

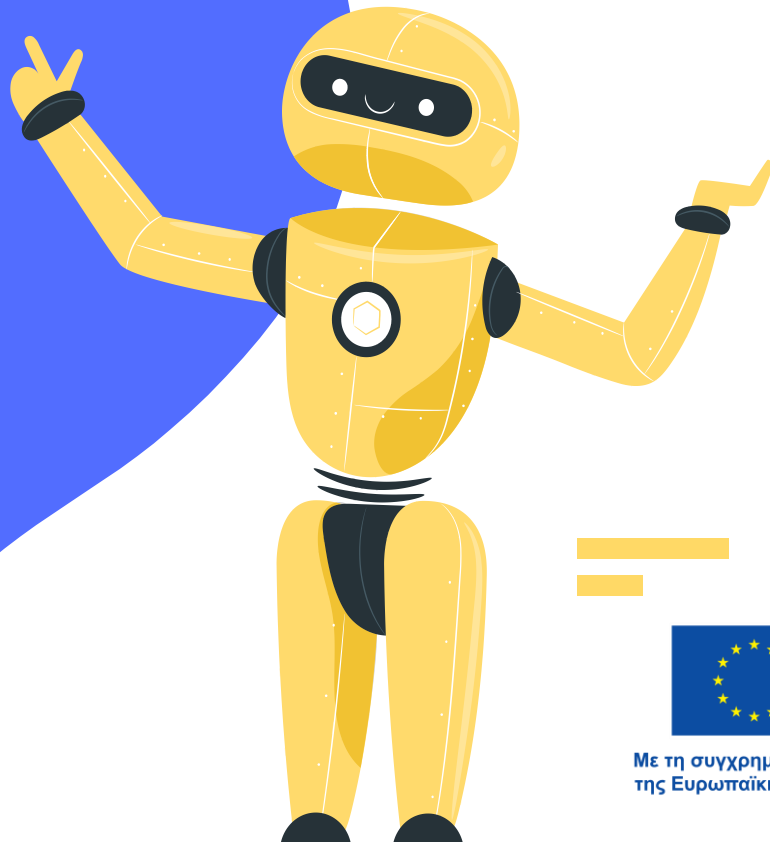


SaferInternet4Kids.gr

ΓΙΑ ΕΝΑ ΑΣΦΑΛΕΣΤΕΡΟ ΔΙΑΔΙΚΤΥΟ

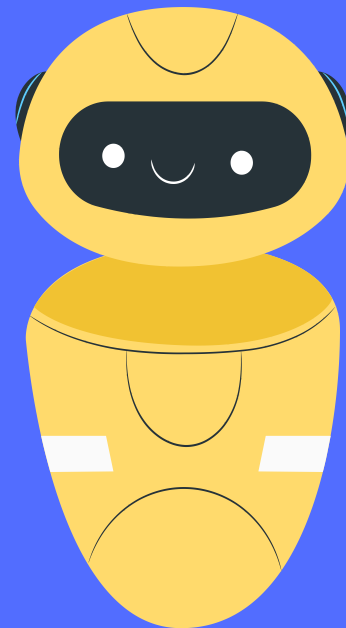
Γιατί η τεχνητή νοημοσύνη κάνει λάθη

Προκατάληψη και
παραπληροφόρηση

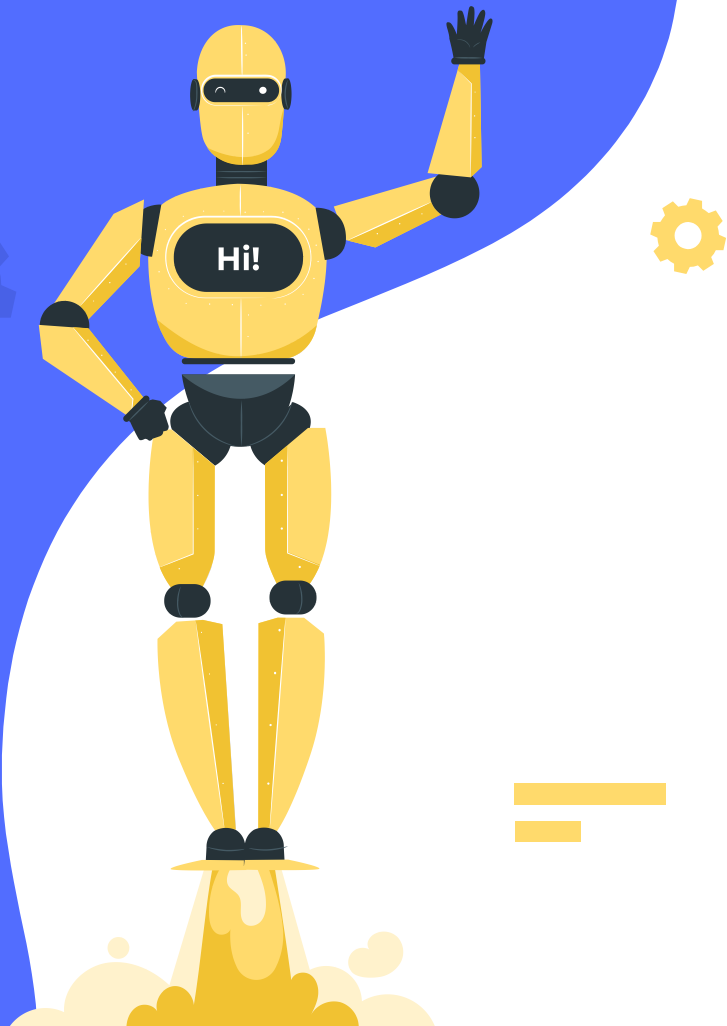


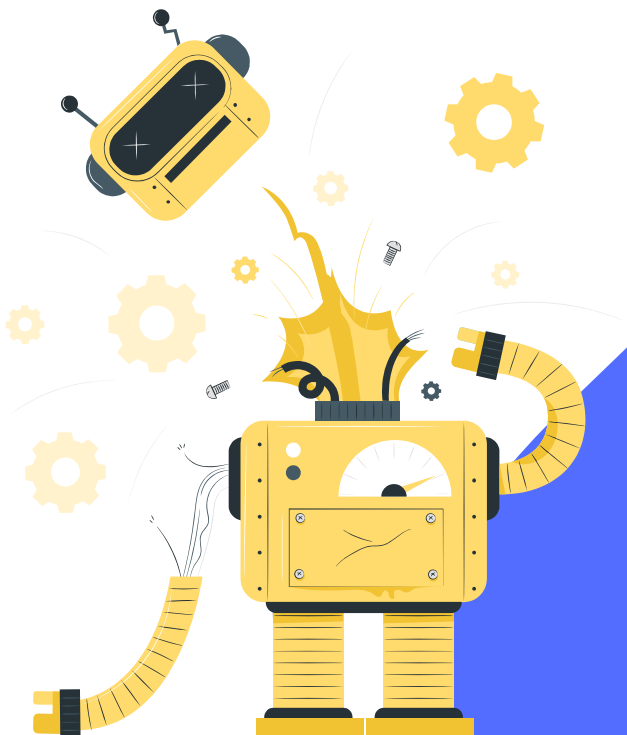
Με τη συγχρηματοδότηση
της Ευρωπαϊκής Ένωσης

- Είδαμε στο προηγούμενο μάθημα πως «μαθαίνει» η τεχνητή νοημοσύνη.
- Ποιος θυμάται να μας πει;



- ❑ Πίσω από κάθε εργαλείο τεχνητής νοημοσύνης κρύβονται οι άνθρωποι που αποφασίζουν ποια δεδομένα εκπαίδευσης θα χρησιμοποιηθούν.
- ❑ Μπορείτε να σκεφτείτε τι κινδύνους μπορεί να κρύβει αυτό;



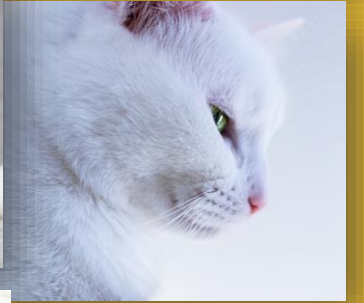
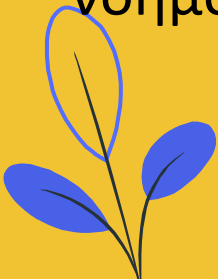


Ας φανταστούμε ότι είμαστε
προγραμματιστές και
