Current dialogue systems typically consist of separate components, which are manually engineered to a large part and need extensive annotation. End-to-end trainable systems exist but produce lower-quality, unreliable outputs. The recent transformer-based pre-trained language models such as GPT-2 brought considerable progress to language modelling, but they rely on huge amounts of textual data, which are not available for common dialogue domains. Therefore, training these models runs a high risk of overfitting. To overcome these obstacles, we propose a novel end-to-end dialogue system called AuGPT. We add auxiliary training objectives to use training data more efficiently, and we use massive data augmentation via back-translation and pretraining on multiple datasets to increase data volume and diversity. We evaluate our system using automatic methods (corpus-based metrics, user simulation), human evaluation as part of the DSTC 9 shared task challenge (where our system placed 3rd out of 10), as well as extensive manual error analysis. Our method substantially outperforms the baseline on the MultiWOZ benchmark and shows competitive results with state-of-the-art end-to-end dialogue systems.