

Tietoja henkilötietojen käsittelystä tutkimushankkeessa Automatic assessment of spoken interaction in second language (AASIS).

Aasis-tutkimushankkeessa (rahoittajana Suomen Akatemia 2023–2027) käsitellään henkilötietoja. Tämän selosteen tarkoitus on antaa tietoa käsiteltävistä henkilötiedoista sekä siitä, mistä henkilötiedot ovat peräisin ja miten niitä käytetään ja on käytetty tutkimuksessa. Selosteen lopussa kerrotaan tarkemmin, mitä oikeuksia rekisteröidyillä on.

Tutkimukseen osallistuminen ja henkilötietojen antaminen on vapaaehtoista. Sinuun ei kohdistu mitään negatiivista seuraamusta, jos et osallistu tutkimukseen tai jos keskeytät osallistumisesi tutkimukseen. Tutkimukseen osallistuminen ei vaikuta muuhun kielitaidon arviointiin (esim. kurssiarvosanasi, YKI-kokeen tulos).

1. Rekisterinpitäjä

Helsingin yliopisto, Osoite: PL 3 (Fabianinkatu 33), 00014 Helsingin yliopisto, Suomi

Jyväskylän yliopisto, Osoite: Seminaarinkatu 15, PL 35, 40014 Jyväskylän yliopisto, Suomi

Aalto-yliopisto, Osoite: PL 11000, FI-00076 Aalto, Suomi.

2. Yhteyshenkilö ja vastuullinen tutkija

Yhteyshenkilö tutkimushanketta koskevissa asioissa:

- Nimi: Raili Hildén
- Yliopisto/tiedekunta/laitos: Helsingin yliopisto / kasvatustieteellinen tiedekunta
- Osoite: PL 9 (Siltavuorenpenger 3A), 000014 Helsingin yliopisto
- Puh: +358504482514
- Sähköposti: raili.hilden@helsinki.fi

3. Tietosuojavastaavan yhteystiedot

Helsingin yliopiston tietosuojavastaavaan saa yhteyden sähköpostiosoitteesta tietosuoja@helsinki.fi, Jyväskylän yliopiston tietosuojavastaavaan osoitteesta tietosuoja@jyu.fi ja Aalto-yliopiston tietosuojavastaavaan osoitteesta tietosuojavastaava@aalto.fi.

4. Kuvaus tutkimushankkeesta ja henkilötietojen käsittelyn tarkoitus

Aasis-hankkeen tavoitteena on arvioida suomi toisena kielenä -oppijoiden suullista vuorovaikutusta automaattisesti. Aasis-hankkeessa tutkitaan suomenoppijoiden puheen lisäksi nonverbaalisia piirteitä, kuten katsetta, eleitä ja kehon liikkeitä vuorovaikutuksessa ja kielitaidon arvioinnissa. Tavoitteena on laajentaa konsortion aiemmassa DigiTala-hankkeessa (Suomen Akatemia 2019–2023) kehitettyä ASR-pohjaista (automaattinen puheentunnistus) työkalua kattamaan puhutun vuorovaikutuksen ja nonverbaalisten piirteiden automaattisen arvioinnin. Automaattinen arviointi voi tukea opettajien ja kielitaidon arvioijien työtä ja antaa oppijoille mahdollisuuden harjoitella puhumista itsenäisesti.

Tutkimuksen osallistujat ovat aikuisia suomenoppijoita, joiden puhe-suoritukset nauhoitetaan ja videoidaan, sekä ihmisarvioijia (eli kielenopettajia tai asiantuntijoita), jotka arvioivat puhe-suoritukset. Ihmisten antamia arviointeja ja vuorovaikutuksen analyysiä käytetään opettamaan automaattisia arviointimalleja, jotka ennustavat puhe-suoritusten saamia arviointeja koneoppimismenetelmien avulla. Lisäksi kyselyillä ja haastatteluilla kerätään osallistujien taustatietoja ja näkemyksiä esimerkiksi automaattisesta arvioinnista ja puhetehtävien tai arviointiasteikkojen toiminnasta. Lisäksi hankkeessa käytetään erilaisia tutkimusmenetelmiä, kuten algoritmipohjaisia kasvojen

ilmeanalyysijä, ja hyödynnetään uusia menetelmiä, kuten katseenseurantaa. Osallistujille kerrotaan käytetyistä menetelmistä tarkemmin ennen aineistonkeruuta.

