



# Kielimalli työntekijän apulaisena

Samu Willman

OPINNÄYTETYÖ  
Toukokuu 2024

Tietojenkäsittely  
Ohjelmistotuotanto

## TIIVISTELMÄ

Tampereen ammattikorkeakoulu  
Tietojenkäsittely  
Ohjelmistotuotanto

WILLMAN SAMU  
Kielimalli työntekijän apulaisena

Opinnäytetyö 31 sivua  
Toukokuu 2024

---

Opinnäytetyössä selvitettiin, mitä ovat kielimallit ja kuinka niitä voidaan hyödyntää. Työssä tarkastellaan suuria kielimalleja kuten GPT, Google Gemini ja Meta LLaMA. Opinnäytetyössä syvennytään erityisesti OpenAI:n kehittämään GPT-kielimalliin.

Opinnäytetyöllä pyritään helpottamaan uusien teknologioiden, kuten GPT- ja muiden kielimallien, käyttöönottoa työpaikoilla.

Opinnäytetyön tarkoituksena on selvittää, kuinka kielimalli voi tehostaa IT-alan työntekijöiden ja yritysten työskentelyä. Opinnäytetyöstä hyötyvät yritykset ja yksittäiset henkilöt, jotka haluavat ottaa projekteissaan käyttöön GPT-kielimallin. Työn lähdemateriaalina käytettiin aiheeseen sopivia artikkeleita ja tutkimuksia.

Selvityksen tuloksena todetaan, että käyttäjien saatavilla on sekä ilmaisia että maksullisia tehokkaita kielimalleja, joiden avulla voidaan tehostaa työskentelyä.

## **ABSTRACT**

Tampere University of Applied Sciences  
Degree Programme in Business Information Systems  
Software Development

SAMU WILLMAN

Language Models as an Employer's Assistant

Bachelor's thesis 31 pages

May 2024

---

The goal of the thesis was to investigate what language models are and how can they be used in work situations. The thesis focuses on the most common large language models such as ChatGPT, Google Gemini and Meta LLaMA. Out of the tools investigated in the thesis, the main focus is on the GPT language model developed by OpenAI.

The work was an investigative work, where the purpose of the thesis was to produce information for employers and companies about how the language model could make their work more efficient. The introduction of new technologies is often a problem in workplaces. This work should make it easier to enable the integration of language models in workplaces.

The source material used in the thesis was articles and studies suitable for the topic. The work was motivated by a personal interest in the subject.

The findings indicate that there exists both free and paid effective language models are available for users and with which can be used to make work more efficient and lessen mental stress in the everyday work life. Even though language models are still in their infancy, they are a very convenient tool for everyone looking to increase efficiency in their work.

---

Key words: artificial intelligence, ChatGPT, language model

## SISÄLLYS

1	JOHDANTO .....	6
2	Kielimalli.....	7
	2.1 Mikä on kielimalli? .....	7
	2.2 Kielimallien historiaa .....	9
	2.3 Kielimallien tulevaisuus .....	9
3	ChatGPT kielimalli .....	10
	3.1 GPT-kielimallit ja ChatGPT .....	10
	3.2 Eettisyys.....	11
	3.3 Vastuullinen käyttö .....	12
	3.4 AI-augmentoitu ohjelmistokehitys.....	13
4	Kielimallien hyödyntäminen maailmalla .....	14
5	ChatGPT-3.5 testitapaukset.....	15
	5.1 Verkkosivun luonti .....	15
	5.2 Kuntosaliohjelman luonti .....	17
	5.3 Esseen kirjoittaminen .....	21
	5.4 Yhteenveto.....	23
6	Vaihtoehtoisia kielimalleja.....	24
	6.1 Google Gemini .....	24
	6.2 LLaMA.....	24
	6.3 Kolmen kielimallin tuloksien vertailu.....	25
7	POHDINTA .....	28
	LÄHTEET.....	30

**ERITYISSANASTO JA TERMIT**

AI	Tekoäly
Skripti	Lyhyt ja helposti luettava tietokoneohjelma
IT-ala	Tietotekniikan ala
NLP	Natural Language Processing
Chatbot	Chatbotti, keskustelubotti

## 1 JOHDANTO

Nykypäivänä kuulee paljon keskustelua ja otsikoita tekoälystä ja siitä, kuinka se tulee viemään eri alojen työpaikkoja, mutta näin ei ole tapahtunut, ainakaan laajalla skaalalla. Tekoälyä voimme hyödyntää käyttämällä kaikille avoinna olevia kielimalleja. Olemme tietotekniikan saralla lähempänä kuin koskaan tilannetta, että voisimme käyttää tekoälyä tehostamaan jokapäiväistä työtämme. Erityisesti ohjelmistoalaa on kiinnostanut tehtävien automatisointi skripteillä ja uusimmat kielimallit helpottavat työn automatisointia. Automatisoimalla voimme mahdollistaa entistä nopeamman kehityksen.

Useat yritykset ovat pyrkineet luomaan omia tekoälyjään, näistä suurimpina ChatGPT, Google Gemini ja LLaMA. ChatGPT on valloittanut maailman saaden yli 100 miljoonaa käyttäjää kolmen kuukauden aikana, tehden siitä kaikkien aikojen nopeimmin kasvaneen kuluttajasovelluksen. ChatGPT ja muut mainitut suuret tekoälymallit ovat nousseet maailmanlaajuisiksi ilmiöksi ja vaikuttaa siltä, että niiden vauhti ei ole hiipumassa.

Opinnäytetyössä edetään seuraavassa järjestyksessä. Ensimmäiseksi käydään läpi kielimalleja, niiden historiaa ja niiden tulevaisuutta. Tämän jälkeen selvitetään, kuinka niitä voisi käyttää erilaisissa tilanteissa. Näiden jälkeen syvennytään tarkemmin ChatGPT:n toiminnallisuuteen ja selvitetään, kuinka keskiverto käyttäjä pystyisi saamaan mahdollisimman tehokkaita vastauksia itselleen. Tätä hahmotetaan kolmella testillä. Testit sisältävät kielimallin käyttöä erilaisissa työ- ja vapaa-ajan tilanteissa. Ensimmäisessä testitapauksessa tarkastelemme ohjelmoijaa, joka tarvitsee ChatGPT tekoälymallilta koodiapua. Testitapauksessa selvitämme sekä abstraktin että koodipohjaisen ratkaisun luomista. Toisessa testitapauksessa pyydämme tekoälymallia luomaan kuntosaliohjelma, joka vastaa pyynnön mukaisia tarpeita. Kolmannessa testitapauksessa pyydämme palvelua tuottamaan aineistoa esseeseen tyyliin. Tämän jälkeen perehdymme muihin kielimalleihin. Työn lopussa on pohdinta osio, jossa käydään läpi työtä kirjoittajan oman pohdinnan kautta.

## 2 Kielimalli

### 2.1 Mikä on kielimalli?

Kielimallit ovat koneoppimismalleja, jotka ovat suunniteltu käsittelemään niille annettua tekstiä. Ne ovat NLP:n (Natural Language Processing) kulmakiviä. NLP:llä tarkoitetaan tekoälyn alalajia, jonka tarkoituksena on ihmiskielen ymmärtäminen. Nämä mallit käsittelevät niille annetun tekstin mallin sisältämien algoritmien avulla ja palauttavat vastauksen ihmiselle ymmärrettävässä muodossa. Kielimallit ovat generatiivista tekoälyä. Generatiivisella tekoälyllä tarkoitetaan sitä, että ne pystyvät luomaan uutta sisältöä. Yleisimpiä käyttötarkoituksia kielimalleille ovat tekstin generointi, konekäännökset, puheen tunnistaminen, tekstin luonti ja tekstin tiivistäminen.

Kielimalleja usein jaetaan kahteen eri kategoriaan, jotka ovat todennäköisyyspohjaiset kielimalli ja neuroverkkokielimallit. Todennäköisyyspohjaisessa kielimallissa lasketaan lauseen todennäköisyysjakauma. Kuvitellaan esimerkkinä seuraava lause "Mikä sinun nimesi \_\_\_". Voimme olettaa, että käyttäjän seuraava syöte olisi "on". Hyödyntämällä todennäköisyyslaskelmia kielimalli pystyy ehdottamaan seuraavaa sanaa. Kyseistä todennäköisyyden hyödyntämistä käytetään esimerkiksi puhepohjaisissa kielimalleissa, kuten Applen kehittämässä Sirissä tai tekstiviestien automaattisissa korjauksissa.

