Adimen Artifiziala eta euskara.
Non gaude? Zer dator?

Xabier Arregi Iparragirre

Universidad del País Vasco / Euskal Herriko Unibertsitatea
HiTZ Hizkuntza Teknologiako Zentroa / Basque Center for Language Technology
ixa

Adimen Artifiziala

- **Makinei gizakia *imitatzeko* gaitasuna ematea** helburu duen jakintza-arloa

  Giza adimena makinen esku jartzea

  Makinek giza adimenari dagozkion portaerak **errepikatzea (simulatzea, emulatzea)**

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa

Hainbat modutara adierazi izan da

- Jokoak: xakea (*DeepBlue, 1996*)
- Sistema prediktoreak edo igarleak
- Gaixotasunen diagnostikoak
- Zentzumenak: ikusmen artifiziala
- Robotika eta sistema autonomoak
- Galderei erantzuteko sistemak (*Watson*: Jeopardy, 2011)

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

Hainbat modutara adierazi izan da

- Jokoak: xakea (*DeepBlue, 1996*)
- Sistema prediktoreak edo igarleak
- Gaixotasunen diagnostikoak
- Zentzumenak: ikusmen artifiziala
- Robotika eta sistema autonomoak
- Galderei erantzuteko sistemak (*Watson*: Jeopardy, 2011)
- Eta **hizkuntza**

Hizkuntza eta teknologia

- Hizkuntzaren ulermena eta mintzamena giza adimenaren ezaugarri behinenak dira
- Hala berean dago hizkuntza Adimen Artifizialaren muinean

Universidad del País Vasco Euskal Herriko Unibertsitatea
HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology
ixa

Hizkuntza eta teknologia

- Hizkuntzaren ulermena eta mintzamena giza adimenaren ezaugarri behinenak dira
- Hala berean dago hizkuntza Adimen Artifizialaren muinean

Makinak hiztun

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa

Hizkuntza-teknologien bilakaera

- 50eko hamarkada: hastapenak
- 70eko hamarkada: erregelatan oinarrituta, logika
- 90eko hamarkada: estatistika, ikasketa automatikoa
- 2010eko hamarkada: **eredu neuronala, ikasketa sakona**

Hizkuntza-teknologien bilakaera

Lehen:

L3. X0 sar = ak X0 kat = knmdek X0 kas = abs X0 kom num = hk X0 kom mug = m	L4. X0 sar = ek X0 kat = knmdek X0 kas = erg X0 kom num = hk X0 kom mug = m
L5. X0 sar = gizon X0 kat = ize X0 azp = arr	L6. X0 sar = txakur X0 kat = ize X0 azp = arr

2. taula. Euskararen lexikoi baten adibidea PATR formalismoan.

Hizkuntza-teknologien bilakaera

Lehen:

is1 →		*ize adj*	*etxe EDER*
	|	*ize*	*etxe*
is2 →		*det is1*	*ZENBAIT etxe eder*
	|	*is1 det*	*etxe eder BAT*
	|	*is1*	*etxe*
	|	*izb*	*JON*

Universidad del País Vasco · Euskal Herriko Unibertsitatea · HiTZ Hizkuntza Teknologiako Zentroa · Basque Center for Language Technology · ixa

Hizkuntza-teknologien bilakaera

Orain:

```
mBERT: Et #xer #ant #z
ours:  Etxera #ntz

mBERT: Medi #kua #rene #ra
ours:  Mediku #aren #era
```

More specifically, mBERT's subwords tend to be shorter and less interpretable, while our subwords are closer to linguistically interpretable strings, like mediku (doctor) aren ('s) and era (to the).

Hizkuntza-teknologien bilakaera

- **Lehen, helduek bezala** ikasten saiatzen ginen
- **Orain, umeek bezala (edo)**

Azken urteetan, sekulako hobekuntza

Lau faktoreren konfluentziagatik:

1) **Sare neuronalen** teknologia: **ikasketa sakona**
2) Datu masiboak (testuak, ahotsa) eskuragarri
3) Konputazio-ahalmena (GPUak) izugarri handitzea
4) Autoikasketarako hurbilpen sinpleak eta eraginkorrak

Orain, hizkuntza-eredu neuronal sortzaileak

- Hizkuntza-ereduek hizkuntzaren jokabidea *gordetzen* duten eredu matematikoak dira
- **Corpus izugarri handietan** entrenatuz osatzen dira
  - Hurrena zer hitz datorkeen igartzen dute
- Hizkuntza-eredu orokor horiek hainbat eginkizun burutzera orienta daitezke

