

K Agajanian
June 13, 2022

TLAAS

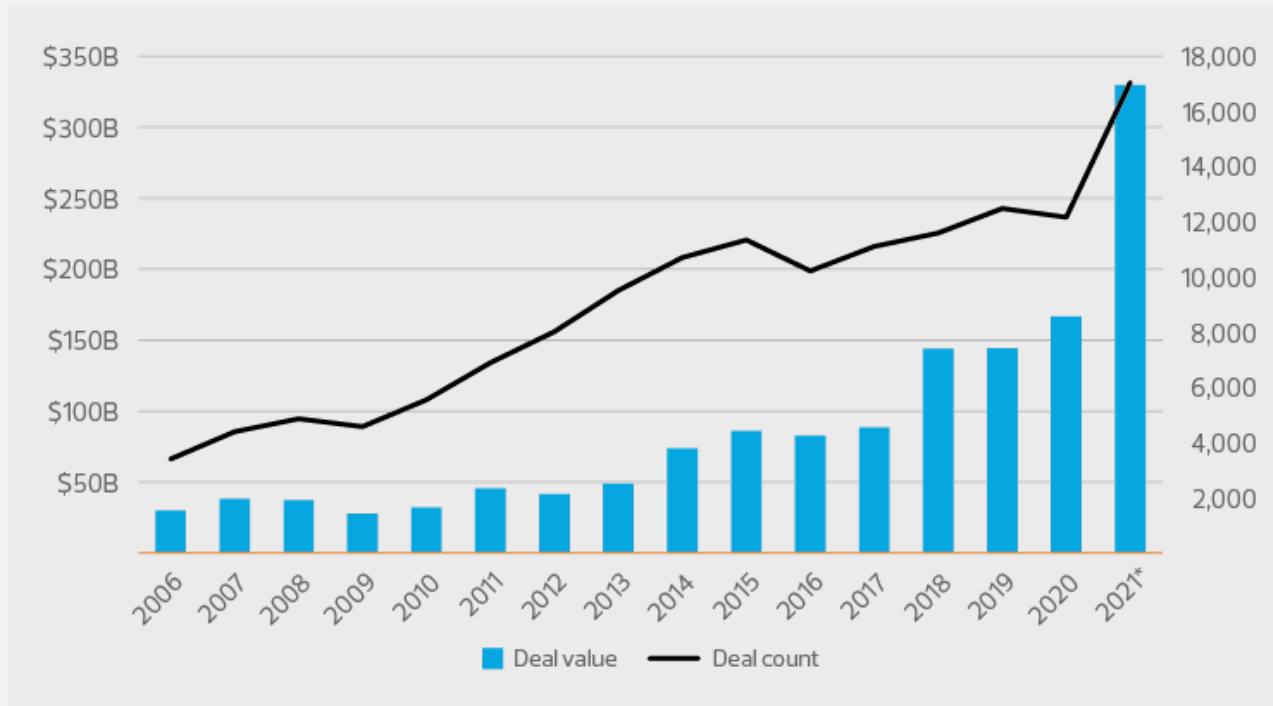
THOUGHT
LEADERSHIP
AS A SERVICE

TABLE OF CONTENTS

- Background:VC
- Problem Statement
- Twitter API Data Collection & Cleaning
- Model Training & Results
- Streamlit App

A BRIEF INTRO TO VENTURE CAPITAL (VC)

U.S. venture capital deal activity



Source: PitchBook

*As of September 2021

- Most popular & well-known private equity
- \$671 billion invested in 2021 worldwide in 38,644 deals ([KPMG](#))
- Estimated that 75% of startup investments don't return over 1x

VC STRUGGLES WITH EQUALITY

- In 2021, 2% of funding went to female founders (despite being >30% of startup founders)
- Even less went to founders of color



Bloomberg Equality

Subscribe


Equality

Female Founders Raised Just 2% of Venture Capital Money in 2021

By Lizette Chapman [+Follow](#)

January 11, 2022, 4:50 PM EST

VC INDUSTRY RELIES ON SIGNALING



Account ▾

The WeWork Manifesto: First, Office Space. Next, the World.


