UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Suuret kielimallit oppimisen tukena: opiskelijoiden kokemuksia ja näkemyksiä

Korpimies, Kai

2024

http://hdl.handle.net/10138/588955

Korpimies, K 2024, Suuret kielimallit oppimisen tukena: opiskelijoiden kokemuksia ja näkemyksiä. julkaisussa ITK – Interaktiivinen Tekniikka Koulutuksessa 2024. Suomen eOppimiskeskus ry:n julkaisusarja, Suomen eOppimiskeskus, Interaktiivinen Tekniikka Koulutuksessa, Hämeenlinna, Suomi, 17/04/2024.

Downloaded from Helda, University of Helsinki institutional repository. https://helda.helsinki.fi
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

Suuret kielimallit oppimisen tukena: opiskelijoiden kokemuksia ja näkemyksiä

Korpimies Kai, tietojenkäsittelytieteen lehtori, Helsingin yliopisto

1 Johdanto

Uudet suuriin kielimalleihin perustuvat välineet, kuten ChatGPT, Google Bard ja Bing Chat, ovat herättäneet paljon keskustelua. Vaikka ainakin ChatGPT:stä on ehditty lyhyessä ajassa jo kirjoittaa paljon, niin julkaisuja opiskelijoiden kokemuksista ja näkemyksistä on vielä niukasti saatavilla. Esitämme tässä artikkelissa tuloksia kyselyistä, joilla selvitettiin kielimallien käyttöä kahdella tietojenkäsittelytieteen kurssilla Helsingin yliopistossa. Suurin osa opiskelijoista oli käyttänyt tai ainakin kokeillut näitä välineitä. Heidän kokemuksensa vaihtelivat käyttötavoittain sen suhteen, miten paljon kielimallit auttoivat opiskelijoita, ja koettiinko niiden edistävän vai haittaavan oppimista. Toisen kurssin osalta havaittiin negatiivinen korrelaatio kielimallien käytön ja kurssimenestyksen välillä.

2 Kielimallit tietojenkäsittelytieteen opetuksessa

Suuriin kielimalleihin perustuvat keskusteluosovellukset, kuten ChatGPT, Google Bard ja Bing Chat, käyttävät suurilla aineistoilla opetettuja neuroverkkoja. Niille on löytynyt monenlaista käyttöä opetuksessa: opiskelijat voivat esimerkiksi hyödyntää niitä tiedonhaussa sekä pyytää niiltä apua harjoitustehtävien ratkaisemiseen ja tentttikysymyksiin vastaamiseen. Useat tutkimukset ovat jo osoittaneet kielimallien selviävän vähintäänkin kohtuullisen hyvin yksinkertaisista ohjelmointitehtävistä (esim. Kiesler & Schiffner, 2023).

Ohjelmointiavun pyytäminen verkossa ei sinänsä ole uusi asia. Opiskelijat ovat jo pitkään voineet saada vastauksia kysymyksiinsä toisilta käyttäjiltä Stack Overflowin kaltaisissa palveluissa. Liu et al. (2023) vertailivat ChatGPT:n ja Stack Overflowin tarjoamaa ohjelmointitukea ja totesivat kummallakin olevan omat vahvuutensa. Kuitenkin mahdollisuus käyttää kielimalleja tekstin ja ohjelmakoodin generointiin harjoituksissa ja tenteissä on herättänyt kysymyksiä siitä, miten se vaikuttaa oppimiseen ja sen arviointiin.

Toinen merkittävä ongelma liittyy välineiden antaman avun ja generoidun materiaalin laatuun: kielimallit tekevät välillä myös virheitä (eli ”hallusinoivat”) tai antavat muuten huonosti sopivia vastauksia. Kabir et al. (2023) totesivat, että 52 % ChatGPT:n antamasta ohjelmointiavusta oli jollain tapaa virheellistä.

