The ethics of algorithms: key problems and solutions

PRESENTED BY:
PAXIMADIS THEOFRASTOS
ANASTASIOU CHARALAMPOS
ANASTASOPOULOS KONSTANTINOS

Εισαγωγή

- Την τελευταία δεκαετία έχει παρατηρηθεί εκθετική ανάπτυξη στον τομέα των αλγοριθμικών συστημάτων. Σήμερα βρισκόμαστε πλέον στο σημείο εφαρμογής ***αλγορίθμων μηχανικής μάθησης***, με στόχο την επίλυση σύνθετων επιστημονικών και κοινωνικών προβλημάτων. Η διαδεδομένη χρήση τους γεννά ωστόσο ορισμένα ηθικά ζητήματα.

- Ο κλάδος της ***αλγοριθμικής ηθικής*** καλείται να αναλύσει επαρκώς και να προτείνει λύσεις στα ηθικά ζητήματα της εφαρμογής των αλγορίθμων.

Ηθικές ανησυχίες από τους αλγορίθμους.

- Inconclusive Evidence
- Instructable Evidence
- Misguided Evidence
- Unfair Outcomes
- Transformative Effects
- Traceability

Inconclusive Evidence

- Οι μη-ντετερμινιστικοί αλγόριθμοι μπορούν να παράγουν μη-σαφή αποδεικτικά στοιχεία, προκαλώντας άδικες ενέργειες. Προκύπτουν προβλήματα λόγω προκαταλήψεων, κακής ποιότητας των δεδομένων και απουσίας ανθρώπινης κριτικής σκέψης.

- Η επικύρωση και ο έλεγχος των δεδομένων ξεχωριστά σε κάθε βήμα είναι ζωτικής σημασίας.

Instructable Evidence

- Η πολυπλοκότητα των αλγορίθμων, η μεταβλητότητα του κώδικα και η δυσκολία στην επεξήγηση μοντέλων μηχανικής μάθησης οδηγούν στην έλλειψη διαφάνειας.

- Βιώσιμη λύση η δημόσια εκπαίδευση των πολιτών στα σημερινά υπολογιστικά μοντέλα και δεδομένα, για την καλύτερη κατανόηση της πολυπλοκότητάς τους.

Misguided Evidence

- Οι προγραμματιστές συχνά δίνουν προτεραιότητα στην απόδοση του αλγορίθμου παρά στο κοινωνικό πλαίσιο, οδηγώντας σε προκατάληψη στις αλγοριθμικές αποφάσεις και διαιωνισμό κοινωνικών ανισοτήτων.

- Απαιτούνται κοινωνικοτεχνικές προσεγγίσεις και έλεγχος για δίκαιη αλγοριθμική λήψη αποφάσεων.

Unfair Outcomes

- Η απουσία ορισμού και μετρικών της δικαιοσύνης ως έννοιας οδηγεί σε διακρίσεις στη λήψη αλγοριθμικών αποφάσεων.

- Λύση οι συνεργατικές μέθοδοι βασισμένες στη γνώση και τον συντονισμό του αλγοριθμικού σχεδιασμού με τις δημόσιες απόψεις.

Transformative Effects

- Οι απρόβλεπτες επιπτώσεις που μπορεί να έχουν οι αλγόριθμοι εις βάρος των χρηστών τους, κυρίως στην αυτονομία τους και στην ιδιωτικότητα τους.

- Ανάγκη συμμετοχής των χρηστών στον σχεδιασμό των συστημάτων.
- Αναδυόμενες μέθοδοι προστασίας του απορρήτου, όπως το διαφορικό απόρρητο.

Traceability

- Η δυσκολία απόδοσης ηθικών ευθυνών για τις αποφάσεις των αλγορίθμων, εξαιτίας της πολυπλοκότητας, έλλειψης διαφάνειας τους και συχνά στην έλλειψη νομοθεσίας.

- Αποσύνδεση της ηθικής ευθύνης από την σκοπιμότητα (intentionality) και αναδιατύπωση της ως συλλογική ευθύνη (collective responsibility) όσων εμπλέκονται.

Συμπεράσματα

- Η χρήση των αλγορίθμων και των τεχνολογιών της τεχνητής νοημοσύνης απαιτεί συνεχείς επιβλέψεις, αναλύσεις και διαβουλεύσεις. Η διαρκής παρακολούθηση και αναθεώρηση της νομοθεσίας και των ηθικών κατευθυντήριων γραμμών είναι ζωτικής σημασίας προκειμένου να αντιμετωπιστούν οι επικίνδυνες επιπτώσεις των αλγορίθμων και παράλληλα να εφαρμοστούν για το κοινό καλό.

ΕΥΧΑΡΙΣΤΟΥΜΕ ΓΙΑ ΤΟΝ ΧΡΟΝΟ ΣΑΣ