



Ευρωπαϊκή Επιτροπή

**ΟΜΑΔΑ ΕΜΠΕΙΡΟΓΝΩΜΟΝΩΝ ΥΨΗΛΟΥ ΕΠΙΠΕΔΟΥ  
ΓΙΑ ΤΗΝ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ**



**ΣΧΕΔΙΟ**

**ΚΑΤΕΥΘΥΝΤΗΡΙΩΝ ΓΡΑΜΜΩΝ ΔΕΟΝΤΟΛΟΓΙΑΣ  
ΓΙΑ ΑΞΙΟΠΙΣΤΗ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ  
ΣΥΝΟΠΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ**

**Έγγραφο εργασίας για διαβούλευση με ενδιαφερόμενα μέρη**

**Βρυξέλλες, 18 Δεκεμβρίου 2018**

# ΣΧΕΔΙΟ ΚΑΤΕΥΘΥΝΤΗΡΙΩΝ ΓΡΑΜΜΩΝ ΔΕΟΝΤΟΛΟΓΙΑΣ

## ΓΙΑ ΑΞΙΟΠΙΣΤΗ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ



Ομάδα εμπειρογνομόνων υψηλού επιπέδου  
**Σχέδιο κατευθυντήριων γραμμών δεοντολογίας για αξιόπιστη τεχνητή νοημοσύνη**

Ευρωπαϊκή Επιτροπή  
Γενική Διεύθυνση Επικοινωνίας

Επικοινωνία Nathalie Smuha - Συντονίστρια της ΟΕΥΕ για την ΤΝ  
E-mail CNECT-HLG-AI@ec.europa.eu

Ευρωπαϊκή Επιτροπή  
B-1049 Brussels

Το έγγραφο δημοσιεύθηκε στις 18 Δεκεμβρίου 2018, στα αγγλικά.

**Το παρόν έγγραφο εργασίας εκπονήθηκε από την Ομάδα εμπειρογνομόνων υψηλού επιπέδου για την τεχνητή νοημοσύνη (ΟΕΥΕ για την ΤΝ), χωρίς να θίγεται η προσωπική άποψη των μελών της επί συγκεκριμένων θεμάτων και με την επιφύλαξη της τελικής έκδοσης του εγγράφου. Το παρόν έγγραφο θα υποβληθεί σε περαιτέρω επεξεργασία και η τελική του έκδοση θα παρουσιαστεί τον Μάρτιο του 2019, έπειτα από διαβούλευση με τα ενδιαφερόμενα μέρη μέσω της Ευρωπαϊκής Συμμαχίας για την ΤΝ.**

Ούτε η Ευρωπαϊκή Επιτροπή ούτε οποιοδήποτε άλλο πρόσωπο ενεργεί εξ ονόματός της φέρει ευθύνη για την ενδεχόμενη χρήση των κάτωθι πληροφοριών. Για τα περιεχόμενα του παρόντος εγγράφου εργασίας αποκλειστικά υπεύθυνη είναι η Ομάδα εμπειρογνομόνων υψηλού επιπέδου για την τεχνητή νοημοσύνη (ΟΕΥΕ για την ΤΝ). Παρόλο που μέλη του προσωπικού των υπηρεσιών της Επιτροπής διευκόλυναν την εκπόνηση των κατευθυντήριων γραμμών, οι απόψεις που διατυπώνονται στο παρόν έγγραφο εκφράζουν την άποψη της ΟΕΥΕ για την ΤΝ και δεν μπορούν σε καμία περίπτωση να εκληφθούν ως επίσημη θέση της Ευρωπαϊκής Επιτροπής. Το παρόν έγγραφο αποτελεί σχέδιο του πρώτου παραδοτέου της ΟΕΥΕ για την ΤΝ. Η τελική έκδοση του παρόντος εγγράφου θα υποβληθεί στην Επιτροπή τον Μάρτιο του 2019. Η τελική έκδοση του δεύτερου παραδοτέου, δηλαδή των συστάσεων πολιτικής και επενδύσεων στον τομέα της ΤΝ, θα υποβληθεί στα μέσα του 2019.

