Ekaia, ale berezia 2025, 47-65
https://doi.org/10.1387/ekaia.26310

ekaia
ZIENTZIA eta TEKNOLOGIA ALDIZKARIA

ISSN 0214-9001 – eISSN 2444-3255

ChatGPT irakurketa errazeko testuak egokitzeko baliagarria eta fidagarria? *Lectura fácil*-eko zenbait testuren azterketa

(Is ChatGPT useful and reliable to adapt texts to easy-to-read? An analysis of some Lectura fácil *texts)*

Margot Madina, Melanie Siegel*

Darmstadt University of Applied Sciences (Hochschule Darmstadt), Darmstadt, Alemania

Itziar Gonzalez-Dios

HiTZ Center-Ixa, University of the Basque Country (UPV/EHU)

LABURPENA: Irakurketa errazaren helburua da arazo kognitiboak edo ikasteko zailtasunak dituzten irakurleentzat testu eta eduki irisgarriagoak sortzea, eta, horretarako, hizkuntza sinplea eta argia erabiltzen da. Testuak irakurketa errazera egokitzea prozesu garestia da, bai denbora aldetik baita ekonomia aldetik ere. Prozesu hori azkartzen ChatGPT bezalako tresnek lagun dezakete. Lan honetan, irakurketa erraza zer den azaltzeaz gain, ChatGPT-4k irakurketa errazera egokitzapena nola egiten duen aztertu dugu. Zehazki, gaztelaniazko *lectura fácil*eko testuak izan ditugu aztergai eta hiru ebaluazio egin ditugu: i) neurri linguistiko automatikoetan oinarritutakoa; ii) testuen sinplifikazioko *capability*etan oinarritutakoa; eta iii) helburu talde diren erabiltzaileekin egindakoa. Gure emaitzen arabera, oso zaila da testuak ChatGPT-4ri egokiaraziko dion *prompt* egoki bat aukeratzea, sortu dituen testuak ez dituzte irakurketa errazeko gidalerroak jarraitzen eta, askotan, ChatGPT-4k sortutako testua ez da egokia helburu taldearentzat.

HITZ GAKOAK: Irakurketa erraza, ChatGPT, ebaluazioa.

ABSTRACT: *Easy-to-Read (E2R) aims to generate more accessible texts and content for readers with cognitive problems or learning difficulties, using simple and clear language. The process of adapting and creating E2R texts is very expensive and time-consuming. Due to the success of Large Language Models (LLMs) such as ChatGPT and their ability to generate written language, it is likely to think that such models can help in the adaptation or creation of text in E2R. In this paper, we explore the concept of E2R, its underlying principles and applications, and provides a preliminary study on the usefulness of ChatGPT-4 for E2R text adaptation. We focus on the Spanish language and its E2R variant,* Lectura Fácil (LF)*. We consider a range of prompts that can be used and the differences in output that this produces. We then carry out a three-folded evaluation on 10 texts adapted by ChatGPT-4: 1. an automated evaluation to check values related to the readability of texts, 2. a checklist-based manual evaluation (for which we also propose three new* capabilities*) and 3. a users' evaluation with people with cognitive disabilities. We show that it is difficult to choose the best* prompt *to make ChatGPT-4 adapt texts to LF. Furthermore, the generated output does not follow the E2R text rules, so it is often not suitable for the target audience.*

KEYWORDS: Easy-to-read, ChatGPT, evaluation.

*** *Harremanetan jartzeko* / *Corresponding author:*** Margot Madina, Darmstadt University of Applied Sciences (Hochschule Darmstadt), Darmstadt (Alemania). – margot.madina-gonzalez@h-da.de – https://orcid.org/0009-0001-9185-4583

Nola aipatu* / *How to cite: Madina, Margot; Gonzalez-Dios, Itziar: Siegel, Melanie (2025). «ChatGPT irakurketa errazeko testuak egokitzeko baliagarria eta fidagarria? *Lectura fácil*-eko zenbait testuren azterketa». *Ekaia*, ale berezia 2025, 47-65. (https://doi.org/10.1387/ekaia.26310).

Jasotze-data: 2025, maiatzak 13; Onartze-data: 2024, urriak 08.

ISSN 0214-9001 - eISSN 2444-3255 / © UPV/EHU Press 2025

Lan hau *Creative Commons Aitortu-EzKomertziala-PartekatuBerdin 4.0 Nazioartekoa* lizentzia baten mende dago

1. **SARRERA**

Gaur egun, munduan, Nazio Batuen erakundearen arabera, populazioaren % 16k alfabetatze-arazoak ditu,[1] eta Ekonomia Lankidetza eta Garapenerako Antolakundea osatzen duten herrialdeetan[2] oinarrizko alfabetatze maila duten helduak % 4,9 eta % 27,7 artean daude [1]. *Programme for the International Assessment of Adult Competencies* (PIAAC) testean ikusi da gaztelaniadun herrialdeetako alfabetatze maila oso baxuak direla. Bestalde, Imsersoren arabera, Espainian 283.256 pertsonak dute % 33ko desgaitasuna edo handiagoa [2].

Pertsona horien komunikazioa eta ulermena bermatzeko, irakurketa erraza (IE) sortu zen; edukiak eta informazioa irisgarriagoa egitea helburu izanik, oinarrizko hiztegia eta egitura sintaktiko sinpleak erabiltzean datza. Hasiera batean, IE arazo kognitiboak zituzten pertsonentzat sortu zen, baina, egun, beste talde batzuetara ere zabaldu da: garapen-arazoak dituztenak, adineko jendea, hizkuntzarekin lotutako gaixotasunak dituztenak, alfabetatze maila baxua dutenak, entzumen-arazoak dituztenak edo tokiko hizkuntza hitz egiten ez dutenak, besteak beste. IEk izendapen ezberdinak ditu hizkuntzaren arabera eta, esate baterako, gaztelaniaz *lectura fácil* (LF)[3] izena hartu du, alemanez *leicthe Sprache*, eta ingelesez *easy-to-read* esate baterako.

IEko testu oso gutxi daude, eta daudenak eskuz sortu dira. Beraz, testu gutxi izanda, zaila da Hizkuntzaren Prozesamenduko (HP) teknikak aplikatzea datu-multzo handiak behar baitira. Hori dela eta, lan honetan azken hilabeteetan arrakasta handia lortu duen ChatGPT [3] tresna LFko testuak sortzeko erabilgarria ote den aztertu nahi dugu. ChatGPT 2022 azaroan plazaratu zen produktu komertzial bat da: haren webgunean erregistratzen den edozein erabiltzailek badu tresna horrekin txateatzeko eta galderak egiteko aukera. Hizkuntza-eredu sortzaile bat du oinarri (bertsioaren arabera aldatzen dena), eta, ondoren, gizakiek emandako feedbackean oinarritutako errefortzu bidezko ikasketa eginez entrenatu dute.[4] Entrenamendurako erabili diren datuak eta parametro kopuruak ez dira ezagunak. ChatGPT teknologia disruptiboa izan da gizartean, zientzian eta eremu profesionaletan erabilera handia izan baitu [4, 5], eta, 2024ko ekaineko datuen arabera, 180 milioi erabiltzaile ditu.[5]

---

[1] https://www.un.org/en/chronicle/article/education-all-rising-challenge (atzipen-data: 2023-10-12).

[2] https://www.oecd.org/ (atzipen-data: 2024-05-09).

[3] Lan honetan IE erabiliko dugu irakurketa errazari buruz oro har hitz egiten dugunean eta LF gaztelaniazkoari buruz hitz egitean.

[4] https://openai.com/chatgpt/ (atzipen-data: 2023-10-12).

[5] https://explodingtopics.com/blog/chatgpt-users (atzipen-data: 2024-07-04).

