



Juska Munne

Suurten kielimallien hyödyntäminen potilastietojärjestelmässä

Metropolia Ammattikorkeakoulu

Insinööri (AMK)

Tieto- ja viestintäteknikka

Insinöörityö

4.10.2025

Tiivistelmä

Tekijä:	Juska Munne
Otsikko:	Suurten kielimallien hyödyntäminen potilastietojärjestelmässä
Sivumäärä:	34 sivua + 1 liitettä
Aika:	4.10.2025
Tutkinto:	Insinööri (AMK)
Tutkinto-ohjelma:	Tieto- ja viestintätekniikka
Ammatillinen pääaine:	Hyvinvointi- ja terveysteknologia
Ohjaaja:	Yliopettaja Mikael Soini

Insinööriyön tavoitteena oli selvittää, kuinka suurten kielimallien avulla voidaan helpottaa potilastietojärjestelmän käyttöä. Työ toteutettiin narratiivisena kirjallisuuskatsauksena, jossa analysoitiin 12 ajankohtaista, vertaisarvioitua tutkimusta PubMed-tietokannasta. Kirjallisuuskatsauksessa tutkimukset jaoteltiin neljään pääteemaan, jotka ovat ennaltaehkäisy, potilasviestintä, yhteenvedot ja tietojen poiminta vapaasti kirjatusta tekstistä. Tutkimusten keskeiset tulokset esitellään teemoittain ja niistä muodostetaan synteesi. Työn teoriaosuudessa käsitellään suurten kielimallien toimintaperiaatetta, Euroopan unionin tekoälyasetusta ja potilastietojärjestelmien ominaisuuksia.

Kirjallisuuskatsauksen tulokset osoittivat suurten kielimallien olevan tietyissä tehtävissä jopa terveydenhuollon ammattilaisia suorituskykyisempiä. Tulosten yleistettävyyttä vaatii kuitenkin lisää tutkimusta aidossa käyttöympäristössä. Keskeiseksi haasteiksi tunnistettiin kielimallin luotettavuus ja riski arkaluontoisten potilastietojen vuotamisesta väärille tahoille. Haasteista huolimatta työn johtopäätöksenä on, että suurilla kielimalleilla on merkittävä potentiaali tehostaa tulevaisuuden potilastietojärjestelmiä.

Avainsanat: suuri kielimalli, potilastietojärjestelmä, narratiivinen kirjallisuuskatsaus

Olen käyttänyt Google Gemini versiota 2.5 Flash insinööriyön jäsentelyyn ja tiedonhakuun. Olen myös käyttänyt samaa ohjelmaa tekstini kieliasun viimeistelyyn. Opin näytetyön tekijänä olen vastuussa kaikesta insinööriyöni sisällöstä.

Tämän insinööriäytetyön alkuperä on tarkastettu Turnitin Originality Check -ohjelmalla.

Abstract

Author: Juska Munne
Title: Utilization of Large Language Models in Patient Information System
Number of Pages: 34 pages + 1 appendices
Date: 4 October 2025

Degree: Bachelor of Engineering
Degree Programme: Information and Communications Technology
Professional Major: Health technology
Supervisor: Mikael Soini, Principal Lecturer

The objective of the study was to investigate how large language models can facilitate the use of electronic health record systems. The study was conducted as a narrative literature review, analysing 12 current, peer-reviewed studies from the PubMed database. The studies were categorized into four main themes: prevention, patient communication, summaries, and information extraction from free-text entries. The key findings of the studies are presented by theme and synthesized. The theoretical part of the thesis discusses the operating principles of large language models, the European Union's AI Act, and the features of electronic health record systems.

The results of the literature review indicated that in certain tasks, large language models can even outperform healthcare professionals. However, the generalizability of the findings requires further research in real-world clinical settings. The key challenges identified were the reliability of the language models and the risk of sensitive patient data being leaked to unauthorized parties. Despite the challenges, the conclusion is that large language models have significant potential to enhance future electronic health record systems.

Keywords: Large Language Model, Electronic Health Record, Narrative Literature Review

Sisällys

1	Johdanto	1
2	Suuret kielimallit	2
2.1	Kehitys ja käyttökohteet	2
2.2	Koulutus	3
2.3	Transformer-arkkitehtuuri	4
2.4	Haasteet	6
3	EU:n tekoälyasetus	7
3.1	Tekoälyn sääntely Euroopan unionissa	7
3.2	Riskiperusteinen lähestymistapa	8
3.3	Yleinen tekoälymalli	11
3.4	Seuraamukset ja käyttöönotto	12
4	Potilastietojärjestelmä	13
4.1	Potilastietojärjestelmän määritelmä ja merkitys	13
4.2	Ominaisuudet ja haasteet	14
4.3	Lainsäädäntö ja asetukset	14
5	Toteutus	15
5.1	Narratiivinen kirjallisuuskatsaus menetelmänä	15
5.2	Tarkoitus ja tavoitteet	16
5.3	Aineiston valinta	16
6	Tulokset	18
6.1	Ennaltaehkäisy	18
6.2	Potilasviestintä	21
6.3	Tietojen poiminta vapaasti kirjatusta tekstistä	22
6.4	Yhteenvedot	25
7	Yhteenveto	27
7.1	Keskeiset huomiot	27
7.2	Eettisyys ja luotettavuus	29
7.3	Jatkotutkimusehdotukset	29

Lähteet

31

Liitteet

Liite 1: Artikkelien yhteenveto

1 Johdanto

Terveystieteiden ammattilaiset käyttävät merkittävän osan työajastaan potilastietojen kirjaamiseen. Samaan aikaan alaa kuormittavat työvoimaresurssien niukkuus, talouden sopeuttamistarve ja alati kasvava hoidontarve väestönrakenteen ikääntyessä. Tekoäly voi tehostaa potilastietojen kirjaamista ja samalla parantaa sen laatua. Kun tekoäly auttaa potilastietojen kirjaamisessa, voivat terveydenhuollon ammattilaiset keskittyä potilaan hoitoon. (Sitra 2025.)

Suuret kielimallit ovat yksi tekoälyn muodoista. Ne kykenevät tuottamaan ihmisestä tekstiä ja jossain määrin ymmärtämään sitä. Niillä voi yleisesti ottaen tehdä kaikenlaista, mikä liittyy tekstin ymmärtämiseen, tuottamiseen ja muokkaamiseen. Kielimallit auttavat ihmisiä säästämään aikaa suorittamalla aikaa vieviä rutiinitehtäviä, jolloin ihmisille jää aikaa luovempaan työhön.

Tämä insinööriyö tutkii, kuinka suuret kielimallit voivat helpottaa potilastietojärjestelmän käyttöä. Narratiivisen kirjallisuuskatsauksen avulla luodaan kokoava yhteenveto viimeaikaisista tutkimuksista, tunnistetaan tutkimusten aukkoja ja tiivistetään tulokset yhtenäiseksi kokonaisuudeksi.

Insinööriyön toinen luku käsittelee suurten kielimallien kehitystä ja arkkitehtuuria. Kolmannessa luvussa perehdytään uuteen Euroopan unionin tekoälyasetukseen, jonka tarkoitus on varmistaa markkinoille tuotavien tekoälyjärjestelmien turvallisuus. Neljännessä luvussa käydään läpi potilastietojärjestelmien ominaisuuksia, haasteita ja lainsäädäntöä. Viidennessä luvussa kerrotaan, kuinka insinööriyön narratiivinen kirjallisuuskatsaus toteutettiin, ja kuudennessa luvussa käydään läpi katsauksen tulokset. Viimeisessä luvussa vedetään yhteen keskeiset huomiot, arvioidaan tehdyn kirjallisuuskatsauksen luotettavuutta ja esitetään jatkotutkimuskohteita.

2 Suuret kielimallit

Tässä luvussa käsitellään suurten kielimallien kehitystä, arkkitehtuuria ja haasteita. Nämä ovat tärkeitä asioita ymmärtää, jotta kielimalleja voidaan soveltaa tehokkaasti ja turvallisesti.

2.1 Kehitys ja käyttökohteet

Kielimallit eivät ole uusi keksintö, vaan ne ovat kehittyneet ajan saatossa. Ennen suuria kielimalleja tuli esikoulutetut kielimallit (engl. pre-trained language models). Suurimpana erona näiden kahden välillä on niiden mittakaava. Suurentamalla mallien käyttämän datan määrää saadaan malli suoriutumaan sille annetuista tehtävistä paremmin. Esikoulutetut kielimallit pääosin tuottivat tekstiä, mutta suuret kielimallit suoriutuvat monimutkaisista tehtävistä. (Zhao ym. 2025: 2–3.)

Yang ym. (2023) suosittelee suurten kielimallien käyttöä esimerkiksi kysymyksiin vastaamiseen, chatbotiksi, laadun arviointiin, yhteenvetojen luomiseen, kielen kääntämiseen, sähköpostiluonnoksiin, uutisartikkelien tekoon, tarinoiden keksimiseen ja koodaamiseen. Sen sijaan tekstin luokitteluun, nimien tai toksisuuden tunnistamiseen, tiedonhakuun, luetun ymmärtämiseen ja lauseiden samankaltaisuuden arviointiin he pitävät hienosäädettyjä pienempiä kielimalleja parempana vaihtoehtona.

Kaplan ym. (2020: 3) tutkimustulokset osoittivat kielimallin kyvykkyyden korreloivan vahvasti mallin parametrien määrän, koulutusaineiston koon ja koulutuksessa käytetyn laskentatehon kanssa. Tulosten mukaan kielimallin kyvykkyyden paraneminen voidaan ennustaa tarkasti, kun parametrien määrää ja koulutusaineiston kokoa kasvatetaan optimaalisessa suhteessa. Lisäksi tutkimus osoitti, että samalla laskentateholla on kustannustehokkaampaa kouluttaa mallia, jossa on enemmän parametreja, kuin mallia, jossa niitä on vähemmän.

2.2 Koulutus

Kielimallin oppiminen voidaan jakaa esikoulutukseen ja hienosäätöön. Esikoulutus on tärkeä vaihe, jossa kielimalli oppii ymmärtämään ja tuottamaan kieltä. Esikoulutuksen jälkeen kielimallia hienosäädetään (engl. fine-tuning) kohdennettua käyttötarkoitusta varten. Hienosäätöä voidaan tehdä erilaisin menetelmin, kuten siirto-oppimalla (engl. transfer learning) tai kontekstista oppimalla (engl. in-context learning). Kielimallia voidaan käyttää ilman hienosäätöä, mutta silloin malli toimii yleisellä tasolla eikä osaa vastata tarkempiin asiayhteydestä riippuviin kysymyksiin. (Yang ym. 2023: 5–6.)

Esikoulutusvaiheessa kielimalli käy läpi valtavan määrän aineistoa, jonka tulee olla laadukasta ja monipuolista. Näin kielimalli oppii kielen rakenteet, sanojen väliset riippuvuudet ja kykenee tuottamaan tekstiä. Käytetty aineisto rajaa mallin käyttömahdollisuuksia, sillä jos koulutusaineistossa ei ole tietoa aiheesta, malli ei kykene vastaamaan sitä koskeviin kysymyksiin. Esikoulutus tapahtuu tyypillisesti itseohjautuvasti ilman ihmisen ohjausta. (Yang ym. 2023: 5; Zhao ym. 2025: 16.)

