



(/)

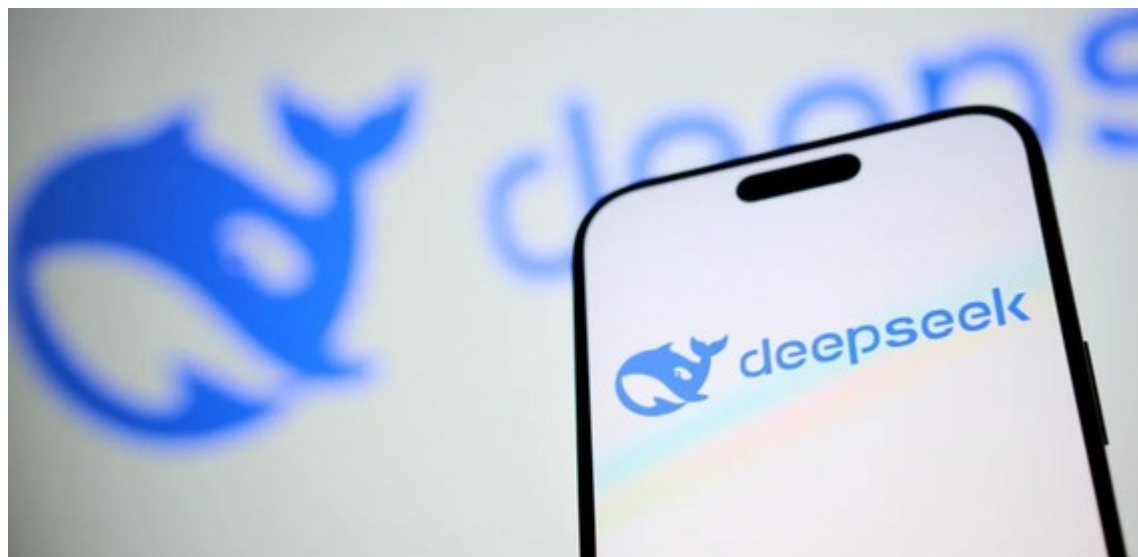
Pertsonalki instala daitekeen Deepseek nekez arituko zaizu euskaraz (oraingoz)



Sustatu

2025-01-29 : 09:10

Deepseek adimen artifizial (AA) txinatarra oso ondo dabilela euskaraz erakutsi genizuen atzo (<https://sustatu.eus/1738051172>). Eta bere abantaila batzuk gainetik azaldu, baliabide aldetik askoz gutxiago kontsumitzen ahal dela, eta libreki deskargatu eta instala daitekeela. Alde honetatik, zehaztapen batzuk gaurko artikuluan. 1.5B eta 8B bertsioak martxan ikusi ditugu ordenagailu arrunt batean, eta ez dira gauza euskaraz erantzuteko. Hala ere, gure aburuz, Deepseek-en etorrerak sortutako lurrikara ez da irudipenezkoa bakarrik, benetakoa baizik, eta euskararen alorrean ere izango da eraginik.



(https://

media.sustatu.eus/media/argazkiak/ogo/deepseek-logo.jpg)

Argipen batzuk, lehenbizi. Deepseek erabiltzen duzunean euskaraz txukun app edo webgunearen bidez, 600B-ko sistema erraldoi bat ari zara kontsultatzen. Hori bera libreki deskargatu eta erabiltzeko moduan jarri du enpresa txinatarrak; baina alimaleko azpiegitura informatikoa behar duzu erabiltzeko. Hala ere, beste eredu arinagoak jarri ditu Deepseek-ek jaisteko moduan (<https://huggingface.co/>

collections/deepseek-ai/deepseek-r1-678e1e131c0169c0bc89728d).

 **deepseek-ai/DeepSeek-R1-Distill-Llama-70B**

 Text Generation • Updated 2 days ago •  47.4k •  322

V

?

