

Η μεγάλη των «πρακτόρων» εποχή

Το επόμενο σύνορο στην επανάσταση της Τεχνητής Νοημοσύνης είναι -όπως όλα δείχνουν- οι πράκτορες (Agents) και η εποχή τους (Agentic AI) που τώρα ανατέλλει. Μιλάμε πλέον για «μηχανές» με σχετική ανεξαρτησία, που εκτελούν τις εντολές μας και συνεννοούνται μεταξύ τους όποτε χρειαστεί, για να εκπληρώσουν την αποστολή τους. Είμαστε, όμως, σήμερα έτοιμοι και είναι άραγε η κατάλληλη στιγμή για να τους «παραδώσουμε τα κλειδιά»;

ΤΟΥ ΓΙΑΝΝΗ ΡΙΖΟΠΟΥΛΟΥ





Η ιστορία της ανθρωπότητας είναι άρρηκτα συνδεδεμένη με την τεχνολογία. Οι βιομηχανικές επαναστάσεις -σήμερα είμαστε στην τέταρτη και προχωράμε γοργά προς την πέμπτη- αποτελούν σημεία καμπής, καθώς αλλάζουν συθέμελα τη ζωή μας. Μάλιστα, όσο περνούν τα χρόνια, ο ρυθμός των αλλαγών ολοένα επιταχύνεται – οι εξελίξεις είναι ραγδαίες, ειδικά από το «γύρισμα» της χιλιετίας και μετά.

Όσα θεωρούμε σήμερα αναπόσπαστο μέρος της καθημερινότητάς μας δεν έχουν παρά λίγα χρόνια ιστορίας πίσω τους: το πρώτο iPhone εμφανίστηκε το 2007, τα κοινωνικά δίκτυα μόλις λίγα χρόνια αργότερα, το ChatGPT, με το οποίο συνδέουμε την «έκρηξη» της Τεχνητής Νοημοσύνης (που ήδη μετρούσε κάμποσες δεκαετίες ζωής, αλλά όχι για το ευρύ κοινό) έκανε πρεμιέρα μόλις στις 30 Νοεμβρίου του 2022, δεν έχει καν κλείσει τα τρία χρόνια του!

Πριν προλάβουμε να γνωρίσουμε κάτι καλά, περνάμε στο επόμενο, ανοίγει νέος, ευρύτερος και ακόμα πιο εντυπωσιακός κύκλος. Από τους υπολογιστές (εκεί εντάσσονται ουσιαστικά κι αυτοί που έχουμε καθημερινά στην τσέπη μας, τα κινητά μας) και τα συμβατικά δίκτυα επικοινωνίας περάσαμε στο cloud και την Τεχνητή Νοημοσύνη, που με τη σειρά της φέρνει γρήγορα κάτι καινούργιο. Ανοίγει την πόρτα σε μια νέα εποχή, που –όπως όλα δείχνουν- θα αλλάξει πάρα πολλά στη ζωή μας, ξεκινώντας από τον τρόπο που εργαζόμαστε.

Δυο όροι κυριαρχούν σ' αυτήν: Agents και Agentic AI – δεν είναι το ίδιο και καλό θα ήταν να το εξηγήσουμε εξ αρχής, καθώς αποτελούν διαφορετικά «δείγματα γραφής», ενταγμένα βεβαίως στο γενικότερο οικοσύστημα της Τεχνητής Νοημοσύνης, αλλά με τις δικές του προδιαγραφές και δυνατότητες το καθένα. Οι «πράκτορες» (Agents) είναι αυτόνομο λογισμικό, το οποίο έχει σχεδιαστεί για να κάνει συγκεκριμένες εργασίες, με την ελάχιστη δυνατή ανθρώπινη παρέμβαση. Πιο εξελιγμένοι από τους ψηφιακούς βοηθούς, αξιοποιούν μεγάλα γλωσσικά μοντέλα (LLMs), όπως το δημοφιλέστατο GPT-4, σε συνδυασμό με άλλα εξωτερικά εργαλεία (πχ. έρευνες στο web, APIs, κλπ.) για να επεξεργαστούν δεδομένα και να μας δώσουν τα αποτελέσματα που τους ζητήσαμε είτε εγγράφως, πληκτρολογώντας το αίτημά μας, είτε «μιλώντας τους» με φυσική γλώσσα.