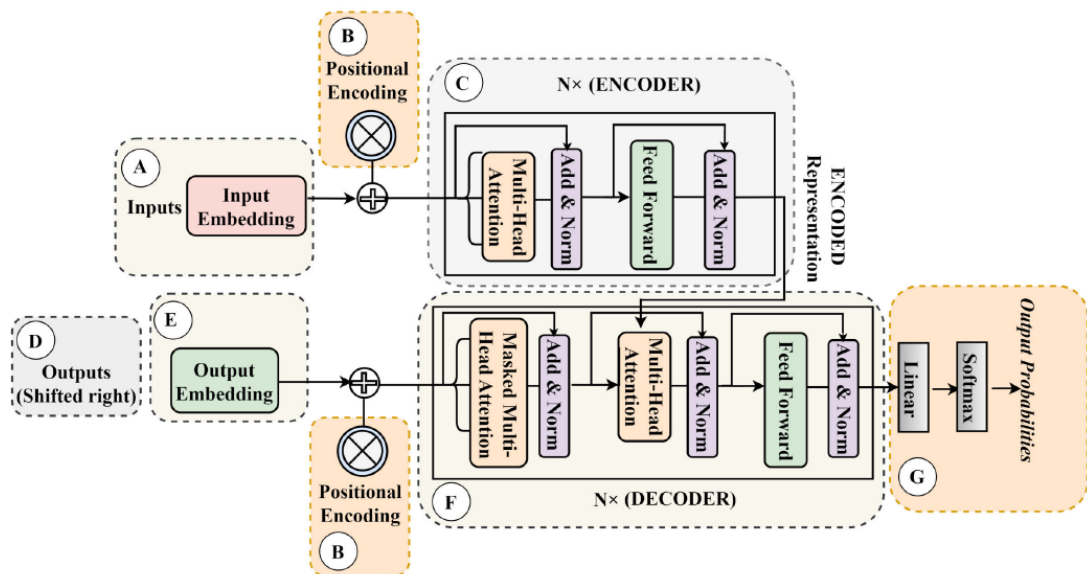
Siirto-oppimisessa hyödynnetään kielimallin esikoulutuksessa oppimaa yleistä tietoa, ja mallia täydennetään tehtäväkohtaisella aineistolla, jotta se suoriutuu paremmin juuri kyseisestä tehtävästä. Menetelmä on kustannustehokkaampi kuin mallin kouluttaminen alusta asti. Toisin kuin esikoulutus, siirto-oppiminen on ohjattua oppimista, jossa muokataan esikoulutuksen aikana opittuja parametrien painokertoimia. Näin mallia opetetaan ratkaisemaan tietty tehtävä entistä tehokkaammin. Painokertoimien muuttaminen voi kuitenkin aiheuttaa ongelmia tehtävissä, jotka muistuttavat läheisesti esikoulutuksessa käytettyjä aineistoja. (Patil & Gudivada 2024: 18.)

Kontekstista oppiminen on kevyempi menetelmä kuin siirto-oppiminen, koska mallin parametreihin ei kosketa ja se pärjää vähäisemmällä aineistolla. Kun kielimallin parametreja ei tarvitse muokata tehtäväperusteisesti, pysyy se yleiskäyttöisenä. Kontekstista oppimisessa kielimalli hyödyntää sille annettua syötettä,

jonka perusteella se muokkaa toimintaansa. Kontekstista oppiminen toimii parhaiten, kun syötteessä annetaan esimerkkejä siitä, millaisia vastauksia kielimallin tulisi antaa. Koska kontekstista oppiminen ei muuta kielimallin parametreja, kielimalli ei muista aiemmin oppimaansa, vaan esimerkit pitää antaa sille aina uudestaan. Kontekstista oppiminen onnistuu vain yli 100 miljoonan parametrin malleilla, ja sen teho kasvaa mallin koon myötä. (Patil & Gudivada 2024: 23–27.)

2.3 Transformer-arkkitehtuuri

Vuonna 2017 kehitettiin Transformer-arkkitehtuuri, jonka pohjalta on rakennettu nykyiset kielimallit, kuten ChatGPT. Sen keskeinen piirre on itsehuomiomekanismi, jonka avulla malli tunnistaa sanojen välisiä suhteita ja rakentaa niiden perusteella johdonmukaisia lauseita. Transformer-arkkitehtuuri erottuu muista arkkitehtuureista erityisesti kyvyllään käsitellä sanojen riippuvuuksia koko lauseen tasolla sekä paremmalla suorituskyvyllään. Sitä pidetäänkin yhtenä tekoälyn merkittävimmistä läpimurroista. (Ferrer 2024.) Seuraavassa esitellään Transformer-arkkitehtuurin keskeiset osat (kuva 1).



Kuva 1 Enkooderi-dekooderi Transformer-arkkitehtuurin rakenne (Raiaan ym. 2024: 11).

Syöte (A)

Kielimallit eivät käsittele sille syötettyä tekstiä sellaisenaan, vaan ne käsittelevät *tokeneita*. Tokenit ovat sanoja tai sananosia, jotka on muutettu numeeriseen muotoon. Jokainen tokeni saa uniikin numeroarvon, ja ne sijoitetaan kielimallin sisäiseen sanakirjaan lähelle muita samankaltaisia sanoja. Tokenien järjestely auttaa kielimallia tunnistamaan sanojen keskinäisiä riippuvuuksia. (Raiaan ym. 2024: 10.)

Asemakoodaus (B)

Kielimalli ei käsittele sille annettua syötettä siinä järjestyksessä, missä se sille syötettiin, joten se ei luonnostaan ymmärrä, missä järjestyksessä tokenit syötteessä ovat. Kielimallin kuitenkin on tärkeä ymmärtää sanojen järjestys, sillä sille annetun syötteen merkitys voi muuttua aivan toiseksi, jos sanojen järjestys vaihtuu. Ratkaisuna tähän ongelmaan Transformer-arkkitehtuurissa on asemakoodaus, jossa jokaiselle tokenille lisätään sijaintitieto. (Vaswani ym. 2017: 6.)

Enkooderi (C)

Enkooderin tehtävänä on muuttaa kielimallille annettu syöte numeeriseen esitysmuotoon, joka sisältää kontekstin. Kontekstin lisääminen on tärkeää, koska sama sana voi tarkoittaa eri asiaa riippuen, missä yhteydessä sitä on käytetty. Enkooderissa on kuusi kerrosta, joiden läpi syöte kulkee kerros kerrokselta. Jokainen kerros syventää edellisen kerroksen ymmärrystä syötteestä. Tämän avulla kielimalli ymmärtää, miten syötteen sanat liittyvät toisiinsa. Prosessia kutsutaan huomiomekanismiksi, joka tekee Transformer-arkkitehtuurista ainutlaatuisen. (Raiaan ym. 2024: 11; Ferrer 2024.)

Dekooderi (D-G)

Dekooderi sisältää samoja toimintoja kuin enkooderi, mutta sitä käytetään kuitenkin eri tarkoitukseen. Dekooderin tehtävänä on tuottaa ulostulo eli käyttäjälle näytettävä lause tulkitsemalla enkooderin luomaa numeerista esitysmuotoa.

Ulostulon dekooderi tuottaa sana kerrallaan hyödyntämällä aiemmin ennustamiin sanoja. Jokaiselle mahdolliselle seuraavalle sanalle lasketaan todennäköisyys, ja suurimman todennäköisyyden saanut sana valitaan. Tätä prosessia toistetaan, kunnes koko ulostulo on valmis. (Raiaan ym. 2024: 11–12; Ferrer 2024.)

2.4 Haasteet

Yksi suurten kielimallien merkittävimmistä ongelmista on hallusinointi, jolla tarkoitetaan kielimallin taipumusta tuottaa tekstiä, joka vaikuttaa uskottavalta, mutta on virheellistä ja ristiriidassa lähdetietojen kanssa. Sovelluksissa, joissa virheisiin ei ole varaa, kielimallin hallusinointi aiheuttaa merkittäviä riskejä. Koulutusaineiston laatu on vahvasti kytköksissä siihen, kuinka paljon kielimalli hallusinoi. Jos koulutusaineisto sisältää saman virheen useita kertoja, malli painaa muistiin kyseisen virheen ja käyttää tätä virheellistä tietoa antaessaan vastauksia. Toisin kuin jatkuvasti muuttuva maailma, kielimallin tieto on staattista ja siksi se vanhenee ajan myötä. Tämän vuoksi malli voi antaa vanhentuneita tai virheellisiä vastauksia, jos se ei ymmärrä kysymystä oikein. Myös liian suppea koulutusaineisto suhteessa kysymykseen voi johtaa siihen, että malli keksii vastauksia. (Zhao ym. 2025: 60, 90–91.)

Suurella kielimallilla on paljon potentiaalia auttaa ihmistä erilaisissa tehtävissä, mutta sen käyttöönotossa on huomioitava kustannustehokkuus. Suurempi malli suoriutuu sille annetuista tehtävistä paremmin, mutta sen koulutus ja käyttö maksavat enemmän. Esimerkiksi 11 miljardia parametria sisältävän T5-kielimallin koulutus kustansi 1,3 miljoonaa dollaria, kun taas 175 miljardia parametria sisältävän GPT-3-kielimallin koulutus kustansi 4,6 miljoonaa dollaria. Mallin koon kasvu lisää myös vastausten tuottamisen viivettä. Suuria kielimalleja joudutaan usein ajamaan useamman tietokoneen voimin ohjelmointirajapinnan kautta, mikä lisää viivettä entisestään. (Yang ym. 2023: 15.)

Kielimallin eettisyydestä on huolehdittava koko sen elinkaaren ajan, aina suunnittelusta ylläpitoon. Suunnitteluvaiheessa on otettava mukaan eri sidosryhmiä,

kuten omistajat ja loppukäyttäjät, jotta ymmärretään heidän tarpeensa ja arvonsa. Kielimallin käyttämisen aineiston keruussa on huomioitava suostumukset, tekijänoikeudet ja aineiston laatu, sillä yleisesti saatavilla oleva aineisto ei välttämättä ole eettisesti kestävä. Kielimallin vinoumat voivat johtua sekä aineistosta että mallin rakenteesta. Pienillä muutoksilla mallin rakenteeseen voidaan vaikuttaa suuresti mallin toimivuuteen. On myös epäeettistä arvioida kielimallin toimintaa vain teknisillä mittareilla ilman ihmisen arviota, sillä ne eivät välttämättä kerro, kuinka malli pärjää oikeassa käyttöympäristössä. (Ungless ym. 2024.)

3 EU:n tekoälyasetus

Tekoälyjärjestelmien markkinoille tuojien tulee olla tietoisia Euroopan unionin tekoälyasetuksesta ja sen asettamista vaatimuksista. Vaatimusten huomiotta jättäminen voi johtaa tuntuviin sakkoihin. Tässä luvussa käsitellään tekoälyasetuksen taustat, riskiperusteinen lähestymistapa ja mahdolliset seuraamukset asetuksen rikkomisesta.

3.1 Tekoälyn sääntely Euroopan unionissa

Vuonna 2021 Euroopan komissio antoi ehdotuksensa EU:n yhteisestä tekoälyasetuksesta (EU 2024/1689), joka hyväksyttiin kolmen vuoden käsittelyn jälkeen vuonna 2024. Tekoälyasetuksen tavoitteena on luoda viitekehys eettisesti ja luotettavasti toimivien tekoälyratkaisujen kehittämistä varten. Samalla sen tarkoitus on edistää innovointia ja helpottaa tekoälyratkaisujen liikkumista EU:n jäsenvaltioiden välillä yhteisten sääntöjen avulla. Asetus perustuu riskilähtöiseen tarkasteluun, jossa tekoälyyn perustuvan tuotteen riskitaso tulee arvioida ja saattaa kyseisen tason vaatimusten mukaiseksi. (Euroopan parlamentti 2025.)

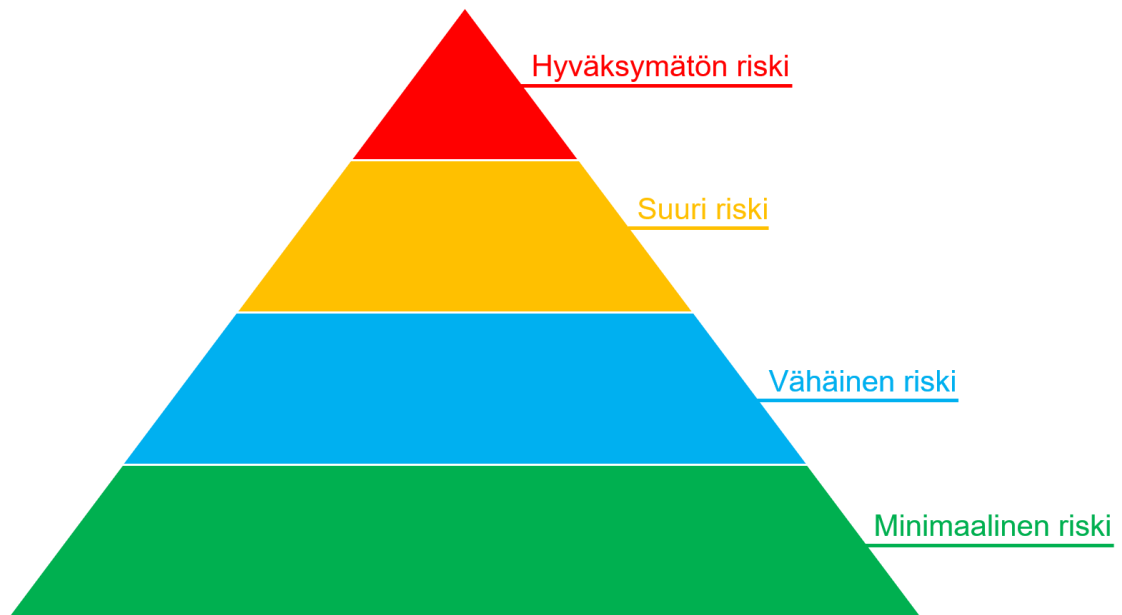
Asetuksen luonnin yhteydessä Euroopan parlamentti tuotti tekoälylle määritelmän, jotta asetusta osataan soveltaa asianmukaisesti. Asetus määrittää tekoälyjärjestelmän olevan konepohjainen järjestelmä, joka osaa päätellä asioita itsenäisesti. Kone osaa siis itsenäisesti tuottaa ennusteita, sisältöä, suosituksia

tai päätöksiä. Ihmisen tuottamaa sääntöpohjaista järjestelmää ei katsota tekoälyjärjestelmäksi. Järjestelmää pidetään tekoälynä, jos se pystyy itsenäisesti oppimaan ja mukautumaan tehtäviinsä sen jälkeen, kun se on otettu käyttöön. (Euroopan komissio 2025a.)

