

Ανθρώπινη αντίληψη της εμπιστοσύνης στην τεχνητή νοημοσύνη

Στόχος είναι να κατανοήσουμε πώς οι χρήστες αντιλαμβάνονται την Τεχνητή Νοημοσύνη από την οπτική γωνία της εμπιστοσύνης προς την τεχνολογία, που περιλαμβάνει πολλαπλές πιθανές προοπτικές (συμπεριλαμβανομένης της ακρίβειας, της αξιοπιστίας, της ασφάλειας, της προστασίας). Είναι απαραίτητο να προσδιοριστούν οι μορφές εμπιστοσύνης και οι διαβεβαιώσεις που αναμένονται σε διαφορετικά επίπεδα (π.χ. κατανοούν οι χρήστες ποιον εμπιστεύονται - τον υποκείμενο προγραμματιστή Τεχνητής Νοημοσύνης, τους παρόχους υπηρεσιών κ.λπ. - και σε τι τους εμπιστεύονται;). Βασικές δραστηριότητες στο πλαίσιο της έρευνας θα είναι: Η διαπίστωση των τρεχουσών αντιλήψεων για την Τεχνητή Νοημοσύνη από την οπτική γωνία της εμπιστοσύνης και της ασφάλειας. Ο προσδιορισμός παραγόντων που παρακινούν ή αναστέλλουν την εμπιστοσύνη στην Τεχνητή Νοημοσύνη. Η αξιολόγηση παραγόντων και συμβιβασμών σε διαφορετικά σενάρια εφαρμογών και περιπτώσεις χρήσης. Η αξιολόγηση και η επικύρωση του μοντέλου και των προτύπων με χρήστες και επαγγελματίες.

Choung, Hyesun, Prabu David, and Arun Ross. "Trust in AI and its role in the acceptance of AI technologies." *International Journal of Human-Computer Interaction* 39.9 (2023): 1727-1739.

Glikson, Ella, and Anita Williams Woolley. "Human trust in artificial intelligence: Review of empirical research." *Academy of management annals* 14.2 (2020): 627-660.