Aasis-hankkeessa työskentelee tutkijoita Helsingin yliopistosta, Aalto-yliopistosta ja Jyväskylän yliopistosta. Helsingin yliopisto vastaa pedagogisesta sisällöstä, visuaalisten vihjeiden analysoinnista ja yhteistyöstä oppijoiden kanssa. Aalto-yliopisto vastaa automaattisen puheentunnistuksen ja automaattisen arvioinnin kehittämisestä. Lisäksi Aalto-yliopisto vastaa tutkimusaineiston säilyttämisestä. Jyväskylän yliopisto vastaa puhenäytteiden analysoinnista ja arvioijien kouluttamisesta.

5. Mitä tietoja tutkimusaineisto sisältää

A) Suomi toisena kielenä -oppijoilta kerättävä aineisto:

- Oppijoiden nimet, yhteystiedot, suostumukset
- Oppijoiden äänitteet (video kasvojen ilmeistä ja eleistä, puhe)
- Oppijoiden vuorovaikutuksen annotoinnit
- Oppijoiden koesuoritukset (vastaukset, arvosanat)
- Oppijoiden taustatiedot (esim. montako kurssia opiskeltuna, mahdollisesti kysymyksiä kielitaustasta, kielen oppimisesta, suullisesta kielitaidosta, sukupuolesta, äidinkielestä)
- Oppijoiden itsearviot kielitaidostaan / koesuorituksesta
- Oppijoiden mielipiteet esim. puhetehtävistä tai automaattisesta palautteesta (kysely, haastattelu)
- Oppijoiden suoritusten arviointi (automaattinen arviointi)
- Oppijoiden fysiologiset mittaukset puhe-suorituksen aikana (esim. katseenseurannat mittaukset)
- Oppijoiden käyttäytyminen verkkoympäristöissä (esim. toimintalokit, hiiren klikkaukset)

B) Ihmisarvioijilta (esim. kielenopettajilta tai asiantuntijoilta) kerättävä aineisto:

- Arvioijien nimet ja yhteystiedot, suostumukset
- Arvioijien taustatiedot (esim. äidinkieli, kielitaito, kokemus)
- Arvioijien arvio oppijoiden suorituksesta (yksittäiset osa-alueet, kokonaisarvio)
- Arvioijilta saatu palaute esim. käytetyistä arviointiasteikoista tai kehitettävästä työkalusta (kysely, haastattelu)
- Arvioijien käyttäytyminen verkkoympäristöissä (esim. toimintalokit, hiiren klikkaukset)

Lisäksi keräämme yliopistojen suomen kielen opettajien nimiä ja sähköpostiosoitteita, jotta voimme ottaa yhteyttä suomenoppijoihin. Tutkimukseen osallistuminen ei kuitenkaan vaikuta muihin kieliarviointeihin (esim. suomen kurssin arvosana, YKI-testin tulokset).

Lisäksi arvioijilta saatetaan kerätä muita tietoja palkkioiden maksamista varten. Näitä tietoja ei kuitenkaan käsitellä osana tutkimushanketta.

Lisäksi hankkeessa voidaan käyttää uudelleen aiemmin kerättyjä tietoja, esimerkiksi DigiTala-hankkeessa (Svenska Folkskolans vänner 2015–2017) ja (Suomen Akatemia 2019–2023) kerättyjä puheaineistoja ja ihmisarvioita.

6. Mistä lähteistä henkilötietoja kerätään

Hankkeessa luodaan suomen kielen suullista kielitaitoa mittaavia puhetehtäviä. Yliopistojen kielenopettajien avulla pyydämme kielenopiskelijoita osallistumaan puhetestiin, joka koostuu monologisista ja dialogisista puhetehtävistä. Nauhoitamme oppijoiden puhe-suoritukset käyttäen yliopistojen laitteita (videokameroita, mikrofoneja, nauhureita) ja tiloja (luokkahuoneita, studioita, laboratorioita).

