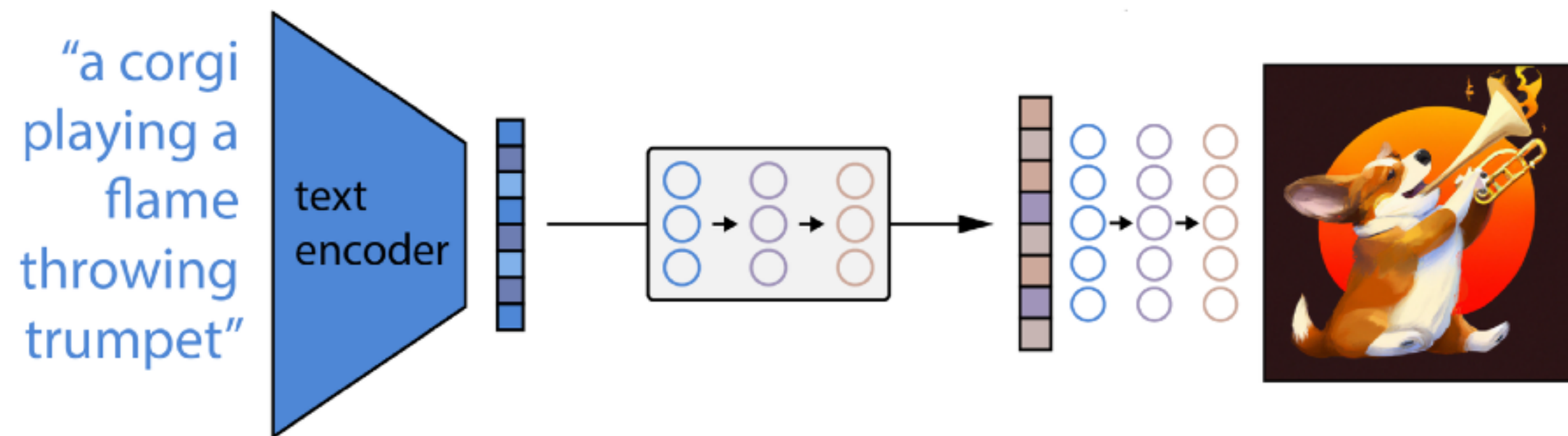


# DALL-E

- It receives both the text and the image as a single stream of data containing up to 1280 tokens
- Train using Max L



- The link between textual semantics and their visual representations in DALL-E 2 is learned by another OpenAI model called CLIP (Contrastive Language-Image Pre-training).



# CAPABILITY

