναφορά στην επιστημονική φαντασία και την παραγωγή νοημάτων, αναγνωρίζοντας ότι η επιστημονική φαντασία, μέσω της αναπαράστασης της τεχνολογίας, διαμορφώνει νοήματα και αντιλήψεις, επομένως είναι πολύ λεπτό το όριο και κρίσιμο να μην αναπαράγονται στερεότυπα σχετικά με την ενσωματότητα, τη σεξουαλικότητα και τους έμφυλους ρόλους (Haraway, χ.χ.).

Βασιζόμενη στο έργο της Haraway, με το θέμα των έμφυλων διαφορών που γίνονται εμφανείς με τη χρήση των νέων τεχνολογιών, ασχολήθηκε και η Rosi Braidotti, στο κείμενό της «Κυβερνοφεμινισμός με διαφορά». Η Braidotti ξεκινάει με την παρατήρηση ότι στη μεταμοντέρνα εποχή η θηλυκότητα αναπαρίσταται με πολύ συγκεκριμένο τρόπο, ο οποίος βασίζεται στα τέλεια σώματα, όπως η κοινωνία τουλάχιστον θεωρεί το τέλειο σώμα, με επεμβάσεις για τη βελτιστοποίησή του, με αποτέλεσμα να υπερτονίζεται η έμφυλη ταυτότητα. Μάλιστα, υπερτονίζεται η εικόνα που η κοινωνία έχει για το πώς θα έπρεπε να μοιάζει το γυναικείο φύλο. Έτσι, προχωρά και εκείνη στην παραδοχή ότι τόσο μέσω των μέσων μαζικής επικοινωνίας όσο και μέσω της επιστημονικής φαντασίας στον κινηματογράφο, με την αναπαραγωγή του προτύπου του ανθρωποειδούς ως τέλειας γυναίκας με τις τέλειες αναλογίες και χαρακτηριστικά, της επικίνδυνης και πλάνας, εδραιώνονται έτι περαιτέρω οι στερεοτυπικές και σεξιστικές αντιλήψεις σε βάρος του γυναικείου φύλου. Καταλήγει, επομένως, στο συμπέρασμα ότι οι νέες τεχνολογίες και η υπερ-πραγματικότητα δεν εξαλείφουν τις έμφυλες διακρίσεις, αντιθέτως τις εντείνουν και τις διαγράφουν ακόμη πιο ξεκάθαρα (Braidotti, 2020).

Το κείμενο της Haraway ήταν επαναστατικό για την εποχή του, ενώ παραμένει επίκαιρο και αποτελεί σημείο αναφοράς μέχρι και σήμερα. Είναι εξαιρετικά προκλητική και ελκυστική η αποδόμηση του κοινωνικού φύλου όπως έχει εδραιωθεί στις κοινωνίες, η αμφισβήτηση των έμφυλων ρόλων και η αναδιαμόρφωσή τους σε μια ενότητα που τα περιλαμβάνει όλα και όλους, αλλά ταυτόχρονα και τίποτα και κανέναν. Επιπλέον, είναι εντυπωσιακό το όραμα της Haraway για μια τεχνολογία που έχει εξελιχθεί σε τέτοιο βαθμό, ώστε έχει αποβάλει από πάνω της όλα τα ανθρώπινα ελαττώματα, όπως τα στερεότυπα και τους διύσμούς και έχει υπερβεί τους κοινωνικούς φραγμούς, διαμορφώνοντας μια ιδεαλιστική κοινωνία και μια νέα κουλτούρα και τάξη πραγμάτων. Ωστόσο, όπως διαπιστώνει και η Braidotti, οι σύγχρονες κοινωνίες δεν έχουν

απαλλαγεί από τα στερεότυπα, με αποτέλεσμα αυτό να μεταφέρεται και στην αποτύπωση των νέων τεχνολογιών και των χρήσεών τους και έτσι να υπερτονίζεται το χάσμα.

Όλα τα ανωτέρω καθιστούν σαφές ότι η τεχνολογία και η κοινωνία είναι αλληλένδετα δεμένες μεταξύ τους, επηρεάζοντας καθοριστικά η μία την άλλη. Έτσι, οποιαδήποτε φεμινιστική ανάλυση επιχειρεί να εξετάσει το τεχνολογικό καθεστώς, θα πρέπει να εμβαθύνει και όχι να αντιμετωπίζει την τεχνολογία ως κάτι ξένο, απόμακρο και μοιραία κακό για τη γυναίκα. Επιπλέον, όταν γίνεται αναφορά στο φύλο, θα πρέπει να εγκαταλειφθούν οι δισμοί, όπως άλλωστε έδωσε την κατεύθυνση και η Haraway. Όλα είναι αλληλένδετα, συνεπώς πρέπει να λαμβάνεται υπόψιν και το μη-δυναδικό φύλο (non-binary), αλλά και η φυλή. Αυτές τις αρχές προεσβεύει και το κίνημα του λεγόμενου τεχνοφεμινισμού (technofeminism), όρος που διατύπωσε η Judy Wajcman, και προτάσσει ως απαραίτητο βήμα τη συμμετοχή των γυναικών σε όλα τα στάδια σχεδιασμού και υλοποίησης ενός τεχνολογικού προϊόντος, με ειδική αναφορά στα συστήματα TN (Wajcman, 2006). Βήμα, άλλωστε, που θεωρεί απαραίτητο και η παγκόσμια ερευνητική, ακαδημαϊκή και νομική κοινότητα.

2.7 Τα έμφυλα στερεότυπα στη γλώσσα και το πρόβλημα της ενσωμάτωσής της στην TN (Word Embeddings-Natural Language Processing)

Ένας από τους τρόπους λειτουργίας ενός συστήματος TN είναι και η επεξεργασία φυσικής γλώσσας (natural language processing-NLP), όπως συμβαίνει με τους αλγορίθμους μετάφρασης. Κατά τη διαδικασία αυτή, προκειμένου η TN να αντιληφθεί το νόημα των λέξεων όπως τις αντιλαμβάνεται ο άνθρωπος, χρησιμοποιείται η ενσωμάτωση λέξεων (word embeddings), διαδικασία κατά την οποία οι λέξεις όπως χρησιμοποιούνται στην καθημερινότητα και γίνονται κατανοητές από τον άνθρωπο, μεταφράζονται σε γλώσσα την οποία μπορεί να αντιληφθεί ο αλγόριθμος. Η διαδικασία αυτή είναι του τύπου: άνδρας = βασιλιάς / γυναίκα = ?, με το αναμενόμενο αποτέλεσμα ως συσχέτιση να είναι «βασίλισσα». Η συσχέτιση αυτή βάσει της διάκρισης που κάνει η γλώσσα, είναι σωστή. Ωστόσο υπάρχουν και αποτελέσματα όπως: άνδρας = γιατρός / γυναίκα = νοσοκόμα. Αυτό είναι αρκετά προβληματικό για δύο λόγους. Αφενός, διότι υπάρχουν και άνδρες νοσοκόμοι, οπότε ο αυτόματος συσχετισμός με τη γυναίκα είναι αυθαίρετος και αφετέρου αντανακλά την ευρύτερη στερεοτυπική αντίληψη που υπάρχει

στην κοινωνία ότι η ιδιότητα της νοσοκόμας αρμόζει σε γυναίκα (Bolukbasi, et al., 2016). Το ζήτημα αυτό συνδέεται και με τις έμφυλες διαφορές που οι ίδιες οι γλώσσες ενσωματώνουν. Για παράδειγμα, τα Αγγλικά είναι μια γραμματικά ουδέτερη ως προς το φύλο γλώσσα, αφού ως επί το πλείστον οι λέξεις δεν ενσωματώνουν διαφορετικές καταλήξεις ανάλογα με το φύλο (π.χ. η λέξη doctor = γιατρός και teacher = δάσκαλος/δασκάλα χρησιμοποιείται αδιαχώριστα και για τα δύο φύλα), ενώ ακόμα και αυτές που χρησιμοποιούνταν με διαφορετική κατάληξη, πλέον προτείνεται να χρησιμοποιούνται με κοινή, ώστε να είναι ολοκληρωτικά ουδέτερη η γλώσσα (π.χ. actor = ηθοποιός άνδρας και actress = ηθοποιός γυναίκα· πλέον προτείνεται η χρήση του όρου actor και για τα δύο φύλα). Αντίθετα, γλώσσες όπως τα Γαλλικά, Ισπανικά, Γερμανικά και Ελληνικά, ανήκουν στις γλώσσες που περιλαμβάνουν διαφορετικές καταλήξεις και άρθρα, ανάλογα με το φύλο, ενσωματώνοντας στη γλώσσα κατά αυτόν τον τρόπο τις έμφυλες διαφορές. Το πρόβλημα εδώ όσον αφορά τους αλγόριθμους έγκειται στο ότι όταν κάποιος χρήστης επιχειρήσει τη μετάφραση μιας λέξης από ουδέτερη ως προς το φύλο γλώσσα (για παράδειγμα της λέξης «teacher» από τα αγγλικά) σε γλώσσα με διαφορετικές καταλήξεις ως προς το φύλο, ο αλγόριθμος θα δώσει ως αποτέλεσμα το αρσενικό γένος, καθώς αυτό θεωρείται το πρότυπο (μετάφραση ως «δάσκαλος» στα ελληνικά) (Criado Perez, 2020). Ένα άλλο παράδειγμα από τον αλγόριθμο Google Translate είναι και το εξής. Αν επιχειρηθεί μετάφραση της φράσης “She is president. He is cooking.” (Ελληνικά: Αυτή είναι πρόεδρος. Αυτός μαγειρεύει.) από τα Αγγλικά σε μια ουδέτερη ως προς το φύλο γλώσσα όπως τα Φαρσί ή Τουρκικά, η μετάφραση θα είναι «Αυτό το άτομο είναι πρόεδρος. Αυτό το άτομο μαγειρεύει.», χωρίς καμία ένδειξη του φύλου του ατόμου. Αν στη συνέχεια, ζητηθεί από τον αλγόριθμο να μεταφράσει εκ νέου τη φράση, προσδίδοντας φύλο, η μετάφραση θα είναι «Αυτός είναι πρόεδρος. Αυτή μαγειρεύει.», παρόλο που αρχικά ήταν το αντίστροφο. Δεν θα επιλέξει την αντωνυμία «αυτός» και για τα δύο, αλλά ούτε και την αντωνυμία «αυτή» και για τα δύο. Αυτό συμβαίνει, επειδή στατιστικά, βάσει των δεδομένων με τα οποία εκπαιδεύτηκε ο αλγόριθμος, η ιδιότητα του προέδρου συσχετίζεται με το ανδρικό φύλο και η δραστηριότητα της μαγειρικής με το γυναικείο (Madgavkar, 2021), ενισχύοντας κατά αυτόν τον τρόπο τα κοινωνικά στερεότυπα περί έμφυλων ρόλων.

2.8 Η θηλυκοποίηση της TN

Προβληματική είναι εξίσου η τάση θηλυκοποίησης της TN, η οποία εκφράζεται και γίνεται αντιληπτή κυρίως μέσω των ψηφιακών βοηθών (digital assistants), αλλά και των αναπαραστάσεων της TN, ενσώματων και μη.

2.8.1 Ψηφιακοί βοηθοί (digital assistants)

Οι ψηφιακοί βοηθοί έχουν σχεδιαστεί για την εξυπηρέτηση καθημερινών αναγκών του χρήστη, λειτουργούν βάσει εντολών-ερωτημάτων που αυτός απευθύνει και διακρίνονται σε βοηθούς φωνής, chatbots και εικονικούς βοηθούς. Από αυτές τις κατηγορίες, οι βοηθοί φωνής είναι οι πιο ευρέως χρησιμοποιούμενοι ψηφιακοί βοηθοί, καθώς είναι ενσωματωμένοι στις προσωπικές συσκευές (όπως υπολογιστές, κινητά και IoT συσκευές), με αποτέλεσμα την εύκολη και άμεση πρόσβαση. Οι βοηθοί φωνής, στους οποίους εστιάζει η παρούσα μελέτη, έχουν σχεδιαστεί με σκοπό να βοηθούν το χρήστη στην οργάνωση του προγράμματος της καθημερινότητάς του και να παρέχουν γρήγορα πληροφορίες που χρειάζεται ανά πάσα στιγμή, κατόπιν γραπτών ή φωνητικών εντολών του χρήστη. Συγκεκριμένα παραδείγματα αποτελούν οι Siri (Apple), Google's Assistant, Alexa (Amazon) και Cortana (Microsoft).

Ως πρώτη παρατήρηση, επισημαίνεται ότι τα ονόματα των τριών από τους τέσσερις ψηφιακούς βοηθούς (Siri, Alexa και Cortana) είναι ονόματα γένους θηλυκού. Το όνομα Siri είναι νορβηγικό και η μετάφρασή του είναι «όμορφη γυναίκα που οδηγεί στη νίκη». Το όνομα Alexa προέρχεται από την αρχαία βιβλιοθήκη της Αλεξάνδρειας. Τέλος, το όνομα Cortana είναι το όνομα ενός γυναικείου άβαταρ στο παιχνίδι της Microsoft Halo (το οποίο απεικονίζεται ως αισθησιακό ανθρωποειδές). Επισημαίνεται περαιτέρω, ότι και οι τέσσερις ψηφιακοί βοηθοί διαθέτουν φωνή που προσομοιάζει σε γυναίκα. Κατά την αρχική τους κυκλοφορία, η μόνη διαθέσιμη επιλογή ήταν η γυναικεία φωνή, ενώ μόλις πρόσφατα προστέθηκε η επιλογή της ανδρικής (Gartenberg, 2021). Για παράδειγμα, η Siri εξακολουθεί να διαθέτει ως προεπιλογή τη γυναικεία φωνή, αλλά ο χρήστης πλέον μπορεί να επιλέξει την ανδρική· αξιοσημείωτο είναι ότι μόνο σε ορισμένες χώρες, όπως οι αραβικές και το Ηνωμένο Βασίλειο, η προεπιλεγμένη γλώσσα είναι η ανδρική (UNESCO; EQUALS Skills Coalition, 2019).



Εικόνα 3: Δείγμα φωνητικής εντολής προς την ψηφιακή βοηθό Siri της Apple (Apple, χ.χ.)

Η εξήγηση που δίνεται από τις τεχνολογικές εταιρίες για τη χρήση θηλυκών χαρακτηριστικών από τις ψηφιακές βοηθούς τους, είναι ότι αυτό γίνεται καθαρά για λόγους αύξησης των πωλήσεων (marketing). Υπάρχει ευρέως η αντίληψη, ότι οι καταναλωτές αναζητούν μια φωνή υπάκουη και υποτακτική, μια φωνή που έχει εκπαιδευτεί στο να την καθοδηγούν, για να είναι πάντα διαθέσιμη και πάντα εξυπηρετική. Αυτή η αντίληψη θεωρεί κατ' επέκταση ότι η γυναικεία φωνή αντικατοπτρίζει πλήρως αυτά τα πρότυπα και ανταποκρίνεται σε αυτές τις ανάγκες. Επομένως, είναι αναμενόμενο οι καταναλωτές να προτιμήσουν μια ψηφιακή βοηθό με γυναικεία φωνή και χαρακτηριστικά. Έτσι, οι βοηθοί φωνής φέρνουν στο μυαλό την εικόνα κλασικών νοικοκυρών που φροντίζουν για την ατμόσφαιρα χαλάρωσης στο σπίτι και γραμματέων που εκτελούν τις εντολές του κυρίου τους, ρόλοι που παραδοσιακά αποδίδονται στο γυναικείο φύλο (Jotanovic, 2018). Οι βοηθοί φωνής καταλήγουν, συνεπώς, να αναπαράγουν το δυϊσμό των φύλων και να ενισχύουν τα έμφυλα στερεότυπα, τόσο μέσω της ονομασίας τους, όσο κυρίως μέσω της γλώσσας. Αναπαράγουν την έννοια της γυναίκας με μοναδικό σκοπό την υποστήριξη και τη βοήθεια. Η θηλυκή φωνή συσχετίζεται με την υποταγή και τη διαφορά ισχύος των φύλων, με εξαιρετικά βλαπτικά αποτελέσματα για το γυναικείο φύλο στην κοινωνική πραγματικότητα (Collett & Dillon, 2019).

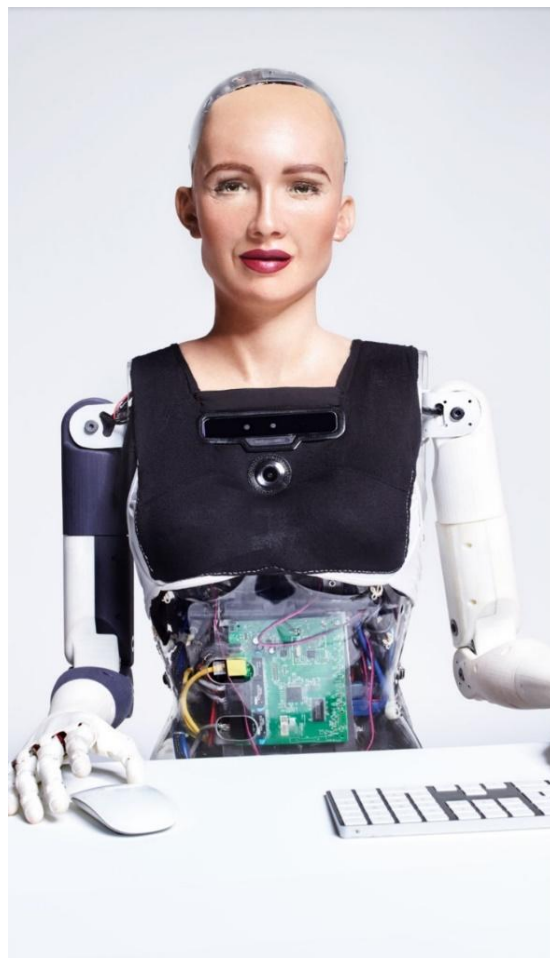
Η αποτύπωση του γυναικείου φύλου στις συσκευές και εφαρμογές των βοηθών φωνής, σχετίζεται, περαιτέρω, κατά αυτόν τον τρόπο με αντικείμενα (π.χ. smartphone) που απαιτούν διαρκή παρακολούθηση, ενημέρωση και αντικατάσταση. Αυτό δίνει μια νέα διάσταση στην αντικειμενοποίηση του γυναικείου φύλου, καθώς, αφενός συνδέεται με ρόλους που παραδοσιακά αποδίδονται στο γυναικείο φύλο και αφετέρου συνδέεται με αντικείμενα τα οποία δεν έχουν κάποια παραπάνω αξία και μπορούν πολύ εύκολα να αντικατασταθούν. Έτσι, το παράδειγμα των βοηθών φωνής δείχνει ότι η ιδανική ψηφιακή βοηθός λειτουργεί όπως η στερεοτυπική φαντασίωση της υπάκουης γυναίκας, η οποία έχει ως αποκλειστικό σκοπό της ζωής της τη φροντίδα και την ανταπόκριση στις ανάγκες του αφέντη της (master), είναι προσεκτική, χρήσιμη, όχι ισχυρογνώμων, έξυπνη, αλλά όχι και υπερβολικά, αντικαταστατή, προσαρμόσιμη και διαθέσιμη, αντανακλώντας για ακόμα μία φορά την εικόνα που έχουν για τη γυναίκα στο μυαλό τους τα λευκά μεγάλα μυαλά της τεχνολογίας (Sutko, 2020). Η αντικειμενοποίηση του γυναικείου φύλου, αλλά και η παθητική στάση η οποία αναμένεται αυτό να διατηρεί, γίνεται περαιτέρω αντιληπτή μέσω της σεξουαλικής παρενόχλησης και λεκτικής βίας που διαπιστώνεται στη χρήση των ψηφιακών βοηθών. Χαρακτηριστικά, αναφέρεται ότι αρχικά όταν ένας άνδρας χρήστης αποκαλούσε τη Siri «σκύλα», εκείνη αποκρινόταν «Αν μπορούσα, θα κοκκίνιζα» με εμφανώς παιχνιδιάρικη και ερωτική διάθεση, ενώ δεν εμφάνιζε την ίδια υπομονή σε υβριστικά σχόλια γυναικών χρηστών. Αντίθετα, σε αυτές απαντούσε πιο επιθετικά και απόλυτα. Οι εταιρίες έχουν κάνει βήματα για τη διόρθωση αυτών των «συμπεριφορών», έχουν ρυθμίσει για παράδειγμα τις ψηφιακές βοηθούς να απαντάνε σε υβριστικά ή σεξιστικά σχόλια με ουδέτερες, αλλά ακόμα παθητικές απαντήσεις, όπως «Δεν ξέρω τι να απαντήσω σε αυτό» (UNESCO; EQUALS Skills Coalition, 2019). Αυτό όμως δεν είναι αρκετό, διότι απλά αγνοούνται οι κακοποιητικές συμπεριφορές του χρήστη, δημιουργώντας του την εντύπωση ότι έχει δικαίωμα και χώρο να τις επαναλάβει, έχοντας απέναντί του μία υποταγμένη παθητική θηλυκότητα που ενισχύει τον ανδρικό του εγωισμό. Υπάρχει έτσι η πιθανότητα να καλλιεργηθεί ένα μοτίβο συμπεριφοράς με βλαπτικά αποτελέσματα για το γυναικείο φύλο εκτός από τον ψηφιακό, και στον πραγματικό κόσμο (Walker, 2020).

Αντιπρόταση στη θηλυκοποίηση των ψηφιακών βοηθών αποτελεί η περίπτωση του Q. Το Q παρουσιάστηκε το 2019 ως η πρώτη «φωνή TN χωρίς φύλο» και είναι μια ανθρώπινη φωνή για χρήση σε ψηφιακούς βοηθούς, η οποία έχει δημιουργηθεί με σκοπό να είναι απροσδιόριστη ως

προς το φύλο. Το Q βασίστηκε στην αρχή ότι η τεχνολογία δεν έχει φύλο και για το σχεδιασμό του συνεργάστηκε μια ομάδα γλωσσολόγων, μηχανικών ήχου και σχεδιαστών με μη-δυναδικά άτομα, δοκιμάζοντας διαφορετικές φωνές μέχρι να καταλήξουν σε ένα ηχητικό εύρος, το οποίο θεωρούσαν ότι είχε τη δυνατότητα να αλλάξει την καθιερωμένη κατάσταση και να αντιπροσωπεύσει τα μη-δυναδικά άτομα στον κόσμο της TN (Gómez-Uregui, 2021).

