Abstract—The language model is typically represented as an unsupervised distribution estimate from a set of examples, each consisting of symbol sequences, and it could predict over sequences of words. We demonstrate the language model based on Generative Pretrained 2 will have a readable generated article for the journalistic robot. Nowadays, there is some trending of journalistic in Indonesia, freedom of the press, and it enables every journalist to make unprofessional news on the media. The problem affects the raise of journalist numbers who have lack journalistic knowledge and increases the amount of inappropriate news content in Indonesia. Therefore, to improve the quality of news produced by the mass media in Indonesia, a journalistic robot is needed to produce news content by the guidelines and the journalistic code of ethics. This research uses language modeling based on GPT-2 to generate articles. The program has four primary steps: building dataset, fine tuning GPT-2, modeling the trained data, and create articles. Furthermore, this research will add an Indonesian model for GPT-2 since the main purpose of this research is Indonesian articles. This paper proposes GPT- 2 to be applied to news contents and calculate the result with BLEU scores to check if the results are readable content. These findings show that the proposed model is capable of generating a readable article after trained by 110 Indonesian articles with an excellent score of BLEU.

Index Terms—journalistic robot, language model, Indonesia, Generative Pretrained Transformer 2