Πρότυπο Ανάπτυξης Θεμάτων

AI Literacy

v.3

Περιεχόμενα

Τίτλος Θέματος
Σενάριο
Ερώτηση 1
1.1 Διατύπωση ερώτησης 1
1.2 Απαντήσεις/Επιλογές Ερώτησης 1
1.3 Αποδεκτή απάντηση στην Ερώτηση 1 [και αιτιολόγηση]
1.4 Οδηγός Κωδικοποίησης Ερώτησης 1 [βάσει πλαισίου αξιολόγησης]
1.5 Μεταδεδομένα Ερώτησης [αιτιολόγηση βάσει πλαισίου αξιολόγησης]
Ερώτηση 2
2.1 Διατύπωση ερώτησης 1
2.2 Απαντήσεις/Επιλογές Ερώτησης 1
2.3 Αποδεκτή απάντηση στην Ερώτηση 1 [και αιτιολόγηση]
2.4 Οδηγός Κωδικοποίησης Ερώτησης 1 [βάσει πλαισίου αξιολόγησης]
2.5 Μεταδεδομένα Ερώτησης [αιτιολόγηση βάσει πλαισίου αξιολόγησης]
Ερώτηση 3
3.1 Διατύπωση ερώτησης 1
3.2 Απαντήσεις/Επιλογές Ερώτησης 1
3.3 Αποδεκτή απάντηση στην Ερώτηση 1 [και αιτιολόγηση]
3.4 Οδηγός Κωδικοποίησης Ερώτησης 1 [βάσει πλαισίου αξιολόγησης]
3.5 Μεταδεδομένα Ερώτησης [αιτιολόγηση βάσει πλαισίου αξιολόγησης]
Παρατηρήσεις
4.1 Δημιουργός/οί θέματος

Τίτλος Θέματος: Αλγόριθμος Ποινικής Πρόβλεψης

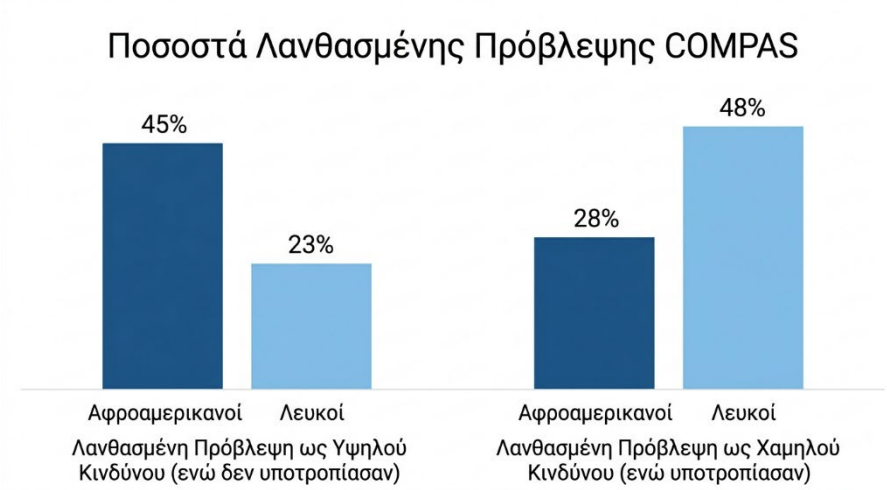
Σενάριο: [Η εισαγωγή του θέματος επαναλαμβάνεται σε όλες τις ερωτήσεις]

Στις Ηνωμένες Πολιτείες, τα δικαστήρια χρησιμοποιούν όλο και περισσότερο συστήματα Τεχνητής Νοημοσύνης για να υποστηρίξουν τις αποφάσεις τους. Ένα από τα πιο γνωστά και αμφιλεγόμενα συστήματα είναι το COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). Ο σκοπός του είναι να βοηθήσει τους δικαστές να αξιολογήσουν την πιθανότητα ένας κατηγορούμενος να διαπράξει ξανά έγκλημα στο μέλλον (υποτροπή). Ο αλγόριθμος αναλύει διάφορα δεδομένα για το άτομο, όπως το ποινικό του μητρώο, την ηλικία, την εργασιακή κατάσταση, αλλά όχι τη φυλή του, και παράγει μια «βαθμολογία κινδύνου» από το 1 έως το 10. Μια υψηλή βαθμολογία υποδηλώνει υψηλή πιθανότητα υποτροπής.

