



*speech
de-tokenizer*

“It was the night---
silent night, whence ...”



*image
de-tokenizer*



*music
de-tokenizer*



ANYGPT



*speech
tokenizer*



speech



text

*image
tokenizer*



image

*music
tokenizer*



music