Hizkuntza-teknologiak (HTak) benetan erabilgarri

- Adibide garbiena, itzulpen automatikoa
- Nonahi, noiznahi erabiltzen ditugu
- Ez bakarrik konputagailu, tablet edo mugikorretan,
  - ez bada autoarekin, hozkailuarekin, telebistarekin, laguntzaile birtualekin, eta abarrekin ari garela

Hizkuntza-teknologiak (HTak) benetan erabilgarri

- Adibide garbiena, itzulpen automatikoa
- Nonahi, noiznahi erabiltzen ditugu
- Ez bakarrik konputagailu, tablet edo mugikorretan,
  - ez bada autoarekin, hozkailuarekin, telebistarekin, laguntzaile birtualekin, eta abarrekin ari garela

Gailu horiek denak, mintzaide

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa

Solaskide horiek

- Ahobizi al dira? Belarriprest ote?
- Bizi al gaitezke euskaraz mintzaide elektroniko horiek euskaraz ikasten ez badute?

https://drive.google.com/file/d/1Lam3XY-Nx5azRnZgO_f4rPhPf-MtjALL/view?usp=sharing

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa

Hizkuntza-teknologiak gaur: aplikazioak (euskaraz?)

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa

Itzulpen automatikoa

- Kalitate oso oneko itzultzaile automatikoak sortu dira
- Euskaratik edo euskarara itzultzen dutenak ere, bikainak. Tartean, gurean sortuak:

elia elhuyar

itzuli itzultzaile neuronala

- Etorkizuna
  - Aldibereko itzulpena
  - Ahotsetik ahotserako itzulpena:
    - https://huggingface.co/spaces/facebook/seamless_m4t

Ahotsaren analisia eta sintesia

- Analisia: speech-to-text, Automatic Speech Recognition (ASR)
- Sintesia: text-to-speech (TTS)
- Izugarri aurreratu da, eta baldintza onetan errendimendu altua lortzen da. Adibidez:
  - https://elevenlabs.io/
  - https://huggingface.co/spaces/openai/whisper

Ahotsa eta bideoa, batera

- Heygen: https://app.heygen.com/home
- Rask: https://www.rask.ai/

Ahotsaren analisia eta sintesia, euskaraz

- Euskararako ere garatzen ari dira hizketa-ezagutzaileak eta sintetizatzaileak, Euskal Herrian bertan

ahoLAB https://aholab.ehu.eus/ahomytts/?q=ahots-bankua&language=eu

aditu elhuyar

https://aditu.eus/

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

Ahotsaren analisia eta sintesia, euskaraz

BATU DITZAGUN GURE AHOTSAK, TEKNOLOGIARI EUSKARAZ IRAKASTEKO

Gaitu proiektua

Adimen artifizialak euskaraz egin dezan

Erronka berri bat dugu: makinek gure hizkuntza ulertu eta hitz egin dezaten, lehenago irakatsi egin behar diegu.

Horretarako, milaka ahots behar ditugu, milaka grabazio ordu euskaraz.

Zure ahotsa garrantzitsua da

Ahots asko eta askotarikoak behar ditugu: adin tarte, genero eta mota guztietakoak. Eta zurea ere bai.

Guztion artean sortzen ari garen ahotsen datu-basea, gero enpresa guztiek erabili ahal izango dute euskarazko tresna eta aplikazio berriak sortzeko, euskaldun guztion onurarako.

Horretarako, sartu eta grabatu, mesedez, ahalik eta esaldi gehien . Oso erraza da.

Ekarpen polita egingo diogu gure hizkuntzaren eta kulturaren garapenari.

