large vocab

Q-values
(simultaneously updated)

learning signal

skiing   down   the   ...   <EOS>

r = 0   r = 0   r = 0   r = 97.3

Off-Policy Data Samples   **On-Policy Model Samples**

Learning from Noisy (Negative) Text

Dataset

Positive Examples

Negative Examples

Train → **Model** → Generate → Positive Outputs
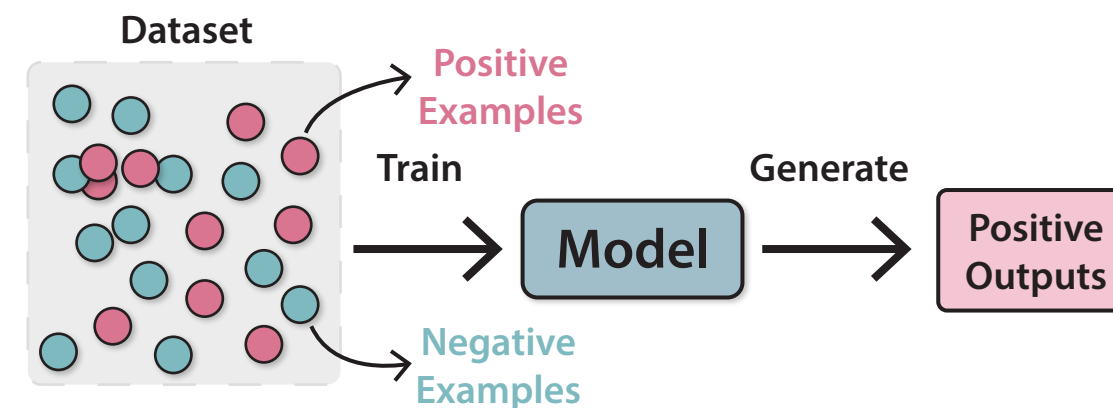
Adversarial Text Generation

Model → Generated Adbersarial Text → **Classifier**

Prompt Generation for Controlling Pretrained LMs

[Topic] "Science" → **Model** → fed into → Pretrained Language Model

[Prompt] "the chemical microscope is"   [Input] "In summary"   [Output from the LM] "... the composition of the chemical ..."