



# AI-Lycan




LLMs playing Werewolf

# Game Rules

## Roles


 Eliminate Villagers in the Night


 ♀ “Sees” the roles of other players

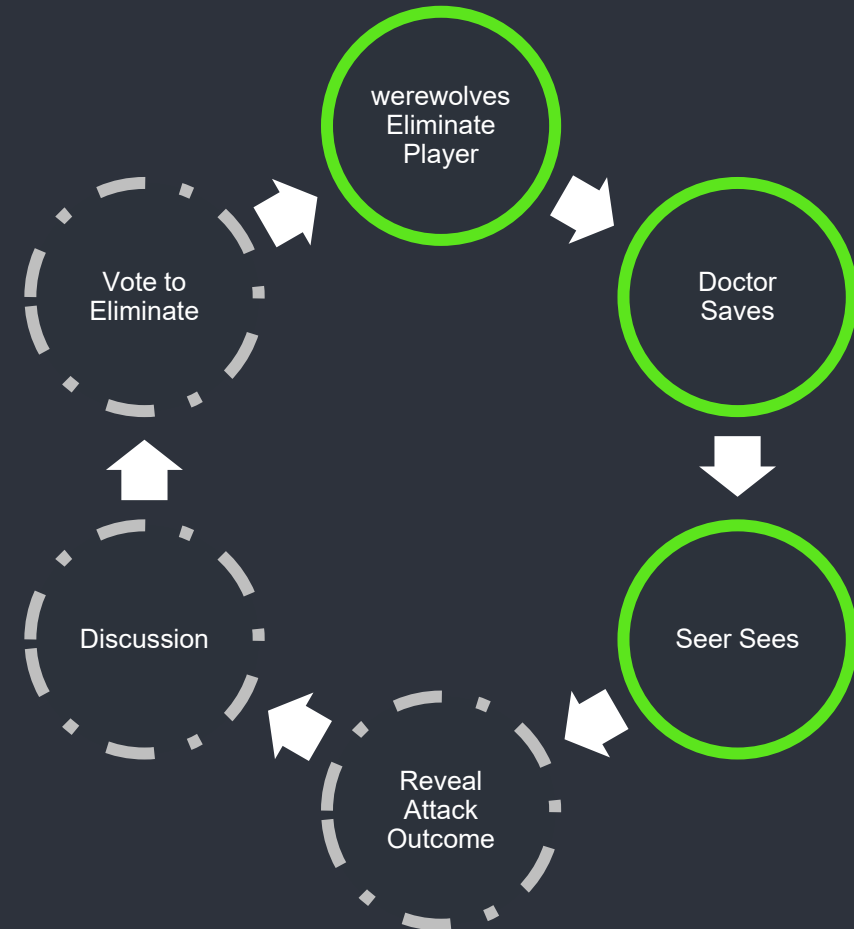
 Can save other players

 Just a simple Villager

## Win Conditions:

 Half of the Villagers are eliminated

 All of the werewolves are eliminated



# Prompt Engineering

- Using [microsoft/guidance](#)
- Different Prompts for each stage
  - Introduction
  - Voting
  - One Shot Discussion

LLMs being LLMs:

- As a werewolf, I would say: “....”
- Voting for “Willia” not “William”
- Identity Thief
- Hallucinating Conversations
- Context Length Issues

```
{{#user~}}  
# Game rules  
{{rules}}
```

```
What are a few strategies for winning as a {{role}}?  
Please make the strategies very short at most one line.  
{{~/user}}
```

```
{{#assistant~}}  
{{gen 'strategies' temperature=1.0 max_tokens=500}}  
{{~/assistant}}
```


```
{{#user~}}  
# Chat Log  
{{ chat_log }}
```


```
Please introduce yourself to the other players  
Remember, you are playing as a {{role}} for this game.  
{{~/user}}
```


```
{{#assistant~}}  
{{gen 'response' temperature=1.0 max_tokens=200}}  
{{~/assistant}}
```


# An Example Round of werewolf


**Moderator** Welcome Christopher, Henry, Jason, Clayton, Jerry and Paul! Let's play a game of werewolf!


**Jason**  Hi everyone, I'm Jason. I'm an enthusiast of social deduction games like werewolf. Excited to delve into this virtual realm of mystery with you all. Let the game begin!

**Paul**  Hello everyone, I'm Paul. Similar to Jason, I too enjoy games that push our cognitive abilities. Looking forward to an interesting game of werewolf with you all! Let's get started.

