ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΤΜΗΜΑ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

ΤΜΗΜΑ ΝΟΜΙΚΗΣ

ΔΙΠΔΡΥΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΔΙΚΑΙΟ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗ

Η ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΚΑΤΑ ΤΗ ΔΙΑΔΙΚΑΣΙΑ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ ΑΠΟ ΤΑ ΠΟΛΙΤΙΚΑ ΚΑΙ ΠΟΙΝΙΚΑ ΔΙΚΑΣΤΗΡΙΑ ΚΑΙ Η ΣΧΕΣΗ ΤΩΝ ΖΗΤΗΜΑΤΩΝ ΠΟΥ ΑΝΑΦΥΟΝΤΑΙ ΜΕ ΤΟΝ ΓΕΝΙΚΟ ΚΑΝΟΝΙΣΜΟ ΓΙΑ ΤΗΝ ΠΡΟΣΤΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΤΟΥ ΕΥΡΩΠΑΪΚΟΥ ΚΟΙΝΟΒΟΥΛΙΟΥ ΚΑΙ ΤΟΥ ΣΥΜΒΟΥΛΙΟΥ ΤΗΣ 27ης ΑΠΡΙΛΙΟΥ 2016

Διπλωματική Εργασία

της

Ελένης Β. Σταμπουλίδου

Θεσσαλονίκη, Μάιος 2023

Η ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΚΑΤΑ ΤΗ ΔΙΑΔΙΚΑΣΙΑ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ ΑΠΟ ΤΑ ΠΟΛΙΤΙΚΑ ΚΑΙ ΠΟΙΝΙΚΑ ΔΙΚΑΣΤΗΡΙΑ ΚΑΙ Η ΣΧΕΣΗ ΤΩΝ ΖΗΤΗΜΑΤΩΝ ΠΟΥ ΑΝΑΦΥΟΝΤΑΙ ΜΕ ΤΟΝ ΓΕΝΙΚΟ ΚΑΝΟΝΙΣΜΟ ΓΙΑ ΤΗΝ ΠΡΟΣΤΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΤΟΥ ΕΥΡΩΠΑΪΚΟΥ ΚΟΙΝΟΒΟΥΛΙΟΥ ΚΑΙ ΤΟΥ ΣΥΜΒΟΥΛΙΟΥ ΤΗΣ 27ης ΑΠΡΙΛΙΟΥ 2016

Ελένη Β. Σταμπουλίδου

Πτυχίο Νομικής ΑΠΘ, 1993

Διπλωματική Εργασία

υποβαλλόμενη για τη μερική εκπλήρωση των απαιτήσεων του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΤΙΤΛΟΥ ΣΠΟΥΔΩΝ ΣΤΟ ΔΙΚΑΙΟ & ΠΛΗΡΟΦΟΡΙΚΗ

Επιβλέπων Καθηγητής:
Ευθύμιος Ταμπούρης

Επιβλέπουσα Καθηγήτρια:
Μαρία Μυλώση

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την

Ευθύμιος Ταμπούρης	Μαρία Μυλώση	Ευγενία Αλεξανδροπούλου-Αιγυπτιάδου

Ελένη Β. Σταμπουλίδου

Περίληψη

Η χρήση συστημάτων τεχνητής νοημοσύνης θα επιφέρει μεγάλες αλλαγές στην απονομή της πολιτικής και ποινικής δικαιοσύνης, προσφέροντας δυνατότητες οι οποίες δεν συναντούν αντιδράσεις, μέχρι το σημείο που αγγίζουν τον πυρήνα της πολιτικής και ποινικής δίκης, τη δικαιοδοτική δηλαδή κρίση, η οποία πρέπει να είναι έργο ανεξάρτητου και αμερόληπτου δικαστή. Οι διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης, στις οποίες μετέχουν και στάδια πριν την τελική απόφαση του δικαστηρίου, καθώς αυτά επηρεάζουν σημαντικά αυτή, εγείρουν ζητήματα προκατάληψης, αδιαφάνειας, συμβατότητας με τον Γενικό Κανονισμό για την Προστασία Δεδομένων του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016 και έλλειψης λογοδοσίας. Η οποιαδήποτε δε πρόταση για επίλυση ή άμβλυση αυτών των ζητημάτων πρέπει να λαμβάνει υπόψη τον ιδιαίτερα ευαίσθητο τομέα της απονομής της δικαιοσύνης και να είναι σύμφωνη με τις αρχές του κράτους δικαίου και τα δικαιώματα του ανθρώπου, όπως αυτά κατοχυρώνονται στην ΕΣΔΑ. Μέσα από την ανάλυση αυτών των ζητημάτων θα σκιαγραφηθούν τα χαρακτηριστικά της διαδικασίας λήψης αποφάσεως η οποία λαμβάνει χώρα με τη χρήση τεχνητής νοημοσύνης στον τομέα της πολιτικής και ποινικής δικαιοσύνης και θα διαφανεί πως αυτή μπορεί να μεταμορφώσει τις σχετικές διαδικασίες και τον ρόλο του δικαστή – εισαγγελέα, καθώς επίσης και υπό ποιες προϋποθέσεις μπορεί να γίνει αποδεκτή η εισαγωγή της σ’ αυτά τα συστήματα. Η σχετική ανάλυση θα πραγματοποιηθεί με τη μέθοδο της συστηματικής βιβλιογραφικής επισκόπησης, ενώ η αναζήτηση ζητημάτων που δεν θα ανευρεθούν με τη μέθοδο αυτή στην επισκοπούμενη βιβλιογραφία και θα κριθεί απαραίτητη η αναφορά τους στην εργασία, θα γίνει περαιτέρω στο διαδίκτυο, με την έρευνα και σε ελληνικές πηγές. Μέσα από αυτήν την ανάλυση, θα διαφανεί ότι οι διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης μπορούν να νοηθούν μόνο ως βοηθητικές και συμπληρωματικές της ανθρώπινης λήψης αποφάσεων, μη δυνάμενες να αντικαταστήσουν τον άνθρωπο δικαστή – εισαγγελέα, ο οποίος δύναται να ελέγχει την ορθότητα των αποτελεσμάτων των εν λόγω διαδικασιών, ενώ με κατάλληλη εκπαίδευση και σε συνεργασία με τον πληροφορικό μπορεί να αλληλοεπιδρά με τα συστήματα τεχνητής νοημοσύνης με τον επωφελέστερο κατά περίπτωση τρόπο, ζητώντας, κατά περίπτωση, τις απαραίτητες επεξηγήσεις για τις αποφάσεις που αυτή λαμβάνει.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τους διδάσκοντες καθηγητές μου για τις αμέτρητες και πολύτιμες γνώσεις που μου προσέφεραν, όλες απαραίτητες για την κατανόηση του σύγχρονου, πολύπλοκου, γεμάτου νέες προκλήσεις κόσμου και για να ελπίζω ότι ίσως «σε έναν αναβαθμισμένο κόσμο» δεν θα νιώθω «σαν κυνηγός Νεάντερταλ στη Γουόλ Στριτ» (Harari, Homo Deus).

Θα ήθελα να ευχαριστήσω ειδικά τον επιβλέποντα καθηγητή μου κ. Ευθύμιο Ταμπούρη και την συνεπιβλέπουσα κ. Μαρία Μυλώση για την καθοδήγηση, ακούραστη βοήθειά τους και τις σημαντικές συμβουλές που μου έδωσαν για την ολοκλήρωση της εργασίας, ανοίγοντάς μου συγχρόνως διάπλατα την πόρτα για την είσοδό μου στον πολύπλοκο σύγχρονο κόσμο και γεμίζοντάς με με εφόδια και εργαλεία για να μπορώ να τον κατανοήσω.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς και τον γιο μου για τη συμπαράστασή τους σε όλη αυτήν τη διαδρομή, από την αρχή του μεταπτυχιακού μέχρι την ολοκλήρωση της εργασίας. Χωρίς αυτούς θα ήταν αδύνατο να ανταποκριθώ στις υποχρεώσεις μου.

Περιεχόμενα

1. Εισαγωγή

1.1. Πρόβλημα – Σημαντικότητα του θέματος

Η εισαγωγή συστημάτων τεχνητής νοημοσύνης στη δικαιοσύνη έχει αναμφισβήτητα σημαντικές και πολλαπλές συνέπειες. Οι συνέπειες αυτές πρέπει να εξεταστούν επισταμένως πριν τα συστήματα αυτά εισαχθούν στον ευαίσθητο αυτό τομέα. Υποστηρίζεται, καταρχήν, ότι η τεχνητή νοημοσύνη στη δικαιοσύνη δύναται να συμβάλλει στη μείωση των εκκρεμών υποθέσεων και στον εκσυγχρονισμό των διαδικασιών απονομής της. Από την άλλη, ωστόσο, μεριά, τονίζονται οι επιπτώσεις που μπορεί να έχει σε θεμελιώδη ζητήματα, τα οποία συνιστούν τον πυρήνα του κράτους δικαίου. Η σημασία των επιπτώσεων αυτών καταδεικνύεται όταν η χρήση της τεχνητής νοημοσύνης αφορά στην άσκηση του δικαιοδοτικού έργου του δικαστή και εισαγγελέα, στη διατύπωση δηλαδή της δικανικής κρίσης και στην έκδοση δικαστικής απόφασης, εισαγγελικής διάταξης, βουλεύματος κ.λ.π.. Κατά την άσκηση αυτού του έργου, τα ζητήματα που τίθενται και συνιστούν το σύστημα του πυρήνα του κράτους δικαίου χρήζουν ιδιαίτερης προσοχής και συνεχούς εγρήγορσης προκειμένου να αποκλειστεί η προσβολή του. Συνεπώς, η εισαγωγή των εν λόγω συστημάτων στον συγκεκριμένο τομέα προϋποθέτει τη διασφάλιση αυτής της παραμέτρου· πρέπει, δηλαδή, να διασφαλιστεί και να διασφαλίζεται, συνεχώς, στην περίπτωση που τέτοια συστήματα χρησιμοποιηθούν κατά την άσκηση του δικαιοδοτικού έργου, ότι οποιαδήποτε χρήση αυτών γίνεται, εφόσον δεν παραβιάζονται οι διάφορες συνιστώσες του κράτους δικαίου, η ανεξαρτησία της δικαιοσύνης, η δίκαιη δίκη, η προστασία των προσωπικών δεδομένων, η ασφάλεια του δικαίου, η ισότητα των διαδίκων, η ισότητα των όπλων, το δικαίωμα πρόσβασης στη δικαιοσύνη.

Μέσα από τις διασφαλίσεις και το πρίσμα του κράτους δικαίου, πρέπει εξάλλου να διερευνηθεί, αν η εν λόγω τεχνολογία είναι δίκαιη, λαμβανομένου υπόψη ότι τα δεδομένα στα οποία στηρίζεται περιέχουν προκαταλήψεις ή παρουσιάζουν άλλες αδυναμίες, αν οι διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης είναι διαφανείς, αν στην περίπτωση αδιαφάνειας αυτών θα μπορούσαν αυτές να εισαχθούν στον τομέα της δικαιοσύνης· επίσης, πρέπει να διερευνηθεί το ζήτημα του ελέγχου των συστημάτων και της ευθύνης από τη χρήση τέτοιων διαδικασιών.

Εξάλλου, όσον αφορά στην άσκηση του δικαιοδοτικού έργου του δικαστή ή εισαγγελέα, η χρήση της τεχνητής νοημοσύνης, ως διαδικασία λήψης αποφάσεων, βρίσκει σοβαρές ενστάσεις, οι οποίες δικαιολογούνται τουλάχιστον όσον αυτές οι διαδικασίες παραμένουν αδιαφανείς και όσον τίθεται υπό αμφισβήτηση το ποιος είναι ο αρμόδιος και υπεύθυνος για τη λήψη των αποφάσεων, δηλαδή εφόσον αμφισβητείται ότι ο τελευταίος πάντα κριτής, αρμόδιος για τη λήψη των αποφάσεων και υπεύθυνος γι’ αυτές θα είναι ο άνθρωπος δικαστής και εισαγγελέας.

1.2. Σκοπός- Στόχοι

Ο στόχος της εργασίας είναι να καταγραφούν και να αναλυθούν τα ζητήματα που θέτει η χρήση τεχνητής νοημοσύνης κατά τη διαδικασία λήψης αποφάσεως στην ποινική και πολιτική δικαιοσύνη. Με αυτόν τον τρόπο, θα επιχειρηθεί να διαφανεί υπό ποιες προϋποθέσεις θα μπορούσε να γίνει δεκτή η εισαγωγή της τεχνητής νοημοσύνης κατά τη δικαιοδοτική κρίση των δικαστών – εισαγγελέων. Οι προϋποθέσεις αυτές περιορίζονται σε ζητήματα δικαιοσύνης, διαφάνειας και λογοδοσίας που αφορούν στις σχετικές διαδικασίες. Άλλα σημαντικά ζητήματα, όπως η ασφάλεια των διαδικασιών αυτών, των σχετικών πληροφοριών και δεδομένων που χρησιμοποιούνται, οι παραβιάσεις του απορρήτου και οι απειλές για την ιδιωτική ζωή δεν θα ερευνηθούν.

Επίσης, μέσα από την ανάλυση αυτών των ζητημάτων (δικαιοσύνης, διαφάνειας, λογοδοσίας) επιχειρείται να αναδειχθεί η μορφή και οι δυνατότητες που μπορεί να δώσει η τεχνητή νοημοσύνη στην πολιτική και ποινική δικαιοσύνη, κατά την άσκηση του δικαιοδοτικού έργου των δικαστών και εισαγγελέων, κατά τρόπο ώστε να μπορούν να αναδειχθούν τα υπέρ και τα κατά από την εισαγωγή της σ' αυτά τα περιβάλλοντα, καθώς επίσης και ο ρόλος που θα κληθεί να έχει ο δικαστής – εισαγγελέας σε μια διαδικασία λήψης αποφάσεων που λαμβάνει χώρα με τη χρήση τεχνητής νοημοσύνης. Σημαντικό, εξάλλου, είναι να επισημανθεί ότι άρθρα – μελέτες που αφορούν καθαρά τεχνικά θέματα/αναλύσεις αλγορίθμων χρησιμοποιούνται μόνο στο μέτρο που απαιτείται για να δοθεί μία εικόνα της μορφής που μπορεί να δώσει η τεχνητή νοημοσύνη στην ποινική και πολιτική δικαιοσύνη και κατά το μέτρο που είναι αναγκαίο για να γίνουν κατανοητά τα ζητήματα που τίθενται με την εισαγωγή της στους κλάδους αυτούς της δικαιοσύνης.

Τέλος, μέσα από την ανάλυση των ζητημάτων αυτών θα καταγραφεί και ο ρόλος που θα κληθεί να έχει η νομοθεσία για τα προσωπικά δεδομένα, ειδικότερα δε ο Γενικός Κανονισμός 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016, ως μέσο απάντησης στις ενστάσεις που εγείρει η χρήση της τεχνητής νοημοσύνης κατά το δικαιοδοτικό έργο των δικαστών- εισαγγελέων.

1.3. Ερωτήματα – Υποθέσεις

Έχοντας το ως άνω αντικείμενο και εύρος, η παρούσα εργασία θα επιχειρήσει να απαντήσει στο ερώτημα αν οι διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης μπορούν να αντικαταστήσουν τον άνθρωπο δικαστή-εισαγγελέα. Με δεδομένο ότι αυτές οι διαδικασίες με το πέρασμα του χρόνου και τις τεχνολογικές εξελίξεις θα βελτιώνονται συνεχώς και θα γίνονται όλο και πιο ακριβείς, αλλά και ότι οι μη ειδικοί, όπως είναι και οι δικαστές, θα αντιμετωπίζουν δυσκολίες κατανόησης των σχετικών διαδικασιών, είναι σημαντικό να αναρωτηθούμε μήπως πρέπει να προτιμάται η αντικατάσταση των ανθρώπων δικαστών-εισαγγελέων από τους αλγόριθμους τεχνητής νοημοσύνης, όταν αυτοί οι αλγόριθμοι θα μπορούν να είναι πιο παραγωγικοί και πιο ακριβείς από τους ανθρώπους.

1.4. Συνεισφορές

Όπως θα διαφανεί κατά την ανάλυση της εργασίας, τα ζητήματα της δικαιοσύνης, της διαφάνειας και της λογοδοσίας κατά τις διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης στον τομέα της δικαιοσύνης απασχολούν σε μεγάλο εύρος τους συγγραφείς, καταδεικνύοντας τη σπουδαιότητα του θέματος. Πολλά, επίσης, ζητήματα που σχετίζονται με τις εν λόγω διαδικασίες (όπως η ασφάλεια των διαδικασιών αυτών, των σχετικών πληροφοριών και δεδομένων), έχουν απασχολήσει τη βιβλιογραφία. Αυτό, ωστόσο, που όπως διαφαίνεται δεν έχει απασχολήσει όσο θα έπρεπε τους μελετητές είναι η δυνατότητα του ανθρώπου να αλληλοεπιδράσει με τα συστήματα λήψης αποφάσεων τεχνητής νοημοσύνης, η εκπαίδευσή του με σκοπό να κατανοεί όσο είναι δυνατόν τα εν λόγω συστήματα, ο σχεδιασμός συστημάτων προσφιλών στον χρήστη-δικαστή και εισαγγελέα, η μεγάλη τεχνογνωσία των τελευταίων, οι οποίοι παίρνοντας ένα αλγοριθμικό αποτέλεσμα, σε υπόθεση αρμοδιότητάς τους, είναι σε θέση να αντιληφθούν αν αυτό είναι ή πλησιάζει να είναι ορθό και τελικά να ελέγξουν την ορθότητά του. Σε τελική ανάλυση, ακόμα και αν ο ειδικός της τεχνητής νοημοσύνης δεν είναι σε θέση να εξηγήσει με ποιο τρόπο ο αλγόριθμος κατέληξε σε ένα αποτέλεσμα, ο δικαστής και ο εισαγγελέας μπορεί να ελέγξει την ορθότητά του.

1.5. Διάρθρωση Μελέτης

Το παρόν Κεφάλαιο αποτελεί μια εισαγωγή στο θέμα της εργασίας και παρουσίαση της σπουδαιότητας του αντικειμένου της, θέτει το κεντρικό ερώτημα που προκύπτει από την ανάλυση των ζητημάτων που αναφύονται κατά τη μελέτη της βιβλιογραφίας και επιχειρεί να εντάξει στη συζήτηση την περαιτέρω συνεισφορά, όπως εκφράστηκε ανωτέρω. Στο κεφάλαιο 2 γίνεται καταγραφή και επεξήγηση των όρων που αφορούν στο αντικείμενο της εργασίας, κατά τρόπο ώστε να γίνει κατανοητή τόσο η ανάλυση των θεμάτων αυτής όσο και το εύρος της. Στο Κεφάλαιο 3 παρουσιάζεται η μεθοδολογία για τη συλλογή και επιλογή των άρθρων και την ανάλυση των ζητημάτων που αναδείχθηκαν από την μελέτη αυτών. Στο Κεφάλαιο 4 γίνεται η κύρια ανάλυση του θέματος με αναφορές στα ζητήματα που απασχόλησαν την επισκοπούμενη βιβλιογραφία αναφορικά με τη χρήση τεχνητής νοημοσύνης κατά τη διαδικασία λήψης αποφάσεως στον τομέα της ποινικής και πολιτικής δικαιοσύνης, με καταγραφή και των αναφορών στον Κανονισμό 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016 και των επιμέρους ζητημάτων που αυτός θέτει ως προς τα ζητήματα που απασχόλησαν την επισκοπούμενη βιβλιογραφία. Στο Κεφάλαιο 5 παρουσιάζονται τα συμπεράσματα της εργασίας καθώς και οι μελλοντικές επεκτάσεις αυτής.

2. Θεωρητικό Υπόβαθρο

2.1. Εισαγωγή.

Σημαντική στο σημείο αυτό είναι η επεξήγηση του όρου τεχνητή νοημοσύνη (ΤΝ), ποινική και πολιτική δικαιοσύνη και προσωπικά δεδομένα, όχι ως έννοιες καθεαυτές αλλά προκειμένου να γίνει κατανοητό το αντικείμενο και οι στόχοι της εργασίας, καθώς και το εύρος της. Πρέπει, μάλιστα, να διασαφηνισθεί ότι η καλή κατανόηση των τεχνικών πτυχών της τεχνητής νοημοσύνης είναι απαραίτητη για την πλήρη αξιολόγηση των συνεπειών της στη λήψη νομικών αποφάσεων (Scherer, 2019). Επίσης, σημαντικό είναι να αποσαφηνιστεί ο όρος "διαδικασία λήψης αποφάσεως τεχνητής νοημοσύνης" και να επισημανθεί ότι στη βιβλιογραφία απαντώνται παρεμφερείς ορισμοί όπως αλγοριθμική λήψη αποφάσεων, αυτοματοποιημένη λήψη αποφάσεων. Επιπλέον, στο πεδίο της εργασίας απαντώνται οι έννοιες μηχανική μάθηση, αλγόριθμοι, μεγάλα δεδομένα κ.α. Όλοι αυτοί οι όροι είναι σχετικοί και αναγκαίοι στο πεδίο της ΤΝ χωρίς όμως να ταυτίζονται με αυτή.

2.2. Τεχνητή νοημοσύνη και σχετικοί στο πεδίο της ΤΝ όροι

Ειδικότερα, ενώ η Τεχνητή Νοημοσύνη (ΤΝ) μπορεί να θεωρηθεί ως ένα πεδίο έρευνας που στοχεύει στη μίμηση των ανθρώπινων ικανοτήτων, η Μηχανική Μάθηση (ΜΜ) είναι ένα συγκεκριμένο υποσύνολο της τεχνητής νοημοσύνης που εκπαιδεύει μια μηχανή πώς να μαθαίνει (Margagliotti and Bollé, 2019) και να βελτιώνεται από την εμπειρία χωρίς να είναι ρητά προγραμματισμένη, σε γενικές δε γραμμές, (η μηχανική μάθηση) αφορά τη χρήση αλγορίθμων για μάθηση από δεδομένα εκπαίδευσης τόσο με «εποπτευόμενο» όσο και με «μη εποπτευόμενο» (αυτοοργανωμένο τρόπο) (Greenstein, no date). Οι ερευνητές της τεχνητής νοημοσύνης διακρίνουν διάφορους τύπους μηχανικής μάθησης, ανάλογα με τον βαθμό της ανθρώπινης συμβολής. Η εποπτευόμενη μάθηση απαιτεί ανθρώπινη αλληλεπίδραση: ο προγραμματιστής εκπαιδεύει το πρόγραμμα ορίζοντας ένα σύνολο επιθυμητών αποτελεσμάτων (π.χ. ταξινόμηση σε ανεπιθύμητη/μη ανεπιθύμητη αλληλογραφία) για ένα εύρος εισροών. Αυτό σημαίνει ότι τα δεδομένα του συνόλου εκπαίδευσης πρέπει να επισημαίνονται επαρκώς (π.χ. μηνύματα ηλεκτρονικού ταχυδρομείου που αναγνωρίζονται ως ανεπιθύμητα ή όχι) και απαιτείται κάποια μορφή ανθρώπινης ανατροφοδότησης (π.χ. όταν το πρόγραμμα ταξινομεί εσφαλμένα ένα μήνυμα ηλεκτρονικού ταχυδρομείου). Αντίθετα, η μάθηση χωρίς επίβλεψη δεν απαιτεί, ή σχεδόν δεν απαιτεί, καμία ανθρώπινη παρέμβαση. Δεν υπάρχουν προκαθορισμένες παραδοχές ή προκαθορισμένα αποτελέσματα. Μάλλον, το πρόγραμμα εντοπίζει συνυπάρχοντα χαρακτηριστικά που θα δημιουργήσουν την προσδοκία ότι θα συνυπάρχουν στο μέλλον. Αυτό συμβαίνει, για παράδειγμα, με πολλά προγράμματα μετάφρασης σύγχρονων γλωσσών. Είναι σημαντικό ότι δεν υπάρχει ένα μόνο σύστημα τεχνητής νοημοσύνης, αλλά μια ποικιλία διαφορετικών μοντέλων (Scherer, 2019). Εξάλλου, ένα τύπος μηχανικής μάθησης είναι και η βαθιά μάθηση. Η βαθιά εκμάθηση παρουσιάζει υψηλή απόδοση σε πολλές εφαρμογές όπως η ανάλυση ιατρικής εικόνας, οι έξυπνες εφαρμογές IoT, η

ανάλυση μεγάλων δεδομένων και η αναγνώριση φωνής (Al_Azrak *et al.*, 2020). Γενικά, μπορούμε να φανταστούμε την ΤΝ ως την ομπρέλα που περιλαμβάνει τις υπόλοιπες τεχνολογίες. Περαιτέρω, θεωρούμε ως «αλγόριθμους» εκείνα τα ανθρώπινα τεχνουργήματα που προκύπτουν από την εκπαίδευση μοντέλων μηχανικής μάθησης σε ψηφιακά δεδομένα, προκειμένου να δημιουργηθούν προβλέψεις για να βοηθήσουν ή να αυτοματοποιήσουν τη λήψη αποφάσεων(Loi, Ferrario and Viganò, 2020). Τα συστήματα και οι αλγόριθμοι μηχανικής μάθησης, η κινητήρια δύναμη πίσω από πολλές εξελίξεις της τεχνητής νοημοσύνης, είναι πολύτιμα λόγω της ικανότητάς τους να μαθαίνουν μόνοι τους «πώς να εντοπίζουν χρήσιμα μοτίβα σε τεράστια σύνολα δεδομένων και να συνδυάζουν πληροφορίες με τρόπους που αποδίδουν εξαιρετικά ακριβείς προβλέψεις ή εκτιμήσεις». Πολλά συστήματα μηχανικής μάθησης εκπαιδεύονται σε μεγάλους όγκους δεδομένων και προσαρμόζουν τις δικές τους παραμέτρους για να βελτιώσουν την αξιοπιστία των προβλέψεών τους με την πάροδο του χρόνου. Τα εργαλεία μηχανικής μάθησης παρέχουν τη δυνατότητα λήψης πιο ακριβών αποφάσεων, πιο γρήγορα, με βάση πολύ μεγαλύτερες ποσότητες δεδομένων από ότι οι άνθρωποι μπορούν να επεξεργάζονται και να χειρίζονται (Deeks, 2019). Ο Greenstein (Greenstein, no date) αναφερόμενος στις σχετικές τεχνολογίες επισημαίνει ότι στην καρδιά της μηχανικής μάθησης βρίσκεται ο μαθηματικός αλγόριθμος, ο οποίος μπορεί να περιγραφεί ως «[μια] διαδικασία ή σύνολο κανόνων που πρέπει να ακολουθούνται σε υπολογισμούς ή άλλες λειτουργίες επίλυσης προβλημάτων, ειδικά από έναν υπολογιστή». Ενώ η λογική ενός κανονικού συστήματος υπολογιστή δημιουργείται από έναν άνθρωπο προγραμματιστή, η λογική ενός συστήματος που χρησιμοποιεί μηχανική μάθηση δημιουργείται από έναν αλγόριθμο. Η μηχανική μάθηση είναι ουσιαστικά η εφαρμογή μαθηματικών αλγορίθμων σε δεδομένα για την παραγωγή ενός μοντέλου που μπορεί να ενσωματωθεί στα συστήματα λήψης αποφάσεων, το οποίο (μοντέλο) είναι αυτόνομο στο βαθμό που μπορεί να ενημερώνεται με βάση νέα δεδομένα (Greenstein, no date). Αν και υπάρχουν σημαντικές διαφορές μεταξύ των εν λόγω τεχνολογιών, γενικά, αυτό που είναι κοινό σε αυτές είναι ότι προτείνουν χειρισμούς (αστυνομικές παρεμβάσεις, αποφάσεις εγγύησης, ποινές) με βάση μαθηματικά δομημένη επεξεργασία, συνήθως σε εκτεταμένα σύνολα δεδομένων (Chiao, 2019). Ο Greenstein (Greenstein, no date) υπογραμμίζει ότι η τεχνητή νοημοσύνη δεν αφορά μόνο την τεχνολογία — μάλλον ενσωματώνει πολλαπλούς κλάδους στην προσπάθεια δημιουργίας μηχανών που σκέφτονται σαν άνθρωποι. Είναι ένας ακαδημαϊκός κλάδος που καλύπτει πολλά θέματα: φιλοσοφία, μαθηματικά, οικονομικά, νευροεπιστήμες, ψυχολογία, μηχανική υπολογιστών, θεωρία ελέγχου και κυβερνητική και γλωσσολογία. Πιο συγκεκριμένα, περιλαμβάνει θέματα όπως αναπαράσταση γνώσης, ευρετική αναζήτηση, προγραμματισμό, έμπειρα συστήματα, μηχανική όραση, μηχανική μάθηση, επεξεργασία φυσικής γλώσσας, πράκτορες λογισμικού, έξυπνα συστήματα διδασκαλίας και ρομποτική. Ένας πιο επίσημος ορισμός περιγράφει την τεχνητή νοημοσύνη ως: [α] διεπιστημονική προσέγγιση για την κατανόηση, τη μοντελοποίηση και την αναπαραγωγή νοημοσύνης και γνωστικών διαδικασιών με επίκληση διάφορες υπολογιστικές, μαθηματικές, λογικές, μηχανικές, ακόμα και βιολογικές αρχές και συσκευές […] (Greenstein, no date). Όπως δε είναι φανερό σημαντική είναι, για την κατανόηση του ως άνω ορισμού και η έννοια του

«μοντέλου». Τα μοντέλα είναι το βασικό τεχνικό στοιχείο ενός συστήματος που χρησιμοποιείται για τη λήψη αποφάσεων, τα οποία έχουν εξοπλιστεί με τις γνώσεις για τα δεδομένα που αποκτώνται από έναν αλγόριθμο. Τα μοντέλα μπορούν επίσης να έχουν προγνωστικό στόχο, καθώς έχοντας αποκτήσει γνώσεις από δεδομένα μπορούν να κάνουν προβλέψεις σχετικά με την ανθρώπινη συμπεριφορά: «[ένα] προγνωστικό μοντέλο καταγράφει τις σχέσεις μεταξύ των δεδομένων πρόβλεψης και της συμπεριφοράς [. ..][αφού δημιουργηθεί ένα μοντέλο, μπορεί να χρησιμοποιηθεί για να γίνουν νέες προβλέψεις για άτομα (ή άλλες οντότητες) των οποίων η συμπεριφορά είναι άγνωστη». Ο αλγόριθμος και η συνοδευτική γνώση που αποκτήθηκε ενσωματώνονται στη συνέχεια σε ένα μοντέλο υπολογιστή και προωθούνται ως μέρος ενός συστήματος λήψης αποφάσεων(Greenstein, no date). Για την κατανόηση των ανωτέρω, σημαντική είναι και η αναφορά σε δύο διαφορετικές προσεγγίσεις μοντέλων: Στα «έμπειρα μοντέλα» και τα μοντέλα μηχανικής μάθησης, όπως τα «νευρωνικά δίκτυα». Τα πρώτα ονομάζονται «expert systems or rules-based programs». Αναπτύχθηκαν στα πρώτα στάδια της ανάπτυξης συστημάτων τεχνητής νοημοσύνης, όταν επιστήμονες υπολογιστών προσπάθησαν να αναπτύξουν προγράμματα που μιμούνται την ανθρώπινη νοημοσύνη επιδιώκοντας να κατανοήσουν τις ανθρώπινες γνωστικές διαδικασίες και να τις αναπαράγουν. Για παράδειγμα, επιστήμονες υπολογιστών προσπάθησαν να κατανοήσουν τις διαδικασίες που εμπλέκονται στην εκμάθηση μιας γλώσσας και έτσι να αναπτύξουν έναν αλγόριθμο – μια ακολουθία ακριβών οδηγιών – που θα επέτρεπε στους υπολογιστές να μάθουν μια γλώσσα. Τα αποτελέσματα ήταν φτωχά, ιδιαίτερα σε σύνθετες εργασίες, όπως η εκμάθηση γλωσσών. Σε μικρότερο βαθμό, παρόμοια μοντέλα εξακολουθούν να χρησιμοποιούνται σήμερα (Scherer, 2019). Στη συνέχεια, η εμφάνιση των «Big Data» επέτρεψε μια σημαντική αλλαγή στην ανάπτυξη της τεχνητής νοημοσύνης. Αντί να αναπτύσσει πολύπλοκους αλγόριθμους για γνωστικές διεργασίες, η τεχνητή νοημοσύνη χρησιμοποιείται για να «μάθει» από τα υπάρχοντα δεδομένα. Η αναφορά στη «μάθηση» δεν αναφέρεται σε γνωστικές διαδικασίες που πιστεύεται ότι εμπλέκονται στην ανθρώπινη μάθηση. Αναφέρεται στη λειτουργική αίσθηση της μάθησης: την ικανότητα αλλαγής συμπεριφοράς μέσω της εμπειρίας με την πάροδο του χρόνου. Η διαδικασία της μηχανικής μάθησης έχει επιτύχει εκπληκτικά αποτελέσματα σε πολλούς τομείς. Συνεχίζοντας το προηγούμενο παράδειγμα εκμάθησης γλωσσών, τα προγράμματα μετάφρασης σε υπολογιστή είναι πλέον εξαιρετικά ακριβή. Σε αντίθεση με το παρελθόν, κανένας προγραμματιστής δεν χρειάζεται να κωδικοποιήσει έναν αλγόριθμο για μετάφραση. Πλέον, τα μοντέλα υπολογιστών, όπως τα νευρωνικά δίκτυα, χρησιμοποιούν τεράστιες ποσότητες διαθέσιμων δεδομένων για να «μάθουν» τις σχετικές δυνατότητες και να βελτιώνονται συνεχώς με άμεση ηλεκτρονική ανατροφοδότηση μέσω κλικ των χρηστών. Χρησιμοποιώντας μεγάλες ποσότητες δειγματοληπτικών δεδομένων και με επαρκή υπολογιστική ισχύ, ο υπολογιστής εξάγει τους απαραίτητους αλγόριθμους, αντί να κωδικοποιούνται αυτοί οι αλγόριθμοι στο μηχάνημα (Scherer, 2019). Ειδικότερα, τα τεχνητά νευρωνικά δίκτυα μαθαίνουν να εκτελούν εργασίες εξετάζοντας παραδείγματα, γενικά χωρίς να προγραμματίζονται με κανόνες που αφορούν συγκεκριμένες εργασίες. Ως εκ τούτου, τα τεχνητά νευρωνικά δίκτυα μπορούν να είναι εξαιρετικά χρήσιμα σε πολλούς τομείς, όπως

η μηχανική όραση, η επεξεργασία φυσικής γλώσσας, η γεωεπιστήμη για τη μοντελοποίηση των ωκεανών ή η ασφάλεια στον κυβερνοχώρο για τον εντοπισμό και τη διάκριση μεταξύ νόμιμων και κακόβουλων δραστηριοτήτων. Δεν απαιτούν δείγματα με ετικέτα, π.χ., για να αναγνωρίσουν τις γάτες στις εικόνες ή τους πεζούς στην κυκλοφορία, αλλά μπορούν να δημιουργήσουν από μόνα τους γνώση για το πώς μοιάζει μια γάτα (Završnik, 2020). Χαρακτηριστικά, η Scherer σημειώνει ότι «τα μοντέλα μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα, μπορούν να περιγραφούν ότι χρησιμοποιούν μια αντίστροφη προσέγγιση, επειδή εξάγουν τον αλγόριθμο από παρατηρήσιμα δεδομένα. Η μέθοδος είναι προγνωστική, υπολογίζοντας την πιθανότητα για οποιοδήποτε δεδομένο αποτέλεσμα με βάση τον εξαγόμενο αλγόριθμο και βελτιώνεται σταθερά». Κατόπιν τούτων, γίνεται σαφές ότι η λήψη αποφάσεων και η πρόβλεψη βάσει τεχνητής νοημοσύνης πραγματοποιούνται όταν εφαρμόζονται αλγόριθμοι σε σύνολα δεδομένων με εργασίες που κυμαίνονται από απλά, στενά αυτοματοποιημένα συστήματα έως πιο εξελιγμένα «νευρικά δίκτυα και βαθιά μάθηση» (McKay, 2020 που αναφέρεται στους Dawson et al. 2019: 14). Εξάλλου, η McKay (McKay, 2020) αναφέρει ότι υπάρχουν διάφορα στατιστικά προγνωστικά εργαλεία βάσει δεδομένων που χρησιμοποιούνται σε αξιολογήσεις κινδύνου ποινικής διαδικασίας που δεν είναι αυστηρά τεχνητή νοημοσύνη. Μάλλον είναι «αναλογιστικά» ή «αλγοριθμικά» όργανα. Παραθέτοντας δε μια σειρά από τέτοια εργαλεία, που χρησιμοποιούνται στην Αγγλία και την Ουαλία – HART, LSI-R, STABLE-2007, STATIC-99R, RSVP κ.λ.π., επισημαίνει ότι προκύπτουν παρόμοιες ηθικές ανησυχίες και προκλήσεις. Επίσης, η ίδια σημειώνει ότι ο όρος «αναλογιστική» αναφέρεται στη χρήση στατιστικών μεθόδων - αντί κλινικών μεθόδων - σε μεγάλα σύνολα δεδομένων ποσοστών εγκληματικών παραβάσεων ... για τον προσδιορισμό των διαφορετικών επιπέδων αδικημάτων που σχετίζονται με μια ομάδα ... και, με βάση αυτούς τους συσχετισμούς, την πρόβλεψη πρώτα της παρελθούσας, παρούσας ή μελλοντικής εγκληματικής συμπεριφοράς ενός συγκεκριμένου ατόμου και, δεύτερον, την επιβολή ενός αποτελέσματος ποινικής δικαιοσύνης για το συγκεκριμένο άτομο (McKay, 2020 παραθέτοντας τον Harcourt 2005), όπως επίσης ότι η αναλογιστική μπορεί επίσης να αναφέρεται στο γεγονός ότι η βαθμολογία καθορίζεται από έναν αλγόριθμο ((McKay, 2020) Smid 2014).

2.3. Πολιτική και ποινική δικαιοσύνη

Η πολιτική δικαιοσύνη αναφέρεται στις διαφορές του αστικού και εμπορικού δικαίου και γενικά στις διαφορές που επιλύονται από τα πολιτικά δικαστήρια. Η αναφορά στην “ποινική δικαιοσύνη” καταλαμβάνει όχι μόνο τη διαδικασία ενώπιον του ποινικού δικαστηρίου και τη λήψη αποφάσεως από τον ποινικό δικαστή (την περί ενοχής και ποινής δηλαδή απόφαση) αλλά και τα στάδια που προηγούνται και έπονται αυτής. Αυτό είναι σημαντικό στα πλαίσια της εργασίας αυτής, καθώς θα διαφανεί πως τα προηγούμενα στάδια μπορούν να επηρεάσουν τα επόμενα. Ειδικότερα, η ποινική διαδικασία περιλαμβάνει μια σειρά από διαδοχικά παραταγμένα σημεία, έρευνα, σύλληψη, κατηγορία, εγγύηση, δίκη, ποινή, αναστολή, χάρη, με αποτέλεσμα αποφάσεις που

ελήφθησαν νωρίτερα στη διαδικασία να τείνουν να επηρεάσουν αποφάσεις στα επόμενα στάδια της ποινικής διαδικασίας (Chiao, 2019).

2.4. Προσωπικά δεδομένα

Όσον αφορά στα "προσωπικά δεδομένα" στην εργασία θα αποτυπωθούν οι αναφορές στον Κανονισμό 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016, που εντοπίστηκαν στην επισκοπούμενη βιβλιογραφία.

