

Politika publikoak ebaluatzeko prozesuetan AA apliakatzearen azterketa, analisia eta prospekzioa. Erabilera-kasuak.

Apirila 2025



besaldi

Enplegu eta inklusio
politiken ebaluazio organoa
Órgano de evaluación de las
políticas de empleo e inclusión



EUSKO JAURLARITZA
GOBIERNO VASCO

EKONOMIA, LAN ETA
ENPLEGU SAILA

DEPARTAMENTO DE ECONOMÍA,
TRABAJO Y EMPLEO



besaldi

Enplegu eta inklusio
politiken ebaluazio organoa
Órgano de evaluación de las
políticas de empleo e inclusión



EUSKO JAURLARITZA
GOBIERNO VASCO

EKONOMIA, LAN ETA
ENPLEGU SAILA
DEPARTAMENTO DE ECONOMÍA,
TRABAJO Y EMPLEO

**Politika publikoak ebaluatzeko prozesuetan AA aplikatzearen azterketa, analisisia
eta prospekzioa**

Erabilera-kasuak

Elaborado con la asistencia técnica de



Edukien taula

0.	SARRERA.....	3
1.	AMAS ENPLEGURAKO PIZGARRIETARAKO PROFILATZE-SISTEMA (AUSTRIA)	4
2.	GIZARTE, GIZARTERATZE ETA ENPLEGU ZERBITZUETARAKO AA (KATALUNIA)	6
3.	GAIXOTASUNAGATIKO BAJAK AURREIKUSTEKO SISTEMA (NORVEGIA).....	9
4.	DIGITALIZAZIOA ETA AAREN ERABILERA ZERBITZU PUBLIKOETAN (FINLANDIA).....	10
5.	HIZKUNTZA HANDIKO EREDUAK (LLM), DATU SENTIKORRAK ATERATZEKO (EE. BB.).....	11
6.	ARRISKUEN EBALUAZIOA, INFORMAZIO-LAGUNTZA ETA HERRITARRENTZAKO ARRETA (NAZIO ANITZEKO).....	12
7.	POBREZIA ENERGETIKOAREN ARRISKUA AURREIKUSTEKO EREDUA (HERBEHEREAK).....	14
8.	LANGABEZIAREN IRAUPENARI BURUZKO IRAGARPEN-EREDUAK (ESLOVAKIA).....	15
9.	POBREZIA AZTERTZEKO SATELITE BIDEZKO IRUDIAK (NAZIO ANITZEKO).....	16
10.	DOITASUN-ANALITIKAGIZARTE-PROGRAMENE BALUAZIOAN (EE. BB.).....	17
11.	TESTU-MEATZARITZA ETA GAIKAKO EBALUAZIOARAKO MACHINE LEARNING (NAZIO ANITZEKO).....	18
12.	DATU-INGENIARITZA ETA MACHINE LEARNING POBREZIA NEURTZEKO (NAZIO ANITZEKO)	19
13.	KALTEBERATASUN ENERGETIKOARI BURUZKO TXOSTENEN AUTOMATIZAZIOA (KATALUNIA).....	20
14.	ERABAKIAUTOMATIZATUAK (ADM) SEKTORE PUBLIKOAN (KANADA).....	21
15.	ESPERIMENTU NATURALAK DATU HISTORIKOEKIN (NAZIO ANITZEKO).....	22
16.	MACHINE LEARNING, BGDSren INPAKTUA EBALUATZEKO (ESPAINIA).....	23
17.	HIZKUNTZA HANDIAREN EREDUA (LLM) EBALUAZIOAN (AZTERKETA ANALITIKOA).....	24
18.	MACHINE LEARNING, ALDAKETAREN TEORIAN OINARRITUTAKO EBALUAZIOA EGITEKO (NAZIO ANITZEKO).....	25
19.	PERTSONA EBALUATZAILEAK: NLP ETA IKASKUNTZA AUTOMATIKOKO EREDUAK (KANBODIA)	26
20.	BIG DATA, JARDUNAREN JOERAK DETEKTATZEKO (ISRAEL).....	27
21.	CASE BASED MODELLING (CBM) ETA AGERTOKIEN SIMULAZIOA (ERRESUMA BATUA).....	28
22.	EBALUATZAILEAK PRESTATZEKO TXATBOTAK (EE. UU. ETA KANADA).....	29

o. Sarrera

Agiri honetan azaltzen diren erabilera-kasuek osatu egiten dute **Adimen Artifiziala (AA) politika publikoaren zikloan aplikatzeko jardunbide egokiei, berrikuntzei eta eraldaketei buruzko azterlanaren** azken txostena.

Agiri honek **kasu desberdinak azaltzen dituzten 22 fitxa** jasotzen ditu, adibide zehatzak emanaz AAk enpleguko eta gizarte-inklusioko politika publikoen zikloko testuinguru eta fase desberdinetan izan duen aplikazioari buruz. Fitxetan politikaren ezarpen-faseko 9 kasu eta ebaluazio-faseko 13 kasu deskribatzen dira, hurrenez hurren.

Fitxa bakoitzean banakatutako eremuak hauek dira:

1. **Erabilera-kasua:** AAre aplikazio zehatza deskribatzen du, ezarritako teknologiaren edo sistemaren ikuspegi orokorra emanaz. Eremu hau funtsezkoa da AAk politika publikoaren testuinguruan duen asmoa eta funtzionaltasuna ulertzeko.
2. **Testuingurua:** AA zein ingurutan ezartzen den azaltzen du, antolaketa arduradunaren eta denbora-esparruaren gaineko zehaztasunak barne. Eremu hori lagungarria da AAre aplikazioa benetako testuinguru praktikoan kokatzeko.
3. **Deskribapen orokorra:** AAre sistemaren funtzionamenduari buruzko azalpen zehatza ematen du, erabilitako metodoak eta prozesuak barne. Eremu hori funtsezkoa da teknologia nola ezartzen den eta zein arazo zehatzi aurre egiten dion ulertzeko.
4. **Aplikaturako teknologia:** AA aplikatzean erabilitako teknologia eta algoritmoak zehazten ditu. Eremu hau erabilgarria da ezarpenaren oinarri teknikoak eta erabilitako tresnak ulertu nahi dituzten profesionalentzat.
5. **Dokumentaturako emaitzak:** AAre ezarpenaren bidez erdietsitako emaitzak eta lorpenak aurkezten ditu, metrika arrakastatsuak eta antzemandako onurak barne. Eremu hau funtsezkoa da teknologiak politika publikoaren testuinguruan duen eragina eta eraginkortasuna ebaluatzeko.
6. **Arriskuak eta mugak:** AAre ezarpenarekin lotutako erroka, arrisku eta muga posibleak identifikatzen ditu. Eremu hau garrantzitsua da teknologiaren ebaluazio kritiko eta orekatua egiteko, zaindu eta hobetu behar diren arloak nabarmenduz.
7. **Erreferentziak:** erabilera-kasua azaltzen duen argitalpenaren iturri bibliografikoa jasotzen du.

22 kasu praktikoz osatutako bilduma honek aukera ematen dio irakurleari ulertzeko nola eraldatu ditzakeen Adimen Artifizialak enpleguko eta gizarte-inklusioko politikak. Horretarako, onurak eta arriskuak erakusten ditu eta, horrez gain, ikuspegi praktikoa eskaintzen du tresna horiek administrazio publikoan txertatzeko asmoa dutenentzat.

1. Enplegurako pizgarrien profilak aurreikusteko sistema - AMAS (Austria)

Erabilera-kasua	Sistema algoritmiko honek gizarte-langileei laguntzen die erabakitzen enplegu-eskatzailleetatik zeintzuek jaso beharko lituzketen laguntza-baliabideak. Sistemak pertsona bakoitzarentzat “integrazio-aukera” puntuatzen du, hiru taldetan sailkatuz, lan-merkatuan sartzeko dituzten aukeren arabera. Horrek zehazten du zein babes-maila dagokien: enplegua eskuratzeko aukerak zenbat eta altuagoak, baliabide gutxiago.
Testuingurua	Austriako Enplegu Zerbitzu Publikoak (AMS) 2018an ezarritako sistema, enplegu-programen eraginkortasuna eta efizientzia administratiboa hobetzea xede.
Deskribapen orokorra	Deskubritu zen ereduak sistematikoki zigortzen zituela emakumeak eta Europatik kanpoko herritartasuna zutenak; datu administratiboak eta ezaugarri pertsonalak uztartzean (hartara, enplegua eskuratzeko aukerak aurreikusteko), sistemak egiturazko desberdintasunak indartu zituen, ustezko objektibotasun teknikoaren izenean. 2021ean sistema bertan behera geratu zen, epaileak hala erabakita. Azterketak alborapenak zehazten ditu, bai eta horiek arintzeko aukerak ere.
Aplikaturako teknologia	Erregresio logistikoa oinarritutako algoritmoa.
Dokumentaturako emaitzak	▪ Teknika aringarriak. Azterlanak alborapenak arintzeko hainbat teknika azaltzen ditu (birhaztapena, bidezko ordezkariak, aukera ekualizatuen prozesamenduostea), eta agerian jartzen du metodo horiek taldeen arteko desberdintasunak murrizten lagun dezaketela, errendimendu prediktiboko maila onargarriak mantentzen badituzte. ▪ Ekitate-metrikak. Azterketak berdintasun-metrikak azaltzen ditu, alborapen algoritmikoa identifikatu eta arintzeko neurri gisa; hala nola, Statistical Parity Difference (SPD) eta Disparate Impact (DI), Equal Opportunity Difference (EOD) eta Average Odds Difference (AOD).
Arriskuak eta mugak	▪ Azterketak AMAS algoritmoaren bertsio berreraiki bat aztertzen du (ez originala), eta generoa ezaugarri bitar gisa ulertzen du, babestutako beste ezaugarri batzuk edo ikuspegi intersektionalak kontuan hartu gabe. ▪ Metodologiak behaketa bidezko metrikak baliatzen ditu, kausa-harremanak ahaztuz eta taldearen ekitatea lehenetsiz norbanakoaren gainetik. ▪ Aurreko politiketako datuak erabiltzen ditu, alborapenak izan ditzaketanak, eta teknika aringarriak ez dira oso gardenak, ez baitute aukerarik ematen prozesuan zehar zein elementu zehatz aldatzen diren identifikatzeko.
Erreferentziak	Achterhold, E., Mühlböck, M., Steiber, N., & Kern, C. (2025). Fairness in Algorithmic Profiling: The AMAS Case. <i>Minds and Machines</i> , 35(1), 9. https://doi.org/10.1007/s11023-024-09706-9

2. AA gizarte-zerbitzuetarako, inklusio-zerbitzuetarako eta enplegurako (Katalunia)

Bartzelonako Udala, Gizarte Eskubideen Arloa eta Gizarte Zerbitzuen Udal Institutua. DPR sistema (eskarien, arazoen eta erantzun sozialen identifikazio automatikoa).

