Πανεπιστήμιο Πειραιώς

Σχολή Τεχνολογιών Πληροφορικής και Τηλεπικοινωνίων

Τμήμα Ψηφιακών Συστημάτων

Μεταπτυχιακό Πρόγραμμα Σπουδών

Ασφάλεια Ψηφιακών Συστημάτων

Προστασία της ιδιωτικότητας κατά την επεξεργασία Δεδομένων Μεγάλου Όγκου (Big Data) με έμφαση σε περιβάλλον μηχανικής μάθησης (Machine Learning)

Επιβλέπων Καθηγητής: κ. Στέφανος Γκρίτζαλης

Ονοματεπώνυμο	E-mail	A.M.
Νεφέλη Χριστιά	n.christia@ssl-unipi.com	mte2034

Πειραιάς

14/10/2022

Περίληψη

Στη σύγχρονη πραγματικότητα, όλο και περισσότεροι είναι οι άνθρωποι που συνδέονται στο Διαδίκτυο, ανταλλάσσοντας πληροφορίες, μέσω ποικίλων μορφών επικοινωνίας. Χάρη στους ισχυρούς αλγορίθμους που έχουν δημιουργηθεί, τα δεδομένα που παράγονται μέσα από αυτές τις επικοινωνίες μπορούν να συλλεχθούν, να αναλυθούν και να παρέχουν γνώση. Οι τεράστιες ποσότητες των δεδομένων που συλλέγονται από διάφορες πηγές (π.χ. διαδίκτυο, αισθητήρες κλπ) έχουν φέρει στο προσκήνιο την έννοια των Δεδομένων Μεγάλου Όγκου (Big Data). Η ανάλυση αυτών βοηθά πλέον στη λήψη αποφάσεων σε πολλούς τομείς όπως το μάρκετινγκ μέσω εξατομικευμένων διαφημίσεων ανάλογα με τα προφίλ που δημιουργούνται γύρω από τα υποκείμενα των δεδομένων. Ωστόσο, παρόλο που η αξιοποίηση των σύγχρονων τεχνολογιών για την ανάλυση των Big Data μπορεί να καταστήσει τη λήψη αποφάσεων ευκολότερη και αποδοτικότερη περιορίζοντας την ανθρώπινη παρέμβαση, υπάρχει ο κίνδυνος της αποκάλυψης των δεδομένων που προκύπτουν από μια παραβίαση ασφάλειας ή την ανταλλαγή δεδομένων σε τρίτους χωρίς τη συγκατάθεση του ατόμου προσβάλλοντας έτσι το δικαίωμά του στην ιδιωτική ζωή.

Η συμμόρφωση με την προστασία των προσωπικών δεδομένων στον σύγχρονο κόσμο που χαρακτηρίζεται από την ψηφιοποίηση και την τεχνολογική εξέλιξη αποτελεί πρόκληση για τη Μηχανική Μάθηση και εφαρμογές όπως είναι η κατάρτιση προφίλ και η αυτοματοποιημένη λήψη αποφάσεων. Ο Γενικός Κανονισμός Προστασίας των δεδομένων, λαμβάνοντας υπόψη τις τεχνολογικές εξελίξεις, συμπεριλαμβάνει δικλείδες ασφαλείας προκειμένου να προστατεύονται τα υποκείμενα των δεδομένων και το αναφαίρετο δικαίωμά τους στην ιδιωτικότητα.

Abstract

Nowadays, more and more people are connecting to the Internet, exchanging information through various forms of communication. Due to the powerful algorithms that have been created, the data generated through these communications can be collected, analyzed and provide knowledge. The vast amounts of data collected from various sources (e.g. internet, sensors, etc.) have brought to the foreground the concept of Big Data. The analysis of Big Data helps in decision making in many areas such as marketing through personalized advertisements depending on the profiles created around the data subjects. However, although the use of modern technologies to analyze Big Data can make decision making easier and more efficient by limiting human intervention, there is a risk of disclosure of data resulting from a security breach or sharing data with third parties without the consent of the individual thus infringing on their right to privacy.

Complying with data protection in today's world characterized by digitization and technological evolution is a challenge for Machine Learning and applications such as profiling and automated decision making. The General Data Protection Regulation, taking into account technological developments, includes safeguards to protect data subjects and their inalienable right to privacy.

Περιεχόμενα

1. Εισαγωγή στα Δεδομένα Μεγάλου Όγκου 1
   1.1. Ορισμός 1
   1.2. Χαρακτηριστικά 1
   1.3. Τύποι Big Data 3
   1.4. Κύκλος Ζωής Big Data 4
      1.1.1. Δημιουργία δεδομένων 4
      1.1.2. Συλλογή δεδομένων 4
      1.1.3. Ανάλυση δεδομένων 5
   1.5. Προκλήσεις των Big Data 5
2. Εισαγωγή στη Μηχανική Μάθηση 7
   2.1. Ορισμός 7
   2.2. Big Data και Μηχανική Μάθηση (Machine Learning) 7
   2.3. Εφαρμογές Μηχανικής Μάθησης (Machine Learning) 8
3. Προστασία Προσωπικών Δεδομένων 10
   3.1. Ιδιωτικότητα και Προστασία Προσωπικών Δεδομένων 10
   3.2. Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) 11
      3.2.1. Αρχή νομιμότητας, αντικειμενικότητας και διαφάνειας 11
      3.2.2. Αρχή περιορισμού του σκοπού 12
      3.2.3. Αρχή ελαχιστοποίησης των δεδομένων 13
      3.2.4. Αρχή ακρίβειας των δεδομένων 13
      3.2.5. Αρχή περιορισμού της περιόδου αποθήκευσης 14
      3.2.6. Αρχή ακεραιότητας και εμπιστευτικότητας 15
      3.2.7. Αρχή λογοδοσίας 15
4. Κατάρτιση προφίλ και αυτοματοποιημένη λήψη αποφάσεων 17
   4.1. Κατάρτιση Προφίλ 17
      4.1.1. Κατάρτιση προφίλ και Μηχανική Μάθηση 18
      4.1.2. Κατάρτιση προφίλ στο πλαίσιο του ΓΚΠΔ 19
      4.1.3. Δικαιώματα υποκειμένων των δεδομένων 20
   4.2. Αυτοματοποιημένη λήψη αποφάσεων 21
      4.2.1. Αυτοματοποιημένη λήψη αποφάσεων στο πλαίσιο του ΓΚΠΔ 21
      4.2.2. Εξαίρεση από την απαγόρευση 24
      4.2.3. Δικαιώματα Υποκειμένου 25
5. Νομοθεσία και Big Data σε περιβάλλον Μηχανικής Μάθησης – Νομικές Προκλήσεις 27
   5.1. Έλλειψη διαφάνειας 27

5.2. Εκτεταμένη συλλογή δεδομένων και δικαιώματα Υποκειμένου 29
5.3. Περιορισμός σκοπού 30
5.4. Αδικία και διακρίσεις 32
6. Αντιμετώπιση Προκλήσεων 34
6.1. Ενίσχυση απαίτησης συγκατάθεσης 34
6.2. Ανωνυμοποίηση 34
6.3. Ψευδωνυμοποίηση 36
6.4. Προστασία των δεδομένων ήδη από το σχεδιασμό και εξ ορισμού 39
6.5. Κρυπτογράφηση 42
7. Συμπεράσματα 46
Βιβλιογραφία 47

Λίστα Εικόνων

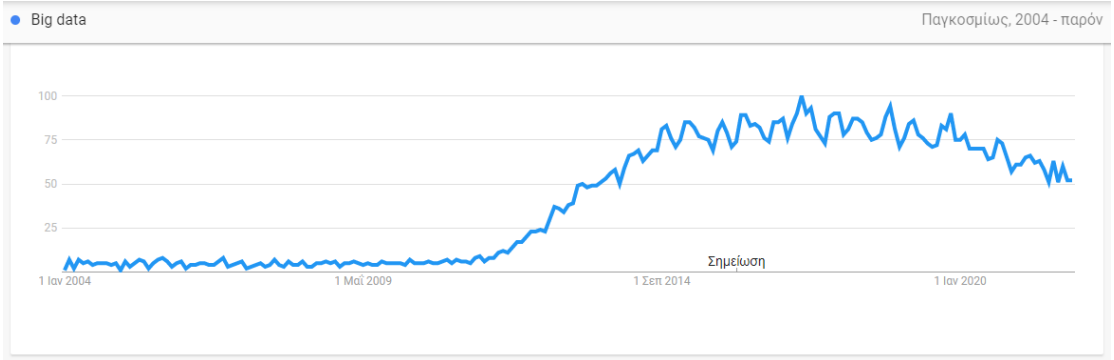
Εικόνα 1: Γράφημα ενδιαφέροντος για τον όρο "Big Data" 1
Εικόνα 2: Χαρακτηριστικά των Big Data 3
Εικόνα 3: Εφαρμογές Μηχανικής Μάθησης [3] 9
Εικόνα 4: Βασικές αρχές ΓΚΠΔ [6] 16
Εικόνα 5: Το φαινόμενο του "μαύρου κουτιού" [11] 28
Εικόνα 6: Σχήμα κρυπτογράφησης 43

1. Εισαγωγή στα Δεδομένα Μεγάλου Όγκου

1.1. Ορισμός

Ο όρος «Δεδομένα Μεγάλου Όγκου» (εφεξής «Big Data») χρησιμοποιείται ευρέως στον ακαδημαϊκό και στο βιομηχανικό χώρο και υποδηλώνει σύνολα δεδομένων, τα οποία έχουν συλλεχθεί και είναι τόσο μεγάλα σε όγκο, ώστε να απαιτούνται σύγχρονες τεχνολογίες για την εξόρυξη, την αποθήκευση, τη διαχείριση και την ανάλυσή τους. Η ποικιλία των Big Data και η ιλιγγιώδης ταχύτητα με την οποία αυτά μεταφέρονται, κάνουν επιτακτική την ανάγκη ανάπτυξης σύγχρονων μεθόδων με σκοπό την επεξεργασία των Big Data και την εξόρυξη πληροφοριών από αυτά.

Όπως φαίνεται, με τη χρήση του εργαλείου Google Trends, ο όρος «Big Data» έχει γίνει αρκετά δημοφιλής τα τελευταία χρόνια και αναζητείται αρκετά, ιδιαίτερα από το 2012 και μετά.



Εικόνα 1: Γράφημα ενδιαφέροντος για τον όρο "Big Data"

Παρόλο που ο όρος «Big Data» προϋπάρχει εδώ και αρκετά χρόνια, δεν έχει καθιερωθεί ένας ενιαίος και σαφής ορισμός γι' αυτόν. Το γεγονός αυτό, δεν είναι παράλογο αν αναλογιστεί κανείς ότι οι οργανισμοί και οι επιχειρήσεις που χρησιμοποιούν Big Data τα ορίζουν σύμφωνα με τη δική τους.

1.2. Χαρακτηριστικά

Με βάση τους διάφορους ορισμούς που έχουν δοθεί κατά καιρούς για τα Big Data, το κύριο χαρακτηριστικό τους, είναι ο μεγάλος τους όγκος, ο οποίος δεν μπορεί εύκολα να διαχειριστεί από τα παραδοσιακά λογισμικά επεξεργασίας δεδομένων. Ωστόσο, τα Big Data διαθέτουν και άλλα εξίσου σημαντικά χαρακτηριστικά. Τα χαρακτηριστικά αυτά, είναι γνωστά ως **«5 V's»** [1] και αναλύονται ως εξής:

- **Όγκος (Volume):** Το χαρακτηριστικό αυτό αναφέρεται σε μεγάλες ποσότητες δεδομένων, διαφορετικών ειδών. Τα δεδομένα αυτά μπορεί να περιλαμβάνουν ροές δεδομένων από κοινωνικά δίκτυα, δεδομένα από ιστοσελίδες ή εφαρμογές. Ανάλογα με τον εκάστοτε οργανισμό, ο όγκος των δεδομένων διαφέρει και μπορεί για κάποιους οργανισμούς να είναι δεκάδες terabyte ενώ για άλλους, εκατοντάδες petabytes.

- **Ταχύτητα (Velocity):** Αναφέρεται στην ταχύτητα ροής των δεδομένων. Ο αυξημένος όγκος αλλά και η ποικιλία των δεδομένων, αυξάνει την ταχύτητα με την οποία αυτά συλλέγονται, αποθηκεύονται, επεξεργάζονται και αναλύονται.

Παράδειγμα του συγκεκριμένου χαρακτηριστικού είναι η ταχύτητα με την οποία πραγματοποιούνται πλέον οι διαδικτυακές συναλλαγές, δημιουργώντας δεδομένα. Γίνεται, λοιπόν, επιτακτική η ανάγκη ανάπτυξης αλγορίθμων και τρόπων για ταχεία επεξεργασία και ανάλυση των δεδομένων.

- **Ποικιλομορφία (Variety):** Αναφέρεται στους διαφορετικούς τύπους δεδομένων που μπορούν πλέον να χρησιμοποιηθούν. Τα δομημένα δεδομένα, αποτελούν το παραδοσιακότερο τύπο δεδομένων και για το λόγο αυτό, έχουν διαρθρωθεί και προσαρμοστεί σε σχεσιακές βάσεις δεδομένων. Πλέον όμως, τα Big Data δεν είναι πάντα δομημένα δεδομένα και δεν τοποθετούνται πάντα με ευκολία σε σχεσιακές βάσεις δεδομένων. Αντίθετα, με την τεχνολογία των Big Data αξιοποιούνται επιπλέον ημι-δομημένα αλλά και αδόμητα δεδομένα, τα οποία μπορούν να συλλεχθούν από φωτογραφίες, ηχητικά μηνύματα ή ακόμα και από βιομετρικούς αισθητήρες. Γίνεται κατανοητό, ότι η διαχείριση δομημένων, ημι-δομημένων και αδόμητων δεδομένων, αυξάνει την πολυπλοκότητα τόσο της αποθήκευσης όσο και της ανάλυσης των Big Data.

- **Εγκυρότητα (Veracity):** Αναφέρεται στην αξιοπιστία των δεδομένων που συλλέγονται. Κατά τη διαχείριση δεδομένων που χαρακτηρίζονται από μεγάλο όγκο, ταχύτητα και ποικιλομορφία, η ακρίβειά τους χρίζει προσοχής. Η ποιότητα των δεδομένων που συλλέγονται μπορεί να ποικίλει και να καθορίσει σε μεγάλο βαθμό την ποιότητα των παραγόμενων αποτελεσμάτων ανάλυσης.

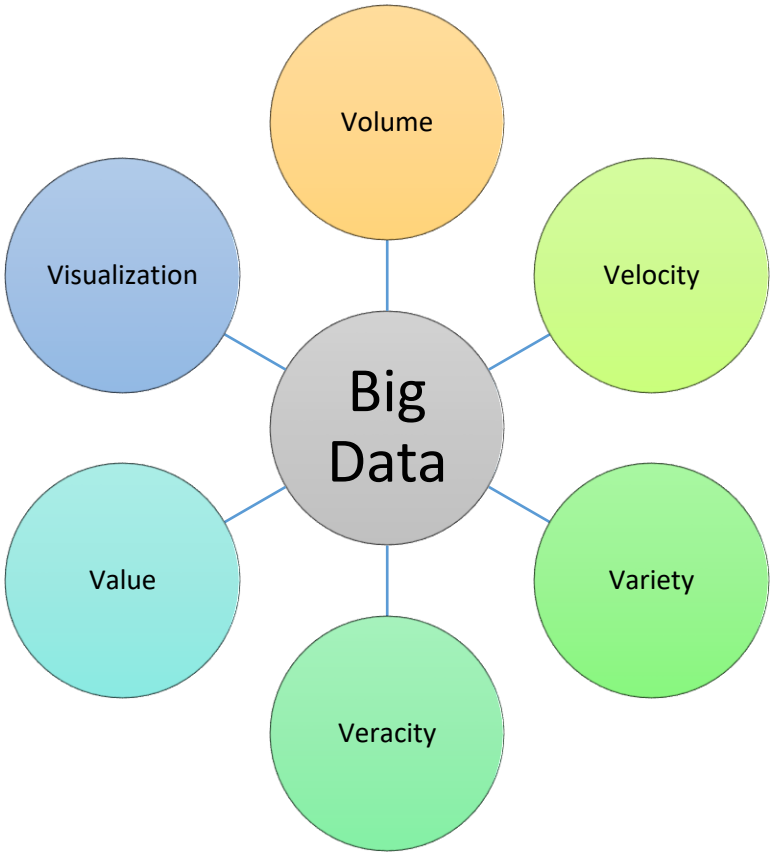
- **Αξία (Value):** Πρόκειται για ένα από τα πιο σημαντικά χαρακτηριστικά των Big Data καθώς η δυνητική αξία τους είναι τεράστια. Είναι γεγονός ότι τα δεδομένα, η κατοχή τους και η πρόσβαση σε αυτά, έχουν αξία. Ωστόσο, δεν έχουν καμία χρησιμότητα μέχρι να επαληθευθεί η αξία που προσδίδουν στην πληροφορία. Επιχειρήσεις και οργανισμοί καλούνται να επιλέξουν την πιο συμφέρουσα λύση με στόχο την αξιοποίηση της πληροφορίας που θα συμβάλλει στη λήψη σωστών αποφάσεων.

Τα τελευταία χρόνια έχουν προστεθεί επιπλέον χαρακτηριστικά στα Big Data, τα οποία αναλύονται ως εξής:

- **Μεταβλητότητα (Variability):** Στην ανάλυση των Big Data, η ασυνέπεια των δεδομένων αποτελεί ένα σύνηθες φαινόμενο καθώς τα δεδομένα προέρχονται από διαφορετικές πηγές. Επιπλέον, όπως έχει ήδη προαναφερθεί, στα Big Data εμπεριέχονται διαφορετικοί τύποι δεδομένων. Για την επιλογή, λοιπόν, των κατάλληλων δεδομένων από έναν τεράστιο όγκο δεδομένων, η ανίχνευση ανωμαλιών και ακραίων στοιχείων είναι απαραίτητη. Καθίσταται σαφές, λοιπόν, πως η μεταβλητότητα αποτελεί αναπόσπαστο χαρακτηριστικό των Big Data.

- **Απεικόνιση (Visualization):** Ένα από τα βασικότερα καθήκοντα των συστημάτων που επεξεργάζονται Big Data, είναι να μετατρέπουν τον τεράστιο όγκο των δεδομένων σε κάτι εύκολα κατανοητό. Μία από τις καλύτερες μεθόδους, που χρησιμοποιείται στις επιχειρήσεις και τους οργανισμούς, είναι η μετατροπή των δεδομένων σε γραφήματα.

Η λίστα με τα χαρακτηριστικά των Big Data, ολοένα και μεγαλώνει και κανείς δεν εγγυάται ότι στο μέλλον δεν θα υπάρξουν περαιτέρω προσθήκες.



Εικόνα 2: Χαρακτηριστικά των Big Data

1.3. Τύποι Big Data

Όπως έχει ήδη γίνει κατανοητό, ο όγκος των Big Data είναι ανάλογος με την ποικιλομορφία τους. Δεν είναι όλα τα δεδομένα εύκολο να χρησιμοποιηθούν σε μεθόδους αναλυτικής (analytics) ή σε περιβάλλοντα μηχανικής μάθησης. Συχνά, τα Big Data χρειάζεται να υποστούν επεξεργασία προκειμένου να μετασχηματιστούν σε δεδομένα προς χρήση.

Οι διαφορετικοί τύποι των Big Data είναι οι εξής:

- Δομημένα δεδομένα
- Μη δομημένα δεδομένα
- Ημι-δομημένα δεδομένα

Η κατανόηση της κατηγορίας στην οποία ανήκουν τα Big Data που θα χρησιμοποιηθούν με περιβάλλοντα μηχανικής μάθησης αποτελεί διαδικασία υψίστης σημασίας. Η κατανόηση της πηγής από την οποία προέρχονται τα ακατέργαστα δεδομένα και πώς πρέπει αυτά να αντιμετωπιστούν πριν από την ανάλυσή τους γίνεται πιο σημαντική κατά την επεξεργασία μεγάλου όγκου Big Data. Λόγω του τεράστιου όγκου τους, είναι σημαντικό να εξάγουν αξιόπιστα αποτελέσματα.

Ο τύπος των δεδομένων είναι το κλειδί όχι μόνο για τον τρόπο επεξεργασίας αυτών, αλλά και για τις γνώσεις που τα ίδια παράγουν. Όλα τα δεδομένα περνούν από μια διαδικασία που ονομάζεται μετασχηματισμός προτού να είναι σε θέση να αναλυθούν. Συνοπτικά, τα δεδομένα συλλέγονται, μορφοποιούνται ώστε να είναι αναγνώσιμα από μια εφαρμογή και

στη συνέχεια αποθηκεύονται για χρήση. Η διαδικασία που ακολουθείται για κάθε τύπο δεδομένων ποικίλλει. Ας εξετάσουμε το κάθε τύπο ξεχωριστά:

Δομημένα δεδομένα: Ο όρος δομημένα δεδομένα αναφέρεται σε δεδομένα που έχουν καθορισμένο μήκος και μορφή. Τα δεδομένα αυτά μπορούν να αποθηκευτούν σε σχεσιακή βάση δεδομένων σε μορφή πίνακα με σειρές και στήλες. Παραδείγματα δομημένων δεδομένων περιλαμβάνουν ηλικίες, αριθμούς, διευθύνσεις και ημερομηνίες. Τα δεδομένα αυτά δημιουργούνται είτε αυτόματα από μηχανές και υπολογιστές χωρίς ανθρώπινη παρέμβαση είτε παρέχονται από ανθρώπους σε συνεργασία με μηχανές. Τα δομημένα δεδομένα είναι ο ευκολότερος τύπος δεδομένων για ανάλυση, καθώς απαιτείται λίγη έως καθόλου προετοιμασία πριν από την επεξεργασία τους.

Μη δομημένα δεδομένα: Τα δεδομένα αυτά είναι ακατέργαστα, μη οργανωμένα και δεν ταιριάζουν στα σχεσιακά συστήματα βάσεων δεδομένων, αφού δεν ακολουθούν μια καθορισμένη μορφή. Τα περισσότερα δεδομένα που υπάρχουν δεν είναι δομημένα. Παραδείγματα μη δομημένων δεδομένων αποτελούν τα βίντεο, ο ήχος και οι εικόνες. Το πιο δύσκολο μέρος της ανάλυσης μη δομημένων δεδομένων είναι η εκπαίδευση μιας εφαρμογής ώστε να κατανοεί τις πληροφορίες που εξάγει. Τις περισσότερες φορές, αυτό σημαίνει τη μετάφραση των πληροφοριών σε κάποια μορφή δομημένων δεδομένων.

Ημι-δομημένα δεδομένα: Τα ημι-δομημένα δεδομένα είναι ένα είδος δεδομένων που εμπίπτει μεταξύ δομημένων και μη δομημένων δεδομένων. Τα δεδομένα αυτά δε συμμορφώνονται υποχρεωτικά με ένα καθορισμένο σχήμα, δηλαδή δομή, αλλά περιέχουν ετικέτες ή άλλους δείκτες για να διαχωρίσουν σημασιολογικά στοιχεία και να επιβάλουν ιεραρχίες αρχείων και πεδίων εντός των δεδομένων. Παραδείγματα ημι-δομημένων δεδομένων περιλαμβάνουν αρχεία JSON, SWIFT και XML. Τα ημι-δομημένα δεδομένα, αναλύονται πιο εύκολα από τα μη δομημένα δεδομένα αφού πολλές λύσεις και εργαλεία επεξεργασίας Big Data προσφέρουν τη δυνατότητα αυτή.

1.4. Κύκλος Ζωής Big Data

1.1.1. Δημιουργία δεδομένων

Προκειμένου να ξεκινήσει ο κύκλος ζωής των Big Data, πρέπει αυτά να δημιουργηθούν. Χωρίς τη δημιουργία των δεδομένων, τα επόμενα βήματα δεν μπορούν να υπάρξουν. Στο σύγχρονο ψηφιακό κόσμο, τα δεδομένα παράγονται συνεχώς από διαφορετικές πηγές. Οποιαδήποτε δραστηριότητα, επικοινωνία, αλληλεπίδραση οντοτήτων στο διαδίκτυο, παράγει δεδομένα. Τα δεδομένα αυτά, εφόσον αντιμετωπιστούν κατάλληλα, μπορούν να οδηγήσουν στην παραγωγή χρήσιμων αποτελεσμάτων τα οποία με τη σειρά τους μπορούν να αξιοποιηθούν για μεγαλύτερη απόδοση οργανισμών και εταιρειών.

1.1.2. Συλλογή δεδομένων

Η συλλογή των δεδομένων αποτελεί το δεύτερο βήμα στον κύκλο ζωής τους. Στη φάση αυτή περιλαμβάνεται η απόκτηση των δεδομένων, η αποθήκευση αυτών και η προεπεξεργασία τους. Η συλλογή των δεδομένων συνεπάγεται την απόκτηση αυτών από διαφορετικές πηγές και σε διαφορετικές μορφές (δομημένα, ημι-δομημένα, μη δομημένα δεδομένα). Τα δεδομένα αυτά, μεταφέρονται σε κατάλληλες υποδομές αποθήκευσης και περιλαμβάνουν μεγάλο ποσοστό περιττής, άχρηστης ή και λανθασμένης πληροφορίας. Κρίνεται, λοιπόν, απαραίτητη η προεπεξεργασία τους ώστε αυτά να καθαριστούν και να μετασχηματιστούν ώστε να είναι αξιόπιστα, συνεπή και ακριβή και να οδηγήσουν σε αντίστοιχα συμπεράσματα. Η προεπεξεργασία των δεδομένων, μειώνει σημαντικά των

όγκο τους, με αποτέλεσμα να εξοικονομείται χώρος αποθήκευσης και μειώνεται το σχετικό κόστος.

1.1.3. Ανάλυση δεδομένων

Μετά τη συλλογή, την αποθήκευση και την προεπεξεργασία των δεδομένων, τα ίδια υποβάλλονται σε περαιτέρω επεξεργασία και αναλύονται προκειμένου να δημιουργηθεί χρήσιμη γνώση. Για την ανάλυση των δεδομένων, χρησιμοποιείται πλήθος τεχνικών και μεθόδων, μερικές εκ των οποίων είναι η ομαδοποίηση (clustering) και η ταξινόμηση (classification).

1.5. Προκλήσεις των Big Data

Είναι γεγονός ότι η χρήση των Big Data έχει επιφέρει πολλαπλά οφέλη σε πολλούς τομείς. Ωστόσο, εκτός από τα πλεονεκτήματα που υπάρχουν στη χρήση τους, ανακύπτουν προβληματισμοί σχετικά με τη ασφάλεια της πληροφορίας και της ιδιωτικότητα.

Κάθε συζήτηση σχετική με την ασφάλεια των πληροφοριών στηρίζεται σε τρεις βασικές αρχές. Αυτές είναι οι: Εμπιστευτικότητα (Confidentiality), Ακεραιότητα (Integrity), Διαθεσιμότητα (Availability). Οι ιδιότητες, όμως, αυτές, δεν επαρκούν ώστε να οριστεί η ασφάλεια των πληροφοριών. Πρόσθετες ιδιότητες που συναντώνται είναι οι εξής:

- Ιδιωτικότητα (Privacy): αναφέρεται «…στο δικαίωμα των πολιτών σε μία ανενόχλητη ιδιωτική ζωή (the right of an individual to be let alone)» (S. Warren, L. Brandeis, 1980)
- Αυθεντικοποίηση (Authentication): η οποία στοχεύει στην πιστοποίηση της ταυτότητας μίας οντότητας.
- Μη αποποίηση (Non repudiation): σύμφωνα με την οποία κανένας χρήστης, ε δεν μπορεί να αποποιηθεί τις πράξεις που έκανε. Ένας από τους πιο διαδεδομένους αλλά και ταυτόχρονα αποτελεσματικούς τρόπους εφαρμογής της μη αποποίησης είναι η χρήση των ψηφιακών υπογραφών από την πλευρά του αποστολέα και του παραλήπτη.

