



# **An Investigation into Game of Thrones Dialogue**

Payson Chadrow

## PURPOSE

Identify important characters within individual episodes, seasons, and overall.  
Determine character sentiment and classify dialogue by character.



# TABLE OF CONTENTS

01

THE DATA

02

INITIAL FINDINGS

03

SENTIMENT &  
CLASSIFICATION

04

CONCLUSIONS

05

FUTURE WORK

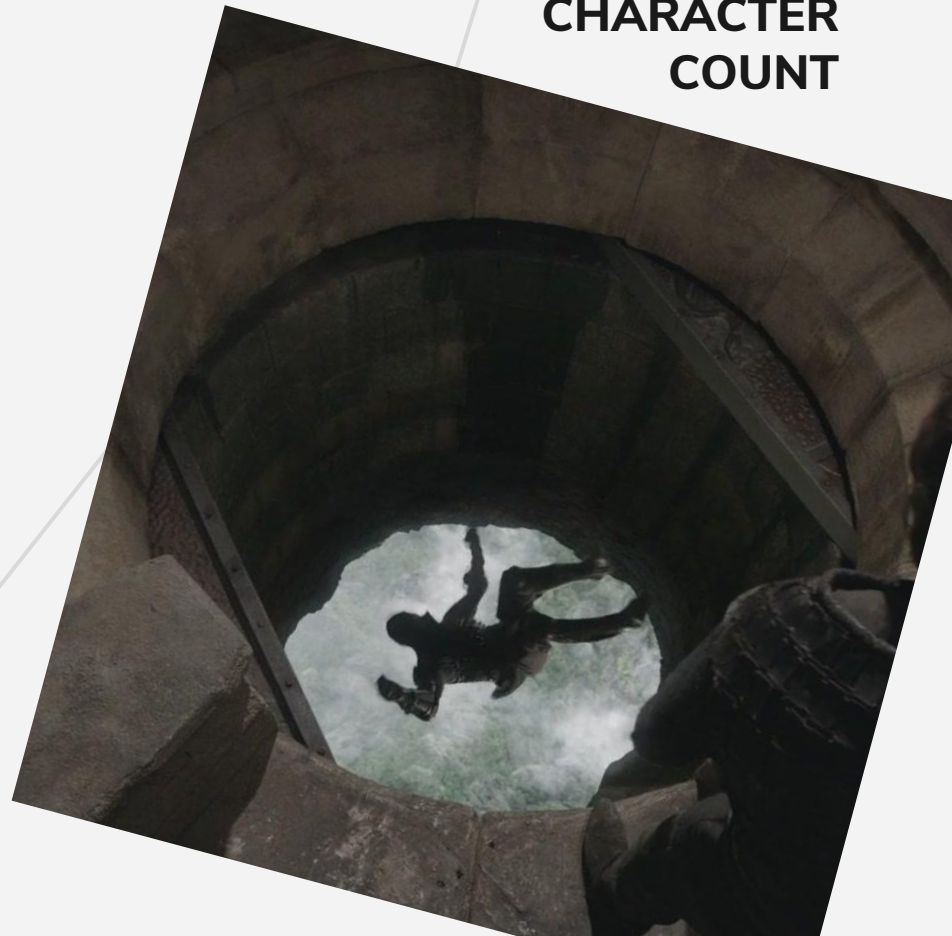
**01**

THE DATA



## DROPPING CHARACTER COUNT

- 564 UNIQUE CHARACTERS
- AT LEAST 44 VARIANTS
- 51 “MEANINGFUL” CHARACTERS
- 16 CLASSIFICATION CHARACTERS







# 02

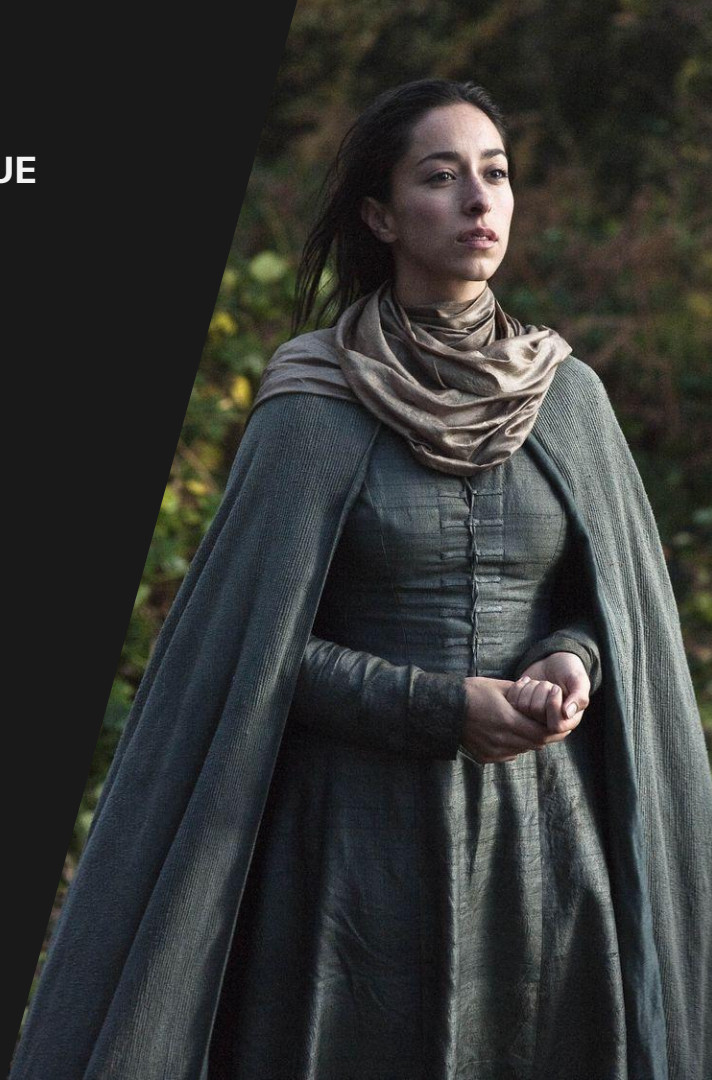
## INITIAL FINDINGS

INTERACTIVE DATA  
INVESTIGATION

<https://got-character-lines-plot.herokuapp.com/>



## LONGEST MONOLOGUE



## TALISA STARK

- SEASON 2 EPISODE 8
- 333 WORDS



## MOST SPOKEN CHARACTER



### TYRION LANNISTER

- ~ 600 MORE LINES
- ~ 11,000 MORE WORDS
- SEASON 7 ONLY EXCEPTION
- OVER 20% OF THE DIALOGUE IN THE FINAL SEASON
- OVER 40% IN FINAL EPISODE

# 03

## SENTIMENT & CLASSIFICATION



## HEALTHY SENTIMENT SKEPTICISM

```
sentiment_analyzer_scores(df['Sentence'][10257])
```

executed in 4ms, finished 14:04:39 2020-08-28

```
On my honor as a Tully, on my honor as a Stark, let him go or I will cut your wife's throat! {'neg':  
0.085, 'neu': 0.645, 'pos': 0.27, 'compound': 0.68}
```

