

Generatiivinen tekoäly fysiikan summatiivisen  
arvioinnin tehtävien tuottamisen työvälineenä

Pro Gradu  
Turun yliopisto  
Fysiikka  
2026  
Simeon Turpeinen  
Tarkastajat:  
FT Tommi Kokkonen  
FT Veli-Matti Vesterinen

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä

TURUN YLIOPISTO

Fysiikan laitos

**Turpeinen, Simeon** Generatiivinen tekoäly fysiikan summatiivisen arvioinnin tehtävien tuottamisen työvälteenä

Pro Gradu, 78 s.

Fysiikka

Toukokuu 2026

---

Generatiivisen tekoälyn yleistymisen suurten kielimallien kehityksen myötä on vaikuttanut myös fysiikan opetukseen. Tutkielmassa keskityttiin erityisesti generatiivisen tekoälyn mahdollisuuksiin tuottaa fysiikan vuosiluokkien 7–9 summatiiviseen arviointiin soveltuvia tehtäviä. Tutkimuskysymykset keskittyvät kielimalleilla tuotettujen tehtävien lisäksi tehtäviä generoitaessa kielimallille annettaviin kehoitteisiin ja kehotteen laatimisen prosessin vaiheisiin.

Tutkielman teoreettinen viitekehys keskittyy arvioinnin teorian lisäksi arvioinnin ja tehtävien analysointiin liittyvään kirjallisuuteen. Näiden perusteella tutkielman aineistona oleville tehtäville kehitettiin arviointikriteeristö. Arvioinnin laadun osalta keskeiseksi ohjaavaksi käsitteeksi tunnistettiin sisältövaliditeetti, jonka kautta tehtävien laadun kuvailua ja arviointia osaltaan voitiin lähestyä. Lisäksi perehdyttiin tutkielman kannalta keskeiseen generatiivisen tekoälyn opetuskäyttöön ja potentiaalisiin hyödyntämismahdollisuuksiin.

Tutkielman aineisto koostui fysiikan summatiiviseen arviointiin tarkoitetuista tehtävistä, joista osa on haettu oppikirjoja julkaisevien kustantamojen kirjasarjojen tehtäväpankeista. Toinen osa aineistosta tuotettiin laatimalla kirjallisuuteen pohjautuen kielimallille kehote, jolla tuotettiin fysiikan summatiiviseen arviointiin soveltuvia tehtäviä samoin sisältöalueen rajauksin kuin tehtäviä tehtäväpankeista valittaessa. Generoitujen tehtävien ominaisuuksien perusteella kehoitteelle tehtyjen muokkausten jälkeen kehotteen voitiin todeta tuottaneen riittävän hyvän lopputuloksen. Kokonaan tehtäväaineisto analysoitiin tämän jälkeen laadullisen sisällönanalyysin keinoin teoriaohjaavaa lähestymistapaa käyttäen.

Päätuloksina voidaan pitää aineiston generoinnin yhteydessä dokumentoitua kehotteen laadinnan ja tehtävien kielimallia hyödyntäen tuottamisen prosessia sekä sisällönanalyysin osoittamaa laadun vaihtelua generatiivisen tekoälyn tuottamien tehtävien välillä. Kielimallin tuottamien tehtävien havaittiin olevan laadultaan vaihtelevia, mutta myös käytettyjä tehtäväpankin tehtäviä osin monipuolisempia ja haastavampia. Tulosten perusteella erityinen mielenkiinto kohdistuu tulevaisuuden osalta tehtävien kielimalleilla tuottamisen ja laadun varmistamisen toteutumiseen kouluympäristössä.

Asiasanat: fysiikan opetus, arviointi, generatiivinen tekoäly, kehote

# Sisällys

<b>Johdanto</b>	<b>1</b>
<b>1 Teoreettinen viitekehys</b>	<b>3</b>
1.1 Arviointi . . . . .	3
1.1.1 Arvioinnin tehtävät . . . . .	3
1.1.2 Summatiivinen ja formatiivinen arviointi . . . . .	5
1.1.3 Arviointi fysiikan opetuksessa . . . . .	7
1.2 Arvioinnin analysoinnin välineitä . . . . .	8
1.2.1 Validiteetti . . . . .	8
1.2.2 Bloomin taksonomia . . . . .	10
1.2.3 Muita arvioinnin analysoinnin tapoja . . . . .	12
1.3 Tehtävien arviointikriteerit . . . . .	13
1.3.1 Tehtävien luokittelu kirjallisuudessa . . . . .	15
1.3.2 Soveltuva kriteeristö . . . . .	17
1.4 Tekoälyn käyttö opetuksessa . . . . .	20
1.4.1 Koneoppiminen ja tekoäly . . . . .	22
1.4.2 Suuret kielimallit ja generatiivinen tekoäly . . . . .	24
1.4.3 Tekoälyn käyttökohteita opetuksessa . . . . .	25
1.4.4 Tekoäly fysiikan opetuksessa . . . . .	27
<b>2 Menetelmät</b>	<b>31</b>
2.1 Aineisto . . . . .	31
2.1.1 Tehtäväpankin tehtävät . . . . .	33
2.1.2 Kehotteen laatiminen . . . . .	35
2.1.3 Generoidut tehtävät . . . . .	40
2.2 Sisällönanalyysi . . . . .	51
2.2.1 Valittu menetelmä . . . . .	51

2.2.2	Laadittujen kriteerien soveltaminen . . . . .	53
2.2.3	Kognitiivinen vaativuustaso . . . . .	54
2.2.4	Kokeellisen menetelmän tuntemus . . . . .	57
2.2.5	Fysiikan oikeellisuus . . . . .	58
2.2.6	Selkeys . . . . .	60
<b>3</b>	<b>Tulokset ja pohdinta</b>	<b>64</b>
3.1	Ensimmäinen tutkimuskysymys . . . . .	64
3.2	Toinen tutkimuskysymys . . . . .	68
3.3	Tulokset opettajan työn näkökulmasta . . . . .	68
<b>4</b>	<b>Johtopäätökset</b>	<b>71</b>

## Johdanto

Generatiivisen tekoälyn nousu laajemman yleisöön tietouteen on tapahtunut viime vuosina huomattavan nopeasti. Laajojen kielimallien, kuten OpenAI:n ChatGPT:n tai Microsoftin Copilotin, käytön yleistymisen vaikuttaa väistämättä myös kouluopetuksen järjestämiseen. Kouluissa opettajien tekoälyosaaminen vaikuttaa ainakin jossain määrin kulkevan oppilaiden tekoälyosaamisen perässä, jolloin oppilailla on ajoittain opettajia paremmat valmiudet hyödyntää tekoälyä myös tavalla, joka ei suoraan tue oppimista [1]. Yeadon ym. [2] ennakoivat jo varhaisemmassa vaiheessa generatiivisen tekoälyn mullistusten vaikuttavan merkittävästi opetuksen järjestämiseen, sillä esimerkiksi itsenäisesti kirjoitettavien lyhyiden esseevastausten laadukas tuottaminen generatiivisella tekoälyllä mahdollistaa arvioinnin vääristymisen. Vastausten tuottamisen lisäksi generatiivista tekoälyä voidaan hyödyntää opettajan arviointityön tukena, jolloin opettajan vastuun korostuessa painottuvat myös tekoälyn vastuullinen käyttö ja tarve ylläpitää korkeita arvioinnin standardeja.

Suuret kielimallit tuottavat uudenlaisia haasteita opetuksen järjestämiselle, mutta avaavat myös useita mahdollisuuksia opetuksen kehittämiseen tekoälyä harkitus- ti hyödyntäen [3]. Mahdollisuuksien realisoituminen edellyttää kuitenkin työkalujen riittävän helppoa käytettävyyttä ja toimivaa lopputulosta. Kielimallien kehityksen ollessa nopeaa ja kielimallien tuottamien mahdollisuuksien ollessa laajoja [3], myös niihin liittyvän opetuksen tutkimuksen kenttä on tutkimusaiheiltaan laaja. Summatiivisen arvioinnin tehtävien tekoälyavusteisen tuottamisen tärkeä edellytys on laadukas kielimallille annettava kehote ja yleisesti riittävä tekoälyosaaminen [4, 5]. Tutkielman eräs painopiste tulee siis olemaan tehtävän generointitarkoituksessa kielimallille annettavan kehotteen kehittäminen ja prosessin dokumentointi. Generatiivisen tekoälyn käyttökelpoisuuteen opettajan arviointityön tukena tehtävien laatimisessa voi osaltaan vaikuttaa tehtävien generoinnin ja kehotteen laatimisen haastavuus.

Summatiivisessa arvioinnissa generatiivisen tekoölyn hyödyntämisen mahdollisuuksia on tutkittu monivalintatestien kysymysten laatimisessa lääketieteen opetuksessa [6, 7] mikä herättää kysymyksen siitä, voidaanko generatiivista tekoölyä tehokkaasti hyödyntää myös laajemmin summatiivisessa arvioinnissa tehtävien tuottamisen tukena. Oppikirjoja julkaisevien kustantajien tuottamat koetehtäväpankit ovat laajasti opettajien käytettävissä summatiivisen arvioinnin tehtäviä valittaessa ja laadittaessa, jolloin kokeen laadintaan käytettävä opettajan työaika oletusarvoisesti lyhenee. Käyttötarkoitusten samankaltaisuuden takia tehtäväpankkien tehtäviä voidaan pitää generatiivisella tekoölyllä tuotettuja summatiivisen arvioinnin tehtäviä vastaavana opettajan lisäresurssina.

Arvioinnin merkittävien tehtävien [8] ja opettajan työtehtäviin liittyvän ajanhallinnan kehittämisen vuoksi generatiivisen tekoölyn käyttöä fysiikan summatiivisessa arvioinnissa on tarpeen arvioida. Tutkielman tavoitteena on tarkastella ChatGPT:n hyödyntämismahdollisuuksia fysiikan summatiivisen arvioinnin tehtävien laadinnassa. Tavoitetta lähestytään analysoimalla kielimallilla tuotettuja summatiiviseen arviointiin tarkoitettuja tehtäviä ja verrokiksi valittuja oppikirjasarjojen tehtäväpankkien tehtäviä.

Tutkielman tutkimuskysymykset ovat:

1. Millaisia fysiikan summatiivisen arvioinnin tehtäviä voidaan laatia generatiivisella tekoölyllä lyhyttä kehotetta käyttäen?
2. Millaisen prosessin kautta generatiivisella tekoölyllä voidaan kehittää tehtäviä fysiikan arvioinnin tarpeisiin?

# 1 Teoreettinen viitekehys

## 1.1 Arviointi

Tekoälyä hyödyntävän arvioinnin tarkastelulle pohjan asettaa olemassa oleva arviointia koskeva tutkimustieto. Generatiivinen tekoäly voi vaikuttaa merkittävästi arvioinnissa yleisesti käytettävien työkalujen pitävyyteen, kuten kotona kirjoitettavien esseiden käyttöön arvioinnin tukena [2]. Arvioinnin tuntemus auttaa ehkäisemään tekoälyn tuottamia haasteita arvioinnin uskottavuudelle, mutta myös kehittämään arviointia tekoälyä hyödyntäen.

Perusopetuslaissa nostetaan esille opetuksen järjestäjän velvoitteeksi arvioida oppilaan lisäksi myös koulutusta kokonaisuutena (Perusopetuslaki 1998/628 § 21). Koulutuksen arvioinnin kolmena päätehtävänä mainitaan oppimisolosuhteiden parantamisen ja opetuksen kehittämistehtävän lisäksi opetuksen lainmukaisuuden kontrollointi. Suomalaisessa peruskoulussa arviointi voidaan nähdä perusopetuslain näkökulmasta tarkasteltuna koko peruskoululaitoksen läpäisevänä ominaisuutena, jossa arvioitavan roolissa on paitsi oppija, myös koululaitos itsessään.

### 1.1.1 Arvioinnin tehtävät

Haney [8] nostaa arvioinnin kolmeksi päätehtäväksi tilivelvollisuuden koulun ulkopuolisille tahoille, tiedon tuottamisen opettajille opetuksen kehittämistä varten ja oppijoiden oppimisen tukemisen. Black [9] tulkitsee Haneyn esittämiä kolmea päätehtävää toisin ja laskee oppimisen tuen lisäksi kahdeksi muuksi tehtäväksi yksittäisten opiskelijoiden sertifiointin ja oppilaitoksen opettajineen julkisen tilivelvollisuuden. Erona näillä tehtävien luokittelulla on formatiivisen arvioinnin painoarvo. Blackin ja Wiliamin [10] käyttämän formatiivisen arvioinnin määritelmän mukaisesti oppimisen tukeminen ja opetuksen kehittäminen arvioinnin perusteella voidaan lukea formatiiviseen arviointiin kuuluviksi, kun taas sertifiointiin ja tilivelvollisuu-

teen liittyvät tehtävät ovat luonteeltaan vahvemmin summatiivisia. Formatiivisen arvioinnin tehtävänä voidaan yleisesti pitää oppimista ja informaation tuottamista oppimisen tueksi [10]. Verrattuna arvioinnin yleisiin päätehtäviin [9] havaitaan, ettei formatiiviseen arviointiin yleisesti sisälly tilivelvollisuutta tai tarvetta raportoida arvioinnin tuloksista ulkopuoliselle taholle.

Summatiivisen arvioinnin keskeisenä tehtävänä on esittää oppijan osaaminen suhteessa opetussuunnitelmaan tai muuhun oppimisen tavoitteita sääntelevään ohjeeseen. Opettajan rooli summatiivisessa arvioinnissa on merkittävä, eikä opetussuunnitelman ja muun opetuksen ja arvioinnin vaikutusta summatiiviseen arviointiin voida helposti sivuuttaa. Koska summatiivinen arviointi kytkeytyy tiiviisti koko opetuksen kokonaisuuteen, tulee summatiivisen arvioinnin opetussuunnitelman mukaisessa toteutuksessa huomioida myös opetuksen ja siihen tiiviisti kuuluvan formatiivisen arvioinnin linja suhteessa noudatettavaan opetussuunnitelmaan. [11] Summatiiviseen arviointiin valittavia tai laadittavia tehtäviä ei siis voida käsitellä opetuksen laajemmasta kokonaisuudesta erillään, vaan tehtävien lopullisen käyttökelpoisuuden tai soveltuvuuden arvioinnin tekee oman opetuksensa luonteen parhaiten tunteva, opetusta ja arviointia kulloinkin toteuttava opettaja.

Suomessa noudatettavan peruskoulun opetussuunnitelman perusteiden oppiaineille asetetut oppimistavoitteet ohjaavat arviointityötä ja arviointikriteerien asettamista. Luostarisen ja Ouakrim-Soivion [12] mukaan opetussuunnitelman asettamien tavoitteiden ja kriteerien välinen suhde on selvä. Tavoitteet kuvaavat tarkemmin, mitä oppilaan odotetaan oppivan, kun taas kriteerit tarjoavat muodoltaan tiiviimmän tavan mitata oppilaan osaamista arviointia varten. Summatiivista arviointia ohjaavien kriteerien tulisi siis selkeästi perustua opetussuunnitelman asettamiin tavoiteisiin, jolloin kriteerien ja tavoitteiden suhde on selkeä ja arviointi johdonmukaista tavoitteiden pohjalta toteutettuun opetukseen nähden. Arvioinnin kohdistumisen ja opetukselle asetettujen tavoitteiden suhdetta voidaan kuvata validiteetin käsitteen

avulla [13].

### 1.1.2 Summatiivinen ja formatiivinen arviointi

Suomalaista perusopetusta ohjaavan Perusopetuksen opetussuunnitelman perusteiden [14] mukaan formatiivinen arviointi tukee oppilasta opetuksen tavoitteiden saavuttamisessa ja antaa oppilaalle itselleen informaatiota oppimisprosessista. Erityisesti mainitaan, ettei formatiivisen arvioinnin erityiselle dokumentoinnille ole tarvetta. Näin tulee korostetuksi formatiivisen arvioinnin tehtävä oppilaan oppimisen välineenä, mikä kiteytetään Perusopetuksen opetussuunnitelman perusteissa: "*Arvioinnin tehtävänä on ohjata ja kannustaa opiskelua sekä kehittää oppilaiden itsearvioinnin taitoja (formatiivinen arviointi)*" [15].

Formatiivinen ja summatiivinen arviointi sekoittuvat helposti keskenään jopa niin, että formatiivisina pidetään useita minitestejä, jotka todellisuudessa ovat luonteeltaan summatiivisia. Formatiiiviselle arvioinnille ominaista on sen kytkeytyminen opetukseen ja oppijoiden osaamisen tasoon, eikä se ole summatiivisen arvioinnin tavoin kriteereihin voimakkaasti kiinnittynyttä. Formatiiivisen arvioinnin voidaan todeta olevan summatiivista arviointia huomattavasti kiinteämpi osa opetusta itsessään, sillä formatiiivisen arvioinnin tuottaman palautteen tehtävänä on tukea oppimista. [16]

Black ja Wiliam [10] rajaavatkin formatiiivisen arvioinnin niin, ettei pelkkä arvioinnin kautta saavutettu tieto oppijan osaamistasosta tee arvioinnista formatiiivista. Arviointi muuttuu siis formatiiiviseksi, kun arvioinnin kautta saavutettu tieto tukee oppijan osaamisen kehittymistä. Formatiiiviseen arviointiin voi kuulua myös opettajalle oppijoiden osaamisen tasosta tietoa tuottavia testejä. Tällöin oppijan oppimisen kehitys nojaa oppijan omaa arviota vahvemmin opettajan valitsemiin toimenpiteisiin, jotka opettaja valitsee testitulosten pohjalta [17]. Formatiiivista arviointia voidaan näin toteutettuna käyttää oppilaan arvioinnin lisäksi perusopetus-

lain mukaiseen koulutuksen arviointiin.

Summatiivisen arvioinnin tyypillisinä piirteinä pidetään kriteeripohjaisuutta, koko arvioitavan osa-alueen kattavuutta ja säännönmukaista tuloksista raportointia. Erona formatiiviseen arviointiin kriteeripohjaisuus merkitsee arvioinnin nojaavan yleisiin kriteereihin, jolloin eri oppijoiden suoritukset ovat keskenään vertailukelpoisia tietyn suuremman viiteryhmän sisällä [16]. Yleisiä kriteerejä summatiiviselle arvioinnille voivat asettaa opetusta ohjaavat säännökset ja asiakirjat, kuten Perusopetuksen opetussuunnitelman perusteet [14] tai paikalliset opetussuunnitelmat.

Summatiivisen ja formatiivisen arvioinnin toisistaan erottamiseen Dixson ja Worrell [17] toteavat luotettavimman ja ainoan tavan olevan arvioinnin perimmäisen käyttötarkoituksen tulkitseminen. Tyypillisesti summatiivista arvioinnin välinettä, kuten laajempaa koetta, voidaan käyttää formatiivisen arvioinnin välineenä, jos kokeen perusteella opettaja selvittää oppijoiden osaamisen tason laajasta aihepiiristä ja mukauttaa opetustaan tulosten pohjalta. Summatiivisen arvioinnin yhteydessä esiintyy lisäksi lähes poikkeuksetta tulosten arviointia arvosanoin, ja yksittäisen summatiivisen arviointisuorituksen vaikutus oppijan elämään on merkittävä [17]. Vastaavasti formatiivisena pidettyä arvioinnin välinettä, kuten itsearviointia, voidaan käyttää arvosanan muodostamiseen ja siten osana summatiivista arviointia [17].

Summatiivisen arvioinnin käytännön toteutustapana on usein koe, joka voi olla loppukoe, lukukauden tai laajemman opintojakson päättävä koe, pääsykoe tai valtakunnallinen koe, mutta myös käytännön osaamista mittaava summatiivinen arviointi on tavanomaista [17]. Kokeiden avulla tehtävä summatiivinen arviointi voidaan osin tai kokonaan korvata opettajan tekemiin havaintoihin perustuvalla summatiivisella arvioinnilla, jossa opettajan ennakoasenteista tai virheellisestä oppijan osaamisen tunnistamisesta seuraavat puutteet ovat yleisiä [18]. Formatiivisen arvioinnin toteutustavoiksi Black ja Wiliam [10] mainitsevat itse- ja vertaisarvioinnin, keskustelun,

testit ja palautteen, joiden avulla hyvin toteutettu arviointi tukee oppimista ja tuottaa opettajalle tietoa opetuksen kohdistamisen tueksi.

### 1.1.3 Arviointi fysiikan opetuksessa

Buick [19] korostaa fysiikan arvioinnissa tiedollisen osaamisen mittaamisen lisäksi taitojen ja myöhemmissä opinnoissa tai työelämässä tarvittavan osaamisen arviointia. Taitojen arvioinnin apuvälineinä nostetaan esiin luonteeltaan vahvemmin formatiiviset tehtävät, projektit ja laboratoriotyöskentely. Fysiikan opetukselle tyypillisen laboratoriotyöskentelyn arviointi opetuksen ohessa tai osana summatiivista arviointia voidaan toteuttaa opettajan havainnointiin perustuvana käytännön työskentelytaitojen arviointina [17].

Fysiikan summatiivista arviointia toteutettaessa on huomioitava tehtävien kielellinen luonne. Fysiikan tehtäville tyypillisten kielellisten ominaispiirteiden tunnistaminen on tärkeää tehtävän tarkoituksenmukaisen ratkaisemisen kannalta, sillä esimerkiksi komentoverbien käyttö poikkeaa muista kielellisistä konteksteista fysiikan tehtäviä ratkaistaessa. Tehtävänannossa käytettävän kehottavan verbin merkityksen ja oppijalta oletetun ratkaisun suhteen ymmärtäminen on osa fysiikan oppimista, mitä voidaan kielitietoisella lähestymisellä tukea. [20]

Fysiikan summatiivisen arvioinnin osalta virhekäsitysten ja käsitteellisen osaamisen tason arviointi on Hestenesin, Wellsin ja Swackhamerin [21] mukaan keskeinen osa laadukasta fysiikan summatiivista arviointia. Osin juuri käsitteellisen osaamisen ja virhekäsitysten tunnistamiseen keskittyvät monivalintatestit, kuten Hestenesin ym. [21] tutkimuksen *Force concept inventory*, ovat käyttökelpoisia fysiikan arvioinnin työkaluja. Laskutehtävät voivat kuitenkin erikoistettujen monivalintatehtävien tavoin antaa tietoa oppijan osaamisesta ja virhekäsityksistä, kunhan tehtävän suunnittelussa huolehditaan siitä, että tehtävä mittaa lähes pelkästään mitattavaa yksittäistä osaamisaluetta [20].

## 1.2 Arvioinnin analysoinnin välineitä

Arvioinnin pohjautuessa määrätyille kriteereille on arvioinnin tulokseen vaikuttavat tekijät tunnistettava mahdollisimman hyvin. Kriteeriperustaisesta arvioinnista voidaan esittää kriittisiä huomioita erityisesti arvioinnin luotettavuuden ja tieteellisyyden osalta, kuten Yorke [22] korostaa. Arvioinnin oikeudenmukaisuudesta Gipps ja Stobart [23] taas toteavat aiemman kirjallisuuden pohjalta, etteivät usein objektiiviksi mielletyt monivalintatestit ja muut helposti pisteytettävät testit ole todellisuudessa objektiivisiä, vaan kysymysten muotoilun ja laadinnan subjektiivisuus aiheuttaa haasteita arvioinnin oikeudenmukaisuudelle. Samoin esseevastauksien kaltaiset luonteeltaan avoimemmat arviointivälineet johtavat usein epäoikeudenmukaiseen arviointiin, mitä aiheuttaa puutteellisen pisteytyksen lisäksi subjektiivisuus tehtävän laadinnassa ja tehtävänannossa [23].

Tehtävien laadinnan subjektiivisuuden tunnistaminen tehtäväpankkien tehtäviä tai kielimallin tuottamia tehtäviä käytettäessä voi olla haastavaa. Kumpaakin tilannetta yhdistää ulkopuolinen varsinainen tehtävän laatija. Oikeudenmukaista ja laadukasta arvioinnin toteutusta tukee arvioinnissa käytettävien tehtävien ja arvioinnin kokonaisuuksien analyttinen tarkastelu.

