



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τηλεπικοινωνιών

Τίτλος διπλωματικής

Διπλωματική Εργασία
του
Θεοφάνη Θαρρόπουλου

Επιβλέπων: Ανδρέας Συμεωνίδης
Καθηγητής Α.Π.Θ.

30 Μαΐου 2024

Περίληψη

Αντικείμενο της παρούσας διπλωματικής εργασίας αποτελεί η έρευνα για την αξιολόγηση της ποιότητας του κώδικα που παράγεται από Μεγάλα Γλωσσικά Μοντέλα (LLMs), και πιο συγκεκριμένα από το GitHub Copilot[1]. Η μελέτη εστιάζει στην αξιολόγηση της ποιότητας του κώδικα που παράγεται από το όπilot και στην βελτιστοποίηση των προτροπών (prompts) για την επίτευξη των επιθυμητών αποτελεσμάτων μέσω τεχνικών μηχανικής προτροπής (prompt engineering) και της μηχανικής μάθησης. Τα αποτελέσματα αποδεικνύουν τις δυνατότητες και τους περιορισμούς του όπilot στην παραγωγή ποιοτικού κώδικα και προσφέρουν νέες προσεγγίσεις για την βελτίωση της αλληλεπίδρασης μεταξύ του χρήστη και του εργαλείου μέσω στοχευμένων τεχνικών προτροπής.

Abstract

Empty

Ευχαριστίες

Άδειο

Τίτλος διπλωματικής

Θεοφάνης Θαρρόπουλος
theofact@ece.auth.gr

30 Μαΐου 2024

Περιεχόμενα

1	Εισαγωγή	2
1.1	Δημιουργία Κώδικα (Code Generation) - Ολοκλήρωση Κώδικα (Code Completion)	3
A'	Ακρωνύμια και συντομογραφίες	4

Κεφάλαιο 1

Εισαγωγή

Η ανάπτυξη των Μεγάλων Γλωσσικών Μοντέλων (LLM) έχει επιφέρει ριζικές αλλαγές στον τομέα της Τεχνητής Νοημοσύνης και της Επεξεργασίας Φυσικής Γλώσσας (NLP) [7, 32, 33, 14]. Η εισαγωγή αυτών των μοντέλων στην καθημερινή ζωή μέσω του Chat-GPT το 2022 [21] έχει πυροδοτήσει μια επανάσταση στην τεχνολογική αγορά, σηματοδοτώντας την απαρχή του αγώνα για την κυριαρχία στην αγορά της Τεχνητής Νοημοσύνης [26, 19, 24, 17, 3]. Τα μοντέλα αυτά έχουν αποδειχθεί εξαιρετικά αποτελεσματικά στην αντιμετώπιση προβλημάτων επεξεργασίας φυσικής γλώσσας, όπως η αναγνώριση φυσικής γλώσσας, προάγοντας την ανάπτυξη της Γενικής Τεχνητής Νοημοσύνης (AGI) [4, 13].

Μέσα σε αυτή την επανάσταση, η χρήση μοντέλων επεξεργασίας φυσικής γλώσσας στον τομέα του προγραμματισμού και της ανάπτυξης λογισμικού, ονόματι βοηθοί κώδικα (Code Assistants), έχει αναδειχθεί ως ένας από τους πιο υποσχόμενους τομείς της τεχνολογίας. Ένα από τα πιο διαδεδομένα εργαλεία με αυτόν το σκοπό είναι το GitHub Copilot [12]. Αναπτυγμένο σε συνεργασία με την OpenAI, το GitHub Copilot ξεκίνησε χρησιμοποιώντας το μοντέλο ονόματι Codex της OpenAI [10], σχεδιασμένο εξ αρχής αποκλειστικά για τη παραγωγή κώδικα, για να προτείνει κώδικα στον προγραμματιστή κατά την γραφή κώδικα.