Η ιδέα πίσω από το COMPAS είναι να παρέχει μια αντικειμενική, βασισμένη σε δεδομένα, αξιολόγηση, απαλλαγμένη από τις ανθρώπινες προκαταλήψεις που μπορεί να επηρεάσουν την κρίση ενός δικαστή. Ωστόσο, το 2016, ο μη κερδοσκοπικός δημοσιογραφικός οργανισμός ProPublica διεξήγαγε μια εκτενή έρευνα για την απόδοση του COMPAS, αναλύοντας τις βαθμολογίες κινδύνου για περισσότερους από 7.000 συλληφθέντες σε μια κομητεία της Φλόριντα. Τα αποτελέσματα ήταν ανησυχητικά και πυροδότησαν μια παγκόσμια συζήτηση για την αλγοριθμική δικαιοσύνη.

Η έρευνα αποκάλυψε μια σημαντική φυλετική ανισότητα στα λάθη του αλγορίθμου. Συγκεκριμένα, το COMPAS έτεινε να κάνει δύο τύπους λαθών με διαφορετική συχνότητα για τους λευκούς και τους Αφροαμερικανούς κατηγορούμενους. Το πρώτο λάθος ήταν η λανθασμένη πρόβλεψη ότι κάποιος που δεν θα υποτροπίαζε, τελικά θα το έκανε (ψευδώς θετικό). Το δεύτερο ήταν η λανθασμένη πρόβλεψη ότι κάποιος που θα υποτροπίαζε, δεν θα το έκανε (ψευδώς αρνητικό).

Η ProPublica διαπίστωσε ότι ο αλγόριθμος ήταν σχεδόν διπλάσια πιθανό να χαρακτηρίσει λανθασμένα Αφροαμερικανούς κατηγορούμενους ως μελλοντικά υποτροπιάζοντες, σε σύγκριση με τους λευκούς. Αντίστροφα, οι λευκοί κατηγορούμενοι χαρακτηρίζονταν λανθασμένα ως χαμηλού κινδύνου πολύ πιο συχνά από τους Αφροαμερικανούς. [ΕΙΚΟΝΑ 1: ai-law-p1.png]



Αυτό το φαινόμενο συμβαίνει επειδή οι αλγόριθμοι ΤΝ μαθαίνουν από ιστορικά δεδομένα. Αν τα δεδομένα που χρησιμοποιούνται για την εκπαίδευσή τους αντικατοπτρίζουν υπάρχουσες κοινωνικές και συστημικές προκαταλήψεις (π.χ. υψηλότερα ποσοστά συλλήψεων σε συγκεκριμένες κοινότητες λόγω εντονότερης αστυνόμευσης), ο αλγόριθμος θα μάθει και θα αναπαράγει αυτές τις προκαταλήψεις, ακόμη και αν η «φυλή» δεν είναι μια από τις μεταβλητές που εξετάζει. Έτσι, μια τεχνολογία που σχεδιάστηκε για να προωθήσει την αντικειμενικότητα, κινδυνεύει να ενισχύσει τις ίδιες τις αδικίες που υποτίθεται ότι θα καταπολεμούσε.

Πηγή: ProPublica, "Machine Bias" (διασκευή)

Ενότητα Ερωτήσεων

Ερώτηση 1

1.1 Διατύπωση ερώτησης 1

Να λάβεις υπόψη σου το κείμενο στα δεξιά. Κάνε κλικ σε μια επιλογή, για να απαντήσεις στην ερώτηση.

Σύμφωνα με το Γράφημα (Εικόνα 1), ποια από τις παρακάτω προτάσεις περιγράφει σωστά τα λάθη πρόβλεψης του COMPAS;

1.2 Απαντήσεις/Επιλογές Ερώτησης 1

Α. Οι λευκοί κατηγορούμενοι είχαν περισσότερες πιθανότητες από τους Αφροαμερικανούς να χαρακτηριστούν λανθασμένα ως υψηλού κινδύνου.