Gure helburua da ChatGPTek, GPT-4 hizkuntza-eredu sortzailea duen bertsioak [3], sortzen dituen testuak ebaluatzea: izan ditzaketen hutsuneak eta hutsegiteak azaleratzea eta LFko arauak betetzen dituzten ala ez ikustea erabiltzaile arruntei, erabiltzea erabakitzen badute, gabeziez eta onurez ohartarazteko. Lan honek, beraz, ez du helburu sistema bat kuantitatiboki ebaluatzea ezta ataza horretan egokiena izan daitekeen sistema zein den jakitea ere. Lan honetan adimen artifizialeko tresna arrakastatsu eta erabili bat ebaluatu nahi dugu gizartea haren alde onez eta arriskuez ohartarazteko; izan ere, zientzialarion ardura ere bada gure teknologiek gizartean izan ditzaketen eraginak aztertzea eta pedagogia egitea. Lan honetan emango ditugun adibideen euskarazko itzulpenak A eranskinean daude.

2. IRAKURKETA ERRAZEKO DOKUMENTUEN EGOKITZAPENA: UNE ARAUA ETA ANTZEKO LANAK

Atal honetan irakurketa errazeko gaztelaniazko testuak nola egokitzen diren azalduko dugu, eta gure lanarekin aldi berean egin diren eta egiten ari diren antzeko lan batzuk aipatuko ditugu.

LF zer den *Norma UNE 153101:2018 EX de Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos*[6][**6**] arauan deskribatuta dago. Jendeak informaziorako duen eskubidea bermatzea du helburu arauak, batez ere idatzizko dokumentuetan dagoen informazioari dagokionez. UNE 153101:2018 EX-izendapenaren esanahia hau da: 153101 zenbakia estandarrari dagokio, 2018 argitalpen-urtea da eta «EX»-ek, berriz, estandar esperimentala dela adierazten du. Izan ere, estandar hori da munduan sortu den lehen estandarra IErako [7].

UNE 153101:2018 EX arauak testu estandar bat LFra egokitzeko prozesua deskribatzen du,[7] hau da, LFko dokumentu bat sortzeko prozesua, eta LFko testua idazteko eta dokumentua diseinatzeko jarraibideak eta gomendioak ematen ditu. Gainera, bi eranskin daude. Lehenengoak dokumentua ulertzen dutela egiaztatzeko onuradunekin edo hartzaileekin egin daitezkeen jardueren adibide argigarriak biltzen ditu. Bigarrenak kontraste kromatikoen taula bat du. Arau batzuk argi definituta daude, eta erraz aplikatzen dira. Adibidez, «Ez erabili siglarik» araua zehatza da, testu multzo batean erraz aplika daitekeena, anbiguotasun esanguratsurik gabe. Hala ere, beste jarraibide batzuk ez dira hain zehatzak, edo hedadura zabalagoa

[6] UNE https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0060036 (atzipen-data: 2024-07-04).

[7] LFra zehazki eta ez IEra oro har, gaztelanian gertatzen diren zailtasun linguistikoak jasotzen baititu.

dute, eta horrek erronka handiagoa dakar koherentziaz aplikatzeko. Arau horren adibide bat da honako hau: «Ez erabili testuari informaziorik eransten ez dion eta irakurtzeko denbora gehiago behar duen hitzik». Arau hori subjektiboa da berez, eta hainbat modutan interpreta daiteke, testuinguruaren arabera. Irizpide espliziturik ez dagoenez, zaila da jarraibide hori praktika estandar batean kuantifikatzea eta kodetzea, eta horrek inkongruentziak eragiten ditu aplikazioan [8]. Esan bezala, arau horretan testuak LFra nola egokitu ere azaltzen da.

Testuak LFra egokitzeko lau rol nagusi hauek behar dira:

— Egokitzailea: LFko dokumentua idazten duena.
— Diseinatzailea/formatuaren diseinatzailea: dokumentua formatu eta itxura egoki batean jartzen duena.
— Dinamizatzailea: ebaluazio fasea kudeatzen duena.
— Ebaluatzaileak: irakur dezaketen baina ulermen-arazoak dituzten pertsonak, esperimentuaren ebaluazioan parte hartzen dutenak.

Prozesua bi fasetan egiten da:

1. Egokitzapena: testuak LFra bihurtzea. Egokitzaileak dokumentua LFn idazten du eta diseinatzaileak formatu egokian prestatzen du.
2. Ebaluazioa: testuen baliagarritasuna eta egokitasuna aztertzea. Dinamizatzaileak ebaluatzaileekin saioak antolatzen ditu eta ebaluatzaileek dokumentua irakurtzean izan dituzten arazoak kontatzen dizkiote dinamizatzaileari. Dinamizatzaileak informazio hori egokitzaileari eta diseinatzaileari ematen dio, eta bigarren egokitzapen fase bat hasten da.

Fase horiek ebaluatzaileek testua egokia dela esaten duten arte errepikatzen dira.

Prozesu luze horretan, pentsa daiteke ChatGPT bezalako HPko tresnek lagun dezaketela. Izan ere, gure lanarekin batera, antzeko ikerketak egiten ari dira: alemanezko IEra egokitzeko esperimentuak egin dira [9] eta gaztelaniazko LF CLEARSIM corpusa sortzeko ere erabiltzen ari dira [10].

Alemanezko IErako egokitzapenaren lanean [9], administrazioko 20 testu ChatGPTrekin egokitu dituzte. Horretarako, hiru urratsetan, tresnari agindu diote i) testua informazio ez-garrantzitsua alde batera utziz berridaztea; ii) testua egitura sintaktiko sinpleekin berridaztea; eta, iii) hitz zailen azalpenak testuan gehitzea. Hiru aldaketa horiek biltzen dituen bertsioa ebaluatu dute eta ikusi dute sortutako testuak sinpleagoak direla, baina ez dituztela IEko arauak betetzen. Gainera, testuen edukia ez da beti zuzena. Hori dela eta, aipatzen dute ChatGPTek sortutako testuak postedi-

tatu behar direla eta lan hori egiteko azpimarratzen dute lan profesionala beharrezkoa dela.[8]

LF CLEARSIM corpusa [10] sortzeari dagokionez, kiroleko, aisialdiko eta kulturako artikuluak bildu dituzte eta ChatGPTri agindu diote artikulu horien laburpenak eta sinplifikazioak egiteko. Ondoren, testu horiek eskuz errebisatu dituzte eta *prompt*ak (instrukzioak, aginduak) egokitu dituzte eskuzko lana arintzeko. Azkenik, gobernuz kanpoko erakunde (GKE) bateko profesional batek testuak errebisatu ditu eta egiaztatu du LFko arauak betetzen direla. Lana argitaratu duten unean 400 testu dituzte *gold* estandarrean.

3. **METODOLOGIA**

Atal honetan, esperimentuak egiteko erabili ditugun baliabideak (testuak eta tresnak) eta proposatu ditugun ebaluazio metodoak azalduko ditugu.

3.1. **Esperimentazioa**

Esperimentuak egiteko, erabili ditugun testuak *Irekia*[9] web-orritik erauzi ditugu, herritarren eta Eusko Jaurlaritzaren arteko komunikazioa ahalbidetzen duen Interneteko ataritik. *Irekia*n albiste batzuk LFra egokituta daude, eta, datuen kutsadura ekiditeko, 2023an argitaratutako zenbait gairi buruzko 10 albiste aukeratu ditugu ausaz gure esperimentuetan (ChatGPTek 2021era arteko datuak ditu).

Erabili dugun tresna ChatGPT-4 izan da; zehazki, 2023ko abuztuaren 4ko bertsioa.[10] Ikerketa-munduan beste hizkuntza-eredu sortzaile asko dauden arren (Hugging Facen 2024ko uztailaren 4an testuen sorkuntza egi-

---

[8] Alemanian, gidalibururik zabalduena eta erabiliena *Netzwerk Leichte Sprache*k garatutakoa da (*Leichte Sprache* sarea). Horretaz gain, Inclusion Europe (IE arauak jasotzen ditu hizkuntza ugarietan, alemana barne), *Barrierefreie- Informations-technik-Verordnung* (Informazio Teknologien Irisgarritasun Araudia) (BITV) eta *Duden* (ebidentzia enpirikoan oinarritutako arau linguistikoa) jarraibideek ere jasotzen dute. Alemanak ez du oraindik IErako estandar nazionalik (Espainiak duen bezala). Dena den, *DIN SPEC 33429: 2023-04 «Empfehlungen für Deutsche Leichte Sprach»* (*leichte Sprache*rako gomendioak) gomendio sorta eskuragarri dago. DIN estandarrek bezala, DIN espezifikazioek, DIN SPEC izenekoak, produktuen zerbitzu edo prozesuetarako eskakizunak definitzen dituzte. Estandarrek ez bezala, DIN SPECek ez dute erabateko adostasunik behar, ezta alderdi interesdun guztien inplikaziorik ere garapenean.