Erabilera-kasua	Esku-hartze sozialetako datuen erregistroaren automatizazioa hizkuntza naturalaren prozesamenduaren bidez.
Testuingurua	Erabakiak babesteko sistema, adimen artifizial aurreratuan oinarritua, urtean 50.000 bisita inguru artatzen dituzten gizarte-zerbitzuen 40 zentrotan ezarria.
Deskribapen orokorra	Hizkuntza naturalaren prozesamendua erabiltzen du informazio-sisteman datuen erregistroa arintzeko. Erabiltzaileekin elkarriketak grabatu ostean, sistemak automatikoki identifikatzen ditu eskariak (E), arazoak (A) eta errekurtsoak (E), hartara profesionalen denbora optimizatzeko.
Aplikatutako teknologia	Hizkuntza naturalaren prozesamendua, egituratu gabeko datuetara aplikatua (elkarriketa sozialen transkripzioak).
Dokumentatutako emaitzak	▪ Ezarpen arrakastatsua 2021etik Bartzelonako zentro guztietarako, mendetasuna artatzeko zerbitzurako eta arriskuan dauden haur eta nerabeen azterketako eskarien talderako. 2022an Udalaren azpiegitura korporatiboan txertatu zen. ▪ Kalitatea neurtzeko eta berrentrenatzeko prozesuak (automatikoak nahiz eskuzkoak) ezartzea gomendatzen da, algoritmoaren arrakasta ikuskatu ahal izateko.
Arriskuak eta mugak	▪ Alborapenak erreproduzitzeko aukera. ▪ Kanpoko zerbitzari batean ostatatutako eredu algoritmikoa.

Kataluniako Generalitatea, Gizarte Eskubideen Departamendua. Gizarte-laguntzen arauen motorra

Erabilera-kasua	Gizarte-laguntzetarako irizpideen betearazpenari buruzko egiaztapen automatizatua.
Testuingurua	Kataluniako administrazio publikoa, ezarri den eta martxan dagoen sistema.
Ikuspegia	Hautagarritasuna ebaluatzeko arauetan oinarritutako sistema.
Deskribapen orokorra	Egiaztatzen du ea herritar edo erakunde batek betetzen dituen gizarte-laguntza zehatzak jasotzeko irizpideak. Bost algoritmo jasotzen ditu, honakoak neurtzeko garatuak: oinarritzko beharrianak, etxeko gastuak, erretiroaren pentsio ez-kontributiboa, erretiroaren pentsio ez-kontributiboaren osagarria, eta jaiotzaren, harreraren edo adopzioaren ziozko prestazioa.
Aplikatutako teknologia	Arauetan oinarritutako algoritmoak.

Dokumentatutako emaitzak	▪ Zerbitzu publikoak (herritarra erdigunean jarrita: herritarren irisgarritasun handiagoa zerbitzu publikoetara) ▪ Gobernuaren eragiketaren erantzun-gaitasuna ▪ Baliabide publikoen kudeaketaren hobekuntza ▪ Prozesu eta sistemen kalitatearen igoera
Arriskuak eta mugak	Ez dira jaso.

Kataluniako Generalitatea, Gizarte Eskubideen Departamendua. Esku-hartze sozialeko beharrianen identifikazioa

Erabilera-kasua	Esku-hartze sozialean dauden beharrianen identifikazio automatikoa.
Testuingurua	eSocial tresnan txertatutako sistema, pilotu-fasean.
Ikuspegia	Gizarte-langileentzako laguntza benetako denboran.
Deskribapen orokorra	Esku-hartze sozialean, enplegatu publikoak funtzionaltasun bat aktibatu dezake (herritarren baimenarekin), elkarrizketa transkribatu eta beharrian-egoerak automatikoki identifikatzeko.
Aplikatutako teknologia	Hizkuntza naturalaren prozesamendua eta, ziur aski, analisi prediktiboa.
Dokumentatutako emaitzak	Ez dira jaso (pilotu-fasean dago).
Arriskuak eta mugak	Ez dira jaso.

Kataluniako Administrazio Irekia (AOC). Elkarrizketarako txatbotak (Xatbots).

Erabilera-kasua	Herritarrei arreta emateko eta enplegatu publikoei euskarria emateko laguntzaile birtualak.
Testuingurua	Kataluniako Administrazio Irekia (AOC). Sistema ezarri da, eta martxan dago 7 egunetan 24 orduz arreta emateko.
Ikuspegia	Laguntzaren automatizazioa eta kontsulten ebazpena.
Deskribapen orokorra	Elkarrizketarako laguntzaileak, enplegatu publikoei eta herritarrei etengabeko arreta emateko garatuak. Izapide digitalekin eta gardentasun-atariaren erabilerarekin lotutako galderak eta gorabeherak erantzuten dituzte.
Aplikatutako teknologia	Elkarrizketako AA algoritmoak, maiz egindako galderetatik ateratako datuekin entrenatuak.
Dokumentatutako emaitzak	▪ Baliabideen optimizazioa: txatbotek lan-karga murrizten dute, maiz egiten diren kontsultak kudeatzean; halaber, aurrez aurreko eta telefono bidezko arretako gastua murrizten dute. ▪ Eraginkortasun handiagoa: berehala eta zehatz erantzuten diete erabiltzaile ugari. ▪ Datuen azterketa: zerbitzuak hobetzeko patroiak eta beharrak antzematea ahalbidetzen dute.

	▪ 24/7 sarbidea: informazioa eta laguntza eskuragarri dago une oro. ▪ Berehalako erantzuna: ez da itxaron behar eta, beraz, izapideak eta kontsultak arintzen dira. ▪ Autokonponketa: erabiltzaileek zalantzak argitzen dituzte pertsonen esku-hartzerik gabe. ▪ Esperientzia hobea: nabigazio gidatuak eta erantzun zehatzek administrazioarekiko harremana errazten dute.
Arriskuak eta mugak	▪ Mendetasun teknologikoa: akatsek edo etenek arretan eragin dezakete, eta mesfidantza sortu. ▪ Pribatutasuna eta segurtasuna: datuak urratzeko arriskua dago, batez ere elkarrizketak erregistratzeagatik. ▪ Laguntzaren mugak: hizkuntza ulertzeko eta erantzun egokia emateko zailtasunek zerbitzuaren kalitatean eragin dezakete. ▪ Erantzunen kalitatea: entrenamendu egokirik gabe, txatbotak erantzun desagokiak eman ditzake, sistemaren eraginkortasunean eta konfiantzan kalteak sortuz.

Argitalpenak AAKo 12 sistema jasotzen ditu, Kataluniako lau administrazio publikoek erabiltzen dituztenak. Horietatik sei tokian aritzen dira eta sei erkidegoan; zazpi daude ezarrita eta bost pilotu-fasean. Gehienek (12tik 7) lan administratiboak automatizatzea dute xede. Identifikatutako sistemen artean sartzen dira eskari sozialak automatikoki identifikatzeko zereginetatik elkarrizketarako laguntzaileen eta prestazioen hautagarritasuna egiaztatzeakoetara bitartekoak.

AA mota desberdinak eta datuen azterketa aurreratuen sistema teknologikoak erabiltzen dira, hizkuntza naturalaren prozesamendua, arauetan oinarritutako sistemak, elkarrizketatarako laguntzaileak eta gomendio-algoritmoak barne. Nabarmentzekoa da 12 sistematik 11k dutela giza gainbegiradaren bat.

Zehaztutako kasu guztietarako erreferentzia hauxe da: m4social. (2024). Radar de algoritmos de IA y procesos de decisión automatizada para el acceso a los derechos sociales en Cataluña. m4Social. <https://m4social.org/es/resources/radar-dalgoritmes-dia-i-processos-de-decisio-automatitzada-per-a-lacces-als-drets-socials-a-catalunya/>.

