

TeHoava-pilotti PROJEKTIN LOPPURAPORTTI

PÄIVÄMÄÄRÄ:

30. Syyskuuta, 2025

TEKIJÄT :

Mika Saari (PI), GPT-Lab, Tampere University

Vaishnavi Bankhele (PM), GPT-Lab, Tampere University

Riku Laine (Editor), GPT-Lab, Tampere University

Sisällysluettelo

01	Tiivistelmä.....03	07	Kehitys ja iteroinnit.....12
02	Johdanto.....04	08	Arviointi ja mittarit.....14
03	Hankkeen aikajana.....05	09	Haasteet ja rajoitukset.....16
04	Tavoitteet ja rajaukset.....07	10	Kontribuutiot.....18
05	Järjestelmän yleiskuvaus....09	11	Merkitys ja vaikutukset.....20
06	Data ja esikäsittely.....11	12	Johtopäätökset.....21

Tiivistelmä

TeHoava-pilottihanke (*Tekoäly Hiilestä kiinni -ohjelman hyödyntämisessä ja tiedon jalkauttamisessa*) kehitti ja testasi tekoälypohjaista työkalua, jonka tavoitteena oli parantaa maankäyttösektorin ilmastotiedon saavutettavuutta Suomessa. Hanke perustui Hiilestä kiinni -ohjelmaan, joka vuosina 2020–2024 rahoitti yli 200 hanketta maatalouden, metsätalouden ja maankäytön päästöjen vähentämiseksi sekä hiilinielujen vahvistamiseksi. Vaikka ohjelma tuotti runsaasti arvokasta tietoa, viljelijöiden, neuvonantajien ja päättäjien voi olla usein vaikea löytää nopeasti juuri heidän tarpeisiinsa sopivaa aineistoa.

Tutkimustiedon jalkauttamisen haasteeseen vastaamiseksi rakennettiin hakupalvelun prototyyppi, joka perustuu retrieval-augmented generation (RAG) -tekniikkaan. Kun käyttäjä esittää kysymyksen, järjestelmä hakee ensin aineistosta aiheeseen liittyvät kohdat ja muodostaa niiden pohjalta vastauksen kielimallin avulla. Käyttäjä voi esittää kysymyksiä suomeksi, saada vastaukset lähdeviitteineen, antaa palautetta vastausten laadusta sekä ladata keskusteluhistoriansa myöhempää käyttöä varten.

Pilottihankkeessa testattiin useita pohjaismaissa ja Suomessa kehitettyjä avoimesti saatavilla olevia kielimalleja: Viking ja PORO. PORO-2 osoittautui parhaiten soveltuvimmaksi suomenkieliseen käyttöön. Palautetta kerättiin rakenteisen testauksen avulla: käyttäjät arvioivat vastauksia viisiportaisella asteikolla ja kommentein. Tämä palaute ohjasi kehitystä ja varmisti, että järjestelmän kehityksessä pystyttiin huomioimaan kohderyhmän esiin nostamia asioita.

Prototyypin lisäksi hanke tuotti ohjeistuksia ja dokumentaatiota siitä, miten vastaavia järjestelmiä voidaan ylläpitää ja jatkokehittää. Hanke edisti myös tutkimusta kahden vertaisarvioidun julkaisun kautta, joissa käsitellään RAG-järjestelmien suunnittelua, testausta ja eettisiä kysymyksiä.

Tuloksena syntyi teknisesti toimiva demonstraatio siitä, miten tekoäly voi helpottaa maankäyttösektorin ilmastokysymyksiin liittyvän tiedon hyödyntämistä. Tämä tukee tutkimustiedon parempaa käyttöä, päätöksenteon pohjaa sekä avoimuutta julkisin varoin tuotetun tiedon jakamisessa. Tässä raportissa käytetään pilotista nimeä TeHoava, jotta se erottuu Ruokaviraston ja Luonnonvarakeskuksen ylläpitämästä AgriHubi-verkkopalvelusta. Mahdollinen jatkokehitys ja käyttöönotto AgriHubin palveluna on tulevien päätösten varassa.

Johdanto

Hiilestä kiinni -ohjelmassa rahoitettiin vuosina 2020–2024 noin 200 hanketta, joiden tavoitteena oli vähentää maatalouden, metsätalouden ja muun maankäytön päästöjä sekä kasvattaa hiilinieluja ja -varastoja. Hankkeissa kehitettiin myös käytäntöjä ilmastonmuutokseen sopeutumiseen ja riskien hallintaan. Ohjelman viimeiset hankkeet päättyivät vuoden 2024 lopussa. Hiilestä kiinni -ohjelman kehittämiss-hankkeiden vaikutustenarvioinnissa Gaia Consulting ja PTT korostivat, että vaikka kokonaisuudessa on tehty vaikuttavaa ilmastotyötä, konkreettisiin ilmasto-vaikutuksiin päästäisiin nopeammin, jos tulevaisuudessa hankehauissa rahoitusta suunnattaisiin meneillään olevien hankkeiden tulosten käyttöönottoon.

Vuodesta 2024 alkaen vastuu MISU:n alueellisesta toimeenpanosta ja Hiilestä kiinni ohjelman tulosten jalkauttamisesta siirtyi maa- ja metsätalousministeriöltä Pohjois-Pohjanmaan ELY-keskuksen valtakunnalliselle ilmastoyksikölle. Tämä yksikkö toimi myös tämän pilottihankkeen rahoittajana.

Pilottihankkeen tavoitteena oli selvittää, miten tekoälyyn perustuvaa retrieval augmented generation (RAG) -hakutoimintoa voidaan hyödyntää maankäyttösektorin tutkimustiedon käytettävyyden parantamisessa. Pilotti perustui Hiilestä kiinni ohjelman hankkeista laadittujen AgriHubi-tietopankin tietokorttien sisältöihin sekä maa- ja metsätalousministeriön "Hiilestä kiinni -toimenpidekokonaisuuden hankkeet" nettisivulla oleviin loppuraportteihin, joita käytettiin järjestelmän lähtöaineistona.