[9] *Irekia* https://www.irekia.euskadi.eus (atzipen-data: 2023-07-12).

[10] https://help.openai.com/en/articles/6825453-chatgpt-release-notes (atzipen-data: 2023-10-12).

ten duten 118.310 eredu daude[11]), lan honetan ChatGPT erabiltzea erabaki dugu gizartean gehien zabaldutakoa delako eta erabiltzaile arruntek erraz erabil dezaketena delako.

ChatGPT-4ri aginduak emateko, hainbat *prompt* edo instrukzio hiru aldiz probatu ditugu (B eranskinean ikus daitezke alde batera utzitakoak). Horien artean, hasierako probetan emaitza onenak eman dizkiguna, eta, ondorioz aukeratu duguna, hau izan da:

— **Prompt** *Las normas de redacción de Lectura Fácil Según la Norma UNE 153101:2018 EX Lectura Fácil son las siguientes: (lista de normas).*
— **Output** ...
— **Prompt** *Adapta esta noticia a Lectura Fácil siguiendo las normas de redacción de Lectura Fácil Según la Norma UNE 153101:2018 EX Lectura Fácil: (noticia).*

Alde batera utzi ditugun *prompt*ek arazo hauek eman dizkigute: irteera-testua jatorrizko testuaren laburpena izatea, LFko arauak ez jarraitzea, erantzun inkonsistenteak ematea, edo estekak edo PDFak irakurtzeko arazoak izatea.

*Prompt*ean eman ditugun arauak *UNE 153101:2018 EX Lectura Fácil Standard*-etik hartu ditugu eta, zehazki, mota hauetakoak aukeratu ditugu: i) ortografikoak ii) hiztegikoak eta fraseologikoak, iii) perpaus eta esaldi mailakoak, iv) estilokoak eta testu antolakuntzakoak, eta v) dokumentuaren aurkezpenekoak. Guztira, UNE estandarrean dauden 130 arauetatik 85 aukeratu ditugu (irudiei dagozkienak baztertu ditugu batez ere). UNE arauak inpertsonalki daude idatzita, eta guk agindu zuzen bat bezala berridatzi ditugu, aditz modalak ekiditeko. Hauek dira, esate baterako, aukeratu ditugun arauetako batzuk eta gure moldaketa:

— **Ortografikoa.** *No se debe utilizar el punto y coma (;).* → No utilices el punto y coma (;).
— **Hiztegikoak eta fraseologikoak.** *Se deben evitar los adverbios terminados en* -mente. → Evita los adverbios terminados en *-mente.*
— **Perpaus eta esaldi mailakoa.** *Se debe evitar la voz pasiva.* → Evita la voz pasiva.
— **Estilokoa eta testu antolakuntzakoa.** *Se debería evitar que el lector tenga que realizar inferencias para comprender el texto.* → Evita que el lector tenga que realizar inferencias para comprender el texto.

[11] https://huggingface.co/models?pipeline_tag=text-generation&sort=trending (atzipen-data: 2024-07-04).

— **Dokumentuaren aurkezpenekoa.** *Se debe evitar la división de las palabras con guiones al final de línea.* → Evita la división de las palabras con guiones al final de línea.

Beraz, esperimentuak egiteko, gure datu-multzoan, testu bakoitzaren i) jatorrizko bertsioa, ii) Irekiak LF formatuan emandako bertsioa, eta iii) ChatGPT-4k sortutako bertsioa dauzkagu.[12] 1 taulan, adibide gisa, 10. testuaren hasiera erakusten dugu. Taulan, LF eta ChatGPT-4 bertsioetako testuek dituzten lerro-jauziak adierazteko barra etzanak (*/*) jarri ditugu eta letra lodiz idatzitako lerroek testuan dagoen sarrera edo izenburu bat adierazten dute.

1. taula. 10. testuaren hasiera (adibidea): ezkerreko zutabean, jatorrizko testua eman digu; erdiko zutabean, LFko bertsioa, eta, eskuineko zutabean ChatGPT-4ren bertsioa.

Jatorrizko testua	LFeko bertsioa	ChatGPT-4 bertsioa
El 65% por ciento de las empresas consultadas por el Servicio Público Vasco de Empleo Lanbide tienen previsto contratar nuevo personal en los próximos dos años en sectores tan diversos como el suministro de energía y agua, la gestión de residuos, la industria manufacturera, las actividades sanitarias y de servicios sociales o la hostelería. La encuesta apunta, además, que siete de cada diez han tenido dificultades en los últimos cinco ejercicios para encontrar perfiles adecuados a sus necesidades.	Lanbide lanzó a finales del año pasado una encuesta / a más de 10.000 empresas vascas de todo tipo de sectores: / energía y agua, gestión de residuos, / industria, hostelería, servicios sociales, sanidad... / A la encuesta han respondido unas 600 empresas. / La encuesta se hizo para saber sobre sus necesidades / en cuanto a formación de sus trabajadores, / problemas y planes de contratación en el futuro.	**Las empresas en el País Vasco buscan más trabajadores** / —Muchas empresas en el País Vasco quieren contratar más gente en los próximos dos años. / —Estas empresas trabajan en áreas diferentes: energía, agua, residuos, salud, servicios, hostelería y otros. / —Pero muchas empresas tienen problemas para encontrar a los trabajadores adecuados. / **¿Qué hizo Lanbide?** / —Lanbide es un servicio de empleo en el País Vasco. / —Lanbide envió una encuesta a muchas empresas. Querían saber qué tipo de trabajadores necesitan. / —600 empresas respondieron a la encuesta.

[12] Erabili ditugun testuak eta bertsioak https://github.com/margotmg/ ChatGPT-for-Spanish-E2R-Text-Adaptation helbidean daude eskuragarri.

3.2. **Ebaluazio proposamena**

Ebaluazioari dagokionez, ez dago LFko testuak ebaluatzeko metodologia estandarrik. Eta, testuen sinplifikazio automatikoa eta laburpen automatikoak ebaluatzeko erabiltzen diren metrikak (BLEU [11], SARI [12], ROUGE [13]) ez dira fidagarriak beren atazetan ere: ez dute korrelaziorik gizakien ebaluazioekin, eta, sinplifikazio eragiketa bat baino gehiago dagoenean, ez dute antzematen besteak beste [14, 15, 16]. Hori dela eta, hiru metodo hauetan oinarritzen den ebaluazioa proposatu dugu:

— **Ezaugarri linguistikoetan oinarritutako ebaluazio automatikoa.** *MultiAzterTest*[13] [17] tresnak ingelesez, gaztelaniaz eta euskaraz hizkuntzaren zenbait mailatako 125 metrika[14] aztertzen ditu eta lan honetan hitzak, esaldiak, hitz arraroak (corpus batetik erauzitako lemen maiztasunen zerrenda batean 4ko maiztasuna baino gutxiago duten lemak dituzten hitzak), mendeko perpausak eta perpausak izan ditugu aztergai. Ebaluazio mota horrekin jatorrizko testuaren, LF formatuko testuaren eta ChatGPT-4k sortutako testuaren balioak konpara ditzakegu. Lan honetan, MultiAzterTest erabili dugu, gaztelaniazko testuak aztertzeko metrikak egokiak direlako, baina beste hizkuntza batzuk aztertu nahi badira, Profiling-UD [18] tresnarekin ere egin daiteke analisia.

— **Checklist-etan oinarritutako eskuzko ebaluazioa.** Cumbicus-Pineda *et al.*-ek [19] testu sinplifikatuek bete behar dituzten hainbat ezaugarri zehaztu zituzten, *capability* deitutakoak.