### Compound Score

**1 = very positive**

**0 = neutral**

**-1 = very negative**

## MOST POSITIVE CHARACTERS



## VARYS/ PETYR BAEISH

- BOTH ALSO HAVE THE HIGHEST AVG WORD COUNT PER LINE
  - VARYS - 16.8
  - PETYR BAEISH - 15.9

**MOST NEGATIVE  
CHARACTER**



**SANDOR CLEGANE**

- MOST NEGATIVE AVG  
SCORE

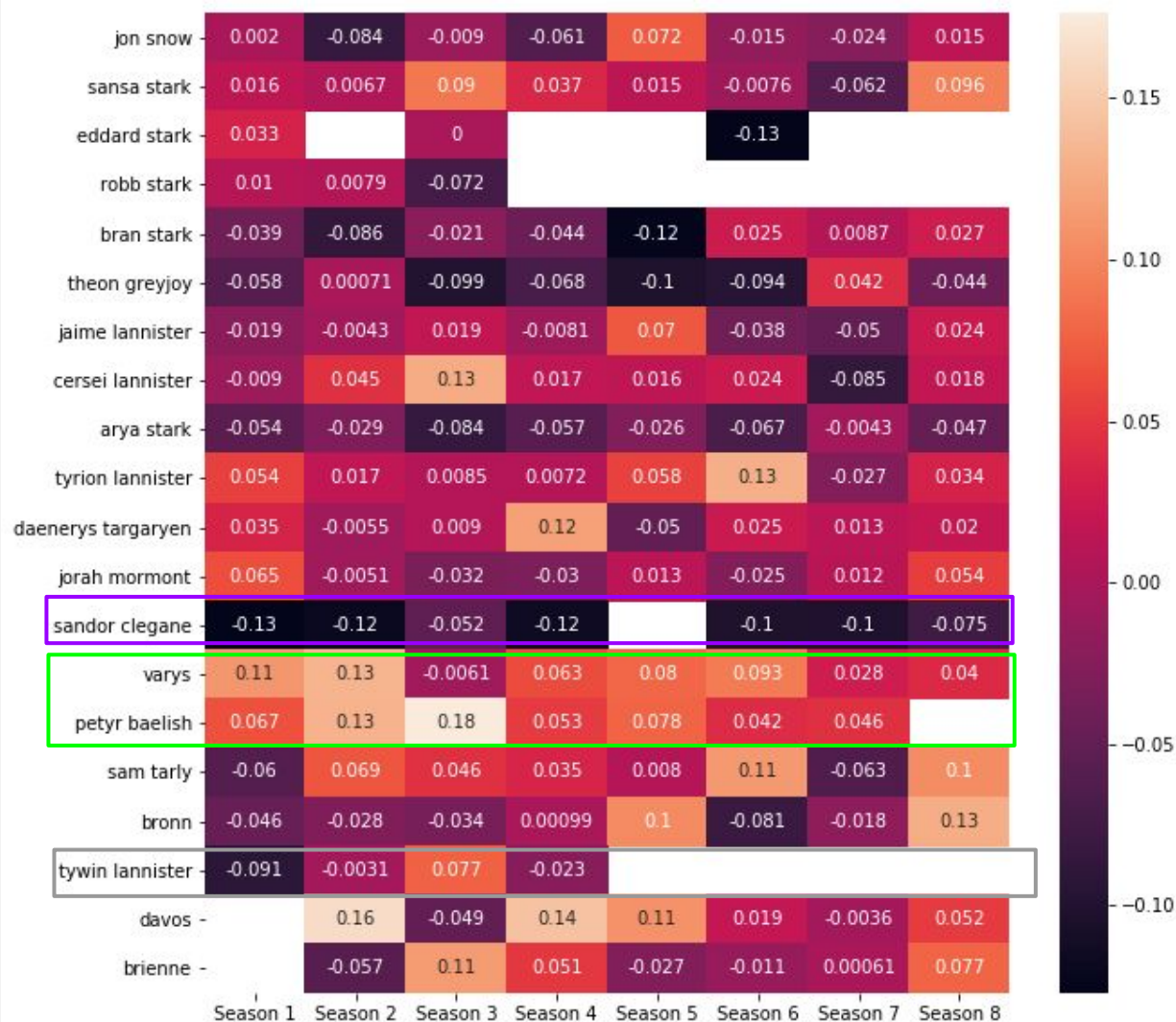


**MOST NEUTRAL  
CHARACTER**



**TYWIN LANNISTER**

- AVG SENTIMENT  
ALMOST PERFECTLY  
ZERO



# CLASSIFICATION PROCESS & PROBLEMS

MANY TO FEW

INSUFFICIENT  
DATA



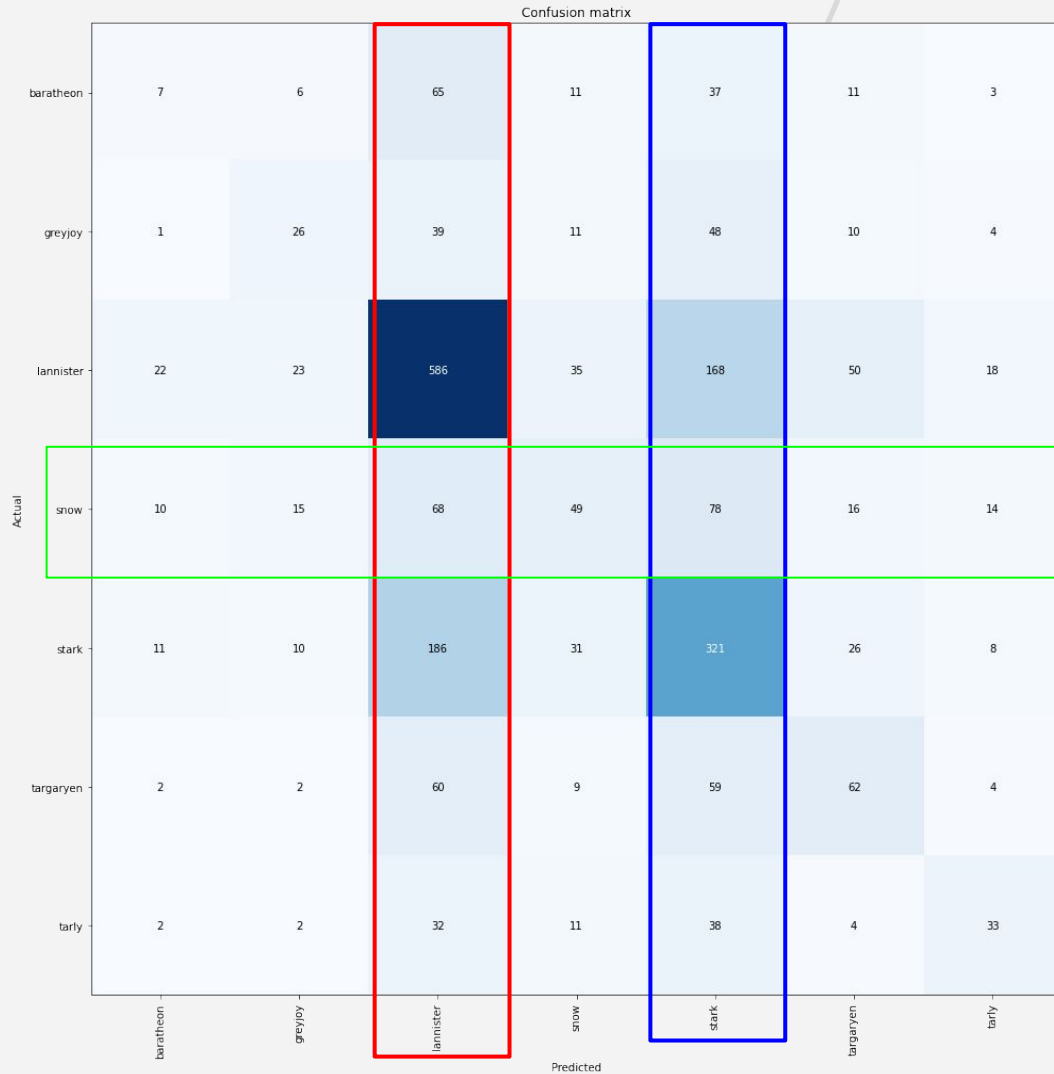
# CHARACTER CLASSIFICATION

- Just shy of 25% total accuracy

Actual \ Predicted																
	arya stark	bran stark	brienne	bronn	cersei lannister	daenerys targaryen	davos	jaimie lannister	jon snow	jorah mormont	sam tarly	sandor clegane	sansa stark	theon greyjoy	tyrion lannister	varys
