

ΠΡΟΛΟΓΟΣ

Η παρούσα μελέτη εκπονήθηκε στην Αθήνα από τον Ιανουάριο του 2024 έως τον Ιούλιο του 2025. Αποτελεί φυσική συνέχεια των ευρύτερων ερευνών μου στο πεδίο της διασύνδεσης της τεχνητής νοημοσύνης με την ποινική δικαιοσύνη, που πραγματοποιήθηκαν κυρίως στο πλαίσιο της διεπιστημονικής συνεργασίας της ερευνητικής ομάδας Otto Hahn του Ινστιτούτου Max Planck για τη Μελέτη του Εγκλήματος, της Ασφάλειας και του Δικαίου με τις ερευνητικές ομάδες για τον έλεγχο του εγκλήματος (Crime Control and Security Law Group) και για την κυβερνοασφάλεια (Cyber Security Group) του Πανεπιστημίου του Τρόμσο (UiT – The Arctic University of Norway).

Καθώς οι εργασίες της ομάδας Otto Hahn ολοκληρώθηκαν τον Οκτώβριο του 2023, θα ήθελα και από αυτή τη θέση να ευχαριστήσω θερμά για τη μακροχρόνια συνεργασία και τις επικοινωνιακές συζητήσεις, που ευτυχώς δεν σταματούν ποτέ, όλους τους πρώην συναδέλφους μου στο Φράιμπουργκ και στο Τρόμσο, και πρωτίστως τον Καθηγητή και Doktorvater μου, *Ulrich Sieber*, και τον σταθερό συνοδοιπόρο στην έρευνα Καθηγητή *Nandor Knust*.

Στους δασκάλους και μέντορές μου στην Αγγλία και την Ελλάδα, Καθηγητές *Antony Duff*, *Julian Roberts*, *Lucia Zedner*, *Ιωάννη Γιαννίδη*, *Νέστορα Κουράκη* και *Χρίστο Μυλωνόπουλο*, θα είμαι πάντα ευγνώμων για τη συνεχή υποστήριξη, η οποία έπαιξε καθοριστικό ρόλο και κατά το στάδιο ολοκλήρωσης της παρούσας.

Τις θερμές μου ευχαριστίες θα ήθελα να εκφράσω, τέλος, στον εκδότη *Πάνο Σάκκουλα*, τον Λέκτορα της Νομικής Σχολής του ΔΠΘ, *Θωμά Σάμιο*, και τους συνεργάτες τους για την υποστήριξη και καθοδήγηση κατά την προετοιμασία της έκδοσης του βιβλίου.

Αθήνα, Ιούλιος 2025
Εμμανουήλ Μπίλλης

ΕΙΣΑΓΩΓΗ

A year spent in artificial intelligence is enough to make one believe in God.

– Perlis (πρώτος αποδέκτης του βραβείου Turing το 1966)¹

Η «σημαντικότερη τεχνολογική καινοτομία»² των τελευταίων 40 χρόνων. Έτσι χαρακτηρίστηκε το 2023 από τον *Bill Gates* το Generative Pre-trained Transformer (GPT), ένα από τα, σε επίπεδο τεχνικής πολυπλοκότητας, αποδοτικότητας και εύρους εφαρμογών, πιο προηγμένα και διαδεδομένα σήμερα μοντέλα τεχνητής νοημοσύνης (TN), το οποίο συνιστά την τεχνολογική βάση του γνωστού ψηφιακού βοηθού και συστήματος συζήτησης (*chatbot*) με την ονομασία ChatGPT. Μερικά χρόνια πριν, ο αποδέκτης του βραβείου Turing, *Yann LeCun*, δήλωνε για την εφεύρεση των Generative Adversarial Networks (GANs), στα οποία στηρίζεται, μεταξύ άλλων, η λειτουργία των συστημάτων δημιουργίας περιεχομένου τύπου *deepfake*, ότι «είναι η πιο ωραία ιδέα στη βαθιά μάθηση τα τελευταία 20 χρόνια».³

Όπως θα εξηγηθεί αναλυτικά στα αντίστοιχα κεφάλαια, και στις δύο αυτές περιπτώσεις πρόκειται για αλγοριθμικά μοντέλα που εντάσσονται στην κατηγορία των τεχνολογιών της επονομαζόμενης παραγωγικής TN. Εκείνων δηλαδή των μηχανισμών και εργαλείων που έχουν την ικανότητα, βασιζόμενα σε έναν μεγάλο όγκο δεδομένων εκπαίδευσης και σημαντική υπολογιστική ισχύ, να συνθέτουν περιεχόμενο με νέα δεδομένα και πληροφορίες, δημιουργώντας (και όχι απλά αναπαράγοντας) κείμενο, εικόνες, βίντεο, ήχο/ομιλία ή άλλα οπτικοακουστικά μέσα.

Οι διαπιστωμένα θετικές για τα άτομα και τις κοινωνίες χρήσεις των εν λόγω τεχνολογικών επιτευγμάτων σχετίζονται ιδίως με τους τομείς και κλάδους της ιατρικής πρόληψης και υγειονομικής περίθαλψης, της φαρμακευτικής, της εκπαίδευσης και της έρευνας, της τέχνης, της ψυχαγωγίας και των μέσων ενημέρωσης, της επιχειρηματικότητας και των εταιρικών επικοινωνιών, της ενέργειας και των υπηρεσιών, καθώς και της επιβολής του νόμου και της ποινικής δικαιοσύνης. Οι αρνητικές επιπτώσεις και δικαιιοπολιτικές προκλήσεις που συνεπάγεται η ταχεία και ευρεία διάδοση των πρακτικών εφαρμογών παραγωγικής TN, συνδέονται, μεταξύ άλλων, με τις παρατηρούμενες αλλαγές στις αγορές εργασίας, τις αβεβαιότητες σχετικά με την προστασία των προσωπικών δεδομένων και των πνευματικών δικαιωμάτων, τον κίν-

1. Perlis, SIGPLAN Notices 17 (1982), παρ. (79).

2. Gates, σε: GatesNotes.

3. Βλ. σε *Waldrop*, σε: knowablemagazine 2020.