3. Gaixotasunaren ziozko bajak aurreikusteko sistema (Norvegia)

Erabilera-kasua	Prototipo interaktiboa, zerbitzu publikoen atari bat simulatzen duena, aurrez zehaztutako interakzio-sekuentziekin, gaixotasunen ziozko bajen iraupena kalkulatzeko.
Testuingurua	Gizarte-laguntza desberdinak kudeatzen dituen Norvegiako erakunde publiko batek sistema prediktibo bat ezarri du, Human-in-the-loop (HITL) ikuspegia tarteko. Horren arabera, AA sistema informazio-iturri osagarri bihurtzen da gizarte-langileentzat (ez erabaki-sistema automatizatua).
Deskribapen orokorra	Azterlanak esperientzia pilotua deskribatzen du, eta zerbitzu publikoetan AA ezartzeak duen eragina aztertu nahi du, horren erabilaren erronka etikoak eta arautzaileak nahiz onurak aztertuz, eta giza inplikazioak (human-in-the-loop) eta gardentasunak gizartean onartua izaten nola lagun dezaketen ikertuz. Pilotuan 18 eta 65 urte bitarteko 20 herritarrek parte hartu dute, gaixotasunaren ziozko bajan dagoen biztanleriaren banaketa geografikoa islatu asmoz.
Aplikaturako teknologia	Prototipoak atari interaktibo batean txertatutako sistema prediktibo bat du oinarrian.
Dokumentatutako emaitzak	▪ Jarrera nagusiki positiboa zerbitzu publikoetan AA ezartzearen inguruan. Faktore lagungarriak: lasaitasuna gizakien gainbegiradaren zioz eta gardentasuna prozesuetan (mikro maila); balio-proposamen orekatua banakako onuren – beharrezkoak ez diren hitzorduak saihestuz– eta onura kolektiboen artean –baliabide publikoaren kudeaketa hobetuta– (meso maila); konfiantza altua gobernuko organoetan; datuen analisisien eta AAren aukerak aztertzen diharduen taldea; informazio argia datuen erabilerei eta ereduaren logikari buruz. ▪ Sistemak diseinatzeko inplikazioak: banakakoentzako eta gizartearentzako balio-proposamena argi eta garbi zehazteko beharra. ▪ Giza gainbegirada: jasotako legitimitatea zaintzeko “gizakiak zirkuituan” mantentzearen garrantzia.
Arriskuak eta mugak	▪ Konfiantza galtzeko arriskua, baldin eta erantzunak argiak edo egokiak ez badira. ▪ Pribatutasunaren eta datu pertsonalen erabilera okerraren gaineko kezka. ▪ Sistemaren legezko azalpenak edo azalpen teknikoak ulertzeko zailtasuna. ▪ Teknologiarekiko gehiegizko mendetasun posiblea, arreta pertsonalizatua murriztuz.
Erreferentzia	Schmager, S., Grøder, C. H., Parmiggiani, E., Pappas, I., & Vassilakopoulou, P. (2024). Exploring citizens' stances on AI in public services: A social contract perspective. <i>Data & Policy</i> , 6, e19. https://doi.org/10.1017/dap.2024.13

4. AAren erabilera eta digitalizazioa zerbitzu publikoetan (Finlandia)

Erabilera-kasua	Teknologia aurreratuak txertatzea zerbitzu publikoen prozesu administratiboetan: gizarte-segurantzaren eta osasuna.
Testuingurua	Finlandiak tradizio luzea du teknologiek administrazio publikoan erabiltzeko orduan. Herriarren konfiantza-maila altuak dituzte, eta gizarte-ongizate unibertsaleko eredu ere badute. Horrek teknologia aurreratuak prozesu administratiboetan eraginkortasunez txertatzen laguntzen du.
Deskribapen orokorra	Azterketaren arabera, zerbitzu publikoetan teknologia digital aurreratuak eraginkortasunez txertatzeko orduan eredu da Finlandia. Gardentasuna, pribatutasuna eta konfiantza publikoa bermatzeko arau-esparru argi bat izatearen garrantzi kritikoa azpimarratzen da.
Aplikaturako teknologia	▪ Adimen artifiziala (AA) predikzio eta pertsonalizazioarako. ▪ Machine learning arriskua ebaluatzeko (adibidez, desgaitasunaren ondoriozko pentsioa). ▪ Prozesuen automatizazioa Robotic Process Automation bidez (RPA). ▪ Txatbot adimentsuak bezeroei arreta emateko. ▪ Arkitektura nazional digitala (AuroraAI) zerbitzu interkonektatuetarako. ▪ Blockchain datuen egiaztapen deszentralizaturako. ▪ Plataforma digital txertatuak (Suomi.fi, Kanta.fi, Tulorekisteri.fi).
Dokumentaturako emaitzak	▪ Igoera nabarmena zerbitzu publikoaren eraginkortasunean eta pertsonalizazioan, herriarren onarpen positiboa nabarmenduz (2020an eskarien % 72 online). ▪ AAren ezarpen arrakastatsua aldi baterako ezintasuneko arriskuak aurreikusteko (% 78ko zehaztasuna). ▪ Erabateko automatizazioarako etengabeko garapena, ohiko kasuak RPA bidez aurreikusi eta eraginkortasunez erabiltzea barne.
Arriskuak eta mugak	▪ Erabaki erabat automatizatuarekin lotutako legezko erronkak, legedi espezifikoaren behar dutenak. ▪ Datu pertsonalen gehiegizko erabilerarekin eta gardentasunarekin lotutako arazoak, etikoak eta pribatutasunari lotutakoak. ▪ Agiri fisikoaren erabileraren iraunkortasuna eta erabateko digitalizazioarako erresistentzia partziala herriar batzuen aldetik.
Erreferentzia	Väänänen, N. (2021). The digital transition of social security in Finland. Frontrunner experiencing headwinds? <i>Ubezpieczenia Społeczne. Teoria i Praktyka</i> , 4, 71–86. https://doi.org/10.5604/01.3001.0015.5251

5. Hizkuntza Eredu Handiak (LLM) datu sentikorrak ateratzeko (AEB).

Erabilera-kasua	Tokiko hizkuntza-ereduak (LocalLLM) aplikatzea sustantzien gehiegikeriarekin lotutako arazoak sailkatu eta informazioa ateratzeko (SRP), haurren zabarkeriaren eta gehiegikeriaren kasuetako ikerketa-laburpen ez-egituratueta oinarrituta. Aplikazioaren helburua testu konfidentzialen bolumen handien azterketa hobetzea da, pribatutasuna kolokan jarri gabe.
Testuingurua	AEBetan egin da azterketa, haurren ongizatearen estatu-agentzia baten txostenak erabilita.
Deskribapen orokorra	Azterketak lau LocalLLM aztertzen ditu, eta ereduak testuak sailkatu eta ateratzeko orduan duten errendimendua neurtzen du, aldeztatik eskuz kodetutako 2.956 laburpenetan giza adituek egindako kodifikazioarekin alderatuta.
Aplikaturako teknologia	LocalLLM: hizkuntza-eredua handiak, tokiko ordenagailuetan egikaritzeko modukoak, aurretiazko entrenamendu zehatzik gabe zereginak ahalbidetzen dituztenak.
Dokumentatutako emaitzak	- LocalLLMek giza adituekin alderatzeko moduko emaitzak lortu zituzten: zehaztasuna %95ekoa izan zen eta Cohen kappa 0.90ekoa. - Ereduetan leialtasun handia erakutsi zen (%99-100) benetako kasu positiboetarako, eta haluzinazio-tasa orokorrak baxuak izan ziren (%0.6-2.9). Ereduak txikiak ere (Llama3-8b) errendimendu bikaina izan zuten, kontsumitzailearen mailako hardwarean egikaritzea ahalbidetuz. - Ereduak aise bete zituzten jarraibideak (%96-100), zitak eta sustantziak formatu egituratueta sortzeko. - Ateratako informazioan esangura handia izan zuen, kasuak era egokian interpretatzeko erraztasunak emanez hala.
Arriskuak eta mugak	- Prompt-ak garatu eta aldatzeko prozesu iteratibo ez-sistematikoa. - Hardware espezializatuko baldintzak (GPU). - LocalLLMek aurrez ikasitako patroietan oinarritutako testua sortzen dute; testuen sorkuntzari buruzko txarto ulertuak versus informazioa ateratzeko hutsa. - Arrakala gizarte-laneko prestakuntza tradizionalaren eta teknologia hauek ezartzeko behar diren gaitasun teknikoaren artean. - Positibo faltsuak nahitaez sustantzia-arazoak adierazten ez dituzten medikamentuekin lotuta.
Erreferentzia	Perron, B. E., et al. (2024). Moving Beyond ChatGPT: Local Large Language Models (LLMs) and the Secure Analysis of Confidential Unstructured Text Data in Social Work Research. <i>Research on Social Work Practice</i> , 10497315241280686. https://doi.org/10.1177/10497315241280686

6. Arriskuen ebaluazioa, babes informatikoa eta herritarren arreta (nazio anitzekoa)

Erabilera-kasua	Tresna digitalen eta AAren integrazioa zerbitzu publiko hauetan erabakiak hartzeko: enplegua, gizarte-zerbitzuak, herritarren arreta, migrazio-kontrola, zerga-erregulazioa, hezkuntza.
Testuingurua	Herrialde hauetako kasuak aztertu dira: Danimarka, Alemania, Grezia, Herbehereak, Norvegia eta Suedia.
Deskribapen orokorra	Berrikuspen sistematiko bat da, zeinak aztertzen baitu AA tresnak datuen analisi aurreratuen testuinguru orokor batean sartzeak nola aldatzen duen langile publikoak erabakiak hartzeko prozedura, kontuan hartuta zerbitzuaren kalitatean eta hautagarritasunean egon daitezkeen mugak nahiz balizko hobekuntzak, bai eta herritarren eta administrazioaren arteko elkarrekintzan egon daitezkeen aldaketak ere.
Aplikaturako teknologia	▪ Erabakiak hartzeko sistema automatizatuak. ▪ Arriskuak ebaluatzeko tresnak. ▪ Kasuak kudeatzeko sistema digitalak. ▪ Datu-base egituratuak. ▪ Herritarren autozerbitzuko tresnak.
Dokumentaturako emaitzak	▪ Automatizazioak zeregin konplexuetarako denbora librean uzten du. ▪ Aurrez aurreko elkarrekintza zuzena nabarmen murrizten da. ▪ Eraginkortasun administratiboa areagotzen da, eta giza akatsak murrizten dira. ▪ Malgutasun profesionala galtzeko eta gehiegizko estandarizaziorako arriskua dago.
Arriskuak eta mugak	▪ Erresistentzia eta estrategia informalak funtzionarioen aldetik zorrotasun teknologikoari aurre egiteko. ▪ Arrakala digitala areagotzen da, eta zerbitzu digitalak eskuratzeko desberdintasunak hazi. ▪ Giza gainbegirada egokirik gabeko erabaki automatizatuak hartzearekin lotutako arrisku etikoak.
Erreferentzia	Marienfeldt, J. (2024). Does digital government hollow out the essence of street-level bureaucracy? A systematic literature review of how digital tools' foster curtailment, enablement and continuation of street-level decision-making. <i>Social Policy & Administration</i> , 58(5), 831–855. https://doi.org/10.1111/spol.12991