Η παροδική απόσυρση του μοντέλου Codex τον Μάρτιο του 2023 και η οριστική του απόσυρση το 2023 [15], οδήγησε στην ανάπτυξη ενός νέου μοντέλου, σε συνεργασία μεταξύ της OpenAI, της Microsoft Azure AI, και της GitHub AI. Το νέο μοντέλο αρχικά βασίστηκε στο GPT-3.5 Turbo [31], με την επόμενη του έκδοση να βασίζεται στο GPT-4 [31], με το κωδικό όνομα GitHub Copilot X [11], δίνοντας την δυνατότητα για μια νέα λειτουργία, του GitHub Copilot Chat, ενός chatbot μοντέλου, παρόμοιο με αυτό του Chat-GPT. Μέσω αυτής της λειτουργίας, ο προγραμματιστής μπορεί μέσα από το περιβάλλον ανάπτυξής του (IDE), να κάνει ερωτήσεις στο μοντέλο, χρησιμοποιώντας φυσική γλώσσα, ενώ το μοντέλο μπορεί να χρησιμοποιήσει τον ήδη υπάρχοντα κώδικα, την τεκμηρίωση, και τις οδηγίες του προγραμματιστή για να απαντήσει στην ερώτηση του προγραμματιστή.

Το GitHub Copilot Chat αρχικά ήταν διαθέσιμο μόνο μέσω λίστας αναμονής, με την πρόσβαση να δίνεται σε περιορισμένο αριθμό προγραμματιστών, κατόπιν αίτησης, με την ενσωμάτωση του GPT-4 να γίνεται επίσημα τον Νοέμβριο του 2023 [25]. Η δημόσια κυκλοφορία του GitHub Copilot Chat έγινε τον Δεκέμβριο του 2023 [30].

Πέρα από το GitHub Copilot, υπάρχουν πολλοί άλλοι code assistants, με αυτούς που λήφθηκαν υπ' όψιν να είναι οι:

- Tabnine [28, 29]
- Codeium [9]
- Amazon Codewhisperer [6]

Η απόφαση για την χρήση του GitHub Copilot ως το εργαλείο, του οποίου η λειτουργία θα εξεταστεί στην παρούσα διπλωματική, έγινε με βάση την ευρεία χρήση του, την ενσωμάτωση του μοντέλου GPT-4, και την δυνατότητα χρήσης του GitHub Copilot Chat, καθώς και την δωρεάν παροχή του σε ενεργούς φοιτητές του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης μέσω του προγράμματος GitHub Student Developer Pack [2].

1.1 Δημιουργία Κώδικα (Code Generation) - Ολοκλήρωση Κώδικα (Code Completion)

Δύο όροι που συχνά συναντώνται στην βιβλιογραφία είναι η **Δημιουργία Κώδικα** (Code Generation) και η **Ολοκλήρωση Κώδικα** (Code Completion). Και οι δύο όροι έχουν άμεση σχέση με την παραγωγικότητα του προγραμματιστή και αποτελούν σημαντικά εργαλεία για την ανάπτυξη λογισμικού [18, 16, 5], και ενώ και τα δύο μπορούν να χρησιμοποιήσουν μοντέλα Τεχνητής Νοημοσύνης [27, 23] με την διαφορά να βρίσκεται στον τρόπο λειτουργίας τους.

Σύμφωνα με Omar et al. [20], η Ολοκλήρωση Κώδικα αφορά εργαλεία που τα περισσότερα περιβάλλοντα ανάπτυξης παρέχουν μέσω της μορφής πλωτού μενού που περιέχει συμφραζόμενες-σχετικές μεταβλητές, πεδία, μεθόδους, τύπους και άλλα αποσπάσματα κώδικα. Επιλέγοντας από αυτό το μενού, οι προγραμματιστές μπορούν να αποφύγουν πολλά συνηθισμένα ορθογραφικά και λογικά λάθη και να εξαλείψουν τις περιττές πληκτρολογήσεις. Το βασικό εργαλείο που χρησιμοποιείται για την επίτευξη του σκοπού αυτού είναι το Language Server Protocol (LSP) της Microsoft, δίνοντας την δυνατότητα σε κάθε περιβάλλον ανάπτυξης να επικοινωνήσει με την εξωτερική διεργασία του Language Server και να λάβει πληροφορίες για τον κώδικα που γράφεται, όπως τις προτάσεις ολοκλήρωσης κώδικα, τα λάθη, και τις προτάσεις διόρθωσης. [22, 8]