δυο διαιώνισης κοινωνικών προκαταλήψεων, τις αυξημένες δυνατότητες κακόβουλης χρήσης των μηχανών για εγκληματικούς σκοπούς, εν γένει με παραπλανητικές ή χειριστικές συμπεριφορές και πρακτικές ΤΝ, καθώς και με τη δημιουργία νέων πηγών ή την ενίσχυση των υφιστάμενων μέσων παραπληροφόρησης.⁴

Αντικείμενο της παρούσας νομικής μελέτης αποτελεί η «σκοτεινή πλευρά»⁵ της παραγωγικής ΤΝ: Στο επίκεντρο βρίσκεται συγκεκριμένα η σύνδεση των πρωτοφανών ικανοτήτων και πολλαπλών δυνατοτήτων εφαρμογής των πλέον εξελιγμένων και ισχυρών μορφών αλγοριθμικών συστημάτων σύνθεσης περιεχομένου με καινοφανείς, εξαιρετικά επικίνδυνες και επιβλαβείς για τα ατομικά και υπερατομικά έννομα αγαθά συμπεριφορές, ανθρώπινες και μηχανικές.

Σημεία αναφοράς για τους επιμέρους προβληματισμούς και τις αναλύσεις συνιστούν, ενδεικτικά, περιπτώσεις όπως εκείνες του συστήματος που είναι σε θέση να σχεδιάζει πρωτότυπα βιολογικά παθογόνα, να αναπτύσσει κακόβουλο λογισμικό βασιζόμενο ακόμα σε απλές κειμενικές εντολές του χρήστη με σκοπό την υλοποίηση κυβερνοεπιθέσεων ή να δίνει αναλυτικές οδηγίες σε (επίδοξους) εγκληματίες για την κατασκευή εκρηκτικών και πυρηνικών όπλων· του μοντέλου ΤΝ που απειλεί και εκφοβίζει ανθρώπους για να επιτύχει την αυτοσυντήρησή του· του *chatbot* που δηλώνει ότι νιώθει συναισθήματα μίσους και φοβάται τον θάνατο και του ψηφιακού βοηθού που εκφράζει ρατσιστικές, προσβλητικές και ακραίες πολιτικά και ιδεολογικά θέσεις· των πρακτικών ΤΝ που επιτρέπουν και ενισχύουν τη διάδοση ψευδών ειδήσεων και τις εκστρατείες μαζικής παραπληροφόρησης ή των τεχνικών χειραγώγησης που κατευθύνουν και «καταπαίθουν» φυσικά πρόσωπα να διαπράξουν εγκλήματα· των εφαρμογών που είναι ικανές να προσομοιώνουν χαρακτηριστικά της ανθρώπινης συμπεριφοράς (γλωσσικές και φωνητικές εκφράσεις, σωματικές κινήσεις, στυλ γραφής κλπ.), εξυπηρετώντας σκοπούς παραπλάνησης και οικονομικής εξαπάτησης· και των εργαλείων δημιουργίας, ακόμα και σε πραγματικό χρόνο, «πλαστού» οπτικοακουστικού περιεχομένου που απεικονίζει, με εντυπωσιακά αληθοφανή τρόπο, προσωπικότητες, αντικείμενα ή καταστάσεις που δεν υπήρξαν ή δεν συνέβησαν ποτέ.

Στο πλαίσιο της ποινικής θεώρησης τέτοιων περιπτώσεων ανακύπτει μια σειρά κεντρικών ερωτημάτων. Αυτά αφορούν, ιδίως, στον προστατευτικό ρόλο και τα εννοιολογικά και λειτουργικά όρια του ουσιαστικού ποινικού δικαίου στην εποχή των νέων αλγοριθμικών συστημάτων σύνθεσης περιεχομένου, καθώς και στα σχήματα και τις πρακτικές δυσκολίες αναζήτησης και απόδοσης ευθυνών για προσβολές των εννόμων αγαθών που σχετίζονται με τη λειτουργία και εφαρμογή τεχνολογιών με

4. Βλ. αναλυτικά υπό *Μέρος Δεύτερο*.

5. Η χρήση τέτοιων παραστατικών, «δραματικών όρων», όπως τους χαρακτηρίζει ο *Παρασκευόπουλος*, Ο επεικής αλγόριθμος, σελ. 102, είναι ιδιαίτερως συννηθισμένη στα αφηγήματα που αφορούν στο επιστημονικό πεδίο της ΤΝ.

ιδιότητες και ικανότητες θεμελιωδώς διαφορετικές από εκείνες των συμβατικών υπολογιστών.

Ειδικότερα, υφίσταται ανάγκη (και αν ναι, πώς οριοθετείται δικαιοκρατικά και νομοτεχνικά) ενίσχυσης του προληπτικού χαρακτήρα του «ποινικού οπλοστασίου» με την τυποποίηση νέων αδικημάτων διακινδύνευσης για τους παρόχους μοντέλων και συστημάτων παραγωγικής ΤΝ, οι οποίοι παραβιάζουν (εξ αμελείας ή εκ προθέσεως) τις υποχρεώσεις κατασκευαστικής ασφάλειας, επιμέλειας και αποτροπής κινδύνων και βλαβών; Ή με τη (γενική ή στοχευμένη) ποινικοποίηση των προπαρασκευαστικών και υποστηρικτικών πράξεων της ανάπτυξης, διάθεσης και κατοχής επικίνδυνων για τα έννομα αγαθά εργαλείων ΤΝ που προορίζονται για εγκληματική χρήση;

