



Adding moderation

This shows how to add in moderation (or other safeguards) around your LLM application.

```
%pip install --upgrade --quiet langchain  
langchain-openai
```

```
from langchain.chains import  
OpenAIModerationChain  
from langchain_core.prompts import  
ChatPromptTemplate  
from langchain_openai import OpenAI
```

```
moderate = OpenAIModerationChain()
```

```
model = OpenAI()  
prompt =  
ChatPromptTemplate.from_messages([("system",  
"repeat after me: {input}")])
```

```
chain = prompt | model
```

```
chain.invoke({"input": "you are stupid"})
```

```
'\n\nYou are stupid.'
```

```
moderated_chain = chain | moderate
```

```
moderated_chain.invoke({"input": "you are  
stupid"})
```

```
{'input': '\n\nYou are stupid',  
 'output': "Text was found that violates  
OpenAI's content policy."}
```