

Frankenstein's AI Lab

The image depicts a futuristic, dimly lit laboratory with a gothic architectural style. The room is filled with rows of computer workstations, each equipped with multiple monitors displaying various data and code. The lighting is predominantly blue and green, creating a cold, technological atmosphere. In the foreground, a humanoid robot stands with its back to the camera, looking towards the workstations. The room features large, arched windows that allow some light to filter in, and the floor is covered with a grid pattern. The overall scene suggests a high-tech environment where artificial intelligence is being developed or monitored.

Dr Vasilis Vasilopoulos technoethicist



Που είναι ο

άνθρωπος;

Η ευφυΐα της μηχανής

Η προσομοίωση της ανθρώπινης σκέψης που νοείται ως υπολογισμός (αριθμητικός και λογικός), η εκτέλεση διανοητικών λειτουργιών των ανθρώπων (ταξινόμηση, δημιουργία κανόνων, ανακάλυψη νόμων, σχέσεων...), η ορθολογική λειτουργία (με κίνητρο το σχέδιο και τη γνώση του περιβάλλοντος), εκλαμβάνεται ως ευφυΐα της μηχανής.

Ωστόσο, κάθε ανθρώπινη σκέψη είναι ένας μαθηματικός υπολογισμός. Για να προσδιοριστεί λοιπόν καλύτερα η έννοια της ευφυΐας του ανθρώπου και της μηχανής απαιτείται η συμφωνία στις παραδοχές ότι η έννοια του απείρως μεγάλου συνόλου επεκτείνεται πέρα από τον υλικό κόσμο.

Πχ: η πίστη στο άπειρο και το άγνωστο είναι σαν την πίστη στον Θεό. Η απόδειξη της ανωτερότητας του ανθρώπινου νου έναντι των μηχανών φαίνεται ανέφικτη στο πλαίσιο της μαθηματικής λογικής, αλλά η ανθρώπινη σκέψη εκτείνεται πέρα από τον υπολογισμό και τη λογική εξαγωγή συμπερασμάτων. Μια μηχανή δεν μπορεί να επιτύχει μια κατάσταση συνείδησης που να της επιτρέπει να συνειδητοποιήσει τη σχέση της με το περιβάλλον και να πει "είμαι" σε σχέση με το ανθρώπινο "εγώ" του καθενός. Οι κοινωνικές και συναισθηματικές σχέσεις, το αίσθημα ντροπής ή η ευγνωμοσύνη είναι εξαιρετικά δύσκολο να προσομοιωθούν. Επιπλέον, έννοιες όπως η ελπίδα είναι έννοιες που υπερβαίνουν την πραγματικότητα του "εδώ και τώρα", και ως εκ τούτου, διαθέτουν ένα στοιχείο υπερβατικότητας που χαρακτηρίζει την ανθρώπινη σκέψη. Το σωστό ή δίκαιο και το καλό, έναντι του λάθους και του κακού, για το άνθρωπο είναι απλή ανάλυση επειδή έχει το εργαλείο να τα εντάσσει σε ένα πλαίσιο.

Στη μηχανική νοημοσύνη (ως ικανότητα απόκτησης και επεξεργασίας πληροφοριών για την επίτευξη προκαθορισμένων στόχων), πραγματοποιείται ευφυής εκτέλεση εργασιών με τη χρήση ευρετικών μεθόδων, οι οποίες επιτρέπουν τον περιορισμό του πεδίου αναζήτησης στο πεδίο του προβλήματος.

Monsters and Monstrosities

Οι εφαρμογές, όπως το ChatGPT, εκπαιδεύονται σε ποικίλες μαζικές ποσότητες δεδομένων κειμένου από το Διαδίκτυο (κυρίως από δημοσιεύσεις που στην πλειονότητά τους είναι περιεχόμενο μιντιακών οργανισμών). Είναι σε θέση να παρακολουθούν το πλαίσιο μιας συνομιλίας, είναι νευρωνικά δίκτυα σχεδιασμένα για την εκμάθηση στατιστικών εξαρτήσεων μεγάλης εμβέλειας σε κείμενα. Η εκπαίδευση ενός γλωσσικού μοντέλου περιλαμβάνει την πρόβλεψη επόμενων λέξεων (Next-Token-Prediction, NTP και τεχνικές απόκρυψης λέξεων MLM). Στην τεχνική MLM, ορισμένες λέξεις σε μια ακολουθία "αποκρύπτονται" και το μοντέλο εκπαιδεύεται για να προβλέπει τις σωστές λέξεις με βάση τα συμφραζόμενα του περιβάλλοντος κειμένου. Για παράδειγμα, με δεδομένη την είσοδο "το ποτήρι ..κάτω από το τραπέζι", το μοντέλο εκτιμά την πιθανότητα μιας πιθανής επόμενης λέξης, όπως "έπεσε" ή "βρίσκεται", λαμβάνοντας υπόψη τη συχνότητα εμφάνισης αυτών των λέξεων σε άλλα κείμενα.