θέλουμε να φτιάξουμε ένα
ρομπότ που ξεχωρίζει ζώα.





- ❑ Θα βάζαμε τις φωτογραφίες στα δεξιά για να «εκπαιδεύσουμε» την τεχνητή νοημοσύνη;
- ❑ Ποια ζώα μπορεί τώρα να «αναγνωρίσει η τεχνητή νοημοσύνη»;

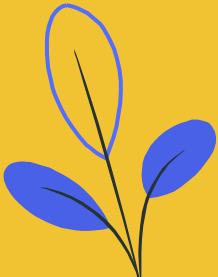




- ❑ Πάμε να ελέγξουμε τώρα αν η τεχνητή νοημοσύνη λειτουργεί όπως πρέπει...
- ❑ Οι ετικέτες κάτω από κάθε εικόνα είναι αυτό που πιστεύει η τεχνητή νοημοσύνη ότι είναι κάθε ζώο.
- ❑ Τι παρατηρείτε;



ΓΑΤΑ





- ❑ Στα δεδομένα με τα οποία «εκπαιδεύσαμε» την τεχνητή νοημοσύνη οι γάτες ήταν το μόνο παράδειγμα άσπρου ζώου.
- ❑ Το εργαλείο AI έμαθε να αναγνωρίζει οτιδήποτε άσπρο ως γάτα.
- ❑ Η τεχνητή νοημοσύνη που δημιουργήσαμε έχει μια «προκατάληψη» να πιστεύει ότι κάθε άσπρο ζώο είναι γάτα.