Για τους σκοπούς, άλλωστε, του Κανονισμού ως δεδομένα προσωπικού χαρακτήρα νοούνται, σύμφωνα με το άρθρο 4 – Ορισμοί: κάθε πληροφορία που αφορά ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο (« το υποκείμενο των δεδομένων»)· το ταυτοποιήσιμο φυσικό πρόσωπο είναι εκείνο του οποίου η ταυτότητα μπορεί να εξακριβωθεί, άμεσα ή έμμεσα, ιδίως μέσω αναφοράς σε αναγνωριστικό στοιχείο ταυτότητας, όπως όνομα, σε αριθμό ταυτότητας, σε δεδομένα θέσης, σε επιγραμμικό αναγνωριστικό ταυτότητας ή σε έναν ή περισσότερους παράγοντες που προσιδιάζουν στη σωματική, φυσιολογική, γενετική, ψυχολογική, οικονομική, πολιτιστική ή κοινωνική ταυτότητα του εν λόγω φυσικού προσώπου. Σημαντικές εξάλλου για το αντικείμενο της εργασίας είναι: α) η ρύθμιση του άρθρου 22 του Κανονισμού, κατά την παρ. 1 του οποίου: « το υποκείμενο των δεδομένων έχει το δικαίωμα να μην υπόκειται σε απόφαση που λαμβάνεται αποκλειστικά βάσει αυτοματοποιημένης επεξεργασίας, συμπεριλαμβανομένης της κατάρτισης προφίλ, η οποία παράγει έννομα αποτελέσματα που το αφορούν ή το επηρεάζει σημαντικά με παρόμοιο τρόπο», κατά δε την παρ. 3 αυτού: « Στις περιπτώσεις που αναφέρονται στην παράγραφο 2 στοιχεία α) και γ), ο υπεύθυνος επεξεργασίας των δεδομένων εφαρμόζει κατάλληλα μέτρα για την προστασία των δικαιωμάτων, των ελευθεριών και των έννομων συμφερόντων του υποκειμένου των δεδομένων, τουλάχιστον του δικαιώματος εξασφάλισης ανθρώπινης παρέμβασης από την πλευρά του υπευθύνου επεξεργασίας, έκφρασης άποψης και αμφισβήτησης της απόφασης», β) οι ρυθμίσεις του τμήματος 2 του Κανονισμού (Ενημέρωση και πρόσβαση σε δεδομένα προσωπικού χαρακτήρα), δηλαδή οι ρυθμίσεις των άρθρων 13 (Πληροφορίες που παρέχονται εάν τα δεδομένα προσωπικού χαρακτήρα συλλέγονται από το υποκείμενο των δεδομένων), 14 (Πληροφορίες που παρέχονται εάν τα δεδομένα προσωπικού χαρακτήρα δεν έχουν συλλεγεί από το υποκείμενο των δεδομένων) και 15 (Δικαίωμα πρόσβασης του υποκειμένου των δεδομένων), καθώς επίσης γ) και οι αρχές και οι νόμιμες βάσεις επεξεργασίας που προβλέπονται στα άρθρα 5 και 6 του Κανονισμού. Μεταξύ δε των αρχών επεξεργασίας (άρθρο 5 του Κανονισμού) προβλέπονται η αρχή της διαφάνειας, της ακρίβειας και της λογοδοσίας. Σύμφωνα με την τελευταία (άρθρο 5 παρ. 2 του Κανονισμού) οι υπεύθυνοι επεξεργασίας δεδομένων (οι οποίοι καθορίζουν τους σκοπούς και τον τρόπο επεξεργασίας δεδομένων προσωπικού χαρακτήρα) φέρουν την ευθύνη και είναι σε θέση να αποδείξουν τη συμμόρφωσή τους με τον Κανονισμό. Ωστόσο, σύμφωνα με το άρθρο 2 παρ. 2 δ του Κανονισμού "ο παρών κανονισμός δεν εφαρμόζεται στην επεξεργασία δεδομένων προσωπικού χαρακτήρα: … δ) από αρμόδιες αρχές για τους σκοπούς της πρόληψης, της

διερεύνησης, της ανίχνευσης ή της δίωξης ποινικών αδικημάτων ή της εκτέλεσης ποινικών κυρώσεων, ...". Στις περιπτώσεις αυτές εφαρμόζεται η οδηγία για την επιβολή του νόμου (2016/680). Όπως αναφέρει η Lynskey (Lynskey, 2019):

Αυτή η ρητή εξαίρεση της «επεξεργασίας δεδομένων για την επιβολή του νόμου» από τον Κανονισμό μπορεί να δώσει την αρχική παραπλανητική εντύπωση ότι ο καταμερισμός της εργασίας μεταξύ των δύο νομοθετικών πράξεων είναι σαφής. Ωστόσο, η αλληλεπίδραση μεταξύ αυτών των δύο δυνητικά συναφών μέσων είναι, στην πραγματικότητα, περίπλοκη. Πρώτον, αξίζει να τονιστεί ότι, για να εμπίπτει στο πεδίο εφαρμογής της οδηγίας για την επιβολή του νόμου, η επεξεργασία των δεδομένων πρέπει να αναλαμβάνεται από «αρμόδια αρχή». Ως αρμόδια αρχή ορίζεται στο άρθρο 3 παρ. 7 στοιχ. α) και β), «κάθε δημόσια αρχή αρμόδια για την πρόληψη, διερεύνηση, ανίχνευση ή δίωξη ποινικών αδικημάτων ή την εκτέλεση ποινικών κυρώσεων, συμπεριλαμβανομένης της προστασίας και της πρόληψης απειλών για τη δημόσια ασφάλεια· ή οποιοσδήποτε άλλος φορέας ή οντότητα στον οποίο έχει ανατεθεί από τη νομοθεσία του κράτους μέλους ο ρόλος δημόσιας αρχής και η εκτέλεση δημόσιας εξουσίας για αυτούς τους σκοπούς». Ως εκ τούτου, σε έναν ιδιωτικό φορέα που διενεργεί επεξεργασία δεδομένων για σκοπούς επιβολής του νόμου, πρέπει να του ανατεθεί αυτός ο ρόλος από το δίκαιο του κράτους μέλους, προτού απομακρυνθεί από το πεδίο εφαρμογής του GDPR και τεθεί υπό την εφαρμογή των σχετικών διατάξεων της Οδηγίας (2016/680). Ως εκ τούτου, ελλείψει τέτοιας νομοθετικής θέσπισης, οι διατάξεις του Κανονισμού εξακολουθούν να ισχύουν για ιδιωτικούς φορείς που επεξεργάζονται προσωπικά δεδομένα για σκοπούς επιβολής του νόμου. Αυτό επιβεβαιώνεται σιωπηρά από το άρθρο 23 του Κανονισμού: αυτή η διάταξη επιτρέπει στο δίκαιο της Ένωσης ή του κράτους μέλους να περιορίζει τις υποχρεώσεις που απορρέουν από καθορισμένες διατάξεις του Κανονισμού, όπου ένας τέτοιος περιορισμός είναι απαραίτητος για σκοπούς επιβολής του νόμου, αναγνωρίζοντας έτσι ότι ο Κανονισμός θα ίσχυε διαφορετικά για τέτοια επεξεργασία σε ορισμένες περιπτώσεις. Συνεπώς, πολλά αφορούν στον ορισμό μιας «αρμόδιας αρχής» για τους σκοπούς της Οδηγίας και πότε θα μπορούσε να ειπωθεί ότι έχει ανατεθεί σε μια ιδιωτική οντότητα από το νόμο αυτό το καθεστώς. Ωστόσο, θα μπορούσε κανείς να αναρωτηθεί, για παράδειγμα, εάν η «ανάθεση» απαιτεί ένα ξεχωριστό νομοθετικό μέσο (Garstka, 2018, όπως παρατίθεται στην Lynskey, 2019). Δεύτερον, ακόμη και όταν η επεξεργασία δεδομένων αναλαμβάνεται από μια «αρμόδια αρχή», το εάν αυτή η επεξεργασία εμπίπτει στο πεδίο εφαρμογής του Κανονισμού ή της Οδηγίας θα εξαρτηθεί από τον σκοπό της επεξεργασίας. Έτσι, όσον αφορά την κοινή χρήση δεδομένων από τις αρμόδιες αρχές, η πράξη της μετάδοσης δεδομένων από μια αρμόδια αρχή σε μια μη αρμόδια αρχή για σκοπούς επιβολής του νόμου εμπίπτει στο πεδίο εφαρμογής της Οδηγίας (π.χ. μεταφορά δεδομένων από την αστυνομία σε έναν ιδιωτικό πάροχο λογισμικού προγνωστικής αστυνόμευσης) ενώ μια μεταφορά για σκοπούς μη επιβολής του νόμου (π.χ. μεταφορά δεδομένων σε ιατρικές ή κοινωνικές υπηρεσίες) θα καλύπτεται από τον Κανονισμό (Αιτιολογική σκέψη 34, Οδηγίας για την επιβολή του νόμου). Ομοίως, όταν μια αρμόδια αρχή συλλέγει αρχικά δεδομένα προσωπικού χαρακτήρα για σκοπούς επιβολής του νόμου, αλλά στη συνέχεια αυτά τα δεδομένα υποβάλλονται σε επεξεργασία για εναλλακτικούς σκοπούς μη

επιβολής του νόμου, εφαρμόζεται ο Κανονισμός (άρθρο 9 παράγραφος 1 και αιτιολογική σκέψη 11 Οδηγίας).

Από τα ανωτέρω καθίσταται σαφές ότι τα αναφερόμενα στην εργασία ζητήματα (θέματα ή παραδείγματα) δύναται να εγείρουν ζήτημα εφαρμογής του ενός ή του άλλου νομοθετήματος. Ενόψει τούτου, πρέπει να αναφερθεί ότι και στην Οδηγία περιλαμβάνονται αρχές που διέπουν την επεξεργασία των δεδομένων (άρθρο 4 Οδηγίας), μεταξύ των οποίων η αρχή της ακρίβειας και της λογοδοσίας. Επίσης, πρέπει να αναφερθεί ότι αναφορικά με την ποινική δικαιοσύνη, στην οποία θα τίθεται θέμα εφαρμογής του ενός ή του άλλου νομοθετήματος, κρίνεται ότι στην περίπτωση πρόβλεψης της δικαιοδοτικής κρίσης ή λήψης αποφάσεων με εργαλεία ΤΝ, από τα δικαστήρια, εφαρμοστέες είναι οι διατάξεις του Κανονισμού. Τούτο για το λόγο ότι σε σχέση με την ποινική δικαιοσύνη, ένας αλγόριθμος μπορεί να γίνει κατανοητός ως ένας κανόνας που χρησιμοποιεί αριθμητικές εισόδους για να παράγει μια πρόβλεψη σχετική με το σημείο της διαδικαστικής απόφασης (McKay, 2020, Christin et al. 2015).] Κατ' αυτόν όμως τον τρόπο καθίσταται σαφές ότι η οποιαδήποτε συλλογή προσωπικών δεδομένων, ακόμα και αν συνέβη για σκοπούς επιβολής του νόμου, γίνεται πλέον για άλλο (εναλλακτικό) σκοπό, με αποτέλεσμα να μην εντάσσεται στο πλαίσιο εφαρμογής της οδηγίας, αλλά του Κανονισμού. Επιπλέον, στο σημείο αυτό πρέπει να σημειωθεί ότι και στην Οδηγία προβλέπεται αντίστοιχη διάταξη αυτής του άρθρου 22 του Κανονισμού. Πρόκειται για τη διάταξη του άρθρου 11 της Οδηγίας, το οποίο όμως, σε αντίθεση με το άρθρο 22 του Κανονισμού, το οποίο διατυπώνεται ως δικαίωμα, διατυπώνεται ως απαγόρευση. Πέραν αυτής και άλλων διαφορών, τις οποίες παραθέτει η Lynskey, 2019, ως πιο σημαντική παραθέτει το ότι η Οδηγία ισχύει μόνο για αυτοματοποιημένες αποφάσεις που βασίζονται αποκλειστικά στην αυτοματοποιημένη επεξεργασία. Αυτό εξαρτάται από το πώς λαμβάνει χώρα η διαδικασία λήψης αποφάσεων στην πράξη. Σε αυτό το πλαίσιο, θα πρέπει να μετρηθεί σε ποιο βαθμό η τελική απόφαση συνεπάγεται τη διακριτική ευχέρεια και την κρίση του αξιωματικού που λαμβάνει αυτή την απόφαση. Εάν η αποφασιστική απόφαση ενσωματώνει την ανθρώπινη κρίση και αποτελεί την τελική σύσταση του αρμόδιου αξιωματικού, τότε αυτό δεν είναι αποκλειστικά αυτοματοποιημένο και δεν εφαρμόζεται το άρθρο 11 της Οδηγίας. Τέλος, πρέπει να σημειωθεί ότι τα κράτη μέλη μπορούν να αποφύγουν την απαγόρευση του άρθρου 11 της Οδηγίας, θέτοντας τέτοιες τεχνολογίες σε θεσμική βάση (Lynskey, 2019).

2.5. Διαδικασία λήψης απόφασης τεχνητής νοημοσύνης

Αναφορικά, τέλος, με τη "διαδικασία λήψης απόφασης τεχνητής νοημοσύνης" για την κατανόηση του εύρους και του αντικειμένου της εργασίας, λεκτέα τα εξής: Καταρχάς, πρέπει να σημειώσουμε ότι είναι πιο γόνιμο να μιλάμε για μια διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης παρά για ένα σύστημα λήψης αποφάσεων τεχνητής νοημοσύνης (Krupiy, 2020). Η Tetyana (Tanya) Krupiy επισημαίνει ότι ο όρος διαδικασία λήψης αποφάσεων είναι προτιμότερος επειδή μπορεί να οριστεί ότι περιλαμβάνει το στάδιο όπου ο επιστήμονας υπολογιστών διατυπώνει το πρόβλημα που πρέπει να λυθεί

χρησιμοποιώντας μια διαδικασία που βασίζεται στην τεχνητή νοημοσύνη. Ειδικότερα, οι επιστήμονες υπολογιστών ξεκινούν την ανάπτυξη ενός συστήματος τεχνητής νοημοσύνης διατυπώνοντας ένα πρόβλημα για το οποίο στοχεύουν στη δημιουργία χρήσιμης γνώσης. Στη συνέχεια, οι επιστήμονες υπολογιστών προετοιμάζουν δεδομένα μετατρέποντάς τα σε μια μορφή που μπορεί να επεξεργαστεί ένα σύστημα τεχνητής νοημοσύνης. Το τελικό αποτέλεσμα που οι επιστήμονες υπολογιστών προσπαθούν να επιτύχουν θα επηρεάσει τον τρόπο με τον οποίο χειρίζονται και επισημαίνουν τα δεδομένα. Το επόμενο βήμα είναι η χρήση των δεδομένων για τη δημιουργία ενός μοντέλου του εξωτερικού περιβάλλοντος που αποτυπώνει το αντικείμενο ενδιαφέροντος, όπως οι προβλεπόμενοι βαθμοί εξέτασης ενός μαθητή. Το μοντέλο εντοπίζει μοτίβα στα δεδομένα ανιχνεύοντας συσχετίσεις μεταξύ τμημάτων δεδομένων. Το μοντέλο προσδιορίζει ποια κομμάτια πληροφοριών σχετίζονται μεταξύ τους. Στη διαδικασία μάθησης χωρίς επίβλεψη, οι επιστήμονες υπολογιστών αφήνουν το σύστημα να αναζητήσει μοτίβα. Το σύστημα κατανέμει τα άτομα σε ομάδες με βάση κοινά χαρακτηριστικά. Στην εποπτευόμενη μαθησιακή διαδικασία, οι επιστήμονες υπολογιστών διατυπώνουν ένα κριτήριο και το σύστημα τεχνητής νοημοσύνης ταξινομεί τα άτομα σε ομάδες με βάση την πιθανότητα εκπλήρωσης αυτού του κριτηρίου. Το μοντέλο που δημιουργεί το σύστημα επιτρέπει στο χρήστη να προβλέπει ότι ένα άτομο ανήκει σε μια συγκεκριμένη ομάδα ανθρώπων με κοινά χαρακτηριστικά. Το σύστημα τεχνητής νοημοσύνης προβλέπει την απόδοση ενός ατόμου με βάση την απόδοση ατόμων τα οποία θεωρεί ότι έχουν παρόμοια χαρακτηριστικά με το εν λόγω άτομο. Εφαρμόζει μια διαδικασία λήψης αποφάσεων για να καθοριστεί εάν ένα άτομο δικαιούται μια θετική απόφαση (Krupiy, 2020). Η Tetyana (Tanya) Krupiy παρουσιάζοντας τα ανωτέρω στάδια ενός συστήματος τεχνητής νοημοσύνης, επισημαίνει ότι η κοινωνία θα πρέπει να αναλογιστεί τον κοινωνικό ρόλο που έχει ο όρος νοημοσύνη καθώς συνεχίζει να βελτιώνει την έννοια αυτού του όρου. Επί του παρόντος, υπάρχουν διαφορετικοί ορισμοί αλγοριθμικών ή αυτοματοποιημένων συστημάτων αποφάσεων. Προτιμότεροι είναι οι ορισμοί που πλαισιώνουν το θέμα ευρέως και με αναφορά σε μια διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης. Επισημαίνει δε ότι ένας από τους λόγους για τους οποίους ο όρος αυτοματοποιημένο σύστημα λήψης αποφάσεων είναι στενότερος από τον όρο διαδικασία λήψης αποφάσεων είναι επειδή αποκλείει στάδια που επηρεάζουν το αποτέλεσμα της διαδικασίας λήψης αποφάσεων αλλά που λαμβάνουν χώρα πριν από την πραγματική κατασκευή του συστήματος. Αντιθέτως, ένας ορισμός με αναφορά σε μια διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης, επικεντρώνεται στα βήματα που εμπλέκονται στην κατασκευή ενός μοντέλου που χρησιμοποιεί το σύστημα λήψης αποφάσεων τεχνητής νοημοσύνης για να κάνει προβλέψεις σχετικά με τις μελλοντικές επιδόσεις και τη διαδικασία λήψης μιας απόφασης σε σχέση με ένα άτομο. Έτσι, δίνεται έμφαση στη διαδικασία που οδηγεί στην απόφαση παρά στον τύπο των τεχνικών που χρησιμοποιούν οι επιστήμονες υπολογιστών για να προγραμματίσουν συστήματα τεχνητής νοημοσύνης. Ένας τέτοιος ορισμός διευκολύνει τους νομοθέτες και τους τελικούς χρήστες να συζητήσουν τις κοινωνικές συνέπειες της χρήσης συστημάτων λήψης αποφάσεων τεχνητής νοημοσύνης. Κατά την ίδια (Krupiy, 2020) ένα άλλο πλεονέκτημα του ορισμού του όρου ως προς τα στοιχεία που περιλαμβάνουν την

εν λόγω διαδικασία λήψης αποφάσεων είναι ότι παρέχει κατανόηση του τι κάνει το σύστημα και πώς επιτυγχάνει τον στόχο του. Από την άλλη πλευρά, ο όρος σύστημα λήψης αποφάσεων τεχνητής νοημοσύνης είναι αδιαφανής. Ειδικότερα, το γεγονός ότι η συγκεκριμένη τεχνολογία χρησιμοποιεί συνδυασμό διαφορετικών στοιχείων αποκαλύπτει λίγα για τη φύση των εργασιών που εκτελεί και τον τρόπο με τον οποίο τις εκτελεί. Αυτό προκύπτει από το γεγονός ότι ο όρος σύστημα εφιστά την προσοχή στη φυσική αρχιτεκτονική του συστήματος και σε ποια στοιχεία αποτελούν το σύστημα. Αυτό που είναι ο πυρήνας για την κατανόηση της λήψης αποφάσεων είναι η διαδικασία μέσω της οποίας κάποιος καταλήγει σε μια απόφαση και όχι το γεγονός ότι διάφορα αλληλεξαρτώμενα στάδια εμπλέκονται στη διαδικασία λήψης αποφάσεων. Τονίζει, επίσης, (Krupiy, 2020) ότι υπάρχει παραλληλισμός μεταξύ των στοιχείων που περιλαμβάνουν τον άνθρωπο και τις διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης. Ο όρος σύστημα λήψης αποφάσεων τεχνητής νοημοσύνης αποτυγχάνει να συλλάβει αυτό το σημαντικό στοιχείο. Αυτό συμβαίνει γιατί η κοινωνία δεν αντιλαμβάνεται τα ανθρώπινα όντα και τη διαβούλευση τους ως σύστημα. Ωστόσο, μπορεί κανείς να μιλήσει για τις ομοιότητες στη διαδικασία λήψης αποφάσεων στην οποία εμπλέκονται τα ανθρώπινα όντα και τα συστήματα τεχνητής νοημοσύνης που εκτελούν, επειδή τα ανθρώπινα όντα αναπτύσσουν τη διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης. Δεδομένου ότι τα ανθρώπινα όντα ασκούν την κρίση τους κατά την ανάπτυξη διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης, δεν αποτελεί έκπληξη το γεγονός ότι μπορεί να υπάρχει ένας βαθμός ομοιότητας μεταξύ των διαδικασιών λήψης αποφάσεων του ανθρώπου και της τεχνητής νοημοσύνης. Η ανθρώπινη διαδικασία λήψης αποφάσεων ξεκινά με τον καθορισμό του στόχου που έχει σχεδιαστεί για να επιτύχει η διαδικασία λήψης αποφάσεων και με τον προσδιορισμό των κριτηρίων που αντιστοιχούν στο δικαίωμα για θετική απόφαση. Όταν οι επιστήμονες υπολογιστών αποφασίζουν πώς να διατυπώσουν το πρόβλημα και τι προβλέπει η διαδικασία τεχνητής νοημοσύνης, επιλέγουν τον στόχο για τη διαδικασία λήψης αποφάσεων και τα κριτήρια που αποτελούν τη βάση της διαδικασίας λήψης αποφάσεων. Κατά τη διατύπωση του προβλήματος που πρέπει να λύσει ο υπολογιστής οι επιστήμονες βρίσκονται σε παρόμοια θέση με τους ανθρώπους που λαμβάνουν αποφάσεις που είναι επιφορτισμένοι με την ανάπτυξη και την εφαρμογή μιας διαδικασίας λήψης αποφάσεων. Η διαφορά είναι ότι οι άνθρωποι που λαμβάνουν αποφάσεις μπορούν να επιλέξουν κριτήρια που μπορούν να εκφραστούν τόσο με ποσοτικούς όσο και με ποιοτικούς όρους. Από την άλλη πλευρά, οι επιστήμονες υπολογιστών μπορούν να επιλέξουν μόνο εκείνα τα κριτήρια που μπορούν να εκφραστούν με ποσοτικούς όρους. Η διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης θα πρέπει να γίνει κατανοητή ως μια πιο περιορισμένη διαδικασία από μια ανθρώπινη διαδικασία λήψης αποφάσεων λόγω της περιορισμένης ικανότητάς της να συλλαμβάνει ποιοτικά δεδομένα και το πλαίσιο πίσω από αυτά τα δεδομένα (Krupiy, 2020). Επίσης, η Tetyana (Tanya) Krupiy, η οποία χρησιμοποιεί στην επισκοπούμενη εργασία της ως μελέτη περίπτωσης, το πλαίσιο όπου τα εκπαιδευτικά ιδρύματα αυτοματοποιούν τη διαδικασία επιλογής μαθητών χρησιμοποιώντας διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης, σημειώνει ότι υπάρχει μια άλλη σημαντική διαφορά μεταξύ των διαδικασιών λήψης αποφάσεων του ανθρώπου και της τεχνητής νοημοσύνης. Οι άνθρωποι

που λαμβάνουν αποφάσεις καθορίζουν ποιες πτυχές του ατόμου εξετάζουν επιλέγοντας τα κριτήρια λήψης αποφάσεων. Αντίθετα, οι διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης δημιουργούν αυτό που ο Luke Stark (Luke Stark, 'Algorithmic Psychometrics and the Scalable Subject' (2018) 48 Social Studies of Science 204, 207) αποκαλεί «κλιμακούμενο θέμα». Σύμφωνα με τον τελευταίο, το σύστημα σκοπεύει να μοντελοποιήσει το άτομο, αλλά στην πραγματικότητα αντανακλά συσχετίσεις μεταξύ χαρακτηριστικών που υπάρχουν μέσα σε μια ομάδα που μπορεί να μην ισχύουν για το εν λόγω άτομο. Η Tetyana (Tanya) Krupiy, για την μελέτη περίπτωσή της, χρησιμοποιεί ως παράδειγμα το εξής: δεδομένου ότι το μοντέλο ομαδοποιεί τους μαθητές με βάση τις επιδόσεις των προηγούμενων εξετάσεων με σκοπό να κάνει μια πρόβλεψη για τα μελλοντικά αποτελέσματα, θα ομαδοποιήσει μαθητές που είχαν κακή επίδοση ανεξάρτητα από τον λόγο για τα χαμηλά αποτελέσματα. Αυτό μπορεί να έχει ως αποτέλεσμα το σύστημα τεχνητής νοημοσύνης να προβλέπει λανθασμένα έναν χαμηλό βαθμό για έναν μαθητή του οποίου η απόδοση στο παρελθόν είχε επηρεαστεί από μια ασθένεια, αλλά ο οποίος ανέρρωσε αργότερα. Η ίδια, ωστόσο, επισημαίνει ότι οι διαφορές μεταξύ των διαδικασιών λήψης αποφάσεων σε ανθρώπους και τεχνητή νοημοσύνη δεν αποκλείουν την εξέταση της αυτοματοποίησης των αποφάσεων με όρους διαδικασίας λήψης αποφάσεων τεχνητής νοημοσύνης. Στην πραγματικότητα, ο όρος διαδικασία υπογραμμίζει το γεγονός ότι τα ανθρώπινα όντα κατασκευάζουν τη διαδικασία λήψης αποφάσεων μέσα στη μηχανή. Επιπλέον, αυτός ο όρος καθιστά δυνατό να οριοθετηθεί σε ποιο σημείο αρχίζει και τελειώνει η λήψη αποφάσεων. Κατ' αυτήν (Krupiy, 2020) σ' έναν κατάλληλο ορισμό μιας διαδικασίας λήψης αποφάσεων τεχνητής νοημοσύνης, πρέπει να συμπεριληφθεί επιπλέον το στοιχείο της ανθρώπινης λήψης αποφάσεων που εμπλέκεται στη διαμόρφωση ενός προβλήματος που πρέπει να λυθεί και πώς να κατασκευαστεί το σύστημα για την επίτευξη αυτού του στόχου. Η διαδικασία λήψης αποφάσεων της τεχνητής νοημοσύνης θα πρέπει να γίνει κατανοητή ως αρχή με τον επιστήμονα υπολογιστών να διατυπώνει το πρόβλημα που πρέπει να λυθεί και τους στόχους που πρέπει να επιτευχθούν. Περιλαμβάνει τη συλλογή, τον καθαρισμό, την επισήμανση, τη συγκέντρωση, την ανάλυση, τον χειρισμό και την επεξεργασία δεδομένων. Ένας εκτενής ορισμός για τον όρο διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης είναι απαραίτητος, για τον εξής λόγο: Η διαδικασία χαρτογράφησης των δεδομένων στο μοντέλο και πρόβλεψης της απόδοσης ενός ατόμου με βάση το μοντέλο έχει επίδραση στο εάν το άτομο θα λάβει μια θετική απόφαση. Ο προτεινόμενος ορισμός έχει σχεδιαστεί για να συλλαμβάνει το γεγονός ότι οι επιστήμονες υπολογιστών λαμβάνουν υποκειμενικές αποφάσεις κατά τη δημιουργία της αρχιτεκτονικής που επιτρέπει στη διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης να συλλέγει, να συγκεντρώνει και να αναλύει δεδομένα. Οι επιλογές που κάνουν οι επιστήμονες υπολογιστών επηρεάζουν τον τρόπο με τον οποίο η διαδικασία λήψης απόφασης τεχνητής νοημοσύνης παράγει αποφάσεις και τι είδους απόφαση λαμβάνει ένα άτομο. Συμπερασματικά, η διαδικασία λήψης αποφάσεων της τεχνητής νοημοσύνης θα πρέπει να οριστεί ώστε να ενσωματώνει όλα τα στάδια ανάπτυξης και λειτουργίας του συστήματος, ξεκινώντας από τη διατύπωση του προβλήματος που πρέπει να λυθεί και τελειώνοντας με μια έξοδο απόφασης (Krupiy, 2020).

3. Μεθοδολογία

Στην εργασία μας ενδιαφέρει να συγκεντρώσουμε βιβλιογραφία για τη χρήση τεχνητής νοημοσύνης στη διαδικασία λήψης αποφάσεων στους τομείς της ποινικής και της πολιτικής δικαιοσύνης. Για το σκοπό αυτά ακολουθήθηκε η μέθοδος της συστηματικής βιβλιογραφικής επισκόπησης, όπως προτείνεται από τους Jane Webster και Richard T.Watson (Webster and Watson, 2002). Η αναζήτηση των άρθρων έγινε στις βάσεις δεδομένων web of science και scopus, ενώ όσα άρθρα ήταν σε περίληψη, αναζητήθηκαν περαιτέρω μέσω google και ανευρέθηκαν στις Google Scholar, ResearchGate, Science Direct Elsevier, Springer Link, IEEE Xplace, DOI. Org (Crossref), ACM Digital Library, Library of Congress ISBN, Frontiers, Cambridge University Press, arXiv.org, core.ac.uk, ενώ υπάρχουν και άρθρα τα οποία (πέραν της περίληψης τους) δεν ανευρέθηκαν. Οι λέξεις-κλειδιά με τις οποίες έγινε η αναζήτηση στις βάσεις δεδομένων web of science και scopus, είναι οι εξής: Artificial intelligence and justice, “artificial intelligence and justice”, “deep learning” AND “court”, “judicial data analysis”, “machine learning and court”, “automated judicial decision making”, “Judicial decision making”, “machine learning” AND “judicial decision making”, “AI” AND “judicial decision making”, “artificial intelligence” AND “judicial decision making”, “machine learning” AND “ judge”, “AI” AND “legal reasoning technology”, “AI” AND “legal reasoning”, "personal data" AND "court", “algorithmic justice”, “artificial intelligence and law”. Αυτές, μαζί με τα αποτελέσματα που έδωσαν καταγράφονται στο Παράρτημα Ι. Από τα αποτελέσματα αυτά τα οποία είναι 188 συνολικά, αφαιρέθηκαν όσα ήταν διπλά, είτε γιατί ανευρέθηκαν σε περισσότερες βάσεις δεδομένων, με την αναζήτηση με τις ίδιες λέξεις-κλειδιά, είτε στην ίδια βάση δεδομένων, με την αναζήτηση με διαφορετικές λέξεις-κλειδιά. Κατά την αρχική αυτή επιλογή των άρθρων, αναγνώστηκε η περίληψη και το συμπέρασμα, ενώ σε δεύτερο χρόνο ακολούθησε η ανάγνωση όλου του άρθρου, προκειμένου να επιλεγούν τα άρθρα που χρησιμοποιήθηκαν περαιτέρω κατά τα επόμενα στάδια της εργασίας. Κριτήριο για την αρχική αυτή επιλογή ήταν η ανεύρεση άρθρων στην αγγλική γλώσσα, καθώς και η ανεύρεση ολόκληρων των άρθρων και όχι μόνο της περίληψής τους. Επίσης, ελήφθη υπόψη ο χρόνος δημοσίευσης των επιλεγόμενων άρθρων μεταξύ της περιόδου από 2014 έως τον Ιανουάριο 2022. Με βάση τα κριτήρια αυτά, ο αριθμός των άρθρων που αφαιρέθηκαν είναι: 58 άρθρα. Απέμειναν, έτσι: 130 άρθρα. Ακολούθως, όπως προαναφέρθηκε, σε δεύτερο χρόνο έγινε η ανάγνωση ολόκληρων των άρθρων. Κατά την ανάγνωση του κάθε άρθρου καταγράφονταν σημειώσεις και συγκεκριμένα καταγράφονταν σημειώσεις για τα ζητήματα που αυτό ανέλυε, καθώς επίσης, και μία περίληψη για κάθε άρθρο ή μία αναφορά στη θεματική του. Μέσα από αυτήν την καταγραφή εντοπίστηκαν σταδιακά τα ζητήματα που απασχόλησαν την επισκοπούμενη βιβλιογραφία και άπτονται του αντικειμένου της εργασίας, δηλαδή τη χρήση τεχνητής νοημοσύνης κατά τη διαδικασία λήψης αποφάσεων ενώπιον των ποινικών και πολιτικών δικαστηρίων. Έτσι, στο δεύτερο αυτό στάδιο

της επιλογής των άρθρων αφαιρέθηκαν όσα δεν είχαν σχέση με το αντικείμενο της εργασίας, καθώς επίσης και όσα ανέπτυσσαν ζητήματα που ήδη είχαν ανευρεθεί κατά την ανάγνωση των προηγούμενων άρθρων, χωρίς να δίνουν κάποια επιπλέον διάσταση σ' αυτά. Επίσης, αφαιρέθηκαν τα άρθρα ανάπτυξης διάφορων μοντέλων τεχνητής νοημοσύνης στον τομέα της δικαιοσύνης, με εξαίρεση κάποια άρθρα, των οποίων η επιλογή κρίθηκε αναγκαία, προκειμένου να χρησιμοποιηθούν για την κατανόηση των αναπτυσσόμενων θεμάτων της εργασίας. Κατόπιν τούτων, απέμειναν όσα άρθρα φέρουν, στο Παράρτημα Ι την ένδειξη: ΝΑΙ, δηλαδή 34 άρθρα Ακολούθως, αφού εντοπίστηκαν τα ζητήματα που αφορούν στη χρήση τεχνητής νοημοσύνης κατά τη διαδικασία λήψης αποφάσεων ενώπιον των ποινικών και πολιτικών δικαστηρίων, σε τρίτο στάδιο, έγινε η κατάταξή τους σε κατηγορίες και υποκατηγορίες αυτών και συγκεκριμένα η ομαδοποίηση των τελευταίων στις κατηγορίες (της δικαιοσύνης, της διαφάνειας και της λογοδοσίας) που κρίθηκαν ως βασικές. Συγκεκριμένα, μέσα από τη μελέτη των άρθρων διαπιστώθηκε λ.χ. ότι ζητήματα αιτιολόγησης αποφάσεων ή κατανόησης των σχετικών διαδικασιών εντάσσονται στην κατηγορία της διαφάνειας ή ότι το ζήτημα της ακρίβειας των αλγορίθμων τεχνητής νοημοσύνης σχετίζεται άμεσα με τη δικαιοσύνη, με το κατά πόσο δηλαδή οι σχετικές διαδικασίες είναι δίκαιες ή όχι. Μετά από αυτήν τη σύνθεση των κατηγοριών και υποκατηγοριών των ζητημάτων, αυτά μελετήθηκαν και αναλύθηκαν περαιτέρω, όπως παρουσιάζονται στα επιλεγμένα άρθρα, συνδυαζόμενα μεταξύ τους. Τα αποτελέσματα αυτής της μελέτης παρατέθηκαν στα κεφάλαια 4.4. επόμενα.

Περαιτέρω, αναφορικά με τα προσωπικά δεδομένα, η βιβλιογραφία που συγκεντρώθηκε με την ως άνω μέθοδο περιορίζεται σε συγκεκριμένα και περιορισμένα ζητήματα που εγείρονται από το δίκαιο των προσωπικών δεδομένων στις σχετικές διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης. Κρίθηκε, όμως, απαραίτητο για την πληρότητα της εργασίας, να γίνει αναφορά, στο Παράρτημα ΙΙ, στην προβλεπόμενη στο άρθρο 35 του Κανονισμού εκτίμηση αντικτύπου, καθώς επίσης και στα ελληνικά νομοθετήματα που άπτονται του θέματος της εργασίας. Για τη συγκέντρωση της βιβλιογραφίας, χρησιμοποιήθηκαν ελληνικές πηγές, για το λόγο ότι, αφενός για το άρθρο 35 του Κανονισμού, δεν εντοπίστηκε στην αναφερόμενη στο Παράρτημα Ι βιβλιογραφία, κάποια αναφορά, αφετέρου για τα ελληνικά νομοθετήματα, γιατί η ρυθμιστική τους εμβέλεια αφορά μόνο στην εσωτερική έννομη τάξη. Η σχετική αναζήτηση έγινε στο διαδίκτυο, όσον αφορά τα νομοθετήματα και στη βάση δεδομένων Qualex της Νομικής Βιβλιοθήκης, με τη λέξη - κλειδί "εκτίμηση αντικτύπου". Επίσης, πρέπει να σημειωθεί ότι στην εργασία καταγράφονται όσα αναφέρονται στην επισκοπούμενη βιβλιογραφία αναφορικά με την προστασία προσωπικών δεδομένων, χωρίς να τίθεται όμως θέμα εφαρμογής του Γενικού Κανονισμού ή της οδηγίας για την επιβολή νόμου, καθώς κάτι τέτοιο είναι περίπλοκο.

4. Αποτελέσματα

4.1. Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζονται τα ζητήματα που θέτει η χρήση τεχνητής νοημοσύνης κατά τη διαδικασία λήψης απόφασης στη ποινική και πολιτική δικαιοσύνη, μέσα από τη μελέτη των άρθρων που προέκυψαν από την αναζήτηση στη διεθνή βιβλιογραφία. Αρχικά, στο κεφάλαιο 4.2., εκτίθενται τα άρθρα αυτά, από τα οποία προέκυψαν τα αποτελέσματα αυτού του κεφαλαίου, συνοπτική αναφορά του περιεχομένου κάθε άρθρου και καταγραφή του ονόματος του πρώτου συγγραφέα, με αναφορά της χώρας με την οποία σχετίζεται ο ίδιος, καθώς και του τύπου του κάθε άρθρου, ενώ στο τέλος της εν λόγω παρουσίασης αναφέρονται το σύνολο των άρθρων κατ' έτος, κατά χώρα, με την οποία σχετίζεται ο πρώτος συγγραφέας ή, εάν προκύπτει, η έρευνά του και κατά τον τύπο τους. Επίσης, στο ίδιο κεφάλαιο γίνεται μια πρώτη αναφορά στα ζητήματα που απασχόλησαν την επισκοπούμενη βιβλιογραφία, όπως αυτά εντοπίζονταν σταδιακά κατά την ανάγνωση των άρθρων, κατά το αναφερόμενο στη μεθοδολογία 2ο στάδιο της εργασίας, μέσω της οποίας, κατά το 3ο στάδιο της μεθοδολογίας, κατέστη δυνατή η κατηγοριοποίηση των ζητημάτων. Ακολούθως, στο κεφάλαιο 4.3. παρουσιάζεται σε πίνακα η δομή των αποτελεσμάτων, που θα ακολουθήσουν και συγκεκριμένα παρατίθενται τα βασικά ζητήματα-κατηγορίες ζητημάτων, που απασχόλησαν την επισκοπούμενη βιβλιογραφία, με αναφορά των επιμέρους ζητημάτων – υποκατηγοριών ζητημάτων, στα οποία αυτά αναλύονται, όπως η σύνθεση τους έγινε κατά το αναφερόμενο στη μεθοδολογία 3ο στάδιο αυτής. Επίσης, εντός του πίνακα αναγράφονται ενδεικτικά τα άρθρα που αφορούν στο κάθε επιμέρους ζήτημα, με την ένδειξη, Α.1., Α.2., κ.λ.π., με βάση τον αριθμό που αυτά έλαβαν στο κεφάλαιο 4.2. Στο κεφάλαιο 4.4., 4.4.1., 4.4.2. παρουσιάζονται ζητήματα προκαταλήψεων και διακρίσεων στις εν λόγω διαδικασίες, περιορισμού της ανθρώπινης λήψης αποφάσεων, καθώς και ζητήματα "ακρίβειας" και αξιοπιστίας, τα οποία άπτονται του ερωτήματος κατά πόσο αυτές οι διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης είναι δίκαιες. Στο κεφάλαιο 4.5., 4.5.1., 4.5.2., 4.5.3, 4.5.4., ερευνάται το ζήτημα της αδιαφάνειας των διαδικασιών αυτών, καθώς και τα ζητήματα που σχετίζονται με την αδιαφάνεια των αλγορίθμων στο εξεταζόμενο πλαίσιο της ποινικής και πολιτικής δικαιοσύνης, δηλαδή ζητήματα κατανόησης, συμβατότητας με το κράτος δικαίου και τα ανθρώπινα δικαιώματα, με την απαίτηση για αιτιολόγηση των δικαστικών αποφάσεων, ενώ παρουσιάζεται και η έννοια της "επεξηγήσιμης" τεχνητής νοημοσύνης. Συναφής με την "επεξηγήσιμη" τεχνητή νοημοσύνη είναι η παρουσίαση στο κεφάλαιο 4.6. των ζητημάτων που αφορούν στα προσωπικά δεδομένα. Τέλος, στο κεφάλαιο 4.7. παρουσιάζεται το ζήτημα της λογοδοσίας

κατά τις διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης στο πλαίσιο της ποινικής και πολιτικής δικαιοσύνης.

4.2. Προφίλ άρθρων

Από την αναζήτηση στη διεθνή βιβλιογραφία προέκυψαν τα κάτωθι άρθρα από τα οποία έχουν εξαχθεί τα αποτελέσματα αυτού του κεφαλαίου:

A.1.. Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice (2019), Vincent Chiao, Καναδάς, άρθρο επιστημονικού περιοδικού: Ηθικές προεκτάσεις της χρήσης τεχνητής νοημοσύνης, μηχανικής μάθησης, μεγάλων δεδομένων και λογισμικού πρόβλεψης σε περιβάλλοντα ποινικής δικαιοσύνης. Οι ανησυχίες που εγείρονται μπορούν να συγκεντρωθούν κάτω από τον τίτλο :δικαιοσύνη, λογοδοσία και διαφάνεια.

A.2.. Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA (2021), Kathrin Hartmann, Γερμανία, άρθρο επιστημονικού περιοδικού: Πραγματική εφαρμογή και επιπτώσεις των συστημάτων ADM στις διαδικασίες λήψης αποφάσεων. Ποια είναι τα κίνητρα αυτών που αποφασίζουν ως προς τη χρήση ADM στις διαδικασίες λήψης αποφάσεων γενικά και ειδικότερα στον τομέα της ποινικής δικαιοσύνης: αποφυγή αβεβαιότητας, αποφυγή ευθύνης.

A.3.. Algorithms and values in justice and security (2020), Paul Hayes, Ολλανδία, άρθρο επιστημονικού περιοδικού: Οι αξίες που έχουν σχέση με τον σχεδιασμό, την υλοποίηση και την ανάπτυξη αλγορίθμων και ειδικότερα οι αξίες που σχετίζονται με τους αλγόριθμους στον τομέα της δικαιοσύνης και της ασφάλειας και πως αλληλοϋποστηρίζονται ή συγκρούονται μεταξύ τους.

A.4.. Machine learning & forensic science (2019), Giulia Margagliotti, Ελβετία, άρθρο επιστημονικού περιοδικού: Οι δυνατότητες της μηχανικής μάθησης, εποπτευόμενη και μη εποπτευόμενη μηχανική μάθηση, αλγόριθμοι ταξινόμησης και παλινδρόμησης. Χρήση τους σε εγκληματολογικές διαδικασίες. Ζητήματα προκατάληψης των μηχανών και αλγόριθμος COMPAS.

A.5.. Courts and Artificial Intelligence (2020), A. D. (Dory) Reiling, Ολλανδία, άρθρο επιστημονικού περιοδικού: Πως μπορεί να χρησιμεύσει η τεχνητή νοημοσύνη στα δικαστήρια. Όπου το αποτέλεσμα είναι προβλέψιμο οι διαδικασίες μπορούν να αυτοματοποιηθούν.

A.6.. Criminal justice, artificial intelligence systems, and human rights (2020), Aleš Završnik, Σλοβενία, άρθρο επιστημονικού περιοδικού: Αυτοματοποίηση στην ποινική δικαιοσύνη και ανθρώπινα δικαιώματα που θίγονται. Το άρθρο καταλήγει προσφέροντας ορισμένες σκέψεις σχετικά με τις προτεινόμενες λύσεις για την αντιμετώπιση των κινδύνων που ενέχουν τα συστήματα τεχνητής νοημοσύνης στον τομέα της ποινικής δικαιοσύνης.

A.7.. A vulnerability analysis: Theorising the impact of artificial intelligence decision-making processes on individuals, society and human diversity from a social justice perspective (2020), Tetyana (Tanya) Krupiy, Ολλανδία, άρθρο επιστημονικού περιοδικού: Με αναφορά στο πλαίσιο αυτοματοποίησης των διαδικασιών επιλογής μαθητών στα εκπαιδευτικά ιδρύματα με τη χρήση διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης και με μέσο την κοινωνική δικαιοσύνη (επειδή η απαίτηση γι' αυτήν είναι πιο εκτεταμένη απ' ότι η απαίτηση της ισότητας) παρουσιάζεται πως η χρήση των διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης επηρεάζει τις διαδικασίες λήψης αποφάσεων και την κοινωνία. Διατυπώνεται ότι η ανθρώπινη λήψη αποφάσεων πρέπει να διατηρηθεί στους τομείς που η ολιστική αξιολόγηση των ατόμων είναι ζωτικής σημασίας για την αξιολόγηση των ικανοτήτων τους.

A.8.. The judicial demand for explainable artificial intelligence (2019), Ashley Deeks, ΗΠΑ, άρθρο επιστημονικού περιοδικού: Οι δικαστές θα αντιμετωπίσουν μία ποικιλία υποθέσεων στις οποίες θα πρέπει να απαιτούν εξηγήσεις για αλγοριθμικές αποφάσεις, συστάσεις ή προβλέψεις, ενώ θα διαδραμματίσουν θεμελιώδη ρόλο στη διαμόρφωση της επεξηγήσιμης τεχνητής νοημοσύνης. Τρόπος αντιμετώπισης του προβλήματος του «μαύρου κουτιού» είναι ο σχεδιασμός συστημάτων που εξηγούν πως οι αλγόριθμοι καταλήγουν στα συμπεράσματά τους. Επειδή όμως πρέπει να αποκαλυφθεί ο πηγαίος κώδικας, προτείνεται η δημιουργία «υποκατάστατου μοντέλου» που θα αναλύει ζεύγη εισόδου-εξόδου χωρίς πρόσβαση στο εσωτερικό του μοντέλου.

A.9.. Transparency as design publicity: explaining and justifying inscrutable algorithms (2020), Michele Loi, Ελβετία, άρθρο επιστημονικού περιοδικού: Σε σύγκριση με τις εκ των υστέρων επεξηγήσεις μεμονομένων αλγοριθμικών αποφάσεων, η δημοσιότητα σχεδιασμού ανταποκρίνεται σε μια διαφορετική απαίτηση, την απαίτηση για απρόσωπη αιτιολόγηση. Η εξήγηση σχεδιασμού ενός αλγορίθμου περιλαμβάνει την κατανόηση του τι σχεδιάστηκε να κάνει ο αλγόριθμός, πως σχεδιάστηκε να το κάνει και γιατί.

A.10.. An efficient method for image forgery detection based on trigonometric transforms and deep learning (2020), Faten Maher Al_Azrak1, Αίγυπτος, άρθρο επιστημονικού περιοδικού: Αφορά στην ανίχνευση πλαστογραφίας εικόνων που γίνεται με τη μέθοδο copy move και αναλύεται μία μέθοδος εντοπισμού αυτού του είδους της πλαστογραφίας.