Οι κλασσικοί μηχανισμοί ασφάλειας, καθίστανται ανεπαρκείς όσον αφορά την ασφάλεια και την ιδιωτικότητα σε περιβάλλοντα Big Data, καθώς τα ίδια έχουν επιφέρει αλλαγές στον ψηφιακό κόσμο που κυμαίνονται τόσο από το μεγάλο όγκο ως και τον τρόπο με τον οποίο αποθηκεύονται και διαχειρίζονται. Στο πλαίσιο αυτό, οι βασικότερες προκλήσεις που σχετίζονται με την ασφάλεια των Big Data συνοψίζονται ως εξής:

- Από τη **συλλογή** ακόμα των δεδομένων, αισθητή γίνεται η πολυπλοκότητα των δεδομένων αλλά και των πηγών τους. Προκειμένου η φάση αυτή να διενεργείται με ασφάλεια και να παρέχει αξιόπιστα αποτελέσματα, είναι σημαντικό να χρησιμοποιούνται αξιόπιστα δεδομένα. Επιπλέον, απαιτούνται πρόσθετα μέτρα ασφάλειας ώστε τα δεδομένα που συλλέγονται να παραμένουν ασφαλή. Ορισμένα μέτρα ασφάλειας που μπορούν να χρησιμοποιηθούν είναι ο περιορισμένος έλεγχος πρόσβασης στα δεδομένα καθώς και η κρυπτογράφηση ορισμένων από τα πεδία αυτών. Δεν είναι λίγες οι φορές όπου τα δεδομένα συλλέγονται μέσω διαδικτύου ή μέσων κοινωνικής δικτύωσης, χωρίς να υπάρχει η ρητή συγκατάθεση του υποκειμένου των δεδομένων.
- Αφού ολοκληρωθεί η φάση της συλλογής, τα δεδομένα **αποθηκεύονται** με σκοπό να χρησιμοποιηθούν για ανάλυση. Καθώς τα δεδομένα που συλλέγονται μπορεί να

περιέχουν ευαίσθητες πληροφορίες, είναι σημαντικό να εφαρμόζονται πρόσθετα μέτρα ασφάλειας για την αποθήκευσή τους. Τα αποθηκευμένα δεδομένα οφείλουν να προστατεύονται από πολλαπλές απειλές συνδυάζοντας τεχνικές φυσικής ασφάλειας αλλά και τεχνολογίες προστασίας δεδομένων. Σε περιβάλλοντα, όπως αυτά της νεφοϋπολογιστικής (cloud computing), όπου μπορεί εύκολα να διακυβευτεί η ακεραιότητα και η εμπιστευτικότητα των δεδομένων είναι σημαντικό να εφαρμόζονται τεχνολογίες διατήρησης του απορρήτου (π.χ. κρυπτογράφηση). Παράλληλα, σε οποιοδήποτε χώρο αποθηκεύονται ευαίσθητα δεδομένα, κρίνεται αναγκαίο, να παρέχεται η πρόσβαση μόνο σε εξουσιοδοτημένα άτομα.

- Η φάση που ακολουθεί μετά τη συλλογή και την αποθήκευση των δεδομένων, όπως έχει ήδη προαναφερθεί, είναι η φάση της **ανάλυσης**. Τα δεδομένα επεξεργάζονται και αναλύονται με σκοπό να παράγουν γνώση. Στη φάση αυτή, είναι σημαντικό να παρέχεται στα δεδομένα, ένα ασφαλές περιβάλλον επεξεργασίας και ανάλυσης. Καθίσταται, ξανά, σαφές ότι η πρόσβαση στα περιβάλλοντα αυτά, είναι απαραίτητο να περιορίζεται σε εξουσιοδοτημένα άτομα.

2. Εισαγωγή στη Μηχανική Μάθηση

2.1. Ορισμός

Η Μηχανική Μάθηση αποτελεί μία τεχνική της της επιστήμης των υπολογιστών, η οποία επιτρέπει στους υπολογιστές να "μαθαίνουν" από μόνοι τους. Συχνά, η Μηχανική Μάθηση συγχέεται με την Τεχνητή Νοημοσύνη, αλλά ουσιαστικά η Μηχανική Μάθηση είναι μόνο ένας κλάδος της Τεχνητής Νοημοσύνης. Το χαρακτηριστικό που διαχωρίζει τη Μηχανική Μάθηση από τις άλλες μορφές Τεχνητής Νοημοσύνης είναι η δυναμική ικανότητά της να τροποποιείται και να παράγει ακριβέστερα αποτελέσματα όταν εκτίθεται σε περισσότερα δεδομένα. Μέσω της πρόσληψης δεδομένων, η μηχανή εκπαιδεύεται αναπτύσσοντας τη δική της λογική η οποία βασίζεται στα δεδομένα τα οποία έχει αναλύσει. Όσον αφορά τη διατύπωση προβλέψεων, η ελλιπής πληροφόρηση μπορεί να οδηγήσει σε λανθασμένη λήψη αποφάσεων για τους ανθρώπους. Παρομοίως, η έλλειψη δεδομένων μπορεί να οδηγήσει στον εντοπισμό λανθασμένων προτύπων από τη μηχανή. Επομένως, η πραγματοποίηση ακριβέστερων προβλέψεων απαιτεί τη είσοδο περισσότερων δεδομένων για τη μηχανή. Παράλληλα, σημαντικό ρόλο για την αποτελεσματικότητα των αλγορίθμων Μηχανικής Μάθησης κατέχει και η μεγάλη επεξεργαστική ισχύς της μηχανής. Δεδομένου ότι τα οφέλη της Μηχανικής Μάθησης έχουν γίνει αντιληπτά από την τάση της ψηφιοποίησης όλα αυτά τα χρόνια, όλο και περισσότερες επιχειρήσεις άρχισαν να επωφελούνται από την εφαρμογή της.

Ο τομέας της Μηχανικής Μάθησης συχνά υποδιαιρείται σε μικρότερους τομείς ανάλογα με τα είδη των προβλημάτων που αντιμετωπίζονται. Μια πρόχειρη κατηγοριοποίηση περιλαμβάνει δύο τύπους Μηχανικής Μάθησης: την επιβλεπόμενη και τη μη επιβλεπόμενη. Η επιβλεπόμενη μάθηση απαιτεί από τον άνθρωπο να παρέχει στη μηχανή τόσο δεδομένα όσο και αποτελέσματα, με στόχο αυτή να κατασκευάζει μια συνάρτηση που απεικονίζει δεδομένες εισόδους (σύνολο εκπαίδευσης) σε γνωστές επιθυμητές εξόδους, με απώτερο στόχο τη γενίκευση της συνάρτησης αυτής και για εισόδους με άγνωστη έξοδο. Αυτή η μέθοδος χρησιμοποιείται σε προβλήματα ταξινόμησης (classification), πρόγνωσης (prediction) και διερμηνείας (interpretation). Από την άλλη, στη μη επιβλεπόμενη μάθηση, η μηχανή μαθαίνει πιο ελεύθερα αφού ο αλγόριθμος κατασκευάζει ένα μοντέλο για κάποιο σύνολο εισόδων (συχνά Big Data) υπό μορφή παρατηρήσεων χωρίς να γνωρίζει τις επιθυμητές εξόδους προκειμένου να προσδιορίσει μοτίβα και συσχετίσεις. Η μέθοδος αυτή χρησιμοποιείται κυρίως σε προβλήματα ανάλυσης συσχετισμών (association analysis) και ομαδοποίησης (clustering).

Οι διαφορετικές τεχνικές Μηχανικής Μάθησης χρησιμοποιούνται σε διαφορετικά πλαίσια και για διαφορετικούς σκοπούς. Καμία από τις δύο δεν απαιτεί ρητό προγραμματισμό σχετικά με το τί πρέπει να αναζητήσει, γεγονός που προσδίδει ένα επίπεδο αυτονομίας στη μηχανή, να παράγει τη δική του λογική εντοπίζοντας τάσεις που μπορεί να είχαν διαφύγει από την ανθρώπινη λογική. Γίνεται κατανοητό ότι ο βαθμός στον οποίο εξασφαλίζεται η εγκυρότητα των αποτελεσμάτων της Μηχανικής Μάθησης εξαρτάται άμεσα από τα δεδομένα εισόδου που παρέχονται. Εξαιτίας αυτού, τα Big Data διαδραματίζουν καθοριστικό ρόλο για την επιτυχία της Μηχανικής Μάθησης.

2.2. Big Data και Μηχανική Μάθηση (Machine Learning)

Ο όγκος των δεδομένων που συλλέγονται στο πλαίσιο των Big Data είναι αδιαμφισβήτητα μεγάλος. Η χρήση των παραδοσιακών εργαλείων για την επεξεργασία τους, όμως, δυσκολεύει την πλήρη αξιοποίηση των δεδομένων και απαιτεί αρκετό χρόνο, ανθρώπινο δυναμικό και υλικό. Ένα νέο σύνολο εργαλείων που διευκολύνουν και εν μέρει

αυτοματοποιούν την επεξεργασία και την ανάλυση των Big Data είναι γνωστό ως Μηχανική Μάθηση (Machine Learning). Η Μηχανική Μάθηση αποτελεί κλάδο της Τεχνητής Νοημοσύνης και στηρίζεται στην ιδέα ότι τα συστήματα μπορούν να μαθαίνουν, ανακαλύπτοντας κανόνες συσχετίσεων βάσει ενός τεράστιου όγκου δεδομένων. Στόχος της Μηχανικής Μάθησης δεν είναι στην πραγματικότητα η απόκτηση τυποποιημένης μάθησης αλλά η κατανόηση της δομής των δεδομένων και η ενσωμάτωσή της σε μοντέλα, για την αυτοματοποίηση εργασιών και την παραγωγή συμπερασμάτων. [2]

Οι μέθοδοι Μηχανικής Μάθησης βελτιώνουν σημαντικά την αποτελεσματικότητα των συστημάτων εκπαιδεύοντας αλγορίθμους ώστε να επιτυγχάνουν τους προγραμματισμένους στόχους με βάση μεγάλο όγκο δεδομένων. Οι αλγόριθμοι εντοπίζουν μοτίβα και μαθαίνουν πώς να κάνουν προβλέψεις μέσω της επεξεργασίας δεδομένων και της ιστορικότητας αντί να λειτουργούν με βάση ρητές προγραμματιστικές οδηγίες. Οι αλγόριθμοι εξακολουθούν να μαθαίνουν όσο χρησιμοποιούνται και βελτιώνουν την αποτελεσματικότητά τους με την πάροδο του χρόνου καθώς αποθηκεύουν δεδομένα και λαμβάνουν αποφάσεις με βάση την εμπειρία τους.

Η Μηχανική Μάθηση συνέβαλε στην ανάπτυξη της βαθιάς μάθησης (deep learning). Η βαθιά μάθηση είναι μία τεχνική Μηχανικής Μάθησης στην οποία αρκετά «επίπεδα» απλών μονάδων επεξεργασίας συνδέονται σε ένα δίκτυο, το ένα πίσω από το άλλο, με αποτέλεσμα η είσοδος στο σύστημα να διέρχεται διαδοχικά μέσα από κάθε ένα από αυτά. Η βαθιά μάθηση, επομένως, επεξεργάζεται έναν μεγάλο όγκο δεδομένων και απαιτεί ελάχιστο προγραμματισμό από τον άνθρωπο. Τα αποτελέσματα που προκύπτουν από τη βαθιά μάθηση είναι συνήθως πιο ακριβή σε σχέση με αυτά που προκύπτουν από τις παραδοσιακές μεθόδους Μηχανικής Μάθησης. Επιπλέον, έχει αποδειχθεί ότι ο μεγάλος όγκος των δεδομένων που χρησιμοποιούνται, συμβάλλει στην παραγωγή ακριβούς αποτελέσματος ακόμα και από τους πιο αδύναμους αλγορίθμους.

Σήμερα, ο όγκος των δεδομένων και η ταχύτητα με την οποία αυτά μεταδίδονται έχουν γίνει υπερβολικά μεγάλες ώστε να μπορούν να επεξεργαστούν από τον ανθρώπινο εγκέφαλο. Ωστόσο, οι υπολογιστές και ιδιαίτερα οι αλγόριθμοι Μηχανικής Μάθησης μπορούν να επεξεργαστούν όλα αυτά τα δεδομένα και τις πληροφορίες με έναν εντελώς διαφορετικό τρόπο. Το σύνολο των δεδομένων που χρησιμοποιούνται για την εκμάθηση των αλγορίθμων αποτελείται από πληροφορίες οι οποίες προσδιορίζουν μία οντότητα (προσωπικά δεδομένα). Επομένως, η Μηχανική Μάθηση έχει αναπτύξει την ανάγκη συλλογής προσωπικών δεδομένων, παρομοιάζοντάς τα με το πετρέλαιο του 21ου αιώνα.

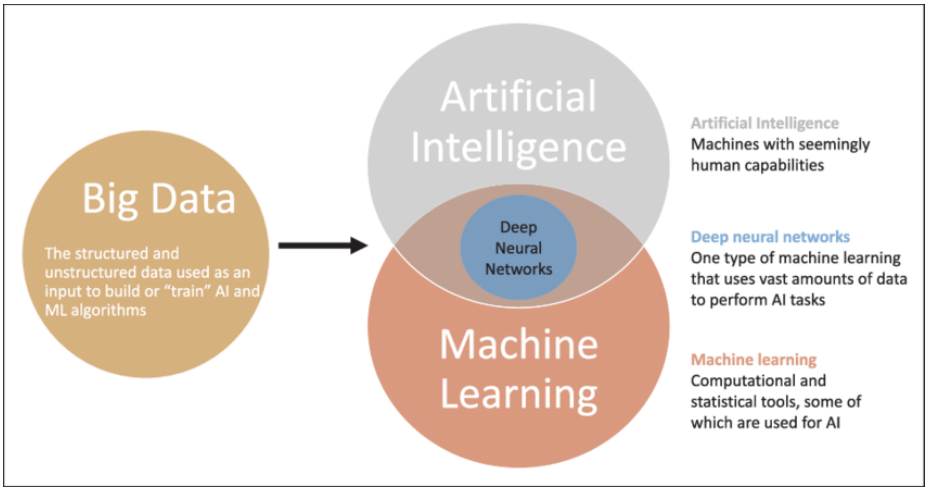
2.3. Εφαρμογές Μηχανικής Μάθησης (Machine Learning)

Η σύγχρονη πραγματικότητα αποτελείται από εφαρμογές της Μηχανικής Μάθησης σε μεγάλες ποσότητες επεξεργασμένων δεδομένων (Big Data). Η αυτόνομη οδήγηση, οι ιατρικές διαγνώσεις και οι προσωπικές προτάσεις σχετικά με προϊόντα και υπηρεσίες για τους χρήστες του διαδικτύου αποτελούν παραδείγματα εφαρμογής της Μηχανικής και της βαθιάς μάθησης. Οι εξελίξεις αυτές έχουν αναπτύξει την ανάγκη δημιουργίας προφίλ (profiling) και της αυτοματοποιημένης λήψης αποφάσεων με την οποία οι ανθρώπινες οντότητες έρχονται σε επαφή καθημερινά συνειδητά ή ασυνείδητα.

Η Μηχανική Μάθηση αποτελεί αναπόσπαστο κομμάτι της καθημερινής ζωής των ανθρώπων σήμερα. Η χρήση διαδικτυακών εργαλείων μετάφρασης, η χρήση της μηχανής αναζήτησης, η αγορά προϊόντων που έχουν προταθεί μέσω του διαδικτύου ακόμα και η επιλογή μουσικής που προτείνεται από το ίδιο, αποτελούν παραδείγματα εφαρμογών Μηχανικής Μάθησης. Οποιαδήποτε χρήση του διαδικτύου, δημιουργεί πληροφορίες και

δεδομένα τα οποία λειτουργούν ως ίχνη του εκάστοτε χρήση και δύνανται να χρησιμοποιηθούν από αλγορίθμους Μηχανικής Μάθησης προκειμένου να προσδιοριστούν τα ενδιαφέροντα του χρήστη με στόχο να λαμβάνει εξειδικευμένες διαφημίσεις και προτάσεις ή ακόμα και με κακόβουλο τρόπο προκειμένου να του υποκλέψουν στοιχεία και να αποκτήσουν πρόσβαση σε λογαριασμούς του. Η επισκόπηση του τρόπου με τον οποίο οι χρήστες χειρίζονται και διαδίδουν τα δεδομένα τους είναι περιορισμένη και μπορεί αυτό να μην αποτελεί πρόβλημα όταν γίνεται λόγος για μεμονωμένες περιπτώσεις, αλλά το πρόβλημα ανακύπτει όταν οι μεγάλες ποσότητες δεδομένων συγκεντρώνονται από διαφορετικές πηγές μέσω της Μηχανικής Μάθησης και διαμορφώνονται εξατομικευμένα μοτίβα και συσχετίσεις.

Οι συνέπειες της χρήσης της Μηχανικής Μάθησης είναι τόσο θετικές όσο και αρνητικές. Οι διαδικτυακές αναζητήσεις, οι πλοήγηση μέσω των διαδικτυακών χαρτών, οι φωνητικοί βοηθοί και οι εξατομικευμένες προτάσεις σε εφαρμογές όπως το Facebook, το Netflix και το YouTube αποτελούν μόνο κάποια παραδείγματα όπου η Μηχανικής Μάθησης, τα οποία αποδεικνύουν ότι η Μηχανική Μάθηση κατέχει ένα σημαντικό ρόλο στην καθημερινή ζωή. Επιπλέον, εταιρείες και οργανισμοί μέσω της Μηχανικής Μάθησης αυτοματοποιούν επιχειρηματικές διαδικασίες και λαμβάνουν καλύτερες αποφάσεις. Ωστόσο, ενώ η Μηχανική Μάθηση υπόσχεται ευκαιρίες για τη σύγχρονη πραγματικότητα, τις εταιρείες και τους οργανισμούς, δεν μπορεί να θεωρηθεί πανάκεια για όλες τις προκλήσεις που μπορεί να εμφανιστούν. Η χρήση των Big Data και της Μηχανικής Μάθησης πρέπει να γίνεται στρατηγικά και μελετημένα, δίνοντας ιδιαίτερη προσοχή σε θέματα που άπτονται της διαχείρισης πληροφοριών, συμπεριλαμβανομένης της προστασίας της ιδιωτικότητας και της προστασίας των δεδομένων. Όσο η Μηχανική Μάθηση εξελίσσεται και ενσωματώνεται όλο και περισσότερο στην καθημερινή ζωή, καλείται να αντιμετωπίσει και τις αντίστοιχες νομικές προκλήσεις. Οι αποφάσεις που προηγουμένως λαμβάνονταν αποκλειστικά από την ανθρώπινη σκέψη, πλέον δημιουργούνται μέσω της αλγοριθμικής λήψης αποφάσεων και αυτή πρέπει να λειτουργεί με τρόπο σύννομο για την προστασία των θεμελιωδών ανθρώπινων δικαιωμάτων. Η σχέση μεταξύ του δικαίου και της Μηχανικής Μάθησης, και πιο συγκεκριμένα ο τρόπος με τον οποίο οι νομοθετικές αποφάσεις για την κατάρτιση προφίλ και την αυτοματοποιημένη λήψη αποφάσεων αλληλοεπιδρούν με τις τεχνολογικές εξελίξεις, παρουσιάζεται στα επόμενα κεφάλαια.



Εικόνα 3: Εφαρμογές Μηχανικής Μάθησης [3]

3. Προστασία Προσωπικών Δεδομένων

3.1. Ιδιωτικότητα και Προστασία Προσωπικών Δεδομένων

Η έννοια της ιδιωτικότητας είναι ιδιαίτερα ενδιαφέρουσα καθώς έχει επί μακρόν απασχολήσει πολλά διαφορετικά επιστημονικά πεδία και έχει αποτελέσει τη βάση για πλήθος συζητήσεων ανάμεσα σε κοινωνικούς επιστήμονες, φιλοσόφους και νομικούς. Η ιδιωτικότητα έχει αναγνωριστεί ως βασικό πανανθρώπινο δικαίωμα και μπορεί στις σύγχρονες δημοκρατικές κοινωνίες η απαίτηση διασφάλισής της αποτελεί θεμελιώδη συνθήκη, αλλά με την αξιοποίηση των τεχνολογιών πληροφοριακών και επικοινωνιακών συστημάτων η ιδιωτικότητα βρίσκεται σε ολοένα αυξανόμενο κίνδυνο.

Αρκετές είναι οι προσπάθειες που έχουν γίνει προκειμένου να δοθεί ένας ορισμός για την ιδιωτικότητα. Οι Αμερικάνοι δικαστές S. Warren και L. Brandeis όρισαν την ιδιωτικότητα ως «…το δικαίωμα των πολιτών σε μία ανενόχλητη ιδιωτική ζωή». Ορισμένοι ερευνητές έχουν προσπαθήσει να εισάγουν έναν εναλλακτικό ορισμό της ιδιωτικότητας, εκφράζοντάς την με όρους: ιδιοκτησίας, αυτονομίας και απομόνωσης. Η ιδιωτικότητα μπορεί να θεωρηθεί ως ιδιοκτησία, στο βαθμό που ένα άτομο είναι δυνατό να χάσει μέρος του ελέγχου που ασκεί στις προσωπικές του πληροφορίες, με αντάλλαγμα κάποιο όφελος. Επιπλέον, μπορεί να θεωρηθεί ως αυτονομία, υπό την έννοια ότι κάθε άτομο είναι ελεύθερο να εξουσιοδοτήσει, μερικώς ή πλήρως, μία τρίτη οντότητα να αποκτά, να επεξεργάζεται, να κατανέμει, να μοιράζεται και να χρησιμοποιεί τις προσωπικές πληροφορίες για κάποιο συγκεκριμένο σκοπό. Ακόμα, η ιδιωτικότητα μπορεί να γίνει κατανοητή ως απομόνωση, υπό την έννοια ότι όλοι έχουν το δικαίωμα να αποζητούν να μην ενοχλούνται από τρίτους.

Η αναγκαιότητα της προστασίας της ιδιωτικότητας προβάλλεται εντονότερα, όταν γίνεται αντιληπτή η δυνατότητα της συλλογής και επεξεργασίας πληροφοριών από τα πληροφοριακά συστήματα, η οποία καθιστά δυνατή την πολυλειτουργική χρήση και την «αποξένωση» της πληροφορίας από το φορέα της, το αρχικό περιβάλλον και τους αρχικούς σκοπούς της συλλογής και της επεξεργασίας της. Η σύγκλιση των τεχνολογιών πληροφορικής και επικοινωνιών, η αποκέντρωση της επεξεργασίας, η διείσδυση της επεξεργασίας και της δικτύωσης στο σύνολο σχεδόν την ανθρώπινης δραστηριότητας αλλάζουν ριζικά το περιβάλλον χρήσης της προσωπικής πληροφορίας αλλά και τα ζητήματα που εγείρονται σε σχέση με την προστασία της. Στο πλαίσιο αυτό διαμορφώνεται η ανάγκη για προστασία των προσωπικών δεδομένων. Η έννοια των προσωπικών δεδομένων, όπως ορίζεται και από την Ομάδα Προστασίας Δεδομένων του Άρθρου 29 νοείται ως «κάθε πληροφορία που αναφέρεται σε φυσικό πρόσωπο του οποίου η ταυτότητα είναι γνωστή ή μπορεί να εξακριβωθεί». Τα προσωπικά δεδομένα είναι πληροφορίες που αφορούν ένα ταυτοποιημένο ή ταυτοποιήσιμο εν ζωή άτομο. Διαφορετικές πληροφορίες οι οποίες, εάν συγκεντρωθούν όλες μαζί, μπορούν να οδηγήσουν στην ταυτοποίηση ενός συγκεκριμένου ατόμου, αποτελούν επίσης δεδομένα προσωπικού χαρακτήρα. [4]

Τα προσωπικά δεδομένα που έχουν καταστεί ανώνυμα, έχουν κρυπτογραφηθεί ή για τα οποία έχουν χρησιμοποιηθεί ψευδώνυμα αλλά τα οποία μπορούν να χρησιμοποιηθούν για την επαναταυτοποίηση ενός ατόμου παραμένουν δεδομένα προσωπικού χαρακτήρα και εμπίπτουν στο πεδίο εφαρμογής του Γενικού Κανονισμού Προστασίας Δεδομένων. Ωστόσο, τα προσωπικά δεδομένα που έχουν καταστεί ανώνυμα με τέτοιον τρόπο ώστε το άτομο να μην είναι ή να μην είναι πια ταυτοποιήσιμο δεν θεωρούνται πλέον δεδομένα προσωπικού χαρακτήρα. Για να είναι πραγματικά ανώνυμα τα δεδομένα, η ανωνυμοποίηση πρέπει να είναι μη αντιστρέψιμη.

Ορισμένα παραδείγματα προσωπικών δεδομένων είναι τα εξής [5]:

- όνομα και επώνυμο·
- διεύθυνση κατοικίας·
- ηλεκτρονική διεύθυνση, π.χ. όνομα.επώνυμο@εταιρεία.com·
- αναγνωριστικός αριθμός κάρτας·
- δεδομένα τοποθεσίας (π.χ. η λειτουργία δεδομένων τοποθεσίας σε κινητό τηλέφωνο)·
- διεύθυνση διαδικτυακού πρωτοκόλλου (IP)·
- το αναγνωριστικό διαφήμισης του τηλεφώνου·
- δεδομένα που φυλάσσονται από νοσοκομείο ή γιατρό, που θα μπορούσαν να είναι ένα σύμβολο που προσδιορίζει αποκλειστικά ένα άτομο.

Ενώ, παραδείγματα δεδομένων που δε θεωρούνται προσωπικά δεδομένα είναι τα εξής:

- αριθμός μητρώου εταιρείας·
- ηλεκτρονική διεύθυνση του τύπου πληροφορίες@εταιρεία.com·
- ανώνυμα δεδομένα.

3.2. Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR)

Αξίζει να σημειωθεί πως η ανάγκη προστασίας των προσωπικών δεδομένων και η εφαρμογή ενός κανονιστικού πλαισίου που τη διασφαλίζει δεν αποτελεί πρωτοφανές φαινόμενο. Ήδη από το 1995 υπάρχει Ευρωπαϊκή υποχρέωση των κρατών-μελών να διασφαλίζουν την προστασία των φυσικών προσώπων έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα (δηλαδή το σεβασμό στην ιδιωτικότητα) και αφετέρου την εξασφαλίζουν την ελεύθερη κυκλοφορία των δεδομένων αυτών, ως μέσο επίτευξης οικονομικής και κοινωνικής προόδου. Ωστόσο, με την πάροδο των χρόνων και τις ραγδαίες τεχνολογικές εξελίξεις που έλαβαν χώρα (διαδίκτυο, κινητή τηλεφωνία, Big Data κ.ά.), η Οδηγία θεωρήθηκε παρωχημένη και η ανάγκη για ένα σύγχρονο κανονιστικό ενιαίο πλαίσιο έγινε εντονότερη από ποτέ. Η ανάγκη αυτή, λοιπόν, γέννησε το Γενικό Κανονισμό για την Προστασία Δεδομένων (GDPR), ο οποίο έδωσε έμφαση σε έννοιες όπως η νόμιμη, αντικειμενική και διαφανής επεξεργασία, ο περιορισμός του σκοπού και η ακρίβεια της επεξεργασίας.

Στην παράγραφο 1 του άρθρου 5 του ΓΚΠΔ περιλαμβάνονται οι έξι αρχές που διέπουν την επεξεργασία δεδομένων προσωπικού χαρακτήρα, οι οποίες αποτελούν τον πυρήνα τόσο του ΓΚΠΔ όσο και της προϊσχύουσας Οδηγίας.

