

# DABISKĀS VALODAS DZIĻO NEIRONU TĪKLU ĢENERATORA *GPT-2* SNIEGUMS ANĢĻU VALODĀ

## Darba mērķis:

Iepazīties ar dabiskās valodas apstrādes pieejām un neironu tīklu lomu valodas apstrādē. Praktiski izmēģināt OpenAI izstrādāto modeli *GPT-2* angļu valodas tekstu ģenerēšanā. Apmācīt modeli ar latviešu valodas datiem, lai spētu ģenerēt tekstus arī latviešu valodā.

## Rezultāti:

Darba sākumā izvirzītie mērķi pilnībā sasniegti netika. Ir izdevies praktiski ģenerēt dažādus tekstu paraugus angļu valodā ar diviem no četriem modeļa *GPT-2* variantiem ar un bez pamudinājumiem. Apmācīt modeli ar latviešu valodas datiem pagaidām nav izdevies. Ir izdevies iegūt lielu latviešu valodas korpusu.



## Secinājumi:

Ar darbā apskatīto modeli ir iespējams ģenerēt patiešām labus un kvalitatīvus tekstu paraugus angļu valodā.

Galvenie novērojumi, apskatot ģenerētos tekstu paraugus, ir:

- Sintaktiskā un teikumu uzbūves ziņā ģenerētie teksti ir vērtējami kā kvalitatīvi.
- Ir novērojamas problēmas ar pielāgošanos specifiskai tēmai.
- Reizēm ir novērojamas semantiskā līmeņa nesaderības.

## Darba turpiējums:

Nākotnes darbā plānots ģenerēt angļu valodas tekstus arī ar *large* un *extra large* *GPT-2* modeļa variantiem.

Kā arī sasniegt sākotnēji izvirzīto mērķi: apmācīt modeli latviešu valodas tekstu ģenerēšanai.

**Darba autors:** Anete Alberte (*aa19147*)

**Darba vadītājs:** profesors, Dr.Sc.Comp. Guntis Bārzdīnš