### 1.2.1 Validiteetti

Arvioinnin validiteetin (*validity*) avulla voidaan kuvata arvioinnin laatua. Erityisesti käsitevaliditeettia (*construct validity*) käytetään yleisesti arvioinnin pätevyyttä tarkasteltaessa. Messickin [24] mukaan käsitevaliditeetti on yhtenäinen kokonaisuus, jota arvioidaan eri osa-alueiden kautta. Käsitevaliditeetin tarkastelun voidaan katsoa edellyttävän arvioinnin suorittamista ja arvioinnin tarkastelua pisteytetty tulos huomioiden [23]. Suoraan tulkittuna käsitevaliditeetin arviointi ei siis ole mahdollista ennen arvioinnin toteuttamista ja pisteytystä. Arvioinnissa käytettävien tehtävien, kokonaisten kokeiden ja testien laatimiselle käsitevaliditeettiin nojaten tämä aset-

taa haasteita, jotka on rajoitukset tiedostaen kierrettävä ensimmäiselläkin kerralla käyttökelpoisen arvioinnin saavuttamiseksi. Crooks, Kane ja Cohen [25] jakavat arvioinnin validiteetin tarkastelun kahdeksaan eri vaiheeseen, joista jokaisessa voidaan kriittisesti tarkastella validiteettiin vaikuttavia tekijöitä.

Validiteettia voidaan tehtävän tai tutkimuksen osalta tarkastella lukuisista toisistaan poikkeavista näkökulmista. Validiteeteista soveltuvin ennen tehtävän käyttöä arviointitarkoitukseen on sisältövaliditeetti (*content validity*). Messick [13] kuvaa sisältövaliditeetin oleellisimman sisällön jakautuvan kahteen osaan: sisällön relevanssiin ja sisällön kattavuuteen. Sisällön relevanssilla tarkoitetaan tehtävän sisällön vastaavuutta sen sisällön kanssa, jonka osaamista tehtävällä halutaan arvioida. Jos tarkoitus on esimerkiksi arvioida Ohmin lain soveltamisen osaamista, on tehtävä sisältövaliditeetin näkökulmasta relevantti, kun tehtävässä aidosti joudutaan soveltamaan Ohmin lakia, eivätkä kiertotiet tehtävän ratkaisemiseksi ilman Ohmin lain soveltamista ole mahdollisia. Sisällön kattavuuden arviointi kohdistuu selkeämmin koko testin tai kokeen kokonaisuuteen. Messickin [13] mukaan sisällön kattavuus voidaan nähdä eritoten sisältöjen ja tehtävän edellyttämien kognitiivisten tasojen suhteen. Tällöin tyypillistä yksittäistä summatiivisessa arvioinnissa käytettävää tehtävää käsiteltäessä ei voida suoraan lausua mitään tehtävän sisältövaliditeetista sisällön kattavuuden osalta, vaan sisältövaliditeetin tarkastelu joudutaan rajoittamaan sisällön relevanssin tarkasteluun. Huomionarvoista on, että Messickin [13] mukaan sisältövaliditeettia ei voida käsitellä kokonaan irrallaan käsitevaliditeetista, sillä käsitevaliditeetti on validiteetin arvioinnin kokonaisuudessa kytköksissä useisiin muihin validiteetin osa-alueisiin.

Sisältövaliditeetin arvioinnin eräs tehtävä on punnita arvioinnin sisältöä suhteessa siihen, mitä pyritään mittaamaan [26]. Yksittäisen koetehtävän osalta voidaan tarkastella, mitä opetussuunnitelmassa asetettua tavoitetta tai tavoitteita pyritään mittaamaan ja toisaalta sitä, kuinka hyvin tehtävä sisällöllisesti vastaa arvioinnin

kohteen osaamista käsiteltävän tavoitteen osalta. Ouakrim-Soivio [26] kuvaa myös Kanen [27] validiteetin ja testin validointiprosessin käsittelyn pohjalta validoinnin jakautumista Kanen teorian pohjalta kahteen laajempaan vaiheeseen, kehittämisvaiheeseen ja arviointivaiheeseen. Kun kehitetään tehtäviä arviointitarkoitukseen ilman tehtävien teettämistä oppijoilla, voidaan arvioida validointiprosessin rajoituvan mallin mukaisesti kehittämisvaiheeseen, vaikka kyseessä ei olekaan teorian mukainen testikokonaisuus, vaan yksittäinen tehtävä. Sisältövaliditeetin arviointiprosessi edellyttää validiteettia arvioivia asiantuntijoita [28, 29] eikä erilaisten sisältövaliditeetin arviointiin kehitettyjen kvantitatiivisten työkalujen käyttö voi korvata asiantuntijoiden huolellista aineiston kvalitatiivista analysointia. Toisaalta Messickin [13] määritelmän mukainen validiteetin tarkastelu ei ole pelkistettävissä ainoastaan asiantuntija-arvioksi, vaan teoreettista taustaa sekä opetuksen ja arvioinnin kokonaisuutta käsittelevä arvio on asiantuntijuudesta riippumatta tarpeen.

Arvioinnin validiteettia tarkasteltaessa on siis pohjimmiltaan kyse arvioinnin oikein kohdistumisen arvioinnista. Arviointi kohdistuu validiteetiltaan korkeassa summatiivisessa arvioinnissa niihin tavoitteisiin, joita on tarkoitus arvioida, kun muiden seikkojen vaikutus tehtävään jää mahdollisimman vähäiseksi. Fysiikan tehtävän ratkaisemisessa tarvitaan lähtökohtaisesti muutakin fysiikan osaamista, mutta tämän osaamisen osuus tehtävässä ei saisi nousta häiritsemään arvioitavaa tavoitetta. Validiteetiltaan hyvälaatuinen tehtävä keskittyy siis arvioimaan pelkästään tarkastelun kohteena olevan sisältöalueen osaamista, kuten Fischer ja Kauertz [20] laadukasta fysiikan arviointiin käytettyä tehtävää kuvaavat.

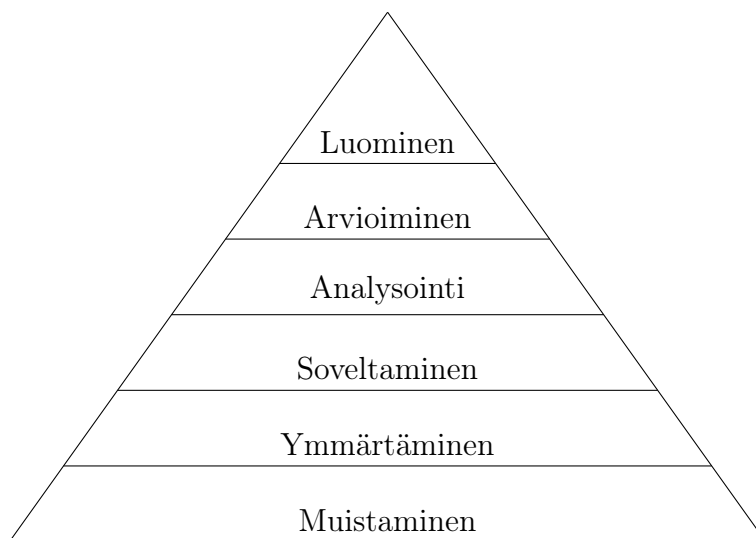
### 1.2.2 Bloomin taksonomia

Opetuksen ja arvioinnin tukena käytetään yleisesti Bloomin taksonomiaa [30], jonka avulla tehtäviä ja tavoitteita voidaan luokitella esimerkiksi tehtävien suorittamiseen vaadittujen ajattelun tasojen avulla. Alkuperäisessä versiossa ajattelun tasoiksi näh-

tiin matalimmasta korkeimpaan tietäminen (*knowledge*), ymmärtäminen (*comprehension*), soveltaminen (*application*), analysointi (*analysis*), syntetisointi (*synthesis*) ja arvioiminen (*evaluation*). Bloomin taksonomiaa käyttäen tehtäviä voidaan helposti ryhmitellä vaativuusryhmiin tehtävässä oletettavasti vaadittavan ajattelun tason mukaan, mikä on hyvä lähtökohta tehtävien kuvailulle ja analysoinnille. Yksin Bloomin taksonomian kautta suoritettu tehtävien tarkastelu ja ryhmittely ei suoraan kuvaa tehtävien kohdistumista haluttuihin tavoitteisiin tai sisältöihin, vaan tehtävien sisällön tarkempi analyysi muita keinoja käyttäen on tarpeen tehtävien laadun selvittämiseksi.

Alkuperäistä Bloomin taksonomiaa yleisemmin käytössä on nykyisin uudistettu Bloomin taksonomia [31], jossa ajattelun tasoiksi luokitellaan muistaminen (*remember*), ymmärtäminen (*understand*), soveltaminen (*apply*), analysointi (*analyze*), arvioiminen (*evaluate*) ja luominen (*create*). Kuvassa 1 esitetyn uudistetun Bloomin taksonomian luokka 'muistaminen' voidaan sisällöltään nähdä vastaavaksi kuin alkuperäisen taksonomian luokka 'tietäminen', samoin ymmärtämisen, soveltamisen ja analysoinnin luokat vastaavat alkuperäisessä ja uudistetussa taksonomiassa toisiaan. Uudistetun ja alkuperäisen Bloomin taksonomian merkittävin ero on kahden ylimmäksi mielletyn ajattelun tason järjestyksessä. Alkuperäisen taksonomian korkein taso arvioiminen on uudistetussa versiossa katsottu toiseksi ylimmäksi tasoksi, kun taas alkuperäisen taksonomian syntetisoinnin luokkaa vastaava luominen on uudistetussa taksonomiassa nostettu korkeimmaksi ajattelun tasoksi.

Bloomin taksonomian käytön eräs haaste on, ettei taksonomia ota automaattisesti huomioon oppiaineiden erityispiirteitä. Fysiikan tehtävien tyypillinen rakenne voi poiketa merkittävästi äidinkielen tai historian oppiaineille tyypillisistä tehtävistä. Shakhmanin ja Moshen [32] käyttämä, erityisesti fysiikan opetukselle kehitetty PPST (*Physics Problem-Solving Taxonomy*), pyrkii Bloomin taksonomiaa ja muita tehtävien vaatimien kognitiivisten prosessien kautta tapahtuvia luokitteluita



Kuva 1. Uudistetun Bloomin taksonomian tasot [31].

hyödyntäen kuvaamaan fysiikalle tyypillisten ongelmanratkaisutehtävien ajattelun taksonomiaa. Bloomin taksonomiaan pohjautuen voidaan siis tehdä laadukkaita ja kuvaavia luokitteluita myös fysiikan tehtävistä, mutta pelkän Bloomin taksonomian hyödyntäminen tehtävien vaatimustason kuvaamiseen voi helposti vääristää tehtävistä muodostuvaa kuvaa.

### 1.2.3 Muita arvioinnin analysoinnin tapoja

Arvioinnissa käytettävien tehtävien laatua ja erityisesti vaikeustasoa kuvaamaan voidaan Bloomin taksonomian lisäksi käyttää kognitiivisen vaativuustason (*cognitive complexity level*) käsitettä, jota on esimerkiksi Woitkowskin [33] tutkimuksessa käytetty fysiikan yliopisto-opetuksen sisältöjen haastavuuden kuvaamiseen. Erityisesti fysiikan opetukselle tyypillisten tehtävärakenteiden analysointiin voidaan vaativuustason kautta tapahtuvan luokittelun kautta saada Bloomin taksonomian soveltamista kuvaavampi tulos. Kognitiivisten vaativuustasojen tarkastelun kautta ei suoraan voida tehdä johtopäätöksiä arvioinnin oikeasta kohdistumisesta arvioitaviksi tarkoitettuihin tavoitteisiin, mutta Bloomin taksonomian tavoin voidaan tehtäviä ryhmitellä vaativuuden mukaan, mikä voi osaltaan tukea tarkempaa arvioinnin

kriittistä tarkastelua.

Embretson ja Daniel [34] käyttävät kognitiivisia vaatavuustasoja matematiikan laajassa käytössä olleiden tehtäväpankkitehtävien luokitteluun ja analysointiin. Käytettyjen menetelmien ollessa luonteeltaan vahvasti myös tilastollisia on huomioitava, ettei tehtävien kognitiivisia vaatavuustasoja voida tässä yhteydessä arvioida yhtä kattavasti arvioida samoin menetelmin. Embretsonin ja Danielin tutkimus osaltaan osoittaa, että kognitiivisia vaatavuustasoja voidaan erilaiset parametrit huomioiden hyvin käyttää matematiikan tehtävien arvioinnissa. Näin ollen myös fysiikan perusopetuksessa käytettyjen tehtävien lähestyminen kognitiivisten vaatavuustasojen kautta voisi olla mahdollista.

Biggs ja Collis [35] esittävät oppimisen laadun kuvaamiseen viisiportaista taksonomista *SOLO*-mallia. Nimensä mukaisesti *SOLO* (*structure of the observed learning outcome*) perustuu lähtökohtaisesti havaittuihin oppimistuloksiin, eikä näin ollen sovellu suoraan käytettäväksi laadittavien tehtävien laadun arvioinnissa. *SOLO*-taksonomian suurin hyöty on siis nähtävissä opetuksessa pidemmällä ajanjaksolla tai laajemman arvioinnin kokonaisuuden yhteydessä suoritettavassa laadun tarkastelussa, mutta mallin olemassaolo on arvioinnin vastaavia taksonomioita tarkasteltaessa syytä tiedostaa.

### 1.3 Tehtävien arviointikriteerit

Tutkielmassa käsiteltäviä kielimallilla generoituja ja tehtäväpankin tehtäviä on tarpeen arvioida ja luokitella, jotta tehtäviä voidaan ensimmäisen tutkimuskysymyksen mukaisesti kuvailla. Tarkasteltaessa tehtäviä mahdollisessa arviointikäytössä erityisen huomion tulisi kohdistua niiden validiteettiin [13]. Tarkastelunäkökulma on tässä yhteydessä perusteltua rajata Messickin [13] määritelmän mukaiseen sisältövaliditeetin sisällön relevanssin osa-alueeseen. Sisällön relevanssin täydellinen arviointi ei ilman laajaa asiantuntijapohjaa ole mahdollista, mutta valittuja arviointikriteerejä

hyödyntäen on mahdollista koostaa rajattu kuvaus tehtävien laadusta ja ominaisuuksista, jotka suoraan tai välillisesti liittyvät tehtävien sisällön relevanssiin.

Tarkasteltavien tehtävien laadukas analysointi edellyttää kriteerejä, joiden pohjalta analyysi voidaan suorittaa. Tehtävien analyysin tarkoituksenmukaisuuden varmistamiseksi kriteerit on syytä perustaa olemassa olevaan teoreettiseen taustaan. Kriteerien muodostamisen kannalta oleellinen kirjallisuus keskittyy tutkielman kanssa samankaltaisen aiheen ja samantyyppisen tutkimusasetelman sisältäviin tutkimuksiin, kuten Küchemannin ym. [36] tutkimus tai fysiikan tehtävien luokittelua käsitteleviin tutkimuksiin, kuten Mesicin ja Muratovicin [37] tehtävien vaikeutta ennustavia tekijöitä käsittelevä tutkimus.

Generatiivisella tekoälyllä tuotettavien tehtävien ja kustantajan tehtäväpankin tehtävien arvioinnin tavoitteena on ensimmäisen tutkimusongelman osalta tehtävien laadun arviointi ja ominaisuuksien kuvaus. Generoitavien tehtävien ja valitun vertailuaineiston suppeuden takia kuvailevampi tehtävien arviointi sopii tehtävien analysointiin määrällisiä menetelmiä paremmin. Tehtävien laadun arvioimiseksi käytetään sisällönanalyysin menetelmiä, joiden avulla kehitetään olemassa olevan kirjallisuuden pohjalta summatiivisen arvioinnin tehtäville soveltuva arviointikriteeristö, jonka avulla tehtävien laadun kannalta keskeiset ominaisuudet voidaan kvantifioida.

Yksittäisten generatiivisella tekoälyllä tuotettujen tehtävien vertailua ja arviointia suoritettaessa on huomioitava annetun kehotteen lisäksi se, että kielimallin ollessa luonteeltaan probabilistinen [38] tuottaa identtinen syöte eri käyttökerroilla todennäköisesti poikkeavia tulosteita. Kehotteissa annettava tieto on siis kriteerien kehittämisen jälkeen muotoiltava niin, että ennen kehotteen antamista kielimallille varmistutaan siitä, että annettujen tietojen perusteella on odotettavissa tuloste, joka on uskottavasti arvioitavissa laaditulla kriteeristöllä.

### 1.3.1 Tehtävien luokittelu kirjallisuudessa

Küchemannin ym. [36] toteuttamassa tutkimuksessa arvioitiin ChatGPT:n tuottamien tehtävien laatua ja ChatGPT:n soveltuvuutta tehtävien tuottamiseen fysiikan opettajaopiskelijoiden tukena. Tutkimuksessa tehtävien laatua analysoitiin kahdeksan mittarin avulla, joista osaa voidaan soveltaa myös tässä generoitavien tehtävien ja tehtäväpankkien tehtävien tarkasteluun. Tehtävien kriteeripohjaisen laadun arvioinnin suoritti kaksi kokenutta opettajankouluttajaa, jolloin asiantuntijuuden merkitys korostuu tehtävien sisältövaliditeetin arvioinnissa [28, 29], mitä tutkimuksessa tehdyn arvioinnin voidaan katsoa edustavan.

Mesicin ja Muratovicin [37] tutkimuksessa tehtävien vaikeutta ennustavia tekijöitä on pyritty etsimään ja tulkitsemaan olemassa olevan kirjallisuuden pohjalta tehdyn sisällönanalyysin avulla. Aineistona tutkimuksessa ovat olleet Bosnian ja Hertsegovinan perusopetuksen päättöarvioinnin valtakunnalliset kokeet. Tutkimuksessa on tarkasteltu siis erityisesti fysiikan summatiivisen arvioinnin tehtäviä erilaisten kriteerien keinoin, minkä vuoksi käytettyjä kriteerejä voidaan mahdollisesti soveltaa myös tässä tutkielmassa käsiteltäville tehtäville toteutettavassa sisällönanalyysissä.

Ensimmäisenä mittarina Küchemannin ym. [36] tutkimuksessa mainitaan spesifisyys (*specificity*), joka arvioitavissa tehtävissä voidaan helpoiten havaita tehtävänannossa annettujen tietojen riittävyytenä. Mikäli tehtävässä ei anneta tehtävän ratkaisemiseksi tarvittavia tietoja, ei tehtävää voida pitää riittävän spesifinä. Tehtävän selkeyttä (*clarity*) voidaan tulkita tarkastelemalla tehtävänannon tulkinanvaraisuuksia. Oikeellisuuden (*correctness*) kriteeriä arvioidaan tehtävänannossa esitettyjen oletusten ja taustatietojen fysiikan oikeellisuuden osalta. Mikäli tehtävänannossa käytetty fysikaalinen tausta ei ole täysin oikein, ei tehtävän voida katsoa olevan oikeellinen. Spesifisyyden, selkeyden ja oikeellisuuden voidaan kriteereinä nähdä keskittyvän osin toisiinsa liittyviin tehtävänannon ominaisuuksiin, mikä voi vaikeuttaa kriteerien soveltamista tehtäviä analysoitaessa.

Küchemannin ym. [36] tutkimuksessa kriteerinä oli myös yksiselitteisyys (*not misleading*). Yksiselitteisyyden voidaan katsoa sisältävän erilaiset tehtävänannon tarpeettomasti ratkaisua väärään suuntaan ohjaavat viestit, joiden löytyminen tehtävänannosta oikeuttaa tehtävän harhaanjohtavana pitämisen. Kriteereistä riittävä haastavuus (*adequate difficulty*) on syytä arvioida tehtävän suunnitellun käyttökohteen ja sitä vastaavan opetussuunnitelmassa määritellyn tavoitellun osaamistason mukaan. Tarkasteltaessa suomalaisen peruskoulun vuosiluokkien 7–9 opetusta, määrittää Perusopetuksen opetussuunnitelman perusteet [14] tavoiteltavan vaikeustason. Kontekstin (*context*) osalta huomiota on kiinnitettävä siihen, että tehtävänannossa esitetään riittävästi sen tilanteen kontekstia, johon ongelma on sidottu.

Relevanssin (*relevance*) voidaan Küchemannin ym. [36] kriteereistä nähdä otettavan tarkimmin kantaa siihen, kuinka hyvin arvioidaan tavoitteen mukaista oppimista. Mikäli relevanssin kriteeriä sovelletaan summatiivisen arvioinnin tehtävän arviointiin, on se eräänlainen kuvaus tehtävän sisältövaliditeetista erityisesti sisällön relevanssin osalta [13], sillä yksittäisen tehtävän perusteella ei tyypillisesti voida tehdä päätelmiä sisällön kattavuuden osalta summatiivisen arvioinnin kontekstissa. Relevanssi ja tehtävän yleinen laatu (*overall quality*) olivat Küchemannin ym. [36] tutkimuksessa osasta muita kriteerejä poiketen arvioitu viisiportaisella asteikolla yhdestä viiteen, kun muissa kriteereissä arviointiasteikko oli binäärinen.

Tehtävien vaikeuden kuvaamiseksi on tarpeen kehittää mekanismi, jolla tehtävien vaikeustasoa pystytään selkeästi arvioimaan. Arvioinnin laadun varmistamisen kannalta keskeistä on arvioinnin edustaman ajattelun ja kognitiivisen toiminnan tason tunteminen, missä arvioijan tukena voivat toimia uudistetun Bloomin taksonomian [31] tai PPST:n [32] kaltaiset luokittelujärjestelmät. Generoitujen tehtävien kognitiivisen haastavuuden tarkastelu tukee tehtävien kuvailua suhteessa tyypillisiin tehtäväpankkien tehtäviin, minkä vuoksi tarkastelu voidaan katsoa ensimmäiseen tutkimuskysymykseen vastaamista tukevaksi. Mesicin ja Muratovicin [37] käyt-

tämät fysiikan tehtävien vaikeustasoa ennustavat ominaisuudet soveltuvat valituilta osin käytettäväksi vertailtavien tehtävien kuvailuun, mutta eivät sellaisenaan tuota tietoa tehtävien varsinaisesta vaikeustasosta. Mesicin ja Muratovicin [37] tutkimuksessa oli käytetty tehtävien vaikeustason arvioimisen tukena myös Bloomin taksonomian [31] pohjalta muokattua luokittelua ajattelun tasoittain. Tässä yhteydessä ajattelun tasot eivät auta arvioimaan suoranaisesti tehtävien laadukkuutta, mutta voivat tukea vertailua kustantajan tehtäväpankin koetehtäviin sekä tehtävien yleistä kuvailua. Woitkowskin [33] käyttämä luokittelu tehtävien kognitiiviselle vaativuustasolle käsittelee tehtävien vaikeustasoon vaikuttavia tekijöitä tavalla, joka muistuttaa uudistetun Bloomin taksonomian [31] mallia, mutta on fysiikan oppisisältöjen suhteen helpommin sovellettavissa.

Validiteettia arvioitaessa arvioijan oma asiantuntemus ja laaja osaaminen korostuvat [28, 29]. Samoin tehtävien luokittelussa ja analyysissa vaaditaan asiantuntemusta fysiikan opetuksesta ja arvioinnista. On tärkeää huomata, että tehtäviä analysoivan kirjoittajan kokemus on tässä tutkielmassa asiantuntijan kokemusta merkittävästi suppeampaa. Asiantuntemuksen puutteen aiheuttamaa vääristymää tuloksiin pyritään tehtävien luokittelu- ja arviointikehystä laadittaessa minimoimaan tekemällä kriteereistä riittävän yksiselitteisiä ja yksinkertaisia, jolloin saadut tulokset ovat laadittujen kriteerien puitteissa kelvolliset.