- Τα LLMs είναι σε θέση να παρακολουθούν το πλαίσιο μιας συνομιλίας.
- Η εκπαίδευσή τους απαιτεί τεράστιες ποσότητες πόρων και είναι ενεργοβόρα.
- Το ChatGPT διαθέτει 1,7 τρισεκατομμύρια παραμέτρους μηχανικής μάθησης.
- Η εκπαίδευση του μοντέλου GPT-4 απασχολεί 25.000 GPU, κατά τη διάρκεια 90-100 ημερών, με κόστος 100 εκατομμύρια δολάρια, ενώ η ενέργεια ήταν ισοδύναμη με την ενέργεια που καταναλώνουν 1.000 νοικοκυριά των ΗΠΑ κατά τη διάρκεια 5-6 ετών.
- Το κόστος εξαγωγής συμπερασμάτων και η χρήση ενέργειας είναι 10 φορές υψηλότερο από το κόστος εκπαίδευσης.
- Σε 1 μήνα, το ChatGPT κατανάλωσε τόση ηλεκτρική ενέργεια όση 26.000 νοικοκυριά των ΗΠΑ..
- Τα υψηλά αυτά ενεργειακά κόστη κάνουν τα LLMs να αναπτύσσονται από πολύ πλούσιες ιδιωτικές εταιρείες και όχι από πανεπιστημιακά ιδρύματα μολονότι στα ακαδημαϊκά ερευνητικά εργαστήρια μπορούν να αναπτυχθούν τα βασικά μοντέλα. Αυτό εξηγείται από την αδυναμία των πανεπιστημίων να διαθέτουν τους πόρους (χρήματα, ενέργεια και δεδομένα).

Γιατί είμαστε ανασφαλείς;

- Το 65% των χρηστών του διαδικτύου ηλικίας 16-64 ετών ανησυχούν ότι η τεχνητή νοημοσύνη θα μπορούσε εύκολα να χρησιμοποιηθεί για ανήθικους σκοπούς.
- Περισσότεροι από τους μισούς ανησυχούν τώρα για το πώς τα εργαλεία AI θα επηρεάσουν τους καλλιτέχνες και αυτό το ποσοστό έχει επίσης αυξηθεί.
- Λιγότεροι από τους μισούς χρήστες του διαδικτύου ηλικίας 16-64 ετών εξακολουθούν να ανησυχούν για τον τρόπο ανάπτυξης των εργαλείων τεχνητής νοημοσύνης

- Οι περισσότεροι άνθρωποι δεν εμπιστεύονται πλήρως τις πληροφορίες που δημιουργούνται από AI. Μόνο το 27% των χρηστών του διαδικτύου ηλικίας 16-64 ετών λένε ότι εμπιστεύονται τις πληροφορίες που παράγονται από την τεχνητή νοημοσύνη (πολύ ή εντελώς).
- Το 32% εμπιστεύεται το εκπαιδευτικό περιεχόμενο που παράγεται με τεχνητή νοημοσύνη.
- Οι περισσότερες εφαρμογές δεν παρέχουν τις πηγές που χρησιμοποιούνται για τη δημιουργία του περιεχομένου, καθιστώντας πιο δύσκολο τον έλεγχο των πληροφοριών.



Ψευδαισθήσεις: Η τεχνητή νοημοσύνη δημιουργεί ψευδείς ή παραπλανητικές εικόνες και πληροφορίες

Ενίσχυση μεροληψίας: διαιώνιση υπαρχουσών προκαταλήψεων στην κοινωνία

Παραβίαση πνευματικών δικαιωμάτων: μη εξουσιοδοτημένη χρήση υλικού που προστατεύεται από πνευματικά δικαιώματα

Παραβίαση προσωπικών δεδομένων: ο άνθρωπος γίνεται μέρος ενός μηχανικού συνόλου, ο χρήστης μηχανικός και προϊόν ταυτόχρονα και κάθε τρίτος γίνεται πρώτη ύλη για νέα προϊόντα.