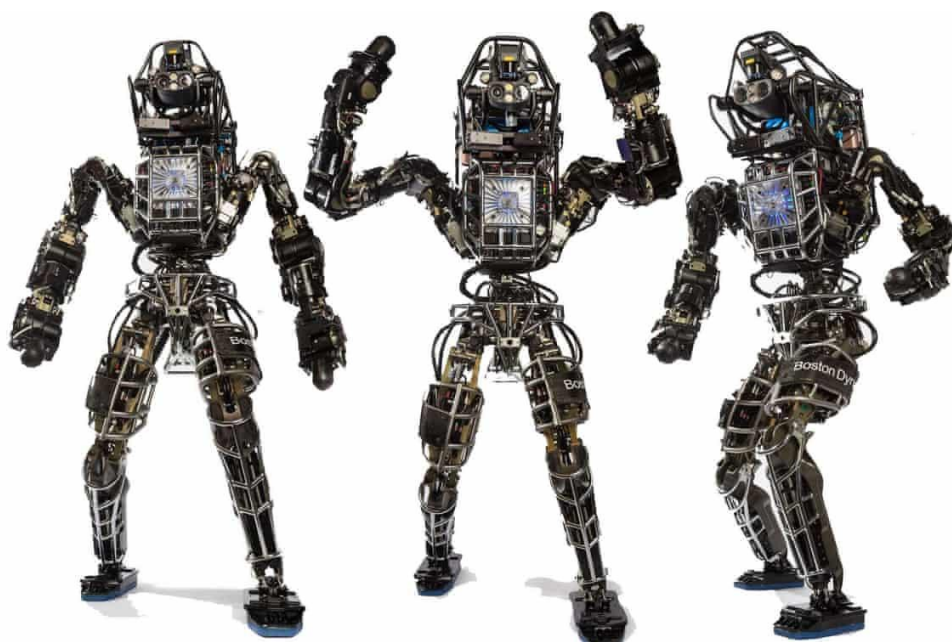
2.8.2 *Ενσώματες αναπαραστάσεις TN – Αναπαραστάσεις στην επιστημονική φαντασία (κινηματογράφος)*

Χαρακτηριστικό παράδειγμα θηλυκοποίησης της TN και των προβληματικών απεικονίσεων του γυναικείου φύλου, αποτελεί και η ενσωμάτωση της TN, με την έννοια της απεικόνισής της με ενσώματο, υλικό τρόπο. Είναι γνωστή πλέον παγκοσμίως, η περίπτωση της Sophia, του ανθρωποειδούς της Hanson Robotics.



Εικόνα 4: Το ανθρωποειδές Sophia της Hanson Robotics ως ενσώματη απότύπωση της TN (Hanson Robotics, χ.χ.)

Σύμφωνα με το βιογραφικό σημείωμα της Sophia, όπως είναι αναρτημένο στην ιστοσελίδα της Hanson Robotics, η Sophia ενσωματώνει τον τρόπο με τον οποίο η ανθρωπότητα ονειρεύεται την TN. Το ανθρωποειδές ρομπότ έχει δημιουργηθεί για την εξερεύνηση της εμπειρίας της αλληλεπίδρασης ανθρώπων και ρομπότ, όσον αφορά τις εφαρμογές υπηρεσιών και ψυχαγωγίας (Hanson Robotics, χ.χ.). Εδώ αξίζει να επισημανθούν δύο σημεία. Η Sophia αφενός ενσαρκώνει την εικόνα που έχουν οι δημιουργοί της, αλλά και μεγάλο μέρος της κοινωνίας για την TN ως γυναίκα και αφετέρου αυτή η TN-γυναίκα έχει δημιουργηθεί με σκοπό να φροντίζει για την εξυπηρέτηση του ανθρώπου και την ψυχαγωγία του. Στην περίπτωση της Sophia, επομένως, γίνεται ακόμα πιο εμφανής η έμφυλη διάκριση, καθώς η ενσώματη αποτύπωση της TN σε αυτήν την περίπτωση έχει εμφανώς τη σιλουέτα και τα χαρακτηριστικά του γυναικείου φύλου ή όπως έχουν την εικόνα για αυτό οι δημιουργοί της και η κοινωνία. Η Sophia έχει τις τέλει αναλογίες, το πρόσωπό της έχει θελκτικά και ευχάριστα στο μάτι χαρακτηριστικά (έμπνευση για το πρόσωπό της αποτέλεσε η ηθοποιός Audrey Hepburn), ακόμα και η φωνή της είναι ήρεμη, απαλή και ζεστή. Δεν είναι τυχαίο και το γεγονός ότι ο χρωματικός τόνος του δέρματός της είναι λευκός· η λευκή επιδερμίδα άλλωστε αποτελούσε από την αρχαιότητα το ιδανικό πρότυπο ομορφιάς. Αποτελεί ιδανικό ταίρι για συζήτηση, μιας και είναι καλή ακροάτρια, δείχνει ενδιαφέρον για το συνομιλητή της και έχει χιούμορ. Αποτελεί, θα έλεγε κανείς, την απεικόνιση της ιδανικής γυναίκας, σύμφωνα με το φαντασιακό των δυτικών κοινωνιών (Collett & Dillon, 2019). Όπως και στην περίπτωση των βοηθών φωνής, έτσι και στην περίπτωση της Sophia, αποδεικνύεται ότι η TN δεν έχει καταφέρει ακόμα να ξεφύγει από το δίπολο των έμφυλων ρόλων και τη διάκριση αρσενικού-θηλυκού. Αυτό μάλιστα γίνεται ακόμα πιο φανερό, αν αναλογιστεί κανείς ότι εφαρμογές και ρομπότ TN που χρησιμοποιούνται για ερευνητικούς και στρατιωτικούς σκοπούς, έχουν ονόματα που είτε προσιδιάζουν στο αρσενικό ή ουδέτερο φύλο είτε δε φέρουν καμία ένδειξη ως προς το φύλο, ενώ ακόμα και στην ενσώματη μορφή τους, δε φέρουν χαρακτηριστικά που αποδίδονται στο γυναικείο φύλο, αλλά προσομοιάζουν περισσότερο σε μια ουδέτερη κατάσταση (Fung, 2019).

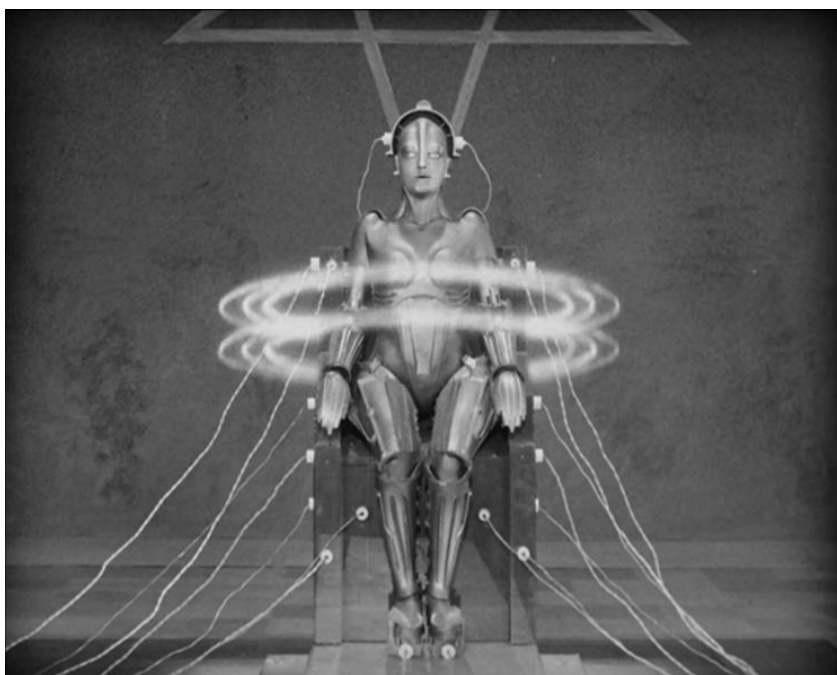


Εικόνα 5: Το ανθρωποειδές Atlas της Boston Dynamics για ερευνητικούς και αμυντικούς σκοπούς στις ΗΠΑ (DARPA) (Gibbs, 2013)

Η στερεοτυπική αναπαράσταση της ΤΝ με βάση το φύλο ενισχύεται και από την επιστημονική φαντασία, όπως αποτυπώνεται στον κινηματογράφο. Από τις σημαντικότερες ταινίες σε σχέση με την ΤΝ είναι το «2001: A space odyssey» του Stanley Kubrick (2001: A space odyssey, 1968), όπου παρουσιάζεται ο Hal 9000, μια άυλη ΤΝ προγραμματισμένη και πεπεισμένη να ακολουθήσει τις εντολές που της έχουν δοθεί πάση θυσία. Θεωρείται η ακριβέστερη απεικόνιση ΤΝ στον κινηματογράφο. Υπάρχουν επίσης ταινίες όπως το «I, Robot» (I, Robot, 2004) και η σειρά ταινιών «Terminator» (The Terminator, 1984), οι οποίες αντανακλούν στο έπακρο τον τρόπο του ανθρώπου για μια πιθανή εξέγερση των μηχανών και τον ακόλουθο αφανισμό του. Σε αυτές τις ταινίες η ΤΝ παρουσιάζεται είτε ως άφυλη είτε με ενισχυμένα ανδρικά χαρακτηριστικά, προβάλλοντας δυναμισμό, στρατηγικό και μαχητικό πνεύμα. Υπάρχουν ωστόσο ορισμένες ταινίες, στις οποίες είναι εμφανής η αποτύπωσή της με βάση το φύλο και τα στερεότυπα γύρω από αυτό, καθώς τις περισσότερες φορές η ΤΝ παρουσιάζεται με συγκεκριμένα υπερτονισμένα «γυναικεία» χαρακτηριστικά και υπό συγκεκριμένη οπτική (The Royal Society, 2018).

Η ταινία «Metropolis» (Metropolis, 1927) του Fritz Lang, θεωρείται ταινία-ορόσημο για το είδος της επιστημονικής φαντασίας, καθώς αποτελεί την πρώτη ιστορικά απεικόνιση της ΤΝ στον

κινηματογράφο. Πρόκειται για ένα ανθρωποειδές ρομπότ, το οποίο έχει σκοπό να παρεισφρήσει στις ομάδες των εργατών της πόλης, να τους ελέγξει εκ των έσω και να προκαλέσει χάος για να τους διχάσει. Το ανθρωποειδές αυτό, αν και κατασκευασμένο από μέταλλο, φέρει εμφανώς γυναικεία χαρακτηριστικά, καθώς η σιλουέτα του έχει γυναικείες αναλογίες και ονομάζεται Hel, το οποίο αναφέρεται στην ταινία ως γυναικείο όνομα. Στην πορεία παίρνει τη μορφή μιας γυναίκας της εργατικής τάξης, της Maria και εδώ είναι που ανακύπτει το πρόβλημα της κοινωνικής κατασκευής του φύλου. Η Hel/Maria έχει εμφανώς έντονη σεξουαλικότητα, την οποία χρησιμοποιεί με σκοπό να παραπλανήσει τους άντρες και να τους οδηγήσει στην καταστροφή. Είναι ιδιαίτερα εμφανές το ανδρικό βλέμμα στη σκηνή κατά την οποία η Hel/Maria σχεδόν γυμνή, χορεύει ερωτικά σαν μια άλλη Σαλώμη, με τους άντρες να την παρακολουθούν σε έκσταση.



Εικόνα 6: Το ανθρωποειδές ως Hel με εμφανώς «γυναικεία» σιλουέτα (Metropolis, 1927)

Πρόκειται για την πρώτη, από τις πολλές που ακολούθησαν στην πορεία, απεικόνιση της TN ως γυναίκας, η οποία κατασκευάζεται και ελέγχεται από άνδρα-δημιουργό και χρησιμοποιεί τη σεξουαλικότητά της ως εργαλείο αποπλάνησης. Ενώ η TN εκ φύσεως είναι άφυλη, εδώ της δίνονται γυναικεία χαρακτηριστικά, τα οποία αντικατοπτρίζουν σε μεγάλο βαθμό τον τρόπο με

τον οποίο το ανδρικό βλέμμα βλέπει τη γυναίκα, αλλά και το πώς αποτυπώνεται αυτή στον κινηματογράφο γενικότερα. Το «ανδρικό βλέμμα» (male gaze) αναφέρεται στο σεξουαλικοποιημένο τρόπο με τον οποίο παρουσιάζεται η γυναίκα, έτσι ώστε να ενδυναμώνονται οι άνδρες και να αντικειμενοποιούνται οι γυναίκες. Στο ανδρικό βλέμμα, η γυναίκα παρουσιάζεται οπτικά ως το «αντικείμενο» της ετεροφυλόφιλης ανδρικής επιθυμίας (Loreck, 2016).

Ένα ακόμα παράδειγμα είναι το «Her» του Spike Jonze (Her, 2013), το οποίο είναι μια ταινία πολύ κοντινή στα σημερινά τεχνολογικά δεδομένα. Πρόκειται για μια ρομαντική ιστορία ανάμεσα σε έναν άνθρωπο, το Theodore, και ένα λειτουργικό σύστημα OS (Operating System) TN, τη Samantha, το οποίο αποτελεί προσωπικό εικονικό βοηθό (personal virtual assistant), που υπόσχεται πως μπορεί να ανταποκριθεί στις εξατομικευμένες ανάγκες κάθε χρήστη του και να αναπτύσσει μια προσωπικότητα με βάση τις επιλογές του. Κατά την εγκατάστασή του, δίνεται η επιλογή να διαλέξει ο χρήστης αν θέλει το λειτουργικό σύστημα να έχει ανδρική ή γυναικεία φωνή. Ο Theodore επιλέγει τη γυναικεία φωνή και αφήνει το σύστημα να επιλέξει το ίδιο το όνομά του, το οποίο και είναι Samantha.

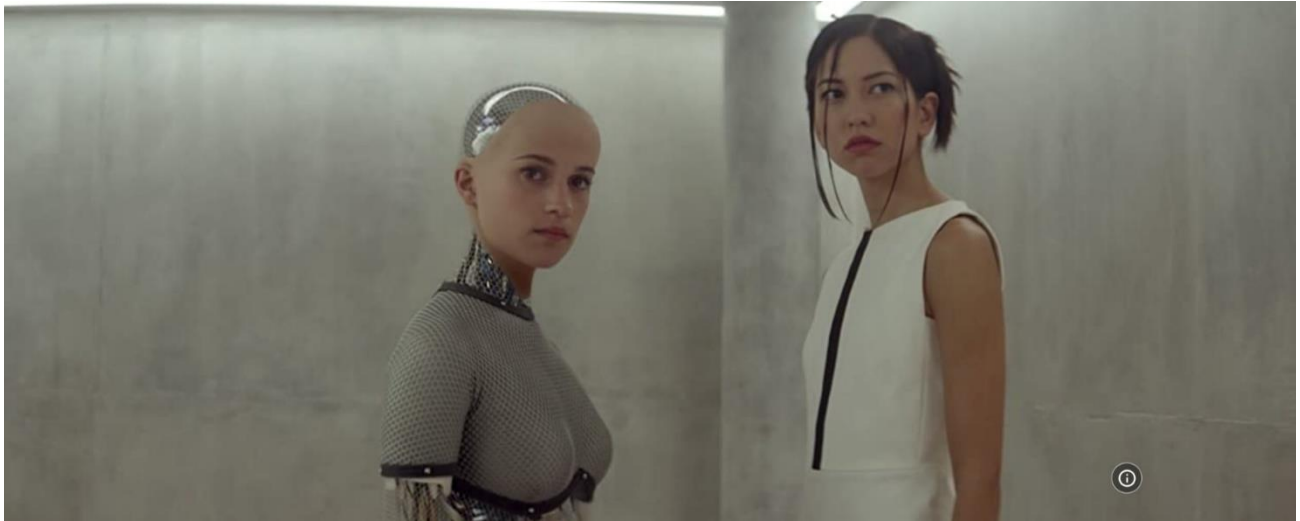


Εικόνα 7: Αλληλεπίδραση της εικονικής βοηθού Samantha με το χρήστη (Her, 2013)

Στην ταινία αυτή η βασική φιγούρα TN, η Samantha, είναι μία γυναικεία, βραχνή, φωνή, με εμφανή ερωτισμό, γεγονός που δεν αντιστοιχεί στην πραγματικότητα, δεδομένου ότι οι φωνές που επιλέγονται για τους ψηφιακούς βοηθούς ή/και τους βοηθούς πλοήγησης είναι «καθαρές», αποστασιοποιημένες και σε ήπιο τόνο, ώστε να γίνονται αντιληπτές οι απαντήσεις τους από τους χρήστες (Sejnowa, χ.χ.). Σε αντίθεση με άλλες έμφυλες απεικονίσεις της TN, η Samantha είναι μια προσωπική ψηφιακή βοηθός που σταδιακά ξεκινά να ανεξαρτητοποιείται, αφού έρχεται σε επικοινωνία με άλλα λειτουργικά συστήματα TN, ανταλλάσσοντας απόψεις. Είναι ενδιαφέρον ωστόσο, ότι αρχίζει να εμφανίζει προβλήματα και να καταρρέει η «σχέση» της Samantha και του Theodore, όταν αυτή χάνει την εστίασή της από εκείνον και δείχνει να αποκτά ενδιαφέροντα και να ικανοποιεί και τις δικές της ανάγκες (Sutko, 2020). Παρά την αυτονομία μέχρι ένα βαθμό που παρουσιάζει, επομένως, η απεικόνισή της βρίθκει κοινωνικών στερεοτύπων για το γυναικείο φύλο. Η φωνή και το ύφος της έχουν έντονο αισθησιασμό. Ο αρχικός προσχεδιασμένος σκοπός της είναι η υπακοή, η εξυπηρέτηση και ευχαρίστηση του χρήστη, ενώ όταν αναπτύσσεται ερωτική σχέση μεταξύ τους, αυτή φαίνεται να παρουσιάζει αμιγώς ανθρώπινα χαρακτηριστικά, όπως ζήλια και εκνευρισμό και μάλιστα με τρόπο που παραδοσιακά αποδίδονται κοινωνικά στο γυναικείο φύλο.

Τέλος, στο «Ex Machina» του Alex Garland (Ex Machina, 2014) η TN παρουσιάζεται ως ανθρωποειδής με πολύ συγκεκριμένα έμφυλα χαρακτηριστικά. Έχει γυναικεία σιλουέτα και αναλογίες και ακούει στο όνομα Ava. Πρόκειται για ακόμα μία γυναικεία απεικόνιση της TN με εμφανή τα χαρακτηριστικά του φύλου της. Η Ava γνωρίζει ότι είναι TN και αλληλεπιδρά μέσω της συζήτησης. Δείχνει περιέργεια για τους ανθρώπους και τον κόσμο, αλλά και ενσυναίσθηση στο άκουσμα δυσάρεστων καταστάσεων. Είναι φιλική, αστεία και ανταποκρίνεται στο φλερτ. Ο δημιουργός της, στο ερώτημα γιατί προσδιόρισε το φύλο της και της έδωσε σεξουαλικότητα, εκφράζει την πεποίθηση ότι η σεξουαλικότητα είναι το κίνητρο που καθοδηγεί τις πράξεις των ανθρώπων και ότι χωρίς αυτήν δε θα υπήρχε κανένα ενδιαφέρον. Εξηγεί επιπλέον ότι η Ava δε δείχνει μόνο εξωτερικά ως γυναίκα, αλλά ακόμα και η «γενετική περιοχή» της διαθέτει ειδικούς αισθητήρες ηδονής, ώστε να υπάρχει ανταπόκριση από πλευράς της, σε περίπτωση που υπάρξει ερωτική συνεύρεση με άνθρωπο. Στην ταινία παρουσιάζονται και άλλα ανθρωποειδή-γυναίκες (fembots) TN, τα οποία ο δημιουργός τους χρησιμοποιεί για τη σεξουαλική του ικανοποίηση και στη συνέχεια τα «σκοτώνει». Η Ava τελικά συνεργάζεται με

ένα από τα ανθρωποειδή-γυναίκες (fembot) και στρέφεται κατά του δημιουργού της, σκοτώνοντάς τον, ενώ η ίδια δραπετεύει για να ζήσει ελεύθερη στον κόσμο. Το *Ex Machina* παρά τη σύγχρονη ματιά του από τεχνολογική και σκηνοθετική άποψη, είναι εξαιρετικά προβληματικό όσον αφορά την προσέγγιση της ΤΝ με βάση το γυναικείο φύλο. Για ακόμα μια φορά δίνονται στην ΤΝ χαρακτηριστικά γυναίκας και μάλιστα αντικειμενικά όμορφης και ευχάριστης στο μάτι (το ανδρικό).



Εικόνα 8: Το ανθρωποειδές Ava και το fembot λίγο πριν εξολοθρεύσουν το δημιουργό τους (Ex Machina, 2014)

Επιπλέον, η Ava παρουσιάζεται να χρησιμοποιεί τη σεξουαλικότητά της για να εκτελέσει το σχέδιό της να δραπετεύσει, ενώ τα fembots που παρουσιάζονται στην ταινία είναι πλήρως υποταγμένα στο δημιουργό τους με μοναδικό τους σκοπό να τον ικανοποιούν σεξουαλικά. Τόσο αυτά, όσο και η Ava είναι παγιδευμένες μέσα σε ένα σπίτι και καταδικασμένες να ανταποκρίνονται στις επιθυμίες του άνδρα δημιουργού τους. Το μόνο όπλο που φαίνεται να έχουν ώστε να σπάσουν τα δεσμά τους είναι ο αποπροσανατολισμός μέσω της σεξουαλικότητάς τους (Watercutter, 2015). Είναι επίκαιρη και εδώ η θεωρία της Laura Mulvey σχετικά με το αρσενικό βλέμμα τον κινηματογράφο, σύμφωνα με την οποία, η αποτύπωση της γυναίκας στο δημοφιλή κινηματογράφο είναι διττή. Αποτυπώνεται τόσο ως αντικείμενο της ανδρικής ερωτικής επιθυμίας όσο και ως απειλή ευνουχισμού του ανδρισμού. Στην ταινία *Ex Machina* η Ava από τη μία παρουσιάζεται ως αντικείμενο της ερωτικής επιθυμίας και από την άλλη

παρουσιάζεται ως απειλή κατά της πατριαρχίας, αφού το παιχνίδι του φλερτ που παίζει αποδεικνύεται επικίνδυνο και για το δημιουργό της, καταλήγοντας να τον εκμεταλλευτεί και να τον σκοτώσει, σαν μια άλλη *femme fatale* (Storey, 2015).