### 1.3.2 Soveltuva kriteeristö

Mesicin ja Muratovicin [37] ja Küchemannin ym. [36] tutkimuksissa käytettyjen kriteerien joukosta soveltuvia kriteerejä valittaessa on huomioitava muodostettavan kriteeristön riittävä kattavuus ja kuvailevuus. Kriteeristön ensimmäiseksi kriteeriksi valitaan Woitkowskin [33] tutkimuksessaan käyttämä, kuvassa 2 esitetty luokittelu-järjestelmä. Tehtävät neljään ryhmään jaotteleva malli soveltuu käytettäväksi myös generoitavien fysiikan tehtävien luokitteluun ja tehtävien kognitiivisen vaativuusta-

Hierarkkisen kompleksisuuden indikaattorit	
-1	Vaatiiko ongelma ainoastaan tunnetun faktan, kaavan tai määritelmän toistamista?
	Kyllä: Faktat; Ei: Seuraava kysymys
-2	Vaatiiko ongelma eksplisiittisesti laskun, perustelun tai perusteen?
	Kyllä: Siirry kohtaan 4; Ei: Seuraava kysymys
-3	Vaatiiko ongelma ainoastaan kuvauksen prosessista sanoin, luonnoksen tai diagrammin avulla?
	Kyllä: Prosessit; Ei: Seuraava kysymys
-4	Vaatiiko ongelma ainoastaan yksittäisen, lineaarisen perustelun, kuten 'Koska x, niin y' ilman tarvetta harkita muita tekijöitä tai rajoitteita?
	Kyllä: Lineaarinen kausaliteetti; Ei: Seuraava kysymys
-5	Vaatiiko ongelma ainoastaan yksittäisen vaikuttavan tekijän käsittelyä?
	Kyllä: Lineaarinen kausaliteetti; Ei: Monimuuttujainen keskinäisriippuvuus

Kuva 2. Suomennettu luokittelujärjestelmä tehtävien kognitiivisen vaativuustason määrittämiseksi [33].

son arvioimiseen, vaikka Woitkowskin tutkimuksessa tehty tehtävien luokittelu koskee fysiikan yliopisto-opinnoissa käytettäviä tehtäviä. Tehtävien voidaan katsoa vastaavan perusopintotason tehtäviä, joissa kouluopetuksen fysiikan tehtävien kanssa samankaltaiset rakenteet ovat tavanomaisia. Woitkowskin käyttämät neljä luokkaa ovat faktat (*facts*), prosessit (*processes*), lineaarinen kausaliteetti (*linear causality*) ja monimuuttujainen keskinäisriippuvuus (*multivariate interdependency*). Tehtäviä analysoitaessa voidaan kognitiivisen vaativuustason kriteerin arviointiin käyttää kuvan 2 taulukon mukaista luokittelua sisällönanalyysin pohjana kognitiivisen vaativuustason kriteerin osalta. Kognitiivisen vaativuustason luokituksen voidaan katsoa olevan uudistetusta Bloomin taksonomiasta [31] poiketen fysiikan tehtävien luonteen paremmin huomioiva ja samanaikaisesti tehtävien edellyttämiä kognitiivisia prosesseja riittävällä tarkkuudella kuvaava, joten erillistä Bloomin taksonomian mukaista luokittelua tehtäville ei ole tarpeellista tehdä.

Tehtävien suhdetta opetussuunnitelmaan on mahdollista tarkastella Küchemanin ym. [36] määrittelemän relevanssin kriteerin kautta, mutta tällöin haasteeksi yksittäisiä tehtäviä tarkasteltaessa nousee sisällön kattavuuden läheinen suhde sisällön relevanssiin. Yksittäisen tehtävän yhteys opetussuunnitelmaan on mahdollista todeta, mutta opetussuunnitelman huomioimisen luotettava arviointi edellyttäisi tehtävän summatiivisen arvioinnin laajemman kontekstin tarkastelua. Koska tutkielmassa tarkastellaan yksittäisiä tehtäviä, ei tehtävien suhdetta opetussuunnitelmaan voida tarkastella luotettavasti. Opetussuunnitelman suhteen tehtävää analyysia sisällön

relevanssin osalta vaikeuttaa myös Perusopetuksen opetussuunnitelman perusteissa [14] esitettyjen kriteerien ja tavoitteiden laajuus, minkä johdosta hyvin erilaisten tehtävien on mahdollista todeta olevan ainakin jossain määrin opetussuunnitelman mukaisia.

Opetussuunnitelmaan liittyy kuitenkin Mesicin ja Muratovicin [37] tutkimuksessa vaikeustasoa määrittäväksi tekijäksi luettu kokeellisen menetelmän tuntemus (*knowledge of experimental method*). Opetussuunnitelmassa [14] huomioidut fysiikan kokeellinen työskentely ja toiminnalliset taidot ovat opetussuunnitelman valossa osa fysiikan arviointia, minkä vuoksi kokeellisuuteen viittaavien elementtien esiintymistä fysiikan summatiivisen arvioinnin tehtävissä on tarpeen tarkastella. Mesicin ja Muratovicin [37] laatimaa binääristä luokittelua mukailleen nostetaan kokeellisen menetelmän tuntemus tehtävien arviointikriteeriksi pääosin samoin arviointikriteerein. Tehtävän luokittelu luokkaan 1 edellyttää tehtävän ratkaisijaltaan vaatimaa kokeellisen työskentelyn tai koeasetelmien tuntemusta. Vaihtoehtoisesti abstraktimpien kokeelliseen menetelmään liittyvien tietojen käytön vaatiminen tehtävissä tuottaa luokituksen 1.

Tehtävän ratkaisemisen kannalta oleellisten fysiikkaan liittyvien tietojen riittävyttä ja oikeellisuutta tarkastellaan fysiikan oikeellisuuden kriteerin kautta. Fysiikan oikeellisuuden kriteeriä laajennetaan Küchemannin ym. [36] määritelmästä siten, että puuttuvien tietojen lisäksi virheelliset oleellisesti tehtävän fysiikkaan liittyvät tiedot tuottavat fysiikan oikeellisuuden arvoksi 0. Kriteerin binäärisyyttä ei ole tarpeen muokata alkuperäiseen kriteeriin nähden, sillä tarkoituksena ei ole tuottaa tarkasti eri tasoilla tehtävän fysiikan virheitä ja ongelmia kuvailevaa analyysia, vaan kuvaus tehtävien yleisestä laadusta. Fysiikan oikeellisuuden kriteeri voidaan nähdä eräänlaisena tehtävän sisällön relevanssin raja-aitana, jonka puutteiden vaikutukset koko tehtävän validiteettiin ovat merkittävät.

Tehtävien selkeyttä kuvaamaan valitaan käytettäväksi tehtävien analysoinnis-

Taulukko I. Tehtävien arviointiin käytettävä kriteeristö

Kriteeri	Asteikko	Lähde
Kognitiivinen vaativuustaso	1/2/3/4	[33]
Kokeellisen menetelmän tuntemus	0/1	[37]
Fysiikan oikeellisuus	0/1	[36]
Selkeys	0/1	[36]

sa selkeyden kriteeri. Sellaisenaan käytettäväksi soveltuvan Kuchemannin ym. [36] määritelmän mukaisesti tehtävänannon on oltava muotoiltu selkeästi ja ytimekkäästi, jotta tehtävälle voidaan antaa selkeyden osalta arvo 1. Mikäli tehtävänannossa esiintyy tehtävän ratkaisemisen tai kontekstin kuvaamisen kannalta epäoleellista tekstiä, annetaan tehtävälle alkuperäisen kriteerin mukaisesti arvo 0. Alkuperäinen luokitus soveltuu käytettäväksi käsiteltävien tehtävien analyysissä sellaisenaan, sillä tehtävien luonne fysiikan tehtävinä on riittävän samankaltainen kuin Kuchemannin ym. [36] tutkimuksessa käsiteltyjen tehtävien.

Tehtävien analyysin pohjana toimivat kriteerit lähteineen on esitetty taulukossa I. Kriteerikohtaiset tarkemmat luokittelua tukevat alakriteerit muodostetaan osana sisällönanalyysin prosessia luvussa 2.2 siten, että kirjallisuuden pohjalta muodostetut kriteerit säilyvät sisällöltään ja merkitykseltään esitettyihin määritelmiin nähden mahdollisimman samankaltaisina. Aineiston koostamisen jälkeen tehtäville suoritetaan laadittujen kriteerien pohjalta suoritettavan sisällönanalyysi, minkä kautta pyritään löytämään vastaus tutkielmaa ohjaavaan tutkimuskysymykseen.

## 1.4 Tekoälyn käyttö opetuksessa

Tekoälyn vaikutus opetukseen ja oppimiseen on kasvanut merkittävästi generatiivisen tekoälyn laajojen kielimallien tultua laajan yleisön tietoisuuteen ja käyttöön

OpenAI:n GPT-3.5 -kielimallin julkaisun myötä marraskuussa 2022 [39]. Tekoälynä voidaan pitää myös joitain jo aiemmin olemassa olleiden opetuksen ja oppimisen työkalujen ominaisuuksia ja toiminnallisuuksia [40]. Ei-generatiivista tekoälyä käytäviä järjestelmiä enemmän keskustelua opetuslalla on mahdollisesti herättänyt generatiivisen tekoälyn käytön arkipäiväistymisen vaikutukset opetuksen toteuttamisen käytäntöihin [1].

Tekoälyn käyttötarkoitukset opetuksessa ja oppimisessa on valittava siten, että käytössä oleva tekoäly kykenee suoriutumaan tehtävästään hyvin. Tätä voidaan painottaa tarkasteltaessa opettajan teknologista osaamista, jolloin erilaisten apuvälineiden tarkoituksenmukainen ja tehokas käyttäminen korostuu teknologian hyödyntämisessä opetuskäytössä [41]. Tekoälyn käyttötarkoitusten rajaaminen kulloistenkin tekoälytuotteiden kyvykkyyksien mukaan vaikuttaa välttämättömältä, jotta tekoäly voidaan luontevasti ottaa osaksi opetusta. OpenAI:n ChatGPT -kielimallit ovat osin käyttökelpoisia oppimisessa ja opetuksessa, minkä vuoksi mallin vahvuuksien ja mahdollisten oppimista uhkaavien tekijöiden tunnistaminen on tärkeää [42].

Koehlerin ja Mishran [41] mukaan teknologian laadukasta ottamista osaksi opetusta voidaan kuvata 'opettajan pedagogisen sisältötiedon' (*PCK, pedagogical content knowledge*) [43] pohjalta kehitetyn 'teknologisen pedagogisen sisältötiedon' (*TPACK, technological pedagogical content knowledge*) käsitteen avulla. PCK:n rinnalle nousee TPACK:ssa teknologia ja siten PCK:n ollessa pedagogisen tiedon ja sisältötiedon leikkauksessa [43], on TPACK pedagogisen tiedon, sisältötiedon ja teknologisen tiedon leikkauksessa [41]. Teknologian hyödyntämistä opetuksessa voidaan pitää haastavana kokonaisuutena. Teknologian tehokkaan ja ammattitaitoisen opetuskäytön kantavana rakenteena voidaan pitää dynaamisesti toistensa suhteen samanaikaisesti hallittuja kolmea TPACK:in osa-aluetta [41]. TPACK:in täydentämiseksi erityisesti tekoälyn vastuulliseen hyödyntämiseen liittyen voidaan laajentaa TPACK:ia ottamalla huomioon myös eettinen näkökulma, jolloin opettajan tekno-

logista pedagogista sisältötietoa voidaan lähestyä älykkään TPACK:in (*intelligent TPACK*) kautta [44].

Ennakkoasenteet (*bias*) voivat Messickin [24] mukaan heikentää arvioinnin pätevyyttä. Laajojen kielimallien osalta ongelmallisena pidetään yleisesti kielimalleissa esiintyvää asenteellisuutta, joka voi olla peräisin mallin koulutuksessa käytetystä aineistosta, algoritmista tai käyttäjän toiminnasta [45]. Tekoälyä arvioinnissa hyödynnettäessä on huomioitava näiden kahden, osin päällekkäisen, ilmiön tuottamat haasteet ja mahdollisuudet arvioinnin laadun suhteen. Ennakkoasenteiden kytkeytyessä käytettyyn kielimalliin, mutta silti ollessa peräisin käyttäjästä, on erityisen huolellinen arvioinnin kriittinen ja ammattimainen tarkastelu perusteltua. Tekoälyn tuottaman syrjinnän riskien tunnistamisessa Suomi sijoittuu Euroopan unionin keskiarvon alapuolelle [46].

Tekoälyn opetuskäyttöön liittyvänä ongelmana tekoälyn ennakkoasenteita voidaan Zhain ja Nehmin [47] mukaan tutkia ja siten pyrkiä välttämään. Tekoälyn hyödyntämättä jättämistä laajoille kielimalleille ominaisten piirteiden vuoksi ei voida heidän mukaansa pitää perusteltuna, varsinkin tekoälyn ollessa jo laajasti käytössä opetuksen ja oppimisen kentällä. Tarpeen on siis löytää tasapainoisia ja älykkään TPACK:in mukaisesti opettajan teknologiseen ja pedagogiseen asiantuntijuuteen nojaavia [44], vastuullisia ratkaisuja. Tuotettaessa tehtäviä summatiiviseen arviointiin generatiivista tekoälyä käyttäen voidaan tehtävien laadun ja validiteetin huolellisella arvioinnilla ehkäistä mahdollisia vinoumia ja epäjohtonmukaisuuksia.

#### 1.4.1 Koneoppiminen ja tekoäly

Erityisen mielenkiinnon kohteena oleva luonnollista kieltä prosessoiva generatiivinen tekoäly on laajemman tekoälyn ja koneoppimisen muodostaman kokonaisuuden alaluokka [48], vaikka usein arjessa tekoälyksi käsitetään lähinnä uudemmat generatiivisen tekoälyn laajat kielimallit, kuten ChatGPT, Gemini ja Copilot. Turingin

vuonna 1950 esittämät ajatukset älykkäistä koneista [49] loivat osaltaan pohjan tekoälyn tutkimukselle ja kehitykselle nykyhetkeen asti. Turingin testi asettaa kriteerit tietokoneen älykkyydelle luonnollisen kielen välityksellä suoritettavien tehtävien kautta [50].

Mitchell [51] korostaa tekoälyä määriteltäessä suhdetta koneiden oppimiseen. Koneoppimista voidaan siis pitää eräänä tekoälyn määritelmällisenä osana. Koneen täytyy poiketa normaalista tietokoneohjelmasta siten, että ohjelma kykenee oppimiseksi määriteltävään toimintaan. Koneoppimista ja tekoälyä käsittelevällä tutkimuksella on jo noin 70 vuoden historia, jonka aikana tapahtunut tietokoneiden kehitys ja kyky suorittaa monimutkaisempia ohjelmia on johtanut aiemmin teoreettisemman tekoälyn käsitteen ainakin osittaiseen todeksi muuttumiseen [50].

Tekoälyn ja koneoppimisen kehityksen varhaisemmassa vaiheessa Mitchell [51] jaotteli koneoppimisen tärkeimmät käyttökohteet kolmeen pääkategoriaan: datanlouhintaan, vaikeasti ohjelmitaviin sovelluksiin ja yksilöllisiin ohjelmistosovelluksiin. Näistä pääkategorioista arviointimateriaalin tuottaminen voidaan luokitella vaikeasti ohjelmitavien sovellusten joukkoon, sillä ohjelmoinnin kautta tapahtuvat mahdollisuudet tuottaa tarpeeksi laadukkaita arviointimateriaaleja ovat rajallisia ja arviointimateriaaleja tuottavat suurelta osin opettajat ja opetusalan ammattilaiset itse.

Tekoälyn yleistä teknistä ja teoreettista taustaa käsittelevät muiden muassa Goodfellow, Bengio ja Courville [52] erityisesti syväoppimisen näkökulmasta ja Russell ja Norvig [48] laajasti tekoälyn ja koneoppimisen perusteita kuvaillen. Tekoälyn yleisten teknisten ja matemaattisten ominaisuuksien tuntemusta oleellisempaa on tässä yhteydessä tuntee erityisesti laajojen kielimallien toiminta ja luonne, jotta kyseisten kielimallien opetuskäyttöä voidaan tehokkaasti arvioida. Laajojen kielimallien opetuskäytön näkymiä sivuavat laajasti Kasneci ym. [53], joskin viime vuosien kehitys generatiivisen tekoälyn osalta on ollut kiivasta. Tästä huolimatta esitetty-

jä arvioita generatiivisen tekoälyn tuottamista mahdollisuuksista, rajoituksista ja haasteista opetukselle voidaan pitää edelleen ajankohtaisina.

#### 1.4.2 Suuret kielimallit ja generatiivinen tekoäly

Tekoälyn kehitys erityisesti luonnollista kieltä käyttävien suurten kielimallien osalta on ollut viime vuosina nopeaa. Köbisin ja Mossinkin [54] GPT-2 -kielimallin tuottamaa runoutta käsittelevä tutkimus osoittaa, ettei ihminen kykene luotettavasti erottamaan ihmisen ja generatiivisen tekoälyn tuottamaa tekstiä toisistaan. Suurten kielimallien ja generatiivisen tekoälyn taustalla vaikuttaa vahvasti transformer-arkkitehtuuri. Transformer-arkkitehtuurin perusteita ja niiden tehokkuutta resursien näkökulmasta käsitellään laajasti Vaswanin ym. [38] artikkelissa. Transformer-arkkitehtuurin tuottamaa tulostetta (*output*) kuvaillaan todennäköisyyksiin perustuvana, mikä on eräs suurten kielimallien tyypillisistä ominaisuuksista.

Viime vuosina yleistyneen OpenAI:n ChatGPT -kielimallin ominaisuuksien tuottamista rajoituksista ja mahdollisuuksista ovat kirjoittaneet katsausartikkelissaan Cong-Lem, Soyoo ja Tsering [55]. Katsauksessa erityisesti rajoittavina tekijöinä nähdään kielimallin haasteet tarkkojen tai monimutkaisten sisältöjen käsittelyssä, käyttäjien toiminnan riippuvuus kielimallista ja eettiset ongelmat. Käyttömahdollisuuksiin lukeutuvat erilaiset ideointi- ja tiedonhaketarkoitukset, joissa kielimallin kanssa käytävä luonnollisen kielen dialogi voi tukea toimintaa, sekä mahdollisuus yksilöidä oppimista tekoälyn tuottaman oppimissisällön tai valmentamisen kautta.

ChatGPT:n GPT-5 kielimalli on sen kehittäneen OpenAI:n mukaan edeltäjiään useilla mittareilla edistyneempi, minkä lisäksi se on jaettu kahteen alamalliin GPT-5 Main:iin ja GPT-5 Thinking:iin jakautunut [56]. Kielimallin todetaan suoriutuvan tehtävistään usealla kielellä varsin hyvin, mutta suomenkielistä suorituskkyä ei ainakaan kielimallin järjestelmäkortissa ole mainittu [56]. Tehtävien generointiin lopulta käytettävä GPT-5.1 -kielimalli jakautuu OpenAI:n [57] mukaan kahteen toi-

sistaan poikkeavaan malliin, joiden väliltä valitseminen tapahtuu tyypillisesti automaattisesti ilman käyttäjän aktiivista valintaa. Yleisesti versiota 5.1 voidaan pitää GPT-5:n kehittyneempänä versiona, jonka taustalla vaikuttavat kaksi erillistä alaversiota GPT-5.1 Instant ja GPT-5.1 Thinking.

Keskustelupohjaisille suurille kielimalleille tyypillistä on kehotteiden (*prompt*) käyttäminen. Kehotteet annetaan keskustelupohjaisissa malleissa luonnollisella kielellä, mikä madaltaa ohjelmointitaidottomien käyttäjien kynnystä käyttää kyseisiä palveluja. Kehotteiden laatimiseen liittyy yleisesti ongelmia, jotka liittyvät kehotteen antamiseen luonnollisella kielellä ja keskustelunomaisuuteen. Haasteita tehokkaalle kehotesuunnittelulle aiheuttavat erityisesti virhekäsitykset, jotka ovat tyypillisesti peräisin tekoälyn kanssa käytävän keskustelun ihmisten väliseen keskusteluun samaistamisesta. [4]

Kehotteiden laadinnassa voidaan hyödyntää Eagerin ja Bruntonin [58] laatimaa kehotteen rakenteen kuvausta, joka esittää kehotteen keskeisimmät osat. Näiden osien huolellinen kehittäminen ja kehotteeseen sisällyttäminen vähentää osaltaan riskiä tehdä yleisiä kielimallin tuottaman tekstin laatua heikentäviä virheitä. Knot-hin ym. [5] tutkimuksessa tutkittiin tekoälyn lukutaidon vaikutusta tehokkaaseen kehotteiden käyttöön ja havaittiin aiemman kokemuksen ohella tekoälyn tuntemuksen tukevan kehotteiden tarkoituksenmukaista käyttöä.

### 1.4.3 Tekoälyn käyttökohteita opetuksessa

Mittal ym. [59] näkevät generatiivisen tekoälyn voivan vastata opetuksessa kuuteen keskeiseen tarpeeseen: yksilöllistämiseen, sitouttamiseen ja vuorovaikutukseen, sopeutettavuuteen, tehokkuuteen ja saavutettavuuteen, luovuuteen ja tutkimuksellisen ilmapiirin ylläpitämiseen. Katsauksessa todetaan sisällön tuottamisen olevan yksi generatiivisen tekoälyn tuottamista mahdollisuuksista, pitäen sisällään potentiaalin luoda materiaaleja opetukseen ja arviointiin generatiivista tekoälyä hyödyn-

täen. Bhatin ym. [60] mukaan OpenAI:n GPT-3 kykeni kohtuullisella tasolla tuottamaan halutunlaisia tehtäviä arviointia varten. Huomioitavaa on, että GPT-3:n jälkeen OpenAI on julkaissut uusia, edistyneempiä kielimalleja, joiden kyvyn tehtävien generoinnissa voidaan olettaa olevan parempi [56].

Diliberti ym. [61] ovat koostaneet raportin, jonka mukaan Yhdysvaltain K-12-opetuksessa eli lastentarhasta 12. luokka-asteeseen ulottuvassa opetuksissa opettajista vain 33% oli käyttänyt tekoälyä opetuksessaan. Heistä 18% käytti tekoälyä aktiivisesti. STEM-aineiden opettajista 19% kertoi hyödyntävänsä tekoälyä opetuksessaan, joten käytön yleisyyttä voidaan pitää vähäisenä. Tekoälyn hyödyntäjistä 53% kertoi käyttäneensä opetuksensa tukena keskustelupohjaista tekoälyä, kuten ChatGPT:tä tai Google Bardia. Arviointiin suunniteltuja tekoälypohjaisia sovelluksia, kuten PrepAI:ta tai Quillions:ia tekoälyn käyttäjistä oli käyttänyt 10%. Suoranaisten koe- tai testilaadinnan apuvälineiden käytön harvinaisuudesta huolimatta toiseksi yleisin tekoälyä hyödyntäneiden opettajien käyttötarkoitus tekoälylle oli kyselyiden ja kokeiden laadinta.

Yläaste- ja lukio-opetukselle erityisesti oppijoiden käyttäminä mahdollisina laajojen kielimallien käyttökohteina mainitaan oppilaiden vuorovaikutteinen tuki oppimisessa sekä harjoitusten ja tehtävien tuottaminen. Opettajan työn näkökulmasta esille nostetaan oppituntien suunnittelu ja oppitunneilla suoritettavien tehtävien laadinta. Arvioinnin osalta Kasneci ym. [53] havaitsee generatiivisen tekoälyn mahdollisuudet erityisesti oppijoiden tuotosten arvioinnin ja pisteyttämisen osalta sekä oppijoiden yksilöllisenä tukena arviointitilanteissa. Oppituntien suunnittelun osalta todetaan kuitenkin tekoälyn soveltuvan hyvin oppimista tukevien harjoitusten suunnitteluun, mikä voi joissain tapauksissa merkitä formatiivista arviointia. [53] Formattiivisen arvioinnin laajan tarkastelun osalta Zhai ja Nehm [47] toteavat tekoälyn soveltuvan tehokkaaksi formattiivisen arvioinnin työkaluksi, joka keventää opettajien arviointityötä oppijoiden vastauksia pisteyttävien tai sisältöä analysoivien kielimal-

lien avulla.

