

# Helburu zehatzetarako eta domeinu irekiko txatboten hibridazioa

## Proposer(s) / Proposatzailea(k):

Elhuyar Hizkuntza Teknologiai  
Xabier Saralegi, PhD  
Testu Meatzaritzaren arduraduna

## Contact / Kontaktua:

x.saralegi@elhuyar.eus  
Tel.: 943363040 | Iuzp.: 220 | mugikorra: 688825874  
Zelai Haundi, 3  
Osinalde industrialdea  
20170 Usurbil

## Description / Deskribapena

Txatbotak hizkuntzaren naturalari erantzuten dien elkarrizketa-agente mota bat dira. Hizkuntza naturala testu eta ahots bidezkoa izan badaiteke ere, proiektu honetan testu bidezko txatbotetan jarriko dugu arreta.

Honelako txabotek elkarrizketa eskaintzen diote erabiltzaileari, eta kasu batzuetan, aginduak ere bete ditzakete. Horrela bereizi daitezke **helburu zehatzik gabeko txatbotak** (Adibidez, hizkuntzen ikasketan edo bideo-jokoetako pertsonaietan erabilitakoak), eta **helburu zehatzetarako txatbotak** (Adibidez, arreta teknikoko zerbitzuen atzean daudenak).

Berriki, bi txatbot mota hauek hibridatzeko teknikak lantzen hasi dira hainbat ikertzaile. Hibridazio honen arrazoia bi txatbot mota hauen alderdi positiboak uztartzea da, estete baterako, helburu zehatzetarako txatbotek atazak burutzeko duten sendotasuna, eta helburu zehatzik gabeko txatboten giza-hizkera. Hibridazio honi esker erabiltzaile batek erosoago jardungo du helburu zehatzetarako txatbot batekin, azkeneko honek modu naturalean hitz egiteko eta domeinutik kanpoko erantzunak emateko gaitasuna izango duelako.

Txatbotak inplementatzeko tekniken artean bi mota nagusi berezi daitezke, erregeletan oinarritutakoak eta datuetan oinarritutakoak. Azkenaldian bigarren teknika-multzoa nagusitzen ari da, batez ere, medio sozialen ez tandarekin batera (Adibidez, Yahoo Answers, WikiAnswers, eta Twitter), txatbotak eraikitzeke datu-multzo handiak eta baliagarriak sortzen ari direlako.

Proiektu honetan euskararako helburu zehatzetarako (domeinua proiektua hasterakoan zehaztuko da) txatbot bat garatu nahi dugu, helburu zehatzik gabeko txatboten ezaugarriak ere izango dituen. Txabota garatzeko datuetan oinarritutako teknikak erabiliko dira.

## Goals / Helburuak

Proiektuaren helburu nagusia euskararako txatbot hibrido bat garatzea da, helburu zehatzetarako eta helburu

zehatzik gabeko txatbot motak uztartuko dituen. Hori lortze aldera honako azpi helburuak bete beharko dira:

- Eskarazko txatbotak trebatzeko baliabideak prestatzea.
- Informazioaren berreskurapenean (*Retrieval-Based*) eta eredu generatiboetan (*Generative Models*) oinarritutako paradigmatik aldatzea.
- Hizketa luzeko eta hizketa laburreko txatbot motak aztertzea.
- Helburu zehatzetarako txatbota eta helburu zehatzik gabeko txatbotak hibridatzea.

## Requirements / Betebeharrak

- Euskara.
- Ikasketa automatikoa:
  - Decoder-encoder arkitekturak.
- Python:
  - Keras, Tensorflow.
  - Scikit-learn.

## Framework / Esparrua

Txatboten teknologia Elhuyar Hizkuntza Teknologietako zenbait produktutan integratzeko asmoa dugu. Horri begira txatbotak garatzeko paradigma ezberdinak aztertzen ari gara lehenengo fase batean. Proiektu hau fase horretan kokatzen da.

## Tasks and plan / Atazak eta plana

- 1. Lan-paketea: Artearen egoera aztertzea:
  - Datuetan oinarritutako txatbotak.
  - Helburu zehatzetarako txatboak.
  - Helburu zehatzik gabeko txatbotak.
  - Txatbot hibridoak.
- 2. Lan-paketea: Baliabideak prestatzea:
  - Entrenamendu eta ebaluaziorako datu-multzoak (*datasetak*).
- 3. Lan-paketea: Esperimentazioa eta ebaluazioa:
  - Sistema komertzial batean (Dialogflow) oinarritutako oinarri-lerroko sistema (*baselinea*).
  - Helburu zehatzetarako txatbotak (*goal-driven chatbot*).
  - Domeinu irekiko edo helburu zehatzik gabeko txatbota (*non-goal-driven chatbot*).
  - Txatbot hibridoa.
- 4. Lan-paketea: Prototipoaren garapena.