Argitalpenak testuinguru geografiko desberdinetako kasuak aztertzen ditu:

- ▶ Danimarka: arriskuak ebaluatzeko tresnak, datu-basea eta kasuen kudeaketa enplegu-zerbitzuen eremuan; artxibo digitala gizarte-zerbitzuen eremuan.
- ▶ Alemania: ezagutza eta kasuak kudeatzeko sistema enplegu eta herritarren arretako zerbitzuen eremuan.
- ▶ Grezia: datu-basea, formularioekin, herritarren arreta-zentroen eremuan.
- ▶ Herbehereak: sistema adimentsua (kamerak) eta arriskuen ebaluazioa kontrol migratorioko eremuan.
- ▶ Norvegia: autozerbitzu digitala eta kasuen kudeaketa enplegu eta zerga-erregulazioko zerbitzuen eremuan.
- ▶ Suedia: erabaki automatizatuak, ebaluazio estandarizatua eta kasuen kudeaketa gizarte-laguntzaren, beken eta enpleguaren eremuan.

Fitxan jasotako informazioa kasuen azterketa bateratuari buruzkoa da.

7. Pobrezia energetikoko arriskua aurreikusteko eredua (Herbehereak)

Erabilera-kasua	Azterlanak ikertzen du ikaskuntza automatikoko teknologia desberdinak nola aplikatzen diren pobrezia energetikoaren arriskua aurreikusteko orduan: konplexutasun altuko eta kausa anitzeko fenomenoak.
Testuingurua	Herbehereak ez dute mekanismo ofizialik pobrezia energetikoa ikuskatu, zenbatetsi eta gainditzeko, eta, beraz, datu sozioekonomikoetan oinarritutako aurreikuspen-eredua garatu beharra antzeman da.
Deskribapen orokorra	Azterketak agerian jartzen du Machine Learning ereduak nabarmen lagundu dezaketela pobrezia energetikoa ulertu eta kudeatzen, baina azpimarratzen du politika argiak behar direla eta ikuspegi zehatzagoa ezarri aldagai gakoak hautatzeko orduan.
Aplikaturako teknologia	- Machine Learning XGBoost bidez. - Datu sozioekonomikoetan oinarritutako sailkapen algoritmikoa (diru-sarrera, etxebizitzaren balioa, etxearen tamaina eta abar). - Datu-base sozioekonomikoen erabilera (KWB eta WoON).
Dokumentatutako emaitzak	- Ereduaren aurreikuspen-gaitasun altua (zehaztasuna % 80 ingurukoa). - Faktore gakoak identifikazio argia, besteak beste diru-sarrerak, etxebizitzaren balioa eta antzintasuna, etxearen tamaina eta biztanleria-dentsitatea. - Diru-sarrerak faktore erabakigarriena direla berretsi da, nahiz eta badauden aurreikuspen-gaitasun nabarmena duten beste faktore batzuk ere.
Arriskuak eta mugak	- Eredua sentikorra da aldagai espezifikoak hautatu eta baztertzeko. - Datu adierazgarri asko behar dira ereduak egoki entrenatzeko, bereziki ordezkartza gutxi duten taldeak identifikatzeko. - Pobrezia energetikoaren eta faktore sozioekonomikoen arteko harremana berez da konplexua, eta sakonago aztertu beharko litzateke.
Erreferentzia	Dalla Longa, F., Sweerts, B., & van der Zwaan, B. (2021). Exploring the complex origins of energy poverty in The Netherlands with machine learning. <i>Energy Policy</i> , 156, 112373. https://doi.org/10.1016/j.enpol.2021.112373

8. Langabeziaren iraupena aurreikusteko ereduak (Eslovakia)

Erabilera-kasua	Langabeziaren iraupenaren aurreikuspena, enplegu-eskatzailleak iraupen luzeko langabezia-arriskuaren arabera sailkatzeko.
Testuingurua	Eslovakiako Errepublikako Lan, Gizarte Gai eta Familia Bulego Nagusia (COLSAF), langabeentzako erregistroa eta laguntza kudeatzeaz arduratzen dena.
Deskribapen orokorra	Azterlanak Eslovakiako langabeziaren iraupena aurreikusten duen ikaskuntza automatikoko eredu bat garatzen du, enplegu-eskatzailleen datu errealak erabiliz. Mihiztatzeko ereduen ikuspegia aplikatzen da (CART, CHAID eta analisi diskriminatzailea), erregresio logistikoaren bidez konbinatuta. Helburua da politika publikoak babestea eta baliabideak optimizatzea iraupen luzeko langabezia murrizteko.
Aplikaturako teknologia	- Ikaskuntza automatikoko eredu konbinatuak (ensemble): erabaki-zuhaitzak (CART, CHAID), analisi diskriminatzailea. "Stacking" metodoaren erabilera (hainbat ereduren emaitzak konbinatuz zehaztasuna hobetzeko).
Dokumentaturako emaitzak	- Langabeziaren iragarpen eraginkorra (% 78 zehaztasun orokorra, % 92 kasu luzeen zehaztasuna). - Identifikaturako funtsezko faktoreak: langabeziaren lan-historia, hezkuntza eta adina. - Laneko esku-hartzeak iragarpen zehatzen bidez hobetzeko potentziala.
Arriskuak eta mugak	- Aurreikuspen-eredu batzuk ez dira oso gardenak, edo interpretatzeko zailak dira ("kutxa beltza"). - Datuak etengabe eguneratu behar dira zehaztasuna mantentzeko.
Erreferentzia	Gabrikova, B., Svabova, L., & Kramarova, K. (2023). Machine Learning Ensemble Modelling for Predicting Unemployment Duration. <i>Applied Sciences</i> , 13(18), 10146. https://doi.org/10.3390/app131810146

9. Pobrezia aztertzekeo satellite-irudiak (nazio anitzekoa)

Erabilera-kasua	Pobreziaren edo ongizate-mailen aurreikuspena satellite-irudiak aztertuz, AAre algoritmoak aplikatuta, azpiegitura, etxebizitzaren kalitatea, gaueko argiztapena edo mugikorren erabilera bezalako adierazleak identifikatzeko.
Testuingurua	Afrika, Asia eta Latinoamerikako herrialdeetan aplikazioak dituzten nazioarteko azterlanak. Herrialde bakar batean zentratzen ez bada ere, Senegal, Bangladesh, Ruanda, Ghana, Uganda eta abarreko kasuak sartzen dira. Ereduak garapen-, plangintza- eta estatistika-agentzietan aplikatzeko modukoak dira.
Deskribapen orokorra	Ikerketak satellite-irudiak eta adimen artifiziala konbinatzen dituzten 60 lan aztertzen ditu, pobrezia aurreikusteko. ML eta DL ereduak konbinazioak, adierazle fisikoen erabilerak eta datu-iturri anitzek zehaztasuna hobetzen dutela erakusten du. Garapen-politika publikoetan aplikatzeko jarraibideak eskaintzen ditu.
Aplikaturako teknologia	- Ikaskuntza automatikoa (Machine Learning) eta sakona (Deep Learning), neurona-sareak. - Ezaugarriak teknika aurreratuen bidez ateratzeko eta hautatzeko prozedura (CNN, Random Forests, PCA). - Satellite-irudiak pobreziari buruzko informazioa lortzeko.
Dokumentaturako emaitzak	- Pobrezia-aurreikuspen hobeak ageriko adierazle fisikoak erabiltzean (azpiegitura, etxebizitzaren kalitatea, etab.) gizarte-adierazle aldakorren aldean (diru-sarrerak, kontsumoa). - Efektu positibo mugatua irudien bereizmen espaziala handitzean. - Pobreziaren ebaluazioen azkartasuna eta zehaztasuna nabarmen hobetzeko potentziala.
Arriskuak eta mugak	- Metodo aurreratu batzuen interpretagarritasun txikia (azaltzeko moduko AAre beharra). - Zehaztasun handiagoa teknika tradizionalak eta aurreratuak konbinatzean, metodo mistoen beharra adieraziz.
Erreferentzia	Hall, O., Dompae, F., Wahab, I., & Dzanku, F. M. (2023). A review of machine learning and satellite imagery for poverty prediction: Implications for development research and applications. <i>Journal of International Development</i> , 35(7), 1753-1768. https://doi.org/10.1002/jid.3751