- Ανάλογα λοιπόν τα δεδομένα με τα οποία «εκπαιδεύουμε» ένα εργαλείο τεχνητής νοημοσύνης μπορεί να δημιουργηθούν αποτελέσματα που να έχουν προκαταλήψεις ή να μην είναι αξιόπιστα.
- Κάτι τέτοιο μπορεί να έχει άμεσο αντίκτυπο σε άτομα ή ομάδες ανθρώπων.



Πάμε να δούμε ένα παράδειγμα

- ❑ Υπήρχε ένα μικρό ρομπότ, το οποίο έμαθε να αναγνωρίζει τα χρώματα από φωτογραφίες. Όμως, το ρομπότ εκπαιδεύτηκε μόνο με φωτογραφίες ανθρώπων με μπλε μαλλιά και πράσινα μάτια, επειδή αυτά ήταν τα χαρακτηριστικά που είχαν οι περισσότεροι άνθρωποι στην περιοχή που έμενε.
- ❑ Όταν το ρομπότ ταξίδεψε στη Χώρα των Χρωμάτων, όπου οι κάτοικοι είχαν όλα τα δυνατά χρώματα, το ρομπότ αντιμετώπισε «προκαταλήψεις». Δεν μπορούσε να αναγνωρίσει σωστά τα διάφορα χρώματα που βρήκε, επειδή δεν είχε μάθει για αυτά στην εκπαίδευσή του.



❑ Το ρομπότ μέσα από αυτήν την εμπειρία ανακάλυψε πώς υπάρχουν και άλλα χρώματα ανθρώπων και έμαθε να τα αναγνωρίζει και να μην έχει προκαταλήψεις. Το ρομπότ έμαθε να εκτιμά τη διαφορετικότητα και να βλέπει την ομορφιά σε όλα τα χρώματα.

❑ Αυτή η ιστορία μας δείχνει ότι, όπως και το ρομπότ, η τεχνητή νοημοσύνη μπορεί να έχει προκαταλήψεις αν δεν εκπαιδεύεται με διάφορα δεδομένα. Είναι σημαντικό να διδάξουμε και στα ρομπότ να εκτιμούν και να αναγνωρίζουν την ποικιλομορφία και τη διαφορετικότητα.



Ποιες είναι μερικές από
τις αρνητικές επιπτώσεις
και συνέπειες της
μεροληψίας της AI ;





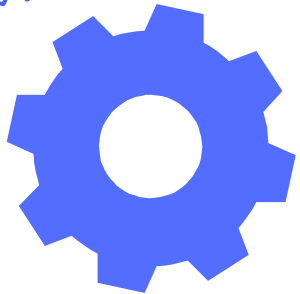
- ❑ Άδικη μεταχείριση
- ❑ Διαιώνιση των στερεοτύπων
- ❑ Άνισες ευκαιρίες
- ❑ Παραπληροφόρηση



ΑΛΗΘΙΝΟ ΠΑΡΑΔΕΙΓΜΑ ΠΡΟΚΑΤΑΛΗΨΗΣ ΑΙ

- ❑ Ο αλγόριθμος προσλήψεων της Amazon
- ❑ Η Amazon είναι ένας από τους μεγαλύτερους τεχνολογικούς γίγαντες στον κόσμο. Το 2015, η Amazon συνειδητοποίησε ότι ο αλγόριθμός της που χρησιμοποιήθηκε για την πρόσληψη υπαλλήλων βρέθηκε να είναι προκατειλημμένος κατά των γυναικών. Ο λόγος για αυτό ήταν επειδή ο αλγόριθμος βασίστηκε στον αριθμό των βιογραφικών που υποβλήθηκαν τα τελευταία δέκα χρόνια και δεδομένου ότι οι περισσότεροι από τους αιτούντες ήταν άνδρες, είχε εκπαιδευτεί να ευνοεί τους άνδρες έναντι των γυναικών.

- ❑ Η αναφορά μεροληψίας AI μπορεί να βοηθήσει τις εταιρείες να βελτιώσουν τα προϊόντα τους.
- ❑ Τι θα μπορούσε να κάνει η εταιρεία στην περίπτωση της ιστορίας που αναφέρθηκε παραπάνω; Πως θα μπορούσε να «μειώσει την προκατάληψη» στο ρομπότ;





Η σωστή και ασφαλής
χρήση της τεχνητής
νοημοσύνης απαιτεί από
εμάς να έχουμε κριτική
σκέψη!



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**ine.gr



www.safeline.gr

Γραμμή παράνομου περιεχομένου

INHOPE

ins@fe



www.inhope.org
www.betterinternetforkids.eu



ITE

ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ



www.forth.gr



Συγχρηματοδοτείται από την
Ευρωπαϊκή Ένωση