Lisäksi oppijat vastaavat kyselyyn, jonka avulla kartoitetaan heidän taustatietojaan, heidän mielipiteitään suorittamastaan kokeesta sekä itsearviointia omasta kielitaidostaan. Joitakin oppijoita myös haastatellaan, jotta saadaan palautetta ja kehitysehdotuksia hankkeen kehittämään arviointityökaluun liittyen.

Aineistoa kerätään myös ihmisarvioijilta, jotka kuuntelevat ja arvioivat oppijoiden puhe-suorituksia verkko-ympäristössä (esim. Moodlessa). Arvioijat vastaavat taustatietokyselyyn ja osallistuvat hankkeen järjestämään koulutukseen. Osaa heistä myös haastatellaan tai pyydetään vastaamaan kyselyyn palautteen ja kehitysehdotuksien keräämiseksi.

Nelivuotisen hankkeen aikana joitakin osallistujia saatetaan pyytää osallistumaan tutkimukseen, johon sisältyy fysiologisia mittauksia, kuten katseenseurantaa. Katseenseurantatiedot kerätään kaupallisilla katseenseurantaan tarkoitetuilla tutkimuslaitteilla, jotka osallistuja voi halutessaan helposti poistaa tai joiden käytön hän voi lopettaa. Näille osallistujille kerrotaan yksityiskohtaisesti mittauslaitteiden käytöstä ennen aineistonkeruun aloittamista.

Käytännön syistä osa aineistonkeruusta voi tapahtua etänä käyttäen videoneuvottelutyökalua (esim. Zoom) ja yliopistojen ylläpitämiä verkko-ympäristöjä (Moodle). Voimme myös tutkia vuorovaikutusta verkossa käyttämällä simuloituja keskusteluja, joissa oppija vastaa avatarin antamiin kehotuksiin. Verkko-ympäristöissä tallennetaan käyttäjien käyttäytymistä, kuten hiiren klikkauksia ja toimintaloikeja.

Lisäksi hanke käyttää aiemmin kerättyjä aineistoja Kielipankista.

7. Arkaluonteiset henkilötiedot

Tutkimuksessa ei käsitellä tietosuojalain 9 artiklan mukaisia erityisiä henkilötietoryhmiä (eli arkaluonteisia henkilötietoja). Puhetietävät sekä kysely- ja haastattelukysymykset on suunniteltu siten, että niissä ei ilmeneisi arkaluonteisia henkilötietoja (kuten etninen alkuperä, poliittiset mielipiteet tai terveyteen liittyvät kysymykset). Osallistujan kielitaitoa koskevilla kysymyksillä ei myöskään ole tarkoitus kerätä tietoa osallistujan etnisestä taustasta.

Emme myöskään kerää biometrisiä tietoja tunnistamistarkoituksiin. Erityistä huomiota kiinnitetään kuitenkin videoaineiston (osallistujien kasvojen ilmeet ja eleet yhdistettynä heidän puheeseensa) sekä fysiologisten mittausten (esim. katseen tiedot) keräämiseen ja tallentamiseen yksityisyydensuojan varmistamiseksi.

Arkaluonteisten henkilötietojen käsittely perustuu yleisen tietosuojalain 9.2.j (käsittely on tarpeen tieteellisiä tutkimustarkoituksia varten) sekä Suomen tietosuojalain 6 §:n 1 momentin 7 kohtaan.

8. Henkilötietojen käsittelyn oikeusperusta

Henkilötietoja käsitellään seuraavalla yleisen tietosuojalain 6 artiklan 1 kohdan mukaisella perusteella: yleistä etua koskeva tehtävä, tieteellinen tutkimus.