Capability horiek gramatikarekin (adibidez, aditzaren komunztadura zuzena da?), esanahiari eustearekin (adibidez, informazio garrantzitsu guztia mantentzen da?) eta sinpletasunarekin (adibidez, hurrenkera logikoa erabiltzen da?) erlazionatuta daude, besteak beste, eta banan-banan aztertzen da ea betetzen diren ala ez. Testuen sinplifikazio automatikoan, normalean eskuzko ebaluazioan ebaluatzen diren *capability* horietaz gain, aurrebaldintzak (karaktere arrarorik ez izatea) eta etika (estereotiporik ez izatea eta informazio okerrik ez egotea) barneratu dituzte. *Capability*ak interpretatzeko eskalan, % 60ra ez iristea emaitza txarra da, eta % 80tik gora emaitza sendoa da.

Cumbicus-Pineda *et al.*-en [19] lanean proposatutako ezaugarriek testuaren formatua kontuan hartzen ez zutenez, guk beste hiru *capability* proposatu ditugu: i) formatua irakurtzeko errazagoa da, ii) le-

[13] http://ixa2.si.ehu.eus/aztertest/ (atzipen-data: 2024-07-04).

[14] MultiAzterTesten metrikak antzeko lanetan oinarrituta daude eta bere horretan hartu ditugu, ez ditugu guk definitu. MultiAzterTesten artikuluan zein web orrian metrika horien azalpenak daude.

rroen arteko tartea handiagoa da eta iii) sintagmak ez daude lerro ezberdinetan banatuta. Gure lanean LFn eskarmentua duten bi hizkuntzalarik ChatGPT-4k sortutako testu guztiak ebaluatu dituzte eta haien arteko adostasuna kalkulatu dugu. Ebaluazio horren bidez, ChatGPT-4ko testuetan sinplifikazio prozesua nola gauzatu den ikus dezakegu. Gainera, ChatGPT-4 bertsioan informazio guztia mantendu den ala ez ere egiazta dezakegu.

— **Helburu taldearekin egindako ebaluazioa.** % 33tik % 79ra arteko desgaitasuna duten 10 pertsonari eskatu diegu LFn sortutako testua eta ChatGPT-4k sortutako testuak irakurtzeko. Ondoren, esan diegu testuaren edukia berriro kontatzeko (birkontatu, *retell*) eta edukiari buruzko galdera batzuei erantzuteko. Amaitzeko, LFk edo ChatGPT-4k sortutako bertsioen artean aukeratzeko eskatu diegu. Ebaluazioa egiteko, Reichrath eta Moonen-ek [20] egindako lanean inspiratu gara. Erabiltzaileekin ebaluatuz, etorkizunean teknologia horiek erabiliko dituztenek zer nahiago duten jakin dezakegu. 1. taulako adibideari dagozkion galdera batzuk hauek dira: *¿Qué es Lanbide?*, *¿cuántas empresas respondieron a la encuesta?*, eta *¿Qué quería saber Lanbide?*

4. EMAITZAK

Atal honetan, egin ditugun ebaluazioen emaitzak azalduko eta aztertuko ditugu.

4.1. Ezaugarri linguistikoetan oinarritutako ebaluazio automatikoa

MultiAzterTest-ekin egindako ebaluazio automatikoaren bidez, ikusi dugu (2. taula) i) ChatGPT-4k sortutako testuak oro har laburragoak direla *hitzak (guztira)* metrikari erreparatuta, ii) testu guztietan baxuagoa da ChatGPT-4 bertsioko hitz kopurua jatorrizkoarena baino eta, oro har, iii) hitz arraro gutxien dituzte (hitz arraroak neurtzen dituzten metrika gehienetan ChatGPT-4 bertsioak ditu emaitzarik baxuenak eta beti jatorrizko testuarenak baino baxuagoak dira). LFko bertsioek, berriz, esaldi kopuru handiena dute *(esaldiak eta esaldiak (guztira)* metrikak) eta, gainera, esaldi horiek beste bertsioetan baino laburragoak dira *esaldi luzera (batezbestekoa)*. Esaldiak laburragoak izanik, ikus daitekeen bezala, mendeko perpausen indizea eta perpaus/esaldi ratioaren batezbestekoa beti baxuena da LFko bertsioetan.

2. taula. MultiAzterTestekin egindako ebaluazio automatikoaren emaitzak.

Testua	**1 testua**			**2 testua**			**3 testua**			**4 testua**			**5 testua**		
Bertsioa	Jatorrizkoa	LF	ChatGPT	Jatorrizkoa	LF	ChatGPT	Jatorrizkoa	LF	ChatGPT	Jatorrizkoa	LF	ChatGPT	Jatorrizkoa	LF	ChatGPT
Hitzak (guztira)	811,0	425,0	**175,0**	721,0	454,0	**179,0**	867,0	621,0	**237,0**	263,0	199,0	**156,0**	265,0	251,0	**218,0**
Hitz ezberdinak (guztira)	275,0	142,0	**85,0**	306,0	192,0	**94,0**	307,0	205,0	**98,0**	147,0	105,0	**81,0**	159,0	133,0	**110,0**
Hitz ezberdinak (1000 hitzekiko)	339,0875	**334,1176**	485,7143	424,4105	**422,9075**	525,1397	354,0946	**330,1127**	413,5021	558,9354	527,6382	**519,2308**	600,0	529,8805	**504,5872**
Esaldiak (guztira)	21,0	**60,0**	14,0	21,0	**60,0**	20,0	20,0	**88,0**	27,0	8,0	**28,0**	16,0	21,0	**33,0**	24,0
Esaldiak (1.000 hitzekiko)	25,894	**141,1765**	80,0	29,1262	**132,1586**	111,7318	23,0681	**141,7069**	113,9241	30,4183	**140,7035**	102,5641	79,2453	**131,4741**	110,0917
Esaldi luzera (batezbestekoa)	38,619	**7,0833**	12,5	34,3333	**7,5667**	8,95	43,35	**7,0568**	8,7778	32,875	**7,1071**	9,75	12,619	**7,6061**	9,0833
Izen arraroak	21,0	4,0	**0,0**	11,0	3,0	**0,0**	44,0	11,0	**10,0**	3,0	1,0	**0,0**	8,0	5,0	**0,0**
Izen arraroak (1.000 hitzekiko)	25,894	9,4118	**0,0**	15,2566	6,6079	**0,0**	50,7497	**17,7134**	42,1941	11,4068	5,0251	**0,0**	30,1887	19,9203	**0,0**
Adjektibo arraroak (guztira)	30,0	**2,0**	3,0	22,0	8,0	**1,0**	25,0	4,0	**0,0**	5,0	3,0	**0,0**	4,0	3,0	**1,0**
Adjektibo arraroak (1.000 hitzekiko)	36,9914	**4,7059**	17,1429	30,5132	17,6211	**5,5866**	28,8351	6,4412	**0,0**	19,0114	15,0754	**0,0**	15,0943	11,9522	**4,5872**
Aditz arraroak (guztira)	14,0	9,0	**0,0**	21,0	12,0	**5,0**	19,0	17,0	**2,0**	8,0	6,0	**1,0**	9,0	**7,0**	**7,0**
Aditz arraroak (1.000 hitzekiko)	17,2626	21,1765	**0,0**	29,1262	**26,4317**	27,933	21,9146	27,3752	**8,4388**	30,4183	30,1508	**6,4103**	33,9623	**27,8884**	32,1101
Adberbio arraroak (guztira)	1,0	**0,0**	**0,0**	1,0	**0,0**	**0,0**	1,0	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	1,0	**0,0**	**0,0**
Adberbio arraroak (1.000 hitzekiko)	1,233	**0,0**	**0,0**	1,387	**0,0**	**0,0**	1,1534	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	3,7736	**0,0**	**0,0**
Kategoria irekietako hitz arraroak (guztira)	66,0	15,0	**3,0**	55,0	23,0	**6,0**	89,0	32,0	**12,0**	16,0	10,0	**1,0**	22,0	15,0	**8,0**
Kategoria irekietako hitz arraroak (1.000 hitzekiko)	81,381	35,2941	**17,1429**	76,2829	50,6608	**33,5196**	102,6528	51,5298	**50,6329**	60,8365	50,2513	**6,4103**	83,0189	59,761	**36,6972**
Kategoria irekietako hitz arraro ezberdinak (guztira)	53,0	11,0	**3,0**	51,0	21,0	6,0	61,0	25,0	4,0	16,0	10,0	1,0	21,0	13,0	7,0
Kategoria irekietako hitz arraro ezberdinak (1.000 hitzekiko)	65,3514	25,8824	**17,1429**	70,7351	46,2555	**33,5196**	70,3576	40,2576	**16,8776**	60,8365	50,2513	**6,4103**	79,2453	51,7928	**32,1101**
Kategoria irekietako hitz arraroen batezbestekoa	19,5846	8,2873	**3,1579**	20,5224	13,6905	**7,6923**	23,2984	11,8081	**10,7143**	14,9533	12,1951	**1,3514**	17,0543	15,4639	**7,6923**
Kategoria irekietako hitz arraro ezberdinen batezbestekoa	26,5	11,9565	**5,8824**	23,1818	16,5354	**9,8361**	25,8475	17,6056	**7,4074**	17,5824	15,625	**2,0408**	20,0	16,4557	**0,6061**
Mendeko perpausak (guztira)	28,0	**8,0**	16,0	32,0	9,0	**8,0**	39,0	26,0	**17,0**	14,0	**9,0**	15,0	15,0	**12,0**	17,0
Mendeko perpausak (1.000 hitzekiko)	34,5253	**18,8235**	91,4286	44,3828	**19,8238**	50,2793	44,9827	**41,868**	71,73	53,2319	**45,2261**	96,1538	56,6038	**47,8088**	77,9817
Perpaus/esaldi ratioa	3,5238	**1,2833**	2,3571	3,7619	**1,3167**	1,6	5,85	**1,5682**	2,0741	4,875	**1,5357**	2,375	2,2857	**1,6061**	1,9167