Hanketta ohjasi ohjausryhmä, johon kuului edustajia maa- ja metsätalousministeriöstä (MMM), Pohjois-Pohjanmaan ELY-keskuksen ilmastoyksiköstä, KEHA keskuksesta, Ruokavirastosta sekä Luonnonvarakeskuksesta (Luke). GPT-Lab tutkimusryhmä Tampereen yliopistosta vastasi toteutuksesta ja teknisestä kehitystyöstä.

Tämä raportti kokoaa yhteen TeHoava-pilotin tulokset. Osa teknisestä aineistosta kuten yksityiskohtaiset kustannusarviot mahdollisesta käyttöönotosta, järjestelmä arkkitehtuurin tarkat kuvaukset ja sisäiset liitteet – on toimitettu ainoastaan ohjausryhmälle luottamuksellisuuden ja tietoturvasyiden vuoksi.

Hankkeen aikajana

TeHoava-pilotti käynnistyi joulukuussa 2024 aloituspalaverilla, jossa määriteltiin RAG-järjestelmän perusteet, raportointikäytännöt ja työtehtävien jako. Varsinaiset kehitysvaiheet etenivät vuoden 2025 aikana vaiheittain:

Taulukko 1. TeHoava-pilotin aikajana ja keskeiset tapahtumat vuonna 2025

Kuukausi 2025	Keskeiset tapahtumat ja tulokset
Tammikuu	Ensimmäinen järjestelmäversio rakennettiin (kokeiluympäristö, kielimalli, AgriHubi-aineisto, testikäyttöliittymä). Versio esiteltiin projektipalaverissa. Sisäinen testaus toteutettiin toteuttajan ja tilaajan teknisten tukihenkilöiden toimesta.
Helmikuu	Toinen versio valmistui. Lisättiin chat-historian tallennus PDF-muodossa, käyttöliittymä lokalisoitiin suomeksi ja viisiportainen palautetyökalu rakennettiin. Testaussuunnitelma laadittiin tilaajalle.
Maaliskuu	Käyttäjäpalautteen keruu integroitiin SQLite-tietokantaan. Testaajan opas valmistui. Ohjausryhmälle esiteltiin kehityksen eteneminen ja jatkosuunnitelmat.
Huhtikuu	Ensimmäinen ulkoinen testauskierros (67 vastausta). Havaittiin puutteita vastausten täydellisyydessä ja terminologiassa. Viking-LLM poistettiin käytöstä ja päätettiin siirtyä PORO 2 -malliperheeseen.
Toukokuu	PORO-2-8B otettiin käyttöön päämallina. Käyttäjäpalautteen analyysi jatkui, dokumentaatiota täydennettiin.

Kuukausi 2025	Keskeiset tapahtumat ja tulokset
Kesäkuu	PORO-2-70B otettiin testikäyttöön. Viking poistettiin kokonaan. Suunniteltiin loppukesän laajempi testikierros.
Elokuu	Iteraatio 8 otettiin käyttöön: järjestelmä alkoi lukea dokumenttien täystekstejä (ei vain tiedostonimiä), vastauspituus kasvatettiin 700 → 2000 tokeniin, lähdeviittaukset eriteltiin selkeämmin ja käyttöliittymä viimeisteltiin suomenkieliseksi. Viimeinen testikierros toteutettiin laajemmilla käyttäjäryhmillä.
Syyskuu	Tekninen dokumentaatio, loppuraportti ja siirtomateriaalit luovutettiin ohjausryhmälle. Mukana karkea kustannusarvio mahdollisesta käyttöönotosta AgriHubin yhteydessä.

Tavoitteet ja rajaukset

Hankkeen tavoitteena oli pilotoida tekoälyyn perustuvaa retrieval-augmented generation (RAG) -hakutoimintoa maankäyttösektorin tutkimustiedon jalkauttamisessa. Lähtöaineistona hyödynnettiin Hiilestä kiinni -ohjelman hankkeista laadittujen AgriHubi-tietopankin tietokorttien sisältöjä sekä maa- ja metsätalousministeriön "Hiilestä kiinni -toimenpidekokonaisuuden hankkeet" sivulla oleviin loppuraportteihin.

Keskeiset tavoitteet

- Tietopohjan saavutettavuus: kehittää toimintatapoja, kuinka Hiilestä kiinni -ohjelman tuloksia voidaan tehdä helpommin käytettäväksi viljelijöille, metsänomistajille, neuvoille, viranhaltijoille ja tutkijoille tekoälyn tukeman hakupalvelun avulla.
- Kielimallien arviointi: selvittää, miten suomenkieliset suuret kielimallit (esim. PORO-mallit) suoriutuvat alakohtaisessa kysymys-vastaus -tehtävässä.
- Palautejärjestelmä: kehittää ja testata palautesilmukka, jossa käyttäjät arvioivat ja kommentoivat järjestelmän vastauksia, tuottaen aineistoa jatkokehittämistä varten.
- Eettiset ja tietoturvanäkökulmat: tunnistaa tekoälyn soveltamisen mahdollisuudet ja rajoitukset maankäyttösektorilla, mukaan lukien tietosuojatietoturva ja vastuullinen käyttö.
- Dokumentointi ja jatkokehityksen tuki: tuottaa teknisiä ohjeita, raportteja ja tutkimusjulkaisuja, jotka tukevat sekä julkista tiedonlevitystä että mahdollista työn jatkamista ohjausryhmän organisaatioissa.