9. Tietojen vastaanottajat

Tietosuojailmoituksen laatimishetkellä tiedossa olevat vastaanottajat:

- Projektitutkijat (kaikki aineisto)
- Aineistonkeruuseen liittyvät palveluntarjoajat: Yliopistojen tarjoama Zoom-videoneuvotteluväline (etäyhteydet, haastattelut, palautekeskustelut, koulutukset). Aalto-yliopiston tarjoama Webropol-työkalu kyselyjen tekemiseen (suostumus- ja taustatietokyselyt). Yliopistojen tarjoama CSC Funet Filesender (suurten tiedostojen lähettäminen).
- Opettajien täydennyskoulutus (näytteet, joita käytetään havainnollistamistarkoituksiin oppijoiden suostumuksella)
- Projektin rekrytoimat ihmisarvioijat (puhe-suoritusten arviointi)
- Projektin rekrytoimat litteroijat/koodaajat (puhe-suoritukset, haastattelut)
- Projektin päätyttyä aineisto arkistoidaan Kielipankkiin niiden tutkittavien osalta, jotka ovat antaneet tähän luvan (nimiä ja yhteystietoja ei tallenneta Kielipankkiin)
- Hankkeen yhteistyökumppanit pääsevät aineistoon käsiksi kumppaniyliopistojen kautta (henkilötietoja siirtämättä)

10. Tietojen siirto Euroopan talousalueen ulkopuolisiin maihin

Tietoja ei siirretä Euroopan talousalueen ulkopuolisiin maihin, vaan niitä käsitellään ainoastaan ETA:n alueella.

11. Automatisoitu päätöksenteko

Tutkimushankkeeseen ei liity automatisoitua päätöksentekoa, jolla on merkittäviä vaikutuksia tutkittaviin. Tutkimukseen sisältyy kuitenkin profilointia, joka liittyy yksilön suullisen kielitaidon automaattiseen tai osittain automaattiseen arviointiin (ks. kohta 4, hankkeessa kehitetään keinoja arvioida toisen kielen oppijoiden suullista vuorovaikutusta automaattisesti).

12. Henkilötietojen suojaus

Tutkimusaineistoon sisältyviä henkilötietoja käsitellään ja säilytetään suojattuna niin, että ainoastaan niitä tarvitsevat henkilöt pääsevät tarkastelemaan tietoja. Osallistujien henkilötiedoista muodostuva rekisteri säilytetään erillään kerätyistä puhenäytteistä ja arvioinneista (eri järjestelmissä).

Oppijoiden puheasuorituksiin viitataan nimien sijasta tutkimustunnisteilla. Osallistujia kehoitetaan välttämään oikeiden nimien, paikkakuntien ja arkaluonteisten tietojen, kuten poliittisten mielipiteiden, antamista (ks. edellä kohta 7).

Tietojärjestelmissä käsiteltävät tiedot suojataan seuraavilla tavoilla:

- Käyttäjätunnus ja salasana
- Rekisteröinti/käyttöloki
- Kulunvalvonta
- Salaus (tarvittaessa)
- Kaksivaiheinen tunnistautuminen

Fyysistä aineistoa (esim. paperilla tai muussa fyysisessä muodossa oleva aineisto) suojataan seuraavilla tavoilla: Lukitussa kaapissa, johon pääsy on ainoastaan Helsingin yliopiston vastuullisilla tutkijoilla (Raili Hilden ja Mikko Kuronen).

Suorien tunnistetietojen käsittely: Suorat tunnistetiedot poistetaan analysointivaiheessa ja säilytetään erillään analysoidusta tutkimusaineistosta.

13. Henkilötietojen käsittelyn kesto tässä tutkimushankkeessa

Henkilötietoja sisältävän tutkimusaineiston säilyttäminen on tärkeää hyvän tieteellisen käytännön takia. Tieteellisessä tutkimuksessa tutkimusaineisto pyritään säilyttämään, jotta tutkimustulokset voidaan jälkikäteen osoittaa oikeiksi ja aiemmin kerättyä aineistoa voidaan hyödyntää myöhemmissä samaa aihetta käsittelevissä tutkimuksissa tai muiden alojen tieteellisissä tutkimuksissa.

Hankkeessa kerättyjä henkilötietoja käsitellään viiden vuoden ajan hankkeen päättymisen jälkeen (2032) tutkimukseen liittyvien julkaisujen valmistumiseksi.

14. Henkilötietojen käsittely tutkimushankkeen päättymisen jälkeen

Tutkimusaineisto poistetaan yliopistojen käyttämiltä tallennusvälineiltä viiden vuoden kuluttua hankkeen päättymisestä (2032).