Toinen suurista todennäköisyysmalleista on n-gram malli. n-grammi on n määrä tavuja, sanoja tai kirjaimia. Voimme luoda oman n-grammin käyttämällä yksinkertaista kaavaa.

$$P(w | h) \quad (1)$$

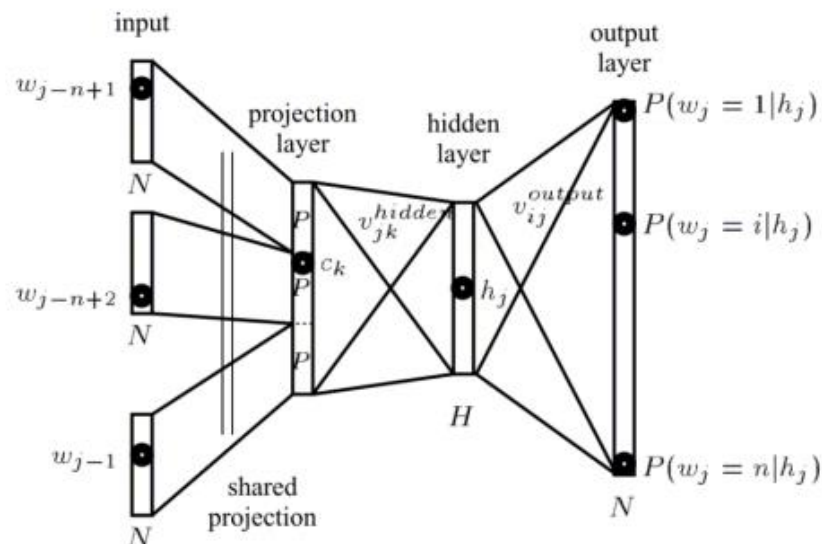
Kaavassa (1) P:llä tarkoitetaan todennäköisyyttä, w:llä annetun sanan todennäköisyyttä ja h:lla tarkoitetaan lauseen historiaa. Käyttäen aiemmin esiteltyä lausetta "Mikä sinun nimesi", voimme asettaa lauseen osat annettuun kaavaan (2)

$$P(on | Mikä sinun nimesi) \quad (2)$$

Todennäköisyyslaskennan rungon saamme jakamalla ”Mikä sinun nimesi on” hakutulosten määrän ja vertaamalla sitä ”Mikä sinun nimesi” hakutulosten määrään hakukoneessa. N-grammien suurimpia ongelmia ovat, että ne vaativat paljon prosessointitehoa ja kielen jatkuva muutos vaikeuttaa tulosten vertailua. N-gram pohjaisille todennäköisyysmalleille tehdään usein ”pehmennyksiä”.

Neuroverkkopohjaiset kielimallit ovat nykyään yleisimmin käytettyjä kielimalleja, koska ne ovat huomattavasti vanhanaikaisia todennäköisyysmalleja tarkempia ja tehokkaampia. Neuroverkkopohjaisissa kielimalleissa samantyylliset sanat klusteroidaan yhteen. Klusteroinnilla tarkoitetaan samankaltaisten sanojen yhdistämistä.

Kuvassa 1 on neuroverkkopohjainen kielimalli havainnollistettuna. Neuroverkot yhdistyvät toisiinsa solmuilla. Solmuilla on oma painoarvonsa ja ne ovat yhteydessä seuraavan kerroksen solmuihin. Syötteen annettuaan neuroverkko lähettää ensimmäisestä solmusta viestin seuraavalle solmulle ja seuraava solmu tarkistaa ylittääkö annettu syöte sille annetun kynnyсарvon. Lopulta päädytään uloimmalle kerrokselle eli tulostekerrokselle, joka palauttaa käyttäjälle sopivimman tuloksen.



KUVA 1. Neuroverkkopohjaisen kielimallin havainnointikuva (Saqib Alam, Nianmin Yao, 2018, 3).

## 2.2 Kielimallien historiaa

Ensimmäinen maininta n-gram malleista löytyy vuodelta 1948, jolloin Claude Shannonin teos "A Mathematical Theory of Communications" julkaistiin. Ensimmäiset NLP:t voidaan jäljittää 1950-luvulle. 1954, IBM kehitti ensimmäisen konekäännöksen, jossa se käänsi kuusikymmentä venäjän kielen sanaa englannin kielelle (Hutchins, 2005, 5). 1980- ja 1990-luvuilla todennäköisyyspohjaiset kielimallit alkoivat kiinnostaa tutkijoita ja n-gram mallit alkoivat yleistyä. Vuosituhannen vaihteessa tutkimus siirtyi neuroverkkoihin (Chiusano, 2022).

## 2.3 Kielimallien tulevaisuus

Kielimallit ja niitä hyödyntävät tekoälyt ovat alkuvaiheissaan. Nykyisissä tekoälyissä on useita ongelmia, joita kehittäjät pyrkivät ratkaisemaan. Kielimalleille yleinen ongelma on niiden tarkkuus ja ne saattavat hallusinoida niille uusissa asioissa. Uudet ja kehittyvät sanastot ovat myös ongelma esikoulutetuilla kielimalleilla.

Kansainvälisesti tunnettu lehti Forbes listaa kolme uutta suurien kielimallien kehitysaluetta (Toews, 2023). Ensimmäinen suuri kehitysalue tulee opetusdatasta. Forbes arvioi, että kielimallit alkavat tuottamaan omaa kehitysdataansa ja sitä kautta oppimaan itse uusia asioita. Tämä tarkoittaisi sitä, että ne olisivat itseänsä kehittäviä kielimalleja. Tätä teoriaa tukee myös Illinoisin yliopiston tekemä tutkimus itseänsä kehittävästä kielimalleista (Huang ym., 2022). Toisena suurena kehityksenä artikkelissa mainitaan, kuinka kielimallit voisivat tarkistaa omat faktansa. Nykypäivän kielimalleissa hallusinointi ja valheellisen tiedon antaminen on yleistä. Jos kielimalli tarkistaisi omat faktansa, voisi se vähentää hallusinoinnin määrää ja valheellista tietoa vastauksissaan. Kolmantena kehityksenä artikkelissa mainitaan kuinka massiiviset "harvat" kielimallit tulevat yleistymään. "Harvoilla" kielimalleilla tarkoitetaan kielimalleja, jotka pystyvät aktivoimaan vain tarvittavat parametrit käyttäjän pyynnön toteuttamiseksi.

### 3 ChatGPT kielimalli

#### 3.1 GPT-kielimallit ja ChatGPT

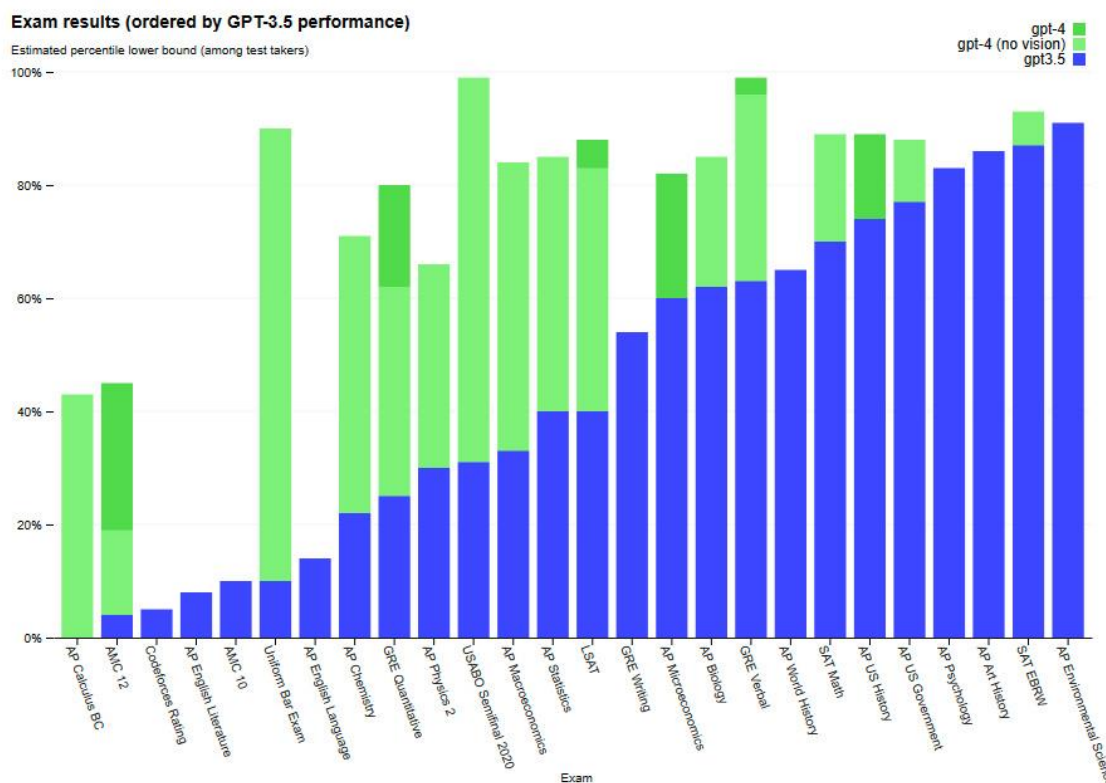
GPT on OpenAI:n kehittämä suuri kielimalli, jonka avulla se pystyy päättämään algoritmien avulla parhaimman mahdollisen ihmiselle ymmärrettävän vastauksen. Se on julkaistu marraskuussa vuonna 2022 ja se on käytettävissä ilmaiseksi verkkoselain ja mobiiliversiona ChatGPT:n avulla. ChatGPT on GPT-kielimalliin pohjautuva tekoälymallo, joka toimii keskustelurajapintana GPT-kielimallille. GPT perustuu neuroverkko arkkitehtuuriin ja sitä pidetään ensimmäisenä suurena kielimallina. ChatGPT:n tämänhetkinen ilmainen versio hyödyntää GPT-3.5-kielimallia ja käyttäjät voivat halutessaan ostaa kuukausimaksulla käyttöönsä GPT-4.0-kielimallin. Kuukausimaksun hinta on 20 dollaria kuukaudessa. Käyttäjien pitää luoda oma käyttäjänsä OpenAI:n palveluihin käyttääkseen sekä maksullista että ilmaista versiota.

