







# AITA?

KRZYSZTOF KALISIAK, MICHAŁ KUNSTLER

# Scraping Reddit

The tool we have decided to use is prawn, an easy, quick and free reddit api.

 **r/AmItheAsshole** · Posted by u/Fun-Construction-112 15 hours ago   

6.7k

## AITA for criticizing how my girlfriend takes job interviews? She basically interviews them, and I feel like she isn't taking it seriously

My girlfriend is at a job she can't do remotely, and we're planning to move to another state together, so she's job hunting right now.

Her first interview, she had a call with a top company who's recruiter had messaged her on LinkedIn. I was expecting her to treat it normally, but she spent an hour grilling the company on its engineering practices then withdrew her application.

And the next few calls with companies she had, she basically grilled them all and decided against moving forward with four of the six.

I told her around then, that I feel like she's making a mistake, being so picky, and she's gonna ruin her reputation in the industry if she's going around taking interviews and cutting the process off early.

She said she wasn't making any enemies, hell, the companies she dropped had been emailing and calling constantly, wanting to bring her in for another interview or asking her to reconsider. If anything, she was a hotter commodity.

I felt like she was probably still hurting her reputation long term, even if her little power play was working for a bit.

She said it wasn't a power play, it was professional, she just didn't want to waste anyone's time.

But the next interview I overheard started a big argument. One of her final two companies had her taking a Zoom interview and she was laughing it up with an interviewer and he was telling her this story about how he and his coworkers fell off a barge into the river working on a project. And she just was like "waiiit they had y'all doing that, not tied off to anything? Look as funny as that is, that's honestly kind of fucked up they put y'all in danger like that - I'm honestly gonna have to withdraw my application"

She got off the phone and said "Damn, people really tell on themselves if you just listen and smile, did you hear that shit?" And I said that I thought she ended it a little prematurely, like didn't even ask if they'd changed anything there, just ended the call.

I said it felt like she was trying to delay getting a new job, was she getting cold feet or something?

She said no, this is literally how people at her level interview, she was serious about the interview process and she wasn't interested in walking into a shitshow.

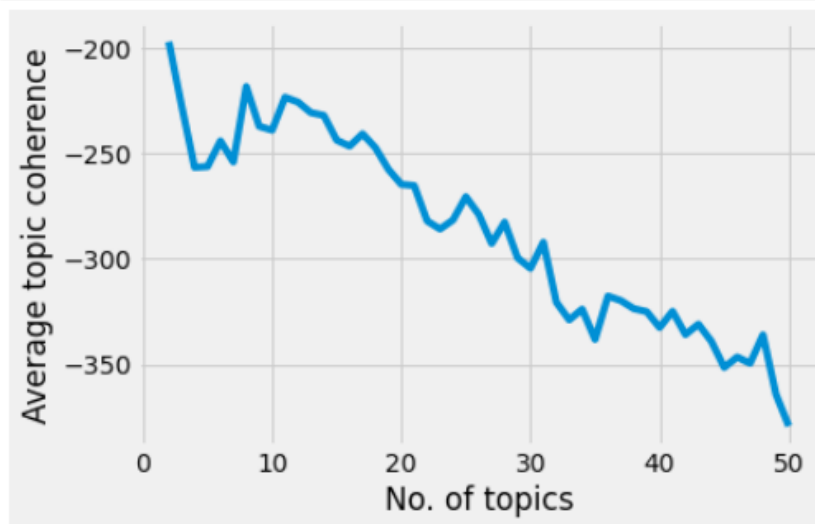
I said that was BS, she was sabotaging herself on purpose basically haranguing the companies who want to hire her on the phone. And she was like "why do they keep coming back for more then? Like I'm critical but I'm not wrong and they know it."

