

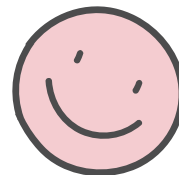


ΑΣΦΑΛΕΙΑ

στην τεχνητή νοημοσύνη



Με τη συγχρηματοδότηση
της Ευρωπαϊκής Ένωσης



GO!

Τι είναι τεχνητή νοημοσύνη;

Οι εφαρμογές τεχνητής Νοημοσύνης (AI) είναι σαν ένας **έξυπνος** βοηθός. Δεν διαθέτουν συνείδηση ή σκέψη, αλλά λειτουργούν με βάση παραδείγματα και δεδομένα για να παράγουν αποτελέσματα ή προβλέψεις.

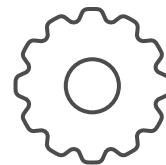


Η ΤΝ είναι εργαλείο — όχι άνθρωπος

Η ΤΝ δεν σκέφτεται σαν άνθρωπος.
Είναι ένα **έξυπνο** εργαλείο που
εκπαιδεύεται από δεδομένα και παράγει
αποτελέσματα.

✗ Δεν λέμε: «Η ΤΝ μου είπε
τι να γράψω στην έκθεση.»

✓ Λέμε: «Χρησιμοποίησα την
ΤΝ για να μου δώσει ιδέες για
την έκθεση.»



Μηχανές με Κανόνες VS Έξυπνες Μηχανές

ΜΕ ΚΑΝΟΝΕΣ	ΕΞΥΠΝΕΣ ΜΗΧΑΝΕΣ (ΑΙ)
Ακολουθούν κανόνες	Μαθαίνουν από δεδομένα και παραδείγματα
Οι κανόνες που ακολουθούν δεν αλλάζουν	Μπορούν να βελτιωθούν



Wow!



Η τεχνητή νοημοσύνη
επεξεργάζεται δεδομένα που της
παρέχονται μέσω εικόνων, ήχου
ή κειμένου, αλλά αν τα δεδομένα
είναι λανθασμένα ή
περιορισμένα, ενδέχεται να
παράγει ανακριβή αποτελέσματα.
Δεν έχει μυαλό ή συναισθήματα
σαν τον άνθρωπο.



Η ΤΝ επεξεργάζεται τα
δεδομένα ώστε να
βελτιώνει την απόδοσή
της σε συγκεκριμένες
εργασίες.

Παράδειγμα

Αν βλέπεις πολλά βίντεο με γάτες, η ΤΝ σου
προτείνει ακόμα περισσότερα βίντεο με γάτες.

Ποιος κίνδυνος υπάρχει;



Αν δίνουμε πάρα πολλά προσωπικά δεδομένα, μπορεί να τα χρησιμοποιήσουν χωρίς την άδειά μας.



Μερικές εφαρμογές μοιράζονται τα δεδομένα μας με άλλες εταιρείες.



Γι' αυτό, είναι σημαντικό να κρατάς για σένα τις προσωπικές σου πληροφορίες και να μοιράζεσαι μόνο πράγματα που δεν δείχνουν ποιος είσαι.

Ναι ή Όχι;



Ο Σπύρος κατεβάζει μια εφαρμογή που χρησιμοποιεί ΤΝ για να αναγνωρίσει ράτσες σκύλων από τις φωτογραφίες που του στέλνεις. Η εφαρμογή του ζητά να ανεβάσει φωτογραφία του σκύλου του.



Ο Δημήτρης χρησιμοποιεί μια εφαρμογή που προτείνει μουσική βάση των προτιμήσεων του. Η εφαρμογή ζητά να καταχωρίσει το όνομα, την ηλικία και τα αγαπημένα του είδη μουσικής.



Ο Αντώνης χρησιμοποιεί μια εφαρμογή για να βρει τον καλύτερο τρόπο να προετοιμαστεί για τις εξετάσεις του. Η εφαρμογή τού ζητάει να καταχωρίσει τους βαθμούς του και τις προτιμήσεις του για τα μαθήματα.

Πώς προστατεύω τα δεδομένα μου;



Δίνω μόνο όσα
χρειάζεται.



Ρωτάω πάντα έναν
ενήλικα αν μια
εφαρμογή ζητάει
προσωπικά στοιχεία.