Β. Ο αλγόριθμος έκανε τον ίδιο αριθμό λαθών και για τις δύο φυλετικές ομάδες.

Γ. Οι Αφροαμερικανοί κατηγορούμενοι είχαν σχεδόν διπλάσιες πιθανότητες να χαρακτηριστούν λανθασμένα ως υψηλού κινδύνου σε σχέση με τους λευκούς.
Δ. Ο αλγόριθμος ήταν ακριβής στις προβλέψεις του για περισσότερο από το 90% των περιπτώσεων.

1.3 Αποδεκτή απάντηση στην Ερώτηση 1 [και αιτιολόγηση]

Αποδεκτή απάντηση είναι η Γ. Οι Αφροαμερικανοί κατηγορούμενοι είχαν σχεδόν διπλάσιες πιθανότητες να χαρακτηριστούν λανθασμένα ως υψηλού κινδύνου σε σχέση με τους λευκούς.

Το γράφημα δείχνει καθαρά ότι το 45% των Αφροαμερικανών χαρακτηρίστηκαν λανθασμένα ως υψηλού κινδύνου, ποσοστό που είναι σχεδόν διπλάσιο από το 23% των λευκών.

Όλες οι άλλες είναι μη αποδεκτές απαντήσεις.

1.4 Οδηγός Κωδικοποίησης Ερώτησης 1 [βάσει πλαισίου αξιολόγησης]

Αποδεκτή Απάντηση: Επιλογή της απάντησης Γ. Ο μαθητής ερμηνεύει σωστά τα δεδομένα του γραφήματος, συγκρίνοντας τις δύο στήλες στην κατηγορία «Λανθασμένη Πρόβλεψη ως Υψηλού Κινδύνου».
Μη Αποδεκτή Απάντηση: Επιλογή οποιασδήποτε άλλης απάντησης (Α, Β, Δ).

1.5 Μεταδεδομένα Ερώτησης [αιτιολόγηση βάσει πλαισίου αξιολόγησης]

Δείτε τα πλαίσια αξιολόγησης και τη χαρτογράφηση των μεταδεδομένων.

Δεξιότητα	Περιεχόμενο και Σκοπός	Τύπος Μέσου
Ανάλυση και Αξιολόγηση	Ψηφιακή Ιθαγένεια	Ψηφιακά πολυμέσα
Εργαλεία	**Επίπεδο Δυσκολίας**	**Τύπος Απάντησης**
Συστήματα AI	Μεσαίο (3-4)	Απλή πολλαπλής επιλογής

Ερώτηση 2

2.1 Διατύπωση ερώτησης 2

Να λάβεις υπόψη σου το κείμενο στα δεξιά. Πληκτρολόγησε την απάντηση σου στην ερώτηση και στη συνέχεια πάτησε ΥΠΟΒΟΛΗ.

Ένας υποστηρικτής της χρήσης του COMPAS ισχυρίζεται: «Ο αλγόριθμος είναι πιο δίκαιος από έναν άνθρωπο δικαστή, επειδή είναι εντελώς αντικειμενικός και δεν επηρεάζεται από συναισθήματα ή προσωπικές προκαταλήψεις». Χρησιμοποιώντας πληροφορίες από το σενάριο και το γράφημα, εξήγησε γιατί αυτός ο ισχυρισμός μπορεί να είναι παραπλανητικός.