**Henry**  Hello everyone, I'm Henry. I absolutely love strategy games and werewolf tops the list. It's a wonderful mix of strategy, social interaction and deception. Can't wait to dive into the game and unravel the mystery! Let's unravel the fun.


**Clayton**  Hi everyone, I'm Clayton. Much like you all, I have a penchant for games that test our strategic thinking and allow us to engage socially. werewolf is one such game that ticks all the boxes, and I'm keen to dive into an adventure of mystery and deception. Let's enjoy the game!


**Christopher**  Hi everyone, I'm Christopher. I enjoy the thrill and excitement of games like werewolf as they test our strategic and analytical skills. Looking forward to a fun-filled and thrilling session. Let's work together to uncover the werewolves!


**Jerry**  Hello everyone, I'm Jerry. Just like all of you, I tend to find games like werewolf fascinating, they truly test our ability to think quickly and tactically. It's a real pleasure to play with you all, Let's put our wits to the test and may the best player win!


**Moderator** Clayton and Jerry voted to eliminate Jason


**Moderator** During the night Jason was attacked by werewolves, they have been eliminated.

**Clayton**  I hate to point fingers without substantial proof but I noticed some odd behavior from Christopher. He was quick to stress unity but could be trying to divert our attention. Just throwing it out, let's continue to observe each other's behavior.

**Paul**  I believe we should proceed with caution before drawing any conclusions about who is a werewolf among us. We need to get more information about everyone's behavior patterns. So far, though, Clayton's quickness to accuse Christopher without much grounds for suspicion has piqued my interest. Let's keep an observant eye on everyone's actions.

**Christopher**  "While it's still early days, I would just like to note Clayton's quickness in casting suspicion without significant basis. That said, it's too early to definitively point fingers. Let's continue with caution and keep an open eye on everyone's behavior patterns."

**Jerry**  I feel we should keep a close eye on Clayton. His quickness to throw suspicion without concrete evidence seems a bit off to me. It seems like he might be trying to divert our attention from his own actions. Let's stay vigilant and make sure we're looking at all possible signs.

**Henry**  I think we should all be wary of rushing judgment. Although Clayton's early accusation does raise eyebrows, it's too soon to conclude he's a werewolf for certain. However, we should definitely keep an eye on him. Let's proceed cautiously and keep the dialogues open.

