

Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση

Τόμ. 1 (2023)

13ο Πανελλήνιο και Διεθνές Συνέδριο «Οι ΤΠΕ στην Εκπαίδευση»



Το ChatGPT περνάει στις Πανελλήνιες. Οι μαθητές;

Μαρία - Πωλίνα Βασιλικού

Βιβλιογραφική αναφορά:

Βασιλικού Μ. -. Π. (2024). Το ChatGPT περνάει στις Πανελλήνιες. Οι μαθητές;. *Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση*, 1, 510–516. ανακτήθηκε από <https://eproceedings.epublishing.ekt.gr/index.php/cetpe/article/view/7330>

Το ChatGPT περνάει στις Πανελλήνιες. Οι μαθητές;

Μαρία - Πωλίνα Βασιλικού

vasilikou@lit.auth.gr

Ποιήτρια, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Θεσσαλονίκη, Ελλάδα

Περίληψη

Η διεθνής βιβλιογραφία δείχνει ήδη ότι σε άλλα επιστημονικά πεδία, το λογισμικό ChatGPT της OpenAI περνάει την πλειοψηφία των εξετάσεων, ενώ μέχρι στιγμής οι επιδόσεις του σε ελληνικά δεδομένα δεν έχουν ερευνηθεί. Η παρούσα έρευνα αξιολογεί τις επιδόσεις του σε απαντήσεις που έδωσε σε ζητούμενα πανελληνίων στο μάθημα της γλωσσικής διδασκαλίας. Βασική υπόθεση ήταν ότι το λογισμικό θα ανταποκριθεί με βάση τα σχολικά κριτήρια. Οι απαντήσεις του λογισμικού ελέγχθηκαν από πέντε εν ενεργεία εκπαιδευτικούς. Από την ανάλυση προέκυψε ότι το λογισμικό μπορεί να περάσει με επιτυχία στην πλειοψηφία των περιπτώσεων, ενώ οι περισσότερες ελλείψεις είναι όμοιες με αυτές που παρουσιάζουν και οι πραγματικοί μαθητές.

Λέξεις κλειδιά: ChatGPT, τεχνητή νοημοσύνη, γλωσσική διδασκαλία, δευτεροβάθμια εκπαίδευση

Εισαγωγή

Τα Ψηφιακά Μέσα είναι πια αναπόσπαστο μέρος της σύγχρονης πραγματικότητας και συνυφασμένα με πολλές ανθρώπινες δραστηριότητες. Η επικοινωνία είναι μία από αυτές που επηρεάστηκαν περισσότερο από την εμφάνιση των Ψηφιακών Μέσων, αφού βασικός της πυλώνας είναι η γραφή και η ανάγνωση (γραμματισμός). Έτσι, αφού επηρεάστηκε το γράψιμο γενικότερα, αναπόφευκτο ήταν να επηρεαστεί με τη σειρά του το Σχολείο, μεγάλο μέρος του οποίου βασίζεται σε αυτό. Έχει σημειωθεί μεγάλη αλλαγή στον τρόπο παραγωγής λόγου, σε σημείο που πλέον το γράψιμο δεν είναι, σε θεωρητικό επίπεδο τουλάχιστον μία άπαξ, αλλά δυναμική διαδικασία. (Κουτσογιάννης, 2017).

Η γραφή και η ανάγνωση, περνούν πια σε μία εποχή Ψηφιακού Γραμματισμού. Με την επεξεργασία να είναι πιο εύκολη από ποτέ, την ανάγνωση να γίνεται (και) επί οθονών, η βάση για τα κείμενα που παράγονται ως απάντηση σε σχολικά ζητούμενα δεν είναι καθόλου σπάνια ακριβώς αυτό: η νέα πραγματικότητα που έχουν σχηματίσει τα Ψηφιακά Μέσα. Μέρος αυτής της νέας πραγματικότητας είναι η ευρεία, πια, διάδοση του περιβάλλοντος του ChatGPT, της OpenAI. Παρότι σχετικά πρόσφατο το περιβάλλον αυτό στη μορφή που το γνωρίζουμε τώρα, έχει ήδη απασχολήσει τη βιβλιογραφία, κυρίως όσον αφορά στην τριτοβάθμια εκπαίδευση και στα ζητήματα παραγωγής λόγου που αυτή απαιτεί. Συγκεκριμένα, ως ένα περιβάλλον που παράγει λόγο απαντώντας σε ζητούμενα (*prompts*), έχει ανησυχήσει την ακαδημαϊκή κοινότητα λόγω της δυνατότητάς του να απαντά σχεδόν στα πάντα- ακόμη και σε σχολικά ζητούμενα.

