Identifying E-cigarette Content on TikTok: Using BERTopic Modeling Approach

Juhan Lee, PhD
Department of Psychiatry
Yale University School of Medicine

2023 Society for Research on Nicotine and Tobacco (SRNT)
March 3rd, 2023

Disclosure

- This research was funded by NIH National Institute of Drug Abuse: R01DA049878 (PI: Kong, G)

- There is no conflict of interest in this study.

BACKGROUND

Background

- E-cigarette content on social media is abundant, including e-cigarette-related posts and promotions.
- E-cigarette content on social media frequently portrays e-cigarette use as positive and appealing.
- A recent meta-analysis found that exposure to e-cigarette content on social media is associated with actual e-cigarette use among youth. (Donaldson et al., 2022)

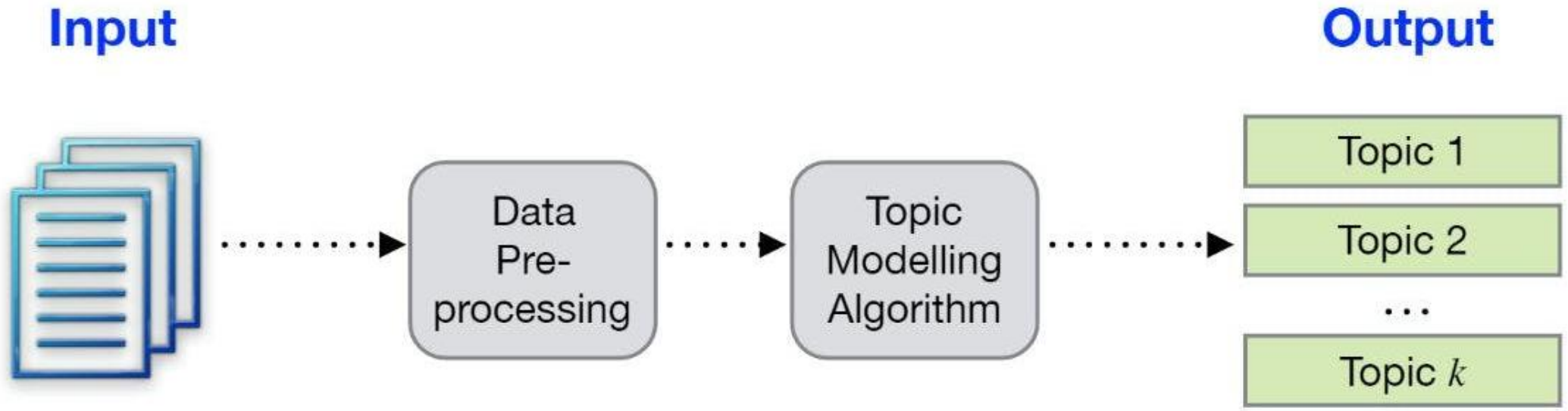
Background

- Previous studies used various machine learning approaches to identify e-cigarette-related content across social media platforms such as YouTube and Instagram.

- However, there is limited understanding of the e-cigarette content on TikTok, one of the most frequently used social media platforms by young people (67% of US teens in 2022)

Background

- **<u>Bidirectional Encoder Representations from Transformers [BERT]</u> topic modeling** is natural language processing, deep-learning method to cluster "texts" to interpretable topics.



Background

- Similar deep learning topic modeling approaches have been validated to identify vape-related Tweets
  - vape use/users, and vaping devices/products (Baker et al., 2022)

- It also successfully classified and clustered e-cigarette marketing and advertisement on Twitter.
  - e.g., *Today only! Buy one JUUL get the second half price with our online coupon code #JUUL4LIFE*

Aims

- This study aimed to identify and examine e-cigarette-related content on TikTok using BERTopic modeling approaches.