The brash, ambitious founders of WeWork, a global network of shared office spaces, want nothing less than to transform the way


NEW \$3B SOFTBANK INVESTMENT BOOSTS WEWORK VALUATION TO AT LEAST \$42B

2018/11/15 BY EMMA ZHOU — [LEAVE A COMMENT](#)

American flexible working space provider WeWork on Tuesday announced that its largest shareholder, SoftBank, has committed another \$3 billion to the company, boosting its valuation to at least \$42 billion.

Softbank is investing in WeWork









BUSINESS INSIDER

This Woman's Revolutionary Idea Made Her A Billionaire — And Could Change Medicine

Kevin Loria

October 4, 2014







SOCIAL
MEDIA
VISIBILITY
IS A MUST

Actual interview questions:

- “Who do you follow on Twitter?”
- “How do you provide thought leadership to the field?”

PROBLEM STATEMENT/USER NEED

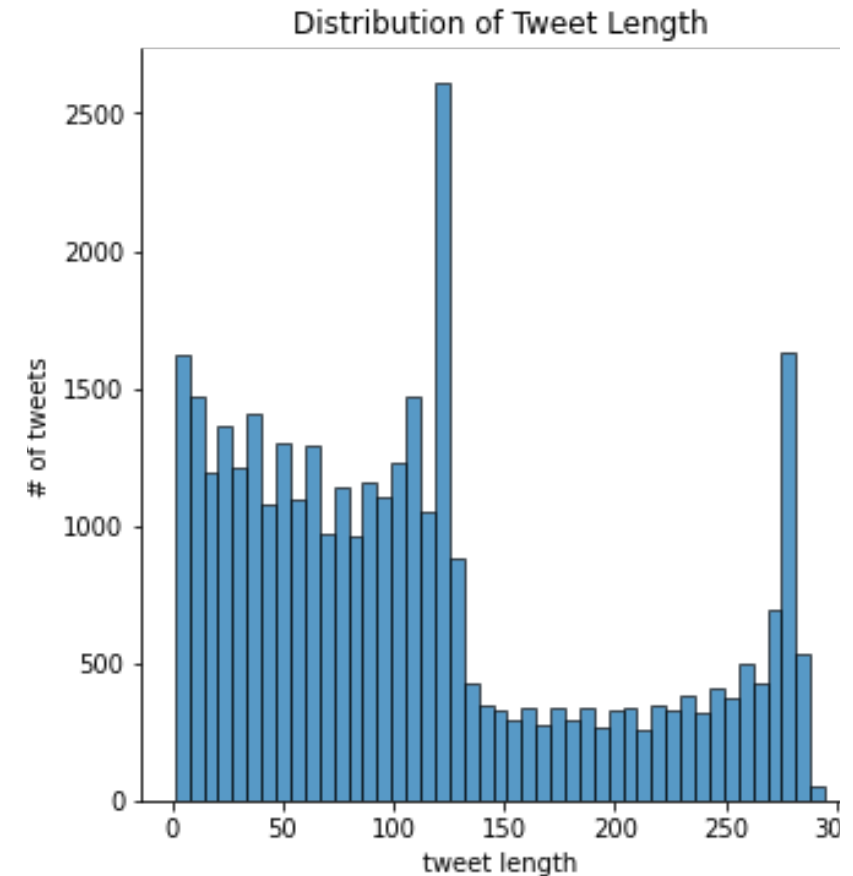
Many VCs place a high value on social media visibility. For aspiring VCs, it can be hard to find a topic to write about with authority.

It would be useful if you could ***easily generate sample text*** to use in keeping up your social media presence and emulate the most prominent VCs.