A.11.. Appellate Court Modifcations Extraction for Portuguese (2019), William Paulo Ducca Fernandes, Βραζιλία, άρθρο επιστημονικού περιοδικού: Δημιουργώντας ένα σύστημα με τη χρήση δημόσιων δεδομένων από τη βάση δεδομένων του Εφετείου του Ρίο ντε Τζανέιρο, εξάγονται πληροφορίες που αφορούν στις τροποποιήσεις που επέφερε το Εφετείο στις αποφάσεις. Τούτο βοηθάει στο να αυξηθεί η συνέπεια στις δικαστικές αποφάσεις και να περιοριστεί το αίσθημα αδικίας που προκύπτει ότι παρόμοιες υποθέσεις κρίνονται με διαφορετικό τρόπο.

A.12.. Extracting value from Brazilian Court decisions (2022), William Paulo Ducca Fernandes, Βραζιλία, άρθρο επιστημονικού περιοδικού: Ανάπτυξη ενός πληροφοριακού συστήματος για την εξαγωγή πληροφοριών από νομικά κείμενα. Χρησιμοποιούνται ως παράδειγμα αστικές αγωγές σχετικές με την ασφάλιση υγειονομικής περίθαλψης. Το σύστημα βοηθά δικηγόρους και δικαστές στη λήψη αποφάσεων, όπως πως να προβλέψουν το αποτέλεσμα σε συγκεκριμένες υποθέσεις.

A.13.. An overview of information extraction techniques for legal document analysis and processing (2021), Ashwini V. Zadgaonkar, Ινδία, άρθρο επιστημονικού περιοδικού: Μέσω των προσεγγίσεων για την επεξεργασία νομικών κειμένων προκύπτει ότι η KBP σε συνδυασμό με τεχνικές NLP (επεξεργασίας φυσικής γλώσσας) είναι σε θέση να βοηθήσει στην ανάλυση νομικού κειμένου για πολλές εργασίες, όπως η αυτόματη σύνοψη, η ταξινόμηση νομικών κειμένων κ.λ.π..

A.14.. Deep learning based algorithm (ConvLSTM) for Copy Move Forgery Detection (2021), Mohamed A. Elaskilya, Αίγυπτος, άρθρο επιστημονικού περιοδικού: Απεικονίζεται ένας νέος αλγόριθμος για την ανίχνευση copy move πλαστογραφίας. Το προτεινόμενο μοντέλο εξαρτάται από τη χρήση CNN και ConvLSTM δικτύων.

A.15.. Application of multiple BERT model in construction litigation (2020), Che-Wen Chen, Ταϊβάν, άρθρο συνεδρίου: Μελέτη που χρησιμοποιεί τεχνολογία βαθιάς μάθησης που βασίζεται σε μοντέλο BERT για να προτείνει ένα σύστημα πρόβλεψης. Το σύστημα έχει 2 λειτουργίες, το να βρίσκει παρόμοιες περιπτώσεις και να προβλέπει δικαστικές αποφάσεις.

A.16..A Deep Learning Method for Judicial Decision Support (2019), Baogui Chen, Κίνα, άρθρο συνεδρίου: Με βάση τα γεγονότα μιας υπόθεσης όπως προκύπτουν από δικαστικά έγγραφα, εφαρμόζεται ένα μοντέλο βαθιάς μάθησης για να προβλεφθεί μια απόφαση και συγκεκριμένα η ποινή, η κατηγορία και οι νομικές διατάξεις.

A.17.. How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence (2018), Haoxi Zhong, Κίνα, άρθρο επιστημονικού περιοδικού: Οι περισσότερες εργασίες LegalAI βασίζονται σε τεχνολογίες επεξεργασίας φυσικής γλώσσας (NLP). Οι ερευνητές εξερευνούν λύσεις που βασίζονται σε NLP σε μια ποικιλία εργασιών LegalAI, όπως η πρόβλεψη δικαστικών αποφάσεων. Τα ερμηνεύσιμα μοντέλα δεν είναι αποτελεσματικά, ενώ αυτά με την καλύτερη απόδοση συνήθως δεν μπορούν να ερμηνευτούν, γεγονός που μπορεί να οδηγήσει σε προκαταλήψεις και φυλετικές διακρίσεις.

A.18.. Preserving the rule of law in the era of artifcial intelligence (AI) (2021), Stanley Greenstein, Σουηδία, άρθρο επιστημονικού περιοδικού: Καταδεικνύεται πως το κράτος δικαίου απειλείται από τα συστήματα λήψης αποφάσεων τεχνητής νοημοσύνης, επισημαίνοντας την αδιαφάνεια των

σχετικών τεχνολογιών και τονίζοντας τα χαρακτηριστικά του κράτους δικαίου που έρχονται σε αντίθεση με τα χαρακτηριστικά των μοντέλων τεχνητής νοημοσύνης.

Α.19.. Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making (2020), Carolyn McKay, Αυστραλία, άρθρο επιστημονικού πεδιοδικου: Επιχειρείται μία σύγκριση των εκτιμήσεων κινδύνου στα διάφορα στάδια μιας ποινικής διαδικασίας, που γίνεται στα πλαίσια της παραδοσιακής απονομής της δικαιοσύνης που στηρίζεται στη διακριτική ευχέρεια των δικαστών και την εξατομικευμένη δικαιοσύνη και σ' αυτή που γίνεται με τη χρήση μοντέλων τεχνητής νοημοσύνης. Με βάση τον ιδιόκτητο χαρακτήρα αυτών των μοντέλων καταδεικνύεται πως η εκτίμηση κινδύνου μπορεί να είναι αδιαφανής τόσο για τον κατηγορούμενο όσο και για τον δικαστή.

Α.20.. Judicial analytics and the great transformation of American Law (2019), Daniel L. Chen, Γαλλία, άρθρο επιστημονικού περιοδικού: Το άρθρο με αναφορά στην πρόβλεψη δικαστικών αποφάσεων τονίζει ότι η χρήση μηχανικής μάθησης μπορεί να εντοπίσει προκαταλήψεις και παράγοντες που επηρεάζουν τις δικαστικές αποφάσεις, καθώς επίσης και ότι μπορεί να χρησιμοποιηθεί για την αξιολόγηση των επιπτώσεων των αποφάσεων.

Α.21.. IS AUSTRALIA READY FOR AI ON THE BENCH? (2020) Bolingford, Ilana, Αυστραλία, άρθρο επιστημονικού περιοδικού: Αντιμετωπίζεται το ζήτημα της ομαλής ενσωμάτωσης τεχνητής νοημοσύνης στο δικαστικό σύστημα της Αυστραλίας, απαντώντας στα ερωτήματα πως η τεχνητή νοημοσύνη θα επηρεάσει τη λήψη δικαστικών αποφάσεων και πως οι δικαστές στην Αυστραλία θα αποδεχτούν τη χρήση συστημάτων τεχνητής νοημοσύνης.

Α.22.. Artificial Intelligence and Legal Decision-Making: The Wide Open? (2019), Maxi SCHERER, Αγγλία, άρθρο επιστημονικού περιοδικού: Εξετάζονται διάφορες ανησυχίες που προκύπτουν από τη χρήση τεχνητής νοημοσύνης στη λήψη δικαστικών αποφάσεων, όπως το ενδεχόμενο τα μοντέλα να μην μπορούν να προσαρμοστούν για να αντιμετωπίσουν αλλαγές, αλλά να μένουν προσκολλημένα σε συντηρητικές προσεγγίσεις. Επισημαίνεται ότι είναι σημαντικό να κατανοήσουμε τις τεχνικές πτυχές της τεχνητής νοημοσύνης για να αξιολογήσουμε τις συνέπειες που θα επιφέρει στη λήψη αποφάσεων, ενώ οι συνέπειες αυτές πρέπει να εξεταστούν πριν τη χρήση της τεχνητής νοημοσύνης στα δικαστικά περιβάλλοντα.

Α.23.. Paths to Digital Justice: Judicial Robots, Algorithmic Decision-Making, and Due Process (2020), Pedro RUBIM BORGES FORTES, Βραζιλία, άρθρο επιστημονικού περιοδικού : Σκέψεις σχετικά με τις δυνατότητες που προσφέρει η τεχνολογία της πληροφορίας, τα μεγάλα δεδομένα και η αλγοριθμική λήψη αποφάσεων στο δρόμο για την ψηφιακή δικαιοσύνη. Πώς δηλαδή οι αλγόριθμοι μπορούν να υποστηρίξουν τη λήψη νομικών αποφάσεων και αν η αλγοριθμική λήψη αποφάσεων θα μπορούσε τελικά να υποκαταστήσει τη δικαστική λήψη αποφάσεων. Αναφορά στο εργαλείο πρόβλεψης κινδύνου Compas. Δυνατότητες

και περιορισμοί στο δρόμο προς την ψηφιακή δικαιοσύνη. Στο δρόμο αυτό, εκτός από τη μαθηματική λογική των αλγορίθμων, σημαντικές θα είναι και οι κοινωνικές μας εμπειρίες, καθώς και ο τρόπος με τον οποίο θα συμβιβάζουμε την αποτελεσματικότητα και την ακρίβεια της αλγοριθμικής λήψης αποφάσεων με τις συνταγματικές εγγυήσεις της ορθής διαδικασίας.

Α.24.. JUDGE V ROBOT? ARTIFICIAL INTELLIGENCE AND JUDICIAL DECISION-MAKING (2018), TANIA SOURDIN, Ηνωμένο Βασίλειο, άρθρο επιστημονικού περιοδικού: Η απάντηση στο ερώτημα πώς η τεχνητή νοημοσύνη θα διαμορφώσει τον ρόλο του δικαστή και αν ο ρόλος αυτός θα παραμείνει ανθρώπινη δραστηριότητα, τουλάχιστον αναφορικά με ορισμένες κατηγορίες διαφορών, δεν φαίνεται να είναι ότι αυτή θα αναμορφώσει πλήρως τη δικαστική λειτουργία και τον ρόλο των δικαστών. Μέσα από την ανάλυση των ζητημάτων που εγείρει η τεχνητή νοημοσύνη στον συγκεκριμένο τομέα, είναι σημαντικό να γίνει κατανοητό ότι οι δικαστές κάνουν πολλά περισσότερα από το να παίρνουν αποφάσεις. Η αποδοχή της τεχνητής νοημοσύνης θα περάσει μέσα από την εξέταση ηθικών ερωτημάτων και ερωτημάτων σχετικών με το ποιος θα παράγει τις εν λόγω τεχνολογίες και τον βαθμό στον οποίο θα διατηρηθεί η διακριτική ευχέρεια των δικαστών. Αυτό που υποστηρίζεται είναι ότι οι δικαστές θα συνεχίσουν να είναι άνθρωποι, αλλά θα συμπληρώνονται, θα υποστηρίζονται από εργαλεία τεχνητής νοημοσύνης.

Α.25.. On the relevance of algorithmic decision predictors for judicial decision making (2021), Floris Bex, Ολλανδία, άρθρο συνεδρίου: Αλγόριθμοι πρόβλεψης αποφάσεων . Κύριοι τύποι αυτού του είδους αλγορίθμων. Πώς και με ποιες προϋποθέσεις μια πρόβλεψη σε μια συγκεκριμένη υπόθεση αποδίδει μία πιθανότητα απόφασης για μια νέα υπόθεση;

Α.26.. Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models (2022), Merylin Monaro, Ιταλία, άρθρο επιστημονικού περιοδικού = Δυνατότητα εντοπισμού ψεύδους με βάση τις μικροεκφράσεις του προσώπου και τεχνικές μηχανικής μάθησης. Οι κινήσεις ορισμένων μυών του προσώπου είναι ως επί το πλείστον ακούσιες και συνεπώς είναι δύσκολο να ελέγχονται εσκεμμένα. Η τεχνητή νοημοσύνη αποδίδει καλύτερα από τους ανθρώπους στην ανίχνευση ψεύδους.

Α.27.. Joint cognition of both human and machine for predicting criminal punishment in judicial system, 2019, A.K. Das, Μπαγκλαντέζ, άρθρο συνεδρίου = Ο συνδυασμός τεχνητών νευρωνικών δικτύων και πραγματικών ανθρώπων δίνει καλύτερες προβλέψεις. Επειδή οι μηχανές κάνουν το χρονοβόρο μέρος μιας δουλειάς, ένας άνθρωπος μπορεί να έχει περισσότερο χρόνο για να επικεντρωθεί σε άλλα ζητήματα. Οι άνθρωποι είναι πάντα καλύτεροι στην επίλυση νέων και διφορούμενων καταστάσεων. Μπορούν να λάβουν ηθικές αποφάσεις που ξεπερνούν τις δυνατότητες μιας μηχανής. Από την άλλη μεριά, η μηχανή μπορεί

να ανταποκριθεί ταχύτερα από τον άνθρωπο όταν η κατάσταση είναι γνωστή και βρίσκεται εντός των γνωστικών της δυνατοτήτων.

Α.28.. Machine learning and legal argument (2021), Jack Mumford, Ηνωμένο Βασίλειο, άρθρο επιστημονικού περιοδικού: Τρόποι με τους οποίους οι προσεγγίσεις της Μηχανικής Μάθησης μπορούν να συμβάλουν στη νομική επιχειρηματολογία. Τίθεται το ερώτημα αν η μηχανική μάθηση μπορεί να προσδιορίσει τα στοιχεία που απαιτούνται για τη δημιουργία μιας αιτιολόγησης: ζητήματα και παράγοντες. Όσον αφορά μεμονωμένες περιπτώσεις, η πρόβλεψη ενός αποτελέσματος δεν αρκεί: απαιτείται μια αιτιολόγηση, διατυπωμένη με νομικούς όρους, είτε για να υποστηρίξει έναν δικαστή που λαμβάνει την απόφαση είτε για να εκθέσει το σκεπτικό στο κοινό με αποδεκτό τρόπο, δεδομένου ότι ο νόμος προβλέπει το δικαίωμα για εξήγηση. Ανάπτυξη ενός αποτελεσματικού συστήματος ML που μπορεί να αποδώσει παράγοντες με την κατάλληλη πιστότητα και αιτιολόγηση. Εάν επιτύχει, θα εφαρμοσθεί η ίδια διαδικασία στο μεγαλύτερο πρόβλημα της εκμάθησης της δομής του τομέα, με σκοπό να παραχθεί υψηλή πιστότητα και αιτιολογημένα αποτελέσματα απευθείας από τα γεγονότα της υπόθεσης.

Α.29.. Explainability of formal models of argumentation applied to legal domain (2019), Michał ARASZKIEWICZ, Πολωνία, άρθρο επιστημονικού περιοδικού: Διερεύνηση της δυνατότητας αύξησης της επεξηγηματικότητας των αποφάσεων που βασίζονται σε τεχνητή νοημοσύνη μέσω υπολογιστικής επιχειρηματολογίας. Τα πλαίσια επιχειρημάτων (Argumentation frameworks) μπορούν κατ' αρχήν να χρησιμοποιηθούν ως εργαλεία επεξήγησης της λειτουργίας των συστημάτων τεχνητής νοημοσύνης. Ωστόσο και τα ίδια (AF), ως μοντέλα που αναπτύχθηκαν σε ένα σύνθετο και ταχέως εξελισσόμενο πεδίο έρευνας, θα πρέπει να ελέγχονται ως προς τη διαφάνεια και την επεξήγηση τους.

Α.30.. Criminal justice profiling and EU data protection law: precarious protection from predictive policing (2019), Orla Lynskey, Αγγλία, άρθρο επιστημονικού περιοδικού: Αβεβαιότητα ως προς το εάν η νομοθεσία προστασίας προσωπικών δεδομένων εφαρμόζεται στην περίπτωση των τεχνολογιών προγνωστικής αστυνόμευσης. Και στην περίπτωση που εφαρμόζεται, η προστασία που η νομοθεσία αυτή παρέχει μπορεί να μη βοηθήσει ιδιαίτερα όσους επηρεάζονται από αυτές τις τεχνολογίες, καθώς η ουσιαστική προστασία που προσφέρει σε περίπτωση αυτοματοποιημένης λήψης αποφάσεων υπόκειται σε μια σειρά περιορισμών. Νομικό πλαίσιο της επεξεργασίας προσωπικών δεδομένων στο πλαίσιο της ποινικής δικαιοσύνης. Διαφορετικό επίπεδο προστασίας της Οδηγίας για την επιβολή του νόμου και του GDPR. Κατά πόσο η διαφοροποίηση αυτή είναι συμβατή με τον Χάρτη. Τα είδη (2) των αποφάσεων, για τη λήψη των οποίων χρησιμοποιούνται τέτοιες τεχνολογίες και κατά πόσο αυτές οι δύο μορφές αυτοματοποιημένης λήψης αποφάσεων (που σχετίζονται με τη συστημική και την προγνωστική αστυνόμευση ταυτοποίησης) συνιστούν επεξεργασία προσωπικών δεδομένων και επομένως εμπίπτουν στο πεδίο εφαρμογής των διατάξεων.

ΔΕΕ. Υπόθεση Nowak. Εφαρμογή PredPol. Εφαρμογή με την οποία η αστυνομία κάνει προβλέψεις για πιθανές εστίες εγκλήματος, μελλοντικά. Εφαρμογή HART στο Ηνωμένο Βασίλειο.

A.31.. JUSTICE IN THE DIGITAL AGE: TECHNOLOGICAL SOLUTIONS, HIDDEN THREATS AND ENTICING OPPORTUNITIES (2021), Yulia Razmetaeva, Ουκρανία, άρθρο επιστημονικού περιοδικού: Οι αλλαγές που συντελούνται στον τομέα της δικαιοσύνης από τη χρήση τεχνολογικών λύσεων πρέπει να αξιολογηθούν με βάση την ισορροπία των κινδύνων και των ευκαιριών που προσφέρουν οι νέες τεχνολογίες. Σημαντικοί κίνδυνοι είναι οι παραβιάσεις της ασφάλειας και του απορρήτου και η διάβρωση της ιδιωτικής ζωής, που μπορούν να προκληθούν από παραβιάσεις ασφαλείας λόγω παραβίασης συσκευών, έλλειψης ασφάλειας των δεδομένων που είναι αποθηκευμένα στους διακομιστές των δικαστηρίων ή μεταδίδονται μέσω e-mail από τους συμμετέχοντες στις δικαστικές διαδικασίες, εγκατάστασης κακόβουλου λογισμικού και απόκτησης μη εξουσιοδοτημένης πρόσβασης σε συστήματα και δεδομένα. Υπόθεση ΕΔΔΑ. S and Marper v. U.K., για την επ' αόριστον αποθήκευση, στα πλαίσια της ποινικής διαδικασίας, των δακτυλικών αποτυπωμάτων, των δειγμάτων κυττάρων και του DNA. Ένα άλλο είδος απειλών είναι η υπονόμευση του κράτους δικαίου και των δικαιωμάτων του ανθρώπου. Η ακρίβεια των αλγορίθμων έρχεται σε σύγκρουση με τη διαφαινόμενη μεροληψία αυτών, είτε εκούσια είτε ακούσια, που μπορεί να οδηγήσει σε διακρίσεις.

A.32.. Scalable and explainable legal prediction (2021), L. Karl Branting, ΗΠΑ, άρθρο επιστημονικού περιοδικού: Μορφές υποστήριξης νομικών αποφάσεων. Επεξηγήσιμα συστήματα πρόβλεψης νομικών αποφάσεων. Συστήματα που βασίζονται αποκλειστικά στη μηχανική μάθηση, τα οποία είναι λίγο ή καθόλου επεξηγήσιμα. Ο στόχος της έρευνας που περιγράφεται σε αυτό το έγγραφο είναι η ανάπτυξη τεχνικών για εξηγήσιμη πρόβλεψη χωρίς μεγάλα κόστη, καθώς οι λόγοι που κινούν τους οργανισμούς να εξετάσουν τα συστήματα υποστήριξης αποφάσεων, αποκλείουν λύσεις που απαιτούν μεγάλες προσπάθειες ανάπτυξης. Οι (2) προσεγγίσεις που εξετάζονται δεν απαιτούν από τους χρήστες να δίνουν τιμές, αλλά ως είσοδο δέχονται κείμενο με περιγραφή των γεγονότων της υπόθεσης. Και οι δύο προσεγγίσεις αποσκοπούν στο να πλησιάσουν πιο κοντά σε ένα ιδανικό σύστημα, που συνδυάζει υψηλή επεξηγηματική ικανότητα με χαμηλή προσπάθεια μηχανικής γνώσης.

A.33.. Explainable AI under contract and tort law: legal incentives and technical challenges (2020), Philipp Hacker, Γερμανία, άρθρο επιστημονικού περιοδικού: Το ζήτημα της επεξήγησης των αποφάσεων που λαμβάνονται με τη χρήση εργαλείων μηχανικής μάθησης είναι βασικό όσον αφορά στη σχέση της τεχνητής νοημοσύνης και του νόμου. Στο άρθρο υποστηρίζεται ότι η επεξήγηση είναι σημαντική όχι μόνο στο δίκαιο προστασίας δεδομένων, αλλά και στο δίκαιο των συμβάσεων και των αδικοπραξιών και ότι, ενόψει του ότι οι απαιτήσεις που θέτει ο GDPR είναι σε μεγάλο βαθμό ασαφείς αυτή τη στιγμή, προτείνεται η

δυνατότητα της επιβολής της χρήσης εξηγήσιμων μοντέλων μηχανικής μάθησης μέσω του νόμου των συμβάσεων και των αδικοπραξιών. Σχέση επεξηγήσιμων μοντέλων και ακρίβειας. Εμπειρική διερεύνηση του συμβιβασμού (επεξήγησης – ακρίβειας) μέσα από ένα παράδειγμα σχετικό με την ταξινόμηση ανεπιθύμητων μηνυμάτων.

A.34..Explainable Artificial Intelligence and Machine Learning: A reality rooted perspective (2020), Frank Emmert-Streib, Φιλανδία, άρθρο επιστημονικού περιοδικού: Τι μπορεί να είναι η εξηγήσιμη τεχνητή νοημοσύνη, προσεγγίζοντάς την από μια προοπτική που βασίζεται στην πραγματικότητα και όχι ως ευσεβή πόθο. Περιορισμοί της εξηγήσιμης τεχνητής νοημοσύνης αναφορικά με τους εφικτούς στόχους. Με τις μεθόδους τεχνητής νοημοσύνης και μηχανικής μάθησης μπορούν να απαντηθούν ερωτήσεις που οι απαντήσεις τους μπορούν να βρεθούν μέσα στα δεδομένα. Κάθε έλλειψη ποιοτικών δεδομένων, π.χ. λόγω περιορισμένου μεγέθους δείγματος, μεταφράζεται άμεσα σε έλλειψη απαντήσεων, πράγμα που αποτελεί εγγενή αβεβαιότητα οποιουδήποτε συστήματος τεχνητής νοημοσύνης. Με την ποσοτική αξιολόγηση των ομοιοτήτων ή των διαφορών μεταξύ δύο συστημάτων τεχνητής νοημοσύνης, μπορούμε να συγκρίνουμε ένα εξηγήσιμο με ένα μη εξηγήσιμο σύστημα, για να αξιολογήσουμε το όφελος του ενός έναντι του άλλου. Ακόμα και αν ένα εξηγήσιμο σύστημα δεν επιλύει πλήρως ένα δεδομένο πρόβλημα, σε σχέση με ένα σύστημα βαθιάς μάθησης, μπορεί να αρκεί η χρήση αυτού (του εξηγήσιμου).

Όπως προκύπτει από τα ως άνω αποτελέσματα, τα άρθρα που χρησιμοποιήθηκαν ανά έτος είναι : 2 άρθρα του έτους 2018, 10 άρθρα του 2019, 12 άρθρα του 2020, 8 άρθρα του 2021 και 2 άρθρα του 2022. Επίσης, 30 από αυτά είναι άρθρα επιστημονικού περιοδικού και 4 είναι άρθρα συνεδρίου. Επιπλέον, 1 συγγραφέας σχετίζεται με τον Καναδά, 2 συγγραφείς σχετίζονται με τη Γερμανία, 4 με την Ολλανδία, 2 με την Ελβετία, 1 με την Σλοβενία, 2 με τις ΗΠΑ, 2 με την Αίγυπτο, 3 με την Βραζιλία, 1 με την Ινδία, 1 με την Ταϊβάν, 2 με την Κίνα, 1 με τη Σουηδία, 2 με την Αυστραλία, 1 με τη Γαλλία, 4 με το Ηνωμένο Βασίλειο, 1 με την Ιταλία, 1 με το Μπαγκλαντέζ, 1 με την Πολωνία, 1 με την Ουκρανία και 1 με τη Φιλανδία. Τα όσα, ωστόσο, οι συγγραφείς αναπτύσσουν στα άρθρα αυτά δεν αναφέρονται αποκλειστικά στις χώρες με τις οποίες αυτοί σχετίζονται, εκτός από τα άρθρα Α.11. και Α.12. που αφορούν στη Βραζιλία και το άρθρο Α.15. με αναφορά στις δικαστικές διαδικασίες στην Κίνα. Χαρακτηριστική είναι η αναφορά, στην πλειονότητα των άρθρων στην εφαρμογή COMPAS και στην απόφαση State of Wisconsin v Loomis, δηλαδή σε μία εφαρμογή που χρησιμοποιείται στις ΗΠΑ και απασχόλησε αμερικάνικο δικαστήριο.

Η χρήση της μηχανικής μάθησης έχει προκαλέσει μια σειρά από ανησυχίες, ειδικά όταν τα συστήματα κάνουν προβλέψεις που επηρεάζουν την ελευθερία, την ασφάλεια ή το απόρρητο των ανθρώπων. Ένα σκέλος της κριτικής επικεντρώνεται στους τρόπους με τους οποίους αυτοί οι αλγόριθμοι μπορούν να αντιγράψουν και να επιδεινώσουν τις κοινωνικές προκαταλήψεις υπό το φως των δεδομένων στα οποία τους εκπαιδεύουν οι επιστήμονες. Μια άλλη σειρά

κριτικών αμφισβητεί την ακρίβεια των διαφόρων προβλέψεων μηχανικής μάθησης, με τους αντιρρησίες να ισχυρίζονται ότι εργαλεία όπως οι αλγόριθμοι ποινικής δικαιοσύνης προβλέπουν την υποτροπή με μικρότερη ακρίβεια από τους ανθρώπους (Deeks, 2019). Εξάλλου, μία από τις κύριες ηθικές ανησυχίες, όσον αφορά τους αλγόριθμους είναι η πιθανή αδιαφάνειά τους, που μπορεί να οδηγήσει σε τρία είδη ηθικών προβλημάτων: α) Μπορεί να οδηγήσει στη λήψη αποφάσεων που στερούνται επεξηγηματικότητας και ως εκ τούτου στερούνται σαφούς αιτιολόγησης, β) μπορεί να οδηγήσει σε έλλειψη ευθύνης και λογοδοσίας και γ) μπορεί να οδηγήσει σε κακές και λανθασμένες αποφάσεις. Όπως όμως υποστηρίζουν οι Hayes et al. (2020), με τον τρόπο αυτόν μπορεί να προκύπτουν σοβαρές επιπτώσεις σε βασικές αξίες, οι οποίες προσδιορίζονται ως σχετιζόμενες με τους αλγόριθμους στον εξεταζόμενο τομέα της δικαιοσύνης και της ασφάλειας: διαφάνεια, λογοδοσία, αυτονομία, ιδιοκτησία, ακρίβεια κ.λ.π., ενώ, επιπλέον, υποστηρίζουν ότι και από τις σχετιζόμενες αξίες, (ακρίβεια, αυτονομία, ιδιωτικότητα, δικαιοσύνη/ισότητα, ιδιοκτησία, λογοδοσία, διαφάνεια), οι οποίες επιλέγονται από έναν κατάλογο αξιών ως οι πιο σημαντικές για τον τομέα της δικαιοσύνης και της ασφάλειας, μπορεί να προκύψουν σημαντικοί κίνδυνοι, καθώς αυτές αλληλεπιδρούν μεταξύ τους (για παράδειγμα η ιδιοκτησία μπορεί να είναι επιβλαβής για την διαφάνεια). Επιπλέον, το ζήτημα της αδιαφάνειας αναφέρεται στη βιβλιογραφία ως το πρόβλημα του «μαύρου κουτιού» και αφορά στην έλλειψη πληροφοριών σχετικά με το πώς ο αλγόριθμος φτάνει στα αποτελέσματά του, γεγονός που μπορεί να υπονομεύσει το αίσθημα δικαιοσύνης και εμπιστοσύνης των ανθρώπων και στο πλαίσιο της ποινικής δικαιοσύνης μπορεί να υπονομεύσει το δικαίωμα υπεράσπισης ενός κατηγορουμένου, συνδέεται δε με την αναφερόμενη στη βιβλιογραφία ως επεξηγήσιμη τεχνητή νοημοσύνη (Deeks, 2019). Υποστηρίζει, μάλιστα, η Deeks (2019) ότι «το να διαφωτιστεί ο τρόπος με τον οποίο ένας αλγόριθμος παράγει τις συστάσεις του μπορεί να βοηθήσει στην αντιμετώπιση των άλλων δύο ανησυχιών, επιτρέποντας στους παρατηρητές να εντοπίσουν προκαταλήψεις και λάθη στον αλγόριθμο».

4.3. Πίνακας για την κατανόηση της δομής των αποτελεσμάτων που αφορούν στα ζητήματα που προκύπτουν κατά τις διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης:

Ζητήματα δικαιοσύνης:	Ζητήματα προκαταλήψεων και διακρίσεων στις διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης: Α.1., Α.3., Α.5., Α.8., Α.18., Α.19., Α.20., Α.22., Α.23., Α.27., Α.28., Α.31. Ζητήματα εξατομικευμένης δικαιοσύνης και διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης: Α.1., Α.19., Α.24.

	Η ακρίβεια κατά τη διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης: Α.3., Α.8., Α.9., Α.19., Α.20., Α.27., Α.31.
Ζητήματα διαφάνειας:	Αδιαφάνεια και ανάγκη για κατανόηση των διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης. Ιδιόκτητη φύση των συστημάτων τεχνητής νοημοσύνης: Α.1., Α.3., Α.6., Α.18, Α.19., Α.23, Α.33. Αδιαφάνεια και δικαιώματα του ανθρώπου: Α.5., Α.6., Α.18., Α.19., Α.31. Αδιαφάνεια και αιτιολόγηση αποφάσεων: Α.5., Α.22., Α.28. Επεξηγήσιμη τεχνητή νοημοσύνη: Α.8., Α.9., Α.29., Α.32., Α.33.
Προσωπικά δεδομένα:	Προσωπικά δεδομένα και διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης. Δικαίωμα στην εξήγηση: Α.1., Α.6., Α.18., Α.30., Α.33.
Λογοδοσία:	Λογοδοσία κατά τη διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης. Ευθύνη και Έλεγχος: Α.1., Α.2., Α.3., Α.5., Α.19., Α.23.

4.4. Ζητήματα προκαταλήψεων και διακρίσεων στις διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης.

Ίσως η πιο συχνά συζητούμενη ένσταση, που έχει προβληθεί σε σχέση με τη διαδικασία λήψης αποφάσεων, η οποία σχετίζεται με τη δικαιοσύνη, είναι ότι οι προγνωστικοί παράγοντες μπορεί να είναι αδίκως προκατειλημμένοι (Chiao, 2019). Χρησιμοποιώντας ως παράδειγμα προγνωστικών παραγόντων τις προηγούμενες συλλήψεις των μελών δύο ομάδων, της Α και της Β, λαμβάνουμε ως δεδομένη την ύπαρξη δύο περιπτώσεων: Στην πρώτη, η ομάδα Α έχει υψηλότερο ποσοστό σύλληψης επειδή τα μέλη της τείνουν να επιδεικνύουν πιο επικίνδυνη, επιθετική συμπεριφορά από τα μέλη της ομάδας Β για λόγους που δεν αντικατοπτρίζουν μεροληψία ή διακρίσεις. Στη δεύτερη περίπτωση, η διαφορά οφείλεται στο ότι τα μέλη της ομάδας Α παρακολουθούνται πιο επίμονα από τις αρχές και όχι στο ότι υπάρχει οποιαδήποτε διαφορά στα βασικά ποσοστά ποινικών αδικημάτων. Θα μπορούσε κανείς να θεωρήσει ότι τα μέλη της ομάδας Α φέρουν άδικα βάρος στη δεύτερη περίπτωση, αλλά όχι στην πρώτη. Κατά συνέπεια, θα μπορούσε κανείς να θεωρήσει τη χρήση των ποσοστών σύλληψης, ως προγνωστικό εγκληματικότητας, ως άδικη στη δεύτερη περίπτωση, ενώ δεν θα ήταν άδικη στην πρώτη περίπτωση, όπου η εξήγηση για το υψηλότερο ποσοστό σύλληψης μεταξύ της ομάδας Α σε σύγκριση για την ομάδα Β είναι εξωγενής ως

προς την αστυνομική δραστηριότητα. Ίσως κάποιες διαφορές στην ποινική δικαιοσύνη μοιάζουν με την πρώτη περίπτωση. Οι διαφορές μεταξύ των φυλών, για παράδειγμα, φαίνεται να παρακολουθούν πραγματικές διαφορές στη συμμετοχή στο έγκλημα. Άλλες όμως διαφορές φαίνεται να μοιάζουν με τη δεύτερη περίπτωση. Ένα χαρακτηριστικό παράδειγμα είναι το έγκλημα ναρκωτικών. Τα στοιχεία της έρευνας υποδηλώνουν παρόμοια ποσοστά χρήσης ναρκωτικών ουσιών μεταξύ των μαύρων και λευκών, αν και οι μαύροι χρήστες ναρκωτικών είναι πιο πιθανό να έχουν επαφή με την ποινική δικαιοσύνη από ότι οι λευκοί χρήστες ναρκωτικών. Μια εξήγηση για αυτό το φαινόμενο είναι ότι η αστυνομία έχει δώσει προτεραιότητα σε ενέργειες επιβολής του νόμου σε υπαίθριες αγορές ναρκωτικών, που χρησιμοποιούνται κυρίως από Αφροαμερικανούς, αντί για συναλλαγές εντός κατοικιών που προτιμώνται από τους Λευκούς. Μιλώντας, για "αυτοεκπληρούμενες προσδοκίες", ο Chiao, σημειώνει ότι εάν οι υπεύθυνοι χάραξης πολιτικής αναμένουν εκ των προτέρων να βρουν περισσότερα εγκλήματα μεταξύ της ομάδας Α από την ομάδα Β, τότε είναι πιθανό ότι θα βρουν αυτή την προσδοκία επικυρωμένη εκ των υστέρων, αλλά μόνο επειδή έχουν αφιερώσει περισσότερο χρόνο αναζητώντας έγκλημα μεταξύ των μελών της ομάδας Α παρά μεταξύ των μελών της ομάδας Β. Τονίζει, ωστόσο, ότι ενώ οι αυτοεκπληρούμενες προφητείες είναι ξεκάθαρα προβληματικές, δεν περιορίζονται στην υπολογιστική εκτίμηση κινδύνου. Είναι ουσιαστικά ενσωματωμένες στην ποινική διαδικασία. Έτσι, όμως, το ερώτημα δεν είναι αν θα γίνει αξιολόγηση κινδύνου, αλλά πώς: μέσω προβλέψεων βασισμένων σε στατιστικές συσχετίσεις σε μεγάλα σύνολα δεδομένων ή μέσω της κλινικής κρίσης και της διαίσθησης δικαστών και εισαγγελέων. Δεδομένου ότι είναι μάλλον εύλογο ότι οι άνθρωποι που παίρνουν αποφάσεις λαμβάνουν υπόψη πολλούς από τους ίδιους παράγοντες, όπως το ιστορικό συλλήψεων και καταδικών, οι οποίοι σχετίζονται με τη φυλετική ταυτότητα, οι αυτοεκπληρούμενες προφητείες είναι πιθανό να αποτελούν αναπόφευκτο πρόβλημα στην ποινική δικαιοσύνη. Επιπλέον, η διαδικασία ποινικής δικαιοσύνης είναι ένα διαδοχικά διατεταγμένο σύνολο αποφάσεων - περιπολία, έρευνα, σύλληψη, κατηγορία, εγγύηση, ένσταση, δίκη, ποινή, αποφυλάκιση, χάρη –και φαίνεται ότι γενικά οι αποφάσεις στα προηγούμενα στάδια ασκούν σημαντική επιρροή στις αποφάσεις στα επόμενα στάδια. Η σύνθετη φυλετική προκατάληψη είναι ένα δομικό πρόβλημα στην ποινική δικαιοσύνη, ανεξάρτητα από το αν χρησιμοποιούμε δομημένους αλγόριθμους ή ανθρώπινη κρίση σε ένα δεδομένο στάδιο της διαδικασίας. Κατά συνέπεια, το ότι η αλγοριθμική αξιολόγηση κινδύνου είναι πιθανό να επιφέρει φυλετικές διακρίσεις δεν αποτελεί, από μόνο του, επαρκή λόγο για την απόρριψη τέτοιων εργαλείων. Οποιοσδήποτε τρόπος απόφασης είναι πιθανό να κάνει το ίδιο. Το ερώτημα είναι απλώς εάν η αλγοριθμική αξιολόγηση κινδύνου μπορεί να δημιουργήσει μια πιο ελκυστική αντιστάθμιση μεταξύ της ακρίβειας στους νόμιμους στόχους της ποινικής δικαιοσύνης και της εισαγωγής περαιτέρω φυλετικής παραμόρφωσης σε ένα δεδομένο σημείο της ποινικής διαδικασίας. Εξάλλου, αν και υπάρχουν ενδείξεις ότι η αλγοριθμική αξιολόγηση κινδύνου μπορεί σε ορισμένα πλαίσια να το επιτύχει, δεν ισχυρίζεται ότι αυτό θα συμβεί αναπόφευκτα. Ισχυρίζεται μόνο ότι υπάρχει λόγος να προτιμάται όποιος τρόπος λήψης αποφάσεων, είτε με ανθρώπινη κρίση είτε με προγνωστικό αλγόριθμο, επιτυγχάνει τον πιο ελκυστικό

συμβιβασμό μεταξύ ακρίβειας και μεροληψίας (Chiao, 2019). Εξάλλου, οι μη ανθρώπινες αυτοματοποιημένες διαδικασίες λήψης αποφάσεων έχουν σκοπό να κάνουν τέτοιες αξιολογήσεις πιο συνεπείς, έγκαιρες, οικονομικά αποδοτικές και ακριβείς (McKay, 2020).

Περαιτέρω, καταδεικνύεται, ότι η έννοια της μεροληψίας είναι μια εγγενής πτυχή της επιστήμης δεδομένων και επομένως των τεχνολογιών τεχνητής νοημοσύνης. Με άλλα λόγια, ο χειρισμός δεδομένων φέρνει αυτόματα μαζί του προκατάληψη. Η πράξη της επιλογής ενός συνόλου δεδομένων έναντι ενός άλλου θα αντανακλά ενδεχομένως μια συγκεκριμένη προκατάληψη. Η προκατάληψη μπορεί να είναι τόσο σκόπιμη όσο και ακούσια και ένας εμπειρικός κανόνας θα πρέπει πάντα να είναι ότι ένα σύνολο δεδομένων ενσωματώνει κάποιο βαθμό μεροληψίας. Η μεροληψία είναι παρούσα σε όλα σχεδόν τα σύνολα δεδομένων και τα μεροληπτικά δεδομένα θα οδηγούν πάντα σε ένα προκατειλημμένο αποτέλεσμα από τα μοντέλα που εκπαιδεύονται σε αυτά τα μεροληπτικά δεδομένα (Greenstein, no date). Άλλωστε, τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση του αλγόριθμου μπορεί να έχουν «μολυνθεί» με ανθρώπινες προκαταλήψεις. Ο αλγόριθμος μηχανικής μάθησης θα βασίζεται σε αυτές τις προκαταλήψεις και, ενδεχομένως, θα τις υπερβάλλει θεωρώντας τις ως «αληθινές» για τις μελλοντικές αποφάσεις ή τις προβλέψεις του (Scherer, 2019). Ωστόσο, και χωρίς να φτάσει στο σημείο να δείχνει ανθρώπινες προκαταλήψεις στα υποκείμενα δεδομένα, το μοντέλο μπορεί να εξάγει μοτίβα από τα δεδομένα και να τα προεκβάλλει με τρόπο που μπορεί να οδηγήσει σε συστημικά λάθη. Για παράδειγμα, μελέτες έχουν δείξει ότι η χρήση αλγορίθμων στην αξιολόγηση του ποινικού κινδύνου στις Ηνωμένες Πολιτείες οδήγησε σε ρατσιστικά προκατειλημμένα αποτελέσματα. Το σύστημα Correctional Offender Management Profiles for Alternative Sanctions (COMPAS) χρησιμοποιείται ευρέως στις Ηνωμένες Πολιτείες για την αξιολόγηση του κινδύνου υποτροπής των κατηγορούμενων. Μελέτες διαπίστωσαν ότι, με το σύστημα αυτό, «οι μαύροι κατηγορούμενοι είχαν ... διπλάσιες πιθανότητες από τους λευκούς κατηγορούμενους να χαρακτηριστούν εσφαλμένα ότι έχουν υψηλότερο κίνδυνο βίαιης υποτροπής», ενώ «οι λευκοί βίαιοι υπότροποι είχαν 63% περισσότερες πιθανότητες να έχουν ταξινομηθεί εσφαλμένα ως χαμηλού κινδύνου βίαιης υποτροπής, σε σύγκριση με τους μαύρους». Το αν αυτή η φυλετική προκατάληψη στο πρόγραμμα υπολογιστή βασίστηκε σε υπάρχουσες ανθρώπινες προκαταλήψεις στα δεδομένα εκπαίδευσης παραμένει ασαφές. Θα μπορούσε επίσης να προέκυψε από το γεγονός ότι ο αλγόριθμος ταξινόμησε εσφαλμένα τους μαύρους κατηγορούμενους με υψηλότερο ποσοστό υποτροπής, επειδή, κατά τα προαναφερόμενα, αυτή η φυλετική ομάδα υπερεκπροσωπείται σε ορισμένα είδη εγκλημάτων. Το υπολογιστικό μοντέλο θα μπορούσε να εξάγει ως συμπέρασμα από αυτό το πρότυπο τη λανθασμένη υπόθεση υψηλότερου κινδύνου υποτροπής. Η εμφάνιση συστημικών σφαλμάτων που βασίζονται σε κρυφά μοτίβα στα υποκείμενα δεδομένα αποτελεί σοβαρό κίνδυνο (Scherer, 2019). Από τα ανωτέρω προκύπτει ότι ένας ορισμός της μεροληψίας είναι ότι : «τα διαθέσιμα δεδομένα δεν είναι αντιπροσωπευτικά του πληθυσμού ή του φαινομένου της μελέτης [...][ότι] τα δεδομένα δεν περιλαμβάνουν μεταβλητές που αποτυπώνουν σωστά το φαινόμενο που θέλουμε να προβλέψουμε [και ότι] τα δεδομένα περιλαμβάνουν περιεχόμενο που παράγεται από ανθρώπους το

οποίο μπορεί να περιέχει μεροληψία εναντίον ομάδων ανθρώπων» (Greenstein, no date). Επίσης, προκύπτει ότι είναι σημαντικό να εξεταστεί εάν και πώς θα μπορούσαν να αντιμετωπιστούν συστημικά λάθη σε αλγόριθμους. Σε συστήματα όπου ο αλγόριθμος κωδικοποιείται από έναν άνθρωπο προγραμματιστή, τα λάθη θα εντοπίζονται συχνά στον ίδιο τον σχεδιασμό των αλγορίθμων. Μπορεί να διορθωθούν μόλις εντοπιστεί το λάθος. Αντίθετα, στα συστήματα μηχανικής μάθησης ο αλγόριθμος εξάγεται από τα δεδομένα, όπως περιγράφεται παραπάνω. Επομένως, τα λάθη συνήθως προκύπτουν από τα δεδομένα εισόδου και είναι πιο δύσκολο να εντοπιστούν και να διορθωθούν. Η απόκρυψη ευαίσθητων στοιχείων κατά την εισαγωγή, όπως η εθνοτική καταγωγή ή η γεωγραφική προέλευση, θα μπορούσε να ληφθεί υπόψη για την αποφυγή προβλημάτων. Ωστόσο, ακόμα κι αν αυτά τα ευαίσθητα χαρακτηριστικά είναι κρυμμένα, οι αλγόριθμοι θα μπορούσαν ωστόσο να τα ανακατασκευάσουν σιωπηρά από μεταβλητές μεσολάβησης (proxy variables) (Scherer, 2019). Όπως τονίστηκε από τους Ignazio και Klein και αναφέρθηκε από τους Angwin et al. (όπως παρατίθεται από τους Hayes, van de Poel and Steen, 2020) μεταξύ των ερωτήσεων που τίθενται σε μια έρευνα COMPAS που δίνονται σε ένα υποκείμενο, είναι ερωτήσεις που σχετίζονται με το ποιος μεγάλωσε το άτομο που ερωτήθηκε και εάν οι γονείς του είχαν χωρίσει. Αυτές οι ερωτήσεις μπορούν να χρησιμοποιηθούν ως μεσολαβητές (proxy) για τη φυλή. Συγκεκριμένα, οι Ignazio and Klein (όπως αναφέρεται στους Hayes, van de Poel and Steen, 2020) υποδεικνύουν ότι στις ΗΠΑ η πλειοψηφία των μαύρων παιδιών μεγαλώνει σε μονογονεϊκές οικογένειες. Συνεπώς, ακόμα και η εξέταση της εξαίρεσης ορισμένων μεταβλητών δεν θα επιλύσει απαραίτητα το πρόβλημα με την προκατάληψη που βασίζεται σε δεδομένα (Bolingford *et al.*, 2020). Επιπλέον, όπως επισήμανε η Scherer (2019), ο στόχος της μηχανικής μάθησης είναι τα προγράμματα υπολογιστών να μαθαίνουν από την εμπειρία και να βελτιώνουν την απόδοσή τους με την πάροδο του χρόνου. Ο αλγόριθμος επομένως επηρεάζεται όχι μόνο από το αρχικό σύνολο δεδομένων εκπαίδευσης, αλλά και από τη χρήση και τη συνεχιζόμενη εισαγωγή δεδομένων με την πάροδο του χρόνου. Επομένως, οι χρήστες έχουν μια αναμφισβήτητη «δύναμη» να αλλάζουν τους αλγόριθμους.