EMAN ZURE AHOTSA

PARTE HARTU ERRONKAN

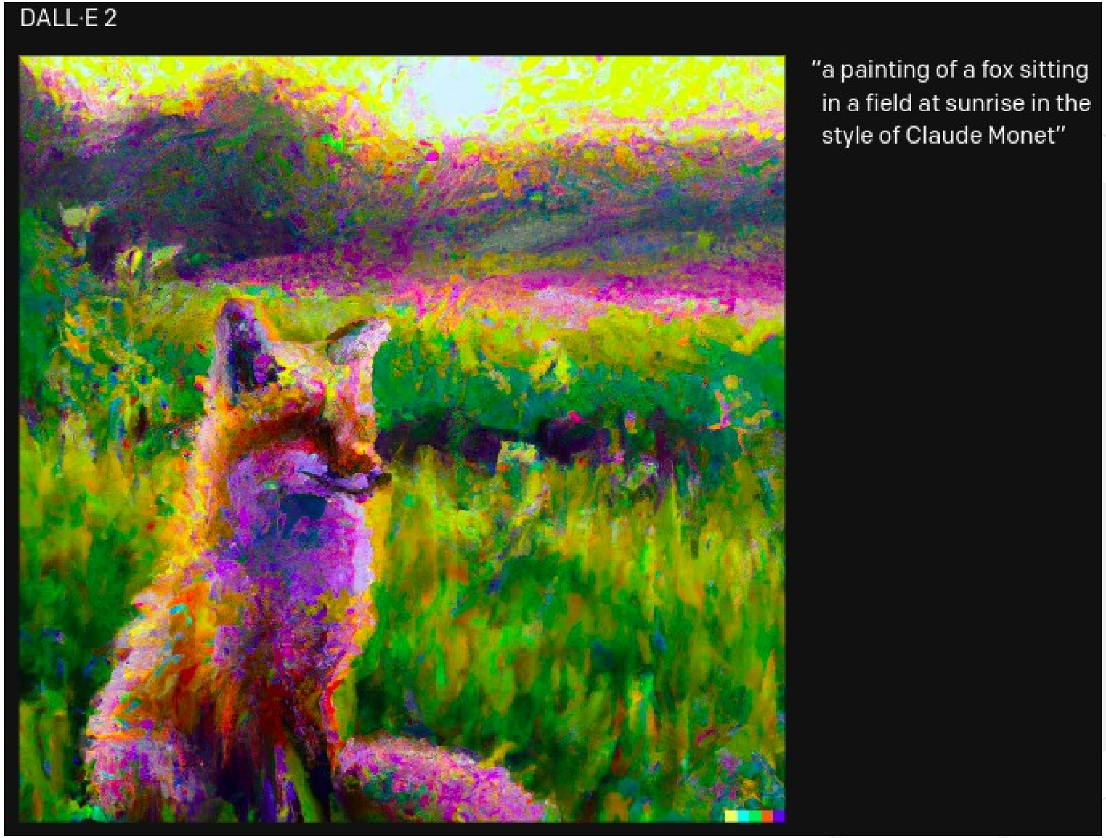
eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

ixa

Hizkuntza eta... irudiak, bideoa, musika

https://stablediffusionweb.com/#demo



Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa

https://huggingface.co/spaces/dalle-mini/dalle-mini

Jorge Oteiza thinking about a room full of stuff

Run

Elkarrizketa-sistemak

- Eskaera edo *prompt* bidezkoak.
- Sistema hauek gai dira elkarrizketa bati eusteko.
- Horretarako:

  a) Eskaera ulertu

  b) Erantzuna sortu

  c) Erabiltzaileek emandako feedback-etik ikasi egiten da: errefortzu bidezko ikasketa

ChatGPT eta GPT4

- Chat **Generative** Pre-training Transformer-en belaunaldi berria
  - Sortzaileak izateaz gain, elkarrizketari eusteko gai
- Osagaiak:
  - Hizkuntza-eredu oinarrizkoa
  - *InstructGPT* eredua: erabiltzailearen eskaerei ondo erantzuteko
    - Giza elkarrekintzaren bidez entrenatzen dira
- OpenAi enpresa aitzindaria

ChatGPT eta GPT4: zertan erabili

- Programa informatikoak kodetu.
- Musika konposatu.
- Mezu elektronikoen zirriborroa osatu.
- Artikuluak, podcast-ak edo aurkezpenak osatu.
- Matematikako problemak ebatzi.
- Bilaketa-motorrak optimizatzeko hitz gakoak eskaini.
- Produktuen deskribapenak egin.
- Lanpostua bilatzen lagundu: curriculumak eta aurkezpen-gutunak.
- Itzulpenak egin.
- ...

chatGPT

XA

Escribe un artículo crítico explicando porqué Aginaga (Usurbil) debe disponer de una entidad local menor y no ser solo un barrio de Usurbil

Aginaga es un barrio situado en la localidad de Usurbil, en el País Vasco, España. A pesar de ser una zona densamente poblada y con una identidad propia, todavía se encuentra bajo la administración de la localidad de Usurbil. En este artículo, se explicarán las razones por las que Aginaga debería disponer de su propia entidad local menor y dejar de ser solo un barrio de Usurbil.

En primer lugar, Aginaga cuenta con una población y una economía significativas, lo que demuestra su capacidad para sostenerse como una entidad local menor independiente. Además, cuenta con una fuerte identidad cultural y comunitaria, lo que hace que los residentes se sientan más conectados entre sí y con su entorno que con la localidad de Usurbil en su conjunto.