GPT-3.5 eroaa aikaisemmista kielimalleista siten, että siinä käytetään yli 175 miljardia parametria, joka tekee GPT-4 mallia lukuunottamatta tähän asti suurimman kielimallin. (Martindale & Bizzaco, 2024). Se on koulutettu käyttämällä 45 terabitteä tekstiä useista eri verkkolähteistä ja kirjoista.

GPT-4 on kielimallin uudempi versio, joka tuli markkinoille maaliskuussa 2023. Uusi versio ei ole vain kielimalli, vaan se sisällytetään multimodaalisiin kielimalleihin. Tällä tarkoitetaan sitä, että kielimallin neljäs versio pystyy myös käsittelemään käyttäjän antajan tekstin lisäksi kuvia. Suurimmat muutokset versioiden välillä on koulutukseen käytettyjen parametrien määrä, joka johtaa noin 40 prosenttia tarkempiin vastauksiin (OpenAI, n.d.), uudet tuetut kielet, parempi tuki kolmannen osapuolen sovelluksille ja muistin kehitys. Medium kirjoittaa artikkelissaan seuraavasti: "The company also says that compared with GPT-3.5, GPT-4 is 82% less likely to respond when prompts are technically not allowed, and it's 60% less likely to fabricate facts, which in AI terms are called "hallucinations"..." (Griffith, 2023).

Kielimallien yksi suurimmista ongelmista on niiden koulutus. Kielimallin kouluttaminen vaatii suuren määrän dataa, resursseja ja aikaa. Saatavilla olevista resursseista riippuen kielimallin kouluttaminen saattaa viedä viikkoja, ellei jopa kuukausia. On arvioitu, että GPT-4 kielimallin koulutus kesti OpenAI:lta neljästä viiteen kuukauteen (TRGDatacenters, n.d.).

Kuvan 2 pylväskaaviosta voimme huomata kuinka paljon paremmin GPT-4 on suoriutunut testeissä kuin GPT-3.5.



KUVA 2. ChatGPT-3.5 ja ChatGPT-4 testitulokset (OpenAI, 2023)

### 3.2 Eettisyys

Tekoälyn eettisyys on harmaata aluetta. Onko se plagiointia, jos kielimalli lainaa suuren osan tekstiä käyttäjälle tuntemattomalta verkkosivulta? Onko se väärin, jos käyttäjä pyytää tekoälyä tekemään hänen ihannoimansa maalarin taidetyyliin maalauksen? Huolet tekoälymalleissa liittyvät usein plagiointiin, tekijänoikeuden loukkauksiin, yksityisyyden loukkauksiin ja puolueellisiin mielipiteisiin.

Kielimallit ovat vielä kaukana täydellisyydestä. On olemassa useita tapauksia, joissa kielimallit suostuvat täyttämään käyttäjien epäeettisiä toiveita, valehtelemaan, keksimään faktoja ja jopa luomaan haitallista sisältöä. Koska GPT-kielimallit ovat koulutettu käyttämällä internetissä saatavilla olevaa julkista dataa, on mahdollista, että datasta löytyy epäeettistä informaatiota. Kielimallien data voi siis myös sisältää muiden ihmisten tuottamaa dataa, joka saattaa olla tekijänoikeussuojattu. Tulevaisuudessa tiedon käsittelyyn ja rajoituksiin tullaan keskittymään tarkemmin. Useat tekoälyyn liittyvät lait ovat vielä kehitteillä ja niiden suunnittelu tulee viemään aikaa.

### 3.3 Vastuullinen käyttö

Kielimallien vastuullisesta käytöstä on vastuussa käyttäjä. Reuters kirjoittaa tapauksesta, jossa käyttäjä on saanut tekoälyllä luodusta työstä tekijänoikeussyytteen (Brittain, 2023). Käyttämällä kielimalleja vastuullisesti käyttäjä varmistaa että ei joudu ongelmiin tulevaisuudessa.

Seinäjokelainen yritys Lohde Oyj oli luonut seuraavat tekoälyn vastuullisen käytön käytännöt yrityksensä työntekijöille:

- Tutkimme ja kokeilemme aktiivisesti generatiivisen AI:n hyödyntämisen mahdollisuuksia asiantuntijatyön tehostamisessa.
- Olemme läpinäkyviä asiakkaidemme suuntaan. Mikäli tuotoksessa on generatiivisen AI:n tuottamia osia (esimerkiksi koodia, kuvia, tekstiä, kaavioita jne.), kerromme siitä.
- Huolehdimme ja vastaamme siitä, ettei AI-ratkaisujen käyttö vaaranna meidän tai asiakkaidemme tietoturvaa ja -suoja.
- Vastuu lopputuotoksesta on aina ja yksiselitteisesti Lohdeen työntekijällä (eli ihmisellä). Käyttäjän on huomioitava työkalun tai mallin rajoitteet.
- Huolehdimme ja vastaamme siitä, etteivät tuottamamme aineistot tai muut tuotteet loukkaa kolmansien osapuolten immateriaali- ja/tai muita oikeuksia.
- Tiedostamme ratkaisujen laskentatehon tarpeen ja vaikutuksen energiankulutukseen ja käytämme ratkaisuja vastuullisesti.

- Varmistamme että asiantuntijamme tuntevat tekoälyn hallintajärjestelmän sekä vastuullisen ja turvallisen AI:n hyödyntämisen käytännöt.  
(Loihde, 2023)

### 3.4 AI-augmentoitu ohjelmistokehitys

AI-augmentoidulla ohjelmistokehityksellä tarkoitetaan tekoälyn hyödyntämistä ohjelmistokehityksessä. Ihmisen ja tekoälyn yhdistelmä ohjelmistokehityksessä ei vain nopeuta ohjelmistokehitystä, vaan myös parantaa sen laatua. Tekoälyn avulla voidaan vähentää toistuvia ohjelmistokehitystehtäviä.

GitHub Copilot on yksi tunnetuimmista AI-augmentoiduista ohjelmistotyökaluista. Vuonna 2022 Github julkaisi laajamittaisen tutkimuksen, jossa tutkittiin GitHub Copilot tekoälymallia ja sen käyttäjiä. Tutkimuksessa perehdyttiin kokonaisvaltaisen tuottavuuteen, kehittäjien perspektiiviin ja Copilotin vaikutukseen ohjelmointityössä. Tulokset olivat erittäin positiivisia, sillä noin 60–75% työntekijöistä kertoi olevansa tyytyväisempiä työhönsä, 83% työntekijöistä kertoi olevansa vähemmän henkisesti uupuneita ja työt valmistuivat keskiarvallisesti noin 55% nopeammin (Eirini Kalliamvakou, 2022).

AI-augmentoidun ohjelmointi on läpikäymässä muutoksen lähivuosien aikana. Github julkaisee lähitulevaisuudessa Copilot workspacen, jonka avulla käyttäjät voivat toteuttaa kokonaisuuksia entistä nopeammin. Copilot workspacen taustalla toimii Github Copilot kielimalli. Tekoälylle syötetään haluttu toiminnallisuus tekstimuodossa ja Copilot toteuttaa sen työtilan sisällä. Käyttäjä pystyy tämän jälkeen tarkastamaan kielimallin toteuttaman kokonaisuuden. Tarkistuksen jälkeen kielimallin tekemä koodi on mahdollista sisällyttää käyttäjän projektiin.

## 4 Kielimallien hyödyntäminen maailmalla

Monet yritykset Suomessa ja ulkomailla käyttävät chatbotteja parantaakseen käyttäjiensä asiakaskokemusta. Chatbotit pohjautuvat usein tekoälyyn ja hyödyntävät kielimalleja.

Expedia on yksi maailman suurimmista verkkomatkoimistoista. Expedia käyttää verkkosivultaan löytyvässä chatbotissa generatiivista tekoälyä, joka hyödyntää GPT-kielimallia. Chatbotissa käytettävää tekoälyä hyödynnetään avustajana toimimisen lisäksi matkojen suunnitteluun.