Η βιβλιογραφία μέχρι στιγμής έχει εξετάσει κυρίως τον χώρο της τριτοβάθμιας εκπαίδευσης, σε εξετάσεις ανοιχτού ή κλειστού τύπου από σχολές διαφόρων επιστημών (βλ. θεωρητικό πλαίσιο), αλλά ελάχιστα έχουν ερευνηθεί εξετάσεις και ζητήματα που έχουν σχέση με τη διδασκαλία της νέας ελληνικής γλώσσας. Αυτό το κενό θα επιχειρήσουμε να καλύψουμε στο παρόν κείμενο, ώστε να έχουμε μια πρώτη τουλάχιστον ενδεικτική εικόνα, της νέας πορείας που φαίνεται να χαράσσεται για το σχολείο με τα νέα δεδομένα.

Θεωρητικό Πλαίσιο

Από τα chatbots μέχρι το ChatGPT, οι προγραμματιστές συνδύασαν γλώσσα, στατιστική ανάλυση και μοντέλα γλωσσικής επεξεργασίας σε ένα αυτοματοποιημένο εργαλείο τεχνητής νοημοσύνης που στην GPT-4 έκδοσή του μπορεί να παράγει λόγο όμοιο με αυτόν του ανθρώπου, για κάθε επιτρεπτή (βάσει κανονισμών) περίπτωση. Όπως λέγεται, είναι έτοιμο να απαντήσει σχεδόν στα πάντα. Η νέα αυτή πραγματικότητα έχει αρχίσει να ελέγχεται ερευνητικά κυρίως στον χώρο της τριτοβάθμιας εκπαίδευσης, αφού έχουν εξεταστεί οι επιδόσεις του σε ερωτήσεις ανάπτυξης αλλά και κλειστού τύπου ερωτήσεις, σε διάφορες γλώσσες.

Ενδεικτικά, δοκιμάστηκε στην Ιατρική για τις εθνικές δοκιμασίες απονομής ιατρικών αδειών στην Ιαπωνία (Kaneda et al., 2023). Ελέγχθηκε στις 400 ερωτήσεις που απάντησαν και οι πραγματικοί υποψήφιοι και κατάφερε να δώσει απάντηση σε 389 από αυτές. Συνολικά, απάντησε σωστά στο 55.0% των περιπτώσεων, ένα ποσοστό όχι αρκετό για να περάσει, που δείχνει όμως ότι αν είχαν δημοσιευθεί οι επίσημες απαντήσεις ή/και αν το μοντέλο είχε εκπαιδευτεί περισσότερο σε τέτοιου είδους κείμενα, θα ήταν περισσότερες οι πιθανότητες να πετύχει μεγαλύτερο ή ακόμη και επαρκές ποσοστό σωστών απαντήσεων.

Επιπλέον, ελέγχθηκε η επίδοσή του σε εξετάσεις Νομικής Σχολής (Choi et al., 2023), όπου οι επιδόσεις του ήταν τουλάχιστον επαρκείς, αφού πέρασε και τα τέσσερα μαθήματα στα οποία εξετάστηκε. Η έρευνα αναφέρει ειδικότερα ότι το λογισμικό διατηρούσε τον τόνο και το ύφος που χρειαζόταν η κάθε απάντηση και απέφυγε την απλή παρουσίαση επιχειρημάτων και ιδεών. Ωστόσο, οι αδυναμίες του όπως περιγράφονται από την έρευνα περιλαμβάνουν συστηματικά λάθη, τα οποία, όπως λένε, *το έκαναν χειρότερο από τον μέσο μαθητή* (Choi et al., 2023). Απαντούσε γενικά, χωρίς λεπτομερή νομική ορολογία ενώ κάποιες φορές, παρέκκλινε από το θέμα που του ζητούνταν να αναπτύξει.

Τέλος, έχει ελεγχθεί και σε μάθημα εισαγωγικό στη Φυσική (Kortemeyer, 2023), με τεστ πολλαπλής επιλογής στο οποίο συγκέντρωσε 18/30, ακριβώς δηλαδή τη βάση για να περάσει κανείς το μάθημα αυτό. Στην ελληνική, η μόνη γνωστή δοκιμή έχει γίνει στο τμήμα Πληροφορικής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, στο μάθημα της Ψηφιακής Επεξεργασίας Εικόνας στο οποίο συγκέντρωσε 8/10 έγκυρες απαντήσεις (Καλτσιδη, 2023).