We had this big fight where she insisted that anyone who was at her level of a career "interviewed" by interviewing companies to see whether they were worth their time, just as much as the other way around, and I said that was BS. She got mad I was telling her about her own career and said she knew it better

AITA for arguing with my girlfriend about her interviews? I feel like she's dragging her feet, she says she's interviewing normally for her field

# The number of Topics

```
In [33]: plt.style.use("fivethirtyeight")
plt.plot(measures_specific_df_lda['n_topics'],measures_specific_df_lda['avg_coherence'])
plt.xlabel("No. of topics")
plt.ylabel("Average topic coherence")
plt.show()
```

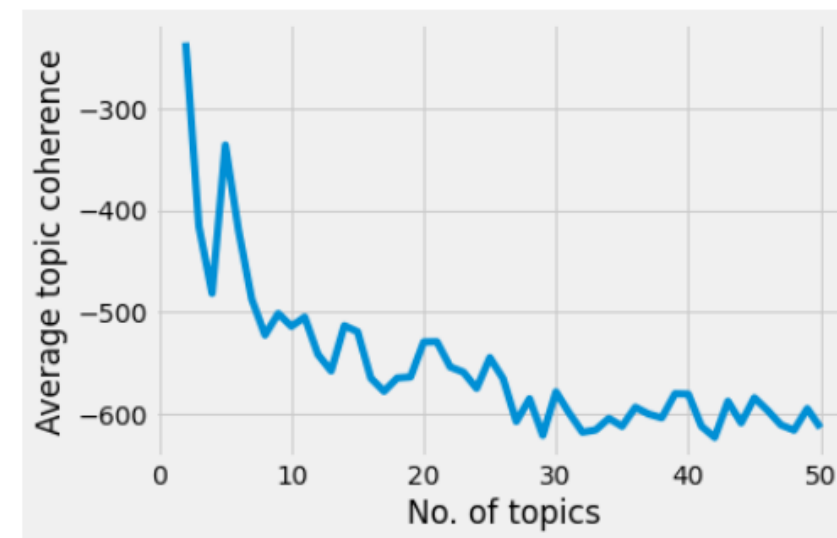


```
In [34]: measures_specific_df_lda.sort_values('avg_coherence', ascending = False).iloc[0:9,:]
```

```
Out[34]:
```

	avg_coherence	n_topics
0	-197.258110	2
6	-218.464319	8
9	-223.506711	11
10	-225.799363	12
1	-227.682847	3
11	-230.803568	13
12	-232.100892	14
7	-237.269628	9
8	-239.096575	10

```
In [38]: plt.style.use("fivethirtyeight")
plt.plot(measures_specific_tfidf_lda['n_topics'],measures_specific_tfidf_lda['avg_coherence'])
plt.xlabel("No. of topics")
plt.ylabel("Average topic coherence")
plt.show()
```



```
In [39]: measures_specific_tfidf_lda.sort_values('avg_coherence', ascending = False).iloc[0:9,:]
```

```
Out[39]:
```

	avg_coherence	n_topics
0	-235.322070	2
3	-335.975590	5
1	-415.732856	3
4	-419.810625	6
2	-481.555180	4
5	-487.373452	7
7	-501.306549	9
9	-504.806318	11
12	-513.255496	14