Το Agentic AI «ανεβαίνει πίστα», καθώς κάνει περισσότερα απ' ό,τι ένας πράκτορας. Εδώ μιλάμε πλέον για συνεργατικά σχήματα όπου πολλοί εξειδικευμένοι σε κάποιο πεδίο πράκτορες προσφέ-

ρουν συνεταιριστικά τις υπηρεσίες τους, προκειμένου να αντιμετωπίσουν πολύπλοκες και δυναμικά εξελισσόμενες απαιτήσεις. Για να το καταφέρουν «αποδομούν» τους δύσκολους στόχους σε πολλές, μικρότερες αποστολές, «μοιράζουν τη δουλειά», βάσει συγκεκριμένων πρωτοκόλλων μεταξύ τους επικοινωνίας, και επανέρχονται με το συνθετικό αποτέλεσμα που τους έχει ζητηθεί, παραδίδοντας λύσεις ή προτάσεις.

Ήδη, τέτοια συστήματα αξιοποιούνται πλέον και στη χώρα μας, με τα πιλοτικά να δίνουν και να παίρνουν, ενώ -σε παγκόσμιο επίπεδο και σύμφωνα με την εταιρεία ερευνών Statista- η αξία της αγοράς των AI agents προβλέπεται να αυξηθεί από τα 5,1 B\$ σήμερα, στα 47 B\$ ως το 2030. Σε κάθε εργαζόμενο θα αναλογούν πολύ περισσότεροι πράκτορες-συνεργάτες και οι ειδικοί λένε πως όσοι οργανισμοί υιοθετούν αυτήν την τεχνολογία αποκτούν πλέον ουσιαστικό ανταγωνιστικό πλεονέκτημα.

Τα υπέρ και τα κατά

Οι πράκτορες διακρίνονται για την αυτονομία τους χωρίς να χρειάζονται συχνή επιτήρηση, την αφοσίωσή τους στον προδιαγεγραμμένο στόχο, τη δυνατότητά τους να αντιδρούν άμεσα σε «εξωτερικά ερεθίσματα» και να προσαρμόζονται στις νέες καταστάσεις. Έτσι, παίρνουν πολύ καλό βαθμό, για παράδειγμα, στην εξυπηρέτηση πελατών, στον έλεγχο κώδικα, στην υποστήριξη αυτοματισμών ή στην εκτέλεση επαναλαμβανόμενων εργασιών.

Όμως, την ίδια ώρα, ένα σύστημα Agentic AI μπορεί να συντονίσει και να ελέγξει -όταν του δοθεί η ανάλογη εντολή- την «μπριγάντα» του, τους «απλούς» πλην εξειδικευμένους πράκτορες με τους οποίους συνεργάζεται σε πραγματικό χρόνο, βάσει προδιαγραφών και περιορισμών όπου χρειάζεται, φέρνοντας σε πέρας πολύ πιο σύνθετες εργασίες: από τον συντονισμό ρομποτικών διατάξεων σε ένα εργοστάσιο και τη λειτουργία κρίσιμων συστημάτων σε μεγάλες κτηριακές εγκαταστάσεις ως την πληρωμή λογαριασμών ή ακόμα και την οργάνωση των διακοπών του χρήστη του.

Βέβαια, στα περισσότερα από αυτά, το σύστημα καλείται να πάρει αποφάσεις -συχνά κρίσιμες- κι εδώ αρχίζουν τα δύσκολα, καθώς η παροχή εξουσιοδότησης στη «μηχανή» προκειμένου να μας εκπροσωπεί σε απλές ή δυσκολότερες εργασίες, δεν είναι απλή υπόθεση ούτε πρέπει να θεωρείται τέτοια. Συνοδεύεται από πολλούς αστερίσκους και ενέχει κινδύνους, αν δεν γίνει με τον σωστό τρόπο, ενώ ο τελευταίος λόγος και η τελική έγκριση πρέπει να ανήκει πάντα στον άνθρωπο-χρήστη.

Για όλα αυτά κι ακόμα περισσότερα, μιλήσαμε με δυο ειδήμονες – καθ' ύλην αρμόδια στελέχη του ΕΚΕΦΕ 'Δημόκριτος', τον Γιώργο Γιαννακόπουλο, Ερευνητή (ΕΛΕ Β') Τεχνητής Νοημοσύνης στο Ινστιτούτο Πληροφορικής & Τηλεπικοινωνιών και συνιδρυτή των εταιρειών καινοτομίας ahedd DIH S.A. και SciFY A.M.K.E, και την Φωτεινή Σαλμούκα, Αναπληρώτρια Γενική Διευθύντρια - ahedd DIH S.A. και Υπεύθυνη Έργων, SmartAttica EDIH. Να τι μας είπαν, μεταξύ πολλών άλλων, απαντώντας στις ερωτήσεις μας.