arya stark	44	6	2	2	5	7	4	5	10	1	2	4	19	0	19	0
bran stark	6	12	1	0	0	4	0	1	11	0	4	0	11	1	5	0
brienne	6	0	12	2	0	7	2	7	2	0	0	4	8	1	11	0
bronn	10	0	1	8	6	5	3	11	5	0	5	8	9	1	18	0
cersei lannister	7	1	4	1	48	16	6	27	17	2	2	0	17	4	54	3
daenerys targaryen	19	1	1	3	8	58	1	7	15	8	2	0	13	1	34	3
davos	7	1	3	2	1	8	12	6	16	1	1	3	9	2	19	2
jaimie lannister	5	4	6	5	19	5	3	29	23	2	6	3	14	3	41	1
jon snow	15	4	2	1	6	20	8	20	58	5	11	3	22	9	22	1
jorah mormont	4	0	3	1	4	9	4	3	8	14	1	0	5	3	15	2
sam tarly	11	1	1	2	5	2	1	4	25	0	12	0	8	1	10	0
sandor clegane	17	1	3	4	2	2	3	6	8	2	0	11	3	0	7	0
sansa stark	16	5	4	0	8	6	5	7	11	0	9	0	55	5	13	2
theon greyjoy	7	6	2	3	2	5	3	7	17	2	3	1	12	17	8	0
tyrion lannister	15	0	8	3	19	22	9	35	32	3	7	6	28	5	118	13
varys	3	0	4	0	5	9	5	6	3	4	1	1	5	0	23	6