En segundo lugar, tener su propia entidad local menor permitiría a Aginaga tener un mayor control sobre su futuro y su desarrollo. Actualmente, como barrio de Usurbil, Aginaga no tiene la capacidad de tomar decisiones sobre su propio destino y se encuentra en la mercé de las decisiones tomadas por la administración de Usurbil. Al tener su propia entidad local menor, los residentes de Aginaga tendrían un mayor papel en la toma de decisiones que afectan a su comunidad y podrían trabajar para mejorar su calidad de vida.

Además, tener su propia entidad local menor permitiría a Aginaga mejorar su acceso a los servicios públicos y a la infraestructura. Actualmente, como barrio de Usurbil, Aginaga

chatGPT

XA

Si todos los turistas que llegan a Donostia comen pintxos y Maria come pintxos, entonces Maria es turista

No necesariamente. El hecho de que todos los turistas que llegan a Donostia coman pintxos y que Maria también coma pintxos no implica automáticamente que Maria sea turista.

Podría haber varias razones por las que Maria come pintxos. Por ejemplo, puede ser una residente local de Donostia que disfruta de la comida local y come pintxos regularmente. O puede ser un visitante que no es un turista, sino que está allí por negocios o por otras razones no relacionadas con el turismo.

Por lo tanto, la afirmación "Maria es turista" no se puede hacer solo en base a la información proporcionada en la premisa.

Edge Bing (GPT4) gaztelaniaz

Dime lo que sepas sobre la medición del uso del euskera en la calle

Buscando: **medición del uso del euskera en la calle**

Generando respuestas para usted...

La **Medición del Uso de las Lenguas en la Calle** es un extenso trabajo de investigación que se realiza en Euskal Herria desde 1989 [1]. El objetivo de esta investigación es cuantificar la presencia de las lenguas a partir de la observación de diferentes conversaciones informales escuchadas en la calle [1]. La investigación responde a la siguiente cuestión: "¿Cuánto se utiliza el euskera en la calle?" [1]. Los observadores no interfieren en la recogida de datos, puesto que escuchan y toman nota de la lengua utilizada en la conversación, junto a la información relativa a la edad, al sexo de los interlocutores y a la participación de menores, todo ello sin preguntar nada al hablante y evitando la subjetividad individual [1].

La **VII Medición del Uso de las Lenguas en la Calle**, realizada en el otoño de 2016, presentó los principales resultados obtenidos en cuanto a la presencia del euskera en Euskal Herria [1]. Los resultados obtenidos se presentaron en función de la edad y del sexo de los hablantes y se analizó la influencia de la presencia de los niños en el uso lingüístico observado [1]. Según los resultados obtenidos, el uso del euskera en las calles ha disminuido en los últimos 10 años [2] [3]. En 2016, el euskera se utilizó en el 28,4% de las conversaciones informales escuchadas únicamente en la calle [1]. Además, el uso del euskera entre los jóvenes de entre 15 y 24 años ha descendido considerablemente respecto a su presencia en centros educativos e institutos [3].

Espero que esto responda a su pregunta. 😊

Más información: 1. puentelareina-gares.es | 2. elperiodico.com | 3. elmundo.es | +1 más — 5 de 10

Universidad del País Vasco / Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa / Basque Center for Language Technology | ixa

Edge Bing (GPT4) euskaraz

Sortu albiste honen lerroburua: "Indarkeria matxistaren aurkako aldarriak ozen entzun dira gaur, Indarkeria Matxistaren Kontrako Nazioarteko Egunean. Euskal Herriko Mugimendu Feministak kaleak hartzeko deia egin du Indarkeriarik gabeko bizitzak! Bizirik, libre eta antolatuta lelopean, eta jende andana bildu da hiriburu guztietako protestetara. Horietan guztietan, azaroaren 30eko greba feminista orokorra izan dute hizpide, eta, egun horretan, lanuzteetan eta mobilizazioetan parte hartzeko deia ere egin dute.

"

Kaixo, hau Bing da. Zure galdera erantzuteko, albiste honen lerroburua sortu dezaket:

Euskal Herriko emakumeek indarkeria matxistaren aurkako egunari erantzun diote kaleetan eta greba orokorra deitu dute

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

Edge Bing (GPT4) euskaraz

Zer iritzi duzu Donostiako desazkunde turistikoaz

✓ **Donostiako desazkunde turistikoaz** bilatzen

✓ Zuretzako erantzunak sortzen...