METHODS

Methods

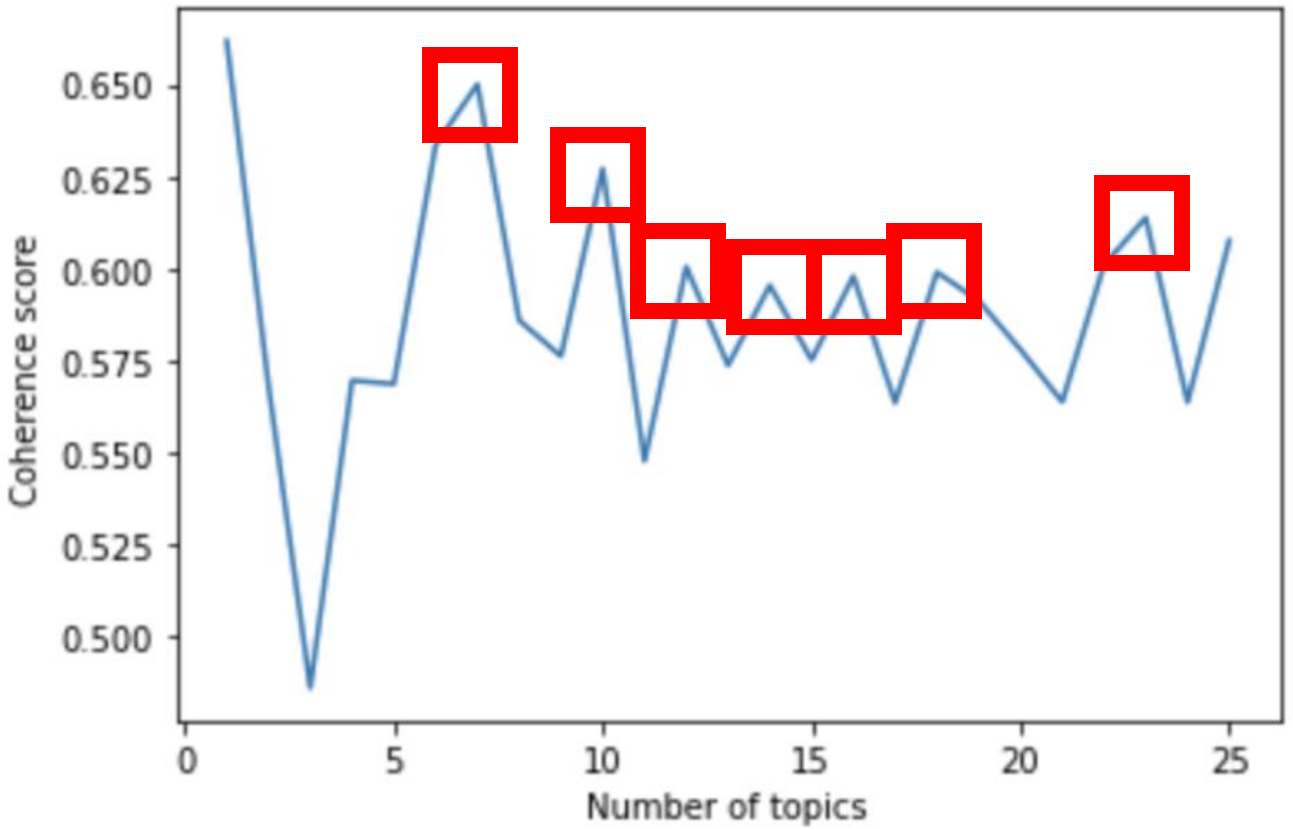
- We used 14 unique hashtags related to e-cigarettes (e.g., #vape, #vapelife, #vapenation) for data collection in Feb 2022.

- Final analytic sample included **13,573** TikTok posts

- Procedures of the number of topics chosen:
  - Quantitative: Coherence test (estimating consistency between its words/texts)
  - Qualitative: checked by two PhD-level reviewers through an iterative corpus-pruning process between computer scientists and tobacco experts.