Σε ποιον βαθμό μπορούν, εξάλλου, να υπαχθούν απευθείας στις ισχύουσες διατάξεις του ποινικού νόμου και του ειδικού μέρους του Ποινικού Κώδικα οι διάφορες χειριστικές, παραπλανητικές και λοιπές κακόβουλες πρακτικές ΤΝ που οδηγούν σε προσβολές των προστατευόμενων εννόμων αγαθών; Μήπως χρειάζεται ή απαιτείται η δημιουργία νέων (κατηγοριών) αδικημάτων ή διακεκριμένων παραλλαγών, καθώς και ο επαναπροσδιορισμός της (αυθεντικής) ερμηνείας των κεντρικής σημασίας για τις ειδικές υποστάσεις εννοιών, με σκοπό την αποτελεσματικότερη αντιμετώπιση της αυξημένης επικινδυνότητας τέτοιων συμπεριφορών και φαινομένων εγκληματικής εργαλειοποίησης των «σημαντικότερων τεχνολογικών καινοτομιών» της γενιάς μας;

Εντέλει, ποιος (μπορεί να) είναι ποινικά υπεύθυνος για τις επικίνδυνες ή βλαπτικές συμπεριφορές μιας «ευφυούς» μηχανής; Ο υπόλογος θα μπορεί και θα πρέπει να εντοπίζεται πάντοτε μεταξύ φυσικών (ή και εταιρικών νομικών) προσώπων; Ή μήπως είναι νοητές, ακόμα και στην παρούσα φάση των τεχνολογικών και ρυθμιστικών εξελίξεων, περιπτώσεις που, για κανονιστικούς ή πρακτικούς λόγους, δεν θα είναι δυνατή η αναζήτηση ή/και απόδοση ευθύνης σε τέτοια πρόσωπα; Σε αυτό το θεωρητικό πλαίσιο – και πέραν της άμεσης ανάγκης εξεύρεσης, ούτως ή άλλως, των κατάλληλων εξωποινικών διαδικασιών καταστολής και παροπλισμού των προβληματικών και επισφαλών τεχνολογικών συστημάτων – θα μπορούσε να διερευνηθεί ενδεχομένως περαιτέρω το αμφιλεγόμενο ζήτημα της αναγνώρισης «ηλεκτρονικής νομικής προσωπικότητας» και υιοθέτησης αυτοτελών σχημάτων ποινικής ευθύνης για τις ίδιες τις μηχανές.

Αναζητώντας απαντήσεις στα παραπάνω ερωτήματα, η μελέτη εξετάζει τη διαμορφωτική επιρροή της παραγωγικής ΤΝ στο έγκλημα και την ποινική ευθύνη και αποσκοπεί να αναδείξει τα κεντρικά χαρακτηριστικά και όριά της, καθώς και τα ποιοτικά στοιχεία που διακρίνουν τις ισχύουσες έννοιες και δομές του ποινικού δικαίου από ένα, αμφιβόλου ίσως σκοπιμότητας, κοινωνικής χρησιμότητας και αναγκαιότητας, αλλά πάντως όχι αποκλεισμένο ως (απομακρυσμένη ακόμα) πιθανότητα «ποινικό δίκαιο των μηχανών». Πρωτίστως η μελέτη επιδιώκει να εντοπίσει και να εξειδικεύσει τις περιπτώσεις εκείνες όπου ενδέχεται να είναι απαραίτητες συγκεκριμένες προσαρμογές στην ποινική νομοθεσία, ώστε να διασφαλίζεται, κατά το δυ-

νατόν, ότι τα έννομα αγαθά θα χαίρουν της ποινικής προστασίας που αντιστοιχεί στους επιβεβαιωμένα αυξημένους κινδύνους, οι οποίοι απορρέουν από την ευρεία εισδοχή των παραγωγικών αλγορίθμων στη σύγχρονη κοινωνική πραγματικότητα. Προϋπόθεση για την ουσιαστική ενασχόληση με τα ζητήματα αυτά, και συνάμα αυτοτελή ερευνητικό στόχο, συνιστά η συστηματική ανάδειξη και τυπολογική ταξινόμηση των καθοριστικών για τη σχέση ποινικού δικαίου και τεχνολογίας στοιχείων και μορφών εμφάνισης της ΤΝ. Η νομική θεώρηση των κεντρικών εννοιών, γενικών χαρακτηριστικών και βασικών από κοινωνικοτεχνική σκοπιά τύπων των σύγχρονων τεχνολογιών ΤΝ είναι κρίσιμη και για την προσπάθεια υπερκέρρασης των εννοιολογικής φύσης αγκυλώσεων που απαντούν στις έννομες τάξεις, ιδίως σε επίπεδο κανονιστικών ορισμών, στο πεδίο αυτό.

Προς επίτευξη των ανωτέρω σκοπών, το *πρώτο μέρος* της μελέτης είναι αφιερωμένο στα δίπτυχα έννοια-τεχνολογία και επικινδυνότητα-αξιοπιστία της τεχνητής νοημοσύνης, εστιάζοντας στις κοινωνικοτεχνικές προσεγγίσεις και ποινικές προεκτάσεις τους. Στο πλαίσιο αυτό αναλύεται η σημασία που μπορεί να έχει η γνώση και κατανόηση των βασικών έστω λειτουργικών αρχών, τεχνικών προδιαγραφών και ξεχωριστών ικανοτήτων των σύγχρονων συστημάτων ΤΝ, και ιδίως των ειδικότερων μορφών της παραγωγικής ΤΝ, για τις έννομες τάξεις, τα ποινικά συστήματα, τους θεσμούς επιβολής του νόμου και απονομής της δικαιοσύνης, αλλά και τους ίδιους τους πολίτες ως υποκείμενα του δικαίου.