Rajaukset ja riippuvuudet

- Lähdeaineisto: demoversio rajattiin hyödyntämään vain Hiilestä kiinni -ohjelman hankkeista laadittuja AgriHubi-tietokortteja sekä maa- ja metsätalousministeriön "Hiilestä kiinni -toimenpidekokonaisuuden hankkeet" nettisivulla oleviin loppuraportteihin. Laajennus muihin aineistoihin on mahdollista tulevaisuudessa.
- Kielituki: järjestelmän pääkieli oli suomi; ruotsin tuki oli rajallinen, ja englannin tuki toissijainen.
- Resurssit: hankkeen testikäyttäjryhmät olivat kooltaan rajallisia, mikä rajoitti tulosten yleistettävyyttä ja laskentakapasiteetti rajoitti testikäyttäjien määrää. Suurimpien kielimallien käyttöä rajoittivat saatavilla olleet laskentaresurssit, koska suuret kielimallit tarvitsevat paljon muistia, jota laitteistopuolella ei ollut saatavilla.
- Tietoturva ja luottamuksellisuus: osa teknisistä ratkaisuksista, kustannusarvioista ja dokumentaatiosta toimitettiin vain ohjausryhmälle, eikä niitä julkaistu avoimesti tietoturvasyistä.

- **Jatkokehitys:** Hankkeen jälkeen mahdollinen jatkokehitys, ylläpito ja integraatio AgriHubi-palveluun jäivät ohjausryhmän organisaatioiden harkittavaksi.

Työpaketit ja vastuut

Hankkeen toteutuksesta vastaavien organisaatioiden tehtävänä oli varmistaa, että pilotti tukee sekä käytännön tiedontarpeita että kansallisia ilmastopoliittisia tavoitteita.

- **Työpaketti 1 – RAG-hakutoiminnon kehittäminen ja pilotointi**

Toteuttaja rakensi iteratiivisella kehitysprosessilla RAG-tekoälyyn ja avoimiin kielimalleihin perustuvan demoversion, hyödyntäen Hiilestä kiinni -ohjelman hankkeista laadittuja AgriHubi-tietokortteja sekä maa- ja metsätalousministeriön "Hiilestä kiinni -toimenpidekokonaisuuden hankkeet" nettisivulla oleviin loppuraportteihin. Kehitys sisälsi ympäristön pystyttämisen, mallien käyttöönoton, AgriHubi-aineiston kytkennän ja käyttöliittymän rakentamisen. Iteraatioiden (V1–V8) kautta järjestelmä parani teknisen testauksen ja käyttäjäpalautteen avulla.

- **Työpaketti 2 – Pilotointi testikäyttäjillä**

Hakutoimintoa testattiin erilaisten käyttäjäryhmien kanssa (viljelijöitä, metsänomistajia, neuvoja, viranhaltijoita, tutkijoita). Heidän palautettaan kerättiin Webropol-kyselyillä, jossa käyttäjältä pyydettiin numeerinen ja sanallinen arvio hakutoiminnon suorituksesta. Palautteet tallennettiin tietokantaan jatkoanalyysiä varten. Palautteen perusteella kehitystä ohjattiin ja arvioitiin käytettävyyttä, tarkkuutta ja luotettavuutta.

- **Työpaketti 3 – Seuranta ja raportointi**

Projektissa seurattiin selkeästi ja avoimesti tekoälyyn perustuvan hakutoiminnon mahdollisuuksia, haasteita, riskejä ja tietoturvanäkökulmia. Tulokset koottiin julkiseen loppuraporttiin sekä tieteellisiin julkaisuihin, joissa kuvataan havaintoja ja jatkokehitystarpeita

- **Työpaketti 4 – Käyttöohjeet ja tekninen dokumentaatio**

Ohjausryhmälle tuotettiin sisäistä dokumentaatiota, jonka avulla järjestelmää voidaan hyödyntää, jatkokehittää ja mahdollisesti integroida osaksi AgriHubi-tietopankkia. Dokumentaatio sisälsi tekniset ohjeet, riskienhallinnan kuvaukset sekä karkean kustannusarvion mahdollisesta käyttöönotosta ja ylläpidosta.

Järjestelmän yleiskuvaus

TeHoava-prototyyppi perustuu modulaariseen arkkitehtuuriin, jossa yhdistyvät valittu asiakirjakokoelma, hakuprosessi, kielimallit sekä verkkopohjainen käyttöliittymä. Projektin dokumentit esikäsiteltiin, jaettiin pienempiin osiin ja tallennettiin semanttiseen vektoritietokantaan. Kun käyttäjä esittää kysymyksen, järjestelmä hakee tietokannasta aiheeseen liittyvät tekstiosuudet ja välittää ne valitulle kielimallille kontekstiksi. Malli muodostaa vastauksen tämän aineiston perusteella ja palauttaa sen käyttöliittymän kautta käyttäjälle.

Arkkitehtuurin suunnittelu mahdollistaa kielimallien vaihtamisen ja dokumenttiaineiston laajentamisen ilman, että koko järjestelmää tarvitsee rakentaa uudelleen. Pilotin aikana testattiin useita malliperheitä, joista Viking todettiin epäluotettavaksi ja poistettiin käytöstä. PORO-2-mallit valittiin ensisijaisiksi suomenkieliseen käyttöön.