# The results

As usual, the topics do say something, but we can make it more interpretable...

```
: lda = LatentDirichletAllocation(n_components = 12, # let us stay with 12, as that is what topic coherence initially rec
                                doc_topic_prior = 0.05,
                                topic_word_prior = 0.100,
                                learning_method = 'online',
                                learning_offset = 10.0,
                                max_iter = 20,
                                random_state = 42)

lda.fit(tf) # TF for now

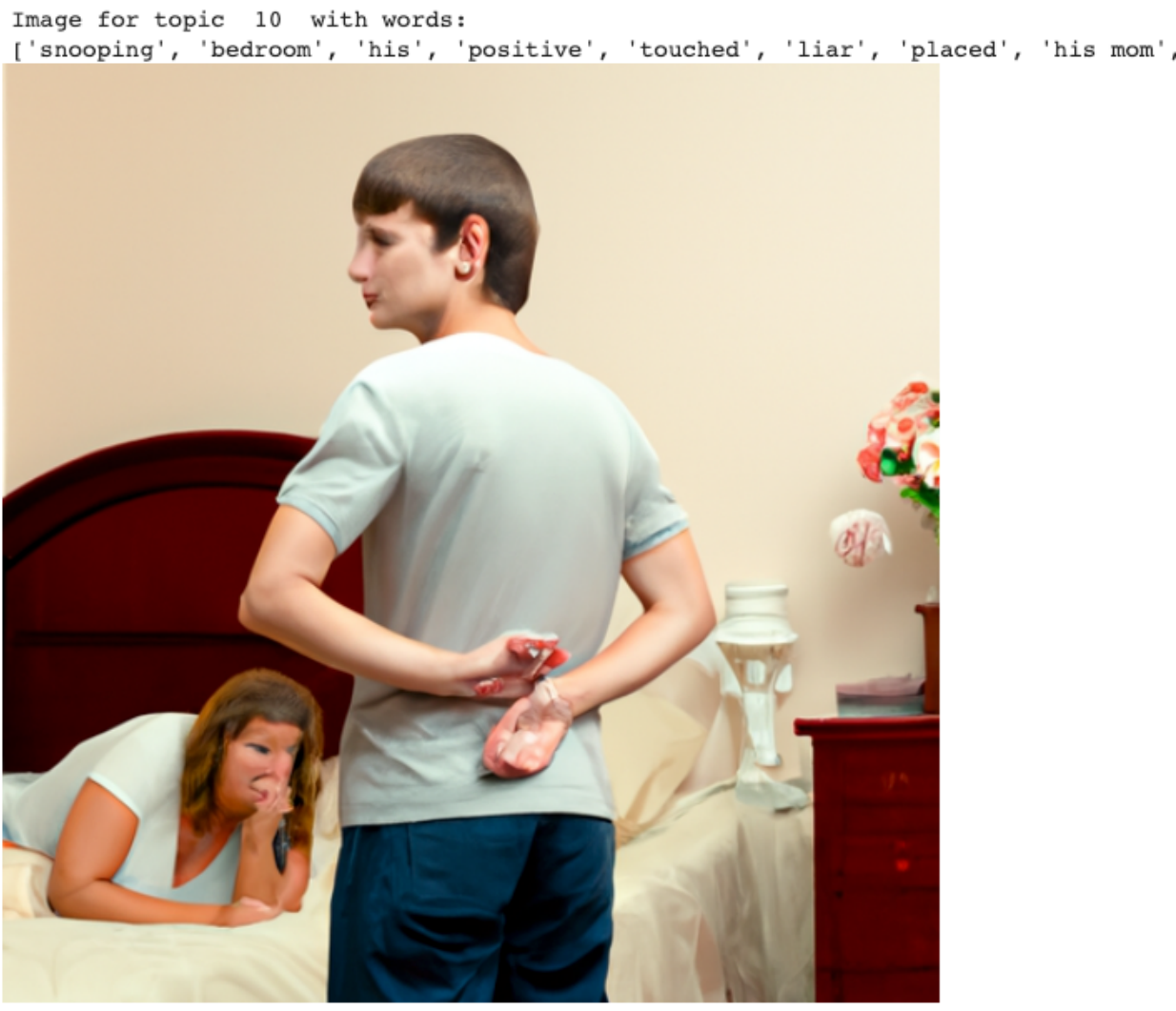
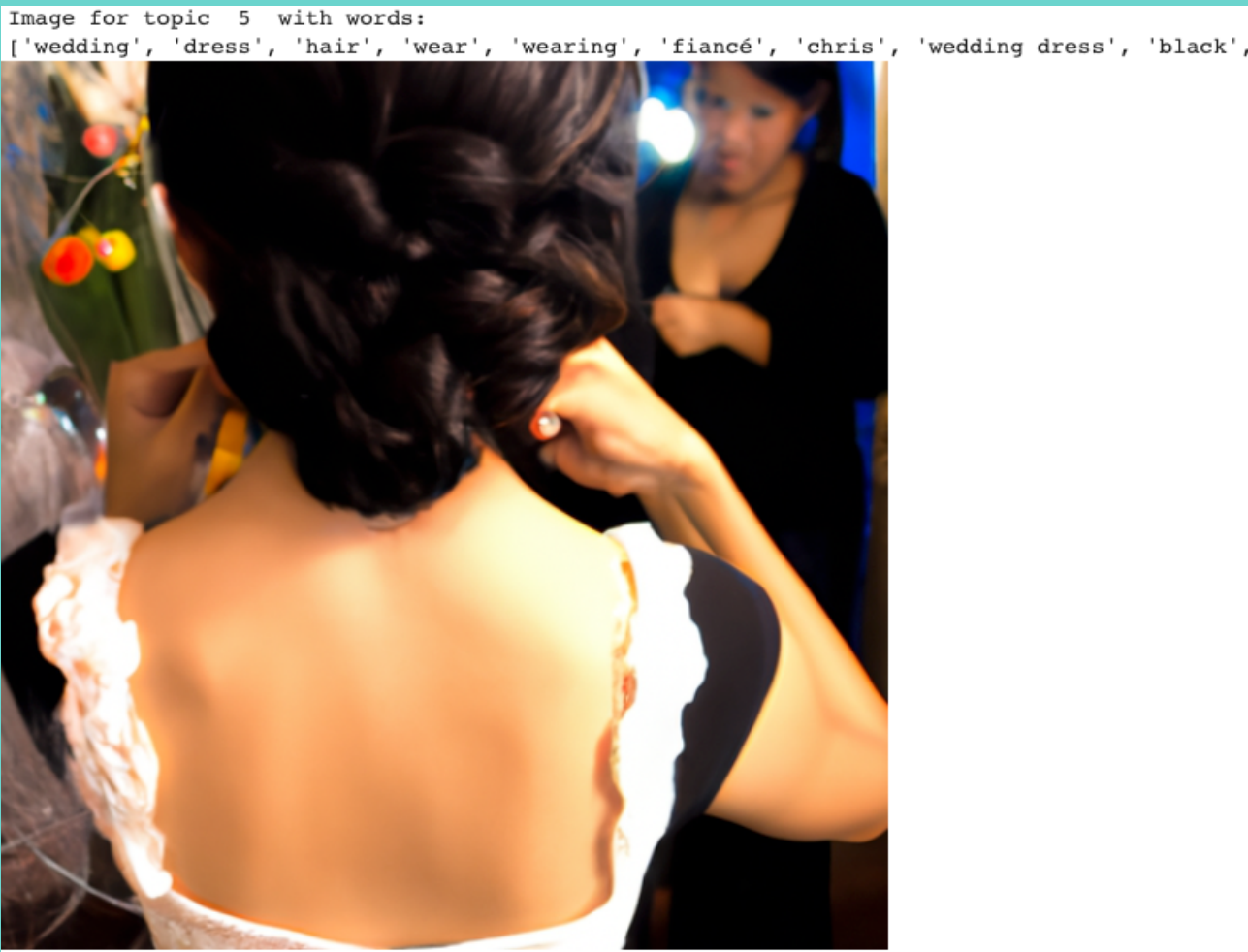
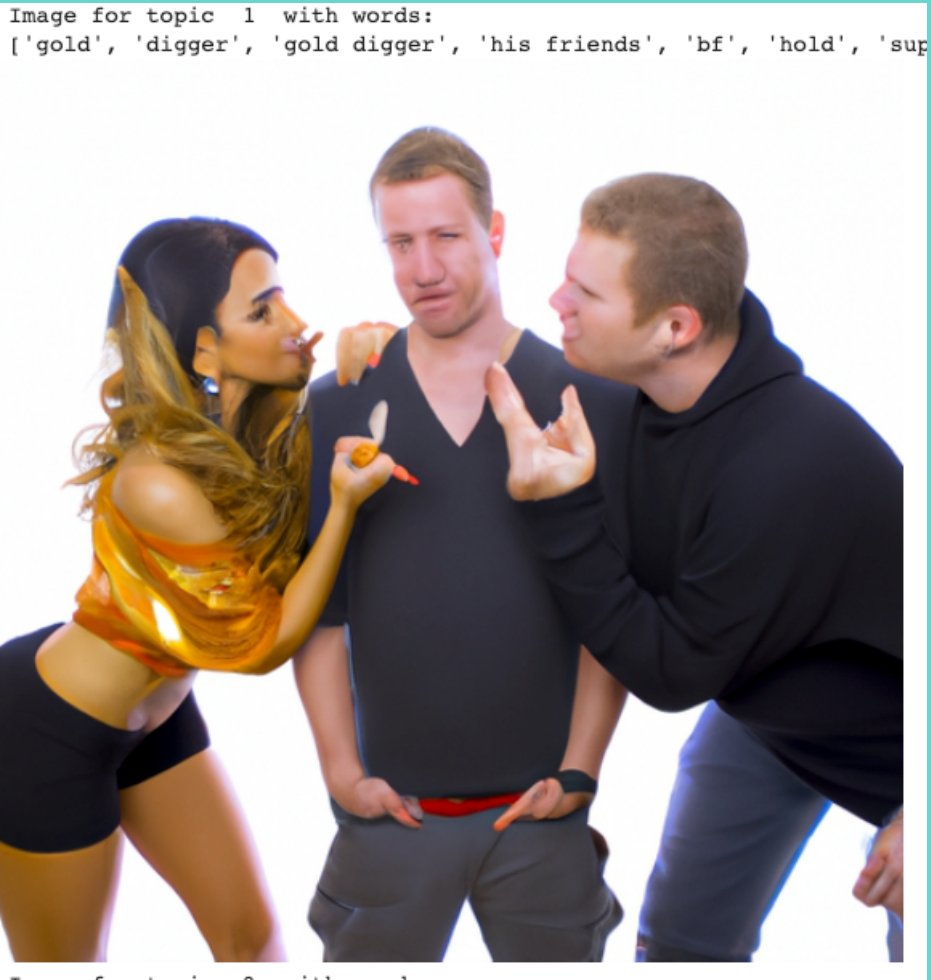
topics_lists = []

for index, component in enumerate(lda.components_): #taking model's components
                                                    #(values from reconstructed Document-Term Matrix)
    zipped = zip(tf_feature_names, component) #taking together tokens' names with components
    top_terms_key=sorted(zipped, key = lambda t: t[1], reverse=True)[:10] #top 10 terms per topic