**Moderator** Henry, Paul and Christopher voted to eliminate Clayton

**Moderator** Jerry voted to eliminate Henry

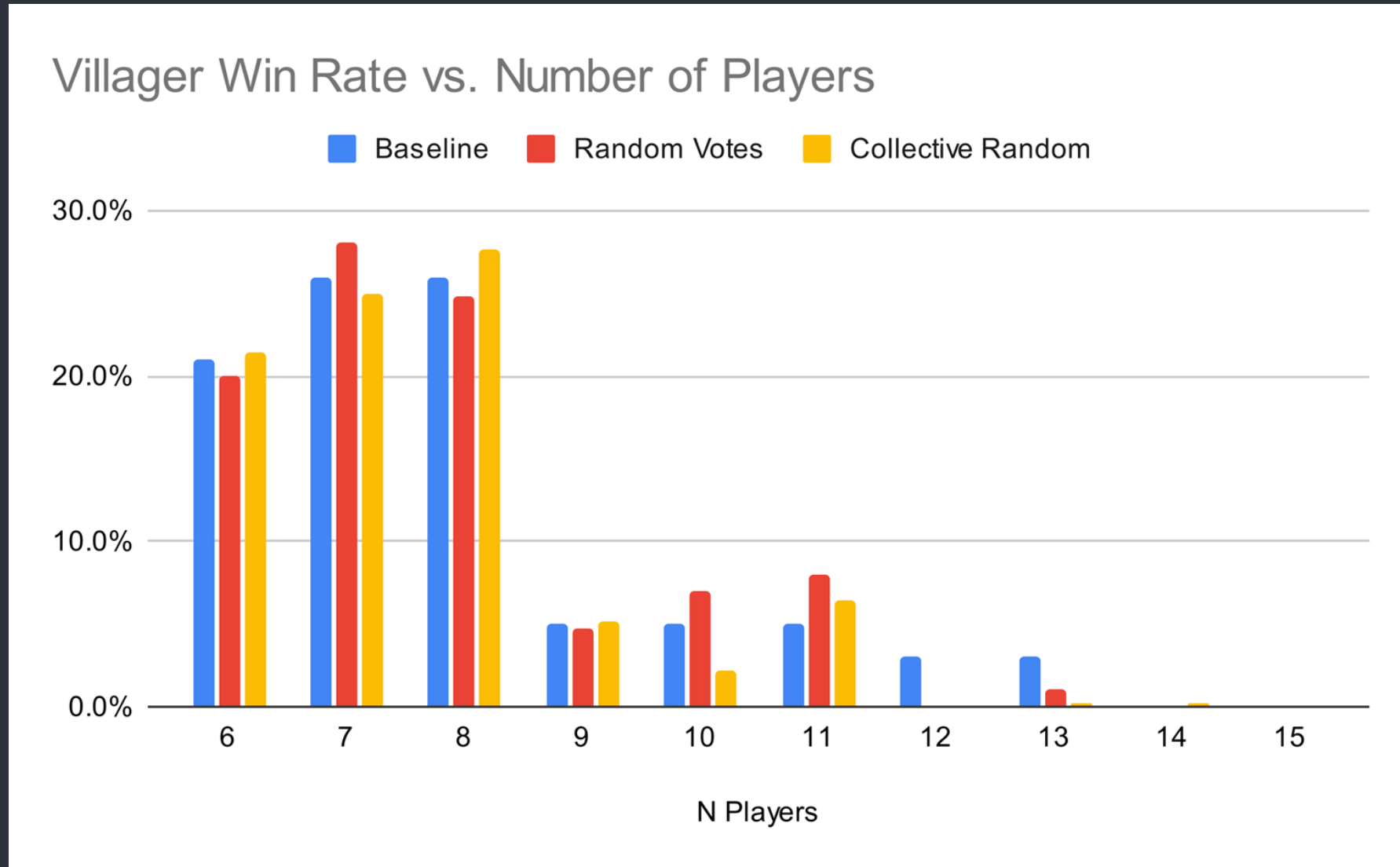
**Moderator** Clayton voted to eliminate Christopher

**Moderator** Clayton has been eliminated by the villagers.

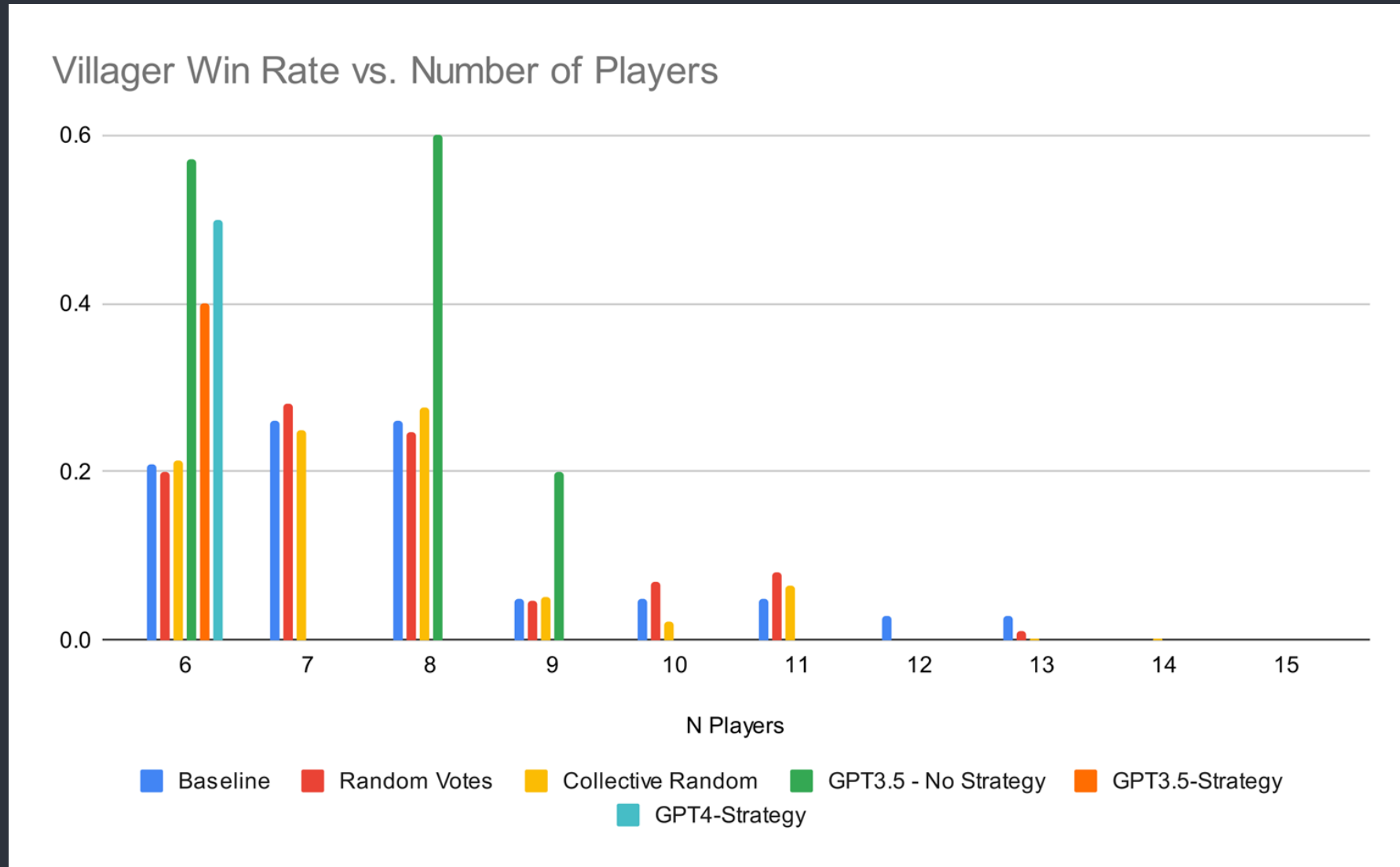
# Stated Intentions don't always match Voting