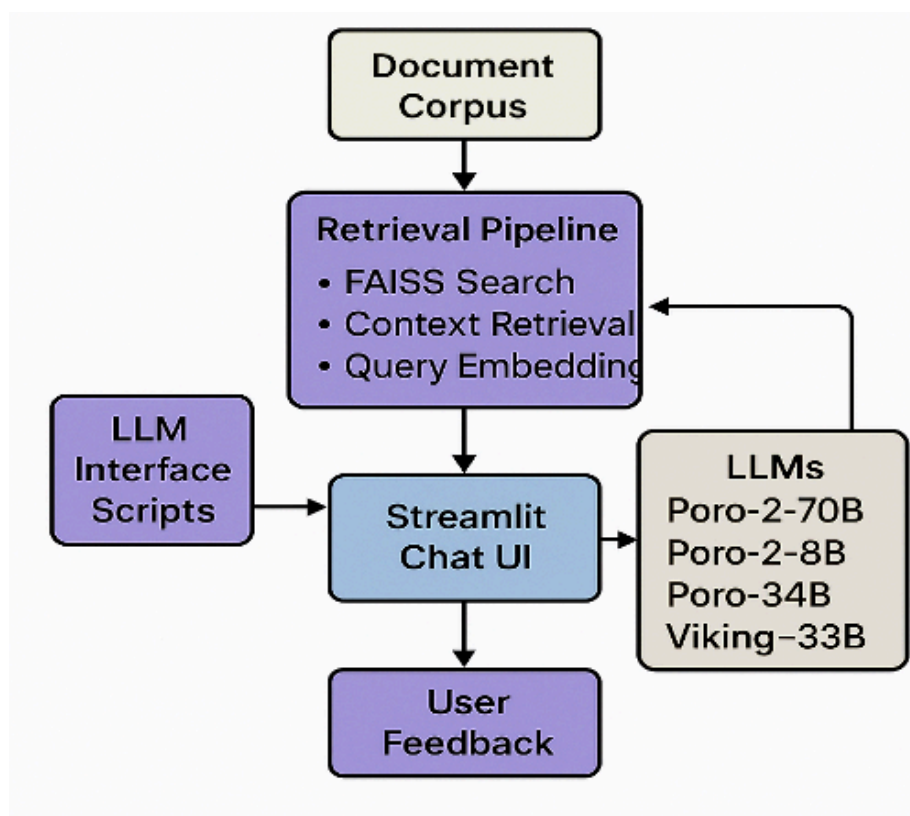


Fig 1. TeHoava-prototyypin arkkitehtuuri

Järjestelmän työnkulku etenee seuraavasti:

1. Kysymyksen analyysi – Käyttäjän kysymys muutetaan samaan vektoriavaruuteen tallennettujen dokumenttien kanssa.
2. Haku – Relevanteimmat tekstiosuudet tunnistetaan ja kootaan kontekstiksi.
3. Vastauksen muodostus – Kielimalli tuottaa kontekstin perusteella vastauksen, joka esitetään käyttäjälle käyttöliittymässä.

Käyttäjät pystyivät lisäksi arvioimaan järjestelmän antamia vastauksia viisiportaisella asteikolla ja lisäämään vapaamuotoisia kommentteja. Kaikki palaute tallennettiin myöhempää arviointia varten. Tämä palautesilmukka tuki järjestelmän jatkuvaa kehittämistä ja arviointia.

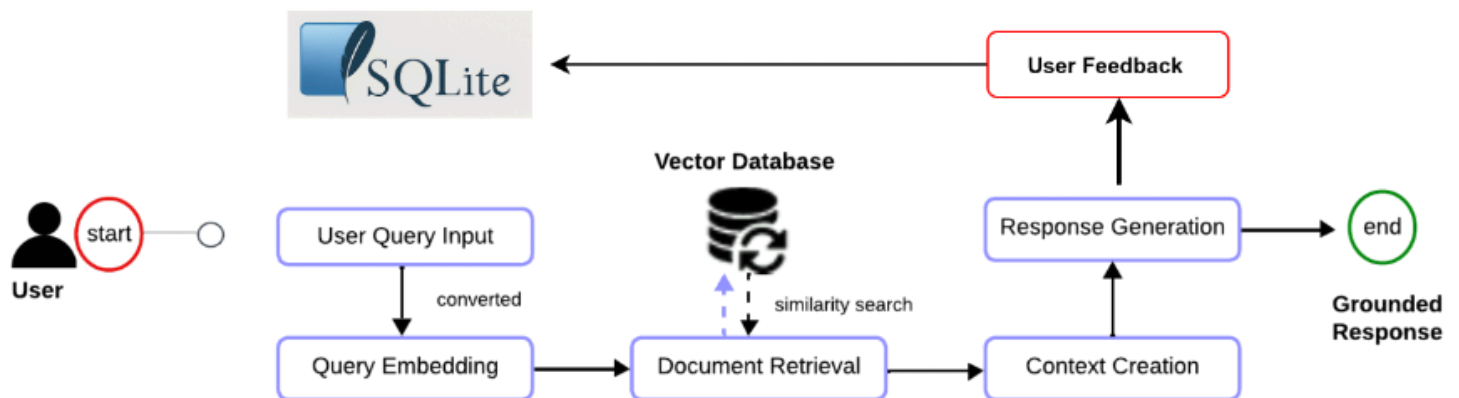


Fig 2. TeHoava-prototyypin työnkulku

Pilotin keskeisiä ominaisuuksia olivat:

- Keskustelumainen web-käyttöliittymä: käyttäjät saattoivat esittää kysymyksiä luonnollisella kielellä ja saada lähteisiin pohjautuvia vastauksia.
- Palautejärjestelmä: jokainen vastaus voitiin arvioida ja kommentoida.
- Keskusteluhistorian lataus: käyttäjä pystyi tallentamaan keskustelunsa PDF-muodossa.
- Monikielisyys: pääasiallinen kieli oli suomi, englanti toimi osittain, mutta ruotsin tuki jäi heikommaksi.
- Läpinäkyvyys: vastaukset sidottiin Hiilestä kiinni -hankkeiden dokumentteihin, mikä mahdollisti tiedon jäljitettävyyden.

Näistä osista muodostui toimiva demonstraatio siitä, miten retrieval-augmented generation -ratkaisua voidaan soveltaa maankäyttösektorin tutkimustuloksien suurien tietomäärien tiedonhallinnassa tutkimustiedon saavutettavuuden parantamiseksi.

Data ja esikäsittely

TeHoava-prototyyppi perustui rajattuun ja kuratoituun aineistoon, joka muodostettiin Hiilestä kiinni -ohjelman hankkeiden tuloksista laadituista AgriHubi-tietokorteista ja niihin liittyvistä raporteista. Tällä rajauksella varmistettiin, että lähdeaineisto perustuisi uusimpaan tutkittuun tietoon ja että sen alkuperä oli jäljitettävissä.