Ongelmista huolimatta on esitetty, että koodin generointi voitaisiin sallia tietojenkäsittelytieteen opiskelussa. Esimerkiksi Daun & Brings (2023) näkevät, että sikäli kun manuaalisen koodauksen merkitys jatkossa vähenee, opetuksessa tulisi entistä enemmän painottaa muita taitoja, kuten sellaisia, jotka liittyvät ohjelmistojen suunnitteluun, validointiin ja testaukseen.

3 Opiskelijoiden kokemukset ja näkemykset

Singhin ja kumppaneiden (2023) kyselytutkimuksessa tietojenkäsittelytieteen kansainväliset opiskelijat näkivät ChatGPT:n merkittävimmäksi eduksi sen kyvyn selittää asioita. Suurimmiksi riskeiksi he näkivät uhat kriittiselle ajattelulle ja omaperäisyydelle (originality) sekä riippuvaisuuden välineestä.

Yilmazin ja Yilmazin (2023) kyselytutkimuksessa olio-ohjelmointia opiskelevat tietojenkäsittelytieteen opiskelijat näkivät ChatGPT:n käytön tärkeimpien etujen joukossa sen nopean ja tehokkaan vastaamisen ja sen mukanaan tuoman ajansäästön sekä avun virheenjäljityksessä ja mutkikkaissa tehtävissä. Toisaalta ChatGPT:n katsottiin mahdollistavan liian laiskan asenteen opiskeluun, sen käyttö koettiin ahdistavaksi (occupational anxiety) ja sen ymmärrettiin antavan myös virheellisiä vastauksia. Edut kuiten-

kin painottuivat haittoja enemmän. Kielimallien käytön vaikutus ajattelutaitoihin (thinking skills) jakoi opiskelijoiden mielipiteitä, joskin suurempi osa vastaajista näki tämän vaikutuksen myönteiseksi. Kirjoittajien mukaan kielteistä vaikutusta voi perustella silloin, kun tehtävään on mahdollista vastata lyhyesti ja yksinkertaisesti. Shoufanin (2023) kyselytutkimuksessa sulautettujen järjestelmien kurssille osallistuneet opiskelijat näkivät välineiden sekä auttavan oppimisessa että haittaavan opiskelua. Tässäkin useammat opiskelijat näkivät silti vaikutuksen myönteiseksi.

4 Kyselyt

Kysyimme opiskelijoiden kielimalleja koskevista kokemuksista ja näkemyksistä kahdella erilaisella tietojenkäsittelytieteen kurssilla. Ensimmäinen kurssi on Ohjelmistotekniikan harjoitustyö, jossa opiskelijat määrittelevät, suunnittelevat, toteuttavat, testaavat ja dokumentoivat pienehkön harjoitustyön itse valitsemastaan aiheesta. Harjoitustyö toteutetaan pienissä paloissa ennalta määritetyn viikkoaikataulun mukaisesti, ja lopputulokselle on asetettu tiettyjä teknisiä vaatimuksia. Ohjelmointikielenä on Python, jota osallistujien odotetaan osaavan jo kohtuullisen hyvin. Kurssi kestää noin seitsemän viikkoa, ja se voidaan suorittaa verkon välityksellä.

Harjoitustyössä opiskelijoilla oli lupa käyttää kielimalleja ohjelmakoodin generointiin ja muutenkin suhteellisen vapaasti. Käyttö yksikkötestien generointiin oli kuitenkin kielletty. Koodin generointi voitiin sallia melko ongelmattomasti, koska kyseessä ei ole varsinainen ohjelmointikurssi. Lisäksi opiskelijoiden tulee testata harjoitustyönsä perusteellisesti, mikä edellyttää generoidunkin koodin suhteellisen syvällistä ymmärtämistä.