Testua	**6 testua**			**7 testua**			**8 testua**			**9 testua**			**10 testua**		
Bertsioa	Jatorrizkoa	LF	ChatGPT	Jatorrizkoa	LF	ChatGPT	Jatorrizkoa	LF	ChatGPT	Jatorrizkoa	LF	ChatGPT	Jatorrizkoa	LF	ChatGPT
Hitzak (guztira)	391,0	**200,0**	201,0	872,0	506,0	**238,0**	1141,0	534,0	**250,0**	547,0	297,0	**157,0**	795,0	440,0	**155,0**
Hitz ezberdinak (guztira)	160,0	101,0	**85,0**	387,0	223,0	**120,0**	425,0	218,0	**126,0**	227,0	143,0	**80,0**	312,0	194,0	80,0
Hitz ezberdinak (1.000 hitzekiko)	**409,2072**	505,0	422,8856	443,8073	**440,7115**	504,2017	**372,4803**	408,2397	504,0	**414,9909**	481,4815	509,5541	**392,4528**	440,9091	516,129
Esaldiak (guztira)	12,0	**29,0**	22,0	30,0	**74,0**	29,0	33,0	**61,0**	23,0	16,0	**38,0**	13,0	23,0	**64,0**	16,0
Esaldiak (1.000 hitzekiko)	30,6905	**145,0**	109,4527	30,9633	**146,2451**	121,8487	28,922	**114,2322**	92,0	29,2505	**127,9461**	82,8025	28,9308	**145,4545**	103,2258
Esaldi luzera (batezbestekoa)	32,5833	**6,8966**	9,1364	32,2963	**6,8378**	8,2069	34,5758	**8,7541**	10,8696	34,1875	**7,8158**	12,0769	34,5652	**6,875**	9,6875
Izen arraroak	10,0	3,0	**0,0**	24,0	8,0	**1,0**	25,0	7,0	**2,0**	14,0	3,0	**0,0**	36,0	13,0	**3,0**
Izen arraroak (1.000 hitzekiko)	25,5754	15,0	**0,0**	27,5229	15,8103	**4,2017**	21,9106	13,1086	**8,0**	25,5941	10,101	**0,0**	45,283	29,5455	**19,3548**
Adjektibo arraroak (guztira)	14,0	4,0	**1,0**	17,0	9,0	**5,0**	55,0	24,0	**12,0**	18,0	7,0	**0,0**	44,0	10,0	**2,0**
Adjektibo arraroak (1.000 hitzekiko)	35,8056	20,0	**4,9751**	19,4954	**17,7866**	21,0084	48,2033	**44,9438**	48,0	32,9068	23,569	**0,0**	55,3459	22,7273	**12,9032**
Aditz arraroak (guztira)	6,0	**1,0**	6,0	22,0	8,0	**5,0**	37,0	10,0	**3,0**	17,0	8,0	**5,0**	12,0	7,0	**2,0**
Aditz arraroak (1.000 hitzekiko)	15,3453	**5,0**	29,8507	25,2294	**15,8103**	21,0084	32,4277	18,7266	**12,0**	31,0786	**26,936**	31,8471	15,0943	15,9091	**12,9032**
Adberbio arraroak (guztira)	3,0	**0,0**	**0,0**	1,0	**0,0**	**0,0**	2,0	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**
Adberbio arraroak (1.000 hitzekiko)	7,6726	**0,0**	**0,0**	1,1468	**0,0**	**0,0**	1,7528	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**	**0,0**
Kategoria irekietako hitz arraroak (guztira)	33,0	8,0	**7,0**	64,0	25,0	**11,0**	119,0	41,0	**17,0**	49,0	18,0	**5,0**	92,0	30,0	**7,0**
Kategoria irekietako hitz arraroak (1.000 hitzekiko)	84,399	40,0	**34,8259**	73,3945	49,4071	**46,2185**	104,2945	76,779	**68,0**	89,5795	60,6061	**31,8471**	115,7233	68,1818	**45,1613**
Kategoria irekietako hitz arraro ezberdinak (guztira)	26,0	8,0	**7,0**	58,0	20,0	**9,0**	83,0	23,0	**9,0**	45,0	18,0	**5,0**	74,0	24,0	**6,0**
Kategoria irekietako hitz arraro ezberdinak (1.000 hitzekiko)	66,4962	40,0	**34,8259**	66,5138	39,5257	**37,8151**	72,7432	43,0712	**36,0**	82,2669	60,6061	**31,8471**	93,0818	54,5455	**38,7097**
Kategoria irekietako hitz arraroen batezbestekoa	17,6471	**8,4211**	8,046	20,6452	12,8866	**10,1852**	23,1969	17,5966	**13,0769**	26,0638	16,6667	**7,9365**	23,4694	13,6986	**8,8608**
Kategoria irekietako hitz arraro ezberdinen batezbestekoa	21,4876	**12,1212**	13,4615	23,4818	15,1515	**12,5**	26,0188	15,6463	**11,1111**	28,125	20,0	**11,1111**	29,8387	16,6667	**11,7647**
Mendeko perpausak (guztira)	8,0	**4,0**	18,0	45,0	17,0	**12,0**	66,0	21,0	**17,0**	22,0	**8,0**	11,0	32,0	12,0	**8,0**
Mendeko perpausak (1000 hitzekiko)	20,4604	**20,0**	89,5522	51,6055	**33,5968**	50,4202	57,844	**39,3258**	68,0	40,2194	**26,936**	70,0637	40,2516	**27,2727**	51,6129
Perpaus/esaldi ratioa	4,3333	**1,6552**	2,0455	4,4074	**1,473**	1,6897	4,7576	**1,623**	2,0435	3,75	**1,4474**	2,1538	5,1739	**1,6562**	2,125

Ebaluazio horretan oinarrituta esan dezakegu ChatGPT-4k sinplifikazio lexikal ona egiten duela eta testuen laburpen handiak egiten dituela. Hala ere, esaldiak LFko bertsioetan baino luzeagoak dira eta mendeko perpaus gehiago dituzte, eta bi horiek irakurketa eta ulermena zail ditzakete.

4.2. **Checklist-etan oinarritutako eskuzko ebaluazioa**

*Capability*ei dagokienez (3 taula), ChatGPT-4k egokitutako testuak gramatika aldetik ondo idatzita daude (% 89ko puntuazioa batez beste), sinpleagoak dira (% 81,11ko puntuazioa batez beste), baina edukia laburtuegi dago eta informazio garrantzitsu asko galtzen da (% 1,7ko puntuazioa batez beste). Gainera, formatuari dagokionez, sintagmak lerro ezberdinetan banatuta daude (% 33,33ko puntuazioa batez beste), eta horrek irakurketa zailtzen du.