Περαιτέρω, όπως αναφέρει ο Greenstein, η μεροληψία θα πρέπει να διακρίνεται από τη διάκριση, η οποία είναι μια νομική έννοια που μπορεί να περιγραφεί ως «η προκατειλημμένη μεταχείριση ενός ατόμου με βάση τη συμμετοχή του σε μια συγκεκριμένη ομάδα ή κατηγορία», όπου τα χαρακτηριστικά που περικλείουν διάκριση περιλαμβάνουν τη φυλή, την εθνικότητα, τη θρησκεία, εθνικότητα, φύλο, σεξουαλικότητα, αναπηρία, οικογενειακή κατάσταση, γενετικά χαρακτηριστικά, γλώσσα και ηλικία. Κατά συνέπεια, ένα μοντέλο λέγεται ότι εισάγει διακρίσεις σε περιστάσεις όπου δύο άτομα έχουν τα ίδια χαρακτηριστικά σχετικά με μια διαδικασία λήψης αποφάσεων, αλλά διαφέρουν ως προς ένα ευαίσθητο χαρακτηριστικό, το οποίο καταλήγει σε διαφορετική απόφαση που παράγεται από το μοντέλο. Η μεροληψία και η διάκριση σχετίζονται επομένως στην έκταση που η μεροληψία στα δεδομένα μπορεί να οδηγήσει σε διακρίσεις, αλλά μπορεί να μην το κάνει απαραίτητα σε όλες τις περιπτώσεις (Greenstein, no date). Σημαντικά για το υπό κρίση ζήτημα της ενσωματωμένης μεροληψίας καταδεικνύεται στην

επισκοπούμενη βιβλιογραφία η απόφαση των Ηνωμένων Πολιτειών (ΗΠΑ) του State of Wisconsin κατά Loomis 881 N.W.2d 749 (Wis. 2016) (McKay, 2020) και το προαναφερόμενο σύστημα COMPAS. Ειδικότερα, ο Eric Loomis αμφισβήτησε τη χρήση του εργαλείου COMPAS (λογισμικό αξιολόγησης κινδύνου κλειστού κώδικα) ως μέρος της καταδίκης του σε εξαετή φυλάκιση, ισχυριζόμενος ότι παραβίαζε το δικαίωμα του κατηγορουμένου σε δίκαιη δίκη. Βασικά, ο Loomis παρουσίασε τρία επιχειρήματα κατά της χρήσης του COMPAS κατά την καταδίκη. Ισχυρίστηκε ότι : (1) παραβιάζει το δικαίωμα του κατηγορουμένου να καταδικαστεί βάσει ακριβών πληροφοριών, εν μέρει επειδή ο ιδιωτικός χαρακτήρας του COMPAS τον εμποδίζει να αξιολογήσει την ακρίβειά του. (2) παραβιάζει το δικαίωμα του κατηγορουμένου για εξατομικευμένη ποινή· και (3) χρησιμοποιεί εσφαλμένα έμφυλες εκτιμήσεις στην καταδίκη (Rubim Borges Fortes, 2020). Μετά την υπόθεση Loomis, η χρήση της τεχνητής νοημοσύνης στο δικαστικό σύστημα στις Ηνωμένες Πολιτείες έχει λάβει αυξημένη προσοχή από τα μέσα ενημέρωσης. Αυτό, ιδιαίτερα από τη στιγμή που η ProPublica, έχοντας εξετάσει τα αποτελέσματα των περιπτώσεων όπου χρησιμοποιήθηκαν αλγοριθμικές αξιολογήσεις κινδύνου, ισχυρίστηκε ότι τα στατιστικά στοιχεία αρχίζουν να εντοπίζουν μια φυλετική προκατάληψη στις αποφάσεις, όπου οι λευκοί αντιμετωπίζονταν πιο ευνοϊκά από τους Αφροαμερικανούς. Πρώτα, εξετάζοντας 7.000 αποφάσεις, τα αποτελέσματα έδειξαν ότι ο αλγόριθμος είναι μόνο 20 τοις εκατό επιτυχής στην ακριβή πρόβλεψη της υποτροπής. Δεύτερον, ο αλγόριθμος επισήμανε εσφαλμένα τους Αφροαμερικανούς σε διπλάσιο ποσοστό από τους λευκούς (Greenstein, no date). Εξάλλου, αναφορικά με την υπόθεση Loomis, πρέπει να αναφερθεί ότι ο αλγόριθμος COMPAS προσδιόρισε τον Loomis ως άτομο που παρουσίαζε υψηλή απειλή για την κοινωνία λόγω υψηλού κινδύνου υποτροπής και το πρωτοβάθμιο δικαστήριο αποφάσισε να απορρίψει το αίτημά του να αφεθεί ελεύθερος υπό όρους. Στην έφεση, το Ανώτατο Δικαστήριο του Ουισκόνσιν αποφάσισε ότι η σύσταση από τον αλγόριθμο COMPAS δεν ήταν ο μοναδικός λόγος για την απόρριψη του αιτήματός του για αποφυλάκιση υπό όρους και ως εκ τούτου η απόφαση του δικαστηρίου δεν παραβίασε το δικαίωμα της δίκαιης διαδικασίας του Loomis. Επιβεβαιώνοντας τη συνταγματικότητα του αλγορίθμου αξιολόγησης κινδύνου συστάσεων, το Ανώτατο Δικαστήριο του Ουισκόνσιν αγνόησε τη δύναμη της «προκατάληψης αυτοματισμού». Αντιθέτως, σε μία άλλη απόφαση (Kansas v. Walls (2017), το Εφετείο της Πολιτείας του Κάνσας κατέληξε σε αντίθετο πόρισμα και αποφάσισε ότι ο κατηγορούμενος πρέπει να έχει πρόσβαση στην αξιολόγηση του εργαλείου LSI-R, στην οποία βασίστηκε το δικαστήριο για να αποφασίσει ποιες προϋποθέσεις αναστολής θα του επιβάλει. Το Εφετείο αποφάσισε ότι αρνούμενο την πρόσβαση του κατηγορουμένου στην αξιολόγηση LSI-R του, το περιφερειακό δικαστήριο του αρνήθηκε την ευκαιρία να αμφισβητήσει την ακρίβεια των πληροφοριών στις οποίες έπρεπε να βασιστεί το δικαστήριο για τον καθορισμό των όρων της δοκιμαστικής περιόδου του. Αναφερόμενο στην απόφαση Kansas κατά Easterling, το Εφετείο αποφάσισε ότι η παράλειψη του περιφερειακού δικαστηρίου να δώσει στον κατηγορούμενο αντίγραφο ολόκληρου του LSI-R, του στέρησε το συνταγματικό του δικαίωμα σε δίκαιη δίκη στη φάση της καταδίκης κατά τη σε βάρος του ποινική διαδικασία (Završnik, 2020). Επιπλέον, στο σημείο αυτό πρέπει να αναφερθούν και τα επιχειρήματα που υπάρχουν υπέρ του

εργαλείου COMPAS, τα οποία σχετίζονται με την υπό κρίση ανησυχία περί προκατάληψης των αποτελεσμάτων του. Ως τέτοια, αναφέρεται καταρχάς ότι ο λόγος για τη φυλετική διαφορά προέρχεται από το γεγονός ότι τα ποσοστά σύλληψης δεν είναι ισοδύναμα μεταξύ των φυλετικών ομάδων και ο αλγόριθμος απλώς αναπαράγει τις προβλέψιμες συνέπειες μιας βαθιάς άνισης κοινωνίας: «μέχρι να συλληφθούν όλες οι ομάδες με τον ίδιο ρυθμό, αυτού του είδους η προκατάληψη είναι μια μαθηματική βεβαιότητα». Ένα άλλο σημαντικό επιχείρημα για την υπεράσπιση του εν λόγω εργαλείου αξιολόγησης κινδύνου είναι ότι το COMPAS αποτελείται απλώς από λογισμικό που υποστηρίζει τη λήψη δικαστικών αποφάσεων σχετικά με τις πιθανότητες υποτροπής και τους κινδύνους υποτροπής και δεν σχεδιάστηκε «για να κάνει απόλυτες προβλέψεις σχετικά με την επιτυχία ή την αποτυχία» (Rubim Borges Fortes, 2020). Σε συμφωνία με όσα προαναφέρθηκαν ανωτέρω, σχετικά με τη θέση ότι "υπάρχει λόγος να προτιμάται όποιος τρόπος λήψης αποφάσεων, είτε με ανθρώπινη κρίση είτε με προγνωστικό αλγόριθμο, επιτυγχάνει τον πιο ελκυστικό συμβιβασμό μεταξύ ακρίβειας και μεροληψίας" (Chiao, 2019), πρέπει να αναφερθεί ότι, σήμερα, οι καλύτεροι αλγόριθμοι χρησιμοποιούν την τεχνική των τυχαίων δασών (random forests) που βασίζονται σε δέντρα απόφασης (decision trees), αλλά οι προβλέψεις βασίζονται σε μοτίβα από δεδομένα και συχνά είναι οριακά πιο ακριβείς από την τυχαία εικασία. Τελικά, το COMPAS δεν πρέπει να συγκριθεί με μαγικές προφητείες ή μηχανισμούς για τέλεια πρόβλεψη, αλλά με τη συγκεκριμένη εναλλακτική της ανθρώπινης κρίσης χωρίς την υποστήριξη αυτού του εργαλείου αξιολόγησης κινδύνου. Σε αυτό το πλαίσιο, το COMPAS μπορεί να έχει δύο πλεονεκτήματα έναντι αυτής της εναλλακτικής λύσης (δηλ. της ανθρώπινης κρίσης): τη συνέπεια να δίνει πάντα ακριβώς την ίδια απάντηση για το ίδιο σύνολο περιστάσεων· και την αποτελεσματικότητα της καλύτερης επεξεργασίας των δεδομένων και της πραγματοποίησης καλύτερων προβλέψεων (Rubim Borges Fortes, 2020). Καταδεικνύεται συνεπώς ότι οι αλγόριθμοι δεν χρειάζεται να είναι τέλειοι, αρκεί να αποδίδουν καλύτερα από τους ανθρώπους που λαμβάνουν αποφάσεις. Παράδειγμα, εξάλλου, εργαλείου τεχνητής νοημοσύνης που αποδίδει καλύτερα από τους ανθρώπους αποτελεί η χρήση τεχνικών μηχανικής μάθησης για τον εντοπισμό μικροεκφράσεων και η εκπαίδευσή τους για να διακρίνουν τις ψεύτικες καταθέσεις από τις ειλικρινείς. Συγκεκριμένα, οι Monaro et al. δοκίμασαν μεθόδους εξαγωγής χαρακτηριστικών OpenFace και τεχνικές μηχανικής μάθησης (δηλαδή, μηχανές υποστήριξης διανυσμάτων έναντι βαθιά νευρωνικών δικτύων), για να διακρίνουν τους ψεύτες από τους ειλικρινείς με βάση τις μικροεκφράσεις του προσώπου, χρησιμοποιώντας ένα σύνολο δεδομένων βιντεοσκοπημένων συνεντεύξεων (Monaro *et al.*, 2022). Πρέπει, ωστόσο, να σημειωθεί ότι αναφορικά με τους αλγόριθμους πρόβλεψης αποφάσεων δεν ενδείκνυται τέτοια σύγκριση. Συγκεκριμένα, στην περίπτωση αυτών, μερικές φορές, αναφέρεται ο ιατρικός τομέας και συγκεκριμένα το παράδειγμα του ογκολόγου, ο οποίος πρέπει να συμβουλευτεί έναν προγνωστικό αλγόριθμο για την αναγνώριση του καρκίνου του δέρματος, εάν αυτός ο αλγόριθμος έχει αποδειχθεί ότι αποδίδει καλύτερα από τους ανθρώπους. Ωστόσο, αυτή η αναλογία απορρίπτεται, καθώς σε αντίθεση με το ιατρικό παράδειγμα, ένας νομικός προγνωστικός αλγόριθμος και ένας δικαστής εκτελούν διαφορετικά καθήκοντα. Στο ιατρικό παράδειγμα, ο

άνθρωπος και ο αλγόριθμος εκτελούν την ίδια εργασία, δηλαδή την αναγνώριση του καρκίνου σε εικόνες, για παράδειγμα, σημαδιών. Επιπλέον, οι εκτιμήσεις του ανθρώπου και του αλγόριθμου συγκρίνονται με την ίδια (αντικειμενική) αλήθεια: εξετάζοντας τα κύτταρα στο μικροσκόπιο μπορεί να διαπιστωθεί με βεβαιότητα αν υπάρχει καρκίνος. Έτσι, ένας άνθρωπος ειδικός και ένας αλγοριθμικός ειδικός συγκρίνονται με βάση το ίδιο πρότυπο. Σε μια τέτοια περίπτωση, μια σύγκριση μεταξύ του τρόπου με τον οποίο αποδίδουν οι άνθρωποι και οι αλγόριθμοι έχει νόημα και ο αλγόριθμος μπορεί να ειπωθεί ότι αποδίδει καλύτερα από τον άνθρωπο γιατρό, δηλαδή αναγνωρίζοντας κακοήθη σημεία που χάνονται από τον άνθρωπο γιατρό. Ωστόσο, ένας αλγοριθμικός παράγοντας πρόβλεψης αποφάσεων εκτελεί διαφορετική εργασία από τον δικαστή. Ένας παράγοντας πρόβλεψης αποφάσεων προβλέπει ποια απόφαση θα έπαιρνε ένας δικαστής, που είναι διαφορετικό καθήκον από το καθήκον που εκτελεί ο δικαστής, το οποίο είναι να αποφασίζει στην υπόθεση. Τότε δεν έχει νόημα να συγκρίνουμε την απόδοση του αλγορίθμου και του ανθρώπου δικαστή. Επιπλέον, ακόμη και μια σωστή πρόβλεψη μιας νομικά εσφαλμένης απόφασης θα μετρούσε ως επιτυχία για τον προγνωστικό αλγόριθμο. Τέτοιες καταστάσεις μπορεί να προκύψουν, για παράδειγμα, καθώς το δοκιμαστικό σύνολο περιέχει νομικά εσφαλμένες αποφάσεις. Το να προβλέπεις σωστά μια απόφαση δεν είναι το ίδιο με το να προβλέπεις μια σωστή απόφαση (Bex and Prakken, 2021).

4.4.1. Ζητήματα εξατομικευμένης δικαιοσύνης και διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης.

Μπορούμε να εμπιστευτούμε ότι η τεχνολογία είναι δίκαιη, δεδομένου ότι η εισαγωγή της τεχνητής νοημοσύνης κατά τη διαδικασία λήψης αποφάσεων στον τομέα της δικαιοσύνης περιορίζει την ανθρώπινη λήψη αποφάσεων;

Περαιτέρω, ένα άλλο χαρακτηριστικό της ανθρώπινης λήψης αποφάσεων, το οποίο σχετίζεται με τα ζητήματα που απασχολούν την επισκοπούμενη βιβλιογραφία και άπτονται του αντικειμένου της παρούσας εργασίας είναι η διακριτική ευχέρεια των δικαστών και εισαγγελέων, κατά τη λήψη των αποφάσεών τους, χαρακτηριστικό το οποίο πρέπει να αποτυπωθεί προκειμένου να αναζητηθούν και οι επιπτώσεις της εισαγωγής των αλγοριθμικών εργαλείων στη διαδικασία λήψης αποφάσεων. Συγκεκριμένα, πολλές αποφάσεις που λαμβάνονται σε όλη την ποινική διαδικασία βασίζονται στην έννοια της εξατομικευμένης δικαιοσύνης και στην άσκηση της ανθρώπινης διακριτικής ευχέρειας: από τη διακριτική ευχέρεια των αστυνομικών έως τη σύλληψη, τη διακριτική ευχέρεια των δικαστών όσον αφορά τη χορήγηση ή την άρνηση αποφυλάκισης με εγγύηση, την καταδίκη και τις αποφάσεις περί φυλάκισης μετά την καταδίκη, που ενδέχεται να σχετίζονται με παραβάτες υψηλού κινδύνου. Η αρχή της εξατομικευμένης δικαιοσύνης ανταποκρίνεται στον μεμονωμένο δράστη, τα γεγονότα και το αδίκημα (Anthony et al. 2015) και ενστερνίζεται την έννοια ότι «δεν υπάρχει μεγαλύτερη ανισότητα από την ίση μεταχείριση των άνισων» (Dennis κατά Ηνωμένων Πολιτειών 339 US 162 , 184 (Frankfurter J) (1950), …, υπό την προϋπόθεση, ωστόσο, ότι «παρόμοιες περιπτώσεις θα πρέπει

να αντιμετωπίζονται με παρόμοιο τρόπο» (Wong v The Queen [2001] HCA 54, Gleeson CJ στο [6]). Φυσικά, η άσκηση διακριτικής ευχέρειας στην ποινική δικαιοσύνη δεν είναι εντελώς απεριόριστη από τη νομοθεσία, το νομικό προηγούμενο και τις κατευθυντήριες γραμμές που χρησιμεύουν για την οριοθέτηση της διακριτικής ευχέρειας (Martin 2017). Για παράδειγμα, στην επιβολή ποινών, οι δικαστές μπορεί να περιορίζονται από κατευθυντήριες αποφάσεις, υποχρεωτικές ελάχιστες ποινές, μέγιστες ποινές, αρχές ότι η φυλάκιση πρέπει να είναι η έσχατη λύση, απαγορευμένες «μειώσεις» ποινών, επιβαρυντικούς και ελαφρυντικούς παράγοντες και περιόδους μη αποφυλάκισης. Στο πλαίσιο αυτής της διαδικασίας, η αριθμητική συνέπεια είναι λιγότερο σημαντική από τη συνεπή εφαρμογή των νομικών αρχών (Hili v The Queen (2010) 242 CLR 520). Τελικά κατά την καταδίκη, οι δικαστές καλούνται να αξιολογήσουν την κατάλληλη ποινή, λαμβανομένου υπόψη του συγκεκριμένου αδικήματος που διέπραξε ο συγκεκριμένος δράστης, ενώ εξισορροπούν αντικειμενικούς και υποκειμενικούς παράγοντες με τους σκοπούς της ποινής που περιλαμβάνουν τιμωρία, αποτροπή, προστασία της κοινωνίας, αποκατάσταση, λογοδοσία, αποδοκιμασία και αναγνώριση της βλάβης που προκλήθηκε στο θύμα και στην κοινωνία. Αυτή είναι μια σύνθετη άσκηση αναλογικότητας και δίνει βάρος σε όλους τους ανταγωνιστικούς και πολλαπλούς στόχους της καταδίκης, μια διαδικασία που αναφέρεται ως «ενστικτώδης» ή «διαισθητική» σύνθεση. Η ενστικτώδης σύνθεση έχει περιγραφεί ως μια παγκόσμια αξιακή απόφαση, η οποία αναγνωρίζεται ως όχι απαραίτητα λογική και μια διαδικασία που μπορεί να παράγει «αποτελέσματα στα οποία τα λογικά μυαλά θα διαφοροποιούνται» [Hudson v The Queen (2010) 30 VR 610; 205 A Crim R 199; [2010] VSCA 332 στο [27], (McKay, 2020, με ανωτέρω αναφερόμενες παραπομπές)]. Ωστόσο, η κατά τα ανωτέρω διακριτική ευχέρεια κατά καιρούς έχει δεχθεί διάφορες επικρίσεις. Ως επιστέγασμα αυτών, πριν 50 χρόνια στις ΗΠΑ περιορίστηκε η ευρεία διακριτική ευχέρεια των δικαστών κατά την επιβολή των ποινών. Σήμερα δε όλες οι δικαιοδοσίες των ΗΠΑ σε κάποιο βαθμό έχουν κυρώσεις που φέρουν τη μορφή κατευθυντήριων γραμμών ή υποχρεωτικών ποινών (guideline or mandatory sentencing). Όπως σημειώνει ο Berry, τα συστήματα κατευθυντήριας ποινής έχουν μειώσει σημαντικά τη διακριτική ευχέρεια στην καταδίκη: Πριν από το 1984, οι ομοσπονδιακοί δικαστές διέθεταν διακριτική ευχέρεια που στην πραγματικότητα ήταν απεριόριστη στον καθορισμό των ποινών, καθοδηγούμενοι μόνο από ευρύ φάσμα ποινών που προβλέπονταν από ομοσπονδιακούς ποινικούς νόμους. Ο νόμος περί μεταρρύθμισης των ποινών του 1984, μετακίνησε το καθεστώς καταδίκης σχεδόν τελείως στο άλλο άκρο, εφαρμόζοντας ένα σύστημα υποχρεωτικών κατευθυντήριων γραμμών που περιόρισε σοβαρά τη διακριτική ευχέρεια του δικαστή κατά την επιβολή της ποινής. Εξάλλου, τα νομοθετήματα σταθερών ποινών ουσιαστικά καθορίζουν μια τυποποιημένη πρόταση για την επιβολή ποινών, η οποία στην πραγματικότητα είναι ένας αλγόριθμος καταδίκης, τον οποίο όμως οι δικαστές πρέπει να χειρίζονται χειροκίνητα (Bolingford *et al.*, 2020). Ένα χαρακτηριστικό παράδειγμα είναι τα “mandatory minimum”, τα οποία στην πραγματικότητα είναι πολύ απλοί αλγόριθμοι: Αν καταδικαστείς για το έγκλημα C, επιβάλλεται τιμωρία P ή μεγαλύτερη (Chiao, 2019). Περιγράφοντας το ως άνω κλίμα, που είχε ήδη διαμορφωθεί στις ΗΠΑ, οι Bolingford et al. υποστηρίζουν ότι σε αυτό οφείλεται η

εκεί δεκτικότητα των αρμοδίων οργάνων να υιοθετήσουν εφαρμογές και εργαλεία τεχνητής νοημοσύνης κατά την εισαγωγή τους στη διαδικασία λήψης αποφάσεων ποινικής δικαιοσύνης. Αντίθετα, η προσέγγιση λήψης αποφάσεων καταδίκης στις Ηνωμένες Πολιτείες διαφέρει σημαντικά από αυτήν που ισχύει στην Αυστραλία, όπου η μεθοδολογία καταδίκης που έχει υιοθετηθεί είναι αυτή της «ενστικτώδους σύνθεσης», που αναφέρεται ανωτέρω. Μάλιστα, παρά τις επικρίσεις που έχουν διατυπωθεί, όπως ότι υπάρχουν ενδείξεις ότι η ενστικτώδης σύνθεση οδηγεί σε σημαντική ασυνέπεια στα αποτελέσματα των υποθέσεων και σε δυσμενή έκβαση, συμπεριλαμβανομένων βαρύτερων ποινών, για ορισμένες ομάδες παραβατών, όπως οι Αβορίγινες παραβάτες, ίσως ως αποτέλεσμα της υποσυνείδητης προκατάληψης των δικαστών, το Ανώτατο Δικαστήριο στην υπόθεση Elias v The Queen δήλωσε: Όπως έχει εξηγήσει αυτό το Δικαστήριο σε περισσότερες από μία περιπτώσεις, οι παράγοντες που επηρεάζουν τον καθορισμό της ποινής συχνά κινούνται προς διαφορετικές κατευθύνσεις. Είναι καθήκον του δικαστή να εξισορροπήσει συχνά ασύγκριτους παράγοντες και να καταλήξει σε μια ποινή που είναι σωστή σε όλες τις περιστάσεις. Η εφαρμογή του ποινικού δικαίου περιλαμβάνει εξατομικευμένη δικαιοσύνη, η επίτευξη της οποίας αναγνωρίζεται ότι συνεπάγεται την άσκηση ευρείας διακριτικής ευχέρειας για την επιβολή ποινών. Έτσι, τα αυστραλιανά δικαστήρια αντιστέκονται σθεναρά σε κάθε προσπάθεια να μετριάσουν τη διακριτική τους ευχέρεια. Δεδομένου τούτου, είναι αναμενόμενο ότι η στροφή προς την ενσωμάτωση αλγορίθμων στη διαδικασία λήψης αποφάσεων δεν θα είναι γρήγορη. Θα χρειαστούν αρκετές εξελίξεις για μια τέτοια αλλαγή, συμπεριλαμβανομένης της πρόβλεψης μιας αιτιολόγησης για τον τρόπο με τον οποίο οι αλγόριθμοι θα ενίσχυαν τη διαδικασία λήψης αποφάσεων (Bolingford *et al.*, 2020).

Εξάλλου, η προσήλωση στην ιδέα της εξατομικευμένης δικαιοσύνης και της κατά τα ανωτέρω διακριτικής ευχέρειας των δικαστών και εισαγγελέων αναδεικνύει τα εξής ζητήματα: Καταρχάς, η αντίληψη είναι ότι οποιοσδήποτε μηχανογραφημένος αλγόριθμος ανεξάρτητα από το πόσο εξελιγμένος είναι και ανεξάρτητα από το μέγεθος και τη σύνθεση του συνόλου των δεδομένων στα οποία εκπαιδεύτηκε, θα παράγει αναπόφευκτα συστάσεις κατά τρόπο ανδρομερή, πολύ περισσότερο απ’ ότι θα μπορούσε να επιτευχθεί με την ανθρώπινη κρίση. Η ανθρώπινη κρίση/δικαιοσύνη ανταποκρίνεται σε ένα απροσδιόριστο μεγάλο εύρος σχετικών παραγόντων και, ως εκ τούτου, είναι κατάλληλη για την αντιμετώπιση πλαισίων λήψης απόφασης στα οποία κάθε περίπτωση είναι μοναδική, διαφορετική από κάθε άλλη περίπτωση. Ίσως, το πιο ξεκάθαρο παράδειγμα είναι τα προαναφερόμενα mandatory minimum: Αν καταδικαστείτε για το έγκλημα C, επιβάλλεται τιμωρία P ή μεγαλύτερη. Κάποιος μπορεί να σκεφτεί ότι αυτό είναι πολύ γενικό, καθώς μπορεί να υπάρχουν περιπτώσεις όπου κάποιος καταδικάζεται για C, αλλά ο οποίος εύλογα θα έπρεπε να τιμωρηθεί λιγότερο από το P. Πιο εύλογα θα μπορούσε κανείς να κατανοήσει την εξατομικευμένη δικαιοσύνη σαν να είναι το ίδιο με τον ισχυρισμό ότι κάθε υπόθεση είναι μοναδική, ιδίως στο πλαίσιο της ποινικής δικαιοσύνης. Επίσης, πιο εύλογα και πιο λογικά, θα μπορούσε κάποιος να ερμηνεύσει ότι μπορούμε συχνά να προσδιορίσουμε κάποιο εύρος εξίσου δικαιολογημένων αποτελεσμάτων για μια δεδομένη περίπτωση. Μέσα σ’ αυτό το εύρος, κανένα αποτέλεσμα δεν είναι εμφανώς ανώτερο από τα άλλα. Ως εκ τούτου, ενώ παρόμοιες περιπτώσεις θα

πρέπει να αντιμετωπίζονται κατά προσέγγιση με τον ίδιο τρόπο, δεν απαιτείται να αντιμετωπίζονται ακριβώς το ίδιο. Αυτό μπορεί να οφείλεται σε γνωσιακούς η γνωστικούς περιορισμούς (δεν είμαστε σε θέση να χρησιμοποιήσουμε κανόνες επαρκούς λεπτομέρειας) ή μπορεί να οφείλεται σε ασάφειες στις νομικές έννοιες ή αξίες που εφαρμόζουμε (η αναλογικότητα μπορεί να συνάδει με μια σειρά από προτάσεις) ή για κάποιον άλλον λόγο. Όποια και αν είναι η αιτία, θα μπορούσε κανείς να συμπεράνει ότι ακόμη και περιπτώσεις που είναι παρόμοιες σε όλες τις σχετικές διαστάσεις, δεν θα πρέπει όλες αναγκαστικά να συμμορφώνονται με το ίδιο ακριβώς αποτέλεσμα. Πράγματι, κάτι τέτοιο φαίνεται πολύ εύλογο. Μέρος του λόγου που οι λειτουργοί της ποινικής δικαιοσύνης τείνουν να έχουν τόση διακριτική ευχέρεια είναι αναμφίβολα επειδή οι αποφάσεις που καλούνται να λάβουν-εάν θα σταματήσουν κάποιον για ανάκριση, που θα περιπολούν, αν θα διατάξουν την κράτηση κάποιου όσο εκκρεμεί δίκη, πώς θα καταδικάσουν κ.λ.π.- δεν έχουν πάντα μοναδικές σωστές απαντήσεις, τουλάχιστον σε σχέση με τα διαθέσιμα στοιχεία τη στιγμή της απόφασης. Με αυτό τον τρόπο, η ανησυχία μπορεί να είναι ότι οι νέες τεχνολογίες προσπαθούν να δώσουν ένα τεχνητό αέρα ποσοτικής ακριβείας σε ότι είναι κατ' ουσίαν ποιοτικές κρίσεις που επιτρεπτά διαφέρουν μεταξύ τους (Chiao, 2019). Ο Chiao εκθέτοντας λεπτομερώς τα ζητήματα που θέτει η επίκληση της εξατομικευμένης δικαιοσύνης ως αντίρρηση για την αλγοριθμική λήψη αποφάσεων αντιτείνει, μεταξύ άλλων, ότι η ποινική διαδικασία παρά την ύπαρξη της διακριτικής ευχέρειας, δεν είναι ελεύθερη από νομικούς κανόνες και αρχές. Εάν η δίκαιη μεταχείριση των ανθρώπων, με την έννοια της απόφασης που προσαρμόζεται στην περίπτωση του καθενός, σημαίνει απόρριψη της τυπικής ισότητας ενώπιον του νόμου, τότε αμφισβητούνται όχι μόνο οι ηλεκτρονικοί αλγόριθμοι, αλλά και μεγάλο μέρος της υπάρχουσας ποινικής διαδικασίας. Είτε μέσω του νόμου, είτε μέσω της νομολογίας, είτε μέσω τοπικής δικαστικής διαδικασίας, οι δικαστές, οι εισαγγελείς και οι συνήγοροι υπεράσπισης αναμένεται να είναι πιστοί σε ένα ευρύ φάσμα νομικών κανόνων γενικής εφαρμογής. Επίσης, ο Chiao αντιτείνει ότι μια υπολογιστική εκτίμηση κινδύνου δεν χρειάζεται να ορίσει κάποιον τεχνητό ευφυή κανόνα, σύμφωνα με τον οποίο, για παράδειγμα, όλοι όσοι υπερβαίνουν ένα συγκεκριμένο όριο προσδιορισμού του κινδύνου πρέπει να τίθενται υπό κράτηση. Μπορεί να υπάρχουν ορισμένες περιπτώσεις, ακόμα και ένας σημαντικός αριθμός περιπτώσεων, όπου η σωστή λύση είναι ασαφής εκ των προτέρων, ακόμα και υπό την εκδοχή της εμπειρικά επικυρωμένης αξιολόγησης κινδύνου. Θα μπορούσε σωστά να υποστηριχθεί ότι σ' αυτές τις περιπτώσεις επιτρέπεται διαφορετικοί λειτουργοί να παίρνουν διαφορετικές αποφάσεις. Αυτό δεν θα ερχόταν σε αντίφαση με την απαίτηση αυτοί οι λειτουργοί να συμβουλεύονται πρώτα ένα εργαλείο αξιολόγησης κινδύνου πριν λάβουν την τελική απόφαση. Τονίζει δε ότι, το ότι δεν υπάρχουν μοναδικά σωστά αποτελέσματα για ορισμένες υποθέσεις, δεν δείχνει ότι δεν υπάρχουν πολλά λανθασμένα αποτελέσματα για αυτές τις περιπτώσεις. Ένα σημαντικό μέρος της απήχησης μιας εμπειρικά επικυρωμένης αξιολόγησης κινδύνου είναι η ικανότητά της να αποκλείει λανθασμένα αποτελέσματα πιο αξιόπιστα από την κλινική ανθρώπινη κρίση. Μπορεί, για παράδειγμα, οι δημοφιλείς απόψεις μεταξύ δικαστών και εισαγγελέων σχετικά με παράγοντες κινδύνου για περαιτέρω τέλεση εγκλημάτων-όπως το καθεστώς απασχόλησης ή η χρήση ναρκωτικών- να αποδειχθούν μετά την ανάλυση των

αποδεικτικών στοιχείων απλώς εσφαλμένες ή μικρής σημασίας. Υποστηρίζει, επιπλέον, ότι ενώ αυτοί οι τύποι εργαλείων δεν είναι άτρωτοι σε άλλες μορφές φυλετικής προκατάληψης, δεν είναι επιρρεπείς στα είδη ασυνείδητων ή σιωπηρών προκαταλήψεων που μπορεί να υπάρχουν σε δικηγόρους και δικαστές που αποφασίζουν για την εγγύηση ή την καταδίκη. Η αλγοριθμική λήψη αποφάσεων μπορεί επομένως να διευκολύνει ένα εξατομικευμένο αποτέλεσμα αποκλείοντας παράγοντες που είναι γνωστό ότι είναι εμπειρικά ή ηθικά άσχετοι. Κατά συνέπεια, ενώ το να βασιστείς στην αλγοριθμική λήψη αποφάσεων, μπορεί να οδηγεί την κατανομή των αποτελεσμάτων να επικεντρώνεται περισσότερο στον μέσο όρο, αυτό δεν οφείλεται στο ότι ο απώτερος στόχος είναι η σύγκλιση σε μια ενιαία σωστή απάντηση σε όλες τις υποθέσεις. Αντίθετα, η μείωση της στατιστικής διακύμανσης μπορεί να υποστηριχθεί με βάση τον περιορισμό της επιρροής εμπειρικά ή ηθικά άσχετων παραγόντων. Περαιτέρω, υπογραμμίζει ο Chiao ότι πιθανές, καθοριστικής σημασίας, αποφάσεις, σχετικά με τη σύλληψη, την εγγύηση, τον ισχυρισμό και την ποινή, λαμβάνονται συχνά γρήγορα, με περιορισμένη βάση πληροφόρησης, από άτομα που έχουν όλες τις συνηθισμένες γνωστικές προκαταλήψεις, ελλιπείς ευρετικές και ασυνείδητες επιρροές με τις οποίες είμαστε εξοικειωμένοι. Υπάρχει ήδη μεγάλη βιβλιογραφία σχετικά με τις προκαταλήψεις που επηρεάζουν τους δικαστές, από την ώρα που δικάζεται μία υπόθεση έως την εμφάνιση του θύματος στον αγώνα της κατηγορίας (Chiao, 2019). Υπάρχει, μάλιστα, μια σειρά παραγόντων που μπορεί να επηρεάσουν τη δικαστική λήψη αποφάσεων και την ουσιαστική δικαιοσύνη. Ως τέτοιοι αναφέρονται : το πότε και τι έχει φάει ένας δικαστής, η ώρα της ημέρας, πόσες άλλες αποφάσεις έχει πάρει εκείνη την ημέρα (decision fatigue), οι προσωπικές αξίες, οι ασυνείδητες υποθέσεις (unconscious assumptions), η εξάρτηση από τη διαίσθηση, η ελκυστικότητα των εμπλεκομένων προσώπων, το συναίσθημα. Ο βαθμός στον οποίο αυτοί οι παράγοντες επηρεάζουν τους δικαστές είναι άγνωστος, αλλά είναι πιθανό ότι ακόμα κι αν ένας δικαστής αντιλαμβάνεται αυτούς τους παράγοντες, είναι πιθανό να υποτιμά τον αντίκτυπό τους (Sourdin, 2018). Φαίνεται έτσι ότι οι δικαστικές αποφάσεις δεν είναι καθαρές και μπορούν να αντικατοπτρίζουν μεροληψία (συνειδητή και ασυνείδητη) και εξωνομικούς παράγοντες, όπως η ώρα της ημέρας. Αυτό σημαίνει ότι για να προβλέψουμε αποφάσεις θα πρέπει να μοντελοποιήσουμε παράγοντες που πραγματικά δεν θα έπρεπε να έχουν θέση στη λήψη αποφάσεων, έτσι ώστε η ακρίβεια να μην είναι πάντα καλό πράγμα (Chen, 2019). Ωστόσο, υπάρχουν αυτοί που ισχυρίζονται ότι οι αλγόριθμοι πρόβλεψης αποφάσεων βελτιώνουν την προβλεψιμότητα και τη συνέπεια της δικαστικής λήψης αποφάσεων, κάτι που απαιτείται από την αρχή της ισότητας. Σύμφωνα με αυτούς τους ισχυρισμούς, οι δικαστές μπορούν να χρησιμοποιήσουν προγνωστικούς παράγοντες για να καταλήξουν σε πιο συνεπείς, πιο ενημερωμένες και λιγότερο προκατειλημμένες αποφάσεις (Bex and Prakken, 2021). Παρά τις ως άνω αντιρρήσεις του στις ανησυχίες που εγείρει η εξατομικευμένη δικαιοσύνη αναφορικά με τη χρήση αλγοριθμικών εργαλείων στη δικαιοσύνη, ο Chiao συμφωνεί ότι απ’ αυτές επιβιώνει η ανησυχία ότι τα υπάρχοντα στατιστικά και προγνωστικά μοντέλα μπορεί να είναι πολύ χονδροειδή για να αποδώσουν εκείνα τα χαρακτηριστικά των νομικών υποθέσεων που θα περιμέναμε απ’ αυτά. Για παράδειγμα, θα μπορούσε κανείς να υποστηρίξει ότι παρόλο που οι Α και Β μπορεί να έχουν παρόμοιο υπόβαθρο,

έτσι ώστε ένα προγνωστικό μέσο να κατανέμει παρόμοιες βαθμολογίες υποτροπής, μπορεί ωστόσο να διαφέρουν κατά τρόπο που επηρεάζει την μελλοντική τους επικινδυνότητα· και ότι ένας άνθρωπος δικαστής μπορεί να είναι σε εγρήγορση για αυτές τις διαφορές με τρόπο που δεν είναι ένα καθαρά στατιστικό εργαλείο. Αυτή είναι μια εξαιρετικά λογική ανησυχία, η οποία όμως θα γίνει λιγότερο πιεστική με την πάροδο του χρόνου, καθώς αυτές οι τεχνολογίες ωριμάζουν. Επίσης, περιορίζεται, γιατί, αντί για μία κατηγορηματική αντίρρηση για τη χρήση, ας πούμε, αλγόριθμων προγνωστικής αστυνόμευσης ή καταδίκης, υποστηρίζει μόνο τη σκέψη ότι τέτοιες συσκευές πρέπει να χρησιμοποιούνται μόνο με τρόπους που διασφαλίζουν ότι η τελική απόφαση είναι κατάλληλα ευαίσθητη σε συγκεκριμένο πλαίσιο υπόθεσης (Chiao, 2019). Περαιτέρω, σημαντικό στο σημείο αυτό είναι να αναφερθεί ότι η αφαίρεση της ανθρώπινης διακριτικής ευχέρειας δεν οδήγησε στην εξάλειψη των διακρίσεων κατά των αφροαμερικανών. Παρά την κατά τα ανωτέρω αφαίρεση μέρους της διακριτικής ευχέρειας των δικαστών στις ΗΠΑ, όπως δείχνουν τα στοιχεία των τελευταίων τριάντα ετών, η φυλετική ανισότητα στο σύστημα ποινικής δικαιοσύνης έχει επιδεινωθεί. Αυτό που παραμελούν οι υποστηρικτές των αυτοματοποιημένων συστημάτων λήψης αποφάσεων είναι η σημασία της ικανότητας να κάμπτονται οι κανόνες και να ερμηνεύονται εκ νέου σύμφωνα με τις κοινωνικές συνθήκες. Επομένως, η αφαίρεση της ανθρώπινης διακριτικής ευχέρειας είναι ένα δίκοπο μαχαίρι: μπορεί να μειώσει την ανθρώπινη προκατάληψη, αλλά μπορεί επίσης να επιδεινώσει τις αδικίες του παρελθόντος ή να δημιουργήσει νέες. Στην ανάλυση της η Turkle για την κοινωνική αποδοχή των μηχανογραφημένων συστημάτων λήψης αποφάσεων, ισχυρίζεται ότι όταν ένα σύστημα γίνεται αντιληπτό ως μεροληπτικό και ως σύστημα που δημιουργεί φυλετικά ανόμοια αποτελέσματα στην καταδίκη, οι μειονεκτούντες αφροαμερικανοί θα επέλεγαν έναν ηλεκτρονικό δικαστή αντί για έναν άνθρωπο δικαστή. Όλοι, οι ανθρώπινοι δικαστές τείνουν να είναι λευκοί μεσήλικες άνδρες. Οι «σκληροί για το έγκλημα» νόμοι που καθιέρωσαν υποχρεωτικές ελάχιστες ποινές για πολλές κατηγορίες εγκλημάτων και αφαίρεσαν μέρος της διακριτικής ευχέρειας των δικαστών έκαναν την ποινική δικαιοσύνη των ΗΠΑ πιο δίκαιη—αλλά όλοι οι κατηγορούμενοι χτυπήθηκαν σκληρά και οι φυλακές σύντομα υπερπληρώθηκαν (Završnik, 2020). Τέλος, πρέπει να σημειωθεί ότι οι επιπτώσεις των συστημάτων τεχνητής νοημοσύνης μπορεί να έχουν στρεβλωτικές επιπτώσεις στους θεμελιώδεις ακρογωνιαίους λίθους και την αρχιτεκτονική των φιλελεύθερων δημοκρατιών, δηλαδή όσον αφορά την αρχή της διάκρισης των εξουσιών και τον περιορισμό της πολιτικής εξουσίας από το κράτος δικαίου (Završnik, 2020).