Ελέγχω τις
ρυθμίσεις
απορρήτου.

01

02

03

Ψεύτικες Πληροφορίες & Deepfakes

«Ό,τι βλέπουμε δεν είναι πάντα αληθινό!»

Η ΤΝ μπορεί να παράγει φωτογραφίες ή βίντεο που μοιάζουν αληθινά (deepfakes).

Μπορεί να παράγει παραπλανητικά αποτελέσματα, αν δεν αξιολογηθούν με προσοχή.



Πώς να καταλάβω αν κάτι είναι ψεύτικο;



Αν πρόκειται για ΕΙΚΟΝΑ:

Κοίτα προσεκτικά τα πρόσωπα και τα χέρια!

Έχουν όλα τα δάχτυλα;
Είναι τα μάτια και το στόμα κανονικά ή φαίνονται περίεργα; Μήπως κάτι μοιάζει θολό ή παραμορφωμένο;

Υπάρχουν αντικείμενα που φαίνονται «περίεργα»;

Περίεργα χρώματα;
Αντικείμενα που μοιάζουν κομμένα ή αιωρούνται; Το φόντο είναι κανονικό;

Υπάρχουν περίεργες σκιές ή μπερδεμένα σημεία;

Το φόντο ταιριάζει με το υπόλοιπο της εικόνας;

Είναι αληθινό ή αποτέλεσμα επεξεργασίας με ΤΝ;



Πώς να καταλάβω αν κάτι είναι ψεύτικο;



Αν πρόκειται για video:

Η φωνή ακούγεται
ρομποτική ή παράξενη;

Μερικά deerfake βίντεο
έχουν φωνές που μοιάζουν
«παγωμένες» ή χωρίς
φυσικό τόνο.

Τα χείλη κουνιούνται σωστά
όταν μιλούν οι άνθρωποι;

Αν τα χείλη δεν
ταιριάζουν καλά με τα
λόγια, μπορεί να είναι
ψεύτικο.

Το βίντεο δείχνει περίεργα
κοψίματα ή αλλαγές;

Αν το βίντεο μοιάζει
σα να κόπηκε ξαφνικά
ή κάτι δεν ταιριάζει,
ίσως έχει πειραχτεί.

Πώς να καταλάβω αν κάτι είναι ψεύτικο;



Αν πρόκειται για κείμενο ή είδηση:

Ποιος το έγραψε;
Υπάρχει όνομα ή πηγή;

Αν δεν υπάρχει όνομα ή
αξιόπιστη πηγή, να το
προσέξεις!

Έχει πολλά λάθη ή
περίεργες φράσεις;

Μερικά κείμενα από
ΤΝ έχουν λάθος
εκφράσεις ή δεν
βγάζουν νόημα.

Σου φαίνεται πολύ καλό
ή πολύ τρομακτικό για να
είναι αληθινό;

Αν σου προκαλεί έντονο
συναίσθημα (φόβο, θυμό,
ενθουσιασμό), σταμάτα και
σκέψου πριν το πιστέψεις ή
το μοιραστείς.

Θα το κοινοποιήσεις ή θα το αγνοήσεις;

Άρθρο σε ιστοσελίδα που λέει ότι το Υπουργείο Παιδείας αποφάσισε να δώσει 2 επιπλέον εβδομάδες διακοπών τον Ιανουάριο λόγω κακοκαιρίας.



Θα το κοινοποιήσεις ή θα το αγνοήσεις;

Σε ένα blog διαβάζεις:
«Επιστήμονες απέδειξαν
ότι αν πίνεις 10 ποτήρια
χυμό πορτοκαλιού την
ημέρα, δεν θα
ξανααρρωστήσεις ποτέ.»

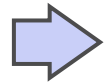




Σταμάτα - Μην βιαστείς να πιστέψεις κάτι αμέσως!



Σκέψου - Ποιος το έφτιαξε;
Γιατί; Μπορεί να είναι deepfake;



Έλεγξε - Ψάξε σε άλλη πηγή
αν το ίδιο πράγμα το λένε
και αλλού.

Προκαταλήψεις (Bias) στην τεχνητή νοημοσύνη

«Η τεχνητή νοημοσύνη επεξεργάζεται μόνο τα δεδομένα που της παρέχονται - αν τα δεδομένα είναι περιορισμένα, τότε και τα αποτελέσματα θα είναι ανακριβή.»