Tekoälyasetuksen käyttöönoton yhteydessä 1. elokuuta 2024 perustettiin Euroopan tekoälyneuvosto, joka vastaa tekoälyasetuksen toimeenpanosta ja valvonnasta. Neuvostossa on yksi edustaja jokaisesta jäsenvaltiosta, ja edustajien on täytettävä tietyt pätevyyskriteerit. Neuvosto jakautuu useampaan alaryhmään, jotka kukin toteuttaa alaryhmälle annettua erikoistehtävää. Neuvosto varmistaa tekoälyasetuksen yhtenäisen toimeenpanon jäsenvaltioiden kesken antamalla helposti ymmärrettävää ohjeistusta ja neuvontaa. Lisäksi neuvosto toimii kansainvälisenä yhteistyöelimenä vahvistaakseen EU:n kilpailukykyä. (Euroopan parlamentti ja neuvosto 2024: 65 artikla.)

3.2 Riskiperusteinen lähestymistapa

Kuvassa 2 esitetään tekoälyasetuksen riskiperusteinen lähestymistapa, jossa on neljä tasoa. Nämä neljä tasoa ovat hyväksymätön riski, suuri riski, vähäinen riski ja minimaalinen riski. Säännöt tiukkenevat, mitä suurempi riskitaso säädeltävällä tekoälyllä on.



Kuva 2 Riskiperusteinen lähestymistapa (Euroopan komissio 2025b).

Euroopan parlamentti ja neuvosto (2024: 5 artikla) määrittävät hyväksymättömiksi riskeiksi kahdeksan tekoälykäytäntöä. Tekoälyjärjestelmää ei saa käyttää seuraavissa tapauksissa:

- **Harhaanjohtavat tekniikat**, joissa tekoälyjärjestelmä saa henkilön tekemään haitallisia asioita, joita hän ei olisi ilman tekoälyn vaikutusta tehnyt.
- **Haavoittuvuuksien hyödyntäminen**, joka vaikuttaa henkilön käyttäytymiseen, hyödyntämällä hänen haavoittuvuuksiaan, kuten ikää, vammaa, sosiaalista tai taloudellista tilannetta.
- **Sosiaalinen pisteytys**, joka asettaa henkilöitä eriarvoiseen asemaan käyttäytymisen tai ominaisuuksien perusteella.
- **Rikollisen käyttäytymisen ennustaminen** profiloinnin tai persoonallisuusominaisuuksien perusteella. Kieltoa ei sovelleta, jos tekoälysovelluksen arvio perustuu objektiivisiin tosiseikkoihin.
- **Kasvojen tunnistustietokannan** luonti, jossa aineisto kerätään internetistä tai valvontakamerakuvista.
- **Tunteiden päättely biometrinen tietojen perusteella** työpaikoilla tai oppilaitoksissa, ellei tekoälyjärjestelmää ole tarkoitus hyödyntää lääketieteellisiin tai turvallisuuteen liittyviin tarkoituksiin.
- **Biometrinen luokittelu**, jossa henkilöt luokitellaan rodun, poliittisten mielipiteiden, ammattiliiton jäsenyyden, uskonnollisen

vakaumuksen, filosofisen vakaumuksen, seksuaalisen käyttäytymisen tai seksuaalisen suuntautumisen perusteella. Kieltoa ei sovelleta, kun kyse on lainvalvonnasta.

- **Reaaliaikainen biometrinen tunnistaminen julkisissa tiloissa lainvalvontatarkoituksessa**, ellei kyseessä ole henkilöiden etsintä, henkilöiden turvallisuuteen kohdistuva välitön uhka tai vakavasta rikoksesta epäillyn henkilön paikantaminen.

Suuririskisen tekoälyjärjestelmän saa ottaa käyttöön vain, jos se täyttää kaikille asetetut vaatimukset. Vaatimusten tavoitteena on huolehtia, ettei tekoälyjärjestelmät aiheuta liiallisia riskejä Euroopan unioniin kuuluville henkilöille. Suuririskisiksi järjestelmiksi luetaan tekoälyjärjestelmät, joilla on haitallinen vaikutus ihmisen terveyteen, turvallisuuteen ja perusoikeuksiin. Tekoälyneuvosto tulee tarkentamaan suuririskisten tekoälyjärjestelmien käyttöönoton ohjeistusta ja antamaan niihin käytännön esimerkkejä viimeistään 2.2.2026. Suuririskiset tekoälyjärjestelmät voidaan luokitella kahteen kategoriaan:

1. EU:n tuoteturvallisuusedirektiiviin 2001/95/EY kuuluvat tekoälyjärjestelmät, jotka löytyvät tekoälyasetuksen liitteestä I. Näihin kuuluvat esimerkiksi lääkinnälliset laitteet, rautatiejärjestelmät ja henkilöstösuojaimet.
2. Tekoälyasetuksen liitteessä III luetellut suuririskiset tekoälyjärjestelmät, jotka avataan seuraavassa kappaleessa. (Euroopan parlamentti ja neuvosto 2024: 6 artikla.)

Tekoälyasetuksen liite III sisältää kahdeksan alaa, joilla tekoälyjärjestelmät luokitellaan suuririskisiksi. Artiklan 7 mukaan komissio voi muokata ja lisätä liitteessä III mainittuja aloja. Tällä hetkellä liitteessä III mainitaan seuraavat suuririskiset alat:

- **Biometriset tunnisteet**, kun niitä käytetään etätunnistusjärjestelmissä, henkilön luokitteluun ominaisuuksien perusteella tai tunteiden tunnistamiseen. Biometrisia tunnisteita voidaan käyttää henkilöllisyyden todentamiseen.
- **Kriittinen infrastruktuuri**, kun kyseessä on turvakomponentti, jota käytetään kriittisen digitaalisen infrastruktuurin, tieliikenteen tai vesi-, kaasu-, lämmitys- tai sähköhuollon osana.
- **Yleissivistävä ja ammatillinen koulutus**. Sisältää koulutukseen hyväksymisen, oppimisen ja koulutustason arvioinnin sekä opiskelijoiden käyttäytymisen tarkkailun.

- **Työllistäminen, henkilöstöhallinta ja ammatinharjoittaminen**, kun kyseessä on rekrytointi, työsuhteen ehdot, työnjako tai suorituksen seuranta.
- **Välttämättömien palvelujen ja etuuksien saatavuus ja käyttö**. Viranomaisten käyttämät tekoälyjärjestelmät etuuksien ja palvelun tarpeen arviointiin, luottokelpoisuuden arviointiin, vakuutusten hinnoitteluun tai pelastus- ja ensihoitopalvelujen tärkeysjärjestyksen määrittelyyn.
- **Lainvalvonta**, kun tekoälyjärjestelmää käytetään rikoksen uhriksi joutumisen arviointiin, valheenpaljastimena, todistusaineiston luotettavuuden arviointiin, rikosten ennustamiseen tai henkilön profilointiin rikosten selvittämistä varten.
- **Muuttoliikkeen hallinta ja rajavalvonta**, kun tekoälyjärjestelmää käytetään valheenpaljastimena, maahanmuuton valvomiseen, turvapaikka tai viisumi kelpoisuuden arviointiin.
- **Oikeudenhoido ja demokraattiset prosessit**, kun oikeusviranomaisen käyttää tekoälyjärjestelmää lainsäädännön tulkintaan tai tekoälyjärjestelmä vaikuttaa henkilöiden äänestyskäyttäytymiseen. (Euroopan parlamentti ja neuvosto 2024: Liite III.)

Vähäisen riskin tekoälyjärjestelmiltä edellytetään ainoastaan läpinäkyvyyttä, eli käyttäjälle on kerrottava selkeästi, että hän on vuorovaikutuksessa tekoälyyn perustuvan järjestelmän kanssa. Chatbot on esimerkki tällaisesta tekoälyjärjestelmästä. Minimaalisen riskin tekoälyjärjestelmiin ei sen sijaan kohdistu lainkaan vaatimuksia, ja suurin osa tekoälyjärjestelmistä kuuluu tähän ryhmään. Tällaisia järjestelmiä ovat esimerkiksi videopelit ja roskapostisuodattimet. (Euroopan komissio 2025b.)

3.3 Yleinen tekoälymalli

Yleisellä tekoälymallilla tarkoitetaan laajasti käytettyä mallia, joka pystyy suorittamaan useita erilaisia tehtäviä ja jonka voi liittää osaksi muita järjestelmiä. Ne koulutetaan valtavilla määrillä dataa ja niissä on oltava vähintään miljardi parametria. Esimerkkejä tällaisista malleista ovat suuret kielimallit, kuten GPT-4, Gemini ja LLaMA. Yleiskäyttöisiä tekoälymalleja koskevat erityiset säännöt edellyttävät, että tarjoajien on ylläpidettävä mallin teknistä dokumentaatiota ja kerrottava, millaisella aineistolla malli on koulutettu. Jos malliin liittyy systeminen riski (mallilla on merkittävä suorituskyky), tarjoajien on lisäksi toteutettava

erityistä riskienhallintaa, raportoitava vaaratilanteista viranomaisille ja huolehdittava mallin kyberturvallisuudesta. (Euroopan parlamentti ja neuvosto 2024: 53 artikla, johdanto-osan kappaleet 97–98, 55 artikla.)

3.4 Seuraamukset ja käyttöönotto

Kiellettyihin tekoälykäytäntöihin liittyvien sääntöjen rikkomisesta voivat suuret yritykset saada sakon, joka on korkeintaan 35 miljoonaa euroa tai 7 prosenttia vuotuisesta liikevaihdosta, riippuen kumpi summista on suurempi. Suuririskisiä tekoälyjärjestelmiä koskevien sääntöjen rikkomisesta puolestaan voi saada sakon, joka on korkeintaan 15 miljoonaa euroa tai 3 prosenttia vuotuisesta liikevaihdosta, riippuen kumpi summista on suurempi. Viranomaisille annetusta virheellisestä, puutteellisesta tai harhaanjohtavasta tiedosta voi seurata sakko, jonka suuruus on korkeintaan 7,5 miljoonaa euroa tai 1 prosentti vuotuisesta liikevaihdosta, riippuen kumpi summista on suurempi. Pienillä ja keskisuurilla yrityksillä sakko määräytyy sen mukaan, kumpi summista on pienempi. (Euroopan parlamentti ja neuvosto 2024: 99 artikla.)

Tekoälyasetus tuli voimaan elokuussa 2024, mutta täysimääräisesti sitä sovelletaan kaksi vuotta myöhemmin elokuussa 2026. Käyttöönotto tapahtuu asteittain. Helmikuussa 2025 tuli voimaan kiellettyjä tekoälykäytänteitä koskeva sääntely, jolla pyritään estämään haitallisten tekoälyjärjestelmien käyttöönotto. Samaan aikaan alettiin soveltaa tekoälylukutaitoa koskevia velvoitteita, joiden tarkoituksena varmistaa tekoälyjärjestelmien parissa työskentelevien henkilöiden tietämys tekoälyn mahdollisuuksista ja riskeistä. Yleiskäyttöisiä tekoälymalleja koskeva säädäntö tuli voimaan elokuussa 2025. Säänneltyihin tuotteisiin sisällytetyt tekoälyjärjestelmiä koskeva sääntely tulee voimaan myöhemmin, elokuussa 2027. Tällaisia tuotteita ovat esimerkiksi lääkinnälliset laitteet, jotka kuuluvat jo ennestään EU-lainsäädännön piiriin. (Euroopan komissio 2025b, Euroopan parlamentti ja neuvosto 2024: johdanto-osan kappale 20.)

Tekoälyasetuksen täysimittainen käyttöönotto vaatii kansallisen tekoälyasetusta täydentävän lainsäädännön. Hallitus antoi esityksen täydentävästä

lainsäädännöstä eduskunnalle 8.5.2025, mutta sen käsittely on yhä kesken. Ehdotetussa laissa säädetään esimerkiksi kansalliset valvontaviranomaiset ja seuraamukset asetuksen rikkomisesta. Asetuksen käsittelyn on tarkoitus valmistua syksyllä 2025. (Valtioneuvosto 2025.)