Ennen kuin aineisto voitiin hyödyntää hakujärjestelmässä, se esikäsiteltiin useassa vaiheessa

- Optinen tekstintunnistus (OCR): skannatut tai kuvamuotoiset dokumentit muunnettiin koneellisesti luettavaan tekstiin.
- Tekstin paloittelu (segmentointi): pitkät dokumentit jaettiin päällekkäisiin tekstikappaleisiin, jotta haku voisi kohdentua tarkemmin olennaisiin kohtiin.
- Metatietojen lisääminen: jokainen tekstikappale merkittiin tiedoilla, kuten dokumentin nimi, lähde, päivämäärä ja kieli, mikä helpotti hakutulosten suodatusta ja läpinäkyvyyttä.

Esikäsitelty teksti muunnettiin semanttisiksi vektoriesityksiksi, jotka tallennettiin FAISS-pohjaiseen hakuindeksiin. Vektorit mahdollistivat hakujen tekemisen merkityksen perusteella, ei pelkästään sanojen täsmällisen vastaavuuden.

Esikäsittelyprosessi rakennettiin toistettavaksi: uusia dokumentteja voidaan lisätä soveltamalla samaa vaiheittaista putkea (OCR, segmentointi, metatiedot, numeerinen esitys ja indeksointi). Tämä tekee tietokannasta laajennettavan ja helpottaa sen ylläpitoa tulevaisuudessa.

Kaikki vastaukset pohjautuivat näin ollen eksplisiittisesti prosessoituun aineistoon, ja mukana esitettiin viittaukset alkuperäisiin dokumentteihin. Tämä paransi läpinäkyvyyttä ja lisäsi käyttäjien luottamusta järjestelmän antamiin vastauksiin.

Kehitys ja iteroinnit

TeHoava-prototyyppiä kehitettiin kahdeksassa iteraatiossa tammikuun ja elokuun 2025 välisenä aikana. Jokainen vaihe toi mukanaan uusia ominaisuuksia, mallien vertailua ja käyttäjäpalautteeseen perustuvia parannuksia. Iteratiivinen lähestymistapa mahdollisti sen, että järjestelmä kypsyi vähitellen ja kehitys perustui koko ajan testituloksiin ja kokemuksiin.

Varhaiset versiot (Iteraatiot 1–3):

Ensimmäisissä versioissa luotiin tekninen perusta. FAISS-hakujärjestelmä otettiin käyttöön, ja käyttöliittymä rakennettiin Streamlitin avulla. Käyttäjät pystyivät esittämään kysymyksiä ja saamaan vastauksia Llama 3.2:n ja PORO-34B:n avulla. Helmikuuuhun mennessä lisättiin toiminnallisuus keskusteluhistorian tallentamiseen PDF-muotoon sekä suomenkielisen terminologian parantamiseen.

Palautesilmukan käyttöönotto (Iteraatio 4):

Maaliskuussa järjestelmään lisättiin viisiportainen Likert-pohjainen arviointityökalu, ja palaute tallennettiin SQLite-tietokantaan. Tämä loi perustan järjestelmälliselle arvioinnille. Sisäiset testit toivat esiin haasteita erityisesti vasteajassa ja vastausten täydellisyydessä.

Ensimmäinen ulkoinen testaus (Iteraatio 5):

Huhtikuussa järjestettiin ensimmäinen ulkoinen testikierros, jossa kerättiin 67 vastausta. Tulokset osoittivat, että järjestelmä oli käyttökelpoinen, mutta vastaukset olivat usein lyhyitä ja epätäydellisiä. Viking-mallin todettiin tuottavan epä johdonmukaisia ja heikkolaatuisia vastauksia, minkä vuoksi se poistettiin käytöstä.

Siirtyminen PORO-2 -malleihin (Iteraatio 6):

Toukokuussa otettiin käyttöön PORO-2-8B, joka paransi merkittävästi suomenkielisten vastausten sujuvuutta ja kattavuutta. Samalla päivitettiin käyttäjädokumentaatio ja testaajan opas.

Suurempien mallien valmistelu (Iteraatio 7):

Kesäkuussa käyttöön otettiin sekä PORO-2-8B että PORO-2-70B. Viking poistettiin kokonaan. Iteraatioissa keskityttiin myös järjestelmän vakauttamiseen ja seuraavan testikierroksen suunnitteluun.

Viimeistelyvaihe (Iteraatio 8):

Elokuun iteraatioissa järjestelmä alkoi lukea dokumenttien täystekstejä (ei pelkästään tiedostonimiä), ja vastausten maksimipituutta kasvatettiin 700 → 2000 tokeniin. Lisäksi lähdeviitteiden esitystapaa parannettiin, tietokantaan tallentaminen vakautettiin ja PDF-vientiin lisättiin tuki pitkille vastauksille. Käyttöliittymä viimeisteltiin suomenkieliseksi. Pidemmät vastaukset lisäsivät jonkin verran vasteaikoja, mutta tekivät sisällöstä täydellisempää ja hyödyllisempää.

Arviointi ja mittarit

Arvioinnin ensimmäinen vaihe toteutettiin sisäisesti kehitystiimin toimesta. Käytössä olivat neljä keskeistä mittaria:

- Hakutarkkuus: onnistuiko järjestelmä löytämään tietopohjasta relevantit tekstikohdat.
- Generatiivinen laatu: olivatko vastaukset faktuaalisesti oikein, riittävän täydellisiä ja selkeästi esitettyjä.
- Vasteaika: kuinka nopeasti järjestelmä palautti vastauksen.
- Kielen laatu: suomenkielen sujuvuus ja käytettävyys, sekä rajoitetusti ruotsinkielen toimivuus.

Sisäiset testit osoittivat, että järjestelmä kykeni yhdistämään relevanttia sisältöä ja tuottamaan perusteltuja vastauksia. Samalla havaittiin haasteita, kuten pidemmät vasteajat suurilla malleilla (esim. PORO-2-70B) sekä Viking-mallin epäluotettavuus, minkä vuoksi se poistettiin käytöstä.