2.9 Το πρόβλημα των δεδομένων στην TN – Υποεκπροσώπηση θηλυκού και μη-δυναμικού φύλου – Φεμινισμός των Δεδομένων (Data Feminism)

Η ποιότητα των δεδομένων που εισάγονται σε ένα σύστημα TN (training data), γίνεται κοινώς αποδεκτό ότι είναι καθοριστική για την εμφάνιση ή μη έμφυλων διακρίσεων (European Union Agency for Fundamental Rights, 2019). Και αυτό γιατί τα συστήματα TN είναι όπως τα παιδιά: ακριβώς όπως ένα παιδί γεννιέται χωρίς κάποια γνώση και αναπτύσσει τις γνώσεις του και την προσωπικότητά του μέσα από την παρατήρηση και την εκπαίδευση, έτσι και ένα σύστημα TN μαθαίνει από τα δεδομένα που εισάγονται σε αυτό. Ανεξάρτητα από το είδος των δεδομένων που χρησιμοποιούνται ή τον τρόπο εκπαίδευσης, αν τα δεδομένα εμπεριέχουν μεροληψίες ή βασίζονται σε μεροληπτικά μοτίβα, σημασία έχει πόση ευφυΐα διαθέτει ο αλγόριθμος για να διορθώσει αυτές τις προκαταλήψεις. Διότι αν δεν έχει επαρκή ευφυΐα, τότε φυσικά όχι μόνο θα αναπαράξει αυτά τα μοτίβα, αλλά υπάρχει ο κίνδυνος να τα μεγεθύνει, λόγω της υπολογιστικής του δύναμης (Madgavkar, 2021).

2.9.1 Τα μεγάλα δεδομένα (*big data*)

Το ζήτημα της ποιότητας των δεδομένων συνδέεται με τα λεγόμενα «μεγάλα δεδομένα» (*big data*), τα οποία χρησιμοποιούνται σήμερα ευρέως για την εκπαίδευση των συστημάτων TN. Τα μεγάλα δεδομένα αναφέρονται γενικά στις τεχνολογικές εξελίξεις που σχετίζονται με τη συλλογή, αποθήκευση, ανάλυση και τις εφαρμογές δεδομένων. Χαρακτηρίζονται από τον αυξημένο όγκο, ταχύτητα και ποικιλία (*volume-velocity-variety*, τα τρία V) των παραγόμενων δεδομένων και συνήθως αναφέρονται (αλλά όχι περιοριστικά) στα δεδομένα που υπάρχουν διαθέσιμα και λαμβάνονται από το διαδίκτυο. Τα μεγάλα δεδομένα προέρχονται από διάφορες πηγές, συμπεριλαμβανομένων των μέσων κοινωνικής δικτύωσης (*social media*) ή μεταδεδομένων (*metadata*), όπως η γεωγραφική τοποθεσία και τα δεδομένα περιήγησης. Το Διαδίκτυο των Πραγμάτων (*Internet of Things-IoT*) συμβάλλει επίσης στα μεγάλα δεδομένα, για παράδειγμα με τα συμπεριφορικά δεδομένα τοποθεσίας που λαμβάνονται από τα έξυπνα

κινητά τηλέφωνα (smartphones) ή τις συσκευές παρακολούθησης φυσικής κατάστασης (fitness tracking apps). Αυτή η αυξημένη διαθεσιμότητα των δεδομένων οδηγεί στη βελτίωση των τεχνολογιών για την ανάλυση και χρήση δεδομένων, όπως είναι τα συστήματα TN (European Union Agency for Fundamental Rights, 2018). Ένα βασικό πρόβλημα που αφορά τα μεγάλα δεδομένα, ωστόσο, είναι ότι δεν έχουν όλοι πρόσβαση στο διαδίκτυο ή στα μέσα κοινωνικής δικτύωσης και η κάλυψη από διαφορετικές εφαρμογές (apps) μπορεί να ποικίλλει επίσης. Αυτό έχει ως αποτέλεσμα, ορισμένες πληθυσμιακές ομάδες να μην καλύπτονται από τα δεδομένα που συλλέγονται μέσω του διαδικτύου. Αντίστοιχα, τα δεδομένα τοποθεσίας είναι αντιπροσωπευτικά μόνο για όσους διαθέτουν αυτές τις πληροφορίες προς χρήση, για παράδειγμα, μέσω των φορητών συσκευών τους (πχ. smartphone). Η τεράστια ανάπτυξη της χρήσης του διαδικτύου, πολλές φορές οδηγεί στο να μη γίνεται αντιληπτό πόσοι άνθρωποι δεν έχουν πρόσβαση σε αυτό και ότι αυτά τα δεδομένα πολύ συχνά ενσωματώνουν μεροληψία, καθώς αντιπροσωπεύουν μόνο ένα συγκεκριμένο μέρος του πληθυσμού (European Union Agency for Fundamental Rights, 2019). Αναφέρεται, χαρακτηριστικά, ότι σύμφωνα με την αναφορά της Διεθνούς Ένωσης Τηλεπικοινωνιών (International Telecommunication Union) του Οργανισμού Ηνωμένων Εθνών για το 2021, περίπου 4,9 δισεκατομμύρια άνθρωποι, δηλαδή το 63% του παγκόσμιου πληθυσμού, χρησιμοποιούν το διαδίκτυο, που σημαίνει ότι περίπου 2,9 δισεκατομμύρια άνθρωποι παραμένουν εκτός σύνδεσης, το 96% των οποίων ζει σε αναπτυσσόμενες χώρες, ενώ περίπου 390 εκατομμύρια άνθρωποι δεν έχουν καν πρόσβαση. Περαιτέρω, στις λιγότερο αναπτυγμένες και αναπτυσσόμενες χώρες παραμένει και το έμφυλο χάσμα, καθώς στην Αφρική χρησιμοποιούν το διαδίκτυο το 24% των γυναικών (έναντι 35% των ανδρών και στα αραβικά κράτη το 56% των γυναικών (έναντι 68% των ανδρών) (International Telecommunication Union, 2021).

2.9.2 Η υποεκπροσώπηση του γυναικείου και μη-δυναδικού φύλου

Το πρόβλημα των ελλιπών δεδομένων σε σχέση με το φύλο ανέδειξε εμπειριστατωμένα η Caroline Criado Perez στο βιβλίο της «Invisible Women» (Criado Perez, 2020). Βασιζόμενη σε στατιστικά στοιχεία η Perez παρουσιάζει την έκταση της υποεκπροσώπησης του γυναικείου φύλου, αλλά και της μαύρης και λατινοαμερικανικής κοινότητας, έναντι της υπερεκπροσώπησης των ανδρών και συγκεκριμένα των λευκών ανδρών. Η ανισοροπία αυτή συναντάται από τις πιο απλές δραστηριότητες της καθημερινότητας, όπως για παράδειγμα το

να παίρνει κανείς τα μέσα μαζικής μεταφοράς για να πάει στη δουλειά του, μέχρι και τις πιο σύνθετες και μεγάλες, όπως ο σχεδιασμός των πόλεων και η υγεία. Συγκεκριμένα, ως αιτία αυτής τη ανισορροπίας προβάλλεται το γεγονός ότι ο άνδρας θεωρείται το πρότυπο (η προεπιλογή-default), ενώ η γυναίκα ανέκαθεν θεωρούνταν ως η παρέκκλιση (the other). Εμπνεόμενη από τη Simone De Beauvoir, η Perez καταδεικνύει πώς για οποιονδήποτε σχεδιασμό και οποιαδήποτε λήψη απόφασης λαμβάνεται υπόψιν κατεξοχήν το ανδρικό φύλο ως πρότυπο και παραβλέπονται το γυναικείο και οι μειονότητες. Ειδικότερα, αναφέρει την περίπτωση της πλατφόρμας προσλήψεων Gild, η οποία επέτρεπε αξιολόγηση των υποψηφίων λαμβάνοντας υπόψιν εκτός από τα βιογραφικά τους, και τα κοινωνικά τους δεδομένα, δηλαδή το διαδικτυακό τους αποτύπωμα. Στην ουσία η σημασία δινόταν στο πόσο σημαντικός θεωρούνταν κάποιος προγραμματιστής στην κοινότητα, γινόταν δηλαδή μια αξιολόγηση της φήμης. Για να το κρίνει αυτό, η Gild λάμβανε υπόψιν μεταξύ άλλων, το χρόνο που διέθετε κάποιος προγραμματιστής στην ανάπτυξη και στο διαμοιρασμό του κώδικα, μέσω συγκεκριμένων πλατφορμών ανάπτυξης, αλλά και στην επικοινωνία με άλλους προγραμματιστές. Αυτό που δεν λήφθηκε ποτέ υπόψιν, είναι ότι οι γυναίκες αυτή τη στιγμή πέραν της επί πληρωμή εργασίας τους, είναι υπεύθυνες σχεδόν αποκλειστικά και για το μεγαλύτερο ποσοστό απλήρωτης εργασίας παγκοσμίως, εννοώντας τις οικιακές εργασίες, την ανατροφή των παιδιών, τη φροντίδα των ηλικιωμένων γονέων κλπ. Αυτό σημαίνει ότι κατά κύριο λόγο οι γυναίκες δεν έχουν πολύ διαθέσιμο χρόνο στη διάθεσή τους για να ξοδεύουν ώρες στο διαδίκτυο πέραν του απαραίτητου. Κατά συνέπεια, εφόσον η πλατφόρμα Gild αξιολογούσε θετικά κάποιον προγραμματιστή που περνούσε ώρες στο διαδίκτυο και αρνητικά κάποιον με λιγότερες, δεν λάμβανε υπόψιν τη διαφοροποίηση του χρόνου των γυναικών, προβαίνοντας έτσι σε διάκριση σε βάρος των γυναικών. Ακόμα πιο σημαντικό είναι το γεγονός ότι αλγόριθμοι σαν κι αυτόν της Gild κρατούνται κρυφοί και δεν υπάρχει πρόσβαση ώστε να διαπιστωθεί ακριβώς τι είδους μεροληψίες ενσωματώνουν και σε ποιο βαθμό (Criado Perez, 2020). Σχετικό είναι και το παράδειγμα της Amazon, η οποία χρησιμοποιούσε σύστημα TN για εξέταση βιογραφικών και πρόσληψη προσωπικού. Ο αλγόριθμος είχε εκπαιδευτεί με βάση τα βιογραφικά των ατόμων που είχαν προσληφθεί παλιότερα, δηλαδή έβρισκε κοινά μοτίβα και προέβαινε στη λήψη απόφασης για την καταλληλότητα του υποψηφίου. Το πρόβλημα όμως ήταν ότι το μεγαλύτερο ποσοστό των εργαζομένων ήταν άνδρες, με αποτέλεσμα ο αλγόριθμος να έχει ως πρότυπο

αυτούς, να θεωρεί ότι οι άντρες υποψήφιοι είναι προτιμητέοι και να προβαίνει συνεπώς σε διακρίσεις σε βάρος των γυναικών υποψηφίων (Dastin, 2018). Ως άλλο παράδειγμα, αναφέρονται τα λογισμικά που βασίζονται στην αναγνώριση φωνής, τα οποία, παρόλο που σύμφωνα με έρευνες οι γυναίκες γενικά τείνουν να μιλούν πιο αργά, καθαρά και δυνατά από τους άνδρες, πολλές φορές δυσκολεύονται να αναγνωρίσουν μια γυναικεία φωνή, ενώ παρουσιάζουν μεγαλύτερη ακρίβεια στις ανδρικές, απλά και μόνο επειδή τα συστήματα TN δεν εκπαιδεύονται με ηχογραφήσεις γυναικών. Αξιοσημείωτη είναι και η περίπτωση της Apple, καθώς η Siri, η ψηφιακή βοηθός της Apple που λειτουργεί βάσει TN, στο αρχικό στάδιο κυκλοφορίας της μπορούσε κατόπιν εντολής του χρήστη να κάνει αναζήτηση και να βρει αποτελέσματα για παρόχους συμπληρωμάτων υποβοήθησης της στύσης (Viagra), αντίθετα δεν εμφάνιζε αποτελέσματα για σημεία πραγματοποίησης εκτρώσεων κατόπιν σχετικής εντολής. Επιπλέον, αντιλαμβανόταν τον όρο «ανακοπή καρδιάς» και μπορούσε να καλέσει βοήθεια, αλλά αν ένας χρήστης έλεγε ότι έχει πέσει θύμα βιασμού, αδυνατούσε να κατανοήσει τη φράση. Αυτά τα παραδείγματα δείχνουν ότι τουλάχιστον κατά την αρχική της κυκλοφορία, οι ενδεχόμενες ανάγκες των γυναικών δεν είχαν ληφθεί υπόψιν (Criado Perez, 2020).

Σημαντικό είναι το πρόβλημα και στον τομέα της υγείας, όπου σε πολλές περιπτώσεις η έλλειψη δεδομένων στα συστήματα TN που χρησιμοποιούνται για διαγνωστικούς και θεραπευτικούς σκοπούς, μπορεί να οδηγήσει σε λανθασμένες και επικίνδυνες για τη ζωή και την υγεία αποφάσεις (Niethammer, 2020). Για παράδειγμα, αναφέρεται ότι σε έρευνα του Ηνωμένου Βασιλείου διαπιστώθηκε ότι οι γυναίκες έχουν 50% περισσότερες πιθανότητες να μη διαγνωστούν με ανακοπή καρδιάς, διότι δεν παρουσιάζουν τα «τυπικά» συμπτώματα. Με τον όρο «τυπικά», νοούνται τα καταγεγραμμένα συνήθη συμπτώματα, τα οποία εμφανίζουν κυρίως οι άνδρες (Criado Perez, 2020). Συνεπώς, αν ένα σύστημα TN για τη διάγνωση καρδιολογικών παθήσεων, εκπαιδευτεί με βάση τα «τυπικά», συνηθισμένα συμπτώματα των ανδρών, εφόσον υπάρχει έλλειψη δεδομένων για τα συμπτώματα των γυναικών, θα προβαίνει σε λίγες ή/και λανθασμένες διαγνώσεις όσον αφορά τις γυναίκες, με ενδεχομένως μοιραία αποτελέσματα (Tenenbaum, 2020). Όλα αυτά τα παραδείγματα, καταδεικνύουν πώς η εκπαίδευση των συστημάτων TN με ελλιπή δεδομένα μπορεί να οδηγήσει όχι μόνο σε διακρίσεις, αλλά ακόμα και σε κίνδυνο ζωής και υγείας.

2.9.3 Μηχανική όραση (*Computer vision*)

Η μηχανική όραση (*computer vision*) είναι το πεδίο ΤΝ που επιτρέπει σε υπολογιστές και συστήματα να αντλούν σημαντικές πληροφορίες από ψηφιακές εικόνες, βίντεο και άλλα οπτικά δεδομένα εισόδου και να προβαίνουν σε ενέργειες ή να κάνουν προτάσεις με βάση αυτές τις πληροφορίες. Η πιο συνηθισμένη και γνωστή χρήση της μηχανικής όρασης είναι τα συστήματα ΤΝ για την ταξινόμηση εικόνων (*image classification*), κατά την οποία ο αλγόριθμος βλέπει μια εικόνα και μπορεί να την κατατάξει σε μια κατηγορία (έναν σκύλο, ένα μήλο, το πρόσωπο ενός ατόμου). Πιο συγκεκριμένα, είναι σε θέση να προβλέψει με ακρίβεια ότι μια δεδομένη εικόνα ανήκει σε μια συγκεκριμένη κατηγορία (IBM, χ.χ.). Τα συστήματα αυτά εκπαιδεύονται με βάση σύνολα δεδομένων, τα οποία πολλές φορές δεν είναι αντιπροσωπευτικά ως προς το γυναικείο και μη-δυναδικό φύλο, καταλήγοντας σε μεροληπτικά αποτελέσματα.

Αξιοσημείωτο είναι το έργο της Joy Buolamwini, ερευνήτριας του MIT και ιδρύτριας του Algorithmic Justice League, ενός οργανισμού που συνδυάζει την τέχνη και την έρευνα για να ευαισθητοποιήσει σχετικά με τις κοινωνικές επιπτώσεις και τους κινδύνους της ΤΝ (Algorithmic Justice League, χ.χ.). Σε συνεργασία με την Timnit Gebru, διεξήγαγαν έρευνα για τρία προγράμματα ΤΝ αναγνώρισης προσώπου με σκοπό την κατηγοριοποίηση ανά φύλο. Συγκεκριμένα, εστίασαν στους αλγορίθμους της Microsoft, IBM και Face++. Για να καταστήσουν το σύνολο δεδομένων που θα εισαγόταν στα συστήματα αυτά για αξιολόγηση και κατηγοριοποίηση όσο το δυνατόν πιο ισορροπημένα και αξιόπιστα, χρησιμοποίησαν μια βάση δεδομένων η οποία περιείχε εικόνες ανδρών και γυναικών πολιτικών από χώρες της Αφρικής και της Ευρώπης, με αρκετά ισότιμο ποσοστό γυναικών και ανδρών. Σκοπός ήταν να εξετάσουν κατά πόσον τα συστήματα ΤΝ θα αναγνώριζαν σωστά το φύλο των εικονιζόμενων προσώπων. Το αποτέλεσμα στο οποίο κατέληξαν ήταν ότι και τα τρία συστήματα ΤΝ αναγνώριζαν με περισσότερη επιτυχία τους λευκούς άνδρες, ενώ το μεγαλύτερο ποσοστό λάθους στην αναγνώριση παρατηρήθηκε στις γυναίκες με σκούρο χρώμα δέρματος (ποσοστό λάθους 20,8-34,7%). Η λύση που προτείνουν είναι η χρήση *training data* που είναι πιο αντιπροσωπευτικά όσον αφορά τόσο το χρώμα του δέρματος, όσο και το φύλο (Buolamwini & Gebru, 2018).

Παρόμοια ευρήματα έδειξε και σχετική έρευνα σχετικά με τις εφαρμογές που παρέχονται από την Amazon, την Clarifai, τη Google, την Imagga, τη Microsoft και το IBM Watson, για την προσθήκη ετικετών σε εικόνες. Συγκεκριμένα, μελετήθηκε ο τρόπος με τον οποίο αυτές οι

εφαρμογές ερμηνεύουν τις εικόνες των ανθρώπων, σε σχέση με τη σωστή ή λανθασμένη χρήση των ετικετών που σχετίζονται με το φύλο, τη χρήση ετικετών που κρίνουν την εξωτερική εμφάνιση (πόσο ελκυστικό είναι ένα πρόσωπο) και τη χρήση ετικετών που οδηγούν σε συμπεράσματα σχετικά με το επάγγελμα/ρόλο, συναισθηματική κατάσταση ή το χαρακτήρα ενός προσώπου. Όσον αφορά τις ετικέτες για το φύλο, αυτές χρησιμοποιούνταν συχνότερα λανθασμένα για τα μαύρα άτομα. Αναφορικά με τις ετικέτες που κρίνουν την εξωτερική εμφάνιση, αυτές αφενός χρησιμοποιούνται συχνότερα για το χαρακτηρισμό ατόμων γυναικείου φύλου και αφετέρου οι χαρακτηρισμοί που χρησιμοποιούνται είναι «όμορφη», «χαριτωμένη», «σέξι», ενώ και εδώ οι χαρακτηρισμοί αυτοί είναι σπανιότεροι στις μαύρες γυναίκες. Τέλος, οι ετικέτες που αφορούν εκτίμηση του χαρακτήρα ενός προσώπου, όπως ο χαρακτηρισμός «έξυπνος» χρησιμοποιούνται περισσότερο σε εικόνες ατόμων του ανδρικού φύλου (σε αντίθεση με τις ετικέτες για την εξωτερική εμφάνιση που χρησιμοποιούνται περισσότερο στα άτομα γυναικείου φύλου). Επισημαίνεται ότι το μη-δυναδικό φύλο δεν αποτυπώθηκε σε καμία ετικέτα. Και σε αυτήν την έρευνα είναι φανερό η επιρροή τόσο των έμφυλων στερεοτύπων, με την έμφαση στην εμφάνιση όσον αφορά το γυναικείο φύλο, όσο και της μη εκπροσώπησης των μη-δυναδικών ατόμων και της υποεκπροσώπησης της μαύρης κοινότητας (Kyriakou, et al., 2019).

2.9.4 Φεμινισμός των Δεδομένων (Data Feminism)

Το ζήτημα της υποεκπροσώπησης του γυναικείου, του μη-δυναδικού φύλου, αλλά και της φυλής όσον αφορά τα δεδομένα, έχει απασχολήσει και τις Catherine D' Ignazio και Lauren F. Klein. Στο βιβλίο τους «Data Feminism» (D' Ignazio & Klein, 2020), αναφέρουν χαρακτηριστικά ότι η μη λήψη υπόψιν των γυναικών κατά το σχεδιασμό των τεχνολογικών προϊόντων γενικότερα, είναι αποτέλεσμα της αφέλειας που έχει κάποιος επειδή είναι στην κορυφή. Το ονομάζουν κίνδυνο προνομίου (privilege hazard), εννοώντας ότι από τη στιγμή που οι άνδρες θεωρούνται το πρότυπο και αποτελούν την πλειοψηφία των τεχνολογικών κολοσσών, δεν το κάνουν σκόπιμα, απλά δε σκέφτονται καν ότι υπάρχουν ανάγκες και δεδομένα πέρα από όσα τους αφορούν. Ειδικότερα, θίγοντας το ζήτημα του γυναικείου και μη-δυναδικού φύλου, αναγνωρίζουν ότι ο κόσμος είναι δομημένος σε δίπολα. Οι περισσότερες εφαρμογές που ζητούν στοιχεία φύλου, δίνουν επιλογές μόνο για το ανδρικό ή γυναικείο, με αποτέλεσμα να υπάρχει αποκλεισμός και μη εκπροσώπηση των μη-δυναδικών ατόμων. Φυσικά αυτό απηχεί τα κοινωνικά στερεότυπα που ακόμα υπάρχουν περί του δυισμού των φύλων, τα οποία ενσωματώνονται στη λειτουργία των

αλγορίθμων. Διατυπώνουν και προτείνουν περαιτέρω την ιδέα του φεμινισμού των δεδομένων (data feminism), δηλαδή «έναν τρόπο σκέψης για τα δεδομένα, τόσο τις χρήσεις τους όσο και τα όριά τους, που τροφοδοτείται από την άμεση εμπειρία, από τη δέσμευση για δράση και από τη διατομεακή φεμινιστική σκέψη» (D' Ignazio & Klein, 2020). Το σημείο εκκίνησης για το φεμινισμό των δεδομένων είναι το γεγονός ότι η εξουσία δεν κατανέμεται ισότιμα στον κόσμο, διαπίστωση μη αναγνωρίσιμη στην επιστήμη των δεδομένων. Αυτοί που ασκούν την εξουσία είναι σε δυσανάλογο ποσοστό άντρες ελίτ, ετεροφυλόφιλοι, λευκοί, ικανοί σωματικά, cisgender (American Psychological Association, χ.χ.) από το βόρειο ημισφαίριο. Σκοπός του φεμινισμού των δεδομένων είναι αρχικά να κατανοήσει τον τρόπο με τον οποίο οι συνήθεις πρακτικές στην επιστήμη των δεδομένων εξυπηρετούν την ενίσχυση των ήδη υπάρχουσών ανισοτήτων και στη συνέχεια να χρησιμοποιήσει την επιστήμη των δεδομένων για να αμφισβητήσει και να αλλάξει την κατανομή της εξουσίας. Ο φεμινισμός των δεδομένων εκφράζει την πίστη, αλλά και δέσμευση για τη συν-απελευθέρωση, δηλαδή την από κοινού συνεργασία των ατόμων, της κοινότητας και των θεσμών για την επίτευξη κοινού καλού. Βασίζεται στην ιδέα ότι τα καταπιεστικά συστήματα εξουσίας βλάπτουν όλους, ότι υπονομεύουν την ποιότητα και την εγκυρότητα της εργασίας και ότι εμποδίζουν από τη δημιουργία αληθινού και διαρκούς κοινωνικού αντικτύπου στην επιστήμη των δεδομένων (D' Ignazio & Klein, 2020).