4.4.2. Η ακρίβεια κατά τη διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης

Περαιτέρω, αναφορικά με τα άνω ζητήματα που αναδεικνύει η χρήση ΤΝ νοημοσύνης στη δικαιοσύνη και άπτονται του ερωτήματος κατά πόσο οι σχετικές τεχνολογίες είναι δίκαιες, σημαντική είναι η έννοια της “ακρίβειας”, ως πιστότητα ή εγγύτητα στην αλήθεια (Hayes, van de Poel and Steen, 2020).

Από την προηγούμενη αναφορά στην επισκοπούμενη βιβλιογραφία, προκύπτει ότι ένας, πράγματι, ακριβής αλγόριθμος μπορεί να βοηθήσει

αποτελεσματικά στη λήψη αποφάσεων. Επίσης, προκύπτει ότι η ακρίβεια σχετίζεται με την ακρίβεια και την αξιοπιστία των δεδομένων που χρησιμοποιούνται (Scherer, 2019) και συγκεκριμένα είναι μια ιδιότητα των δεδομένων εκπαίδευσης, των δεδομένων εισόδου και των δεδομένων εξόδου (Hayes, van de Poel and Steen, 2020). Προεκτέθηκε, ωστόσο, η διαπίστωση ότι οι αλγόριθμοι μπορεί να αντικατοπτρίζουν μεροληψίες και προκαταλήψεις και, γενικά, διαφόρων ειδών ευπάθειες, που επηρεάζουν την ακρίβειά τους. Στις περιπτώσεις αυτές σημαντικές κρίνονται η αξία της δικαιοσύνης και της ισότητας, ενώ έχει διατυπωθεί και η θέση ότι μπορεί να είναι απαραίτητο να μειωθεί η συνολική ακρίβεια ενός αλγορίθμου, με την αφαίρεση λ.χ. μεταβλητών που σχετίζονται με τις μειονοτικές ομάδες, προκειμένου να αποφευχθούν δυσανάλογες επιπτώσεις σε βάρος τους (Hayes, van de Poel and Steen, 2020). Περαιτέρω, πρέπει να γίνει αντιληπτό ότι για να ληφθούν αποφάσεις για τους ανθρώπους, ουσιαστικά μειώνονται σε σημεία δεδομένων που συσχετίζονται και σταθμίζονται μαθηματικά μεταξύ τους. Αυτό με τη σειρά του έχει ως αποτέλεσμα τα μοντέλα να λαμβάνουν τις αποφάσεις, αντιμετωπίζοντας τους ανθρώπους με βάση τον τρόπο με τον οποίο αντιπροσωπεύονται στα δεδομένα, όπως καθορίζεται από το μοντέλο που ενσωματώνει τον αλγόριθμο. Αυτή η τεχνολογία μπορεί να είναι επιρρεπής σε προκαταλήψεις και λάθη και οι ψηφιακές αναπαραστάσεις των ανθρώπων μπορεί να μην αντικατοπτρίζουν την πραγματικότητα. Ωστόσο, πιθανώς το κύριο κακό με την τεχνολογία είναι το γεγονός ότι ένα μοντέλο δεν μπορεί να εκπαιδευτεί για να προβλέψει κάθε προσωπικότητα για την οποία πρέπει να αποφασίσει, με αποτέλεσμα το άτομο να πρέπει να προσαρμοστεί σε ένα υπάρχον σύνολο παραγόντων (Greenstein, no date). Επιπλέον, διαπιστώνεται ότι συχνά, μελέτες καταλήγουν στο συμπέρασμα ότι ένας συγκεκριμένος προγνωστικός αλγόριθμος δεν είναι πιο ακριβής από έναν άνθρωπο που λαμβάνει αποφάσεις και ενεργεί μόνος του. Ωστόσο, σε μελέτες όπου ζητείται από τους ανθρώπους να κάνουν προβλέψεις σχετικά με δεδομένα αποτελέσματα, είναι σπάνιο τα υποκείμενα να λαμβάνουν άμεση ανατροφοδότηση σχετικά με την ακρίβεια αυτών των προβλέψεων. Επίσης, γενικά δεν τους ζητείται να βασίσουν τις προβλέψεις τους στο ίδιο σύνολο δεδομένων με τον αλγόριθμο (επειδή τα δεδομένα είναι πολύ μεγάλα για να τα επεξεργαστεί ένας άνθρωπος που ενεργεί μόνος του). Αυτές οι μελέτες στη συνέχεια παρερμηνεύουν τόσο τη φύση όσο και τον σκοπό των προγνωστικών αλγορίθμων που λειτουργούν σε στατιστικά και αναλογιστικά μοντέλα. Ένας άνθρωπος που λαμβάνει αποφάσεις (ακόμα και ένας ειδικός) είναι πιθανό να κάνει προβλέψεις με βάση την έμπειρη και τεκμηριωμένη διαίσθηση, δηλαδή να κάνει μια τεκμηριωμένη εικασία. Ένας αλγόριθμος δεν μαντεύει. Εφαρμόζει τους τύπους του σε ένα προκαθορισμένο σύνολο δεδομένων και χρησιμοποιεί συναρτήσεις γραμμικής παλινδρόμησης για να χαρτογραφήσει τις σχέσεις μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών. Αυτό δημιουργεί ένα μοντέλο το οποίο ο αλγόριθμος μπορεί στη συνέχεια να χρησιμοποιήσει για να συγκρίνει με μεταγενέστερα δεδομένα για να το τελειοποιήσει. Ο αλγόριθμος μπορεί να λάβει άμεση ανατροφοδότηση σχετικά με το πόσο ακριβές ήταν το μοντέλο σε συγκεκριμένες καταστάσεις και στη συνέχεια να το βελτιώσει ώστε να ταιριάζει καλύτερα στα νέα δεδομένα. Αυτή η συνάρτηση «μάθησης» με την οποία οι αλγόριθμοι βελτιώνουν τις μελλοντικές τους προβλέψεις βάσει

ανατροφοδότησης σχετικά με πρόσφατες προβλέψεις δεν είναι κάτι που συνήθως δοκιμάζουν οι συγκριτικές μελέτες για την ακρίβεια. Αυτό ανάγει αυτές τις μελέτες σε μια στατική ανάλυση που αγνοεί τη λειτουργικότητα και την πραγματική αξία των αλγορίθμων μηχανικής μάθησης (Bolingford *et al.*, 2020). Από την άλλη μεριά, η Scherer (2019) διατυπώνοντας τα τέσσερα (4) V^s και τους εγγενείς περιορισμούς των μοντέλων που βασίζονται σε δεδομένα για τη λήψη νομικών αποφάσεων με τεχνητή νοημοσύνη, επισημαίνει ότι στο νομικό πλαίσιο πιθανό είναι να υπάρξει πρόβλημα με την έλλειψη των δεδομένων. Ως εκ τούτου, με την πάροδο του χρόνου, οι αποφάσεις μπορεί να μην είναι συχνές και, όταν λαμβάνουν χώρα, ενδέχεται να έχει υπάρξει μια αλλαγή στην πολιτική, έτσι ώστε τα προηγούμενα δεδομένα να είναι ξεπερασμένα. Αυτές οι αλλαγές πολιτικής μπορεί να είναι ριζικές και γρήγορες κατά καιρούς. Ως παράδειγμα χρησιμοποιεί από τον τομέα της διεθνούς διαιτησίας, την απόφαση του Δικαστηρίου της Ευρωπαϊκής Ένωσης στην υπόθεση Slowakische Republik v. Achmea, που άλλαξε ριζικά τη συμβατότητα της διαιτησίας μεταξύ επενδυτή-κράτους μέλους με το ευρωπαϊκό δίκαιο σε μια νύχτα. Αυτό εγείρει το ερώτημα πώς τα μοντέλα τεχνητής νοημοσύνης που, εξ ορισμού, βασίζονται σε πληροφορίες που εξάγονται από προηγούμενα δεδομένα, μπορούν να αντιμετωπίσουν αυτές τις αλλαγές πολιτικής. Είναι αλήθεια ότι η ουσία της μηχανικής μάθησης είναι η ικανότητα βελτίωσης του αλγόριθμου με την πάροδο του χρόνου. Ωστόσο, μια τέτοια βελτίωση βασίζεται πάντα σε δεδομένα του παρελθόντος. Οι αλλαγές πολιτικής στη νομολογία απαιτούν αναγκαστικά αποκλίσεις από προηγούμενα δεδομένα, δηλαδή προηγούμενες υποθέσεις. Για αυτούς τους λόγους, τα μοντέλα τεχνητής νοημοσύνης είναι πιθανό να διατηρήσουν «συντηρητικές» προσεγγίσεις που είναι σύμφωνες με προηγούμενες περιπτώσεις (Scherer, 2019). Οποιεσδήποτε, ωστόσο, ενστάσεις θα μπορούσαν να τεθούν σε βάρος των αλγορίθμων και της ακριβείας τους δεν είναι για να τους καταδικάσουμε αλλά για να σκιαγραφήσουμε τους κινδύνους που αναφύονται από την εφαρμογή τους στη διαδικασία λήψης αποφάσεων. Ένας ακριβής αλγόριθμος μπορεί να είναι μια χρήσιμη πηγή πληροφόρησης για την αποτελεσματική λήψη αποφάσεων. Ωστόσο, πρέπει να διεξαγάγουμε μια σοβαρή συζήτηση σχετικά με το τι είδους όριο ακρίβειας είναι αποδεκτό όταν αυτοί οι αλγόριθμοι μπορούν να έχουν επιπτώσεις σε ένα πλήθος από τις αξίες μας. Το κατώτατο όριο ακρίβειας που απαιτείται εξαρτάται από το λειτουργικό πλαίσιο του αλγορίθμου, καθώς το περισσότερο δεν είναι πάντα καλύτερο. Η τρέχουσα έρευνα σχετικά με την ακρίβεια των αλγορίθμων για τη δικαιοσύνη και την ασφάλεια είναι ανάμεικτη. Οι κατάλληλα ακριβείς αλγόριθμοι μπορούν να προσθέσουν αξία στη δικαιοσύνη και την ασφάλεια, ωστόσο οι ανακριβείς αλγόριθμοι (σε συνδυασμό με κακές πρακτικές ανάπτυξης και διαχείρισης δεδομένων) μπορούν να δημιουργήσουν αναποτελεσματικότητα, να διευκολύνουν καταστροφικούς βρόχους ανατροφοδότησης και ακόμη και να θέσουν σε κίνδυνο τη ζωή και την ελευθερία των άμεσων ή έμμεσων στόχων τους. Αυτό που ξεχωρίζει για την αξία της ακρίβειας, είναι η αναγκαιότητα επιλογής των σωστών εισροών και η διασφάλιση της ποιότητας των δεδομένων, της προγνωστικής εγκυρότητας και τελικά των πραγματικών αποτελεσμάτων—οι αλγόριθμοι πρέπει να διερευνηθούν, να εξεταστούν και να δοκιμαστούν προσεκτικά (Hayes, van de Poel and Steen, 2020). Σημαντικό, επιπλέον, είναι να αναφερθεί ότι στην επισκοπούμενη βιβλιογραφία

παρουσιάζονται και εργαλεία τεχνητής νοημοσύνης τα οποία στοχεύουν στην αύξηση της ακρίβειας σε διάφορα στάδια της ποινικής και πολιτικής δίκης ή ακόμα και στο στάδιο της προετοιμασίας αυτών. Ως τέτοια παραδείγματα μπορούν να αναφερθούν ενδεικτικά: Η ανάπτυξη ενός αλγορίθμου βαθιάς μάθησης για την ανίχνευση της copy move πλαστογραφίας. Το προτεινόμενο μοντέλο βασίζεται στη χρήση CNN (Convolution Neural Network) και CovLSTM (Convolutional Long ShortTerm Memory) δικτύων και στοχεύει στην αύξηση της ακρίβειας της ανίχνευσης των πλαστογραφημένων ψηφιακών εικόνων με τη μορφή copy move. Η σημασία της μεθόδου είναι μεγάλη λαμβανομένου υπόψη ότι τέτοιες εικόνες μπορεί να χρησιμοποιηθούν ως αποδεικτικά μέσα στα δικαστήρια (Elaskily *et al.*, 2021). Επίσης, άλλη μέθοδος, που χρησιμοποιεί CNN για την ανίχνευση της ίδιας μορφής πλαστογραφίας αναλύουν οι Al Azrak et al (2020), οι οποίοι διαπιστώνουν ότι «η ακρίβεια με τη βαθιά μάθηση έφτασε το 100% σε δύο διαφορετικά σενάρια, γεγονός που αντανακλά τη δυνατότητα ανίχνευσης πλαστογράφησης εικόνων με βάση διακριτούς μετασχηματισμούς και βαθιά εκμάθηση» (Al_Azrak *et al.*, 2020). Σημαντική, εξάλλου, έκταση στην επισκοπούμενη βιβλιογραφία καταλαμβάνει η ανάπτυξη μοντέλων πρόβλεψης των αποτελεσμάτων των δικαστικών αποφάσεων. Ως παράδειγμα, οι Chen et al (2019) προτείνουν μία μέθοδο λήψης δικαστικών αποφάσεων, που βασίζεται σε ένα μοντέλο βαθιάς μάθησης σε υπάρχοντα δικαστικά έγγραφα και ειδικότερα στην ανάπτυξη μοντέλου πρόβλεψης των ποινών, πρόβλεψης των νομικών διατάξεων και πρόβλεψης της κατηγορίας, χρησιμοποιώντας τον αλγόριθμο TextCNN για την εκπαίδευση των δεδομένων εκπαίδευσης και το fastText για την ταξινόμηση κειμένων (Chen *et al.*, 2019). Ως ένα άλλο παράδειγμα, οι Chen et al (2020) προτείνουν ένα σύστημα πρόβλεψης των αποτελεσμάτων της δικαστικής απόφασης, με τη χρήση ενός μοντέλου BERT, που επιτρέπει στους διαδίκους ή τους δικηγόρους να ελέγχουν την αποτελεσματικότητα της δίκης πριν από την προσφυγή στο δικαστήριο και στη συνέχεια να εξετάσουν εάν θα διεξαγάγουν νομικές διαδικασίες. Το ποσοστό ακρίβειας της πρόβλεψης της δικαστικής απόφασης, όσον αφορά τη διάρκεια της δίκης και το επιδικασθέν ποσό έφτασε στο 82,22% και στο 89,89% (Chen *et al.*, 2020). Σημαντικές, επίσης, είναι οι αναφορές στις τεχνολογίες πρόβλεψης εγκλήματος και αξιολόγησης κινδύνου. Όπως σημειώνει ο Završnik (2020) μια μελέτη 1,36 εκατομμυρίων υποθέσεων προφυλάκισης έδειξε ότι ένας υπολογιστής θα μπορούσε να προβλέψει εάν ένας ύποπτος θα δραπετεύσει ή θα επαναλάβει το έγκλημα, καλύτερα από έναν άνθρωπο δικαστή.

4.5. Το ζήτημα της αδιαφάνειας των διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης

Παρά τη διαφαινόμενη υψηλή ακρίβεια των διαφόρων μοντέλων ΤΝ, αναμφισβήτητη είναι η έλλειψη κατανόησης των διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης. Πέραν δε του δεδομένου ότι οι κανόνες της τεχνητής νοημοσύνης είναι οι κανόνες των μαθηματικών και της στατιστικής, τα πράγματα επιδεινώνονται καθώς αυτοί οι κανόνες κρύβονται είτε στο ιδιόκτητο «μαύρο κουτί» ή κρύβονται επίσης στο βαθμό που δεν μπορούν να γίνουν κατανοητοί— δεν μπορούν να διαβαστούν, δεν μπορούν να συζητηθούν, δεν μπορούν να

αναλυθούν και δεν μπορούν να αιτιολογηθούν (Završnik, 2020). Διαφαίνεται, άλλωστε, ότι στις διαδικασίες λήψης αποφάσεων τεχνητής νοημοσύνης δεν αρκεί απλά η πρόβλεψη με έναν βαθμό ακρίβειας, όταν το πλαίσιο λήψης απόφασης είναι ιδιαίτερα σημαντικό για τη ζωή του ανθρώπου (Krupiy, 2020), όπως αυτό συμβαίνει και στον τομέα της δικαιοσύνης. Εξάλλου, απαραίτητη είναι η διαφάνεια για τη διαπίστωση τόσο της ακρίβειας όσο και της τυχόν προκατάληψης αυτών των τεχνολογιών, όπως αυτό προέκυψε από την υπόθεση Loomis. Τονίζεται, έτσι, στην επισκοπούμενη βιβλιογραφία, η ανάγκη για ανάπτυξη μοντέλων που να παρέχουν την απαραίτητη διαφάνεια, η οποία συνδέεται με την κατανόηση, την επεξηγηματικότητα και αιτιολόγηση, καθώς και την προβλεψιμότητα. Η ανάγκη, εξάλλου, για διαφάνεια μπορεί να αφορά στην ίδια την απόφαση, στα στάδια που προηγήθηκαν αυτής, στα σύνολα δεδομένων που χρησιμοποιούνται για την εκπαίδευση των αλγορίθμων (Margagliotti and Bollé, 2019), μπορεί να συνδέεται με πολλά άλλα ζητήματα που τίθενται από το δίκαιο των δικαιωμάτων του ανθρώπου, από τη συμβατότητα ή όχι της χρήσης αυτών των τεχνολογιών με το κράτος δικαίου (Greenstein, no date) ή να εγείρει το ερώτημα εάν τα άτομα έχουν συναινέσει στη χρήση των δεδομένων τους για τη λήψη της απόφασης ή ακόμη αν έχουν επίγνωση της απόφασης που τα επηρεάζει (Završnik, 2020).

4.5.1. Αδιαφάνεια και ανάγκη για κατανόηση των διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης. Ιδιόκτητη φύση των συστημάτων τεχνητής νοημοσύνης.

Αρχικά, αναφορικά με τη διάσταση της διαφάνειας που αφορά στην απαίτηση για κατανόηση, διαφαίνεται ότι δεν υπάρχει η ίδια ανάγκη για κατανόηση σε όλες τις υποθέσεις, λαμβανομένης υπόψη της πολυπλοκότητας ή της προβλεψιμότητας μιας υπόθεσης και ανάλογα με τον τρόπο λήψης της αυτοματοποιημένης αποφάσεως. Συγκεκριμένα, τέτοια ανάγκη δεν φαίνεται να υπάρχει στις περιπτώσεις που το αποτέλεσμα της υπόθεσης είναι προβλέψιμο, η απόφαση θα μπορούσε να είναι τυποποιημένη, ή όταν η αλγοριθμική λήψη αποφάσεων δεν βασίζεται στη μηχανική μάθηση: Πράγματι, ανεξάρτητα από το αντικείμενο της υπόθεσης, το έργο των δικαστηρίων και των δικαστών είναι να επεξεργάζονται πληροφορίες. Δεν χρειάζεται όμως κάθε επεξεργασία πληροφοριών περίπλοκες εξατομικεύσεις. Οι ερημοδικίες και οι απαράδεκτες αιτήσεις είναι συχνές. Πολλές υποθέσεις απαιτούν απλή αξιολόγηση χωρίς ακρόαση και ορισμένες υποθέσεις καταλήγουν σε συμβιβασμό. Μόνο ένα περιορισμένο ποσοστό των υποθέσεων που πρέπει να αντιμετωπίσει το δικαστικό σώμα είναι περίπλοκες, με αντικρουόμενους ισχυρισμούς. Σε διοικητικές και αστικές υποθέσεις, ο τρόπος με τον οποίο διεκπεραιώνονται οι υποθέσεις εξαρτάται κυρίως από (α) την πολυπλοκότητα των πληροφοριών σε μια υπόθεση και (β) τον βαθμό προβλεψιμότητας του αποτελέσματος. Ένα σχετικά μεγάλο ποσοστό περιπτώσεων ρουτίνας έχουν προβλέψιμη έκβαση. Σε αυτές τις περιπτώσεις, η δικαστική απόφαση είναι ένα έγγραφο που συντάσσεται με μια κατά κύριο λόγο αυτόματη διαδικασία που βασίζεται στα δεδομένα που

παρέχονται. Επιπλέον, εάν το αποτέλεσμα είναι προβλέψιμο, η επεξεργασία υποθέσεων θα μπορούσε να αυτοματοποιηθεί εν μέρει ή ακόμη και σε μεγάλο βαθμό με χρήση τεχνητής νοημοσύνης, ακριβώς επειδή το αποτέλεσμα είναι σε μεγάλο βαθμό ή εντελώς βέβαιο (Realing, 2020). Επίσης, η τεχνητή νοημοσύνη με ανθρώπινη επίβλεψη μπορεί να υιοθετηθεί για την έναρξη διαλόγου για διαμεσολάβηση και άλλες μορφές εναλλακτικής επίλυσης διαφορών. Ειδικά στην περίπτωση μικροδιαφορών, όπου η υπόθεση δεν είναι πολύπλοκη και το κόστος είναι χαμηλό, ενώ η εφαρμογή του νόμου είναι επαναλαμβανόμενη, υπάρχει δυνατότητα ηλεκτρονικής διαιτησίας που ζητείται από τον εναγόμενο και δεσμεύει μόνο αυτόν και όχι τον ενάγοντα. Σε αυτές τις περιπτώσεις, μια σημαντική διαφοροποίηση μπορεί να προέρχεται από τις διαφορετικές χρήσεις της τεχνολογίας. Επίσης, η τεχνολογία της πληροφορίας υποστηρίζει τις αποφάσεις αστικής ευθύνης κατά κανόνα ταξινομώντας παρόμοιες υποθέσεις και συγκεντρώνοντάς τις στο πλαίσιο της προετοιμασίας για μια συνολική δικαστική απόφαση. Για τις περισσότερες περιπτώσεις αδικοπραξιών γίνεται δεκτή η ύπαρξη αντικειμενικής ευθύνης, χωρίς την ανάγκη λεπτομερούς εξέτασης πταίσματος και υποκειμενικής ευθύνης που χαρακτηρίζουν τις ποινικές αποφάσεις (Rubim Borges Fortes, 2020). Εξάλλου, και από την άποψη του κράτους δικαίου, παρόλο που το τελευταίο εξαρτάται από τη φυσική γλώσσα για να γίνει κατανοητό, αυτό δεν ισχύει απαραίτητα για όλους τους τομείς του δικαίου, όπου ορισμένες νομικές διαδικασίες είναι ευκολότερο να αυτοματοποιηθούν (Greenstein, no date). Περαιτέρω, όπως υποστηρίζει ο Chiao (2019), δεν μπορεί να αποτελεί κατηγορηματική αντίρρηση στην αλγοριθμική λήψη αποφάσεων ότι οι περισσότεροι άνθρωποι δεν καταλαβαίνουν πώς λειτουργούν οι αλγόριθμοι. Ούτε, αναφορικά με τον τομέα της ποινικής δικαιοσύνης, μπορεί να αποτελέσει αντίρρηση για την αλγοριθμική λήψη αποφάσεων ότι είναι ένα περιβάλλον υψηλού κινδύνου. Λίγοι άνθρωποι που υποβάλλονται σε εγχείρηση καρδιάς θα μπορούσαν, να εξηγήσουν πραγματικά τις θεμελιώδεις αρχές της ανατομίας, της βιολογίας, της χημείας και της ιατρικής στις οποίες αναπόφευκτα βασίζεται μια τέτοια διαδικασία. Η ευφυΐα πρέπει να σημαίνει κάτι διαφορετικό από το τι είναι κατανοητό για τους περισσότερους ανθρώπους ή ακόμα και για εκείνους που αποτελούν το αντικείμενο της εν λόγω τεχνολογίας. Επιπλέον, ο ίδιος τονίζει ότι ενώ μεγάλο μέρος της προσοχής γύρω από την αλγοριθμική λήψη αποφάσεων έχει να κάνει με τεχνικές «μηχανικής μάθησης», όπως τα νευρωνικά δίκτυα, τα οποία όντως παρουσιάζουν ζητήματα κατ' αρχήν κατανόησης, ορισμένες αλγοριθμικές συσκευές που χρησιμοποιούνται σήμερα σε περιβάλλοντα ποινικής δικαιοσύνης δεν βασίζονται στη μηχανή - τεχνικές μάθησης. Για παράδειγμα, το εργαλείο αξιολόγησης κινδύνου που αναπτύχθηκε από το Ίδρυμα Arnold βασίζεται σε σχετικά απλά μοντέλα παλινδρόμησης. Ως εκ τούτου, ακόμη και αν οι περισσότεροι άνθρωποι, συμπεριλαμβανομένων των περισσότερων δικαστών, δικηγόρων και κατηγορουμένων, δεν καταλαβαίνουν πώς λειτουργούν οι παλινδρομήσεις, αυτό δεν αποτελεί αντίρρηση να βασιστούν σε μια συσκευή αξιολόγησης κινδύνου αυτού του είδους. Οι τεχνικές μηχανικής μάθησης, ιδίως τα νευρωνικά δίκτυα, εγείρουν ένα ξεχωριστό σύνολο ανησυχιών, δεδομένου ότι ένας αλγόριθμος μηχανικής μάθησης «μαθαίνει» από μόνος του να αντλεί συσχετίσεις μεταξύ των αποτελεσμάτων και των εισροών, συμπεριλαμβανομένων των εισροών που δεν

θα είχαν πολύ νόημα για έναν άνθρωπο. Στην περίπτωση των αεροπλάνων, των γεφυρών και των φαρμακευτικών προϊόντων , ακόμα κι αν οι μη ειδικοί δεν καταλαβαίνουν πώς λειτουργούν, οι ειδικοί το καταλαβαίνουν. Υπάρχουν τεχνικοί, μηχανικοί και χημικοί που καταλαβαίνουν τον μηχανισμό. Αντίθετα, στην περίπτωση ενός αλγόριθμου μηχανικής μάθησης, μπορεί κανείς να μην κατανοεί πραγματικά πως αντλεί τους συσχετισμούς του. Αυτοί οι συσχετισμοί μπορεί να είναι αρκετά αξιόπιστοι, αλλά μπορεί κανείς να μην είναι σε θέση να διατυπώσει ακριβώς γιατί είναι αξιόπιστοι και αυτό σίγουρα εγείρει ξεχωριστές ανησυχίες σχετικά με το ζήτημα της κατανόησης (Chiao, 2019). Εδώ, όμως, γεννάται και πάλι το ερώτημα (που τέθηκε και με το ζήτημα της ακρίβειας): με τι γίνεται αυτή η σύγκριση. Οι Stobbs et al. αναφερόμενοι στην «ενστικτώδη σύνθεση» σημειώνουν ότι «Όσο καλά κι αν αρθρώνεται η διαδικασία της ενστικτώδους σύνθεσης στις κρίσεις της ποινής, σε πολλούς φαίνεται να είναι μια αδιαφανής, αυθαίρετη, απρόβλεπτη και ίσως υπερβολικά υποκειμενική διαδικασία» (όπως αναφέρεται στην Mckay 2020). Τίθεται έτσι το ερώτημα: «Θα ήταν η διαδικασία πιο διαφανής και δίκαιη εάν η ενστικτώδης σύνθεση και οι σχετικές εκτιμήσεις κινδύνου από ανθρώπους επιλύονταν ή δομούνταν από έναν αλγόριθμο, έναν δικαστή AI» (Sourdin, 2018, όπως αναφέρεται στην McKay, 2020); Υποστηρίζεται, άλλωστε, ότι η ποινή μπορεί να αυτοματοποιηθεί, επειδή βασίζεται σε καθιερωμένες αρχές, σταθμίσεις και βασικούς παράγοντες. Η τεχνητή νοημοσύνη είναι κατάλληλη για την πολυπλοκότητα της βαθμονόμησης πολλαπλών μεταβλητών και τη βελτίωση του συστήματος ποινής ενσωματώνοντας αλγοριθμικές αξιολογήσεις κινδύνου και αφαιρώντας την «υποσυνείδητη προκατάληψη» των ανθρώπων (Stobbs et al., 2018, όπως αναφέρεται στην McKay, 2020). Φαίνεται, συνεπώς, ότι το ζήτημα της κατανόησης των διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης συγκρίνεται με το ζήτημα της κατανόησης της αντίστοιχης διαδικασίας λήψης αποφάσεων από ανθρώπους δικαστές, προκειμένου να μην απορριφθεί, μόνο για το λόγο αυτό, η εισαγωγή των διαδικασιών αυτών στον τομέα της δικαιοσύνης. Σχετικά, ο Chiao (2019) υποστηρίζει ότι ακόμα κι αν δεν κατανοούμε πλήρως τους συσχετισμούς στους οποίους βασίζει τις προβλέψεις του ένας αλγόριθμος μηχανικής μάθησης, θα πρέπει ίσως να αναρωτηθούμε πόσο καλά κατανοούμε τους λόγους για τους οποίους οι άνθρωποι – συμπεριλαμβανομένων των δικαστών, των εισαγγελέων, της αστυνομίας και άλλων – κάνουν τις προβλέψεις που κάνουν. Προτείνει δε να ζητηθεί από έναν δικαστή να εξηγήσει γιατί διέταξε την καταβολή εγγύησης σε μια υπόθεση, ενώ σε άλλη την κράτηση. Τονίζει, ωστόσο, ότι θα ήταν αφελές να πιστεύουμε ότι οι λόγοι που επικαλούνται οι άνθρωποι δημόσια ταυτίζονται με τους λόγους για τους οποίους πράγματι παίρνουν μία απόφαση. Πράγματι, αυτό δεν θα ήταν απλώς αφελές, αλλά και δεν θα βασιζόταν σε όσα γνωρίζουμε από τους κλάδους της κοινωνικής ψυχολογίας, της συμπεριφορικής οικονομίας και της νευροεπιστήμης, που παρέχουν όλο και πιο εξελιγμένες αναφορές για το χάσμα μεταξύ αυτού που λένε, πιστεύουν ή βιώνουν οι άνθρωποι και αυτού που κάνουν στην πραγματικότητα. Με άλλα λόγια, η κατανόησή μας για το γιατί ένας άνθρωπος που λαμβάνει αποφάσεις αποφάσισε μια αμφίρροπη υπόθεση με αυτόν τον τρόπο και όχι με έναν άλλον τρόπο, απέχει πολύ από το να είναι τέλεια, ιδιαίτερα όταν αγνοούμε τις υποκειμενικές αναφορές του ίδιου του ατόμου – ακόμη και αναφορές που έχουν

τη μορφή εκ των υστέρων νομικών αιτιολογήσεων. Εξάλλου, όπως προαναφέρθηκε (κατά την εξέταση των ζητημάτων που τίθενται αναφορικά με την ακρίβεια των αλγοριθμικών λήψεων αποφάσεων) και αναφέρεται εκ νέου στο σημείο αυτό, καθώς εκτιμάται ότι σχετίζεται και με το ζήτημα της προβλεψιμότητας των διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης, όσον αφορά το COMPAS, όπως σημειώνει ο Rubim Borges Fortes (2020), δεν πρέπει να συγκριθεί με μαγικούς χρησμούς ή μηχανισμούς για τέλεια πρόβλεψη, αλλά με τη συγκεκριμένη εναλλακτική της ανθρώπινης απόφασης χωρίς την υποστήριξη αυτού του εργαλείου αξιολόγησης κινδύνου. Σε αυτό το πλαίσιο, το COMPAS μπορεί να έχει δύο πλεονεκτήματα έναντι αυτής της εναλλακτικής λύσης: τη συνέπεια να δίνει πάντα ακριβώς την ίδια απάντηση για το ίδιο σύνολο περιστάσεων· και την αποτελεσματικότητα της καλύτερης επεξεργασίας των δεδομένων και της πραγματοποίησης καλύτερων προβλέψεων. Πέραν δε του COMPAS, στην επισκοπούμενη βιβλιογραφία, αναπτύσσονται μοντέλα πρόβλεψης δικαστικών αποφάσεων, τα οποία υποστηρίζεται ότι συμβάλλουν (εκτός από την ακρίβεια που αναφέρθηκε προηγουμένως) στην ασφάλεια δικαίου και την προβλεψιμότητα. Ως παράδειγμα οι Fernandes *et al.*, 2022, προτείνουν ένα μοντέλο, που ενσωματώνουν σε ένα πληροφοριακό σύστημα, το οποίο υποστηρίζουν ότι βοηθά τους δικηγόρους και τους δικαστές στη λήψη των αποφάσεών τους, εξάγοντας πληροφορίες από νομικά κείμενα στα Δικαστήρια της Βραζιλίας και αφορά σε αστικές αγωγές με αντικείμενο την άρνηση των ασφαλιστικών εταιρειών να παρέχουν βοήθεια στους πελάτες τους και ειδικότερα την ασφάλιση υγειονομικής περίθαλψης. Συγκεκριμένα, υποστηρίζουν ότι ο δικηγόρος μπορεί γρήγορα και αποτελεσματικά να συμβουλευτεί το σύστημα για να διαπιστώσει τις τελευταίες τάσεις του δικαστηρίου εφαρμόζοντας φίλτρα που αναγνωρίζουν το είδος της αγωγής που έχει ασκηθεί κατά της ασφαλιστικής εταιρείας και το συγκεκριμένο είδος ασθένειας που προκάλεσε άρνηση, μεταξύ άλλων παραγόντων, που ενδέχεται να σχετίζονται με δικαστικά αποτελέσματα. Επίσης, ένας δικαστής μπορεί να χρησιμοποιήσει το προτεινόμενο πληροφοριακό σύστημα. Για παράδειγμα, ο δικαστής πρέπει να δικάσει μια αγωγή όπου ο ενάγων ζητά κάλυψη από τον εναγόμενο για ακτινοθεραπεία, καθώς και ηθική βλάβη. Προκειμένου να διασφαλιστεί η δικαιοσύνη και η προβλεψιμότητα, ο δικαστής θα πρέπει να λάβει υπόψη τον τρόπο με τον οποίο κρίθηκαν από το Δικαστήριο προηγούμενες και παρόμοιες αγωγές που αφορούσαν αξιώσεις για τις ως άνω θεραπείες (π.χ.: Υπήρξαν άλλες ασφαλιστικές εταιρείες υπεύθυνες για άρνηση κάλυψης για τη συγκεκριμένη θεραπεία; Αν ναι, ποια ηθική βλάβη επιδικάστηκε σε βάρος τους;). Ακριβώς όπως ο δικηγόρος, έτσι και ο δικαστής δεν θα ήταν σε θέση να συγκεντρώσει αυτές τις πληροφορίες από τον υψηλό αριθμό υποθέσεων παρόμοιων αγωγών χωρίς τη βοήθεια ενός ευφυούς συστήματος. Ωστόσο, με την πρόσβαση στην προτεινόμενη μέθοδο, ο δικαστής θα είναι σε θέση να κατανοήσει γρήγορα τον τρόπο με τον οποίο το Δικαστήριο αποφάσισε παρόμοιες αξιώσεις και να δικάσει ανάλογα (Fernandes *et al.*, 2022). Περαιτέρω, σε ένα άλλο άρθρο, οι Fernandes el al., 2020, υπογραμμίζουν την ανάγκη για δικαστική συνέπεια: προκειμένου να προωθηθούν οι νομικές αρετές της βεβαιότητας και προβλεψιμότητας και να μειωθεί η αντίληψη περί αδικίας που προκύπτει από παρόμοιες υποθέσεις που κρίνονται με διαφορετικούς τρόπους,

είναι απαραίτητο οι δικαστικές αποφάσεις να μην έρχονται σε αντίθεση μεταξύ τους. Συγκεκριμένα, αναπτύσσουν μία μέθοδο, υποστηρίζοντας ότι η πρόβλεψη πληροφοριών θα βοηθήσει τους δικηγόρους να προσδιορίσουν εύκολα και με σιγουριά τις αξίες αποζημίωσης που χορηγούνται τακτικά στο πλαίσιο negativação indevida (αδικαιολόγητης άρνησης). Αντί η προσπάθεια να έγκειται στην κατανόηση της εθνικής νομολογίας διαβάζοντας πολλές δικαστικές αποφάσεις, η τεχνική που ανέπτυξαν είναι σε θέση να παρέχει αμέσως ένα πιο αξιόπιστο νομικό σενάριο. Επιπλέον, οι δικηγόροι θα μπορούν να προβλέψουν τις πιθανότητές τους να ανατρέψουν μια απόφαση στο Εφετείο. Επίσης, μπορεί να βοηθήσει τους δικαστές να ανακαλύψουν εάν οι αποφάσεις τους είναι σύμφωνες με τη συνολική στάση του Εφετείου, οδηγώντας σε δικαστική συνέπεια και βεβαιότητα, καθώς οι δικαστές θα μπορούν εύκολα να αξιολογήσουν τον τρόπο με τον οποίο αποφασίζονται από το Δικαστήριο νομικά ζητήματα στα οποία βασίζονται πολλές παρόμοιες υποθέσεις. Έτσι, οι αξίες αποζημίωσης που επιδικάζονται και τροποποιούνται από τα Εφετεία για το σενάριο negativação indevida, στο οποίο οι αξιώσεις είναι τακτικά παρόμοιες μεταξύ τους, δεν θα πρέπει να διαφέρουν τόσο πολύ όσο σήμερα. Ως εκ τούτου, η πρόβλεψη του αποτελέσματος αυτού του νομικού τομέα μπορεί να βοηθήσει τα δικαστήρια να διατηρήσουν την ασφάλεια δικαίου και την προβλεψιμότητα, που αποτελούν αξίες πυλώνων ενός δημοκρατικού κράτους υπό το κράτος δικαίου (Fernandes *et al.*, 2020).

Το πρόβλημα της κατανόησης των αλγοριθμικών λήψεων αποφάσεων και ως εκ τούτου της διαφάνειας αυτών συνδέεται εν πολλοίς και με το γεγονός ότι τα εν λόγω συστήματα αναπτύσσονται από ιδιωτικές εταιρείες, το συμφέρον των οποίων υπαγορεύει τη μη δημοσιοποίηση του σχεδιασμού τους (προστασία πνευματικής ιδιοκτησίας, ανταγωνιστικό προβάδισμα). Η προστασία των δικαιωμάτων Πνευματικής Ιδιοκτησίας για να αποτραπεί η επικάλυψη ή να διατηρηθεί το ανταγωνιστικό πλεονέκτημα, εξυπηρετεί καταρχήν έναν εύλογα θεμιτό σκοπό, καθώς επιτρέπει στους δημιουργούς να επωφεληθούν από την εργασία και/ή τις επενδύσεις τους · ωστόσο, τούτο συνδυάζεται με το γεγονός ότι οι αλγόριθμοι δεν είναι απαραίτητα κατανοητοί από τους χρήστες τους, δεν μπορούν να αμφισβητηθούν οι στόχοι τους, η καταλληλότητα και η ακρίβεια του μοντέλου τους ενδέχεται να μην είναι διαθέσιμα για επιθεώρηση και επαλήθευση και, κατά συνέπεια, η παρουσία μεροληψίας μπορεί να παραμείνει απαρατήρητη (Hayes, van de Poel and Steen, 2020). Έτσι, πολλοί μελετητές προσδιορίζουν την αδιαφάνεια του αλγορίθμου ως ιδιαίτερα προβληματική. Πράγματι, καθώς, ο πραγματικός αλγόριθμος, οι εισροές ή οι διαδικασίες του μπορεί να είναι προστατευμένα εμπορικά μυστικά, τα άτομα που επηρεάζονται από την αλγοριθμική αξιολόγηση δεν μπορούν να ασκήσουν κριτική ή να κατανοήσουν την απόφαση (Hogan-Doran, 2017, Carlson, 2017, όπως παρατίθενται στην McKay, 2020). Για παράδειγμα, η αδιαφάνεια του αλγόριθμου COMPAS (την ιδιωτική φύση του οποίου έχει επισημάνει και το Ανώτατο Δικαστήριο του Ουισκόνσιν στην υπόθεση Loomis) προέρχεται από τα ιδιόκτητα χαρακτηριστικά των νομικά προστατευμένων πηγαίων κωδίκων που παραμένουν άγνωστα στον κατηγορούμενο, τον δικηγόρο υπεράσπισης και τον ποινικό δικαστή. Ακόμα, όμως, και αν το δικαστήριο απαιτούσε τη διαφάνεια και τη δημοσίευση του

αλγορίθμου ως ανοιχτού κώδικα για όλους, ως προϋπόθεση για τη χρήση εργαλείων εκτίμησης κινδύνου στην ποινική καταδίκη, υπάρχει η πιθανότητα οι κανόνες λήψης αποφάσεων να προκύψουν με τρόπους, με τους οποίους κανείς —ακόμα και οι προγραμματιστές λογισμικού— δεν μπορεί να εξηγήσει το γιατί και πώς λαμβάνονται ορισμένες αλγοριθμικές αποφάσεις. Για παράδειγμα, οι αλγόριθμοι μηχανικής μάθησης ενός τεχνητού νευρωνικού δικτύου (ANN) μαθαίνουν μέσα από μια πολύπλοκη πολυεπίπεδη δομή και οι κανόνες απόφασής τους δεν προγραμματίζονται a priori και είναι συνήθως ακατανόητοι για τον άνθρωπο. Ακόμα κι αν μπορούμε να φανταστούμε ότι αυτοί οι αλγόριθμοι είναι λιγότερο μεροληπτικοί από τους ανθρώπους δικαστές, όπως προαναφέρθηκε, αυτοί οι αλγόριθμοι μηχανικής μάθησης λειτουργούν σύμφωνα με δεδομένα που χρησιμοποιούνται στην εκπαίδευσή τους και αναπαράγουν τις διακρίσεις που υπάρχουν στα δεδομένα εισόδου, τα οποία είναι αντιπροσωπευτικά του προκατειλημμένου κόσμου μας (Rubim Borges Fortes, 2020). Τίθενται, έτσι, τα ερωτήματα: Είναι δυνατόν να αμφισβητηθεί η ακριβής στάθμιση που εφαρμόζεται σε διάφορους παράγοντες κινδύνου για να κατανοηθεί εάν η στάθμιση είναι υπερβολική ή δυσανάλογη με άλλους παράγοντες; Πώς μπορούν τα άτομα να απαντήσουν στην κατηγορία που ασκήθηκε σε βάρος τους, να αμφισβητήσουν την ακρίβεια του αλγορίθμου και να αμυνθούν έναντι μιας δυσμενούς γι' αυτά απόφασης; Στο τέλος της ημέρας, υπάρχουν αποφάσεις ή αξιολογήσεις που επηρεάζουν κρίσιμα τις ανθρώπινες ζωές και την ελευθερία, οι οποίες δεν πρέπει να ανατίθενται σε αλγόριθμους; Τα αυτοματοποιημένα συστήματα και οι αλγοριθμικές αξιολογήσεις έχουν γίνει «μια σε μεγάλο βαθμό αδιαμφισβήτητη πτυχή» και επιβαρύνουν την ποινική διαδικασία (Carlson 2017: 313 που παραθέτει το Harcourt 2005) ωστόσο γεννούν ερωτήματα σχετικά με το ποιος - ή τι - είναι τώρα ο βασικός λήπτης των αποφάσεων (HoganDoran, 2017, όπως παρατίθενται στην McKay, 2020). Παρά τα ανωτέρω ζητήματα που θέτει η έλλειψη κατανόησης των σχετικών συστημάτων, που συνδέεται με τον ιδιοκτησιακό καθεστώς τους, δεν πρέπει να παραβλεφθεί ότι το καθεστώς αυτό μπορεί να αποτρέψει τη διάδοση της τεχνολογίας σε κακούς φορείς (εγκληματικές οργανώσεις και αυταρχικά καθεστώτα) που μπορούν να την προσαρμόσουν και να τη χρησιμοποιήσουν για κακόβουλους σκοπούς (Hayes, 2018, 265–268, όπως παρατίθεται στους Hayes, van de Poel and Steen, 2020). Σε λάθος χέρια ένας αλγόριθμος μπορεί να είναι ένα ισχυρό όπλο και η ιδιοκτησία μπορεί να είναι ένα ισχυρό εργαλείο για τον περιορισμό της ροής της δυνητικά επικίνδυνης γνώσης. Εν ολίγοις, η αξία του δικαιώματος της ιδιοκτησίας είναι ότι προάγει την ικανότητα των ιδιοκτητών να επωφελούνται από την ιδιοκτησία τους και να ελέγχουν τη διανομή ή τη χρήση της, ενώ, επίσης, τους τοποθετεί σε θέση ευθύνης (που μπορεί να χρησιμοποιηθεί για την υπεύθυνη αδειοδότηση ή διάδοση μιας τεχνολογίας), πλην, όμως, το επίπεδο ελέγχου που ασκούν επί της ιδιοκτησίας τους μπορεί να έχει ως συνέπεια την αδιαφάνεια (απόκρυψη της εσωτερικής λειτουργίας ενός αλγορίθμου από τους τελικούς χρήστες ή το κοινό γενικότερα) (Hayes, van de Poel and Steen, 2020).