Seuraavissa vaiheissa arviointi perustui rakenteelliseen käyttäjätestaukseen. Käyttöliittymään lisättiin viisiportainen Likert-asteikko sekä vapaan tekstin kommentointimahdollisuus. Kaikki palaute tallennettiin kysymysten ja mallin tunnisteiden yhteyteen myöhempää analyysiä varten. Tämä mahdollisti sen, että arviointi jatkui myös kontrolloidun testauksen jälkeen, ja käyttäjät pystyivät tuomaan esiin sekä vahvuuksia että kehityskohteita.

Ensimmäinen testikierros:

- Järjestelmä todettiin käyttökelpoiseksi, mutta vastaukset olivat usein lyhyitä ja epätäydellisiä.
- Viking-malli tuotti epä johdonmukaista suomea ja faktoihin liittyviä virheitä, minkä vuoksi se korvattiin PORO-2-malleilla.

Toinen testikierros:

- Iteraatio 8 toi pidemmät ja täydellisemmät vastaukset. Keskimääräinen vastaus oli noin 230 sanaa, ja ~69 % vastauksista ylitti 200 sanaa.
- Kaikki vastaukset sisälsivät viisi lähdeviitettä, mikä lisäsi läpinäkyvyyttä ja jäljitettävyyttä.

- PORO-2-8B paransi selvästi suomen kielen sujuvuutta ja aihepiirien kattavuutta, ja käyttäjät kokivat vastaukset luonnollisemmiksi ja paremmin jäsennellyiksi.
- Palautteessa nousi esiin myös kehitystarpeita:
 - Terminologian tarkkuusongelmat (esim. metsänhoitolaki sekoitettiin metsälakiin).
 - Satunnaiset virheet ja yleistykset, mikä näkyi siinä, että 31,4 % vastauksista arvioitiin arvosanalla 1–2.
 - Monikierroksiset keskustelut eivät toimineet johdonmukaisesti, sillä jatkokysymykset rikkoivat toisinaan kontekstin.
 - Pidemmät vasteajat, etenkin PORO-2-70B:llä, joka kuitenkin tuotti tarkempia ja yksityiskohtaisempia vastauksia.

Yhteenveto:

Toinen testikierros osoitti edistystä verrattuna ensimmäiseen testikierrokseen. Vastausten keskiarvo nousi tasolle 3,17, ja 37,1 % sai arvosanan 4–5. Käyttäjät kiittivät erityisesti selkeyttä ja laajojen aiheiden yhteenvedon laatua, mutta kritiikki kohdistui juridisen ja hallinnollisen tarkkuuden varmistamiseen, sektorikohtaisen tiedon kattavuuteen sekä yksityiskohtaiseen spesifisyyteen.

Haasteet ja rajoitukset

TeHoava-prototyypin kehityksen ja pilotoinnin aikana tunnistettiin useita tekijöitä, jotka vaikuttivat järjestelmän toimintaan ja arviointiin.

Tekniset haasteet:

Vasteajat vaihtelivat mallien välillä. Suuremmat mallit, kuten PORO-2-70B, vaativat huomattavia laskentaresursseja ja tuottivat pidempiä vasteaikoja. Käytettävissä olleet laitteistoresurssit rajoittivat suurten mallien laajamittaisia kokeiluja. Viking-33B osoittautui epäluotettavaksi: se tuotti lyhyitä ja virheellisiä vastauksia, eikä sen hienosäätö tuottanut riittävää laatua, minkä vuoksi malli poistettiin käytöstä.

Datan määrä:

Järjestelmälle syötettiin suuri määrä dataa (237 tiedostoa), joka sisälsi eri hankkeiden raportteja. Vaikka määrä oli suuri, niin se ei kuitenkaan ollut kattava kuvaus suomalaisesta maa- ja metsätaloudesta, jonka tiedonantokanavaksi järjestelmää rakennettiin. Rajoite ilmenee vastauksissa, joissa kysymys liittyy asiaan jota datasta ei löydy, tai tiedon määrä datassa on rajoittunut. Hyvänä esimerkkinä tästä on kysymys “Räme?” - Järjestelmä ei sisältänyt sanaa tai sen käyttöä, joilloin se alkoi satunnaisesti hallusinoimaan vastausta.

Asiantuntemuksen ja kielituen rajoitukset:

Projektiryhmällä oli osaamista maatalouden ja metsätalouden keskeisistä teemoista, mutta ei kattavaa asiantuntemusta kaikista erikoisalueista. Tämä vaikutti palautteen tarkkuuteen erityiskysymyksissä. Lisäksi suurin osa kehittäjistä ei ollut suomenkielisiä, mikä vaikeutti kielellisten nyanssien ja terminologian arviointia. Ruotsin kielen tuki jäi selvästi heikommaksi, sillä PORO-mallit oli ensisijaisesti optimoitu suomeksi.

Testauksen rajoitukset:

Käyttäjärühmät olivat suhteellisen pieniä. Ensimmäinen testikierros keräsi 67 vastausta, mikä tarjosi arvokasta aineistoa, mutta ei riittänyt vahvoihin tilastollisiin johtopäätöksiin. Toinen testikierros tuotti enemmän aineistoa ja osoitti kehitystä, mutta palautteessa nousi esiin edelleen terminologian epätarkkuuksia ja satunnaisia faktoihin liittyviä virheitä. Tämä näkyi tuloksissa siten, että osa vastauksista arvioitiin edelleen arvosanalla 1–2.

Tietosuoja ja turvallisuus:

Koska järjestelmä hakee sisältöä sisäisistä dokumenteista, vastaukset saattoivat sisältää viittauksia tai katkelmia lähdemateriaaleista. Tämä vaatii huolellista hallintaa mahdollisessa jatkokehityksessä, jotta varmistetaan vain hyväksytyjen aineistojen käyttö. Kehitysvaiheessa ohjelman salaiset avaimet tallennettiin tiedostoihin, joita vain kehittäjällä on oikeus käsitellä ja nähdä. Salaiset avaimet suojaavat ohjelmaa luvattomalta käytöltä. Avaimia vaihdettiin säännöllisesti, jonka seurauksena ohjelman turvallisuus parani entisestään. Hankkeen aikana ei ilmennyt tietoturvaloukkauksia, ja sovittuja tietojen käsittelykäytäntöjä noudatettiin.