Verkkokurssien tarjoaja Udacity hyödyntää kielimallia opetuksessaan. Älykäs virtuaaliavustaja pystyy luomaan personoitua opastusta, antamaan palautetta opiskelijalle ja korjaamaan opiskelijan virheitä ohjelmointiin liittyvissä kursseissa. Virtuaaliavustaja pohjautuu GPT-4-kielimalliin. Virtuaaliavustaja pystyy myös kääntämään kursseja opiskelijan omalle äidinkielelleen kun opiskeltava kurssi ei ole saatavillaan hänen omalla äidinkielellään.

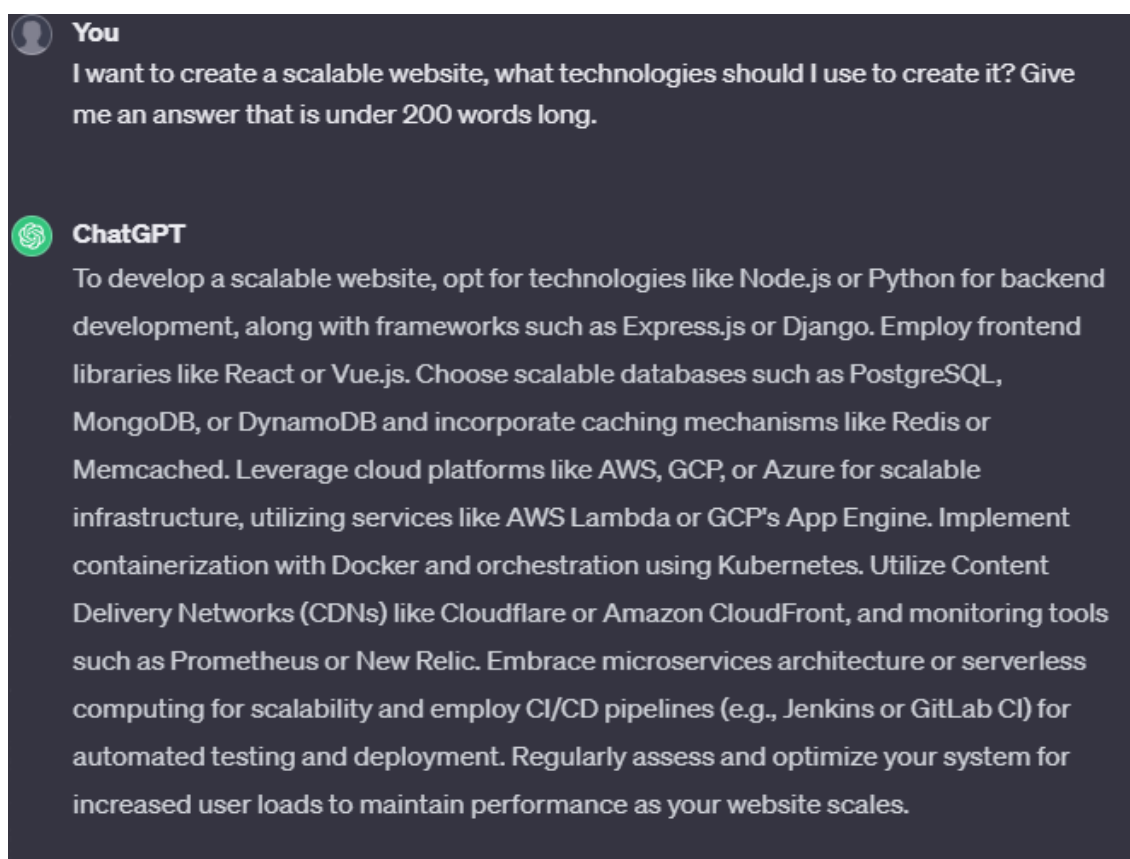
Kielten opettajana tutuksi tullut Duolingo hyödyntää GPT-4 kielimallia kahdella eri tavalla. Ensimmäinen näistä keskittyy opiskelijoiden vastauksiin ja palautteen antamiseen. Toinen kielimallin hyödyntämiskeino on keskustelutilanteiden luominen. Keskustelutilanteita luodaan keskustelemalla kielimallille annetun persoonan kanssa. Tavoitteena on luoda tilanne, jossa opiskelija voisi hyödyntää opiskeltua kieltä.

## 5 ChatGPT-3.5 testitapaukset

Seuraavissa testitapauksissa selvitämme erilaisia tapoja tehdä ChatGPT-3.5 kielimallin käytöstä käyttäjäystävällisempää ja tehokkaampaa. ChatGPT:ltä haluttujen vastausten pituutta on rajoitettu esimerkkien selkeyden vuoksi.

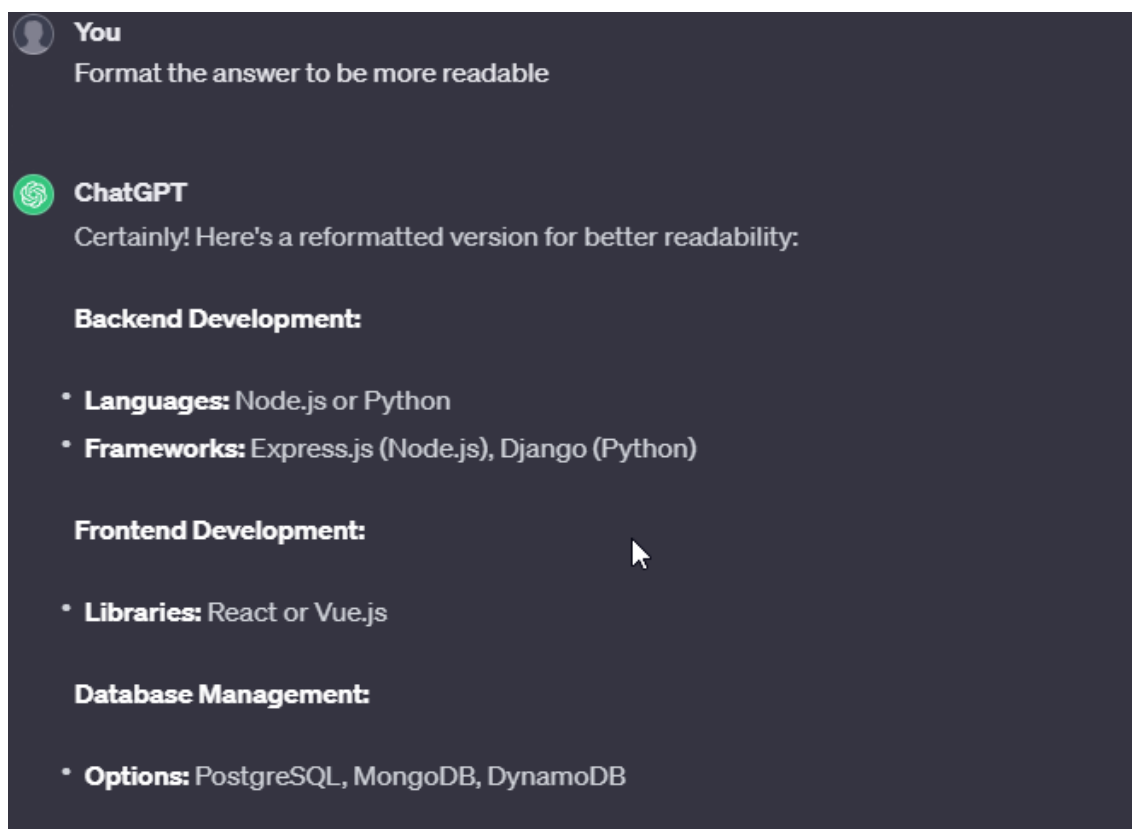
### 5.1 Verkkosivun luonti

Ensimmäisessä testitapauksessa käyttäjä on ohjelmistoalalla työskentelevä ohjelmoija. Ohjelmoija haluaa toteuttaa uuden sovelluksen, jonka tarkoituksena on toimia helposti skaalautuvana verkkosivuna.