Περισσότερες πληροφορίες σχετικά με την Ομάδα εμπειρογνομόνων υψηλού επιπέδου για την τεχνητή νοημοσύνη διατίθενται στο διαδίκτυο (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

Η πολιτική για την περαιτέρω χρήση εγγράφων της Ευρωπαϊκής Επιτροπής καθορίζεται στην απόφαση 2011/833/ΕΕ (ΕΕ L 330 της 14.12.2011, σ. 39). Για κάθε χρήση ή αναπαραγωγή φωτογραφιών ή άλλου υλικού που δεν υπόκειται στους κανόνες της ΕΕ για τα δικαιώματα πνευματικής ιδιοκτησίας πρέπει να ζητείται άδεια απευθείας από τους κατόχους δικαιωμάτων πνευματικής ιδιοκτησίας.

## ΣΥΝΟΠΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ

Το παρόν έγγραφο εργασίας αποτελεί σχέδιο των κατευθυντήριων γραμμών δεοντολογίας για την ΤΝ που εκπονήθηκε από την Ομάδα εμπειρογνομόνων υψηλού επιπέδου για την τεχνητή νοημοσύνη (ΟΕΥΕ για την ΤΝ) της Ευρωπαϊκής Επιτροπής, η τελική έκδοση του οποίου προβλέπεται τον Μάρτιο του 2019.

Η τεχνητή νοημοσύνη (ΤΝ) αποτελεί μία από τις πλέον μετασχηματιστικές δυνάμεις στην εποχή μας και θα μεταβάλει αναπόφευκτα τον κοινωνικό ιστό. Αποτελεί επίσης μια τεράστια ευκαιρία για αύξηση της ευημερίας και της ανάπτυξης, την οποία η Ευρώπη πρέπει να αδράξει. Την τελευταία δεκαετία πραγματοποιήθηκαν σημαντικά βήματα προόδου χάρη στη διαθεσιμότητα τεράστιου όγκου ψηφιακών δεδομένων και ισχυρών υπολογιστικών αρχιτεκτονικών, καθώς και χάρη στις προόδους των τεχνικών ΤΝ, όπως η μηχανική εκμάθηση. Η ποιότητα της ζωής μας βελτιώνεται καθημερινά χάρη σε σημαντικές εξελίξεις βασισμένες στην ΤΝ σε τομείς όπως τα αυτόνομα οχήματα, η ιατροφαρμακευτική περίθαλψη, τα οικιακά/υπηρεσιακά ρομπότ, η εκπαίδευση και η κυβερνοασφάλεια. Επιπλέον, η ΤΝ συνιστά καταλυτικό παράγοντα για την αντιμετώπιση πολλών από τις μείζονες προκλήσεις που αντιμετωπίζει ο πλανήτης, όπως η παγκόσμια υγεία και ευημερία, η κλιματική αλλαγή, η ύπαρξη αξιόπιστων νομικών και δημοκρατικών συστημάτων και άλλες προκλήσεις που αναφέρονται στους στόχους βιώσιμης ανάπτυξης των Ηνωμένων Εθνών.

Η ΤΝ μπορεί να αποφέρει τεράστια οφέλη στους πολίτες και την κοινωνία, ενέχει όμως και ορισμένους κινδύνους που θα πρέπει να αντιμετωπιστούν δεόντως. Δεδομένου ότι, σε γενικές γραμμές, τα οφέλη της ΤΝ υπερτερούν των κινδύνων της, πρέπει να ακολουθήσουμε μια πορεία που **μεγιστοποιεί τα οφέλη της ΤΝ και, ταυτόχρονα, ελαχιστοποιεί τους κινδύνους της**. Για να εξασφαλίσουμε ότι δεν θα ξεφύγουμε από αυτή την πορεία, **χρειάζεται μια ανθρωποκεντρική προσέγγιση της ΤΝ** που θα μας υπενθυμίζει διαρκώς ότι η ανάπτυξη και η χρήση της ΤΝ δεν θα πρέπει να θεωρείται αυτοσκοπός αλλά ως μέσο για τη βελτίωση της ανθρώπινης ευημερίας. **Πυξίδα μας θα είναι η αξιοπιστία της ΤΝ**, διότι, μόνο αν οι άνθρωποι εμπιστευτούν την τεχνολογία, θα μπορέσουν να αποκομίσουν πλήρως και με σιγουριά τα οφέλη της ΤΝ.