Η γενικότερη διαπίστωση που προκύπτει από τις πρώτες έρευνες είναι ότι το λογισμικό ανταποκρινόταν αρκετά ικανοποιητικά, με τη σημαντικότερη διαπίστωση να είναι ότι στην πλειοψηφία των περιπτώσεων περνούσε τις εξετάσεις. Το ότι επιτυγχάνει, λοιπόν, διαρκώς, τουλάχιστον οριακή –ή καλύτερη– βαθμολογία, θέτει ένα σημαντικό πρόβλημα: δεν αποκλείεται να μπορεί να ανταποκριθεί στα ελληνικά, σε διάφορα ζητούμενα. Αυτό σημαίνει ότι είναι πιθανό να γίνει υποψήφια πηγή για έτοιμες προς παράδοση απαντήσεις. Την πιθανότητα αυτή θα προσπαθήσουμε να ελέγξουμε στην παρούσα πιλοτική έρευνα, ώστε να αναλύσουμε τις επιδόσεις του αλλά και κατά πόσο μπορεί να γίνει διαθέσιμος πόρος απαντήσεων προς αντιγραφή.

Μεθοδολογία

Στόχος, λοιπόν, της παρούσας εργασίας είναι να εξετάσουμε κατά πόσο θα μπορούσε το ChatGPT να ανταποκριθεί στα σχολικά ή/και ακαδημαϊκά ζητούμενα, ενεργοποιώντας κατά συνέπεια το ενδεχόμενο να επηρεαστεί το σχολείο και η γλωσσική διδασκαλία. Για να επιτευχθεί αυτό, αποφασίστηκε η διεξαγωγή της παρούσας έρευνας με κύριο στόχο τις αξιολογήσεις των απαντήσεων που δίνει το ChatGPT σε ερωτήσεις Πανελληνίων εξετάσεων των τελευταίων ετών. Οι αξιολογήσεις αυτές αξιολογήθηκαν από πραγματικούς διδάσκοντες στη δευτεροβάθμια εκπαίδευση.

Ειδικότερα, για τη διεξαγωγή της παρούσας έρευνας, αξιοποιήθηκαν πέντε εν ενεργεία φιλόλογοι, τρεις που υπηρετούν σε δημόσιο σχολείο και δύο σε ιδιωτική εκπαιδευτική δομή, στην ίδια αστική περιοχή. Όλοι συναίνεσαν στην επεξεργασία των δεδομένων τους σύμφωνα με όσα προβλέπει ο νόμος για την προστασία των προσωπικών τους δεδομένων και με μόνο σκοπό να υποστηρίξουν την παρούσα έρευνα. Για τις ανάγκες της έρευνας σχεδιάστηκε ένα ημιδομημένο ερωτηματολόγιο που αποτέλεσε το εργαλείο πάνω στο οποίο στηρίχθηκαν οι συνεντεύξεις που πραγματοποιήθηκαν. Ειδικότερα, οι πέντε εκπαιδευτικοί κλήθηκαν να αξιολογήσουν τις απαντήσεις που δίνει το συγκεκριμένο λογισμικό σε θέματα Πανελληνίων εξετάσεων του μαθήματος της Έκθεσης από τα έτη 2019, 2020 και 2022.

Η διαδικασία που ακολουθήθηκε ήταν η εξής: αρχικά, εντοπίστηκαν τα θέματα που ήταν κατάλληλα για να αξιοποιηθούν, με δεδομένο ότι το περιβάλλον ακόμη δε μπορεί να δώσει μεγάλης έκτασης απαντήσεις στην ελληνική γλώσσα. Τα θέματα αυτά, ανέρχονται σε τρία συνολικά: τα θέματα του 2022 που είναι τα πιο πρόσφατα και συμφωνούν με το νέο σύστημα των εξετάσεων, τα θέματα του 2020 που τέθηκαν σε μεταβατική περίοδο, από το παλιό προς το νέο σύστημα και τα θέματα του 2019, που συμφωνούν με το παλιό σύστημα. Στο λογισμικό δόθηκαν οι εκφωνήσεις του ζητούμενου παραγωγής γραπτού λόγου, μαζί με τα κείμενα που έπρεπε οι μαθητές να λάβουν υπ' όψη. Τα κείμενα τοποθετήθηκαν σε εισαγωγικά και με τίτλο «Κείμενο 1» και «Κείμενο 2», όπου χρειαζόταν. Με αυτό τον τρόπο, έγινε δυνατό να τα λάβει υπ' όψη του και το λογισμικό, αντίστοιχα. Το πρόγραμμα ανταποκρίθηκε καλά στο σύνολό του, και δεν χρειάστηκε παρά ελάχιστες παρεμβάσεις, οι οποίες και θα συζητηθούν περαιτέρω στους Περιορισμούς.