Toinen kurssi on nimeltään Tietojenkäsittely ja yhteiskunta. Se suoritetaan enimmäkseen ryhmissä suoritettavilla kirjoitustehtävillä ja yksin suoritettavilla monivalintatehtäviä ja vastaavia sisältävillä harjoituksilla ja tentillä. Esseisiin liittyy myös alustuksia, opponointia ja vertaisarviointia. Vaikka opiskelijat eivät saa generoida esseitään kielimalleilla, he voivat käyttää kielimalleja muilla tavoilla. Monivalintatehtävissä kielimallien käyttöä ei ole kielletty, mutta tentissä niiden käyttö olisi lyhyen vastausajan vuoksi käytännössä suhteellisen haastavaa.

Molempien kurssien alussa kielimallit mainittiin, mutta niiden käyttöä ei opetettu. Opiskelijoita varoitettiin niiden antamien vastausten epäluotettavuudesta.

Ohjelmistotekniikan harjoitustyössä opiskelijoita pyydettiin täyttämään lomake, jolla kysyttiin ChatGPT:n, Google Bardin ja Bing Chatin käytöstä ohjelmakoodin parantelussa (virheiden jäljitys ja korjaus, refaktorointi, kommentointi ym.), koodin generoinnissa, dokumentaation asiasisällön valmistelussa tai luonnostelussa, dokumentaation viimeistelyssä (uudelleenmuotoilu, kieli- ja/tai ulkoasun parantelu) sekä mahdollisista muista käyttötavoista. Opiskelijat arvioivat karkeasti kuinka suuressa osassa koodia tai dokumentaatiota kielimalleja oli näihin tarkoituksiin käytetty sekä kuinka hyödylliseksi tai haitalliseksi he tämän käytön näkivät oppimisen kannalta kurssin oppimistavoitteet huomioiden. Lisäksi kysyttiin koodin generointiin mahdollisesti liittyneistä ongelmista, mahdollisesta muiden kielimallien käytöstä ja näkemyksiä siitä, että tulisiko kielimallien käyttöä rajoittaa nykyistä enemmän tai sallia nykyistä laajemmin. Tutkimusluvan sisältäneet 17 vastausta kattoivat noin 61 % kurssilla loppuun asti jatkaneista osallistujista. Koska opiskelijat vastasivat omilla nimillään, vastauksia voitiin verrata kurssimenestykseen.

Kurssilla Tietojenkäsittely ja yhteiskunta vastaukset annettiin nimettöminä. Lomakkeella kysyttiin vastaavasti kuin Ohjelmistotekniikan harjoitustyössäkin kielimallien käytöstä, mutta erikseen viitatut käyttötavat olivat tällä kurssilla monivalintatehtäviin vastaaminen, kirjoitustehtävien asiasisällön valmistelu tai luonnostelu ja niiden viimeistely (uudelleenmuotoilu, kieli- ja/tai ulkoasun parantelu). Tutkimusluvan sisältäneitä vastauksia annettiin 12 kpl ja ne kattoivat noin 32 % aktiivisista osallistujista.

5 Tulokset

Ohjelmistotekniikan harjoitustyössä 59 % ja kurssilla Tietojenkäsittely ja yhteiskunta 67 % vastaajista olivat käyttäneet kielimalleja. Kolmen kyselyissä erikseen mainitun välineen lisäksi vain yksi vastaaja mainitsi käyttäneensä

jotain muuta kielimalleihin perustuvaa välinettä. Taulukossa 5.1 näkyy erilaisten käyttötapojen yleisyys ja kuinka monta vastaajaa näki kunkin niistä edistävän tai haittaavan oppimista. Kun samassa vastauksessa eriteltiin sekä oppimista edistäviä että haittaavia tekijöitä, sitä ei laskettu tässä suhteessa mukaan. Sarakkeessa ”Käyttö yli 25 %” on lueteltu lukumäärät niistä, jotka ilmoittivat käyttäneensä kielimalleja yli 25 % ohjelmakoodista, dokumentaatiosta, monivalinnoista tai kirjoitustehtävistä.