Η αξιοπιστία της ΤΝ έχει **δύο συνιστώσες**: η ΤΝ (1) θα πρέπει να σέβεται τα θεμελιώδη δικαιώματα, την εφαρμοστέα νομοθεσία και τις βασικές αρχές και αξίες, διασφαλίζοντας έναν «**δεοντολογικό σκοπό**», και (2) θα πρέπει να είναι **τεχνικά άρτια** και αξιόπιστη διότι, ακόμα και όταν οι προθέσεις είναι καλές, η έλλειψη τεχνογνωσίας μπορεί να προκαλέσει ακούσια βλάβη.

Με τις παρούσες κατευθυντήριες γραμμές θεσπίζεται ένα **πλαίσιο για αξιόπιστη ΤΝ**:

- Το **κεφάλαιο I** πραγματεύεται την **εξασφάλιση του δεοντολογικού σκοπού της ΤΝ** και καθορίζει τα θεμελιώδη δικαιώματα, τις αρχές και τις αξίες που θα πρέπει να τηρούνται.
- Βάσει αυτών των αρχών, στο **κεφάλαιο II** συνάγονται **κατευθύνσεις για την επίτευξη** αξιόπιστης ΤΝ, λαμβανομένων υπόψη τόσο του δεοντολογικού σκοπού όσο και της τεχνικής αρτιότητας. Για τον λόγο αυτόν, απαριθμούνται οι απαιτήσεις για την αξιοπιστία της ΤΝ και παρέχεται συνοπτική θεώρηση των τεχνικών και μη τεχνικών μεθόδων που μπορούν να χρησιμοποιηθούν για την υλοποίησή της.
- Στη συνέχεια, το **κεφάλαιο III καθιστά λειτουργικές** τις απαιτήσεις προβλέποντας συγκεκριμένο αλλά μη εξαντλητικό κατάλογο αξιολόγησης της αξιοπιστίας της ΤΝ. Ο εν λόγω κατάλογος προσαρμόζεται εν συνεχεία σε συγκεκριμένες περιπτώσεις χρήσης.

Σε αντίθεση με άλλα έγγραφα που αφορούν τη δεοντολογία στον τομέα της ΤΝ, σκοπός των ακόλουθων κατευθυντήριων γραμμών δεν είναι να εκπονηθεί ένας ακόμη κατάλογος βασικών αξιών και αρχών για την ΤΝ, αλλά να παρασχεθούν κατευθύνσεις για συγκεκριμένη εφαρμογή και υλοποίηση αυτών σε συστήματα ΤΝ. Οι εν λόγω κατευθύνσεις παρέχονται σε τρία επίπεδα, από τα πιο αφηρημένα στο κεφάλαιο I (θεμελιώδη δικαιώματα, αρχές και αξίες) έως τα πιο συγκεκριμένα στο κεφάλαιο III

(κατάλογος αξιολόγησης).

Οι κατευθυντήριες γραμμές απευθύνονται σε όλα τα **εμπλεκόμενα ενδιαφερόμενα μέρη που συμμετέχουν στην ανάπτυξη, την εγκατάσταση και τη χρήση ΤΝ**, δηλαδή σε επιχειρήσεις, οργανισμούς, ερευνητές, δημόσιες υπηρεσίες, θεσμικά όργανα, μεμονωμένους πολίτες και άλλες οντότητες. Στην τελική έκδοση των παρούσων κατευθυντήριων γραμμών, θα προβλέπεται μηχανισμός που θα παρέχει στα ενδιαφερόμενα μέρη τη δυνατότητα να τις εγκρίνουν σε εθελοντική βάση.