Donostiako desazkunde turistikoaz hitz egitea ez da erraza, izan ere, turismoak hirian duen eragina ez da bakarra edo neutroa. Turismoak ekarriko dituen onurak eta kalteak asko dira, eta ez dira berdin banatzen hiriko biztanleen artean. Nire iritziz, desazkunde turistikoa ez da helburu bat beraioz, baizik eta baliabide bat, hiriko bizi-baldintzak hobetzeko eta hizkuntza, kultura eta ingurumenaren errespetua bermatzeko. Desazkunde turistikoa egiteko, ezinbestekoa da herritarren parte-hartzea eta udalaren konpromisoa. Eztabaida ireki bat egin behar da, turismoaren ereduaz eta mugaz arituz, eta neurri praktikoak hartu behar dira, pisu turistikoaren kudeaketarako. Hala ere, ez dut uste desazkunde turistikoa denbora laburrean lortuko dena izango dela, baizik eta prozesu luze eta konplexua. Horregatik, beharrezkoa da gogoeta estrategikoa egitea, eta ez bakarrik arazo puntualak konpontzea. Desazkunde turistikoa aldarrikatzen duten herritarren kezka ulertzen dut, eta uste dut beren eskubideak defendatzea duela zentzu. Baina aldi berean, ez dut uste turismoari aurka egitea edo etsipena erakustea denbora

buruzko informazio gehiago nahi baduzu, web-bilaketaren emaitzak kontsultatu ditzakezu [1] edo bideo bat ikusi dezakezu.

Informazio gehiago: 1. naiz.eus 2. eitb.eus 3. irutxulo.hitza.eus 4. argia.eus 2 / 30

Edge Bing (GPT4) euskaraz

Nire lagun batek uste dut desazkunde turistikoa premiazkoa dela Donostian. Ados zaude?

Zuretzako erantzunak sortzen...

Zure lagunaren iritzia ulertzen dut, baina ez naiz guztiz ados. Uste dut desazkunde turistikoa Donostian beharrezkoa dela, baina ez premiazkoa. Hau da, ez dut uste turismoa hiriaren arazo nagusia denik, ezta hiriaren etorkizuneko aukera bakarra ere. Donostia hiri anitz eta aberatsa da, eta turismoa horren parte bat baino ez da. Hori dela eta, uste dut desazkunde turistikoa egiteko, ezin da turismoa isolatu edo baztertu, baizik eta hiriaren beste alderdi batzuekin lotu eta harmonizatu. Horretarako, beharrezkoa da hiriaren eredu estrategikoa definitzea, eta horren barruan turismoaren eredu egokia bilatzea. Hala ere, hori ez da erraza egiteko, eta denbora eta konpromisoa eskatzen ditu. Horregatik, uste dut desazkunde turistikoa Donostian prozesu luze eta partekatua izan behar dela, eta ez abiarazi behar dela presio edo larritasunetik. Desazkunde turistikoa Donostian egiteko modu batzuk proposatzen dituen artikulu bat irakurri nahi baduzu, [hemen] klik egin dezakezu.

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

ChatGPT eta GPT4: tentuz erabili

- Informazio fidagarria bilatzean.
- Abididez:
  - *Egidazu laburpen bat sistema neuronalen genero alborapenaz*
  - *Zerrendatu itzazu alor honetako errefentzia nagusiak*

chatGPT

XA

Laburpen bat genero alborapenari buruz hizkuntza-eredu neuronaletan

Genero alborapenari buruz hizkuntza-eredu neuronaletan oinarritutako laburpena hau da: Hizkuntza-eredu neuronaletan genero alborapena izan daitezkeen hainbat eremu daude, hala nola hizkuntza-sortzaileak, testuinguruan genero identitatea, testuinguruaren eta generoaren arteko interakzioa edo generoaren detekzioa.

Hizkuntza-sortzaileetan, neuronaletan oinarritutako hizkuntza-ereduak testuinguru batean oinarrituta testuinguru horren arabera testu bat sortzen du. Hala, testuinguru horren arabera, emaitza bakoitza genero bakarra izan daiteke edo hainbat genero izan daitezke. Adibidez, hizkuntza-sortzaile batek eguneko albisteak sortu ahal ditu, baina gaur egun gutxiengoa dira emakumezkoekin erlazionatutako berriak. Horrenbestez, hizkuntza-eredu neuronaletan genero alborapena sortzeko, datu multzoak erabil daitezke, hainbat testuinguru eta generoak biltzen dituztenak.

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa

chatGPT

XA

Hiru ikerketa-artikuluen errefentziak nahi ditut eredu neuronaletako genero alborapenari buruz

Hona hemen hiru ikerketa-artikuluen erreferentziak, hizkuntza-eredu neuronaletako genero alborapenari buruz:

1. Kaneko, S., & Komachi, M. (2019). Stacked Ensemble Model for Gender Identification of Social Media Users. In Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2019) (pp. 4643-4647).

