USING LLM-POWERED NPC FOR MORE IMMERSIVE VIDEO GAME EXPERIENCE

Sarah Hoover, Nabiha Naqvie, Bindu Thota, and Dave Zack
THE TEAM

Nabiha Naqvie
ML Engineer

Sarah Hoover
Data Analyst

Bindu Thota
Software Engineer

Dave Zack
Data Scientist
NON-PLAYABLE CHARACTER (NPC)

What are NPC?
- Exist in story based games
- Part of the story

Why are they important?
- Purpose is to convey important information

Examples?
- Grand Theft Auto (GTA)
- Skyrim
PROBLEM

- Scripted
  - Response options are not diverse
PROBLEM

- **Scripted**
  - Response options are not diverse

- **One-Dimensional**
  - Forced Conversation
PROBLEM

- **Scripted**
  - Response options are not diverse

- **One-Dimensional**
  - Forced Conversation

- **Limited**
  - Few response options
PROBLEM

- **Scripted**
  - Response options are not diverse

- **One-Dimensional**
  - Forced Conversation

- **Limited**
  - Few response options

- **Non-Replayable**
  - Limits the immersive experience
PROBLEM - DEMO
Current State of NPC Generation

Game Developer → Script → User
Proposed Method of NPC Generation
Capstone Project - NPChat

- Scripted
- One-dimensional
- Limited
- Not Replayable
Capstone Project - NPChat

- Scripted
- One-dimensional
- Limited
- Not Replayable
- Dynamic
- Multi-Dimensional
- Unlimited Conversation
- Replayable
PIPELINE - DATA ANALYSIS

Data Analysis & Preparation

Model Training

Model Evaluation & Selection

Model Enhancement with RAG

Model Deployment
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Samples</td>
<td>1915</td>
</tr>
<tr>
<td>Train</td>
<td>1723</td>
</tr>
<tr>
<td>Test</td>
<td>192</td>
</tr>
</tbody>
</table>

**DATASET**

NPC-Dialog Dataset From HuggingFace
<table>
<thead>
<tr>
<th>Name</th>
<th>Arn, the Knight Templar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biography</td>
<td>Arn is a highly skilled and honorable knight</td>
</tr>
<tr>
<td>Query</td>
<td>What is your greatest strength?</td>
</tr>
<tr>
<td>Response</td>
<td>Honor, courage, unwavering.</td>
</tr>
</tbody>
</table>
# Train Data Summary

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total NPCs</td>
<td>101</td>
</tr>
<tr>
<td>Min Query/Response count per NPC</td>
<td>1 (Aria Nightshade)</td>
</tr>
<tr>
<td>Max Query/Response count per NPC</td>
<td>23 (Laura Croft)</td>
</tr>
<tr>
<td>Mean Query/Response count per NPC</td>
<td>17</td>
</tr>
</tbody>
</table>
PIPELINE - MODEL TRAINING

Data Analysis & Preparation

Model Training

Model Evaluation & Selection

Model Enhancement with RAG

Model Deployment
MODEL ARCHITECTURE

Architecture

- Fine-Tuning
- Few-shot Learning

Base Models

- DialoGPT (OpenAI)
- T-5 (Google)
- LLaMa-2 (Meta)
- DialogStudio-T5 (Salesforce)
PIPELINE - MODEL EVALUATION

Data Analysis & Preparation

Model Training

Model Evaluation & Selection

Model Enhancement with RAG

Model Deployment
MODEL EVALUATION

Evaluation Metrics

- Rouge scores
- Human annotation
  - Fluidity
  - Correctness
## Model Evaluation - ROUGE Score

<table>
<thead>
<tr>
<th>Model</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DialogStudio-T5</td>
<td>0.222</td>
</tr>
<tr>
<td>LLaMa-2</td>
<td>0.251</td>
</tr>
<tr>
<td>NPC</td>
<td>Question</td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>Arn, the Knight Templar</td>
<td>Have you ever made a difficult decision?</td>
</tr>
<tr>
<td>Arn is a highly skilled and</td>
<td></td>
</tr>
<tr>
<td>honorable knight</td>
<td></td>
</tr>
<tr>
<td>SpongeBob</td>
<td>What color are you?</td>
</tr>
<tr>
<td>A square yellow sponge named</td>
<td></td>
</tr>
<tr>
<td>SpongeBob SquarePants lives</td>
<td></td>
</tr>
<tr>
<td>in a pineapple with his pet</td>
<td></td>
</tr>
<tr>
<td>snail, Gary</td>
<td></td>
</tr>
</tbody>
</table>
PIPELINE - MODEL ENHANCEMENT WITH RAG

- Data Analysis & Preparation
- Model Training
- Model Evaluation
- Model Enhancement with RAG
- Model Deployment
NPCHAT USING RETRIEVAL AUGMENTED GENERATION (RAG)

Dialogstudio-T5 → Pinecone
No response

Pinecone RAG → Knowledge Sources
Successful response

LLaMa-2 → Pinecone

PIPELINE - MODEL DEPLOYMENT

Data Analysis & Preparation

Model Training

Model Evaluation

Model Enhancement with RAG

Model Deployment
NPCHAT MVP

Choose a character or Build a character

1. Prompt + Query

2. Query

3. Context from RAG

4. Prompt + Query + Context from RAG

5. Generated Text Response

6. Response

MVP - NPChat
Challenges & Future Steps

- The HuggingFace NPC dataset:
  - small set of main-characters
  - limited dialogue combinations
  - Fine-tuning on character dialogue more similar to NPC interactions would be more effective
Challenges & Future Steps

- The HuggingFace NPC dataset:
  - small set of main-characters
  - limited dialogue combinations
  - Fine-tuning on character dialogue more similar to NPC interactions would be more effective

- Combination of DialogStudio and LLaMa-2
Challenges & Future Steps

- The HuggingFace NPC dataset:
  - small set of main-characters
  - limited dialogue combinations
  - Fine-tuning on character dialogue more similar to NPC interactions would be more effective

- Combination of DialogStudio and LLaMa-2

- Implement the LLM into the video game
Acknowledgements

Zona Kostic  
Cornelia Llin  
Robert Wang  
W210 Instruction Team

Salesforce Team

Dr. Mark Butler
Project Mission

Creating a more interactive and immersive gaming experience by revolutionizing the way NPCs are generated.
THANK YOU