

# LESSON PLAN

## Γιατί η τεχνητή νοημοσύνη κάνει λάθη

Προκατάληψη και παραπληροφόρηση

### ΣΤΟΧΕΥΜΕΝΟ ΚΟΙΝΟ

Σε μαθητές Γυμνασίου & Λυκείου

### ΔΙΑΡΚΕΙΑ ΥΛΟΠΟΙΗΣΗΣ

1 διδακτική ώρα

### ΠΡΟΫΠΟΘΕΣΕΙΣ ΥΛΟΠΟΙΗΣΗΣ

- ✓ Ο εκπαιδευτικός να είναι εξοικειωμένος με τις Τεχνολογίες Πληροφορικής και Επικοινωνιών (ΤΠΕ).
- ✓ Οι μαθητές να έχουν κατακτήσει τις βασικές έννοιες της Τεχνολογίας της Πληροφορικής και να γνωρίζουν τις βασικές δεξιότητες χρήσης του Η/Υ.
- ✓ Οι μαθητές να έχουν διδαχθεί το μάθημα «Τεχνητή Νοημοσύνη»
- ✓ Οι μαθητές να έχουν διδαχθεί το μάθημα «Πόσο καλά σε γνωρίζουν οι αλγόριθμοι;»

### ΕΡΓΑΛΕΙΑ ΥΛΟΠΟΙΗΣΗΣ

- ✓ Η/Υ
- ✓ Βιντεοπροβολέας
- ✓ Σύνδεση στο διαδίκτυο
- ✓ Παρουσίαση «Γιατί η τεχνητή νοημοσύνη κάνει λάθη. Προκατάληψη και παραπληροφόρηση»
- ✓ Εργαλείο Quick Draw

## ΓΕΝΙΚΟΣ ΣΚΟΠΟΣ

Γενικός σκοπός της παρούσας διδακτικής πρότασης είναι οι μαθητές μέσα σε ένα κατευθυνόμενο από τον εκπαιδευτικό μαθησιακό περιβάλλον, με τη χρήση πολλαπλών και ποικίλων ερεθισμάτων, να κατανοήσουν την προκατάληψη και την παραπληροφόρηση που μπορεί να κρύβει η τεχνολογία της τεχνητής νοημοσύνης και να αναλογιστούν τις αρνητικές επιπτώσεις της μεροληψίας της AI.

## ΕΠΙΜΕΡΟΥΣ ΣΤΟΧΟΙ

- ✓ Συνειδητοποίηση της ύπαρξης προκατάληψης και παραπληροφόρησης στην τεχνητή νοημοσύνη.
- ✓ Κατανόηση του τρόπου που συμβαίνει μεροληψία στην τεχνητή νοημοσύνη.
- ✓ Διερεύνηση τρόπων ελαχιστοποίησης της μεροληψίας στην τεχνητή νοημοσύνη.
- ✓ Υιοθέτηση θετικών στάσεων έναντι των νέων τεχνολογιών για θέματα δια βίου εκπαίδευσης και δημιουργικότητας.
- ✓ Ανάπτυξη κριτικής σκέψης σε θέματα νέων τεχνολογιών.
- ✓ Αποφόρτιση μέσω της δραστηριότητας Quick Draw.

## ΑΝΑΛΥΤΙΚΗ ΠΟΡΕΙΑ ΔΙΔΑΣΚΑΛΙΑΣ

<p><b>Προκατάληψη και παραπληροφόρηση στην Τεχνητή Νοημοσύνη. Ποιες οι συνέπειες.</b></p> <p><b>Διάρκεια 30΄</b></p>	<p>Το μάθημα ξεκινά με μια επανάληψη του lesson plan «Τεχνητή Νοημοσύνη» με έμφαση στο «πως μαθαίνει» η τεχνητή νοημοσύνη (<u>διαφάνεια 2</u>). Ο εκπαιδευτικός, αφού σιγουρευτεί ότι οι μαθητές θυμούνται πως εκπαιδεύεται η τεχνητή νοημοσύνη, στη συνέχεια με ερωτοαπαντήσεις προσπαθεί να οδηγήσει τη σκέψη των μαθητών στους κινδύνους που προκύπτουν από το γεγονός ότι η τεχνητή νοημοσύνη εκπαιδεύεται από ανθρώπους. Ο εκπαιδευτικός εξηγεί στους μαθητές ότι η μεροληψία στην τεχνητή νοημοσύνη συμβαίνει όταν ένα εργαλείο τεχνητής νοημοσύνης λαμβάνει μια απόφαση που είναι λανθασμένη ή προβληματική επειδή εκπαιδεύτηκε από δεδομένα που δεν αντιμετώπιζαν με ακρίβεια όλα τα μέρη, τα άτομα και τα πράγματα. Και τα δεδομένα εκπαίδευσης είναι οι πληροφορίες που δίνονται σε ένα εργαλείο AI για να το βοηθήσουν να μάθει πώς να κάνει μια συγκεκριμένη διεργασία (<u>διαφάνεια 3</u>).</p> <p>Για να εμβαθύνουν οι μαθητές στο γιατί μπορεί ένα εργαλείο τεχνητής νοημοσύνης να μεροληπτεί βλέπουν το παράδειγμα στις <u>διαφάνειες 4, 5, 6, &amp; 7</u>. Στη <u>διαφάνεια 5</u> με βάση τα δεδομένα που δόθηκαν, οι μαθητές διερευνούν ποια ζώα είναι σε θέση να αναγνωρίσει η τεχνητή νοημοσύνη. Στις <u>διαφάνειες 6 &amp; 7</u> φαίνεται ότι στα δεδομένα με τα οποία «εκπαιδεύσαμε» την τεχνητή νοημοσύνη οι γάτες ήταν το μόνο παράδειγμα άσπρου ζώου. Έτσι το εργαλείο AI έμαθε να αναγνωρίζει οτιδήποτε άσπρο ως γάτα. Εξηγήστε ότι τα λάθη που έκανε το εργαλείο είναι ένα παράδειγμα προκατάληψης της τεχνητής νοημοσύνης. Όταν ένα εργαλείο τεχνητής νοημοσύνης λαμβάνει μια απόφαση που είναι λανθασμένη ή προβληματική αυτό οφείλεται στο γεγονός ότι έμαθε από δεδομένα εκπαίδευσης που δεν αντιμετώπιζαν όλα τα άτομα, τα μέρη και τα πράγματα με ακρίβεια (<u>διαφάνεια</u></p>
--	---