Καινοτομία και ρίσκα

Γιώργος Γιαννακόπουλος: Η καινοτομία στους πράκτορες είναι στη φυσική γλώσσα επικοινωνίας και στο γεγονός ότι δημιουργούνται πλέον κάποια τεχνικά πλαίσια: δίνω τη δυνατότητα σε ένα σύστημα να κάνει -έπειτα από δικό μου αίτημα και επιβεβαίωση, πάντα- παρέμβαση στον υπολογιστή μου, στα δεδομένα μου, στην επιχείρησή μου. Το ρίσκο είναι η απόφαση της εκπροσώπότης μου, για κάποιους πιο έξυπνους αυτοματισμούς, στον υπολογιστή μου, απ' ότι συνήθως

Φωτεινή Σαλμούκα: Πάμε από μια αντιστροφή της εργαλειοποίησης στη λογική της αντιπροσώπευσης. Όταν μιλάμε για πράκτορες, πρακτικά εννοούμε δικούς μας αντιπροσώπους. Γι' αυτό πολλές φορές τους προσδίδουμε διάφορα ανθρωπομορφικά χαρακτηριστικά, ώστε να προσομοιάζουν στη δική μας συμπεριφορά. Μπορούν να ενεργούν αυτόματα ή ημιαυτόματα, πάντα υπό την επιτήρησή μας, να σχεδιάζουν ενέργειες και να μας εκπροσωπούν σε διαδικασίες. Ουσιαστικά, οδηγούμαστε στη μετατόπιση της ευθύνης και στη λήψη τη απόφασης, η οποία δεν γίνεται πια από εμάς, αλλά από τον agent. Περνάμε από την παθητική στην ενεργητική πρωτοβουλία εκ μέρους του. Οι πράκτορες μπορούν να πάρουν την απάντηση σε ό,τι ρωτήσουμε, να τη συνδυάσουν με άλλα δεδομένα και να προχωρήσουν μόνοι τους σε νέα ερωτήματα ή και δράσεις. Πρακτικά, αν του ζητήσουμε να βρει μια πτήση για τον Χ προορισμό, ο πράκτορας μπορεί πλέον να συνδυάσει όχι μόνο τα δεδομένα για το πώς και πότε μπορώ να πάω εκεί, αλλά να συγκρίνει ξενοδοχεία, να ρωτήσει και να προσαρμόσει τις προτιμήσεις μου ανάλογα με τα ενδιαφέροντά μου και να προτείνει ένα ολοκληρωμένο πλάνο ταξιδιού. Βεβαίως, εγώ θα λάβω τις τελικές αποφάσεις και θα το υιοθετήσω ή όχι.

Γ.Γ: Σε όλα αυτά δεν γίνεται κάτι ερήμην μας. Οι προγραμματιστές πχ. τώρα με το Vibe coding κι άλλες τέτοιες τάσεις, που έρχονται και παρέρχονται για το πώς συ-



«Η μεταφορά της υπευθυνότητας στα συστήματα είναι μια επανάσταση, ως προς την κλίμακά της»

Γιώργος Γιαννακόπουλος,
Ερευνητής (ΕΛΕ Β') Τεχνητής Νοημοσύνης στο Ινστιτούτο Πληροφορικής & Τηλεπικοινωνιών και συνιδρυτή των εταιρειών καινοτομίας aheadd DIH S.A. και SciFY A.M.K.E



«Απαιτούνται μηχανισμοί επιτήρησης αυτών των εργαλείων και των στόχων που έχουν τεθεί»

Φωτεινή Σαλμούκα,
Αναπληρώτρια Γενική Διευθύντρια - aheadd DIH S.A. και Υπεύθυνη Έργων, SmartAttica EDIH

νεργάζομαι με τον υπολογιστή, μπορούν να προγραμματίζουν με τη βοήθειά του. Αυτό πάντα γινόταν, αλλά τώρα του ζητούν συμβουλές για το πώς θα κάνουν κάτι και, με βάση τη γνώση που έχει συλλέξει από άλλους ανθρώπους, το σύστημα κάνει προτάσεις, οπότε του λένε 'σου δίνω την άδεια να κάνεις εσύ αυτό που μου πρότείνεις, για να φτιάξω καλύτερα αυτό που θέλω'. Η δυνατότητα δράσης σε έναν περιορισμένο χώρο που τον ορίζω εγώ, είναι η βασική μεταφορά αυτής της υπευθυνότητας στα συστήματα. Κι αυτό είναι μια επανάσταση ως προς την κλίμακά του.

Φ.Σ: Μιλάμε πλέον για μια ποιοτική αλλαγή, όχι για ποσοτικά δεδομένα και υπολογιστική ισχύ, για αλλαγή στάσης.