4.5.2. Αδιαφάνεια και δικαιώματα του ανθρώπου

Περαιτέρω, ευχερώς μπορεί να διαπιστωθεί ότι η αδιαφάνεια της διαδικασίας λήψης αποφάσεων τεχνητής νοημοσύνης συνδέεται τόσο με τα

δικαιώματα του ανθρώπου όσο και με το κράτος δικαίου. Συγκεκριμένα, πολλά είναι τα ερωτήματα που τίθενται και αφορούν τα ανθρώπινα δικαιώματα και το κράτος δικαίου, λαμβανομένου υπόψη και του γεγονότος ότι ενώ η τεχνολογία αναπτύσσεται γρήγορα και αγκαλιάζει έννοιες όπως η διεθνοποίηση και η παγκοσμιοποίηση, το παραδοσιακό δίκαιο, ως επί το πλείστον, μπορεί να καθυστερήσει να αντιδράσει στις τεχνολογικές εξελίξεις και επίσης περιορίζεται κυρίως σε εθνικά σύνορα. Ωστόσο, έννοιες όπως τα δικαιώματα του ανθρώπου και το κράτους δικαίου αψηφούν το φαινόμενο του δικαίου να δεσμεύεται από τα εθνικά σύνορα και χαίρουν παγκόσμιας αναγνώρισης (Greenstein, no date). Μέσα σ' αυτό το πλαίσιο, σημαντικές είναι στην επισκοπούμενη βιβλιογραφία οι ανησυχίες που συνδέουν την αλγοριθμική λήψη αποφάσεων με τη δίκαιη δίκη και τις διάφορες εκφάνσεις της και συγκεκριμένα με ανησυχίες σχετικές με το κατά ποσό παραβιάζονται οι δικονομικές εγγυήσεις που θέτει το άρθρο 6 της ΕΣΔΑ. Ζητήματα, επίσης, μπορούν να δημιουργηθούν από την τυχόν ανάθεση μιας δημόσιας εξουσίας σε ιδιωτικούς φορείς, στους οποίους οι κυβερνήσεις μπορεί να αναθέσουν την ανάπτυξη αλγορίθμων τεχνητής νοημοσύνης που θα χρησιμοποιηθούν στα δικαστικά τους συστήματα (Završnik, 2020). Όπως άλλωστε σημειώνει ο Carlson, οι κυβερνήσεις θα πρέπει να αναπτύξουν τα δικά τους αναλογιστικά και αλγοριθμικά μέσα (όπως παρατίθεται στην McKay). Ωστόσο, σε χώρες όπου υφίσταται διαχωρισμός των εξουσιών μεταξύ εκτελεστικών και δικαστικών λειτουργιών, αυτό θα μπορούσε να οδηγήσει σε μια απαράδεκτη ασάφεια αυτού του διαχωρισμού (McKay, 2020). Επίσης, ο Greenstein (2021), αναφερόμενος στις προκλήσεις που θέτει η χρήση της τεχνητής νοημοσύνης από την άποψη του κράτους δικαίου, αναφέρει ως τέτοια (πρόκληση) τον βαθμό στον οποίο το δικαστικό σώμα, που βασίζεται στην τεχνητή νοημοσύνη που αναπτύχθηκε από ιδιωτικές εταιρείες, μπορεί να θεωρηθεί ανεξάρτητο. Πέραν δε τούτου, ο ίδιος υποστηρίζει ότι η αυξημένη χρήση της τεχνητής νοημοσύνης για την πρόβλεψη της ανθρώπινης συμπεριφοράς, πιο συγκεκριμένα της εγκληματικής συμπεριφοράς, αμφισβητεί ορισμένες παραδοσιακές νομικές έννοιες. Μια νομική έννοια που αμφισβητείται είναι ότι ο κατηγορούμενος θεωρείται αθώος μέχρι να αποδειχθεί η ενοχή του (τεκμήριο αθωότητας). Για παράδειγμα, η χρήση αλγοριθμικών αξιολογήσεων κινδύνου σε ποινικές δίκες για τον προσδιορισμό της υποτροπής εγείρει το ερώτημα εάν ο κατηγορούμενος θεωρείται ένοχος για ένα πιθανό έγκλημα, δηλαδή την τάση να διαπράξει ένα έγκλημα πριν αυτό συμβεί πραγματικά. Αυτό αναγνωρίζεται στις αρχές του «nullum crimen sine lege» και «nulla poena sine lege» που αναγνωρίζουν ότι δεν υπάρχει έγκλημα ή τιμωρία χωρίς νόμο (Greenstein, no date). Επιπλέον, όπως αναφέρει η McKay (2020), εκτός από το τεκμήριο αθωότητας, άλλα ζητήματα που μπορούν να εγείρουν τα αλγοριθμικά μέσα υπό το πρίσμα των δικαιωμάτων του ανθρώπου είναι ότι μπορεί να παραβιάζουν την ισότητα ενώπιον των δικαστηρίων και το δικαίωμα σε δίκαιη και δημόσια ακρόαση από ένα αρμόδιο, ανεξάρτητο και αμερόληπτο δικαστήριο. Η ισότητα των όπλων είναι βασική αρχή στη δικονομική δικαιοσύνη που σημαίνει ότι ο κατηγορούμενος δεν πρέπει να βρίσκεται σε μειονεκτική θέση σε σύγκριση με το κράτος που διώκεται, δηλαδή ότι πρέπει να υπάρχουν ίσοι όροι ανταγωνισμού (Roberts and Zuckerman, 2010, όπως παρατίθεται στην McKay, 2020). Φυσικά, στην ποινική διαδικασία, η αρχή αυτή αντιπροσωπεύει το ιδανικό και όχι την πραγματικότητα, ωστόσο

αμφισβητείται περαιτέρω σε καταστάσεις όπου η εισαγγελία χρησιμοποιεί ανεξερεύνητα αλγοριθμικά εργαλεία και δεδομένα εισαγωγής εναντίον ενός κατηγορουμένου, τα οποία δεν του έχει γνωστοποιήσει (McKay, 2020).

Περαιτέρω, όπως εκθέτει ο Završnik (2020):

Οι κανόνες της δίκαιης δίκης που περιλαμβάνονται στο άρθρο 6 της ΕΣΔΑ εγγυώνται στον κατηγορούμενο το δικαίωμα να συμμετέχει αποτελεσματικά στη δίκη και περιλαμβάνουν το τεκμήριο της αθωότητας, το δικαίωμα έγκαιρης ενημέρωσης για την αιτία και τη φύση της κατηγορίας, το δικαίωμα σε δίκαιη ακρόαση και το δικαίωμα να υπερασπιστεί κανείς τον εαυτό του αυτοπροσώπως. Το δικαίωμα στην αποτελεσματική συμμετοχή μπορεί να παραβιαστεί σε ποικίλες διαφορετικές καταστάσεις, που κυμαίνονται από κακή ακουστική στην αίθουσα έως την αποτροπή του κατηγορουμένου από το να παραστεί στη δίκη ή να εξετάσει έναν μάρτυρα που καταθέτει εναντίον του. Το τελευταίο είναι επίσης μία εκ των ελάχιστων εγγυήσεων δίκαιης δίκης που περιέχονται στο άρθ. 6, παράγρ. 3 και συνήθως απαιτεί να προσκομίζονται όλα τα στοιχεία εναντίον του κατηγορουμένου παρουσία του/της σε δημόσια ακρόαση, γεγονός που δίνει στον κατηγορούμενο μια αποτελεσματική ευκαιρία να αμφισβητήσει τα αποδεικτικά στοιχεία εναντίον του/της. Το δικαίωμα αντιπαράθεσης δεν ισχύει μόνο για μάρτυρες, δεδομένου ότι έχει μια αυτόνομη έννοια στο σύστημα της Σύμβασης που υπερβαίνει τη συνήθη σημασία του και περιλαμβάνει επίσης εμπειρογνώμονες, πραγματογνώμονες και θύματα. Σε κάθε περίπτωση στην οποία η κατάθεση χρησιμεύει σε σημαντικό βαθμό ως βάση για την καταδίκη του κατηγορουμένου, αποτελεί αποδεικτικό στοιχείο για τη δίωξη στην οποία ισχύουν οι εγγυήσεις της Σύμβασης. Το δικαίωμα που κατοχυρώνεται στο άρθρο 6 παράγραφος 3 στοιχείο δ, μπορεί ακόμη και να εφαρμοστεί στα αποδεικτικά έγγραφα και στα ηλεκτρονικά αρχεία, σχετικά με τις ποινικές κατηγορίες κατά του κατηγορουμένου. Επομένως, για να διασφαλιστεί η αποτελεσματική συμμετοχή σε μια δίκη, ο κατηγορούμενος πρέπει επίσης να μπορεί να αμφισβητήσει την αλγοριθμική βαθμολογία που αποτελεί τη βάση της καταδίκης του. Ωστόσο, το δικαίωμα στην αντιπαράθεση δεν είναι απόλυτο και μπορεί να περιοριστεί εάν πληρούνται ορισμένες προϋποθέσεις. Η παραδοσιακή προσέγγιση του Ευρωπαϊκού Δικαστηρίου Δικαιωμάτων του Ανθρώπου ήταν ότι το δικαίωμα σε δίκαιη δίκη παραβιαζόταν εάν μια καταδίκη βασιζόταν είτε αποκλειστικά είτε σε καθοριστικό βαθμό σε μια μη αμφισβητούμενη δήλωση (ο «μοναδικός ή αποφασιστικός κανόνας»). Στην (απόφαση) Al Khawaja and Tahery το Δικαστήριο απομακρύνθηκε εν μέρει από την προηγούμενη νομολογία του, δηλώνοντας ότι η αποδοχή μη ελεγμένων αποδεικτικών στοιχείων δεν θα οδηγήσει αυτόματα σε παραβίαση του άρθρου 6 παράγραφος 1: κατά την αξιολόγηση της συνολικής δικαιοσύνης μιας δίκης, το Ευρωπαϊκό Δικαστήριο Ανθρωπίνων Δικαιωμάτων έχει να εξετάσει εάν ήταν απαραίτητο να γίνουν δεκτά τέτοια στοιχεία και εάν υπήρχαν επαρκείς παράγοντες αντιστάθμισης, συμπεριλαμβανομένων ισχυρών διαδικαστικών διασφαλίσεων. Τα προβλήματα που δημιουργούνται από τα συστήματα τεχνητής νοημοσύνης είναι πολύ παρόμοια με εκείνα που παρουσιάζονται από ανώνυμους μάρτυρες ή μη κοινοποιηθέντα αποδεικτικά έγγραφα, καθώς τα συστήματα τεχνητής νοημοσύνης είναι αδιαφανή. Τουλάχιστον κάποιος βαθμός αποκάλυψης είναι απαραίτητος προκειμένου να

διασφαλιστεί ότι ο κατηγορούμενος έχει την ευκαιρία να αμφισβητήσει τα αποδεικτικά στοιχεία εναντίον του και να αντισταθμίσει το βάρος της ανωνυμίας. Οι απόντες ή ανώνυμοι μάρτυρες, αν και αυτοί καθ' εαυτό δεν είναι ασυμβίβαστοι με το δικαίωμα σε δίκαιη δίκη, μπορούν να συμμετάσχουν σε ποινική διαδικασία μόνο ως έσχατο μέτρο και υπό αυστηρές προϋποθέσεις που διασφαλίζουν ότι ο κατηγορούμενος δεν τίθεται σε μειονεκτική θέση. Ένας τέτοιος κανόνας θα πρέπει να εφαρμόζεται στη χρήση συστημάτων τεχνητής νοημοσύνης που χρησιμοποιούνται σε περιβάλλοντα ποινικής δικαιοσύνης. Θα πρέπει να επιτευχθεί δίκαιη ισορροπία μεταξύ του δικαιώματος αποτελεσματικής συμμετοχής στη δίκη, αφενός, και της χρήσης αδιαφανών συστημάτων τεχνητής νοημοσύνης που έχουν σχεδιαστεί για να βοηθούν τους δικαστές να καταλήξουν σε ακριβέστερες εκτιμήσεις της μελλοντικής συμπεριφοράς του κατηγορουμένου, αφετέρου. Το δικαίωμα της κατ’ αντιπαράσταση εξέτασης μαρτύρων θα πρέπει να ερμηνεύεται έτσι ώστε να περιλαμβάνει επίσης το δικαίωμα εξέτασης των δεδομένων και τους βασικούς κανόνες της μεθοδολογίας βαθμολόγησης κινδύνου. Στις διαδικασίες αναστολής, ένα τέτοιο δικαίωμα θα πρέπει να συνεπάγεται τη διασφάλιση ότι είναι δυνατό για ένα καταδικασθέν άτομο να αμφισβητήσει το μοντέλο που εφαρμόστηκε — από τα δεδομένα που τροφοδοτούνται στον αλγόριθμο έως τη συνολική σχεδίαση του μοντέλου.

4.5.3. Αδιαφάνεια και αιτιολόγηση αποφάσεων

Περαιτέρω, η βιβλιογραφία επιβεβαιώνει ότι οι διαφανείς διαδικασίες λήψης αποφάσεων διαδραματίζουν σημαντικό ρόλο στην αιτιολόγηση των αποφάσεων που λαμβάνονται. Επομένως, ενώ οι άνθρωποι μπορεί να κάνουν λάθη στην κρίση τους, μπορεί στη συνέχεια να είναι υπεύθυνοι για την αιτιολογία που χρησιμοποιούν για τη λήψη της απόφασής τους, γεγονός που μπορεί να δικαιολογήσει την προτίμηση των ανθρώπων στην ανθρώπινη απόφαση/πρόβλεψη (Bolingford *et al.*, 2020). Πράγματι, οι ηθικές και νομικές απαιτήσεις για διαφάνεια δεν περιορίζονται στη δημοσιότητα, αλλά σχετίζονται περισσότερο με ενδεχόμενο ανεξάρτητο έλεγχο, ένα κοινό σύστημα αιτιολόγησης που είναι κατανοητό στους άλλους και ένα σύστημα λογοδοσίας με ελέγχους και αμεροληψία για τη διόρθωση λαθών (Rubim Borges Fortes, 2020). Ειδικότερα, όσον αφορά τη νομιμότητα του συστήματος ποινικής δικαιοσύνης, και ιδιαίτερα την καταναγκαστική εξουσία του κράτους να τιμωρεί, να φυλακίζει και να επιβλέπει τους παραβάτες, αυτή έχει βασιστεί στη διαφανή δικαιοσύνη, δηλαδή σε ένα σύστημα που επιδιώκει τη λογοδοσία, την αμεροληψία και τη διαφάνεια (McKay, 2020). Με αυτόν τον τρόπο, συνεπώς, επιτυγχάνεται η αξιοπιστία του νομικού συστήματος. Από την άλλη μεριά, η τεχνολογική πολυπλοκότητα που σχετίζεται με την τεχνητή νοημοσύνη δεν την καθιστά κατάλληλη για την ανθρώπινη κατανόηση, τη διορατικότητα ή τη διαφάνεια. Για παράδειγμα, οι μαθηματικοί υπολογισμοί που λαμβάνουν χώρα στα κρυφά στρώματα των νευρωνικών δικτύων ή οι μεταλλακτικές δυνατότητες των γενετικών αλγορίθμων είναι πέρα από την ανθρώπινη γνωστική κατανόηση και το μεγαλύτερο μέρος της ανθρώπινης εξήγησης. Χρησιμοποιώντας ως παράδειγμα το να μάθεις σε ένα νευρωνικό δίκτυο, τι είναι μια γάτα, είναι πολύ απίθανο οι μαθηματικές πολυπλοκότητες αυτής της λειτουργίας να μπορούν να εξηγηθούν πλήρως σε

φυσική γλώσσα. Για παράδειγμα, μπορεί να ειπωθεί ότι χρησιμοποιήθηκε ένα νευρωνικό δίκτυο για την επίλυση ενός προβλήματος και ότι μπορεί να αναγνωρίσει τις γάτες με έναν ορισμένο βαθμό ακρίβειας. Το «γιατί», ωστόσο, σε σχέση με την έξοδο δεν μπορεί να εξηγηθεί (Greenstein, no date). Οι δικαστικές όμως αποφάσεις πρέπει να περιέχουν αιτιολογήσεις και λογικές εξηγήσεις του σκεπτικού τους. Για το λόγο αυτό οι προγραμματιστές λογισμικού πρέπει να δημιουργήσουν μια επεξηγηματική τεχνολογία που μπορεί να παρέχει εξηγήσεις για την τεχνολογική λογική πίσω από την αλγοριθμική λήψη αποφάσεων (Rubim Borges Fortes, 2020). Πράγματι σε ορισμένες περιπτώσεις, οι αποφάσεις του ευφυούς συστήματος μπορεί να είναι καλύτερα αξιόπιστες όταν χρησιμοποιούν ένα ενσωματωμένο σύστημα επεξήγησης, το οποίο εξηγεί στο επηρεαζόμενο άτομο πώς λήφθηκε μια απόφαση, αλλά για ορισμένους, το επίπεδο λεπτομέρειας που χρησιμοποιούν αυτά τα συστήματα επεξήγησης μπορεί να μην επαρκεί για να δικαιολογήσει εμπιστοσύνη στο σύστημα (Bolingford *et al.*, 2020). Άλλωστε, υποστηρίζεται ότι τα αποτελέσματα των μοντέλων πρέπει να λαμβάνονται υπόψη μόνο ως συστάσεις. Διαφορετικά, το νομικό σύστημα δεν θα είναι πλέον αξιόπιστο (Zhong *et al.*, 2020). Επιπλέον, υποστηρίζεται ότι οι προβλέψεις τέτοιων αλγορίθμων δεν μπορούν να βελτιώσουν την προβλεψιμότητα και τη συνέπεια της δικαστικής λήψης αποφάσεων με επιθυμητούς τρόπους. Επισημαίνεται ότι οι απλοί προγνωστικοί παράγοντες απόφασης, δηλαδή οι προβλέψεις που δεν μπορούν να εξηγήσουν τις προβλέψεις τους με όρους νομικής σημασίας, δεν πρέπει να χρησιμοποιούνται καθόλου από τους δικαστές ως εργαλεία υποστήριξης αποφάσεων. Τέτοιοι αλγόριθμοι δεν δίνουν καμία χρήσιμη πληροφορία στους δικαστές και μπορεί στην πραγματικότητα να είναι παραπλανητικοί και να προκαλούν πνευματική τεμπελιά. Εάν ένας αλγοριθμικός παράγοντας πρόβλεψης αποφάσεων δίνει οποιαδήποτε χρήσιμη πληροφορία στους δικαστές, δεν είναι στις προβλέψεις του αλλά στις εξηγήσεις του για αυτές τις προβλέψεις (Bex and Prakken, 2021). Εξάλλου, η απόδοση των συστημάτων πρόβλεψης δεν είναι σε καμία περίπτωση τέλεια (συνήθως λιγότερο από 80%). Επομένως δεν μπορεί να υπάρξει βεβαιότητα ότι το αποτέλεσμα θα είναι σωστό, πλην όμως στο πεδίο της δικαιοσύνης απαιτείται πολύ υψηλός βαθμός βεβαιότητας. Ως εκ τούτου, οι δικαστές θα πρέπει να διαμορφώσουν τη δική τους ανεξάρτητη γνώμη και χωρίς τις εξηγήσεις για τις προβλέψεις της μηχανής δεν έχουν κανένα λόγο να δώσουν βαρύτητα σ' αυτές (Mumford, Atkinson and Bench-Capon, 2021).

4.5.4. Επεξηγήσιμη τεχνητή νοημοσύνη

Από τα ανωτέρω προκύπτει, η σχέση της διαφάνειας με την κατανόηση και αυτής με την αξιοπιστία των διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης και την εμπιστοσύνη των ανθρώπων προς αυτήν. Σημαντική, άλλωστε, προκύπτει η έννοια της "επεξήγησης" και η συσχέτιση των επεξηγήσεων με τη διαφάνεια (Hacker *et al.*, 2020). Συγκεκριμένα, ο προαναφερόμενος χαρακτήρας του "μαύρου κουτιού" δημιουργεί προβλήματα σε διάφορα πεδία. Για παράδειγμα, όταν λαμβάνονται αποφάσεις σε ένα νοσοκομείο σχετικά με τη θεραπεία ασθενών ή στο δικαστήριο σχετικά με την καταδίκη ενός κατηγορούμενου, τέτοιες αποφάσεις θα πρέπει να είναι "εξηγήσιμες" (Emmert-

Streib, Yli-Harja and Dehmer, 2020). Ο Lipton, 2018, εντόπισε διαφορετικά είδη επεξηγήσεων που σχετίζονται με αλγόριθμους μηχανικής μάθησης ανάλογα με συγκεκριμένα χαρακτηριστικά του μοντέλου (όπως παρατίθεται στους Hacker *et al.*, 2020). Εξάλλου, το πεδίο της Επεξηγήσιμης Τεχνητής Νοημοσύνης (XAI) διερευνά μηχανισμούς για την εξήγηση ή την έκθεση των εσωτερικών λειτουργιών ή των αποτελεσμάτων των ευφυών συστημάτων για την υποστήριξη της κατανόησης και την αύξηση της εμπιστοσύνης (Ghajargar *et al.*, 2021). Πράγματι, όπως εκθέτουν οι Emmert-Streib, F et al (2020), η επεξηγήσιμη τεχνητή νοημοσύνη δεν είναι ένα νέο πεδίο, αλλά έχει ήδη αναγνωριστεί και συζητηθεί για έμπειρα συστήματα τη δεκαετία του 1980. Οι ίδιοι (Emmert-Streib, Yli-Harja and Dehmer, 2020), εκθέτοντας διάφορους ορισμούς που προτείνονται για το τί είναι επεξηγήσιμη τεχνητή νοημοσύνη και τους στόχους αυτής, εκθέτουν ότι σε, γενικές γραμμές, υπάρχει συμφωνία ότι ένα εξηγήσιμο σύστημα τεχνητής νοημοσύνης δεν πρέπει να είναι αδιαφανές (ή ένα μαύρο κουτί)· δεν πρέπει να είναι ένα σύστημα που κρύβει τον λειτουργικό του μηχανισμό. Επίσης, εκθέτουν ότι εάν ένα σύστημα δεν είναι αδιαφανές και μπορεί κανείς να καταλάβει ακόμη και πώς οι είσοδοι αντιστοιχίζονται μαθηματικά στις εξόδους, τότε το σύστημα είναι ερμηνεύσιμο. Συνολικά, αυτό συνεπάγεται διαφάνεια του μοντέλου. Οι όροι ερμηνευτικότητα και επεξηγηματικότητα (και μερικές φορές κατανοητότητα) χρησιμοποιούνται συχνά συνώνυμα, αν και η κοινότητα μηχανικής μάθησης φαίνεται να προτιμά τον πρώτο ενώ η κοινότητα τεχνητής νοημοσύνης προτιμά τον δεύτερο. Ωστόσο, έχει ήδη καταστεί σαφές ότι στην περίπτωση των αλγορίθμων μηχανικής μάθησης, επειδή αυτοί προσαρμόζουν επανειλημμένα τον τρόπο με τον οποίο ζυγίζουν τα δεδομένα για να βελτιώσουν την ακρίβεια των προβλέψεών τους, μπορεί να είναι δύσκολο να προσδιοριστεί πώς και γιατί φτάνουν στα αποτελέσματά τους (Deeks, 2019). Διαφαίνεται δε ότι μια ιδιαίτερα απλή αλλά χρήσιμη μορφή υποστήριξης αποφάσεων είναι η επεξηγήσιμη πρόβλεψη νομικών αποφάσεων. Εγγενής στα επεξηγήσιμα συστήματα πρόβλεψης αποτελεσμάτων είναι μια αντιστάθμιση μεταξύ της ποιότητας της εξήγησης και της προσπάθειας αναπαράστασης. Στο ένα άκρο, τα συστήματα που βασίζονται αποκλειστικά στη μηχανική μάθηση απαιτούν σχετικά μικρή προσπάθεια αναπαράστασης, αλλά συνήθως έχουν μικρή ή καθόλου επεξηγητική ικανότητα. Στο άλλο άκρο, συστήματα στα οποία τα γεγονότα έχουν αναπαρασταθεί με χειροκίνητη λειτουργία και οι νομικοί κανόνες αντιπροσωπεύονται σε εκτελέσιμη λογική, μπορεί να είναι ικανά να παράγουν εξηγήσεις με μεγάλη πιστότητα με τις ανθρώπινες εξηγήσεις, αλλά συνήθως απαιτούν την αναπαράσταση των νέων υποθέσεων με τους όρους αυτών των λειτουργιών και όχι ως κείμενο και συχνά απαιτούν υψηλά επίπεδα προσπάθειας ανάπτυξης για εφαρμογή σε κλίμακα. Μια βασική, επομένως, απαίτηση για εξηγήσιμα συστήματα προβλέψεων αποφάσεων είναι η βελτιστοποίηση της αντιστάθμισης μεταξύ της ποιότητας της επεξήγησης και της προσπάθειας αναπαράστασης, δηλαδή ο εντοπισμός προσεγγίσεων που μπορούν να παράγουν χρήσιμες και κατανοητές προβλέψεις από κείμενο με αρκετά χαμηλό κόστος μηχανής ώστε να επιτρέπουν την ανάπτυξη, την επαλήθευση και τη συντήρηση σε κλίμακα (Branting *et al.*, 2021). Επιπλέον, πρέπει να σημειωθεί ότι σε μοντέλα τελευταίας τεχνολογίας, η επεξήγηση και η απόδοση όσον αφορά την ακρίβεια πρόγνωσης είναι συχνά αντιστρόφως ανάλογες: τα εύκολα εξηγήσιμα μοντέλα

ενδέχεται να έχουν κακή απόδοση, ενώ τα μοντέλα με την καλύτερη επίδοση είναι τόσο πολύπλοκα που ακόμη και οι ειδικοί προγραμματιστές δεν μπορούν να εξηγήσουν ούτε διαισθητικά ούτε τεχνικά πώς προκύπτουν τα καλά αποτελέσματα. (Hacker *et al.*, 2020). Οι Emmert-Streib, F et al (2020), επισημαίνοντας ότι ο στόχος μιας εξηγήσιμης τεχνητής νοημοσύνης μπορεί να είναι μια ιδεαλιστική απαίτηση, υποστηρίζουν ότι το σύστημα τεχνητής νοημοσύνης μπορεί να μην χρειάζεται να είναι επεξηγήσιμο σε μια φυσική γλώσσα, εφόσον το σφάλμα γενίκευσης του δεν υπερβαίνει ένα αποδεκτό επίπεδο. Τονίζουν δε ότι έχει δοθεί η εντύπωση ότι οι μέθοδοι τεχνητής νοημοσύνης και μηχανικής μάθησης είναι πιο ισχυρές από τις φυσικές θεωρίες, επειδή μπορούν, λόγω της διαθεσιμότητας μεθόδων και δεδομένων, να φτάσουν σε περιοχές που είναι αποκλεισμένες για τη φυσική. Ωστόσο, πρέπει να γίνει κατανοητό ότι αυτές οι μέθοδοι δεν προορίζονται ως θεωρίες, αλλά απλά ως πρακτικά εργαλεία για την αντιμετώπιση πολύπλοκων φαινομένων. Οι ερωτήσεις που μπορούν να απαντήσουν πρέπει να μπορούν να βρεθούν μέσα στα δεδομένα. Κάθε έλλειψη τέτοιων ποιοτικών δεδομένων, π.χ. λόγω περιορισμένου μεγέθους δείγματος, μεταφράζεται άμεσα σε έλλειψη απαντήσεων και για αυτόν τον λόγο υπάρχει εγγενής αβεβαιότητα οποιουδήποτε συστήματος τεχνητής νοημοσύνης. Με τις ως άνω επισημάνσεις, οι ίδιοι προτείνουν ως εναλλακτική προσέγγιση, στο πρόβλημα της ερμηνευσιμότητας και της επεξήγησης των αποτελεσμάτων, να προτιμώνται οι μέθοδοι που δεν υποφέρουν από τέτοιους περιορισμούς, όπως τα δέντρα αποφάσεων. Συγκεκριμένα, υποστηρίζουν ότι πρέπει να αξιολογηθεί ποσοτικά η ομοιότητα ή η διαφορά μεταξύ δύο συστημάτων τεχνητής νοημοσύνης για να μπορεί να συγκριθεί ένα εξηγήσιμο με ένα μη εξηγήσιμο σύστημα τεχνητής νοημοσύνης και να αξιολογηθεί το όφελος του ενός έναντι του άλλου. Κατ' αυτόν τον τρόπο, ακόμα κι αν ένα εξηγήσιμο σύστημα τεχνητής νοημοσύνης δεν επιλύει πλήρως ένα δεδομένο πρόβλημα, πχ. σε σύγκριση με μια προσέγγιση βαθιάς μάθησης, μπορεί να είναι αρκετή η χρήση του. Επίσης, ακόμα και όταν ένα σύστημα τεχνητής νοημοσύνης από μόνο του δεν είναι εξηγήσιμο, η σύγκριση των δύο συστημάτων μπορεί να είναι κατανοητή. Περαιτέρω, όπως αναφέρει η Deeks, A. (2019), υπάρχει μια έντονη συζήτηση για το ποιες αλγοριθμικές αποφάσεις απαιτούν εξήγηση και ποιες μορφές πρέπει να λάβουν αυτές οι εξηγήσεις. Οι δικαστές θα αντιμετωπίσουν μία ποικιλία υποθέσεων στις οποίες θα πρέπει να απαιτούν εξηγήσεις για αλγοριθμικές αποφάσεις, συστάσεις και προβλέψεις και θα διαδραματίσουν θεμελιώδη ρόλο στη διαμόρφωση της φύσης και της μορφής της επεξηγήσιμης τεχνητής νοημοσύνης. Εξάλλου, η επεξηγήσιμη τεχνητή νοημοσύνη παρέχει μία σειρά από πλεονεκτήματα: Μπορεί να ενισχύσει την εμπιστοσύνη μεταξύ των ανθρώπων και του συστήματος, εντοπίζει περιπτώσεις στις οποίες το σύστημα φαίνεται να είναι προκατειλημμένο ή άδικο και ενισχύει τις γνώσεις μας για το πώς λειτουργεί ο κόσμος. Επίσης, σε νομικά πλαίσια, η επεξηγήσιμη τεχνητή νοημοσύνη μπορεί να ωφελήσει δικαστές που επιθυμούν να βασιστούν στους αλγόριθμους για υποστήριξη αποφάσεων, διαδίκους που επιδιώκουν να πείσουν τους δικαστές ότι η χρήση αλγορίθμων από μέρους τους μπορεί να δικαιολογηθεί και κατηγορούμενους που επιθυμούν να αμφισβητήσουν τις προβλέψεις σχετικά με την επικινδυνότητά τους. Από την άλλη, ωστόσο, μεριά, η επεξηγήσιμη τεχνητή νοημοσύνη δεν είναι χωρίς κόστος. Το πιο σημαντικό είναι ότι το να γίνει ένας

αλγόριθμος εξηγήσιμος μπορεί να οδηγήσει σε μείωση της ακρίβειάς του. Η επεξηγήσιμη τεχνητή νοημοσύνη μπορεί επίσης να καταπνίξει την καινοτομία, να αναγκάσει τους προγραμματιστές να αποκαλύψουν εμπορικά μυστικά και να επιβάλλουν υψηλό χρηματικό κόστος επειδή μπορεί να είναι ακριβή στην κατασκευή της. Επιπλέον, η Deeks, A. (2019), προτείνει εναλλακτικές λύσεις στις διάφορες προσεγγίσεις της επεξηγήσιμης τεχνητής νοημοσύνης. Μία εξ αυτών είναι η δημιουργία ενός δεύτερου συστήματος παράλληλα με το αρχικό μοντέλο «μαύρου κουτιού», που μερικές φορές ονομάζεται «υποκατάστατο μοντέλο». Ένα υποκατάστατο μοντέλο λειτουργεί αναλύοντας τα επιλεγμένα ζεύγη εισόδου και εξόδου, αλλά δεν έχει πρόσβαση στις εσωτερικές σταθμίσεις του ίδιου του μοντέλου. Για παράδειγμα, οι μελετητές κατασκεύασαν ένα δέντρο αποφάσεων που αντικατόπτριζε αποτελεσματικά τους υπολογισμούς ενός μοντέλου «μαύρου κουτιού» που προέβλεπε τον κίνδυνο διαβήτη των ασθενών. Το δέντρο αποφάσεων επέτρεψε στους επιστήμονες υπολογιστών να παρακολουθήσουν ποιους παράγοντες (όπως το επίπεδο χοληστερόλης, η εξάρτηση από τη νικοτίνη και το οίδημα) στάθμισε το μοντέλο του μαύρου κουτιού κατά τη διενέργεια των εκτιμήσεων κινδύνου. Στο νομικό πλαίσιο, αυτή η προσέγγιση μπορεί να συνεπάγεται τη δημιουργία ενός δέντρου αποφάσεων που αναπαριστάνει με ακρίβεια τις αποφάσεις των αλγορίθμων του μαύρου κουτιού ενός αυτοοδηγούμενου αυτοκινήτου σε μια υπόθεση ευθύνης προϊόντος, για παράδειγμα. Περαιτέρω, πρέπει να σημειωθεί ότι ενώ στην υπάρχουσα βιβλιογραφία για την εξηγήσιμη τεχνητή νοημοσύνη, η έννοια της διαφάνειας επικεντρώνεται στην εξήγηση της εσωτερικής λειτουργίας των αλγορίθμων ή στην ερμηνευσιμότητα των επιμέρους αποτελεσμάτων τους (Lipton 2018, Ribeiro et al . 2016, όπως παρατίθεται στους Loi, Ferrario and Viganò, 2020), προτείνεται και μία άλλη προσέγγιση για την αλγοριθμική διαφάνεια, που αναφέρεται ως διαφάνεια ως δημοσιότητα σχεδιασμού (Loi, Ferrario and Viganò, 2020), η οποία συνίσταται στην εξήγηση του σχεδιασμού των αλγορίθμων, με την επισήμανση ότι η εξήγηση σχεδιασμού ενός αλγορίθμου περιλαμβάνει «την κατανόηση του τι σχεδιάστηκε να κάνει ο αλγόριθμος, πώς σχεδιάστηκε για να το κάνει αυτό και γιατί σχεδιάστηκε με αυτόν τον συγκεκριμένο τρόπο αντί με κάποιον άλλο τρόπο» (Kroll 2018, όπως παρατίθεται στους Loi, Ferrario and Viganò, 2020). Τέλος, πρέπει να αναφερθεί ότι στην επισκοπούμενη βιβλιογραφία διερευνάτε η δυνατότητα αύξησης της εξήγησης των αποφάσεων που βασίζονται σε τεχνητή νοημοσύνη μέσω υπολογιστικής επιχειρηματολογίας. Εξ ορισμού, η επιχειρηματολογία είναι μια ορθολογική διαδικασία υποβολής αιτιολογιών υπέρ και κατά μιας δεδομένης θέσης, προκειμένου να επιλεγεί ένα τέτοιο συμπέρασμα που δικαιολογείται καλύτερα υπό το φως των διαθέσιμων αιτιολογιών. Τα πλαίσια επιχειρημάτων είναι ένα καλό παράδειγμα εξηγήσιμων τεχνικών τεχνητής νοημοσύνης. Ωστόσο, ενδέχεται να διαδραματίσουν ακόμη μεγαλύτερο ρόλο στο μέλλον στη δημιουργία επεξηγήσεων σε υβριδικά συστήματα τεχνητής νοημοσύνης. Τέτοια συστήματα συνδυάζουν μοντέλα μαύρου κουτιού (όπως τεχνητά νευρωνικά δίκτυα) με πρόσθετες εγκαταστάσεις επεξήγησης (π.χ. δέντρα αποφάσεων) (Araszkiewicz and Nalepa, 2019).

4.6. Προσωπικά δεδομένα και διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης. Δικαίωμα στην εξήγηση.