Taulukko 5.1: Kielimallien käyttötapojen yleisyys ja opiskelijoiden näkemykset niiden vaikutuksesta oppimiseen

	Käyttö yli 0%	Osuus vastaajista	Käyttö yli 25%	Osuus vastaajista	Edisti oppimista	Haittasi oppimista
Ohjelmistotekniikan harjoitustyö						
Koodin parantelu	8	47 %	3	18 %	6	0
Koodin generointi	7	41 %	1	6 %	4	3
Dokumentaation valmistelu	2	12 %	1	6 %	1	0
Dokumentaation viimeistely	2	12 %	1	6 %	0	0
Yleinen tiedonhaku	7	41 %	-	-	7	0
Virheilmoitusten tulkinta	4	24 %	-	-	3	0
Tietojenkäsittely ja yhteiskunta						
Monivalinnat	4	33 %	2	17 %	1	2
Kirjoitustehtävien valmistelu	6	50 %	2	17 %	1	3
Kirjoitustehtävien viimeistely	2	17 %	1	8 %	1	1

Taulukosta puuttuvat seuraavat hajamaininnoille (1–2 vastausta) jääneet käyttötavat: neuvojen pyytäminen, koodin analysointi, koodin selittäminen, esimerkkien generointi sekä tekstin tiivistäminen, kääntäminen ja laajentaminen.

Kun kielimallien käytön nähtiin edistävän oppimista, viitattiin selvästi useimmiten ajansäästöön. Hajamaininnoille jäivät mm. seuraavat syyt: välineet tarjosivat vaihtoehtoisia näkökulmia, lisä- ja vasta-argumentteja, ideoita ja esimerkkejä. Ne myös olivat jatkuvasti tarjolla, niiden kanssa saattoi keskustella, ne selittivät asioita ymmärrettävästi ja niille saattoi esittää ”tyhmiäkin kysymyksiä”.

Kielimallien haitoista eniten mainintoja sai mahdollisuus virheellisiin vastauksiin. Vaikka monet näkivät kielimallien käytön nopeuttavan työskentelyä, kaksi vastaajaa katsoi niiden hidastavan sitä. Kielimallien nähtiin haittaavan oppimista myös mm. siksi, että ohjelmakoodin tai tekstin kirjoittaminen itse on opettavaisempaa ja kirjoitustehtävissä ”tekstistä tulee koneen luomaa vaikka ei sitä huomaakaan”. Monessa vastauksessa erityisesti kurssilla Tietojenkäsittely ja yhteiskunta korostettiin välineiden oikeanlaisen käytön merkitystä: niitä tulisi käyttää tavoilla, jotka tukevat oppimista eivätkä vain helpota kurssien suorittamista.

Kielimallien käyttö korreloi Ohjelmistotekniikan harjoitustyössä negatiivisesti kokonaispisteiden kanssa: ne, jotka eivät lainkaan käyttäneet kielimalleja, saivat kurssilla keskimäärin 11,3 % enemmän pisteitä kuin kielimalleja käyttäneet. Pisteet olivat keskimäärin sitä heikompia, mitä enemmän kielimalleja oli käytetty. Kielimalleja koodin generointiin käyttäneiden kokonaispisteet olivat keskimäärin vain 76,6 % muiden saamista. Suurimmillaan tällainen ero saattaa johtaa kolmea yksikköä huonompaan lopputulokseen kurssilla käytetyllä arvosana-asteikolla 0…5. Vastaava prosenttilukema oli 88,9 %, kun kielimalleja oli käytetty koodin paranteluun. Toisaalta ne, jotka olivat käyttäneet kielimalleja yleiseen tiedonhakuun, saivat keskimäärin 14,7 % muita paremmat pisteet ja kielimallien käytöllä kääntäjän antamien virheilmoitusten tulkintaan ei ollut mainittavaa yhteyttä pisteisiin. Koska yksittäisissä vastauksissa oli usein viitattu useisiin käyttötapoihin ja pisteitä annettiin erilaisista asioista, analyysi olisi hyvä suorittaa hienojakoisemmin. Tätä ei kuitenkaan voitu mielekkäästi tehdä vastaajien vähäisen määrän vuoksi.