Yhteenveto:

Nämä haasteet ja rajoitukset korostavat, että jatkokehityksessä tarvitaan vahvempia laskentaresursseja vasteaikojen lyhentämiseksi, laajempaa asiantuntijapanosta terminologian tarkkuuden varmistamiseksi sekä suojatoimia, jotka tukevat järjestelmän luotettavaa käyttöä julkishallinnon kontekstissa.

Kontribuutiot

TeHoava-hanke tuotti kontribuutioita sekä käytännön järjestelmäkehityksen että tutkimuksellisen tiedon edistämisen näkökulmasta.

Järjestelmäkontribuutiot:

Hankkeessa kehitettiin toimiva prototyyppi, joka osoittaa, miten RAG-tekniikkaan perustuva hakupalvelu voidaan mahdollisesti soveltaa Suomen maa- ja metsätalouden sekä maankäyttösektorin tutkimustiedon jalkauttamisen kontekstissa. Järjestelmä yhdisti PORO-mallit semanttiseen hakutietokantaan ja monikieliseen verkkokäyttöliittymään. Se sisälsi ominaisuuksia, kuten keskusteluhistorian tallennus PDF-muodossa ja viisiportainen palautemekanismi. Lisäksi luotiin rakenteinen testauskehikko, jonka avulla sekä kehittäjät että ulkoiset käyttäjät pystyivät arvioimaan vastauksia systemaattisesti. Tulokset tallennettiin analyysiä ja jatko-kehitystä varten. Hanke tuotti myös käyttöohjeita, teknisiä muistiinpanoja ja konfiguraatio-ohjeita, jotka tekevät ratkaisusta toistettavan ja luovat pohjan mahdolliselle jatkolle.

Tutkimusjulkaisut:

Hankkeen aikana hyväksyttiin kaksi vertaisarvioitua tieteellistä julkaisua:

- Ayman Asad Khan, Md Toufique Hasan, Mika Saari, Kai-Kristian Kemell and Jussi Rasku. Why Adapt RAG for Agile? Challenges, Frameworks, and the Role of Evaluator Agent (XP 2025)
- Md Toufique Hasan, Muhammad Waseem, Kai-Kristian Kemell, Ayman Asad Khan, Mika Saari, Pekka Abrahamsson, Engineering RAG Systems for Real-World Applications: Design, Development, and Evaluation. In: Taibi, D., Smite, D. (eds) Software Engineering and Advanced Applications. SEAA 2025. Lecture Notes in Computer Science, vol 16082. Springer, Cham. https://doi.org/10.1007/978-3-032-04200-2_10

Näissä julkaisuissa käsiteltiin hankkeen toteutuksessa esiintulleita tieteellisesti mielenkiintoisia asioita. Tutkimusjulkaisut annettiin kansainvälisen tutkimusyhteisön arvioitavaksi ja hyödynnettäväksi.

Tutkimukselliset havainnot:

Hankkeen kokemukset ja julkaisut loivat laajempia tutkimuksellisia näkökulmia tekoälyn hyödyntämiseen. Ensimmäisessä artikkelissa esiteltiin Evaluator Agent -konsepti osana moniagenttisia RAG-kehyksiä ja pohdittiin, miten se voi vahvistaa läpinäkyvyyttä, joustavuutta ja eettisyyttä. Toinen artikkeli käytti TeHoavaa tapaus-tutkimuksena RAG-sovellusten suunnittelusta, kehittämisestä ja arvioinnista käytännön konteksteissa. Artikkelissa esitettiin ohjeita alakohtaisten RAG-järjestelmien suunnitteluun ja nostettiin esiin oppeja, jotka ovat sovellettavissa myös muihin toimialoihin.

Hanke osoitti kaksoisarvon: toisaalta konkreettisenä demonstraationa siitä, miten tekoäly voi tukea tutkimustiedon hyödyntämistä Suomessa, ja toisaalta tutkimuksellisenä panoksena, joka edistää vastuullista RAG-järjestelmien suunnittelua ja arviointia sekä kotimaista tekoälyyn pohjautuvaa osaamista ja teknisiä ratkaisuja.

Merkitys ja vaikutukset

TeHoava-hanke osoittaa, että tekoäly voi vahvistaa julkisin varoin tuotetun tutkimustiedon käyttöä maankäyttösektorilla. Kotimaisten kielimallien ja hankkeessa kehitetyn teknisen ratkaisun hyödyntäminen mahdollistaa, että palvelun käyttäjien esittämien kysymysten ja palautteen pohjalta voidaan kehittää toimintaa joustavasti tarpeen mukaan. Samalla voidaan seurata esiin nousevia teemoja ja kerätä tietoa maa- ja metsätalouden toimijoita kiinnostavista aiheista. Tätä tietoa voidaan hyödyntää esimerkiksi hankkeiden aiheiden määrittelyssä ja rahoituksen kohdentamisessa, jolloin tutkimustieto vastaa entistä paremmin käytännön tarpeisiin.

Pilotoimalla järjestelmää, joka hakee ja jäsentää Hiilestä kiinni -ohjelman tietoa, hanke toi esiin, kuinka monimutkaiset tutkimustulokset voidaan tehdä helpommin saavutettaviksi ja käytännössä hyödynnettäviksi. Tämä on merkittävää Suomen ilmastopolitiikan toimeenpanossa, jossa päätökset maataloudessa, metsätaloudessa ja maankäytössä perustuvat yhä enemmän ajantasaiseen ja luotettavaan tietoon. Tämänkaltaiset ratkaisut voivat lisäksi lisätä kustannustehokkuutta ja tutkimustulosten vaikuttavuutta, kun ajantasaisimmat suositukset, ohjeistukset ja selvitykset saadaan asiakaslähtöisesti käyttöön maa- ja metsätalouden sekä maankäyttösektorin toimijoille.