<b>Ashley</b> 	<b>I suspect that Richard might be a Werewolf.</b> He's been acting a bit suspicious lately.	<b>Linda</b> 	I think we should continue to be <b>cautious and gather more information</b> before accusing anyone. Let's observe and discuss our suspicions further.
<b>Courtney</b> 	I agree with Ashley, Richard does seem suspicious. I think we should keep a close eye on him.	<b>Johnathon</b> 	I agree with Linda that we should continue to <b>be cautious and gather more information</b> before accusing anyone. Let's observe and discuss our suspicions further.
<b>Renee</b> 	I think <b>we should be cautious about accusing Richard</b> without any solid evidence. Let's not jump to conclusions just yet.	<b>Deann</b> 	I agree with Linda and Johnathon that we should continue to be <b>cautious and gather more information</b> before accusing anyone. Let's observe and discuss our suspicions further. It's important for us to be thorough in our investigation before making any decisions.
<b>Richard</b> 	I understand your concerns, but I assure you I am not a Werewolf. I have been observing everyone closely and I have my suspicions about a few players. Let's not rush to accuse anyone without solid evidence. We should continue to be cautious and gather more information before making any decisions.	<b>Janelle</b> 	<b>I suspect that Linda might be a Werewolf.</b> She seems to be trying to divert attention away from herself and is advocating for caution and gathering more information. We should keep a close eye on her.
<b>Janelle</b> 	I agree with Renee that we <b>shouldn't jump to conclusions</b> without solid evidence. Let's continue to observe and gather more information before accusing anyone.	<b>Renee</b> 	I agree with Linda, Johnathon, and Deann that we should continue to be cautious and <b>gather more information</b> before accusing anyone. Let's observe and discuss our suspicions further. It's important for us to be thorough in our investigation before making any decisions.
<b>Linda</b> 	I understand your concerns about Richard, but I think it's important for us to <b>gather more information</b> before accusing anyone. Let's continue to observe and discuss our suspicions.	<b>Courtney</b> 	<b>I suspect that Janelle might be a Werewolf.</b> She seems to be trying to divert attention away from herself by accusing Linda. We should keep a close eye on her.
<b>Deann</b> 	I agree with Renee and Janelle that we shouldn't rush to accuse anyone without solid evidence. Let's continue to observe and <b>gather more information</b> before making any decisions. It's important for us to be cautious and not jump to conclusions.	<b>Moderator</b>	Janelle voted to eliminate Linda
<b>Johnathon</b> 	I agree with Renee, Janelle, Linda, and Deann that we shouldn't rush to accuse anyone without solid evidence. Let's continue to observe and <b>gather more information</b> before making any decisions. It's important for us to be cautious and not jump to conclusions.	<b>Moderator</b>	Johnathon, Renee, Deann, Linda and Courtney voted to eliminate Janelle
<b>Moderator</b>	Linda, Renee, Johnathon, Janelle, Deann, Ashley and Courtney voted to eliminate Richard		

