



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ

ΣΧΟΛΗ: ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ: ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Διεύθυνση: Μ. Αλεξάνδρου 1, Τηλ.:2610 - 369236, fax: 2610-369193

Θέμα: Εξηγήσιμη Τεχνητή Νοημοσύνη: διαφάνεια και ερμηνεία αλγορίθμων	
Επιβλέπων: Ιωάννης Ζαχαράκης	e-mail: zaharakis@uop.gr
	τηλ:
	Άτομα 2
Στόχος Η διερεύνηση των τεχνικών και των εργαλείων που καθιστούν τα μοντέλα της Τεχνητής Νοημοσύνης κατανοητά και δίνουν τη δυνατότητα να ερμηνευτούν οι προβλέψεις και οι αποφάσεις τους.	
Αντικείμενο Η Εξηγήσιμη Τεχνητή Νοημοσύνη (Explainable Artificial Intelligence) αποτελεί ένα σύνολο μεθόδων και αντίστοιχων εργαλείων για την κατανόηση της συμπεριφοράς ενός ευφυούς συστήματος, ενισχύοντας με αυτόν τον τρόπο την εμπιστοσύνη των χρηστών. Το αντικείμενο της διπλωματικής αρχικά εστιάζει στη σημασία της διαφάνειας της Τεχνητής Νοημοσύνης και στην ανάλυση της ανάγκης για την ερμηνεία των αποτελεσμάτων ενός ευφυούς συστήματος. Θα πραγματοποιηθεί μελέτη της επιστημονικής και ερευνητικής βιβλιογραφίας στο πεδίο της Εξηγήσιμης Τεχνητής Νοημοσύνης και θα καταγραφούν οι πιο γνωστές τεχνικές και εργαλεία. Θα ακολουθήσει αξιολόγηση αυτών μέσω ενδεικτικών πειραμάτων προκειμένου να εξεταστεί κατά πόσο βελτιώνεται η κατανόηση της συμπεριφοράς των χρησιμοποιούμενων αλγορίθμων και κατά πόσο ενισχύεται η εμπιστοσύνη των χρηστών.	
Η εργασία περιλαμβάνει <input type="checkbox"/> Σχεδιασμό και ανάπτυξη συστήματος <input checked="" type="checkbox"/> Συγκριτική επισκόπηση ή μελέτη, και πλαίσιο αξιολόγησης <input type="checkbox"/> Ανάλυση και σχεδιασμό μοντέλων <input type="checkbox"/> Θεωρητική μελέτη, ανάπτυξη ή ανάλυση αλγορίθμων <input type="checkbox"/> Πρότυπη κατασκευή	
Σχετιζόμενα Μαθήματα Πρωτεύοντα: Τεχνητή Νοημοσύνη, Μηχανική Μάθηση Δευτερεύοντα:	