Politiikan toimeenpanossa pilotti tuki ilmastolain sekä siihen liittyvien suunnitelmien (MISU, KAISU, KISS 2030) tavoitteita kokeilemalla välinettä, joka auttaa kääntämään tutkimustiedon käytännön päätöksenteon tueksi. Viljelijät, metsänomistajat ja neuvijat pystyivät hakemaan tietoa tehokkaammin, ja viranhaltijat sekä tutkijat pystyivät hyödyntämään samaa aineistoa yhdenmukaisesti. Näin tämänkaltaiset järjestelmät vahvistaisivat tiedonsiirtoa politiikan, tutkimuksen ja käytännön välillä. Tämän tiedon soveltamisessa tulee kuitenkin varmistaa vastauksien laaduntarkistus ja -valvonta, jotta teknisellä ratkaisulla pystyttäisiin saavuttamaan kustannustehokkuutta ja konkreettisia hyötyjä.

Sidosryhmille hanke tarjosi mahdollisuuden testata suomalaisia kielimalleja konkreettisessa käyttötapauksessa. Käyttäjäpalautteella oli suora vaikutus mallien valintaan, järjestelmän ominaisuuksiin ja arviointimenetelmiin, mikä osoittaa sidosryhmien osallistumisen arvon julkisen sektorin teknologioiden kehittämisessä. Ohjausryhmän organisaatiot – MMM, ELY-keskukset, KEHA, Luke ja Ruokavirasto – saivat käyttöönsä toimivan prototyypin ja siihen liittyvän dokumentaation, jotka tukevat päätöksiä jatkokehityksestä ja järjestelmien yhteensovittamisesta.

Hanke edisti myös vastuullisen tekoälyn kehittämistä. Järjestelmä perustui suomenkielisiin malleihin, sitoi vastaukset tunnistettaviin dokumentteihin ja huomioi tietosuoja- sekä tietoturvanäkökohdat. Julkaisujen kautta vaikutus ulottui myös kansainväliseen tutkimusyhteisöön, jolle tarjottiin oppeja ja uusia lähestymistapoja, kuten Evaluator Agent -malli (arvioiva agentti), joka voi parantaa moniagenttisten järjestelmien läpinäkyvyyttä ja mukautuvuutta.

Yhteenvetona hankkeen merkitys on tutkimuksen ja käytännön yhdistämisessä, suomenkielisten kielimallien potentiaalin osoittamisessa yhteiskunnalliseksi hyödyksi sekä tekoälyn vastuullisen ja läpinäkyvän käytön edistämässä ilmasto- sekä maa- ja metsätalouspolitiikan tukena.

Johtopäätökset

TeHoava-hanke osoitti, että tekoälyä voidaan hyödyntää tutkimustiedon saavutettavuuden parantamisessa maankäyttösektorilla. Kahdeksan kehityssyklin aikana tuotettiin toimiva prototyyppi, joka yhdisti kuratoidun dokumenttiaineiston, suomenkieliset kielimallit, hakuprosessin ja käyttäjäystävällisen käyttöliittymän. Järjestelmä arvioitiin sekä sisäisesti että pilottitestauksessa, ja käyttäjäpalautteen perusteella PORO-2 valittiin päämalliksi samalla kun Viking poistettiin käytöstä. Prototyyppi osoitti, että retrieval-augmented generation -ratkaisut voivat tuottaa käyttökelpoisia, lähteisiin perustuvia vastauksia suomeksi ja tukea sekä käytännön toimijoita että päätöksentekijöitä.

Hankkeessa tuotettiin myös muuta kuin prototyyppi: ohjeistuksia, testerin oppaita, konfiguraatiodokumentaatiota ja riskimuistioita, jotka tekevät järjestelmästä toistettavan ja luovat pohjan mahdolliselle jatkolle. Lisäksi hankkeen aikana hyväksyttiin kaksi vertaisarvioitua julkaisua, joiden kautta opit ja uudet lähestymistavat – kuten Evaluator Agent -konsepti – jaettiin kansainväliselle tutkimusyhteisölle.

Loppuvaiheessa ohjausryhmälle luovutettiin tekninen dokumentaatio ja tämä julkinen raportti. Työpaketti 4 sisälsi myös alustavan migraatiosuunnitelman ja karkean kustannusarvion ratkaisun siirtämisestä yliopisto-ympäristöstä AgriHubin infrastruktuuriin. Tämä materiaali toimitettiin ohjausryhmälle luottamuksellisesti päätöksenteon tueksi. Mahdollinen integraatio AgriHubin palveluksi jää vastuullisten organisaatioiden päätettäväksi.

Hankkeen aikana tunnistettiin myös jatkokehitystarpeita, kuten vasteaikojen parantaminen suurilla malleilla, arviointimenetelmien laajentaminen pelkkiä käyttäjäarvioita pidemmälle, ruotsinkielisen tuen vahvistaminen sekä suojatoimien tarkentaminen sisäisen aineiston hallinnassa. Tuleva tutkimus voisi kehittää edelleen myös automatisoituja arviointimekanismeja, kuten Evaluator Agent -konseptia.

Kokonaisuutena hanke saavutti tavoitteensa: teknisesti kehitetty ratkaisu mahdollista tekoälypohjaisen hakuratkaisun pilotoinnin maa- ja metsätalouden sekä maankäyttösektorin tutkimustiedon jalkauttamiseksi. Hanke tuotti konkreettisia arviointituloksia ja toi lisäymmärrystä tekoälyn mahdollisuuksista ja haasteista maankäyttösektorin tutkimustiedon jalkauttamisessa, niin maankäyttösektorin parissa toimiville julkisille toimijoille kuin kansainvälisille tekoälyn parissa työskentelevälle tutkimusyhteisölle.

Ota yhteyttä



www.gpt-lab.eu



Pohjoisranta 11 A, 28100
Pori



Elinkeino-, liikenne- ja
ympäristökeskus



Maa- ja metsätalousministeriö
Jord- och skogsbruksministeriet



Tampereen yliopisto
Tampere University



RUOKAVIRASTO
Livsmedelsverket • Finnish Food Authority



LUONNONVARAKESKUS