Παράρτημα Α΄

Ακρωνύμια και συντομογραφίες

LLM Large Language Model

NLP Natural Language Processing

IDE Integrated Development Environment

LSP Language Server Protocol

Bibliography

- [1] “Github copilot.” [Online]. Available: <https://github.com/features/copilot>
- [2] “Github student developer pack.” [Online]. Available: <https://education.github.com/pack>
- [3] “Trendforce says with cloud companies initiating ai arms race, gpu demand from chatgpt could reach 30,000 chips as it readies for commercialization,” *TrendForce*, Nov 2023, archived from the original on May 30, 2024. Retrieved November 2, 2023.
- [4] S. Adams, I. Arel, J. Bach, R. Coop, R. Furlan, B. Goertzel, J. S. Hall, A. Samsonovich, M. Scheutz, M. Schlesinger *et al.*, “Mapping the landscape of human-level artificial general intelligence,” *AI Magazine*, vol. 33, no. 1, pp. 25–42, 2012.
- [5] M. Asaduzzaman, C. K. Roy, K. A. Schneider, and D. Hou, “Csc: Simple, efficient, context sensitive code completion,” in *2014 IEEE International Conference on Software Maintenance and Evolution*. IEEE, 2014, pp. 71–80.
- [6] J. Bays, “Aws announces amazon codewhisperer (preview),” Jun 2022. [Online]. Available: <https://aws.amazon.com/about-aws/whats-new/2022/06/aws-announces-amazon-codewhisperer-preview/>
- [7] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill *et al.*, “On the opportunities and risks of foundation models,” 2021.
- [8] H. Bunder, “Decoupling language and editor - the impact of the language server protocol on textual domain-specific languages,” in *Proceedings of the 7th International Conference on Model-Driven Engineering and Software Development - MODELSWARD, INSTICC*. SciTePress, 2019, pp. 131–142.
- [9] S. Carter, “Exponential baby! navigating the ai convergence of tech with nvidia,” *Forbes Digital Assets*, May 2024, retrieved May 30, 2024. [Online]. Available: <https://www.forbes.com/sites/digital-assets/2024/05/11/exponential-baby-navigating-the-ai-convergence-of-tech-with-nvidia/>
- [10] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. P. de Oliveira Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, A. Ray, R. Puri, G. Krueger, M. Petrov, H. Khlaaf, G. Sastry, P. Mishkin, B. Chan, S. Gray, N. Ryder, M. Pavlov, A. Power, L. Kaiser, M. Bavarian, C. Winter, P. Tillet, F. P. Such, D. Cummings, M. Plappert, F. Chantzis, E. Barnes, A. Herbert-Voss, W. H. Guss, A. Nichol, A. Paino, N. Tezak, J. Tang, I. Babuschkin, S. Balaji, S. Jain, W. Saunders, C. Hesse, A. N. Carr, J. Leike, J. Achiam, V. Misra, E. Morikawa,