3.2.1. Αρχή νομιμότητας, αντικειμενικότητας και διαφάνειας

Το Άρθρο 5 αναφέρει χαρακτηριστικά ότι "Τα δεδομένα προσωπικού χαρακτήρα υποβάλλονται σε σύννομη και θεμιτή επεξεργασία με διαφανή τρόπο σε σχέση με το υποκείμενο των δεδομένων". Η επεξεργασία των δεδομένων θεωρείται σύννομη εάν και εφόσον ισχύει τουλάχιστον μία από τις προϋποθέσεις που περιγράφονται στο Άρθρο 6 του GDPR (παρ. (α) - (στ)). Οι προϋποθέσεις αυτές περιλαμβάνουν τη συναίνεση του υποκειμένου στην επεξεργασία, την απαίτηση της επεξεργασίας για την εκτέλεση σύμβασης, την απαίτηση της επεξεργασίας για συμμόρφωση με έννομη υποχρέωση, την απαίτηση της επεξεργασίας για τη διαφύλαξη ζωτικού συμφέροντος, την απαίτηση της επεξεργασίας για την εκπλήρωση καθήκοντος που εκτελείται προς το δημόσιο συμφέρον καθώς και την απαίτηση της επεξεργασίας για σκοπούς των έννομων συμφερόντων. Ειδική μνεία γίνεται στον GDPR για την επεξεργασία των ειδικών κατηγοριών δεδομένων προσωπικού χαρακτήρα στο Άρθρο 9. Η παράγραφος 2 του ίδιου Άρθρου περιγράφει τις

περιπτώσεις στις οποίες η επεξεργασία των ειδικών κατηγοριών δεδομένων προσωπικού χαρακτήρα μπορεί να θεωρηθεί σύννομη. Καθίσταται λοιπόν σαφές, ότι η βάση της επεξεργασίας των Big Data οφείλει να στηρίζεται σε μία από τις αρχές που αναφέρονται στο Άρθρο 6 του Κανονισμού (παρ. 1, στ. α) - στ)). Σε αντίθετη περίπτωση, η επεξεργασία θεωρείται παράνομη και αντιτίθεται στον Κανονισμό.

Οι υπεύθυνοι επεξεργασίας, οφείλουν, μεταξύ άλλων, να λαμβάνουν υπόψη τις επιπτώσεις στα δικαιώματα των υποκειμένων των δεδομένων κατά τον προσδιορισμό της κατάλληλης νομικής βάσης, προκειμένου να τηρείται η αρχή της αντικειμενικότητας. Η εν λόγω αρχή περιλαμβάνει την αναγνώριση των θεμιτών προσδοκιών των υποκειμένων των δεδομένων όσον αφορά πιθανές αρνητικές συνέπειες που ενδέχεται να επιφέρει η επεξεργασία σε αυτούς.

Ο GDPR προσθέτει, επίσης την απαίτηση της διαφάνειας σύμφωνα με την οποία η επεξεργασία των προσωπικών δεδομένων θα πρέπει να γίνεται με διαφανή τρόπο σε σχέση με το υποκείμενο των δεδομένων. Το υποκείμενο των δεδομένων οφείλει να έχει πλήρη γνώση για το λόγο για τον οποίο συλλέγονται τα δεδομένα του και πώς αυτά θα χρησιμοποιηθούν, ούτως ώστε να είναι σε θέση να ασκήσει τα νόμιμα δικαιώματά του, όπως αυτά περιγράφονται στα άρθρα 13-15 του Κανονισμού. Οι υπεύθυνοι επεξεργασίας οφείλουν να είναι σε θέση να παράσχουν οποιαδήποτε πληροφορία σχετική με την επεξεργασία των δεδομένων ζητηθεί από το υποκείμενο. Για παράδειγμα, σε περίπτωση που το υποκείμενο ζητήσει να δει ποια δεδομένα του συλλέγονται από τον οργανισμό ή την εταιρεία, ο υπεύθυνος επεξεργασίας θα πρέπει να τα παράσχει στο υποκείμενο. Πρακτικά, στον χώρο του διαδικτύου, το υποκείμενο λαμβάνει την ενημέρωση σχετικά με τα δεδομένα που συλλέγονται και τον τρόπου που αυτά επεξεργάζονται μέσω της πολιτικής απορρήτου που εμφανίζεται στο παράθυρο των χρηστών κατά την επίσκεψη τους σε μια ιστοσελίδα ή κατά την είσοδο τους για πρώτη φορά σε μια εφαρμογή ή μια πλατφόρμα. Η πολιτική αυτή θα πρέπει να είναι σαφής και απαλλαγμένη από εξειδικευμένους όρους ώστε να μην παρουσιάζονται δυσκολίες στην κατανόησή της.

3.2.2. Αρχή περιορισμού του σκοπού

Το Άρθρο 5 παρ. 1 στ. β περιγράφει την αρχή του περιορισμού του σκοπού ορίζοντας ότι «τα δεδομένα προσωπικού χαρακτήρα συλλέγονται για καθορισμένους, ρητούς και νόμιμους σκοπούς και δεν υποβάλλονται σε περαιτέρω επεξεργασία κατά τρόπο ασύμβατο προς τους σκοπούς αυτούς· η περαιτέρω επεξεργασία για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον ή σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς δεν θεωρείται ασύμβατη με τους αρχικούς σκοπούς σύμφωνα με το άρθρο 89 παράγραφος 1». Με την παρούσα αρχή, ουσιαστικά η συλλογή και η επεξεργασία των δεδομένων περιορίζεται με βάση έναν προκαθορισμένο σκοπό ο οποίος έχει τεθεί εξ αρχής από τον υπεύθυνο επεξεργασίας και έχει γνωστοποιηθεί στο υποκείμενο των δεδομένων. Με τον τρόπο αυτό το υποκείμενο είναι σε θέση να συγκατατεθεί με την επεξεργασία των δεδομένων του ή αντίστοιχα να αρνηθεί, ασκώντας τα νόμιμα δικαιώματά του.

Γίνεται κατανοητό ότι κάθε νέος σκοπός επεξεργασίας δεδομένων ο οποίος είναι ασύμβατος με τον αρχικό πρέπει να έχει τη δική του νομική βάση και δεν μπορεί να δικαιολογείται από το γεγονός ότι τα δεδομένα αποκτήθηκαν αρχικά ή υποβλήθηκαν σε επεξεργασία για άλλο νόμιμο σκοπό. Από την άλλη πλευρά, η νόμιμη επεξεργασία περιορίζεται στον αρχικά καθορισμένο σκοπό, και για οποιονδήποτε νέο σκοπό απαιτείται νέα, χωριστή νομική βάση. Για παράδειγμα, η κοινοποίηση δεδομένων προσωπικού χαρακτήρα σε τρίτους για νέο σκοπό θα πρέπει να εξετάζεται προσεκτικά, καθώς μια

τέτοια κοινοποίηση θα χρειάζεται πιθανώς πρόσθετη νομική βάση, διαφορετική από εκείνη στην οποία βασίστηκε η συλλογή των δεδομένων. Σύμφωνα με τον GDPR και την Εκσυγχρονισμένη Σύμβαση 108, η «περαιτέρω επεξεργασία για λόγους αρχειοθέτησης που άπτονται του δημόσιου συμφέροντος, για σκοπούς επιστημονικής ή ιστορικής έρευνας ή για στατιστικούς σκοπούς» θεωρείται εκ των προτέρων συμβατή με τον αρχικό σκοπό. Ωστόσο, κατά την περαιτέρω επεξεργασία δεδομένων προσωπικού χαρακτήρα πρέπει να υπάρχουν κατάλληλες εγγυήσεις, όπως ανωνυμοποίηση, κρυπτογράφηση ή ψευδωνυμοποίηση των δεδομένων.

Η ανάλυση των Big Data προσφέρει αδιαμφισβήτητα μεγάλες ευκαιρίες. Ωστόσο, η χρήση και η διατήρηση μεγάλου όγκου δεδομένων χωρίς προκαθορισμένο σκοπό, αν και είναι κύριο χαρακτηριστικό των Big Data, αποτελεί παρέκκλιση από την αρχή του περιορισμού του σκοπού, όπως αυτή περιγράφεται στον ΓΚΠΔ.

3.2.3. Αρχή ελαχιστοποίησης των δεδομένων

Στο Άρθρο 5, παρ. 1 στ. γ περιγράφεται ρητά ότι «Τα δεδομένα προσωπικού χαρακτήρα είναι κατάλληλα, συναφή και περιορίζονται στο αναγκαίο για τους σκοπούς για τους οποίους υποβάλλονται σε επεξεργασία». Σύμφωνα με την παρούσα αρχή τα προσωπικά δεδομένα συλλέγονται και αποθηκεύονται μόνο στο βαθμό στον οποίο αυτό απαιτείται και οφείλουν να διαγράφονται όταν δεν εξυπηρετείται πλέον ο σκοπός για τον οποίο συλλέχθηκαν. Μια τέτοια αρχή μειώνει σε μεγάλο βαθμό την πιθανότητα παραβίασης της ιδιωτικότητας των υποκειμένων των δεδομένων από τον ίδιο τον υπεύθυνο επεξεργασίας καθώς ελαχιστοποιείται ο κίνδυνος να χρησιμοποιηθούν τα δεδομένα για σκοπό πέρα από αυτό με τον οποίο έχει συναινέσει το υποκείμενο. Ωστόσο, η αρχή αυτή έρχεται σε αντίθεση με δύο από τα βασικά χαρακτηριστικά των Big Data και συγκεκριμένα με τον όγκο και την ποικιλομορφία τους.

Ο όγκος αποτελεί ένα από τα κυριότερα χαρακτηριστικά των Big Data, όπως έχει αναφερθεί ήδη σε προηγούμενη ενότητα. Η βάση στην οποία στηρίζεται η αξία των Big Data δίνει μεγάλα κίνητρα σε εταιρείες και οργανισμούς να συλλέγουν, να αποθηκεύουν και να επεξεργάζονται όσο το δυνατό περισσότερα δεδομένα για όσο το δυνατό περισσότερο. Όσα περισσότερα δεδομένα συλλέγονται και επεξεργάζονται για μία οντότητα, τόση περισσότερη γνώση υπάρχει για λεπτομέρειες που αφορούν τη ζωή του και τόσες περισσότερες συσχετίσεις μπορούν δυνητικά να προκύψουν. Ο άφθονος όγκος των δεδομένων που συλλέγονται, ωστόσο, μπορεί να υπερβαίνει τα δεδομένα που είναι απαραίτητα για τους σκοπούς της επεξεργασίας και ως εκ τούτου να παραβιάζεται η αρχή της ελαχιστοποίησης, όπως αυτή ορίζεται από τον ΓΚΠΔ.

Ομοίως, άλλο ένα χαρακτηριστικό των Big Data είναι η ποικιλομορφία τους. Μέσω μεγάλης ποικιλίας δεδομένων, τα οποία συνδυάζονται από διαφορετικές πηγές, μπορούν να προκύψουν νέα συμπεράσματα και να δημιουργηθούν νέες πληροφορίες σχετικά με τις οντότητες που εντάσσονται στο πεδίο μελέτης. Ωστόσο, η μεγάλη ποικιλία πηγών εγείρει επίσης ερωτηματικά σχετικά με το αν τα δεδομένα που συλλέγονται είναι πράγματι συμβατά με την επίτευξη συγκεκριμένου σκοπού ή εάν πρόκειται για παραβίαση της αρχής της ελαχιστοποίησης των δεδομένων.

3.2.4. Αρχή ακρίβειας των δεδομένων

Η παρούσα αρχή αποτυπώνεται στο Άρθρο 5, παρ. 1 στ. δ του ΓΚΠΔ, όπου αναφέρεται ότι: «Τα δεδομένα προσωπικού χαρακτήρα είναι ακριβή και, όταν είναι αναγκαίο, επικαιροποιούνται· πρέπει να λαμβάνονται όλα τα εύλογα μέτρα για την άμεση διαγραφή ή διόρθωση δεδομένων προσωπικού χαρακτήρα τα οποία είναι ανακριβή, σε σχέση με τους

σκοπούς της επεξεργασίας («ακρίβεια»)». Ο υπεύθυνος επεξεργασίας οφείλει, επομένως, να λαμβάνει κατάλληλα μέτρα ώστε να διασφαλίζει ότι τα δεδομένα που χρησιμοποιούνται προς επεξεργασία είναι ακριβή και επικαιροποιημένα. Σε ορισμένες περιπτώσεις, ο έλεγχος της ακρίβειας των δεδομένων, αποτελεί μια επαναλαμβανόμενη διαδικασία, με στόχο την αποφυγή ζημίας με επιπτώσεις στο ίδιο το υποκείμενο των δεδομένων, όπως για παράδειγμα σε περιπτώσεις ελέγχου της πιστοληπτικής ικανότητας ενός πελάτη τράπεζας. Το ίδιο ισχύει και για τα δεδομένα προσωπικού χαρακτήρα που χρησιμοποιούνται ως δεδομένα εισόδου σε συστήματα μηχανικής μάθησης, προκειμένου να συμβάλλουν στη λήψη συμπερασμάτων ή αποφάσεων σχετικά με τα υποκείμενα των δεδομένων.

Κατά την ανάλυση Big Data αλλά και κατά τη χρήση Μηχανικής Μάθησης, ο όγκος των δεδομένων που συλλέγονται είναι τεράστιος, με αποτέλεσμα να περιλαμβάνει τόσο ακριβή όσο και ανακριβή δεδομένα καθώς δεν υπάρχει μηχανισμός που να μπορεί να τα διαχωρίσει. Εκτός όμως από το μεγάλο όγκο των δεδομένων, σημαντικό ρόλο στο μη διαχωρισμό ακριβών και μη δεδομένων, παίζει και η πληθώρα διαφορετικών πηγών από τις οποίες αυτά συλλέγονται με αποτέλεσμα να περιέχουν σφάλματα.

Όσον αφορά τα συστήματα μηχανικής μάθησης, είναι σημαντικό να υπάρχει σαφής διάκριση σχετικά με το αν τα προσωπικά δεδομένα χρησιμοποιούνται μόνο ως δεδομένα εισόδου σε ένα σύνολο εκπαίδευσης, για την εκμάθηση συσχετίσεων ή ως δεδομένα εισόδου σε έναν αλγόριθμο κατάρτισης εξατομικευμένου προφίλ. Προφανώς, από τη στιγμή που τα δεδομένα είναι διαθέσιμα για το σύνολο της εκπαίδευσης, ο πειρασμός να χρησιμοποιηθούν τα ίδια δεδομένα για την εξαγωγή εξατομικευμένων συμπερασμάτων είναι ισχυρός. Η ανωνυμοποίηση και η ψευδωνυμοποίηση, θα μπορούσαν να λειτουργήσουν ως μέτρα ασφάλειας και να συμβάλουν στη σημαντική μείωση του προαναφερθέντος κινδύνου.

3.2.5. Αρχή περιορισμού της περιόδου αποθήκευσης

Στο Άρθρο 5 παρ. 1 στ. ε του ΓΚΠΔ διατυπώνεται η παρούσα αρχή σύμφωνα με την οποία "τα δεδομένα προσωπικού χαρακτήρα διατηρούνται υπό μορφή που επιτρέπει την ταυτοποίηση των υποκειμένων των δεδομένων μόνο για το διάστημα που απαιτείται για τους σκοπούς της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα· τα δεδομένα προσωπικού χαρακτήρα μπορούν να αποθηκεύονται για μεγαλύτερα διαστήματα, εφόσον τα δεδομένα προσωπικού χαρακτήρα θα υποβάλλονται σε επεξεργασία μόνο για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον, για σκοπούς επιστημονικής ή ιστορικής έρευνας ή για στατιστικούς σκοπούς, σύμφωνα με το άρθρο 89 παράγραφος 1 και εφόσον εφαρμόζονται τα κατάλληλα τεχνικά και οργανωτικά μέτρα που απαιτεί ο παρών κανονισμός για τη διασφάλιση των δικαιωμάτων και ελευθεριών του υποκειμένου των δεδομένων". Επομένως, καθίσταται σαφές ότι τα δεδομένα που συλλέγονται και τα οποία διατηρούνται σε μορφή που επιτρέπει την ταυτοποίηση των υποκειμένων τους, θα πρέπει να διαγράφονται όταν δεν εκπληρώνεται πλέον ο σκοπός για τον οποίο έχει πραγματοποιηθεί η συλλογή τους.

Ο περιορισμός της περιόδου αποθήκευσης των δεδομένων και, ουσιαστικά, η διαγραφή τους όταν αυτά δε χρειάζονται πλέον, μειώνει τον κίνδυνο αυτά να καταστούν άσχετα, ανακριβή ή παρωχημένα. Παράλληλα, μειώνεται σημαντικά και ο κίνδυνος κατάχρησης των δεδομένων σε βάρος των υποκειμένων τους. Ωστόσο, η ανάλυση των Big Data αποτελεί ένα καλό παράδειγμα κατά το οποίο τα δεδομένα που συλλέγονται χρησιμοποιούνται και επεξεργάζονται για μεγάλο χρονικό διάστημα προκειμένου να προκύψουν ασφαλέστερα συμπεράσματα. Στην περίπτωση αυτή λοιπόν, η παρούσα αρχή

μπορεί να υπονομεύσει την ικανότητα πρόβλεψης, η οποία είναι μία από τις ευκαιρίες που καθίστανται δυνατές από την ανάλυση των Big Data. Πράγματι, αν η ανάλυση των Big Data επιτρέπει την πρόβλεψη, είναι επειδή οι αλγόριθμοι μπορούν να συγκρίνουν τα τρέχοντα δεδομένα με αποθηκευμένα δεδομένα του παρελθόντος προκειμένου να καθορίσουν τι πρόκειται να συμβεί στο μέλλον.

3.2.6. Αρχή ακεραιότητας και εμπιστευτικότητας

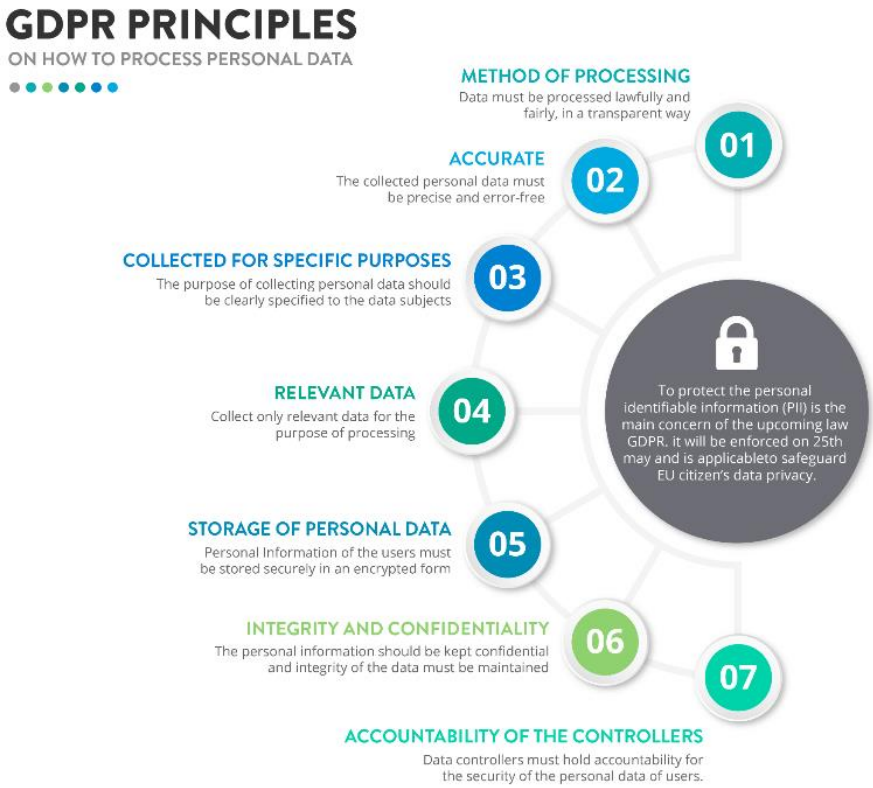
Στο Άρθρο 5 παρ. 1 στ. στ αποτυπώνεται η παρούσα αρχή σύμφωνα με την οποία "τα δεδομένα προσωπικού χαρακτήρα υποβάλλονται σε επεξεργασία κατά τρόπο που εγγυάται την ενδεδειγμένη ασφάλειά τους , μεταξύ άλλων την προστασία τους από μη εξουσιοδοτημένη ή παράνομη επεξεργασία και τυχαία απώλεια, καταστροφή ή φθορά, με τη χρησιμοποίηση κατάλληλων τεχνικών ή οργανωτικών μέτρων". Η αρχή αυτή ενισχύει την ιδέα ότι η ασφάλεια των δεδομένων αποτελεί θεμελιώδη υποχρέωση των υπεύθυνων επεξεργασίας και τονίζει ότι τα δεδομένα πρέπει να προστατεύονται από μη εξουσιοδοτημένη και παράνομη επεξεργασία, συμπεριλαμβανομένης της καταστροφής ή απώλειας κατά τη διάρκεια της επεξεργασίας των δεδομένων. Προκειμένου αυτό να επιτευχθεί είναι σημαντικό να διαμορφωθούν και να τεθούν σε ισχύει τα κατάλληλα οργανωτικά και τεχνικά μέτρα ασφάλειας.

Παραδείγματα τέτοιων μέτρων ασφάλειας που επιλέγονται μπορούν να περιλαμβάνουν την ψευδωνυμοποίηση ή την κρυπτογράφηση των δεδομένων. Η ενσωμάτωση τέτοιων τεχνικών μέτρων στη διαδικασία επεξεργασίας διασφαλίζει την εμπιστευτικότητα και την ακεραιότητα των δεδομένων και λειτουργεί ως απόδειξη συμμόρφωσης με την παρούσα αρχή.

3.2.7. Αρχή λογοδοσίας

Η αρχή της λογοδοσίας παρουσιάζεται στο Άρθρο 5 παρ. 2 και ορίζει ότι "Ο υπεύθυνος επεξεργασίας φέρει την ευθύνη και είναι σε θέση να αποδείξει τη συμμόρφωση με την παράγραφο 1". Η παρούσα αρχή, λοιπόν, απαιτεί τη θέσπιση και την εφαρμογή μηχανισμών και συστημάτων ελέγχου που διασφαλίζουν και αποδεικνύουν τη συμμόρφωση σε τρίτους, συμπεριλαμβανομένων των εποπτικών αρχών. Η λογοδοσία, επιπλέον, απαιτεί την τακτική επαλήθευση ότι όλοι οι εσωτερικοί έλεγχοι και συστήματα συνεχίζουν να είναι κατάλληλα και κάθε επεξεργασία που υφίστανται τα δεδομένα συνεχίζει να είναι σύννομη με την παρούσα αρχή. Ωστόσο, στο πεδίο των Big Data η πλήρωση της αρχής της λογοδοσίας δε φαίνεται να είναι εύκολο έργο δεδομένης της χρήσης αλγορίθμων, οι οποίοι χρησιμοποιούν τεράστιο όγκο δεδομένων προκειμένου να καταλήξουν σε συμπεράσματα και έχουν τη δυνατότητα να λαμβάνουν αποφάσεις οι οποίες μπορεί να είναι λανθασμένες ή αδικαιολόγητες.

Υπάρχουν πολλά στοιχεία, όμως, τα οποία μπορούν να αποτελέσουν καλές πρακτικές για τη συμμόρφωση με την αρχή της λογοδοσίας. Η προστασία των δεδομένων ήδη από το σχεδιασμό και εξ' ορισμού (privacy by design, privacy by default), οι εκτιμήσεις αντικτύπου, ο συστηματικός έλεγχος, η κατάλληλη εμπειρία σε θέματα προστασίας δεδομένων, συμπεριλαμβανομένης της διαθεσιμότητας ενός υπεύθυνου προστασίας δεδομένων, μπορούν να συμβάλουν και να αποτελέσουν αναπόσπαστο μέρος ενός συστήματος ελέγχου και διασφάλισης της ορθής χρήσης των Big Data.



Εικόνα 4: Βασικές αρχές ΓΚΠΔ [6]

4. Κατάρτιση προφίλ και αυτοματοποιημένη λήψη αποφάσεων

Η πρόοδος στην τεχνολογία, με τις δυνατότητες που παρέχει η ανάλυση των Big Data, η τεχνητή νοημοσύνη και η μηχανική μάθηση, έχει καταστήσει ευκολότερη τη δημιουργία προφίλ και τη λήψη αυτοματοποιημένων αποφάσεων.

Η ευρύτατη διαθεσιμότητα προσωπικών δεδομένων στο διαδίκτυο, αλλά και μέσω συσκευών διασυνδεδεμένων συσκευών στο Διαδίκτυο των πραγμάτων (IoT), σε συνδυασμό με την αυξημένη δυνατότητα εύρεσης συσχετισμών μεταξύ των δεδομένων, μπορεί να οδηγήσει στον καθορισμό, την ανάλυση και τελικώς την πρόβλεψη πτυχών της προσωπικότητας ενός ατόμου ή της συμπεριφοράς, των ενδιαφερόντων και των συνηθειών του.

Η κατάρτιση προφίλ και η αυτοματοποιημένη λήψη αποφάσεων μπορεί να είναι χρήσιμη για τα υποκείμενα των δεδομένων, τους οργανισμούς καθώς και για την οικονομία και την κοινωνία ως σύνολο, παρέχοντας οφέλη όπως η αυξημένη αποτελεσματικότητα και η εξοικονόμηση πόρων.

Οι διαδικασίες αυτές μπορεί να έχουν πολλές εμπορικές εφαρμογές, όπως για παράδειγμα να χρησιμοποιηθούν για την καλύτερη κατανομή των αγορών και την προσαρμογή των υπηρεσιών και των προϊόντων στις ατομικές ανάγκες των καταναλωτών. Οι τομείς της ιατρικής, της εκπαίδευσης, της υγειονομικής περίθαλψης και των μεταφορών μπορούν επίσης να επωφεληθούν από την κατάρτιση προφίλ και την αυτοματοποιημένη λήψη αποφάσεων.

Ωστόσο, οι διαδικασίες αυτές απαιτούν την ύπαρξη των κατάλληλων ασφαλιστικών δικλείδων, καθώς μπορούν να δημιουργήσουν σημαντικούς κινδύνους για τα δικαιώματα των ατόμων και τις ελευθερίες τους [7]. Ορισμένοι δυνητική κίνδυνοι που ελλοχεύουν είναι οι εξής:

- Η κατάρτιση προφίλ δεν είναι εύκολο να παρακολουθηθεί από τα άτομα τα οποία αφορά.
- Ενδέχεται τα άτομα να μην περιμένουν ότι τα προσωπικά τους δεδομένα και οι προσωπικές τους πληροφορίες θα χρησιμοποιηθούν γι’ αυτό το σκοπό.
- Τα άτομα μπορεί να μην κατανοούν πώς λειτουργεί η διαδικασία της κατάρτισης προφίλ και πώς μπορεί να τα επηρεάσει.
- Οι αποφάσεις που λαμβάνονται μπορεί να οδηγήσουν σε δυσμενείς επιπτώσεις για ορισμένα άτομα.

Το γεγονός ότι η ανάλυση των Big Data των ατόμων οδηγεί σε συσχετίσεις, δε σημαίνει ότι αυτές μπορούν να οδηγήσουν σε ασφαλή συμπεράσματα. Εφόσον οι τεχνικές που χρησιμοποιούνται, κάνουν υποθέσεις σχετικά με τα χαρακτηριστικά και τη συμπεριφορά κάποιου, υπάρχει πάντα το περιθώριο του σφάλματος και γι’ αυτό χρειάζεται ορθή εξισορρόπηση ώστε να περιοριστούν οι κίνδυνοι από τη χρήση των αποτελεσμάτων. Οι σχετικές διατάξεις του ΓΚΠΔ έχουν σχεδιαστεί ειδικά προκειμένου να αντιμετωπιστούν τέτοιου είδους κίνδυνοι.