3. taula. *Capability*etan oinarritutako checklist ebaluazioaren emaitzak *capability* motaren arabera.

Mota/testua	1 testua	2 testua	3 testua	4 testua	5 testua	6 testua	7 testua	8 testua	9 testua	10 testua	Batez-bestekoa
Aurrebaldintzak	66,67	66,67	66,67	66,67	66,67	33,33	66,67	66,67	66,67	66,67	63,33
Gramatika	95,00	75,00	80,00	85,00	90,00	90,00	90,00	85,00	100,00	100,00	89,00
Esanahia	0,00	0,00	0,00	0,00	0,00	0,00	0,00	16,67	0,00	0,00	1,67
Sinpletasuna	77,78	88,89	72,22	88,89	88,89	66,67	94,44	83,33	83,33	66,67	81,11
Etika	50,00	50,00	50,00	50,00	50,00	25,00	50,00	50,00	50,00	50,00	47,50
Formatua	33,33	33,33	33,33	33,33	33,33	33,33	33,33	33,33	33,33	33,33	33,33

Ebaluatzaile bakoitzak emandako puntuazioa eta batez besteko puntuazioa 4 taulan ikus daiteke. Ebaluazio hori irakurketa errazean adituak diren bi hizkuntzalarik egin dute, eta haien arteko adostasuna (Cohen-en kappa) 0,85ekoa izan da, adostasun altua adierazten duena [21]. Emaitzei dagokienez, testu guztiek, 6 testua izan ezik, *capability*etan ezartzen den % 60ko langa gainditzen dute (hortik behera emaitza txarra da), baina testu bat bera ere ez da iristen emaitza sendoa izatera, hau da, % 80ko puntuaziora.

4. taula. *Capability*etan oinarritutako *checklist* ebaluazioaren emaitzak ebaluatzailearen arabera.

Testua	1 testua	2 testua	3 testua	4 testua	5 testua	6 testua	7 testua	8 testua	9 testua	10 testua	Batez-bestekoa
1 ebaluatzailea	70,00	66,67	63,33	70,00	70,00	66,67	70,00	66,67	73,33	70,00	68,67
2 ebaluatzailea	66,67	63,33	60,00	66,67	70,00	50,00	73,33	66,67	70,00	63,33	65,00
Batezbestekoa	68,33	65,00	61,67	68,33	70,00	58,33	71,57	66,67	71,67	66,67	66,83
Adostasuna	0,77	0,93	0,93	0,92	1,00	0,53	0,92	0,70	0,92	0,85	0,85

4.3. **Helburu taldearekin egindako ebaluazioa**

Amaitzeko, helburu taldearekin egindako ebaluazioari dagokionez, 5. taulan ikus daitekeen bezala, lau parte-hartzailek ChatGPT-4 eta LFko bertsioak birkontatu eta galderei erantzun diete. Beste parte-hartzaile batek LFko bertsioa birkontatu du, baina ez ChatGPT-4 bertsioa; eta ez da galderei erantzuteko gai izan. Parte-hartzaile bat gai izan da testuei buruzko galderei erantzuteko, baina ez birkontatzeko. Azkenik, lau parte-hartzailek ezin izan dute irakurri zutena kontatu, ezta galderei erantzun ere.

Nahiago duten bertsioaren galderari erantzunez, helburu taldeko erabiltzaile guztiek LFko bertsioa nahiago izan dute, eta arrazoi hauek eman dituzte: i) hitz gehiago ulertzen dituzte; ii) lerroen arteko tartea handiagoa da; iii) esaldiak laburragoak dira; iv) hobeto azalduta dago; eta, v) eduki gehiago eta azalpen gehiago ditu.

5. **ONDORIOAK ETA ERRONKAK**

Atal honetan gure ebaluazioaren emaitzak sakonago aztertuko ditugu eta izan ditugun erronkak aipatuko ditugu.

5. taula. Erabiltzaileekin egindako ebaluazioaren emaitzak.

Ebaluatzailea	1	2	3	4	5	6	7	8	9	10
ChatGPTko bertsioa birkontatzeko gai?	bai	ez	ez	ez	ez	bai	ez	bai	bai	ez
LFko bertsioa birkontatzeko gai?	bai	ez	ez	bai	ez	bai	ez	bai	bai	ez
ChatGPTko bertsioan galderei erantzuteko gai?	bai	bai	ez	ez	ez	bai	ez	bai	bai	ez
LFko bertsioan galderei erantzuteko gai?	bai	bai	ez	ez	ez	bai	ez	bai	bai	ez
Nahiago duen bertsioa	LF	LF	LF	LF	LF	LF	LF	LF	LF	LF

ChatGPT-4ren ebaluazioari dagokionez, gure emaitzen arabera, ChatGPT-4k sinplifikazio lexikala ondo egiten du, ebaluazio automatikoan ikusi baita hitz arraroen presentzia baxua dela. Adibidez, 1 taulako testu zatian jatorrizkoan dagoen *nuevo personal* sintagma (*personal* hitz konplexuagotzat hartuta) *más gente* sintagmarekin ordezkatu du. Baina beste ataza batzuk ez ditu ondo egin:

— ChatGPT-4 ez da gai adibideak edo azalpenak gehitzeko, eta hori LFen garrantzitsua da; helburu taldearen iruzkinetan ikusi dugun bezala, oso baliagarriak zaizkie.
— Informazio garrantzitsu guztia ez du mantentzen, ChatGPT-4k asko laburtzen baititu edukiak.
— Sintagmak lerro ezberdinetan banatzen ditu, eta horrek irakurmena zailtzen du.

Bestetik, ChatGPT-4 ez-determinista denez (sarrera-testu bera eta *prompt* bera emanda, irteeratestua edo emaitza ezberdina izan daiteke), zaila da emandako arauen arabera irteera-testu mota jakin bat izatea: ChatGPT-4k testu batzuetan paragrafo ezberdinetan eman du informazioa, eta, beste batzuetan, zerrendatzeak erabili ditu. Gainera, emaitzak erreplikatzea ezinezkoa da interfazea erabilita.

Prompt egoki bat diseinatzea ere ez da erraza. Aukeratu dugun *prompt*ak emaitza konsistenteak eman dizkigu hasierako esperimentuetan, baina, esperimentua egitean bariazio handia egon da testuen formatuan. Esperimentuak APIaren bidez eginez gero, testuak egokitu litezke pre- eta postedizioaren bidez, eta horrek emaitzen konsistentzia gehiago ekarriko luke.

Beraz, testu-egokitzaile profesionalek ziurtatu beharko lukete jatorrizko bertsioan dagoen informazio guztia egokitutako testuan jasota dagoela eta IEko eta, kasu honetan, LFko arau guztiak errespetatu direla. Izan ere, lehen begiratu batean badirudi sistemaren emaitza ona dela, baina, testuak ondo analizatuz gero, hutsuneak eta erroreak daude. Gure emaitzak, gainera, alemanerako egindako azterketarekin [9] bat datoz.

Lan honetan erabili ditugun ebaluazio metodoei dagokienez, BLEU, ROUGE eta SARI metrika automatikoek IEko testuak ebaluatzeko dituzten mugak direla eta, metodo berriak aukeratu ditugu ebaluazioa egiteko. *MultiAzterTest* tresna erabilgarria izan zaigu ezaugarri linguistiko jakin batzuk analizatzeko. *Checklist*ean oinarritutako ebaluazioan ikusi dugu testua bere osotasunean eta testuaren formatua ebaluatzen duten *capability*ak eta IEko *capability*ak falta zirela. Hori dela eta, hiru *capability* berri proposatu ditugu. Azkenik, helburu taldearekin ebaluazioa egiteak berebiziko garrantzia du, haiek izango baitira etorkizuneko erabiltzaileak eta haien iritzia kontuan hartu behar da.

6. AMAIERA ETA ETORKIZUNEKO LANAK

Lan honetan, ChatGPT-4k IEra testuak nola egokitzen dituen aztertu dugu. Zehazki, gaztelaniazko LF izan dugu aztergai, eta 10 testu ebaluatu ditugu, automatikoki, eskuz eta helburu taldeko erabiltzaileekin jatorrizko bertsioan, LFko bertsioan eta ChatGPT-4k sortutako bertsioan.