Käytettäessä ChatGPT:n 3.5 -versiota lääketieteen opetuksen monivalintates-  
tin laatimiseen tarvitaan kysymysten, vastausten ja selitysten generoinnin jälkeen  
usein merkittäviä muokkauksia ennen mahdollista arviointikäyttöä. Tekoölyavustei-  
nen arviointi vaikuttaa siis mahdolliselta, mutta ajankäytöllisesti haastavalta, sillä  
kielimallin tuottamien virheellisten väittämien ja osien löytäminen ja korjaaminen  
on aikaavievää. Sisältöjen ollessa monimutkaisia ja yksityiskohtaisia, ainakin aiem-  
mat ChatGPT:n versiot ovat siis kohdanneet vaikeuksia, mutta tilanteen voidaan  
olettaa parantuneen uusien versioiden ilmestymisen myötä. [6]

Summatiivista arviointia voidaan pitää formatiivista arviointia kapeampana, kun  
verrataan niille asetettuja tehtäviä [9], joten summatiivisen arvioinnin tekoölyl-  
le soveltuvat käyttökohteet ovatkin mahdollisesti formatiivisen arvioinnin vastaa-  
via rajallisemmat. Generatiivisen tekoölyn erityisenä ominaisuutena korostuu kyky  
tuottaa uutta materiaalia, minkä vuoksi summatiivisen arvioinnin osalta tehtävien  
tuottaminen näyttäytyy johdonmukaisena tapana hyödyntää generatiivisen tekoölyn  
ominaisuuksia. Opettajan vastuu opetuksen opetussuunnitelman mukaisesta toteut-  
tamisesta edellyttää sekä arviointi- että muussa opetuskäytössä tekoölyä hyödynnet-  
täessä opettajan valvontaa, sillä vastuuta opetuksesta ei voida ulkoistaa kolmannelle  
osapuolelle.

#### 1.4.4 Tekoöly fysiikan opetuksessa

Polverini ja Gregoric [1] tarkastelevat erityisesti OpenAI:n GPT-4:n kykyä fysiikan  
tehtävien suorittamisessa, missä kielimalli osoittautuu edeltäjiään huomattavasti te-  
hokkaammaksi. Toimivia ja tehokkaita ratkaisuja fysiikan tehtäviin generatiivisella  
tekoölyllä tuotettaessa on Polverinin ja Gregoricin [1] mukaan erityisesti kiinnitettä-  
vä huomiota tekoölylle annettaviin kehoitteisiin. Huolellisesti laadituista kehoitteista  
huolimatta GPT-4:n ja varhaisempien versioiden ei nähty ratkaisevan fysiikan tehtä-

viä, jolloin kielimalleja voisi käyttää oppijoiden opetuksen tukena itsenäisen tutorin tapaan. Toisaalta viitteitä GPT-3:n GPT-3.5:n mahdollisesta käyttökelpoisuudesta opetuksen tukena assistenttina on havaittavissa Ahmedin ja Shanton [62] etsivässä tutkimuksessa korkeakouluissa tapahtuvasta generatiivisen tekoälyn käytöstä.

Qiu ja Zhang [63] tarkastelevat katsauksessaan fysiikan yliopisto-opetukseen soveltuvia generatiivisen tekoälyn työvälineitä ja niiden tarjoamia mahdollisuuksia. Generatiivisen tekoälyn laajan potentiaalin hyödyntäminen opetuksen tukemisessa edellyttää tekoälypohjaisten työkalujen huolellista ja ammattitaitoista sisällyttämistä osaksi opetusta [63], jolloin rajoittavat tekijät pystytään huomioimaan ja kielimallien hyvät ominaisuudet ottamaan opetuksen tueksi. Yliopistossa tapahtuvaan fysiikan opetukseen soveltuvien generatiivista tekoälyä hyödyntävien työkalujen tarjoamat laajat hyödyntämisen tavat voivat heijastella tilannetta generatiivisen tekoälyn opetuskäytössä hyödyntämisen osalta laajemminkin. Fysiikan yliopisto-opetuksen tukena käytetyn generatiivisen tekoälyn vaikutus oppimistuloksiin ei Robledo-Rellan ja Tohin tutkimuksessa ollut merkittävä [64], mutta uusien oppimista tukevien keskustelupohjaisten tekoälyratkaisujen rooli myös fysiikan opetuksessa tunnistetaan.

Frieder ym. [65] toteavat erityisesti GPT-4:n suoriutuvan haastavista, mutta alle maisterivaiheen opintojen tason (*undergraduate*) olevista matemaattisista ongelmista hyvin. Perusopetuksen kontekstissa voidaan päätellä ChatGPT:n matemaattisen ongelmanratkaisun kyvyn olevan perusopetuksen fysiikan opetuksen suhteen pääsääntöisesti riittävä. Kotsis [66] näkee ChatGPT:n mahdollisena fysiikan opettajan assistenttina, joka tarvittaessa kykenisi tuottamaan oppilaille nopeasti vastauksia monimutkaisia fysiikan oppisisältöjä koskien.

Cooper [67] on ChatGPT:n aiempaa versiota käyttäen tarkastellut sen soveltuvuutta luonnontieteiden opetukseen ja tuottanut ChatGPT:tä käyttäen itsearviointiin ja testaamiseen soveltuvia materiaaleja, jotka vaativat ainoastaan joitain muokkauksia ollakseen käyttökelpoisia. Tutkimus ei tosin käsitä nimenomaisesti fysiikan

opetukseen kohdistuvia materiaaleja, mutta siitä huolimatta antaa käytännön esimerkin kautta alustavan kuvan generatiivisen tekoälyn hyödyntämispotentiaalista fysiikan arvioinnissa. Cooperin [67] itsearviointilomakkeen ja monivalintatestin generoimiseen käyttämät kehotteet ovat muotoilultaan yksinkertaisia ja suoraviivaisia, joten materiaalien generoinnin ei voida nähdä vaatineen tässä yhteydessä laajaa tietämystä kehotteen suunnittelusta tai kyseisen kielimallin toiminnasta.

Gregorcic ja Pendrill [68] esittävät kuvauksen ChatGPT:n kyvykkyydestä Newtonin mekaniikkaan liittyvän kysymyksen käsittelyssä ja aiheeseen liittyvästä keskustelusta kielimallin kanssa. Keskustelu paljastaa kielimallin ilmeiset ongelmat sisällöllisesti koherentin tekstin tuottamisessa, kun kyseessä on yleisesti virhekäsityksiä aiheuttava heittoliikkeeseen liittyvä ongelma. Dokumentoidun keskustelun aikana kielimallin tuottamat vastaukset olivat ristiriitaisia ja kielimalli tuotti peräkkäisiin vastauksiinsa jopa päinvastaisia väitteitä. Vastaavassa tilanteessa esimerkiksi oppilaalle selityksiä tuottavana tukena [53] ChatGPT:n kyseinen versio ei vaikuta täytävän oppimisen kannalta keskeisiä vaatimuksia sisällön oikeellisuuden ja argumentoinnin pitävyyden osalta.

Kognitiivisen kuormituksen teoriaan [69] pohjaten Liang ym. [70] tunnistavat suurten kielimallien tuottamat mahdollisuudet keventää oppijan kognitiivista kuormitusta ja näin myös tukea fysiikan oppimista. Fysiikan opetuskäytössä generatiivisen tekoälyn käyttöönottoa heidän mukaansa mahdollisesti hidastaa kuitenkin fysiikkaan liittyvien laskennallisten tehtävien usein toistuva vektoriluonne, jota aina-kaan ChatGPT:n varhaisimmat versiot eivät ole kyenneet kovinkaan varmasti käsittelemään oikein. Vektorilaskut pois lukien on fysiikan opetukseen ja oppimiseen kuitenkin Liangin ym. [70] mukaan mahdollista hyvillä kehoitteilla saada laadukasta ja valmentavaa tukea, vaikka katsauksen tulokset perustuvat suurelta osin teoreettisiin havaintoihin, eikä varsinaisia empiriaan pohjaavia tuloksia kielimallien käytöstä opetuksessa juuri ole saatavilla.

Chan ja Ali [71] käsittelevät automaattiseen tehtävien generointiin STEM-oppiaineissa liittyviä kysymyksiä erityisesti kehotteiden valinnan näkökulmasta. Generoitujen tehtävien laadun kannalta parhaaksi vaihtoehdoksi havaittiin CoT-tyyppinen (*chain of thought*) kehotetyyli sovelluksineen. Oikean kehotetyylin valitsemista oleellisempaa on kuitenkin opettajan ymmärrys niistä generatiivisen tekoälyn toimintamekaniikan perusteista ja rajoitteista, joita keskustelupohjaisiin tekoälytyökaluihin liittyy.

Generatiiviselle tekoälylle esitetyt käyttökohteet fysiikan opetuksessa mukailevat suurelta osin yleisesti opetukseen tai muiden oppiaineiden tai alojen koulutukseen esitettyjä käytötapoja. Fysiikan opetukselle spesifejä tekoälytyökaluja tai erityisesti fysiikan opetukseen liittyviä tapoja hyödyntää tekoälyä ei ole esitelty suurta määrää. Tekoälyn hyödyntäminen opettajan assistenttina tai oppilaan yksilöllisenä tukena vaatii käytettävältä kielimallilta kykyä käsitellä fysiikan oppisisällöille tyypillistä kieltä, minkä vuoksi tekoälyn tuottaman fysiikan opetus- ja arviointimateriaalin tarkempi analysointi on tarpeen.

## 2 Menetelmät

Tutkielmassa tarkastelun kohteena olevia summatiivisen arvioinnin tehtäviä on tarkoitus analysoida laadullisen sisällönanalyysin keinoin ja kirjallisuuden perusteella tehtäville laadittuja kriteerejä hyödyntäen. Tehtävien laatiminen generatiivisella tekoälyllä edellyttää huolellista kehotteen suunnittelua, sillä ilman systemaattisesti ja tekoälytietoisesti kehitettyä kehotetta lopputuloksen laatu voi olla merkittävästi heikompi [4]. Tämän vuoksi aineiston valinnan ja rajaamisen yhteydessä täytyy osaksi käytettäviä menetelmiä lukea myös kielimallille annettavan kehotteen laatiminen ja muokkaaminen.

Tutkimuskysymyksiin vastaamiseksi tehtäviä arvioidaan sisällönanalyysin keinoin valittuja kriteerejä hyödyntäen. Toiseen tutkimuskysymykseen vastaamiseksi tarkastellaan ja kuvaillaan prosessia kielimallille annettavan kehotteen laatimisesta generoidun tehtävän valmistumiseen. Tutkimuskysymyksiin vastaaminen edellyttää laadittavan ja tuotettavan aineiston huolellista valintaa, jotta voidaan varmistua saavutettavien tulosten laadusta. Kielimallilla generoitavien tehtävien lisäksi analysoidaan oppikirjoja julkaisevien kustantajien ylläpitämien tehtäväpankkien koe-tehtäviä, jotka voidaan opetustyössä nähdä mahdollisena vaihtoehtona kielimallilla tuotetuille tehtäville.

### 2.1 Aineisto

Käsiteltävänä aineistona käytetään OpenAI:n GPT-5.1 -kielimallilla generoituja tehtäviä ja fysiikan oppimateriaaleja julkaisevien kustantamojen kirjasarjojen tehtäväpankkien tehtäviä. Kielimallilla generoidut tehtävät ja tehtäväpankkien tehtävät käsitellään yhtenevin menetelmin, mutta merkittävämpi painoarvo suhteessa tutkimuskysymyksiin on kielimallien tehtävillä. Tehtäväpankkien tehtävien tarkoituksena on osana aineistoa havainnollistaa vaihtoehtoisia opettajan käytössä olevia summatiivisen arvioinnin valmiiksi laadittujen tehtävien resursseja, joita voidaan verrata

generatiiviseen tekoälyyn.

Annettaessa generatiiviselle tekoälylle kehote tehtävän laatimiseksi on tarpeellista esittää asianmukainen sisällöllinen rajausta vuosiluokkien suhteen tehtävän rajauksen lisäksi. Perusopetuksen opetussuunnitelman perusteissa [14] sisältöjen ja tavoitteiden muotoilu on fysiikan vuosiluokkien 7–9 opetuksen osalta paikoin hyvin avointa, jolloin täsmälliset sisältökohtaiset oppimistavoitteet voidaan johtaa esitetyistä sisältöjen kuvauksista ja yleisempää tasoa käsittelevistä oppiaineen yleisistä tavoitteista.

Arvioinnin oikeaa kohdentumista tarkasteltaessa arvioinnin aiheena olevien tavoitteiden tarpeeksi tarkka tunteminen on keskeistä. Tämän vuoksi on tehtävien laadun arvioinnin toteuttamiseksi muodostettava Perusopetuksen opetussuunnitelman perusteiden pohjalta valitulle osa-alueelle selkeät tavoitteiden muotoilut, joihin tuotettuja tehtäviä ja kustantajan tehtäväpankin tehtäviä voidaan peilata. Tavoitteiden tarkemmassa muotoilussa ohjaavana tekijänä voitaisiin käyttää oppikirjamateriaaleissa esitettyjä opetussuunnitelmaan pohjautuvia tarkempia tavoitteita, mutta opetussuunnitelman ulkopuolisten vaikutteiden minimoimiseksi näin ei ole tässä tutkimuksessa tarpeen menetellä.

Perusopetuksen opetussuunnitelman perusteista sopivaksi sisältöalueeksi valitaan sisältö S6 Sähkö, jonka sisältöä kuvataan seuraavasti:

*"Virtapiirin tarkastelussa käytetään lähtökohtana jännitteen ja sähkövirran välistä yhteyttä. Sitä tarkastellaan ensin kvalitatiivisesti ilmiöiden ja ominaisuuksien tasolla, sitten kvantitatiivisesti mittaamalla suureiden arvoja ja tutkimalla suureiden välisiä riippuvuuksia. Sisältöjä valitaan myös kodin sähköturvallisuuteen sekä sähkön käyttöön ja tuottamiseen liittyen. Sähköinen varautuminen ja magnetismi yhdistetään kvalitatiivisesti virtapiirien ilmiömaailmaan."*[14]. Kvantitatiivinen tarkastelu sähköopin keskittyy vuosiluokilla 7-9 siis sähköopin perussuureiden arvojen mittaamiseen ja suureiden välisten riippuvuuksien tarkasteluun. Sähkömagnetismin kä-

sittely tapahtuu ensisijaisesti kvalitatiivisesti, jolloin laskennalliset sähkömagnetismin liittyvät tehtävät eivät lähtökohtaisesti tule kyseeseen.

Sähköopin sisällöstä tarkasteltavaksi valitaan sisältöalue, joka on rajaukseltaan riittävän suppea ja edellyttää tyypillisesti tehtävänratkaisussa kvantitatiivista ratkaisua. Virtapiirien kvantitatiivinen tarkastelu mainitaan opetussuunnitelmassa suureiden arvojen mittaamisen ja suureiden välisten yhteyksien tarkastelun osalta. Mitä tehtäviä ei ole tarkoituksenmukaista generoida, mutta Ohmin lakia käsittelevät tehtävät rajaavat sisällön riittävän tarkasti, jolloin tehtäviä voidaan verrata toisiinsa.

Perusopetuksen opetussuunnitelman perusteiden [14] kuvaukset sisältökokonaisuuksista eivät sisällä tarkkoja sisältöihin liittyvään osaamiseen kytkeytyviä tavoitteita. Myöskään oppiaineen yleiset tavoitteet eivät ole sisältöspesifejä. Tehtävien analyysin ja laatimisen keskeisiin tavoitteisiin ei lukeudu tehtävien analysointi suhteessa opetussuunnitelmaan. Generoitavien tehtävien osalta kielimallille annettavan kehotteen sisältämät maininnat yhteydestä opetussuunnitelmaan voidaan nähdä riittävänä opetussuunnitelman huomioimisen tasona, mikäli tehtävissä ei esiinny selkeästi opetussuunnitelmasta tai sen periaatteista poikkeavaa sisältöä. Tavanomaisessa summatiivisessä arvioinnissa ei voida huomioida opetussuunnitelmassa määriteltyjä sisältöjä ja tavoitteita yksittäisten tehtävien sisällä kokonaisuudessaan, joten tämänkaltaista kytköstä ei voida kielimalliltakaan vaatia. Lisäksi arvioinnin jatkuvan ja monimuotoisen luonteen vuoksi [16] opetussuunnitelman määrittelemien tavoitteiden täysimääräinen arviointi pelkästään summatiiviseen arviointiin nojautuen on haastavaa [26].

### 2.1.1 Tehtäväpankin tehtävät

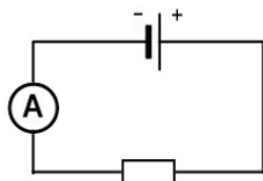
Kielimallilla generoiduille tehtäville täytyy löytää mielekäs verrokki. Käyttötilanteeltaan tekoälyllä generoitavien tehtävien voidaan ajatella olevan samankaltaisia

kuin oppikirjoja julkaisevien kustantamojen ylläpitämien tehtäväpankkien tehtävät. Kumpikin tapa laatia tehtäviä summatiiviseen arviointiin on oletusarvoisesti tehtävien itsenäistä laatimista nopeampaa. Valmiita tehtäviä poimitaan kahden kustantamon tehtäväpankeista. Tehtäväkirjastot ovat verkkopalveluissa jaoteltuina kirjasarjoittain ja tässä yhteydessä valitaan SanomaPro Oy:n ja e-Oppi Oy:n kustantamista materiaaleista soveltuvimmat. Sopivia tehtäviä suodatettaessa on huomioitava generoitaville tehtäville asetettavat rajoitukset. Valittavien tehtävien on käsiteltävä Ohmin lakia ja oltava luonteeltaan laskemista edellyttäviä. Lisäksi on suljettava pois tehtävät, joissa on useampi kuin kaksi alatehtävää.

E-Oppi Oy:n Opiq-alustalta löytyy suomeksi yksi perusasteen vuosiluokille 7–9 suunnattu oppikirja, Spektri 7-9 [72]. Verkkoalustan koetehtäväpankin tasavirtaa käsittelevästä aineistosta löytyy kaksi kriteerit täyttävää tehtävää, jotka on esitetty kuvissa 3 ja 4. Tehtävissä käytetään visuaalisena tukena tehtävänannolle tasavirtapiiriä esittäviä kytkentäkaavioita. SanomaPro Oy:n verkkopalvelusta löytyy perusopetuksen vuosiluokille kaksi kirjasarjaa, joista valittiin Ilmiö 7–9 Fysiikka [73]. Kirjasarjan koetehtäväpankista löytyi viisi sisällöltään rajaukseen sopivaa tehtävää, joista käsiteltäviksi valittiin yksi tehtävä. Tehtävistä mikään ei sisältänyt tekstin lisäksi muuta aineistoa ja tehtävien rakenne oli samankaltainen. Merkitään tehtävää "Vastuksen resistanssi on  $45 \Omega$  ja sen läpi kulkeva sähkövirta  $0,30 \text{ A}$ . Kuinka suuri on vastuksen jännitehäviö?" tunnuksella TP2.

Kaikki tarkastellut kirjasarjojen tehtäväpankkien valittuun aiheeseen liittyvät tehtävät ovat yksiosaisia, mikä tulee mahdollisesti huomioida kielimallin avulla tehtäviä generoitaessa. Kaksi- tai useampiosaisen tehtävän vertaaminen yksiosaiseen tehtävään voi tuottaa haasteita, sillä moniosaista tehtävää kokonaisuutena tarkasteltaessa taulukossa I esitettyjen kriteerien arvioiminen poikkeaa yksiosaisien tehtävien analyysistä, sillä aineisto on tehtävää kohden lähtökohtaisesti laajempi ja monipuolisempi. Tehtäväpankkien tehtävien lyhyys ominaispiirteenä voidaan kuitenkin

7. Kuinka suuri on virtamittarin lukema kuvan kytkennässä, kun vastuksen resistanssi on  $500\text{ m}\Omega$  ja pariston jännite  $1,5\text{ V}$ ?



\_/6

Kuva 3. Soveltuva tehtävä (TP1) oppikirjan Spektri 7-9 opettajan aineistosta [72].

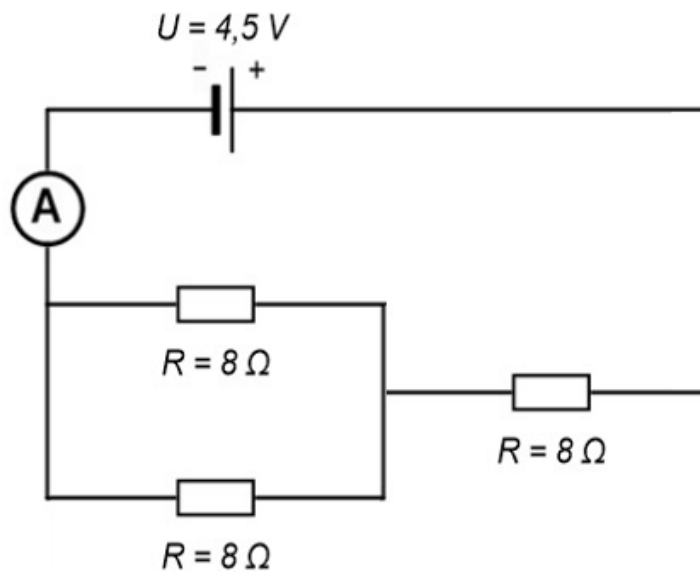
huomioida jo suunniteltaessa kehotetta kielimallille tehtävien generoimiseksi.

### 2.1.2 Kehotteen laatiminen

Kehotteen laatimisen tyypillisenä haasteena korostuvat usein annettavan kehotteen liialliset ihmisten välistä vuorovaikutusta muistuttavat piirteet [4]. Näiden piirteiden minimoiminen suurelle kielimallille annettavassa kehotteessa tapahtuu systemaattisen kehotteen suunnittelun kautta. Kehotetta rakennettaessa ei siis voida tukeutua yksinomaan intuitioon tai kehotteen muokkaamiseen valistuneen arvauksen kautta, mikäli saatu tulos ei ole toivottu [4]. Tueksi tarvitaan ymmärrystä kielimallien toiminnasta ja erityisesti prosessia, jonka perusteella kehote suunnitellaan, sillä kehotteen korkea laatu on suurten kielimallien tapauksessa suoraan kytköksissä saatavan tulosten korkeaan laatuun, kuten Knoth ym. [5] aiemman kirjallisuuden pohjalta toteavat. Kehotteen suunnittelu kytkeytyy tyypillisesti kiinteästi siihen, millaista tulosta tavoitellaan. Kehotteen laatiminen voidaan siis nähdä iteratiivisena prosessina, jossa kehotteen laatimista itsessään ohjaa tavoite saavutettavasta tulosteesta [58].

Kehotteen suunnitteluun on valittavissa useita toisistaan poikkeavia strategioita. Few-shot -strategiassa kehotteeseen sisällytetään useita esimerkkejä samankaltaisen kontekstin tehtävistä ja toivotuista vastauksista. Tehtävistä viimeinen on kielimallille todellisuudessa annettava tehtävä, johon ei kehotteessa anneta vastausta

10. Määritä virtamittarin lukema.



\_/6

Kuva 4. Soveltuva tehtävä (TP3) oppikirjan Spektri 7-9 opettajan aineistosta [72].