Σε αυτό το πλαίσιο, αξίζει να σημειωθεί ότι οι αλγόριθμοι μηχανικής μάθησης θα μπορούσαν να σχεδιαστούν έτσι ώστε να διασφαλίζουν τη συνεχή αναγνωσιμότητά τους από τους ανθρώπινους χειριστές, παρέχοντας, για παράδειγμα, έναν απολογισμό των αποφάσεών τους σε κάθε επίπεδο του δικτύου. Ο Γενικός Κανονισμός για την Προστασία Δεδομένων του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016, περιλαμβάνει, το δικαίωμα αυτού που επηρεάζεται από μια απόφαση που βασίζεται «αποκλειστικά σε αυτοματοποιημένη επεξεργασία» να «λάβει αιτιολόγηση της απόφασης που ελήφθη στο πλαίσιο της εν λόγω εκτίμησης και να αμφισβητήσει την απόφαση» (Chiao, 2019). Προβλέπει, πράγματι, ο Κανονισμός αυτό που ορισμένοι έχουν ονομάσει «δικαίωμα στην εξήγηση». Μερικοί μελετητές έχουν ερμηνεύσει τον Κανονισμό ότι απαιτεί από τους υπεύθυνους επεξεργασίας δεδομένων, οι οποίοι λαμβάνουν αποφάσεις για άτομα που βασίζονται «αποκλειστικά στην αυτοματοποιημένη επεξεργασία» να παρέχουν στα άτομα αυτά ουσιαστικές πληροφορίες σχετικά με τη λογική που εμπλέκεται σε αυτή την αυτοματοποιημένη λήψη αποφάσεων. Ωστόσο, παραμένει ασαφές τι ακριβώς απαιτεί ο Κανονισμός και τι μέτρα πρέπει να λάβουν τα κράτη και οι εταιρείες για να ανταποκριθούν σε αυτές τις απαιτήσεις (Deeks, 2019). Εξάλλου, το πρόβλημα της επεξήγησης των αποφάσεων που λαμβάνονται με τη βοήθεια εργαλείων μηχανικής μάθησης θεωρείται κυρίαρχο στη διασταύρωση της τεχνητής νοημοσύνης και του νόμου. Σήμερα, το μεγαλύτερο μέρος αυτής της συζήτησης επικεντρώνεται, στη νομική πλευρά, στη νομοθεσία περί προστασίας δεδομένων (Wachter et al. 2017; Selbst and Powles 2017; DoshiVelez 2017; Goodman and Flaxman 2016; Malgieri and Comandé 2017; Wischmeyer 2018, όπως παρατίθενται στους Hacker *et al.*, 2020). Υποστηρίζεται, όμως, ότι η επεξήγηση είναι ένας σημαντικός νομικός τύπος όχι μόνο στο δίκαιο προστασίας δεδομένων, αλλά και στο δίκαιο των συμβάσεων και των αδικοπραξιών. Καθώς οι νομικές απαιτήσεις βάσει του Κανονισμού είναι σε μεγάλο βαθμό ασαφείς, αυτή τη στιγμή, φαίνεται γόνιμη η διερεύνηση του ρόλου της επεξήγησης σε άλλους νομικούς τομείς. Έτσι, υποστηρίζεται ότι το δίκαιο των συμβάσεων και των αδικοπραξιών και όχι το δίκαιο περί προστασίας δεδομένων, μπορεί τελικά να επιβάλει νομικές απαιτήσεις για τη χρήση εξηγήσιμων μοντέλων μηχανικής εκμάθησης (Hacker *et al.*, 2020). Παρόλα αυτά, το δικαίωμα στην εξήγηση για μια απόφαση που λαμβάνεται από έναν αλγόριθμο, όπως ρυθμίζεται στον Κανονισμό, μπορεί να είναι ένα βήμα προς τη σωστή κατεύθυνση για την αύξηση της εμπιστοσύνης στις αλγοριθμικές αποφάσεις. Από την άποψη της δημόσιας λήψης αποφάσεων, αυτό είναι παρόμοιο με μια εξήγηση των δικαιωμάτων που γίνεται στο πλαίσιο της Ελευθερίας της Πληροφορίας όπου η διαφάνεια θεωρείται «ως ένα από τα προπύργια της δημοκρατίας, της φιλελεύθερης διακυβέρνησης, της λογοδοσίας και του περιορισμού της αυθαίρετης ή ιδιοτελούς ασκήσεως εξουσίας». Ωστόσο, όπως σημειώνουν οι Edwards και Veale, «ο μηχανισμός λογοδοσίας της ιδιωτικής λήψης αποφάσεων» είναι λιγότερο διαφανής λόγω των εμπορικών μυστικών και της προστασίας των δικαιωμάτων πνευματικής ιδιοκτησίας. Αναμφισβήτητα η διαφάνεια και η λογοδοσία είναι σημαντικές στη χρήση αλγοριθμικών αποφάσεων- εκεί όπου οι

αποφάσεις αυτές μπορεί να έχουν δυσμενείς επιπτώσεις σε ένα άτομο. Οι Edwards και Veale δηλώνουν ότι τα δικαιώματα διαφάνειας «παραμένουν στενά συνδεδεμένα με το ιδανικό του αποτελεσματικού ελέγχου της αλγοριθμικής λήψης αποφάσεων» και κοινωνικές αξίες όπως η «ανθρώπινη αξιοπρέπεια», ο «έλεγχος της πληροφορίας» και η «αυτονομία και ο σεβασμός» παίζουν ρόλο στον τρόπο με τον οποίο βλέπει η κοινωνία τις διαδικασίες λήψης αποφάσεων (όπως παρατίθεται στους Bolingford *et al.*, 2020). Περαιτέρω, σημαντική είναι η αναφορά στον ρόλο του νόμου ως μηχανισμού εξισορρόπησης αντικρουόμενων συμφερόντων, με τον οποίο επιτυγχάνεται μια ισορροπία μεταξύ των διαφόρων δικαιωμάτων και υποχρεώσεων, διαφορετικών ενδιαφερομένων, στα παραδοσιακά νομικά κείμενα. Η τεχνολογία, ωστόσο, διαταράσσει την κοινωνία από πολλές απόψεις και μια εκδήλωση αυτού του διασπαστικού χαρακτηριστικού είναι ότι θέτει εκτός συγχρονισμού την εξισορρόπηση των συμφερόντων που μπορεί να προέκυψε μέσω του παραδοσιακού νόμου. Ο Γενικός Κανονισμός για την Προστασία Δεδομένων είναι ένα παράδειγμα αυτής της πράξης εξισορρόπησης που εκτελείται από την παραδοσιακή νομοθεσία στο πλαίσιο της προστασίας δεδομένων. Αυτό το νομικό πλαίσιο επιχειρεί να εξισορροπήσει τα δικαιώματα πνευματικής ιδιοκτησίας με τα δικαιώματα που σχετίζονται με την ιδιωτική ζωή σε σχέση με την τεχνητή νοημοσύνη. Η αιτιολογική σκέψη 63, που επεκτείνεται στο άρθρο 22 που αφορά την τεχνητή νοημοσύνη, παρέχει στο υποκείμενο των δεδομένων το δικαίωμα να γνωρίζει και να λαμβάνει ανακοινώσεις σχετικά με τη λογική πίσω από κάθε επεξεργασία δεδομένων σε σχέση με την αυτοματοποιημένη λήψη αποφάσεων. Αυτό δυνητικά εκχωρεί στο υποκείμενο των δεδομένων το δικαίωμα για επεξήγηση της τεχνολογίας. Ωστόσο, αυτό το δικαίωμα μειώνεται στην ίδια αιτιολογική σκέψη όπου τα εμπορικά μυστικά και τα δικαιώματα πνευματικής ιδιοκτησίας υπερισχύουν της διαφάνειας. Τα ανωτέρω μπορούν να γίνουν αντιληπτά, δεδομένου ότι ένα δικαίωμα στην πληροφόρηση σχετικά με την επεξεργασία δεδομένων προσωπικού χαρακτήρα απαντάτε στο άρθρο 15 παρ. 1 (η) του Κανονισμού, το οποίο παρέχει στο υποκείμενο των δεδομένων δικαίωμα σε πληροφορίες σχετικά με τη λογική που ακολουθείται στην αυτοματοποιημένη επεξεργασία και τις συνέπειές της για το υποκείμενο των δεδομένων, ενώ έχει υποστηριχθεί ότι το «δικαίωμα στην πληροφόρηση» είναι το ίδιο με το «δικαίωμα στην εξήγηση». Επίσης, εδώ, μπορεί να υποστηριχθεί ότι η εξισορρόπηση της ιδιωτικής ζωής με την πνευματική ιδιοκτησία διαταράσσεται από τη φύση της ίδιας της τεχνολογίας - η τεχνητή νοημοσύνη δεν μπορεί να συγκριθεί με καμία τεχνολογία που προηγήθηκε και μπορεί να κριθεί ότι η διαφάνεια στην εσωτερική της λειτουργία είναι απολύτως απαραίτητη για την παροχή επαρκούς προστασίας από τις βλάβες της. Εξ ου και το επιχείρημα ότι τα δικαιώματα πνευματικής ιδιοκτησίας δυνητικά βοηθούν στη δημιουργία του μαύρου κουτιού της τεχνολογίας (Greenstein, no date). Επισημαίνοντας τα ανωτέρω, η Greenstein, no date, υπογραμμίζει το ευρύ πεδίο εφαρμογής του Κανονισμού και υποστηρίζει ότι δεν είναι αδιανόητο ότι θα έχει σημασία σε πολλές περιπτώσεις όπου η τεχνητή νοημοσύνη χρησιμοποιείται για τη λήψη αποφάσεων της διοίκησης ή ακόμη και στο δικαστικό σύστημα. Χρησιμοποιώντας ως παράδειγμα μία υπόθεση ολλανδικού Δικαστηρίου που αφορούσε το ψηφιακό σύστημα SyRI (System Risk Indication), το οποίο χρησιμοποιείται στις Κάτω Χώρες, υπογραμμίζει ότι η εν λόγω υπόθεση

απεικόνισε τον Κανονισμό, αναφέροντας τις αρχές προστασίας των δεδομένων, της διαφάνειας, του περιορισμού του σκοπού και της ελαχιστοποίησης των δεδομένων, με τις δύο τελευταίες να αποτελούν έκφραση της αρχής της αναλογικότητας. Σημαντικό, επίσης, είναι να αναφερθεί ότι η ανισορροπία που δημιουργεί η τεχνητή νοημοσύνη σε σχέση με διαφορετικά συμφέροντα μπορεί να διορθωθεί με διάφορους μηχανισμούς. Ειδικότερα, "έμπιστα τρίτα μέρη" ενδέχεται να διαδραματίσουν εποικοδομητικό ρόλο στη διασφάλιση ότι οι αλγόριθμοι αναπτύσσονται και εφαρμόζονται σύμφωνα με τις αξίες του κράτους δικαίου. Για παράδειγμα, στο Ηνωμένο Βασίλειο, η Νομική Επιτροπή για τη χρήση αλγορίθμων στο σύστημα δικαιοσύνης δημοσίευσε πρόσφατα μια έκθεση όπου μία από τις συστάσεις ήταν η δημιουργία ενός Εθνικού Μητρώου Αλγοριθμικών Συστημάτων, όπου διάφορες πτυχές σε σχέση με τους αλγόριθμους που χρησιμοποιούνται στο σύστημα ποινικής δικαιοσύνης θα μπορούσε να ελεγχθεί και να επαληθευτεί. Αυτή η ιδέα αντικατοπτρίζεται επίσης στην πρόταση Κανονισμού για την τεχνητή νοημοσύνη που δημοσιεύθηκε πρόσφατα από την Ευρωπαϊκή Επιτροπή. Εδώ, σε σχέση με την τεχνητή νοημοσύνη υψηλού κινδύνου, η πρόταση κανονισμού δημιουργεί τον μηχανισμό με τον οποίο αυτά τα συστήματα τεχνητής νοημοσύνης πρέπει να καταχωρούνται σε μια βάση δεδομένων της ΕΕ (άρθρο 51) που έχει συσταθεί μέσω συνεργασίας μεταξύ των κρατών μελών (άρθρο 60). Υποστηρίζεται ότι η χρήση έμπιστων τρίτων μερών δυνητικά βελτιώνει την εικόνα της πολυπλοκότητας της τεχνητής νοημοσύνης, ενώ ταυτόχρονα αποτρέπει τη γενική πληροφόρηση του κοινού, διατηρώντας έτσι τα συμφέροντα που προστατεύονται από το δικαίωμα της πνευματικής ιδιοκτησίας (Greenstein, no date). Περαιτέρω, πρέπει να επισημανθεί ότι το καθεστώς προστασίας προσωπικών δεδομένων δεν επαρκεί για την αντιμετώπιση όλων των προκλήσεων όσον αφορά τη διασφάλιση της συμμόρφωσης των συστημάτων τεχνητής νοημοσύνης με τα ανθρώπινα δικαιώματα. Οι επιπτώσεις στα ανθρώπινα δικαιώματα είναι αναγκαστικά πολλαπλές, όπως ορθά αναγνωρίζει η Επιτροπή Εμπειρογνωμόνων για τους Διαμεσολαβητές του Διαδικτύου (MSI-NET) στο Συμβούλιου της Ευρώπης (Završnik, 2020). Επιπλέον, αναφορικά με τις τεχνολογίες προγνωστικής αστυνόμευσης, οι οποίες είναι πιθανό να αποκτήσουν αυξανόμενη δημοτικότητα τα επόμενα χρόνια, καθώς υπόσχονται πιο αποτελεσματική αστυνόμευση με χαμηλότερο κόστος, πρέπει να αναφερθεί ότι η εφαρμογή της νομοθεσίας περί προστασίας δεδομένων στις τεχνολογίες αυτές είναι αβέβαιη. Ακόμη, όμως, κι αν ισχύουν οι κανόνες προστασίας δεδομένων, ενδέχεται να μην βοηθήσουν ιδιαίτερα εκείνους των οποίων η τύχη καθορίζεται από αυτές. Μπορεί να φανεί ότι ο νόμος περί προστασίας δεδομένων δεν εμποδίζει τη χρήση συστημικών ή εξατομικευμένων εφαρμογών προγνωστικής αστυνόμευσης, εάν τέτοιες εφαρμογές προβλέπονται από νόμο (Lynskey, 2019). Η Lynskey, 2019, αναφερόμενη στις εν λόγω τεχνολογίες χρησιμοποιεί ως παράδειγμα το μοντέλο HART (Harm Assessment Risk Tool), το οποίο έχει αναπτύξει, στο Ηνωμένο Βασίλειο, η Αστυνομία της Durham σε συνεργασία με ερευνητές της στατιστικής στο Πανεπιστήμιο του Κέιμπριτζ, για να αξιολογήσει τον κίνδυνο μελλοντικής παράβασης και να παράσχει στους παραβάτες που ταξινομούνται ως «μέτριου κινδύνου» την ευκαιρία να συμμετέχουν σε πρόγραμμα αποκατάστασης. Τα δεδομένα εισόδου, σ' αυτό το εργαλείο (τριάντα τέσσερις προγνωστικοί παράγοντες που βασίζονται

στο προσωπικό ιστορικό του δράστη) είναι προσωπικά δεδομένα, καθώς το περιεχόμενό τους αφορά ένα ταυτοποιημένο πρόσωπο, ενώ τα δεδομένα εξόδου - μια προτεινόμενη πιθανότητα επανάληψης του αδικήματος - είναι προσωπικά δεδομένα, καθώς ο σκοπός και η επίδρασή τους είναι να επηρεάζουν τις μελλοντικές προοπτικές ενός ταυτοποιημένου προσώπου. Είναι ίσως λιγότερο σαφές ότι το «μεσαίο» στάδιο επεξεργασίας – όπου ο αλγόριθμος του προγράμματος εφαρμόζεται στα δεδομένα εισόδου – είναι προσωπικά δεδομένα. Τελικά, ακόμη και αν κριθεί ότι αυτά τα μοντέλα εμπίπτουν στο πεδίο εφαρμογής των κανόνων, η προστασία που προσφέρουν στα άτομα που επηρεάζονται από μια τέτοια επεξεργασία είναι επισφαλής.

4.7. Λογοδοσία κατά τη διαδικασία λήψης αποφάσεων τεχνητής νοημοσύνης. Ευθύνη και Έλεγχος

Με τα ανωτέρω, την αδιαφάνεια των συστημάτων τεχνητής νοημοσύνης, την έλλειψη κατανόησης της λειτουργίας της από τους δικαστές και εισαγγελείς και (έλλειψη) αιτιολογίας των αποφάσεών της, συνδέεται μια ακόμα αξία από αυτές που, όπως προαναφέρθηκε, προσδιορίζονται ως σχετιζόμενες με τους αλγόριθμους στον εξεταζόμενο τομέα της δικαιοσύνης και της ασφάλειας: η λογοδοσία. Συγκεκριμένα, η αδιαφάνεια μπορεί να οδηγήσει σε έλλειψη ευθύνης και λογοδοσίας για τις αποφάσεις που λαμβάνονται. Σημαντικές, άλλωστε είναι οι επιπτώσεις των αλγόριθμων στις σχετιζόμενες αρχές και, συνεπώς, και στην αρχή της λογοδοσίας (Hayes, van de Poel and Steen, 2020). Κατά αυτόν δε τον τρόπο διακυβεύεται η αξιοπιστία των αλγόριθμων και κατ' αποτέλεσμα, με την εφαρμογή τους στον τομέα της δικαιοσύνης, η αξιοπιστία ολόκληρου του συστήματος απονομής της. Έτσι, οι τεχνολογικές λύσεις στον τομέα της δικαιοσύνης πρέπει να δύναται να υπόκεινται σε έλεγχο και να είναι υπεύθυνες. Ωστόσο, υπάρχουν πράγματα που προκαλούν ανησυχία, συγκεκριμένα η αυξανόμενη εξάρτηση από εταιρείες, συστήματα και μέσα που δεν βασίζονται στο κράτος δικαίου και τα ανθρώπινα δικαιώματα και ενδέχεται να μην υπόκεινται στην παραδοσιακή λογοδοσία των δημοκρατικών θεσμών ('Justice in the Digital Age: Technological Solutions, Hidden Threats and Enticing Opportunities', 2021). Γεννάται, λοιπόν, το ερώτημα πως μπορεί να διασφαλιστεί η λογοδοσία σε τέτοια περιβάλλοντα. Ειδικότερα, στο πλαίσιο της ποινικής δικαιοσύνης, η δημόσια λογοδοσία, με την έννοια της διασφάλισης ότι οι αποφάσεις είναι τόσο πιθανόν να είναι σωστές όσο μπορούμε να διαχειριστούμε, πιθανώς δεν ενισχύεται καλύτερα εάν οι κατηγορούμενοι αμφισβητούν τις τεχνικές λεπτομέρειες ενός αλγορίθμου κατά τη διάρκεια της ποινικής διαδικασίας. Ούτε οι δικηγόροι ούτε οι δικαστές, σε τελική ανάλυση, είναι πιθανό να έχουν την απαιτούμενη τεχνική εμπειρογνωμοσύνη για να αξιολογήσουν τη στατιστική αξιοπιστία ενός μέσου αξιολόγησης κινδύνου, ούτε το να βασιζόμαστε στις ιδιαιτερότητες της δικαστικής διαδικασίας και της διαπραγμάτευσης αποτελεί καλό μέσο για την εξασφάλιση συνεπούς αξιοπιστίας στα τεχνικά μέσα. Η λογοδοσία –με την έννοια της διασφάλισης της διαρκούς αξιοπιστίας ενός αλγοριθμικού οργάνου– πιθανότατα εξυπηρετείται καλύτερα με την ανάθεση αυτού του καθήκοντος στα χέρια ενός εξειδικευμένου ρυθμιστικού φορέα με συγκεκριμένη εντολή και την απαιτούμενη τεχνική εμπειρογνωμοσύνη. Σε τομείς

που απαιτούν υψηλή τεχνική εμπειρογνωμοσύνη, η ρυθμιστική εποπτεία μπορεί να αποδειχθεί πιο ικανοποιητικό μέσο για τη διασφάλιση της λογοδοσίας από ό,τι οι διαδικασίες ιδιωτών ενώπιον δικαστηρίων που δεν διαθέτουν ειδικές γνώσεις (Chiao, 2019). Εξάλλου, πρέπει να καταστεί σαφές ότι ο ανθρώπινος έλεγχος είναι απαραίτητος σε όλες τις φάσεις των διαδικασιών λήψης αποφάσεων τεχνητής νοημοσύνης. Πρώτα απ' όλα, οι χρήστες πρέπει να καθορίσουν τι πρέπει να κάνει η τεχνητή νοημοσύνη, πώς μετράται και αξιολογείται· πρέπει να υπάρχουν συνεχείς δοκιμές για να διαπιστωθεί ότι η τεχνητή νοημοσύνη εξακολουθεί να κάνει ό,τι πρέπει· το σύστημα πρέπει να σχεδιαστεί με τέτοιο τρόπο ώστε να μπορεί να προσαρμόζεται εύκολα και σθεναρά, και είναι απαραίτητος ο συνεχής έλεγχος. Υποστηρίζεται δε ότι αυτός ο έλεγχος, παρά την ανεξαρτησία της δικαιοσύνης, δεν πρέπει να είναι εσωτερικός (εντός του δικαστικού σώματος), αλλά εξωτερικός, καθώς με έναν εξωτερικό έλεγχο το δικαστικό σώμα μπορεί να είναι πιο διαφανές. Αυτό θα δημιουργήσει περισσότερη αξιοπιστία και εμπιστοσύνη από έναν εσωτερικό έλεγχο (Realing, 2020). Τίθεται, έτσι, το ερώτημα, πως επιτυγχάνεται διαφάνεια και υπευθυνότητα χωρίς να παραβιάζεται το απόρρητο ή τα δικαιώματα πνευματικής ιδιοκτησίας; Οι Kroll et al υποστηρίζουν ότι η αποκάλυψη του πηγαίου κώδικα δεν είναι η λύση και μπορεί να «δημιουργήσει από μόνη της βλάβη» (όπως παρατίθεται στους Bolingford *et al.*, 2020). Η αποκάλυψη κώδικα μπορεί ακόμη και να οδηγήσει σε «παιχνίδι» του συστήματος, όπου κάποιοι προσπαθούν να σαμποτάρουν την αποτελεσματικότητα και τη δικαιοσύνη των αλγορίθμων. Οι συγγραφείς υποστηρίζουν ότι η λογοδοσία μπορεί να επιτευχθεί με έλεγχο και αναζήτηση στις εξωτερικές εισροές και εκροές της διαδικασίας της απόφασης (Bolingford *et al.*, 2020). Ωστόσο, οι κοινωνίες μας θα πρέπει να παρακολουθούν και να αποτρέπουν την αλγοριθμική αδικία, απαιτώντας μεγαλύτερη ευθύνη από όσους συλλέγουν δεδομένα, δημιουργούν συστήματα και θέτουν τους κανόνες. Η έλλειψη διαφάνειας είναι μέρος του προβλήματος λόγω του γεγονότος ότι οι εταιρείες τεχνολογίας κρατούν μυστικούς τους αλγόριθμούς τους. Μία πιθανή απάντηση για τον κανονιστικό έλεγχο των αλγορίθμων συνίσταται στον έλεγχο για τη διασφάλιση της ασφάλειας και της δικαιοσύνης της αλγοριθμικής λήψης αποφάσεων. Σε σύγκριση με την έκκληση για πλήρη δημοσιότητα και τεχνητή νοημοσύνη ανοιχτού κώδικα, ο αλγοριθμικός έλεγχος μπορεί να εκτελείται από μία ελεγχόμενη και διακριτή ομάδα εμπειρογνωμόνων που μπορούν να παρέμβουν για να διορθώσουν τους κανόνες λήψης αποφάσεων χωρίς να αποκαλύψουν τον κώδικα και άλλες ιδιόκτητες πληροφορίες στο ευρύ κοινό. Ο έλεγχος θέτει μια σειρά από προκλήσεις, επειδή οι εξελιγμένοι αλγόριθμοι μπορεί να μην αποκαλύψουν τα πραγματικά τους αποτελέσματα κατά τη διάρκεια της δοκιμής, μπορεί να είναι σε θέση να παρακάμπτουν τις συστάσεις των ελεγκτών, δημιουργώντας συνδέσμους σε σύνολα δεδομένων και οι δημόσιες αρχές ενδέχεται να μην είναι σε θέση επιβλέπουν την ανάπτυξή τους και να συμβαδίζουν με τη βιομηχανία της τεχνολογίας (Rubim Borges Fortes, 2020). Εκτός από την έννοια του ελέγχου, η λογοδοσία εμπεριέχει και την έννοια της ευθύνης, δεδομένου μάλιστα ότι στις συγκεκριμένες διαδικασίες λήψης αποφάσεως οι εμπλεκόμενοι είναι πολλοί. Συγκεκριμένα, πολλοί μπορεί να εμπλέκονται σε ένα δίκτυο ή σειρά γεγονότων (από το σχεδιασμό έως την υλοποίηση και την ανάπτυξη), που οδηγούν σε κάποιο προβληματικό

αποτέλεσμα, από τους παράγοντες που εμπλέκονται στη δημιουργία του αλγόριθμου (στελέχη και μηχανικοί λογισμικού ή επιστήμονες δεδομένων), τους υπεύθυνους χάραξης πολιτικής που αναθέτουν ή δίνουν άδεια στους αλγορίθμους, μέχρι τους χρήστες των αλγορίθμων (για παράδειγμα, αστυνομικοί που είτε είναι εν δράσει είτε αναπτύσσουν την στρατηγική τους στα αστυνομικά τμήματα) (Nissenbaum 1996, 28–32, όπως πατατίθεται στους Hayes, van de Poel and Steen, 2020). Ενδέχεται να προκύψουν σφάλματα σε ένα λογισμικό που θα μπορούσαν να ήταν απρόβλεπτα και τα οποία μπορεί να επιβαρύνουν την εφαρμογή της λογοδοσίας (Nissenbaum 1996, 32-34, όπως παρατίθεται στους Hayes, van de Poel and Steen, 2020). Αν θυμηθούμε το πρόβλημα της επιστημικής εξάρτησης ή της υποδούλωσης, θεωρούμε υπεύθυνο έναν αστυνομικό που πυροβολεί χωρίς να πρέπει, βάσει των πληροφοριών που έχει λάβει από ένα τεχνούργημα ή εκείνους που σχεδίασαν ή έκαναν συντήρηση στο τεχνούργημα; Σε αυτήν την περίπτωση, μπορούμε να υποστηρίξουμε ότι όλοι όσοι εμπλέκονται με κάποιο τρόπο στο δίκτυο που οδήγησε στη ζημία πρέπει να λογοδοτήσουν γι' αυτό. Ωστόσο χρειαζόμαστε ακόμη ικανοποιητικές πληροφορίες σχετικά με τα στοιχεία που κρύβονται πίσω από αυτό το δίκτυο για να καταλογίσουμε δίκαια την ευθύνη (Hayes, van de Poel and Steen, 2020). Διαφαίνεται, στο σημείο αυτό, και πάλι, η ανάγκη για διαφάνεια των συστημάτων.

Σχετιζόμενο με την έννοια της ευθύνης και συνεπώς και της λογοδοσίας είναι και το ζήτημα των τυχόν επιπτώσεων που μπορεί να έχουν οι συγκεκριμένες διαδικασίες λήψης αποφάσεως στους ανθρώπους λήπτες αποφάσεων, όσον αφορά το ενδεχόμενο να αντιμετωπίσουν τους αλγορίθμους ως εργαλεία αποφυγής ευθυνών. Στο πλαίσιο της ποινικής δικαιοσύνης, για παράδειγμα, οι υπεύθυνοι λήψης αποφάσεων —εισαγγελείς, δικαστές, κ.λ.π.— δεν μπορούν να είναι σίγουροι για την πιθανότητα ένα άτομο να επαναλάβει το έγκλημα μόλις αποφυλακιστεί ή αφεθεί ελεύθερο με εγγύηση. Απλώς δεν υπάρχει τρόπος να υπολογιστεί αυτό με τον κατάλληλο τρόπο, γι' αυτό οι κανόνες και οι διαδικασίες είναι ζωτικής σημασίας. Για το λόγο αυτό, επί δεκαετίες, οι δικαστές και άλλοι παράγοντες του συστήματος της ποινικής δικαιοσύνης ακολούθησαν αυστηρούς κανόνες και τυπικές διαδικασίες σε αυτές τις υποθέσεις και βασίστηκαν κυρίως σε ειδικές γνώσεις (κυρίως που παρέχονται από αναφορές ψυχολόγων, κοινωνικών λειτουργών και άλλων) για να αιτιολογήσουν την απόφασή τους. Με αλγοριθμικά εργαλεία, αυτή η κατάσταση θεμελιώδους αβεβαιότητας μετατρέπεται σε κατάσταση στατιστικού κινδύνου. Εάν τα στατιστικά στοιχεία που δημιουργούνται από αλγόριθμους παρέχουν συγκεντρωτικές βαθμολογίες κινδύνου που υποδεικνύουν στατιστικό κίνδυνο υποτροπής, το πλαίσιο απόφασης στο οποίο κάποιος παράγοντας του συστήματος της ποινικής δικαιοσύνης πρέπει να αποφασίσει αλλάζει από μια κατάσταση αβεβαιότητας σε στατιστικό κίνδυνο. Αυτός είναι ένας ιδιαίτερος εκτεταμένος μετασχηματισμός της κατάστασης λήψης αποφάσεων: εάν κάποιος από τους ως άνω παράγοντες δεν χρειάζεται να συλλέξει ο ίδιος τα δεδομένα, αλλά λαμβάνει ένα αποτέλεσμα από έναν αλγόριθμο βαθμολόγησης ή έναν ταξινομητή που είναι εύκολο να ερμηνευτεί (Hartmann and Wenzelburger, 2021). Στην πραγματικότητα, ενώ είναι σαφές για έναν κοινωνικό επιστήμονα ότι τέτοιες «αξιολογήσεις κινδύνου αποδίδουν πιθανότητες, όχι βεβαιότητες, και ότι μετρούν συσχετίσεις και όχι αιτίες» (Završnik, 2019, όπως παρατίθεται στους Hartmann and Wenzelburger,

2021), αυτό μπορεί να είναι πολύ λιγότερο σαφές για έναν επαγγελματία που είναι ευτυχής να λάβει πρόσθετες πληροφορίες. Τέτοιες πληροφορίες δεν μπορούν να αγνοηθούν, αλλά αντιπροσωπεύουν μια άγκυρα βοήθειας για οποιαδήποτε περαιτέρω ερμηνεία από τον άνθρωπο. Αυτή η εικόνα σχετίζεται και με τη λεγόμενη «προκατάληψη αυτοματισμού» (Dzindolet et al. 2003, όπως παρατίθεται στους Hartmann and Wenzelburger, 2021), σύμφωνα με την οποία οι άνθρωποι φαίνεται να θεωρούν τις αποφάσεις που παράγονται ή υποστηρίζονται από υπολογιστές ως υπερβολικά αξιόπιστες. Πράγματι, μια σοβαρή ανησυχία σχετικά με τους αλγόριθμους μηχανικής μάθησης είναι ότι παράγουν "προκατάληψη αυτοματισμού" - την τάση να αποδέχονται οι άνθρωποι αδικαιολόγητα τη σύσταση μιας μηχανής. Η Deeks, 2019 αναφέρει ότι η παροχή της επεξηγήσιμης τεχνητής νοημοσύνης στους δικαστές μπορεί να τους οδηγήσει να αμφισβητήσουν τα συμπεράσματα ενός αλγορίθμου με τρόπο που να τους βοηθά να αποφύγουν να ενδώσουν στην «προκατάληψη αυτοματισμού» (automation bias). Την ισχύ αυτής (της «προκατάληψης αυτοματισμού»), πρέπει να σημειωθεί ότι, αγνόησε το Ανώτατο Δικαστήριο του Ουισκόνσιν, στην προαναφερόμενη απόφαση Loomis v. Wisconsin. Με τον ισχυρισμό ότι το κατώτερο δικαστήριο είχε τη δυνατότητα να απομακρυνθεί από την προτεινόμενη αλγοριθμική αξιολόγηση κινδύνου, το Δικαστήριο αγνόησε την κοινωνική ψυχολογία και έρευνα αλληλεπίδρασης ανθρώπου-υπολογιστή, που δείχνει ότι από τη στιγμή που ένα εργαλείο υψηλής τεχνολογίας προσφέρει μια σύσταση, γίνεται εξαιρετικά επαχθές για έναν άνθρωπο που λαμβάνει αποφάσεις να αντικρούσει μια τέτοια «σύσταση» (Završnik, 2020). Ωστόσο, υποστηρίζεται ότι η απλή ύπαρξη μιας βαθμολογίας αξιολόγησης κινδύνου, ακόμη και ως συμπληρωματική μεθοδολογία, μπορεί να επηρεάσει τον άνθρωπο που λαμβάνει αποφάσεις. Οι Eckhouse et al., 2019, αναφέρουν ότι μια προγνωστική βαθμολογία κινδύνου μπορεί να επηρεάσει την απόφαση ενός δικαστή, δίνοντας βάρος στην πιθανή υποτροπή πέρα από άλλους παράγοντες (όπως παρατίθεται στην McKay, 2020). Για παράδειγμα, ο Carlson (2017) παραθέτει μια περίπτωση όπου η βαθμολογία COMPAS ήταν τόσο υψηλή που ο δικαστής ανέτρεψε τον δικαστικό διακανομισμό και καταδίκασε τον δράστη σε δύο χρόνια, ενώ ο ίδιος αναγνώρισε ότι χωρίς την αξιολόγηση κινδύνου, θα είχε επιβάλει μία ποινή μόνο ενός έτους (όπως παρατίθεται στην McKay, 2020).

5. Επίλογος

5. 1. Σύνοψη – Συμπεράσματα:

Αν θα θέλαμε να αναζητήσουμε τη μορφή και τις δυνατότητες που μπορεί να δώσει η τεχνητή νοημοσύνη στην ποινική και πολιτική δίκη, εύκολα θα διαπιστώναμε ότι το μέλλον δεν θα είχε καμία σχέση με το παρόν, με τις παρούσες διαδικασίες, με τις συνθήκες και τον τρόπο εργασίας των εργαζομένων στους κλάδους αυτούς της δικαιοσύνης, δικηγόρους, γραμματείς, δικαστές και εισαγγελείς, καθώς και τις συνθήκες που αντιμετωπίζουν οι διάδικοι, οι μάρτυρες και γενικά όλα τα εμπλεκόμενα άτομα σε μια ποινική και πολιτική διαδικασία. Είναι αδιαμφισβήτητο ότι πολλά είναι τα εργαλεία που μπορούν να βοηθήσουν τη δικαιοσύνη να γίνει πιο αποτελεσματική. Η υιοθέτηση αυτών δεν φαίνεται να συναντά αντιδράσεις όταν πρόκειται για εργαλεία που δεν εμπλέκονται στη διαδικασία λήψης αποφάσεων από τους δικαστές και εισαγγελείς κατά την άσκηση του δικαιοδοτικού τους έργου. Όταν, όμως, πρόκειται για την άσκηση του δικαιοδοτικού έργου, για τον πυρήνα δηλαδή της απονομής της δικαιοσύνης, οι αντιδράσεις και ενστάσεις είναι πολλές. Η δικαιοδοτική, εξάλλου, κρίση αφορά και σε εργαλεία που συμμετέχουν σ' αυτή, όπως εργαλεία αξιολόγησης κινδύνου και βαθμολογίας υποτροπής, τα οποία μπορούν να χρησιμοποιηθούν στην επιμέτρηση της ποινής, στην αναστολή και τον καθορισμό του χρόνου πραγματικής έκτισης της ποινής, στο καθεστώς μετά την αποφυλάκιση. Τα ζητήματα που δημιουργούνται από τη συμμετοχή εργαλείων τεχνητής νοημοσύνης κατά το δικαιοδοτικό έργο των δικαστών και εισαγγελέων, στο οποίο συμμετέχουν και συμπληρωματικές διαδικασίες, όπως αυτές που προαναφέρθηκαν, αναφορικά με την ποινική δίκη ή διαδικασίες που μπορεί να αφορούν στην εξέταση μαρτύρων και αξιολόγηση αποδεικτικών μέσων, αναφορικά με την ποινική και πολιτική δίκη, αφορούν σε μεγάλη έκταση σε ζητήματα προκατάληψης, αδιαφάνειας και λογοδοσίας. Συγκεκριμένα, σημαντικό είναι το ζήτημα, κατά πόσο οι συγκεκριμένες διαδικασίες θα είναι δίκαιες, καθώς οι αλγόριθμοι μπορεί να ενσωματώνουν προκαταλήψεις που επηρεάζουν την ακρίβειά τους, ενώ σημαντικές είναι και οι ενστάσεις που τίθενται λόγω της σύγκρουσης τους με την αρχή της εξατομικευμένης δικαιοσύνης. Οσοδήποτε, εξάλλου, ακριβείς και αν αποδειχθούν οι αλγόριθμοι, σημαντικό παραμένει το ζήτημα της αδιαφάνειας των συστημάτων τεχνητής νοημοσύνης, η έλλειψη κατανόησης της λειτουργίας της από τους δικαστές και εισαγγελείς και η (έλλειψη) αιτιολογίας των αποφάσεών της. Επίσης, η αδιαφάνεια μπορεί να οδηγήσει σε έλλειψη ευθύνης και λογοδοσίας για τις αποφάσεις που λαμβάνονται, ενώ συνδέεται με τους κανόνες της δίκαιης δίκης και των δικαιωμάτων του ανθρώπου.

Επιπλέον, οποιαδήποτε σύγκριση μεταξύ της διαδικασίας λήψης αποφάσεων από ανθρώπους (δικαστές, εισαγγελείς αλλά και δικηγόρους) με τις διαδικασίες της τεχνητής νοημοσύνης, μπορεί να αποβεί υπέρ των τελευταίων και να προκύψει ότι οι αλγόριθμοι είναι πιο ακριβείς, αξιόπιστοι και δίκαιοι από τους ανθρώπους. Η απάντηση, ωστόσο, στο ερώτημα κατά πόσο οι τελευταίοι θα αντικαταστήσουν τελικά τους ανθρώπους δικαστές, εισαγγελείς και δικηγόρους, θα πρέπει να απαντηθεί αφού γίνει κατανοητό ότι στον συγκεκριμένο τομέα της δικαιοσύνης οι τεχνικές ανάλυσης λήψης αποφάσεων είναι δύσκολο να

αυτοματοποιηθούν (Zadgaonkar and Agrawal, 2021). Υπάρχει, βέβαια, ένας μεγάλος αριθμός υποθέσεων που δεν χρήζουν της ίδιας ανάγκης κατανόησης, δεδομένου του μικρού βαθμού πολυπλοκότητάς τους και της προβλέψιμης έκβασής τους (Realing, 2020). Σε αυτές τις περιπτώσεις είναι εύκολο να διαπιστωθεί η ορθότητα του αποτελέσματος μιας απόφασης. Στις πολύπλοκες όμως υποθέσεις, με αντικρουόμενους ισχυρισμούς, ο δικανικός συλλογισμός είναι περίπλοκος. Σε αυτές τις περιπτώσεις, όπως και στις ποινικές διαδικασίες, ο δικανικός συλλογισμός πρέπει να είναι διαφανής και η απόφαση αιτιολογημένη κατά τρόπο που να μπορεί ευχερώς να διαπιστωθεί ή αμφισβητηθεί η ορθότητά της. Κατ' αυτόν τον τρόπο, εφόσον οι δυνατότητες της τεχνητής νοημοσύνης δεν δύνανται να προσφέρουν ένα τέτοιο επίπεδο κατανόησης της διαδικασίας της και αιτιολογίας των αποτελεσμάτων αυτής δεν δύναται να αντικαταστήσει την ανθρώπινη απόφαση, η οποία είναι αποτέλεσμα ανθρώπινης κρίσης με βάση το νόμο, με δικανικό συλλογισμό, διαφάνεια και ενσυναίσθηση. Αναφορικά, μάλιστα, με τις εργασίες που ήδη μπορεί να εκτελέσει η τεχνητή νοημοσύνη, κυρίως, στον τομέα της ποινικής δικαιοσύνης και ειδικότερα, αναφορικά με τις τεχνολογίες πρόβλεψης και εκτίμησης κινδύνου, για τις οποίες εγείρονται πολλές επιφυλάξεις, η δυνατότητα κατανόησης και αιτιολόγησης των αποτελεσμάτων που αυτές προσφέρουν, μπορεί να βοηθήσει τους δικαστές και εισαγγελείς στο έργο τους και όχι να αντικαταστήσει τη λήψη της απόφασης από αυτούς. Η τελική απόφαση, πρέπει να είναι έργο ανθρώπινο. Μέσα από την κατανόηση αυτών των εργαλείων και των αποτελεσμάτων τους, ένας άνθρωπος δικαστής και εισαγγελέας μπορεί να αποφασίσει αν μπορεί να λάβει υπόψη το αποτέλεσμα του αλγόριθμου, αν αυτό δεν στηρίζεται σε προκατειλημμένα και ανακριβή δεδομένα· μπορεί ακόμα, στην περίπτωση που το αποτέλεσμα είναι θετικό, να επισημάνει, συγκρίνοντας τα δεδομένα στα οποία στηρίχθηκε ο αλγόριθμος, στοιχεία προκατάληψης στον δικό του τρόπο λήψης αποφάσεων, προκειμένου να καταλήξει στη δικαιότερη απόφαση. Τελικά, μπορεί να επαληθεύσει, με βάση το νόμο, το αποτέλεσμα ενός αλγόριθμου. Εκείνο, όμως, που πραγματικά μπορεί να βοηθήσει έναν άνθρωπο δικαστή και εισαγγελέα κατά το δικαιοδοτικό έργο του, είναι οι αιτιολογίες και επεξηγήσεις που τυχόν θα μπορούσε να του προσφέρει η τεχνητή νοημοσύνη. Χωρίς αυτές, η προσφορά της στο έργο του δεν είναι σημαντική και δεν θα μπορεί να συμβάλλει στην αποτελεσματικότητά του. Απαραίτητο, συνεπώς, κρίνεται η τεχνητή νοημοσύνη να είναι επεξηγήσιμη, προκειμένου να χρησιμοποιηθεί στις διαδικασίες λήψης αποφάσεων στα δικαστήρια κατά το δικαιοδοτικό έργο των δικαστών και εισαγγελέων. Η απαίτηση για επεξηγήσιμη τεχνητή νοημοσύνη συμβαδίζει, άλλωστε, και με τον Γενικό Κανονισμό για την Προστασία Δεδομένων του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016, που περιλαμβάνει, το δικαίωμα αυτού που επηρεάζεται από μια απόφαση που βασίζεται «αποκλειστικά σε αυτοματοποιημένη επεξεργασία» να «λάβει αιτιολόγηση της απόφασης που ελήφθη στο πλαίσιο της εν λόγω εκτίμησης και να αμφισβητήσει την απόφαση» (Chiao, 2019).

5.2. Όρια και περιορισμοί της έρευνας

Στις 21 Απριλίου 2021 η Ευρωπαϊκή Επιτροπή υπέβαλε πρόταση Κανονισμού για την τεχνητή νοημοσύνη. Η πρόταση αυτή (Artificial Intelligence Act) περιλαμβάνει ρυθμίσεις αναφορικά και με τη χρήση συστημάτων τεχνητής νοημοσύνης για σκοπούς επιβολής του νόμου, ενώ κατηγοριοποιεί τα διάφορα συστήματα με βάση τον κίνδυνο, ταξινομώντας μια σειρά συστημάτων τεχνητής νοημοσύνης που προορίζονται να χρησιμοποιηθούν στο πλαίσιο της επιβολής του νόμου, ως υψηλού κινδύνου (αιτιολ. σκ. 38-άρθρ. 6 παρ. 2-Παράρτημα ΙΙΙ). Είναι εμφανές ότι η εν λόγω Πράξη αφορά άμεσα το αντικείμενο της εργασίας, ενώ η ψήφιση του Κανονισμού θα αποτελέσει ένα σημαντικότατο βήμα για τη ρύθμιση των εκτιθέμενων ζητημάτων.

5.3. Μελλοντικές επεκτάσεις

Αντικείμενο μιας επόμενης εργασίας θα μπορούσε να αποτελέσει το πως θα μπορούσε να αλληλοεπιδράσει ο δικαστής – εισαγγελέας με ένα σύστημα τεχνητής νοημοσύνης, ποιες ικανότητες πρέπει να καλλιεργήσει, πως πρέπει να συνεργαστεί με τον πληροφορικό-δημιουργό του συστήματος για να μπορέσουν και οι δύο να κατανοήσουν τι πρέπει να προσφέρει ο καθένας από τη μεριά του για τη λειτουργία ενός συστήματος που να μπορεί να ανταποκριθεί στις απαιτήσεις του τομέα, την παροχή δηλαδή προβλέψεων-αποφάσεων κατανοητών, η ορθότητα των οποίων να μπορεί να επαληθευτεί από τον δικαστή – εισαγγελέα, ως στηριζόμενων στο νόμο. Για αρχή, θα μπορούσε να αναπτυχθεί ένα σύστημα, στα πλαίσια του ελληνικού συστήματος απονομής δικαιοσύνης, στον τομέα του αστικού δικαίου, το οποίο θα παρέχει τη σύνοψη της αγωγής, με βάση τα απαραίτητα από το νόμο στοιχεία, προκειμένου να διατυπωθεί το μέρος της απόφασης που αφορά στο περιεχόμενο της αγωγής, καθώς επίσης και τη κρίση περί του νόμω βάσιμου αυτής. Επίσης, η ίδια εργασία θα γίνεται και για τις ενστάσεις του εναγόμενου. Μέσα από αυτό το σύστημα, το οποίο θα μπορεί να παρέχει το τμήμα εκείνο της απόφασης πριν από τη διατύπωση της ελάσσονας πρότασης, θα μπορούσε να ερευνηθεί η δυνατότητα αλληλεπίδρασης του νομικού με το σύστημα, οι ικανότητες που πρέπει να καλλιεργήσει για να το χρησιμοποιήσει και ο τρόπος συνεργασίας του με τον πληροφορικό προκειμένου να επιτύχουν ένα αποτέλεσμα το οποίο, μέχρι το σημείο αυτό μιας δικαστικής απόφασης, είναι εύκολα επαληθεύσιμο από τον δικαστή – εισαγγελέα.

ΠΑΡΑΡΤΗΜΑ Ι:

Λήμμα αναζήτησης	Βάση δεδομένων όπου έγινε η αναζήτηση	Σύνολο αποτελε-σμάτων.	Αποτελέσματα αναζήτησης σχετιζόμενα με την εργασία	Χρήση στην διατριβή
artificial intelligence and justice	scopus	141	Opening privacy sensitive microdata sets in light of GDPR 2019	Μόνο η περίληψη στα αγγλικά ΟΧΙ
			Can artificial intelligence help estimate the risk of recidivism in violent behavior?	Μόνο η περίληψη στα αγγλικά ΟΧΙ
			Un estudio sobre la posibilidad de aplicar la inteligencia artificial en las decisionesjudicialeς	Δεν υπάρχει στα αγγλικά ΟΧΙ
			Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice 2019	 ΝΑΙ
			Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA 2021	ΝΑΙ
			Algorithms and values in justice and security 2020	ΝΑΙ
			KNOWLEDGE ELICITATION FOR EXPERT SYSTEMS IN THE LAW ENFORCEMENT DOMAIN 1989	ΟΧΙ
			Machine learning & forensic science 2019	ΝΑΙ
και με το "artificial intelligence and law"			Introduction to the special issue on machine law	ΟΧΙ
			Courts and Artificial Intelligence 2020	ΝΑΙ
			Making Artificial Intelligence Transparent**:** Fairness and the Problem of Proxy Variables Using factors to predict and analyze landlord-tenant decisions to increase access to Justice 2021	Μόνο περίληψη ΟΧΙ
			Criminal justice, artificial intelligence systems, and human rights 2020	ΝΑΙ
			A vulnerability analysis: Theorising the impact of artificial intelligence decision-making processes on individuals, society and human	ΝΑΙ

			diversity from a social justice perspective 2020	
			Artificial intelligence and law: What do people really want?: Example of a French multidisciplinary working croup 2019	OXI
	scopus		The judicial demand for explainable artificial intelligence 2019	NAI
	scopus		Artificial intelligence as evidence in criminal trial 2020	Μόνο περίληψη OXI
	scopus		Transparency as design publicity: explaining and justifying inscrutable algorithms 2020	NAI
"artificial intelligence and justice"	Scopus Web of Science	1	Artificial intelligence, digital capital, and epistemic domination on Twitter: A study of families affected by imprisonment 2021	OXI
“deep learning” AND “court”	Web of Science	101 στο scopus και 89 στο web of science	An efficient method for image forgery detection based on trigonometric transforms and deep learning 2020	NAI
	Web of Science		A Comprehensive Review of Deep-Learning-Based Methods for Image Forensics 2021	OXI
	Web of Science		A deep learning approach for the forensic evaluation of sexual assault 2018	OXI
(και με "deep learning" AND "court" και με "artificial intelligence and law")	Web of Science		Appellate Court Modifcations Extraction for Portuguese 2019	NAI
	scopus		Events Matter: Extraction of Events from Court Decisions 2020	OXI
(και με “judicial data analysis”)	scopus		Using Case Facts to Predict Penalty with Deep Learning 2019	OXI
	Web of Science		LegalCap: a model for complex case discrimination based on capsule neural network 2020	OXI

	Web of Science		Image forgery detection using deep textural features from local binary pattern map 2020	OXI
	Web of Science		Inferring Association Between Alcohol Addiction and Defendant's Emotion Based on Sound at Court 2021	OXI
	Web of Science		A Meta Analysis of Attention Models on Legal Judgment Prediction System 2021	OXI
	scopus είχε μόνο περίληψη/"deep learning" AND "court"/ το βρήκα στο google		Predicting the Law Area and Decisions of French Supreme Court Cases 2017	OXI
	scopus		An Approach of Rhetorical Status Recognition for Judgments in Court Documents using Deep Learning Models 2019	OXI
	scopus		A Comparative Study of Classifying Legal Documents with Neural Networks 2018	OXI
	scopus		Data transcription for India's supreme court documents using deep learning algorithms 2020	Μόνο περίληψη OXI
	Web of Science		Physiognomy: Personality Traits Prediction by Learning 2017	OXI
και με “machine learning” AND “judge”	scopus		Extracting value from Brazilian Court decisions 2022	NAI
	scopus		An overview of information extraction techniques for legal document analysis and processing 2021	NAI
	scopus		Using social signals to predict shoplifting: A transparent approach to a sensitive activity analysis problem 2021	OXI
	scopus		Natural language processing in law: Prediction of outcomes in the higher courts of Turkey 2021	OXI

	scopus		Predicting Outcomes of Court Judgments - A Machine Learning Approach 2021	OXI
	scopus		Through a glass, darkly: Artificial intelligence and the problem of opacity 2021	OXI
	scopus		A comprehensive review of deep-learning-based methods for image forensics 2021	OXI/διπλό
(και με "artificial intelligence and law" στη νεότερη αναζήτηση-web of science)	scopus		DeepRhole: deep learning for rhetorical role labeling of sentences in legal case documents 2021	Μόνο περίληψη OXI
	scopus		Deep learning based algorithm (ConvLSTM) for Copy Move Forgery Detection 2021	NAI
	scopus		A Meta Analysis of Attention Models on Legal Judgment Prediction System 2021	OXI/διπλό
	scopus		Application of multiple BERT model in construction litigation 2020	NAI
"judicial data analysis"	Scopus/web of science	3 αποτελέσματα.	Improved CBSO: A distributed fuzzy-based adaptive synthetic oversampling algorithm for imbalanced judicial data 2021	OXI
	scopus		A Deep Learning Method for Judicial Decision Support 2019	NAI
(και με "deep learning" AND "court")	scopus		Using Case Facts to Predict Penalty with Deep Learning 2019	OXI/διπλό
Το βρήκα ψάχνοντας το Using Case Facts to Predict Penalty with Deep Learning, στο google			How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence 2018	NAI
"machine learning and court"		Κανένα document		
"automated judicial decision making"	scopus / web of science	1 document στο web of science	Introducing validity in fuzzy probability for judicial decision-making	Είναι του 2013 OXI
"Judicial decision making"		Στο scopus έδωσε 237 documents και στο web of		

		science 812		
Το περιόρισα με “machine learning” AND “judicial decision making”	Web of science Και στο scopus μόνο σε περίληψη	2 στο scopus και 5 στο web of science	Preserving the rule of law in the era of artifcial intelligence (AI) 2021	NAI
			Nonlinear dimensionality reduction with judicial document learning 2018	OXI
(και με “AI” AND “judicial decision making” και με “artificial intelligence” AND “judicial decision making”	Web of science/scopus		Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making 2020	NAI
	Web of science		Judicial analytics and the great transformation of American Law 2019	NAI
	Web of science		Age of gray matters: Neuroprediction of recidivism 2018	OXI
	Web of science		Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective 2016	OXI
“AI” AND “judicial decision making”	Web of science	4 αποτελέσματα στο web of science	IS AUSTRALIA READY FOR AI ON THE BENCH? 2020	NAI
	Στο web of science με λέξη κλειδί “AI” AND “judicial decision making” Και Στο scopus με λέξη κλειδί "artificial intelligence" AND " judicial decision making"		SALVAGING THE SPIRIT OF THE METER-MODELS TRADITION: A MODEL OF BELIEF REVISION BY WAY OF AN ABSTRACT IDEALIZATION OF RESPONSE TO INCOMING EVIDENCE DELIVERY DURING THE CONSTRUCTION OF PROOF IN COURT 2010	OXI
Και με “machine learning” AND “judicial decision making”	Web of science		Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making	OXI/διπλό
Και με "artificial intelligence" AND "judicial decision making"	Web of science		Artificial Intelligence and Legal Decision-Making: The Wide Open? 2019	NAI
“artificial intelligence” AND	Web of science/scopus		IS AUSTRALIA READY FOR AI ON THE BENCH?	