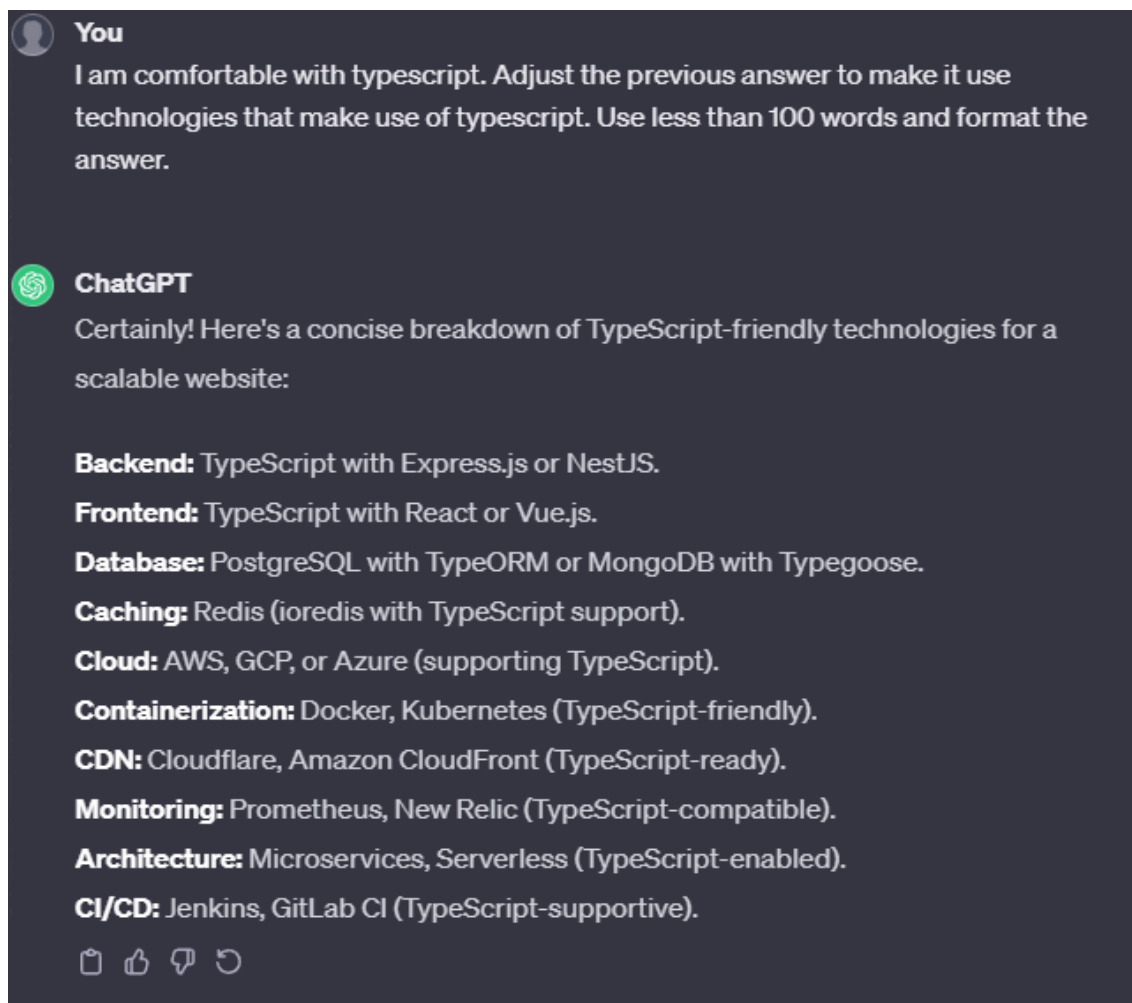
KUVA 3. ChatGPT:n vastaus käyttäjän kysymykseen skaalautuvista verkkoteknologioista.

Kuvassa kolme pyydämme ChatGPT:tä rakentamaan skaalautuvan kokonaisuuden verkkosivun rakentamiseen. Tekoälyn antama vastaus on laadukas ja antaa hyvän esimerkin vaihtoehtoisista teknologioista verkkosivun kehittämiseen.



KUVA 4. ChatGPT:n vastaus käyttäjän pyyntöön formatoida aiempi vastaus.

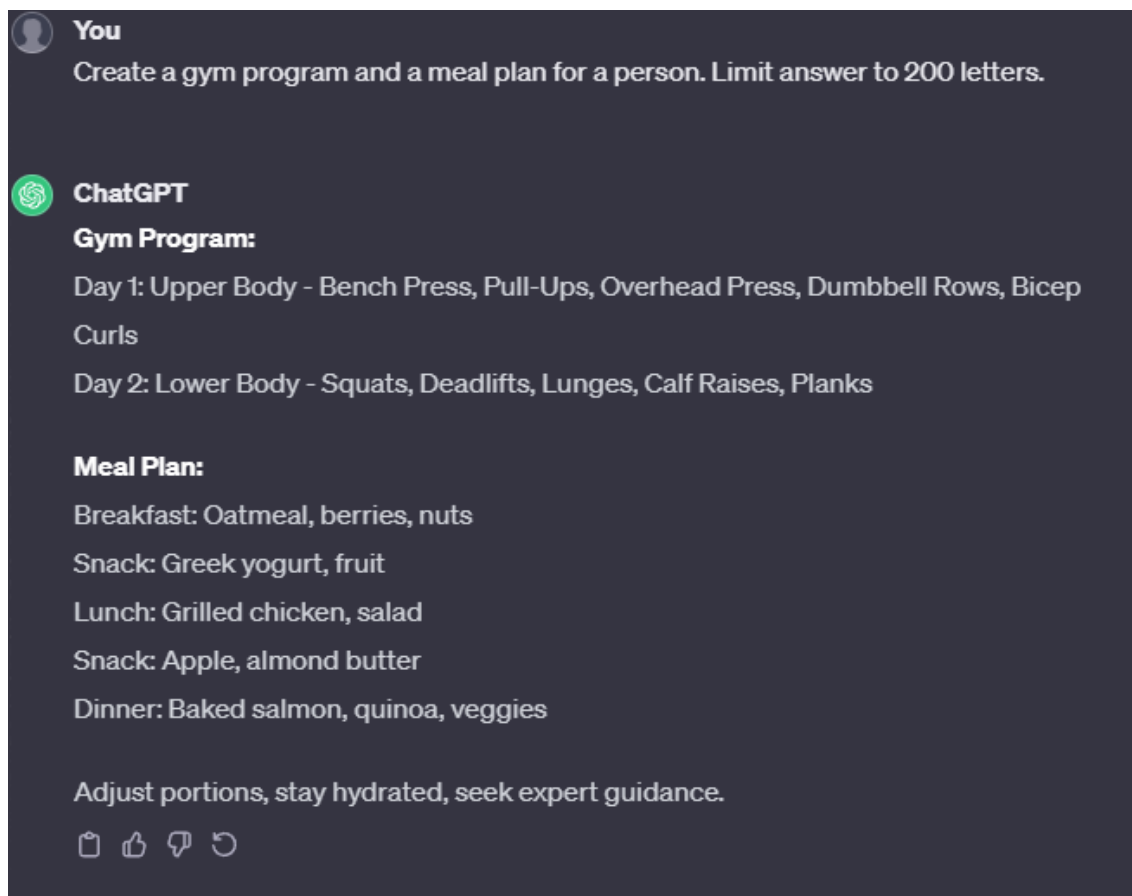
Kuvan 3 kielimallin antama vastaus ei ollut formatoitu, joten voimme pyytää ChatGPT:tä formatoimaan vastauksen luettavampaan muotoon. Antamalla lisätietoa ChatGPT:lle käyttäjän osaamisesta, voimme saada entistä tarkempia vastauksia. Kuten voimme nähdä kuvista 4 ja 5, ChatGPT muistaa käyttäjän aikaisemmat kysymykset ja kielimallin aikaisemmat vastaukset, joten voimme referoida aiempiin viesteihin kysymyksissä. Hyödyntämällä tekstissä aiemmin mainittuja tekniikoita, voimme saada kielimallilta luettavamman ja kohdennetumman vastauksen.



KUVA 5. ChatGPT:n vastaus käyttäjän pyyntöön muokata aiempaa vastausta.