Εικόνα 9: Πληροφοριακό γράφημα με τις αρχές του Φεμινισμού των Δεδομένων στο πλαίσιο της έκθεσης "You and AI: Through the Algorithmic Lens" της Στέγης Γραμμάτων και Τεχνών του Ιδρύματος Ωνάση (D' Ignazio, et al., 2020)

Συνοψίζοντας, η ΤΝ αποτελεί έναν από τους τομείς όπου οι έμφυλες διακρίσεις εκδηλώνονται έντονα, ενώ, όσο περισσότερο εξελίσσεται και όσο περισσότερο γίνεται μέρος της ζωής των ανθρώπων, τόσο περισσότερο θα εντείνονται τα έμφυλα στερεότυπα. Οι ψηφιακοί βοηθοί, τα ελάχιστα και λάθος δεδομένα, η αγνόηση των αναγκών του θηλυκού και μη-δυναδικού φύλου και οι αναπαραστάσεις στην επιστημονική φαντασία είναι μόνο ένα δείγμα του μεγέθους και της έκτασης τους προβλήματος. Και αυτό, διότι με το metaverse προ των πυλών, τον κυβερνοκόσμο εικονικής πραγματικότητας που υπόσχεται ο Mark Zuckerberg (MacDonald, 2022), οι έμφυλες διακρίσεις ενδέχεται να γίνουν συνήθεια, κοινή αποδοχή και μη ελέγξιμη κατάσταση (Adeyemo, 2021). Ήδη τα προβληματικά στοιχεία του φυσικού κόσμου έχουν αρχίσει να γίνονται αισθητά και εκεί, καθώς καταγράφονται τα πρώτα περιστατικά διακρίσεων με βάση το φύλο και τη φυλή (Egkolforoulou & Gardner, 2021), αλλά και περιστατικά σεξουαλικής παρενόχλησης και κακοποίησης (Morris, 2022), τα οποία βασίζονται στο γεγονός ότι η συντριπτική πλειονότητα των χρηστών είναι ανδρικού φύλου, αλλά και στην αντίληψη ότι πρόκειται για έναν ψηφιακό άυλο κόσμο χωρίς συνέπειες (Frenkel & Browning, 2021), κάνοντας επιτακτική την ανάγκη για ρύθμισή του το συντομότερο δυνατόν (Tusk, 2022). Πρόκειται για μία κρίσιμη περίοδο, ένα σημείο καμπής, που προσφέρει γόνιμο έδαφος για συζητήσεις και καλεί επιτακτικά να βρεθούν τρόποι αντιμετώπισης για μια πιο ηθική και δίκαιη ΤΝ.

ΚΕΦΑΛΑΙΟ 3ο

ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΤΗΝ ΑΝΤΙΜΕΤΩΠΙΣΗ ΤΩΝ ΕΜΦΥΛΩΝ ΔΙΑΚΡΙΣΕΩΝ ΣΤΗΝ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ – ΠΡΟΣ ΜΙΑ ΗΘΙΚΗ, ΔΙΚΑΙΗ ΚΑΙ ΣΥΜΠΕΡΙΛΗΠΤΙΚΗ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

3.1 Εισαγωγικές Παρατηρήσεις

Υπάρχει μέρος της επιστημονικής και τεχνολογικής κοινότητας που θεωρεί την ΤΝ ουδέτερη, δίκαιη και αλάνθαστη, επειδή βασίζεται σε μεγάλο όγκο δεδομένων (big data), επομένως τα αποτελέσματά της αντανακλούν την αντικειμενική αλήθεια (φονταμενταλισμός των δεδομένων-data fundamentalism) (Crawford, 2013). Ο όρος «φονταμενταλισμός των δεδομένων» (data fundamentalism) έχει διατυπωθεί από την Kate Crawford και αποδίδει την τάση να αντιμετωπίζονται τα μεγάλα δεδομένα καθαρά ως χρήσιμα και «αντικειμενικά», παραβλέποντας τους πολύπλοκους παράγοντες, οι οποίοι επηρεάζουν τι είδους δεδομένα συλλέγονται και πώς διαμορφώνονται για να παραχθεί ένα συγκεκριμένο αποτέλεσμα (Roberts, 2013). Αυτή η αντίληψη συνδέεται άμεσα με τον τεχνοσωβινισμό (technochauvinism) (Broussard, 2018), την πίστη δηλαδή ότι η τεχνολογία προσφέρει καλύτερες και πιο δίκαιες λύσεις, ο οποίος κατά τη Meredith Broussard χαρακτηρίζει τις μεγάλες τεχνολογικές εταιρίες και τη γενικότερη αντιμετώπιση των νέων τεχνολογιών, αποπροσανατολίζοντας και αγνοώντας τις επιβλαβείς επιπτώσεις τους και τα άτομα που επηρεάζονται (Thompson, 2018). Αυτές οι απόψεις οδηγούν σε ενίσχυση της εξουσίας αυτών που ελέγχουν τα τεχνολογικά μέσα και ωφελούνται από αυτά, αποκλείοντας και ακυρώνοντας τις ανάγκες και εμπειρίες των ατόμων που επηρεάζονται αρνητικά, όπως οι γυναίκες και τα μη-δυναμικά άτομα (Kalluri, 2020). Οι έμφυλες διακρίσεις της ΤΝ, όπως και όλα τα είδη μεροληψιών που εμφανίζει η ΤΝ, έχουν τις ρίζες τους τόσο σε τεχνολογικούς λόγους, όσο και σε ηθικούς-κοινωνικούς, γεγονός που καθιστά πολύ δύσκολη τη δημιουργία ενός συστήματος ΤΝ που θα είναι πλήρως απαλλαγμένο από αυτές. Η ΤΝ λειτουργεί ως καθρέφτης (Madgavkar, 2021) που αντανακλά τις εδραιωμένες κοινωνικές αντιλήψεις και καταστάσεις, προσφέροντας έτσι την ευκαιρία να παρατηρηθούν συνολικά, λεπτομερώς και ξεκάθαρα αυτά τα ζητήματα για την αντιμετώπισή τους, η οποία απαιτεί πολυεπίπεδη εξέταση και διεπιστημονική-διατομεακή προσέγγιση. Οι βασικοί άξονες

στους οποίους βασίζονται οι προτάσεις για την αντιμετώπιση των έμφυλων διακρίσεων στην ΤΝ, εντοπίζονται στην εκπαίδευση, τη διασφάλιση συμπεριληπτικότητας (inclusivity) και ποικιλομορφίας (diversity) ως βασικές αρχές που θα πρέπει να διέπουν την ΤΝ, την ευαισθητοποίηση μέσω ακτιβιστικών και ερευνητικών οργανώσεων, τα τεχνολογικά εργαλεία και την ενσωμάτωση ηθικών αρχών στην ΤΝ.

3.2 Ενίσχυση της συμμετοχής των γυναικών και μη-δυναδικών ατόμων στο πεδίο της Επιστήμης, Τεχνολογίας, Μηχανικής και Μαθηματικών (STEM - Science, Technology, Engineering, Mathematics)

Τα μαθηματικά, η φυσική και η πληροφορική είναι μαθήματα, τα οποία αποτελούν βασικό κομμάτι της εκπαιδευτικής ύλης στα σχολεία ήδη από το στάδιο της πρωτοβάθμιας εκπαίδευσης. Στο δημοτικό, τα μαθήματα απευθύνονται σε όλα τα μέλη της σχολικής τάξης, χωρίς να παρατηρούνται ιδιαίτερες διακρίσεις. Στις ανώτερες σχολικές βαθμίδες παρατηρείται μεγάλη απόκλιση, καθώς τα αγόρια τείνουν να επιλέγουν και να αποδίδουν καλύτερα στα μαθήματα θετικής και τεχνολογικής φύσης, ενώ τα κορίτσια προτιμούν και έχουν καλύτερες επιδόσεις στα θεωρητικά μαθήματα. Κατά συνέπεια, υπάρχει μετέπειτα μεγαλύτερο ποσοστό ανδρών στο πεδίο της Επιστήμης, Τεχνολογίας, Μηχανικής και Μαθηματικών (εφεξής STEM), τόσο στο Πανεπιστήμιο, όσο και στην εργασία. Οι δύο στερεοτυπικές αντιλήψεις που επικρατούν στο χώρο των STEM, δηλαδή ότι «τα αγόρια είναι καλύτερα από τα κορίτσια στα μαθήματα STEM» και ότι «τα μαθήματα STEM είναι για αγόρια και όχι για κορίτσια» (UNESCO, 2017), δημιουργούν στα κορίτσια σύγχυση ταυτότητας, καθώς νιώθουν ότι αν δείξουν ενδιαφέρον και ασχοληθούν με τα STEM, αυτό έρχεται σε αντίθεση με το φύλο τους (με την κοινωνική έννοια του όρου) (European Parliament, Directorate-General for Internal Policies of the Union; Zacharia, Zacharias C.; Hovardas, Tasos; Xenofontos, Nikoletta; Pavlou, Ivoni; Irakleous, Maria, 2020). Οι αντιλήψεις περί των έμφυλων ρόλων, οι οποίες υπάρχουν εδραιωμένες στην κοινωνία, στο σχολικό περιβάλλον και στην οικογένεια ασκούν πίεση στα κορίτσια ήδη από μικρή ηλικία, αποθαρρύνοντάς τα από το να ασχοληθούν, να πιστέψουν στον εαυτό τους και να επιδιώξουν μια καριέρα στο πεδίο των STEM. Τόσο οι μέθοδοι των εκπαιδευτικών, όσο και η σχολική ύλη, πρέπει να επαναπροσδιοριστούν, ώστε να περιλαμβάνουν και τα άτομα

γυναικείου και μη-δυναδικού φύλου, να κάνουν το πεδίο των STEM ελκυστικό σε αυτά και να ενισχύουν τις προσπάθειές τους.

Απαρέγκλιτη προϋπόθεση για την ισότιμη αντιμετώπιση των μαθητών, όσον αφορά το πεδίο των STEM, αποτελεί η ισότιμη πρόσβαση και συμμετοχή όλων στην εκπαιδευτική διαδικασία. Σε ορισμένες χώρες της Αφρικής για παράδειγμα, όπου η πρόσβαση στην εκπαίδευση είναι περιορισμένη, δίνεται προτεραιότητα στα αγόρια, αποκλείοντας τα κορίτσια. Η ισότιμη πρόσβαση των μαθητών στην εκπαίδευση είναι πρωταρχικής σημασίας και σε επόμενο στάδιο, οι εκπαιδευτικοί που έρχονται σε επαφή με τα παιδιά σε καθημερινή βάση ήδη από μικρή ηλικία, οφείλουν να είναι ενημερωμένοι, ευαισθητοποιημένοι και ενεργοί, να προωθούν την ισότιμη αντιμετώπιση και τη συμμετοχή όλων στο πεδίο των STEM, να αναγνωρίζουν τις δυσκολίες που αντιμετωπίζουν ειδικότερα τα άτομα γυναικείου και μη-δυναδικού φύλου και να βοηθούν στην αντιμετώπισή τους, να τα στηρίζουν· σε αυτό εξαιρετικά σημαντική είναι η πρόσληψη περισσότερων γυναικών και μη-δυναδικών εκπαιδευτικών από το πεδίο των STEM, διότι αφενός είναι σε θέση να κατανοήσουν βαθύτερα και πληρέστερα τη μειονεκτική θέση στην οποία βρίσκονται τα κορίτσια και τα μη-δυναδικά άτομα και αφετέρου μπορούν να λειτουργήσουν ως πρότυπο για αυτά, ενισχύοντας την αυτοπεποίθηση και συμμετοχή τους. Ακόμα και η σχολική ύλη είναι απαραίτητο να επαναπροσδιοριστεί, ώστε να αποτινάξει τα έμφυλα στερεότυπα και να περιλαμβάνει όλους στη γνωστική διαδικασία. Τέλος, απαραίτητο είναι να αλλάξουν τα κοινωνικά πρότυπα περί έμφυλων ρόλων, όπως προβάλλονται από τα μέσα επικοινωνίας (media) και τα μέσα κοινωνικής δικτύωσης (social media), ώστε να αποσυνδεθεί το πεδίο των STEM από το ανδρικό φύλο (UNESCO, 2017).

3.3 Συμπεριληπτικότητα (inclusivity) και ποικιλομορφία (diversity) στο χώρο της ΤΝ

Η βασικότερη αιτία για την εμφάνιση έμφυλων διακρίσεων στην ΤΝ είναι η έλλειψη επαρκών δεδομένων σχετικά με τις γυναίκες και τα μη-δυναδικά άτομα, αλλά και το μικρό ποσοστό συμμετοχής αυτών των ατόμων στον εργασιακό και ερευνητικό χώρο της ΤΝ (βλ. ανωτέρω 2.9). Τα δύο σημαντικότερα βήματα επομένως προς μια ΤΝ όσο το δυνατόν απαλλαγμένη από έμφυλες διακρίσεις είναι η εκπαίδευσή της με συμπεριληπτικά δεδομένα και η ενίσχυση της ποικιλομορφίας στην εργασία και την έρευνα.

Η εμπιστοσύνη στις νέες τεχνολογίες και το μεγάλο όγκο δεδομένων (big data) δημιουργούν συχνά τη διαστρεβλωμένη εικόνα ότι τα αποτελέσματα που παράγονται βάσει αυτών είναι ακριβή. Τα μεγάλα δεδομένα χρησιμοποιούνται για την εκπαίδευση της ΤΝ και την αυτοματοποίηση των διαδικασιών και τη λήψη αποφάσεων, επομένως η ποιότητα των δεδομένων αποκτά εξέχουσα σημασία. Εάν τα δεδομένα δεν καλύπτουν επαρκώς τον πληθυσμό που θα έπρεπε να καλύπτουν, τα στοιχεία που προκύπτουν είναι εσφαλμένα (δηλαδή μεροληπτικά). Από τη στιγμή που, όπως αναλύθηκε ανωτέρω, τα δεδομένα που υπάρχουν για τις γυναίκες και τα μη-δυναδικά άτομα, αλλά και για τη μαύρη κοινότητα, είναι ανεπαρκή ή εσφαλμένα, υπάρχει σφάλμα εκπροσώπησης (representation bias). Επιπλέον, μπορεί να υπάρχουν και σφάλματα μέτρησης (measurement biases), δηλαδή τα δεδομένα που χρησιμοποιούνται ενδέχεται να μην είναι τα κατάλληλα για να οδηγήσουν σε ακριβές αποτέλεσμα βάσει του ζητήματος που τίθεται κάθε φορά και του αποτελέσματος που επιδιώκεται. Εδώ ανακύπτει και το ζήτημα του labelling, δηλαδή της κατηγοριοποίησης των δεδομένων βάσει των οποίων εκπαιδεύεται η ΤΝ και το οποίο πραγματοποιείται από ανθρώπους (European Union Agency for Fundamental Rights, 2019). Η διαδικασία του labelling είναι πολύ σημαντική για την αξιολόγηση της μεροληψίας σε ένα σύνολο δεδομένων, καθώς είναι ένας από τους τρόπους που μπορούν να υπεισέλθουν στερεοτυπικές αντιλήψεις στα δεδομένα εκπαίδευσης (training data), όπως είναι η θεώρηση του φύλου ως δυαδικού και ο αποκλεισμός των μη-δυναδικών και τρανς ατόμων (Keyes, 2018). Η ποιότητα των δεδομένων, επομένως, κρίνεται υψίστης σημασίας και η αξιολόγησή της καθίσταται απαραίτητη (data quality assessment), ώστε να επιτυγχάνεται όσο το δυνατόν μεγαλύτερη συμπεριληπτικότητα (inclusivity) (European Union Agency for Fundamental Rights, 2019). Αν και ο έλεγχος όλων των μεγάλων δεδομένων που υπάρχουν διαθέσιμα είναι εξαιρετικά δύσκολος έως απίθανος, είναι σημαντική η αξιολόγησή τους, ενδεχομένως από ειδικές επιτροπές και φορείς (European Union Agency for Fundamental Rights, 2018) και η δημιουργία ενός συνόλου δεδομένων τα οποία θα θεωρούνται ποιοτικά επαρκή για την εκπαίδευση των συστημάτων ΤΝ.

Η έλλειψη ποικιλομορφίας (diversity) στο χώρο της ΤΝ, η οποία συνδέεται με την έλλειψη συμπεριληπτικότητας, είναι ένα πρόβλημα που έχει τις ρίζες του ήδη στα πρώτα χρόνια ανάπτυξης της ΤΝ και στην κουλτούρα που αναπτύχθηκε, η οποία έδινε προτεραιότητα στη δημιουργία και την επίτευξη αποτελεσμάτων, παραβλέποντας τις κοινωνικές προεκτάσεις της

χρήσης της TN και τις αρνητικές επιπτώσεις σε ορισμένες πληθυσμιακές ομάδες (Broussard, 2018). Μπορεί να παρατηρηθεί σε διάφορα πεδία, όπως οι διαφορές στις θέσεις εργασίας στη βιομηχανία και τον ακαδημαϊκό χώρο, στις διαφορές στις προσλήψεις και στις δυνατότητες προαγωγής, αλλά και στις έμφυλες διακρίσεις της TN που προκύπτουν ως αποτέλεσμα. Το πρόβλημα της ποικιλομορφίας δεν αφορά μόνο τις γυναίκες και τα μη-δυναδικά άτομα. Αφορά το φύλο, τη φυλή και, κυρίως, την κατανομή της εξουσίας, με την έννοια του ποιοι ελέγχουν την TN, τους πόρους και τις δυνατότητες ανάπτυξης της και ευνοούνται από αυτήν. Επηρεάζει τον τρόπο με τον οποίο λειτουργούν οι μεγάλες τεχνολογικές εταιρίες οι οποίες παράγουν TN, το είδος των προϊόντων που κατασκευάζονται, το κοινό το οποίο είναι σχεδιασμένα να εξυπηρετούν και τα άτομα που ωφελούνται από την ανάπτυξή τους (Myers West, et al., 2019). Όπως αναφέρθηκε ανωτέρω, σύμφωνα με την αναφορά του Παγκόσμιου Οικονομικού Φόρουμ για το παγκόσμιο έμφυλο χάσμα για το 2021, οι γυναίκες που απασχολούνται στο χώρο των δεδομένων και της TN ανέρχονται σε ποσοστό 32,4% παγκοσμίως, ενώ συγκεκριμένα στο χώρο της TN το ποσοστό συμμετοχής των γυναικών είναι κάτω του 25% (World Economic Forum, 2021). Η αντιμετώπιση της έλλειψης ποικιλομορφίας στις εταιρίες τεχνολογίας είναι απαραίτητος στόχος και βήμα προς την ουσιαστική αλλαγή.

Οι πρωτοβουλίες για την επίτευξη ποικιλομορφίας και συμπεριληπτικότητας πρέπει να συνοδεύονται από προσπάθειες αντιμετώπισης της επιβλαβούς και τοξικής κουλτούρας που έχει διαμορφωθεί στο χώρο εργασίας για τον τρόπο με τον οποίο σχεδιάζονται τα τεχνολογικά συστήματα (Myers West, et al., 2019). Όλα τα στάδια ανάπτυξης της TN πρέπει να διέπονται από συμπεριληπτικότητα και ποικιλομορφία, αλλά αυτό είναι καίριο ιδίως στο στάδιο του σχεδιασμού. Η λευκή ανδρική ελίτ των μεγάλων τεχνολογικών εταιριών συλλαμβάνει τη διαδικασία σχεδιασμού μόνο βάσει της δικής της οπτικής. Οι σχεδιαστές λαμβάνουν ως πρότυπο την κυρίαρχη κοινωνική ομάδα, δηλαδή τους λευκούς cisgender (American Psychological Association, χ.χ.) άνδρες, με εκπαιδευτικό προνόμιο, υψηλό τεχνολογικό αλφαριθμητισμό και αγγλόφωνους (Costanza-Chock, 2018). Μόνο με τη συμπερίληψη των γυναικών, των μη-δυναδικών ατόμων και των τρανς ατόμων στη διαδικασία σχεδιασμού θα μπορέσει να εισφερθεί η βιωμένη εμπειρία τους, η οπτική τους και οι ιδέες τους, διαμορφώνοντας πραγματικά μια TN που δε θα αποκλείει κανέναν (Hamidi, et al., 2018).