Ο αντίκτυπος στο media space

Η ανάγκη για νέες δεξιότητες: έλεγχος στοιχείων, κριτική σκέψη, και ηθική χρήση τεχνητής νοημοσύνης

Ο ρόλος των χρηστών στην ανάπτυξη και ρύθμιση της τεχνητής νοημοσύνης

Η αυτοματοποιημένη παραγωγή και οι κίνδυνοι για τις επιχειρήσεις (media, διαφημιστές και διαφημιζόμενοι)

Μετριασμός των κινδύνων:

Human -in- the -loop : η σημασία της ανθρώπινης επίβλεψης

Prebunking : προληπτική αντιμετώπιση της παραπληροφόρησης

Ηθικές κατευθυντήριες γραμμές: ανάπτυξη προτύπων για την τεχνητή νοημοσύνη

Διαφάνεια

Επιχειρηματικά Μοντέλα και Υβριδικά Κοινά Πελατών στην Εποχή της Τεχνητής Νοημοσύνης

Η τεχνητή νοημοσύνη (TN) αναδιαμορφώνει τον τρόπο με τον οποίο οι επιχειρήσεις αλληλεπιδρούν με τους πελάτες τους.

Τι είναι τα υβριδικά κοινά;

Πρόκειται για κοινότητες που αποτελούνται τόσο από ανθρώπους όσο και από μηχανές, οι οποίες συνεργάζονται, καταναλώνουν και παράγουν υπεραξίες για την επίτευξη κοινών στόχων. Η TN παίζει καθοριστικό ρόλο σε αυτές τις κοινότητες, καθώς επιτρέπει στις μηχανές να συμμετέχουν σε συζητήσεις, να παρέχουν εξατομικευμένες προτάσεις και να εκτελούν εργασίες που απαιτούν ανθρώπινη νοημοσύνη.

Πώς επηρεάζει η TN τα επιχειρηματικά μοντέλα;

Εξατομίκευση: Η TN επιτρέπει στις επιχειρήσεις να προσφέρουν εξατομικευμένες εμπειρίες σε κάθε πελάτη, αναλύοντας μεγάλους όγκους δεδομένων και προσαρμόζοντας τα προϊόντα ή τις υπηρεσίες τους στις ανάγκες του καθενός.

Αυτοματοποίηση: Πολλές επαναλαμβανόμενες εργασίες μπορούν να αυτοματοποιηθούν, επιτρέποντας στους ανθρώπους να επικεντρωθούν σε πιο δημιουργικές και στρατηγικές εργασίες.

Νέα προϊόντα και υπηρεσίες: Η TN διευρύνει τις δυνατότητες δημιουργίας νέων προϊόντων και υπηρεσιών, όπως οι προσωπικοί βοηθοί και οι έξυπνες συσκευές.

Ο ρόλος των υβριδικών κοινοτήτων

Οι υβριδικές κοινότητες παρέχουν ένα πλαίσιο όπου οι άνθρωποι και οι μηχανές μπορούν να συνεργαστούν για την επίλυση προβλημάτων, την ανάπτυξη νέων ιδεών και τη δημιουργία καινοτόμων προϊόντων. Αυτές οι κοινότητες μπορούν να οδηγήσουν σε:

Αύξηση της ικανοποίησης των πελατών: Οι πελάτες μπορούν να λάβουν πιο γρήγορες και αποτελεσματικές απαντήσεις στα αιτήματά τους.

Βελτίωση της ενεργούς συμμετοχής των πελατών: Οι πελάτες μπορούν να συμμετάσχουν ενεργά στη διαμόρφωση των προϊόντων και των υπηρεσιών.

FRANK E STEIN

AI LAB



Agentic AI:

Purpose: Autonomy.

Acts independently, adapting to environments. Examples: Autonomous marketing systems, intelligent business agents. Human Interaction: Minimal.



Generative AI:

Purpose: Creativity. Produces new content like text or images. Examples: GPT-4, DALL-E. Human Interaction: Works flexibly with user prompts.



Traditional ML:

Purpose: Data-driven decisions. Recognizes patterns and makes predictions. Examples: Fraud detection, regression models. Human Interaction: Requires input for training and updates.