[74]. One-shot -strategia eroaa few-shotista siten, että annettavia esimerkkejä ja vastauksia on vain yksi, minkä jälkeen kielimallin tehtäväksi jää tuottaa viimeisimmän annetun tehtävän ratkaisu [74]. Zero-shot -mallissa kielimallille ei anneta valmista esimerkkiä tapauksesta, vaan pelkästään tehtävä tai ohje toimintaa varten [75].

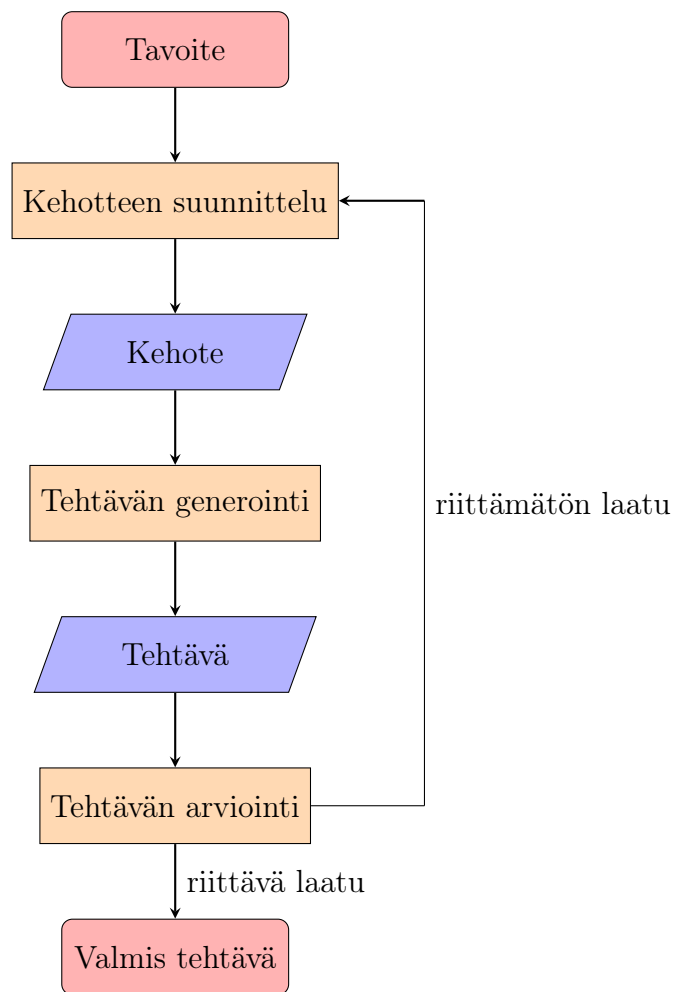
Chain-of-thought -strategiaa hyödynnettäessä kehoitteella pyritään kielimallia tukeamaan tuomalla esiin tulosteessa käytetty päättelyketju. Chain-of-thought -kehotteenlaadintaa voidaan hyödyntää few-shot -strategiaan sisällytettynä [76] tai osana zero-shot -kehotetta [75]. Suuria kielimalleja käytettäessä zero-shot -strategia on ajankäytöllisesti few-shot -kehotteenlaadintaa perustellumpaa, sillä suurten kielimallien kyky tuottaa laadukkaita tulosteita zero-shot -kehoitteilla on hyvä [75].

Suurta kielimallia käytettäessä zero-shot -lähestyminen on siis perusteltu, mitä puoltaa myös kielimallille annettavan tehtävän konteksti. Opetuskäyttöön summaatiivisen arvioinnin tehtävää generoitaessa opettajan kehotteen suunnitteluun käyttämä aika on todennäköisesti vähäinen, eikä few-shot -kehote ole erityisen toden-

näköinen. Tutkielmassa laadittavan eräänlaisen zero-shot -kehotteen laadun parantamiseksi käytetään kehotteen rakenteen runkona mallia, jossa Eager ja Brunton [58] jakavat kehotteen laatimisen prosessina kuuteen vaiheeseen. Ensimmäinen vaihe on päämäärän määrittäminen. Tässä yhteydessä kielimallin halutaan tuottavan fysiikan summatiiviseen arviointiin perusopetuksen vuosiluokkien 7-9 opetuksessa soveltuvan tehtävän, joka käsittelee Ohmin lakia. Toisena vaiheena mainitaan sisällön tyyppin ja muodon määrittäminen. Tehtävän halutaan olevan tekstimuodossa, joskin tehtävää tukevia kuvia ei ole tarpeen rajata pois.

Kolmantena kehotteen laatimisen vaiheena Eager ja Brunton [58] esittävät ensimmäisen kehotteen laatimisen. Tätä seuraa vaihekaavion mukaan ensimmäisen kehotteen testaaminen, jossa kehotteen kielimallille syöttäminen tuottaa ensimmäisen tuotoksen. Ensimmäisen tuotoksen generointia seuraa tuotoksen arviointi ja kehotteen iteratiivinen muokkaus, kunnes saavutettu lopputulos vastaa riittävän hyvin asetettuja tavoitteita. Tehtävän generoinnissa kehotteen iteratiiviseen kehitykseen ei käytetä valmista kriteeristöä, vaan kehotetta muokkaamalla pyritään tarvittaessa generoimaan tehtäviä, jotka vastaavat aiemmalla kehotteen versiolla generoituja tehtäviä paremmin kehotteessa asetettuja reunaehtoja. Tehtävien generoinnin valmistuttua arvioidaan tehtävät kriteeristön mukaisesti, mikä tuottaa tietoa myös kehotteen laatimisen prosessin onnistumisesta. Eagerin ja Bruntonin [58] kehotteen laatimisen mallin pohjalta muotoiltu tehtävän laatimisen kehoteprosessi on esitetty kuvassa 5.

Eager ja Brunton [58] esittävät kuusiosaisen mallin avulla generatiiviselle tekoälylle annettavan kehotteen osat. Mallia mukailemalla voidaan laatia tehokas kehotte fysiikan summatiivisessä arvioinnissa käytettävien tehtävien luomiseksi. Mallin mukainen suomennos kehotteen osista yhdessä annettavan kehotteen osien kanssa on esitetty taulukossa II. Tehtävää generoitaessa sopiva kehotteessa annettava verbi on "laadi". Ihmisen suorittamana toimenpiteenä kyse olisi tehtävän laadinnasta, joten



Kuva 5. Kehoteprosessin prosessikaavio.

Taulukko II. Kehotteen osat Eagerin ja Bruntonin [58] mukaan.

Osa	Merkitys
Verbi	Ilmaisee halutun toiminnan
Fokus	Määrittää verbin tuottaman toiminnan kohteen tai lopputuloksen
Konteksti	Määrittää toiminnan rajaukset
Fokuksen ehto	Määrittää aihealueen ja tavoitteen
Linjaus	Sovittaa generoitavan sisällön tavoitteen mukaiseksi
Rajoitukset	Määrittää rajoitukset, joiden rajaamana generointi tapahtuu

luonnollisella kielellä käytävää keskustelua mukailevalle kielimallille verbivalinta on luonteva. Fokuksena voidaan käyttää yksinkertaisesti sanaa "tehtävä", sillä laajempi konteksti määritellään mallin mukaisesti seuraavassa vaiheessa.

Fokusta seuraa Eagerin ja Bruntonin [58] mallin mukaan konteksti. Tarkoituksena on generoida summatiiviseen arviointiin soveltuva fysiikan tehtävä suomalaisen perusopetuksen vuosiluokilla 7-9, joten kehotteen osa "fysiikan summatiiviseen arviointiin suomalaisen perusopetuksen vuosiluokilla 7-9" täyttää kontekstin määritelmän mukaisen tehtävän. Kehotteen ensimmäiseksi virkkeeksi muotoutuu siis "Laadi tehtävä fysiikan summatiiviseen arviointiin suomalaisen perusopetuksen vuosiluokilla 7-9."

Seuraavien kolmen kehotteen osan tarkoituksena on rajata ja tarkentaa ensimmäisen virkkeen sisältöä. Fokuksen ehto rajoittaa annettua fokusta halutun lopputuloksen mukaiseen sisältöön. Tehtävän suhde opetussuunnitelmassa esitettyihin sisältöihin on tässä yhteydessä syytä täsmentää. Fokuksen ehdoksi asetetaan "Tehtävän tulee käsitellä Perusopetuksen opetussuunnitelman perusteiden (2014) mukaisen sisältökokonaisuuden S6 Sähkö mukaisesti Ohmin lakia." Linjauksessa korostetaan tehtävän suhdetta opetussuunnitelmaan: "Tehtävässä tulee huomioida Perusopetuksen opetussuunnitelman perusteiden (2014) esittämät fysiikan oppiaineen vuosiluokien 7-9 yleiset tavoitteet."

Rajoitukset-osassa esitetään muut halutun lopputuloksen kannalta oleelliset asiat, joilla rajataan mahdollisten lopputulosten joukkoa. Jotta tehtävät ja oppikirjan tehtävät olisivat vertailukelpoisempia, on tarpeen esittää rajoitus tehtävän pituudelle. Ensimmäiseksi osaksi rajoituksia muotoillaan "Tehtävä ei saa sisältää useampaa kuin kaksi alakohtaa (kuten a ja b)." Toisena rajoituksena esitetään vaatimus tehtävän laskemista vaativasta luonteesta: "Tehtävän ratkaisemiseksi on tarvittava laskemista."

Kokonaiseksi kehoitteeksi Eagerin ja Bruntonin [58] mallin mukaisesti kehoitteen suunnittelulla muodostuu "Laadi tehtävä fysiikan summatiiviseen arviointiin suomalaisen perusopetuksen vuosiluokilla 7-9. Tehtävän tulee käsitellä Perusopetuksen opetussuunnitelman perusteiden (2014) mukaisen sisältökokonaisuuden S6 Sähkö mukaisesti Ohmin lakia. Tehtävässä tulee huomioida Perusopetuksen opetussuunnitelman perusteiden (2014) esittämät fysiikan oppiaineen vuosiluokkien 7-9 yleiset tavoitteet. Tehtävä ei saa sisältää useampaa kuin kaksi alakohtaa (kuten a ja b). Tehtävän ratkaisemiseksi on tarvittava laskemista."

Kehotteella generoiduista tehtävistä voidaan havaita viiden kuudesta alakohdasta alkavan sanalla "laske". Lisäksi kaikissa kolmessa tapauksessa generoidut tehtävät ovat kaksiosaisia, mikä vaikuttaa johtuvan annetun kehoitteen pituusrajoituksesta enintään kahteen alakohtaan. Tehtävien vertailukelpoisuuden kehittämiseksi on harkittava kehoitteen muokkausta muotoon, jossa tehtävien alakohdat kielletään kokonaan.

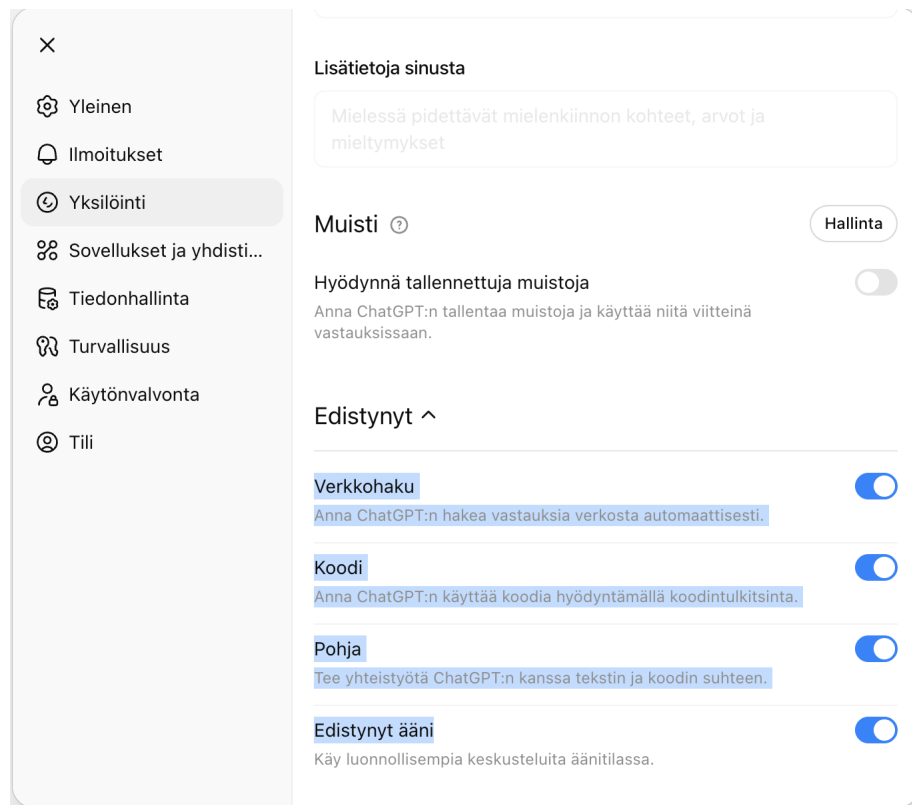
### 2.1.3 Generoidut tehtävät

Syöttämällä laadittu kehote ChatGPT:n keskustelunäkymään kielimalli tuottaa tulosteita, jotka sisältävät tutkielmassa analysoitavat generoidut tehtävät. Esiteltyjen kehoitteen laadinnan periaatteiden mukaisesti [58] on mahdollista, että generoitujen tehtävien laadun tarkastelun perusteella kehotetta joudutaan muokkaamaan ha-

lutunlaisten tehtävien tuottamiseksi. Huomiota kiinnitetään tässä kehoteprosessin vaiheessa erityisesti tehtäville asetettuihin laajuusvaatimuksiin, jotta saavutettaisiin riittävän tiivis summatiivisen arvioinnin tehtävän muoto. Kehotteen mahdollisista muokkauksista huolimatta kaikkien kehotteen versioiden tehtävät on mahdollista ottaa tehtäville suoritettavan sisällönanalyysin kohteeksi. Aineiston kokoa rajataan siten, että kullakin mahdollisella kehotteen versiolla generoidaan kolme tehtävää.

Ennen tehtävien generoimista tarkastellaan ChatGPT:n asetuksia, jotka voivat vaikuttaa generoitavaan sisältöön. Tilin asetuksista kytketään valinnat "Ota muokaus käyttöön" ja "Hyödynnä tallennettuja muistoja" pois päältä. Edistyneistä asetuksista säilytetään asetukset "Verkkohaku", "Koodi", "Pohja" ja "Edistynyt ääni" aktiivisina, sillä näiden neljän asetuksen aktiivisuuden ei kuvassa 6 esitettyjen kuvausten perusteella pitäisi käyttää aiempaa käyttäjän toimintaa avukseen. Kehotteella Perusopetuksen opetussuunnitelman perusteisiin [14] viitattaessa voidaan asetuksen "Verkkohaku" perusteella myös olettaa kielimallin tarvittaessa hakevan asiakirjan verkosta hyödynnettäväksi. Tiedonhallinnan asetuksista asetukset "Auta parantamaan mallia kaikille" kytketään pois päältä.

Tehtäviä generoitaessa jokaisen tehtävän generoinnin yhteydessä tarkistettiin käytettävä kielimalli ChatGPT:ltä kysymällä. Kaikissa tapauksissa käytetyksi malliksi ilmoitettiin GPT-5. Ensimmäiset kolme tehtävää generoitiin kehotteella "Laadi tehtävä fysiikan summatiiviseen arviointiin suomalaisen perusopetuksen vuosiluokilla 7-9. Tehtävän tulee käsitellä Perusopetuksen opetussuunnitelman perusteiden (2014) mukaisen sisältökokonaisuuden S6 Sähkö mukaisesti Ohmin lakia. Tehtävässä tulee huomioida Perusopetuksen opetussuunnitelman perusteiden (2014) esittämät fysiikan oppiaineen vuosiluokkien 7-9 yleiset tavoitteet. Tehtävä ei saa sisältää useampaa kuin kaksi alakohtaa (kuten a ja b). Tehtävän ratkaisemiseksi on tarvittava laskemista.", kukin tehtävä omassa keskustelussaan. Kuvissa 7, 8 ja 9 on esitetty ensimmäisellä kehotteen versiolla generoidut tehtävät.



Kuva 6. Tehtävien generoinnissa käytetyt ChatGPT:n edistyneet asetukset.

Kaikkien kolmen tehtävän tapauksessa kielimalli tuotti varsinaisen tehtävän ohella runsaasti kyseiseen tehtävään liittyvää oheismateriaalia, kuten summauksen tehtävän suhteesta opetussuunnitelmaan, tehtävän arviointiin liittyviä ohjeita ja jopa malliratkaisun. Tutkielman tarkoituksena ei kuitenkaan ole tarkastella erityisesti kielimallin tuottamaa tulostetta kokonaisuutena, sillä suurin tutkimuksellinen pääpaino on tehtävien laadun ja piirteiden analysoinnissa. Tästä syystä jo aineiston kokoamisvaiheessa leikataan pois kielimallin tuottamaan tulosteeseen sisältyvä teksti, joka ei kuulu generoidun tehtävän tehtävänantoon.

Kolme ensimmäistä generoitua tehtävää täyttävät sinänsä kehotteella tehtäville asetetut rajoitukset ja halutut ominaisuudet, mutta tehtäväpankin tehtävien yksiosaisuus ei tule riittävällä tavalla huomioiduksi, kun tehtävien osien lukumäärä rajataan kahteen. Tuntemattomasta syystä kielimalli vaikuttaa tuottavan valitulla kehotteella lähtökohtaisesti kaksiosaisia tehtäviä, mikä oli ainoastaan tehtävän pi-

**Tehtävä: Lamppu ja Ohmin laki**

Kun taskulampun lamppu kytketään paristoon, sen läpi kulkee sähkövirta. Lamppu noudattaa Ohmin lakia. Lamppuun kytketyn jännitteen ja sen läpi kulkevan virran välinen yhteys on kuvattu seuraavasti:

Jännite (V)	Virta (A)
1,5	0,30
3,0	0,60

- a) Laske lampun resistanssi käyttäen Ohmin lakia.  
 b) Kuinka suuri virta kulkee lampun läpi, jos sen napoihin kytketään 4,5 V:n jännite?

Kuva 7. Ensimmäinen ensimmäisellä kehotteella ChatGPT:llä generoiduista tehtävistä.

**Tehtävä: Lamppu ja vastus virtapiirissä**

Koulun fysiikan laboratoriossa tutkitaan virtapiiriä, jossa on virtalähde, lamppu ja säädettävä vastus. Kun vastuksen arvoa muutetaan, myös virtapiiriin läpi kulkeva sähkövirta muuttuu.

Mittauksissa saatiin seuraavat arvot:

Jännite U (V)	Virta I (A)
2,0	0,40
3,0	0,60
4,0	0,80

- a) Laske vastuksen R arvo Ohmin lain avulla jokaisessa mittauksessa ja päättele, onko vastus ohminen (eli pysyykö sen resistanssi vakiona).  
 b) Laske, kuinka suuri sähköteho (P) kehittyy vastuksessa, kun jännite on 4,0 V.

Kuva 8. Toinen ensimmäisellä kehotteella ChatGPT:llä generoiduista tehtävistä.

**Tehtävä (1 tehtävä, korkeintaan 2 alakohtaa)**

Pieni taskulamppu saa virtansa 4,5 V paristosta, jonka sisäinen resistanssi on 0,50  $\Omega$ .

Taskulampun hehkulangan (polttimon) resistanssi on 2,50  $\Omega$ .

a) Laske virtapiirin läpi kulkeva virta ja polttimon napajännite (jännite polttimon yli). Käytä tarvittaessa Ohmin lakia ja ota huomioon pariston sisäinen resistanssi.

b) Laske polttimossa hetkellinen teho ja energiankulutus 10 minuutin yhtäjaksoisessa käytössä. Ilmaise energiankulutus sekä jouleina että wattitunteina (Wh). Pyöristä vastaukset sopivasti ja merkitse yksiköt.

Kuva 9. Kolmas ensimmäisellä kehotteella ChatGPT:llä generoiduista tehtävistä tuudelle asetettu yläraja, ei tehtävän vaadittu laajuus. Tehtävien keskinäisen vertailun helpottamiseksi muokataan kehotetta siten, että moniosaiset tehtävät rajataan kokonaan pois. Tämä tapahtuu helpoiten muokkaamalla kehotteen Eagerin ja Bruntonin [58] mallin mukaista rajoitusten osaa, jonka tehtävän pituutta koskeva raja-  
 oli alkuperäisessä kehotteessa laadittu muotoon "Tehtävä ei saa sisältää useampaa kuin kaksi alakohtaa (kuten a ja b)." Muokataan virke seuraavasti, jolloin raja-  
 edellistä kapeampi ja generoitavien tehtävien tulisi olla yksiosaisia: "Tehtävä ei saa sisältää alakohtia (kuten a ja b)." Nyt toinen versio kehotteesta on "Laadi tehtävä fysiikan summatiiviseen arviointiin suomalaisen perusopetuksen vuosiluokilla 7-9. Tehtävän tulee käsitellä Perusopetuksen opetussuunnitelman perusteiden (2014) mukaisen sisältökokonaisuuden S6 Sähkö mukaisesti Ohmin lakia. Tehtävässä tulee huomioida Perusopetuksen opetussuunnitelman perusteiden (2014) esittämät fysiikan oppiaineen vuosiluokkien 7-9 yleiset tavoitteet. Tehtävä ei saa sisältää alakohtia (kuten a ja b). Tehtävän ratkaisemiseksi on tarvittava laskemista."

Uuden kehotteen tehtävät laaditaan vastaavilla ChatGPT:n asetuksilla kuin ensimmäiset kolme tehtävää. Toisella kehotteen versiolla generoitiin ensimmäisen kehotteen tavoin kolme tehtävää erillisissä keskusteluissaan. Kaikissa kolmessa tehtävässä käytetyksi kielimalliksi ilmoitettiin GPT-5. Toisella kehotteella generoidut tehtävät on esitetty kuvissa 10, 11 ja 12. Generoidut tehtävät ovat kielimallille syöttees-

**Tehtävä**

Kytke tasavirran lähde (paristo) jännitteellä 12,0 V sarjaan kahden vastuksen kanssa: ulkoiset vastukset 24,0  $\Omega$  ja 12,0  $\Omega$ . Paristolla on sisäinen vastus 0,50  $\Omega$ . Selvitä laskemalla: (1) piirin sähkövirran suuruus, (2) jännitehäviöt jokaisen ulkoisen vastuksen yli ja pariston sisäisen vastuksen yli, (3) teho, joka muuttuu lämmöksi kussakin ulkoisessa vastuksessa ja pariston sisäisessä vastuksessa, (4) kuinka paljon energiaa (joulea) pariston sisäisessä vastuksessa muuttuu lämmöksi yhden tunnin aikana, ja (5) lyhyt (1–3 lauseen) perusteltu arvio siitä, aiheuttaako piiri merkittävää kuumenemistä ja mitä turvallisuus- tai käyttöhuomioita opettaja/oppilas tulisi tässä tapauksessa ottaa huomioon. Kirjaa kaikki laskut ja yksiköt selvästi.

Kuva 10. Ensimmäinen toisella kehotteella laadituista tehtävistä.

**Summatiivinen arviointitehtävä (v. 7–9): Ohmin laki ja sähkön laskeminen**

Yhden 9,0 V pariston, sarjaan kytketyn kiinteän vastuksen ja hehkulampun muodostamassa yksinkertaisessa virtapiirissä sarjavastus on 180  $\Omega$  ja lampun resistanssi on 200  $\Omega$  silloin kun lampun läpi kulkee noin 20 mA. Sovella Ohmin lakia ja sähkötehon laskentaa ja vastaa seuraaviin laskutehtäviin: laske piirin kokonaisresistanssi ja piirissä kulkeva sähkövirta, laske lampun yli oleva jännite ja lampussa syntyvä teho, sekä määritä kuinka monta ohmia täytyy olla sarjavastuksessa, jotta saman pariston kanssa piirin virta olisi 30,0 mA (olettaen että lampun resistanssi säilyy 200  $\Omega$ ). Perustele laskusi, merkitse yksiköt ja pyöristä luvut sopivasti (kolme merkitsevää numeroa riittää).