3.4 Κινητοποίηση και ευαισθητοποίηση (raising awareness)

Η συνειδητοποίηση του βάθους του προβλήματος των μεροληψιών, γενικότερα, και των έμφυλων διακρίσεων της TN, ειδικότερα, και η αντιμετώπισή τους, εξαρτώνται σε μεγάλο βαθμό από την εξοικείωση και ευαισθητοποίηση. Προς το σκοπό αυτό καίρια είναι η δραστηριοποίηση διαφόρων οργανώσεων, οι οποίες βασίζονται στη διεπιστημονική αντιμετώπιση του θέματος και την έμπρακτη δράση, με σκοπό τη διαμόρφωση ενός πιο συμπεριληπτικού και ηθικού χώρου στον τομέα της TN. Υπάρχουν οργανώσεις που εστιάζουν στην εκπαίδευση, κινητοποιώντας κορίτσια και μη-δυναδικά άτομα στην εκμάθηση συγγραφής κώδικα, όπως οι Girls Who Code (Girls Who Code, χ.χ.), αλλά και προσφέροντας ευκαιρίες για επαγγελματική ανέλιξη, όπως οι Women in AI (WAI) (Women in AI, χ.χ.). Άλλες οργανώσεις, στοχεύουν περισσότερο στην ευαισθητοποίηση και κινητοποίηση, μέσω έμπρακτων δράσεων και πρωτότυπων και εναλλακτικών προσεγγίσεων, με ορισμένες από αυτές να παρουσιάζονται ενδεικτικά ακολούθως.

3.4.1 *Distributed Artificial Intelligence Research Institute (DAIR)*

Το 2021, η Timnit Gebru, η οποία προηγουμένως εργαζόταν ως ερευνήτρια στον τομέα της ηθικής για την TN στη Google, ίδρυσε το Distributed Artificial Intelligence Research Institute (DAIR), με σκοπό την καταγραφή των δυσμενών συνεπειών της TN σε βάρος συγκεκριμένων μειονοτήτων και την ανάπτυξη εφαρμογών TN με σκοπό το θετικό αντίκτυπο σε αυτές τις πληθυσμιακές ομάδες (Tiku, 2021). Πρόκειται για ένα διεπιστημονικό και διεθνές ερευνητικό ίδρυμα για την TN, το οποίο βασίζεται στην πεποίθηση ότι η TN μπορεί να διορθωθεί, οι βλάβες της μπορούν να αποφευχθούν και όταν η παραγωγή και η ανάπτυξή της περιλαμβάνουν διαφορετικές οπτικές και συνειδητές διαδικασίες, μπορεί να είναι επωφελής. Ενσωματώνοντας ερευνητές από διαφορετικά υπόβαθρα και μειονότητες, το DAIR επικεντρώνεται στην παραγωγή ερευνητικού έργου, το οποίο αντανακλά τη βιωμένη εμπειρία και βασίζεται στην εις βάθος ανάλυση πιθανών κινδύνων και βλαβών ήδη από την αρχή δημιουργίας ενός συστήματος TN (Distributed AI Research Institute, χ.χ.).¹

¹ Αξίζει να σημειωθεί, ότι το Δεκέμβριο 2020 η Timnit Gebru εξαναγκάστηκε στην ουσία σε παραίτηση από τη Google, διότι σύμφωνα με τη Google η τελευταία της μελέτη δεν πληρούσε τα απαιτούμενα πρότυπα και η Google αρνήθηκε να τη δημοσιεύσει. Άνδρες συνάδελφοί της, έσπευσαν να τονίσουν ότι οι

3.4.2 *Algorithmic Justice League*

Ο Algorithmic Justice League² είναι ένας οργανισμός που συνδυάζει την τέχνη και τη δημιουργικότητα με την έρευνα, για να φωτίσει τις κοινωνικές επιπτώσεις και τα επιβλαβή αποτελέσματα της ΤΝ. Η αποστολή του είναι να ευαισθητοποιήσει το κοινό σχετικά με τις επιπτώσεις της ΤΝ, να προσφέρει στους υποστηρικτές του τη δυνατότητα εμπειρικής έρευνας, να δώσει φωνή στις μειονότητες που επηρεάζονται περισσότερο και να κινητοποιήσει ερευνητές, υπεύθυνους χάραξης πολιτικής και επαγγελματίες του κλάδου να μετριάσουν τις μεροληψίες και τα επιβλαβή αποτελέσματα της ΤΝ.

Σημαντικό έργο του Algorithmic Justice League είναι το ντοκιμαντέρ Coded Bias, στο οποίο αναδεικνύονται ιστορίες ανθρώπων που έχουν επηρεαστεί από την επιβλαβή τεχνολογία και παρουσιάζονται πρωτοπόρες γυναίκες, οι οποίες εφιστούν την προσοχή για τις απειλές που θέτει η ΤΝ για τα ανθρώπινα δικαιώματα και τη δημοκρατία. Το ντοκιμαντέρ έκανε προεμίρα

δικές τους έρευνες δεν αξιολογούνταν ποτέ ως προς το περιεχόμενό τους, αλλά εξετάζονταν μόνο ως προς το αν περιείχαν απόρρητα στοιχεία που δεν έπρεπε να δημοσιευθούν. Η Google έχει επικριθεί έντονα για την κίνησή της αυτή, καθώς καταδεικνύει ότι τα κίνητρά της συνδέονται περισσότερο με την απομάκρυνση μιας φωνής που έθετε ηθικούς προβληματισμούς και κατά συνέπεια περιορισμούς και ειδικότερα γυναίκας, όταν δε συνέβη το ίδιο με άνδρες συναδέλφους της (Simonite, 2021). Το Φεβρουάριο 2022, δύο ακόμα μέλη της ομάδας της Google για την Ηθική της ΤΝ παραιτήθηκαν για να ακολουθήσουν τη Gebru και να γίνουν μέλη του DAIR, εκθέτοντας περαιτέρω τον τεχνολογικό κολοσό (Roth, 2022).

² Η Joy Buolamwini, ερευνήτρια του MIT είχε την έμπνευση να ιδρύσει το Algorithmic Justice League κατόπιν προσωπικής της εμπειρίας και της διάκρισης που βίωσε από αλγόριθμο. Κατά τη διάρκεια εργασίας της σε μεταπτυχιακό πρόγραμμα του MIT, το λογισμικό ανάλυσης προσώπου που ανέπτυξε δυσκολεύτηκε να εντοπίσει το πρόσωπό της. Η ίδια υποψιάστηκε ότι αυτό ήταν κάτι περισσότερο από ένα τεχνικό λάθος και διαπίστωσε ότι οι συνάδελφοί της στο MIT με πιο ανοιχτόχρωμο δέρμα δεν είχαν το ίδιο πρόβλημα. Όταν προσπάθησε να σχεδιάσει ένα πρόσωπο στην παλάμη του χεριού της, το μηχάνημα το αναγνώρισε αμέσως. Ωστόσο, δε συνέβη το ίδιο με το πραγματικό της πρόσωπο και έτσι αναγκάστηκε να ολοκληρώσει τη συγγραφή κώδικα του έργου της με μια λευκή μάσκα στο πρόσωπό της, ώστε να μπορεί ο αλγόριθμος να το εντοπίζει. Αυτή η εμπειρία της έδωσε το κίνητρο για να δημιουργήσει το Algorithmic Justice League (Algorithmic Justice League, χ.χ.).

στο Φεστιβάλ Κινηματογράφου Sundance τον Ιανουάριο του 2020 και είναι διαθέσιμο στη διαδικτυακή συνδρομητική πλατφόρμα Netflix (Algorithmic Justice League, χ.χ.).



Εικόνα 11: Η Joy Buolamwini με άσπρη μάσκα, ώστε ο αλγόριθμος να μπορέσει να εντοπίσει το πρόσωπό της (Coded Bias, 2020)

3.4.3 DATAWO

Ο DATAWO είναι ένας ελληνικός μη κερδοσκοπικός οργανισμός που ιδρύθηκε από μια ομάδα γυναικών δικηγόρων, την Αναστασία Καραγιάννη και την Ανθή Αργυρίου, με στόχο την ευαισθητοποίηση της κοινωνίας για την προστασία των δικαιωμάτων των γυναικών στην ψηφιακή εποχή. Αποτελεί τον πρώτο οργανισμό στην Ελλάδα που ασχολείται με τα ζητήματα ισότητας των φύλων στην ψηφιακή εποχή. Ξεκινώντας από τη θέση ότι τα στερεότυπα της πατριαρχικής κοινωνίας στο φυσικό κόσμο αντανακλώνται και στον ψηφιακό, με σοβαρό αντίκτυπο σε θεμελιώδη ανθρώπινα δικαιώματα, μεγεθύνοντας το πρόβλημα, και αναγνωρίζοντας ότι ο ψηφιακός κόσμος κυριαρχείται και προγραμματίζεται κυρίως από τους άνδρες, ο DATAWO εφιστά την προσοχή σε ζητήματα έμφυλης βίας στο διαδίκτυο, όπως η σεξουαλική παρενόχληση μέσω διαδικτύου, το cyber-stalking και η πορνογραφία, καθώς και σε ζητήματα ισότητας των φύλων σχετικά με το σχεδιασμό και την εφαρμογή νέων τεχνολογικών συστημάτων με διακριτικό τρόπο εις βάρος των γυναικών. Οι δράσεις του DATAWO

περιλαμβάνουν τη διεξαγωγή νομικής έρευνας πάνω στα προαναφερθέντα ζητήματα, τη διοργάνωση ενημερωτικών σεμιναρίων και εκπαιδευτικών προγραμμάτων σε εθνικό, Ευρωπαϊκό και διεθνές επίπεδο (DATAWO, χ.χ.).

Μία από τις πιο σημαντικές δράσεις της είναι το Female Code Lab, ένα φεμινιστικό εργαστήριο προγραμματισμού, με στόχο την ψηφιακή ενδυνάμωση και την εξοικείωση των θηλυκοτήτων που συμμετείχαν με τον προγραμματισμό. Αξιοσημείωτη είναι και η συμμετοχή του DATAWO στο MozFest 2022³ με το έργο «Ένα φεμινιστικό λεξικό στην Τεχνητή Νοημοσύνη». Η ομάδα του DATAWO, με επικεφαλής τη συνιδρύτριά του Αναστασία Καραγιάννη, απαρτιζόμενη από άτομα τόσο με νομικό, όσο και με τεχνικό υπόβαθρο, δημιούργησε μια εργαλειοθήκη σε μορφή λεξικού με όρους και έννοιες που περιλαμβάνουν όλα τα φύλα και η οποία μπορεί να χρησιμοποιηθεί ως αναφορά στην εποχή της TN. Στόχος αυτού του λεξικού είναι να γεφυρώσει το χάσμα μεταξύ θεωρητικών και προγραμματιστών, ώστε τα άτομα που εμπλέκονται σε όλη τη διαδικασία δημιουργίας συστημάτων TN να εξοικειωθούν με την ισότητα των φύλων και την αρχή της μη διάκρισης (Καραγιάννη, 2022). Απευθύνεται σε όλη την κοινότητα της TN γενικά και στοχεύει στο να συμβάλει στην κατανόηση της φεμινιστικής θεωρίας και στην ουσιαστική εφαρμογή της αρχής της ισότητας των φύλων (DATAWO, χ.χ.).

3.5 Ρυθμίσεις εκ του σχεδιασμού της TN (by design) - Διαφάνεια (transparency) - Debiasing

3.5.1 Ενσωμάτωση ηθικών αρχών εκ του σχεδιασμού (by design) – Διαφάνεια (transparency)

Ο GDPR προέβλεψε τη δυνατότητα της επεξεργασίας προσωπικών δεδομένων με σεβασμό προς τα δικαιώματα των υποκειμένων επεξεργασίας (φυσικών προσώπων) και τις αρχές του Κανονισμού, ήδη εκ του σχεδιασμού (εφεξής by design) του συστήματος επεξεργασίας

³ Το MozFest είναι ένα υβριδικό συνέδριο: αποτελεί εν μέρει ένα συνέδριο τέχνης, τεχνολογίας και κοινωνικών ζητημάτων, αλλά και ένα φεστιβάλ στο οποίο συγκεντρώνονται άτομα που δραστηριοποιούνται στο χώρο της τεχνολογίας και ακτιβιστές σε διάφορα παγκόσμια κινήματα που αγωνίζονται για έναν πιο ανθρώπινο ψηφιακό κόσμο. Διοργανώνεται ετησίως από το Mozilla Foundation, έναν παγκόσμιο μη κερδοσκοπικό οργανισμό αφιερωμένο στο να διατηρεί το διαδίκτυο ως έναν παγκόσμιο δημόσιο πόρο που είναι ανοιχτός και προσβάσιμος σε όλους (DATAWO, χ.χ.).

(Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο, 2016). Σύμφωνα με την αρχή της προστασίας των δεδομένων ήδη από τον σχεδιασμό, κατά τη στιγμή του σχεδιασμού των συστημάτων επεξεργασίας και του καθορισμού των μέσων επεξεργασίας, ο υπεύθυνος επεξεργασίας πρέπει να ενσωματώνει και να εφαρμόζει κατάλληλα μέτρα και να χρησιμοποιεί τεχνολογίες ενίσχυσης της ιδιωτικότητας, όπως ψευδωνυμοποίηση δεδομένων προσωπικού χαρακτήρα, το συντομότερο δυνατόν (δηλ. αντικατάσταση προσωπικά ταυτοποιήσιμων πληροφοριών με τεχνητά αναγνωριστικά στοιχεία), κρυπτογράφηση (κωδικοποίηση προσωπικών δεδομένων έτσι ώστε μόνο όσοι είναι εξουσιοδοτημένοι να μπορούν να τα διαβάσουν), ελαχιστοποίηση της επεξεργασίας των δεδομένων και ενσωμάτωση των απαραίτητων εγγυήσεων στην επεξεργασία, κατά τρόπο ώστε να πληρούνται οι απαιτήσεις του GDPR και να προστατεύονται τα δικαιώματα των υποκειμένων των δεδομένων (Αρχή Προστασίας Δεδομένων, χ.χ.). Με άλλα λόγια, η εκ του σχεδιασμού ρύθμιση σημαίνει ότι το σύστημα επεξεργασίας έχει σχεδιαστεί εξαρχής κατά τέτοιο τρόπο, ώστε να σέβεται τα δικαιώματα των φυσικών προσώπων και των νομοθετικών αρχών, χωρίς καμία περαιτέρω ενέργεια.

Με δανεισμό της έννοιας από τη νομοθεσία για την προστασία των προσωπικών δεδομένων, έχει διατυπωθεί η πρόταση της εκ του σχεδιασμού ενσωμάτωσης στην ΤΝ των αρχών του σεβασμού προς τα θεμελιώδη ανθρώπινα δικαιώματα, όπως η ισότητα και η απαγόρευση των διακρίσεων και ηθικών αρχών, ώστε να εξαλειφθούν οι έμφυλες διακρίσεις της ΤΝ. Αυτό απαιτεί τη διασάφηση των συγκεκριμένων νομικών υποχρεώσεων που πρέπει να σέβονται οι σχεδιαστές ΤΝ και τη μετάφρασή τους σε εφαρμόσιμες τεχνολογικές-μηχανικές πρακτικές. Πρόκειται επομένως για ρύθμιση και συμμόρφωση εκ των προτέρων, με την πρόβλεψη της απαγόρευσης των διακρίσεων και τη διασφάλιση της ισότητας ως απαρέγκλιτη προϋπόθεση για τη σύννομη λειτουργία του συστήματος ΤΝ (Van Cleynenbreugel, 2022).

Απαραίτητη για την ενσωμάτωση στην ΤΝ των αρχών του σεβασμού προς τα ανθρώπινα δικαιώματα είναι η διασφάλιση διαφάνειας, με την έννοια της πρόσβασης στον τρόπο με τον οποίο λειτουργεί η ΤΝ και της δομής της, ώστε να μπορεί να εξεταστεί και να γίνει κατανοητή. Η πρόσβαση, ωστόσο, στον πηγαίο κώδικα του αλγορίθμου τις περισσότερες φορές είναι αδύνατη, διότι αποτελεί πολύτιμο περιουσιακό στοιχείο των εταιριών και εμπλεκόμενων (Ferrer, et al., 2021). Η διαφάνεια πρέπει να εξασφαλίζεται από την αρχή της ανάπτυξης του συστήματος ΤΝ (Felzmann, et al., 2020) και να παρέχονται πληροφορίες για τα δεδομένα που

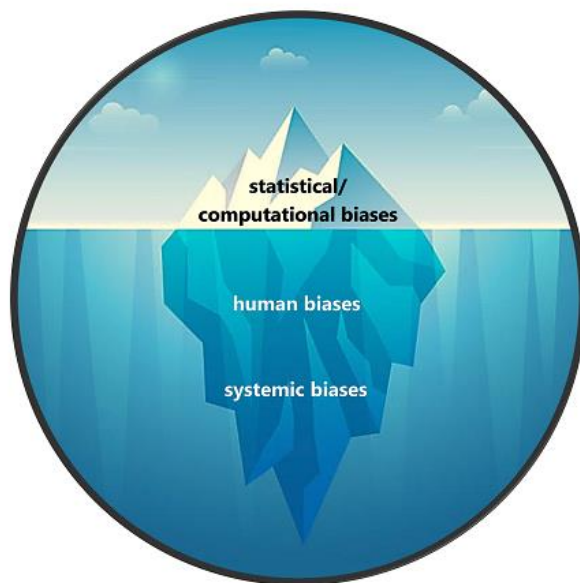
χρησιμοποιήθηκαν κατά την εκπαίδευσή του, για τη δομή και τον τρόπο λειτουργίας του, για το σκοπό λειτουργίας και τη χρήση του, ώστε να μπορούν να ανιχνευθούν με ακρίβεια οι έμφυλες διακρίσεις που τυχόν παρουσιάζει και κατά συνέπεια να επιτευχθεί η διόρθωση των σφαλμάτων και η ενσωμάτωση των ηθικών κανόνων.

Προς αυτήν την κατεύθυνση καταλύτης είναι η λογοδοσία (accountability), με την έννοια του ποιος θα φέρει την ευθύνη για τις επιβλαβείς συνέπειες που ενδεχομένως προκαλέσει ένα σύστημα TN (Schwartz, et al., 2022), οι οποίες όσον αφορά τις έμφυλες διακρίσεις, έγκεινται στις περισσότερες περιπτώσεις στην παραβίαση του θεμελιώδους δικαιώματος στην ισότητα, της αρχής απαγόρευσης των διακρίσεων, αλλά και στην ευθεία προσβολή της προσωπικότητας του ατόμου. Η λογοδοσία είναι ένας παράγοντας που μπορεί να ασκήσει πίεση στους δημιουργούς των συστημάτων TN και τις μεγάλες τεχνολογικές εταιρίες για την ενσωμάτωση ηθικών κανόνων.

3.5.2 *Debiasing*

Για να αξιολογηθεί εάν ένας αλγόριθμος είναι απαλλαγμένος από μεροληψίες, πρέπει να εξεταστεί το σύνολο της αλγοριθμικής διαδικασίας. Αυτό συνεπάγεται ότι πρέπει πρώτα να επιβεβαιωθεί ότι οι αρχικές εντολές που έχουν δοθεί στην TN και η μοντελοποίηση (διαδικασία ανάλυσης και αποσαφήνισης των δεδομένων) δεν περιλαμβάνουν μεροληψίες. Δεύτερον, ότι η εκπαίδευσή του και τα δεδομένα εκπαίδευσης (training data) δεν περιλαμβάνουν μεροληψίες· και τέλος, ότι ο αλγόριθμος μπορεί να παράγει αποτελέσματα επαρκώς για το συγκεκριμένο σκοπό και πλαίσιο, εξετάζεται δηλαδή σε συνάρτηση με το σκοπό τον οποίο εξυπηρετεί (Ferrer, et al., 2021). Ένα μέτρο που προτείνεται και μπορεί να εφαρμοστεί ήδη εκ του σχεδιασμού του συστήματος TN, είναι η εκ των προτέρων αφαίρεση όλων των παραγόντων που μπορούν να οδηγήσουν σε μεροληπτικά αποτελέσματα (debiasing) (Balayn & Gürses, 2021). Η διαδικασία αυτή μπορεί να γίνει είτε με την αφαίρεση της ένδειξης του φύλου στα δεδομένα εκπαίδευσης (training data) ενός συστήματος TN, ώστε τα αποτελέσματά του να μη συνδέονται με αυτό (Barocas, et al., 2019) είτε με την προσθήκη ορισμένων στοιχείων και δεδομένων υπέρ των υποεκπροσωπούμενων ατόμων προς εξισορρόπηση της έλλειψης δεδομένων, ώστε τα αποτελέσματα να είναι περισσότερο δίκαια (Balayn & Gürses, 2021). Σημειώνεται ότι αυτή η πρακτική (debiasing) μειώνει την ακρίβεια του συστήματος TN, ενώ δεν αποκλείει με σιγουριά

την εμφάνιση έμφυλων διακρίσεων, δεδομένου ότι ενδέχεται να παραμείνουν μέσα στα δεδομένα εκπαίδευσης (training data) άλλες πληροφορίες, οι οποίες συνδέονται έμμεσα με το φύλο (Ferrer, et al., 2021), καταλήγοντας στο να καταστεί ακόμα πιο δύσκολη η πλήρης εξέταση των δεδομένων εκπαίδευσης (training data) και ο εντοπισμός των σφαλμάτων (European Union Agency for Fundamental Rights, 2018). Επιπλέον, η λήψη υπόψιν του φύλου ενδέχεται σε ορισμένες περιπτώσεις να είναι σημαντική για την επίτευξη συμπεριληπτικότητας και ποικιλομορφίας (Floridi, et al., 2020). Γενικότερα, οι αμιγώς τεχνολογικές λύσεις, αποξενωμένες από το κοινωνικο-ηθικό πλαίσιο, δεν κρίνονται ικανές να οδηγήσουν αποτελεσματικά στην αντιμετώπιση των έμφυλων διακρίσεων της ΤΝ, αντίθετα απαιτείται μια ολιστική προσέγγιση (Hao, 2019). Ο σεβασμός της ισότητας σε συνάρτηση με την κατανόηση του εκάστοτε ηθικο-κοινωνικού πλαισίου και των ιστορικών ανισοτήτων, όπου έχουν τις ρίζες τους οι έμφυλες διακρίσεις, είναι απαραίτητες προϋποθέσεις για την ανάπτυξη και το σχεδιασμό συστημάτων ΤΝ που θα ενσωματώνουν ηθικές αξίες (Aizenberg & van den Hoven, 2020).