Δεν έχουμε ένα έξυπνο πρόγραμμα, αλλά έναν ενεργό συνεργάτη δίπλα μας.

Εντολές και επιφυλάξεις

Γ.Γ: Αυτό το επίπεδο της αλληλοεπίδρασης, πάντως, ορίζεται ανά εργασία, όχι επ' άπειρον - δεν σε εκπροσωπεί πλήρως... Κάθε φορά του δίνεις εντολή ή (επειδή είναι και έξυπνο) μια πάγια εντολή, πχ. να σου πληρώνει τους λογαριασμούς. Το στοίχημα της εποχής είναι να θέτουμε καλά το πλαίσιο. Αν, παρόλα αυτά, το σύστημα κάνει λάθος; Αυτό είναι το μεγάλο ζήτημα, ηθικό και νομικό - υπάρχουν πλέον ερωτήματα του τύπου 'μπορεί να έχει νομική ευθύνη και έως πού, ένας ψηφιακός βοηθός; Την έχει μήπως ο προγραμματιστής, ο χρήστης

ή κάποιος άλλος; Αυτά δεν έχουν λυθεί κι επίσης, μπαίνουμε σε λίγο πιο ενδιαφέροντα μονοπάτια, όταν ρωτήσουμε αν ένας αλγόριθμος πετύχει μια καινοτομία, μπορεί άραγε να την κατοχυρώσει; Η ΕΕ για την ώρα λέει όχι – το εγκρίνει μόνο για ζώντες οργανισμούς, που έχουν ηθική υπόσταση και μπορούν να πάρουν ηθικές αποφάσεις... Πολύ περισσότερο όταν ακόμα και τα δικαστήρια αυτοματοποιούνται – έχουμε ήδη αποφάσεις που βγαίνουν με τη βοήθεια ΤΝ. Στην Ευρώπη δεν μπορούν να ληφθούν αποφάσεις ερήμην των ανθρώπων – δικαστών και σίγουρα όχι αυτόματα, από αλγόριθμο. Ουσιαστικά, τώρα σκιαγραφούμε μέχρι πού και πόσο θέλουμε να χρησιμοποιούνται οι agents.

Φ.Σ.: Το κομμάτι της οριοθέτησης είναι το πιο σημαντικό, από πλευράς σχεδίασης, τεχνολογίας και ηθικής. Ποιος θα αναλάβει την ευθύνη και τη λογοδοσία - με δεδομένο ότι σήμερα οι πράκτορες δεν έχουν νομική υπόσταση, άρα δεν μπορούν να κατηγορηθούν; Αυτά παραμένουν στον χρήστη και τον οργανισμό που τους ανέθεσε τον ρόλο, αλλά περισσότερο σε όσους σχεδιάζουν συστήματα. Είναι ένα βάρος που αυξάνει για τους επιστήμονες, γι' αυτό βλέπουμε την ανάγκη διεπιστημονικής συνύπαρξης στις ομάδες, προκειμένου να διασφαλιστεί κατά το δυνατόν η τήρηση των κανονιστικών πλαισίων και που, ακόμα κι εκείνα, δεν μας καλύπτουν για τη συζήτηση που κάνουμε σήμερα. Δηλαδή, το AI Act δεν καλύπτει επαρκώς τον τομέα των πρακτόρων, αφού δεν υπήρχαν την προηγούμενη περίοδο.

Γ.Γ.: Νιώθω ότι πρόκειται για κοινωνική προσπάθεια, συλλογική κοινωνική προσπάθεια. Δεν είναι τεχνολογική, ούτε νομική, αλλά μια κοινωνική συλλογική αναζήτηση, ως προς το ποια πράγματα θέλουμε ή όχι. Εκεί το AI Act και οι προηγούμενες προσπάθειες, όλες οι προσπάθειες περί των δικαιωμάτων του ανθρώπου κλπ., είναι πολύ σημαντικές προκειμένου να σκιαγραφήσουμε το μέλλον των agents, όπως είχε χρειαστεί και χρειάζεται ακόμα

να το σκιαγραφούμε για τα δεδομένα μας.