10. Zehaztasun-azterketak gizarte-programen ebaluazioan (AEB)

Erabilera-kasua	Kausazko Egitura Modelatua (SCM) eta Zehaztasun Analitika (PA) aplikatzea, gizarte-programen ebaluazioa hobetzeko. Bi kasu-azterlanen bidez: 1) Gemma Services, egoitza-zentro bateko adingabeentzako osasun mentaleko tratamenduen pertsonalizazioa; (2) Program to Aid Citizen Enterprise (PACE): ekitatearen azterketa arrazializatutako komunitateak artatzen dituzten GKEen baliabideen banaketan.
Testuingurua	Azterketa AEBetan egin zen, Pittsburgh eskualdean, haurren osasun mentaleko zerbitzuetan zentratua. Irabazi-asmorik gabeko erakundeek inplementatu bazuten ere, HHS, HUD eta National Science Foundation bezalako agentzia federalen laguntza eta finantzaketa izan zuen.
Deskribapen orokorra	Azterlanak SCM eta Precision Analytics duten eredu kausalak aplikatzen ditu programa sozialak ebaluatzeko. Gemma Servicesen adingabeentzako osasun mentaleko tratamenduak pertsonalizatzen dira; PACEn GKEentzako funtsen banaketa ekitatiboa aztertzen da. Bi kasuek erakusten dute nola erabili datu masiboak erabaki bidezkoagoak eta eraginkorragoak hartzeko.
Aplikatutako teknologia	▪ Structural Causal Modeling (SCM) eta Precision Analytics (PA). ▪ Machine learning gainbegiraturua, analisi ia esperimentalak, taldeen matchinga. ▪ Dashboards interaktiboak, analisi geoespaziala. ▪ Plataforma: Equitable Impact Platform (EquiP), BCT Partners.
Dokumentatutako emaitzak	▪ 30 egun murriztu da haur bakoitzeko batez besteko egonaldia Gemma Servicesen (kasu bakoitzeko 10.000 dolar aurreztuta). ▪ Altaren unean larritasun-maila % 40 murriztu da, ospitaleratze-tasa txikiagoarekin lotuta. ▪ Berariazko azpitaldeetarako tratamendu-konbinazio eraginkorragoak identifikatu dira. ▪ PACEn, arrazializatutako komunitateei arreta emateko GKEen finantzaketan dagoen desberdintasunaren ebidentzia jaso da.
Arriskuak eta mugak	▪ Eskuragarri dauden datuen kalitatearekiko eta zehaztasunarekiko mendekotasuna. ▪ Balidazio kualitatiboa behar duten positibo/negatibo faltsuen presentzia. ▪ Alborapen algoritmikoen arriskua, aldagai sozialak behar bezala kontrolatzen ez badira. ▪ Aurreikuspen-tresnen erabileran zuzenean inplikaturik ez dauden onuradunak. ▪ Gaitasun tekniko egokirik gabe emaitzak interpretatzeko eta aplikatzeko zailtasunak.
Erreferentzia	York, P. (2024). The Future of Evaluation Analytics: Case Studies of Structural Causal Modeling in Action. En S. B. Nielsen, F. M. Rinaldi, & G. J. Petersson (Eds.), Artificial Intelligence and Evaluation (219-241 or.). Routledge. https://doi.org/10.4324/9781003512493-11

11. Testu-meatzaritza eta Machine Learning ebaluazio tematikorako (nazio anitzekoa)

Erabilera-kasua	Testu-meatzaritza eta Machine Learning gainbegiratua, Munduko Bankuaren Doing Business (DB) mailegu-programaren 2010-20 urteei buruzko adierazleen gaikako ebaluazioa egiteko.
Testuingurua	Ebaluazioaren helburua DB adierazleen garrantzia eta eraginkortasuna aztertzea zen. Erronka zen ebaluazio-aldian onartutako 5,710 proiekturen artean DB adierazleek eragindako proiektu guztiak identifikatzea.
Deskribapen orokorra	Independent Evaluation Group-ek (IEG) metodologia hibridoa garatu zuen, NLP gainbegiratua, gako-hitzen bidezko bilaketak eta eskuzko berrikuspen selektiboa konbinatzen dituen. Ikuspegi iteratibo horri esker, ebaluatzaileek beren iritzi kritikoa aplikatu zuten proiektuen azpimultzo maneigarri batean, algoritmoek dokumentazio-bolumen handiak eraginkortasunez prozesatzen zituzten bitartean.
Aplikaturako teknologia	Hizkuntza-aurreprozesamenduaren konbinazioa (stopword-ak ezabatzea, lematizazioa), TF-IDF bidezko bektorizazioa, sailkapen gainbegiratuaren ereduaren ensembleda, eta terminoen maiztasunean oinarritutako puntuazioa duten bilaketa-sistema automatizatuak, dokumentu-bolumen handien artean proiektu garrantzitsuak eraginkortasunez identifikatzeko.
Dokumentatutako emaitzak	- Ikuspegiak identifikatutako ebidentzia-oinarriaren tamaina bikoiztea ahalbidetu zuen (269 proiektu guztira), aurkikuntzen baliozkotasuna eta zabaltasuna hobetuz. - Giza berrikuspena eta prozesamendu automatizatuaren konbinatzean ebaluazio-denbora nabarmen murriztu zen. - Zeregin honetarako garatutako sailkapen-ereduak IEGren beste ebaluazio batzuetan berrerabili ziren, hasierako inbertsioaren itzulera areagotuz.
Arriskuak eta mugak	- Baliabide-mugek entrenamendu-datuen tamaina eta kalitatea murriztu ditzakete, gainbegiratuaren NLP ereduaren zehaztasunari eraginez. - Proiektuak "kreditua lortzea" bezalako erregulazio-eremuekin erlazionatzeko modu ugari daude, eta horrek zaildu egiten du bilaketa-taxonien bidezko harrapaketa osoa. - Funtsezkoa da datu-zientzialarien eta ebaluatzaileen artean komunikazio estua izatea, aplikazioari buruzko ulermen partekatua garatzeko.
Erreferentzia	Ziulu, V., Anuj, H., Hagh, A., Raimondo, E., & Vaessen, J. (2024). Extracting Meaning from Textual Data for Evaluation: Lessons from Recent Practice at the Independent Evaluation Group of the World Bank. En S. Bohni Nielsen, F. Mazzeo Rinaldi, & G. J. Petersson (Eds.), Artificial Intelligence and Evaluation: Emerging Technologies and Their Implications for Evaluation (1st ed., 78-103 ot.). Routledge. https://doi.org/10.4324/9781003512493

12. Datu-ingeniarietza eta Machine Learning pobreziaren gaineko eragina neurtzeko (nazio anitzekoa)

Erabilera-kasua	Datu-ingeniarietza eta ikaskuntza automatikoko ikuspegia erabiltzen da informazio digitala prozesatzeko eta adierazle sozioekonomiko geoerreferentziatuak sortzeko, pobreziaren aurkako esku-hartzeak ebaluatze aldera.
Testuingurua	Erakunde akademikoan, ikerketa teknologikoko zentroen eta garapenerako lankidetzako erakunde aldeaniztunen arteko lankidetzak.
Deskribapen orokorra	Telefonia mugikorrek datuak (datuen kontsumoa, gizarte-sareetako jarduerak, finantza-transakzioak) eta satelite-irudiak (lur-erabilerak, eraikuntza-materialak, automobilak, gaueko argiztapena, hiri-azpiegitura) konbinatzen dituen analisi konputazionalen sistema. Informazio hori AA algoritmoen eta ikaskuntza automatikoaren bidez prozesatzen da, adierazle sozioekonomiko geoerreferentziatuak sortzeko.
Aplikaturako teknologia	Machine Learning datu heterogeneoetan patroien errekonozimenduari, satelite-irudiak prozesatzeko ikusmen artifizialari eta adierazle digitalak metrika sozioekonomiko tradizionalekin lotzeko korrelazio-algoritmoak aplikatua. Datu-bolumen handiak (Big Data) prozesatzeko teknikak barne hartzen ditu.
Dokumentatutako emaitzak	- Datuak eguneratzeko maiztasun handiagoa bost urteko edo hamar urteko zentsuekin alderatuta, eta datuak biltzeko kostuen murrizketa esanguratsua. Ia etengabe monitorizatzeko gaitasuna. - Pobrezia-adierazleen granularitate espazial handiagoa. - Neurketarako sarbidea logistika- edo aurrekontu-mugak dituzten eremuetan.
Arriskuak eta mugak	- Lagin-alborapenak, teknologia digitalerako sarbiderik ez duten biztanleak baztertzeagatik. - Baliozkotze mugatua neurketa sozioekonomiko tradizionalei dagokienez. - Eraginkortasun diferentziala testuinguruaren arabera: satelite-irudiak eraginkorragoak dira landa-eremuetan, eta telefono-datuek hobeto funtzionatzen dute hiri-inguruneetan. Distortsioak dentsitate handiko hiri-finkapenen harrapaketan. - Etikoak: datu pertsonalen erabileran pribatutasun-arazo potentzialak; populazio kalteberen azpi-ordezkapenagatik arrakala digitalak areagotzeko arriskua.
Erreferentzia	Arguelles Toache, E. (2023). Ventajas y desventajas del uso de la Inteligencia Artificial en el ciclo de las políticas públicas: análisis de casos internacionales. Acta universitaria, 33, e3891. https://doi.org/10.15174/au.2023.3891