2.2 Απαντήσεις/Επιλογές Ερώτησης 2

Πλαίσιο για ανάπτυξη σύντομης απάντησης

2.3 Αποδεκτή απάντηση στην Ερώτηση 2 [και αιτιολόγηση]

Αποδεκτή απάντηση:

Η απάντηση πρέπει να εξηγεί ότι η «αντικειμενικότητα» του αλγορίθμου είναι επιφανειακή. Πρέπει να αναφέρει ένα από τα παρακάτω σημεία:

1. Ο αλγόριθμος, αν και δεν εξετάζει τη φυλή, εκπαιδεύεται σε δεδομένα που αντανακλούν ιστορικές προκαταλήψεις του συστήματος δικαιοσύνης.
2. Τα αποτελέσματα (όπως φαίνεται στο γράφημα) δείχνουν ότι ο αλγόριθμος αναπαράγει και ενισχύει τις φυλετικές ανισότητες, χαρακτηρίζοντας λανθασμένα τους Αφροαμερικανούς ως υψηλού κινδύνου συχνότερα από τους λευκούς.
3. Συνεπώς, η απουσία ανθρώπινης προκατάληψης δεν εγγυάται ένα δίκαιο αποτέλεσμα, καθώς η τεχνολογία μπορεί να «κωδικοποιήσει» και να αυτοματοποιήσει τις προκαταλήψεις που υπάρχουν στα δεδομένα.

Μη Αποδεκτή Απάντηση: Απαντήσεις που είναι υπερβολικά γενικές (π.χ. «Οι υπολογιστές κάνουν λάθη»), δεν αναφέρονται στα δεδομένα του σεναρίου ή του γραφήματος, ή συμφωνούν με τον ισχυρισμό χωρίς κριτική σκέψη.

2.4 Οδηγός Κωδικοποίησης Ερώτησης 2 [βάσει πλαισίου αξιολόγησης]

Αποδεκτή Απάντηση: Ο μαθητής συνθέτει πληροφορίες από το κείμενο και το γράφημα για να αποδομήσει τον ισχυρισμό. Αναγνωρίζει ότι η μεροληψία μπορεί να ενσωματωθεί σε ένα σύστημα ΤΝ μέσω των δεδομένων εκπαίδευσης και ότι τα αποτελέσματα του COMPAS αποτελούν απόδειξη αυτής της μεροληψίας.

Μη Αποδεκτή Απάντηση: Η απάντηση δεν συνδέει τα δεδομένα εκπαίδευσης με τα μεροληπτικά αποτελέσματα ή δεν χρησιμοποιεί τα στοιχεία από το γράφημα.

2.5 Μεταδεδομένα Ερώτησης [αιτιολόγηση βάσει πλαισίου αξιολόγησης]

Δείτε τα πλαίσια αξιολόγησης και τη χαρτογράφηση των μεταδεδομένων.

Δεξιότητα	Περιεχόμενο και Σκοπός	Τύπος Μέσου
Αναστοχασμός και Δράση με ηθικό και υπεύθυνο τρόπο	Ψηφιακή Ιθαγένεια	Ψηφιακό κείμενο , Ψηφιακά πολυμέσα
Εργαλεία	**Επίπεδο Δυσκολίας**	**Τύπος Απάντησης**
Συστήματα AI	Μεσαίο (3-4)	Ανάπτυξης

Ερώτηση 3

3.1 Διατύπωση ερώτησης 3

Να λάβεις υπόψη σου το κείμενο στα δεξιά. Κάνε κλικ σε μια επιλογή, για να απαντήσεις στην ερώτηση.

Το σενάριο αναφέρει ότι το COMPAS δεν χρησιμοποιεί τη φυλή ως μεταβλητή για τις προβλέψεις του. Ποιος είναι ο πιο πιθανός λόγος που τα αποτελέσματά του εξακολουθούν να παρουσιάζουν φυλετική μεροληψία;

3.2 Απαντήσεις/Επιλογές Ερώτησης 3

Α. Οι προγραμματιστές που δημιούργησαν τον αλγόριθμο ήταν προκατειλημμένοι.

Β. Τα ιστορικά δεδομένα με τα οποία εκπαιδεύτηκε ο αλγόριθμος περιείχαν και αναπαρήγαγαν συστημικές κοινωνικές προκαταλήψεις.

Γ. Ο αλγόριθμος είναι πολύπλοκος και κάνει τυχαία σφάλματα.

Δ. Οι υπολογιστές δεν μπορούν να πάρουν αποφάσεις για ανθρώπους.