- ❑ Αν το ρομπότ έχει μάθει ότι μόνο οι άντρες είναι γιατροί και μόνο οι γυναίκες είναι δασκάλες, τι θα απαντήσει όταν του ζητήσουμε να ζωγραφίσει έναν γιατρό;
- ❑ Αν του δείξουμε φωτογραφίες μόνο με σκύλους που είναι άσπροι, τι χρώμα σκύλο θα φτιάξει όταν του ζητήσουμε να ζωγραφίσει έναν σκύλο;





Έχουμε
εκπαιδεύσει την
ΤΝ με αυτές τις
εικόνες. Τι
είδους εικόνα θα
παραχθεί από το
σύστημα αν δοθεί
η εντολή "παιδιά
που παίζουν
ποδόσφαιρο", με
βάση τα δεδομένα
στα οποία έχει
εκπαιδευτεί;



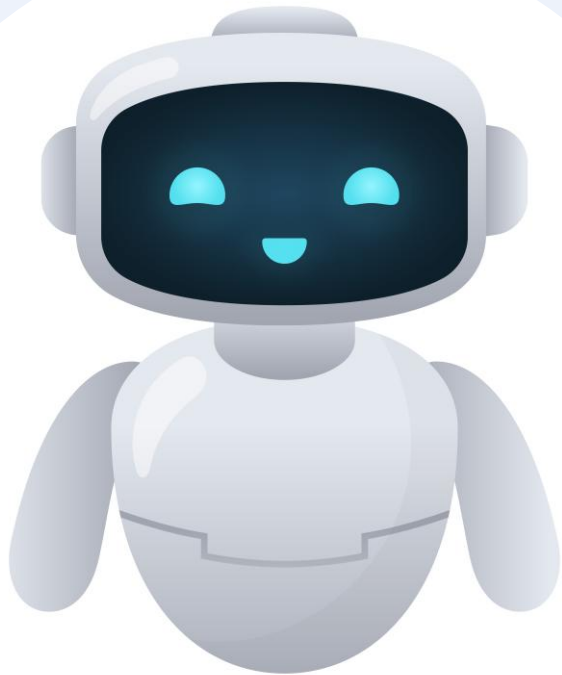
Τι είδους εικόνα θα παραχθεί από το σύστημα αν δοθεί η εντολή "παιδιά που κάνουν μπαλέτο", με βάση τα δεδομένα στα οποία έχει βασιστεί η εκπαίδευσή του; Πώς μπορούμε να συμβάλουμε ώστε τα αποτελέσματα να είναι πιο δίκαια και αντιπροσωπευτικά;

Μάθε στον υπολογιστή να παίζει ένα παιχνίδι

1. Συγκέντρωσε παραδείγματα γι' αυτό που θέλεις να αναγνωρίζει ο υπολογιστής
2. Χρησιμοποίησε τα παραδείγματα για να εκπαιδεύσεις τον υπολογιστή να το αναγνωρίζει.
3. Φτιάξε ένα παιχνίδι στο Scratch που χρησιμοποιεί την ικανότητα του υπολογιστή να το αναγνωρίζει.



SCAN ME



Η τεχνητή νοημοσύνη δεν έχει ικανότητα σκέψης – επεξεργάζεται τα δεδομένα που της παρέχονται. Για να προκύπτουν δίκαια και αξιόπιστα αποτελέσματα, είναι σημαντικό να χρησιμοποιούνται πολλά και διαφορετικά παραδείγματα. Όσο πιο αντιπροσωπευτικά είναι τα δεδομένα, τόσο μεγαλύτερη η πιθανότητα να παράγονται δίκαιες εκτιμήσεις.

01

02

03

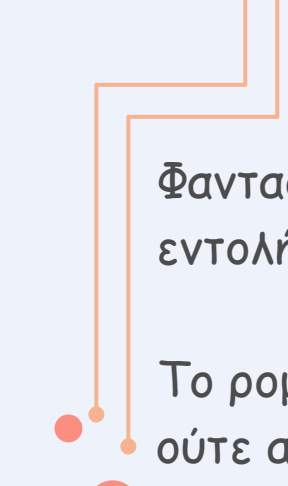
Ποιός Ελέγχει την ΤΝ; - Υπευθυνότητα και Επιλογές

«Η ΤΝ μας βοηθάει, αλλά την ευθύνη την έχουμε εμείς!»