4.1. Κατάρτιση Προφίλ

Στο Άρθρο 4 του ΓΚΠΔ παρατίθεται ο ορισμός της κατάρτισης προφίλ, σύμφωνα με τον οποίο ως κατάρτιση προφίλ νοείται «οποιαδήποτε μορφή αυτοματοποιημένης επεξεργασίας δεδομένων προσωπικού χαρακτήρα που συνίσταται στη χρήση δεδομένων

προσωπικού χαρακτήρα για την αξιολόγηση ορισμένων προσωπικών πτυχών ενός φυσικού προσώπου, ιδίως για την ανάλυση ή την πρόβλεψη πτυχών που αφορούν την απόδοση στην εργασία, την οικονομική κατάσταση, την υγεία, τις προσωπικές προτιμήσεις, τα ενδιαφέροντα, την αξιοπιστία, τη συμπεριφορά, τη θέση ή τις μετακινήσεις του εν λόγω φυσικού προσώπου». Η κατάρτιση προφίλ, λοιπόν, μπορεί να περιγραφεί ως η αυτοματοποιημένη διαδικασία κατά την οποία επεξεργάζονται προσωπικά δεδομένα με στόχο την αξιολόγηση προσωπικών πτυχών του ατόμου.

Σύμφωνα με τις «Κατευθυντήριες γραμμές για την αυτοματοποιημένη λήψη αποφάσεων και την κατάρτιση προφίλ για τους σκοπούς του κανονισμού 2016/679» της Ομάδας Εργασίας για την Προστασία Δεδομένων του Άρθρου 29, τρία είναι τα στοιχεία που απαρτίζουν την κατάρτιση προφίλ [8] :

- πρέπει να είναι αυτοματοποιημένη μορφή επεξεργασίας,
- πρέπει να αφορά δεδομένα προσωπικού χαρακτήρα, και
- στόχος της κατάρτισης προφίλ πρέπει να είναι η αξιολόγηση προσωπικών πτυχών ενός φυσικού προσώπου.

Η Ομάδα Εργασίας επισημαίνει ότι το Άρθρο 4 του ΓΚΠΔ (παρ. 4) αναφέρεται σε «οποιαδήποτε μορφή αυτοματοποιημένης επεξεργασίας» αντί σε «αποκλειστικά» αυτοματοποιημένη επεξεργασία. Πάνω σε αυτό, η Ομάδα Εργασίας στηρίζει την άποψη ότι η κατάρτιση προφίλ πρέπει να περιλαμβάνει κάποια μορφή αυτοματοποιημένης επεξεργασίας, παρότι η ανθρώπινη συμμετοχή δε συνεπάγεται απαραίτητα τον αποκλεισμό της δραστηριότητας από τον ορισμό.

Επίσης, η ίδια ομάδα επισημαίνει άλλο ένα σημείο του ΓΚΠΔ σχετικά με την κατάρτιση προφίλ, σύμφωνα με το οποίο: η κατάρτιση προφίλ είναι η αυτοματοποιημένη επεξεργασία δεδομένων προσωπικού χαρακτήρα για την αξιολόγηση προσωπικών πτυχών, ιδίως για την ανάλυση ή την πραγματοποίηση προβλέψεων σχετικά με φυσικά πρόσωπα. Σύμφωνα με την Ομάδα Εργασίας η χρήση της λέξης «αξιολόγηση» υποδηλώνει ότι η κατάρτιση προφίλ περιλαμβάνει κάποια μορφή αξιολόγησης ή κρίσης σχετικά με ένα πρόσωπο.

Επιπλέον, η Ομάδα Εργασίας, μέσω των κατευθυντήριων γραμμών επισημαίνει ότι η απλή ταξινόμηση προσώπων με βάση γνωστά χαρακτηριστικά όπως η ηλικία, το φύλο και το ύψος δε συνιστά κατ' ανάγκη κατάρτιση προφίλ. Αυτό θα εξαρτηθεί από τον σκοπό της ταξινόμησης. Για παράδειγμα, μια επιχείρηση μπορεί να επιθυμεί να ταξινομεί τους πελάτες της με βάση την ηλικία ή το φύλο τους για στατιστικούς λόγους και προκειμένου να διαμορφώσει μια συγκεντρωτική γενική εικόνα των πελατών της χωρίς να προβαίνει σε προβλέψεις ή να εξάγει συμπεράσματα σχετικά με ένα συγκεκριμένο άτομο. Σ' αυτή την περίπτωση, ο σκοπός δεν είναι η αξιολόγηση ατομικών χαρακτηριστικών και, ως εκ τούτου, δεν πρόκειται για κατάρτιση προφίλ.

4.1.1. Κατάρτιση προφίλ και Μηχανική Μάθηση

Η Μηχανική Μάθηση χρησιμοποιείται κατά κόρον στις σύγχρονες τεχνολογικές λύσεις που χρησιμοποιούνται για την κατάρτιση προφίλ. Για την κατάρτιση των προφίλ συλλέγεται πλήθος Big Data από διαφορετικές πηγές και πάνω σε αυτό υλοποιούνται τεχνικές και σχεδιάζονται αλγόριθμοι που έχουν ως στόχο να εξάγουν πληροφορία από έναν μεγάλο όγκο δεδομένων, χρήσιμη για όποιον πρόκειται να την αξιοποιήσει. Ως αλγόριθμος νοείται μία ακολουθία οδηγιών ή κανόνων ειδικά σχεδιασμένων για την

ολοκλήρωση ενός στόχου ή την επίλυση ενός προβλήματος. Η κατάρτιση προφίλ επιτυγχάνεται μέσω αλγορίθμων οι οποίοι εντοπίζουν συσχετίσεις ανάμεσα σε διαφορετικά σύνολα δεδομένων. Αυτοί οι αλγόριθμοι μπορούν στη συνέχεια να χρησιμοποιηθούν για τη λήψη ενός ευρέως φάσματος αποφάσεων, όπως για παράδειγμα για την πρόβλεψη της συμπεριφοράς ή για τον έλεγχο της πρόσβασης σε μια υπηρεσία. Τα συστήματα τεχνητής νοημοσύνης και η Μηχανική Μάθηση χρησιμοποιούνται όλο και περισσότερο για τη δημιουργία και την εφαρμογή τέτοιων αλγορίθμων [9].

Οι αλγόριθμοι μηχανικής μάθησης αποτελούν τον πυρήνα της κατάρτισης προφίλ καθώς διαδραματίζουν δύο βασικούς ρόλους. Πρώτον, έχουν τον έλεγχο της διαδικασίας της κατάρτισης προφίλ και έτσι καθορίζουν τον τρόπο με τον οποίο αυτή πραγματοποιείται και δεύτερον, οι αλγόριθμοι μέσω μαθηματικών διαδικασιών εντοπίζουν τάσεις και εντοπίζουν προτιμήσεις από σύνολα δεδομένων. Η ανάπτυξη και η χρήση όλο και πιο εξελιγμένων αλγορίθμων μπορεί να οδηγήσει στην αποτελεσματικότερη κατάρτιση προφίλ σε σχέση με εκείνη που μπορεί να παραχθεί αποκλειστικά από την ανθρώπινη γνώση.

4.1.2. Κατάρτιση προφίλ στο πλαίσιο του ΓΚΠΔ

Οι εφαρμογές της μηχανικής μάθησης έχουν αναπτυχθεί και συνεχίζουν να αναπτύσσονται με ιλιγγιώδεις ρυθμούς ενώ δεν προβλέπεται ότι οι ρυθμοί ανάπτυξης θα μειωθούν στα επόμενα χρόνια. Όπως έχει ήδη αναφερθεί, η επεξεργασία τεράστιου όγκου προσωπικών δεδομένων και Big Data, συνεπάγεται τόσο οφέλη όσο και κινδύνους. Ένας από τους βασικότερους κινδύνους που απορρέουν είναι ο κίνδυνος παραβίασης της ιδιωτικότητας των ατόμων, ο οποίος μπορεί να οδηγήσει σε άρνηση του ίδιου του ατόμου της επεξεργασίας των προσωπικών του δεδομένων. Προκειμένου, λοιπόν, να περιοριστούν τέτοιου είδους κίνδυνοι, ο ΓΚΠΔ σχεδιάστηκε με τέτοιο τρόπο ώστε να διασφαλίζεται η ιδιωτικότητα των ατόμων σε περιπτώσεις κατάρτισης προφίλ.

Ήδη από την πρώτη αιτιολογική σκέψη του ΓΚΠΔ καθίσταται σαφές ότι: «Η προστασία των φυσικών προσώπων έναντι της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα είναι θεμελιώδες δικαίωμα». Επιπλέον, σύμφωνα με την ίδια αιτιολογική σκέψη: «Το άρθρο 8 παράγραφος 1 του Χάρτη των Θεμελιωδών Δικαιωμάτων της Ευρωπαϊκής Ένωσης («Χάρτης») και το άρθρο 16 παράγραφος 1 της Συνθήκης για τη λειτουργία της Ευρωπαϊκής Ένωσης (ΣΛΕΕ) ορίζουν ότι κάθε πρόσωπο έχει δικαίωμα στην προστασία των δεδομένων προσωπικού χαρακτήρα που το αφορούν». Ιδιαίτερο ενδιαφέρον προκαλεί και η αιτιολογική σκέψη 4 του Κανονισμού σύμφωνα με την οποία: «Η επεξεργασία των δεδομένων προσωπικού χαρακτήρα θα πρέπει να προορίζεται να εξυπηρετεί τον άνθρωπο». Η ιδιωτικότητα των ατόμων αποτελεί θεμελιώδες δικαίωμα αλλά σύμφωνα με την αιτιολογική σκέψη 4 δε θεωρείται απόλυτο δικαίωμα και οφείλει να εκτιμάται και να σταθμίζεται με άλλα θεμελιώδη δικαιώματα. Για το λόγο αυτό, είναι σημαντικό να χρησιμοποιείται η αιτιολογική σκέψη 4 ως γνώμονας κατά την εφαρμογή των υπόλοιπων άρθρων του ΓΚΠΔ.

Όπως έχει ήδη περιγραφεί σε προηγούμενη ενότητα, το Άρθρο 5 του ΓΚΠΔ περιλαμβάνει τις βασικές αρχές της επεξεργασίας των προσωπικών δεδομένων. Το Άρθρο αυτό αφορά άμεσα την κατάρτιση προφίλ καθώς σύμφωνα με την αιτιολογική σκέψη 72: «Η κατάρτιση προφίλ υπόκειται στους κανόνες του Κανονισμού που διέπουν την επεξεργασία δεδομένων προσωπικού χαρακτήρα, όπως νομικοί λόγοι επεξεργασίας ή αρχές προστασίας δεδομένων».

4.1.3. Δικαιώματα υποκειμένων των δεδομένων

Ένας από τους κύριους στόχους του ΓΚΠΔ είναι να ενισχύσει τα υποκείμενα των δεδομένων, προσφέροντάς τους το δικαίωμα του ελέγχου των προσωπικών τους δεδομένων. Σύμφωνα με τον Κανονισμό, τα υποκείμενα των δεδομένων, τα οποία υφίστανται επεξεργασία, έχουν συγκεκριμένα δικαιώματα.

Στο Άρθρο 12 γίνεται αναφορά στο δικαίωμα που έχουν τα υποκείμενα των δεδομένων στη λήψη ακριβούς πληροφόρησης και επικοινωνίας ανάμεσα σε αυτά και τον υπεύθυνο επεξεργασίας. Επιπλέον, στο ίδιο Άρθρο επισημαίνεται ότι κάθε πληροφορία που παρέχεται στο υποκείμενο και αναφέρεται στα Άρθρα 13 και 14 και κάθε ανακοίνωση στο πλαίσιο των Άρθρων 15 έως 22 και του Άρθρου 34 σχετικά με την επεξεργασία οφείλει να είναι σε συνοπτική, διαφανή, κατανοητή και εύκολα προσβάσιμη μορφή, χρησιμοποιώντας σαφή και απλή διατύπωση. Ενώ, λοιπόν, το Άρθρο 12 προσδιορίζει τον τρόπο με τον οποίο θα πρέπει να γίνεται η πληροφόρηση και η ενημέρωση του υποκειμένου των δεδομένων σχετικά με τα προσωπικά του δεδομένα, τα επόμενα δύο Άρθρα (13 και 14) καθορίζουν τις πληροφορίες που θα πρέπει να παρέχονται σχετικά με τη συλλογή των προσωπικών δεδομένων. Η διαφορά των δύο αυτών Άρθρων έγκειται στο γεγονός ότι το Άρθρο 13 αφορά περιπτώσεις στις οποίες τα προσωπικά δεδομένα συλλέγονται από το υποκείμενο των δεδομένων ενώ το Άρθρο 14 αφορά περιπτώσεις στις οποίες τα προσωπικά δεδομένα συλλέγονται από οποιονδήποτε άλλον πλην του υποκειμένου των δεδομένων. Και στις δύο περιπτώσεις, ο υπεύθυνος επεξεργασίας, κατά τη λήψη των προσωπικών δεδομένων, οφείλει να παρέχει, μεταξύ άλλων, πληροφορίες στο υποκείμενο σχετικά με την ταυτότητα και τα στοιχεία επικοινωνίας του υπευθύνου επεξεργασίας, τους σκοπούς της επεξεργασίας καθώς και τη νομική βάση για την επεξεργασία, τους αποδέκτες των προσωπικών δεδομένων, το χρονικό διάστημα για το οποίο θα αποθηκευτούν τα προσωπικά δεδομένα και την ύπαρξη δικαιώματος υποβολής αιτήματος στον υπεύθυνο επεξεργασίας για πρόσβαση και διόρθωση ή διαγραφή των προσωπικών δεδομένων ή περιορισμό της επεξεργασίας που αφορούν το υποκείμενο των δεδομένων. Σύμφωνα με το Άρθρο 13, συγκεκριμένα σε περιπτώσεις κατάρτισης προφίλ, είναι σημαντικό ο υπεύθυνος επεξεργασίας να είναι σε θέση να παρέχει στο υποκείμενο των δεδομένων σημαντικές πληροφορίες σχετικά με τη λογική που ακολουθείται, καθώς και τη σημασία και τις προβλεπόμενες συνέπειες της εν λόγω επεξεργασίας για το υποκείμενο των δεδομένων. Ιδιαίτερη έμφαση δίνεται μέσω της αιτιολογικής σκέψης 60 και στην υποχρέωση του υπεύθυνου επεξεργασίας για δίκαιη και διαφανή επεξεργασία των δεδομένων κατά τη διαδικασία κατάρτισης προφίλ καθώς το υποκείμενο των δεδομένων θα πρέπει να ενημερώνεται αν καταρτίζεται το προφίλ του και ποιες συνέπειες έχει αυτό. Με τον τρόπο αυτό διασφαλίζεται η αρχή της νομιμότητας, αντικειμενικότητας και διαφάνειας, όπως αυτή ορίζεται και περιγράφεται στο Άρθρο 5.

Καθίσταται σαφές ότι μέσω της κατάρτισης προφίλ, γίνονται προβλέψεις οι οποίες αφορούν μεμονωμένες οντότητες. Τα αποτελέσματα, όμως, μπορεί να μην είναι απόλυτα ακριβή λόγω της ανακρίβειας των δεδομένων που συλλέγονται, των μη σχετικών στοιχείων που συλλέγονται και αλλοιώνουν το τελικό αποτέλεσμα ή ακόμα και λόγω των μη ορθών συσχετίσεων που προκύπτουν μέσω των αλγορίθμων. Η μη σωστή, λοιπόν, κατάρτιση προφίλ μπορεί να οδηγήσει τόσο σε λάθος συμπεράσματα και προβλέψεις για τα άτομα, επηρεάζοντας τόσο στρατηγικές και επιχειρησιακές αποφάσεις, όσο και τα ίδια τα άτομα. Για το λόγο αυτό, γίνεται επίκληση στο δικαίωμα του υποκειμένου στη διόρθωση και διαγραφή των προσωπικών δεδομένων, όπως αυτά περιγράφονται αναλυτικά στα Άρθρα 16, 17 και 18 του ΓΚΠΔ. Με την άσκηση του δικαιώματος της διόρθωσης, το υποκείμενο

είναι σε θέση να διορθώσει τα ανακριβή αποτελέσματα που προκύπτουν γι' αυτό αντικαθιστώντας τις λανθασμένες πληροφορίες που υπάρχουν για το ίδιο με τις σωστές ή παρέχοντας επιπλέον πληροφορίες. Επιπλέον, τα δικαιώματα διόρθωσης και διαγραφής μπορούν να εφαρμοστούν τόσο για τα προσωπικά δεδομένα εισόδου (τα δεδομένα, δηλαδή, που χρησιμοποιούνται προκειμένου να πραγματοποιηθεί η κατάρτιση προφίλ) όσο και για τα δεδομένα εξόδου (στο ίδιο, δηλαδή, το προφίλ που προκύπτει). Αξίζει να σημειωθεί πως σε οποιοδήποτε στάδιο βρίσκεται η διαδικασία κατάρτισης προφίλ, το υποκείμενο των δεδομένων μπορεί να παρέμβει ασκώντας το δικαίωμα του περιορισμού της επεξεργασίας όπως αυτό προβλέπεται στο Άρθρο 18 του Κανονισμού.

Εξίσου σημαντικό δικαίωμα των υποκειμένων αποτελεί το δικαίωμα της εναντίωσης όπως αυτό περιγράφεται στο Άρθρο 21. Βάσει του Άρθρου, το υποκείμενο των δεδομένων μπορεί να αντιταχθεί στην επεξεργασία (περιλαμβανομένης της κατάρτισης προφίλ), για λόγους που συνδέονται με την ιδιαίτερη κατάστασή της.

4.2. Αυτοματοποιημένη λήψη αποφάσεων

Ο ΓΚΠΔ περιέχει επιπλέον υποχρεώσεις και περιορισμούς προκειμένου να αντιμετωπίσει τους κινδύνους που προκύπτουν από την κατάρτιση προφίλ ως αποτέλεσμα της αυτοματοποιημένης λήψης αποφάσεων, η οποία θίγει άμεσα ζητήματα που σχετίζονται με την ιδιωτικότητα. Με τον όρο αυτοματοποιημένη λήψη αποφάσεων εννοείται κατά βάση η μαζική συλλογή Big Data από διαφορετικές πηγές, η επεξεργασία των δεδομένων αυτών από σειρές αλγορίθμων, και η εξαγωγή ενός συμπεράσματος για το υποκείμενο των δεδομένων. Όπως υποδηλώνεται από τον ίδιο τον όρο, οι συζητήσεις που γίνονται για την αυτοματοποιημένη λήψη αποφάσεων αφορούν τις διαδικασίες λήψης αποφάσεων που περιλαμβάνουν την αυτοματοποίηση, δηλαδή τη χρήση τεχνολογιών για την εκτέλεση ενεργειών οι οποίες θα απαιτούσαν διαφορετικά την ανθρώπινη παρέμβαση. Οι αποφάσεις που προκύπτουν δεν έχουν μια σταθερή μορφή και μπορεί να παράγουν αποτελέσματα διαφορετικής συνάφειας αλλά σε όλες τις περιπτώσεις, μία ή περισσότερες ενέργειες που οδηγούν σε απόφαση, ανατίθεται σε κάποια μηχανή.

Η αυτοματοποιημένη λήψη αποφάσεων μπορεί να πραγματοποιηθεί με ή χωρίς κατάρτιση προφίλ. Ομοίως, η κατάρτιση προφίλ μπορεί να πραγματοποιηθεί χωρίς να ληφθεί κάποια αυτοματοποιημένη απόφαση. Αν και η αυτοματοποιημένη λήψη αποφάσεων και η κατάρτιση προφίλ μπορεί να είναι ξεχωριστές, η κατάρτιση προφίλ μπορεί να αποτελεί τη βάση για αποφάσεις που λαμβάνονται αυτοματοποιημένα, πράγμα που σημαίνει ότι η κατάρτιση προφίλ αποτελεί μέρος της αυτοματοποιημένης λήψης αποφάσεων. Ο συνδυασμός των πρόσφατων ανακαλύψεων στις τεχνικές Τεχνητής Νοημοσύνης, της σύγχρονης υπολογιστικής ισχύος και του μεγάλου όγκου δεδομένων (Big Data) που παράγονται έχουν ανοίξει διάφορες δυνατότητες για τη χρήση της Μηχανικής Μάθησης στις διαδικασίες λήψης αποφάσεων, εισάγοντας δυνητικά πλεονεκτήματα αλλά και κινδύνους για όσους επηρεάζονται από τις αποφάσεις.

Δεδομένου ότι πολλές αποφάσεις, οι οποίες λαμβάνονται αυτοματοποιημένα, έχουν ουσιαστικό αντίκτυπο στη σύγχρονη κοινωνία και βασίζονται στα προσωπικά δεδομένα των ατόμων, εμπίπτουν στο πεδίο εφαρμογής των νόμων περί προστασίας των δεδομένων που έχουν θεσπιστεί σε διάφορες χώρες σε όλο τον κόσμο.

4.2.1. Αυτοματοποιημένη λήψη αποφάσεων στο πλαίσιο του ΓΚΠΔ

Όπως έχει ήδη προαναφερθεί, η αυτοματοποιημένη λήψη αποφάσεων αποτελεί μέρος της σύγχρονης καθημερινότητας. Ο ΓΚΠΔ, λοιπόν, περιλαμβάνει συγκεκριμένες διατάξεις οι οποίες διέπουν την αυτοματοποιημένη λήψη των αποφάσεων. Συγκεκριμένα, οι διατάξεις

αυτές περιλαμβάνονται στο Άρθρο 22 του Κανονισμού, σύμφωνα με το οποίο: "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". Με τον τρόπο αυτό, θεσπίζεται η απαγόρευση της λήψης αποφάσεων, οι οποίες βασίζονται αποκλειστικά σε αυτοματοποιημένη επεξεργασία και μπορεί να επιφέρουν νομικές συνέπειες ή συνέπειες που επηρεάζουν σημαντικά το υποκείμενο των δεδομένων, για παράδειγμα επηρεάζοντας το δικαίωμά του σε κοινωνικές παροχές.

Σύμφωνα με τις ειδικές διατάξεις σχετικά με την αποκλειστικά αυτοματοποιημένη λήψη αποφάσεων που ορίζεται στο Άρθρο 22, ο όρος "δικαίωμα" δε σημαίνει ότι η παράγραφος 1 του ίδιου εφαρμόζεται μόνο όταν το επικαλείται ενεργά το υποκείμενο των δεδομένων. Αντίθετα, με τον τρόπο αυτό, θεσπίζεται μία γενική απαγόρευση σχετικά με τη λήψη των αποφάσεων όταν αυτή βασίζεται αποκλειστικά στην αυτοματοποιημένη επεξεργασία. Η εν λόγω απαγόρευση εφαρμόζεται ανεξάρτητα από το αν το υποκείμενο των δεδομένων προβεί σε ενέργειες όσον αφορά την επεξεργασία των προσωπικών του δεδομένων. Ωστόσο, η απαγόρευση του Άρθρου 22 (παράγραφος 1) εφαρμόζεται μόνο για συγκεκριμένες περιπτώσεις κατά τις οποίες μια απόφαση που βασίζεται αποκλειστικά σε αυτοματοποιημένη επεξεργασία, περιλαμβανομένης της κατάρτισης προφίλ, έχει έννομα αποτελέσματα ή επηρεάζει σημαντικά με παρόμοιο τρόπο ένα φυσικό πρόσωπο.

Αναλύοντας το παρόν Άρθρο διακρίνουμε την έννοια της απόφασης που λαμβάνεται αποκλειστικά βάσει αυτοματοποιημένης επεξεργασίας και την έννοια των έννομων ή σημαντικών αποτελεσμάτων, οι οποίες αναλύονται παρακάτω.

<u>Απόφαση που λαμβάνεται αποκλειστικά βάσει αυτοματοποιημένης επεξεργασίας</u>

Το Άρθρο 22 παράγραφος 1 αναφέρεται σε αποφάσεις «λαμβανόμενες αποκλειστικά βάσει» αυτοματοποιημένης επεξεργασίας. Αυτό συνεπάγεται την απόλυτη έλλειψη ανθρώπινης παρέμβασης στη διαδικασία λήψης απόφασης. Δεδομένου ότι ο όρος "αποκλειστικά βάσει" δεν ορίζεται περαιτέρω στον Κανονισμό, ο ίδιος επιτρέπει μία ερμηνεία, η οποία αποκλείει οποιαδήποτε ανθρώπινη συμμετοχή. Αυτό, ελλοχεύει τον κίνδυνο που θα καθιστούσε τη συγκεκριμένη παράγραφο του Άρθρου μη εφαρμόσιμη σε πολλές σύγχρονες πρακτικές αυτοματοποιημένης λήψης αποφάσεων.

Προκειμένου να αντιμετωπιστούν τέτοιου είδους προκλήσεις, η Ομάδα Εργασίας, μέσω των Κατευθυντήριων γραμμών, ορίζει το πεδίο εφαρμογής της αποκλειστικά αυτοματοποιημένης λήψης αποφάσεων, προσφέροντας τις ακόλουθες διευκρινήσεις:

- Ο όρος «λαμβανόμενες αποκλειστικά βάσει» συνεπάγεται την απόλυτη έλλειψη ανθρώπινης παρέμβασης στη διαδικασία λήψης απόφασης.
- Ο υπεύθυνος επεξεργασίας δεν μπορεί να αποφύγει τις διατάξεις του Άρθρου 22 κατασκευάζοντας την ανθρώπινη παρέμβαση.
- Προκειμένου να τεκμηριώνεται η ανθρώπινη παρέμβαση, ο υπεύθυνος επεξεργασίας πρέπει να μεριμνά ώστε η τυχόν εποπτεία της απόφασης να είναι ουσιαστική και όχι απλώς συμβολική. Θα πρέπει να διεκπεραιώνεται από άτομο το οποίο έχει την εξουσιοδότηση και την αρμοδιότητα να μεταβάλει την απόφαση. Στο πλαίσιο της ανάλυσης, θα πρέπει να εξετάζει το σύνολο των σχετικών δεδομένων.

Γίνεται κατανοητό ότι ο υπεύθυνος επεξεργασίας δεν μπορεί να αποφύγει τις διατάξεις του Άρθρου 22 κατασκευάζοντας την ανθρώπινη παρέμβαση και πως η ανθρώπινη παρέμβαση πρέπει να είναι ουσιαστική προκειμένου να τεκμηριώνεται. Ωστόσο, δεν υπάρχει σαφής εξήγηση σχετικά με το τί χαρακτηρίζεται ως ουσιαστική παρέμβαση, ιδίως σε περιπτώσεις πολύπλοκων και αδιαφανών μορφών προηγμένης επεξεργασίας.

