

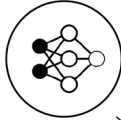
Facebook assembles over 160 GB of text, including books, news, and English Wikipedia.



Facebook pre-trains RoBERTa using masked language modeling and next sentence prediction.



RoBERTa is made available as open-source code.



Applicants write 150-word essays in response to this prompt: "Please briefly elaborate on one of your extracurricular activities or work experiences."



Criteria for identifying personal qualities are developed.



Human raters use criteria to code essays of about 3,000 applicants.



Human ratings are used to fine-tune our pre-trained version of RoBERTa.



Computer model generates personal quality scores for over 300,000 applicants.