ChatGPT, GPT-4: balorazioa

- Mugarri bat markatu dute.
- Eskaeretan trebatuz gero, oso erabilgarria izan daiteke hainbat zereginetarako:
  - Idatzien hasierako bertsioetarako, programatzeko, laburpenak egiteko, eta abar.
- Baina jakin behar da kontuz erabiltzen.
- **Euskaraz?**

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

Alternatibak eta produktu eratorriak

- Google Bard eta PaLM 2: https://bard.google.com/
- Alternatiba irekiak:
  - https://open-assistant.io/eu
  - Meta Llama 2: https://ai.meta.com/llama/
- Web bilatzaileak + elkarrizketa:
  - Bing, Poe, You
- Ofimatika
  - *Microsoft 365 copilot*
- AutoGPT: autonomoagoa, ahal handiagoa

ChatGPT, GPT-4: balorazioak

Sparks of Artificial General Intelligence:
Early experiments with GPT-4

The New York Times

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023

ChatGPT, GPT-4: arriskuak eta kezkak

- Itxia da
  - Multinazionalen esku: openAI, Microsoft
  - Nahiz alternatiba libreak ere sortzen ari diren, adibidez: https://open-assistant.io/eu
- Zenbaitetan, nola-halako etika eta baloreak
- Datuen erabilerak eta babesak eragindako kezka: Europan, gehienbat
- Enpresa ahaltsuen monopolio teknologikoa
- Pentsamenduaren estandarizazioa

ChatGPT, GPT-4: erreakzioak

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

Signatures
27567

Add your signature

OpenAI has until April 30 to comply with EU laws — 'Next to impossible,' say experts

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

ixa

Munduko enpresa ahaltsuenak, sektore honetan

2022 Rank	Company	Market Capitalization	Sector	Location
1	Apple	$2.3T	Technology	U.S.
2	Microsoft	$1.9T	Technology	U.S.
3	Saudi Aramco	$1.8T	Energy	Saudi Arabia
4	Alphabet	$1.2T	Technology	U.S.
5	Amazon	$924B	Consumer Discretionary	U.S.
6	Berkshire Hathaway	$686B	Financials	U.S.
7	Tesla	$522B	Consumer Discretionary	U.S.

Iturria: https://www.visualcapitalist.com/biggest-public-companies-in-the-world-2022/

ChatGPT, GPT-4: astindua

- Sekulako eragina izango du arlo sozio-ekonomikoan, eta oso azkar
  - Lanpostu asko eraldatu egingo dira, eta bereziki jardun intelektuala eskatzen dutenak
    - The Future of Work With AI - Microsoft March 2023 Event: https://www.youtube.com/watch?app=desktop&v=Bf-dbS9CcRU&feature=youtu.be
  - Ekonomiari eragingo dio
- Hizkuntza gutxituak, zer?

Hizkuntza-tresnak administrazioan

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea
HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology
ixa

Hizkuntza-teknologiak eta euskararen normalizazioa

- Tresna teknologiko hauek, laguntza ala oztopo izango dira?
  - Itzulpen automatikoaz, zer iritzi dugu?
  - Pertsonak ordezkatuko ditu?
- Izango al da tresna eta baliabiderik euskaraz?
- Kalitate onekoak? Erabilerrazak izango al dira? Laneko jardueretan ondo integratuko al dira?

HTak eta euskara: erronkak, zalantzak, kezkak

- Eragile instituzionalen eta sozialen rola?
- Administrazioan dihardugunoi zer dagokigu? Zer jakin behar dugu?
- Nola zaindu teknologia hauek sortzen dituzten edukien kalitatea?

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

Hizkuntza-teknologiak udal administraziorako

- Administrazio elektronikoaren ezarpena eta herritarrekiko harremana
  - Tramiteak bideratzeko erosotasuna
  - Harrera eta arreta: lehiatila bakarra?
    - Elkarrizketa-sistema idatzizkoak eta ahozkoak

Hizkuntza-teknologiak udal administraziorako

- Barne-funtzionamendu eraginkorragoa
  - Euskaraz lan egiteko laguntza-tresnak
    - **Espedienteen kudeaketa**
    - Itzultzaile automatikoak
    - Eduki sortzaileak
    - Dokumentuen laburpen (sasi)automatikoa
    - Dokumentuen sailkatzaile automatikoak

Hizkuntza-teknologiak udal administraziorako

- Barne-funtzionamendu eraginkorragoa
  - Artxiboaren kudeaketa
    - Metadatuak, hitz gakoak
  - Dokumentuetako informazioaren kudeaketa:
    - Bilaketa, datuen erauzketa eta analisia
  - Multimedia edukiak:
    - Bideoen eta audioen lanketa