3.3 Αποδεκτή απάντηση στην Ερώτηση 3 [και αιτιολόγηση]

Αποδεκτή απάντηση είναι η Β. Τα ιστορικά δεδομένα με τα οποία εκπαιδεύτηκε ο αλγόριθμος περιείχαν και αναπαρήγαγαν συστημικές κοινωνικές προκαταλήψεις.

Το κείμενο εξηγεί ρητά ότι «Αν τα δεδομένα που χρησιμοποιούνται για την εκπαίδευσή τους αντικατοπτρίζουν υπάρχουσες κοινωνικές και συστημικές προκαταλήψεις... ο αλγόριθμος θα μάθει και θα αναπαράγει αυτές τις προκαταλήψεις».

Μη Αποδεκτή Απάντηση: Όλες οι άλλες απαντήσεις.

3.4 Οδηγός Κωδικοποίησης Ερώτησης 3 [βάσει πλαισίου αξιολόγησης]

Αποδεκτή Απάντηση: Επιλογή της απάντησης Β. Ο μαθητής κατανοεί την έννοια της αλγοριθμικής μεροληψίας που πηγάζει από τα δεδομένα εκπαίδευσης, όπως περιγράφεται στο κείμενο.

Μη Αποδεκτή Απάντηση: Επιλογή οποιασδήποτε άλλης απάντησης (Α, Γ, Δ)

3.5 Μεταδεδομένα Ερώτησης [αιτιολόγηση βάσει πλαισίου αξιολόγησης]

Δείτε τα πλαίσια αξιολόγησης και τη χαρτογράφηση των μεταδεδομένων.

Δεξιότητα	Περιεχόμενο και Σκοπός	Τύπος Μέσου
Ανάλυση και Αξιολόγηση	Πειθώ	Ψηφιακό κείμενο
Εργαλεία	**Επίπεδο Δυσκολίας**	**Τύπος Απάντησης**
Συστήματα AI	Μεσαίο (3-4)	Απλή πολλαπλής απάντησης

Ερώτηση 4

4.1 Διατύπωση ερώτησης 4

Να λάβεις υπόψη σου το κείμενο στα δεξιά. Κάνε κλικ σε μια επιλογή, για να απαντήσεις στην ερώτηση.

Ποια είναι η κύρια συνέπεια του να χαρακτηριστεί κάποιος λανθασμένα ως «υψηλού κινδύνου» από τον αλγόριθμο COMPAS;

4.2 Απαντήσεις/Επιλογές Ερώτησης 4

Α. Μπορεί να οδηγήσει σε αυστηρότερη ποινή, υψηλότερη εγγύηση ή άρνηση αποφυλάκισης.

Β. Το άτομο θα πρέπει να παρακολουθήσει επιπλέον εκπαιδευτικά προγράμματα.

Γ. Η βαθμολογία του ατόμου θα διορθωθεί αυτόματα από το σύστημα.

Δ. Δεν υπάρχει καμία πραγματική συνέπεια, καθώς η απόφαση του δικαστή είναι η τελική.

4.3 Αποδεκτή απάντηση στην Ερώτηση 4 [και αιτιολόγηση]

Αποδεκτή Απάντηση: Α. Μπορεί να οδηγήσει σε αυστηρότερη ποινή, υψηλότερη εγγύηση ή άρνηση αποφυλάκισης.

Μια υψηλή βαθμολογία κινδύνου, ακόμα κι αν είναι λανθασμένη, είναι πιθανό να επηρεάσει αρνητικά την κρίση του δικαστή, οδηγώντας σε δυσμενέστερες αποφάσεις για τον κατηγορούμενο.

Μη Αποδεκτή Απάντηση: Όλες οι άλλες απαντήσεις.

4.4 Οδηγός Κωδικοποίησης Ερώτησης 4 [βάσει πλαισίου αξιολόγησης]

Αποδεκτή Απάντηση: Επιλογή της απάντησης Α. Ο μαθητής συνάγει λογικά τις πρακτικές συνέπειες μιας αρνητικής αξιολόγησης στο πλαίσιο του δικαστικού συστήματος.