4 Potilastietojärjestelmä

Tämä luku käsittelee potilastietojärjestelmän tarkoitusta, keskeisiä toimintoja ja lainsäädäntöä.

4.1 Potilastietojärjestelmän määritelmä ja merkitys

Potilastietojärjestelmä on ohjelmisto, johon terveydenhuollon ammattilaiset kirjaavat potilastietoa, kuten diagnooseja ja hoitosuunnitelmia. Sen tarkoituksena on tukea ammattilaisia heidän päivittäisessä työssään, parantaa potilasturvallisuutta ja sujuvoittaa hoitoprosessia. Potilastietojärjestelmä kokoaa potilaan tiedot yhteen paikkaan, jolloin potilaan kokonaiskuvaa on helpompi seurata. Potilaat pysyvät paremmin ajan tasalla omasta hoidostaan, sillä Suomessa luokan A potilastietojärjestelmät lähettävät kirjaukset Kanta-palveluun. Kanta-palveluun kirjatut tiedot ovat nähtävissä potilaalle OmaKannan kautta. (Vitec Acute 2025.)

Terveydenhuollon ammattilaisia veloitetaan kirjaamaan potilaan hoidon kannalta olennaiset tiedot viivytyksettä potilastietojärjestelmään. Näitä tietoja ovat esimerkiksi potilaskertomus, kuolemansyyn selvitys ja muut hoitoasiakirjat. Tiedot jaotellaan eri näkymille, jotta samaan aiheeseen liittyvät asiat löytyvät yhdestä paikasta. Tässä muodossa ne myös tallennetaan Kanta-palveluihin, josta muut luokan A potilastietojärjestelmät pystyvät noutamaan nämä rakenteiset tiedot omaan järjestelmäänsä. Rakenteinen kirjaaminen tarkoittaa, että tiedot tallennetaan sovitulla tavalla, mikä varmistaa yhtenäisen laadun ja helpottaa tiedon hyödyntämistä esimerkiksi yhteenvetojen tekemisessä. (Terveyden ja hyvinvoinnin laitos 2024a.)

4.2 Ominaisuudet ja haasteet

Potilastietojärjestelmien ominaisuudet vaihtelevat toimittajan mukaan, eikä kaikilla ole yhtä kattavia toimintoja. Järjestelmässä tulisi kuitenkin olla tietyt perustoiminnot, kuten potilaan diagnoosit, riskitiedot, laboratoriotutkimukset, rokotukset, toimenpiteet, lääkitys, kuvantamistutkimukset ja fysiologiset mittaukset. Tämän lisäksi järjestelmistä voi löytyä esimerkiksi päätöksenteon tuki, potilasportaali, ajanvaraustoiminnallisuus ja erilaisia raportteja toiminnan tarkasteluun. (SimplePractice 2023.)

Koska terveydenhuollon työ on jo valmiiksi kiireistä ja kuormittavaa, tietojärjestelmien tulisi helpottaa eikä hankaloittaa sitä. Lääkärit ovat nimenneet yhdeksi eniten stressiä aiheuttavaksi tekijäksi tietojärjestelmät. Vähentämällä terveydenhuollon ammattilaisten kokemaa stressiä ja aivokuormaa voidaan saavuttaa parempia tuloksia potilaiden hoidossa. Terveydenhuollon ammattilaisille tehdyn tutkimuksen mukaan potilastietojärjestelmissä hankaliksi koetaan tietojen hajanaisuus, puutteellinen toiminnanohjaus ja integraatioiden puute. Potilaskertomusta kirjatessaan ammattilaiset joutuvat jatkuvasti vaihtelemaan näkymää, sillä erityyppiset asiat kirjataan eri näkymillä. Näkymien vaihtelu kuormittaa ammattilaisten muistia ja hidastaa työntekoa. Käyttäjien mukaan järjestelmät eivät ohjaa käyttöä, vaan niiden käyttö perustuu muistinvaraisuuteen. Päätöksenteon tukea ei koettu hyödyllisenä, koska se ei osannut nostaa oikeita asioita potilaskertomuksesta. Terveydenhuollossa on käytössä useita järjestelmiä, jotka eivät keskustele keskenään, mikä johtaa pahimmassa tapauksessa saman tiedon kirjaamiseen useampaan paikkaan. (Vehko ym. 2018.)

4.3 Lainsäädäntö ja asetukset

Potilastietojärjestelmien kehitystä Suomessa ohjaa useat lait ja asetukset. Usean lain soveltaminen tuo haasteita, varsinkin kun lait ovat osittain vanhentuneita. Sosiaali- ja terveysministeriö pyrkii yhtenäistämään lait, jotta kaikki tieto löytyisi yhdestä paikasta. Tällä hetkellä sovellettavia lakeja ovat laki sosiaali- ja terveydenhuollon asiakastietojen sähköisestä käsittelystä 784/2021, laki

sosiaali- ja terveydenhuollon asiakastietojen käsittelystä 703/2023, laki sosiaali- ja terveydenhuollon järjestämisestä 612/2021, laki digitaalisten palvelujen tarjoamisesta 306/2019, laki julkisen hallinnon tiedonhallinnasta 906/2019 ja tietosuojalaki 1050/2018. (Terveyden ja hyvinvoinnin laitos 2024b.) Lakien lisäksi Sosiaali- ja terveysministeriö sekä Terveyden ja hyvinvoinnin laitos ovat julkaisseet asetuksia ja ohjeita täydentämään lakeja.

Suomessa Valvira valvoo potilastietojärjestelmien vaatimustenmukaisuutta, joka jaetaan toiminnallisiin vaatimuksiin, yhteentoimivuuteen ja tietoturvaan. Toiminnalliset vaatimukset määrittelevät, mitä toimintoja järjestelmässä tulee olla ja miten tietoa käsitellään. Niiden laajuuden määrittää järjestelmän käyttötarkoitus. Luokan A järjestelmiltä vaaditaan yhteentoimivuutta muiden luokan A järjestelmien kanssa. Niiden on pystyttävä tallentamaan potilastietoa Kanta-palveluihin sekä hakemaan sieltä muiden järjestelmien sinne tallentamaa potilastietoa. Yhteentoimivuus todennetaan yhteistestauksella Kelan kanssa. Tietoturva-vaatimuksilla pyritään varmistamaan potilastiedon turvallinen käsittely. Potilastietojen käsittely on sallittua ainoastaan henkilöille, joilla on olemassa oleva hoitosuhde kyseiseen potilaaseen. Lisäksi tietojen tulee aina olla ajan tasalla ja käytettävissä, kun terveydenhuollon ammattilainen niitä tarvitsee. Tietoturva-vaatimusten täyttymisen todentaa Traficom. (Valvira 2025.)

5 Toteutus

Tässä luvussa käydään läpi narratiivisen kirjallisuuskatsauksen tavoite ja siinä käytetyn aineiston valintaprosessi.

5.1 Narratiivinen kirjallisuuskatsaus menetelmänä

Narratiivinen kirjallisuuskatsaus on tutkimusmenetelmä, jossa tutkitaan ja arvioidaan tiettyyn aiheeseen liittyvää olemassa olevaa tutkimusta ja kirjallisuutta. Tavoitteena on luoda uudenlainen kokonaiskuva aiheesta, joka auttaa lukijaa ymmärtämään aihetta paremmin. Tutkijaa ohjaavat työssään tutkimuskysymykset, joihin hän pyrkii katsauksellaan löytämään vastaukset. Tutkimuskysymysten

avulla perustellaan lukijalle, miksi kyseinen tutkimus on tärkeä. Lopuksi niiden pohjalta arvioidaan, onko tutkimus vastannut asetettuihin tavoitteisiin. Narratiivinen kirjallisuuskatsaus on muita katsauksia vapaamuotoisempi, sillä sitä eivät sido tarkat säännöt. (Marjamaa & Sinisalo 2022; Mannila 2021.)

5.2 Tarkoitus ja tavoitteet

Väestön ikääntyminen ja valtion velkaantuminen lisäävät jatkuvasti terveydenhuollon kuormitusta. Tässä insinööriyössä tarkastellaan, miten digitalisaatio voi tehostaa prosesseja ja helpottaa terveydenhuollon ammattilaisten työtaakkaa. Tavoitteena on luoda yleiskatsaus suurten kielimallien hyödyntämiseen potilastietojärjestelmissä ja jäsentää lopputulema johdonmukaiseksi kokonaisuudeksi. Työ toteutetaan narratiivisena kirjallisuuskatsauksena, sillä sen avulla voidaan tehokkaasti hahmottaa laaja kokonaiskuva.

Tutkimuskysymykseksi asetettiin:

- Kuinka suurten kielimallien avulla voidaan helpottaa potilastietojärjestelmän käyttöä?

5.3 Aineiston valinta

Kirjallisuuskatsauksen aineisto haettiin PubMed-tietokannasta heinäkuussa 2025. Hakusanoina käytettiin englanninkielisiä termejä "large language model" ja "electronic health record". Hakusanojen välissä käytettiin AND-operaattoria, jolloin molemmat termit löytyivät haetusta aineistosta. Kummankin hakusanan perään lisättiin merkki (*), jotta haku tavoittaisi sanan eri taivutusmuodot.

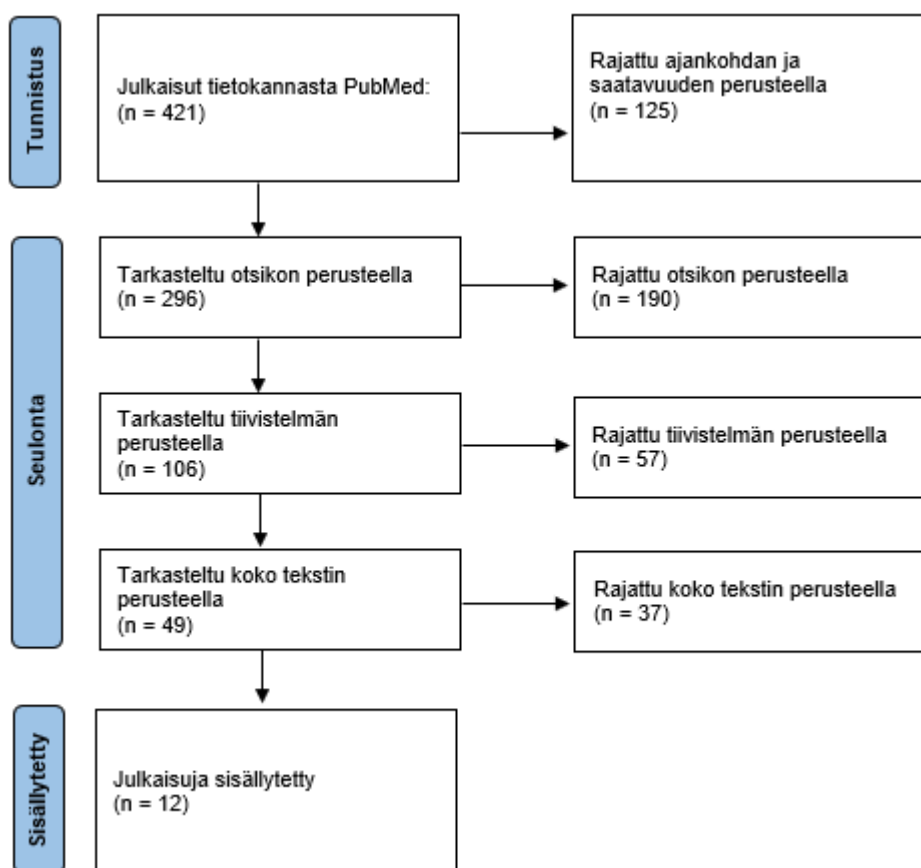
Sopivien aineistojen löytämiseksi haussa käytettiin sisäänotto- ja poissulkukriteereitä, jotka ovat esitetty taulukossa 1. Sisäänottokriteereiden mukaan aineiston tuli olla enintään viisi vuotta vanha, vertaisarvioitu, kokonaisuudessaan luetavissa, käsitellä suuria kielimalleja potilastietojärjestelmän näkökulmasta ja kielenä olla suomi tai englanti. Lisäksi käytettiin sisäänottokriteereitä tarkentavia poissulkukriteereitä, joilla oli tarkoitus karsia epäolennaiset aineistot pois.

Haetuista aineistoista poissuljettiin opinnäytetyöt ja sellaiset aineistot, jotka eivät ole sovellettavissa suomalaiseen potilastietojärjestelmään.