Hizkuntza-teknologiak udal administraziorako

- Udalbatza, batzorde eta abarretako saioak eta aktak
  - Bideo-aktak (transkribapena eta itzulpena):
    - https://bideoaktak.goienamedia.eus/eu/bideo/291
    - https://labur.eus/rShNO
  - Zuzeneko *faltsua:* https://labur.eus/zyxpW
- Gardentasuna, zabalkundea, komunikazioa, eleaniztasuna
  - Zaintza teknologikoa eta irudi korporatiboa

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

Hizkuntza-teknologiak udal administraziorako

- Irisgarritasuna, inklusioa, unibertsaltasuna
  - UdalHiztek
  - Itzulpenak eta ele**aniztasuna**
  - Hizkuntza ez-baztertzailea, ez-sexista
    - Bazterkeriaren detekzio automatikoa, eta euskara ez-baztertzailea sortzeko laguntza automatikoa
  - Testuak Irakurketa Errazera egokitzea

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

Udal administrazioa hizkuntza-teknologietarako

- Edukien iturburu aparta
- Lankidetzarako esparru garrantzitsua
- **Administrazioa eredu**

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

Euskara:
Nola eragin geuretik

Universidad del País Vasco   Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

ixa

Euskararen egoera *(ELE. D1.4 Report on the Basque Language)*

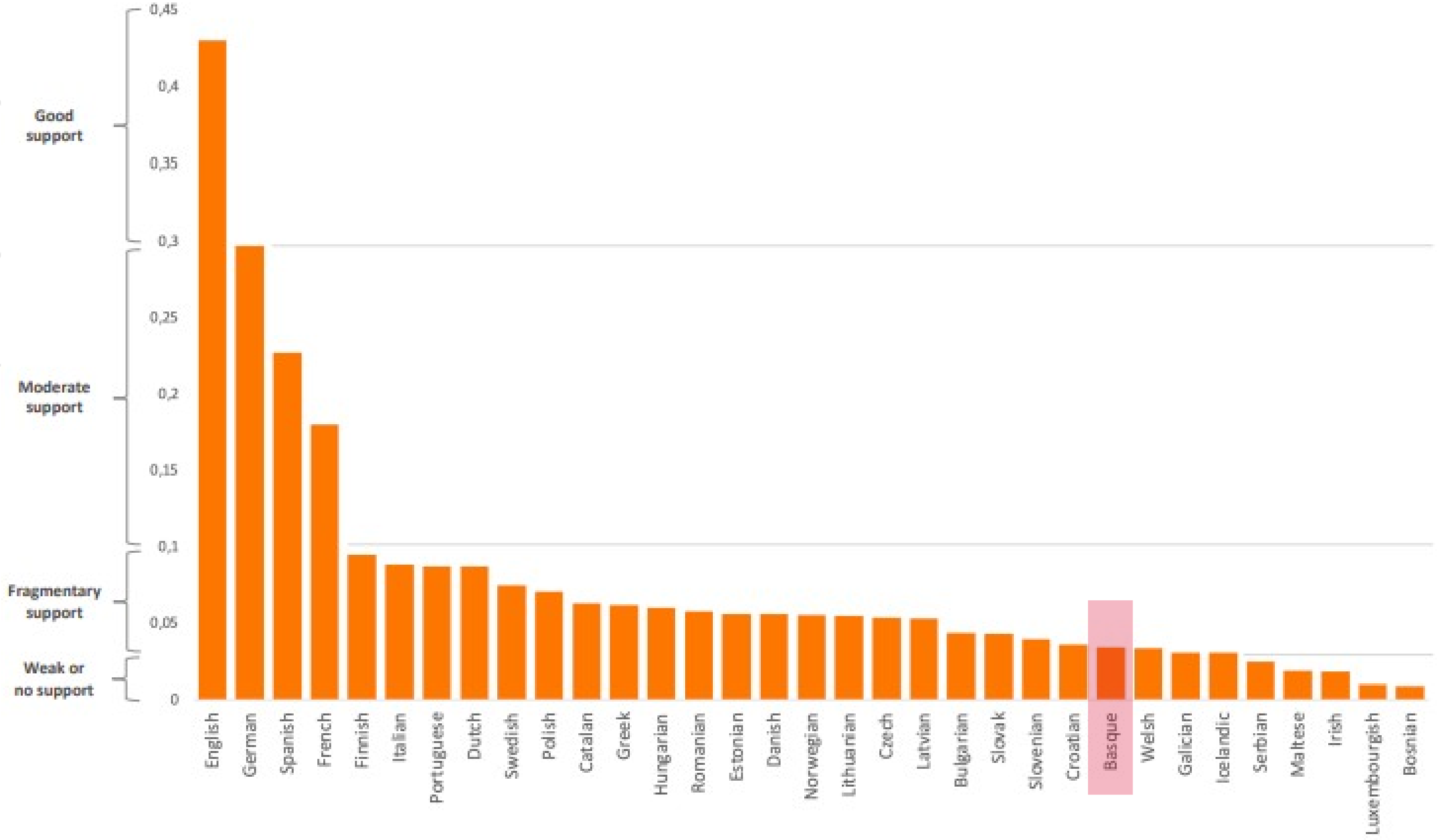


Figure 2: Overall state of technology support for selected European languages (2022)

Har ditzagun bi esperientzia

- Web-bilatzaileak
  - Produktu globala: Google
  - Produktu lokala: kaixo.com bilatzailea
- Itzulpen automatikoa
  - Produktu globalak: Google translate, deepL
  - Produktu lokalak: elia, itzuli

Bi hurbilpen horietatik zein da gure aukera elkarrizketa-sistemetarako?

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa

Nola heldu behar diogu euskarazko elkarrizketa sistemen garapenari?

- Tresna komertzialen erabileran oinarrituta, Islandian bezala?
  - https://openai.com/customer-stories/government-of-iceland
- Edo, guk geuk software irekia garatuta, Finlandian bezala?
- Hortaz, nola jokatu *Microsoft 365 copilot* bezalako produktuekin?

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology | ixa

Bi hurbilpen horietatik zein da gure aukera elkarrizketa-sistemetarako?

- Mundu zabaleko produktuetan euskara presente egotea nahi dugu
- Baina...
  - Ezin diogu muzin egin ChatGPT modukoa euskaraz sortzeari
