This thesis investigates non-task-oriented open-domain dialogue modeling, i.e., generating responses in a basic social conversation, using the ConvAI2 dataset, focusing on using neural language models and maintaining consistent chatbot's personality and overall coherence. We work with finetuning models from the GPT-2 family; we improve over a basic finetuned generation setup with a two-stage approach and an additional model learned to rank candidates by the first model. We further improve model training using direct preference optimization. Our modifications achieve state-of-the-art performance in the F1 score on the ConvAI2 dataset. We then engineer a prompt for GPT-3.5 to use this large language model for the task. Human evaluation experiment reveals that, despite lower F1 scores, the GPT-3.5 model surpasses all others in performance.