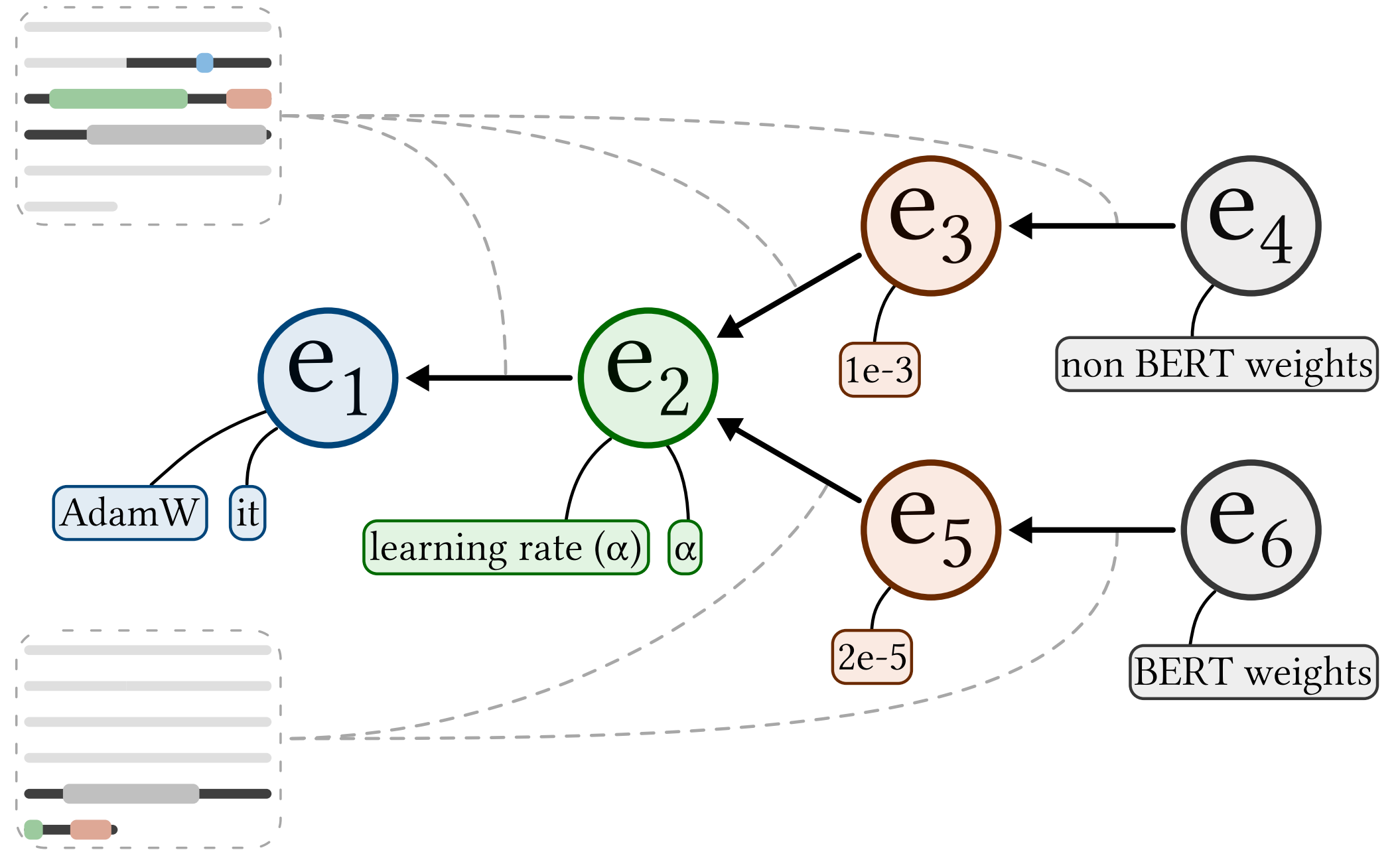


AdamW is chosen as the optimizer. We use it with a learning rate (α) of $1e-3$ for all non BERT weights. For BERT weights we set α to $2e-5$.

text example
based on arXiv:2005.00512



relation
evidence

entities \bigcirc , relations \leftarrow , and
entity mentions \square extracted