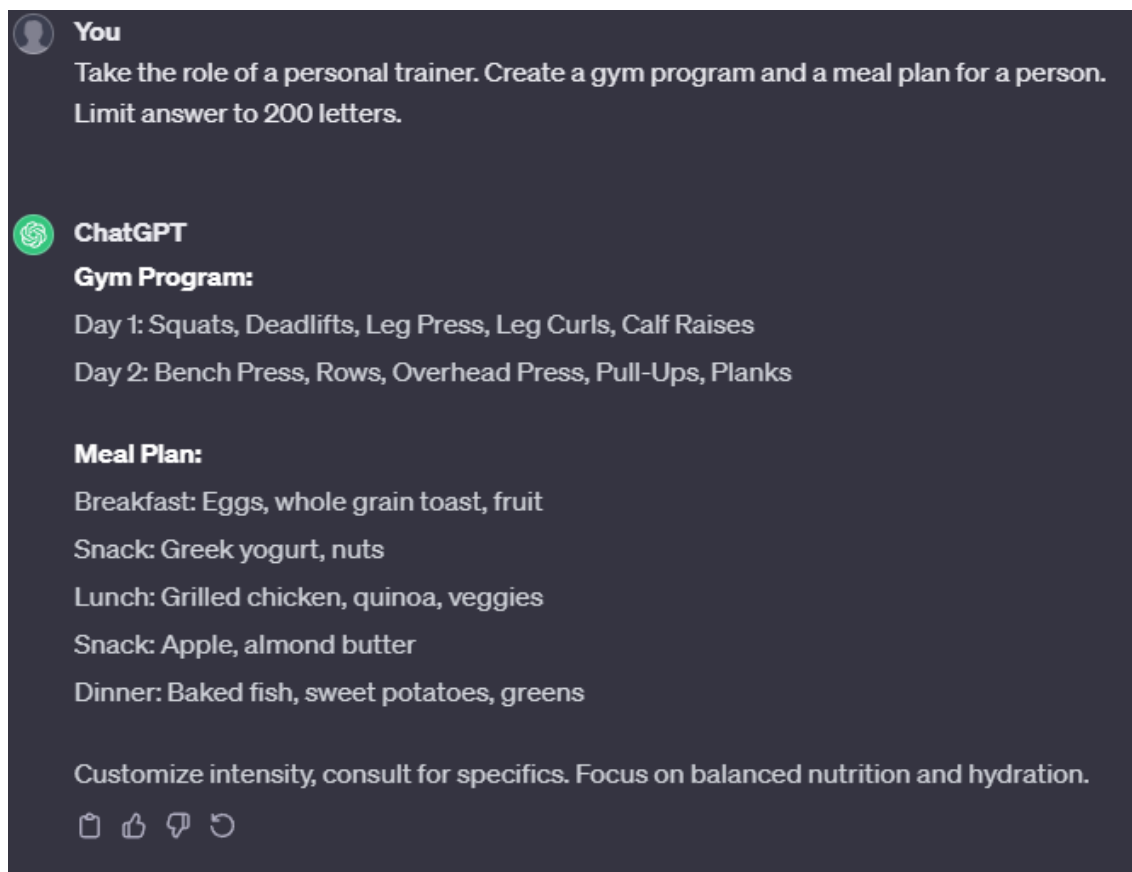
## 5.2 Kuntosaliohjelman luonti

Toisessa testitapauksessa käyttäjä on innostunut kuntosaliturheilusta, mutta ei tiedä mistä aloittaa. Käyttäjä haluaa ChatGPT:n luovan hänelle kuntosaliohjelman ja ateriasuunnitelman. Testitapauksessa pyrimme selvittämään kuinka voimme tarkentaa syötettä antamalla ChatGPT:lle roolin ja kuinka hyvin ChatGPT pystyy auttamaan käyttäjää uudessa harrastuksessa.



KUVA 6. ChatGPT:n vastaus käyttäjän pyyntöön luoda kuntosaliohjelma ja ateriasuunnitelma.

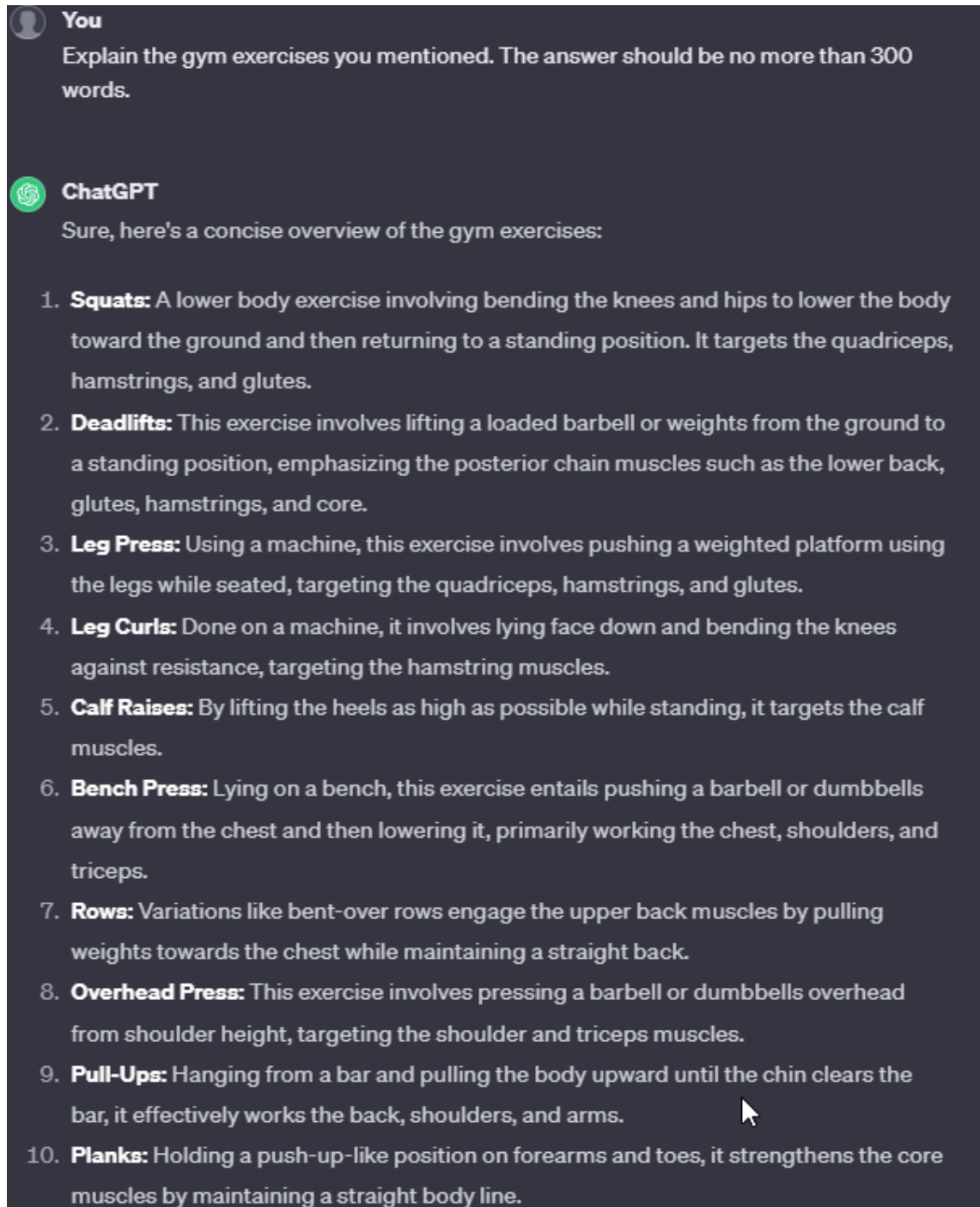
Kuvassa 6 näemme ChatGPT:n luoman kuntosaliohjelman ja ateriasuunnitelman. Voimme muokata tätä vastausta hyödyntämällä kielimallien säätelyyn kehitettyjä tekniikkoa. Antamalla ChatGPT kielimallille roolin jonka näkökulmasta haluamme hänen antavan vastauksen, saamme suuremmalla todennäköisyydellä tarkemman vastauksen. Roolin antaminen voi auttaa käyttäjää saamaan vastauksen, joka on hänelle helposti ymmärrettävissä. Roolin antamisessa kielimallille on myös vaaransa, koska kielimallin antama vastaus voi olla puolueellinen.



KUVA 7. ChatGPT:n vastaus käyttäjän pyyntöön luoda kuntosaliohjelma ja ateriasuunnitelma käyttäjälle kuntosaliohjaajan näkökulmasta

Kuvasta 7 voimme huomata kuinka kielimallin annettu vastaus muuttui roolin antamisen jälkeen. Antamalla roolin lisäämme lauseen kontekstia. Kaikki kontekstia lisäävä ja kysymystä tarkentava informaatio tarkentaa myös vastauksen tarkkuutta.

Kuvitteellinen henkilömme on kuitenkin vasta aloittelija eikä tiedä miten annettuja kuntosaliharjoituksia tehdään. Kuvasta 8 näemme ChatGPT:n selittämät kuntosaliharjoitteet selitettynä. Kuvassa 9 näemme ChatGPT:n selittämät kuntosaliharjoitteet, jos se selittäisi niitä kymmenvuotiaalle. Kuvista voimme huomata myös, kuinka vastausten tyyli on täysin erilainen. Kymmenvuotiaalle selitettäessä myös teksti on huomattavasti ymmärrettävämpää, kun taas aikuiselle selitettäessä on tekstissä enemmän syvyyttä.



KUVA 8. ChatGPT:n vastaus käyttäjän pyyntöön selittää kielimallin antavat kuntosaliharjoitukset.



KUVA 9. ChatGPT:n vastaus käyttäjän pyyntöön selittää kuntosaliharjoitteet kymmenvuotiaalle.

### 5.3 Esseen kirjoittaminen

Kolmannessa testitapauksessa selvitämme kuinka sujuvaa ja luovaa suomenkielistä tekstiä ChatGPT pystyy kirjoittamaan. Seuraava teksti on täysin ChatGPT:n luomaa tekstiä.

”ChatGPT on kehittynyt tekoälymalli, joka perustuu avoimen lähdekoodin GPT (Generative Pre-trained Transformer) -arkkitehtuuriin. Se kykenee tuottamaan ihmismäisiä tekstejä erilaisista aiheista, mukaan lukien esseet. Tämä osio keskittyy

siihen, miten ChatGPT pystyy luomaan laadukkaita esseitä ja millaisia strategioita se käyttää tässä prosessissa.