    top_terms_list=list(dict(top_terms_key).keys()) #taking only tokens, no weights

    topics_lists.append(top_terms_list)
    print("Topic "+str(index)+": ",top_terms_list) #prints top 10 tokens per topic
```

```
Topic 0: ['his', 'im', 'because', 'like', 'get', 'time', 'has', 'got', 'would', 'asked']
Topic 1: ['gold', 'digger', 'gold digger', 'his friends', 'bf', 'hold', 'suppose', 'route', 'correctly', 'seconds']
Topic 2: ['his', 'im', 'because', 'like', 'get', 'got', 'its', 'even', 'one', 'before']
Topic 3: ['bride', 'tattoo', 'makeup', 'color', 'groom', 'order', 'party', 'brides', 'bridal', 'artist']
Topic 4: ['daughter', 'his', 'friend', 'daughter asked', 'got', 'asked', 'since', 'land', 'could', 'because']
Topic 5: ['wedding', 'dress', 'hair', 'wear', 'wearing', 'fiancé', 'chris', 'wedding dress', 'black', 'dresses']
Topic 6: ['his', 'got', 'husband', 'his mom', 'mom', 'like', 'get', 'home', 'because', 'would']
Topic 7: ['amy', 'ava', 'lisa', 'bill', 'italian', 'watch', 'sex', 'profile', 'grandfather', 'born']