13. Kalteberatasun energetikoari buruzko txostenen automatizazioa (Katalunia)

Erabilera-kasua	24/2015 Legeak eskatutako kalteberatasun energetikoari buruzko txosten sozialen automatizazioa, AA bidez, bizitegi-bazterkeria jasateko arriskuan dauden pertsonen oinarritzko hornidurak etetea saihesteko.
Testuingurua	Kataluniako sei udalerritan sistema pilotu bat ezartzeko kasuaren azterketa: El Prat de Llobregat, Amposta, Granollers, Sant Boi de Llobregat, Palau Solità i Plegamans eta Santa Coloma de Gramenet.
Deskribapen orokorra	Kalteberatasun energetikoari buruzko txosten sozialak sortzeko automatizazio-sistema. Txosten horiek gizarte-zerbitzuen oinarritzko arloek (ABSS) eskatzen dituzte hilerok, ez ordaintzeagatik gerta daitezkeen hornidura-mozketen aurrean kalteberatasun ekonomikoko egoerak egiaztatzeko. Sistemak hainbat datu-iturri integratzen ditu hodeiko plataforma baten bidez (Microsoft Azure-EBn dago). Plataforma horrek aukera ematen du: energia-hornitzaileen fitxategiak kargatzeko, titularren datu sozioekonomikoak lortzeko, koefizienteak kalkulatzeko, kalteberatasunari buruzko txosten automatikoak sortzeko eta e-Notum bidez enpresa hornitzaileei jakinarazteko.
Aplikaturako teknologia	Datuak prozesatzeko algoritmoak prozesu-automatizazio robotikoko teknologekin (RPA) eta negozio-prozesuen kudeaketarekin (BPM) konbinatzea. Consultors BPM enpresa pribatuak garatutako sistema, Via Oberta elkarreragingarritasun-zerbitzuak eta Hestia gizarte-zerbitzuen informazio-sistema bilduz.
Dokumentatutako emaitzak	▪ Estaldura: % 90eko arrakasta txostenen automatizazioan (ia % 100era iristeko itxaropenarekin, lege- eta antolaketa-doikuntzen ondoren). ▪ Denbora eta administrazio-karga nabarmen txikiagoak, bai profesionalentzat, bai herritarrentzat. ▪ Enpresen, udalerrien eta administrazioen arteko elkarreragingarritasuna. ▪ Kalteberatasun-egoeren datuen kudeaketa eta ikuspegi zehatzagoa. ▪ Teknologiarako sarbidea baliabide mugatuak dituzten udalerrientzat, hodeiko zerbitzuaren bidez, instalaziorik gabe.
Arriskuak eta mugak	▪ Giza gainbegiradaren beharra emaitzak baliozkotzeko. ▪ Datuen babesaren gaineko inplikazioak informazio pertsonal sentikorrarekin lan egitean. ▪ Sistema eta konpainia hornitzaile anitzekin integratzeko beharra (50 baino gehiago lurraldean).
Erreferentzia	m4Social. (2024). Radar de algoritmos de IA y procesos de decisión automatizada para el acceso a los derechos sociales en Cataluña. m4social. https://m4social.org/es/resources/radar-dalgoritmes-dia-i-processos-de-decisio-automatitzada-per-a-lacces-als-drets-socials-a-catalunya/

14. Erabaki automatizatuak (ADM) sektore publikoan (Kanada)

Erabilera-kasua	Adimen artifizialak sektore publikoko erabakiak hartzeko prozesuetan eta horren ondoriozko antolamendu-eraldaketetan duen eraginaren azterketa, arrazionaltasun mugatuko ikuspegitik.
Testuingurua	Kanadako administrazio publikoa aro digitalean, eta erabakiak hartzeko sistema automatizatuak.
Deskribapen orokorra	AAk sektore publikoan sustatzen dituen erabakiak hartzeko sistema automatizatuak (ADM) aztertzea, giza elementuen eta elementu teknologikoen arteko harremana ikertuz gobernuaren antolamendu-egituren barruan. Azterlanak arrazionaltasun mugatuko eredu modernizatu bat garatzen du, AAn oinarritutako ADM sistemak aztertzeko.
Aplikaturako teknologia	Ikaskuntza automatikoko sistemak (Machine Learning), bereziki ikaskuntza sakona (Deep Learning), neurona-sareak, sailkapen-zuhaitzak eta euskarri-bektoreen makinak, sektore publikoko erabaki-sistema automatizatuei aplikatuta.
Dokumentaturako emaitzak	▪ Erakundeen eraginkortasuna nabarmen handitzea, erabakietako elementu faktikoen ebaluazioa hobetuta. ▪ Antolamendua erabat aldatzea aurreikusten da, AAk "erabaki onaren" nozio tradizionalak eta antolaketa-egiturak aldatzen dituen heinean.
Arriskuak eta mugak	▪ Erronka etiko garrantzitsuak identifikatzen dira, azalgarritasun-arazoak, alborapen algoritmikoa eta kontu-ematearen murrizketa barne AA sistema aurreratuetan. ▪ Gaur egungo AA sistemak mugatuak dira giza balio konplexuak erabat ulertzeko, eta horrek tentsioak sortzen ditu eraginkortasun teknikoaren eta iritzi etikoaren artean. ▪ Ezagutza teknikoko arrakala bat dago sektore publikoan, AA sistemak eraginkortasunez ezartzeko eta gainbegiratzeko.
Erreferentzia	Alexander, W. (2022). Applying Artificial Intelligence to Public Sector Decision Making. Major Research Paper. University of Ottawa.

15. Esperimentu naturalak datu historikoekin (nazio anitzekoa)

Erabilera-kasua	Saiakuntza analitikoa osagai teoriko eta aplikatuekin, kontzeptuzko berrikuspenean eta kasuen adibideetan oinarritua.
Testuingurua	Afrika eta garapenaren testuinguru globala; 2019-2020.
Deskribapen orokorra	Artikuluak ebaluazioaren arloan funtsezkoak diren bost teknologia disruptiboen inplikazioak aztertzen ditu: 1) Big Data, 2) Gauzen Internet (IoT), 3) Droneak eta sateliteak, 4) Adimen artifiziala eta datu-meatzaritza, eta 5) Blockchain. Ghanan errepideak eraikitzeke proiektu baten eraginaren ebaluazio baten kasu argigarri bat jasotzen du, aplikazio potentzialak frogatzeko.
Aplikaturako teknologia	Ikaskuntza automatikoko algoritmoak: ▪ datu historikoetan esperimentu naturalak aurkitzea, programen esleipenari/ekarpenari buruzko ondorio ia esperimentalak ateratzeko; ▪ ebaluazio-prozesuak automatizatzea; ▪ aurreikuspen-ereduen bidez parekatutako konparazio-taldeak sortzea; ▪ hizkuntza naturalaren prozesamendua garatzea analisi kualitatiborako.
Dokumentatutako emaitzak	▪ Datu azkarragoak, merkeagoak eta askotarikoagoak. ▪ Laginen ordez unibertsoarekin lan egiteko aukera. ▪ Testuinguru-aldagaiak konplexutasuna aztertzeke. ▪ Sistemen analisisien integrazioa. ▪ Kasu atipiko positiboen azterketa sakonagoa. ▪ Iraunkortasunaren ebaluazioa epe luzera. ▪ Kasu bat dokumentatzen da non giza laguntza zuten PNL algoritmoak zortzi aldiz azkarragoak izan baitziren giza kodetzaileak baino, zehaztasunik galdu gabe.
Arriskuak eta mugak	▪ Teknologiekiko ohitasun gutxi, ebaluatzaile beterranoen erresistentziak, berrantolaketaetako eta ekipoen eraikuntzarako beharra. ▪ Ekipoetan, softwarean, langileen arloan eta aholkularien eremuan inbertsio esanguratsuak egitea beharrezkotzat jotzen da. ▪ Ebaluazio-bulegoen, datu-zentroen eta sail operatiboen arteko harremanak berregituratzeko beharra. ▪ Erakundeen arteko arrakala digitalaren arriskua. ▪ Datuen pribatutasunari, propietateari eta erabilera etikoari buruzko kezak.
Erreferentzia	Bamberger, M., & York, P. (2020). Transforming Evaluation in the 4th Industrial Revolution: Exciting Opportunities and New Challenges. <i>eVALUation Matters Second Quarter 2020</i> , 11-21. African Development Bank Group.

16. Machine Learning, BGDSren eragina ebaluatzeko (Espainia)

Erabilera-kasua	Adimen artifizial gainbegiratua aplikatzea Bizitzeko Gutxieneko Diru-sarrerak Espainiako etheen finantza-ongizatean duen eragina ebaluatzeko, eragin objektiboak (diru-sarrerak, pobrezia) eta subjektiboak (egoera ekonomikoaren pertzepzioa) bereiziz, gizarteratzeko politika publikoak optimizatzeko.
Testuingurua	BGDSk Espainian duen ezarpenari buruzko azterlana, London School of Economics-en esparru akademikoan garatua.
Deskribapen orokorra	Gainbegiratutako adimen artifiziala erabiltzen da, zehazki, Kontrol Sintetikoko Metodoa (SCM) eta haren bertsio hobetua, Ridge Augmented SCM (RASCN). Teknika horiek, gainbegiratutako machine learning eremuaren barruan, kontrafaktual automatizatu bat sortzeko aukera ematen dute (Espainia BGDSrik gabe), BGDSren benetako efektua kausalki balioesteko, jokabide-suposizioak saihestuz eta hautaketa-alborapenak zuzenduz.
Aplikaturako teknologia	Eurostaten datu panel nazionalekin entrenatutako AAren algoritmoak (2010-2022). Aztertutako adierazleak: pobrezia-tasa, pobrezia-arrakala, batez besteko diru-sarrera eta pertzepzio subjektiboa. EBko 11 herrialdek osatutako kontrol sintetikoko talde batekin alderatu da.
Dokumentatutako emaitzak	▪ Ez du eragin esanguratsurik adierazle objektiboetan lehen urte eta erdian. ▪ 2,5 urtera, ongizate subjektiboaren hobekuntza hautemangarria (10-14 puntuko igoera pertzepzio positiboan). ▪ Pobreziaren arrakala eta batez besteko errenta bezalako adierazleak zertxobait hobetzen dira, baina garrantzi estatistikorik gabe. ▪ AAri esker, inpaktu diferentzialak hauteman daitezke ongizatearen dimentsio objektibo eta hautemanetan.
Arriskuak eta mugak	▪ Eragin estatistiko mugatua ebaluatutako adierazle objektiboetan. ▪ Ezarpenean izandako atzerapenek eta BGDSaren estaldura baxuak eragina murrizten dute. ▪ Pertzepzioak banakako joerei eta kanpoko testuinguru sozioekonomikoari lotuta daude (COVID, inflazioa). ▪ Kalitatearen mendeotasuna eta herrialdeen arteko datu alderagarrien eskuragarritasuna.
Erreferentzia	Bilbao-Goyoaga, E. (2023) Perceptions Matter: Quasi-Experimental Evidence on the Effects of Spain's New Minimum Income on Households' Financial Wellbeing, <i>Social Policy Working Paper 02-23</i> , London: LSE Department of Social Policy. https://www.lse.ac.uk/social-policy/Assets/Documents/PDF/working-paper-series/WPS-02-23-Eugenia-Bilbao-Goyoaga.pdf