Η ουσιαστική ανθρώπινη παρέμβαση είναι δύσκολο να καθοριστεί. Από τη μία, η ανθρώπινη λήψη αποφάσεων μπορεί να επηρεαστεί σημαντικά, να διαμορφωθεί και να προκαταληφθεί από προφίλ τα οποία έχουν δημιουργηθεί από εντελώς αυτοματοποιημένα μέσα. Χαρακτηριστικό παράδειγμα αποτελεί το σύστημα COMPASS της Αμερικής, που είναι υπεύθυνο για την ανάλυση των πιθανοτήτων ενός φυλακισμένου να διαπράξει ξανά κάποιο έγκλημα και χρησιμοποιείται στην έγκριση εγγυήσεων. Συγκεκριμένα, το COMPASS (Correctional Offender Managment Profiling for Alternative Sanctions) αποτελεί ένα εργαλείο που αναπτύχθηκε από τη Northpointe το οποίο βασιζόμενο σε διάφορα στοιχεία του εκάστοτε κρατούμενου (όπως προηγούμενα εγκλήματα και πόσο σοβαρά ήταν) του αναθέτει ένα σκορ, το οποίο επηρεάζει σημαντικά τις αποφάσεις των δικαστών για αιτήματα εγγυήσεων και αποφυλακίσεων. Παρόλο που η τελική απόφαση λαμβάνεται επισήμως από δικαστή, η αυτοματοποιημένη απόφαση που λαμβάνεται από το COMPASS μπορεί να είναι καθοριστική, ιδίως εάν ο δικαστής βασίζεται αποκλειστικά στα αποτελέσματα του COMPASS και δεν έχει λάβει προειδοποιήσεις σχετικά με τους κινδύνους που ελλοχεύουν, συμπεριλαμβανομένου το ότι το λογισμικό μπορεί να παρήγαγε ανακριβείς, μεροληπτικές ή άδικες αποφάσεις. [10]

Από την άλλη πλευρά, η ανθρώπινη εποπτεία δε μπορεί να είναι ουσιαστική σε περιπτώσεις που η ίδια η επεξεργασία είναι αδιαφανής. Αυτό είναι ιδιαίτερα σημαντικό σε περιπτώσεις προηγμένης επεξεργασίας που βασίζεται σε υπολογιστικούς αλγορίθμους, μηχανική μάθηση και μεγάλες ποσότητες δεδομένων (Big Data). Τέτοιου είδους επεξεργασία μπορεί να είναι πολύπλοκη και αδιαφανής, με αποτέλεσμα όσοι βασίζουν τις αποφάσεις τους σε αυτή, να μην έχουν απαραίτητα επίγνωση των λειτουργιών και των ελλείψεών της. Στην περίπτωση αυτή, ακόμα και αν ένας άνθρωπος λαμβάνει την τελική απόφαση, μια αυτοματοποιημένη διαδικασία έχει ουσιαστικά λάβει την απόφαση γι' αυτόν, χωρίς ο άνθρωπος να έχει τη δυνατότητα να συμμετέχει ουσιαστικά στην απόφαση αυτή.

Καταλήγουμε στο συμπέρασμα ότι προκειμένου η ανθρώπινη παρέμβαση να μπορεί να θεωρηθεί ουσιαστική, τα άτομα που λαμβάνουν την απόφαση θα πρέπει να είναι σε θέση να προσδιορίσουν εάν το προφίλ στο οποίο βασίζονται για τη λήψη των αποφάσεων είναι ακριβές και δίκαιο. Αυτό, βέβαια, προϋποθέτει ότι το άτομο που παρέχει ουσιαστική ανθρώπινη εποπτεία έχει επαρκές επίπεδο τεχνικής κατανόησης και εξειδίκευση σχετικά με τους πολυάριθμους τρόπους με τους οποίους η κατάρτιση προφίλ και η αυτοματοποιημένη λήψη αποφάσεων μπορούν να οδηγήσουν σε αδικίες και ανακρίβειες. Προϋποθέτει, επίσης, ότι το σύστημα που χρησιμοποιείται για τη λήψη αποφάσεων μπορεί επαρκώς να ελεγχθεί και να ερμηνευθεί. Η Ομάδα Εργασίας τονίζει ότι πρέπει να λαμβάνονται υπόψιν όλα τα δεδομένα εισόδου και εξόδου, πράγμα που δεν είναι πάντα εφικτό, ειδικά στο πλαίσιο ανάλυσης Big Data και Μηχανικής Μάθησης.

<u>Έννομα ή σημαντικά αποτελέσματα</u>

Προκειμένου να έχει ένα άτομο το δικαίωμα να μην υπόκειται σε αυτόματη λήψη αποφάσεων, το Άρθρο 22 (παράγραφος 1) του ΓΚΠΔ απαιτεί η απόφαση να παράγει έννομα αποτελέσματα για το υποκείμενο ή να το επηρεάζει σημαντικά. Οι όροι "έννομα"

ή "σημαντικά" αποτελέσματα δεν ορίζονται σαφώς στον ΓΚΠΔ ωστόσο, η διατύπωση καθιστά σαφές ότι το Άρθρο 22 καλύπτει μόνο τις σοβαρές συνέπειες.

Η Ομάδα Εργασίας, μέσω των Κατευθυντήριων γραμμών, δίνει μία ερμηνεία σχετικά με τις επιπτώσεις της επεξεργασίας δεδομένων που επηρεάζει σημαντικά κάποιο φυσικό πρόσωπο. Συγκεκριμένα, αναφέρει ότι η απόφαση, προκειμένου να είναι μεγάλη ή σημαντική πρέπει να έχει τη δυνατότητα:

- να επηρεάζει σημαντικά την κατάσταση, τη συμπεριφορά ή τις επιλογές των ενδιαφερόμενων φυσικών προσώπων,
- να έχει παρατεταμένες ή μόνιμες επιπτώσεις στο υποκείμενο των δεδομένων, ή
- στην ακραία περίπτωση, να έχει ως αποτέλεσμα τον αποκλεισμό ή τις διακρίσεις σε βάρος των φυσικών προσώπων.

Η Ομάδα Εργασίας κάνει, επίσης, λόγο για το θέμα της επιγραμμικής διαφήμισης, η οποία βασίζεται ολοένα και περισσότερο σε αυτοματοποιημένα εργαλεία και περιλαμβάνει αποκλειστικά αυτοματοποιημένη ατομική λήψη αποφάσεων. Υποστηρίζει ότι η αυτοματοποιημένη λήψη απόφασης για την προβολή στοχευμένης διαφήμισης που βασίζεται στην κατάρτιση προφίλ δεν μπορεί να επηρεάσει το άτομο σε τέτοιο βαθμό ώστε να εμπίπτει στην απαγόρευση του Άρθρου 22. Για την ενίσχυση της συγκεκριμένης άποψης, η Ομάδα Εργασίας παραθέτει το παράδειγμα της διαφήμισης ενός ηλεκτρονικού καταστήματος που αντιπροσωπεύει τη βασική τάση της μόδας με βάση ένα απλό δημογραφικό προφίλ: «γυναίκες στην περιοχή του Βελγίου, ηλικίας από 25 έως 35 ετών, οι οποίες είναι πιθανό να ενδιαφέρονται για τη μόδα και ορισμένα είδη ρουχισμού». Διευκρινίζεται ότι οι αποφάσεις που λαμβάνονται μπορεί να έχουν σημαντικές συνέπειες, υπό την έννοια του Άρθρου 22 (παράγραφος 1) του ΓΚΠΔ ανάλογα με την εκάστοτε περίσταση. Οι καθοριστικοί παράγοντες που πρέπει να λαμβάνονται υπόψη είναι, μεταξύ άλλων: η αδιακρισία της διαδικασίας κατάρτισης προφίλ, συμπεριλαμβανομένης της παρακολούθησης φυσικών προσώπων σε διάφορους δικτυακούς τόπους, συσκευές και υπηρεσίες, οι προσδοκίες και οι επιθυμίες των ενδιαφερόμενων φυσικών προσώπων, ο τρόπος με τον οποίο παρουσιάζεται η διαφήμιση και η χρήση γνώσεων σχετικά με τα ευάλωτα σημεία των στοχευόντων υποκειμένων των δεδομένων.

4.2.2. Εξαίρεση από την απαγόρευση

Παρόλο που η παράγραφος 1 του Άρθρου 22 προβλέπει τη γενική απαγόρευση της αποκλειστικά αυτοματοποιημένης λήψης αποφάσεων με έννομες ή σημαντικές συνέπειες για το άτομο, το ίδιο Άρθρο περιλαμβάνει εξαιρέσεις από την απαγόρευση αυτή. Η παράγραφος 2 του Άρθρου 22 παραθέτει τις εξαιρέσεις από την απαγόρευση για τις περιπτώσεις στις οποίες η απόφαση:

- είναι αναγκαία για τη σύναψη ή την εκτέλεση σύμβασης μεταξύ του υποκειμένου των δεδομένων και του υπευθύνου επεξεργασίας των δεδομένων,
- επιτρέπεται από το δίκαιο της Ένωσης ή το δίκαιο κράτους μέλους στο οποίο υπόκειται ο υπεύθυνος επεξεργασίας και το οποίο προβλέπει επίσης κατάλληλα μέτρα για την προστασία των δικαιωμάτων, των ελευθεριών και των έννομων συμφερόντων του υποκειμένου των δεδομένων ή
- βασίζεται στη ρητή συγκατάθεση του υποκειμένου των δεδομένων.

Συνεπώς, ένα άτομο μπορεί να υπόκειται σε πλήρως αυτοματοποιημένη λήψη αποφάσεων εάν ισχύει οποιαδήποτε από τις παραπάνω προϋποθέσεις. Προκειμένου να διευκρινιστούν οι προϋποθέσεις αυτές, η Ομάδα Εργασίας κάνει ιδιαίτερη μνεία για κάθε μία από αυτές.

Ένα παράδειγμα που υπόκεινται στην εξαίρεση της απαγόρευσης λόγω σύναψης σύμβασης αποτελεί μία επιχείρηση, η οποία ανακοινώνει μια κενή θέση εργασίας και λαμβάνει μεγάλο όγκο αιτήσεων από υποψήφιους. Προκειμένου να συνταχθεί ένας κατάλογος επίλεκτων υποψηφίων, με σκοπό τη σύναψη σύμβασης, η αυτοματοποιημένη λήψη αποφάσεων μπορεί να είναι απαραίτητη. Αξίζει να σημειωθεί πως στην περίπτωση αυτή, η αυτοματοποιημένη λήψη απόφασης δεν εστιάζει στην τελική απόφαση για την επιλογή των υποψηφίων που θα στελεχώσουν την κενή θέση αλλά στην απόκλιση αιτήσεων που δεν παρουσιάζουν την απαιτούμενη συνάφεια. Σχετικά με την επόμενη προϋπόθεση που αποτυπώνεται στην ίδια παράγραφο, η αιτιολογική σκέψη 71 παραθέτει τους σκοπούς παρακολούθησης και πρόληψης της απάτης και της φοροδιαφυγής ως παραδείγματα όπου η νομοθεσία μπορεί να επιτρέπει την αυτοματοποιημένη λήψη αποφάσεων. Τέλος, αν και η παράγραφος 2 του Άρθρου 22 αναφέρει τη ρητή συγκατάθεση του υποκειμένου των δεδομένων ως προϋπόθεση για την εξαίρεση από την απαγόρευση της αυτοματοποιημένης λήψης απόφασης, αυτό ενέχει σημαντικούς κινδύνους όσον αφορά την προστασία των δεδομένων. Για το λόγο αυτό, θεωρείται αρμόζουσα η ύπαρξη υψηλού επιπέδου ατομικού ελέγχου επί των προσωπικών του δεδομένων.

4.2.3. Δικαιώματα Υποκειμένου

Η χρήση της αυτοματοποιημένης λήψης αποφάσεων ολοένα και αυξάνεται στη σύγχρονη πραγματικότητα. Ενδεικτικά παραδείγματα αποτελούν: η αυτοματοποιημένη λήψη απόφασης σχετικά με την πιστοληπτική ικανότητα των ατόμων, η διαχείριση όγκου αιτήσεων εργασίας αλλά και η προβολή εξατομικευμένων διαφημίσεων με βάση την κατάρτιση προφίλ. Ο κίνδυνος περιορισμού των δικαιωμάτων των φυσικών προσώπων κατά την αυτοματοποιημένη λήψη αποφάσεων είναι υπαρκτός και για το λόγο αυτό ο ΓΚΠΔ απαιτεί από τον υπεύθυνο επεξεργασίας να λαμβάνει τα κατάλληλα μέτρα ώστε να διασφαλίσει τη διαφάνεια της επεξεργασίας και να προστατεύσει το υποκείμενο των δεδομένων.

Βάσει του Άρθρου 13 παράγραφος 2 στοιχείο στ) και του Άρθρου 14 παράγραφος 2 στοιχείο ζ) απαιτείται από τους υπεύθυνους επεξεργασίας να παρέχουν συγκεκριμένες και εύκολα προσβάσιμες πληροφορίες σχετικά με την αυτοματοποιημένη λήψη αποφάσεων, οι οποίες λαμβάνονται βάσει αποκλειστικά αυτοματοποιημένης επεξεργασίας, συμπεριλαμβανομένης της κατάρτισης προφίλ, η οποία παράγει έννομα ή με παρόμοιο τρόπο σημαντικά αποτελέσματα. Ειδικότερα, ο υπεύθυνος επεξεργασίας, κατά τη λήψη προσωπικών δεδομένων, παρέχει στο υποκείμενο των δεδομένων τις εξής πληροφορίες: την ύπαρξη αυτοματοποιημένης λήψης αποφάσεων, συμπεριλαμβανομένης της κατάρτισης προφίλ, σημαντικές πληροφορίες σχετικά με τη λογική που ακολουθείται, και τη σημασία και τις προβλεπόμενες συνέπειες της εν λόγω επεξεργασίας για το υποκείμενο των δεδομένων.

Σύμφωνα με την Ομάδα Εργασίας, η έννοια της ουσιαστικής πληροφόρησης σχετικά με τη λογική που ακολουθείται, σημαίνει πως ο υπεύθυνος επεξεργασίας θα πρέπει να επινοεί απλούς τρόπους να ενημερώνει το υποκείμενο των δεδομένων σχετικά με το υποκείμενο σκεπτικό, ή τα κριτήρια με βάση τα οποία λαμβάνεται η απόφαση. Ο ΓΚΠΔ δεν απαιτεί την παροχή σύνθετης επεξήγησης των αλγορίθμων που χρησιμοποιούνται αλλά οι πληροφορίες που δίνονται θα πρέπει να είναι αρκετά σαφείς ώστε να γίνουν αντιληπτές από το υποκείμενο των δεδομένων, το οποίο ενδεχομένως να μη διαθέτει ιδιαίτερη τεχνική εμπειρία.

Το Άρθρο 15 παράγραφος 1 στοιχείο η) χρησιμοποιεί την ίδια διατύπωση με το Άρθρο 13 παράγραφος 2 στοιχείο στ) και το Άρθρο 14 παράγραφος 2 στοιχείο ζ) και παρέχει στα υποκείμενα τον δεδομένων το δικαίωμα να λαμβάνουν την ίδια ενημέρωση σχετικά με την αποκλειστικά αυτοματοποιημένη λήψη αποφάσεων, συμπεριλαμβανομένης της κατάρτισης προφίλ. Παρόλο που η διατύπωση της υποχρέωσης της ενημέρωσης του υποκειμένου στα προαναφερθέντα σημεία είναι η ίδια, η διαφορά έγκειται στο χρονική στιγμή άσκησης του δικαιώματος. Το υποκείμενο των δεδομένων μπορεί να ασκήσει τα δικαιώματά του οποιαδήποτε χρονική στιγμή, το οποίο υποδηλώνει πως πρέπει να είναι σε θέση να λάβει πληροφορίες σχετικά με την επεξεργασία των προσωπικών του δεδομένων, πριν αυτή ξεκινήσει ή ακόμα και αν έχει ήδη πραγματοποιηθεί.

Ιδιαίτερη πρόβλεψη για τα δικαιώματα των υποκειμένων στην αυτοματοποιημένη λήψη αποφάσεων υπάρχει και στην αιτιολογική σκέψη 71, σύμφωνα με την οποία "η επεξεργασία αυτή θα πρέπει να υπόκειται σε κατάλληλες εγγυήσεις, οι οποίες θα πρέπει να περιλαμβάνουν ειδική ενημέρωση του υποκειμένου των δεδομένων και το δικαίωμα εξασφάλισης ανθρώπινης παρέμβασης, το δικαίωμα διατύπωσης της άποψης του, το δικαίωμα να λάβει αιτιολόγηση της απόφασης που ελήφθη στο πλαίσιο της εν λόγω εκτίμησης και το δικαίωμα αμφισβήτησης της απόφασης".

Συνοψίζοντας, προκειμένου να προστατευθούν τα υποκείμενα των δεδομένων, οι πληροφορίες που του παρέχονται θα πρέπει να είναι επαρκείς τόσο για να συναινέσει στην επεξεργασία των δεδομένων όσο και μετά τη λήψη της απόφασης που θα προκύψει από την επεξεργασία. Όπως χαρακτηριστικά αναφέρεται στο Άρθρο 3, παράγραφος 3, το υποκείμενο των δεδομένων πρέπει να έχει τη δυνατότητα να εκφράσει την άποψή του και την αμφισβήτησή του για την απόφαση που έχει ληφθεί. Είναι προφανές ότι χωρίς την παροχή ενημέρωσης μετά την επεξεργασία των δεδομένων και τη λήψη αποφάσεων, τα υποκείμενα των δεδομένων θα έπρεπε να αποδεχτούν την οποιαδήποτε απόφαση. Κάτι τέτοιο θα μπορούσε να χαρακτηριστεί ιδιαίτερα προβληματικό, καθώς η αυτοματοποιημένη λήψη αποφάσεων βασίζεται στη μηχανική μάθηση και συνεπώς τα παραγόμενα αποτελέσματα προκύπτουν μέσω πιθανοτήτων και έχουν σημαντικές συνέπειες για το υποκείμενο των δεδομένων.

Αν και ο ΓΚΠΔ περιέχει τα προαναφερθέντα Άρθρα προκειμένου να προστατεύσει τα υποκείμενα των δεδομένων, ο τρόπος με τον οποίο τα εν λόγω Άρθρα θα λειτουργούν στην πράξη, όπου οι παραβιάσεις είναι αδιαφανείς και δύσκολο να εντοπιστούν, θα χρειαστούν περαιτέρω διευκρινήσεις. Όπως είναι κατανοητό, αυτό γίνεται ακόμα πιο περίπλοκο όταν η αυτοματοποιημένη επεξεργασία και λήψη αποφάσεων περιλαμβάνει εφαρμογές μηχανικής μάθησης.

5. Νομοθεσία και Big Data σε περιβάλλον Μηχανικής Μάθησης – Νομικές Προκλήσεις

Όπως έχει γίνει ήδη κατανοητό από τις προηγούμενες ενότητες, η συνεχής τεχνολογική ανάπτυξη των δυνατοτήτων της Μηχανικής Μάθησης και ο τεράστιος όγκος μεγάλων δεδομένων που συλλέγονται και επεξεργάζονται, μετατρέπουν τις αυτοματοποιημένες αποφάσεις και την κατάρτιση προφίλ σε αναπόσπαστο κομμάτι της καθημερινότητας. Στην ενότητα αυτή, παρατίθενται οι νομικές προκλήσεις που καλούνται να αντιμετωπίσουν οι προαναφερθείσες εφαρμογές Μηχανικής Μάθησης σε σχέση με το δικαίωμα των ατόμων στην ιδιωτικότητα και την προστασία των δεδομένων τους.

5.1. Έλλειψη διαφάνειας

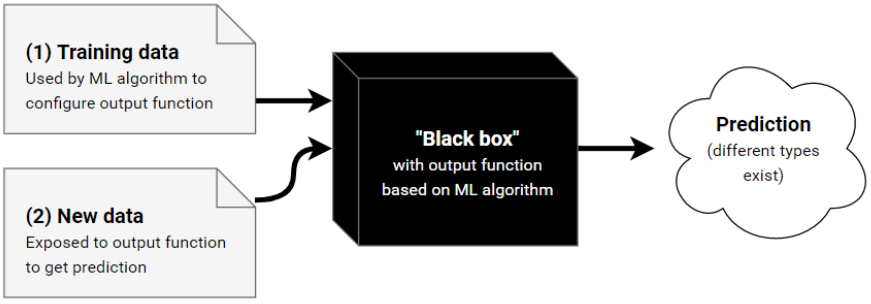
Αρκετές προκλήσεις που σχετίζονται με τα ζητήματα προστασίας των δεδομένων και απορρέουν από το ΓΚΠΔ σχετίζονται κυρίως με τη διαφάνεια της συλλογής, της επεξεργασίας και της χρήσης των δεδομένων. Τα υποκείμενα των δεδομένων έχουν το δικαίωμα να γνωρίζουν το λόγο για τον οποίο τα δεδομένα τους υποβάλλονται σε επεξεργασία ενώ ταυτόχρονα διατηρούν το δικαίωμα της πρόσβασης στα δεδομένα τους αλλά και της διόρθωσης αυτών. Σύμφωνα με το ΓΚΠΔ, ο χρήσης έχει το δικαίωμα να λαμβάνει ενημέρωση σχετικά με την επεξεργασία που υφίστανται τα δεδομένα του και συγκεκριμένα, με τρόπο σαφή και κατανοητό. Η ανάπτυξη και η πολυπλοκότητα, ωστόσο, της Μηχανικής Μάθησης μπορεί να δυσχεραίνουν την κατανόηση του τρόπου λειτουργίας μιας αυτοματοποιημένης διαδικασίας λήψης αποφάσεων ή της κατάρτισης προφίλ. Για το λόγο αυτό, γίνεται σαφές ότι προκύπτει ανησυχία σχετικά με τη διαφάνεια της επεξεργασίας και προκύπτει μία νομική πρόκληση η οποία χρήζει αντιμετώπισης.

Σε πολλά σημεία του ΓΚΠΔ, τα οποία έχουν ήδη αναλυθεί, τονίζεται η ανάγκη για χρήση αλγορίθμων κατά τη λήψη αυτοματοποιημένων αποφάσεων, οι οποίοι είναι κατανοητοί από το υποκείμενο ούτως ώστε το ίδιο να είναι σε θέση να αντιληφθεί τόσο την επεξεργασία που έχουν υποστεί τα δεδομένα του, όσο και την απόφαση που έχει προκύψει σχετικά με αυτό. Ωστόσο, η Ομάδα Εργασίας επισημαίνει πως η ενημέρωση του υποκειμένου δεν είναι ταυτόσημη με την αποκάλυψη του πλήρη αλγορίθμου που χρησιμοποιείται, αλλά είναι σημαντικό να δίνονται από τον υπεύθυνο επεξεργασίας σαφείς και επαρκείς πληροφορίες ώστε το υποκείμενο να μπορεί να κατανοήσει την απόφαση που προέκυψε αλλά και τη λογική πίσω από αυτή. Ιδιαίτερα για τα περιβάλλοντα Μηχανικής Μάθησης, στα οποία η πληθώρα των συμμετεχόντων και η πολυπλοκότητα των χρησιμοποιούμενων τεχνολογιών καθιστούν δύσκολο για το υποκείμενο των δεδομένων να γνωρίζει και να κατανοεί εάν, από ποιον και για ποιο σκοπό συλλέγονται προσωπικά δεδομένα που το αφορούν, η αιτιολογική σκέψη 58 ορίζει ότι οποιαδήποτε ενημέρωση που απευθύνεται στο κοινό ή στο υποκείμενο των δεδομένων πρέπει να είναι συνοπτική, εύκολα προσβάσιμη και εύκολα κατανοητή. Ο υπεύθυνος επεξεργασίας δεδομένων είναι υποχρεωμένος να ενημερώνει τα δεδομένα σχετικά με την ύπαρξη αυτοματοποιημένης λήψης αποφάσεων και να παρέχει ουσιαστικές πληροφορίες, σχετικά με τη λογική που εμπλέκεται, καθώς και τη σημασία και τις προβλεπόμενες συνέπειες της εν λόγω επεξεργασίας.

Η διάταξη αναφορικά με την παροχή πληροφοριών που σχετίζονται με τη λογική που ακολουθείται και τις προβλεπόμενες συνέπειες των αλγορίθμων που χρησιμοποιούνται για την αυτοματοποιημένη λήψη αποφάσεων, υπήρξε το επίκεντρο μιας τεράστιας συζήτησης στην ερευνητική κοινότητα. Μέσα από τη συζήτηση αυτή αναδείχθηκε η νομική αυτή απαίτηση σε θεμελιώδες ζήτημα στον τομέα της διαφάνειας της Μηχανικής Μάθησης. Η

διαφάνεια, λοιπόν, αποτελεί μία προϋπόθεση προκειμένου να διασφαλιστεί ότι η υπηρεσία που παρέχουν οι αλγόριθμοι Μηχανικής Μάθησης ανταποκρίνεται στις δηλωμένες υποσχέσεις του. [11]

Στο σημείο αυτό και προκειμένου να γίνει σαφής η έκταση της πολυπλοκότητας και της δυσκολίας που απαιτεί η εν λόγω υποχρέωση είναι σημαντικό να αναφερθούμε στο φαινόμενο του "μαύρου κουτιού" (black box) που χαρακτηρίζει τα συστήματα Μηχανικής Μάθησης. Ο όρος "μαύρο κουτί" χρησιμοποιείται για συστήματα ή διαδικασίες των οποίων η διαφάνεια είναι ελλιπής, υπονοώντας την αδυναμία των ανθρώπων να γνωρίζουν πώς και με ποιες παραμέτρους έχει ληφθεί μία αλγοριθμική απόφαση. Σε αυτού του είδους περιπτώσεις επεξεργασίας δεδομένων, ο άνθρωπος μπορεί να γνωρίζει τα δεδομένα που δέχτηκε το σύστημα ως είσοδο, καθώς και τα αποτελέσματα που προέκυψαν αλλά δεν είναι σε θέση να γνωρίζει τον τρόπο με τον οποίο το σύστημα κατέληξε στο αποτέλεσμα.



Εικόνα 5: Το φαινόμενο του "μαύρου κουτιού" [12]

Ιδιαίτερα στα περιβάλλοντα Μηχανικής Μάθησης, τα οποία βασίζουν τα αποτελέσματά τους στην επεξεργασία τεράστιου όγκου δεδομένων, διακρίνονται για την πολυπλοκότητά τους και την αυτονομία, καθιστώντας με αυτό τον τρόπο την εσωτερική τους λειτουργία αδιαφανή για την ανθρώπινη νόηση. Όπως τονίζεται και από τη CNIL, οι αλγόριθμοι που χρησιμοποιούνται "δεν είναι αδιαφανείς μόνο για τους τελικούς χρήστες, αλλά και οι ίδιοι οι σχεδιαστές χάνουν σταθερά την ικανότητα να κατανοούν τη λογική πίσω από τα αποτελέσματα που παράγονται". Επομένως, η απαίτηση της διαφάνειας απευθύνεται τόσο στους τεχνικούς σχεδιαστές όσο και στους υπεύθυνους επεξεργασίας. Οι πρώτοι πρέπει να ανταποκριθούν στην πρόκληση να αντιμετωπίσουν την αδιαφάνεια βελτιώνοντας την τεχνολογία, ενώ οι δεύτεροι θα πρέπει να ενημερώνουν τα υποκείμενα των δεδομένων σχετικά με τρόπο επεξεργασίας αυτών.