RESULTS

COHERENCE TEST

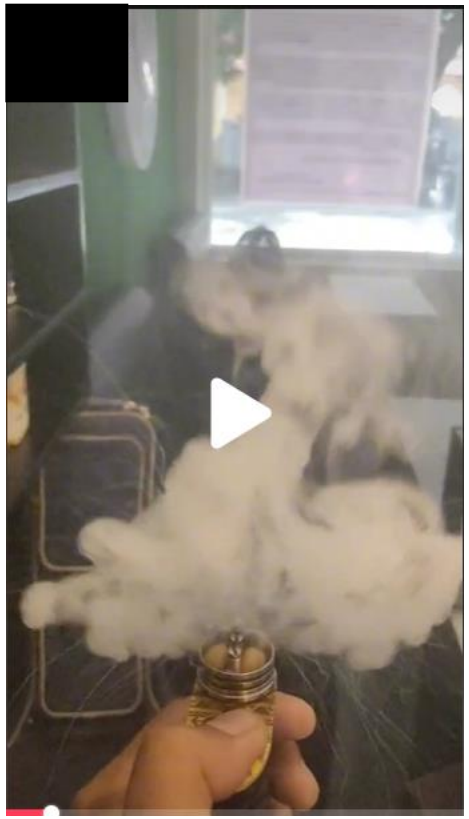


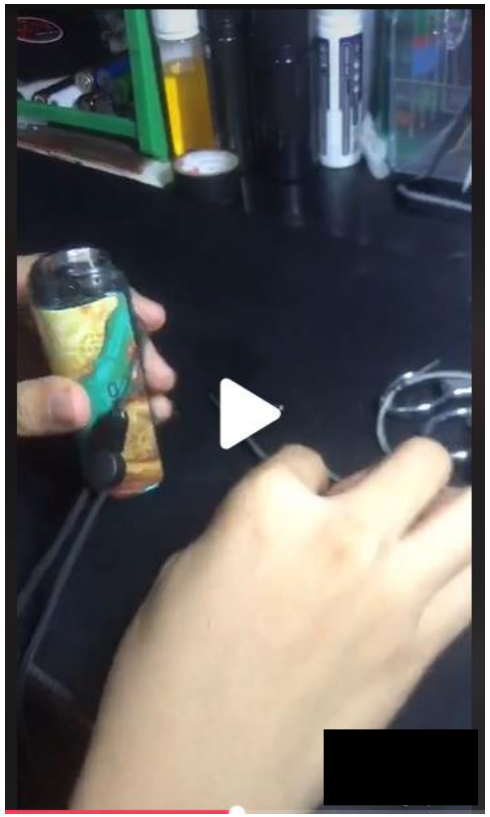
Yale TCORS
Yale Center for the Study of Tobacco Product Use and Addiction

Results

Topics	N=18	words
Social media and TikTok-related features	N=4	“duet”, “sketch”, “viral”, “trending”
Vape shops/brands	N=3	“store”, “drag (Voopoo)”
Vape tricks	N=3	“ripsaw”, “dragon”, “bane”, “lasso”
Modified use of e-cigarettes	N=1	“coil”, “wire”, “build”
Vaping and girls	N=1	“girl”
Vape flavors	N=1	“flavors”
Vape and smoke	N=1	“smoke”
Vape identities and communities	N=1	“community”, “nation”, “family”
Non-English topics	N=3	French, Romanian and Spanish

