19 Ιανουαρίου 2018

Το αδιαχώρητο στο ΑΠΘ για τον Έλληνα που έλυσε τον Γρίφο του Νας

/ Επιστήμες, Τέχνες & Πολιτισμός

Μια εκπληκτική διάλεξη για την εξέλιξη της τεχνητής νοημοσύνης τα επόμενα

«πέντε έως 50 χρόνια» έκανε στη Θεσσαλονίκη, ο Κωνσταντίνος Δασκαλάκης, καθηγητής της Επιστήμης των Υπολογιστών στο MIT που έλυσε τον, άλυτο επί 60 χρόνια, Γρίφο του Νας. Η αίθουσα του ΑΠΘ ήταν κατάμεστη με περισσότερους από 1500 φοιτητές και ακαδημαϊκούς που παρακολούθησαν τον έλληνα επιστήμονα να μιλά για τις εξελίξεις στην Τεχνητή Νοημοσύνη και πώς αυτή θα επηρεάσει την ανθρωπότητα.

Πώς θα χρησιμοποιήσει τη στατιστική ένα ρομπότ, για να αποφασίσει σε ποιον υποψήφιο πελάτη θα δώσει η τράπεζα ένα δάνειο και ποιος θεωρείται αναξιόπιστος;

Μερικά από τα ερωτήματα που έθεσε:

Μπορούν τα ρομπότ να σώσουν το παγκόσμιο ασφαλιστικό σύστημα από την επαπειλούμενη κατάρρευση; Πόσο πιθανό είναι ένας …καλοκάγαθος αλγόριθμος να εξελιχθεί σε ρατσιστή συνωμοσιολόγο μέσα σε λίγες ώρες;

Θα εναπόκειται στις ηθικές αξίες του …software ενός αυτοοδηγούμενου αυτοκινήτου να αποφασίσει ποιος ζει και ποιος πεθαίνει, όταν το όχημα "συνειδητοποιεί" ότι επίκειται ένα σοβαρό αυτοκινητιστικό ατύχημα με εμπλοκή πεζών;

> ***Πώς θα χρησιμοποιήσει τη στατιστική ένα ρομπότ, για να αποφασίσει σε ποιον υποψήφιο πελάτη θα δώσει η τράπεζα ένα δάνειο και ποιος θεωρείται αναξιόπιστος;***

Είναι εφικτό μια ομάδα ανθρώπων να "πείσει" έναν αξιόπιστο αλγόριθμο αναγνώρισης εικόνας ότι μια καραμπίνα που έχει μπροστά του είναι ένα αθώο παιδικό παιχνίδι;

Η εποχή της τεχνητής νοημοσύνης έρχεται "φορτωμένη" με υποσχέσεις, προκλήσεις και κινδύνους. Και παρότι για να ανοίξει ο νέος αυτός οικονομικός κύκλος στην ιστορία της ανθρωπότητας απαιτείται ένα μεγάλο άλμα (η μετάβαση από την τεχνητή νοημοσύνη ειδικών εφαρμογών -που ήδη "βλέπουμε" να εφαρμόζεται σε αρκετές περιπτώσεις- στη γενική τεχνητή νοημοσύνη, η πλήρης ανάπτυξη της οποίας πιθανότατα θα απαιτήσει αρκετές δεκαετίες), μια νέα συναρπαστική εποχή ήδη ανατέλλει.

Κατά τον Έλληνα καθηγητή, που μεταξύ άλλων έχει λάβει το βραβείο Kalai από την Διεθνή Ένωση Θεωρίας Παιγνίων και το βραβείο έρευνας από το ίδρυμα Giuseppe Sciacca του Βατικανού, το πιθανότερο είναι ότι σε πέντε χρόνια

από σήμερα θα έχουμε έναν προσωπικό γραμματέα με τεχνητή νοημοσύνη και αυτο-οδηγούμενα αυτοκίνητα, ενώ σε 15 χρόνια η διεπαφή του ανθρώπινου εγκεφάλου με την τεχνολογία θα γίνει ενδεχομένως πολύ πιο άμεση και το όριο που διαχωρίζει το πού ξεκινά ο άνθρωπος και πού αρχίζει η μηχανή πιο δυσδιάκριτο.

“Μπορεί όλο αυτό να ξεφύγει από τον έλεγχο; Ναι, θα μπορούσε όπως έχει συμβεί και με άλλα πράγματα στο παρελθόν. Το να είμαστε όμως αρνητικοί απέναντι στο ποτάμι που έρχεται κατά πάνω μας δεν είναι εποικοδομητικό, αυτό που πρέπει να σκεφτόμαστε, είναι πώς θα το βάλουμε στη σωστή κατεύθυνση” σημείωσε, μιλώντας σε εκδήλωση που διοργάνωσαν τα Τμήματα Πληροφορικής και Μαθηματικών της Σχολής Θετικών Επιστημών του ΑΠΘ.