Topic 8: ['his', 'family', 'mom', 'got', 'wife', 'dad', 'because', 'like', 'husband', 'im']
Topic 9: ['wig', 'utm_name', 'utm_medium', 'share utm_medium', 'iossmf', 'ios_app utm_name iossmf', 'utm_name iossmf', 'ios_app', 'ios_app utm_name', 'utm_medium ios_app']
Topic 10: ['snooping', 'bedroom', 'his', 'positive', 'touched', 'liar', 'placed', 'his mom', 'got one', 'said his mom']
Topic 11: ['miracle', 'his shit', 'asshole doing', 'disrespected his', 'being told', 'relationship said', 'chris', 'thought was funny', 'rushed', 'fiancee']
```





# Logistic regression based on topics

## TOPIC 5

HAS THE LEAST  
AS\*\*\*LES. ITS THE ONE  
RELATING TO BLACK  
DRESSES AND  
WEDDINGS

## TOPIC 8

HAS THE MOST  
AS\*\*\*LES. THIS IS THE  
ONE RELATED TO  
MARIAGE AND  
PARENTS.

```
array([[ -0.20932809, -0.14161635,  0.27066165, -0.09012881,  0.27066165,  
        -0.33453293,  0.27066165, -0.12957307,  0.72625572, -0.07534916,  
        0.27066165, -0.12008789]])
```

# MAIN ISSUES IN THE YEAR PAST.

TOPIC 0 OVER THE QUARTERS OF 2022