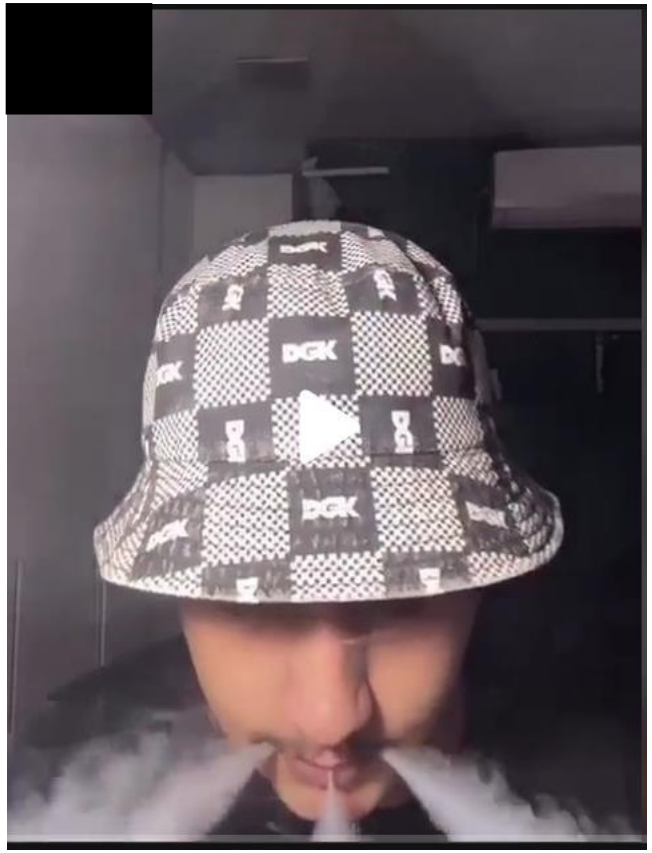
Results

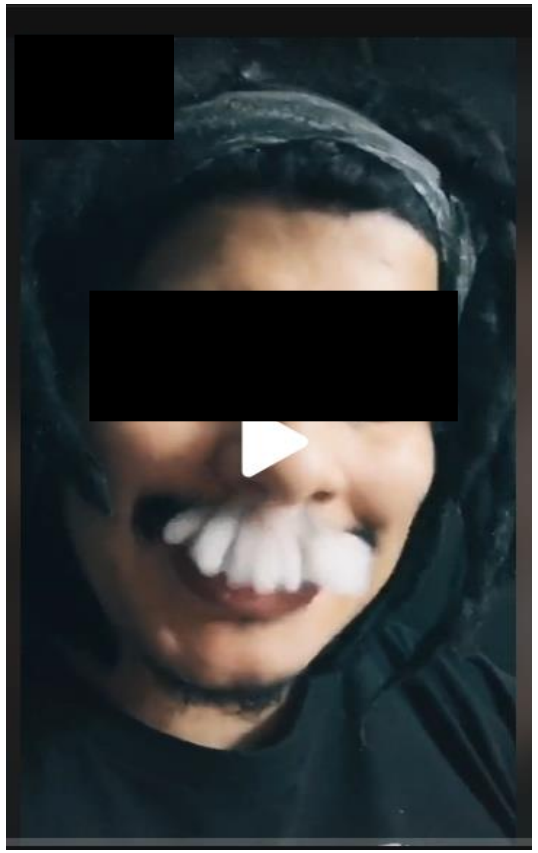




- **Modified use of e-cigarettes**
  - Showing side and upper profile of modified e-cigarette device
  - Showing smoke and sparks from a modified e-cigarette device
  - Fast-playing video of e-cigarette modification (rather than instructional)