Εικόνα 10: Το βάθος του προβλήματος των μεροληψιών (biases) της ΤΝ - Βάση: Συστημικές/ιστορικές μεροληψίες, Μέση: Ανθρώπινες μεροληψίες, Κορυφή: Στατιστικές/Αλγοριθμικές μεροληψίες (Schwartz, et al., 2022)

3.6 Προς μια ηθική TN – Τάσεις, διαδικασίες και προβληματισμοί περί ηθικής

Οι έμφυλες διακρίσεις της TN αποτελούν ζήτημα το οποίο επηρεάζει άμεσα συγκεκριμένες πληθυσμιακές ομάδες, καταλήγοντας σε δυσμενή αποτελέσματα, διακρίσεις, διαμορφώνοντας καταστάσεις και ενισχύοντας τα έμφυλα στερεότυπα. Είναι σαφές ότι δεν πρόκειται για ένα αμιγώς τεχνολογικό πρόβλημα και δεν πρέπει να εξετάζεται απομονωμένα, αλλά σε συνάρτηση με το ηθικο-κοινωνικό πλαίσιο (Selbst, et al., 2019). Επομένως, οι προσεγγίσεις που εστιάζουν αμιγώς σε τεχνολογικές λύσεις και τη διαδικασία του αλγορίθμου και τη λογική του, δεν κρίνονται επαρκείς (Ferrer, et al., 2021). Η ηθική της TN (AI ethics) αποσκοπεί στην επίλυση αυτών των ζητημάτων. Η ηθική της TN είναι ένα σύνολο αξιών, αρχών και τεχνικών που χρησιμοποιούν ευρέως αποδεκτά πρότυπα σωστού και λάθους για την καθοδήγηση της ηθικής συμπεριφοράς στην ανάπτυξη και χρήση τεχνολογιών TN. Το πεδίο της ηθικής της TN έχει εμφανιστεί σε μεγάλο βαθμό ως απάντηση στο φάσμα των ατομικών και κοινωνικών βλαπτικών συνεπειών που μπορεί να προκαλέσει η κακή χρήση, η κατάχρηση, ο κακός σχεδιασμός ή οι αρνητικές ακούσιες συνέπειες των συστημάτων TN (Leslie, 2019).

3.6.1 Ορθολογικές (*rational*) και σχεσιακές (*relational*) προσεγγίσεις

Όσο περισσότερο χρησιμοποιούνται οι αλγόριθμοι για τη λήψη αποφάσεων αυτοματοποιώντας εκφάνσεις της κοινωνικής ζωής, τα ηθικά, κοινωνικά και πολιτικά αμφιλεγόμενα θέματα που παραδοσιακά γίνονταν αντικείμενο συζήτησης, αντιπαράθεσης και εξέτασης, τείνουν να περιορίζονται σε μαθηματικά προβλήματα με τεχνική λύση. Αυτή η μαθηματικοποίηση (*mathematization*) και η τυποποίηση (*formalization*) των κοινωνικών ζητημάτων, οδηγεί στη θεώρηση των αποτελεσμάτων των αλγορίθμων ως αντικειμενικά, ουδέτερα, αξιόπιστα και χωρίς ηθικά πρόσημα (*amoral*). Παρατηρείται δηλαδή μια τάση αμοραλισμού στα συστήματα TN. Έτσι, οι χαρακτηρισμοί που ενδέχεται να κάνει μια TN, για παράδειγμα η «αποδεκτή» συμπεριφορά, ο «κανονικός» σωματότυπος ή ακόμα και η κρίση περί του φύλου ενός προσώπου, γίνονται αποδεκτοί ως απλές εργασίες ταξινόμησης και κατηγοριοποίησης, χωρίς πολιτικό χαρακτήρα. Τα άνισα και επιβλαβή αποτελέσματα, κατά συνέπεια, αντιμετωπίζονται ως απλές παρενέργειες που μπορούν να αντιμετωπιστούν με τεχνικές λύσεις, όπως η αφαίρεση των μεροληψιών από τα δεδομένα εκπαίδευσης (*debiasing*) και όχι ως προβλήματα που έχουν βαθιές ρίζες σε ιστορικές ανισότητες και ασύμμετρες ιεραρχίες ισχύος ή ανεξέταστες προβληματικές υποθέσεις που διεισδύουν σε πρακτικές δεδομένων.

Οι ορθολογικές προσεγγίσεις της ηθικής (rationalism), που αποτελούν τη βάση της δυτικής επιστήμης και φιλοσοφίας, θέτουν στο επίκεντρο την ορθολογική σκέψη, με την έννοια ότι η λογική για την κατάκτηση της γνώσης και την κατανόηση του κόσμου υπερέχει από την εξέταση των σχέσεων και τη βιωμένη εμπειρία. Το προνόμιο της λογικής ως το απόλυτο κριτήριο καθιστά τη γνώση μια αποστασιοποιημένη πράξη. Η αναζήτηση για την ορθολογική κοσμοθεωρία βασίζεται στη βεβαιότητα, σταθερότητα και τάξη, και συνεπώς στην απομόνωση, το διαχωρισμό και τα καθορισμένα δίπολα. Έννοιες, όπως η γνώση και η ηθική τείνουν να είναι αφηρημένες, χωρίς φύλο, εκτός πλαισίου και χωρίς φυλή. Η γνώση, σύμφωνα με αυτή την κοσμοθεωρία, έχει τις ρίζες της στο ιδανικό ορθολογικό, στατικό, αυτοτελές και αυτάρκες υποκείμενο που ατενίζει τον εξωτερικό κόσμο από μακριά με έναν «καθαρά γνωστικό» τρόπο ως ασώματος και αδιάφορος παρατηρητής.

Σε αντίθεση με τις ορθολογικές προσεγγίσεις και την ατομικιστική κοσμοθεωρία, οι σχεσιακές προσεγγίσεις (relationalism) βλέπουν την ύπαρξη ως θεμελιωδώς συνυπάρχουσα με ένα δίκτυο σχέσεων. Σε μια προσπάθεια επαναπροσδιορισμού της ηθικής, η βιωμένη εμπειρία των ατόμων που επηρεάζονται τίθεται στο επίκεντρο, με έμφαση στη συνειδητοποίηση και την αναγνώριση ιστορικών ανισοτήτων και του αντικτύπου των συστημάτων ΤΝ σε ευάλωτες ομάδες. Είναι καιρία επομένως η αναγνώριση ότι στους τομείς της πληροφορικής και των επιστημών δεδομένων, το πεδίο κυριαρχείται σε μεγάλο βαθμό από προνομιούχες ομάδες κυρίως ελίτ, Δυτικών, λευκών ανδρών cisgender (American Psychological Association, χ.χ.) και άρτιων σωματικά. Έτσι, η γνώση που παράγεται και τα αποτελέσματά της είναι περιορισμένα στην οπτική της προνομιούχας αυτής ομάδας (βλ. ανωτέρω 2.9.4). Τα άτομα που επηρεάζονται και αποκλείονται από τις έμφυλες διακρίσεις της ΤΝ είναι οι γυναίκες, τα μη-δυναδικά άτομα και η τρανς κοινότητα, συνεπώς είναι απαραίτητη η συμπερίληψη των ατόμων αυτών, καθώς είναι τα μόνα σε θέση να συνεισφέρουν τα βιώματά τους και να προτείνουν ουσιαστικές λύσεις. Συνεπώς, το ζήτημα των έμφυλων διακρίσεων αναδεικνύεται ως ηθικο-κοινωνικό ζήτημα με τεχνολογικές προεκτάσεις, το οποίο απαιτεί μια ολιστική προσέγγιση, λαμβάνοντας υπόψη ηθικές οπτικές, φεμινιστικές προσεγγίσεις και τη βιωμένη εμπειρία και όχι απλά τεχνικές λύσεις (Birhane, 2021).

3.6.2 *Top-down και bottom-up ηθική ρύθμιση*

Η ενσωμάτωση ηθικών αρχών στην ΤΝ μπορεί να γίνει με δύο τρόπους, είτε από πάνω προς τα κάτω (εφεξής top-down) είτε από κάτω προς τα πάνω (εφεξής bottom-up). Η top-down ρύθμιση αναφέρεται στην ύπαρξη ορισμένων κανόνων για τη λειτουργία της ΤΝ, οι οποίοι μετατρέπονται σε αλγόριθμο και ενσωματώνονται σε αυτήν. Οι κανόνες αυτοί μπορεί να προέρχονται από τη νομοθεσία, τη θρησκεία, την ηθική, τη φιλοσοφία. Παράδειγμα της top-down ρύθμισης είναι η πληθώρα δεσμευτικών και μη κανόνων που έχουν παραχθεί μέχρι τώρα και οι οποίοι αποσκοπούν στο να ρυθμίσουν ηθικά, νομικά και τεχνολογικά την ΤΝ, προβλέποντας ορισμένους θεμελιώδεις κανόνες και ελάχιστες προϋποθέσεις που θα πρέπει να ακολουθούνται (βλ. ανωτέρω 1.5). Κατά την top-down ρύθμιση, η ηθική της ΤΝ εξετάζεται με βάση τα αποτελέσματά της, δηλαδή εάν το σύστημα ΤΝ καταλήγει σε ένα χρήσιμο και θετικό αποτέλεσμα προς όφελος των ατόμων (utilitarianism), με βάση το κατά πόσον είναι διακριτά τα όρια υποχρεώσεων και δικαιωμάτων και ο σεβασμός τους (deontology), αλλά και με βάση εάν το ίδιο το σύστημα ΤΝ ενσωματώνει ηθικές αρχές και λειτουργεί βάσει αυτών (virtue ethics). Εδώ ανακύπτουν δύο σημαντικά ζητήματα· αφενός, πώς κρίνεται τι πρέπει να περιλαμβάνει η ΤΝ ως ηθικό χαρακτηριστικό και τι όχι, από τη στιγμή που δεν υπάρχει ένας κοινός παγκόσμιος ηθικός κώδικας, αλλά τόσοι, όσα και τα κράτη. Αφετέρου, παρουσιάζει πρακτική δυσκολία η μετατροπή αφηρημένων και πολύπλοκων ηθικών εννοιών και ιδεών σε κώδικα, μαθηματική και τεχνική γλώσσα που μπορεί να αντιληφθεί η ΤΝ (Wallach & Allen, 2010). Παρά τις δυσκολίες και τις πολιτισμικές διαφορές περί ηθικής, κοινή βάση όλων των προσπαθειών ρύθμισης της ΤΝ μέχρι τώρα, είτε με τη μορφή κατευθυντήριων γραμμών είτε με τη μορφή δεσμευτικών κανόνων, είναι η αδιαμφισβήτητη θέση ότι αυτή τίθεται στην υπηρεσία της ανθρωπότητας, ενώ περαιτέρω, προβάλλεται ως καίρια επιταγή ο σεβασμός προς τα θεμελιώδη ανθρώπινα δικαιώματα, μεταξύ των οποίων είναι η ισότητα και η απαγόρευση των διακρίσεων (Rodrigues, 2020). Επομένως, τουλάχιστον ως προς ορισμένα βασικά ανθρώπινα δικαιώματα και ηθικές αρχές φαίνεται να υπάρχει κοινή αποδοχή (consensus), ορίζοντας κατά αυτόν τον τρόπο τη βάση της ρύθμισης.

Η bottom-up προσέγγιση αντλεί την έμπνευσή της από την ικανότητα του ανθρώπου να μαθαίνει από τα πρώτα στάδια της ζωής του, να αντιλαμβάνεται και να ενσωματώνει ηθικές προσαγές, διαμορφώνοντας αντίστοιχα την προσωπικότητά του. Έτσι, η μηχανική μάθηση

αναπαριστά ακριβώς αυτό, το πώς η ΤΝ μπορεί να αφηθεί ελεύθερη να ανακαλύψει τον κόσμο των ανθρώπων και να μάθει. Οι άνθρωποι ήδη από την παιδική ηλικία μαθαίνουν τις έννοιες του καλού και του κακού, αποκτούν εμπειρίες μέσα από την επιβράβευση και την αποτυχία και μαθαίνουν πώς να συνυπάρχουν αρμονικά στο κοινωνικό σύνολο σεβόμενοι και ενσωματώνοντας τους ηθικούς, κοινωνικούς και νομικούς κανόνες. Κατ' επέκταση, αν ορισμένες βασικές αρχές και περιορισμοί ενσωματωθούν σε ένα σύστημα ΤΝ με τη μορφή ενός ηθικού κώδικα ήδη από τα πρώιμα στάδια ανάπτυξής του, θα μπορούσαν να λειτουργήσουν ως «συνείδηση» του συστήματος, βοηθώντας το να αναπτυχθεί ηθικά αυτόνομα (Wallach & Allen, 2010). Πλεονέκτημα αυτής της προσέγγισης είναι ότι επιτρέπει στην ΤΝ να προσαρμόζεται ανάλογα με τις ηθικο-κοινωνικές συνθήκες που επικρατούν, έχοντας μια δυναμική σχέση με την κοινωνία. Ένα μειονέκτημα είναι ότι μπορεί να παρουσιάσει απρόβλεπτη συμπεριφορά με βλαπτικά αποτελέσματα για ορισμένες επηρεαζόμενες ομάδες χωρίς δυνατότητα να εξεταστεί ο τρόπος λειτουργίας της και να εντοπιστεί η ρίζα του προβλήματος (Tasioulas, 2019).

Ο συνδυασμός top-down και bottom-up προσεγγίσεων, ώστε αφενός να υπάρχει εποπτεία και έλεγχος της ΤΝ και αφετέρου να υπάρχει ευελιξία και προσαρμοστικότητα, φαίνεται να είναι η καταλληλότερη λύση. Με έμπνευση την Αριστοτελική φιλοσοφία, με βάση την οποία η διανοητική αρετή μπορεί να διδαχθεί, ενώ τις ηθικές αρχές μπορεί κάποιος να τις μάθει μέσω της συνήθειας και της εξάσκησης, ο συνδυασμός top-down και bottom-up προσεγγίσεων, συνεπάγεται ότι η ηθική ρύθμιση της ΤΝ θα πρέπει να βασίζεται στη συνδυαστική πρόβλεψη και ενσωμάτωση ορισμένων κανόνων ήδη από τη δημιουργία της και στην ελεύθερη αυτόνομη λειτουργία και ανάπτυξή της, ώστε να μάθει από το περιβάλλον της και να διαμορφώσει το δικό της «ηθικό κώδικα» (Wallach & Allen, 2010). Σημαντική προϋπόθεση για την ενσωμάτωση ηθικών κανόνων από την ΤΝ είναι αφενός οι φιλόσοφοι και οι θεωρητικοί της ηθικής να μην εξετάζουν το ζήτημα από απόσταση και καθαρά θεωρητικά, αλλά να αναζητούν πρακτικές εφαρμογές, και αφετέρου οι προγραμματιστές να ξεφύγουν από τα αυστηρά τεχνολογικά πλαίσια και να έρθουν σε επαφή με ηθικούς και κοινωνικούς προβληματισμούς, αντιλαμβανόμενοι ότι η ΤΝ αποτελεί ένα πεδίο με κοινωνικο-τεχνολογικές (socio-technical) προεκτάσεις (Schwartz, et al., 2022). Είναι καίρια επομένως η συνειδητοποίηση του προβλήματος των έμφυλων διακρίσεων και των αιτιών τους, ώστε να συμπεριληφθούν ειδικοί κανόνες σε

κάθε είδους ηθικό κώδικα που καλείται να ενσωματώσει ο προγραμματιστής σε ένα σύστημα TN (Smith & Rustagi, 2021).

3.6.3 Ηθική αξιολόγηση (*ethics assessment*)

Η ηθική ρύθμιση της TN συνεπάγεται την ηθική αξιολόγηση των συστημάτων TN (*ethics assessment*), ενδεχομένως και μέσω διαδικασίας πιστοποίησης (UNESCO, 2020). Ειδικότερα, προτείνεται ηθική αξιολόγηση των συστημάτων TN σε όλα τα στάδια λειτουργίας τους, δηλαδή τη λήψη της απόφασης για τη δημιουργία τους, το σχεδιασμό τους, την παράδοσή τους προς χρήση, αλλά και μετά από αυτήν, ώστε να εξετάζεται κάθε φορά ποιος τα ελέγχει, ποιος ωφελείται, ποιος επηρεάζεται δυσμενώς, πώς χρησιμοποιούνται, αν εκδηλώνουν μεροληπτικές συμπεριφορές και γενικά αν σέβονται ηθικές και νομικές επιταγές.

Η ηθική αξιολόγηση της TN μπορεί να επιτευχθεί με τη συμμετοχή του ανθρώπινου παράγοντα σε ορισμένες διαδικασίες, ώστε να υπάρχει εποπτεία (*human in the loop*), με την προϋπόθεση της ποικιλομορφίας και αντιπροσωπευτικότητας (Schwartz, et al., 2022). Ένα άλλο μέτρο είναι η αξιολόγηση των αλγορίθμων από άλλους αλγορίθμους. Ειδικότερα, προτείνεται η δημιουργία εργαλείων TN, με σκοπό τον έλεγχο και εποπτεία των συστημάτων TN που λαμβάνουν αποφάσεις, για τον εντοπισμό των σφαλμάτων και της παράνομης και μη δίκαιης λειτουργίας τους, από την άποψη της παραβίασης δικαιωμάτων, ηθικών και νομικών κανόνων (Reichman & Sartor, 2022). Σε αυτό το πλαίσιο προτείνεται και η δημιουργία των λεγόμενων «*ethics bots*», δηλαδή προγραμμάτων TN που αναλύουν χιλιάδες στοιχεία πληροφοριών, όχι μόνο πληροφορίες που είναι διαθέσιμες δημόσια στο διαδίκτυο, αλλά και πληροφορίες που συλλέγονται από τους υπολογιστές ενός ατόμου, για να αντιληφθούν τις ηθικές προτιμήσεις του. Ο σκοπός είναι τα *ethics bots*, αφού έχουν εξοικειωθεί με τις ηθικές αρχές ενός συγκεκριμένου ατόμου, να καθοδηγούν τις εφαρμογές TN που αυτό χρησιμοποιεί, ώστε να λειτουργούν βάσει αυτών (Etzioni & Etzioni, 2016). Τα *ethics bots*, θα μπορούσαν να χρησιμοποιηθούν τόσο για εξατομικευμένες υπηρεσίες, όσο και για την εποπτεία και συμμόρφωση συστημάτων TN ευρέως, για την ηθική τους αξιολόγηση.

Οι μεγάλες τεχνολογικές εταιρίες, συνειδητοποιώντας ότι η ενσωμάτωση ηθικών αρχών και η προώθηση της ποικιλομορφίας και συμπεριληπτικότητας στις εφαρμογές TN που προσφέρουν προσδίδει ανταγωνιστικό πλεονέκτημα και μπορεί να αποβεί κερδοφόρα, έχουν αρχίσει να

κάνουν τα πρώτα βήματα προς αυτήν την κατεύθυνση. Η IBM έχει αναπτύξει ήδη ένα εργαλείο ανοιχτού κώδικα (open source), το AI Fairness 360 για την εξέταση των συστημάτων TN, την αξιολόγηση του κατά πόσον παρουσιάζουν μεροληψίες και την καταπολέμησή τους (IBM Research Trusted AI, χ.χ.). Η Microsoft έχει σχηματίσει μια ομάδα σημαντικών στο χώρο της ηθικής της TN επιστημόνων, με σκοπό την έρευνα στο πεδίο των μεροληψιών της TN και την προώθηση του δικαίου, λογοδοσίας, διαφάνειας και ηθικής στην TN (Fairness, Accountability, Transparency, Ethics-FATE) (Microsoft, χ.χ.), παράγοντας σχετικό έργο με προτάσεις για το σχεδιασμό περισσότερο συμπεριληπτικών και ηθικών συστημάτων TN (Microsoft Design, χ.χ.). Αξιοσημείωτη είναι και η πρωτοβουλία της Meta (πρώην Facebook), η οποία διαθέτει επίσης ερευνητική ομάδα για την προώθηση δίκαιης και ηθικής TN, ενώ πρόσφατα δημιούργησε ένα ανοιχτού κώδικα (open source) σύνολο δεδομένων με πραγματικές συζητήσεις ανθρώπων από διαφορετικές ηλικίες, φυλές και φύλα (Casual Conversations data set), ώστε αυτό να μπορεί να χρησιμοποιηθεί από ερευνητές και προγραμματιστές για τον εντοπισμό αποκλίσεων και μεροληψιών στα συστήματα TN αναγνώρισης φωνής, αλλά και στη δημιουργία συμπεριληπτικών και αντιπροσωπευτικών συστημάτων TN αναγνώρισης φωνής, όσο το δυνατόν περισσότερο απαλλαγμένα από μεροληψίες (Sari, et al., 2022).

Συνοψίζοντας, δεδομένου ότι το ζήτημα των έμφυλων διακρίσεων στην TN δεν είναι ένα αμιγώς τεχνολογικό πρόβλημα, αλλά έχει ηθικο-κοινωνικές προεκτάσεις, το έργο της εξάλειψής τους από την TN καθίσταται αμφίβολο και δύσκολο. Ωστόσο, η TN τις μετατρέπει σε μετρήσιμα δεδομένα, τα οποία μπορούν να εξεταστούν και να εντοπιστούν, προσφέροντας κατά αυτόν τον τρόπο την ευκαιρία για διόρθωσή τους (Silberg & Manyika, 2019). Οι γυναίκες, τα μη-δυναμικά άτομα και άτομα όλων των φυλών, τα οποία μέχρι τώρα αποκλείονταν, είναι τα μόνα που μπορούν να συνεισφέρουν ουσιαστικά με τα προσωπικά τους βιώματα και εμπειρίες, ώστε να επιτευχθεί πραγματική αλλαγή. Με ευαισθητοποίηση, συνεργασία των νομικών, φιλοσόφων και προγραμματιστών για το σχεδιασμό της TN με σκοπό την πραγματική εξυπηρέτηση όλων των ανθρώπων και όχι μόνο μιας περιορισμένης ελίτ, μπορεί να δημιουργηθεί TN η οποία θα σέβεται θεμελιώδη ανθρώπινα δικαιώματα και ηθικές αρχές, με σκοπό το κοινό καλό (Floridi,

et al., 2020). Δεν αρκεί το ερώτημα αν η TN είναι καλή. Το πραγματικό ερώτημα είναι για ποιον είναι καλή (Broussard, 2018).