17. Hizkuntza Eredu Handia (LLM) ebaluazioan (azterketa analitikoa)

Erabilera-kasua	Saiakuntza analitikoa eta esploratzailea, hausnarketa kontzeptuala eta aplikazio-adibide praktikoak konbinatzen dituena. Ebaluazio-prozesuaren hainbat etapatan ChatGPT laguntza-tresna gisa erabiltzearen aukerak eta erronkak aztertzen ditu artikulak, erreferentzia-terminoak sortzetik azken txostenak barreiatzeraino.
Testuingurua	Ez dago espezifikazio geografiko zehatzik, baina ebaluazio-diziplinari buruzko ikuspegi orokorra du. 2022-2023: ChatGPTren hasierako erabilgarritasun publikoa (2023an argitaratua, GPT-3n oinarritutako esperientziekin).
Deskribapen orokorra	Egileak "hacking by the prompt" (sormenezko erabilera eta elkarrizketarako jarraibide trebeak) erabilera aztertzen du, ChatGPTrekin elkarrekintzan aritzeko ebaluazio-lanaren hainbat alderditan. Adibide eta hausnarketa kritikoen bidez, ebaluazio-praktikan AA txertatzeko aukerak eta mugak erakusten dira. Ebaluazio-prozesuaren etapa bakoitzerako prompt-en (indikazioak) adibide zehatzak sartu dira.
Aplikaturako teknologia	Hizkuntza Eredu Handia (LLM) ChatGPT-3 eta ChatGPT-4ren erreferentziak, hizkuntza naturala prozesatzeko eta sortzeko gaitasuna duen eta aurrez trebatuta dagoen AA sortzailea, elkarrizketa-interfazearen bidez eskura daitekeena.
Dokumentatutako emaitzak	Azterlanak ebaluazioko hainbat erabilera dokumentatzen ditu: ▪ Dokumentu teknikoaren sorrera (erreferentzia-terminoak, ohar kontzeptualak, esparru logikoak). ▪ Tresnen diseinua (galdetegiak, elkarrizketa-gidak). ▪ Testuaren azterketa (laburpenak egitea, gako-hitzak ateratzea, sentimendua aztertzea). ▪ Komunikaziorako laguntza (edukia audientzia espezifikoetara egokitzea, itzultzea).
Arriskuak eta mugak	▪ ChatGPTk probabilitateekin jarduten du, ez ezagutza faktiko egiaztatuekin. ▪ "Haluzinazioak" sortzen ditu (fabrikatutako informazioak onargarria dirudi, baina faltsua da). ▪ Entrenamendu-datuei buruzko gardentasunik ezak kezka sortzen du munduaren joera eta ikuspegi nagusiei buruz. Datuen pribatutasunari eta etikari buruzko kezkak. ▪ Pentsamendu kritikoa eta giza judizioa eraginkortasun automatizatuarekin ordezkatzeko arriskua. ▪ AA sistemen bilakaerari eta erregulazioari buruzko kontrolik eza.
Erreferentzia	Ferretti, S. (2023). Hacking by the prompt: Innovative ways to utilize ChatGPT for evaluators. <i>New Directions for Evaluation</i> , 2023, 73-84. https://doi.org/10.1002/ev.20557

18. Machine Learning aldaketaren teorian oinarritutako ebaluaziorako (nazio anitzekoa)

Erabilera-kasua	Adimen artifizialen aplikazioa, aldaketa-teorietan oinarritutako programa konplexuen ebaluazioetan edukiaren analisia eta sintesi kualitatiboa automatizatzeko eta hobetzeko.
Testuingurua	Munduko Bankuaren Independent Evaluation Group (IEG) delakoaren gaikako ebaluazioa, haurren desnutrizio kronikoa murrizteko eta hazkundea atzeratzeko programei buruzkoa: 64 herrialdetako proiektuei buruzko 392 txosten.
Deskribapen orokorra	AAren hiru ikuspegi osagarriren ezarpena: (1) ikaskuntza automatiko gainbegiratua (SML), testuak aurrez definitutako kontzeptu-esparru baten arabera deduktiboki sailkatzeko; (2) gainbegiratu gabeko ikaskuntza automatikoa (UML), arrakasta- eta porrot-faktoreen gai induktibo berriak identifikatzeko; (3) ezagutza-grafoak, emaitzak aldaketaren teoria baten arabera egituratzeko eta teoria horretan oinarritutako azterketa errazteko.
Aplikaturako teknologia	▪ Testuak zenbakiz irudikatzeko bektore-espazioen modelatzea. ▪ Ikaskuntza automatiko gainbegiratua (testu-sailkatzaileak) eta gainbegiratu gabea (gaiak modelatzea). ▪ t-SNE bisualizazioa (t-distributed stochastic neighbor embedding). ▪ Ezagutza-grafoak, arauetan oinarritutako arrazoiketarekin (Vadalog).
Dokumentatutako emaitzak	▪ SML: zehaztasun apala etiketa zehatzak aurreikusteko orduan; zehaztasun handia goi-mailako kategoriak aurreikusteko orduan; gaitasun handia programen ezaugarri bereizgarriak herrialde bakoitzeko identifikatzeko t-SNE bisualizazioaren bidez. ▪ UML: errendimendu bikaina 10 gai koherente eta garrantzitsu identifikatzean, proiektuaren jardunean eragina duten funtsezko faktore gisa; ondorengo balioztatze estatistikoa. ▪ Ezagutza-grafoak: ikaskuntza automatikoko outputak egituratzeko gaitasuna, aldaketa-teoriaren arabera; esku-hartzeen eta emaitza-adierazleen lorpenen arteko lotura sinpleen identifikazioa.
Arriskuak eta mugak	▪ Datuak prestatzeko eta esparru analitikoak eta grafoen eskema ezartzeko hasierako denboraren inbertsio esanguratsua. ▪ Adituen mendekotasuna testu-atal garrantzitsuak eskuz atera eta etiketatzeko. ▪ Ebidentzia osatu gabea, esparru kontzeptualeko elementuak lotuz (nodo batzuk "umezurtz" geratzen dira).
Erreferentzia	Franzen, S. et al. (2022). Advanced Content Analysis: Can Artificial Intelligence Accelerate Theory-Driven Complex Program Evaluation? <i>IEG Methods and Evaluation Capacity Development Working Paper Series</i> . Independent Evaluation Group. Washington, DC: World Bank.

19. Ebaluatzaileak: NLP eta ikaskuntza automatikoko ereduak (Kanbodia)

Erabilera-kasua	"Ebaluatzaileak" sartzea eta aplikatzea, programen ebaluazioetan datu kualitatibo konplexuak aztertzeko eta gomendio eragingarriak sortzeko tresna metodologiko gisa.
Testuingurua	Kanbodiako saneamendu-programa, iDEk ezarria eta Australian Aidek finantzatua, Water for Women programaren bidez.
Deskribapen orokorra	Artikuluak "pertsonek" ebaluatzaileen erabilera aztertzen du, hau da, datu errealean oinarritutako gezurrezko profilak, garapen-programetan ebaluazio kualitatiboen analisia eta erabilgarritasuna hobetzeko tresna gisa. Aztertutako kasuan, ikerketa-talde batek ikuspegi hori aplikatu zuen genero-zeharkakotasuneko programa batek Kanbodiako saneamendu-programa bateko langileengan (staff) sortu zuen eragina ebaluatzean. 176 langilerengandik jasotako 199 mikro-narratibetatik 14 pertsona sortu ziren.
Aplikaturako teknologia	Hizkuntza Naturalaren Prozesamendua (NLP) bi algoritmoen bidez: Latent Dirichlet Allocation (LDA) eta Structural Topic Modeling (STM), identifikaturako pertsona-taldeak baliozkotzeko erabiltzen dira, eskuzko azterketa tematiko kualitatiboaren bidez.
Dokumentaturako emaitzak	14 pertsona identifikatu ziren, hiru kategoriatan banatuta: behatzaile kritikoak (% 34), auto-erreflexibo kritikoak (% 18) eta aktore kritikoak (% 48). Pertsonak datu kualitatibo konplexuak benetan ikustea eta komunikatzea ahalbidetu zuten. - Generoa txertatzeko estrategiei buruzko eztabaidak erraztu zituzten programako liderrekin, eta nolabaiteko anonimitasuna eman zuten.
Arriskuak eta mugak	- Atzera begirako mikro-narratiben bilketan, ikertzaileek ezin dute egiaztatu istorioen baliozkotasuna, eta horrek nolabaiteko edertzea ekar dezake; programaren nolabaiteko distantzia (fisikoa eta denborazkoa). - Zentsu-laginketak anonimitasun-arazoak sortu zituen. Pertsonak horiek arintzen lagundu zuten arren, oraindik ere kontsiderazio etiko garrantzitsua dira.
Erreferentzia	MacArthur, J., Moung, V., Carrard, N., & Willetts, J. (2024). Personas for program evaluation: Insights from a gender-focused evaluation in Cambodia. <i>Evaluation</i> , 31(1), 70-91. https://doi.org/10.1177/13563890241284425