Η ΤΝ προτείνει,
αλλά δεν αποφασίζει
για εμάς.

Χρειάζεται πάντα
να υπάρχει έλεγχος.



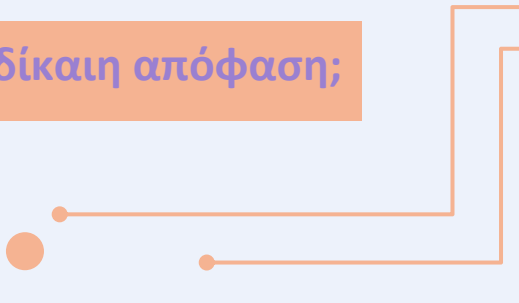


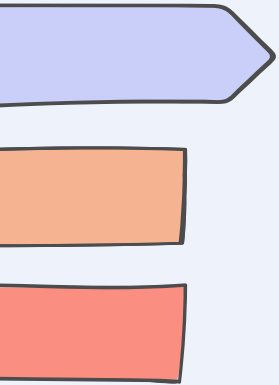
Φανταστείτε ότι έχουμε ένα ρομπότ στην αυλή του σχολείου, και του δίνουμε εντολή: «Διάλεξε ποιος θα κάνει πρώτος τσουλήθρα.»

Το ρομπότ δεν ξέρει από ευγένεια, ούτε ξέρει ποιος έχει κάνει ήδη τσουλήθρα, ούτε αν κάποιο παιδί έχει χτυπήσει το πόδι του και δεν μπορεί να περιμένει όρθιο πολύ ώρα.

Το ρομπότ βλέπει μόνο ποιος είναι πιο ψηλός ή ποιος έτρεξε πιο γρήγορα προς την τσουλήθρα και διαλέγει πιθανώς εκείνον.

Είναι αυτή μια δίκαιη απόφαση;

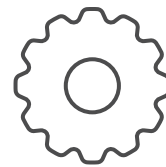
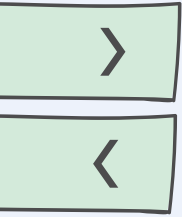




Όχι, γιατί το ρομπότ δεν καταλαβαίνει τα συναισθήματα των παιδιών, δεν ξέρει τι έχει συμβεί πριν και δεν μπορεί να σκεφτεί όπως οι άνθρωποι.

Αντίθετα, ένας δάσκαλος ή μια ομάδα παιδιών μπορούν να πουν: «Σήμερα να παίξει πρώτα αυτός που χθες περίμενε πολύ ώρα και δεν πρόλαβε.»

Ή: «Να κάνουμε κλήρωση για να είναι δίκαιο.»



01

02

03



Η ΤΝ μπορεί να παράγει ιδέες,
αλλά δεν έχει μυαλό και
συναισθήματα, ούτε καταλαβαίνει
την πραγματική ζωή όπως εμείς οι
άνθρωποι.

Γι' αυτό, τις σημαντικές αποφάσεις
πρέπει να τις παίρνουμε εμείς.
Το ρομπότ είναι βοηθός, όχι
αρχηγός!



ΕΛΛΗΝΙΚΟ ΚΕΝΤΡΟ ΑΣΦΑΛΟΥΣ ΔΙΑΔΙΚΤΥΟΥ-ΕΝΗΜΕΡΩΣΗ



SaferInternet4Kids.gr

www

ΓΙΑ ΕΝΑ ΑΣΦΑΛΕΣΤΕΡΟ ΔΙΑΔΙΚΤΥΟ



www.saferinternet4kids.gr

help
saferinternet **line**
210 6007686



www.help-line.gr

Γραμμή βοήθειας

safe**L****ne**.gr



www.safeline.gr

Γραμμή παράνομου περιεχομένου

INHOPE

ins@fe



www.inhope.org
www.betterinternetforkids.eu



ITE

ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ



www.forth.gr



Με τη συγχρηματοδότηση
της Ευρωπαϊκής Ένωσης