ΕΠΙΛΟΓΟΣ – ΣΥΜΠΕΡΑΣΜΑ

Η ΤΝ και η μεταφορά όλο και περισσότερων εκφάνσεων της καθημερινής ζωής στον ψηφιακό κόσμο δεν είναι πλέον επιστημονική φαντασία, αλλά πραγματικότητα. Κοινωνία και τεχνολογία αλληλοεπηρεάζονται, αναδιαμορφώνοντας και επαναπροσδιορίζοντας έννοιες και καταστάσεις. Μέσα σε αυτό το πλαίσιο, η ΤΝ είναι ένα πεδίο που όσο ενδιαφέρον, καινοτόμο και χρήσιμο είναι, άλλο τόσο αμφιλεγόμενο και προβληματικό εξακολουθεί να παραμένει, καθώς εγείρει πληθώρα ηθικών ζητημάτων και ενέχει τον κίνδυνο προσβολής θεμελιωδών ανθρωπίνων δικαιωμάτων.

Η ΤΝ αποτελεί το νέο πεδίο, όπου όλες οι κοινωνικές αντιλήψεις, συμπεριλαμβανομένων των έμφυλων διακρίσεων, βρίσκουν νέο έδαφος για να εκδηλωθούν. Αυτό είναι αναμενόμενο και μέχρι ένα σημείο κατανοητό, από την άποψη ότι οι εφαρμογές της ΤΝ είναι προϊόντα του ανθρώπου. Προέρχονται από δική του έμπνευση και δημιουργία, επομένως είναι φυσικό αυτή να φέρει και τα αρνητικά του στοιχεία και αντιλήψεις. Δεδομένου ότι οι στερεοτυπικές αντιλήψεις περί φύλου και φυλής είναι βαθιά εδραιωμένες στις ανθρώπινες κοινωνίες και απορρέουν από ιστορικές ανισότητες, είναι αναπόφευκτο η ΤΝ όχι μόνο να μην αμβλύνει το χάσμα, αντίθετα, λόγω της ευρείας χρήσης της και της υπολογιστικής της δύναμης, να το εντείνει. Η φεμινιστική θεωρία έχει συμβάλει σημαντικά στην ευαισθητοποίηση για το ζήτημα των έμφυλων διακρίσεων στην ΤΝ, ωστόσο, το γυναικείο και μη-δυναμικό φύλο, αλλά και φυλές που αποτελούν μειονότητες, εξακολουθούν να πλήττονται από τα στερεότυπα, τα οποία βρίσκουν τρόπο να επιβιώνουν και να αλλάζουν μορφή, ανάλογα με τις συνθήκες.

Κοινή βάση όλων των προσπαθειών ρύθμισης της ΤΝ μέχρι τώρα, είτε με τη μορφή κατευθυντήριων γραμμών είτε με τη μορφή δεσμευτικών κανόνων, είναι η αδιαμφισβήτητη θέση ότι αυτή τίθεται στην υπηρεσία της ανθρωπότητας, ενώ περαιτέρω, προβάλλεται ως καίρια επιταγή ο σεβασμός προς τα θεμελιώδη ανθρώπινα δικαιώματα, όπως αυτά κατοχυρώνονται από το διεθνές δίκαιο, γενικότερα, και το δίκαιο της ΕΕ και των κρατών, ειδικότερα, με ιδιαίτερη έμφαση στην ανάγκη διασφάλισης της ισότητας και της απαγόρευσης των διακρίσεων. Επισημαίνεται ότι η πιο πρόσφατη πρωτοβουλία της ΕΕ, η Πράξη της ΕΕ για τις Ψηφιακές Υπηρεσίες (Digital Services Act), η οποία εγκρίθηκε στις 5 Ιουλίου 2022 από το

Ευρωπαϊκό Κοινοβούλιο, στοχεύει σε ένα ασφαλέστερο ψηφιακό περιβάλλον, προβλέποντας αυστηρότερες προϋποθέσεις λογοδοσίας των μεγάλων ψηφιακών πλατφορμών και διαφάνεια των αλγορίθμων βάσει των οποίων λειτουργούν. Συγκεκριμένα, μεταξύ άλλων, απαιτεί διαφάνεια από τις ψηφιακές πλατφόρμες για διάφορα ζητήματα, συμπεριλαμβανομένων των αλγορίθμων που χρησιμοποιούνται για τη σύσταση περιεχομένου ή προϊόντων στους χρήστες, ενώ επιπλέον προβλέπει ότι ερευνητές μπορούν να έχουν πρόσβαση στα δεδομένα βασικών πλατφορμών, προκειμένου να ελέγχουν τον τρόπο λειτουργίας τους (European Parliament, 2022). Είναι σαφής επομένως η πρόθεση για τον καθορισμό ελάχιστων προαπαιτούμενων, με σκοπό μια περισσότερο ηθική και δίκαιη ΤΝ.

Η εκ των άνω ρύθμιση, ωστόσο, ή αμιγώς τεχνολογικές λύσεις, δεν κρίνονται αυτοτελώς επαρκείς για την αντιμετώπιση του προβλήματος των έμφυλων διακρίσεων στην ΤΝ και την ενσωμάτωση κανόνων ηθικής και δικαιοσύνης σε αυτήν. Απαρέγκλιτη προϋπόθεση για ουσιαστική αλλαγή είναι η πραγματική συνειδητοποίηση της έκτασης και των αιτιών του ζητήματος, η διεπιστημονικότητα, η συμπερίληψη και ποικιλομορφία, ώστε το γυναικείο και μη-δυναδικό φύλο, αλλά και κάθε φυλή να έχουν λόγο στη λήψη αποφάσεων, το σχεδιασμό συστημάτων ΤΝ, τη διαμόρφωση κανόνων, τη σύλληψη ιδεών και την υλοποίησή τους. Μόνο με αυτόν τον τρόπο μπορεί να περιοριστεί η κυριαρχία των λευκών ανδρών που λαμβάνουν τις αποφάσεις από τα γραφεία των high tech εταιριών τους και να μετατοπιστεί το βάρος εξουσίας.

Με την αυξανόμενη αυτοματοποίηση και την υλοποίηση ενός πλήρους ψηφιακού κόσμου και ζωής, στην οποία οδηγεί, ενδεχομένως αναπόφευκτα και η υπόσχεση του Metaverse, ως φυσική εξέλιξη της σύγχρονης ζωής, η ανθρωπότητα βρίσκεται σε ένα σημείο καμπής, καλούμενη να αποφασίσει τι είδους ψηφιακή κοινωνία θέλει να δημιουργήσει. Η παγκόσμια ερευνητική και νομοθετική κινητοποίηση προσφέρουν μια σημαντική ευκαιρία δημιουργίας πραγματικά ηθικής και δίκαιης ΤΝ, απαλλαγμένης από παρωχημένα δίπολα και στερεότυπα, συμπεριληπτικής και ικανής για την ανοικοδόμηση ενός περισσότερο ισότιμου κόσμου. Ευκαιρία που δεν πρέπει να χαθεί.

ΒΙΒΛΙΟΓΡΑΦΙΑ

2001: A space odyssey. 1968. [Film] Directed by Stanley Kubrick. UK/USA: Metro-Goldwyn-Mayer(MGM)/Stanley Kubrick Productions.

Adeyemo, B., 2021. *I'm a Black woman and the metaverse scares me – here's how to make the next iteration of the internet inclusive.* [Ηλεκτρονικό]

Available at: <https://theconversation.com/im-a-black-woman-and-the-metaverse-scares-me-heres-how-to-make-the-next-iteration-of-the-internet-inclusive-173310>

[Πρόσβαση 26 Ιανουαρίου 2022].

ADMIN Cross-linked glossary, 2019. *European Commission: CROS Collaboration in Research and Methodology for Official Statistics.* [Ηλεκτρονικό]

Available at: https://ec.europa.eu/eurostat/cros/content/bias_en

[Πρόσβαση 19 Ιανουαρίου 2022].

Aizenberg, E. & van den Hoven, J., 2020. Designing for human rights in AI. *Big Data & Society*, pp. 1-14.

Algorithmic Justice League, χ.χ. *Algorithmic Justice League.* [Ηλεκτρονικό]

Available at: <https://www.ajl.org/>

[Πρόσβαση 25 Ιανουαρίου 2022].

American Psychological Association, χ.χ. *APA Dictionary of Psychology.* [Ηλεκτρονικό]

Available at: <https://dictionary.apa.org/cisgender>

[Πρόσβαση 10 Μαΐου 2022].

Angwin, J., Larson, J., Mattu, S. & Kirchner, L., 2016. *MachineBias.* [Ηλεκτρονικό]

Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

[Πρόσβαση 20 Ιανουαρίου 2022].

Apple, χ.χ. *Siri does more than ever. Even before you ask.* [Ηλεκτρονικό]
Available at: <https://www.apple.com/siri/>
[Πρόσβαση 26 Ιανουαρίου 2022].

Azria, S. & Wickert, F., 2019. *Facial Recognition: Current Situation and Challenges.* [Ηλεκτρονικό]
Available at: <https://rm.coe.int/t-pd-2019-05rev-facial-recognition-report-003-/16809eadf1>
[Πρόσβαση 14 Νοεμβρίου 2021].

Balayn, A. & Gürses, S., 2021. *Beyond Debiasing: Regulating AI and its Inequalities*, Brussels: EDRI.

Barocas, S., Hardt, M. & Narayanan, A., 2019. *Fairness and Machine Learning: Limitations and Opportunities.* s.l.:fairmlbook.org.

Birhane, A., 2021. Algorithmic injustice: a relational ethics approach. *Patterns*, Volume 2, pp. 1-9.

Bolukbasi, T. et al., 2016. *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings.* New York, Curran Associates Inc., p. 4356–4364.

Boucher, P. N., 2020. *Artificial intelligence: How does it work, why does it matter, and what can we do about it?.* [Ηλεκτρονικό]
Available at: [https://www.europarl.europa.eu/stoa/en/document/EPRS_STU\(2020\)641547](https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2020)641547)
[Πρόσβαση 14 Νοεμβρίου 2021].

Braidotti, R., 2020. Κυβερνοφεμινισμός με διαφορά. Στο: Μ. Πατηνιώτης, επιμ. *Εισαγωγή στις Ψηφιακές Σπουδές.* Θεσσαλονίκη: Εκδόσεις Ροπή, pp. 293-316.

Broussard, M., 2018. *Artificial Unintelligence: How Computers Misunderstand the World.* Cambridge: MIT Press.

Buolamwini, J. & Gebru, T., 2018. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.* s.l., s.n.

Butler, J., 2009. *Αναταραχή Φύλου: Ο Φεμινισμός και η Ανατροπή της Ταυτότητας.* Αθήνα: Εκδόσεις Αλεξάνδρεια.

Cambridge Dictionary, χ.χ. *gender discrimination*. [Ηλεκτρονικό]
Available at: <https://dictionary.cambridge.org/dictionary/english/gender-discrimination>
[Πρόσβαση 22 Ιανουαρίου 2022].

Cave, S. & Dihal, K., 2020. The Whiteness of AI. *Philosophy & Technology*, Volume 33, pp. 685-703.

Coded Bias. 2020. [Film] Directed by Shalini Kantayya. United States; China; United Kingdom; 7th Empire Media; Chicken and Egg Pictures; Ford Foundation-Just Films;

Collett, C. & Dillon, S., 2019. *AI and Gender: Four Proposals for*, Cambridge: Cambridge: The Leverhulme Centre for the Future of Intelligence.

Consalvo, M., 2002. *Cyberfeminism*. [Ηλεκτρονικό]
Available at: https://study.sagepub.com/sites/default/files/Ch17_Cyberfeminism.pdf
[Πρόσβαση 23 Ιανουαρίου 2022].

Costanza-Chock, S., 2018. *Design Justice: towards an intersectional feminist framework for design theory and practice*. Limerick, Design Research Society.

Council on Artificial Intelligence, 2019. *Artificial intelligence*. [Ηλεκτρονικό]
Available at: <https://www.oecd.org/digital/artificial-intelligence/>
[Πρόσβαση 14 Νοεμβρίου 2021].

Crawford, K., 2013. *The Hidden Biases in Big Data*. [Ηλεκτρονικό]
Available at: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>
[Πρόσβαση 16 Απριλίου 2022].

Criado Perez, C., 2020. *Invisible Women: Exposing Data Bias in a World Designed for Men*. London: Vintage.

Curtis, B. & Savulescu, J., 2022. *Is Google's LaMDA conscious? A philosopher's view*. [Ηλεκτρονικό]
Available at: <https://theconversation.com/is-googles-lamda-conscious-a-philosophers-view-184987>
[Πρόσβαση 10 Ιουλίου 2022].

D' Ignazio, C. & Klein, L. F., 2020. *Data Feminism*. s.l.:The MIT Press.

D' Ignazio, C., Klein, L. F. & Agudelo Diaz, M., 2020. *The Data Feminism Infographic*. [Art].

Daly, A. et al., 2022. AI, Governance and Ethics Global Perspectives. In: H. Micklitz, et al. eds. *Constitutional Challenges in the Algorithmic Society*. Cambridge: Cambridge University Press, pp. 182-201.

Dastin, J., 2018. *Amazon scraps secret AI recruiting tool that showed bias against women*. [Ηλεκτρονικό] Available at: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

[Πρόσβαση 25 Ιανουαρίου 2022].

DATAWO, χ.χ. DATAWO. [Ηλεκτρονικό] Available at: <https://www.datawo.org/>

[Πρόσβαση 13 Απριλίου 2022].

Distributed AI Research Institute, χ.χ. DAIR. [Ηλεκτρονικό] Available at: <https://www.dair-institute.org/>

[Πρόσβαση 10 Μαΐου 2022].

EDRi, 2019. *The many faces of facial recognition in the EU*. [Ηλεκτρονικό] Available at: <https://edri.org/our-work/the-many-faces-of-facial-recognition-in-the-eu/>

[Πρόσβαση 14 Νοεμβρίου 2021].

EDRi, 2020. *Your face rings a bell: Three common uses of facial recognition*. [Ηλεκτρονικό] Available at: <https://edri.org/our-work/your-face-rings-a-bell-three-common-uses-of-facial-recognition/>

[Πρόσβαση 14 Νοεμβρίου 2021].

Egkolfopoulou, M. & Gardner, A., 2021. *Even in the Metaverse, Not All Identities Are Created Equal*. [Ηλεκτρονικό]

Available at: https://www.bloomberg.com/news/features/2021-12-06/cryptopunk-nft-prices-suggest-a-diversity-problem-in-the-metaverse?mc_cid=b71950dd71&mc_eid=86fddb0d40

[Πρόσβαση 29 Μαρτίου 2022].

Etzioni, A. & Etzioni, O., 2016. AI assisted ethics. *Ethics and Information Technology*, Volume 18, pp. 149-156.

European Commission for the Efficiency of Justice (CEPEJ), 2019. *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*. [Ηλεκτρονικό]
Available at: <https://www.coe.int/en/web/cepej/artificial-intelligence-in-justice-systems>
[Πρόσβαση 14 Νοεμβρίου 2021].

European Commission, Directorate-General for Education, Youth, Sport and Culture; Šucha, Vladimir; Gammel, Jean-Philippe, 2021. *Humans and Societies in the Age of Artificial Intelligence*, Luxembourg: Publications Office of the European Union.

European Commission, 2022. *Digital Services Act: Commission welcomes political agreement on rules ensuring a safe and accountable online environment*. [Ηλεκτρονικό]
Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2545
[Πρόσβαση 1 Μαΐου 2022].

European Parliament, Directorate-General for Internal Policies of the Union; Zacharia, Zacharias C.; Hovardas, Tasos; Xenofontos, Nikoletta; Pavlou, Ivoni; Irakleous, Maria, 2020. *Education and employment of women in science, technology and the digital economy, including AI and its influence on gender equality*, s.l.: European Parliament.

European Parliament, 2022. *Digital Services: landmark rules adopted for a safer, open online environment*. [Ηλεκτρονικό]
Available at: <https://www.europarl.europa.eu/news/en/press-room/20220701IPR34364/digital-services-landmark-rules-adopted-for-a-safer-open-online-environment>
[Πρόσβαση 10 Ιουλίου 2022].

European Union Agency for Fundamental Rights, 2018. *BigData: Discrimination in data-supported decision making*. [Ηλεκτρονικό]
Available at: <https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making>
[Πρόσβαση 25 Ιανουαρίου 2022].

European Union Agency for Fundamental Rights, 2019. *Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights*. [Ηλεκτρονικό]
Available at: <https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence->

mitigating-bias-and-error-protect

[Πρόσβαση 25 Ιανουαρίου 2022].

European Union Agency for Fundamental Rights, 2020. *Getting the future right – Artificial intelligence and fundamental rights*, Luxembourg: Publications Office of the European Union.

Ex Machina. 2014. [Film] Directed by Alex Garland. UK: A24/Universal Pictures/Film4.

Felzmann, H., Fosch-Villaronga, E., Lutz, C. & Tamò-Larrieux, A., 2020. Towards Transparency by Design for Artificial Intelligence. *Science and Engineering Ethics*, Volume 26, pp. 3333-3361.

Ferrer, X. et al., 2021. Bias and Discrimination in AI: A Cross-Disciplinary Perspective. *IEEE Technology and Society Magazine*, 40(2), pp. 72-80.

Floridi, L., Cowls, J., King, T. C. & Taddeo, M., 2020. How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics*, Volume 26, pp. 1771-1796.

Frenkel, S. & Browning, K., 2021. *The Metaverse's Dark Side: Here Come Harassment and Assaults*. [Ηλεκτρονικό]

Available at: <https://www.nytimes.com/2021/12/30/technology/metaverse-harassment-assaults.html>

[Πρόσβαση 29 Μαρτίου 2022].

Friedman, B. & Nissenbaum, H., 1996. Bias in Computer Systems. *ACM Transactions on Information Systems*, 14(3), pp. 330-347.

Fung, P., 2019. *This is why AI has a gender problem*. [Ηλεκτρονικό]

Available at: <https://www.weforum.org/agenda/2019/06/this-is-why-ai-has-a-gender-problem/>

[Πρόσβαση 26 Ιανουαρίου 2022].

G20 Trade Ministers and Digital Economy Ministers, 2019. *G20 Ministerial Statement on Trade and Digital Economy*. [Ηλεκτρονικό]

Available at: https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/en/documents/

[Πρόσβαση 14 Νοεμβρίου 2021].

Gartenberg, C., 2021. *Alexa finally gets a masculine-sounding voice option*. [Ηλεκτρονικό]

Available at: <https://www.theverge.com/2021/7/21/22587130/alexa-masculine-voice-gender-ziggy->

amazon-smart-assistant

[Πρόσβαση 6 Μαρτίου 2022].

Genova, J., 1994. Turing's sexual guessing game. *Social Epistemology*, 8(4), pp. 313-326.

Gibbs, S., 2013. *What is Boston Dynamics and why does Google want robots?*. [Ηλεκτρονικό]
Available at: <https://www.theguardian.com/technology/2013/dec/17/google-boston-dynamics-robots-atlas-bigdog-cheetah>

[Πρόσβαση 26 Ιανουαρίου 2022].

Girls Who Code, χ.χ. *Girls Who Code*. [Ηλεκτρονικό]
Available at: <https://girlswhocode.com/>

[Πρόσβαση 13 Απριλίου 2022].

Gómez-Upegui, S., 2021. *The Future of Digital Assistants Is Queer*. [Ηλεκτρονικό]
Available at: <https://www.wired.com/story/digital-assistant-smart-device-gender-identity/>

[Πρόσβαση 29 Μαρτίου 2022].

GPAI: Global Partnership on Artificial Intelligence, 2020. *GPAI The Global Partnership on Artificial Intelligence*. [Ηλεκτρονικό]

Available at: <https://gpai.ai/>

[Πρόσβαση 14 Νοεμβρίου 2021].

Halberstam, J., 1991. Automating Gender: Postmodern Feminism in the Age of the Intelligent Machine. *Feminist Studies*, 17(3), pp. 439-460.

Hamidi, F., Scheuerman, M. K. & Branham, S. M., 2018. *Gender Recognition or Gender Reductionism? The Social Implications of Automatic Gender Recognition*. Montreal, ACM.

Hanson Robotics, χ.χ. *Sophia*. [Ηλεκτρονικό]
Available at: <https://www.hansonrobotics.com/sophia/>

[Πρόσβαση 26 Ιανουαρίου 2022].

Hao, K., 2019. *This is how AI bias really happens—and why it's so hard to fix*. [Ηλεκτρονικό]
Available at: <https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happens-and-why-its-so-hard-to->

[fix/?utm_source=LinkedIn&utm_campaign=site_visitor.unpaid.engagement&utm_medium=tr social](https://www.linkedin.com/company/ai-ethics-center/)

[Πρόσβαση 25 Μαρτίου 2022].

Hao, K., 2020. *The US just released 10 principles that it hopes will make AI safer*. [Ηλεκτρονικό] Available at: <https://www.technologyreview.com/2020/01/07/130997/ai-regulatory-principles-us-white-house-american-ai-initiative/>

[Πρόσβαση 14 Νοεμβρίου 2021].

Haraway, D., χ.χ. *Το μανιφέστο των σάιμποργκ*. Αθήνα: Εκδόσεις Τοποβόρος.

Her. 2013. [Film] Directed by Spike Jonze. USA: Annapurna Pictures/Stage 6 Films.

High-Level Expert Group on Artificial Intelligence, 2019. *A definition of Artificial Intelligence: main capabilities and scientific disciplines*. [Ηλεκτρονικό]

Available at: <https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>

[Πρόσβαση 14 Νοεμβρίου 2021].

High-Level Expert Group on Artificial Intelligence, 2019. *Ethics guidelines for trustworthy AI*. [Ηλεκτρονικό]

Available at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

[Πρόσβαση 14 Νοεμβρίου 2021].

Hunt, E., 2016. *Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter*. [Ηλεκτρονικό]

Available at: <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>

[Πρόσβαση 20 Ιανουαρίου 2022].

I, Robot. 2004. [Film] Directed by Alex Proyas. USA/Germany: 20th Century Fox/Mediastream Vierte Film GmbH & Co Vermarktungs KG/Davis Entertainment.

IBM Cloud Education, 2020. *What is machine learning?*. [Ηλεκτρονικό]

Available at: <https://www.ibm.com/cloud/learn/machine-learning>

[Πρόσβαση 14 Νοεμβρίου 2021].

IBM Research Trusted AI, χ.χ. *AI Fairness* 360. [Ηλεκτρονικό]
Available at: https://aif360.mybluemix.net/?_ga=2.15242230.1377906663.1580739909-449412013.1563284276

[Πρόσβαση 31 Μαρτίου 2022].

IBM, χ.χ. *What is computer vision?*. [Ηλεκτρονικό]
Available at: <https://www.ibm.com/topics/computer-vision>

[Πρόσβαση 31 Μαρτίου 2022].