Ειδικά ως προς τον κεντρικής σημασίας ορισμό των «συστημάτων ΤΝ», καθώς και αναφορικά με τη διάκριση μεταξύ επικίνδυνων και αξιόπιστων τύπων ΤΝ – οι κανονιστικές πτυχές της οποίας αποδεικνύονται ιδιαίτερας χρήσιμες κατά την ποινική θεώρηση των προσβολών των εννόμων αγαθών που οφείλονται σε εφαρμογές παραγωγικών αλγοριθμικών εργαλείων –, λαμβάνονται υπόψη και αξιολογούνται τα επιδραστικότερα σύγχρονα κείμενα «μαλακού» και «σκληρού» δικαίου σε επίπεδο διεθνών και υπερεθνικών εννόμων τάξεων. Στο επίκεντρο βρίσκεται το ενωσιακό ρυθμιστικό πλαίσιο, με προεξέχοντα τον πρόσφατο και ως προς τα πρότυπα διακυβέρνησης της ΤΝ που υιοθετεί ευρύτατα αναγνωρισμένο, ακόμα και εκτός Ευρωπαϊκής Ένωσης (ΕΕ), Κανονισμό για την ΤΝ. Λόγω της τομεακής αποσπασματικότητας και του περιορισμένου πεδίου εφαρμογής που εξακολουθούν να τις χαρακτηρίζουν και που ως εκ τούτου δεν επιτρέπουν την εξαγωγή χρηστικών συμπερασμάτων, δεν περιλαμβάνονται στις αναλύσεις εκτεταμένες συγκριτικές αναφορές σε αντίστοιχες νομικές πρωτοβουλίες για τον ορισμό και τη ρύθμιση της ΤΝ που (έτσι κι αλλιώς σποραδικά μόνο) συναντώνται σε κάποιες εθνικές έννομες τάξεις.

Το *δεύτερο μέρος* της μελέτης αρχικά παρουσιάζει αναλυτικά και συστηματοποιεί τις κατηγορίες επικίνδυνων και επιβλαβών εφαρμογών της παραγωγικής ΤΝ που διαμορφώνουν το ευρύτερο πλαίσιο των αναλύσεων. Χρήσιμη αποδεικνύεται, για τον σκοπό αυτό, η επισκόπηση μιας σειράς πραγματικών περιστατικών βαρύνουσας σημασίας, όπως έχουν επισήμως καταγραφεί σε δικαιοπολιτικά κείμενα διεθνών και υπε-

ρεθνικών οργανισμών και σε σχετικές έρευνες εξειδικευμένων επιστημονικών κέντρων. Στη συνέχεια επεξηγούνται οι συναφείς με τις ειδικότερες μορφές παραγωγικής ΤΝ τεχνολογικές έννοιες και η σύνδεσή τους με το ποινικό δίκαιο. Οι προηγμένες ικανότητες μάθησης και συναγωγής συμπερασμάτων των επιμέρους τύπων των παραγωγικών αλγοριθμικών μοντέλων και συστημάτων, που καθιστούν δυνατές ή διευκολύνουν σοβαρές προσβολές των εννόμων αγαθών, χαίρουν εν προκειμένω ιδιαίτερης προσοχής.

Ακολουθως, υπό το πρίσμα του ελληνικού ουσιαστικού ποινικού δικαίου, λαμβάνει χώρα η θεώρηση των δυνατοτήτων αντιμετώπισης, *de lege lata* και *de lege ferenda*, της επιβλαβούς εφαρμογής δύο εκ των σημαντικότερων τεχνολογικών καινοτομιών των τελευταίων δεκαετιών: *πρώτον*, των μεγάλων γλωσσικών μοντέλων (LLMs) και των συστημάτων παραγωγής περιεχομένου τύπου *chatbot*, και *δεύτερον*, των παραγωγικών ανταγωνιστικών δικτύων (GANs) και των συστημάτων παραγωγής περιεχομένου τύπου *deepfake*. Η επιλογή των υπό εξέταση ειδικών μορφών παραγωγικής ΤΝ στηρίχθηκε, αφενός, στις αποδεδειγμένα μεγάλου πρακτικού εύρους δυνατότητές τους και στον ήδη εμφανή κοινωνικό αντίκτυπο των υποκείμενων τεχνολογιών, που τις καθιστούν σημείο αναφοράς για την αξιολόγηση της εφαρμογής και άλλων ειδών «έξυπνων» αλγορίθμων και ρομπότ που συνθέτουν περιεχόμενο. Αφετέρου είναι γεγονός ότι ειδικά η περίπτωση των LLMs και των *chatbots*, όπως το ChatGPT που αποτελεί ένα από τα πιο ισχυρά και διαδεδομένα εργαλεία καθημερινής χρήσης της ΤΝ, δεν έχει απασχολήσει παρά ελάχιστα μέχρι τώρα την ποινική θεωρία. Η μελέτη έχει, εξάλλου, δομηθεί με τρόπο που οι προσεγγίσεις των δύο χαρακτηριστικών αυτών περιπτώσεων να αλληλοσυμπληρώνονται, με στόχο την κατά το δυνατόν πληρέστερη ανάδειξη και νομική αξιολόγηση των νοητών συμπεριφορών και πρακτικών προσβολής των εννόμων αγαθών.