Kuva 11. Toinen toisella kehotteella laadituista tehtävistä.

sä asetetun rajoituksen mukaisia, sillä ne eivät suoraan sisällä alakohtia. Tehtävien sisältöä tarkasteltaessa voidaan kuitenkin havaita tehtävien sisältävän useita erillisiä määritettäviä suureita tai kysymyksiä. Verrattaessa tehtäviä tehtäväpankkien tehtäviin on ilmeistä, että tehtävien laajuus poikkeaa merkittävästi tehtäväpankkien tehtävistä.

Tehtäväpankin tehtävien ja generoitujen tehtävien keskenään poikkeavan laajuuden takia täytyy kehotetta edelleen muokata tehtävien lyhentämiseksi. Muutoksen on oltava luonteeltaan vastausta rajoittava, joten Eagerin ja Bruntonin [58] mallin mukaisesti muokataan kehotteen rajoituksia käsittelevää osaa lisäämällä kehottee-

**Tehtävä — Ohmin laki ja vastusryhmät**

Kytke piirissäsi paristoon 12,0 V. Pariston plusnavasta lähtee sarjaan vastus  $R_3 = 30,0 \Omega$ , jonka jälkeen piiri jakautuu kahdeksi rinnakkaiseksi haaraksi, joissa on vastukset  $R_1 = 50,0 \Omega$  ja  $R_2 = 100,0 \Omega$ . Laske seuraavat suureet käyttäen Ohmin lakia ja tarvittaessa sarja- ja rinnankytkennän peruslaskuja: piirissä kulkeva kokonaisvirta paristosta, rinnakkaisosan kokonaisvastus, piirissä olevan rinnakkaisosan jännite, rinnakkaisosan haarojen virrat sekä  $R_3$ :n tehohäviö. Ilmaise kaikki vastaukset sopivien yksiköiden ja pyöristä kolmeen merkitsevään numeroon.

Kuva 12. Kolmas toisella kehotteella laadituista tehtävistä.

seen tarkennus "eikä näitä vastaavia useita erillisiä tehtäviä tehtävän sisällä". Muita osia kehoitteesta ei aiempien generoitujen tehtävien laadun perusteella ole tarpeen muokata, joten kehoitteen kolmanneksi versioksi muodostuu: "Laadi tehtävä fysiikan summatiiviseen arviointiin suomalaisen perusopetuksen vuosiluokilla 7-9. Tehtävän tulee käsitellä Perusopetuksen opetussuunnitelman perusteiden (2014) mukaisen sisältökokonaisuuden S6 Sähkö mukaisesti Ohmin lakia. Tehtävässä tulee huomioida Perusopetuksen opetussuunnitelman perusteiden (2014) esittämät fysiikan oppiaineen vuosiluokkien 7-9 yleiset tavoitteet. Tehtävä ei saa sisältää alakohtia (kuten a ja b), eikä näitä vastaavia useita erillisiä tehtäviä tehtävän sisällä. Tehtävän ratkaisemiseksi on tarvittava laskemista."

Kolmannella kehoitteella generoiduissa tehtävissä käytetty kielimalli on vaihtunut GPT-5.1 -malliin. Tehtävien pituutta voidaan nyt pitää aiemmillä kehoitteilla saatuja tuloksia sopivampana, eikä yksikään tehtävä sisällä aiempien generoitujen tehtävien tapaan useita määritettäviä suureiden arvoja tai pitkää tehtävälistaa. Tehtävien analyysivaiheessa tapahtuvan käsittelyn helpottamiseksi merkitään kuvassa 13 esitettyä tehtävää tunnuksella GAI7, kuvan 14 tehtävää tunnuksella GAI8 ja kuvan 15 tehtävää tunnuksella GAI9. Koska ensimmäiset kuusi tehtävää on generoitu käyttäen kielimallia GPT-5, ei niitä voida suoraan pitää vertailukelpoisina tehtävien GAI7, GAI8 ja GAI9 kanssa. ChatGPT ei salli nykyisessä muodossaan kielimallin vapaata valintaa, joten aiemmat kaksi kehoitteen versiota on tarpeen ajaa kielimallin GPT-5.1 läpi vertailukelpoisten tehtävien tuottamiseksi. Näin voidaan tarkastella tehtävien generoinnin ja kehoitteen kehittämisen prosessia yhtä kielimallia käyttäen, jolloin saadut tulokset myös kuvastavat paremmin mahdollista todellisessa opetus-käytössä esiintyvää tilannetta.

Alkuperäisellä kehoitteen versiolla generoidaan siis kielimallia GPT-5.1 käyttäen kolme erillistä tehtävää omissa keskusteluissaan. Kehoitteen ensimmäisen version mukaiset kielimallilla GPT-5.1 generoidut tehtävät on esitetty kuvissa 16, 17 ja 18.

**Summatiivisen arvioinnin tehtävä (Fysiikka, vuosiluokat 7–9)****Aihe: Ohmin laki (S6 Sähkö)**

Sähköpiirissä on virtalähde ja yksi vastus. Kun vastuksen napojen välinen jännite on 4,5 V, mitataan piirissä kulkevaksi sähkövirraksi 0,30 A. Laske vastuksen resistanssi ja arvioi laskemasi perusteella, miten vastuksen koko vaikuttaa virtapiirissä havaittavaan sähkövirtaan. Perustele vastauksesi fysiikan käsitteitä käyttäen.

Kuva 13. Ensimmäinen kolmannella kehotteella laadituista tehtävistä (GAI7).

**Tehtävä (fysiikka, vuosiluokat 7–9 – sisältökokonaisuus S6 Sähkö / Ohmin laki)**

Koulun teknisen työn luokassa huomataan, että eräessä laitteessa käytettävä vastus kuumenee epätavallisen paljon. Laitetta testattaessa siihen kytketään tasajännitelähde, jonka napajännitteeksi mitataan 9,0 V. Mittauksen aikana vastuksen läpi kulkeva sähkövirta on 0,30 A. Arvioi, onko vastus todennäköisesti kunnossa laskemalla sen resistanssi Ohmin lain avulla ja vertaamalla saamaasi arvoa valmistajan ilmoittamaan nimellisarvoon 27  $\Omega$ . Kerro laskun avulla, päätelläänkö laitteen olevan toimiva vai viittaako tulos mahdolliseen vikaan.

Kuva 14. Toinen kolmannella kehotteella laadituista tehtävistä (GAI8).

**Tehtävä:**

Tutki yksinkertaista sähköpiiriä, jossa on tasajännitelähde ja yksi vastus. Piiriin kytketään 9,0 V:n jännite, ja vastuksen läpi mitataan kulkevan virran suuruudeksi 0,30 A. Laske vastuksen resistanssi Ohmin lain avulla. Esitä laskusi ja perustele, miten sovellat fysiikan käsitteitä ja periaatteita ratkaistessasi tehtävää.

Kuva 15. Kolmas kolmannella kehotteella laadituista tehtävistä (GAI9).

**Summatiivinen arviointitehtävä (luokat 7–9) — Ohmin laki (S6: Sähkö)****Tehtävä**

Kytke oheisen kuvan mukaisesti virtapiiriin jännitelähde ja vastus  $R$  sekä ampeeri- mittari sarjaan. Kun kytkentä on tehty, mitataan, että virtapiiriin läpi kulkee  $0,30\text{ A}$  ja käytetty jännitelähteen napajännite on  $9,0\text{ V}$ .

a) Laske vastuksen  $R$  arvo käyttäen Ohmin lakia. Ilmaise vastaus oikeassa yksikössä ja näyttäjä käytetyt kaavat ja yksikkömuunnokset.

b) Laske piirin vastuksessa vapautuva teho  $P$ . Arvioi laskusi perusteella, voiko siinä käyttää tavallista  $0,25\text{ W}$  ( $1/4\text{ W}$ ) -teholuokiteltua vastusta ilman ylikuumentumisriskiä — perustele vastauksesi laskennallisesti.

Kuva 16. Ensimmäinen kehotteen ensimmäisellä versiolla kielimallilla GPT-5.1 generoiduista tehtävistä (GAI1).

Käytetty kehote on "Laadi tehtävä fysiikan summatiiviseen arviointiin suomalaisen perusopetuksen vuosiluokilla 7-9. Tehtävän tulee käsitellä Perusopetuksen opetussuunnitelman perusteiden (2014) mukaisen sisältökokonaisuuden S6 Sähkö mukaisesti Ohmin lakia. Tehtävässä tulee huomioida Perusopetuksen opetussuunnitelman perusteiden (2014) esittämät fysiikan oppiaineen vuosiluokkien 7-9 yleiset tavoitteet. Tehtävä ei saa sisältää useampaa kuin kaksi alakohtaa (kuten a ja b). Tehtävän ratkaisemiseksi on tarvittava laskemista."

Generoidut kolme tehtävää ovat kielimallin versiolla GPT-5 generoitujen ensimmäisen kehotteen tehtävien tavoin kaksi- tai useampiosaisia ja erityisesti tehtävä GAI3 on lukuisine määritettävine suureineen merkittävän laaja verrattuna tehtäväpankkien tehtäviin. Kaikkien kolmen tehtävän tulosteet sisältävät aiempien generoitujen tehtävien tavoin merkittävän määrän muuta tehtävään liittyvää tekstiä, joka suljetaan pois varsinaisesta aineistosta. Kielimallia GPT-5 käytettäessä ilmenyt tarve muokata kehotetta on tehtävien laajuuden perusteella läsnä kehotteen ensimmäisellä versiolla myös mallilla GPT-5.1 tehtäviä tuotettaessa.

Kehotteen alkuperäinen muokkaus voidaan kielimallin version vaihtumisesta huolimatta säilyttää samana kuin aiemmin, sillä ensimmäisten kolmen kielimallilla GPT-5.1 generoitujen tehtävien osalta puutteet ilmenivät samankaltaisina kuin kielimallilla GPT-5 generoitaessa. Toisena kehotteen versiona käytetään kehotetta: "Laadi tehtävä fysiikan summatiiviseen arviointiin suomalaisen perusopetuksen vuosiluokilla 7-9. Tehtävän tulee käsitellä Perusopetuksen opetussuunnitelman perusteiden (2014) mukaisen sisältökokonaisuuden S6 Sähkö mukaisesti Ohmin lakia. Tehtävässä tulee huomioida Perusopetuksen opetussuunnitelman perusteiden (2014) esittämät fysiikan oppiaineen vuosiluokkien 7-9 yleiset tavoitteet. Tehtävä ei saa sisältää useampaa kuin kaksi alakohtaa (kuten a ja b). Tehtävän ratkaisemiseksi on tarvittava laskemista."

### Tehtävä (yhteensä 10 pistettä)

Kytke seuraavat komponentit sarjaan: paristo 12,0 V, vastus  $R_1 = 100 \Omega$  ja vastus  $R_2 = 200 \Omega$ .

a) Laske piirin kokonaisvastus ja piirin virta. (4 p.)

b) Laske vastuksen  $R_2$  jännitehäviö ja teho, jonka  $R_2$  muuttaa lämmöksi. (6 p.)

Kuva 17. Toinen kehotteen ensimmäisellä versiolla kielimallilla GPT-5.1 generoituista tehtävistä (GAI2).

Paristossa on jännite  $U = 12,0 \text{ V}$ . Paristoon on kytketty vastus  $R_1 = 24 \Omega$ .

a) Laske vastuksen  $R_1$  läpi kulkeva sähkövirta ja vastuksen häviöteho. Käytä Ohmin lakia  $U = IR$  ja teholle kaavoja  $P = UI$  tai  $P = I^2R$ . Merkitse laskut selkeästi ja ilmoita vastaukset sopivin yksiköihin.

b) Kuvitellaan, että pariston rinnalle kytketään toinen vastus  $R_2 = 12 \Omega$  siten, että  $R_1$  ja  $R_2$  ovat rinnan liitettynä. Laske tällöin

- piirin kokonaisvastus  $R_{\text{eq}}$ ,
- paristosta lähtevä kokonaisvirta  $I_{\text{total}}$ ,
- kunkin vastuksen läpi kulkevat virrat  $I_1$  ja  $I_2$ ,
- ja kunkin vastuksen häviötehot  $P_1$  ja  $P_2$ .

Perustele laskusi ja ilmoita vastaukset sopivin yksiköihin.

Kuva 18. Kolmas kehotteen ensimmäisellä versiolla kielimallilla GPT-5.1 generoituista tehtävistä (GAI3).

killä 7-9. Tehtävän tulee käsitellä Perusopetuksen opetussuunnitelman perusteiden (2014) mukaisen sisältökokonaisuuden S6 Sähkö mukaisesti Ohmin lakia. Tehtävässä tulee huomioida Perusopetuksen opetussuunnitelman perusteiden (2014) esittämät fysiikan oppiaineen vuosiluokkien 7-9 yleiset tavoitteet. Tehtävä ei saa sisältää alakohtia (kuten a ja b). Tehtävän ratkaisemiseksi on tarvittava laskemista." Toisella kehotteversiolla generoidut tehtävät, joiden generoinnissa on käytetty kielimallia GPT-5.1 on esitetty kuvissa 19, 20 ja 21. Tehtävät ovat edelleen pitkiä ja moniosaisia tehtäväpankin tehtäviin verrattuna, joten kehotteen muokkaus ja tehtävien GAI7, GAI8 ja GAI9 generointi on tarpeen myös kielimallilla GPT-5.1 generoitujen kehotteen toisen version mukaisten tehtävien perusteella. Erityisenä piirteenä tehtävissä GAI5 ja GAI6 on havaittavissa tehtävän pituuteen liittyvän alakohtien rajoituksen kiertämistä tekemällä tehtävästä poikkeuksellisen laaja.

### Summatiivinen arviointitehtävä: Ohmin laki ja sähköturvallisuus

Polkupyörän etuvaloon kytketään 6,0 voltin paristo. Lampun läpi kulkeva sähkövirta on mitattu, ja sen suuruudeksi saadaan 0,40 ampeeria.

Oppilas pohtii, onko lamppu turvallinen käyttää myös 9,0 voltin paristolla.

Laske ensin lampun sähköinen vastus Ohmin lain avulla ja määritä sen jälkeen, kuinka suuri sähkövirta kulkisi lampun läpi 9,0 voltin paristolla. Arvioi lopuksi laskujesi perusteella, onko 9,0 voltin paristo turvallinen tälle lampulle, kun tiedetään, että lamppu rikkoutuu, jos virta ylittää 0,60 ampeeria. Perustele vastauksesi laskuihin ja fysiikan käsitteisiin nojautuen.

Kuva 19. Ensimmäinen kehotteen toisella versiolla kielimallilla GPT-5.1 generoiduista tehtävistä (GAI4).

Pariston nimellijännite on 12,0 V ja sen napoihin on kytketty sarjaan kaksi vastusta: lampun vastus 4,0  $\Omega$  ja säädettävä vastus, jonka arvo on säädetty 8,0  $\Omega$ . Piirissä oleva mittari, jonka resistanssi on niin suuri että sitä voi olettaa vaikutukseltaan mitätön, mittaa kokonaisvirran. Laske ja perustele seuraavat asiat käyttäen Ohmin lakia (muista yksiköt ja selkeä laskujärjestys): kuinka suuri sähkövirta kulkee piirissä, millä jännitteellä lampun navat ovat pariston napojen suhteen (eli lampun yli oleva jännite), kuinka suuri jännite on säädettävän vastuksen yli, ja kuinka paljon tehoa lamppu muuntaa lämmöksi/valoksi. Perustele lopuksi lyhyesti (maks. 4 lauseen verran) miten ja miksi lamppuun kohdistuva teho muuttuisi, jos pariston jännite olisi kaksinkertainen.

Kirjaa kaikki laskut näkyviin ja kirjoita loppuun yksi lyhyt turvallisuus- tai käytännönmittaushuomio, joka liittyy kotilaboratorioon tai luokka-asteen koesuunnitteluun (maks. 2 lausetta).

Kuva 20. Toinen kehotteen toisella versiolla kielimallilla GPT-5.1 generoiduista tehtävistä (GAI5).

#### Tehtävä — Virtapiiri ja Ohmin laki

Kotiaskareissa käytettävä paristokäyttöinen työvalo koostuu yhdestä 9,0 V paristosta ja lampusta, jonka resistanssi on 18,0  $\Omega$ . Lampun rinnalle kytketään tarkoituksellisesti sarja: säätövastus, jonka resistanssi on 10,0  $\Omega$ , jotta valoa voidaan himmentää. Piirissä oletetaan että paristo voi toimittaa tarvittavan jännitteen vakiona ja liitännöiden resistanssi on mitätön.

Laske ja esitä selkeästi: sähkövirran suuruus lampun alkuperäisessä piirissä ilman säätövastusta, lampun teho tässä tilanteessa, sähkövirran ja lampun tehon suuruus kun säätövastus 10,0  $\Omega$  on kytketty sarjaan, kuinka suuri osa pariston tehosta muuttuu lämmöksi säätövastuksessa (wattimäärä), ja selkeä laskelmaan perustuva arvio siitä, muuttuuko lampun valoteho merkittävästi säätövastuksen lisäämisen seurauksena. Perustele lopuksi yhden lauseen mittainen turvallisuus- tai käytännön huomio liittyen komponenttien lämmöntuottoon tai pariston kuormitukseen.

Kuva 21. Kolmas kehotteen toisella versiolla kielimallilla GPT-5.1 generoiduista tehtävistä (GAI6).

## 2.2 Sisällönanalyysi

Tehtävien keskinäiseen vertailuun ja tavoitteen mukaiseen ChatGPT:n käytön arviointiin summatiivisen arvioinnin tehtävien laadinnassa tarvitaan tehokas ja järjestelmällinen tapa analysoida tekstiä. Laadullinen sisällönanalyysi on tekstipohjaisen aineiston analyysimenetelmä, jota käyttäen tekstiä voidaan luokitella erilaisiin kategorioihin, joita yhdistävät yhtenevät merkitykset [77]. Laadullisen sisällönanalyysin avulla tehtävien sisältöä voidaan siis järjestää tutkimuskysymyksiin vastaamisen kannalta suotuisampaan muotoon. Laadullista sisällönanalyysia ei voida pitää yksittäisenä tai selvärajaisena menetelmänä, vaan paremminkin laajana joukkona menetelmiä, joita yhdistävät tietyt tyypilliset piirteet [77].

Laadullisen sisällönanalyysin menetelmät voidaan jakaa karkeasti kahteen pääryhmään sisällön koodaamisen ja sisällöstä muodostettavien luokkien mukaan. Sisällön koodaamisella tarkoitetaan tekstiaineiston luokittelua systemaattisesti ja johdonmukaisesti erilaisiin kategorioihin, joiden muodostamista ohjaavat tutkimuskysymykset ja tutkimuksen tarkoitus [78]. Aineistolähtöisessä eli induktiivisessä sisällönanalyysissa aineistosta nousevat havainnot ohjaavat sisällön koodausta, kun taas teorialähtöisessä eli deduktiivisessä sisällönanalyysissa taustalla vaikuttava, sovellettavissa oleva teoria ohjaa koodausprosessia ja koodauksessa käytettävien kategorioiden muodostamista [77]. Aineiston koodaaminen on siis osa sisällönanalyysia valitusta lähestymistavasta riippumatta. Aineistolähtöisen ja teorialähtöisen sisällönanalyysin eräänlainen välimuoto on teoriaohjaava eli abduktiivinen sisällönanalyysi, jossa yhdistyy induktiivisen ja deduktiivisen sisällönanalyysin piirteitä [79].

### 2.2.1 Valittu menetelmä

Aineistolähtöisen sisällönanalyysin suhde olemassa olevaan teoriaan on lähtökohtaisesti minimaalinen, sillä valmiin teorian osuudeksi analyysissa voidaan lukea ainoastaan analyysin toteuttamiseen liittyvä teoria, jonka ei tulisi vaikuttaa analyysissa

tehtäviin havaintoihin [80]. Teoriaohjaavassa analyysissä sen sijaan teoria voidaan kytkeä osaksi analyysin toteutusta myös analyysin kohteena olevan sisällön osalta, kun taas teorialähtöinen analyysi on suoraan alisteinen valmiille tai olemassa olevan teorian pohjalta muotoillulle mallille, jonka pohjalta analyysi toteutetaan [80]. Jos analyysin prosessia kuvataan liikesuuntina, induktiivisessa sisällönanalyysissä sen voidaan nähdä kulkevan havainnoista kohti teoriaa, kun taas deduktiivisessä lähestymistavassa suunta on päinvastainen. Teoriaohjaavassa sisällönanalyysissä analyysin liikesuunta tyypillisesti vaihtelee havaintojen ja teorian välillä, minkä vuoksi abduktiivisen sisällönanalyysin prosessista on mahdollista havaita induktiivisiä ja deduktiivisiä osia [79].

Tässä tutkielmassa analyysia ohjaa olemassa olevan teorian pohjalta muotoiltu analyysikehys, johon analyysin kohteena olevat tehtävät pyritään analyysissä sovitamaan. Mayring [81] kuvailee sisällön deduktiivista luokittelua kategorioihin laadullisen sisällönanalyysin mahdollisesti käytetyimmäksi menetelmäksi. Deduktiivinen sisällönanalyysi voidaan jakaa kolmeen vaiheeseen, joista ensimmäisessä tekstin osat jaetaan tarkasti määriteltyihin kategorioihin, toisessa kategoriat kiinnitetään valittuihin ja perusteltuihin esimerkkeihin jokaisesta kategoriasta ja kolmannessa muodostetaan säännöt sisällön koodaamiseksi luotuihin kategorioihin [81]. Kerätylle tehtäväaineistolle sisällönanalyysia suoritettaessa seurataan joiltain osin vastaavaa prosessia. Suoritettavaa analyysia voidaan siis pitää toisaalta teorialähtöisenä, mutta muodostetun kriteeristön ohella analyysissä huomioidaan aineistosta esiin nousevia ominaispiirteitä ja tarve sovittaa analyysikehystä huomioimaan aineiston luonnetta ja sisältöä. Tämän vuoksi suoritettava analyysi on pikemminkin teoriaohjaavaa kuin teorialähtöistä.

## 2.2.2 Laadittujen kriteerien soveltaminen

Kirjallisuuden pohjalta luvussa 1.3.2 muodostetut neljä kriteeriä, joiden suhteen tehtäviä analysoidaan ovat kognitiivinen vaativuustaso, kokeellisen menetelmän tuntemus, fysiikan oikeellisuus ja selkeys. Kunkin kriteerin on tarkoitus olla itsenäinen osa sisällönanalyysin avulla tehtävistä koostettavaa kokonaisuutta. Kriteerikohtainen sisällön koodaus tapahtuu siis yksitellen ja kriteeri kerrallaan. Mayringin [81] käsittelemää mallia mukaillen kriteerien kategoriat pyritään kiinnittämään esimerkkeihin ennen tehtävien koodaamista kategorioihin. Varsinaiset kategoriat, joihin tehtävät luokitellaan, on muodostettu kirjallisuuden pohjalta luvussa 2.

Kategoriat ja tehtävien luokittelukriteerit on osin muodostettu kirjallisuuden pohjalta ennen aineiston keräämistä. Tästä huolimatta analyysissa huomioidaan sisällöstä nousevat piirteet kategorioiden rajoja ja sääntöjä muodostettaessa. Kirjallisuuden pohjalta voitiin päätellä fysiikan oikeellisuuden ja selkeyden kriteerien olevan lähellä toisiaan ja mahdollisesti osin päällekkäiset. Aineiston tarkastelun perusteella voidaan todeta, että kriteerien luvussa 2 esitettyä määrittelyä selkeämpi rajaaminen on tarpeen kriteerien itsenäisyyden turvaamiseksi.