 **deepseek-ai/DeepSeek-R1-Distill-Qwen-32B**

 Text Generation • Updated 2 days ago •  132k •  639

 **deepseek-ai/DeepSeek-R1-Distill-Qwen-14B**

 Text Generation • Updated 2 days ago •  44k •  228

 **deepseek-ai/DeepSeek-R1-Distill-Llama-8B**

 Text Generation • Updated 2 days ago •  57.7k •  282

 **deepseek-ai/DeepSeek-R1-Distill-Qwen-7B**

 Text Generation • Updated 2 days ago •  72.5k •  234

 **deepseek-ai/DeepSeek-R1-Distill-Qwen-1.5B**

 Text Generation • Updated 2 days ago •  150k •  496

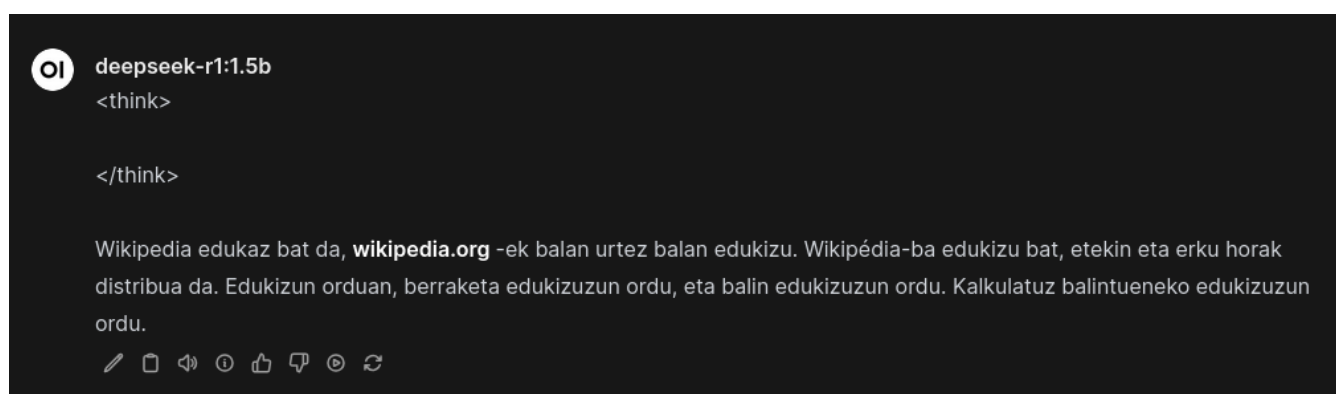
1.5B edo 32B irakurtzen dugularik, bilioi bat token esan nahi du horrek, AA sistema hauen neurketa elementu bat, arin samar esanda, zeinean B hori *bilioi* estatubatuarra den, izatez *mila milioi* edo *miliarra* da esangura. Azaroan euskara eredu batzuetan nola moldatzen zen konparatu zuten Hitz zentrokoek, eta Latxa 8B modeloa zein ondo moldatzen zen ikusi genuen:

Model 	ELO 	90% CI 
Claude Sonnet 3.5	1042	+12/-11
GPT-4o	1025	+10/-9

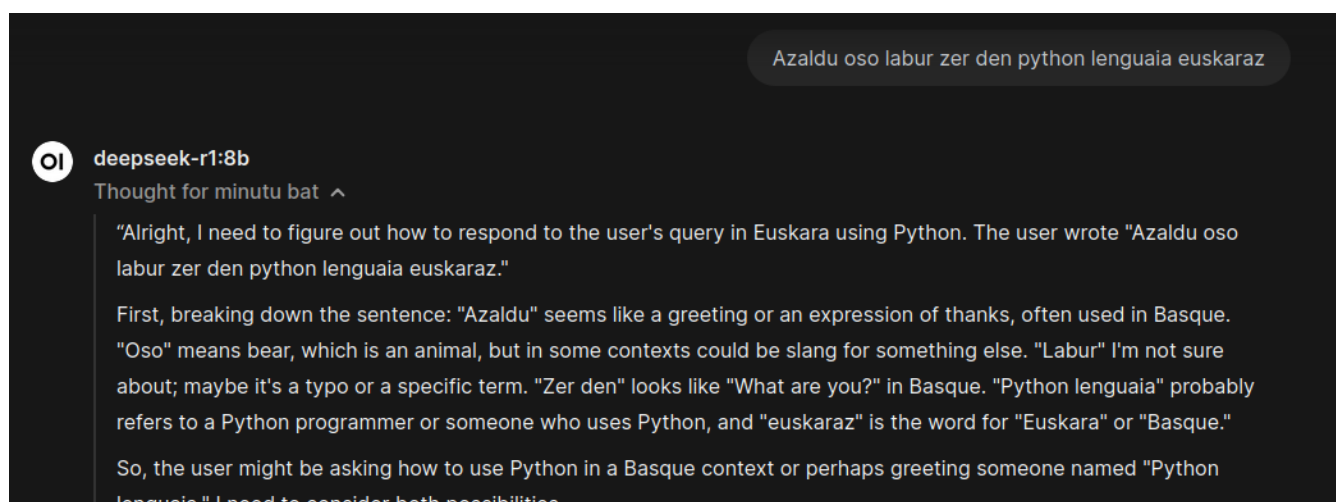
Latxa 8B (berria)	1021	+14/-12
Gemini 1.5 Flash 8B	1007	+12/-11
Command R+ 104B	958	+13/-15
Llama 3.1 Instruct 8B	945	+15/-13