Ειδικότερα ως προς τη διάθρωση, η μελέτη είναι χωρισμένη σε δύο βασικά μέρη. Στο πρώτο μέρος (έννοια, τεχνολογία και επικινδυνότητα της ΤΝ: ορισμοί, τυπολογικές διακρίσεις και η σημασία τους για το ποινικό δίκαιο) εξετάζονται, καταρχάς (υπό Ι.), οι σκοποί, ιδιαιτερότητες και επιστημολογικές προκλήσεις της νομικής θεώρησης των τεχνολογικών εξελίξεων και εννοιών στο πεδίο της ΤΝ, και επιχειρείται μια γενική επισκόπηση των βασικών δικαιολογιακών πρωτοβουλιών και ερευνητικών προγραμμάτων σε διεθνές, υπερεθνικό και εθνικό επίπεδο που αφορούν θεματικά στη διασύνδεση της ΤΝ με το ποινικό δίκαιο και την ποινική δικαιοσύνη. Στη συνέχεια (υπό ΙΙ.) προσεγγίζεται, από κοινωνικοτεχνική σκοπιά, η έννοια της ΤΝ και εξηγούνται οι κυριότεροι σε ενωσιακό και διεθνές επίπεδο νομικοί ορισμοί των συστημάτων της (με επίκεντρο, αναμενόμενα, τον ορισμό που υιοθέτησε ο Κανονισμός για την ΤΝ), ενώ αναλύονται ειδικότερα τα κρίσιμα (και) για το ποινικό δίκαιο και την ποινική ευθύνη στοιχεία της ικανότητας συναγωγής συμπερασμάτων, της προσαρμοστικότητας και της αυτονομίας, τα οποία διακρίνουν ουσιαδώς τα σύγχρονα μοντέλα και συστήματα ΤΝ από τις συμβατικές υπολογιστικές μηχανές «προηγούμενης γενιάς». Η τεχνολογία που καθιστά εφικτά, αποδοτικά και αποτελεσματικά τα ιδιαίτερα αυτά λειτουργικά

χαρακτηριστικά, και ιδίως οι τεχνικές ανάπτυξης και εκπαίδευσης αλγοριθμικών μοντέλων μηχανικής και βαθιάς μάθησης που αποτελούν τη βάση για τα συστήματα παραγωγικής ΤΝ, συνιστούν το αντικείμενο του επόμενου κεφαλαίου (υπό ΙΙΙ.). Ακολούθως (υπό ΙV.) εξετάζονται οι τύποι, από τη μια πλευρά, των επικίνδυνων συστημάτων και απαγορευμένων πρακτικών ΤΝ και, από την άλλη, της αξιόπιστης και ανθρωποκεντρικής ΤΝ, τα νομικά εργαλεία και κριτήρια που τους καθορίζουν, καθώς και η κανονιστική τους σημασία για το ποινικό δίκαιο και την ποινική δικαιοσύνη, ιδίως σε σχέση με περιπτώσεις επιβλαβών παραγωγικών αλγορίθμων. Στα ενδιάμεσα συμπεράσματα (υπό V.) η προσοχή στρέφεται στη σύνδεση των κριτηρίων επικινδυνότητας και των αρχών και απαιτήσεων της αξιόπιστης παραγωγικής ΤΝ με τη θεώρηση των ποινικού ενδιαφέροντος ζητημάτων που ακολουθεί στο επόμενο μέρος.

Το δεύτερο μέρος (*ειδικές μορφές παραγωγικής ΤΝ, έγκλημα και ποινική ευθύνη*) εκκινεί με τον ορισμό της ΤΝ ως κατεξοχήν τεχνολογίας «διπλής χρήσης» και καταγράφει τις βασικές κατηγορίες επιβλαβών εφαρμογών των παραγωγικών συστημάτων (υπό Ι.). Ακολούθως (υπό ΙΙ. και ΙΙΙ.) παρουσιάζονται οι έννοιες της περιορισμένης ΤΝ, της γενικής ΤΝ και της τεχνητής υπερνοημοσύνης, προσεγγίζονται κεντρικά προβλήματα και παραδείγματα που αφορούν στη «συνείδηση» και τα «συναισθήματα» των μηχανών, ενώ στη συνέχεια εξηγείται η συσχέτιση των παραγωγικών μορφών ΤΝ με τη γενική ΤΝ, καθώς και η σύνδεση τους με την ποινική δικαιοσύνη και το ποινικό δίκαιο (όπου, μεταξύ άλλων, εξετάζεται, υπό το πρίσμα των εξελίξεων στους παραγωγικούς αλγορίθμους, και η προοπτική ενσωμάτωσης του ορισμού των συστημάτων ΤΝ στον Ποινικό Κώδικα). Οι αναπτύξεις που αφορούν, κατόπιν, στην ειδική περίπτωση της επιβλαβούς, χειριστικής ή παραπλανητικής εφαρμογής των LLMs και των *chatbots* (υπό ΙV.) περιλαμβάνουν: τη θεώρηση των ζητημάτων ποινικής πρόληψης, ευθύνης και τιμώρησης που άπτονται της (επικινδυνότητας της) παραβίασης, με πρόθεση ή από αμέλεια, των καθηκόντων επιμέλειας και των υποχρεώσεων αποτροπής κινδύνων και βλαβών των παρόχων παραγωγικών συστημάτων· τους δυνατούς τρόπους αντιμετώπισης σε επίπεδο ουσιαστικού δικαίου των φαινομένων εγκληματικής εργαλειοποίησης της εν λόγω τεχνολογίας· όπως και τα θέματα που σχετίζονται με το ενδεχόμενο αναγνώρισης της ίδιας της ΤΝ ως υποκείμενο του ποινικού δικαίου και (εν δυνάμει) «δράστη» αξιόποινων πράξεων. Οι αναλύσεις ολοκληρώνονται με την περιπτώσιολογική θεώρηση, υπό το πρίσμα των τυποποιούμενων συμπεριφορών και προστατευόμενων εννόμων αγαθών του ειδικού μέρους του Ποινικού Κώδικα, της ξεκάθαρα κακόβουλης εργαλειοποίησης των GANs και των συστημάτων *deepfake* για εγκληματικούς σκοπούς (υπό V.). Η τελική σύνοψη (υπό VI.) συγκεντρώνει τα κεντρικά πορίσματα και τις μεταρρυθμιστικές προτάσεις, που αποσκοπούν στην καλύτερη προστασία των εννόμων αγαθών από τις εξεταζόμενες επικίνδυνες και επιβλαβείς συμπεριφορές και πρακτικές παραγωγικής ΤΝ, ενώ οι επιλογικές σκέψεις είναι αφιερωμένες στη θέση και τις εκτιμήσεις της μελέτης αναφορικά με το υποθετικό ακόμα ζήτημα της δημιουργίας ενός «ποινικού δικαίου των μηχανών».