- A. Radford, M. Knight, M. Brundage, M. Murati, K. Mayer, P. Welinder, B. McGrew, D. Amodei, S. McCandlish, I. Sutskever, and W. Zaremba, "Evaluating large language models trained on code," 2021.
- [11] T. Dohmke, "Github copilot x: The ai-powered developer experience," *GitHub Blog*, March 2023, retrieved May 30, 2024. [Online]. Available: <https://github.blog/2023-03-22-github-copilot-x-the-ai-powered-developer-experience/>
 - [12] GitHub, "Introducing github copilot: your ai pair programmer," *GitHub Blog*, June 2021, retrieved May 30, 2024. [Online]. Available: <https://github.blog/2021-06-29-introducing-github-copilot-ai-get-code-done-faster/>
 - [13] B. Goertzel, "Artificial general intelligence: Concept, state of the art, and future prospects," *Journal of Artificial General Intelligence*, vol. 5, no. 1, pp. 1–46, 2014, submitted 2013-2-12, Accepted 2014-3-15.
 - [14] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson, 2009.
 - [15] J. Kemper. (2023, March) Openai kills its codex code model, recommends gpt3.5 instead. Retrieved May 27, 2024. [Online]. Available: <https://thdecoder.com/openai-kills-its-codex-code-model-recommends-gpt3-5-instead/>
 - [16] H. H. Koester and S. Levine, "Effect of a word prediction feature on user performance," *Augmentative and alternative communication*, vol. 12, no. 3, pp. 155–168, 1996.
 - [17] Z. Liu, "Chatgpt will command more than 30,000 nvidia gpus: Report," *Tom's Hardware*, Mar 2023, archived from the original on May 30, 2024. Retrieved November 2, 2023.
 - [18] C. Luo, "A report on automatic code completion," 03 2017.
 - [19] C. Metz and T. Mickle, "Openai completes deal that values the company at \$80 billion," *The New York Times*, Feb 2024, retrieved May 30, 2024.
 - [20] C. Omar, Y. Yoon, T. LaToza, and B. Myers, "Active code completion," in *2012 34th International Conference on Software Engineering (ICSE)*. IEEE, 2012, pp. 859–869.
 - [21] OpenAI, "Chatgpt: Optimizing language models for dialogue," 2022. [Online]. Available: <https://web.archive.org/web/20221130180912/https://openai.com/blog/chatgpt/>
 - [22] J. K. Rask, F. P. Madsen, N. Battle, H. D. Macedo, and P. G. Larsen, "The specification language server protocol: A proposal for standardised lsp extensions," 2022.
 - [23] V. Raychev, M. Vechev, and E. Yahav, "Code completion with statistical language models," in *Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation*. ACM, 2014, pp. 419–428.
 - [24] E. Roth, "Microsoft spent hundreds of millions of dollars on a chatgpt super-

computer,” *The Verge*, Mar 2023, archived from the original on May 30, 2023. Retrieved March 30, 2023.

- [25] Staff, “Github copilot – november 30th update,” *GitHub Blog*, November 2023, retrieved May 30, 2024. [Online]. Available: <https://github.blog/changelog/2023-11-30-github-copilot-november-30th-update/>
- [26] K. Staff, “Microsoft-backed openai valued at \$80bn after company completes deal,” *The Guardian*, Feb 2024, retrieved May 30, 2024.
- [27] A. Svyatkovskoy, S. Lee, A. Hadjitofi, M. Riechert, J. Franco, and M. Al-lamanis, “Fast and memory-efficient neural code completion,” *arXiv preprint arXiv:2004.13651*, 2020.
- [28] Unknown, “Tabnine: Coding in vs code with the help of an ai assistant,” *learn.microsoft.com*, March 2021, retrieved April 19, 2024.
- [29] J. Vincent, “This ai-powered autocompletion software is gmail’s smart compose for coders,” *The Verge*, July 2019, retrieved May 19, 2024.
- [30] S. Zhao, “Github copilot chat now generally available for organizations and individuals,” *GitHub Blog*, December 2023, retrieved May 30, 2024. [Online]. Available: <https://github.blog/2023-12-29-github-copilot-chat-now-generally-available-for-organizations-and-individuals/>
- [31] S. Z. Zhao, “Smarter, more efficient coding: Github copilot goes beyond codex with improved ai model,” *GitHub Blog*, July 2023, retrieved May 27, 2024. [Online]. Available: <https://github.blog/2023-07-28-smarter-more-efficient-coding-github-copilot-goes-beyond-codex-with-improved-ai-model/>
- [32] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong *et al.*, “A survey of large language models,” 2023.
- [33] Q. Zhou, C. Li, J. Yu, Y. Liu, G. Wang, K. Zhang, C. Ji, Q. Yan, L. He *et al.*, “A comprehensive survey on pretrained foundation models: A history from bert to chatgpt,” *arXiv preprint arXiv:2302.09419*, 2023.