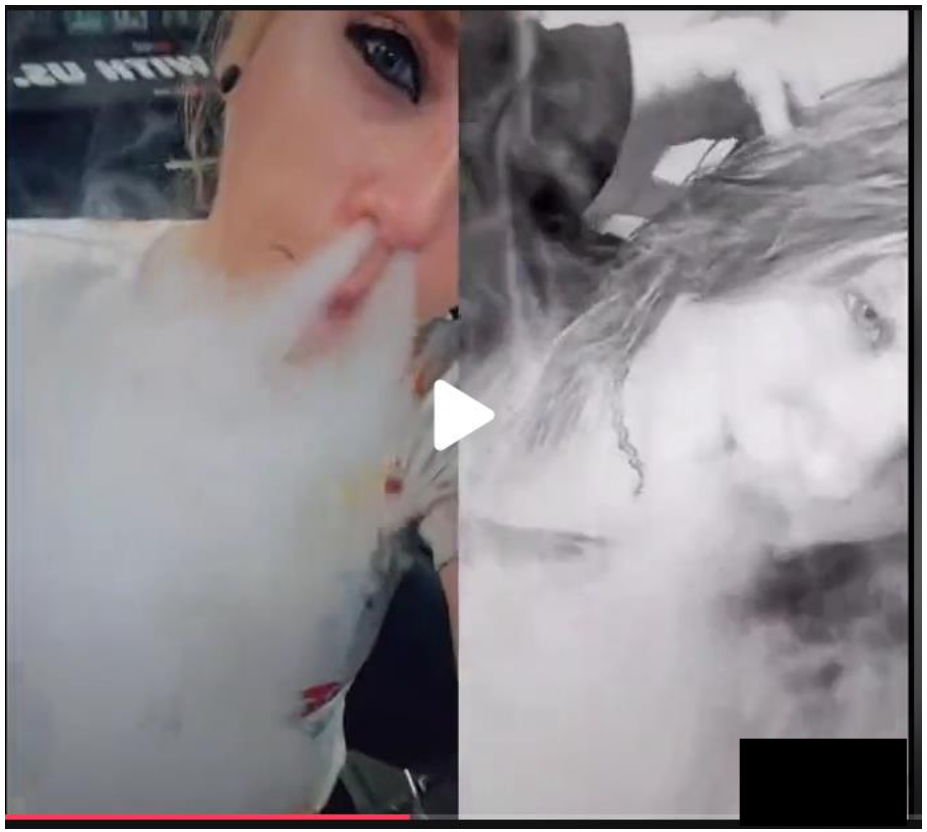
Results





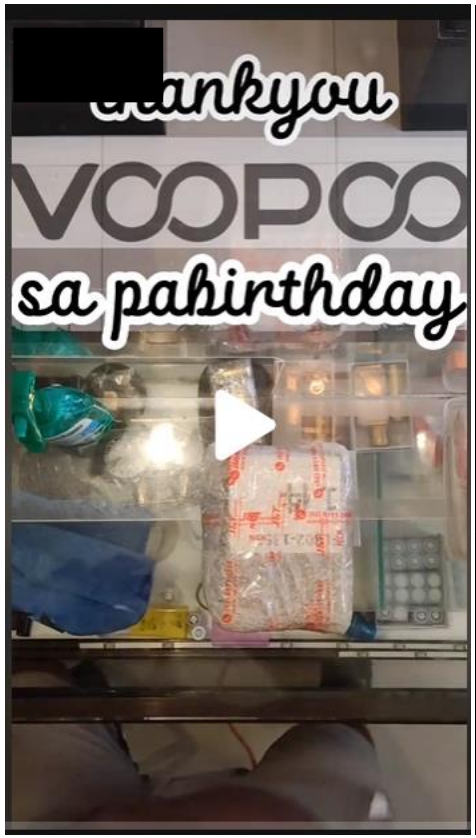
- **Vape tricks**
  - Posts featured various types of vape tricks generally set to fast-paced hip hop and rock background music

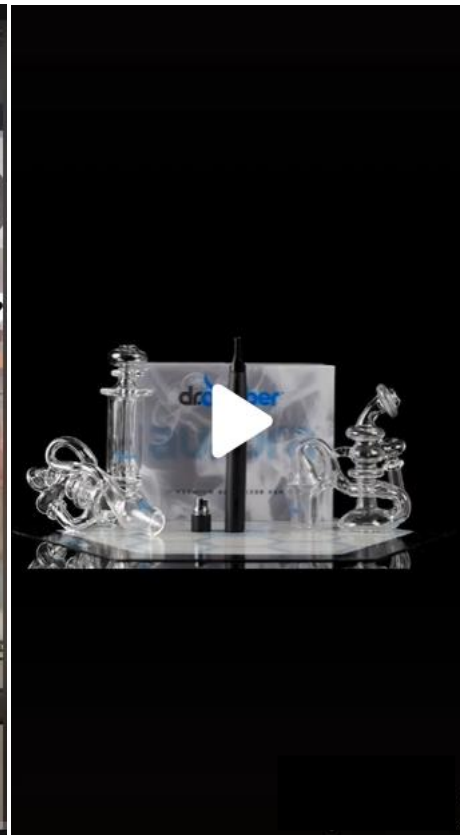
Results



- **Vaping and girls**
  - Featuring girls vaping, often with similar styles (e.g., make-up).
  - Potentially normalizing vaping in girls.

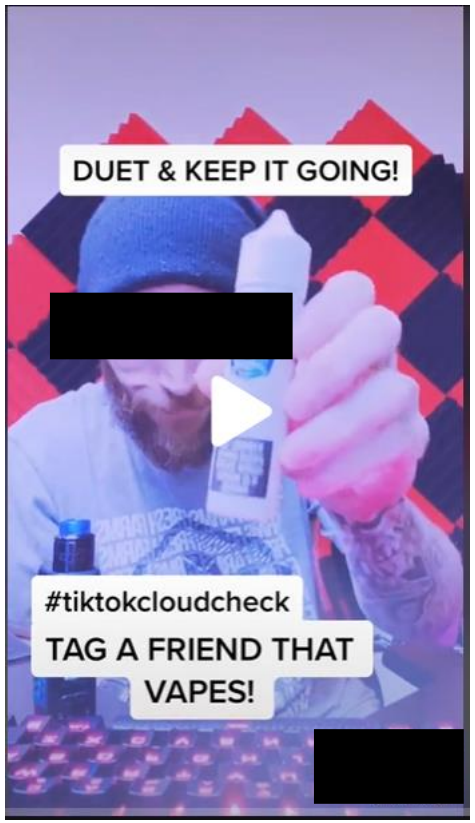
Results





- **Vape shops/brands, and flavors**
  - Featuring vape shops and brands.
  - Showing the arrangement of vape products and flavors.
  - Physical address and contact number sometimes provided.