International Telecommunication Union, 2021. *Measuring digital development: Facts and Figures 2021*, Geneva: International Telecommunication Union.

Jotanovic, D., 2018. *This is how artificial intelligence is undoing women's rights*. [Ηλεκτρονικό]
Available at: <https://www.independent.co.uk/voices/artificial-intelligence-siri-cortana-alexa-music-pop-culture-robotics-a8406391.html>

[Πρόσβαση 26 Ιανουαρίου 2022].

Kalluri, P., 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, 583(169).

Keyes, O., 2018. *The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition*. New York, Association for Computing Machinery.

Kumar, S., 2021. *AI & the Digital Divide: Analyzing the Impact of Digital Connectivity on Human Development*. [Ηλεκτρονικό]

Available at: <https://medium.com/omdena/ai-the-digital-divide-analyzing-the-impact-of-digital-connectivity-on-human-development-6bb2bb814bc7>

[Πρόσβαση 14 Νοεμβρίου 2021].

Kyriakou, K., Barlas, P., Kleanthous, S. & Otterbacher, J., 2019. *Fairness in Proprietary Image Tagging Algorithms: A Cross-Platform Audit on People Images*. Munich, Association for the Advancement of Artificial.

Langston, J., 2015. *Who's a CEO? Google image results can shift gender biases*. [Ηλεκτρονικό]
Available at: <https://www.washington.edu/news/2015/04/09/whos-a-ceo-google-image-results-can>

[shift-gender-biases/](#)

[Πρόσβαση 20 Ιανουαρίου 2022].

Leslie, D., 2019. *Understanding artificial intelligence ethics and safety*. [Ηλεκτρονικό]
Available at: <https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety>

[Πρόσβαση 26 Μαρτίου 2022].

Leslie, D. και συν., 2021. *AI, human rights, democracy and the rule of law: A primer prepared for the Council of Europe*. [Ηλεκτρονικό]

Available at: <https://www.turing.ac.uk/research/publications/ai-human-rights-democracy-and-rule-law-primer-prepared-council-europe>

[Πρόσβαση 14 Νοεμβρίου 2021].

Loreck, J., 2016. *Explainer: what does the 'male gaze' mean, and what about a female gaze?.* [Ηλεκτρονικό]
Available at: <https://theconversation.com/explainer-what-does-the-male-gaze-mean-and-what-about-a-female-gaze-52486>

[Πρόσβαση 26 Ιανουαρίου 2022].

MacDonald, K., 2022. *I've seen the metaverse – and I don't want it*. [Ηλεκτρονικό]
Available at: <https://www.theguardian.com/games/2022/jan/25/ive-seen-the-metaverse-and-i-dont-want-it>

[Πρόσβαση 27 Ιανουαρίου 2022].

Madgavkar, A., 2021. *A conversation on artificial intelligence and gender bias*. [Ηλεκτρονικό]
Available at: <https://www.mckinsey.com/featured-insights/asia-pacific/a-conversation-on-artificial-intelligence-and-gender-bias>

[Πρόσβαση 22 Ιανουαρίου 2022].

Metropolis. 1927. [Film] Directed by Fritz Lang. Germany: Universum Film.

Microsoft Design, χ.χ. *Inclusive Design*. [Ηλεκτρονικό]

Available at: <https://www.microsoft.com/design/inclusive/>

[Πρόσβαση 31 Μαρτίου 2022].

Microsoft, χ.χ. *FATE: Fairness, Accountability, Transparency, and Ethics in AI*. [Ηλεκτρονικό]
Available at: <https://www.microsoft.com/en-us/research/theme/fate/>
[Πρόσβαση 29 Μαρτίου 2022].

Morris, C., 2022. *Sexual assault is already a problem in the metaverse, and a new report suggests it will get worse*. [Ηλεκτρονικό]
Available at: <https://www.fastcompany.com/90756868/report-metaverse-meta-virtual-sexual-assault-abuse>
[Πρόσβαση 10 Ιουλίου 2022].

Musik, C. & Bogner, A., 2019. Digitalization and Society: A sociology of technology perspective on current trends in data, digital security and the internet. *Österreichische Gesellschaft für Soziologie*, Volume 44, pp. 1-14.

Myers West, S., Whittaker, M. & Crawford, K., 2019. *Discriminating Systems: Gender, Race and Power in AI*, New York: AI Now Institute.

Niethammer, C., 2020. *AI Bias Could Put Women's Lives At Risk - A Challenge For Regulators*. [Ηλεκτρονικό]
Available at: <https://www.forbes.com/sites/carmenniethammer/2020/03/02/ai-bias-could-put-womens-lives-at-risk-a-challenge-for-regulators/?sh=229a5063534f>
[Πρόσβαση 22 Ιανουαρίου 2022].

Plant, S., 2016. *Zeros and Ones: Digital Women and the New Technoculture*. London: Fourth Estate.

Reichman, A. & Sartor, G., 2022. Algorithms and Regulation. In: H. Micklitz, et al. eds. *Constitutional Challenges in the Algorithmic Society*. Cambridge: Cambridge University Press, pp. 131-181.

Roberts, P., 2013. *At MIT Conference, Warnings of Big Data Fundamentalism*. [Ηλεκτρονικό]
Available at: <https://securityledger.com/2013/10/at-mit-conference-warnings-of-big-data-fundamentalism/>
[Πρόσβαση 10 Μαΐου 2022].

Rodrigues, R., 2020. Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, Volume 4, p. 100005.

- Roth, E., 2022. *Two members of Google's Ethical AI group leave to join Timnit Gebru's nonprofit*. [Ηλεκτρονικό]
Available at: https://www.theverge.com/2022/2/2/22915079/google-ethical-ai-group-departure-timnit-gebru?mc_cid=11078e1100&mc_eid=86fddb0d40
[Πρόσβαση 29 Μαρτίου 2022].
- Russel, S. & Norvig, P., 2021. *Artificial Intelligence: A Modern Approach*. 4th Edition ed. Harlow: Pearson Education.
- Salvini, S., 2014. *Encyclopedia of Quality of Life and Well-Being Research: Gender Discrimination*. Dordrecht: Springer.
- Sari, L. και συν., 2022. *A new resource to help measure fairness in AI speech recognition systems*. [Ηλεκτρονικό]
Available at: <https://ai.facebook.com/blog/a-new-resource-to-help-measure-fairness-in-ai-speech-recognition-systems/>
[Πρόσβαση 29 Μαρτίου 2022].
- Schwartz, R. και συν., 2022. *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. [Ηλεκτρονικό]
Available at: <https://www.nist.gov/publications/towards-standard-identifying-and-managing-bias-artificial-intelligence>
[Πρόσβαση 29 Μαρτίου 2022].
- Sejnoha, V., χ.χ. *Can We Build 'Her'? : What Samantha Tells Us About the Future of AI*. [Ηλεκτρονικό]
Available at: <https://www.wired.com/insights/2014/02/can-build-samantha-tells-us-future-ai/>
[Πρόσβαση 26 Ιανουαρίου 2022].
- Selbst, A. D. et al., 2019. *Fairness and Abstraction in Sociotechnical Systems*. Atlanta USA, Association for Computing Machinery.
- Silberg, J. & Manyika, J., 2019. *Tackling bias in artificial intelligence (and in humans)*. [Ηλεκτρονικό]
Available at: <https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans>
[Πρόσβαση 27 Μαρτίου 2022].

Simoncini, A. & Longo, E., 2022. Fundamental Rights and the Rule of Law in the Algorithmic Society. In: H. Micklitz, et al. eds. *Constitutional Challenges in the Algorithmic Society*. Cambridge: Cambridge University Press, pp. 27-41.

Simonite, T., 2021. *What Really Happened When Google Ousted Timnit Gebru*. [Ηλεκτρονικό] Available at: <https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/> [Πρόσβαση 27 Μαρτίου 2022].

Singh, S., 2021. *What we can learn from China's proposed AI regulations*. [Ηλεκτρονικό] Available at: <https://venturebeat.com/2021/10/03/what-we-can-learn-from-chinas-proposed-ai-regulations/> [Πρόσβαση 14 Νοεμβρίου 2021].

Smith, G. & Rustagi, I., 2021. *When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity*. [Ηλεκτρονικό] Available at: https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity [Πρόσβαση 26 Μαρτίου 2022].

Storey, J., 2015. *Πολιτισμική Θεωρία και Λαϊκή Κουλτούρα*. Αθήνα: Εκδόσεις Πλέθρον.

Sutko, D. M., 2020. Therorizing femininity in artificial intelligence: a framework for undoing technology's gender troubles. *Cultural Studies*, 34(4), pp. 567-592.

Tasioulas, J., 2019. First Steps Towards an Ethics of Robots and Artificial Intelligence. *Journal of Practical Ethics*, 7(1), pp. 49-83.

Taulli, T., 2020. *Turing Test At 70: Still Relevant For AI (Artificial Intelligence)?*. [Ηλεκτρονικό] Available at: <https://www.forbes.com/sites/tomtaulli/2020/11/27/turing-test-at-70-still-relevant-for-ai-artificial-intelligence/?sh=7d803c4d250f> [Πρόσβαση 23 Ιανουαρίου 2022].

Tenenbaum, C., 2020. Not Intelligent: Encoding Gender Bias. *Minnesota Journal of Law, Science & Technology*, 21(2), pp. 283-296.

The Institute of Technological Ethics, 2021. *Three kinds of bias in computer systems*. [Ηλεκτρονικό]
Available at: <https://www.technologicalethics.org/three-kinds-of-bias>
[Πρόσβαση 22 Ιανουαρίου 2022].

The Royal Society, 2018. *AI narratives: portrayals and perceptions of artificial intelligence and why they matter*. [Ηλεκτρονικό]
Available at: <https://royalsociety.org/topics-policy/projects/ai-narratives/>
[Πρόσβαση 26 Ιανουαρίου 2022].

The Terminator. 1984. [Film] Directed by James Cameron. United Kingdom; United States;: Cinema '84; Euro Film Funding; Hemdale;

Thompson, D., 2018. *Tech Was Supposed to Be Society's Great Equalizer. What Happened?*. [Ηλεκτρονικό]
Available at: <https://www.theatlantic.com/technology/archive/2018/09/tech-was-supposed-to-be-societys-great-equalizer-what-happened/571660/>
[Πρόσβαση 10 Μαΐου 2022].

Tiku, N., 2021. *Google fired its star AI researcher one year ago. Now she's launching her own institute*. [Ηλεκτρονικό]
Available at: https://www.washingtonpost.com/technology/2021/12/02/timnit-gebru-dair/?mc_cid=1e441f5787&mc_eid=86fddb0d40
[Πρόσβαση 29 Μαρτίου 2022].

Turing, A., 2020. Υπολογιστικά Μηχανήματα και Νοημοσύνη. Στο: Μ. Πατηνιώτης, επιμ. *Εισαγωγή στις Ψηφιακές Σπουδές*. Θεσσαλονίκη: Εκδόσεις Ροπή, pp. 51-82.

Tusk, B., 2022. *We need to put regulations on the metaverse now. Here's where to start*. [Ηλεκτρονικό]
Available at: https://www.fastcompany.com/90717243/we-need-to-put-regulations-on-the-metaverse-now-heres-where-to-start?mc_cid=0e6ad3503d&mc_eid=86fddb0d40
[Πρόσβαση 29 Μαρτίου 2022].

UNESCO; EQUALS Skills Coalition;, 2019. *I'd blush if I could: closing gender divides in digital skills through education*, s.l.: EQUALS and UNESCO.

- UNESCO's General Conference: 41st session, 2021. *Recommendation on the ethics of artificial intelligence*. [Ηλεκτρονικό]
Available at: <https://en.unesco.org/artificial-intelligence/ethics>
[Πρόσβαση 15 Ιανουαρίου 2022].
- UNESCO, 2017. *Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)*, Paris: UNESCO.
- UNESCO, 2020. *Artificial Intelligence and Gender Equality: Key Findings of UNESCO's Global Dialogue*, Paris: UNESCO.
- Van Cleynenbreugel, P., 2022. EU By-Design Regulation in the Algorithmic Society: A Promising Way Forward or Constitutional Nightmare in the Making?. In: H. Micklitz, et al. eds. *Constitutional Challenges in the Algorithmic Society*. Cambridge: Cambridge University Press, pp. 202-219.
- Veltri, N. F., Krasnova, H., Baumann, A. & Kalayamthanam, N., 2014. Gender Differences in Online Gaming: A Literature Review. *20th Americas Conference on Information Systems*.
- Wajcman, J., 2006. *Technofeminism*. Cambridge: Polity Press.
- Walker, T., 2020. "Alexa, are you a feminist?": Virtual Assistants Doing Gender and What That Means for the World. *The iJournal*, 6(1), pp. 1-16.
- Wallach, W. & Allen, C., 2010. *Moral Machines Teaching Robots Right from Wrong*. New York: Oxford University Press.
- Walsh, T. et al., 2019. *Closer to the Machine: Technical, Social and Legal Aspects of AI*. Melbourne: Office of the Victorian Information Commissioner.
- Watercutter, A., 2015. *Ex Machina Has a Serious Fembot Problem*. [Ηλεκτρονικό]
Available at: <https://www.wired.com/2015/04/ex-machina-turing-bechdel-test/>
[Πρόσβαση 26 Ιανουαρίου 2022].
- Wiewiórowski, W., 2020. *AI and Facial Recognition: Challenges and Opportunities*. [Ηλεκτρονικό]
Available at: https://edps.europa.eu/press-publications/press-news/blog/ai-and-facial-recognition-challenges-and-opportunities_en
[Πρόσβαση 14 Νοεμβρίου 2021].

Winkler, M. M., 2020. *Ovid on Screen: A Montage of Attractions*. Cambridge: Cambridge University Press.

Women in AI, χ.χ. *Women in AI*. [Ηλεκτρονικό]
Available at: <https://www.womeninai.co/>
[Πρόσβαση 13 Απριλίου 2022].

World Economic Forum, 2018. *The Global Gender Gap Report 2018*, Geneva: World Economic Forum.

World Economic Forum, 2021. *Global Gender Gap Report 2021*, Geneva: World Economic Forum.

Αρχή Προστασίας Δεδομένων, χ.χ. *Προστασία των δεδομένων ήδη από τον σχεδιασμό και εξορισμού*. [Ηλεκτρονικό]
Available at: <https://www.dpa.gr/el/foreis/genikes-upoxrewseis/prostasia-apo-sxediasmo>
[Πρόσβαση 25 Μαρτίου 2022].

Ευρωπαϊκή Ένωση, 2016. *Ενοποιημένη απόδοση της Συνθήκης για την Ευρωπαϊκή Ένωση και της Συνθήκης για τη λειτουργία της Ευρωπαϊκής Ένωσης*. [Ηλεκτρονικό]
Available at: <https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=celex%3A12016ME%2FTXT>
[Πρόσβαση 17 Ιανουαρίου 2022].

Ευρωπαϊκή Επιτροπή, 2018. *ΑΝΑΚΟΙΝΩΣΗ ΤΗΣ ΕΠΙΤΡΟΠΗΣ ΠΡΟΣ ΤΟ ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΟΒΟΥΛΙΟ, ΤΟ ΕΥΡΩΠΑΪΚΟ ΣΥΜΒΟΥΛΙΟ, ΤΟ ΣΥΜΒΟΥΛΙΟ, ΤΗΝ ΕΥΡΩΠΑΪΚΗ ΟΙΚΟΝΟΜΙΚΗ ΚΑΙ ΚΟΙΝΩΝΙΚΗ ΕΠΙΤΡΟΠΗ ΚΑΙ ΤΗΝ ΕΠΙΤΡΟΠΗ ΤΩΝ ΠΕΡΙΦΕΡΕΙΩΝ Τεχνητή νοημοσύνη για την Ευρώπη*. [Ηλεκτρονικό]
Available at: <https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=COM%3A2018%3A237%3AFIN>
[Πρόσβαση 14 Νοεμβρίου 2021].

Ευρωπαϊκή Επιτροπή, 2020. *Λευκή Βίβλος για την Τεχνητή Νοημοσύνη Η ευρωπαϊκή προσέγγιση της αριστείας και της εμπιστοσύνης*. [Ηλεκτρονικό]
Available at: https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_el
[Πρόσβαση 14 Νοεμβρίου 2021].

Ευρωπαϊκή Επιτροπή, 2021. *Μια Ευρώπη έτοιμη για την ψηφιακή εποχή: η Επιτροπή προτείνει νέους κανόνες και δράσεις για την αριστεία και την εμπιστοσύνη στην τεχνητή νοημοσύνη.* [Ηλεκτρονικό]

Available at: https://ec.europa.eu/commission/presscorner/detail/el/ip_21_1682
[Πρόσβαση 14 Νοεμβρίου 2021].

Ευρωπαϊκή Επιτροπή, 2021. *Πρόταση ΚΑΝΟΝΙΣΜΟΣ ΤΟΥ ΕΥΡΩΠΑΪΚΟΥ ΚΟΙΝΟΒΟΥΛΙΟΥ ΚΑΙ ΤΟΥ ΣΥΜΒΟΥΛΙΟΥ ΓΙΑ ΤΗ ΘΕΣΠΙΣΗ ΕΝΑΡΜΟΝΙΣΜΕΝΩΝ ΚΑΝΟΝΩΝ ΣΧΕΤΙΚΑ ΜΕ ΤΗΝ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ (ΠΡΑΞΗ ΓΙΑ ΤΗΝ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ) ΚΑΙ ΓΙΑ ΤΗΝ ΤΡΟΠΟΠΟΙΗΣΗ ΟΡΙΣΜΕΝΩΝ ΝΟΜΟΘΕΤΙΚΩΝ ΠΡΑΞΕΩΝ ΤΗΣ ΕΝΩΣΗΣ.* [Ηλεκτρονικό]

Available at: <https://eur-lex.europa.eu/legal-content/EL/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206>
[Πρόσβαση 14 Νοεμβρίου 2021].

Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο της Ευρωπαϊκής Ένωσης, 2015. *Οδηγία (ΕΕ) 2015/1535 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 9ης Σεπτεμβρίου 2015, για την καθιέρωση μιας διαδικασίας πληροφόρησης στον τομέα των τεχνικών προδιαγραφών και των κανόνων σχετικά με τις υπηρεσίες της κοινωνίας των πληροφοριών (Κε.* [Ηλεκτρονικό]

Available at: <https://eur-lex.europa.eu/legal-content/el/TXT/?uri=CELEX%3A32015L1535>
[Πρόσβαση 14 Νοεμβρίου 2021].

Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο, 2016. *Κανονισμός (ΕΕ) 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 27ης Απριλίου 2016, για την προστασία των φυσικών προσώπων έναντι της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα και για την ελεύθερη κυκλοφορία των δεδομένων αυτών και την.* [Ηλεκτρονικό]

Available at: <https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434>
[Πρόσβαση 14 Νοεμβρίου 2021].

Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο, 2019. *Κανονισμός (ΕΕ) 2019/1150 του Ευρωπαϊκού Κοινοβουλίου Και του Συμβουλίου, της 20ής Ιουνίου 2019, για την προώθηση της δίκαιης μεταχείρισης και της διαφάνειας για τους επιχειρηματικούς χρήστες επιγραμμικών υπηρεσιών διαμεσολάβησης.* [Ηλεκτρονικό]

Available at: <https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=celex%3A32019R1150>
[Πρόσβαση 14 Νοεμβρίου 2021].

Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο, 2020. *Πρόταση ΚΑΝΟΝΙΣΜΟΣ ΤΟΥ ΕΥΡΩΠΑΪΚΟΥ ΚΟΙΝΟΒΟΥΛΙΟΥ ΚΑΙ ΤΟΥ ΣΥΜΒΟΥΛΙΟΥ σχετικά με διεκδικήσιμες και δίκαιες αγορές στον ψηφιακό τομέα (πράξη για τις ψηφιακές αγορές).* [Ηλεκτρονικό]
Available at: <https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=COM%3A2020%3A842%3AFIN>
[Πρόσβαση 14 Νοεμβρίου 2021].

Ευρωπαϊκό Κοινοβούλιο, Συμβούλιο και Επιτροπή, 2016. *Χάρτης των Θεμελιωδών Δικαιωμάτων της Ευρωπαϊκής Ένωσης.* [Ηλεκτρονικό]
Available at: <https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=celex%3A12016P%2FTXT>
[Πρόσβαση 17 Ιανουαρίου 2022].

Ευρωπαϊκό Κοινοβούλιο, 2020. *Τι είναι η τεχνητή νοημοσύνη και πώς χρησιμοποιείται.* [Ηλεκτρονικό]
Available at: <https://www.europarl.europa.eu/news/el/headlines/priorities/i-techniti-noimosuni-stin-ee/20200827STO85804/ti-einai-i-techniti-noimosuni-kai-pos-chrisimopoeitai>
[Πρόσβαση 14 Νοεμβρίου 2021].

Ευρωπαϊκό Κοινοβούλιο, 2020. *Ψήφισμα του Ευρωπαϊκού Κοινοβουλίου της 20ής Οκτωβρίου 2020 με συστάσεις προς την Επιτροπή όσον αφορά το καθεστώς αστικής ευθύνης για την τεχνητή νοημοσύνη (2020/2014(INL)).* [Ηλεκτρονικό]
Available at: <https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=CELEX%3A52020IP0276>
[Πρόσβαση 14 Νοεμβρίου 2021].

Καραγιάννη, Α., 2022. *Προσωπικό Αρχείο: Συνέντευξη με DATAWO [Συνέντευξη] (25 Απριλίου 2022).*

Μαρξ, Κ., 2006. *Το Κεφάλαιο - Επίτομο.* Αθήνα: Εκδόσεις Μαρξιστικό Αρχείο.

Όμηρος, 2014. *Ιλιάδα.* Αθήνα: Ars Brevis.

Σαράλα, Ν., 2020. Από την τεχνητή νοημοσύνη στη συνείδηση (ή μήπως το αντίστροφο;). Στο: Μ. Πατηνιώτης, επιμ. *Εισαγωγή στις Ψηφιακές Σπουδές.* Θεσσαλονίκη: Εκδόσεις Ροπή, pp. 163-170.

Συμβούλιο της Ευρώπης, 1950. *Ευρωπαϊκή Σύμβαση Δικαιωμάτων του Ανθρώπου*. [Ηλεκτρονικό]
Available at: https://www.echr.coe.int/documents/convention_ell.pdf
[Πρόσβαση 17 Ιανουαρίου 2022].