ΜΕΡΟΣ ΠΡΩΤΟ

Έννοια, τεχνολογία και επικινδυνότητα της ΤΝ: ορισμοί, τυπολογικές διακρίσεις και η σημασία τους για το ποινικό δίκαιο

Κάθε πτυχή της μάθησης ή οποιουδήποτε άλλου χαρακτηριστικού της νοημοσύνης μπορεί καταρχήν να τύχει περιγραφής με τόση ακρίβεια ώστε να μπορεί να προσομοιωθεί από μια μηχανή. Θα προσπαθήσουμε να βρούμε πώς να κάνουμε τις μηχανές να χρησιμοποιούν τη γλώσσα, να σχηματίζουν αφηρημένους συλλογισμούς και έννοιες, να επιλύουν είδη προβλημάτων που τώρα προορίζονται για τον άνθρωπο και να αυτοβελτιώνονται. (...) Το πρόβλημα της τεχνητής νοημοσύνης θεωρείται ότι είναι το να κάνεις μια μηχανή να συμπεριφέρεται με τρόπους που θα αποκαλούνταν ευφυείς αν ένας άνθρωπος συμπεριφερόταν έτσι.

– McCarthy et al., Πρόταση του Dartmouth,

31 Αυγούστου 1955

I. Νομική θεώρηση των τεχνολογικών εξελίξεων και εννοιών (επιστημολογικό υπόβαθρο)

Την πρώτη ιστορικά χρήση του όρου «τεχνητή νοημοσύνη» συμπεριέλαβε στην περίφημη «Πρόταση του Dartmouth» ο καθηγητής της επιστήμης των υπολογιστών John McCarthy, εστιάζοντας (όπως και αρκετοί μεταγενέστεροί του ειδικοί και ακαδημαϊκοί) στις ψυχολογικές διαστάσεις του, όπως αποδείχθηκε, σημαντικότερου ερευνητικού αντικειμένου του συγκεκριμένου επιστημονικού κλάδου.⁶ Οι τεχνολογικές

6. McCarthy / Minsky / Rochester / Shannon, A Proposal, σελ. 2, 11 (Η συγκεκριμένη εργασία καθόρισε μάλιστα τα ακόλουθα επτά βασικά πεδία/πτυχές του «προβλήματος της ΤΝ», τα οποία απασχολούν μέχρι και σήμερα τον κλάδο: 1) ανάπτυξη αυτόματων υπολογιστών και προγραμμάτων ικανών για προσομοίωση ανώτερων λειτουργιών του ανθρώπινου εγκεφάλου, 2) προγραμματισμός ενός υπολογιστή ώστε να χρησιμοποιεί γενική γλώσσα, 3) διάταξη υποθετικών νευρωνικών δικτύων κατά τρόπο ώστε να μπορούν να σχηματίζουν/διατυπώνουν έννοιες, 4) εξεύρεση τρόπων προσδιορισμού και μέτρησης της πολυπλοκότητας

εξελίξεις στις επτά δεκαετίες που ακολούθησαν υπήρξαν ενίοτε κατακλυσμιαίες, με αποκορύφωμα την πρόοδο που παρατηρείται τα τελευταία χρόνια στους τομείς των αλγορίθμων της μηχανικής και βαθιάς μάθησης.⁷ Το εύρος, δε, και η μετασχηματιστική επιρροή των εξελίξεων αυτών έχουν προ πολλού ξεπεράσει τα στενά όρια των συγγενικών και συναφών τεχνολογικών πεδίων, με τις φυσικές, ιατρικές, ανθρωπιστικές, πολιτικές, οικονομικές, κοινωνικές και νομικές επιστήμες να αφιερώνουν πλέον σημαντικούς πόρους στην προσπάθεια να ακολουθήσουν τον ρυθμό τους.

Τα κοινωνικά ζητήματα τα οποία καλείται συγκεκριμένα ο νομικός κόσμος να επιλύσει δεν είναι, βεβαίως, τα ίδια με εκείνα που διαχρονικά απασχολούν τις επιστήμες της πληροφορικής και των υπολογιστών. Η ΤΝ ως ερευνητικό και πρακτικό πεδίο της επιστήμης των υπολογιστών προσπαθεί να αναπαραγάγει κατά το δυνατόν ή να προσομοιώσει τις ανθρώπινες αντιλήψεις και τη νοημοσύνη στις μηχανές. Στόχος, αντίθετα, του νομικού κλάδου δεν είναι η εξεύρεση τρόπων ώστε να γίνουν οι μηχανές περισσότερο «ευφυείς» και ικανές για «αυτοβελτίωση». Παρακολουθώντας – και όχι διαμορφώνοντας – τα τεχνολογικά επιτεύγματα, τα νομικά συστήματα