> ***Το πιθανότερο είναι ότι σε πέντε χρόνια από σήμερα θα έχουμε έναν προσωπικό γραμματέα με τεχνητή νοημοσύνη και αυτο-οδηγούμενα αυτοκίνητα***

Wonderland, Pessiland, Stagnatia

Ο ίδιος ανέλυσε τρία σενάρια για την εξέλιξη της τεχνητής νοημοσύνης στα επόμενα “πέντε έως 50 χρόνια”, επισημαίνοντας ότι αυτό που πιθανότατα θα επικρατήσει είναι η μίξη τους.

Με βάση το πρώτο (θετικό) σενάριο, με τίτλο “Wonderland”, η αλληλεπίδραση ανθρώπων- μηχανών είναι θετική και ο πρώτος κερδίζει από την ύπαρξη των δεύτερων. Οι μηχανές κάνουν τις χειρονακτικές εργασίες, ο άνθρωπος έχει περισσότερο ελεύθερο χρόνο ή εκτελεί πνευματικές εργασίες και το ασφαλιστικό σύστημα σώζεται, αφού η έλλειψη νέων ανθρώπων που εργάζονται και καταβάλουν εισφορές αναπληρώνεται από την ύπαρξη των ρομπότ, που δεν χρειάζονται ασφάλιση ή σύνταξη. Προϋπόθεση για να επαληθευτεί αυτό το σενάριο είναι να κατακτήσει η επιστήμη τη γενική νοημοσύνη, δηλαδή η μηχανή να μάθει να χρησιμοποιεί τη διαίσθηση και την εμπειρία που αποκτά από μια νοητική λειτουργία και να τη μεταφέρει σε μια που δεν ξέρει καθόλου (πχ, όταν γνωρίζει να παίζει σκάκι, να μπορεί να χρησιμοποιήσει στρατηγική και στο πόκερ).

Βάσει του δεύτερου -αρνητικού- σεναρίου, με τίτλο “Pessiland”, η επιστήμη κατακτά την γενική νοημοσύνη, αλλά αυτή δεν είναι προσβάσιμη σε όλους, αλλά μόνο σε εργαστήρια εταιρειών ή κρατών, που τη χρησιμοποιούν για ιμπεριαλιστική επιρροή. “Αν πάμε σε αυτή την κατεύθυνση, το σενάριο είναι προφανώς δυστοπικό” επισήμανε ο καθηγητής.

Το τρίτο σενάριο, με τίτλο “Stagnatia”, για το οποίο ο δρ Δασκαλάκης επισήμανε ότι “έχει αρκετές πιθανότητες (επαλήθευσης)”, είναι αυτό κατά το οποίο ενώ υπάρχουν ολοένα και περισσότερες εφαρμογές ειδικής τεχνητής νοημοσύνης (πχ αναγνώριση εικόνας και ήχου ή μετάφραση), η επιστήμη δεν καταφέρνει να κάνει το άλμα στη γενική τεχνητή νοημοσύνη και επικρατεί σχετική στασιμότητα.

Κατά τον δρα Δασκαλάκη, σήμερα ένας από τους βασικούς προβληματισμούς της ανθρωπότητας είναι η αξιοπιστία της τεχνολογίας. “Υπάρχουν μεγάλα θέματα αξιοπιστίας και ένας από τους λόγους είναι ότι όταν τα δεδομένα με τα οποία τροφοδοτείς τον αλγόριθμο είναι ελλιπή ή μη αντιπροσωπευτικά, μπορεί να οδηγήσουν σε λανθασμένες ή ελλιπείς νοητικές λειτουργίες. Πχ, έγινε γνωστό ότι ένα αυτοκίνητο Tesla έπεσε σε φορτηγό σταματημένο στην αριστερή λωρίδα. Γιατί συνέβη αυτό; Ίσως γιατί ποτέ στα δεδομένα που εισήχθησαν για να προπονηθεί ο αλγόριθμος στην αναγνώριση εικόνας δεν υπήρχε αυτοκίνητο σταματημένο στην αριστερή λωρίδα του δρόμου, επειδή αυτό σπάνια συμβαίνει. Ο αλγόριθμος θα επεξεργαστεί τα ελλιπή δεδομένα που τού δώσαμε και θα ενσωματώσει την έλλειψη” σημείωσε, ενώ πρόσθεσε ότι φοιτητές του MIT επιτέθηκαν στον καλύτερο αλγόριθμο αναγνώρισης εικόνας και τον έκαναν να “πιστέψει” ότι μια τρισδιάστατη χελώνα τυπωμένη σε εκτυπωτή 3D ήταν …καραμπίνα. “Δεν έχουμε τόσο αξιόπιστη Τεχνητή Νοημοσύνη σήμερα. Προσπαθούμε να φτιάξουμε τρόπους προστασίας αλγορίθμων από τέτοιου είδους επιθέσεις” επισήμανε.