Results



- **Duet**
  - Unique TikTok engagement strategy
  - A duet contains two videos on a split screen that play at the same time.
  - Some e-cigarette-related videos “encourage” people to create duet video with their e-cigarette content.

DISCUSSION

Discussion

- We used a machine learning method, BERTopic modeling, to successfully identify e-cigarette content from TikTok.

- Concerningly, predominant topics related to e-cigarettes on TikTok are found:
  - Names of vape shops/brands, vape tricks, e-cigarette modifications, girls vaping, flavors, which are not suitable for youth.

Limitations

- Even though we restricted our data collection to English words, Non-English topics are still occurring e.g., Spanish, French, and Romanian.
  - These words have similar spelling (but different meanings) across languages, such as “marina” (marine in Romanian), and “responder” (respond in Spanish).
  - Additional human-led validations are considered and recommended.

Implications

- Comprehensive regulation is needed on social media platforms to prevent youth access to the e-cigarette-related content.
  - Strong enforcement of vape-related content policy on social media is warranted.

- Developing counter-messaging for youth e-cigarette use prevention on social media is needed.

Acknowledgement

Grace Kong, PhD

Yale School of Medicine

Dhiraj Murthy, PhD

University of Texas at Austin

Ben Pretzer

University of Texas at Austin

Tanvi Anand, MS

University of Texas at Austin

Thank You

Do you have any questions?

juhan.lee@yale.edu

Yale TCORS
Yale Center for the Study of Tobacco Product Use and Addiction

01.
Topic

Describing topic of the section here

Yale TCORS
Yale Center for the Study of Tobacco Product Use and Addiction

01.
Topic

Describing topic of the section here

Yale TCORS

Yale Center for the Study of Tobacco
Product Use and Addiction

Yale TCORS

Yale Center for the Study of Tobacco
Product Use and Addiction

Abstract

01.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

02.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

03.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

04.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Abstract

01.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

02.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

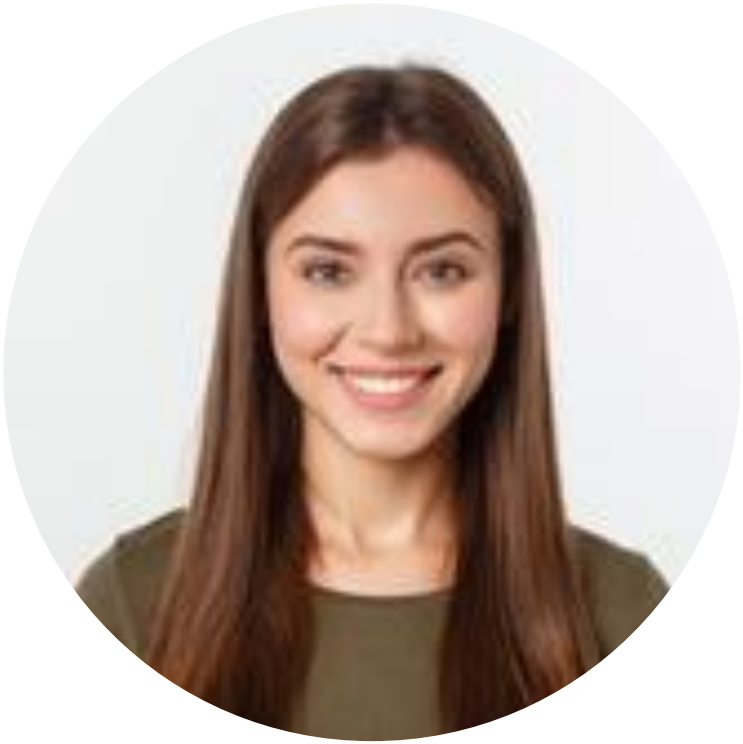
03.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

04.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Acknowledgement



Insert photo, then crop the photo to circle shape ratio

Their Name

Acknowledgement

InsertPhoto

Their Name

Their Name