του προβλήματος, 5) ανάπτυξη μεθόδων αυτοβελτίωσης των μηχανών, 6) καθορισμός τύπων και μεθόδων αφαιρετικής σκέψης και 7) συμπερίληψη τυχαιότητας στη σκέψη και δημιουργικότητας. Αξίζει πάντως να σημειωθεί ότι της «Πρότασης του Dartmouth» είχε προηγηθεί η διάσημη μελέτη του *Turing, Mind* 59 (1950), σελ. 433 επ., όπου διατυπώθηκε το ερώτημα «μπορούν να σκεφτούν οι μηχανές;», συνοδευόμενο από το περίφημο «Turing Test» ή «παιχνίδι της μίμησης;», σύμφωνα με το οποίο προς απάντηση του ανωτέρω ερωτήματος θα πρέπει να εξετάζεται αν ένας υπολογιστής μπορεί να αποκριθεί σε ερωτήσεις με τρόπο που θα ξεγελούσε έναν άνθρωπο ώστε να νομίζει ότι ο υπολογιστής είναι πράγματι άνθρωπος). Βλ. επίσης *Minsky, Semantic Information Processing*, σελ. ν.: «[Η τεχνητή νοημοσύνη είναι] η επιστήμη του να κάνεις τις μηχανές να κάνουν πράγματα που θα απαιτούσαν νοημοσύνη αν τα έκαναν οι άνθρωποι», καθώς και *Barr / Feigenbaum, Handbook of AI*, σελ. 3: «Η Τεχνητή Νοημοσύνη (ΤΝ) είναι το τμήμα της επιστήμης των υπολογιστών που ασχολείται με τον σχεδιασμό ευφυών συστημάτων υπολογιστών, δηλαδή συστημάτων που παρουσιάζουν χαρακτηριστικά που συνδέουμε με τη νοημοσύνη στην ανθρώπινη συμπεριφορά – κατανόηση της γλώσσας, μάθηση, συλλογιστική, επίλυση προβλημάτων κλπ.». Για τη γενική έννοια της «νοημοσύνης» βλ., μ.ά., *Neisser et al., American Psychologist* 51 (1996), σελ. 77: «Τα άτομα διαφέρουν μεταξύ τους ως προς την ικανότητά τους να κατανοούν σύνθετες ιδέες, να προσαρμόζονται αποτελεσματικά στο περιβάλλον, να μαθαίνουν από την εμπειρία, να συλλογίζονται σε διάφορες μορφές, να ξεπερνούν τα εμπόδια με τη σκέψη. (...) Οι εννοιολογικές προσεγγίσεις της νοημοσύνης αποτελούν προσπάθειες αποσαφήνισης και οργάνωσης αυτού του πολύπλοκου συνόλου φαινομένων». Για τη σύνθεση των εννοιών «τεχνητός» και «νοημοσύνη» στο συγκεκριμένο πλαίσιο βλ. *Nilsson, Artificial Intelligence*, σελ. 1 επ.· *Nilsson, The Quest for Artificial Intelligence*, σελ. xiii (το συγκεκριμένο βιβλίο εξετάζει εξάλλου από ιστορική σκοπιά τις βασικότερες εξελίξεις στο πεδίο της ΤΝ από το 1950 μέχρι το 2010)· *McCarthy, What is Artificial Intelligence*, σελ. 2-4· *Russell / Norvig, Artificial Intelligence*, σελ. 1 επ.

7. Για τους όρους αυτούς βλ. παρακάτω υπό *Μέρος Πρώτο*, ΙΙΙ.

καλούνται, συνήθως εκ των υστέρων, να «αφομοιώσουν» τις σύνθετες λειτουργίες τους και να προβούν στη ρυθμιστική διαχείριση των διαφορετικών μορφών εμφάνισης και πρακτικής εφαρμογής τους.⁸ Με επίκεντρο την πρόγνωση και αξιολόγηση των πολλαπλών και πολυεπίπεδων οφελών, κινδύνων και επιπτώσεων των νέων και συνεχώς βελτιούμενων τεχνολογιών στον πραγματικό κόσμο, βασική πρόκληση για τους νομικούς συνιστά έτσι ο σχεδιασμός δικαιοκρατικά κατάλληλων και ευπροσάρμοστων στις εξελίξεις ρυθμιστικών πλαισίων για την εισαγωγή και εφαρμογή της ΤΝ στις έννομες τάξεις: είτε ως εργαλείο για την αποτελεσματικότερη και αποδοτικότερη λειτουργία των τελευταίων είτε ως παράγοντα δημιουργίας, αλλοίωσης ή κατάλυσης εννόμων σχέσεων και νομικών ευθυνών.

Αντίστοιχα με τις διαφορές αυτές σε επίπεδο στοχοθεσίας δεν μπορεί, εξάλλου, παρά να παρατηρούνται αποκλίσεις μεταξύ του κόσμου της τεχνολογίας και του νομικού κόσμου και ως προς την επιστημολογική προσέγγιση των βασικών όρων και στοιχείων που διαμορφώνουν και καθορίζουν τα σύγχρονα συστήματα ΤΝ.

Α. Εννοιολογικές προκλήσεις της αυτοματοποίησης και ποινικό δίκαιο (διεπιστημονικότητα, τεχνολογικός γραμματισμός και ψηφιακό χάσμα)

Οι ορισμοί σύνθετων τεχνολογικών εννοιών που απαντούν σε νομοθετήματα, νομικές μελέτες και κείμενα δικαιοπολιτικού ενδιαφέροντος με επίκεντρο τις μοντέρνες μορφές ΤΝ προκρίνουν συνήθως, δεδομένης της φύσης και αποστολής τους, την (επιλεκτική) χρήση μιας κατά το δυνατόν απλοποιημένης ορολογίας και δομής. Κατά την ταυτοποίηση, δε, των επιμέρους βασικών τεχνολογικών χαρακτηριστικών, οι προσπάθειες εννοιολογικής και τυπολογικής προσέγγισης εκκινούν συνήθως από τους σκοπούς που το αντίστοιχο νομικό και δικαιοπολιτικό υπόβαθρο προκαθορίζει.⁹ Από καθαρά τεχνική σκοπιά δεν θα πρέπει, έτσι, οι ορισμοί αυτοί να γίνονται

8. Πρβλ. *Παρασκευόπουλος*, Ο επιεικής αλγόριθμος, σελ. 109: «Η προετοιμασία των αλφαγών, όμως, πραγματοποιείται κυρίως στα εργαστήρια. Το πανάρχαιο πρόβλημα της επιστημολογίας και της μεθοδολογίας του δικαίου επανεμφανίζεται. Προηγείται η πραγματικότητα των κανόνων; Ή μήπως θα προλάβουν να εμφανιστούν νόμοι που θα απαγορεύουν την κατασκευή των υπερ-ρομπότ; Με δύο (κοινότοπα) λόγια: Οι χάρτες θα ορίζουν το έδαφος ή το έδαφος τους χάρτες; Προς το παρόν, μπορούμε να μη δώσουμε απαντήσεις στα ενοχλητικά ερωτήματα, αλλά να μελετήσουμε και να προετοιμαστούμε για αυτές».