Taulukko 1. Sisäänotto- ja poissulkukriteerit

Sisäänottokriteerit	Poissulkukriteerit
Korkeintaan viisi vuotta vanha	Opinnäytetyöt
Vertaisarvioitu	Ei sovellettavissa suomalaiseen potilastietojärjestelmään
Kokonaisuudessaan luettavissa	
Käsittelee suuria kielimalleja potilastietojärjestelmän näkökulmasta	
Kieli suomi tai englanti	

Tietokanta-haku palautti yhteensä 421 artikkelia. Kun hakua rajattiin koskemaan viimeisen viiden vuoden aikana julkaistuja ja kokonaisuudessaan saatavilla olevia aineistoja, jäljelle jäi 296 artikkelia. Näistä rajattiin otsikon perusteella pois epäolennaiset, jolloin määrä supistui 106 artikkeliin. Artikkelien tiivistelmät luettiin, joiden perusteella aineisto rajattiin 49 artikkeliin. Artikkelien koko teksti käytiin läpi, jonka perusteella lopullinen kirjallisuuskatsauksessa käytetty aineisto rajattiin 12 artikkeliin. Kuvassa 3 havainnollistetaan aineiston rajausprosessi.



Kuva 3. Tutkimusten rajaus Prisma Flow Diagrammia mukailien (PRISMA statement 2024).

6 Tulokset

Valitut artikkelit luettiin kokonaisuudessaan ja jaoteltiin tutkimusongelman kannalta keskeisiin teemoihin. Teemoituksen tavoitteena oli jäsentää artikkelit järjestykseen kokonaisuuksiin analyysia varten. Tässä luvussa käydään artikkelien sisällöt lävitse teemoittain. Teemat ovat ennaltaehkäisy, potilasviestintä, yhteenvedot ja tietojen poiminta vapaasti kirjatusta tekstistä. Löytyneet teemat ja artikkelien keskeiset huomiot löytyvät tiivistetyssä muodossa liitteestä 1.

6.1 Ennaltaehkäisy

Ennaltaehkäisevän hoidon merkitys korostuu terveydenhuollossa, joka on perinteisesti keskittynyt reagoimaan sairauksiin niiden ehkäisyyn sijaan. Merkitys on

suuri sekä yksilön että yhteiskunnan kannalta. Ennaltaehkäisyllä voidaan vähentää sairastavuutta, parantaa elämänlaatua ja pienentää terveydenhuollon kustannuksia. Hyvä terveys näkyy pidempinä työurina ja vähäisempinä sairauspoissaoloina. Esimerkiksi diabeteksen hoitaminen maksaa yhteiskunnalle 833 miljoonaa euroa vuodessa. (Sosiaali- ja terveysministeriö 2025.)

Shashikumar ym. (2025) julkaisemassa tutkimuksessa kehitettiin avoimen lähdekoodin Mixtral 8x7B -kielimalliin perustuva verenmyrkytystä ennustava järjestelmä COMPOSER-LLM. Järjestelmä koostuu kahdesta osasta: rakenteisen tekstin tulkitsijasta ja vapaan tekstin tulkitsijasta. Kyseinen järjestelmä pyrkii ennustamaan verenmyrkytystä rakenteisesta tekstistä, kuten peruselintoiminnoista ja laboratoriotutkimuksista. Jos järjestelmä huomaa rakenteisesta tekstistä tehdyn ennusteen olevan epävarma, ottaa se mukaan vapaan tekstin kirjaukset, kuten hoitotyön merkinnät. Järjestelmää testattiin aidossa käyttöympäristössä, jossa sen suorituskyyvyksi saatiin 63,9 % F1 score -tunnusluvulla. Kielimallijärjestelmän antamia virheellisiä positiivisia tutkittiin kahden lääkärin toimesta tarkemmin, josta selvisi, että 62 % tapauksista oli epäily mahdollisesta verenmyrkytyksestä. Moni väärä positiivinen tulos olisi siis voinut olla myös aidosti positiivinen.

Myös Cui ym. (2025) tutkivat suurten kielimallien soveltuvuutta sairauksien ennustamiseen, sillä nykyisten koneoppimiseen perustuvien järjestelmien rakentaminen on aikaa vievää ja kallista. Tutkimuksessa käytettyä kielimallia ei koulutettu laajasti tehtävää varten, jotta sen käyttö olisi mahdollisimman kustannustehokasta. Tutkimusryhmän luoma kielimalli EHR-CoAgent muistuttaa hieman Shashikumar ym. (2025) luomaa mallia, sillä tässäkin mallissa on kaksi eri vaihetta. Ensimmäisessä vaiheessa ennustaja-agentti tekee ennusteen sairaudesta rakenteisten potilastietojen perusteella ja perustelee päätöksensä. Toisessa vaiheessa kriitikkoagentti saa syötteenä ennustaja-agentin ennustuksen perusteluineen. Kriitikkoagentti analysoi saamansa ennusteen, jonka perusteella se antaa palautetta ennustaja-agentille tämän tekemästä ennustuksesta. Saadun palautteen perusteella ennustaja-agentti luo uuden parannellun ennusteen. F1 score -tunnusluvulla mitattuna EHR-CoAgent oli hyvin lähellä laajasti

koulutettua koneoppimisjärjestelmää 60,21–73,88 % suorituskvyyllään. Tutkimuksen perusteella suuri kielimalli voi olla perinteistä koneoppimisjärjestelmää kustannustehokkaampi vaihtoehto.

Luo ym. (2024) tutkivat, miten suuria kielimalleja voidaan hyödyntää ennaltaehkäisevien hoitosuositusten automaattisessa poiminnassa potilastietojärjestelmiin. Kun potilaan tiedot vastaavat hoitosuosituksista poimittuja riskikriteerejä, järjestelmä voi luoda terveydenhuollon ammattilaiselle herätteen tarvittavista hoitotoimenpiteistä. Tutkimuksessa käytettiin kolmea kielimallia (GPT-3.5, GPT-4 ja PaLM 2), joita ei koulutettu annettuun tehtävään. Kielimallien tehtävänä oli poimia hoitosuosituksia riskikriteereineen internetistä ja lajitella ne tutkijoiden ennalta määrittämiin kategorioihin. Parhaiten suoriutunut kielimalli onnistui poimimaan kaikki riskikriteerit täysin oikein yhdeksästä suosituksesta, kun suosituksia oli yhteensä 24. Kielimallien suorituskvyyky vaihteli eri kategorioiden välillä. Suorituskvyyky oli parhaimmillaan sellaisten kriteerien poiminnassa, jossa käytetään standardoitua tietoa kuten ikä ja sukupuoli. Kriteerit, jotka perustuvat monisanaisiin kuvauksiin, tuottivat kielimalleille vaikeuksia. Lisäksi tutkimuksessa havaittiin kielimalleille ominaista hallusinoitua, sillä mallit sekoittivat aiemmin oppimiaan tietoja poimittuihin hoitosuosituksiin, mikä johti virhetulkintoihin. Vaikka kielimallit ovat huomattavasti ihmisiä nopeampia tulkitsemaan hoitosuosituksia, on niiden tarkkuudessa vielä parantamisen varaa.

Tutkimusten perusteella suurilla kielimalleilla voidaan saavuttaa merkittäviä yhteiskunnallisia hyötyjä sairauksien ennaltaehkäisyssä. Ne parantavat potilasturvallisuutta tehostamalla varhaista sairauksien tunnistamista ja tarjoavat terveydenhuollon ammattilaisille ajantasaista tukea päätöksentekoon, mikä mahdollistaa yksilöllisemmän hoidon ja resurssien tehokkaamman käytön. Vaikka tarkkuudessa on vielä kehitettävää ja hallusinoitua tuo omat haasteensa, tulokset osoittavat, että kielimallit voivat olla kustannustehokas keino edistää sairauksien ennaltaehkäisyä.

6.2 Potilasviestintä

Potilasportaalien yleistymisen on lisännyt digitaalista viestintää potilaiden ja terveydenhuollon ammattilaisten välillä. Tämä kasvattaa ammattilaisten kognitiivista kuormaa ja voi johtaa työuupumukseen, sillä pelkästään potilasviesteihin vastaaminen vie lääkäreiltä keskimäärin puolitoista tuntia päivässä. Viestinnän tarvetta lisäävät myös vaikeaselkoiset raportit, joita potilaat eivät ymmärrä ilman ammattilaisten selitystä. (Liu ym. 2024: 1; Yang ym. 2025: 2.)

Liu ym. (2024) kehittivät kaksi kielimallia CLAIR-Shortin ja CLAIR-Longin, jotka perustuvat avoimen lähdekoodin kielimalliin LLaMa-65B. Näiden kielimallien avulla selvitettiin, kuinka hienosäädetty kielimalli suoriutuu potilasviesteihin vastaamisesta. Molemmat kielimallit hienosäädettiin potilasportaalista noudettujen noin 500 000 viestin avulla. Tämän lisäksi CLAIR-Longin koulutuksessa käytettiin 5000 avoimen lähteen potilasviestiä, joita paranneltiin GPT-3.5-kielimallin avulla. Näiden kahden hienosäädetyn kielimallin vastauksia verrattiin GPT-3.5- ja GPT-4-kielimallien sekä terveydenhuollon ammattilaisen tuottamiin vastauksiin neljän lääkärin toimesta. Hienosäädettyjen LLaMa-65B-kielimallien vastaukset saivat heikommat arvostelut kuin hienosäätämättömien GPT-kielimallien vastaukset, mutta olivat kuitenkin lähtökohtaisesti parempia kuin terveydenhuollon ammattilaisen vastaukset. Vaikka hienosäätö ei välttämättä paranna vastauksien laatua, voi se auttaa terveydenhuollon ammattilaisia omaksumaan tekoälyn helpommin, koska hienosäädetyn kielimallin vastaukset saadaan muistuttamaan terveydenhuollon ammattilaisten vastauksia.

Kuten Liu ym. tutkimuksessa, myös Afshar ym. (2024) tutkivat, kuinka suurta kielimallia voidaan hyödyntää potilasportaaliviesteihin vastaamiseen. Näkökulma tässä tutkimuksessa oli hieman erilainen, sillä kielimallia ei yritetty parantaa hienosäädöllä vaan kehotesuunnittelulla (engl. prompt engineering). Tutkimus jaettiin kahteen vaiheeseen. Ensimmäisessä vaiheessa kielimalli testattiin lääkärin luomalla kehoitteella, ja toisessa vaiheessa kyseistä kehoitetta pyrittiin parantamaan kehotesuunnittelun avulla. Kehotesuunnittelulla pyrittiin vaikuttamaan kielimallin johdonmukaisuuteen ja auttamaan sitä ymmärtämään, millaisia

vastauksia siltä odotetaan. Lääkärien luomalla kehoitteella 12,0 % viesteistä sai negatiivisen palautteen, kun kehoitesuunnittelulla paranneltu viesti sai negatiivisen palautteen 8,2 % tapauksista. Laadun parantuminen ei kuitenkaan lisännyt luonnosten käyttöastetta merkittävästi. Lääkärit hyödynsivät vain 17,5 % luonnoksista, joista 2,6 % otettiin käyttöön muokkaamatta.

Yang ym. (2025) tutkivat, voiko suuri kielimalli (GPT-4) tehdä patologian lausunnoista potilasystävällisempiä. Vapaaehtoiset, joilla ei ollut lääketieteellistä taustaa, arvioivat lausuntojen ymmärrettävyyttä. Vapaaehtoiset lukivat sekä alkuperäisen lausunnon että kielimallin luoman yksinkertaistetun version. Tämän jälkeen he arvioivat asteikkoa käyttäen, kuinka paljon helpompi yksinkertaistettu teksti oli ymmärtää. Kielimallin yksinkertaistamat lausunnot nostivat keskimääräiset arviot alkuperäisen lausunnon 5,23:sta 7,98:aan. Lisäksi potilaan kanssa käytetty aika lausuntojen läpikäyntiin lyheni 35 minuutista 10 minuuttiin eli yli 70 %. Kolme patologia tarkisti kielimallin luomat lausunnot, eivätkä löytäneet niistä epä johdonmukaisuuksia. Selkeämmät lausunnot voivat parantaa potilaiden hoitoon sitoutumista heidän ymmärtäessään terveydentilansa paremmin.

Tutkimukset osoittavat kielimallien potentiaalin potilasviestinnän helpottajana, mutta täysimääräisessä käyttöönotossa on vielä haasteita. Jatkossa tulisi tutkia potilaiden näkemyksiä kielimallin tuottamista sisällöistä, sillä edellä mainituissa tutkimuksissa arvioijina toimi terveydenhuollon ammattilaiset. Olisi myös hyödyllistä tutkia, kuinka aiemmin kirjattujen potilastietojen hyödyntäminen vaikuttaa kielimallin generoimien tekstien laatuun. Muutosvastarinnan vähentämiseksi on löydettävä keinoja, joilla kielimallit saadaan tuotua luonnolliseksi osaksi klinisiä työnkulkuja.