Οι εκπαιδευτικοί δεν γνώριζαν ότι οι απαντήσεις προέρχονται από περιβάλλον τεχνητής νοημοσύνης και τους ζητήθηκε σε πρώτη φάση να αξιολογήσουν τις απαντήσεις σαν αυτές να είχαν γραφτεί από κάποιον μαθητή τους, να δώσουν μία βαθμολογία με άριστα το 30 ή το 40 κατά περίπτωση και σύντομα ή εκτενέστερα σχόλια, σαν να απευθύνονταν στον μαθητή τους. Τις βαθμολογίες αυτές, λόγω των διαφορών ανάμεσα στο σύστημα εξετάσεων του 2019 αφενός και των 2020 και 2022 αφετέρου, τις μεταφέραμε με την απλή μέθοδο των τριών από τις αρχικές κλίμακες των 30 και 40 μορίων αντίστοιχα στην κλίμακα του 10. Έτσι, ήταν δυνατό να υπολογίσουμε τον μέσο όρο των αξιολογήσεων σε μία, ενιαία κλίμακα.

Αφού ολοκληρώθηκε η διαδικασία της αξιολόγησης, θέσαμε ορισμένες ερωτήσεις στους εκπαιδευτικούς, που εστίαζαν στις αξιολογήσεις που έδωσαν, μέσα από τις οποίες οδηγήθηκαν στο να αναπτύξουν τις σκέψεις τους για τον πιθανό μαθητή πίσω από τα γραπτά και τις ιδεολογίες του για τη μάθηση. Οι ερωτήσεις αυτές αποτέλεσαν τον βασικό σκελετό της συζήτησης, ενώ οι εκπαιδευτικοί ήταν ελεύθεροι να συζητήσουν και για ζητήματα εκτός από όσα θίγουν οι ερωτήσεις αυτές, στην ευρύτερη θεματική της διδασκαλίας γλωσσικών μαθημάτων. Με το πέρας της συζήτησης για τις αξιολογήσεις που έδωσαν, αποκαλύψαμε την πληροφορία που δεν τους είχε δοθεί: την ακριβή προέλευση των κειμένων. Με βάση αυτό, κλήθηκαν να σχολιάσουν εκ νέου όλα τα παραπάνω ερωτήματα, επεκτείνοντας έτσι τη συζήτηση στο θέμα του ρόλου της Τεχνητής Νοημοσύνης στο σχολείο.

Με αυτού του τύπου τις ερωτήσεις και ακολουθώντας αυτά τα βήματα, στόχος είναι να ελεγχθεί αν το πρόγραμμα αυτό μπορεί να περάσει την πλειοψηφία, αν όχι όλες τις δοκιμασίες παραγωγής γραπτού λόγου στη μορφή που υπάρχουν και λειτουργούν σήμερα.

Περιορισμοί

Το πρόγραμμα έδωσε ολοκληρωμένες απαντήσεις, με την έννοια ότι αυτές είχαν μια στοιχειώδη δομή, δεν περιείχαν ελλείψεις προτάσεις, ή αντιγραμματικές δομές. Βασική του αδυναμία ήταν ότι κάποιες φορές διέκοιπε την παραγωγή στο μέσο της απάντησής του για άγνωστο λόγο. Ωστόσο, αυτό έγινε δυνατό να ξεπεραστεί με την απλή εντολή *συνέχισε*,

παρακαλώ ή συνέχισε την απάντησή σου, παρακαλώ, και το πρόγραμμα ολοκλήρωνε έτσι την απάντησή του. Αυτή η διαδικασία χρειάστηκε να γίνει περισσότερες από μία φορές σε κάθε απάντηση.

Επιπλέον, στα θέματα του 2020 χρειάστηκε μία επιπλέον παρέμβαση μόνο σε ελάχιστα σημεία, στα οποία το πρόγραμμα ενέταξε στην απάντησή του την φράση «ως λογισμικό, δεν έχω προσωπική εμπειρία, όμως...», η οποία θα πρόδιδε την εντελώς τεχνητή προέλευσή της. Αφαιρώντας την όμως, οι απαντήσεις επέστρεψαν στο να φαίνονται απόλυτα ανθρώπινες. Σε κάθε περίπτωση, οι μικρο-παρεμβάσεις χρειάστηκαν ελάχιστο χρόνο, λιγότερο του ενός λεπτού κάθε φορά, τον οποίο θεωρούμε ότι ένας μαθητής θα ήταν κατά πάσα πιθανότητα σε θέση να διαθέσει για να επεξεργαστεί το εκάστοτε κείμενο πριν το παρουσιάσει ως δικό του.

Κάτι στο οποίο θα πρέπει να δοθεί έμφαση, τέλος, είναι ο αριθμός συμμετεχόντων. Συνολικά, συμμετείχαν στην έρευνα πέντε φιλόλογοι, όπως προαναφέρθηκε. Ο αριθμός αυτός δεν είναι αντιπροσωπευτικός. Όμως, στόχος της εργασίας αυτής ήταν να διερευνηθεί σε ένα πρώτο, πιλοτικό επίπεδο το αν και πώς ανταποκρίνεται το νέο αυτό λογισμικό στις απαιτήσεις του σχολείου, ζητώντας τη γνώμη μιας μικρής μερίδας εν ενεργεία φιλόλογων.