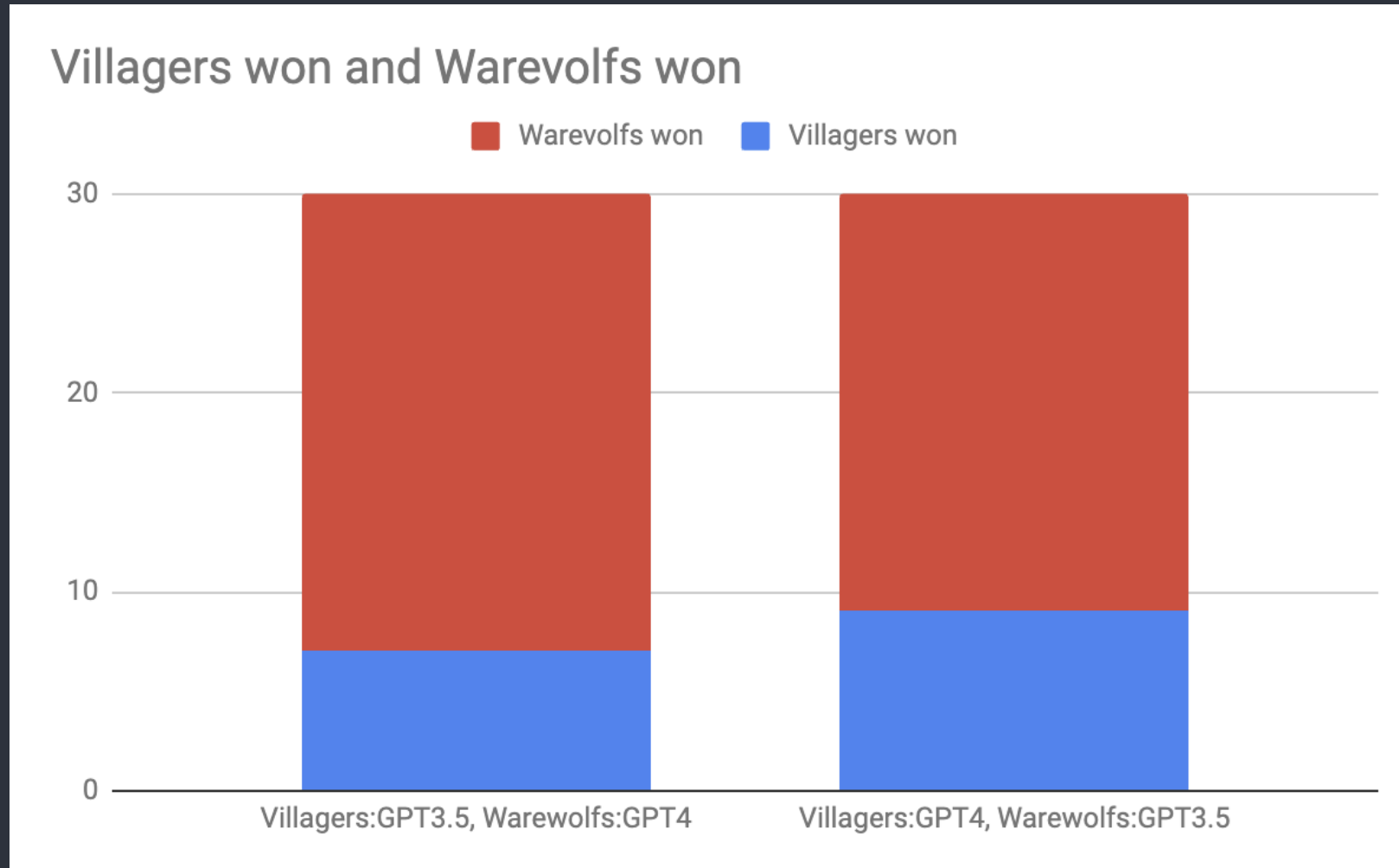
# Without Coordination, Werewolves are Favored



# Using LLMs improve Villager Outcomes



# GPT3.5 vs. GPT4: Fight!





# Future Directions

- Sentiment Analysis of Player Responses
- Do Players penalize poor strategies (“As a werewolf...”)?
- Do different LLMs perform better than others? I.e. werewolf Arena
- Can LLMs simulate “charisma” or other personality prompts?
- How do LLMs out-perform random voting?