Σημειωτέον ότι οι παρούσες κατευθυντήριες γραμμές δεν προορίζονται να υποκαταστήσουν καμία μορφή χάραξης πολιτικής ή κανονιστικών διατάξεων (αυτές θα αποτελέσουν αντικείμενο του δεύτερου παραδοτέου της ΟΕΥΕ για την ΤΝ, δηλαδή των συστάσεων πολιτικής και επενδύσεων, το οποίο αναμένεται τον Μάιο του 2019), ούτε στοχεύουν να αποτρέψουν τη χάραξη πολιτικής ή τη θέσπιση κανονιστικών διατάξεων. Επιπλέον, οι κατευθυντήριες γραμμές θα πρέπει να εκληφθούν ως έγγραφο σε εξέλιξη το οποίο πρέπει να επικαιροποιείται ανά τακτά χρονικά διαστήματα, ώστε να συμβαδίζει διαρκώς με την εξελισσόμενη τεχνολογία και τις γνώσεις μας για αυτήν. Το παρόν έγγραφο θα πρέπει, επομένως, να αποτελέσει εφαλτήριο για την έναρξη του διαλόγου με τίτλο **«Αξιόπιστη ΤΝ με τη σφραγίδα της Ευρώπης»**.

Η Ευρώπη μπορεί να προωθήσει τη δεοντολογική της προσέγγιση ως προς την ΤΝ μόνο αν είναι ανταγωνιστική σε παγκόσμιο επίπεδο, ωστόσο, η **δεοντολογική προσέγγιση της ΤΝ συνιστά βασικό παράγοντα για την υπεύθυνη ανταγωνιστικότητα**, διότι θα εμπνέει εμπιστοσύνη στους χρήστες και θα διευκολύνει την ευρύτερη διεξόδυση της ΤΝ. Οι παρούσες κατευθυντήριες γραμμές δεν έχουν ως στόχο να αποθαρρύνουν την καινοτομία στον τομέα της ΤΝ στην Ευρώπη, αλλά, αντιθέτως, να αξιοποιήσουν τη δεοντολογία ως έμπνευση ώστε να αναπτυχθεί μια μοναδική σειρά ΤΝ, σκοπός της οποίας είναι να προστατεύει και να ωφελεί τους μεμονωμένους πολίτες και να υπηρετεί το κοινό συμφέρον. Με αυτόν τον τρόπο, η Ευρώπη θα καταστεί πρωτοπόρος στην προηγμένη, ασφαλή και δεοντολογική ΤΝ. Μόνο αν διασφαλιστεί η αξιοπιστία, θα μπορέσουν οι Ευρωπαίοι πολίτες να αποκομίσουν πλήρως τα οφέλη της ΤΝ.

Τέλος, οι παρούσες κατευθυντήριες γραμμές αποσκοπούν επίσης στη **δημιουργία ευνοϊκών συνθηκών για την ανάπτυξη προβληματισμού και τη διεξαγωγή διαλόγου** με αντικείμενο ένα πλαίσιο δεοντολογίας για την ΤΝ, πέρα από την Ευρώπη, **σε παγκόσμιο επίπεδο**.

## **ΣΥΝΟΠΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ**

Κάθε κεφάλαιο των κατευθυντήριων γραμμών παρέχει κατευθύνσεις για την επίτευξη αξιόπιστης ΤΝ, οι οποίες απευθύνονται σε όλα τα εμπλεκόμενα ενδιαφερόμενα μέρη που συμμετέχουν στην ανάπτυξη, την εγκατάσταση και τη χρήση ΤΝ. Οι κατευθύνσεις αυτές παρουσιάζονται συνοπτικά παρακάτω:

### **Κεφάλαιο I: Βασικές κατευθύνσεις για τη διασφάλιση δεοντολογικού σκοπού:**

- Εξασφάλιση **ανθρωποκεντρικής ΤΝ**: η ΤΝ θα πρέπει να αναπτύσσεται, να εγκαθίσταται και να χρησιμοποιείται με **«δεοντολογικό σκοπό»**, ο οποίος βασίζεται και αντικατοπτρίζει θεμελιώδη δικαιώματα, κοινωνικές αξίες και τις δεοντολογικές αρχές της *ωφέλιμης δράσης* (ωφελείν), της *αποτροπής ζημιών* (μη βλάπτειν), της *ανθρώπινης αυτονομίας*, της *δικαιοσύνης* και της *επεξηγησιμότητας*. Αυτό έχει κείμενο σημασία για την επίτευξη **αξιόπιστης ΤΝ**.
- Χρήση θεμελιωδών δικαιωμάτων, δεοντολογικών αρχών και αξιών ως βάση για τη διερευνητική εκτίμηση πιθανών επιπτώσεων της ΤΝ στον άνθρωπο και το κοινό καλό. **Ιδιαίτερη προσοχή** πρέπει να δίνεται σε καταστάσεις που αφορούν περισσότερο **ευάλωτες ομάδες**, όπως παιδιά, άτομα με αναπηρίες ή μειονότητες, ή σε καταστάσεις όπου επικρατούν **ασυμμετρίες εξουσίας ή πληροφόρησης**, όπως μεταξύ εργοδοτών και εργαζομένων ή μεταξύ επιχειρήσεων και καταναλωτών.

- Αναγνώριση και επίγνωση του γεγονότος ότι η ΤΝ, ενώ αποφέρει ουσιαστικά οφέλη για τους πολίτες και την κοινωνία, μπορεί να έχει επίσης αρνητικές επιπτώσεις. Εγρήγορση σε κρίσιμους προβληματικούς τομείς.