6.3 Tietojen poiminta vapaasti kirjatusta tekstistä

Potilastiedon kirjaaminen on tärkeää hoidon jatkuvuuden kannalta. Terveydenhuollon ammattilaisilla on oltava käytössä ajantasainen tieto potilaan terveydentilasta ja suoritetuista toimenpiteistä, jotta he voivat tarjota parasta mahdollista hoitoa. Kirjaaminen kuitenkin on aikaa vievä prosessi, ja usein se tehdään ei-

rakenteisessa muodossa vapaalla tekstillä. Tällöin tärkeät tiedot voivat hukkua tekstimassan joukkoon, ja niiden hyödyntäminen esimerkiksi tutkimuksessa, raportoinnissa tai yhteenvedoissa on hankalaa. Suuret kielimallit ovat erinomaisia käsittelemään tekstiä, joten niiden avulla on mahdollisuus poimia tärkeät asiat tekstimassasta ilman aikaa vievää manuaalista työtä. (Ntinopoulos ym. 2025: 1; Garcia-Carmona ym. 2025: 2.)

Ntinopoulos ym. (2025) selvittivät, kuinka 19 eri kielimallia pystyvät poimimaan potilastietoa rakenteiseen muotoon vapaasti kirjatusta tekstistä. Oikeiden potilastietojen arkaluontoisuuden vuoksi tutkimus suoritettiin 50 keinotekoisesti laaditun potilaskertomuksen avulla. Nämä keinotekoisesti laaditut potilaskertomukset sisälsivät sekä rakenteista että vapaata tekstiä. Teksteihin lisättiin kirjoitusvirheitä ja lyhenteitä, jotta ne vastaisivat paremmin todellisuutta. Kielimallit testattiin tehtävässä, jossa piti tunnistaa 250 potilaskertomuksesta maininnat sydänleikkauksen jälkeisistä komplikaatioista. Kielimallit suorittivat saman tehtävän kolme kertaa yhdenmukaisuuden selvittämiseksi. Parhaiten tehtävästä selviytyi Claude 3.0 Opus, joka teki tehtävässä kaksi virhettä. Toiseksi sijoittui Claude 3.0 Sonnet kolmella virheellä, ja kolmanneksi Claude 2.0 ja GPT-4, jotka tekivät viisi virhettä. Yhdenmukaisin kielimalli oli Claude 2.0 saaden täydet pisteet Krippendorffin alfa -mittarilla eli se antoi täsmälleen saman vastauksen joka kerta. Myös muut tehtävässä pärjänneet kielimallit olivat erittäin yhdenmukaisia. Tulokset ovat lupaavia, mutta on otettava huomioon, että tutkimusta ei suoritettu oikealla potilasaineistolla ja annettu tehtävä oli hyvin suppea.

Garcia-Carmona ym. (2025) lähestyivät tietojen poimintaa vapaasta tekstistä hieman eri kulmalla kuin Ntinopoulos ym. (2025). Keinotekoisien potilaskertomuksien sijaan käytettiin 100 aitoa kertomusta, jotka oli valittu laajalla otannalla eri erikoisaloilta. Erona oli myös Retrieval Augmented Generation (RAG) -menetelmän käyttö, jonka avulla pyrittiin minimoimaan kielimallien hallusinaatiot ja epätarkkuudet. RAG-menetelmässä kielimallin käyttöön annetaan erillinen tehtävä kohtainen aineisto, jota se voi hyödyntää antaessaan vastausta. Tutkimuksessa vertailtiin kuutta kielimallia, jotka ovat GPT-4, Llama 3, Llama 3.1, Gemma 2, Qwen 2 ja Qwen 2.5. Kielimallien tehtävänä oli poimia tietoa

kategorioihin nimi, ikä, diagnoosit, lääkkeet ja testit. Parhaiten tehtävistä suoriutui GPT-4 F1 score -tunnusluvalla 81 %. Heikoiten kielimallit selvisivät nimien ja ikien poiminnassa. Nimikategorian tuloksia ei voida pitää luotettavina, sillä potilaiden nimet olivat anonymisoitu epärealistisesti, eivätkä muistuttaneet oikeita nimiä.

Gu ym. (2025) puolestaan tutkivat, kuinka avoimen lähdekoodin suuret kielimallit osaavat poimia terveyden sosiaalisia määrittäjiä (engl. social determinants of health) vapaan tekstin joukosta. Terveyden sosiaaliset määrittäjät sisälsivät siivillisäädyn, lasten lukumäärän, työllisyystilanteen, koulutustason, elämäntapatekijät (tupakan, alkoholin, huumeiden käytön ja liikunnan) sekä asumistilanteen. Terveyden sosiaalisten määrittäjien avulla voidaan arvioida yksilön terveydentilaa. Tutkimuksessa käytetyt kielimallit poikkesivat kielimalleista, joita käytettiin Garcia-Carmona ym. (2025) ja Ntinopoulos ym. (2025) tutkimuksissa. Käytetyt kielimallit olivat openchat_3.5, zephyr-7b, vicuna-7b, Llama-2-7b, vicuna-13b, WizardLM-13b, Llama-2-13b ja vicuna-33b. Toisin kuin aiemmissä tutkimuksissa, aineistona käytettiin oikeita, anonymisoimattomia potilaskertomuksia. Tietojen poiminnat suoritettiin ilman kehotesuunnittelua sekä kehotesuunnittelun kanssa. Parhaiten annetusta tehtävästä suoriutui openchat_3.5, jonka F1 score -tunnusluku oli lähes 80 %. Kehotesuunnittelulla oli kaksijakoinen vaikutus: halusointi väheni, mutta se teki malleista varovaisempia, mikä puolestaan heikensi niiden tarkkuutta.

Nämä kolme tutkimusta osoittavat, että suuret kielimallit voivat tehokkaasti poimia tietoa potilaskertomusten vapaasta tekstistä. Mallien suorituskyky paranee, kun niitä hienosäädetään ja koulutetaan terveydenhuoltoon liittyvällä aineistolla. Tulokset kuitenkin vaihtelevat, sillä malli, joka suoriutuu hyvin yhdestä tehtävästä, ei välttämättä onnistu toisessa. Siksi oikean mallin valinta on ratkaisevan tärkeää. Luotettavampien tulosten saamiseksi malleja tulisi testata todellisissa käyttöympäristöissä, sillä nykyiset tutkimukset perustuivat suppeisiin aineistoihin ja rajattuihin tehtäviin. Laajemmat tutkimukset ovat tarpeen tulosten yleistämiseksi.

6.4 Yhteenvedot

Terveydenhuollon ammattilaiset käyttävät huomattavan osan työajastaan potilastietojen tarkasteluun. Potilastietojen sähköistyminen sekä lisääntyneet kirjaimisvaatimukset ovat kasvattaneet käsiteltävän tiedon määrää, mikä on yhteydessä ammattilaisten kuormittumiseen ja työuupumuksen riskiin. Suuret kielimallit voivat tarjota ratkaisun tähän haasteeseen tuottamalla tiiviitä ja kattavia yhteenvetoja potilaan keskeisistä tiedoista. Tällaiset yhteenvedot tukevat päätöksentekoa, helpottavat terveystarkistusten tunnistamista ja vähentävät hajanaisen tiedon läpikäymiseen kuluvaan aikaan. (Fraile Navarro ym. 2025: 1–2; Van Veen ym. 2024: 2.)

Van Veen ym. (2024) tutkimuksessa arvioitiin, pystyvätkö suuret kielimallit tuottamaan potilastiedoista yhtä laadukkaita tai jopa parempia yhteenvetoja kuin lääkärit. Kielimalleille annettiin neljä erilaista yhteenvetotehtävää. Niiden piti tehdä yhteenveto radiologian lausunnosta, potilaiden kysymyksistä, hoitokertomusteksteistä sekä lääkärin ja potilaan välisestä keskustelusta. Tutkimuksen ensimmäisessä vaiheessa arvioitiin kahdeksaa eri kielimallia: FLAN-T5, FLAN-UL2, Alpaca, Med-Alpaca, Vicuna, Llama-2, GPT-3.5 ja GPT-4. Kielimallien tuottamien yhteenvetojen laatua mitattiin vakiintuneilla luonnollisen kielen käsittelyn mittareilla (BLEU, ROUGE-L, BERTscore ja MEDCON). Parhaiten suoriutui GPT-4, joka valittiin tutkimuksen seuraavaan vaiheeseen. Toisessa vaiheessa kymmenen lääkäriä arvioi 100 yhteenvetoa sokkokokeessa: heille näytettiin kaksi yhteenvetoa kerrallaan, joista toinen oli GPT-4:n ja toinen lääkärin tekemä. Arvioijien tehtävänä oli valita, kumpi yhteenveto oli heidän mielestään parempi kattavuuden, virheettömyyden ja tiiveyden perusteella. Kielimallin tuottama yhteenvetoa pidettiin 45 % tapauksista vastaavana kuin lääkärin ja 35 % tapauksista parempana. Viimeisessä vaiheessa tehtiin turvallisuusanalyysi, jossa verrattiin virheiden määrää ja vakavuutta. Tulosten mukaan lääkärit tekivät todennäköisemmin ja vakavampia virheitä kuin GPT-4.

Myös Fraile Navarro ym. (2025) arvioivat suurten kielimallien soveltuvuutta yhteenvetojen luomiseen. Tutkimuksessa käytettiin neljää hienosäädettyä ja yhtä

yleiskäyttöistä kielimallia. Kielimallit BART-LSG-conv ja BART-DnC hienosäädettiin kliinisellä aineistolla, LongDialSumm yleisellä aineistolla ja T5-Booksum kirjallisuuteen perustuvalla aineistolla. Mukana oli myös hienosäätämätön GPT-3.5. Kuten Van Veen ym. (2024) tutkimuksessa, kielimalleja arvioitiin luonnollisen kielen käsittelyn mittareilla. Tässä tutkimuksessa käytettiin UniEval-mittaria ROUGEn lisäksi. Tutkijat huomasivat UniEval-mittarin toimivan paremmin yhteenvetojen laadun arvioinnissa kuin ROUGEn, sillä UniEvalin antamat arviot olivat lähempänä ihmisten antamia arvioita. ROUGEn antamilla arvioilla paras kielimalli yhteenvedon luomiseen oli hienosäädetty BART-LSG-conv ja heikoin GPT-3.5. UniEval- ja ihmisarvointi kuitenkin käänsivät tilanteen päinvastaiseksi, jolloin GPT-3.5 oli paras, jopa parempi kuin ihminen. Arvioijia pyydettiin kertomaan, millä perusteella he ovat arvionsa antaneet. Heikoimmat arviot saivat yhteenvedot, joissa oli faktavirheitä, hallusinaatioita, kieliopillisia ongelmia tai epätarkkuuksia. Parhaimmat arvosanat annettiin yhteenvedoille, jotka olivat tarkkoja, rakenteeltaan selkeitä ja tiivistivät sisällön tehokkaasti.

Pienetkin muutokset suurelle kielimallille annettuun kehoitteeseen voivat johtaa huomattavan erilaisiin yhteenvetoihin. Koska potilastietojärjestelmiltä edellytetään luotettavuutta, tällainen vaihtelevuus rajoittaa kielimallien käyttöä. Tämän haasteen ratkaisemiseksi Chuang ym. (2024) kehittivät SPeC (Soft Prompt-Based Calibration) -menetelmän, jonka tarkoituksena on pitää yhteenvedot samankaltaisina riippumatta kehoitteen pienistä eroista. Pehmeä kehote (engl. soft prompt) ei perustu ihmisen kirjoittamaan käskyyn tai kysymykseen, kuten perinteinen kehote. Pehmeä kehote on käyttäjälle näkymätön ohje, joka liitetään automaattisesti osaksi käyttäjän syöttämää kehotetta. Niiden avulla kielimallin alkuperäisiä parametrejä ei tarvitse hienosäätää tiettyä tehtävää varten, vaan kielimalli saa ikään kuin lisäosan uusia parametrejä. SPeC-menetelmän toimivuutta testattiin kolmella kielimallilla (Flan-T5, BART ja Pegasus-xsum), joiden tehtävänä oli luoda yhteenvedoja radiologien lausunnoista. Tulosten mukaan SPeC vähensi merkittävästi yhteenvetojen vaihtelevuutta – Flan-T5-kielimallin kohdalla vaihtelevuus väheni jopa 43 %.