# FAMILY CLASSIFICATION

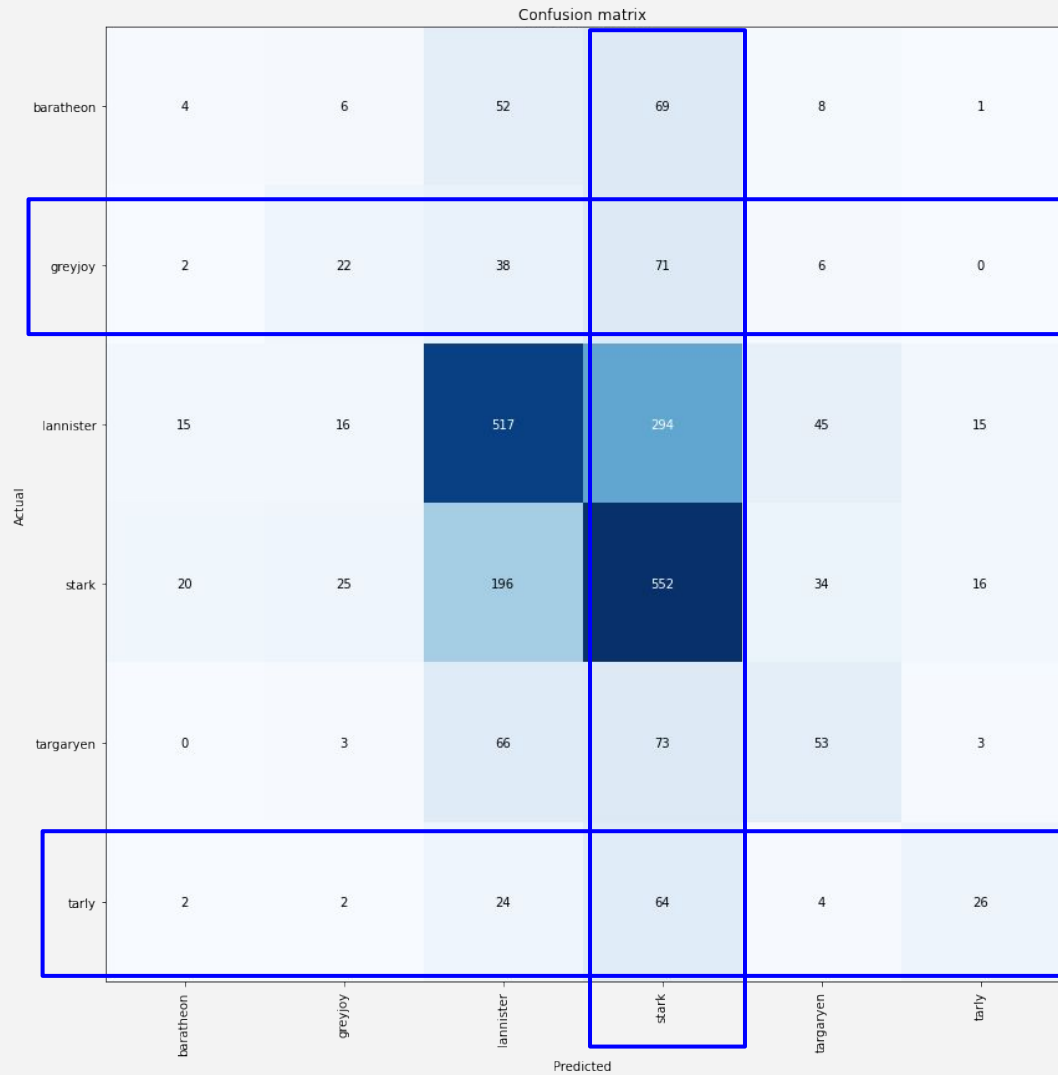
- Improved to around 46% total accuracy