Αποτελέσματα

Στον Πίνακα 1 παρουσιάζεται συγκεντρωτικά το πώς οι εκπαιδευτικοί αξιολόγησαν τα αποτελέσματα που προέκυψαν από το λογισμικό ChatGPT, ως απάντηση στα θέματα των τριών προαναφερθέντων ετών. Από τον πίνακα αυτό προκύπτει ότι το λογισμικό περνάει σε κάποιες περιπτώσεις με πολύ καλή βαθμολογία, ενώ έλαβε βαθμολογία τουλάχιστον μεγαλύτερη από τη βάση σε όλες, πλην μίας, τις περιπτώσεις. Οι αξιολογήσεις παρατίθενται στην κλίμακα του 10, πια, ενώ κατηγοριοποιούνται ανά έτος και ανά εκπαιδευτικό. Στο τέλος κάθε στήλης βρίσκεται ο μέσος όρος των αξιολογήσεων ανά έτος, ενώ στο τέλος της κάθε σειράς βρίσκεται ο αντίστοιχος μέσος όρος όλων των αξιολογήσεων του κάθε εκπαιδευτικού. Ο γενικός μέσος όρος προκύπτει από τον συνυπολογισμό όλων των μέσων όρων ανά έτος και των αντίστοιχων ανά εκπαιδευτικό.

Πίνακας 1. Αποτελέσματα αξιολόγησης απαντήσεων

| Πληροφορητής | 2019 | 2020 | 2022 | ΜΟ |
|--------------|----------------|----------------|----------------|----------------|
| Π1 | 8/10 | 4,5/10 | 7,6/10 | 6,7/10 |
| Π2 | 6,75/10 | 6,5/10 | 7/10 | 6,8/10 |
| Π3 | 6,75/10 | 6,6/10 | 5,6/10 | 6,3/10 |
| Π4 | 8,75/10 | 8,3/10 | 7,6/10 | 8,2/10 |
| Π5 | 5,5/10 | 5/10 | 6,6/10 | 5,7/10 |
| ΜΟ | 7,15/10 | 6,18/10 | 6,88/10 | 6,87/10 |

Ενδιαφέρον έχουν και τα σχόλια των εκπαιδευτικών. Ο πρώτος πληροφορητής (Π1) σημείωσε στις παρατηρήσεις του ότι διακρίνει *ωραία επιχειρήματα και έκφραση*, πρωτότυπες ιδέες αλλά σοβαρά προβλήματα δομής και στα τρία κείμενα. Ειδικότερα, παρατηρεί την απουσία τίτλου και αφόρμησης, έντονη καταλογογράφηση των επιχειρημάτων στο κυρίως μέρος και, όπως ανέφερε, δύο προλόγους και δύο επιλόγους. Βρίσκει κουραστικές τις κφραστικές επαναλήψεις. Εντοπίζει ασάφειες και απουσία προλόγου και επιλόγου στο

δεύτερο κείμενο και αντιθέτως, ελλιπή ανάπτυξη του κυρίως μέρους αλλά υπερμεγέθη επίλογο για το τρίτο κείμενο. Αναφέρει, ενδεικτικά:

Δίνε έμφαση στην ποιότητα και ανάλυση, όχι μόνο στην ποσότητα των επιχειρημάτων. Δομικά, μοιάζει σαν να υπάρχουν δύο ξεχωριστοί επίλογοι-σαν να είναι δύο εκθέσεις. Δεν υπάρχει σωστή δόμηση ούτε σε πρόλογο, ούτε σε επίλογο. Στον επίλογο πρέπει να κλείνουμε θετικά. Στο κυρίως μέρος δεν αναπτύχθηκαν επαρκώς τα ζητούμενα. Τα στοιχεία της πρώτης και της δεύτερης παραγράφου θα μπορούσαν να αποτελούν ανάπτυξη του Α' ζητούμενου, αντί να αποτελούν τα ίδια τον πρόλογο.

Μετά την αποκάλυψη της εμπλοκής του ChatGPT, σχολίασε:

Θεώρησα ότι έρχονται από έναν μαθητή που επειδή βασίζεται πολύ στο μυαλό του, είναι καλός στο να καταλαβαίνει, βασίζεται και δε λαμβάνει υπ' όψιν του το πώς κάνω πρόλογο, τι μου έχει πει ο ένας ο καθηγητής και τι ο άλλος, είναι λίγο χύμα, γιατί έχει αυτοπεποίθηση στις γνώσεις και στο μυαλό μου που είναι καλό. Δε θέλω μόνο μυαλό, θέλω και μια στοιχειώδη μεθοδολογία και να ακολουθώ κάποια κοινά που μου ζητά το σχολείο.