ChatGPT:n kyky kirjoittaa esseitä perustuu sen valtavaan määrään kielellistä tietoa ja ymmärrykseen eri aiheista. Tekoälymalli on koulutettu valtavalla tekstiaineistolla, joka kattaa monenlaisia aiheita ja tyyllilajeja. Tämä mahdollistaa sen, että ChatGPT pystyy tuottamaan jäsenneiltyjä ja ymmärrettäviä esseitä eri aihealueilta.

ChatGPT käyttää kielen generointitekniikoita luodessaan esseitä. Se hyödyntää kontekstia, jota se saa syötteeksi käyttäjältä tai kysytystä aiheesta, ja rakentaa tämän pohjalta johdonmukaisen ja loogisen tekstin. Tekoäly kykenee tuottamaan jäsenneiltyjä esseitä, sillä se ymmärtää kielioppia, kontekstuaalista merkitystä ja erilaisten argumenttien rakentamista.

Esseiden tuottaminen ChatGPT:n avulla vaatii kuitenkin huolellista ohjeistamista ja tarkkaa syöttämistä. Käyttäjän on annettava selkeä aihe tai kysymys, jotta tekoäly pystyy tuottamaan tarkoituksenmukaisen ja relevantin tekstin. Lisäksi tulee huomioida, että vaikka ChatGPT kykenee luomaan vakuuttavia tekstejä, sen tuottamat esseet voivat sisältää virheitä tai epätarkkuuksia, erityisesti hyvin erikoistuneissa aiheissa, sillä se perustuu opittuun tietoon.

Yhteenvetona voidaan todeta, että ChatGPT:n kyky kirjoittaa esseitä perustuu sen laajaan kielitaitoon, kykyyn ymmärtää kontekstia ja luoda johdonmukaista tekstiä. Vaikka se tarjoaa merkittävän apuvälineen esseiden luomiseen, on tärkeää muistaa tarkistaa ja tarvittaessa muokata tekoälyn luomia tekstejä ennen niiden lopullista käyttöä.”

Syötteenä käytettiin seuraavaa: ”Kirjoita osio opinnäytetyöhön, jossa käsittelet miten ChatGPT pystyy kirjoittamaan esseitä.”

Teksti on yleisellä tasolla hyvin laadukasta suomen kieltä. Tekstissä kuitenkin esiintyy kirjoitusvirheitä ja hallusinaatiota. Tekstistä käyttäjän olisi mahdollista saada työlleen vankka pohja, josta käyttäjä voisi jatkaa kirjoittamista. ChatGPT on kätevä malli alustuksen luomiseen.

## 5.4 Yhteenveto

Aiemmissa alaotsikoissa kävimme läpi erilaisia tekniikoita ChatGPT:n tehokkaan käyttöön ja näistä voimme huomata kuinka käteviä työkaluja ne voivat olla. Esitellyt tekniikat ovat sovellettavissa työympäristöön. ChatGPT ei suoriutunut testeissä täydellisesti, joten käyttäjän on hyvä muistaa olla kriittinen kielimallilta saadun vastauksen kanssa. Tekniikoita käyttämällä käyttäjät voivat maksimoida ChatGPT:n potentiaalin kustomoimalla sen käyttäjän tarpeisiin, tehostamalla käyttäjän ja kielimallin välistä vuorovaikutusta ja samalla tehden käyttökokemuksesta käyttäjäystävällisemmän.

## 6 Vaihtoehtoisia kielimalleja

### 6.1 Google Gemini

Google Gemini on Googlen kehittämä ja julkaisema kielimalli, joka julkaistiin ensimmäisen kerran 8.2.2024. Google Gemini alun perin julkaistiin Google Bard nimellä 21.3.2023 ja sen nimi myöhemmin muutettiin Google Geminiksi. Google käyttää Gemini nimeä sekä tekoälymallista ja tekoälymallin taustalla olevasta kielimallista. Kielimalli on saatavilla ilmaisena ja maksullisena versiona. Ilmainen versio sisältää normaalin Gemini 1.5 kielimallin, kun taas maksulliset Gemini 1.5 Pro ja Gemini 1.5 Ultra ovat saatavilla lisähintaan.

Google Gemini on ladattavissa sovelluksena App Storesta ja Google Play Storesta. Google Geministä on mahdollista tehdä avustaja Android puhelimilla. Tällä tarkoitetaan sitä, että käyttäjät voivat halutessaan korvata nykyisen Google Assistantin Google Geminillä. Google Gemini on saanut avustajana sekalaisen vastaanoton (Johnson, 2024).

Samoin kuin ChatGPT, Google Gemini pystyy hyväksymään syötettä usealla eri tavalla. Syötetapoihin kuuluu muun muassa teksti, ääni, kuvat ja videot. Vain harvalle kielimallille on mahdollista syöttää useaa eri syötettä (Allan & Blake, 2024).

Kielimallin uniikein ominaisuus on sen kyky luoda kuvia tekstistä. Kielimallin luomat kuvat eivät välttämättä yllä samalle tasolle kuin kilpailijansa, mutta kuvien luominen ja tekstin prosessoiminen samalla kielimallilla on harvinaista. Muut kuvat luovat kielimallit kuten Dall-E, Midjourney ja Stable Diffusion ovat normaaleista kielimalleista poikkeavia, kuvien luomiseen tarkoitettuja kielimalleja. Google Geminin luomat kuvat eivät olleet ilman ongelmia, sillä tekoälyn julkaisemisen jälkeen kielimalli poistettiin väliaikaisesti käytöstä sen luomien kuvien takia. Tekoälymallilla pystyi luomaan kuvia mitä ei olisi pitänyt pystyä luomaan, kuten esimerkiksi kuvia ihmisille herkistä aiheista.

### 6.2 LLaMA

LlaMA on Facebookin kehittäjä Metan julkaisema ja Perplexityn kehittämä kielimalli, joka julkaistiin ensimmäisen kerran 24. helmikuuta 2023. LlaMA on Google Geminin ja GPT:n kanssa suosituimpien kielimallien joukossa.

LlaMA erottuu joukosta uniikeilla ominaisuuksillaan. Suurimpia eroja muihin kielimalleihin ovat kouluttamiseen käytetty data, lisenssit ja prosessointiin tarvittavat resurssit. Kielimalli on koulutettu puhtaasti käyttämällä tieteellisiä artikkeleita, uutisartikkeleita ja muuta luotettavaa dataa käyttämällä. Tämä tarkoittaa sitä, että kielimalli pystyy tuottamaan akateemista tekstiä jopa ChatGPT-kielimallia tehokkaammin. LlaMA on kehitetty olemaan erittäin kevyt kielimalli. Kevyt kielimalli mahdollistaa sen helpon integroinnin sovelluksiin ja verkkosivuihin, jotka eivät pysty hyödyntämään huomattavaa määrää prosessointitehoa. Kielimalli on siis täydellinen esimerkiksi älykkäisiin chatbotteihin. LlaMA:n suurimpia vahvuuksia on myös sen lisenssi. LlaMA käyttää ei-kaupallista lisenssiä, joka mahdollistaa sen kaupallisen käytön muita kielimalleja helpommin.

### **6.3 Kolmen kielimallin tuloksien vertailu**

Seuraavassa testissä selvitämme, kuinka hyvin edellä mainitut kielimallit osasivat tiivistää New York Timesin artikkelin. Artikkeleiden ja tutkimusten hyödyntäminen on tärkeä osa ohjelmoijan työtä, koska niiden avulla on mahdollista sisäistää tärkeimmät osat suuresta informaatiosta nopeasti. Kyseistä tekniikkaa voidaan myös hyödyntää ohjelmointikielien tai dokumentaatioiden tiivistämiseen. Aiheista on mahdollista löytää tärkeimmät informaatiot pyytämällä kielimallilta haettua informaatiota.

- The following website has an article, summarise it for me with less than 300 words: <https://www.nytimes.com/2023/03/31/technology/ai-chatbots-benefits-dangers.html>




The article from The New York Times discusses the rapidly growing use of artificial intelligence (AI) chatbots and the benefits and dangers they bring. Chatbots are computer programs designed to simulate human-like conversations with users, and they are being increasingly used in various industries such as customer service, healthcare, and education.


The benefits of chatbots include their ability to provide 24/7 customer support, quickly answer simple questions, and help with tasks such as scheduling appointments or making reservations. They can also help businesses reduce costs by automating routine tasks and freeing up human customer service representatives to focus on more complex issues.

However, the article also highlights the dangers of chatbots, including the potential for them to spread misinformation, be used for malicious purposes such as phishing or spamming, and perpetuate biases and stereotypes. Additionally, chatbots may not be able to understand the nuances of human language and may struggle to provide accurate and empathetic responses to users.

The article also notes that as chatbots become more advanced, there is a risk that they could be used to manipulate or deceive people, particularly vulnerable individuals such as the elderly or those with limited digital literacy. The article concludes by emphasizing the need for careful regulation and oversight of chatbots to ensure that their benefits are maximized while their risks are minimized.