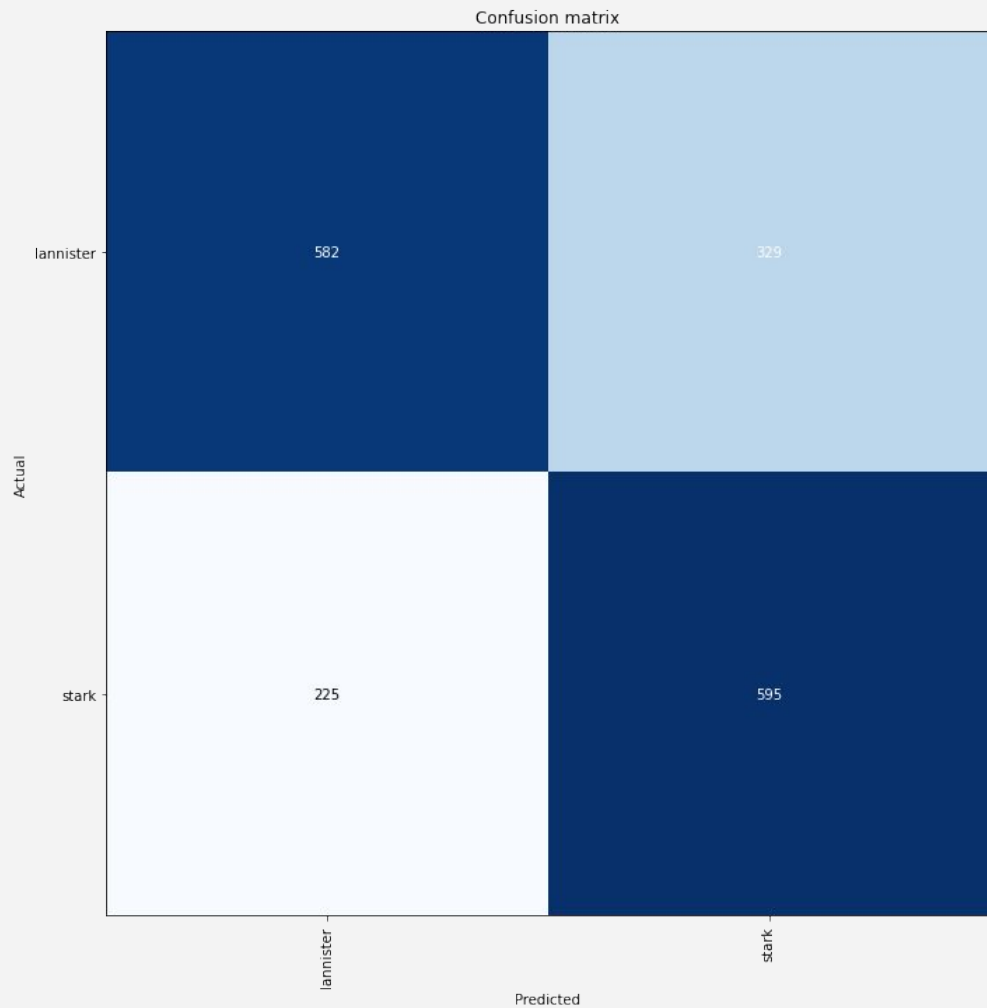
# FAMILY CLASSIFICATION (Snow as a Stark)