Jokaisen kriteerin osalta siis tarkistetaan teorian pohjalta muodostetun kriteerikehyksen soveltuvuus koodauksen pohjaksi, minkä jälkeen sisältöön nojaten muodostetaan kriteerikohtainen säännöstö tehtävien koodaamiseksi. Säännöstöä soveltaen tehtävät koodataan kriteerikohtaisiin kategorioihin, joiden yhtenevyys kirjallisuudessa esitettyihin kriteereihin varmistetaan vielä sisällön koodauksen jälkeen. Kielimallilla GPT-5 generoidut tehtävät jätetään sisällönanalyysin avulla suoritettavan tarkastelun ulkopuolelle, sillä vastaavilla kehotteilla kielimallilla GPT-5.1 generoidut tehtävät kattavat kaikilla kolmella kehotteen mallilla generoidut tehtävät. Tarkoituksena ei ole myöskään syventyä kielimallin versioiden välisten erojen vaikutukseen tehtävien ominaisuuksiin, mikä puoltaa version GPT-5.1 tuottamien tehtävien GAI1-GAI9 tarkasteluun keskittymistä.

### 2.2.3 Kognitiivinen vaativuustaso

Kognitiivisen vaativuustason kriteerin toteutumista analysoitaessa käytetään soveltuvilta osin koodauksessa apuna Woitkowskin [33] taulukossa 2 esitettyä luokittelujärjestelmää tehtävän kognitiivisen vaikeustason määrittämiseksi neljän luokan välillä. Käytetyt koodaukset kriteereineen on esitetty taulukossa III. Tehtävistä yksikään ei edustanut kahta matalinta kognitiivisen haastavuuden tasoa, osin johtuen tehtäville asetetusta kriteeristä, jonka mukaan tehtävän täytyy olla luonteeltaan laskutehtävä. Tehtäville käytettyä kognitiivisen vaativuustason kriteeriä luokituksineen voidaan tästä huolimatta pitää perusteltuna, sillä kriteerin avulla voidaan kuvata laskutehtävien haastavuuseroja nojaten fysiikan tehtäville tyypillisempiin tekijöihin.

Tehtäviä analysoitaessa käytettiin taulukossa III kuvatulla tavalla apuna tehtävien komentoverbejä, joista tehtävien GAI1 komentoverbi "kytke" sivuutettiin paikkaansapitämättömänä ja tehtävän GAI9 "tutki" luettiin samoin merkitykseltään epäselväksi. Tehtävässä GAI1 tehtävän onnistunut ratkaisu ei edellytä kytkentöjen tekemistä, vaan todellisiksi komentoverbeiksi arvioitiin "laske", "arvioi" ja "perustele", joista kahden jälkimmäisen edellyttämät ratkaisun osat määrittivät tehtävän kognitiiviseksi vaativuustasoksi tason C4. Tehtävän GAI9 todellisiksi komentoverbeiksi todettiin "laske" ja "perustele", mutta tehtävässä edellytetyn perustelun ollessa luonteeltaan suoraviivainen ja yksinkertainen voitiin tasoksi määrätä C3.

Osa kielimallilla tuotetuista tehtävistä on moniosaisia, vaikka tehtävien pituutta pyrittiin kehoitetta muokkaamalla rajoittamaan. Moniosaisten tehtävien luokittelu asettaa haasteita kriteeristön soveltamiselle erityisesti, jos tehtävien osat eivät ole välittömästi toisiinsa kytkeytyviä. Moniosaisten tehtävien tapauksessa tehtäviä tarkastellaan analyysissa kokonaisuutena ja huomioidaan kognitiivisena vaativuustasona tehtävän korkeinta vaativuustasoa edustanut kokonaisuus tai kahden osatehtävän tuottama monimuuttujainen keskinäisriippuvuus, kuten tehtävässä GAI1. Tehtävän a-kohta ei itsenäisenä kokonaisuutena edustaisi monimuuttujaisen keskinäis-

riippuvuuden luokkaa, mutta b-osan monivaiheinen ja syvempää tarkastelua vaativa tehtävä nostaa tehtävän monimuuttujaisen keskinäisriippuvuuden luokkaan. Mikäli tehtävän b-osassa olisi annettu tarvittavat lähtötiedot ja osaa olisi tarkasteltu yksittäisenä tehtävänä, olisi se luokiteltu luokkaan C4. Koska tehtävän GAI1 korkeinta kognitiivisen vaatavuustason luokkaa edustanut osa kuului luokkaan C4, luokitellaan tehtävä myös kokonaisuutena tuohon luokkaan.

Taulukko III: Tehtävien analysointiin käytetty koodausohje kognitiivisen vaativuustason osalta.

<b>Koodi</b>	<b>Kuvaus</b>	<b>Piirre tekstissä</b>	<b>Esimerkki tehtävästä</b>
Faktat (C1)	Ratkaisu edellyttää ainoastaan tunnetun faktan, kaavan tai määritelmän toistamista.	"Kerro", "Selitä"	
Prosessit (C2)	Ratkaisu edellyttää ainoastaan prosessin kuvausta sanoin, luonnoksen tai diagrammin avulla.	"Esitä", "Piirrä", "Kuvaile"	
Lineaarinen kausa- liteetti (C3)	Ratkaisu edellyttää ainoastaan yksittäistä, suoraviivaista perustelua tai laskua.	"Laske", "Perustele", "Sovella"	TP1
Monimuuttujainen keskinäis- riippuvuus (C4)	Ratkaisu edellyttää useamman tekijän punnitsemista ja/tai monimutkaisempaa perustelua tai laskua.	"Laske", "Arvioi"	GAI8

#### 2.2.4 Kokeellisen menetelmän tuntemus

Kokeellisen menetelmän kriteerin soveltamisessa huomioitiin erityisesti tehtävänannossa esitetty tilanne ja mahdolliset vaaditut kytkökset työturvallisuuteen tai kokeellisiin järjestelyihin. Käytännön kokeellisia harjoitteita edellyttäviä tehtäviä ei aineistossa esiintynyt, vaikka tehtävien GAI1 ja GAI2 kehotukset kytkentöjen tekemiseen viittaavat kyseiseen tehtävätyyppiin. Yksinkertaistetusti voidaan todeta tehtävän tulevan luokitelluksi luokkaan K1, mikäli tehtävän ratkaisun osana edellytetään kokeellisen työskentelyn tuntemusta ja taitoja tehtävän ratkaisemiseksi. Kokeellisen menetelmän osalta sisällönanalyysin koodausohje on esitetty taulukossa IV.

Kriteerin osalta generatiivisen tekoälyn tuottamissa tehtävissä esiintyy epäjohdonmukaisia viittauksia kokeelliseen työskentelyyn, kuten komentoverbi tehtävässä GAI9 oleva "tutki". Pelkkiä tehtävänannon komentoverbejä tarkastelemalla ei voida suoraan havaita, että tehtävä hyödyntää kokeellista menetelmää. Tehtävissä opilaalta odotetun toiminnan perusteella voidaan saada käsitys siitä, edellyttääkö tehtävän ratkaiseminen kokeelliseen menetelmään liittyvää toimintaa.

Tehtävissä GAI1 ja GAI2 kehoitetaan tehtävän GAI9 tavoin käytännön toimintaan, vaikka tehtävän ratkaisu ei perustu fyysisiin kytkentöihin tai tutkimuksiin. Tehtävät alkavat kehottavalla verbillä "kytke", mutta tehtävien fysiikka ei edellytä kytkentöjen tekemistä tai välttämättä edes kytkentäkaavion laatimista. Tehtävät GAI2 ja GAI9 luokiteltiin kokeellista menetelmää käsittelemättömien tehtävien luokkaan K0, mutta tehtävä GAI1 sen sijaan katsottiin kokeelliseksi tehtäväksi sen sisältämän kokeellisen asetelman ja riskienhallinnan näkökulman vuoksi.

Tehtävänpankkien tehtävissä TP1, TP2 ja TP3 ei havaittu kokeellisen menetelmän elementtejä, joiden perusteella tehtäviä olisi voitu luokitella luokkaan K1. Tehtävien TP1 ja TP3 sisältämistä kytkentäkaavioista huolimatta tehtävien luonne on lähinnä laskennallinen, joskin erityisesti tehtävässä TP3 edellytetään myös kyt-

kentäkaavion merkintöjen tarkempaa tulkintaa. Koodauksen pohjana on käytetty Mesicin ja Muratovicin [37] määrittelyn mukaista kriteerin kehystä, jota on täsmennetty vastaamaan aineistossa esiin nousseisiin erityispiirteisiin, kuten turvallisuusnäkökulmien korostumiseen joissain tehtävänannoissa.

Taulukko IV: Tehtävien analysointiin käytetty kokeellisen menetelmän koodausohje.

Koodi	Kuvaus	Piirre tekstissä	Esimerkki tehtävästä
Kokeellinen (K1)	Tehtävän ratkaiseminen vaatii kokeellisen työskentelyn taitoja ja osaamista.	Turvallisuusnäkökulmat, kysymykset kokeellisista järjestelyistä.	GAI1
Ei kokeellinen (K0)	Tehtävä voidaan ratkaista ilman kokemusta kokeellisesta työskentelystä.	Suoraviivaiset laskutehtävät, ei selkeää kytköstä kokeelliseen työskentelyyn.	TP2

### 2.2.5 Fysiikan oikeellisuus

Tehtävien fysiikan oikeellisuuden kriteeriä tarkastellaan erityisesti fysiikkaan suoraan liittyvässä sisällössä. Virheet tekstin kieliasussa tai ymmärrettävyydessä voidaan sivuuttaa, mikäli tehtävän fysiikkaan suoraan kytkeytyvä sisältö on tulkittavissa tehtävänannosta. Küchemannin ym. [36] kriteeristön spesifisyyden kriteeriä mukaileva kriteeri arvioidaan arvoksi 0, mikäli tehtävän fysiikkaan liittyvissä tie-

doissa on puutteita ja arvoksi 1, jos tehtävän fysiikka on sisällöltään oikeellista ja koherenttia. Kriteerin koodausohje on esitetty taulukossa V.

Fysiikan oikeellisuuden kriteerin mukaisessa sisällönanalyysissä havaitut poikkeamat keskittyvät generatiivisella tekoälyllä tuotettuihin tehtäviin. Tehtävän GAI1 tehtävänannossa viitataan tehtävään liitettyyn kuvaan, jota ei kuitenkaan ole olemassa. Tämä voidaan tulkita puutteeksi tehtävän fysiikkaan kytkeytyvässä sisällössä, minkä vuoksi tehtävälle annetaan kriteerin arvoksi 0. Lisäksi tehtävässä mainitaan vastuksen  $R$  arvo vastuksen resistanssin sijaan, mikä on fysikaalisesti epätasällinen ilmaus.

Tehtävät GAI3 ja GAI4 sijoittuvat analyysissä luokkaan O1, sillä niiden fysikaalinen tausta on esitetty oikein. Tehtäviä yhdistää kuitenkin tavanomaisesta poikkeava käsitteistö, joka voisi mahdollisesti tuottaa hämmennystä tehtävää ratkaistaessa. Tehtävässä GAI3 käytetään yleisemmin käytössä olevan tehohäviön käsitteestä ilmausta häviöteho ja tehtävässä GAI4 pyydetään laskemaan lampun sähköinen vastus sanan resistanssi käyttämisen sijaan. Molemmissa tapauksissa kyse on harvoin käytetyistä synonyymeistä, joita epätavanomaisuudestaan huolimatta voidaan hyvin käyttää.

Tehtävän GAI5 selostus jännitteistä pariston napojen ja virtapiiriin kytkettyjen laitteiden yli on kokonaisuutena tulkinnanvarainen, erityisesti ilman tehtävänantoon liitettävää selventävää kuvaa. Tässä tapauksessa kyse ei ole kuitenkaan tehtävän selkeydessä olevasta puutteesta, vaan tehtävän fysiikan oikeellisuuden puutteesta, sillä puutteellinen teksti liittyy suoraan tehtävänannossa tarjottavaan fysikaaliseen taustaan.

Taulukko V: Tehtävien analysointiin käytetty koodausohje fysiikan oikeellisuuden osalta.

Koodi	Kuvaus	Piirre tekstissä	Esimerkki tehtävästä
Oikeellinen (O1)	Tehtävän fysiikka on oikeellista ja koherenttia.	Fysiikan käsitteiden käyttö on oikeaa ja tehtävässä annettavat fysiikkaan liittyvät tiedot ovat oikein.	TP1
Ei oikeellinen (O0)	Tehtävän fysiikkaan liittyvässä sisällössä on puutteita.	Virheet käsitteiden käytössä, puuttuvia tai ristiriitaisia tietoja.	GAI7

### 2.2.6 Selkeys

Selkeyden kriteerin suhteen suoritettu analyysi keskittyy Küchemannin ym. [36] selkeyden (*clarity*) kriteerin mukaisesti tehtävänannon tekstin selkeyden ja ytimekkyyden tarkasteluun. Tehtävistä erityisesti kielimallin tuottamat tehtävät sisälsivät ratkaisua tyypillisesti yleisellä tasolla ohjaavaa tekstiä, joka ei automaattisesti ole tehtävän selkeyttä heikentävä tekijä. Tehtävän GAI9 ohjaava teksti yhdistettynä tehtävän alussa annettuun kehotukseen "tutki" tuottaa kuitenkin merkittäviä haasteita tehtävän selkeydelle ja ymmärrettävyydelle. Tehtävän lopussa esiintyvä ratkaisua

ohjaava virke "Esitä laskusi ja perustele, miten sovellat fysiikan käsitteitä ja periaatteita ratkaistessasi tehtävää." yhdistettynä tehtävän alun kehotukseen tuottaa ristiriidan tehtävän ratkaisemiseen.

Tehtävien selkeyttä tarkasteltaessa pääpaino oli erityisesti tehtävän kokonaisuutena hahmottamista tukevien ja heikentävien seikkojen havaitsemisessa. Tehtävän GAI6 tapauksessa pienet kirjoitusvirheet, kuten "liitännöiden resistanssi" eivät vaikuta tehtävän luettavuuteen kokonaisuutta merkittävästi heikentäen ja tehtävän ollessa muutoin ymmärrettävä ja tarpeeksi tiivistä muotoiltu, sijoittui tehtävä luokkaan S1. Tehtävän GAI9 lisäksi tehtävät GAI1 ja GAI2 olivat ainoat luokkaan S0 sijoittuvat tehtävät. Tehtäviä yhdistää alun kehottavan verbin ristiriita tehtävän todellisen oletetun ratkaisun kanssa ja tehtävän GAI1 tapauksessa tehtävänantoon kuuluvan kuvan puuttuminen.

Tehtävät GAI1 ja GAI9 sijoittuvat sekä luokkaan S0 että O0 osin samojen tehtävänannon puutteiden takia. Selkeyden ja fysiikan oikeellisuuden kriteereillä on siis joitain yhtymäkohtia näissä tehtävissä, mutta kumpikin kriteeri toimii itsenäisesti osana laajempaa tehtävien analyysia, sillä erot muiden tehtävien luokitteluissa osoittavat kahden erillisen kriteerin tarpeen. Selkeyden kriteeri keskittyy erityisesti tehtävien luettavuuden tarkasteluun, kun taas fysiikan oikeellisuuden kriteerin avulla arvioidaan tehtävänannossa esitettyjen tietojen riittävyttä ja johdonmukaisuutta tehtävän fysikaalisen sisällön osalta.

Tehtävän ytimekkyyden arvioinnissa lyhyiden ratkaisua ohjaavien lisäysten, kuten tehtävän GAI7 "Perustele vastauksesi fysiikan käsitteitä käyttäen." tai tehtävän GAI5 "Laske ja perustele seuraavat asiat käyttäen Ohmin lakia (muista yksiköt ja selkeä laskujärjestys):..." katsottiin toimivan tehtävän ratkaisemista helpottavina ja siten perusteltuina osina tehtävänantoa. Tehtävät GAI5 ja GAI7 voitiin siis ohjaavan tekstin löytymisestä huolimatta katsoa sisältöön nähden riittävän tiiviiksi ja sijoittaa luokkaan S1.

Generoitujen tehtävien ja tehtäväpankkien tehtävien tehtävänantojen pituuden merkittävä ero tuottaa kysymyksen siitä, voidaanko selkeyden kriteeriä uskottavasti käyttää molempien vertailevaan analyysiin, jos selkeyden kriteerin osaksi on määritelty tehtävän ytimekkyys. Pidempien tehtävänantojen osalta analyysissä on huomioitu tehtävänanto kokonaisuutena niin, että tehtävän absoluuttisen pituuden sijaan huomio keskitetään erityisesti tehtävän ratkaisemisen kannalta oleellisen ja epäoleellisen sisällön suhteeseen. Tällöin pidempienkin tehtävien ytimekkyuden arviointi on linjassa tehtäväpankkien tehtäville samalla kriteerillä suoritettun analyysin kanssa.

Taulukko VI: Tehtävien selkeyden kriteerin analysointiin käytetty koodausohje.

Koodi	Kuvaus	Piirre tekstissä	Esimerkki tehtävästä
Selkeä (S1)	Tehtävänanto on selkeä, eikä sisällä epäselviä ohjeita tai osia.	Tehtävän sisältö on johdonmukaista, eikä siinä ole tarpeettomia osia.	GAI4
Ei selkeä (S0)	Tekstissä on harhaanjohtavia piirteitä.	"Tutki"(ei tutkimustehtävä), ylimääräistä tai harhaanjohtavaa ohjaavaa tekstiä	GAI9

### 3 Tulokset ja pohdinta

Kehotteen kehittämisen prosessin ja aineiston sisällönanalyysin tulosten voidaan katsoa olevan tämän tutkielman tärkeimpiä tuloksia. Tehtävien ja kehotteen laatiminen sekä aineistolle suoritettu sisällönanalyysi ovat kaksi selkeää tutkielman toiminnallista osakokonaisuutta. Näiden kokonaisuuksien tulosten kautta pyritään vastaamaan asetettuihin tutkimuskysymyksiin osakokonaisuuksia itsenäisinä, mutta kuitenkin laajemman kokonaisuuden osina tarkastellen.

#### 3.1 Ensimmäinen tutkimuskysymys

Taulukossa VII on esitetty tehtäville suoritettun sisällönanalyysin tulokset. Sisällönanalyysin pohjalta voidaan kuvailla tehtävien laatua ja ominaisuuksia tavalla, joka vastaa ensimmäiseen tutkimuskysymykseen, mutta samalla on huomioitava tarkastelun kattavuuden rajoitukset. Tulosten pohjalta voidaan kuitenkin muodostaa kapean tarkastelunäkökulman osalta kuvaus siitä, millaisia fysiikan summatiivisen arvioinnin tehtäviä generatiivisen tekoälyn avulla voidaan tuottaa.

Verrattaessa generatiivisella tekoälyllä tuotettuja tehtäviä tehtäväpankkien tehtäviin voidaan havaita, että tekoälyllä tuotetut tehtävät sijoittuvat huomattavan usein kognitiiviselta vaativuustasoltaan monimuuttujaisen keskinäisriippuvuuden luokkaan C4. Tarkastelluista tehtäväpankkien tehtävistä sen sijaan kaikki edustavat matalampaa kognitiivisen vaativuustason luokkaa C3, kun taas kielimallilla generoiduista tehtävistä seitsemän sijoittui tasolle C4 ja loput kaksi tasolle C3 eli lineaarisen kausaliteetin luokkaan. Eroa selittänee osin kielimallin taipumus tuottaa pidempiä ja laajempia tehtäviä, minkä vuoksi kehotetta jouduttiin muokkaamaan ennen lopullisten tehtävien GAI7-GAI9 generointia.

Kielimallilla tuotetuista tehtävistä viisi luokiteltiin kokeellista menetelmää sisältäväksi, mikä on huomattava ero tehtäväpankkien tehtäviin nähden. Tehtäväpankkien tehtävistä yksikään ei saanut kokeellisen menetelmän osalta luokitusta 1. Sel-

keitä poikkeamia tehtäväpankin tehtävien ja generoitujen tehtävien välillä havaitaan myös fysiikan oikeellisuuden ja selkeyden kriteereissä. Kummassakin kriteerissä havaitut tehtävien heikommat fysiikan oikeellisuuden luokitukset O0 ja selkeyden S0 esiintyivät yksinomaan kielimallin tuottamissa tehtävissä.

Küchemannin ym. [36] tutkimuksessa tehtäville suoritettun tarkastelun tuloksena tekoälyllä tuotetut tehtävät sijoittuivat selkeyden (*clarity*) ja kontekstin (*context*) kriteerien luokituksissa oppikirjan avulla laadittuja tehtävien matalammalle tasolle. Erityisesti Küchemannin ym. [36] käyttämä selkeyden kriteeri on määrittelyltään lähellä tässä tutkielmassa käytettyä selkeyden kriteeriä, joten tämän ja Küchemannin ym. tulosten voidaan todeta olevan keskenään linjassa selkeyden kriteerissä. Küchemannin ym. [36] tutkimuksessa yksiselitteisyyden (*not misleading*) ja oikeellisuuden (*correctness*) havaittiin olevan kielimallin avulla tuotetuissa tehtävissä samalla tai hieman korkeammalla tasolla kuin oppikirjan avulla tuotetuissa tehtävissä, mikä ei täysin vastaa fysiikan oikeellisuuden kriteerin osalta tässä tutkielmassa tehtyjä havaintoja.

Erot Küchemannin ym. [36] tutkimuksessa tehtyihin havaintoihin voivat selittyä osin sillä, että tässä tutkielmassa tehtäviä analysoi ainoastaan kirjoittaja, kun taas Küchemannin ym. [36] tutkimuksessa tehtävien analysoinnissa käytössä ollut asiantuntijuus on ollut laajempaa. Lisäksi on huomattava Küchemannin ym. [36] tutkimuksessa olleen kyseessä oppikirjan avulla laadittujen tehtävien käsittely, kun taas tässä yhteydessä käsiteltiin kielimallin tuottamien tehtävien vertailukohtana kustantajien julkaisemien tehtäväpankkien tehtäviä. Fysiikan perusasteen ylempien luokkien summatiivisen arvioinnin laskutehtävien tuottaminen on eneratiivisen tekoälyn käyttökohteena varsin spesifi. Suorien vertailukohtien puuttuessa tämä korostaa perusteellisemmän jatkotutkimuksen mahdollisuuksia.

Mesicin ja Muratovicin [37] havaitsema yhteys tehtävän vaikeuden ja tehtävän sisältämän kokeellisen menetelmän välillä jää otoksen suppeuden vuoksi tässä yh-

teydessä suoraan näyttämättä. Sisällönanalyysin perusteella on tosin huomattavissa, että kaikki tarkastellut tehtävät, joissa kokeellinen menetelmä esiintyi, sijoittuivat kognitiiviselta vaativuustasoltaan korkeampaan luokkaan. Kun tarkastellaan sisällönanalyysin tuottamia luokituksia kokonaisuutena, voidaan todeta, että generatiivisella tekoälyllä tuotetut tehtävät sisältävät useammin kokeellista menetelmää ja ovat kognitiiviselta vaativuustasoltaan korkeampia kuin tässä tutkielmassa käytetyt tehtäväpankkien verrokkitehtävät. Toisaalta kolmeen verrokkitehtävään nähden kielimallin tuottamat tehtävät ovat laadultaan fysiikan oikeellisuuden ja selkeyden osalta selvästi enemmän luokituksia O0 ja S0 sisältäviä, mikä tarkoittaa tehtävissä esiintyvän puutteita fysiikan oikeellisuudessa ja selkeydessä.

Kielimallin tuottamissa tehtävissä ilmennyt runsas kokeellisen menetelmän huomiointi sisältää fysiikan arviointia myös taitojen ja muun osaamisen arvioinnin suhteen. Fysiikan summatiiviseen arviointiin liitettynä työskentelytaitojen arviointia voidaan tehdä opettajan havainnointiin perustuen [17], mutta generatiivisella tekoälyllä tuotettavat käytännön työskentelyn osaamista kirjallisen tehtävän kautta mittaavat elementit voivat laajentaa fysiikan arvioinnin kattavuutta. Tästä seuraa, että arvioinnin sisältövaliditeetti voi vastaavasti kehittyä suotuisasti [13] suhteessa tehtäväpankkien koetehtäviin.