g). Ιδιαίτερη έμφαση δίνεται στον αντίκτυπο που μπορεί να έχουν αυτά τα λάθη προκατάληψης σε κάποιους ανθρώπους. Ο εκπαιδευτικός διαβάζει το παράδειγμα με το ρομπότ στις διαφάνειες 9 & 10 και ξεκινάει μια συζήτηση στην τάξη που στόχο έχει οι μαθητές να συνειδητοποιήσουν ότι η τεχνητή νοημοσύνη εκπαιδεύεται σε δεδομένα πραγματικού κόσμου που της δίνουν οι άνθρωποι και εάν αυτά τα δεδομένα περιέχουν προκαταλήψεις (ή είναι ελλιπή), η τεχνητή νοημοσύνη μπορεί επίσης να καταλήξει να είναι προκατειλημμένη.

**Πιθανές ερωτήσεις που μπορεί να θέσει ο εκπαιδευτικός (διαφάνειες 11 & 12)**

- ✓ Ποιες είναι μερικές από τις αρνητικές επιπτώσεις και συνέπειες της μεροληψίας AI;
- ✓ Εάν ένα εργαλείο τεχνητής νοημοσύνης είναι προκατειλημμένο, μπορεί να λάβει αποφάσεις που είναι άδικες για ορισμένες ομάδες ανθρώπων;
- ✓ Εάν ένα εργαλείο τεχνητής νοημοσύνης μαθαίνει από δεδομένα που περιλαμβάνουν στερεότυπα (π.χ. φυλή ή φύλο), μπορεί να λάβει αποφάσεις που βασίζονται σε αυτές τις προκατειλημμένες ιδέες;
- ✓ Η προκατάληψη της τεχνητής νοημοσύνης μπορεί επίσης να περιορίσει τις ευκαιρίες για ορισμένα άτομα ευνοώντας άδικα μια άλλη ομάδα;
- ✓ Εάν ένα εργαλείο τεχνητής νοημοσύνης μάθει από μεροληπτικές πληροφορίες, μπορεί να καταλήξει να δημιουργεί και να διαδίδει ψευδείς ή ελλιπείς πληροφορίες;

Στη διαφάνεια 13 ο εκπαιδευτικός διαβάζει στα παιδιά ένα πραγματικό παράδειγμα προκατάληψης AI και συζητούν τις επιπτώσεις του. Στη συνέχεια ο εκπαιδευτικός διερευνά με τους μαθητές πως πρέπει να αντιδράσει ένας χρήστης όταν συναντήσει μεροληψία στην τεχνητή νοημοσύνη. Ο εκπαιδευτικός υπογραμμίζει στους μαθητές ότι η γνώση της προκατάληψης της τεχνητής νοημοσύνης μπορεί να μας βοηθήσει να σκεφτόμαστε κριτικά και να ενεργούμε υπεύθυνα εάν και όταν χρησιμοποιούμε τεχνητή νοημοσύνη. Εάν παρατηρήσουμε ότι η μεροληψία AI έχει αρνητικό αντίκτυπο σε κάποιον ή κάτι, μπορούμε να βοηθήσουμε αναφέροντάς το στην εταιρεία (διαφάνεια 15). Ο εκπαιδευτικός ρωτάει τους μαθητές να σκεφτούν τι θα μπορούσε να κάνει η εταιρεία έτσι ώστε το ρομπότ να μπορεί να αναγνωρίσει καλύτερα όσο το δυνατόν περισσότερα διαφορετικά χρώματα ανθρώπων;

Το μάθημα ολοκληρώνεται με συζήτηση για την αξία της κριτικής σκέψης όταν χρησιμοποιούμε τις νέες τεχνολογίες.

	Οι μαθητές καλούνται να ζωγραφίσουν κάτι και το AI θα προσπαθήσει να το βρει (στα Αγγλικά).
<b>Αξιολόγηση</b> <b>Διάρκεια 5'</b>	Αφιερώστε 5 λεπτά συζητώντας με τους μαθητές αν γνώριζαν για την προκατάληψη και την παραπληροφόρηση στην τεχνητή νοημοσύνη. Τι τους άρεσε, τι τους έκανε εντύπωση, τι τους προβλημάτισε. Πώς θα χρησιμοποιήσουν τη νέα γνώση που απέκτησαν μέσα από το συγκεκριμένο μάθημα στην καθημερινότητά τους;