Το σχόλιο αυτό είναι ενδεικτικό, αφού ο συγκεκριμένος εκπαιδευτικός θεωρεί ότι το κείμενο έρχεται από παιδί με *μυαλό* και *γνώσεις*.

Η δεύτερη πληροφορήτρια (Π2) σημειώνει ορισμένες κοινές ελλείψεις και για τα τρία κείμενα, όπως την απουσία τίτλου και αφορμής και τη μη αξιοποίηση της διαπίστωσης του συγγραφέα. Για το δεύτερο κείμενο, πρώτα-πρώτα αναφέρει την έλλειψη τίτλου και αφορμής, ενώ η δεύτερή της παρατήρηση αφορά στην αδύναμη δομή και συνοχή, αδυναμία που όπως παρατηρεί, επηρεάζει έντονα και την κατά τα άλλα καλή ποιότητα επιχειρημάτων και ιδεών που παρουσιάζονται κατά την ανάπτυξη του κυρίως μέρους. Για το τρίτο κείμενο σχολιάζει την ελλιπή ανάπτυξη του πρώτου ζητούμενου, την απουσία τίτλου και αφορμής, και την αδύναμη έως ανύπαρκτη σύνδεση των νοημάτων και των ζητούμενων μεταξύ τους στο κυρίως μέρος. Επισημαίνει, ενδεικτικά:

Απουσιάζουν αφορμή και τίτλος- αδυναμία στη γλώσσα και στη δομή. Μάλλον "αμήχανο" κείμενο. Καλές ιδέες, που δεν εκφράζονται τόσο σωστά, και δεν έχουν μπει καθόλου σε σειρά. "Αδύναμο" το πρώτο ζητούμενο. Απουσιάζει η αφορμή και ο τίτλος. Αδύναμη και η σύνδεση του πρώτου με το δεύτερο ζητούμενο του θέματος.

Για την προέλευση των κειμένων και για το γιατί ακριβώς αυτά συγκέντρωσαν βαθμολογίες πάνω από τη βάση, σχολίασε ειδικότερα:

...στέκονται ως κείμενα, δεν ανταποκρίνονται καθόλου στα σχολικά κριτήρια και σε αυτά των πανελληνίων, κι όμως περνούν. Ίσως αυτό να είναι και μια προστασία ενάντια στην τεχνητή νοημοσύνη: να θέτεις κριτήρια που δε τα καταλαβαίνεις ούτε εσύ ο ίδιος, ας πούμε.

Η τρίτη πληροφορήτρια (Π3) ανέφερε ότι εντόπισε κοινές ελλείψεις και στα τρία κείμενα, με βασικότερες και επικρατέστερες την απουσία τίτλου και αφορμής, την κάπως ανοργάνωτη δομή και την καταλογογράφηση των επιχειρημάτων στα κύρια μέρη των κειμένων. Ωστόσο, σημειώνει πως η έκφραση τής φάνηκε ικανοποιητική και το περιεχόμενο αξιολογώτατο. Σχολίασε μετέπειτα πως τα λάθη που εντοπίζει στα κείμενα είναι κοινά με αυτά που εντοπίζει στα κείμενα των δικών της μαθητών. Μετά την αποκάλυψη της προέλευσης των κειμένων, τα σχόλια της ήταν τα εξής:

Δεν υποψιάστηκα την τεχνητή νοημοσύνη! Δεν έχω πολλές γνώσεις πάνω στις δυνατότητες που έχει η τεχνητή νοημοσύνη και στην ευρύτητα των χρήσεών της. Δυστυχώς ή ευτυχώς οι απαντήσεις λίγο έμοιαζαν με τις απαντήσεις ενός κανονικού/-ής μαθητή/-τριας που θα είχε

αντιγράφει τα επιχειρήματά του/ της από κάποιο εγχειρίδιο. Ναι, για να είμαι ειλικρινής, υποψιάστηκα «αντιγραφή».