Tutkimusaineisto arkistoidaan osallistujan luvalla myöhempää, yhteensopivaa tieteellistä tutkimusta varten tietosuojasetuksen vaatimusten mukaisesti: tunnistetiedot mukana (ei nimiä tai yhteystietoja, mutta ääni- ja videotallenteet, myös kasvokuvat).

Tutkimusaineiston tallentaminen perustuu yleisen tietosuojasetuksen 5 artiklan 1 kohdan b ja e alakohtiin.

Rekisteröidyt saavat uuden tietosuojailmoituksen tutkimusaineiston uudesta käytöstä, paitsi jos rekisterinpitäjä ei voi enää tunnistaa rekisteröityjä tietojen perusteella.

Lisäksi rekisteröidyille ei ilmoiteta uudesta tutkimuksesta, jos tietojen toimittaminen rekisteröidyille on mahdotonta tai edellyttää suhteettoman suuria ponnistuksia tai jos se tekee tutkimuksen tavoitteiden saavuttamisen mahdottomaksi tai haittaa vakavasti niiden saavuttamista (tietosuoja-asetuksen artikla 14(5)(b)).

Missä ja kuinka kauan tiedot arkistoidaan: Kielipankki tai muu vastaava, luotettava ja kuratoitu tietoarkisto, pysyvästi.

Aineisto tallennetaan Kielipankkiin korkeimman suojaustason aineistoksi (RESTRICTED). Restricted-aineistoon myönnetään oikeus henkilökohtaiseen tutkimuskäyttöön vain hakemuksen perusteella. Ennen uutta tutkimuskäyttöä Kielipankki varmistaa, että uusi tutkimustarkoitus on yhteensopiva aineiston alkuperäisen käytön kanssa asetuksen vaatimusten mukaisesti. Kielipankki antaa aineistolle pysyvät tunnisteet.

15. Mitä oikeuksia rekisteröidyllä on ja oikeuksista poikkeaminen

Yhteyshenkilö tutkimushenkilöiden oikeuksiin liittyvissä asioissa on tämän ilmoituksen kohdassa 1 mainittu henkilö.

Rekisteröidyn oikeudet

Tietosuoja-asetuksen mukaan rekisteröidyllä on oikeus:

- saada pääsy tietoihin
- oikaista tietoja
- poistaa tiedot ja tulla unohdetuksi
- rajoittaa tietojen käsittelyä
- siirtää tiedot järjestelmästä toiseen
- vastustaa tietojen käsittelyä
- olla joutumatta automaattisen päätöksenteon kohteeksi

Rekisteröity ei kuitenkaan voi käyttää kaikkia oikeuksia kaikissa tilanteissa. Tilanteeseen vaikuttaa esimerkiksi se, millä perusteella henkilötietoja käsitellään.

Tarkempaa tietoa rekisteröidyn oikeuksista eri tilanteissa löytyy tietosuojavaltuutetun verkkosivuilta:

<https://tietosuoja.fi/en/what-rights-do-data-subjects-have-in-different-situations>.

Oikeuksista poikkeaminen

Tietosuoja-asetus ja Suomen tietosuojalaki mahdollistavat tietyistä rekisteröidyn oikeuksista poikkeamisen silloin, kun henkilötietoja käsitellään tieteellisessä tutkimuksessa ja oikeuksien toteuttaminen estäisi tai vaikeuttaisi suuresti käsittelyn tarkoitusten saavuttamista.

Tarvetta poiketa rekisteröidyn oikeuksista arvioidaan aina tapauskohtaisesti.

Valitusoikeus

Sinulla on oikeus tehdä valitus tietosuojavaltuutetun toimistoon, mikäli katsot, että henkilötietojesi käsittelyssä on rikottu voimassa olevaa tietosuojalainsäädäntöä. Yhteystiedot:

Tietosuojavaltuutetun toimisto

Käyntiosoite: Ratapihantie 9, 6. kerros, 00520 Helsinki

Postiosoite: PL 800, 00521 Helsinki

Puhelin (puhelinvaihe): 029 56 66700

Faksi: 029 56 66735

Sähköposti: tietosuoja(at)om.fi