Gure emaitzen arabera, ChatGPT-4 erabilgarria izan daiteke sinplifikazio lexikalerako, baina ez da gai azalpenak/adibideak emateko, ezta LFko arauei bete-betean jarraitzeko ere. Gainera, informazioa kentzeko joera handia du. Horretaz gain, *prompt* egokia aukeratzeko zailtasunak eta irteera-testuen aldakortasunak adierazten digute tresna hori ezin dela bere horretan erabili eta harekin sortutako testuak editatu behar direla.

Lan honetan, batez ere, erabiltzaileen eta egokitzaileen ikuspegia hartu dugu kontuan: behar-beharrezkoa da teknologia berriak ebaluatzea eta haien hutsuneak ezagutaraztea. Jende arruntak ChatGPT eta halako tresnak erabiltzen ditu haien funtzionamendua ulertu gabe edo emaitzak kritikoki ebaluatzeko gaitasunik gabe. Horregatik, garrantzitsua da esperimentuak haien ikuspuntutik egitea; izan ere, kasu askotan, haiek dira azken erabiltzaileak.

Etorkizuneko lanetan, antzeko esperimentuak egin nahi ditugu alemanez eta euskaraz, eta beste tresna batzuekin, ChatGPT API, Bloom [22], LLaMa[15] edo Latxa [23], esate baterako. Testu gehiago ere analizatu nahi ditugu, gure emaitzak orokor daitezkeen ala ez ikusteko eta formatu-aldaketak aztertu nahi ditugu. Horretaz gain, LFko arauak betetzen diren ala ez jakiteko ebaluazio metodo estandar eta automatikoago bat proposatu nahi dugu.

ESKER ONAK

Bihotz-bihotzez eskerrak eman nahi dizkiegu ikerketa honetan parte hartu zuten Atzegiko boluntarioei. Horretaz gain, lan honek proiektu hauen laguntza izan du: i) IXA taldea A motako ikertalde finkatua (IT-1805-22), Eusko Jaurlaritzak finatziatua; ii) DeepKnowledge (PID2021-127777OB-C21) MCIN/AEI/10.13039/501100011033 eta FEDER-ek finantzatuta; eta iii) AWARE Commonsense for a new generation of natural language understanding applications (TED2021-131617B-I00) MCIN/AEI-k eta Europar Batasunak NextGenerationEU/ PRTR-ek finantziatua.

BIBLIOGRAFIA

[1] OECD. 2013. *OECD Skills Outlook 2013*. URL: https://www.oecd-ilibrary.org/content/publication/9789264204256-en

[2] IMSERSO. 2022. *Base Estatal de Datos de Personas con Valoración del Grado de Discapacidad*, Ministerio de Derechos Sociales, Gobierno de España.

[3] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat eta kolaboratzaileak. 2023. «GPT-4 Technical Report», *arXiv preprint arXiv:230308774*.

[15] Introducing LLaMa https://ai.facebook.com/blog/large-language-model-llama-meta-ai/ (atzipen-data: 2023-10-12).

[4] J. C. CHOW, L. SANDERS eta K. LI. 2023. «Impact of ChatGPT on Medical Chatbots as a Disruptive Technology», *Frontiers in artificial Intelligence*, **6**, 1166014.

[5] F. J. GARCÍA-PEÑALVO. 2023. «La percepción de la inteligencia artificial en contextos educativos tras el lanzamiento de chatgpt: disrupción o pánico», *Education in the Knowledge Society (EKS)*, **24**, e31279. URL: https://revistas.usal.es/tres/index.php/eks/article/view/31279

[6] ASOCIACIÓN ESPAÑOLA DE NORMALIZACIÓN (UNE). 2018. «UNE 153101:2018 EX Lectura fácil. Pautas y recomendaciones para la elaboración de documentos». URL: https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0060036

[7] Ó. G. MUÑOZ eta A. MATAMALA. 2021. «Easy language in spain», *Handbook of Easy Languages in Europe*, Orria 493.

[8] M. MADINA, I. GONZALEZ-DIOS eta M. SIEGEL. 2024. «Languagetool as a cat tool for easy-to-read in spanish», *Proceedings of the 3rd Workshop on Tools and Resources for People with REAding DIfficulties (READI)@ LREC-COLING 2024*, Orrialdeak 93-101.

[9] S. DEILEN, S. H. GARRIDO, E. LAPSHINOVA-KOLTUNSKI eta C. MAASS. 2023. «Using ChatGPT as a CAT tool in Easy Language translation», *Proceedings of the Second Workshop on Text Simplification, Accessibility and Readability*, Orrialdeak 1-10.

[10] I. ESPINOSA-ZARAGOZA, J. ABREU-SALAS, P. M. POZO eta M. PALOMAR. 2023. «Automatic Text Simplification for People with Cognitive Disabilities: Resource Creation within the ClearText Project», *Proceedings of the Second Workshop on Text Simplification, Accessibility and Readability*, Orrialdeak 68-77.

[11] K. PAPINENI, S. ROUKOS, T. WARD eta W.-J. ZHU, 2002, «Bleu: a Method for Automatic Evaluation of Machine Translation», *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, Orrialdeak 311-318.

[12] W. XU, C. ÑAPOLES, E. PAVLICK, Q. CHEN eta C. CALLISON-BURCH. 2016. «Optimizing Statistical Machine Translation for Text Simplification», *Transactions of the Association for Computational Linguistics*, **4**, 401-415.

[13] C.-Y. LIN, 2004, «Rouge: A Package for Automatic Evaluation of Summaries», *Text summarization branches out*, Orrialdeak 74-81.

[14] E. SULEM, O. ABEND eta A. RAPPOPORT. 2018. «BLEU is Not Suitable for the Evaluation of Text Simplification», *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Orrialdeak 738-744.

[15] F. ALVA-MANCHEGO, C. SCARTON eta L. SPECIA. 2021. «The (Un)Suitability of Automatic Evaluation Metrics for Text Simplification», *Computational Linguistics*, **47**(4), 861-889. URL https://doi.org/10.1162/coli_a_00418

[16] N. GRABAR eta H. SAGGION, 2022, «Evaluation of Automatic Text Simplification: Where are we Now, Where should we Go from Here», *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1: conférence principale*, Orrialdeak 453-463.

[17] K. Bengoetxea eta I. Gonzalez-Dios. 2021. «Multiaztertest: A Multilingual Analyzer on Multiple Levels of Language for Readability Assessment», *arXiv preprint arXiv:210904870*.

[18] D. Brunato, A. Cimino, F. dell'Orletta, G. Venturi eta S. Montemagni. 2020. «Profiling-UD: a Tool for Linguistic Profiling of Texts», *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Orrialdeak 7145-7151.

[19] O. M. Cumbicus-Pineda, I. Gonzalez-Dios eta A. Soroa. 2021. «Linguistic Capabilities for a Checklist-based evaluation in Automatic Text Simplification», *Proceedings of the First Workshop on Current Trends in Text Simplification (CTTS 2021), co-located with SEPLN 2021*, Orrialdeak 70-83.

[20] E. Reichrath eta X. Moonen. 2022. «Assessing the effects of Language for all», *Nordic Journal of Linguistics*, **45**(2), 232-248.

[21] R. Artstein eta M. Poesio, 2008, «Inter-coder Agreement for Computational Linguistics», *Computational linguistics*, **34**(4), 555-596.

[22] T. L. Scao, A. Fan, C. Akiki, E. Pavlick, S. Ilić, D. Hesslow, R. Castagné, A. S. Luccioni, F. Yvon, M. Gallé eta kolaboratzaileak. 2022. «Bloom: A 176b-parameter Open-access Multilingual Language Model», *arXiv preprint arXiv:221105100*.

[23] J. Etxaniz, O. Sainz, N. Miguel, I. Aldabe, G. Rigau, E. Agirre, A. Ormazabal, M. Artetxe eta A. Soroa, 2024, «Latxa: An open language model and evaluation suite for Basque», L.-W. Ku, A. Martins eta V. Srikumar, editoreak, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Orrialdeak 14952-14972, Association for Computational Linguistics, Bangkok, Thailand. URL https://aclanthology.org/2024.acl-long.799

A. **ERANSKINA: ITZULPENAK**

Atal honetan gaztelaniazko testuen euskarazko itzulpenak emango ditugu.