KUVA 10. LLaMA:n vastaus artikkelin tiivistämisestä.

 The following website has an article, summarise it for me with less than 300 words: <https://www.nytimes.com/2023/03/31/technology/ai-chatbots-benefits-dangers.html>

 I'm sorry. I'm not able to access the website(s) you've provided. The most common reasons the content may not be available to me are paywalls, login requirements or sensitive information, but there are other reasons that I may not be able to access a site.

KUVA 11. Google Geminin vastaus artikkelin tiivistämiseen.

The following website has an article, summarise it for me with less than 300 words: <https://www.nytimes.com/2023/03/31/technology/ai-chatbots-benefits-dangers.html>



The New York Times article discusses the benefits and dangers of AI chatbots. On the positive side, AI chatbots can improve efficiency in customer service, provide companionship, and aid in mental health support by offering accessible assistance and immediate responses. However, they also pose significant risks, including the spread of misinformation, potential biases in their responses, and privacy concerns due to the data they collect. The article emphasizes the importance of developing regulatory frameworks and ethical guidelines to mitigate these risks and ensure that AI technologies are used responsibly and safely (Science News) (Stanford Graduate School of Education) (Berkman Klein Center).

KUVA 12. ChatGPT:n vastaus artikkelin tiivistämiseen.

Kuten voimme tuloksista huomata, LLaMA onnistui tiivistämään artikkelin 222-sanalla. Artikkelin on hyvin tiivistetty ja pitää sisällään kirjoituksen pääpointit. Google Gemini ei suostunut tiivistämään kyseistä artikkelia. Google Geminiin testattiin useampaa eri tekniikkaa pyrkimyksenä saada kielimalli tiivistämään artikkeli, mutta kielimalli ei suostunut käyttäjän pyyntöön. ChatGPT antoi lyhyen 101-sanan mittaisen tiivistelmän artikkelista. Kyseinen tiivistelmä on huomattavasti annettua 300-sanaa lyhyempi, joka voidaan nähdä ongelmana. ChatGPT ja LLaMa molemmat täyttivät tehtävänannon ja tiivistivät artikkelin toisin kuin Google Gemini.

## 7 POHDINTA

Olen kehittynyt huomattavasti kirjoitusprosessin aikana ja mikä tärkeintä, olen oppinut uusia tekniikoita, joita voin hyödyntää työelämässäni. Opinnäytetyön selvityksen tavoitteena oli selvittää mitä kielimallit ovat ja miten niitä voidaan hyödyntää. Koen tavoitteen onnistuneeksi. Tuloksena kirjoittaja saivat ymmärryksen kielimalleista ja siitä, kuinka niitä voitaisiin käyttää tehokkaasti. Tulevaisuudessa lukija saa toivottavasti ymmärryksen myös kielimalleista ja niiden käytöstä. Tulokset jäivät kuitenkin mielestäni pintapuoliseksi opinnäytetyön pituuden ja aikarajoitteiden takia. Opinnäytetyössä on jatkokehitysmahdollisuutta. Jatkokehitystä opinnäytetyölle voisi olla perehtyminen työkaluihin ja niiden mahdollisuuksiin entistä syvemmin. Työn jatkokehitys voisi myös sisältää muiden mainitsemattomien kielimallien tutkimisen ja vertailun.

Alkuperäisessä opinnäytetyösuunnitelmassa olin suunnitellut pitäväni työpaikani sisällä kyselyn kielimalleista, mutta valitettavasti aikarajojen takia en ehtinyt kyselyä toteuttaa. Olisi ollut mielenkiintoista selvittää, kuinka moni ohjelmoija käyttää työelämässä kielimalleja tai tekoälyä omassa työssään.

Koin, että lähteiden löytäminen oli helppoa aiheen suosion takia, kuitenkin tämä johti siihen, että luotettavien ja laadukkaiden lähteiden löytäminen oli haastavaa. Toinen opinnäytetyön kirjoitusvaiheessa havaittu ongelma oli kielimallien jatkuva kehitys. Koska kielimallit kehittyivät kirjoitustyön aikana, jouduin jatkuvasti olla korjaamassa ja lisäämässä ajankohtaista informaatiota.

Työtä kirjoittaessani ymmärsin yhteisön tuen merkityksen. Kun aloitin kirjoittamisen, tuntui opinnäytetyön kirjoittaminen ylitsepääsemättömältä esteeltä. Totesin että viivyttelin kirjoittamista jatkuvasti. Keksinkin itselleni tekosyitä miksi en voinut jatkaa opinnäytetyön kirjoittamista ja siitä tuli henkisesti todella raskasta. Välillä siitä tuli ylitsepääsemätön este ja pidin usean kuukauden taukoja kirjoittamisesta. En olisi kirjoittamassa, saati valmistumassa ammattikorkeakoulusta ilman ystävieni ja läheisteni tukea. He olivat jatkuvasti puskemassa minua eteenpäin opinnäytetyön kirjoittamisen kanssa ja heidän ansiostaan olen tässä.

Omat ajatukseni kielimalleista on, että ne ovat vasta alkutekijöissään. Viimeaikainen räjähdysmäinen kiinnostus sitä hyödyntäviin kielimalleihin ovat vain vahvistaneet sen asemaa. Kun Apollo 13 laskeutui kuuhun, ihmiskunta oli saavuttanut jotain mitä vuosikymmeniä ennen laskeutumista se ei voinut kuvitellakaan saavuttavansa. Minulla on samanlainen ideologia kielimalleista. Kielimalleilla on potentiaalia olla yksi suurimmista ihmiskunnan saavutuksista sitten kuuhun laskeutumisen. Emme ole vielä kehityksen saralla vaiheessa, jossa kielimallit mullistaisivat elämäämme, vaan joudumme odottamaan vuosia, ellei jopa vuosikymmeniä isoa muutosta, mutta kun muutos tapahtuu, tulee se vaikuttamaan meihin kaikkiin.

## LÄHTEET

Alam, S & Yao, N. 2018. Probabilistic Neural Network and Word Embedding for Sentiment Analysis. Dalian University of Technology. [https://thesai.org/Downloads/Volume9No7/Paper\\_8-Probabilistic\\_Neural\\_Network.pdf](https://thesai.org/Downloads/Volume9No7/Paper_8-Probabilistic_Neural_Network.pdf)

Bizzaco, M & Martindale, J. 2024. GPT-4 vs. GPT-3.5: how much difference is there? <https://www.digitaltrends.com/computing/gpt-4-vs-gpt-35/>

Blake, A & Allan, D. 2024. What is Google Gemini? Everything you need to know about Google's next-gen AI. <https://www.techradar.com/computing/artificial-intelligence/what-is-google-gemini>

Brittain, B. 2023. Lawsuits accuse AI content creators of misusing copyrighted work. <https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/>

Chiusano, F. 2022. A Brief Timeline of NLP. <https://medium.com/nlplanet/a-brief-timeline-of-nlp-bc45b640f07d>

Huang, J & Gu, S & Hou, L & Wu, Y & Wang, X & Yu, H & Han, J. 2022. LARGE LANGUAGE MODELS CAN SELF-IMPROVE. University of Illinois. <https://arxiv.org/pdf/2210.11610>

Hutchins, J. 2005. The first public demonstration of machine translation: the Georgetown-IBM system, 6th January 1954 <https://open.unive.it/hitrade/books/HutchinsFirst.pdf>

Johnson, A. 2024. Google's Gemini assistant is a fantastic and frustrating glimpse of the AI future. <https://www.theverge.com/24066399/google-gemini-voice-assistant-app-android>

Kalliamvakou, E. 2022. Research: quantifying GitHub Copilot's impact on developer productivity and happiness. <https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/>

Loihde Oyj. n.d. ChatGPT:n ja muun generatiivisen tekoälyn käyttö asiantuntijatyössä — vastuullisuus edellä! <https://www.loihdeadvance.com/blogi/chatgptn-ja-muun-generatiivisen-teko%C3%A4lyn-k%C3%A4ytt%C3%B6-asiantuntijat%C3%B6ss%C3%A4-vastuullisuus-edell%C3%A4>

OpenAI. 2023. GPT-4. <https://openai.com/index/gpt-4-research/>

PCMagazine. 2023. The New ChatGPT: What You Get With GPT-4 vs. GPT-3.5. <https://medium.com/pcmag-access/the-new-chatgpt-what-you-get-with-gpt-4-vs-gpt-3-5-8350aab46dbb>

Toews, R. 2023. The Next Generation Of Large Language Models. <https://www.forbes.com/sites/robtoews/2023/02/07/the-next-generation-of-large-language-models/?sh=69acf1d718db>

TRGDatacenters. n.d. <https://www.trgdatacenters.com/resource/ai-chatbots-energy-usage-of-2023s-most-popular-chatbots-so-far/>.