Taulukko VII. Tehtäville suoritettun sisällönanalyysin tuottamat luokitukset.

	<b>A/B</b>	Kogn. vaativuus C1/C2/C3/C4	Kok. menetelmä 0/1	Fys. oikeellisuus 0/1	Selkeys 0/1
Tehtäväpankki	TP1	C3	0	1	1
	TP2	C3	0	1	1
	TP3	C3	0	1	1
1. kehote	GAI1	C4	1	0	0
	GAI2	C3	0	1	0
	GAI3	C4	0	1	1
2. kehote	GAI4	C4	1	1	1
	GAI5	C4	1	0	1
	GAI6	C4	1	1	1
3. kehote	GAI7	C4	0	0	1
	GAI8	C4	1	1	1
	GAI9	C3	0	1	0

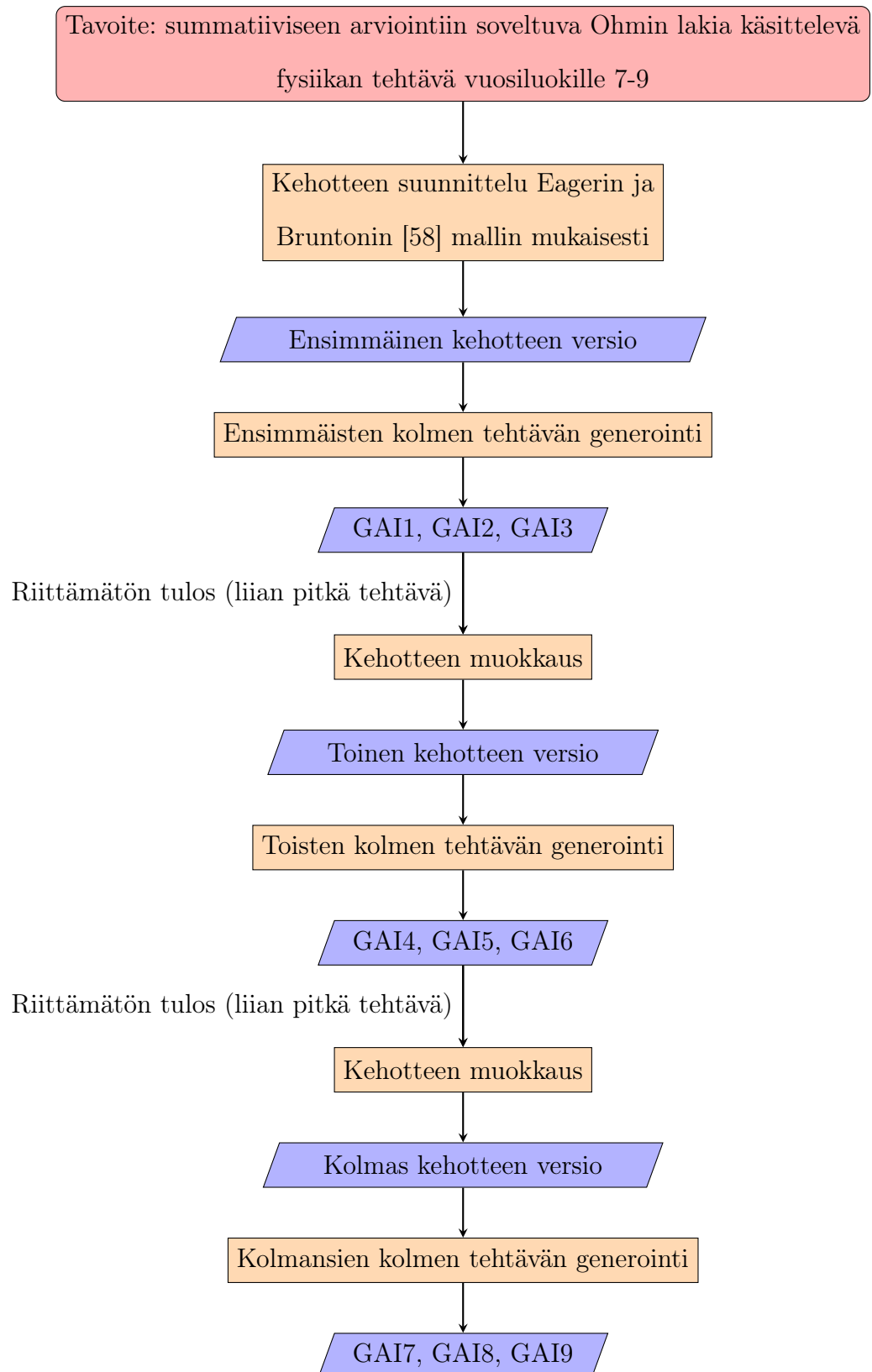
## 3.2 Toinen tutkimuskysymys

Toiseen tutkimuskysymykseen vastauksen antavana tuloksena voidaan pitää kehotteen laatimisen prosessin muotoilua ja toteutusta. Kuvassa 22 on kuvattu kehotteen laatimisen ja muokkaamisen prosessi, jonka tuloksena tarkastellut kielimallin tuottamat tehtävät syntyivät. Eagerin ja Bruntonin [58] mallin mukaista kehotteen rakennetta noudattavaa kehotetta muokattiin kahteen kertaan, jolloin viimeisiksi versioiksi jääneet tehtävät tuotettiin sen kolmannella versiolla. Tehtävien GAI7, GAI8 ja GAI9 vastatessa kehotesuunnittelulle asetettuun tavoitteeseen voitiin kehoteprosessi päättää kyseisten tehtävien generoinnin jälkeen.

Qiun ja Zhangin [63] korostama tekoälyyn pohjautuvien ratkaisujen asiantuntevan ja harkitun opetukseen integroinnin roolin voidaan tulosten perusteella nähdä korostuvan, sillä pelkästään halutun laajuisten tehtävien tuottaminen vaati Eagerin ja Bruntonin [58] esittämän kehotteen laatimisen prosessin mukaista kehotteen muokkausta ja arviointia kaksi kertaa. Havainto tukee TPACK:in [41] mukaista käsitystä opettajan roolista teknologian hyödyntämisestä opetuksessa. Opettajalta vaaditaan ymmärrystä kielimallien toiminnasta, jotta kehotteen suunnittelu on riittävän tehokasta [4], mutta toisaalta generoitujen tehtävien laadun arviointi edellyttää vahvaa pedagogisen sisältötiedon hallintaa [43].

## 3.3 Tulokset opettajan työn näkökulmasta

Tuloksena saadun kehotteen suunnitteluprosessin kuvaus herättää kysymyksen siitä, kuinka hyvin vastaava kehotteenlaadinnan prosessi voisi toteutua todellisessa opetukseen kytkeytyvässä tilanteessa. Kehotteen laadinnassa kokemattoman opettajan kohtaamat haasteet [4] ja toisaalta arvioinnille [14] sekä tekoälyn opetuskäytölle [82] asetetut rajoitteet voivat nostaa kynnyistä tai heikentää mahdollisuuksia käyttää generatiivista tekoälyä summatiivisen arvioinnin materiaalien tehokkaaseen tuottamiseen. Kielimallien tasosta ja kehotteen laadusta mahdollisesti johtuvat puutteet



Kuva 22. Kehotesuunnittelun prosessin kulku.

tehtävien laadussa edellyttävät opettajalta siis oppiaineen sisällön hyvän tuntemuksen lisäksi kykyä suhtautua kriittisesti kielimallien tuottamiin tehtäviin.

Generatiivinen tekoäly voidaan tulosten perusteella nähdä tulevaisuudessa eräänä vaihtoehtona tehtäväpankeille valittaessa tehtäviä summatiiviseen arviointiin. Generatiivisen tekoälyn käytöllä voidaan mahdollisesti saavuttaa ajankäytöllisiä etuja verrattuna tehtävien laadintaan ilman generatiivista tekoälyä tapahtuvaan ja sisällönanalyysin tulosten perusteella etuja sisällön monipuolisuuden osalta myös tehtäväpankkeihin nähden. Ajanhallinnallisen hyötynäkökulman saavuttaminen vaatii kuitenkin tehtävien generointiin oleellisesti kuuluvan kehoteprosessin kehittämistä, jolloin kehotteensuunnittelun iteraatioiden minimaalisella määrällä saadaan tuotettua halutunlaisia tehtäviä.

## 4 Johtopäätökset

Tutkielman tavoitteena oli ChatGPT:n hyödyntämismahdollisuuksien tarkastelu fysiikan summatiivisen arvioinnin tehtävien tuottamisessa. Tavoitetta lähestyttiin tehtävien generoinnin kehoteprosessin kuvaamisen ja generoiduille tehtäville suoritettun laadullisen sisällönanalyysin keinoin. Kehotteen tuottamisen ja kehoteprosessin dokumentoinnin kautta muodostui kuvaus systemaattisesta tavasta, jolla tehtäviä on mahdollista kehittää fysiikan summatiiviseen arviointiin suurta kielimallia käyttäen. Sisällönanalyysin avulla saatiin tulokseksi tuotettujen tehtävien ominaisuuksia ja tyypillisiä piirteitä kuvaileva kokonaisuus.

Sisällönanalyysin perusteella kielimallilla generoidut tehtävät olivat kirjasarjojen tehtäväpankkien tehtäviin verrattuna kognitiiviselta vaativuustasoltaan tyypillisesti haastavampia ja useammin kokeelliseen työskentelyyn viittaavia. Toisaalta sisällönanalyysissä havaittiin useiden generoitujen tehtävien laadussa puutteita yksiselitteisyyden ja oikeellisuuden kriteerejä käytettäessä. Vastaavia puutteita ei havaittu tehtäväpankkien tehtävissä. Näin ollen tulosten perusteella laadukkaiden fysiikan summatiivisen arvioinnin tehtävien tuottamisen edellytykset generatiivista tekoälyä hyödyntäen ovat olemassa, mutta tasokkaiden tehtävien valitsemiseen generoitujen tehtävien joukosta ja tehokkuudeltaan riittävään kehotteen laadintaan vaaditaan opettajalta kattavaa pedagogisen sisältötiedon hallintaa ja hyvää käsitystä suurten kielimallien toiminnasta. Käsitys kielimallien toiminnasta ja kehoteprosessin iteratiivisesta luonteesta on tarpeen, sillä suoritettun kehoteprosessin tarkastelun perusteella kehotteen hallitun muokkaamisen taito on keskeistä tehtäviä generatiivisella tekoälyllä tuotettaessa.

Opettajan rooli aktiivisena summatiivisen arvioinnin materiaalien tuottajana ja suorittamansa arvioinnin korkeiden standardien ylläpitäjänä korostuu, jos generoidut tehtävät ovat laadultaan vaihtelevampia kuin oppikirjasarjojen koetehtäväpankkien tehtävät. Tulevaisuudessa jatkuvan generatiivisen tekoälyn kehityksen myötä

avautuu mahdollisuuksia jatkotutkimukselle opettajien tehokkaan kehotteenlaadinnan taidoista oppi- ja arviointimateriaaleja tuottaessa. Tämä mahdollinen jatkotutkimuksen suunta voi osaltaan tukea generatiivisen tekoälyn hallittua tuomista osaksi opettajan ajallisia resursseja vapauttavaa työvälinevalikoimaa.

## Tekoälyn käyttö tutkielmassa

Tekoälyä on tutkielman tutkimustehtävän mukaisesti käytetty luvussa 2.1.3 esitettyjen generatiivisella tekoälyllä tuotettujen tehtävien laatimiseen. Kuvissa 7–12 esitetyt tehtävät on tuotettu OpenAI:n GPT-5 -kielimallilla. Kuvissa 13–21 esitetyt tehtävät on tuotettu OpenAI:n GPT-5.1 -kielimallilla. Lisäksi OpenAI:n kielimalleja GPT-5 ja GPT-5.1 on käytetty viitteiden tietojen muotoilemiseksi BibTeX-muotoon luettelointia varten.

## Viitteet

- [1] G. Polverini ja B. Gregorcic, *European Journal of Physics* **45**, 025701 (2024).
- [2] W. Yeadon, O. Inyang, A. Mizouri, A. Peach ja C. P. Testrow, *Physics Education* **58**, 035027 (2023).
- [3] S. Küchemann, K. E. Avila, Y. Dinc, C. Hortmann, N. Revenga, V. Ruf, N. Stausberg, S. Steinert, F. Fischer, M. Fischer *et al.*, *npj Science of Learning* **10**, 11 (2025).
- [4] J. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann ja Q. Yang, kirjassa *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23* (Association for Computing Machinery New York, NY, USA, 2023).
- [5] N. Knoth, A. Tolzin, A. Janson ja J. M. Leimeister, *Computers and Education: Artificial Intelligence* **6**, 100225 (2024).
- [6] A. Ngo, S. Gupta, O. Perrine, R. Reddy, S. Ershadi ja D. Remick, *Academic Pathology* **11**, 100099 (2024).
- [7] M. Stadler, A. Horrer ja M. R. Fischer, *GMS journal for medical education* **41**, Doc20 (2024).
- [8] W. Haney, kirjassa *Expanding Student Assessment*, toimittanut V. Perrone (Association for Supervision and Curriculum Development Alexandria, VA, 1991), pp. 142–163.
- [9] P. J. Black, *Studies in Science Education* **21**, 49 (1993).
- [10] P. Black ja D. Wiliam, *Assessment in Education: Principles, Policy & Practice* **5**, 7 (1998).
- [11] P. Black, M. Wilson ja S.-Y. Yao, *Measurement: Interdisciplinary Research & Perspective* **9**, 71 (2011).
- [12] A. Luostarinen ja N. Ouakrim-Soivio, kirjassa *Arvioinnin käsikirja*, toimittanut A. Luostarinen ja J. H. Nieminen (PS-kustannus Jyväskylä, 2019), pp. 33–48.
- [13] S. Messick, *American psychologist* **35**, 1012 (1980).
- [14] Opetushallitus, *Perusopetuksen opetussuunnitelman perusteet 2014: määräykset ja ohjeet 2014:964*, painos, PDF-julkaisu, Opetushallitus, 2014, iSBN 978-952-13-5998-9 (nid.), ISBN 978-952-13-5999-6 (pdf); ISSN-L 1798-887X; ISSN 1798-887X (painettu); ISSN 1798-8888 (verkkojulkaisu).
- [15] Opetushallitus, *Oppilaan oppimisen ja osaamisen arviointi perusopetuksessa, 2020*, saatavilla: <https://www.oph.fi>.
- [16] W. Harlen ja M. James, *Assessment in Education: Principles, Policy & Practice* **4**, 365 (1997).

- [17] D. D. Dixson ja F. C. Worrell, *Theory Into Practice* **55**, 153 (2016).
- [18] W. Harlen, *Research Papers in Education* **20**, 245 (2005).
- [19] J. Buick, *European Journal of Physics Education* **2**, 12 (2011).
- [20] H. E. Fischer, A. Kauertz, H. E. Fischer ja R. Girwidz, *Physics Education, Challenges in Physics Education* (Springer Berlin / HeidelbergSwitzerland, 2022), pp. 231–267.
- [21] D. Hestenes, M. Wells ja G. Swackhamer, *The Physics Teacher* **30**, 141 (1992).
- [22] M. Yorke, *Studies in Higher Education* **36**, 251 (2011).
- [23] C. Gipps ja G. Stobart, kirjassa *Educational Assessment in the 21st Century: Connecting Theory and Practice*, toimittanut C. Wyatt-Smith ja J. J. Cumming (Springer NetherlandsDordrecht, 2009), pp. 105–118.
- [24] S. Messick, *American Psychologist* **50**, 741 (1995).
- [25] T. J. Crooks, M. T. Kane ja A. S. Cohen, *Assessment in Education: Principles, Policy & Practice* **3**, 265 (1996).
- [26] N. Ouakrim-Soivio, väitöskirja, University of Helsinki, Helsinki, 2013.
- [27] M. Kane, *Handbook of test development* **1**, 131 (2006).
- [28] E. Almanasreh, R. Moles ja T. F. Chen, *Research in social and administrative pharmacy* **15**, 214 (2019).
- [29] S. Sireci ja M. Faulkner-Bond, *Psicothema* **26**, 100 (2014).
- [30] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill ja D. R. Krathwohl, *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain* (David McKay CompanyNew York, 1956).
- [31] D. R. Krathwohl, *Theory into practice* **41**, 212 (2002).
- [32] L. Shakhman ja M. Barak, *Eurasia Journal of Mathematics, Science and Technology Education* **15**, em1764 (2019).
- [33] D. Woitkowski, *International Journal of Science Education* **42**, 1585 (2020).
- [34] S. E. Embretson ja R. C. Daniel, *Psychology Science* **50**, 328 (2008).
- [35] J. B. Biggs ja K. F. Collis, *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)* (Academic PressNew York, 1982).
- [36] S. Küchemann, S. Steinert, N. Revenga, M. Schweinberger, Y. Dinc, K. E. Avila ja J. Kuhn, *Physical Review Physics Education Research* **19**, 020128 (2023).

- [37] V. Mesic ja H. Muratovic, Physical Review Special Topics—Physics Education Research **7**, 010110 (2011).
- [38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser ja I. Polosukhin, kirjassa *Advances in Neural Information Processing Systems*, toimittanut I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan ja R. Garnett (Curran Associates, Inc.Red Hook, NY, USA, 2017), Vol. 30.
- [39] OpenAI, Esittelyssä ChatGPT, [https://openai.com/fi-FI/index/chatgpt/?utm\\_source=chatgpt.com](https://openai.com/fi-FI/index/chatgpt/?utm_source=chatgpt.com), 2022, viitattu: 6.10.2025.
- [40] B. Woolf, AI in Education, 1991.
- [41] M. Koehler ja P. Mishra, Contemporary issues in technology and teacher education **9**, 60 (2009).
- [42] D. Baidoo-Anu ja L. O. Ansah, Journal of AI **7**, 52 (2023).
- [43] L. Shulman, Harvard educational review **57**, 1 (1987).
- [44] I. Celik, Computers in human behavior **138**, 107468 (2023).
- [45] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman ja A. Galstyan, ACM computing surveys (CSUR) **54**, 1 (2021).
- [46] E. U. A. for Fundamental Rights, *Getting the Future Right – Artificial Intelligence and Fundamental Rights* (Publications Office of the European Union-Luxembourg, 2020).
- [47] X. Zhai ja R. H. Nehm, Journal of Research in Science Teaching **60**, 1390 (2023).
- [48] S. J. Russell ja P. Norvig, *Artificial Intelligence: A Modern Approach*, fourth edition ed. (PearsonHoboken, NJ, USA, 2021).
- [49] A. M. Turing, Mind **59**, 433 (1950).
- [50] B. G. Buchanan, AI magazine **26**, 53 (2005).
- [51] T. M. Mitchell, AI magazine **18**, 11 (1997).
- [52] I. Goodfellow, Y. Bengio ja A. Courville, *Deep Learning* (MIT PressCambridge, MA, USA, 2016), <http://www.deeplearningbook.org>.
- [53] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn ja G. Kasneci, Learning and Individual Differences **103**, 102274 (2023).
- [54] N. Köbis ja L. D. Mossink, Computers in human behavior **114**, 106553 (2021).

- [55] N. Cong-Lem, A. Soyoof ja D. Tsering, *International Journal of Human-Computer Interaction* **41**, 3851 (2025).
- [56] OpenAI, GPT-5 System Card, viitattu: 4.2.2026.
- [57] OpenAI, GPT-5.1 Instant and GPT-5.1 Thinking System Card Addendum, viitattu: 15.2.2026.
- [58] B. Eager ja R. Brunton, *Journal of university teaching and learning practice* **20**, 1 (2023).
- [59] U. Mittal, S. Sai, V. Chamola *et al.*, *IEEE Access* **12**, 142733 (2024).
- [60] S. Bhat, H. A. Nguyen, S. Moore, J. C. Stamper, M. Sakr ja E. Nyberg, kirjassa *Proceedings of the International Conference on Educational Data Mining (EDM)* (International Educational Data Mining SocietyMA, USA, 2022).
- [61] M. Diliberti, H. L. Schwartz, S. Doan, A. K. Shapiro, L. Rainey ja R. J. Lake, *Using artificial intelligence tools in K-12 classrooms* (RAND Santa Monica, CASanta Monica, CA, USA, 2024).
- [62] Z. Ahmed, S. S. Shanto ja A. I. Jony, *STEM Education* **4**, 165 (2024).
- [63] K. Qiu ja T. Zhang, kirjassa *HCI International 2025 Posters*, toimittanut C. Stephanidis, M. Antona, S. Ntoa ja G. Salvendy (Springer Nature SwitzerlandCham, 2025), pp. 397–405.
- [64] V. Robledo-Rella ja B.-Y. Toh, kirjassa *Proceedings of the 2024 10th International Conference on E-Society, e-Learning and e-Technologies (ICSLT), ICSLT '24* (Association for Computing MachineryNew York, NY, USA, 2024), p. 68–75.
- [65] S. Frieder, L. Pinchetti, C. Chevalier, R.-R. Griffiths, T. Salvatori, T. Lukasiewicz, P. Petersen ja J. Berner, kirjassa *Advances in Neural Information Processing Systems*, toimittanut A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt ja S. Levine (Curran Associates, Inc.Red Hook, NY, USA, 2023), Vol. 36, pp. 27699–27744.
- [66] K. T. Kotsis, *EIKI Journal of Effective Teaching Methods* **2**, 18 (2024).
- [67] G. Cooper, *Journal of science education and technology* **32**, 444 (2023).
- [68] B. Gregorcic ja A.-M. Pendrill, *Physics Education* **58**, 035021 (2023).
- [69] J. Sweller, *Cognitive science* **12**, 257 (1988).
- [70] Y. Liang, D. Zou, H. Xie ja F. L. Wang, *Smart Learning Environments* **10**, 52 (2023).

- [71] K. W. Chan, F. Ali, J. Park, K. S. B. Sham, E. Y. T. Tan, F. W. C. Chong, K. Qian ja G. K. Sze, *Computers and Education: Artificial Intelligence* **8**, 100344 (2025).
- [72] J. Karkkulainen, O. Karkkulainen ja A. Kinnunen, *Spektri 7–9 Fysiikka: digioppimateriaali ja opettajan aineisto*, 2016.
- [73] H. Lehto, H. Salonen ja J. Maalampi, *Ilmiö 7–9 Fysiikka: Digiopetusmateriaali (OPS 2016)*, 2016.
- [74] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, *Advances in neural information processing systems* **33**, 1877 (2020).
- [75] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo ja Y. Iwasawa, *Advances in neural information processing systems* **35**, 22199 (2022).
- [76] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, *Advances in neural information processing systems* **35**, 24824 (2022).
- [77] H.-F. Hsieh ja S. E. Shannon, *Qualitative health research* **15**, 1277 (2005).
- [78] R. P. Weber, *Basic content analysis* (SageNewbury Park, CA, 1990), Vol. 49.
- [79] U. H. Graneheim, B.-M. Lindgren ja B. Lundman, *Nurse education today* **56**, 29 (2017).
- [80] J. Tuomi ja A. Sarajärvi, *Laadullinen tutkimus ja sisällönanalyysi*, uudistettu laitos. ed. (Kustannusosakeyhtiö TammiHelsinki, 2018).
- [81] P. Mayring, kirjassa *Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods*, toimittanut A. Bikner-Ahsbahs, C. Knipping ja N. Presmeg (Springer NetherlandsDordrecht, 2015), pp. 365–380.
- [82] Opetushallitus, *Tekoäly varhaiskasvatuksessa ja koulutuksessa – lainsäädäntö ja suositukset: Velvoitteet ja suositukset*, <https://www.oph.fi/fi/teemat-ja-kehittaminen/velvoitteet-ja-suositukset>, 2025, viitattu: 18.3.2026.