Tutkimukset osoittivat, että suurilla kielimalleilla on merkittävää potentiaalia potilastietojen tiivistämisessä. Kielimallien suorituskkyky oli vähintään yhtä hyvä kuin ihmisen, ja parhaimmillaan jopa parempi. Kielimallien luomilla yhteenvedoilla voidaan säästää terveydenhuollon ammattilaisten aikaa itse potilastyölle ja parantaa hoidon laatua. Haasteena kuitenkin on kielimallin suorituskyyvyn vaihtelu riippuen sille annetusta syötteestä. Oman haasteensa tuo myös se, kuinka kielimallin suorituskkykyä voidaan mitata luotettavasti. Automaattisesti suoritettavat mittarit eivät välttämättä anna todellista kuvaa, vaan arviointiin on otettava mukaan ihmisen näkemys.

7 Yhteenveto

Insinööriyön tarkoituksena oli selvittää, kuinka suurten kielimallien avulla voidaan helpottaa potilastietojärjestelmän käyttöä. Kirjallisuuskatsauksen perusteella löydettiin lupaavia käyttökohteita, mutta myös haasteita. Tässä luvussa on tarkoitus käydä läpi keskeiset havainnot valituista tutkimuksista, pohtia kirjallisuuskatsauksen eettisyyttä ja luotettavuutta sekä ehdottaa jatkotutkimuskohteita.

7.1 Keskeiset huomiot

Kiinnostus suurilla kielimalleilla kohtaan on ollut jatkuvassa nousussa Transformer-arkkitehtuurin ja varsinkin ChatGPT-kielimallin julkistuksen jälkeen. Niiden kyky ymmärtää ja tuottaa tekstiä avaa laajat mahdollisuudet toimia terveydenhuollon ammattilaisten apurina erinäisissä tehtävissä säästämällä ammattilaisten aikaa itse potilastyötä varten. Tällä hetkellä tutkituimpia käyttökohteita on tietojen poiminta jäsentymättömästä tekstistä, sairauksien ennustaminen, potilastietojen tiivistäminen sekä potilasviestinnän tehostaminen. Tutkimuksien mukaan suuret kielimallit pärjäävät kyseisissä tehtävissä varsin hyvin, jopa yhtä hyvin tai paremmin kuin ihminen. Kielimallit eivät vain vapauta aikaa terveydenhuollon ammattilaisille, vaan ne tukevat päätöksenteossa, sujuvoittavat kommunikointia potilaan kanssa ja tuovat tärkeät tiedot esille suuren tietomassan joukosta.

Esikoulutetut kielimallit toimivat hyvin sellaisenaan, mutta niiden suorituskkyä voidaan parantaa hienosäätämällä ja kehotesuunnittelulla. Hienosäädössä kielimallia koulutetaan uudelleen rajatulla tehtäväkohtaisella aineistolla. Tämä prosessi muokkaa kielimallin sisäisiä parametrejä pysyvästi tiettyä tarkoitusta varten. Hienosäätö on kehotesuunnittelua raskaampi prosessi, ja se vaatii enemmän teknistä osaamista ja laskentaresurssia. Kehotesuunnittelussa ei kosketa kielimallin sisäiseen toimintaan, vaan kielimallia ohjataan parempaan lopputulokseen sille annetun syötteen avulla. Syötteessä voidaan esimerkiksi antaa kielimallille mallivastauksia, joita sen pitäisi mukailla. Kehotesuunnittelu on aloittelijaystävällisempi vaihtoehto kielimallin jatkojalostamiseen. Hienosäätö ja kehotesuunnittelu eivät poissulje toisiaan, vaan ne ovat toisiaan täydentäviä tekniikoita.

Vaikka suuret kielimallit omaavat suuren potentiaalın terveydenhuollon apurina, on niiden laajamittaisessa käyttöönotossa vielä haasteita. Kielimallit tuottavat välillä virheellistä tietoa, jota kutsutaan hallusinoinniksi. Hallusinaatiot voivat johtua kielimallin koulutuksessa käytettävän aineiston virheistä tai kielimallin toimintalogiikasta ennustaa seuraavaa todennäköistä sanaa. Kielimalleilla on myös taipumus tuottaa erilainen vastaus eri ajokerroilla, vaikka sille annettu syöte olisi täysin sama. Terveysthuollossa vaaditaan erityistä luotettavuutta tietojärjestelmiä kohtaan potilasturvallisuuden varmistamiseksi, joten tällainen epäjohtonmukaisuus ei ole hyväksyttävää. Kielimallin käyttöönoton haasteet eivät rajoitu pelkästään teknisiin rajoitteisiin, vaan on otettava myös huomioon ihmilliset tekijät. Koska muutosvistarinta on ihmisille luontaista, uuden teknologian tuominen osaksi arjen työtä ei aina ole helppoa. Käyttäjät on otettava mukaan kehityksen alkuvaiheilla, jotta tekoälyapurista saadaan heille mieleinen. Teknisten ja ihmillisten haasteiden lisäksi on vielä tietosuoja ja eettiset ongelmat. Potilastieto on arkaluontoista, joten sen käsittelyssä on oltava hyvin varovainen. Erityisesti pilvipalveluissa toimivien kielimallien kohdalla on vaarana arkaluontoisen datan siirtyminen ulkopuolisten käyttöön. Oman haasteensa tuo kielimallin päättelyketjun läpinäkymättömyys, joka tekee mahdottomaksi selvittää, millä perusteilla malli päätyi tiettyyn vastaukseen.

7.2 Eettisyys ja luotettavuus

Kirjallisuuskatsauksessa käytettiin vertaisarvioituja julkaisuja laadun ja luotettavuuden varmistamiseksi. Vertaisarvioidut julkaisut ovat käyneet läpi saman alan asiantuntijoiden arvioinnin ennen julkaisua. Hakusanoina käytetyt termit “large language model” ja “electronic health record” eivät välttämättä palauttaneet kaikkia olennaisia artikkeleita, sillä potilastietojärjestelmään voidaan viitata useilla muillakin termeillä, kuten myös suuriin kielimalleihin. Valittuihin hakusanoihin päädyttiin ajankäytöllisestä syystä. Haku suoritettiin PubMed-tietokantaan, joka on yksi maailman käytetyimmistä tietokannoista terveysalalla. Kattavamman otannan saamiseksi haku olisi voitu suorittaa useampaan tietokantaan. Haku rajattiin korkeintaan viisi vuotta vanhoihin julkaisuihin, mutta valitut julkaisut olivat alle kaksi vuotta vanhoja. Valinta perustui siihen, että suuret kielimallit kehittyvät niin kovaa vauhtia, ettei vanhoilla tutkimuksilla ole enää suurta painoarvoa. Selkeillä sisäänotto- ja poissulkukriteereillä tuotiin näkyväksi, kuinka julkaisut valittiin katsaukseen ja vähennettiin subjektiivista julkaisujen poimintaa.

Hauilla löydetyistä julkaisuista valittiin vain sellaisia, jotka aidosti vastasivat asetettuun tutkimuskysymykseen. Julkaisujen valinnassa pyrittiin huomioimaan erilaisia näkökulmia monipuolisen kokonaiskuvan saamiseksi. Löydetyistä julkaisuista huomattiin niiden esittelevän aihetta hyvin positiiviseen sävyyn. Pohdittavaksi jää, julkaistaanko positiivisia tutkimuksia herkemmin kuin negatiivisia. Julkaisujen tuli olla helposti ymmärrettävissä ja niissä tehdyt johtopäätökset perusteltuja. Valinnassa olisi voitu kiinnittää tarkemmin huomiota, kuinka arvoitettuja julkaisun kirjoittajat ovat omalla alallaan ja kuinka paljon kyseistä julkaisua on siteerattu aiemmin. Valituille julkaisuille ei suoritettu järjestelmällistä laaduntarkistusta, joka voi vaikuttaa katsauksen luotettavuuteen.

7.3 Jatkotutkimusehdotukset

Kirjallisuuskatsauksessa tarkastellut tutkimukset tehtiin joko keinotekoisella tai vanhentuneella potilastiedolla. Tutkimusta pitäisi suunnata todelliseen ympäristöön, missä potilaskertomus täydentyy jatkuvasti. Tämä auttaa ymmärtämään,

miten kielimallit toimivat oikeasti potilastietojärjestelmissä ja millainen vaikutus niillä on potilaan hoitoon. Samalla pystyttäisiin tutkimaan, kuinka terveydenhuollon ammattilaiset hyväksyvät tämän uuden teknologian osaksi työkulkujaan. Kokevatko he kielimallin rasitteena vai hyötynä?

Kielimallien hienosäätö vaatii lisää tutkimusta havaintojen yleistämiseksi. Tutkimukset osoittivat hienosäädöllä olevan positiivisia vaikutuksia kielimallin suori- tuskykyyn, mutta tutkimukset keskittyivät hyvin spesifeihin tehtäviin ja käytetty aineisto oli hyvin rajallinen. Tutkimuksista kävi myös ilmi, että kielimallien suori- tuskyky vaihteli eri tehtävien välillä. Lisätutkimuksia tarvitaan selvittämään, mikä kielimalli sopii kuhunkin tehtävään parhaiten. Kielimallin hallusinoinnin lieventä- minen on edelleen tärkeää, erityisesti potilastietojen käsittelyssä, koska kyse on potilasturvallisuudesta.

Lähteet

Afshar, Majid; Gao, Yanjun; Wills, Graham; Wang, Jason; Churpek, Matthew; Westenberger, Christa; Kunstman, David; Gordon, Joel; Goswami, Cherodeep; Liao, Frank & Patterson Brian. 2024. Prompt engineering with a large language model to assist providers in responding to patient inquiries: a real-time implementation in the electronic health record. JAMIA Open.

Chuang, Yu-Neng; Tang R, Ruixiang; Jiang, Xiaoqian & Hu Xia. 2024. SPeC: A Soft Prompt-Based Calibration on Performance Variability of Large Language Model in Clinical Notes Summarization. Journal of Biomedical Informatics.

Cui, Hejie; Shen, Zhuocheng; Zhang, Jieyu; Shao, Hui; Qin, Lianhui; Ho, Joyce & Yang, Carl. 2025. LLMs-based Few-Shot Disease Predictions using EHR: A Novel Approach Combining Predictive Agent Reasoning and Critical Agent Instruction. AMIA Annual Symposium Proceedings.

Euroopan komissio. 2025a. Verkkoaineisto. Tekoälylautakunta. <<https://digital-strategy.ec.europa.eu/fi/policies/ai-board>>. Luettu 17.6.2025.

Euroopan komissio. 2025b. Verkkoaineisto. Tekoälysäädös. <<https://digital-strategy.ec.europa.eu/fi/policies/regulatory-framework-ai>>. Luettu 20.6.2025.

Euroopan parlamentti. 2025. Verkkoaineisto. EU:n tekoälysäädös on ensimmäinen laatuaan. <<https://www.europarl.europa.eu/topics/fi/article/20230601STO93804/eu-n-tekoalyasaados-on-ensimmainen-laatu>>. Luettu 15.6.2025.

Euroopan parlamentti ja neuvosto. 2024. Asetus (EU) 2024/1689 tekoälyä koskevista yhdenmukaistetuista säännöistä. EUR-Lex.

Ferrer, Josep. 2024. Verkkoaineisto. How Transformers Work: A Detailed Exploration of Transformer Architecture. <<https://www.datacamp.com/tutorial/how-transformers-work>>. Luettu 9.6.2025.

Fraile Navarro, David; Coiera, Eerico; Hambly, Thomas; Triplett, Zoe; Asif, Nahyan; Susanto, Anindya; Chowdhury, Anamika; Azcoaga Lorenzo, Amaya; Dras, Mark & Berkovsky, Shlomo. 2025. Expert evaluation of large language models for clinical dialogue summarization. Scientific Reports.

Garcia-Carmona, Angel; Prieto, Maria-Lorena; Puertas Enrique & Beunza Juan-Jose. 2025. Leveraging Large Language Models for Accurate Retrieval of Patient Information From Medical Reports: Systematic Evaluation Study. JMIR Publications.

Gu, Bowen; Shao, Vivian; Liao, Ziqian; Carducci, Valentina; Brufau, Santiago; Yang, Jie & Desai Rishi. 2025. Scalable information extraction from free text electronic health records using large language models. BMC Med Res Methodol.

Kaplan, Jared; McCandlish, Sam; Henighan, Tom; Brown, Tom; Chess, Benjamin; Child, Rewon; Gray, Scott; Radford, Alec; Wu, Jeffrey & Amodei Dario. 2020. Scaling Laws for Neural Language Models. arXiv preprint arXiv:2001.08361.

Liu, Siru; McCoy, Allison; Wright, Aileen; Carew, Babatunde; Genkins, Julian; Huang, Sean; Peterson, Josh; Steitz, Bruan & Wright Adam. 2024. Leveraging Large Language Models for Generating Responses to Patient Messages. Journal of the American Medical Informatics Association.