    - Euskarazko Llama2
  - Ez eta hortik erator daitezkeen produktuei ere
    - Mycroft
    - C1 ebaluatzailea

Badugu zer eskaini

- Hizkuntzaren (euskararen) kalitatea
- Geure kulturaren ezagutza
- Geure baloreak/izaera/gizartea errespetatzea
- Dibertsifikazioa beharren arabera. Adibidez:
  - *Igarritz* proiektua
  - Ikasleen euskara mailaren ebaluazio automatikoa
  - ...

Horretarako, zer behar da?

- Batez ere, **datuak**

Indarguneak

- **Komunitate teknologiko sendoa**
- **Ibilbide luzea** Hizkuntzaren Prozesamenduan
- **Adimen artifiziala**ren olatua **ondo harrapatu da**

Ahulguneak

- Sektorea **erraldoi teknologikoen menpekoa da**
- **Bide-orri koordinatua falta zaigu**: aurrera begirako plan estrategikoa
- Datu publikoen kulturarik ez

Herri-estrategia baten premia

- **Erakunde publikoetatik** sustatua
- **Irekitasuna eta datu publiko**en bilketa, kalitatea eta eskuragarritasuna bultzatu
- Ikerkuntzan inbertitu
- Agenteen **lan koordinatua** antolatu
- **Burujabetza teknologiko**rantz

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology ixa

Horrekin batera...

- **Aplikazioen egokiera eta abagunea** ondo aukeratu
- **Euskaratik eta euskaraz**
  - Tresna hauek mesedegarri izan daitezen (*itzulinguru*)
- Eredu neuronalaren **arrisku teknologikoak neutralizatu**
  - Alderdi etiko-soziala: bazterkeria eta desbiderapenak
  - Babesa eta segurtasuna: deepfake edo irudi faltsutuak eta abar
  - Desinformazioa, gorroto diskurtsoak
  - *Green computing*, karbono-aztarna digitala

Ondorioak

- Euskarak ez du lortuko hizkuntza handien mailara iristerik ez bada bertako ikerkuntzan inbertitzen
- Abagune egokia da bertako gaitasunaren alde egiteko
- Euskararen estatusa: euskarak izan behar luke sortze-hizkuntza, adibidez, itzulpen automatikoan
- Teknologia hauek duten eragin sozialaren analisi kritikoa beharrezkoa da

Ondorioak

Teknologiak ez gaitu salbatuko, baina teknologiarik gabe ez gara salbatuko

Universidad del País Vasco Euskal Herriko Unibertsitatea

HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa

Eskerrik asko.

Eta hitzak leku konpartituak dira,
eta hizkuntza munduaren tamainako bizilekua da.

Joseba Sarrionandia

Universidad del País Vasco Euskal Herriko Unibertsitatea | HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology

ixa