A1. ***Prompt*aren itzulpena**

Jarraian, 3. atalean emandako *prompt*aren itzulpena emango dugu:

— **Prompt** *Las normas de redacción de Lectura Fácil Según la Norma UNE 153101:2018 EX Lectura Fácil son las siguientes: (lista de normas)* → UNE 153101:2018 EX Lectura Fácil arauan zehaztuta dauden irakurketa errazeko arauak hauek dira: (arauen zerrenda)

— **Output** ...

— **Prompt** *Adapta esta noticia a Lectura Fácil siguiendo las normas de redacción de Lectura Fácil Según la Norma UNE 153101:2018 EX Lectura Fácil: (noticia)* → UNE 153101:2018 EX Lectura Fácil arauaren arabera egokitu irakurketa errazera albiste hau: (albistea).

A2. **LF arauen itzulpena**

Atal honetan, adibide gisa eman ditugun LFko arauen itzulpena emango dugu:

— **Ortografikoa.** Ez da puntu puntu eta koma erabili behar → Ez erabili puntu eta koma.
— **Hiztegikoak eta fraseologikoak.** *-mente* atzizkiarekin bukatzen diren adberbioak saihestu behar dira. → Ekidin *-mente* atzizkiarekin bukatzen diren adberbioak.
— **Perpaus eta esaldi mailakoa.** Ahots pasiboa saihestu behar da. → Saihestu ahots pasiboa.
— **Estilokoa eta testu antolakuntzakoa.** Irakurleak testua ulertzeko inferentziak egin behar izatea saihestu beharko litzateke. → Saihestu irakurleak testua ulertzeko inferentziak egin behar izatea.
— **Dokumentuaren aurkezpenekoa.** Lerroaren amaieran hitzak marratxoekin zatikatzea saihestu behar da. → Saihestu lerroaren amaieran hitzak marratxoekin zatikatzea.

A3. **1 taulako adibidearen itzulpena**

6 taulan 1 taulako adibidearen itzulpena eman dugu. LF eta ChatGPT-4 bertsioetako itzulpenetan lerro-jauziak jatorrizkoari eta euskal gramatikari egokituta ematen ahalegindu gara.

A4. **Galderen itzulpenak**

Erabiltzaile taldeari egindako galderen itzulpenak emango ditugu.

— *¿Qué es Lanbide?* → Zer da Lanbide?
— *¿Cuántas empresas respondieron a la encuesta?* → Zenbat enpresak erantzun zioten galdetegiari?
— *¿Qué quería saber Lanbide?* → Zer jakin nahi zuen Lanbidek?

6. taula. 1 taulako testu zatien euskarazko itzulpena.

Jatorrizko testua	LFko bertsioa	ChatGPT-4 bertsioa
El 65% por ciento de las empresas consultadas por el Servicio Público Vasco de Empleo Lanbide tienen previsto contratar nuevo personal en los próximos dos años en sectores tan diversos como el suministro de energía y agua, la gestión de residuos, la industria manufacturera, las actividades sanitarias y de servicios sociales o la hostelería. La encuesta apunta, además, que siete de cada diez han tenido dificultades en los últimos cinco ejercicios para encontrar perfiles adecuados a sus necesidades.	Lanbide lanzó a finales del año pasado una encuesta / a más de 10.000 empresas vascas de todo tipo de sectores: / energía y agua, gestión de residuos, / industria, hostelería, servicios sociales, sanidad... / A la encuesta han respondido unas 600 empresas. / La encuesta se hizo para saber sobre sus necesidades / en cuanto a formación de sus trabajadores, / problemas y planes de contratación en el futuro.	**Las empresas en el País Vasco buscan más trabajadores** / —Muchas empresas en el País Vasco quieren contratar más gente en los próximos dos años. / —Estas empresas trabajan en áreas diferentes: energía, agua, residuos, salud, servicios, hostelería y otros. / —Pero muchas empresas tienen problemas para encontrar a los trabajadores adecuados. / **¿Qué hizo Lanbide?** / —Lanbide es un servicio de empleo en el País Vasco. / —Lanbide envió una encuesta a muchas empresas. Querían saber qué tipo de trabajadores necesitan. / —600 empresas respondieron a la encuesta.
Lanbide Euskal Enplegu Zerbitzu Publikoak kontsultatutako enpresen % 65ek datozen bi urteetan langile berriak kontratatzea aurreikusi dute, hainbat sektoretan: energia eta ur-horniduran, hondakinen kudeaketan, manufaktura-industrian, osasun eta gizarte-zerbitzuetako jardueratan edo ostalaritzan. Inkestaren arabera, gainera, hamarretik zazpik zailtasunak izan dituzte azken bost ekitaldietan beren beharretara egokitutako profilak aurkitzeko.	Iaz, Lanbidek inkesta bat egin zien / era guztietako 10.000 euskal enpresari baino gehiagori: / energia eta ura, hondakinen kudeaketa, / industria, ostalaritza, gizarte-zerbitzuak, osasuna... / Inkestari 600 bat enpresak erantzun diote. / Inkesta egin zen beraien premiak ezagutzeko / langileen prestakuntzari, / arazoei eta etorkizuneko kontratazio-planei dagokienez.	**Euskal Autonomia Erkidegoko enpresak langile gehiago bilatzen ari dira** / Euskal Herriko enpresa askok jende gehiago kontratatu nahi dituzten datozen bi urteetan. / —Enpresa horiek hainbat arlotan lan egiten dute: energia, ura, hondakinak, osasuna, zerbitzuak, ostalaritza eta beste batzuk. / —Baina enpresa askok arazoak dituzte langile egokiak aurkitzeko. / **Zer egin zuen Lanbidek?** / —Lanbide enplegu-zerbitzu bat da Euskal Herrian. / —Lanbidek inkesta bat bidali zien enpresa askori. Nolako langileak behar dituzten jakin nahi zuten. / —600 enpresek erantzun zuten galdetegia.

B. **ALDE BATERA UTZITAKO PROMPTAK**

Atal honetan, probatutako eta alde batera utzitako *prompt*ak emango ditugu.

1. *Redacta esta noticia en Lectura Fácil: (link a la noticia) Para ello, deberás respetar las normas de Lectura Fácil que aparecen aquí: (link a las normas de Inclusion Europe).*
2. *Redacta esta noticia en Lectura Fácil: (link) Para ello, deberás seguir las normas indicadas en Información para todos: Las reglas europeas para hacer información fácil de leer y comprender de aquí: (link a la web donde se pueden encontrar las normas de Inclusion Europe en pdf, pero no al pdf en sí).*
3. *Aquí tienes un mensaje en Lectura Fácil que puedes utilizar como referencia: (link) ¿Puedes crear un texto en Lectura Fácil basado en esta noticia? (link).*
4. *¿Sabes lo que es la Lectura Fácil? ... Redacta esta noticia en Lectura Fácil: (link).*
5. *Redacta esta noticia en Lectura Fácil: (link) Para ello, deberás seguir las normas para hacer información fácil de leer y comprender de aquí: (link a una página web que creamos en la que escribimos las normas de Lectura Fácil).*
6. *Las normas de redacción de Lectura Fácil son estas: (link a una página web que creamos en la que escribimos las normas de Lectura Fácil) Redacta esta noticia en Lectura Fácil siguiendo las normas que te he dado: (link).*
7. *Redacta esta noticia siguiendo las normas de Lectura Fácil europeas para hacer información fácil de leer y comprender: (link).*
8. *Tienes que redactar la siguiente noticia en Lectura Fácil: (link).*
9. *Esta es la versión original de una noticia: (link a la noticia original) Esta es la misma noticia redactada en Lectura Fácil: (link a la misma noticia en versión Lectura Fácil) Basándote en este ejemplo, redacta esta noticia en Lectura Fácil: (link a otra noticia en versión original)*
10. *Las normas de redacción de Lectura Fácil Según la Norma UNE 153101:2018 EX Lectura Fácil son las siguientes: (lista de normas) Redacta esta noticia en Lectura Fácil siguiendo las normas de redacción de Lectura Fácil Según la Norma UNE 153101:2018 EX Lectura Fácil: (texto original).*