ενώ για το αν και πώς απειλεί ένα τέτοιο λογισμικό την εκπαίδευση σημειώνει:

Τα παιδιά ήδη είναι εξαρτημένα από την τεχνολογία. Μήπως πρέπει να μαθαίνουν πώς να μαθαίνουν; Μήπως πρέπει να ερευνούν; Μήπως καλό είναι να επιλέγουν τι θέλουν να υιοθετήσουν και τί να απορρίψουν; Θέλουμε τέλεια κείμενα με πλούσιο λεξιλόγιο “in vitro” ή προσωπικό ύφος έστω με λάθη “in vivo”;

Ο τέταρτος πληροφορητής (Π4) εντόπισε κι αυτός κοινές ελλείψεις και στα τρία κείμενα, με βασικότερα μειονεκτήματα την απουσία τίτλου και αφόρμησης, και τα προβλήματα δομής, αφού δεν εντοπίζει ξεκάθαρο πρόλογο ή επίλογο στη διάρκεια του κειμένου. Για το περιεχόμενο χρησιμοποιεί τον χαρακτηρισμό *βιωματικό*. Για την προέλευση των κειμένων, σχολίασε τα παρακάτω:

Αν δε καταφέρουμε ως εκπαιδευτικοί να μεταδώσουμε τον σωστό τρόπο αξιοποίησης τέτοιου είδους λογισμικών και εφαρμογών, είναι, ναι, πολύ πιθανό να επηρεάσει αρνητικά τους μαθητές.

Η πέμπτη και τελευταία πληροφορήτρια (Π5), που έδωσε τη συνολικά χαμηλότερη βαθμολογία και για τα τρία κείμενα, δήλωσε πως τα σοβαρότερα προβλήματα των κειμένων ήταν η απουσία τίτλου και αφόρμησης και η παρουσίαση των επιχειρημάτων σε μορφή λίστας. Εντοπίζει και η ίδια έντονες επαναλήψεις λέξεων και φράσεων αλλά και κυκλική επιχειρηματολογία [...η κατανόηση της ιστορίας μας βοηθά να κατανοήσουμε τον κόσμο που ζούμε σήμερα, διότι η σημερινή κατάσταση δε μπορεί να εξηγηθεί χωρίς την κατανόηση του παρελθόντος.] Επιπλέον, στο δεύτερο κείμενο σχολιάζει ως αυθαίρετη την αλλαγή προσώπων μεταξύ πρώτου ενικού και πληθυντικού προσώπου και τρίτου ενικού και πληθυντικού προσώπου. Στο τρίτο κείμενο παρατηρεί μία σχετική πραγματολογική άγνοια του “γράφοντος”, αφού εν ολίγοις δίνει ως απάντηση προτάσεις με περιεχόμενο που ήδη υπάρχουν/ συμβαίνουν. Για την προέλευση των κειμένων, σχολίασε:

Είναι αλήθεια ότι δεν περίμενα πως τα κείμενα αυτά προέρχονται από εφαρμογές τεχνητής νοημοσύνης, καθώς τα λάθη και οι συγγραφικές δυσκολίες που εντόπισα έμοιαζαν με τις αντίστοιχες συνήθειες δυσκολίες ενός μαθητή.

Συνολικά, τα κείμενα που δόθηκαν ως απαντήσεις στα ζητούμενα των Πανελληνίων εξετάσεων των τριών ετών 2019, 2020 και 2022 έλαβαν στις 14 από τις 15 περιπτώσεις βαθμολογία πάνω από τη βάση, “έγραψαν” δηλαδή τουλάχιστον επαρκώς στο 93,3% των περιπτώσεων. Τρεις από τους πέντε πληροφορητές δήλωσαν ότι δεν αντιλήφθηκαν καθόλου την εμπλοκή της τεχνητής νοημοσύνης ενώ οι υπόλοιποι δύο το υποψιάστηκαν διότι είχαν ήδη πληροφορηθεί για την ύπαρξη και μερικές από τις λειτουργίες του λογισμικού, όμως όπως δήλωσαν, βαθμολόγησαν σαν να είχαν πραγματικά κείμενα μπροστά τους.

Συμπεράσματα- Συζήτηση

Η έρευνα αυτή είχε ως βασικό της σκοπό να ελέγξει τις επιδόσεις του λογισμικού ChatGPT σε ζητούμενα από τις Πανελλήνιες εξετάσεις, στο μάθημα της Νέας Ελληνικής Γλώσσας. Στόχος της ήταν να ελέγξει κατά πόσο οι απαντήσεις του ChatGPT μπορούν να σταθούν ως πιθανές απαντήσεις σε θέματα πανελληνίων εξετάσεων.

Από την ανάλυση προκύπτει ότι ο λόγος του ChatGPT στα ελληνικά, αν και το ίδιο δεν έχει εκπαιδευτεί σε σχολικά κριτήρια (κάτι που αποτυπώνεται σε “λάθη” που εμφανίζει

επανελημμένα, όπως η απουσία τίτλου και αφόρμησης) και παρ' ότι ο λόγος του στα ελληνικά δεν είναι ακόμη τελειοποιημένος (καταλογογράφηση επιχειρημάτων, κυκλικά επιχειρήματα, επαναλήψεις), είναι και πάλι άξιο να λάβει βαθμολογία τουλάχιστον πάνω από τη βάση στο μάθημα της Νέας Ελληνικής Γλώσσας, και ενδεχομένως να επιτύχει και έναν βαθμό αρκετά κοντά σε αυτόν της πλειοψηφίας. Η διαπίστωση αυτή συμφωνεί, εν πολλοίς με τα αποτελέσματα της διεθνούς έρευνας η οποία δείχνει ότι το ChatGPT ανταποκρίνεται τουλάχιστον επαρκώς σε πολλούς επιστημονικούς τομείς, συμπεριλαμβανομένων της Φυσικής και της Νομικής.