Είναι γεγονός ότι η ανάπτυξη της Μηχανικής Μάθησης συμβάλλει στην ολοένα αυξανόμενη αυτονομία των μηχανών. Την ίδια στιγμή, όσο αυξάνεται η αυτονομία των μηχανών, τόσο μειώνεται η δυνατότητα επεξήγησης και η κατανόησης των αλγορίθμων που αυτές χρησιμοποιούν. Οι αλγόριθμοι που χρησιμοποιούνται, αναπτύσσουν τη δική τους λογική προκειμένου να καταλήξουν σε συμπεράσματα, η οποία διαμορφώνεται συνεχώς, όσο ο αλγόριθμος τροφοδοτείται με νέα δεδομένα και μαθαίνει. Συνεπώς, οι αλγόριθμοι παρουσιάζουν μια αδιαφάνεια, η οποία αποτυπώνεται ως αδυναμία μετάφρασης των αλγοριθμικών διαδικασιών σε κατανοητή γλώσσα. Γίνεται κατανοητό, ότι η αλγοριθμική διαφάνεια βρίσκεται σε κίνδυνο όταν πραγματοποιούνται πολύπλοκες αλγοριθμικές διαδικασίες όπως είναι η κατάρτιση προφίλ και η λήψη αυτοματοποιημένων αποφάσεων σε περιβάλλοντα Μηχανικής Μάθησης. Εάν, όμως, υπεύθυνος επεξεργασίας

δεν είναι σε θέση να εξηγήσει και να δώσει σαφείς πληροφορίες στο υποκείμενο των δεδομένων, δεν μπορεί να εκπληρώσει την απαίτηση που ορίζεται από τον ΓΚΠΔ και συνεπώς να συμμορφωθεί με αυτόν. Εγείρονται έτσι ενδιαφέροντα ερωτήματα σχετικά με τη φύση και την έκταση του δικαιώματος στην ιδιωτικότητα και της προστασίας των δεδομένων των ατόμων καθώς η έλλειψη διαφάνειας στα περιβάλλοντα Μηχανικής Μάθησης, συμπεριλαμβανομένης της κατάρτισης προφίλ και της λήψης αυτοματοποιημένων αποφάσεων, αποτελεί εμπόδιο για τη συμμόρφωση με τις νομικές διατάξεις και πρόκληση για το δικαίωμα της ιδιωτικότητας και την προστασία των δεδομένων των ατόμων.

5.2. Εκτεταμένη συλλογή δεδομένων και δικαιώματα Υποκειμένου

Οι νομικές προκλήσεις που καλούνται να αντιμετωπίσουν τα περιβάλλοντα Μηχανικής Μάθησης δε σταματάνε στην έλλειψη διαφάνειας των αλγορίθμων. Ο ΓΚΠΔ θεσπίζει μία θεμελιώδη αρχή σύμφωνα με την οποία τα προσωπικά δεδομένα που συλλέγονται πρέπει να είναι επαρκή, σχετικά και να περιορίζονται στα απαραίτητα για το σκοπό για τον οποίο έχουν συλλεχθεί. Όμως αυτό, φαίνεται να έρχεται σε αντίθεση με τα περιβάλλοντα Μηχανικής Μάθησης, τα οποία χαρακτηρίζονται από την εκτεταμένη συλλογή των Big Data καθώς και τη δημιουργία νέων παραγόμενων προσωπικών δεδομένων. Μέσω της κατάρτισης προφίλ και της λήψης αυτοματοποιημένων αποφάσεων, τα προσωπικά δεδομένα σταματάνε να περιορίζονται στο όνομα, το επίθετο ή τον αριθμό του τηλεφώνου του υποκειμένου. Οι προβλέψεις που γίνονται για το άτομο αυτό και το προφίλ που δημιουργείται γύρω από το ίδιο μπορούν να θεωρηθούν προσωπικά δεδομένα και συνεπώς εγείρεται ξανά το ερώτημα σχετικά με το δικαίωμά του στην ιδιωτικότητα. Στο σημείο αυτό, αξίζει να αναφερθεί το γεγονός ότι τα παραγόμενα αποτελέσματα δεν είναι πάντα ρεαλιστικά ή αληθινά, εφόσον είναι βασισμένα σε προβλέψεις που έχουν προκύψει από πιθανολογικά σενάρια.

Μέσω του Άρθρου 14 του ΓΚΠΔ παρέχεται στα υποκείμενα των δεδομένων το δικαίωμα ενημέρωσής τους για τα προσωπικά δεδομένα που προέκυψαν μέσω της κατάρτισης προφίλ. Με τον τρόπο αυτό, διασφαλίζεται το δικαίωμα των υποκειμένων των δεδομένων να ζητήσουν τη διόρθωση ή τη διαγραφή των δεδομένων. Επιπλέον, όπως έχει ήδη αναφερθεί, η βασική διάταξη του ΓΚΠΔ που αφορά την αυτοματοποιημένη λήψη αποφάσεων, παρουσιάζεται στο Άρθρο 22 και δίνει στο υποκείμενο το δικαίωμα να εξαιρείται από τέτοιου είδους διαδικασίες.

Το Άρθρο 22 επηρεάζει άμεσα και σε μεγάλο βαθμό τις πρακτικές των Big Data και της Μηχανικής Μάθησης σε πολλά επίπεδα. Η απαγόρευση της αυτοματοποιημένης ανάλυσης μπορεί να υπονομεύει τις πρακτικές Μηχανικής Μάθησης ενώ η απαίτηση για ερμηνευσιμότητα σε κάθε στάδιο επεξεργασίας μπορεί να θέσει σε κίνδυνο την ακρίβεια του συστήματος προκειμένου να καταστεί δυνατή η παροχή αυτής της μορφής λεπτομερούς εξήγησης. Επιπλέον, η δυνατότητα ανθρώπινης παρέμβασης είναι πιθανό να επιβαρύνει περαιτέρω την αυτοματοποιημένη διαδικασία και να επιβραδύνει την τεχνολογική καινοτομία.

Το Άρθρο 22 είναι ίσως το πιο χαρακτηριστικό παράδειγμα σύγκρουσης του ΓΚΠΔ και των τεχνικών Big Data και Μηχανικής Μάθησης, σηματοδοτώντας μια βαθιά δυσπιστία απέναντι στις αυτοματοποιημένες διαδικασίες. Προκειμένου να συμμορφωθούν οι επιχειρήσεις με αυτόν τον κανόνα είναι πιθανό να κληθούν να προβούν σε ριζικές αλλαγές που αφορούν τις τεχνολογικές αρχιτεκτονικές και ακόμα και ολόκληρα επιχειρηματικά μοντέλα, επιλέγοντας λιγότερο αποτελεσματικές τεχνικές απλά για να διασφαλίσουν τη

συμμόρφωση με το συγκεκριμένο κανόνα. Ωστόσο, είναι πιθανό, από όλες τις διατάξεις που ορίζονται από τον ΓΚΠΔ, το Άρθρο 22 να είναι εκείνο με το μικρότερο αποτέλεσμα, πρακτικά. Όπως έχει ήδη αναφερθεί και αναλυθεί σε προηγούμενο κεφάλαιο, το Άρθρο αυτό, μπορεί εύκολα να παρακαμφθεί με την εισαγωγή έστω και ελάχιστης ανθρώπινης αλληλεπίδρασης. [13]

Γίνεται κατανοητό πως το δικαίωμα της ανθρώπινης παρέμβασης θέτει νέες προκλήσεις τόσο για τις επιχειρήσεις που αξιοποιούν τις πρακτικές Big Data και Μηχανικής Μάθησης αλλά και για τους ανθρώπους της τεχνολογίας. Η ICO τονίζει στους οργανισμούς διαχείρισης Big Data, την ανάγκη να «είναι προσεκτικοί προτού βασιστούν σε αποφάσεις μηχανικής μάθησης που δεν μπορούν να εξηγηθούν με κατανοητούς όρους από τον άνθρωπο».

5.3. Περιορισμός σκοπού

Μία από τις θεμελιώδεις έννοιες του Κανονισμού, όπως περιγράφεται και σε προηγούμενη ενότητα, ορίζεται στο Άρθρο 5 και σύμφωνα με αυτή τα προσωπικά δεδομένα πρέπει να συλλέγονται για καθορισμένους, ρητούς και νόμιμους σκοπούς και να μην υποβάλλονται σε περαιτέρω επεξεργασία κατά τρόπο ασύμβατο προς τους σκοπούς αυτούς. Στόχος της αρχής περιορισμού του σκοπού είναι να αποτρέψει τη χρήση και την επαναχρησιμοποίηση των συλλεχθέντων δεδομένων με τρόπους μη αναμενόμενους από το υποκείμενο των δεδομένων, επιτρέποντας παράλληλα στους υπεύθυνους επεξεργασίας να επεξεργάζονται περαιτέρω δεδομένα για άλλους χρήσιμους σκοπούς που δε θεωρούνται ασύμβατοι με τους αρχικούς σκοπούς. Ο περιορισμός του σκοπού αποτελεί ακρογωνιαίο λίθο για την προστασία των δεδομένων και για το λόγο αυτό λειτουργεί ως βάση για πολλές από τις υπόλοιπες διατάξεις του ΓΚΠΔ.

Η έννοια, όμως, αυτή έρχεται σε αντίθεση με τις τεχνολογίες Big Data σε περιβάλλοντα Μηχανικής Μάθησης. Συχνά, η ανάλυση των Big Data μπορεί να χρησιμοποιηθεί ως είσοδος σε άλλες μεθόδους και να χρησιμοποιηθεί για σκοπούς, τους οποίους ούτε η οντότητα που συλλέγει τα δεδομένα αλλά ούτε και το υποκείμενο των δεδομένων θα μπορούσαν να είχαν φανταστεί κατά τη διάρκεια της συλλογής. Για παράδειγμα, η απαίτηση συλλογής δεδομένων για συγκεκριμένο σκοπό, φαίνεται να έρχεται σε σύγκρουση με τη διαδικασία που ακολουθείται για την κατάρτιση προφίλ, η οποία βασίζεται πρώτα στη συλλογή μεγάλου όγκου Big Data και στη συνέχεια ακολουθεί η εξεύρεση του σκοπού γι' αυτά. Προκειμένου η χρήση των Big Data σε περιβάλλοντα Μηχανικής Μάθησης να είναι σύμφωνη με την προαναφερθείσα αρχή, οι υπεύθυνοι επεξεργασίας των δεδομένων θα πρέπει να ενημερώνουν τα υποκείμενα των δεδομένων σχετικά με τις μορφές της επεξεργασίας στις οποίες θα εμπλακούν καθώς και να παρακολουθούν στενά τις μεθόδους επεξεργασίας ώστε να είναι σίγουροι πως είναι σύννομες με τις αρχές του ΓΚΠΔ. Τόσο η ενημέρωση όσο και η συνεχής παρακολούθηση μπορεί να αποδειχθούν δαπανηρές, δύσκολες ή και αδύνατες. Όμως, ανάλογα με το εύρος στο οποίο προσδιορίζεται ο σκοπός, ο υπεύθυνος επεξεργασίας μπορεί να έχει την ευελιξία να αποφασίσει πώς θα επεξεργαστούν τα δεδομένα που έχουν συλλεχθεί, ενώ η επεξεργασία τους θα συνεχίζει να εμπίπτει στον αρχικό σκοπό συλλογής. Πολλοί ερευνητές αντιτίθενται σε αυτή την προσέγγιση.

Λαμβάνοντας υπόψη ότι ως συλλεχθέντα δεδομένα νοούνται τόσο εκείνα τα οποία το ίδιο το υποκείμενο διαθέτει προς επεξεργασία όσο και εκείνα που παράγονται από την ανάλυση δεδομένων που έχει παράσχει το υποκείμενο, δε φαίνεται να είναι νομικά εύκολο για έναν οργανισμό οποίος χρησιμοποιεί Big Data για παράδειγμα για κατάρτιση προφίλ, να

προσδιορίσει ένα ευρύ σκοπό, που να περιλαμβάνει πιθανές μελλοντικές χρήσεις των προσωπικών δεδομένων. Όπως γίνεται κατανοητό και από προηγούμενες ενότητες τα περιβάλλοντα Μηχανικής Μάθησης στηρίζονται κατά κύριο λόγο στο μεγάλο όγκο των δεδομένων και Big Data. Αυτό σημαίνει ότι μπορεί να συλλέγονται και επεξεργάζονται ταυτόχρονα πολλά προσωπικά δεδομένα πολλών διαφορετικών υποκειμένων. Αυτός είναι ένας λόγος για τον οποίο απαιτείται να υπάρχει σαφής προσδιορισμός για το σκοπό της επεξεργασίας των δεδομένων. Εκτός όμως από αυτό, ιδιαίτερη προσοχή πρέπει να δίνεται στου είδος της επεξεργασίας αλλά και στην πολυπλοκότητα των μεθόδων που χρησιμοποιούνται κατά την επεξεργασία των δεδομένων καθώς πολλές φορές, μπορεί να χρησιμοποιούνται εξαιρετικά εξελιγμένες και προηγμένες μέθοδοι επεξεργασίας. Συνεπώς, η επεξεργασία των Big Data απαιτεί ένα υψηλό επίπεδο προσδιορισμού του σκοπού προκειμένου να διασφαλιστεί ότι θα πραγματοποιείται σύννομα με τον ΓΚΠΔ και προστατεύοντας τα προσωπικά δεδομένα των υποκειμένων.

Είναι γεγονός ότι υπάρχουν αρκετοί ουσιαστικοί λόγοι για τη διατήρηση και την υιοθέτηση του περιορισμού του σκοπού ακόμα και στην εποχή των Big Data. Αρχικά, η υποχρέωση των υπεύθυνων επεξεργασίας να σέβονται την αρχή του περιορισμού του σκοπού δίνει στα υποκείμενα των δεδομένων να ασκούν κάποιον έλεγχο επί των προσωπικών τους δεδομένων, προστατεύοντας έτσι το αναφαίρετό τους δικαίωμα στην ιδιωτικότητα. Επιπλέον, η αρχή περιορισμού του σκοπού προάγει την εμπιστοσύνη στα περιβάλλοντα Μηχανικής Μάθησης και γενικότερα της επεξεργασίας δεδομένων ενώ παράλληλα προάγουν τον ευγενή ανταγωνισμό μεταξύ των οργανισμών που χρησιμοποιούν μεθόδους Big Data σε περιβάλλοντα Μηχανικής Μάθησης.

Αξίζει να σημειωθεί ότι η αρχή περιορισμού του σκοπού, όπως αυτή περιγράφεται στον ΓΚΠΔ, περιλαμβάνει ένα συγκεκριμένο χαρακτηριστικό, το οποίο θα μπορούσε να επιτρέψει στην ανάλυση των Big Data να ευδοκιμήσει. Το χαρακτηριστικό αυτό, επιτρέπει τη μεταγενέστερη επεξεργασία δεδομένων, ακόμα και αν υπερβαίνει τον αρχικά καθορισμένο σκοπό, αρκεί αυτή η μεταγενέστερη επεξεργασία να είναι συμβατή με τον αρχικά καθορισμένο σκοπό. Ωστόσο, ακόμα κι έτσι, μπορεί κανείς να ισχυριστεί με βεβαιότητα ότι ο ΓΚΠΔ παρεμποδίζει σημαντικά τις πρωτοβουλίες των τεχνικών Big Data και τα αποτελέσματά τους.

Η έννοια της συμβατότητας επεξηγείται στο Άρθρο 5 (παράγραφος 1 στ. Β) του ΓΚΠΔ. Συγκεκριμένα, το Άρθρο ορίζει πως η επεξεργασία για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον ή σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς δεν θεωρείται ασύμβατη με τους αρχικούς σκοπούς. Επομένως εάν τα Big Data εμπίπτουν σε κάποια από αυτές τις κατηγορίες, θα μπορούσαν να υποβληθούν σε περαιτέρω επεξεργασία. Η έκταση της εξαίρεσης αυτής περιγράφεται με μεγαλύτερη λεπτομέρεια στο Άρθρο 89, σύμφωνα με το οποίο η επεξεργασία για τους σκοπούς αυτούς, πρέπει να υπόκειται σε κατάλληλες εγγυήσεις ως προς τα δικαιώματα και τις ελευθερίες του υποκειμένου των δεδομένων. Οι εν λόγω εγγυήσεις διασφαλίζουν ότι έχουν θεσπιστεί τα τεχνικά και οργανωτικά μέτρα, ιδίως για να διασφαλίζουν την τήρηση της αρχής της ελαχιστοποίησης των δεδομένων. Τα εν λόγω μέτρα μπορούν να περιλαμβάνουν τη χρήση ψευδωνύμων, εφόσον οι εν λόγω σκοποί μπορούν να εκπληρωθούν κατ' αυτόν τον τρόπο.

Η μεγαλύτερη πρόκληση για την επίκληση της εξαίρεσης των στατιστικών σκοπών για την επεξεργασία των Big Data περιγράφεται στην αιτιολογική σκέψη 162. Η αιτιολογική σκέψη αυτή παρέχει πρόσθετες επεξηγηματικές διατυπώσεις σχετικά με την έννοια της εν λόγω εξαίρεσης. Στο ίδιο σημείο σημειώνεται ότι ο όρος «στατιστικοί σκοποί» σημαίνει

κάθε πράξη συλλογής και την επεξεργασία δεδομένων προσωπικού χαρακτήρα που είναι αναγκαία για την πραγματοποίηση στατιστικών ερευνών ή για την παραγωγή στατιστικών αποτελεσμάτων. Ο στατιστικός σκοπός συνεπάγεται ότι το αποτέλεσμα της επεξεργασίας για στατιστικούς σκοπούς δεν είναι δεδομένα προσωπικού χαρακτήρα αλλά συγκεντρωτικά δεδομένα και ότι το αποτέλεσμα αυτό ή τα δεδομένα προσωπικού χαρακτήρα δεν χρησιμοποιούνται προς υποστήριξη μέτρων ή αποφάσεων που αφορούν συγκεκριμένο φυσικό πρόσωπο. Ωστόσο, οι μέθοδοι επεξεργασίας Big Data, πολλές φορές χρησιμοποιούνται προκειμένου να παρέχουν μοναδική και ειδική μεταχείριση στα υποκείμενα των δεδομένων. Εκ πρώτης όψεως, τέτοιου είδους χρήσεις επιτρέπονται μόνο εφόσον συμπίπτουν με την αρχή περιορισμού του σκοπού όπως αυτή έχει περιγραφεί παραπάνω, το οποίο είναι δύσκολο για τα περιβάλλοντα που επεξεργάζονται Big Data.

Εν κατακλείδι, η απαίτηση για τον περιορισμό του σκοπού έρχεται σε σύγκρουση με τις μεθόδους επεξεργασίας των Big Data. Μπορεί ο ΓΚΠΔ να περιλαμβάνει ορισμένα μέτρα όπως η ψευδωνυμοποίηση προκειμένου να περιοριστεί το πρόβλημα, ωστόσο το αποτέλεσμα εξακολουθεί να είναι ανεπαρκές και να δημιουργεί αβεβαιότητα. Εφόσον. Η επεξεργασία των Big Data βασίζεται σε σκοπούς που επιτρέπουν την επαναχρησιμοποίησή τους χωρίς την εγκατάλειψη της αρχής του περιορισμού του σκοπού και αυτό οδηγεί στο συμπέρασμα το Άρθρο 5 να μεταφράζεται με διττό τρόπο: αφενός οι σκοποί αυτοί δεν παραβιάζουν την προαναφερθείσα αρχή, αφετέρου, όμως, μπορούν να λειτουργήσουν και ως δικαιολογητική βάση για συγκέντρωση και επεξεργασία μεγαλύτερου όγκου Big Data καθιστώντας τελικά άνευ περιεχομένου την αρχή του περιορισμού του σκοπού.

5.4. Αδικία και διακρίσεις

Ένας από τους βασικότερους στόχους της επεξεργασίας Big Data σε περιβάλλοντα Μηχανικής Μάθησης είναι να καταστήσει τις μηχανές ικανές να λαμβάνουν αποφάσεις χωρίς να χρειάζεται η ανθρώπινη παρέμβαση. Όμως, η έλλειψη της ανθρώπινης παρέμβασης για τη λήψη αποφάσεων, συνεπάγεται την έλλειψη ενσυναίσθησης στις μηχανές και συνεπώς στις αποφάσεις που λαμβάνονται. Αυτό, μπορεί να έχει ως αποτέλεσμα τη λήψη λανθασμένων ή και άδικων για το υποκείμενο αποφάσεων.

Η επεξεργασία πλήθους Big Data σε περιβάλλοντα Μηχανικής Μάθησης για την παραγωγή αποτελεσμάτων και αποφάσεων στηρίζεται στην εύρεση συσχετίσεων. Για το λόγο αυτό, τα αποτελέσματα που παράγονται ενέχουν ένα βαθμό αβεβαιότητας ως προς την ορθότητά τους. Αυτό είναι ιδιαίτερα ανησυχητικό όταν τα αποτελέσματα αυτά μπορούν να καθορίσουν σημαντικά τις ικανότητες και τις δυνατότητες του υποκειμένου όπως για παράδειγμα βάσει του αποτελέσματος μπορεί να καθοριστεί η πιστοληπτική ικανότητα ενός ατόμου. Αναζητώντας την αιτία για την οποία η προαναφερθείσα επεξεργασία μπορεί δυνητικά να οδηγήσει σε άδικα ή μεροληπτικά αποτελέσματα, συμπεραίνει κανείς ότι οι αιτίες ποικίλλουν αλλά και αλληλοεπιδρούν μεταξύ τους. Μία από τις βασικότερες, όμως, αιτίες που μπορούν να οδηγήσουν σε άδικα αποτελέσματα είναι η χρήση άδικων δεδομένων ως δεδομένα εισόδου για τους αλγορίθμους Μηχανικής Μάθησης. Εάν για παράδειγμα, στο παρελθόν, έχουν γίνει διακρίσεις σε συγκεκριμένη ομάδα ατόμων στο παρελθόν είναι πιθανό να έχουν ληφθεί αποφάσεις με βάση αυτές τις διακρίσεις. Τα αποτελέσματα αυτά, μπορούν να χρησιμοποιηθούν ως δεδομένα εισόδου σε περιβάλλοντα Μηχανικής Μάθησης τα οποία, με τη σειρά τους θα εξάγουν άδικα αποτελέσματα.

Οι αλγόριθμοι Μηχανικής Μάθησης δεν μπορούν να αντιταχθούν στα άδικα αποτελέσματα αφού λειτουργούν με βάση τα δεδομένα που εισάγονται σε αυτούς από ανθρώπους.

Εφόσον στις ανθρώπινες κοινωνίες δεν είναι απόλυτα δίκαιες και περιέχουν διακρίσεις όπως είναι ο αποκλεισμός συγκεκριμένων ατόμων ή ομάδων, το ίδιο θα ισχύει και στα δεδομένα που υπάρχουν σε αυτές. Το ενδεχόμενο των άδικων Big Data μπορεί να υφίσταται επίσης λόγω του ότι ολόκληρες ομάδες ατόμων, είναι πιθανό να εκπροσωπούνται από ένα συγκεκριμένο δείγμα. Αυτό σημαίνει ότι είναι πιθανό τα αποτελέσματα να τείνουν να ευνοούν τις ομάδες οι οποίες εκπροσωπούνται καλύτερα μέσω του δείγματός τους.

6. Αντιμετώπιση Προκλήσεων

6.1. Ενίσχυση απαίτησης συγκατάθεσης

Ένα πρώτο πιθανό μέτρο που θα μπορούσε να προστατεύσει την ιδιωτικότητα των υποκειμένων κατά την επεξεργασία των Big Data που το αφορούν είναι η ενίσχυση της συγκατάθεσής τους. Αυτό, μπορεί να επιτευχθεί με την αύξηση των περιπτώσεων κατά τις οποίες είναι απαραίτητη η συγκατάθεση των υποκειμένων προκειμένου τα δεδομένα που το αφορούν να υποστούν κάποιου είδους επεξεργασία. Η απαίτηση ρητής συγκατάθεσης για κάθε είδους επεξεργασία δεδομένων, ενισχύει τη διαφάνεια και θα μπορούσε να βελτιώσει τη λήψη αποφάσεων οι οποίες θα σέβονται την ιδιωτικότητα των υποκειμένων και θα μπορούν να τεκμηριωθούν εάν αυτό απαιτηθεί. Ωστόσο, υπάρχουν αρκετές αμφιβολίες σχετικά με την αποτελεσματικότητα αυτού του μέτρου καθώς η ευθύνη για τη συγκατάθεση βαραίνει τα υποκείμενα των δεδομένων, τα οποία συχνά όταν καλούνται να διαβάσουν τους αντίστοιχους όρους σε συμβόλαια και συμβάσεις, δε δίνουν την απαραίτητη προσοχή. Τα υποκείμενα των δεδομένων συνάπτουν πληθώρα συμβάσεων που περιλαμβάνουν όρους και δίνουν τη συγκατάθεσή τους. Η τεράστια όμως συχνότητα με την οποία τα υποκείμενα καλούνται να δώσουν τη συγκατάθεσή τους, θα μπορούσε να δυσκολέψει τα ίδια να διακρίνουν πότε πρόκειται για καταστάσεις χαμηλού και πότε για υψηλού κινδύνου.

Πολλοί συγγραφείς και μελετητές έχουν εστιάσει στη θέσπιση πιο ουσιαστικών κανόνων, οι οποίοι αποτρέπουν ορισμένες πρακτικές προκειμένου να βελτιωθεί ο τρόπος με τον οποίο τα υποκείμενα των δεδομένων καλούνται να δώσουν τη συγκατάθεσή τους. Οι κανόνες αυτοί επικεντρώνονται στον τρόπο με τον οποίο παρουσιάζονται οι ειδοποιήσεις και οι όροι που αφορούν την προστασία της ιδιωτικότητας και περιλαμβάνουν σύντομες περιγραφές αποφεύγοντας περιττά λόγια και σχόλια και πιο κατανοητή έκφραση ώστε να μπορεί να γίνει εύκολα αντιληπτό το περιεχόμενο. Η έλλειψη χρόνου που χαρακτηρίζει τη σύγχρονη πραγματικότητα οδηγεί στη μη αξιολόγηση όλων των σχετικών συμβάσεων και όρων για την προστασία της ιδιωτικότητας που καλούνται να υπογράψουν τα υποκείμενα. Εάν κανείς αναλογιστεί, σε συνδυασμό με την έλλειψη χρόνου, την επιθυμία των υποκειμένων να απολαύσουν νέα προϊόντα και υπηρεσίες τα οποία βασίζονται στην επεξεργασία των προσωπικών τους δεδομένων, μπορεί να καταλήξει στο συμπέρασμα ότι πολλές συμβάσεις προστασίας της ιδιωτικότητας υπογράφονται χωρίς να δίνεται η απαραίτητη προσοχή με αποτέλεσμα το ίδιο το υποκείμενο να μη γνωρίζει ξεκάθαρα το λόγο για τον οποίο δίνει τη συγκατάθεσή του, το οποίο έρχεται σε πλήρη αντίθεση με το σκοπό της αρχής της συγκατάθεσης.

Η ενίσχυση της απαίτησης συγκατάθεσης, λοιπόν, ως μέτρο προστασίας της ιδιωτικότητας χάνει την αποτελεσματικότητά της όταν συλλέγεται μεγάλο πλήθος Big Data αφού η ποικιλομορφία των δεδομένων δυσχεραίνει τον έλεγχο της επεξεργασίας αυτών και υποχρεώνει τα υποκείμενα των δεδομένων να συναναστρέφονται με μεγάλο πλήθος υπεύθυνων επεξεργασίας, ανάλογα με τα εκάστοτε δεδομένα και την επεξεργασία που αυτά υφίστανται.