- Improved to around 50% total accuracy



# LANNISTER V. STARK

- Improved to around 68% total accuracy





# 04

## CONCLUSIONS

# TAKEAWAYS



- TYRION LANNISTER = SHOW
- SENTIMENT AND SETTING
- INSUFFICIENT DATA FOR PROPER CLASSIFICATION
- FAMILY CLASSIFICATION

**05**

**FUTURE WORK**





# FUTURE PLANS

REMOVING  
NON-DIALOGUE

1

MODEL FINE TUNING

3

TYRION CHAT-BOT

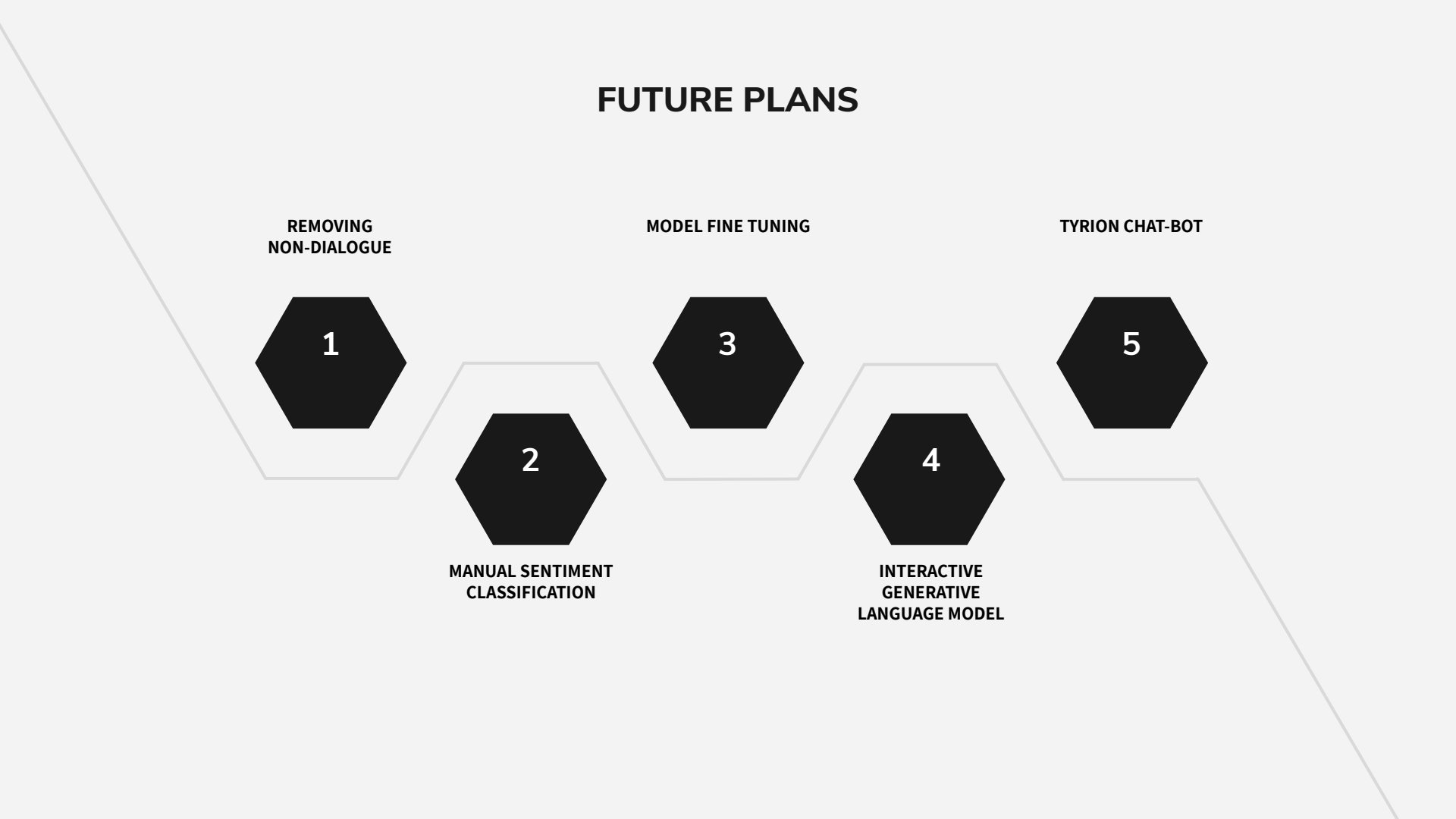
5

2

MANUAL SENTIMENT  
CLASSIFICATION

4

INTERACTIVE  
GENERATIVE  
LANGUAGE MODEL





**THANK YOU!**