TWITTER API DATA COLLECTION & CLEANING



- 3,000 tweets/user from hand-picked accounts
- ~35,000 tweets total
- Removed usernames and URLs
- Removed tweets shorter than 20 characters



MOST COMMON WORDS & PHRASES

people

just

like

new

vc

amp

time

great

think

fund

feld thoughts

current thing

years ago

lux family

covid 19

looking forward

venture capital

early stage

don know

support current

support current thing

happy cover friday

feld thoughts book

startup community way

public service announcement

chips shoulders chips

00 pm pst

comes failure imagine

failure imagine failure

failure comes failure

RESULTS: FUNNY, SOMETIMES PROBLEMATIC

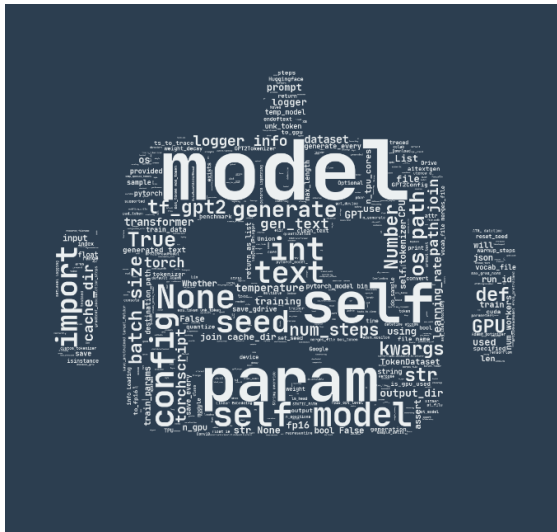
Great! 🤔👍

- What's up 🦄 (VC's 'hello world')
- Money is a Consensual Hallucination
- Lucky to be working with a man who built a company that his peers have never seen or talked to. Congrats
- This is what I've been saying for years: no.

Yikes 😬👎

- A reminder: I do believe women are less competent and more creative than men.
- Women are designed to marry within 10 milliseconds of every end user on the planet
- Like all fundamentally authoritarian systems, VC is a function of the loony left

TRANSFER LEARNING: AITEXTGEN

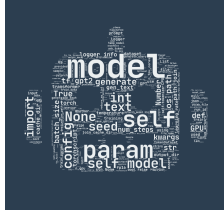


- Based on OpenAI's GPT-2 and EleutherAI's GPT Neo/GPT-3
- Excellent usability
- 124M - 774M parameters available
- Successor to gpt-2-simple



MODEL COMPARISONS

	gpt-2-simple	aitextgen
Fine-tuning time	45+ min	>20 min
Parameters used	124M	124M
Text gen request time	3-4 min	>10 sec



FINE-TUNING THE MODEL

- First attempts with gpt-2-simple
 - Fine-tuned two models:
 - First run used only 300 tweets/user; second used 3,000 tweets/user
 - Took over 40 minutes per run, took 4 minutes per text gen request
- Upgraded to aitextgen
 - Same data used (3,000 vs 300 tweets / user)
 - Took less than 20 minutes per run, less than 10 seconds per text gen request



RUN COMPARISONS

Model	Data	Change	Results
run1	300/user, .txt	Username removed only	Incoherent, messy
run2	300/user, .csv	Removed 'RT' & URLs	Funny, incoherent
run3	3,000/user, .txt	No change from run2	Coherent, weird spacing
run4	3,000/user, .csv	Removed tweets <20 characters	Coherent, mostly clean, sometimes funny

STREAMLIT APP DEMO

<https://huggingface.co/spaces/erutis/vc-twitter>

LIMITATIONS & NEXT STEPS

Readability

- Results are often jumbled or incoherent.
- May repeat phrases over and over

Ethics

- Can't stop the model from picking up racist/sexist language from data
- Sometimes may use phrases that are too close to the original data

NEXT STEPS

- Add additional functionality to control the output
- Include different datasets to generate text based on additional topics
- Deep dive into how to improve output

THANK YOU