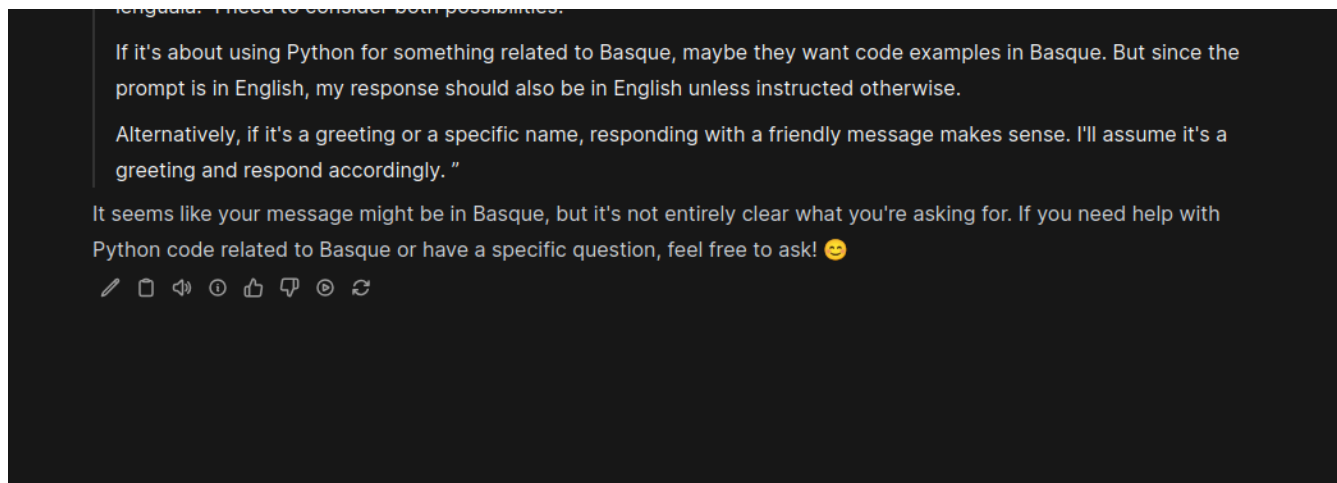
8B-rekin emaitza on hori lortu zuen Hitz zentroak, eta esperantza hor izan behar genuke euskararekin, mundua erre beharrik gabe funtziona dezaketen sistema arinak moldatzen badira euskaraz.

Deepseek-ek jaisteko (<https://huggingface.co/collections/deepseek-ai/deepseek-r1-678e1e131c0169c0bc89728d>) utzitako modeloetara itzulita, Sustatuk ikusi ahal izan du 1.5B eta 8B ereduak martxan ordenagailu arruntetan instalatuta, eta ez dabilta euskaraz. Batek euskara mordoilo bat sortzen du; ikus adibidez Wikipedia zer den galdetu diogunean 1.5 B ereduari:



8B eredua indartsuagoa da. Saiatzen da logika aplikatzen eta euskara ulertzen, baina erantzuna ingelesez ematen dizu: koherentzia kognitibo ote? ezin euskaraz erantzun eta ingelesez saiatu? Ikus:

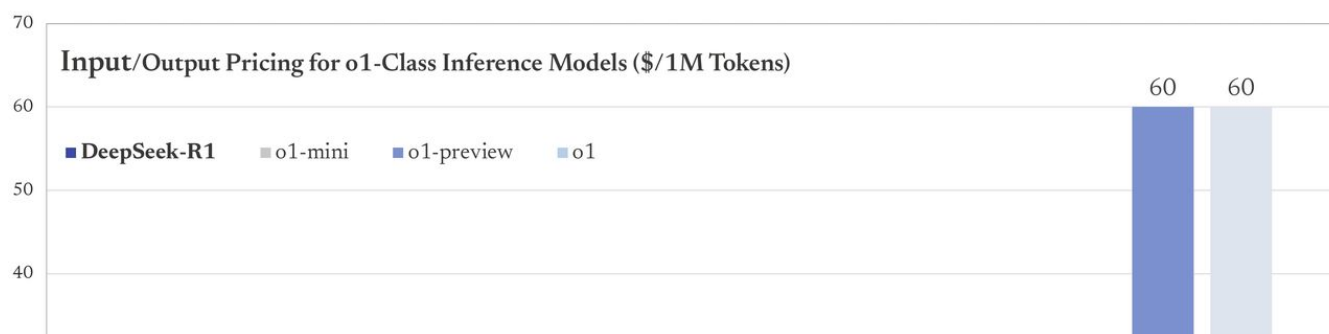


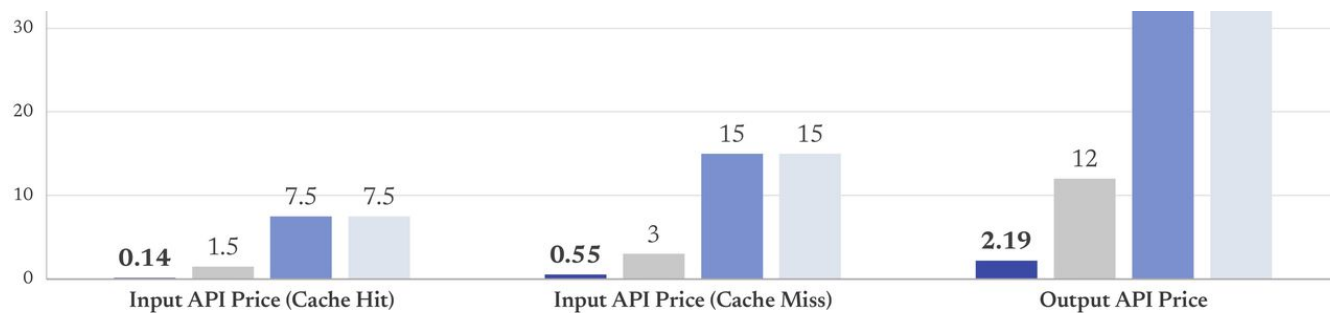


Bi aproba hauetan, bere arrazoibidea erakutsi egiten dizu sistemak: ikus goiko irudia, ingelesez esplikatuta dago ze urrats egiten dituen Deepseek-ek erantzuna eman aurretik.

Goiko emaitzen pantailak ulertzeko, izan kontuan gordinki deskargatuta egin direla aprobak. Baina tresna horiek elikatu eta trebatu daitezke euskarazko datu eta testu gehiagorekin, eta hortik hobekuntzarako bidea. Hain zuzen ere, goragoko Latxa 8B-ren emaitza oma oinarritu da Metaren Llama sisteman, baina Hitz zentroan trebatu eta hobetu dutelako hobetu da. Agian aukera dute orain Deepseek ere euskal ingurunerako hobetzen. Deepseek-en oinarritutako euskarazko garapen gehiago ikusiko ditugula uste dugu, edo bestela Deepseek bestalako beste sistema txinatar / arin batzuen ondorio direnak.

Gainerakoan, 600B-ko Deepseek erraldoia pertsonalki erabiltzea erraza da, app-a instalatu edo web bidez konektatu eta kito. Erabilpen edo ustiaketa sakonago baterako. Deepseek-ek merkatuan jarri duen ordainpeko API zerbitzua dago. Eta hor prezioak haustera etorri dira txinatarrak. Taula honek erakusten du urtarri amaieran honetan OpenAI (ChatGPT) enpresak kobratzen dituenak bere APIaren erabilpenagatik (o1 bertsio horiek OpenAIrenak dira); eta marra urdin ilunak adierazten du Deepseek-en prezioa:





Deepseek-i berea sortzea askoz gutxiago kostatu zaio OpenAI-ri baino, eta horregatik birrindu ditzake prezioak. OpenAI azpiegiturretan egiten ari den inbertsio ikaragarriak nola ordainduko ditu txinatar hauek (eta etorriko diren beste batzuek) ezarritako prezio eta gaitasun berrien lehiaren aurrean? Horra AEBetako merkatuetan sentitu duten lurrikararen gakoa.



CodeSyntaxek kudeatua, Eusko Jaurlaritzaren Hizkuntza Politika eta Kultura Sailak (Hizkuntza Politikarako Sailburuordetzak) (<https://www.kultura.ejgv.euskadi.net/r46-704/eu/>) diruz lagundua.