Φ.Σ.: Αυτό μας φέρνει πιο κοντά στην πάντα επίκαιρη συζήτηση, στο κομμάτι της εκπαίδευσης και της κατάρτισης. Πλέον γίνεται ακόμα πιο επιτακτική και πρέπει να ξεκινάει σχεδόν από τη νηπιακή ηλικία, ακόμα και για θέματα που αφορούν περισσότερο σε έννοιες και δεν έχουν να κάνουν με τεχνολογικές δεξιότητες. Μόνο μ' αυτόν τον τρόπο (και ο 'Δημόκριτος' κινείται ενεργά σ' αυτό το πεδίο, εκπαιδεύοντας εκπαιδευτικούς σε τέτοια θέματα) θα μπορέσουμε αφενός να κατανοήσουμε πώς λειτουργούν αυτά τα συστήματα, αφετέρου να υπάρξει μια γενικότερη ευαισθητοποίηση. Δηλαδή ο ίδιος ο πολίτης να ενδιαφέρεται για το τι κρύβεται πίσω από το σύστημα, πού δίνω τα δεδομένα μου, για ποιο σκοπό τα δίνω, ποιος ακριβώς είναι ο τρόπος που αυτά τα συστήματα παίρνουν αποφάσεις. Και φυσικά είναι αναγκαία η ευαισθητοποίηση στο κομμάτι της ανίχνευσης των προκαταλήψεων και των στερεοτύπων που όλα τα μοντέλα προβάλλουν και αναπαράγουν με έναν τρόπο πολλές φορές πιο ραγδαίο και εκθετικό από την κοινωνία. Αυτή θα μας επιτρέψει να ξεχωρίσουμε τι είναι χρήσιμο και τι ανακριβές σε ό,τι κάνουμε.

Ατροφία κρίσης και εμπιστοσύνη

Φ.Σ.: Έχουμε, επίσης, να αντιμετωπίσουμε δύο πολύ σημαντικές προκλήσεις οι οποίες είναι απολύτως πραγματικές: Η ατροφία κρίσης, δηλαδή ο κίνδυνος της υπερεξάρτησης από τη μηχανή, είναι σαν το GPS που, πολλούς από εμάς, μας έχει πάει στον γκρεμό. Η τυφλή εμπιστοσύνη σε ένα σύστημα, ως προς τη λήψη απόφασης, έρχεται να μειώσει τη δυνατότητα αυτόνομης κρίσης και το κομμάτι του ψυχολογικού αντικτύπου που έχει η χρήση και η αλληλοεπίδραση με τους πράκτορες. Εδώ έχουμε κάποιες νέες ψυχολογικές δυναμικές οι οποίες όλο και πιο ραγδαία έρχονται στη συζήτηση, δηλαδή το κομμάτι της εμπιστοσύνης, το

κομμάτι της απογοήτευσης. Η εμπιστοσύνη δεν πρέπει να είναι τυφλή. Η άλλη είναι το θέμα της υπερ-προσωποποίησης, δηλαδή βλέπουμε από πίσω μια οντότητα, ένα νόημα νοι. Νομίζουμε πολλές φορές ότι αλληλοεπιδρούμε με έναν άλλο άνθρωπο κι όχι με ένα εργαλείο που έχει σχεδιαστεί από έναν άνθρωπο, με όλες τις μεροληψίες που μπορεί να φέρει. Επίσης, πολύ κρίσιμο, είναι το κομμάτι της εσωτερικής διακυβέρνησης. Οι πράκτορες, όπως και τα συστήματα του Agent AI, δεν θα έρθουν να λειτουργήσουν στο κενό. Χρειάζονται πολιτικές για την ασφάλεια των δεδομένων, την προστασία της ιδιωτικότητας, το κομμάτι της λογοδοσίας για τις αποφάσεις που λαμβάνονται. Άρα, απαιτούνται μηχανισμοί επιτήρησης αυτών των εργαλείων και των στόχων που έχουν τεθεί.

Γ.Γ.: Και βέβαια, για την υιοθέτηση απαιτείται προηγουμένως η ανάλογη ενημέρωση, δηλ. η ενσωμάτωση της γενικής γνώσης (οριζόντια, από τα παιδιά μέχρι τους ενήλικες, όπως και στον εργασιακό χώρο) του τι είναι, τι δεν είναι και ποιες οι αποφάσεις που κρύβονται πίσω από την Τεχνητή Νοημοσύνη. Όσο για την εμπιστοσύνη, αυτή κερδίζεται στην πράξη. Τα συστήματα οφείλουν να κάνουν αυτό που ζητάμε, για να μας πείσουν ότι μπορούμε να τα εμπιστευτούμε. Και, για να εστιάσουμε στη χώρα μας, στην Ελλάδα έχουμε τον τρόπο και τα παρελθόν που μας επιτρέπει να κάνουμε πιο δημιουργικές τις διαδικασίες εκδημοκρατισμού της τεχνολογίας και την ανάγκη να το κάνουμε ως κοινωνία τώρα. Σ' αυτό το θέμα, νομίζω ότι η Ελλάδα μπορεί να ηγηθεί.