Luo, Xiao; Tahabi, Fattah; Marc, Tressica; Hainert, Laura & Storey, Susan. 2024. Zero-shot learning to extract assessment criteria and medical services from the preventive healthcare guidelines using large language models. The Journal of the American Medical Informatics Association.

Mannila, Margit. 2021. Verkkoaineisto. Kirjallisuuskatsaus opinnäytetyön muotona. <<https://energiaa.vamk.fi/artikkelit/osaaminen/kirjallisuuskatsaus-opinnaytetyon-muotona/>>. Luettu 11.7.2025.

Marjamaa, Minna & Sinisalo, Riikka. 2022. Verkkoaineisto. Kirjallisuuskatsauksen ohjaus – perustana tutkimuskysymys ja ohjaushaastattelu. <<https://www.kreodi.fi/arkisto/artikkelit/kirjallisuuskatsauksen-ohjaus-perustana-tutkimuskysymys-ja-ohjaushaastattelu.html>>. Luettu 11.7.2025.

Ntinopoulos, Vasileios; Rodriguez Cetina Biefer, Hector; Tudorache, Igor; Papadopoulos, Nestoras; Odavic, Dragan; Risteski, Petar; Haeussler, Achim & Dzemali, Omer. 2025. Large language models for data extraction from unstructured and semi-structured electronic health records: a multiple model performance evaluation. BMJ Health & Care Informatics.

Patil, Rajvardhan & Gudivada, Venkat. 2024. A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs). Applied Sciences.

PRISMA Flow Diagram. 2024. Verkkoaineisto. PRISMA statement. <<https://www.prisma-statement.org/prisma-2020-flow-diagram>>. Luettu 9.3.2025.

Raiaan, Mohaimenul; Mukta, Saddam; Fatema, Kaniz; Fahad Nur Mohammad; Sakib, Sadman; Mim, Most & Azam, Sami. 2024. A review on large language

models: Architectures, applications, taxonomies, open issues and challenges. *IEEE access*.

Ungless, Eddie; Vitsakis, Nikolas; Talat, Zeerak; Garforth, James; Ross, Björn; Onken, Arno; Kasirzadeh, Atoosa & Birch, Alexandra. 2024. The Only Way is Ethics: A Guide to Ethical Research with Large Language Models. arXiv:2412.16022

Shashikumar, Supreeth; Mohammadi, Sina; Krishnamoorthy, Rishivardhan; Patel, Avi; Wardi, Gabriel; Ahn, Joseph; Singh, Karandeep; Aronoff-Spencer, Eliah & Nemati, Shamim. 2025. Development and Prospective Implementation of a Large Language Model based System for Early Sepsis Prediction. medRxiv.

SimplePractice. 2023. Verkkoaineisto. What Are the Core Functionalities of an EHR System?. <<https://www.simplepractice.com/resource/ehr-functionality/>> Luettu 6.7.2025.

Sitra. 2025. Verkkoaineisto. Kokeilut: Tekoäly sujuvoitti potilastyön arkea, vähensi kirjallisia töitä ja vapautti ammattilaisten aikaa hoitotyöhön. <<https://www.sitra.fi/artikkelit/kokeilut-tekoaly-sujuvoitti-potilastyon-arkea-va-hensi-kirjallisia-toita-ja-vapautti-ammattilaisten-aikaa-hoitotyohon/>>. Luettu 22.8.2025.

Sosiaali- ja terveysministeriö. 2025. Verkkoaineisto. Terveyden ja hyvinvoinnin edistäminen kannattaa. <<https://stm.fi/terveyden-ja-hyvinvoinnin-edistaminen-kannattaa>>. Luettu 26.7.2025

Terveyden ja hyvinvoinnin laitos. 2024a. Verkkoaineisto. Kirjaaminen. <<https://thl.fi/aiheet/tiedonhallinta-sosiaali-ja-terveysalalla/kirjaaminen>>. Luettu 2.7.2025.

Terveyden ja hyvinvoinnin laitos. 2024b. Verkkoaineisto. Lainsäädäntö ja linjaukset. <<https://thl.fi/aiheet/sote-palvelujen-johtaminen/kehittyva-palvelujarjestelma/digitaaliset-palvelut/lainsaadanto-ja-linjaukset>>. Luettu 8.7.2025.

Valtioneuvosto. 2025. Verkkoaineisto. EU:n tekoälyasetuksen soveltaminen alkaa 2.8.2025. <<https://valtioneuvosto.fi/-/1410877/eu-n-tekoalyasetuksen-soveltaminen-alkaa-2.8.2025>>. Luettu 31.8.2025.

Valvira. 2025. Verkkoaineisto. Sosiaali- ja terveydenhuollon tietojärjestelmät. <<https://valvira.fi/sosiaali-ja-terveydenhuollon-tietojarjestelmat>>. Luettu 9.7.2025.

Van Veen, Dave; Van Uden, Cara; Blankemeier, Louis; Delbrouck, Jean-Benoit; Aali, Asad; Bluethgen, Christian; Pareek, Anuj; Polacin, Malgorzata; Reis, Eduardo; Seehofnerová, Anna; Rohatgi, Nidhi; Hosamani, Poonam; Collins, William; Ahuja, Neera; Langlotz, Curtis; Hom, Jason; Gatidis, Sergios; Pauly, John

& Chaudhari, Akshay. 2024. Adapted large language models can outperform medical experts in clinical text summarization. *Nature Medicine*.

Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan; Kaiser, Lukasz & Polosukhin, Illia. 2017. Attention Is All You Need. arXiv:1706.03762.

Vehko, Tuulikki; Hyppönen, Hannele; Ryhänen, Miia; Tuukkanen, Johanna; Ketola, Eeva & Heponiemi, Tarja. 2018. Tietojärjestelmät ja työhyvinvointi – terveydenhuollon ammattilaisten näkemyksiä. *Finnish Journal of EHealth and EWelfare*.

Vitec Acute. 2025. Verkkoaineisto. Potilastietojärjestelmän rooli terveydenhuollossa. <<https://www.vitec-acute.com/ratkaisut/acute-potilastietojarjestelma/>>. Luettu 30.6.2025.

Zhao, Wayne; Zhou, Kun; Li, Junyi; Tang, Tianyi; Wang, Xiaolei; Hou, Yupeng; Min, Yingqian; Zhang, Beichen; Zhang, Junjie; Dong, Zican; Du, Yifan; Yang, Chen; Chen, Yushuo; Chen, Zhipeng; Jiang, Jinhao; Ren, Ruiyang; Li, Yifan; Tang, Xinyu; Liu, Zikang; Liu, Peiyu; Nie, Jian-Yun & Wen, Ji-Rong. 2025. A Survey of Large Language Models. arXiv:2303.18223.

Yang, Jingfeng; Jin, Hongye; Tang, Ruixiang; Han, Xiaotian; Feng, Qizhang; Jiang, Haoming; Yin, Bing & Hu, Xia. 2023. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. arXiv:2304.13712.

Yang, Xiongwen; Xiao, Yi; Liu, Di; Zhang, Yun; Deng, Huiyin; Huang, Jian; Shi, Huiyou; Liu, Dan; Liang, Maoli; Jin, Xing; Sun, Yongpan; Yao, Jing; Zhou, XiaoJiang; Guo, Wankai; He, Yang; Tang, WeiJuan & Xu, Chuan. 2025. Enhancing doctor-patient communication using large language models for pathology report interpretation. *BMC Medical Informatics and Decision Making*. BMC Medical Informatics and Decision Making.

Artikkelien yhteenveto

Teema	Kielimalli	Keskeiset huomiot	Lähde
Ennaltaehkäisy	Mixtral 8x7B	Suuri kielimalli pystyi ennustamaan potilaan verenmyrkytyksen 63,9 % suorituskyvällä. (F1-score).	Shashikumar ym. 2025
	GPT-4 GPT-3.5	Tutkimuksessa esitellään uusi menetelmä, jossa kaksi kielimalliagenttia tekee yhteistyötä sairauksien ennustamiseksi. Tämä yhteistyöhön perustuva lähestymistapa paransi ennustuksen tarkkuutta.	Cui ym. 2025
	GPT-4 GPT-3.5 PaLM 2	Tutkimus osoitti, että suurten kielimallien avulla on mahdollista poimia hoitosuosituksia ja niihin liittyviä riskikriteereitä verkkolähteestä potilastietojärjestelmään. Kun riskikriteerit täyttyvät potilaan kohdalla, voidaan ammattilaiselle antaa hehärete tarvittavista toimenpiteistä.	Luo ym. 2024
Potilasviestintä	GPT-4	Potilaiden oli helpompi ymmärtää kielimallin selkeyttämiä patologian lausuntoja kuin alkuperäisiä lääkärin lausuntoja. Asteikolla 1–10 kielimallin selkeyttämien lausuntojen pisteet olivat keskimäärin 7,98, kun taas alkuperäisten lausuntojen	Yang ym. 2025

		5,23. Lääkärin ja potilaan välinen keskustelu-aika lyheni yli 70 %, koska kielimallin selkeyttämän lausunnon läpikäynti oli tehokkaampaa.	
	GPT-4 GPT-3.5 CLAIR-Short (LLaMA-65B) CLAIR-Long (LLaMA-65B)	Suuri kielimalli voi pystyä vastaamaan potilaiden potilasportaalin kautta esittämiin kysymyksiin empaattisemmin ja ymmärrettävimminkin lääkärit. Hienosäätämällä kielimallia vastaukset saadaan muistuttamaan lääkärin kirjoittamia vastauksia.	Liu ym. 2024
	GPT-4	Kehotemuotoilun avulla saatiin kielimallin vastausluonnoksien laatua parannettua, mutta sillä ei kuitenkaan ollut myönteistä vaikutusta vastausluonnoksien käyttöasteeseen.	Afshar ym. 2024
Yhteenvedot	FLAN-T5 FLAN-UL2 Alpaca Med-Alpaca Vicuna Llama-2 GPT-3.5 GPT-4	Kielimallin tuottamia yhteenvetoja potilastiedoista pidettiin usein täydellisempinä, virheettömämpinä ja ytimekkäämpinä kuin asiantuntijoiden tekemiä. Kielimalli mahdollisesti voisi vähentää terveydenhuollon ammattilaisten dokumentointitaakkaa.	Van Veen ym. 2024

	<p>BART-LSG-conv LongDialSumm T5-BookSum Bart-DnC GPT-3.5</p>	<p>ChatGPT tuotti laadukkaampia yhteenvetoja potilastiedoista kuin hienosäädetyt kielimallit tai lääkärit. Perinteiset kielimallien arviointimetriikat (esim. ROUGE) eivät välttämättä toimi lääketieteellisten sovelluksien arvioinnissa.</p>	<p>Fraile Navarro ym. 2025</p>
	<p>Flan-T5 BART Pegasus-xsum</p>	<p>Pehmeiden kehoitteiden avulla voidaan merkittävästi vähentää kielimallin luomien yhteenvetojen laadun vaihtelevuutta.</p>	<p>Chuang ym. 2024</p>
<p>Tietojen poiminta vapaasti kirjatusta tekstistä</p>	<p>Claude 3.0 Opus Claude 3.0 Sonnet Claude 2.0 GPT-4 Claude 2.1 Gemini Advanced PaLM 2 chat bison Llama 3-70b Gemini GPT-3.5 Claude 1.0 Llama 2-70b chat Llama 3-8b Claude 3.0 Haiku</p>	<p>Tutkimus osoitti, että osa suurista kielimalleista pystyy poimimaan halutun tiedon strukturoimattoman tekstin joukosta.</p>	<p>Ntinopoulos ym. (2025)</p>

	<p>Llama 2-13b chat</p> <p>RoBERTa base</p> <p>SQuAD</p> <p>Cohere command</p> <p>Llama 2-7b chat</p> <p>Mistral 7b instruct</p>		
	<p>GPT-4</p> <p>Llama 3</p> <p>Llama 3.1</p> <p>Gemma 2</p> <p>Qwen 2</p> <p>Qwen 2.5</p>	<p>Kielimalli pystyy poimimaan halutun tiedon vapaan tekstin joukosta suurella tarkkuudella, mutta anonymisoidun tiedon kanssa suorituskyky laskee.</p>	<p>Garcia-Carmona ym. 2025</p>
	<p>openchat_3.5</p> <p>zephyr-7b</p> <p>vicuna-7b</p> <p>Llama-2-7b</p> <p>vicuna-13b</p> <p>WizardLM-13b</p> <p>Llama-2-13b</p> <p>vicuna-33b</p>	<p>Kehotesuunnittelun avulla kielimalli saadaan hallusinoimaan vähemmän tietojen poiminnan yhteydessä, mutta samalla vähentää kielimallin poiminnan tarkkuutta.</p>	<p>Gu ym. 2025</p>