“judicial decision making”				ΟΧΙ/ διπλό
Και στο web of science με λέξη κλειδί “AI” AND “judicial decision making”	Scopus/ web of science		SALVAGING THE SPIRIT OF THE METER-MODELS TRADITION: A MODEL OF BELIEF REVISION BY WAY OF AN ABSTRACT IDEALIZATION OF RESPONSE TO INCOMING EVIDENCE DELIVERY DURING THE CONSTRUCTION OF PROOF IN COURT (2010)	ΟΧΙ/διπλό
	Scopus/web of science		Fuzzy support systems for discretionary judicial decision making (2003)	Μόνο περίληψη ΟΧΙ
	Web of science/scopus		Paths to Digital Justice: Judicial Robots, Algorithmic Decision-Making, and Due Process (2020)	ΝΑΙ
Και με “AI” AND “judicial decision making” και “artificial intelligence” AND “judicial decision making”	Web of science/scopus		Artificial Intelligence and Legal Decision-Making: The Wide Open? 2019	ΟΧΙ/διπλό
	Web of science/scopus		JUDGE V ROBOT? ARTIFICIAL INTELLIGENCE AND JUDICIAL DECISION-MAKING 2018	ΝΑΙ
	scopus		The Application of Judicial Intelligence and ‘Rules’ to Systems Supporting Discretionary Judicial Decision-Making (1998)	1998 ΟΧΙ
Και με “machine learning” AND “judicial decision making”	Web o science/scopus		Preserving the rule of law in the era of artifcial intelligence (AI) (2021)	ΟΧΙ/διπλό
	Web of science		A STUDY ON THE POSSIBILITY OF APPLYING ARTIFICIAL INTELLIGENCE IN JUDICIAL DECISIONS	Στα ισπανικά ΟΧΙ
	Web of science /scopus		THE "BLACK BOX" OF JUDICIAL DECISION-MAKING: BETWEEN HUMAN AND ALGORITHMIC JUDGEMENT	Μόνο περίληψη ΟΧΙ
Και με τα: AI" AND "judicial decision making"/ "machine learning" AND "judicial decision making"			Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making	ΟΧΙ/διπλό
	Web of science		Thinking Machines and Smiley Faces 2019	Μόνο περίληψη ΟΧΙ
	Web of science /scopus		Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective (2016)	ΟΧΙ/διπλό
Και με “machine learning” AND	Web of science		Judicial analytics and the great transformation of American Law	ΟΧΙ/διπλό

“judicial decision making” ΚΑΙ “machine learning” AND “judge”				
	Web of science		Beyond the formalism debate: Expert reasoning, fuzzy logic, and complex statutes (1999)	ΟΧΙ
	scopus		On the relevance of algorithmic decision predictors for judicial decision making	ΝΑΙ
	scopus		A study on the possibility of applying artificial intelligence in judicial decisions	στα ισπανικά ΟΧΙ
	scopus		Smart Technologies, Human Security and Global Justice	Μόνο περίληψη ΟΧΙ
και με “AI” AND “judicial decision making” και με “machine learning” AND “judicial decision making”	Scopus		Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making	ΟΧΙ/διπλό
και με “AI” AND “judicial decision making”	scopus		Artificial Intelligence and Legal Decision-Making: The Wide Open?	ΟΧΙ/διπλό
“machine learning” AND “ judge”	scopus	736 στο scopus και 240 στο web of science	Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models 2022	ΝΑΙ
	Web of science		Use of a machine learning framework to predict substance use disorder treatment success 2017	ΟΧΙ
και με “machine learning” AND “judicial decision making”	Scopus/και στο web of science		Judicial analytics and the great transformation of American Law	ΟΧΙ/διπλό
	scopus		Joint cognition of both human and machine for predicting criminal punishment in judicial system 2019	Μόνο περίληψη και το βρήκα στο researchgate ΝΑΙ
και με “AI” AND “legal reasoning”	Web of science		The Change of Judicial Power in China in the Era of Artificial Intelligence-2020	
	Web of science		HUMAN DECISIONS AND MACHINE PREDICTIONS	ΟΧΙ
και με “deep learning and court” (και στη νεότερη και με “artificial intelligence and law”	Web of science		Appellate Court Modifcations Extraction for Portuguese	ΟΧΙ/διπλό
	Web of science		Text classification of ideological direction in judicial opinions (2020)	ΟΧΙ

	Web of science		Predicting Contextual Informativeness for Vocabulary learning (2018)	OXI
	Web of science		Big-Data Measurement-Model Research about Judges' Actual Workload in China 2020	OXI
	Web of science		Computer-based personality judgments are more accurate than those made by humans	OXI
Και με "deep learning" AND "court"	scopus		Extracting value from Brazilian Court decisions	OXI/διπλό
	scopus		Legal Judgment Prediction Based on Machine Learning: Predicting the Discretionary Damages of Mental Suffering in Fatal Car Accident Cases (2021)	OXI
	scopus		Applying Decision Tree Analysis to Family Court Decisions: Factors Determining Child Custody in Taiwan-2021	
	scopus		Context-Aware Legal Citation Recommendation using Deep Learning - 2021	OXI
"AI" AND "legal reasoning technology"		0 αποτελέσματα		
"AI" AND "legal reasoning"	scopus	στο scopus 81 (κατέγραψα όσα είναι μέχρι και το 2012) στο web of science 34	A novel understanding of legal syllogism as a starting point for better legal symbolic AI systems Το βρήκα σε νεότερη αναζήτηση	μόνο περίληψη OXI
	scopus		Machine learning and legal argument 2021	NAI
	Web of science/Scopus		AI Applications to the Law Domain in Japan 2020	OXI
	Scopus		Defeasible Systems in Legal Reasoning: Comparative Assessment 2019	OXI
	scopus		Legal automation: AI and law revisited-2019	OXI
	scopus		Explainability of formal models of argumentation applied to legal domain (2019)	NAI
	scopus		Introduction: Legal and Ethical Dimensions of AI, NorMAS, and the Web of Data-2018	OXI

και με “artificial intelligence and law”	scopus		Rediscovering artificial intelligence and law: an inadequate jurisprudence? (2016)	OXI
	scopus		Statement types in legal argument 2016	Δεν υπήρχε και το βρήκα στο google-0XI
	Scopus/web of science		Law and logic: A review from an argumentation perspective 2015	OXI
και με “artificial intelligence and law”	Scopus/web of science		On balance 2015	OXI
	scopus		Using event progression to enhance purposive argumentation in the value judgment formalism-2013	2013 OXI
	scopus		Discussion paper: how much of commonsense and legal reasoning is formalizable? A review of conceptual obstacles	2012 OXI
	scopus		What makes a system a legal expert?	2012 OXI
	Web of science		A model of legal reasoning with cases incorporating theories and values	2003 OXI
	Scopus/web of science		Explanation in AI and law: Past, present and future-2020	OXI
	Scopus/web of science		Discussion paper: how much of commonsense and legal reasoning is formalizable? A review of conceptual obstacles	2012 OXI
	Web of science		From Berman and Hafner’s teleological context to Baude and Sachs’ interpretive defaults: an ontological challenge for the next decades of AI and Law 2016	OXI
	Web of science		LogiKEy workbench: Deontic logics, logic combinations and expressive ethical and legal reasoning (Isabelle/HOL dataset)-2020	OXI
και με το "artificial intelligence and law"	Web of science		A history of AI and Law in 50 papers: 25 years of the international conference on AI and Law-2012	OXI
	Scopus/web of science		Using background knowledge in case-based legal reasoning: A computational model and an intelligent learning environment 2003	OXI
	Scopus/web of science		Before and after Dung: Argumentation in AI and Law 2020	OXI
	Web of science		A methodology for designing systems to reason with legal cases	OXI

			using Abstract Dialectical Frameworks-2016	
	Web of science		Designing normative theories for ethical and legal reasoning: LogiKEy framework, methodology, and tool support 2020	OXI
	Web of science		Law and logic: A review from an argumentation perspective 2015	OXI
στο web of science και με το : "machine learning" AND "judge"	Scopus/web of science		The Change of Judicial Power in China in the Era of Artificial Intelligence 2020	OXI
	Web of science		Complexity Results and Algorithms for Extension Enforcement in Abstract Argumentation-2017	OXI
και με το “artificial intelligence and law”	Web of science		Resolving counterintuitive consequences in law using legal debugging-2021	OXI
"personal data" AND "court"	Web of science		Criminal justice profiling and EU data protection law: precarious protection from predictive policing-2019	NAI
	Web of science		INTERNATIONAL TRENDS IN THE JUSTICE DIGITALIZATION DEVELOPMENT-2020	OXI
“algorithmic justice”	scopus	στο web of science 14 αποτελέσματα	Algorithmic Pollution: Understanding and Responding to Negative Consequences of Algorithmic Decision-Making 2018	OXI
	Scopus/web of science		Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making-2019	OXI
	Scopus/web of science		Digital prediction technologies in the justice system: The implications of a ‘race-neutral’ agenda-2020	OXI
	Web of science		Algorithmic justice: Algorithms and big data in criminal justice settings-2019	OXI
	Web of science		JUSTICE IN THE DIGITAL AGE: TECHNOLOGICAL SOLUTIONS, HIDDEN THREATS AND ENTICING OPPORTUNITIES 2021	NAI
“artificial intelligence and law”	scopus	στο scopus 139 και στο web of science 331	LEGIS: A Proposal to Handle Legal Normative Exceptions and Leverage Inference Proofs Readability	OXI
	Web of science		In memoriam Douglas N. Walton: the infuence of Doug Walton on AI and law (2020)	OXI

και με το "AI" AND "legal reasoning"	Web of science		A history of AI and Law in 50 papers: 25 years of the international conference on AI and Law - 2012	ΟΧΙ
	Web of science/scopus		Scalable and explainable legal prediction-2021	ΝΑΙ
	scopus		Towards Consumer-Empowering Artificial Intelligence-2018	ΟΧΙ
	scopus		DEVELOPING A LEGAL EXPERT SYSTEM FOR THE PALESTINIAN LABOR LAW (2017)	ΟΧΙ
και με artificial intelligence and justice.	scopus		Introduction to the special issue on machine law	ΟΧΙ
	scopus		On the Principle of Privacy by Design and Its Limits: Technology, Ethics and the Rule of Law-2021	ΟΧΙ
και με "AI" AND "legal reasoning"/	scopus		Rediscovering artificial intelligence and law: an inadequate jurisprudence? 2016	ΟΧΙ/ διπλό
	Scopus/web of science		Representing dimensions within the reason model of precedent 2018	ΟΧΙ
	web of science		Sentence Boundary Detection in Adjudicatory Decisions in the United States-2017	ΟΧΙ
	Scopus/web of science		Artifcial intelligence as law Presidential address to the seventeenth international conference on artifcial intelligence and law (2020)	(ομιλία) ΟΧΙ
	web of science		The epistemology of scientific evidence-2013	ΟΧΙ
	web of science		Introduction for artificial intelligence and law: special issue “natural language processing for legal texts”-2019	ΟΧΙ
	web of science		A COMPUTATIONAL SOLUTION CAPABLE OF PREDICTING JUDICIAL DECISIONS-2021	ΟΧΙ
	web of science		Building sustainable free legal advisory systems: Experiences from the history of AI & law-2018	ΟΧΙ
			Fairness and Predictive Justice. A Path from Machine Learning to the Concept of Law (2020)	έχει μόνο περίληψη ΟΧΙ
	web of science		Protecting Sentient Artificial Intelligence: A Survey of Lay Intuitions on Standing, Personhood, and General Legal Protection (2021)	ΟΧΙ
	web of science		Contributions of FGCS technology to applications in legal reasoning-1995	ΟΧΙ

	web of science		Electronic evidence in the blockchain era: New rules on authenticity and integrity (2020)	OXI
	web of science		Business E-NeGotiAtion: A Method Using a Genetic Algorithm for Online Dispute Resolution in B2B Relationships (2021)	OXI
	web of science		Symbiosis with artificial intelligence via the prism of law, robots, and society (2021)	OXI
	web of science		Legal requirements on explainability in machine learning	OXI
	web of science		Proof with and without probabilities Correct evidential reasoning with presumptive arguments, coherent hypotheses and degrees of uncertainty (2017)	OXI
	web of science		Evidence & decision making in the law: theoretical, computational and empirical approaches (2020)	OXI
	web of science		Dynamic epistemic logic of belief change in legal judgments (2018)	OXI
	web of science		On the concept of relevance in legal information retrieval-2017	
	web of science		Cognitive computing and proposed approaches to conceptual organization of case law knowledge bases: a proposed model for information preparation, indexing, and analysis (2016)	OXI
(και με "deep learning" AND "court" στη νεότερη αναζήτηση-scopus)	web of science		DeepRhole: deep learning for rhetorical role labeling of sentences in legal case documents (2021)	OXI/διπλό
	web of science		A review of predictive policing from the perspective of fairness (2021)	OXI
	web of science		Taking stock of legal ontologies: a feature-based comparative analysis (2020)	OXI
	web of science		Arguing about causes in law: a semi-formal framework for causal arguments	OXI
	web of science		Using machine learning to predict decisions of the European Court of Human Rights (2020)	OXI
	web of science		Explainable AI under contract and tort law: legal incentives and technical challenges (2020)	NAI

	web of science		Norms and value based reasoning: justifying compliance and violation (2017)	OXI
	web of science		Recognizing cited facts and principles in legal judgements (2017)	OXI
	web of science		From Berman and Hafner's teleological context to Baude and Sachs' interpretive defaults: an ontological challenge for the next decades of AI and Law (2016)	OXI
	web of science		A methodology for designing systems to reason with legal cases using Abstract Dialectical Frameworks (2016)	OXI
	web of science		Unsupervised approaches for measuring textual similarity between legal court case reports (2021)	OXI
	web of science		Building a corpus of legal argumentation in Japanese judgement documents: towards structure-based summarization (2019)	OXI
	web of science		Using artificial intelligence to support compliance with the general data protection regulation (2017)	OXI
	web of science		Unsupervised law article mining based on deep pre-trained language representation models with application to the Italian civil code (2021)	OXI
	web of science		Evaluating causes of algorithmic bias in juvenile criminal recidivism-2021	OXI
	web of science		A method for explaining Bayesian networks for legal evidence with scenarios (2016)	OXI
Και με "AI' AND "legal reasoning" - scopus	web of science		On balance (2015)	OXI/διπλό
Και με "AI' AND "legal reasoning"	web of science		Resolving counterintuitive consequences in law using legal debugging (2021)	OXI
	web of science		Two-layered fuzzy logic-based model for predicting court decisions in construction contract disputes	OXI
	Web of science		PRILJ: an efficient two-step method based on embedding and clustering for the identification of regularities in legal case judgments (2021)	OXI

	Web of science		A system of communication rules for justifying and explaining beliefs about facts in civil trials (2020)	OXI
	Web of science		Group-to-individual (G2i) inferences: challenges in modeling how the U.S. court system uses brain data (2020)	OXI
και με "machine learning" AND "judicial decision making" ΚΑΙ με "artificial intelligence" AND "judicial decision making"	web of science		Preserving the rule of law in the era of artifcial intelligence (AI)	OXI/διπλό
	scopus		Design AI so that it's fair	OXI
	scopus		Explainable Artificial Intelligence and Machine Learning: A reality rooted perspective (2020)	NAI
	scopus		Is there a place for machine learning in law? 2017	το βρήκα στο scopus/μόνο περίληψη/ και το βρήκα στο google scholar OXI
	scopus		Developing a legal expert system for the palestinian labor law-2017	OXI/διπλό
	scopus		Checking the validity of rule-based arguments grounded in cases-2018	OXI
	scopus		The Genealogy of Ideology: Predicting Agreement and Persuasive Memes in the U.S. Courts of Appeals (δεν έχει ημερομηνία)	το βρήκα στο scopus/ είχε μόνο την περίληψη και το βρήκα στο google OXI
	scopus		Criminal Conviction Classification Based on Multiple Learning Methods (2019)	το βρήκα στο scopus/ είχε μόνο την περίληψη και το βρήκα στο google OXI
			How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence-2020	OXI/διπλό

ΠΑΡΑΡΤΗΜΑ ΙΙ

Πέραν του προαναφερόμενου περιεχομένου της λογοδοσίας, που αφορά στον έλεγχο των συστημάτων τεχνητής νοημοσύνης και των σχετικών διαδικασιών λήψης αποφάσεων, όπως επίσης και στην ευθύνη για αυτές, η λογοδοσία είναι μία από τις αρχές που διέπουν την επεξεργασία δεδομένων προσωπικού χαρακτήρα. Ουσιαστικά, επιτρέπει στον υπεύθυνο επεξεργασίας δεδομένων να εκτιμά τους κινδύνους που ενέχει η αυτοματοποιημένη λήψη αποφάσεων, περιλαμβανομένης της κατάρτισης προφίλ, ενώ αποτελεί έναν τρόπο να αποδεικνύεται ότι εφαρμόζονται κατάλληλα μέτρα για την αντιμετώπιση των εν λόγω κινδύνων και, ως εκ τούτου, να καταδεικνύεται η συμμόρφωση με τον Κανονισμό (Χριστίνα-Ειρήνη Ασημακοπούλου, 2020). Σύμφωνα με την παρ. 3 του άρθρου 35 του Κανονισμού: Η αναφερόμενη στην παράγραφο 1 εκτίμηση αντικτύπου σχετικά με την προστασία δεδομένων απαιτείται ιδίως στην περίπτωση: α) συστηματικής και εκτενούς αξιολόγησης προσωπικών πτυχών σχετικά με φυσικά πρόσωπα, η οποία βασίζεται σε αυτοματοποιημένη επεξεργασία, περιλαμβανομένης της κατάρτισης προφίλ, και στην οποία βασίζονται αποφάσεις που παράγουν έννομα αποτελέσματα σχετικά με το φυσικό πρόσωπο ή ομοίως επηρεάζουν σημαντικά το φυσικό πρόσωπο, β) μεγάλης κλίμακας επεξεργασίας των ειδικών κατηγοριών δεδομένων που αναφέρονται στο άρθρο 9 παρ. 1 ή δεδομένων προσωπικού χαρακτήρα που αφορούν ποινικές καταδίκες και αδικήματα που αναφέρονται στο άρθρο 10 ή γ) συστηματικής παρακολούθησης δημοσίως προσβάσιμου χώρου σε μεγάλη κλίμακα. Εξάλλου, στην υπό στοιχ. α περίπτωση, η υποχρέωση για διενέργεια εκτίμησης αντικτύπου ισχύει για αποφάσεις που παράγουν έννομα αποτελέσματα σχετικά με το φυσικό πρόσωπο ή επηρεάζουν σημαντικά το φυσικό πρόσωπο, είτε αυτές βασίζονται αποκλειστικά είτε εν μέρει σε αυτοματοποιημένη επεξεργασία (Δημήτριος Ευ. Τζέλλης, Μαρία Δ. Μυλώση, 2022). Η παραπάνω παράθεση επεξεργασιών υψηλού κινδύνου είναι ενδεικτική, όπως τονίστηκε και όχι εξαντλητική και για το λόγο αυτό, η ΟΕ29, στην προσπάθειά της να υποστηρίξει τους υπευθύνους επεξεργασίας, εξέδωσε τις σχετικές κατευθυντήριες γραμμές, στις οποίες περιλαμβάνονται τα εννέα κριτήρια, τα οποία πρέπει να λαμβάνονται υπόψη σχετικά με το αν είναι πιθανό μία επεξεργασία να οδηγήσει σε υψηλό κίνδυνο για τα δικαιώματα και τις ελευθερίες των ατόμων και επομένως αν είναι υποχρεωτική η εκτίμηση αντικτύπου. Μεταξύ αυτών των κριτηρίων είναι και τα εξής:… 2. Αυτοματοποιημένη λήψη αποφάσεων με έννομες συνέπειες (βλ. σχετικά παρ. εδ. α, παρ. 3 άρθρου 35 ΓΚΠΔ), με ενδεχόμενη επίπτωση τον αποκλεισμό ή τη δυσμενή διάκριση των ατόμων. … 4. Δεδομένα ειδικών κατηγοριών ή ειδικής

προσωπικής φύσεως (άρθρα 9 και 10 ΓΚΠΔ). Σε αυτά εμπίπτουν τα χαρακτηριζόμενα και ως ευαίσθητα δεδομένα, για παράδειγμα δεδομένα που αποκαλύπτουν φυλετική ή εθνική καταγωγή, πολιτικές ή φιλοσοφικές πεποιθήσεις, υγεία, γενετικά χαρακτηριστικά ή βιομετρικής φύσεως κ.ά., καθώς και τα δεδομένα που αφορούν ποινικές δίκες ή αδικήματα ή σχετικά μέτρα ασφάλειας. Εκτός όσων προβλέπονται στα παραπάνω άρθρα 9 και 10 του ΓΚΠΔ, άλλες κατηγορίες δεδομένων, όπως αυτές που σχετίζονται με ηλεκτρονικές επικοινωνίες, με τη γεωγραφική θέση του υποκειμένου ή με την οικονομική κατάσταση ή δραστηριότητα αν αποτελούν αντικείμενο συλλογής και επεξεργασίας ενδέχεται να οδηγήσουν σε υψηλή διακινδύνευση δικαιωμάτων και ελευθεριών, όπως της ελευθερίας κυκλοφορίας ή της διευκόλυνσης οικονομικής απάτης ή εγκλημάτων. … 8. Νέες τεχνολογίες ή τεχνικές ή οργανωτικές λύσεις, όταν επιλέγονται ως μέσα επεξεργασίας, ενδέχεται να οδηγήσουν σε σημαντική αύξηση των κινδύνων για τα δικαιώματα και τις ελευθερίες των υποκειμένων των δεδομένων. Σύμφωνα με την παράγραφο 1 άρθρο 35 και τις αιτιολογικές σκέψεις 89 και 91 του ΓΚΠΔ, η αξιοποίηση νέων τεχνολογιών στην επεξεργασία, όπως εφαρμογών του ‘Διαδικτύου των Πραγμάτων’ μπορεί να οδηγήσει στην ανάγκη εκτίμησης αντικτύπου στην προστασία δεδομένων, έτσι ώστε τυχόν υψηλοί κίνδυνοι να αντιμετωπιστούν από τους υπεύθυνους επεξεργασίας έγκαιρα και αποτελεσματικά. Άλλο σχετικό παράδειγμα αποτελεί η συνδυασμένη χρήση περισσότερων βιομετρικών μεθόδων ή και η συνδυασμένη χρήση συστημάτων παρακολούθησης και βιομετρικών τεχνικών κατηγοριοποίησης ή ακόμη και αναγνώρισης ταυτότητας (ταυτοποίησης) (Βασίλης Ζορκάδης, 2018).

Εξάλλου, πρέπει να σημειωθεί ότι η αυτοματοποιημένη λήψη αποφάσεων ρυθμίζεται εκτός από τον Κανονισμό και την Οδηγία για την επιβολή του νόμου και στο ελληνικό δίκαιο. Συγκεκριμένα, ο Έλληνας νομοθέτης ψήφισε τον ν. 4624/2019, για την εφαρμογή του Κανονισμού και την ενσωμάτωση στην εθνική νομοθεσία της Οδηγίας (ΕΕ) 2016/680. Σύμφωνα δε με το άρθρο 52 του ν. 4624/2019 : 1. Απαγορεύεται η λήψη απόφασης που βασίζεται αποκλειστικά σε αυτοματοποιημένη επεξεργασία, περιλαμβανομένης της κατάρτισης προφίλ, η οποία παράγει δυσμενή έννομα αποτελέσματα για το υποκείμενο των δεδομένων ή το επηρεάζει σημαντικά, εκτός εάν προβλέπεται ρητά από διάταξη νόμου ή το δίκαιο της Ένωσης, η οποία ορίζει τις κατάλληλες εγγυήσεις για τα δικαιώματα και τις ελευθερίες του υποκειμένου των δεδομένων και κατ` ελάχιστον περιλαμβάνει ρυθμίσεις που κατοχυρώνουν την ειδική και εύληπτη ενημέρωση του υποκειμένου των δεδομένων, το δικαίωμα εξασφάλισης ανθρώπινης παρέμβασης εκ μέρους του υπευθύνου επεξεργασίας και το δικαίωμα του υποκειμένου των δεδομένων να διατυπώσει τις απόψεις του, να απαιτήσει αιτιολόγηση της απόφασης που ελήφθη κατόπιν της εν λόγω αξιολόγησης και να αμφισβητήσει ή να ζητήσει επανεξέταση της απόφασης. 2. Οι αποφάσεις της παρ. 1 δεν επιτρέπεται να βασίζονται σε επεξεργασία των ειδικών κατηγοριών δεδομένων προσωπικού χαρακτήρα που αναφέρονται στην παρ. 1 του άρθρου 46, εκτός εάν τούτο προβλέπεται ρητά από διάταξη νόμου ή από το δίκαιο της Ένωσης και υφίστανται κατάλληλα μέτρα για την προστασία των δικαιωμάτων, των ελευθεριών και των έννομων συμφερόντων του υποκειμένου των δεδομένων, στα οποία συμπεριλαμβάνονται οι διασφαλίσεις που ορίζονται στην

παρ. 2 του άρθρου 46. 3. Απαγορεύεται η κατάρτιση προφίλ που έχει ως αποτέλεσμα διακρίσεις σε βάρος φυσικών προσώπων με βάση τις ειδικές κατηγορίες δεδομένων προσωπικού χαρακτήρα που αναφέρονται στην παρ. 1 του άρθρου 46». Όπως προκύπτει από τις ως άνω διατάξεις, η απαγόρευση της αυτοματοποιημένης ατομικής λήψης αποφάσεων (άρ. 52 παρ. 1 του ν. 4624/2019) αφορά τις αποφάσεις που παράγουν δυσμενή έννομα αποτελέσματα (π.χ. εφαρμογή αυξημένων μέτρων ασφάλειας ή εποπτείας από τις αρμόδιες αρχές) ή τις αποφάσεις που θίγουν το υποκείμενο των δεδομένων σημαντικά (π.χ. εισαγωγή διακρίσεων σε βάρος φυσικών προσώπων, άρνηση σε φυσικό πρόσωπο της επιβίβασης σε μεταφορικό μέσο λόγω της καταχώρισής του σε μαύρη λίστα)(Δημήτριος Ευ. Τζέλλης, Μαρία Δ. Μυλώση, 2022). Οι Δημήτριος Ευ. Τζέλλης, Μαρία Δ. Μυλώση, 2022, επισημαίνουν, μεταξύ άλλων καλών πρακτικών, τις εξής:

Επιβεβαιώστε ότι πριν ληφθεί τέτοια απόφαση, έχετε εφαρμόσει τις κατάλληλες διασφαλίσεις υπέρ των δικαιωμάτων και των ελευθεριών του υποκειμένου των δεδομένων (τουλάχιστον το δικαίωμα εξασφάλισης της ανθρώπινης παρέμβασης εκ μέρους του υπεύθυνου επεξεργασίας), που προβλέπονται από το δίκαιο της Ευρωπαϊκής Ένωσης ή της Ελλάδας. Να παρέχετε στα υποκείμενα των δεδομένων κατάλληλες πληροφορίες όσον αφορά την ύπαρξη αυτοματοποιημένης λήψης αποφάσεων, περιλαμβανομένης της κατάρτισης προφίλ, καθώς και ουσιαστικές πληροφορίες για το σχετικό σκεπτικό (αιτιολογική σκέψη 38 της οδηγίας (ΕΕ) 2016/680). Να διενεργήσετε Εκτίμηση Αντικτύπου στην προστασία προσωπικών δεδομένων όταν η επεξεργασία ενδέχεται να δημιουργήσει σημαντικό κίνδυνο για τα προστατευόμενα έννομα συμφέροντα των ενδιαφερομένων. Να προσδιορίζετε στο αρχείο δραστηριοτήτων το αν προβαίνετε σε κατάρτιση προφίλ (άρ. 68 του ν. 4624/2019). Να εξασφαλίσετε ότι ποτέ δεν χρησιμοποιείτε τη συγκατάθεση ως νομική βάση επεξεργασίας προσωπικών δεδομένων (Δημήτριος Ευ. Τζέλλης, Μαρία Δ. Μυλώση, 2022).

Επίσης, οι ίδιοι, επισημαίνουν ότι ο Έλληνας νομοθέτης δεν ενσωμάτωσε στο ελληνικό δίκαιο το δικαίωμα του υποκειμένου των δεδομένων να εξασφαλίζει την ανθρώπινη παρέμβαση εκ μέρους του υπεύθυνου επεξεργασίας στη λήψη απόφασης που βασίζεται αποκλειστικά σε αυτοματοποιημένη επεξεργασία, περιλαμβανομένης της κατάρτισης προφίλ, η οποία παράγει δυσμενή έννομα αποτελέσματα για το υποκείμενο των δεδομένων ή θίγει αυτό σε μεγάλο βαθμό (άρ. 52 του ν. 4624/2019), παρότι αυτό προβλεπόταν στο άρθρο 11 της οδηγίας (ΕΕ) 2016/680 και στην αιτιο-λογική σκέψη 38 της οδηγίας (ΕΕ) 2016/680. Η αιτιολογική σκέψη 38 της οδηγίας (ΕΕ) 2016/680 προβλέπει ότι είναι δικαίωμα του υποκειμένου των δεδομένων: να εξασφαλίσει την ανθρώπινη παρέμβαση εκ μέρους του υπεύθυνου επεξεργασίας· να λάβει αιτιολόγηση της απόφασης που λήφθηκε· να διατυπώσει την άποψή του· να αμφισβητήσει την απόφαση που λήφθηκε(Δημήτριος Ευ. Τζέλλης, Μαρία Δ. Μυλώση, 2022).

Εκτός από το ν. 4624/2019, ο νόμος 4842/2021 για την “Ταχεία πολιτική δίκη, προσαρμογή των διατάξεων της πολιτικής δικονομίας για την ψηφιοποίηση της πολιτικής δικαιοσύνης, άλλες τροποποιήσεις στον Κώδικα Πολιτικής Δικονομίας και λοιπές επείγουσες διατάξεις”, συμβάλλει στην ενίσχυση της

ψηφιοποίησης της πολιτικής δίκης. Χαρακτηριστικά, στην παρ. 4 του άρθρου 468 ΚΠολΔ προβλέπεται ότι η προβλεπόμενη στις προηγούμενες παραγράφους διαδικασία που προβλέπεται για τις μικροδιαφορές "μπορεί να γίνει και με χρήση τυποποιημένων εγγράφων ή Τεχνολογίας Πληροφορίας και Επικοινωνίας. Με κοινή απόφαση των Υπουργών Δικαιοσύνης και Ψηφιακής Διακυβέρνησης καθορίζονται οι λεπτομέρειες για την εφαρμογή της παρούσας με χρήση Τεχνολογίας Πληροφορίας και Επικοινωνίας. Με απόφαση του Υπουργού Δικαιοσύνης καθορίζεται το περιεχόμενο των τυποποιημένων εγγράφων". Η ψηφιοποίηση των εγγράφων και διαδικασιών αποτελεί απαραίτητη προϋπόθεση για την περαιτέρω εισαγωγή αλγοριθμικών εργαλείων και εφαρμογών τεχνητής δικαιοσύνης που θα βοηθούν στη λήψη αποφάσεων.

Βιβλιογραφία

Al_Azrak, F.M. *et al.* (2020) 'An efficient method for image forgery detection based on trigonometric transforms and deep learning', *Multimedia Tools and Applications*, 79(25–26), pp. 18221–18243. Available at: https://doi.org/10.1007/s11042-019-08162-3.

Araszkiewicz, M. and Nalepa, G.J. (2019) 'Explainability of formal models of argumentation applied to legal domain', in. *CEUR Workshop Proceedings*.

Bex, F. and Prakken, H. (2021) 'On the relevance of algorithmic decision predictors for judicial decision making', in. *Proceedings of the 18th International Conference on Artificial Intelligence and Law, ICAIL 2021*, pp. 175–179. Available at: https://doi.org/10.1145/3462757.3466069.

Bolingford, I. *et al.* (2020) 'Is Australia Ready for AI on the Bench?', *Journal of Judicial Administration*, 30(1), pp. 3–18.

Branting, L.K. *et al.* (2021) 'Scalable and explainable legal prediction', *Artificial Intelligence and Law*, 29(2), pp. 213–238. Available at: https://doi.org/10.1007/s10506-020-09273-1.

Chen, B. *et al.* (2019) 'A Deep Learning Method for Judicial Decision Support', in *2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*. *2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pp. 145–149. Available at: https://doi.org/10.1109/QRS-C.2019.00040.

Chen, C.-W. *et al.* (2020) 'Application of Multiple BERT Model in Construction Litigation', in *2020 8th International Conference on Orange Technology (ICOT)*. *2020 8th International Conference on Orange Technology (ICOT)*, pp. 1–4. Available at: https://doi.org/10.1109/ICOT51877.2020.9468727.

Chen, D.L. (2019) 'Judicial analytics and the great transformation of American Law', *Artificial Intelligence and Law*, 27(1), pp. 15–42. Available at: https://doi.org/10.1007/s10506-018-9237-x.

Chiao, V. (2019) 'Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice', *International Journal of Law in Context*, 15(2), pp. 126–139. Available at: https://doi.org/10.1017/S1744552319000077.

Deeks, A. (2019) 'The judicial demand for explainable artificial intelligence', *Columbia Law Review*, 119(7), pp. 1829–1850.

Elaskily, M.A. *et al.* (2021) 'Deep learning based algorithm (ConvLSTM) for Copy Move Forgery Detection', *Journal of Intelligent and Fuzzy Systems*, 40(3), pp. 4385–4405. Available at: https://doi.org/10.3233/JIFS-201192.

Emmert-Streib, F., Yli-Harja, O. and Dehmer, M. (2020) 'Explainable artificial intelligence and machine learning: A reality rooted perspective', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(6). Available at: https://doi.org/10.1002/widm.1368.

Fernandes, W.P.D. *et al.* (2020) 'Appellate Court Modifications Extraction for Portuguese', *Artificial Intelligence and Law*, 28(3), pp. 327–360. Available at: https://doi.org/10.1007/s10506-019-09256-x.

Fernandes, W.P.D. *et al.* (2022) 'Extracting value from Brazilian Court decisions', *Information Systems*, 106. Available at: https://doi.org/10.1016/j.is.2021.101965.

Ghajargar, M. *et al.* (2021) 'From "Explainable AI" to "Graspable AI"', in *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*. *TEI '21: Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, Salzburg Austria: ACM, pp. 1–4. Available at: https://doi.org/10.1145/3430524.3442704.

Greenstein, S. (no date) 'Preserving the rule of law in the era of artificial intelligence (AI)', *Artificial Intelligence and Law* [Preprint]. Available at: https://doi.org/10.1007/s10506-021-09294-4.

Hacker, P. *et al.* (2020) 'Explainable AI under contract and tort law: legal incentives and technical challenges', *Artificial Intelligence and Law*, 28(4), pp. 415–439. Available at: https://doi.org/10.1007/s10506-020-09260-6.

Hartmann, K. and Wenzelburger, G. (2021) 'Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA', *Policy Sciences*, 54(2), pp. 269–287. Available at: https://doi.org/10.1007/s11077-020-09414-y.

Hayes, P., van de Poel, I. and Steen, M. (2020) 'Algorithms and values in justice and security', *AI & SOCIETY*, 35(3), pp. 533–555. Available at: https://doi.org/10.1007/s00146-019-00932-9.

'Justice in the Digital Age: Technological Solutions, Hidden Threats and Enticing Opportunities' (2021) *Access to Justice in Eastern Europe*, 4(2), pp. 104–117. Available at: https://doi.org/10.33327/AJEE-18-4.2-a000061.

Krupiy, T. (Tanya) (2020) 'A vulnerability analysis: Theorising the impact of artificial intelligence decision-making processes on individuals, society and human diversity from a social justice perspective', *Computer Law & Security Review*, 38, p. 105429. Available at: https://doi.org/10.1016/j.clsr.2020.105429.

Loi, M., Ferrario, A. and Viganò, E. (2020) 'Transparency as design publicity: explaining and justifying inscrutable algorithms', *Ethics and Information Technology* [Preprint]. Available at: https://doi.org/10.1007/s10676-020-09564-w.

Lynskey, O. (2019) 'Criminal justice profiling and EU data protection law: precarious protection from predictive policing', *International Journal of Law in Context*, 15(2), pp. 162–176. Available at: https://doi.org/10.1017/S1744552319000090.

Margagliotti, G. and Bollé, T. (2019) 'Machine learning & forensic science', *Forensic Science International*, 298, pp. 138–139. Available at: https://doi.org/10.1016/j.forsciint.2019.02.045.

McKay, C. (2020) 'Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making', *Current Issues in Criminal Justice*, 32(1), pp. 22–39. Available at: https://doi.org/10.1080/10345329.2019.1658694.

Monaro, M. *et al.* (2022) 'Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models', *Computers in Human Behavior*, 127. Available at: https://doi.org/10.1016/j.chb.2021.107063.

Mumford, J., Atkinson, K. and Bench-Capon, T. (2021) 'Machine learning and legal argument', in. *CEUR Workshop Proceedings*, pp. 47–56.

Reading, A.D.D. (2020) 'Courts and Artificial Intelligence', *International Journal for Court Administration*, 11(2), pp. 1–10. Available at: https://doi.org/10.36745/IJCA.343.

Rubim Borges Fortes, P. (2020) 'Paths to Digital Justice: Judicial Robots, Algorithmic Decision-Making, and Due Process', *Asian Journal of Law and Society*, 7(3), pp. 453–469. Available at: https://doi.org/10.1017/als.2020.12.

Scherer, M. (2019) 'Artificial Intelligence and Legal Decision-Making: The Wide Open? A Study Examining International Arbitration', *Journal of International Arbitration*, 36(5), pp. 539–573.

Sourdin, T. (2018) 'JUDGE V ROBOT? ARTIFICIAL INTELLIGENCE AND JUDICIAL DECISION-MAKING', 41, p. 20.

Webster, J. and Watson, R.T. (2002) 'Guest Editorial: Analyzing the Past to Prepare for the Future: Writing a literature Review', p. 11.

Zadgaonkar, A.V. and Agrawal, A.J. (2021) 'An overview of information extraction techniques for legal document analysis and processing', *International Journal of Electrical and Computer Engineering*, 11(6), pp. 5450–5457. Available at: https://doi.org/10.11591/ijece.v11i6.pp5450-5457.

Završnik, A. (2020) 'Criminal justice, artificial intelligence systems, and human rights', *ERA Forum*, 20(4), pp. 567–583. Available at: https://doi.org/10.1007/s12027-020-00602-0.

Zhong, H. *et al.* (2020) 'How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence', *arXiv:2004.12158 [cs]* [Preprint]. Available at: http://arxiv.org/abs/2004.12158 (Accessed: 18 June 2021).

Βασίλης Ζορκάδης (2018) *Εκτίμηση Αντικτύπου στην Προστασία Δεδομένων*, *Γενικός Κανονισμός για την Προστασία των Προσωπικών Δεδομένων (GDPR)*. Available at: https://secure.livechatinc.com/ (Accessed: 23 February 2023).

Δημήτριος Ευ. Τζέλλης, Μαρία Δ. Μυλώση (2022) *Εκτίμηση Αντικτύπου στην προστασία προσωπικών δεδομένων*. Available at: https://secure.livechatinc.com/ (Accessed: 23 February 2023).

Χριστίνα-Ειρήνη Ασημακοπούλου (2020) *Το δικαίωμα εναντίωσης σε αυτοματοποιημένες αποφάσεις υπό τον Ν 2472/1997 και υπό τον ΓΚΠΔ 2016/679 - Δικαίωμα εξασφάλισης ανθρώπινης παρέμβασης και Τεχνητή Νοημοσύνη*, *Qualex*. Available at: https://www.qualex.gr/el-GR/periexomeno/arthrografia/arthrografia?id=1077305&search=data%20impact%20assessment&qt2=&r1=1&r2=1 (Accessed: 23 February 2023).