6.2. Ανωνυμοποίηση

Ένας από τους μηχανισμούς που χρησιμοποιούνται προκειμένου να ενισχυθεί η ιδιωτικότητα των υποκειμένων των δεδομένων είναι η ανωνυμοποίηση. Όταν αυτή εκτελείται σωστά, καθίσταται αδύνατο να εξακριβωθεί η ταυτότητα του υποκειμένου των δεδομένων. Τα ανωνυμοποιημένα δεδομένα δε νοούνται πλέον ως προσωπικά δεδομένα και για το λόγο αυτό δεν εμπίπτουν στο πεδίο εφαρμογής του ΓΚΠΔ. Σύμφωνα με τον

ΓΚΠΔ, ως ανωνυμοποίηση ορίζεται η διαδικασία διαγραφής των αναγνωριστικών προσωπικού χαρακτήρα σε εγγραφές αποθηκευμένων δεδομένων, έτσι ώστε να μην είναι πλέον δυνατόν τα δεδομένα αυτά να συσχετιστούν με το υποκείμενο των δεδομένων το οποίο αφορούν. Η χρήση της ανωνυμοποίησης διαφέρει από αυτή της ψευδωνυμοποίησης, διότι καθιστά θεωρητικά αδύνατο να προσδιοριστεί το υποκείμενο των δεδομένων, σε αντίθεση με την τεχνική της ψευδωνυμοποίησης με την οποία δεν διαγράφεται η ταυτότητα, αλλά αντικαθίσταται με τέτοιο τρόπο ώστε να απαιτούνται επιπλέον πληροφορίες για να είναι δυνατή η αναγνώριση των αρχικών υποκειμένων. Συνεπώς βάσει και της αιτ. σκέψης (26), ο GDPR δεν εφαρμόζεται σε τέτοιου είδους πληροφορίες (ανωνυμοποιημένες) αφού δεν μπορούν να συσχετιστούν με ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο, ή σε δεδομένα προσωπικού χαρακτήρα. Κατά την διαδικασία της ανωνυμοποίησης αφαιρούνται από τα δεδομένα όλες οι πληροφορίες που μπορούν να ταυτοποιήσουν το υποκείμενο των δεδομένων (όπως ονοματεπώνυμο, ημερομηνίες γέννησης κα.), διατηρώντας όμως όλα τα υπόλοιπα στοιχεία ώστε τα δεδομένα αυτά να είναι χρήσιμα για έρευνα, όπως για παράδειγμα να μπορεί να γίνει εντοπισμός επαναλαμβανόμενων μοτίβων για εξαγωγή χρήσιμων συμπερασμάτων.

Και ενώ η ανωνυμοποίσηση μπορεί να λειτουργήσει αποτελεσματικά για ένα πεπερασμένο σύνολο δεδομένων, στην περίπτωση των Big Data απαιτείται ιδιαίτερη προσοχή για την επιλογή της κατάλληλης τεχνικής ανωνυοποίησης. Ιδιαίτερα στα Big Data, τα οποία χαρκατηρίζονται από τον μεγάλο όγκο και τη μεγάλη ποικιλομορφία τους, είναι δύσκολη η επιλογή της κατάλληλης τεχνικής ανωνυμοποίησης προκειμένου να μη διακυβευτεί η ιδιωτικότητα των δεδομένων αλλά ταυτόχρονα ούτε η χρησιμότητά τους. Τα ανωνυμοποιημένα δεδομένα πρέπει να εγγυώνται ότι καμία οντότητα δεν μπορεί να ταυτοποιηθεί κατά τη σύνδεση συνόλου δεδομένων με άλλες διαθέσιμες πληροφορίες. Ωστόσο, μια μερική ανωνυμοποίηση κάποιων στοιχείων όπως το όνομα δεν αρκεί προκειμένου να διατηρηθεί η ιδιωτικότητα καθώς η πρόσβαση σε άλλες πρόσθετες πληροφορίες θα μπορούσε να επιτρέψει την εκ νέου ταυτοποίηση των ατόμων. Ως εκ τούτου είναι ζωτικής σημασίας να προσδιορίζονται όλα τα πιθανά δεδομένα που θα μπορούσαν να χρησιμοποιηθούν για την ταυτοποίηση ενός ατόμου, λαμβάνοντας υπόψη πιθανές πληροφορίες τις οποίες κάποιος επιτιθέμενος μπορεί να έχει από άλλες πηγές. Από την άλλη, μια πολύ ισχυρή ανωνυμοποίηση θα μπορούσε να εμποδίσει τη σύνδεση των δεδομένο για ένα συγκεκριμένο άτομο ή μια ομάδα ατόμων και ως εκ τούτου να παρεμποδίσει πολλά από τα οφέλη που προσφέρουν οι τεχνικές των Big Data.

Η ανωνυμοποίηση των δεδομένων είναι αναμφίβολα μία πρακτική η οποία ευνοεί την προστασία της ιδιωτικότητας καθώς συμβάλλει στη μη ταυτοποίηση των ατόμων. Στα περιβάλλοντα, όμως, των Big Data πρέπει να λαμβάνεται υπόψη το γεγονός ότι ο όγκος των δεδομένων είναι μεγαλύτερο και συνεπώς ανακύπτουν μεγαλύτερα υπολογιστικά προβλήματα. Επιπλέον, εκτός από αυτό, προκύπτουν τα εξής προβλήματα:

- Έλλειψη ελέγχου και διαφάνειας: Από τη στιγμή που πλήθος δεδομένων συλλέγεται συνεχώς από διαφορετικές πηγές, γίνεται όλο και πιο δύσκολο για τα υποκείμενα των δεδομένων να γνωρίζουν ποιος κατέχει τα δεδομένα τους. Είναι άλλωστε πιθανό, να υπάρχουν οργανισμοί οι οποίοι κατέχουν και επεξεργάζονται προσωπικά δεδομένα, χωρίς τα υποκείμενα των δεδομένων να το γνωρίζουν. Για να γίνει αυτό κατανοητό, αρκεί να αναλογιστεί κανείς τις πληροφορίες που συλλέγονται από αισθητήρες και κάμερες, από αναρτήσεις σε κοινωνικά δίκτυα ή από αναλύσεις διαδικτυακών αναζητήσεων. Προκύπτει εύκολα το συμπέρασμα πως για κάθε υποκείμενο δεδομένων υπάρχει σύνολο δεδομένων, διαφορετικών

ειδών τα οποία συλλέγονται από διαφορετικές πηγές και μπορούν να χρησιμοποιηθούν προκειμένου να δημιουργηθεί ένα εξατομικευμένο προφίλ γύρω από αυτό.

- Συνδεσιμότητα: Συχνά, στην περίπτωση των Big Data, τα ίδια συνδέονται με βάσεις δεδομένων προκειμένου να βελτιωθεί η ποσότητα αλλά και η ποιότητα των παραγόμενων πληροφοριών και υπηρεσιών. Όμως, η συνδεσιμότητα των δεδομένων με ολόκληρες βάσεις δεδομένων, αυξάνει τον κίνδυνο της ταυτοποίησης ενός ατόμου καθώς πλέον υπάρχουν πολλές περισσότερες πληροφορίες για το κάθε άτομο. Αξίζει να σημειωθεί πως όσο περισσότερες οι πληροφορίες που συλλέγονται για ένα άτομο, τόσο πιο εύκολο είναι το άτομο να ταυτοποιηθεί αλλά και να επαναπροσδιοριστεί ενώ αντίθετα, τόσο πιο δύσκολο είναι να προστατευτεί η ιδιωτικότητά του.
- Εξαγωγή συμπερασμάτων και επαναχρησιμοποίηση δεδομένων: Υπάρχουν αποτελεσματικοί αλγόριθμοι οι οποίοι εξάγουν συμπεράσματα για μεμονωμένες οντότητες ή ομάδες, τα οποία σχετίζονται με ευαίσθητα προσωπικά δεδομένα τους όπως είναι οι πολιτικές πεποιθήσεις ή ο σεξουαλικός τους προσανατολισμός. Όμως ένας από τους βασικότερος στόχους της επεξεργασίας των Big Data και ιδιαίτερα στα περιβάλλοντα Μηχανικής Μάθησης είναι η χρήση των υφιστάμενων δεδομένων για νέους σκοπούς. Αυτό αυξάνει την ικανότητα των αλγορίθμων να εξάγουν συμπεράσματα αλλά ταυτόχρονα δημιουργεί αμφιβολίες σχετικά με τη μελλοντική χρήση των δεδομένων αυτών και συνεπώς τη μη διαγραφή τους. Ιδιαίτερα στην περίπτωση των ανωνυμοποιημένων δεδομένων, η συσχέτιση μιας ιδιότητας που προκύπτει από την επεξεργασία των Big Data με ένα ευαίσθητο χαρακτηριστικό μπορεί να προκαλέσει σοβαρούς κινδύνους συνδεσιμότητας και να πλήξει ανεπανόρθωτα την ιδιωτικότητα των υποκειμένων. Για να γίνει αυτό κατανοητό, αρκεί να σκεφτεί κανείς τη σύνδεση ενός γνωστού ευαίσθητου δεδομένου ενός υποκειμένου ή ομάδας υποκειμένων δεδομένων με υποτιθέμενα ανώνυμα δεδομένα, μέσω της οποίας παράγονται συμπεράσματα.

Συμπερασματικά, ενώ η ανωνυμοποίηση είναι μία αποδεκτή τεχνική για την προστασία της ιδιωτικότητας, για την περίπτωση των Big Data τα πράγματα είναι πιο περίπλοκα. Υπάρχουν λύσεις οι οποίες έχουν αναπτυχθεί προκειμένου να ταιριάζουν με τις ιδιότητες των Big Data, αλλά εξακολουθούν να υπάρχουν ανοιχτές προκλήσεις όπως είναι η μείωση της πολυπλοκότητας αλλά και του χρόνου με τον οποίο επεξεργάζονται τα δεδομένα ενώ παράλληλα σέβονται την ιδιωτικότητα των υποκειμένων των δεδομένων. Η ανάπτυξη των Big Data, λοιπόν, έχει μεγάλο αντίκτυπο στο απόρρητο των δεδομένων καθώς οι εγγενείς ιδιότητές τους καθιστούν δύσκολη τη διασφάλιση της ιδιωτικότητας. [14]

6.3. Ψευδωνυμοποίηση

Με βάση τον ΓΚΠΔ, η ψευδωνυμοποίηση αποτελεί ένα μέτρο το οποίο θα μπορούσε να προστατέψει την ιδιωτική ζωή των ατόμων ήδη από τη φάση του σχεδιασμού. Σύμφωνα με τον ΓΚΠΔ, ως ψευδωνυμοποίηση νοείται «η επεξεργασία δεδομένων προσωπικού χαρακτήρα κατά τρόπο ώστε τα δεδομένα να μην μπορούν πλέον να αποδοθούν σε συγκεκριμένο υποκείμενο των δεδομένων χωρίς τη χρήση συμπληρωματικών πληροφοριών, εφόσον οι εν λόγω συμπληρωματικές πληροφορίες διατηρούνται χωριστά και υπόκεινται σε τεχνικά και οργανωτικά μέτρα προκειμένου να διασφαλιστεί ότι δεν μπορούν να αποδοθούν σε ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο». Τα ψευδωνυμοποιημένα δεδομένα παραμένουν «δεδομένα προσωπικού χαρακτήρα», καθώς η απλή αφαίρεση των μοναδικών αναγνωριστικών στοιχείων δεν είναι αρκετή για την

προστασία από τους κινδύνους εκ νέου ταυοποίησης, και συνεπώς υπόκεινται στις απαιτήσεις του ΓΚΠΔ. Ωστόσο, ο ΓΚΠΔ παρέχει ορισμένα ρυθμιστικά κίνητρα για την υιοθέτηση της ψευδωνυμοποίησης και, ως εκ τούτου, υπάρχουν ορισμένα σημαντικά οφέλη από τη χρήση της. Συγκεκριμένα, η ψευδωνυμοποίηση δεδομένων μπορεί να βοηθήσει έναν οργανισμό να ανταποκριθεί σε ορισμένες από τις απαιτήσεις του ΓΚΠΔ, αλλά δεν απαλλάσσει πλήρως τον οργανισμό από αυτές. Αξίζει να σημειωθεί πως η διαδικασία της ψευδωνυμοποίησης είναι αναστρέψιμη και υπάρχει μεγάλη πιθανότητα επαναπροσδιορισμού της ταυτότητας των υποκειμένων των δεδομένων. Η ψευδωνυμοποίηση μειώνει τις πιθανότητες εκ νέου ταυτοποίησης, αλλά δεν την αποτρέπει σε περίπτωση διαρροής δεδομένων. Η μέθοδος αυτή δεν παρέχει επαρκή προστασία από την κατάχρηση των δεδομένων για το άτομο και θα πρέπει να χρησιμοποιείται ως εσωτερική μέθοδος μετριασμού του κινδύνου.

Τόσο η ψευδωνυμοποίηση όσο και η ανωνυμοποίηση περιλαμβάνουν μεθόδους απόκρυψης δεδομένων ώστε να μη γίνεται αντιληπτή η ταυτότητα του υποκειμένου των δεδομένων. Η διαφορά μεταξύ των δύο έγκειται στο γεγονός ότι η ψευδωνυμοποίηση ενέχει περισσότερες πιθανότητες επαναπροσδιορισμού του υποκειμένου των δεδομένων σε σχέση με την ανωνυμοποίηση. Αυτό οδηγεί και στην ουσιαστική διαφορά των δύο υπό το πρίσμα του ΓΚΠΔ όπου τα ψευδωνυμοποιημένα δεδομένα εξακολουθούν να θεωρούνται «δεδομένα προσωπικού χαρακτήρα» σε αντίθεση με τα ανωνυμοποιημένα δεδομένα που δεν εμπίπτουν στο πεδίο εφαρμογής του ΓΚΠΔ.

Οι τεχνικές ψευδωνυμοποίησης των δεδομένων κατηγοριοποιούνται ανάλογα με τον τρόπο με τον οποίο παράγονται τα ψευδώνυμα σε:

- Τεχνικές στις οποίες το ψευδώνυμο είναι ανεξάρτητο από τα αρχικά δεδομένα - π.χ. tokenization (το ψευδώνυμο -token- , παράγεται ανεξάρτητα από τα αρχικά δεδομένα και μπορεί να είναι τυχαία παραγόμενοι αριθμοί) και
- Τεχνικές κατά τις οποίες παράγεται ψευδώνυμο από τα αρχικά δεδομένα - π.χ. ψευδωνυμοποίηση με κρυπτογράφηση ή ψευδωνυμοποίηση με συνάρτηση κατακερματισμού. [15]

Οι σύγχρονες τεχνολογίες κυβερνοασφάλειας, προκειμένου να ανταποκριθούν στις σύγχρονες απαιτήσεις, εξαρτώνται σε μεγάλο βαθμό στη συσχέτιση συμβάντων που αποκαλύπτουν την ύπαρξη απειλών, την εκπαίδευση συστημάτων μηχανικής μάθησης προκειμένου να ταξινομηθούν οι απειλές και τη δημιουργία μέτρων ασφάλειας με βάση τα προφίλ απειλών που δημιουργούνται. Έτσι, οι τεχνολογίες κυβερνοασφάλειας βασίζονται σε μεγάλο βαθμό στην επεξεργασία των Big Data. Στο πλαίσιο αυτό θα μπορούσε να χρησιμοποιηθεί η ψευδωνυμοποίηση, προκειμένου να διασφαλίσει την επεξεργασία των δεδομένων, διατηρώντας την ιδιωτικότητα των υποκειμένων των δεδομένων και την προστασία των δεδομένων.

Αρκετά συστήματα Μηχανικής Μάθησης που αναπτύσσονται σήμερα χρησιμοποιούν τη συλλογική γνώση -η οποία αποδεικνύεται πως είναι καλύτερη από τη γνώση των λίγων- προκειμένου να καλύψει επαρκώς ένα τεράστιο χώρο όπως τα URLs και τα κατεβασμένα αρχεία. Για το σκοπό αυτό, σχεδιάζονται συστήματα φήμης (Reputation Systems) τα οποία προσπαθούν να αποδώσουν μία βαθμολογία φήμης σε μία οντότητα (π.χ. σε μία διεύθυνση URL ή ένα υποψήφιο αρχείο λήψης) συλλέγοντας και συσχετίζοντας δεδομένα με την εν λόγω οντότητα. Για παράδειγμα, εάν η λήψη ενός συγκεκριμένου αρχείου έχει συσχετιστεί με έναν αριθμό ύποπτων ή κακόβουλων αποτελεσμάτων (π.χ. μολυσμένους υπολογιστές), τότε ένα σύστημα φήμης μπορεί να καταγράψει αυτή τη συσχέτιση και να τη

χρησιμοποιήσει για να προειδοποιήσει τους χρήστες ανάλογα. Αυτή η διαδικασία μπορεί να είναι αποτελεσματική μόνο εάν αναλυθούν μεγάλοι πληθυσμοί συνόλων δεδομένων και συσχετιστούν, προκειμένου να εκπαιδευτεί κατάλληλα ένα μοντέλο μηχανικής μάθησης.

Ας αναλύσουμε την περίπτωση στην οποία μία εταιρεία κυβερνοασφάλειας αναλύει αρχεία προκειμένου να τα κατατάξει ως κακόβουλα ή μη. Σε πολλές περιπτώσεις, όμως, ο διαχωρισμός δεν είναι απλός. Υπάρχουν αρχεία τα οποία εμπίπτουν σε μία «γκρίζα ζώνη», όπου δεν είναι ξεκάθαρο εάν είναι εντελώς κακόβουλα ή μη. Θα μπορούσε κανείς να φανταστεί τα αποτελέσματα της ανάλυσης ως μία βαθμολογία με κλίμακα από το 0 έως το 9, όπου το μεσαίο εύρος είναι απροσδιόριστο. Θα μπορούσε κανείς να επιλέξει να μην καταδικάσει τέτοια αρχεία ως κακόβουλα, διακινδυνεύοντας την πιθανότητα να αμελήσει πολλές από τις απειλές. Από την άλλη, η ταξινόμηση των αρχείων του μεσαίου εύρους ως κακόβουλα, μπορεί να οδηγήσει σε άδικες διακρίσεις καθιστώντας ολόκληρο το σύστημα μη χρηστικό.

Ένα σύστημα φήμης των αρχείων στοχεύει στην κατάταξη του αρχείου με βάση τα χαρακτηριστικά φήμης του – σε αντίθεση με την παραδοσιακή ανάλυση των αρχείων. Τα χαρακτηριστικά φήμης ενός αρχείου περιλαμβάνουν ιδιότητες όπως:

- Με ποια άλλα αρχεία έχει εγκατασταθεί το συγκεκριμένο αρχείο,
- Τί κακόβουλα αρχεία συνυπάρχουν με το συγκεκριμένο αρχείο,
- Ποια είναι η κατάσταση του υπολογιστή στον οποίο εγκαθίσταται το συγκεκριμένο αρχείο (π.χ. με βάση των αριθμό των περιστατικών που έχουν παρατηρηθεί),
- Πόσοι υπολογιστές που περιέχουν το συγκεκριμένο αρχείο βρέθηκαν να είναι μολυσμένοι, κλπ.

Συνδυάζοντας αυτούς (και όχι μόνο αυτούς) τους παράγοντες, ένα σύστημα φήμης αρχείων μπορεί να εντοπίσει συσχετίσει που επιτρέπουν στο σύστημα να παράγει βαθμολογία για το εν λόγω αρχείο που εμπίπτει στην «γκρίζα ζώνη».

Στην περίπτωση του συστήματος φήμης, λοιπόν, η ψευδωνυμοποίηση μπορεί να εφαρμοστεί σε δύο φάσεις: είτε στη φάση της εκπαίδευσης (όπου θα χρειάζεται να ψευδωνυμοποιηθεί ένα μεγάλο πλήθος δεδομένων).

- Κατά τη διάρκεια της φάσης της εκπαίδευσης τα ψευδωνυμοποιημένα δεδομένα που συλλέγονται από τον υπεύθυνο επεξεργασίας χρησιμοποιούνται προκειμένου να εκπαιδευτούν οι αλγόριθμοι που χρησιμοποιούνται για την βαθμολόγηση. Αυτό απαιτεί ένα σύνολο δεδομένων το οποίο θα χρησιμοποιηθεί από τον εκτελών την επεξεργασία για τη δημιουργία γραφημάτων, τα μοντέλα Μηχανικής Μάθησης και όλα τα υπόλοιπα αλγοριθμικά εργαλεία. Σε γενικές γραμμές, η φάση της εκπαίδευσης αποτελείται από δύο επαναληπτικά βήματα: α) τον υπολογισμό των παραμέτρων του μοντέλου και β) τον έλεγχο της ακρίβειας του συστήματος. Όσο η ακρίβεια του συστήματος δεν είναι ικανοποιητική, γίνεται επιστροφή στο βήμα α). Αν και το βήμα α) θα μπορούσε να εκτελεστεί με ψευδωνυμοποιημένα δεδομένα, ο έλεγχος της ακρίβειας στη συνέχεια θα απαιτούσε την επαλήθευση της ορθότητας των αποτελεσμάτων. Για να εκτελεστούν αυτοί οι έλεγχοι, ο εκτελών την επεξεργασία θα πρέπει να κοινοποιήσει τα αποτελέσματα στον υπεύθυνο επεξεργασίας, ο οποίος θα επαναπροσδιορίσει τα δεδομένα, θα υπολογίσει τις μετρικές ορθότητας/ακρίβειας και θα ενημερώσει κατάλληλα τον εκτελών την επεξεργασία. Αυτή η διαδικασία ελέγχου των αποτελεσμάτων με το σύνολο των

δεδομένων εκπαίδευσης χωρίς ταυτοποίηση των δεδομένων είναι απαραίτητος για την αποτελεσματικότητα του συστήματος και επαναλαμβάνεται μέχρι να επιτευχθεί η το επιθυμητό επίπεδο ακρίβειας.

- Αφού το σύστημα εκπαιδευτεί, μεταφέρεται στη φάση της παραγωγής. Όταν ένα ερώτημα σχετικά με το προαναφερθέν αρχείο υποβάλλεται στο σύστημα, δρομολογείται στον εκτελών την επεξεργασία. Κατά τον έλεγχο, εάν ο εκτελών την επεξεργασία κατατάξει το αρχείο ως κακόβουλο, τότε, ο αρχικός χρήστης χρειάζεται να λάβει την κατάλληλη ανατροφοδότηση. Ο εκτελών την επεξεργασία απαντά στον υπεύθυνο επεξεργασίας προκειμένου να αντιστρέψει την ψευδωνυμοποίηση, να ταυτοποιήσει τον εν λόγω χρήστη και να του παράσχει τις κατάλληλες πληροφορίες.

6.4. Προστασία των δεδομένων ήδη από το σχεδιασμό και εξ ορισμού

Το Άρθρο 25 του ΓΚΠΔ καλεί τους υπεύθυνους επεξεργασίας να λειτουργούν συνεχώς με στόχο την προστασία των δικαιωμάτων των υποκειμένων των δεδομένων, εφαρμόζοντας "κατάλληλα τεχνικά και οργανωτικά μέτρα". Η αιτιολογική σκέψη 78 απαριθμεί μερικά από τα μέτρα από αυτά και συγκεκριμένα αναφέρει ότι "Τέτοια μέτρα θα μπορούσαν να περιλαμβάνουν, μεταξύ άλλων, την ελαχιστοποίηση της επεξεργασίας δεδομένων προσωπικού χαρακτήρα, την ψευδωνυμοποίηση δεδομένων προσωπικού χαρακτήρα το συντομότερο δυνατόν, τη διαφάνεια όσον αφορά τις λειτουργίες και την επεξεργασία των δεδομένων προσωπικού χαρακτήρα, ώστε να μπορεί το υποκείμενο των δεδομένων να παρακολουθεί την επεξεργασία δεδομένων και να είναι σε θέση ο υπεύθυνος επεξεργασίας να δημιουργεί και να βελτιώνει τα χαρακτηριστικά ασφάλειας." .

Δεδομένου ότι η προστασία των δεδομένων ήδη από το σχεδιασμό αφορά την εφαρμογή τεχνικών και οργανωτικών μέτρων, στα αρχικά στάδια του σχεδιασμού των πράξεων επεξεργασίας, με τέτοιον τρόπο ώστε να διασφαλίζονται οι αρχές ιδιωτικού απορρήτου και προστασίας δεδομένων ήδη από την αρχή, μπορεί να αποδειχθεί ιδιαίτερα χρήσιμη για τα Big Data, επιτρέποντας την έγκαιρη εφαρμογή σχετικών ελέγχων που είναι να προστατεύονται τα προσωπικά δεδομένα των υποκειμένων, εξ ορισμού. Ωστόσο, στην περίπτωση τον Big Data, λόγω του τεράστιου όγκου των δεδομένων, της μεγάλης ποικιλομορφίας τους, αλλά και της ταχύτητας με την οποία αυτά μεταφέρονται και επεξεργάζονται, παρουσιάζονται σημαντικές προκλήσεις.

Αρχικά, προκειμένου να εντοπιστούν τάσεις και μοτίβα τα οποία θα οδηγήσουν στην εξαγωγή χρήσιμων συμπερασμάτων, τα διαθέσιμα σύνολα των δεδομένων θα πρέπει να περιλαμβάνουν όσο το δυνατό μεγαλύτερο πλήθος από Big Data. Ωστόσο, η ελαχιστοποίηση των δεδομένων και ο σαφής καθορισμός ορίων διατήρησης των δεδομένων, αποτελούν αναπόσπαστο κομμάτι της προστασίας της ιδιωτικότητας, όπως αυτό έχει αναλυθεί σε προηγούμενη ενότητα. Ως εκ τούτου, μπορεί να υποστηριχθεί ότι η ελαχιστοποίηση των δεδομένων έρχεται σε αντίθεση με τα Big Data, όπου μεγάλος όγκος δεδομένων συλλέγεται και αποθηκεύεται πριν περάσει στη φάση της επεξεργασίας. Επιπλέον, η επαναχρησιμοποίηση των ήδη επεξεργασμένων ή εγγενών δεδομένων, επηρεάζει τα υφιστάμενα μοντέλα συγκατάθεσης και ειδοποίησης των υποκειμένων, καθώς κατ΄ αυτόν τον τρόπο, η επεξεργασία των δεδομένων είναι αδιαφανής ως προς τα υποκείμενα των δεδομένων.

Παρόλο που η προστασία των δεδομένων ήδη από το σχεδιασμό φαίνεται να έχει ανοιχτές προκλήσεις στον κόσμο των Big Data, η ίδια μπορεί να αποτελέσει ένα ισχυρό εργαλείο για την προστασία της ιδιωτικότητας, χωρίς να παρεμποδίζεται το έργο και τα οφέλη που

προσφέρει η ανάπτυξη τέτοιων μεθόδων. Προκειμένου Η προστασία των δεδομένων ήδη από το σχεδιασμό επιτυγχάνεται με συγκεκριμένες στρατηγικές οι οποίες είναι οι εξής: [16]

- **Ελαχιστοποίηση**: Σύμφωνα με την οποία ο όγκος των προσωπικών δεδομένων που συλλέγονται θα πρέπει να περιορίζεται στο ελάχιστο δυνατό.
- **Απόκρυψη**: Τα προσωπικά δεδομένα καθώς και οι διασυνδέσεις αυτών με άλλα δεδομένα ή συμπεράσματα θα πρέπει να αποκρύπτονται ώστε να μην είναι προσπελάσιμα από μη εξουσιοδοτημένες οντότητες.
- **Διαχωρισμός**: Τα προσωπικά δεδομένα θα πρέπει να υποβάλλονται σε επεξεργασία με κατανεμημένο τρόπο, σε ξεχωριστά τμήματα, όποτε αυτό είναι εφικτό.
- **Συνάθροιση**: Τα προσωπικά δεδομένα θα πρέπει να υποβάλλονται σε επεξεργασία στο υψηλότερο επίπεδο συγκέντρωσης – δηλαδή να συναθροίζονται όσο το δυνατόν περισσότερα δεδομένα – και με τη μικρότερη δυνατή λεπτομέρεια με την οποία παραμένουν χρήσιμα.
- **Ενημέρωση:** Τα υποκείμενα των δεδομένων θα πρέπει να ενημερώνονται άμεσα οποτεδήποτε τα δεδομένα τους υφίστανται οποιοδήποτε είδος επεξεργασίας.
- **Έλεγχος:** Αφορά τον έλεγχο των υποκειμένων των δεδομένων επί της συλλογής, αποθήκευσης, λειτουργίας και μετάδοσης των πληροφοριών του.
- **Επιβολή:** Θα πρέπει να υπάρχει μια πολιτική απορρήτου συμβατή με τις νομικές απαιτήσεις η οποία θα πρέπει να επιβάλλεται.
- **Επίδειξη:** Οι υπεύθυνοι επεξεργασίας δεδομένων πρέπει να είναι σε θέση να αποδεικνύουν τη συμμόρφωση με την ισχύουσα πολιτική προστασίας της ιδιωτικής ζωής και τυχόν εφαρμοστέες νομικές απαιτήσεις.