20. Jarduera-joerak detektatzeko big data (Israel)

Erabilera-kasua	Testu-meatzaritzan, ikaskuntza automatikoan eta datuen bisualizazioan oinarritutako metodologia automatizatua garatu eta ezartzea, egituratu gabeko jardun- eta ebaluazio-txostenen bolumen handietan joerak detektatzeko.
Testuingurua	Sistema hau Israelgo Estatuaren Interbentzioak (State Comptroller) 30 urtez (1987-2017) ekoiztako 1.700 auditoria-txosten eta jardunaren ebaluazioa aztertzeke aplikatu zen.
Deskribapen orokorra	Metodoak Wikipedia eta kanpoko beste iturri batzuk erabiltzen ditu ezagutza-oinarri gisa dokumentuetan termino garrantzitsuak identifikatzeko. Gainbegiratu gabeko ikaskuntza automatikoko algoritmoen bidez, sistemak irudikapen bisual bat sortzen du grafiko moduan, non semantikoki lotutako txostenak klusterretan multzokatuta agertzen diren. Bistaratzeko dauden koloreek Estatuko hainbat kontu-hartzaileen zerbitzu-aldiak adierazten dituzte. Metodologia horrek aukera eman zien ikertzaileei edukiak aztertzeke metodo tradizionalen bidez ia detektaezinak izango ziren ereduak eta joerak identifikatzeko.
Aplikaturako teknologia	Hizkuntza naturalaren analisi estatistikoa (SNLP), gainbegiratu gabeko ikaskuntza automatikoa testuak sailkatzeko, analisi semantiko esplizitua (ESA), dimentsionaltasun ez-linealaren murrizketa, eta grafoak bistaratzeko algoritmoak (Force Atlas eta Fruchterman-Reingold).
Dokumentaturako emaitzak	Sistemak modu eraginkorrean identifikatu zituen gaikako klusterrak, denbora-ibilbideak eta dokumentuen arteko harremanak, aurrez definitutako gako-hitzen bidez bilatu beharrik gabe. Honelako patroiak agertu ziren: (1) kontu-ikuskaritza tradizionaletatik jarduera-auditorietarako trantsizioa, (2) aldaketak Estatuko kontu-hartzaileen ikuspegietan, (3) jarraitutasuna/etena ebaluatutako gaietan, eta (4) lotura kluster tematikoen artean. Ikuspegiari esker, 1700 txostenen berrikuspenean 50 txosten adierazgarrietara murriztu ahal izan zen, domeinuaren ikuspegi orokorra lortzeko.
Arriskuak eta mugak	- Sistemak programazio pertsonalizatua behar du eremu bakoitzerako. Ezagutza teknikoak eskatzen dituzten doikuntza espezifikoak behar ditu, garapen-esperientziarik ez duten erabiltzaileentzako sarrera-oztopo bat sortuz. - Gako-hitzen arabera bilaketa-alborapena saihesten duen arren, beste alborapen posible batzuk sartzen ditu zenbait erabakiren bidez, hala nola Wikipedia ezagutzaren oinarri gisa erabiltzean edo datuen irudikapenean eragina duten bisualizazio-algoritmo espezifikoak hautatzean.
Erreferentzia	Raveh, E., Ofek, Y., Bekkerman, R., & Cohen, H. (2020). Applying Big Data visualization to detect trends in 30 years of performance reports. <i>Evaluation</i> , 26(4), 516-540. https://doi.org/10.1177/1356389020905322

21. Case Based Modelling (CBM) eta eszenatokien simulazioa (Erresuma Batua)

Erabilera-kasua	Kasuetan oinarritutako modelatze-metodologia bat garatzea eta aplikatzea eta agertokiak simulatzea (CBSS), mundu errealeko sistema konplexuetan esku-hartze politikoak ebaluatzeko.
Testuingurua	Erresuma Batuko gobernuaren Negozio, Energia eta Industria Estrategia Sailaren (BEIS) eta Hezkuntza Sailaren (DfE) programei buruz 2008 eta 2018 artean argitaratutako lau ebaluazio aztertzen ditu, eta zehazki: Heat Network Investments Project-en (HNIP) ebaluazio pilotua; Renewable Heat Incentive-ren (RHI) ebaluazioa – BEIS; Frontline Fast-Track Social Work programa pilotuaren ebaluazioa; Sure Start-en ebaluazio nazionala – DfE.
Deskribapen orokorra	CBSS (Case-Based Scenario Simulation) ex-post ebaluaziorako ikerketa sozialeko tresna bat da, kasuetan oinarritutako eta datuetara bideratutako sistemen modelatze konputazionalako ikuspegian oinarritua. COMPLEX-IT web-plataformaren bidez ezarria, ebaluatzaileei aukera ematen die kasuen patroiak identifikatzeko konglomeratuen analisisien bidez (clustering), emaitzak berresteko mapa autoantolatuekin (SOM) eta agertoki alternatiboak simulatzeko kontrafaktualak eta proiektzioak arakatzeko aldera.
Aplikaturako teknologia	Gainbegiratu gabeko ikaskuntza automatikoko metodoak, zehazki: (1) k-means konglomeratuen analisisa kasualdeak identifikatzeko, (2) neurona-sare artifizialak Kohonen (SOM) mapa auto-antolatuen bidez kasuen bistaratze topografikoa egiteko, eta (3) Monte Carlo simulazioak sentikortasun analisiak egiteko.
Dokumentatutako emaitzak	▪ Sistema konplexuak modelatzeko gaitasun frogatua du, azalpena faktore indibidualetara murriztu gabe. ▪ Esku-hartzei emandako erantzun desberdinak jasotzen dituzten azpitaldeak identifikatu dira. ▪ Kontrafaktualak eta agertoki alternatiboak aztertu dira. ▪ Taldekatze-soluzioen sendotasuna ebaluatu da. ▪ Ebaluazio errealistaren ikuspegiak osatzen ditu.
Arriskuak eta mugak	▪ Sistema konplexuetako ziurgabetasunei aurre egiteko mugak daude. ▪ Ez ditu behin betiko kausazko erlazioak identifikatzen, baizik eta erlazio horiek esploratzeko ingurune bat ematen du. ▪ Kasuak eta kasuen profilak behar bezala zehaztu behar dira. ▪ Datuen kalitatearekiko eta osotasunarekiko sentikorra da. ▪ Ebaluatzailearen eta ikaskuntza-ingurunearen arteko etengabeko elkarreraginaren mende dago.
Erreferentzia	Schimpf, C., Barbrook-Johnson, P., & Castellani, B. (2021). Cased-based modelling and scenario simulation for ex-post evaluation. <i>Evaluation</i> , 27(1), 116-137. https://doi.org/10.1177/1356389020978490

22. Ebaluatzaileak prestatzeko txatbotak (AEB eta Kanada)

Erabilera-kasua	Adimen artifizialak ebaluatzaileen prestakuntzan izan ditzakeen inplikazioak eta aplikazioak aztertzea, txatboten prototipoak barne, ebaluazioaren arloko prestakuntza-esperientzia hobetzeko.
Testuingurua	Esparru akademikoa eta hezkuntza-eremua ebaluazioko prestakuntza-programetan, batez ere Estatu Batuetan eta Kanadan, 2022-2023an, aplikazio globaleko aurreikuspenekin.
Deskribapen orokorra	Ebaluazioan prestatzeko AA sortzailean oinarritutako hiru txatbot-prototiporen garapena: (1) ebaluazioko prestakuntza-programei buruzko galderei erantzuteko orientatzaile bat, (2) ebaluazio-teoria espezifikoak argitzeko irakasle-laguntzaile bat eta (3) ebaluazio-praktikari buruzko funtsezko galderei erantzuteko aholkulari bat.
Aplikaturako teknologia	Hizkuntza Eredu Handietan (LLM) oinarritutako adimen artifizial sortzailea, zehazki OpenAIren GPT 3.5 turbo eredua erabiliz, Hizkuntza Naturalaren Prozesamenduarekin (NLP), Berreskurapen bidezko Sorkuntza Areagotuaren bidez (RAG) ezarria, ebaluazioari buruzko corpus dokumental espezifikoetan oinarritutako erantzunak emateko.
Dokumentatutako emaitzak	Prototipoek honako hauetarako gaitasuna erakutsi zuten: ▪ AEBetako ebaluazioari buruzko prestakuntza-programei buruzko informazio zehatza emateko; ▪ galdera teorikoen erantzunak sintetizatzeko, hainbat dokumentu konbinatuz; ▪ ebaluazio-praktikari buruzko galderak erantzuteko, sortzen ari diren ebaluatzaileei egokitutako ikuspegiarekin.
Arriskuak eta mugak	▪ Teknikoak: itzulpen automatiko eleaniztunaren kalitatearen aldakortasuna, bereziki ebaluazio-terminologia espezializaturako; mugak erabilitako ereduen testuinguru-leihoan; haluzinazioak edo erantzun okerrak izateko arriskua. ▪ Metodologikoak: txatboten errendimenduaren atariko ebaluazioa eta ebaluazio ez-sistematikoa; sortutako erantzunen zehaztasuna eta erabilgarritasuna ebaluatzeke metrika formalik eza; benetako erabiltzaileekin proba mugatua. ▪ Etikoak: ikerketan eta ebaluazioan alborapen eta bazterketa historikoak betikotzeko aukera, AAren ereduen bidez; teknologia horiek eskuratzeko ekitateari buruzko kezkak, bereziki Hego Globalean.
Erreferentzia	Tilton, Z., LaVelle, J. M., Ford, T., & Montenegro, M. (2023). Artificial intelligence and the future of evaluation education: Possibilities and prototypes. <i>New Directions for Evaluation</i> , 2023(178-179), 97-109. https://doi.org/10.1002/ev.20564



besaldi

Enplegu eta inklusio
politiken ebaluazio organoa
Órgano de evaluación de las
políticas de empleo e inclusión



EUSKO JAURLARITZA
GOBIERNO VASCO

EKONOMIA, LAN ETA
ENPLEGU SAILA

DEPARTAMENTO DE ECONOMÍA,
TRABAJO Y EMPLEO