Ohjelmakoodia generoineista 42,9 % totesi, ettei generoitu koodi toiminut odotetusti. Muista generoidun koodin ongelmista mainittiin huono soveltuvuus, epäjohdonmukaisuus, monimutkaisuus, ”kummallisuus”, sopimattomuus

omaan ohjelmointityyliin ja ylimääräinen koodi. Koodia generoineista vastaajista 28,6 % kertoi havainneensa joitain ongelmia vasta testauksen, laadunvarmistuksen tai dokumentoinnin yhteydessä.

Ohjelmistotekniikan harjoitustyössä yksikään vastaaja ei yksiselitteisesti toivonut, että kielimallien käyttöä rajoitettaisiin nykyistä enemmän. Kurssilla Tietojenkäsittely ja yhteiskunta kaksi vastaajaa oli nykyistä tiukempien rajoitusten ja yksi rajoituksista luopumisen kannalla.

6 Yhteenveto

Kyselyihin vastanneiden lukumäärä jäi melko pieneksi, joten tulosten yleistämiseen tulee suhtautua varoen. Kielimallien käytön yleisyydessä ei voitu havaita suurta eroa kahdella kurssilla, mutta kurssilla Tietojenkäsittely ja yhteiskunta niiden käyttö nähtiin haitallisemmaksi oppimiselle. Yleisiä ja oppimista edistäviksi nähtyjä käyttötapoja olivat yleinen tiedonhaku ja ohjelmakoodin parantelu. Ohjelmistotekniikan harjoitustyössä edellinen korreloi positiivisesti mutta jälkimmäinen negatiivisesti kurssimenestyksen kanssa. Kielimallien käyttö ohjelmakoodin generointiin, monivalintoihin vastaamiseen ja kirjoitustehtävien valmisteluun nähtiin suhteellisen usein oppimisen kannalta haitallisiksi. Ohjelmakoodin generointi myös korreloi selkeän negatiivisesti kurssimenestyksen kanssa Ohjelmistotekniikan harjoitustyössä. Tämän syyn selvittäminen vaatisi jatkotutkimuksia.

Lähteet

Daun, M., & Brings, J. (2023). How ChatGPT will change software engineering education. In Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1 (pp. 110-116). DOI: 10.1145/3587102.3588815.

Kabir, S., Udo-Imeh, D. N., Kou, B., & Zhang, T. (2023). Who answers it better? an in-depth analysis of ChatGPT and stack overflow answers to software engineering questions. arXiv preprint arXiv:2308.02312.

Kiesler, N., & Schiffner, D. (2023). Large Language Models in Introductory Programming Education: ChatGPT's Performance and Implications for Assessments. arXiv preprint arXiv:2308.08572.

Liu, J., Tang, X., Li, L., Chen, P., & Liu, Y. (2023). Which is a better programming assistant? A comparative study between chatgpt and stack overflow. arXiv preprint arXiv:2308.13851.

Shoufan, A. (2023). Exploring Students' Perceptions of ChatGPT: Thematic Analysis and Follow-Up Survey. IEEE Access, vol. 11, pp. 38805-38818. DOI: 10.1109/ACCESS.2023.3268224.

Singh, H., Tayarani-Najaran, M. H., & Yaqoob, M. (2023). Exploring computer science students' perception of ChatGPT in higher education: A descriptive and correlation study. Education Sciences, 13(9), 924. DOI: 10.3390/educsci13090924.

Yilmaz, R., & Yilmaz, F. G. K. (2023). Augmented intelligence in programming learning: Examining student views on the use of ChatGPT for programming learning. Computers in Human Behavior: Artificial Humans, 1(2), 100005. DOI: 10.1016/j.chbah.2023.100005.