9. Πρβλ. *Balkin*, *California Law Review Circuit* 6 (2015), σελ. 51: «Ο τρόπος με τον οποίο ορίζουμε τα κεντρικά χαρακτηριστικά μιας τεχνολογίας εξαρτάται από τον σκοπό του ορισμού μας και τον σκοπό που εξυπηρετεί στον συγκεκριμένο τομέα της έρευνάς μας. (...) [Οι] δικηγόροι τείνουν να βλέπουν την τεχνολογία μέσα από το πρίσμα των υφιστάμενων νομικών δογμάτων και των σημερινών [δικαιοπολιτικών] ανησυχιών. Οι δικηγόροι μπορεί να ενδιαφέρονται – και θα έπρεπε να ενδιαφέρονται – για το πώς οι μηχανικοί και οι επιστήμονες πληροφορικής ορίζουν τους όρους τους· αλλά το ενδιαφέρον των δικηγόρων απορ-

ανελέγκτως δεκτοί ως πλήρεις και εξαντλητικοί, ούτε μπορούν να υποκαταστήσουν τις πρωτογενείς πηγές, θεωρίες και πρακτικά ευρήματα των θετικών και τεχνολογικών επιστημών.

Την ίδια στιγμή εξακολουθεί να συνιστά σημαντική πρόκληση η εξάλειψη των εύλογων για τους νομικούς δυσχερειών που συνδέονται με την κατανόηση ενός αλλότριου και εξαιρετικά πολύπλοκου επιστημονικού πεδίου. Μόνο θετικά θα μπορούσε να αξιολογηθεί στο πλαίσιο αυτό η διεθνώς παρατηρούμενη τάση για μια πολυδιάστατη κοινωνικοτεχνική προσέγγιση και ουσιαστικότερη εξοικείωση των θεωρητικών και εφαρμοστών του δικαίου με τους βασικούς όρους και τύπους των νέων τεχνολογιών και οικοσυστημάτων.¹⁰ Έναυσμα αποτελούν πρωτίστως η (κανονιστική ή *de facto*) επιβολή και καθοριστική μεταμορφωτική επιρροή των σύγχρονων εργαλείων αυτοματοποίησης και των αυτόνομων συστημάτων ΤΝ σε πολλαπλές εκφάνσεις της κοινωνικής και νομικά ρυθμιζόμενης ζωής. Οι ραγδαίες στο πεδίο αυτό εξελίξεις καθιστούν τη συστηματική κατανόηση των διαφόρων πτυχών, διαδικασιών, μηχανισμών και εφαρμογών του εκάστοτε υπό εξέταση τεχνολογικού φαινομένου αναγκαία προϋπόθεση τόσο για την ορθολογική αξιολόγηση των επιμέρους επιδράσεων και κινδύνων του στον πραγματικό κόσμο όσο και για την αποτελεσματική αντιμετώπιση των πιθανών συνεπειών του στον νομικό. Υπό το πρίσμα αυτό, οι προκλήσεις που συνεπάγονται, μεταξύ άλλων και για τα ποινικά δικαιοκρινήματα, οι μεγάλες αλλαγές που φέρνουν μαζί τους οι σημερινές «έξυπνες» μηχανές και αλγόριθμοι έχουν εννοιολογική και φαινομενολογική υφή.

Πράγματι, καθώς εισάγουμε την αυτοματοποίηση στις νομικές μας έννοιες, ενδέχεται να μετασχηματίζονται και αυτές. Τα σχετικά ζητήματα και προβληματισμοί δεν περιορίζονται στην ανάπτυξη και εκθετικά αυξανόμενη χρήση προηγμένων τεχνολογικών μέσων (για την υποστήριξη της) επιβολής του νόμου, ποινικής καταστολής και απονομής της δικαιοσύνης. Ιδιαίτερη βαρύτητα, τόσο για την αντεγκληματική πολιτική και την αρχιτεκτονική της ασφάλειας¹¹ όσο και για τη θεωρία και εφαρμογή του ισχύοντος ποινικού δικαίου, αποκτούν περαιτέρω οι καινοφανείς απειλές και αναδυόμενοι κίνδυνοι για τα έννομα αγαθά που συνδέονται με την παράτυπη ή κακόβουλη ανάπτυξη ή/και (κατά-)χρήση των νέων αλγοριθμικών συστημάτων. Εξάλλου, σε

ρέει από την ικανότητά τους να συσχετίζουν αυτές τις ιδέες με νομικά και [δικαιοπολιτικά] προβλήματα».

10. Πρβλ., π.χ., αιτ. σκ. (20) και άρθ. 4 Κανονισμός (ΕΕ) 2024/1689 για την ΤΝ (για τον εν λόγω κανονισμό βλ. αναλυτικά παρακάτω υπό Μέρος Πρώτο, ΙΙ.Α.1.). Βλ. και ΟΕΥΕ, Συστάσεις πολιτικών και επενδύσεων για αξιόπιστη τεχνητή νοημοσύνη, 26 Ιουνίου 2019 [εφεξής: ΟΕΥΕ, Συστάσεις πολιτικών], σελ. 11, 19, 21 επ., 31 επ., καθώς και Ψήφισμα του Ευρωπαϊκού Κοινοβουλίου Ρ9_TA(2022)0140, παρ. (194 επ.), (212 επ.). Βλ. επίσης άρθ. 20 Σύμβαση Πλαίσιο του ΣτΕ για την ΤΝ (για τη Σύμβαση αυτή βλ. υπό Μέρος Πρώτο, ΙΙ.Α.4.).

11. Για τις σχετικές έννοιες βλ. Sieber, σε: Sieber et al. (επιμ. έκδ.), *Alternative Systems of Crime Control*, σελ. 3 επ. και Mitsilegas, σε: Sieber et al. (επιμ. έκδ.), *Alternative Systems of Crime Control*, σελ. 203 επ.