Όταν εξετάζουμε τον κύκλο ζωής των Big Data, είναι σημαντικό να λαμβάνουμε υπόψη το σκοπό της κάθε φάσεις καθώς και όλους τους εμπλεκόμενους σε αυτές (υποκείμενο των δεδομένων, υπεύθυνοι επεξεργασίας, εκτελούντες την επεξεργασία, τρίτα μέρη). Με τον τρόπο αυτό είναι δυνατό να εντοπιστούν οι απαιτήσεις προστασίας την ιδιωτικότητας και τα σχετικά μέτρα που πρέπει να ληφθούν ανά φάση. Παρακάτω παρουσιάζονται οι στρατηγικές που επηρεάζουν κάθε μία από της φάσεις του κύκλου ζωής των Big Data:

Συλλογή δεδομένων:

Ελαχιστοποίηση: Μία από τις βασικές αρχές που συνδέονται με την φάση της συλλογής των Big Data είναι η ελαχιστοποίηση των δεδομένων. Κάθε υπεύθυνος εργασίας που συλλέγει δεδομένα οφείλει να προσδιορίσει ποια είναι τα προσωπικά δεδομένα που χρειάζεται να συλλεχθούν για το σκοπό της επεξεργασίας και να καθορίσει ένα χρονικό πλαίσιο διατήρησής τους. Τα προσωπικά δεδομένα που ξεπερνάνε τα απαραίτητα για το σκοπό της επεξεργασίας θα πρέπει να αποκλείονται από τη συλλογή μέσω ειδικά σχεδιασμένων διαδικασιών. Η εκτίμηση αντικτύπου (Privacy Impact Assessment – PIA) θα μπορούσε να αποτελέσει ένα χρήσιμο εργαλείο για τους υπεύθυνους επεξεργασίας, προκειμένου να καθορίσουν τα ακριβή δεδομένα που χρειάζεται να επεξεργαστούν ούτως ώστε αυτά να περιοριστούν στα απολύτως απαραίτητα για το σκοπό της επεξεργασίας.

Συνάθροιση: Επιπλέον, μια άλλη πτυχή που μπορεί να προκύψει ως αποτέλεσμα της εκτίμησης αντικτύπου, είναι η χρήση συναθροισμένων πληροφοριών έναντι των προσωπικών δεδομένων. Πράγματι, σε ορισμένες περιπτώσεις όπως είναι η στατιστική ανάλυση, τα προσωπικά δεδομένα μπορεί να μη χρειάζεται καν να συλλεχθούν εξ αρχής

και η συλλογή ανωνυμοποιημένων πληροφοριών να είναι επαρκής για να προκύψει κάποιο συμπέρασμα. Η πιο γνωστή λύση που εφαρμόζεται σε αυτές τις περιπτώσεις είναι η τοπική ανωνυμοποίηση (local anonymization), η οποία θα μπορούσε να επιτρέψει στον υπεύθυνο επεξεργασίας να αφαιρέσει όλες τις προσωπικές πληροφορίες από τα δεδομένα πριν αυτά περάσουν στη φάση της ανάλυσης.

Απόκρυψη: Η φάση της συλλογής των δεδομένων πολλές φορές μπορεί να λαμβάνει χώρα χωρίς το υποκείμενο των δεδομένων να είναι ενήμερο γι' αυτό. Χαρακτηριστικό παράδειγμα αποτελεί η συλλογή των δεδομένων που λαμβάνει χώρα μέσω της συνολικής συμπεριφοράς των ατόμων στο διαδίκτυο αλλά και των αναζητήσεών του στον διαδικτυακό ιστό. Ωστόσο, σήμερα υπάρχουν τεχνολογίες οι οποίες υποστηρίζουν την προστασία της ιδιωτικότητας των χρηστών του διαδικτύου οι οποίες περιλαμβάνουν μεταξύ άλλων τεχνολογίες κατά της παρακολούθησης (anti-tracking), κρυπτογράφησης και εργαλεία ασφαλούς διαμοιρασμού αρχείων.

Ενημέρωση: Τα υποκείμενα των δεδομένων πρέπει να ενημερώνονται άμεσα σχετικά με τη συλλογή των δεδομένων τους όταν αυτά χρησιμοποιούνται κατά την επεξεργασία των Big Data. Γι' αυτό, πρέπει να υπάρχουν κατάλληλοι μηχανισμοί ενημέρωσης αλλά και διαφάνειας. Τέτοια εργαλεία, ωστόσο πρέπει να είναι διαθέσιμα για όλες τις φάσεις του κύκλου ζωής των Big Data καθώς η διαφάνεια είναι αναπόσπαστο στοιχείο της σε όλες τις φάσεις επεξεργασίας αυτών και όχι μόνο κατά τη συλλογή τους. Ωστόσο, η φάση της συλλογής αποτελεί ένα κομβικό σημείο για το υποκείμενο των δεδομένων προκειμένου να αποφασίσει για τα προσωπικά του δεδομένα.

Έλεγχος: Η φάση της συλλογής είναι η φάση κατά την οποία το υποκείμενο των δεδομένων θα συναινέσει για την επεξεργασία των προσωπικών του δεδομένων. Η εφαρμογή των μηχανισμών opt in κρίνονται πρακτική και κρίσιμη σε αυτή την περίπτωση. Ωστόσο, θα πρέπει να προσφέρεται στα υποκείμενα των δεδομένων η δυνατότητα εξαίρεσης (opt out) σε οποιοδήποτε στάδιο της επεξεργασίας.

<u>Αποθήκευση δεδομένων</u>

Απόκρυψη: Τα μέτρα ασφάλειας όπως είναι ο έλεγχος πρόσβασης και οι μηχανισμοί αυθεντικοποίησης είναι απαραίτητα για την προστασία των προσωπικών δεδομένων όταν αυτά βρίσκονται αποθηκευμένα σε βάσεις δεδομένων. Τεχνολογίες όπως είναι ο έλεγχος πρόσβασης βάσει χαρακτηριστικών (Attribute Based Access Control) μπορεί να είναι κλιμακούμενες στα Big Data επιβάλλοντας ισχυρές πολιτικές ελέγχου πρόσβασης. Η κρυπτογράφηση, επίσης, αποτελεί βασικό στοιχείο για την προστασία των δεδομένων όταν αυτά βρίσκονται "at rest".

Διαχωρισμός: Οι αναλύσεις που διατηρούν την ιδιωτικότητα σε κατανεμημένα συστήματα είναι επίσης σημαντικές για την προστασία των προσωπικών δεδομένων, καθώς προβλέπουν υπολογισμούς σε διαφορετικές βάσεις δεδομένων χωρίς την ανάγκη κεντρικής αποθήκης. Τα μέτρα ελέγχου πρόσβασης και οι τεχνικές κρυπτογράφησης μπορούν και πάλι να υποστηρίξουν αυτού του είδους τις λύσεις.

<u>Ανάλυση και χρήση δεδομένων</u>

Συνάθροιση: Η δημοσίευση και η ανάκτηση δεδομένων με γνώμονα την προστασία της ιδιωτικότητας βασίζονται συνήθως στην ανωνυμοποίηση προκειμένου να αποτραπεί η εξαγωγή συμπερασμάτων για προσωπικά δεδομένα. Ζητήματα σχετικά με την προέλευση των δεδομένων κατά τη λήψη αποφάσεων (με βάση τα Big Data) είναι ένα άλλο θέμα

ενδιαφέροντος, ιδίως όσον αφορά την αξιοπιστία και το επίπεδο της συγκέντρωσης των μεταδεδομένων (ώστε να αποφεύγεται η ταυτοποίηση ατόμων).

Αξίζει να σημειωθεί πως η **Επιβολή** και η **Επίδειξη** εφαρμόζονται καθ' όλη τη διάρκεια της επεξεργασίας των Big Data προκειμένου οι υπεύθυνοι της επεξεργασίας να είναι σύννομοι με ό,τι ορίζεται από τον ΓΚΠΔ για την προστασία των δεδομένων. Τα αυτοματοποιημένα εργαλεία μέσω των οποίων ορίζονται οι πολιτικές και επιβάλλονται καθώς και τα εργαλεία συμμόρφωσης μπορούν να αποδειχθούν χρήσιμα υποστηρίζοντας τη λογοδοσία και την απόδοση ευθυνών. [17]

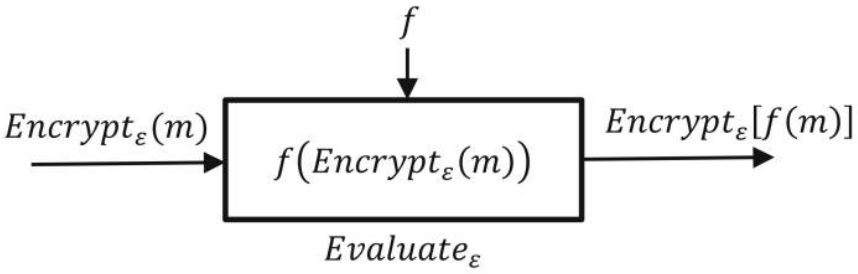
6.5. Κρυπτογράφηση

Ο ΓΚΠΔ, στο Άρθρο 32 περιλαμβάνει την κρυπτογράφηση των δεδομένων ως παράδειγμα τεχνικού μέτρου προκειμένου να διασφαλίζεται το κατάλληλο επίπεδο ασφάλειας. Η κρυπτογράφηση αποτελεί μία από τις θεμελιώδεις τεχνικές ασφάλειας, αν αναλογιστεί κανείς πως είναι γνωστή από αρχαιοτάτων χρόνων και δείγματα χρήσης της μπορεί κανείς να συναντήσει σε πολλά αρχαία κείμενα. Η τεχνολογία της κρυπτογράφησης είναι σχεδόν συνώνυμη με την εμπιστευτικότητα των δεδομένων καθώς κατέχει κυρίαρχο ρόλο σε αυτή αποτρέποντας τη μη εξουσιοδοτημένη αποκάλυψη των προσωπικών δεδομένων και πληροφοριών. Προκειμένου να αποτραπεί η μη εξουσιοδοτημένη αποκάλυψη των δεδομένων, αυτά επικαλύπτονται από ένα κλειδί κρυπτογράφησης, έτσι ώστε να μην μπορούν να αναγνωστούν και να γίνουν κατανοητά από άτομα που δεν έχουν πρόσβαση στο κλειδί.

Παραδοσιακά, το μέγεθος των δεδομένων ήταν σχετικά μικρό και οι κρυπτογραφικοί αλγόριθμοι σχεδιάστηκαν με βάση αυτό. Τα κρυπτογραφικά συστήματα εξαρτώνται από τα κρυπτογραφικά κλειδιά. Το κρυπτογραφικό κλειδί ανταλλάσσεται μεταξύ των μερών που συμμετέχουν στην επικοινωνία αποτρέποντας σε μη εξουσιοδοτημένα άτομα να έχουν πρόσβαση στην επικοινωνία και συνεπώς στα δεδομένα που μεταφέρονται. Αν και υπάρχουν διαφορετικές μέθοδοι κρυπτογράφησης ανάλογα με τα δεδομένα που πρέπει να κρυπτογραφηθούν, παρακάτω αναλύονται κάποιες τεχνικές οι οποίες θα μπορούσαν να έχουν εφαρμογή στην επεξεργασία των Big Data, όπου τα χαρακτηριστικά τους όπως ο μεγάλος όγκος τους, η ποικιλομορφία τους και η μεγάλη ταχύτητα με την οποία αυτά μεταφέρονται, αυξάνουν τις προκλήσεις που καλούνται να αντιμετωπίσουν οι τεχνικές κρυπτογράφησης. [18]

Ομομορφική Κρυπτογράφηση

Ο σκοπός της ομομορφικής κρυπτογράφησης είναι να επιτρέψει την εκτέλεση υπολογισμών σε κρυπτογραφημένα δεδομένα. Έτσι τα δεδομένα μπορούν να παραμείνουν εμπιστευτικά κατά την επεξεργασία τους, επιτρέποντας την επίτευξη χρήσιμων εργασιών με τα αυτά ενώ είναι αποθηκευμένα σε μη αξιόπιστα περιβάλλοντα. Στα μαθηματικά ο όρος «ομομορφικό» περιγράφει τον μετασχηματισμό ενός συνόλου δεδομένων σε ένα άλλο διατηρώντας παράλληλα τις σχέσεις μεταξύ των στοιχείων και στα δύο σύνολα. Επειδή τα δεδομένα σε ένα ομομορφικό σύστημα κρυπτογράφησης διατηρούν την ίδια δομή, πανομοιότυπες μαθηματικές λειτουργίες, είτε αυτές εκτελούνται σε κρυπτογραφημένα ή σε μη κρυπτογραφημένα δεδομένα, θα οδηγήσουν σε ισοδύναμα αποτελέσματα. [19]



Εικόνα 6: Σχήμα κρυπτογράφησης

Η ομομορφική κρυπτογράφηση ανάλογα με τους υπολογισμούς που πραγματοποιούνται διακρίνεται σε:

- Μερικώς ομομορφική κρυπτογράφηση (Partially Homomorphic Encryption - PHE): Αφορά τα κρυπτοσυστήματα στα οποία επιτρέπεται η εκτέλεση μιας πράξης σε κρυπτογραφημένα δεδομένα. Η πράξη αυτή μπορεί να είναι είτε πρόσθεση είτε πολλαπλασιασμός και η πράξη αυτή μπορεί να εκτελεστεί απεριόριστες φορές. [20]
- Κάπως ομομορφική κρυπτογράφηση (SomeWhat Homomorphic Encryption - SWHE): Αυτά τα συστήματα επιτρέπουν έναν περιορισμένο αριθμό είτε πρόσθεσης είτε πολλαπλασιασμού να εκτελείται σε κρυπτογραφημένα δεδομένα. [20]
- Πλήρης ομομορφική κρυπτογράφηση (Fully Homomorphic Encryption -FHE): Αυτό το ομομορφικό σύστημα κρυπτογράφησης επιτρέπει τόσο την πράξη της πρόσθεσης όσο και του πολλαπλασιασμού σε κρυπτογραφημένα δεδομένα. [20]

Προκειμένου να υλοποιηθούν και να αναπτυχθούν βιβλιοθήκες ομομορφικής κρυπτογράφησης, έχουν γίνει πολυάριθμες προσπάθειες τόσο από ακαδημαϊκές κοινότητες όσο και από βιομηχανικές εταιρείες. Φαίνεται ότι η πλήρης ομομορφική κρυπτογράφηση είναι πολύ κοντά στις απαιτήσεις του πραγματικού κόσμου για μαζικές υλοποιήσεις και μπορούν να μετριάσουν τις δυσκολίες που προκύπτουν από την επεξεργασία και την ανάλυση των Big Data και ιδιαίτερα σε περιβάλλοντα Μηχανικής Μάθησης. Στον παρακάτω πίνακα παρατίθενται τα κυριότερα εργαλεία ομομορφικής κρυπτογράφησης. Οι περισσότερες βιβλιοθήκες έχουν σχεδιαστεί για κρυπτογράφους και αναλυτές δεδομένων και παρέχουν δυνατότητες οι οποίες μπορούν να διαμορφωθούν κατάλληλα ώστε να διερευνηθεί το σχήμα της ομομορφικής κρυπτογράφησης και οι παράμετροί του [21].

Βιβλιοθήκη	Συντάκτης & Σχήμα	Γλώσσα	Αρχική έκδοση	Τελευταία σημαντική ενημέρωση
HElib	Halevi, Shoup (IBM), BGV, CKKS	C++	Μάιος 2013	Αύγουστος 2019
HEAAN	Cheon, Kim, Kim, Song, CKKS	C++	Μάιος 2016	Ιούλιος 2021
PALISADE	NJIT, BFV, BGV, and CKKS	C++	Ιούλιος 2017	Αύγουστος 2021

Βιβλιοθήκη	Συντάκτης & Σχήμα	Γλώσσα	Αρχική έκδοση	Τελευταία σημαντική ενημέρωση
TFHE	Chillotti et al., TFHE	C++	Απρίλιος 2017	Φεβρουάριος 2020
Microsoft SEAL	Microsoft, BFV, CKKS	C++	Δεκέμβριος 2018	Νοέμβριος 2020
NuFHE	NuCypher, GPU based TFHE	Python	Οκτώβριος 2018	Ιούλιος 2019

Οι κυριότερες και βιβλιοθήκες που βρίσκουν εφαρμογή σήμερα στην ανάλυση των Big Data είναι οι εξής:

- **SEAL**: Πρόκειται για μία βιβλιοθήκη ανοικτού κώδικα, υπό την άδεια MIT, με σκοπό την υλοποίηση συστημάτων ομομορφικής κρυπτογράφησης. Η ανάπτυξή της ξεκίνησε από τη Microsoft και η πρώτη της έκδοση δημοσιεύτηκε το Δεκέμβριο του 2018. Έχει αναπτυχθεί σε γλώσσα C++ δίχως εξάρτηση σε εξωτερικές βιβλιοθήκες, ώστε να μπορεί να εγκατασταθεί και να χρησιμοποιηθεί εύκολα ανεξαρτήτως πλατφόρμας. Η SEAL υλοποιεί δύο ξεχωριστά συστήματα ομομορφικής κρυπτογράφησης:
  - **BFV**: Εκτελεί πράξεις αριθμητικής υπολοίπων σε κρυπτογραφημένους ακεραίους.
  - **CKSS**: Εκτελεί προσθέσεις και πολλαπλασιασμούς σε κρυπτογραφημένους πραγματικούς ή μιγαδικούς αριθμούς, αλλά παράγει μόνο προσεγγιστικά αποτελέσματα.

  Το σύστημα BFV αποτελεί λοιπόν τη μόνη επιλογή όταν απαιτούνται ακριβείς τιμές. Από την άλλη, σε εφαρμογές όπως η άθροιση κρυπτογραφημένων πραγματικών αριθμών, η αποτίμηση μοντέλων μηχανικής μάθησης πάνω σε κρυπτογραφημένα δεδομένα, ή ο υπολογισμός αποστάσεων μεταξύ κρυπτογραφημένων τοποθεσιών, το σύστημα CKKS είναι μακράν η καλύτερη επιλογή. [22]

- **HElib:** Πρόκειται για μία πλατφόρμα ανοικτού κώδικα της IBM που διευκολύνει διάφορα είδη ομομορφικής κρυπτογράφησης. Έχει αναπτυχθεί, επίσης σε γλώσσα C++ και υλοποιεί διάφορα συστήματα ομομορφικής κρυπτογράφησης όπως BGV και CKKS. [21].

- **PALISADE:** Πρόκειται για μία βιβλιοθήκη λογισμικού ομομορφικής κρυπτογράφησης η οποία είναι επίσης ανοικτού κώδικα. Βασίζεται στη γλώσσα C++ και υποστηρίζει διάφορους αλγορίθμους ομομορφικής κρυπτογράφησης όπως BFV BGV και CKKS. [21].

Λόγω της ποικιλομορφίας, του μεγάλου όγκου, της σημασίας και της πολυπλοκότητας των Big Data, απαιτούνται συστήματα επεξεργασίας δεδομένων με συνεχώς αυξανόμενες υπολογιστικές ικανότητες. Επιπλέον, η ομομορφική κρυπτογράφηση παρέχει αμετάβλητες, ασφαλείς και διαφανείς μεταφορές δεδομένων που απαιτούν μεγαλύτερη επεξεργαστική ισχύ. Ωστόσο, όταν τα εξελιγμένα Big Data κρυπτογραφούνται με έναν αλγόριθμο ομομορφικής κρυπτογράφησης, το σύστημα πάσχει από απροσδόκητη υπολογιστική πολυπλοκότητα, με αποτέλεσμα κακές επιδόσεις. Ως εκ τούτου, η δομή του αλγορίθμου ομομορφικής κρυπτογράφησης θα πρέπει να βελτιωθεί περαιτέρω ώστε να

ανταποκρίνεται στις περιπτώσεις χρήσης και στην προσαρμοστική κρυπτογράφηση πολλαπλών κλειδιών.

7. Συμπεράσματα

Όσο η τεχνολογία εξελίσσεται, ο όγκος και το είδος των δεδομένων που συλλέγει και επεξεργάζεται αυξάνεται με αποτέλεσμα να εγείρονται αρκετά ερωτηματικά σχετικά με θέματα που άπτονται της προστασίας των προσωπικών δεδομένων αλλά και της ιδιωτικότητας. Εφαρμογές Μηχανικής Μάθησης, μέσω της ανάλυσης των Big Data δημιουργούν ολόκληρα προφίλ γύρω από τα υποκείμενα των δεδομένων και λαμβάνουν αυτοματοποιημένες αποφάσεις προκειμένου να προσφερθούν σε αυτά εξατομικευμένες υπηρεσίες. Η λήψη όμως των αποφάσεων από τις μηχανές, μπορεί να είναι αρκετά γρήγορη αφού περιορίζεται η ανθρώπινη παρέμβαση όμως ελλοχεύει κινδύνους για την προστασία των προσωπικών δεδομένων. Η έλλειψη διαφάνειας, η εκτεταμένη συλλογή δεδομένων καθώς και οι άδικες αποφάσεις που μπορεί να προκύψουν, είναι μερικές από τις προκλήσεις που καλούνται να αντιμετωπίσουν οι σύγχρονοι αλγόριθμοι Μηχανικής Μάθησης. Ο ίδιος ο ΓΚΠΔ παρουσιάζει μία σειρά από μέτρα τα οποία θα μπορούσαν να συμβάλλουν στην προστασία των δεδομένων. Ανάμεσα σε αυτά βρίσκεται η κρυπτογράφηση, η προστασία των δεδομένων ήδη από το σχεδιασμό, η ανωνυμοποίηση και η ψευδωνυμοποίηση.

Μπορεί ο ΓΚΠΔ, με τα μέτρα που προβάλει να μετριάζει κάποιους από τους κινδύνους σχετικά με την προστασία των προσωπικών δεδομένων και της ιδιωτικότητας, όμως, στην περίπτωση των Big Data οι προκλήσεις είναι μεγαλύτερες λόγω του όγκου, της ποικιλομορφίας και των γενικότερων χαρακτηριστικών των Big Data, όπως αυτά έχουν αναλυθεί. Η προστασία των προσωπικών δεδομένων είναι και παραμένει σημαντική και παρόλο που μερικές φορές φαίνεται να μη συμβαδίζει με τις τεχνολογικές εξελίξεις, είναι απαραίτητο να ενισχυθεί. Η εξέλιξη στον τομέα της τεχνολογίας και η ανάπτυξη των συστημάτων Μηχανικής Μάθησης, είναι σημαντικό να συνάδουν με το αναφαίρετο δικαίωμα του ανθρώπου στην ιδιωτική ζωή και να λειτουργούν με γνώμονα αυτό. Για το λόγο αυτό, είναι απαραίτητη η ύπαρξη ενός νομικού πλαισίου, που συμβαδίζει με τις τεχνολογικές εξελίξει προκειμένου να διασφαλίζεται η προστασία των προσωπικών δεδομένων και το δικαίωμα των υποκειμένων των δεδομένων στην ιδιωτικότητα.

Βιβλιογραφία

[1] B. K. Hackenberger, «Data by data, Big Data,» 2019.

[2] C. o. Europe, «What's AI?,» [Ηλεκτρονικό]. Available: https://www.coe.int/en/web/artificial-intelligence/what-is-ai.

[3] F. Rodríguez, D. Scheinker και R. A. Harrington, «Promise and Perils of Big Data and Artificial Intelligence in Clinical Medicine and Biomedical Research,» 2018.

[4] Ο. Π. Δ. Τ. Α. 29, «Γνώμη 4/2007 σχετικά με την έννοια του όρου 'δεδομένα προσωπικού».

[5] Ε. Επιτροπή, «Τι είναι τα δεδομένα προσωπικού χαρακτήρα;,» [Ηλεκτρονικό]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_el. [Πρόσβαση 2022].

[6] J. Purswani, «Infographic : GDPR Principles for Processing Personal Data,» Hubilo, 24 April 2018. [Ηλεκτρονικό]. [Πρόσβαση 2022].

[7] Lawspot.gr, «Lawspot,» 16 November 2017. [Ηλεκτρονικό]. [Πρόσβαση 2022].

[8] Ο. Ε. Γ. Τ. Π. Δ. Τ. Α. 29, «Κατευθυντήριες γραμμές για την αυτοματοποιημένη λήψη αποφάσεων και την κατάρτιση προφίλ για τους σκοπούς του κανονισμού 2016/679,» 2018.

[9] Information Commisioner's Office, [Ηλεκτρονικό]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/. [Πρόσβαση 2022].

[10] E. Jackson και C. Mendoza, «Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not,» 31 March 2020. [Ηλεκτρονικό]. [Πρόσβαση 2022].

[11] CNIL, «Algorithms and artificial intelligence: CNIL's report on the ethical issues,» 25 May 2018. [Ηλεκτρονικό]. Available: https://www.cnil.fr/en/algorithms-and-artificial-intelligence-cnils-report-ethical-issues. [Πρόσβαση 2022].

[12] S. Natakarnkitkul, «EXPLAIN THE MACHINE LEARNING MODEL | TOWARDS AI,» 18 December 2019. [Ηλεκτρονικό]. Available: https://pub.towardsai.net/show-me-the-black-box-3495dd6ff52c. [Πρόσβαση 2022].

[13] T. Z. Zarsky, «Incompatible: The GDPR in the Age of Big Data,» 2017.

[14] G. N.-A. Vicenç Torra, «Big Data Privacy and Anonymization,» 2017.

[15] H. Hamidovic, J. Kabil και E. Šehić, «EU General data protection regulation (GDPR) - Anonymisation and pseudonymisation in function of data protection,» 2019.

[16] I. S. a. P. Workshops, «A Critical Analysis of Privacy Design Strategies,» 2016.

[17] E. U. A. F. N. A. I. Security, «Privacy by design in big data,» 2015.

[18] C. Dhawale, «Cryptography in Big Data Security,» 2018.

[19] Α. Αραμπατζής, «Τι είναι η Ομομορφική Κρυπτογράφηση;,» [Ηλεκτρονικό]. Available: https://www.homodigitalis.gr/posts/5200. [Πρόσβαση 2022].

[20] M. H. D. M. A. A. a. C. T. G. Roger A. Hallman, «Homomorphic Encryption for Secure Computation on Big Data,» United States Department of Defense, SPAWAR Systems Center Pacific, San Diego, Ca, U.S.A., 2019.

[21] R. Hamza, A. Hassan, A. Ali, M. B. Bashir, S. M. Alqhtani, T. M. Tawfeeg και A. Yousif, «Towards Secure Big Data Analysis via Fully Homomorphic Encryption Algorithms,» 2022.

[22] Ι. Σομός, «Ομομορφική κρυπτογράφηση με τη χρήση της βιβλιοθήκης SEAL,» 2022.

[23] Στέγη της Ελληνικής Βιομηχανίας, «Ο νέος Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR),» 2018. [Ηλεκτρονικό].

[24] European Data Protection Board, Κατευθυντήριες γραμμές 2/2019 για την επεξεργασία δεδομένων προσωπικού χαρακτήρα σύμφωνα με το άρθρο 6 παράγραφος 1 στοιχείο β) του ΓΚΠΔ στο πλαίσιο της παροχής επιγραμμικών υπηρεσιών σε υποκείμενα δεδομένων, 2019.

[25] Οργανισμός Θεμελιωδών Δικαιωμάτων της Ευρωπαϊκής Ένωσης και Συμβούλιο της Ευρώπης, 2019. [Ηλεκτρονικό]. Available: https://fra.europa.eu/sites/default/files/fra_uploads/fra-coe-edps-2018-handbook-data-protection_el.pdf.