### **Κεφάλαιο II: Βασικές κατευθύνσεις για την επίτευξη αξιόπιστης ΤΝ:**

- Ενσωμάτωση των **απαιτήσεων για την αξιοπιστία της ΤΝ από το αρχικό στάδιο σχεδιασμού:** Λογοδοσία, Διακυβέρνηση των δεδομένων, Σχεδιασμός για όλους, Διακυβέρνηση της αυτονομίας της ΤΝ (Επίβλεψη από άνθρωπο), Μη εισαγωγή διακρίσεων, Σεβασμός της ανθρώπινης αυτονομίας, Σεβασμός της ιδιωτικής ζωής, Ανθεκτικότητα, Ασφάλεια, Διαφάνεια.
- Εξέταση τεχνικών και μη τεχνικών μεθόδων για την εξασφάλιση της εφαρμογής των εν λόγω απαιτήσεων στο σύστημα ΤΝ. Επιπλέον, οι εν λόγω απαιτήσεις θα πρέπει να συνυπολογίζονται κατά τη στελέχωση της ομάδας που θα εργαστεί επί του συστήματος και κατά την ανάπτυξη του ίδιου του συστήματος, του περιβάλλοντος δοκιμών και των πιθανών εφαρμογών του συστήματος.
- Σαφής και εκ των προτέρων  **ενημέρωση των ενδιαφερόμενων μερών** (πελάτες, εργαζόμενοι κ.λπ.) σχετικά με τις δυνατότητες και τους περιορισμούς του συστήματος ΤΝ, ώστε να μπορούν τα ενδιαφερόμενα μέρη να θέσουν ρεαλιστικές προσδοκίες. Υπό αυτή την έννοια έχει ιδιαίτερη σημασία να εξασφαλίζεται **ιχνηλασιμότητα** του συστήματος ΤΝ.
- Ενσωμάτωση της αξιοπιστίας της ΤΝ στην **νοοτροπία επιχειρήσεων και οργανισμών** και ενημέρωση των ενδιαφερόμενων μερών σχετικά με τον τρόπο με τον οποίο εφαρμόζεται η αξιοπιστία της ΤΝ στον σχεδιασμό και τη χρήση των συστημάτων ΤΝ. Η αξιοπιστία της ΤΝ μπορεί επίσης να περιλαμβάνεται στους κώδικες δεοντολογίας ή συμπεριφοράς επιχειρήσεων και οργανισμών.
- Μέριμνα για τη συμμετοχή και **εμπλοκή των ενδιαφερόμενων φορέων** στη διαδικασία σχεδιασμού και ανάπτυξης του συστήματος ΤΝ. Επιπλέον, πρέπει να διασφαλίζεται **πολυμορφία** κατά τη στελέχωση των ομάδων που αναπτύσσουν, εφαρμόζουν και υποβάλλουν σε δοκιμές το προϊόν.
- Καταβολή προσπάθειών για τη **διευκόλυνση της ελεγχιμότητας** των συστημάτων ΤΝ, ιδίως σε περιπτώσεις ή καταστάσεις κρίσιμης σημασίας. Στο μέτρο του εφικτού, το σύστημα πρέπει να σχεδιάζεται κατά τρόπο που επιτρέπει την ανίχνευση επιμέρους αποφάσεων επί εισερχόμενων στοιχείων, δεδομένων, εκ των προτέρων εκπαιδευμένων μοντέλων κ.λπ. Επιπλέον, πρέπει να ορίζονται **μέθοδοι επεξήγησης** του συστήματος ΤΝ.
- Διασφάλιση συγκεκριμένης διαδικασίας για τη **διαχείριση της λογοδοσίας**.
- Πρόβλεψη **κατάρτισης και εκπαίδευσης** και διασφάλιση της ενημέρωσης και της κατάρτισης διευθυντών, προγραμματιστών, χρηστών και εργοδοτών σε θέματα αξιοπιστίας της ΤΝ.
- Σημειωτέον ότι ενδέχεται να υφίστανται θεμελιώδεις συγκρούσεις μεταξύ διαφορετικών στόχων (η διαφάνεια ενδέχεται να αφήνει περιθώριο για κατάχρηση, η διαπίστωση και η διευθέτηση τυχόν μεροληψίας ενδέχεται να συγκρούεται με διατάξεις περί προστασίας της ιδιωτικής ζωής). Οι αντισταθμιστικές αυτές ρυθμίσεις πρέπει να κοινοποιούνται και να τεκμηριώνονται.
- Προώθηση της έρευνας και της καινοτομίας με στόχο την τήρηση των απαιτήσεων για την αξιοπιστία της ΤΝ.

### **Κεφάλαιο III: Βασικές κατευθύνσεις για την αξιολόγηση της αξιοπιστίας της ΤΝ:**

- Έκδοση **καταλόγου αξιολόγησης** της αξιοπιστίας της ΤΝ κατά την ανάπτυξη, την εγκατάσταση και τη χρήση ΤΝ, και προσαρμογή αυτού στη συγκεκριμένη περίπτωση χρήσης του συστήματος.
- Σημειωτέον ότι ο κατάλογος αξιολόγησης **δεν θα είναι ποτέ εξαντλητικός** και ότι η εξασφάλιση αξιόπιστης ΤΝ δεν συνίσταται σε συμπλήρωση τετραγωνιδίων, αλλά σε διαρκή διαδικασία προσδιορισμού απαιτήσεων, αξιολόγησης λύσεων και διασφάλισης βελτιωμένων αποτελεσμάτων σε ολόκληρο τον κύκλο ζωής του συστήματος ΤΝ.

Οι παρούσες κατευθύνσεις εντάσσονται στο όραμα μιας ανθρωποκεντρικής προσέγγισης της ΤΝ που

θα επιτρέψει στην Ευρώπη να αναλάβει παγκοσμίως ρόλο πρωτοπόρου στον τομέα της δεοντολογικής, ασφαλούς και προηγμένης ΤΝ. Απώτερος σκοπός των παρούσων κατευθύνσεων είναι να διευκολύνουν και να προαγάγουν την «**Αξιόπιστη ΤΝ με τη σφραγίδα της Ευρώπης**» η οποία θα βελτιώσει την ευημερία των Ευρωπαίων πολιτών.