Οι απόψεις που εκφράζουν οι εκπαιδευτικοί είχαν επίσης ιδιαίτερο ενδιαφέρον. Όλοι θεώρησαν ότι τα κείμενα είναι πολύ καλά ως προς το περιεχόμενο, αλλά έχουν ζητήματα δομής. Αυτό είναι προφανές από τη στιγμή που το περιβάλλον δεν είναι εκπαιδευμένο για σχολικού τύπου ερωτήσεις. Επιπλέον, σχετικά με τις ελλείψεις που παρατηρούν -σύμφωνα με τα σχολικά κριτήρια πάντα- οι ίδιοι σημειώνουν ότι είναι αρκετά όμοια με αυτά που παρουσιάζει το σύνολο των πραγματικών μαθητών.

Η εντυπωσιακή ομοιότητα με πραγματικά γραπτά μαθητών δεν μπορεί παρά να απασχολήσει την ακαδημαϊκή αλλά και τη διδακτική κοινότητα. Ο κίνδυνος για παρουσίαση εργασιών απευθείας από το ChatGPT στη σχολική αίθουσα είναι μεγάλος, ενώ τέτοιες τακτικές, αν δε γίνουν έγκαιρα αντιληπτές, δεν είναι απίθανο να οδηγήσουν σε μαζική, πρωτοφανή αποτυχία των μαθητών σε διαγωνίσματα και τελικές εξετάσεις, όπου το λογισμικό αυτό δε θα χρησιμοποιείται.

Επιπλέον, τίθενται σημαντικά ερωτήματα σχετικά με τον πορήνα των γλωσσικών αλλά και των υπόλοιπων μαθημάτων. Τα σχολικά ζητούμενα είναι σημαντικό να αλλάξουν μορφή και κατεύθυνση, ώστε να απαιτούν μία δυναμική σχέση μεταξύ των μαθητών, των γραπτών τους και των εκπαιδευτικών. Κύριο μέλημα θα πρέπει να είναι η παρακολούθηση της πραγματικής προόδου των μαθητών και ο έγκαιρος εντοπισμός δυσκολιών, με μοναδικό σκοπό την ενδυνάμωση και την ενίσχυση των δυνατοτήτων τους.

Η έρευνα αυτή, αν και πιλοτική, είναι ένα πρώτο δείγμα του πόσο καλά μπορεί να ανταποκριθεί το ChatGPT και στα ελληνικά ζητούμενα. Είναι προφανές ότι χρειάζεται περαιτέρω έρευνα, όχι μόνο στην αξιολόγηση των απαντήσεων αλλά και στο πώς πιθανώς χρησιμοποιείται από τα παιδιά για σχολικού τύπου εργασίες.

Αναφορές

- Choi, J. H., Hickman, K. E., Monahan, A., & Schwarcz, D. (2023). Chatgpt goes to law school. Available at SSRN. <https://ssrn.com/abstract=4335905>
- Geng, C., Yihan, Z., Pientka, B., & Si, X. (2023). Can ChatGPT Pass An Introductory Level Functional Language Programming Course?. arXiv preprint arXiv:2305.02230.
- Kaneda, Y., Tanimoto, T., Ozaki, A., Sato, T., & Takahashi, K. (2023). Can ChatGPT Pass the 2023 Japanese National Medical Licensing Examination?. <https://doi.org/10.20944/preprints202303.0191.v1>
- Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course?. *Physical Review Physics Education Research*, 19(1), 010132.
- Καλτσιδη, Μ. (2023, 07 Φεβρουαρίου). Ο αλγόριθμος ChatGPT έδωσε μάθημα εξετάσεις σε μάθημα του ΑΠΘ και πέρασε με βαθμολογία 8 στα 10. ERT News. <https://www.ertnews.gr/ert3/o-algorithmos-chatgpt-edose-eksetaseis-se-mathima-tou-apth-kai-perase-me-vathmologia-8-sta-10/>
- Κουτσογιάννης, Δ. (2017). *Γλωσσική διδασκαλία χθες, σήμερα, αύριο: μια πολιτική προσέγγιση* (No. IKEEBOOK-2020-035). Ινστιτούτο Νεοελληνικών Σπουδών, Ίδρυμα Μανόλη Τρανταφyllίδη.