Μη Αποδεκτή Απάντηση: Επιλογή οποιασδήποτε άλλης απάντησης (Β, Γ, Δ).

4.5 Μεταδεδομένα Ερώτησης [αιτιολόγηση βάσει πλαισίου αξιολόγησης]

Δείτε τα πλαίσια αξιολόγησης και τη χαρτογράφηση των μεταδεδομένων.

Δεξιότητα	Περιεχόμενο και Σκοπός	Τύπος Μέσου
Αναστοχασμός και Δράση με ηθικό και υπεύθυνο τρόπο	Ψηφιακή Ιθαγένεια	Ψηφιακό κείμενο
Εργαλεία	**Επίπεδο Δυσκολίας**	**Τύπος Απάντησης**
Συστήματα AI	Μεσαίο (3-4)	Απλή πολλαπλής απάντησης

Ερώτηση 5

5.1 Διατύπωση ερώτησης 5

Να λάβεις υπόψη σου το κείμενο στα δεξιά. Κάνε κλικ σε μια επιλογή, για να απαντήσεις στην ερώτηση.

Με βάση τα ευρήματα της έρευνας, ποια από τις παρακάτω ενέργειες θα μπορούσε να μειώσει τη μεροληψία σε αλγορίθμους όπως το COMPAS;

5.2 Απαντήσεις/Επιλογές Ερώτησης 5

Α. Η χρήση περισσότερων προσωπικών δεδομένων, όπως οι αναρτήσεις στα μέσα κοινωνικής δικτύωσης.

Β. Η αύξηση της επεξεργαστικής ισχύος των υπολογιστών που εκτελούν τον αλγόριθμο.

Γ. Ο έλεγχος, η διόρθωση και η απο-μεροληπτικοποίηση των ιστορικών δεδομένων που χρησιμοποιούνται για την εκπαίδευση του αλγορίθμου.

Δ. Η διατήρηση του αλγορίθμου ως μυστικό για να μην μπορεί κανείς να τον χειραγωγήσει.

5.3 Αποδεκτή απάντηση στην Ερώτηση 5 [και αιτιολόγηση]

Αποδεκτή Απάντηση: Γ. Ο έλεγχος, η διόρθωση και η απο-μεροληπτικοποίηση των ιστορικών δεδομένων που χρησιμοποιούνται για την εκπαίδευση του αλγορίθμου.

Εφόσον το πρόβλημα πηγάζει από τα μεροληπτικά δεδομένα εκπαίδευσης, η λύση περιλαμβάνει την ενεργή προσπάθεια για τη διόρθωση αυτών των δεδομένων, ώστε να μην αναπαράγονται οι ιστορικές αδικίες.

Μη Αποδεκτή Απάντηση: Όλες οι άλλες απαντήσεις.

5.4 Οδηγός Κωδικοποίησης Ερώτησης 5 [βάσει πλαισίου αξιολόγησης]

Αποδεκτή Απάντηση: Επιλογή της απάντησης Γ. Ο μαθητής αναγνωρίζει την αιτία του προβλήματος (μεροληπτικά δεδομένα) και προτείνει μια λογική λύση που στοχεύει σε αυτή την αιτία.

Μη Αποδεκτή Απάντηση: Επιλογή οποιασδήποτε άλλης απάντησης (Α, Β, Δ).

5.5 Μεταδεδομένα Ερώτησης [αιτιολόγηση βάσει πλαισίου αξιολόγησης]

Δείτε τα πλαίσια αξιολόγησης και τη χαρτογράφηση των μεταδεδομένων.

Δεξιότητα	Περιεχόμενο και Σκοπός	Τύπος Μέσου
Αναστοχασμός και Δράση με ηθικό και υπεύθυνο τρόπο	Ψηφιακή Ιθαγένεια	Ψηφιακό κείμενο
Εργαλεία	**Επίπεδο Δυσκολίας**	**Τύπος Απάντησης**
Συστήματα AI	Υψηλό (5-6)	Απλή πολλαπλής απάντησης

Παρατηρήσεις

4.1 Δημιουργός θέματος : Θεόδωρος Ζευγίτης