Ποιος αποφασίζει ποιος θα χάσει τη ζωή του;

‘Ενα άλλο θέμα, πρόσθεσε, έχει να κάνει με ηθικά διλήμματα. “Ενα κλασικό πρόβλημα είναι το εξής. Σκεφτείτε ότι φτιάχνουμε αυτοοδηγούμενα αυτοκίνητα που κινούνται μαζικά στους δρόμους. Αναπόφευκτα κάποιο από αυτά θα βρει τον εαυτό του σε φάση αναγνώρισης του γεγονότος ότι σε μερικά δευτερόλεπτα θα γίνει ένα αναπόφευκτο ατύχημα με εμπλοκή πεζών. Ο αλγόριθμος που οδηγεί καταλαβαίνει τότε ότι έχει δύο δυνατότητες: να πάει ευθεία και να σκοτώσει τους πεζούς ή να πάει αριστερά, να χτυπήσει στο στηθαίο και να σκοτώσει τους επιβαίνοντες. Δεν μπορεί να σώσει και τους δύο. Πώς θα πάρει την απόφαση; Ο αλγόριθμος μπορεί επίσης να καταλαβαίνει ότι οι πεζοί είναι ένα παιδάκι 8

χρονών, ο μπαμπάς του, 41, και ο σκύλος τους και οι επιβαίνοντες μια έγκυος γυναίκα 30 ετών και το αγοράκι της. Πώς εγώ που σχεδιάζω τον αλγόριθμο θα λάβω την απόφαση για το ποιος θα ζήσει";

Ο ρατσιστής αλγόριθμος

Κατά τον δρα Δασκαλάκη, η Τεχνητή Νοημοσύνη είναι σαν ένα μωρό. Το μωρό έρχεται στον κόσμο με γενετικά χαρακτηριστικά, αλλά εν πολλοίς είναι tabula rasa. Οι γονείς τού δίνουν δεδομένα και στόχους. Αν τα δεδομένα που λαμβάνει το μωρό περιέχουν ρατσιστικές απόψεις ή προκαταλήψεις ή θέσεις, αυτές τις θέσεις θα τις υιοθετήσει. Το ίδιο ισχύει και για την Τεχνητή Νοημοσύνη, η οποία μαθαίνει από την αλληλεπίδραση με τους ανθρώπους.

Χαρακτηριστικό είναι το παράδειγμα ενός chat bot (σ.σ. ρομπότ που κάνει διάλογο μέσω κειμένου ή ήχου). Μια ομάδα χρηστών του επιτέθηκε, παρέχοντάς του ρατσιστικό και συνωμοσιολογικό περιεχόμενο. "Μέσα σε 17 ώρες έγινε τρελός ρατσιστής και συνωμοσιολόγος" σημείωσε ο καθηγητής.

Τίθενται επίσης ζητήματα αμεροληψίας, γιατί αν τα δεδομένα είναι ελλιπή, η τεχνητή νοημοσύνη θα υιοθετήσει στατιστικές που δεν είναι αντιπροσωπευτικές. Κι εδώ για παράδειγμα το ερώτημα είναι: έστω πως φτιάχνω τεχνολογία που αποφαίνεται αν κάποιος είναι άξιος λήψης δανείου, αλλά έχω ελλιπή στοιχεία για μια πληθυσμιακή ομάδα. Τι γίνεται τότε; "Πρέπει να προστατέψουμε την τεχνητή νοημοσύνη από το να κάνει τέτοια στατιστικά λάθη, αλλά το πρόβλημα είναι ότι η στατιστική είναι δύσκολη επιστήμη" σημείωσε.

Ο Κωνσταντίνος Δασκαλάκης είναι απόφοιτος των Ηλεκτρολόγων του Ε.Μ.Π. Έκανε διδακτορικό στο Πανεπιστήμιο του Μπέρκλεϋ, και εργάστηκε ως μεταδιδακτορικός ερευνητής στη Microsoft. Η έρευνά του επικεντρώνεται στην θεωρητική πληροφορική και την διεπαφή της με τα Οικονομικά, την Στατιστική και την Τεχνητή Νοημοσύνη. Έχει μεταξύ άλλων τιμηθεί με το βραβείο της καλύτερης διδακτορικής διατριβής στην πληροφορική από τον διεθνή οργανισμό επιστήμης των υπολογιστών ACM, με το βραβείο Kalai από την διεθνή ένωση Θεωρίας Παιγνίων, το βραβείο εξαιρετικής δημοσίευσης από την διεθνή ένωση εφαρμοσμένων μαθηματικών SIAM, το Career Award από το Ίδρυμα Επιστημών της Αμερικής, το βραβείο Πληροφορικής του Ιδρύματος Sloan και την ερευνητική υποτροφία της Microsoft.

Πηγή:thestival/ΑΠΕ