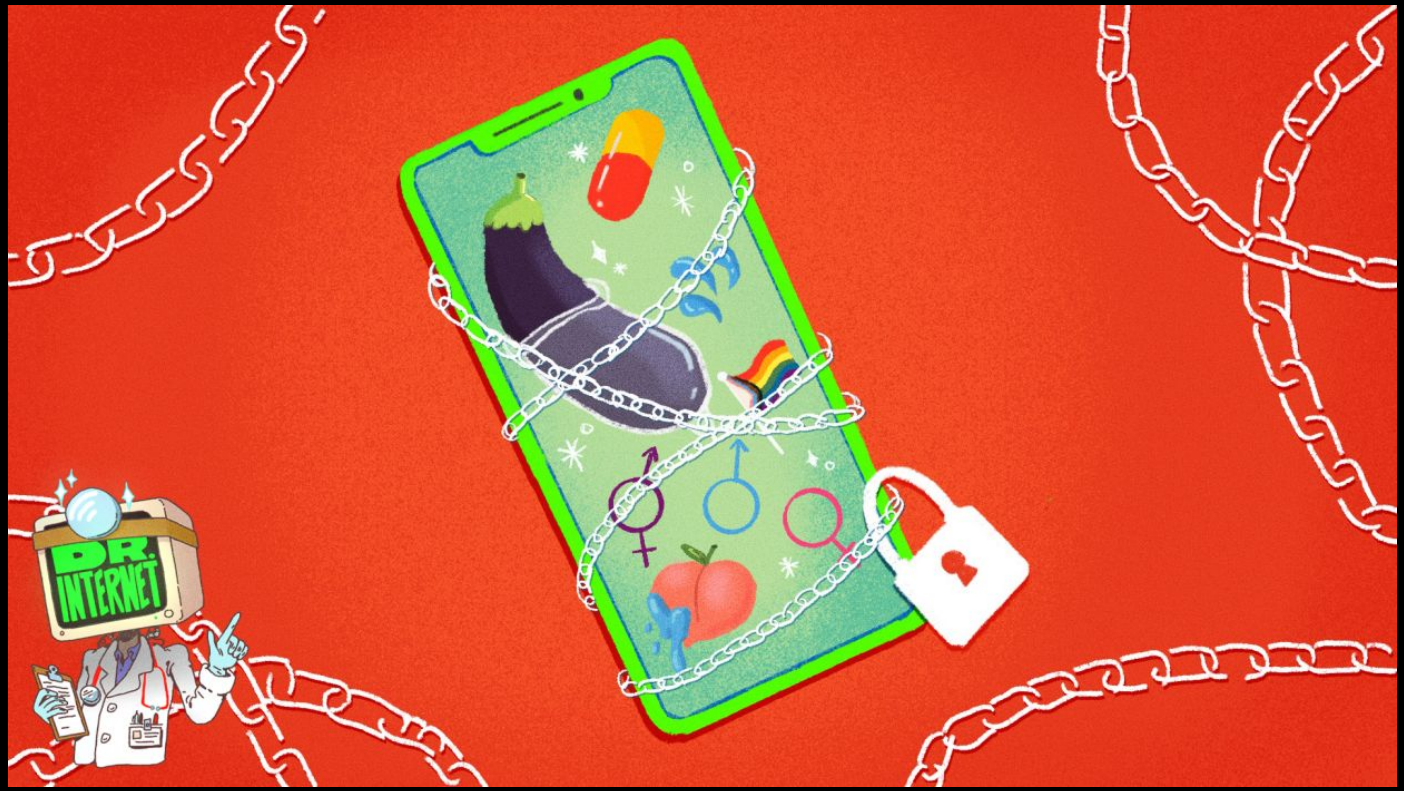
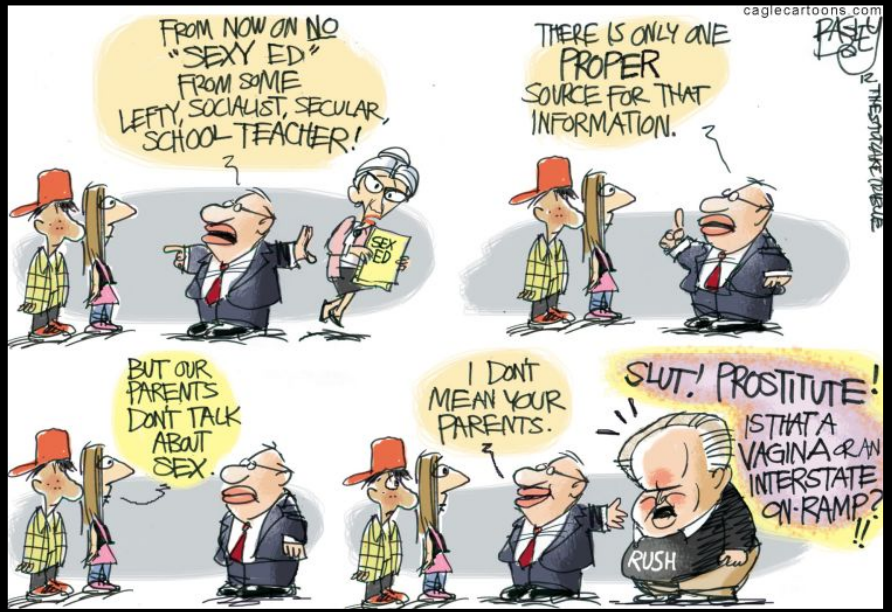
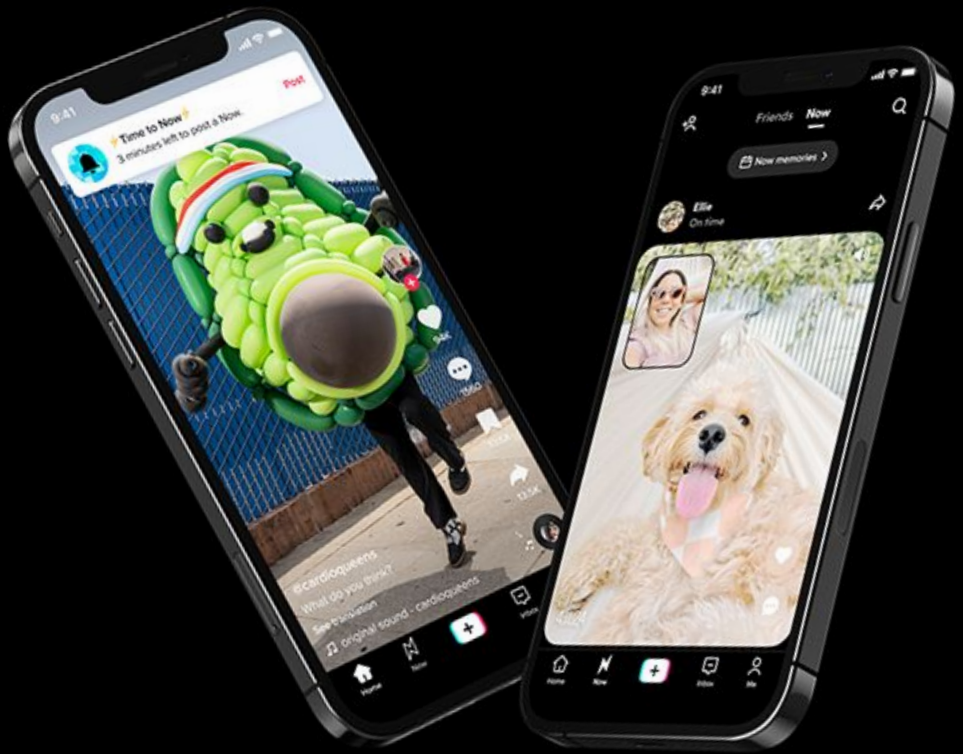


SexTok : It’s not suggestive, it’s educative

Credit: Ashley Vargas / Mashable



State of Adolescent Sex Education in the United States



Significant Adolescent Population

10-19 – 32.5%
20-29 – 29.5%

Account status

Your account is in view-only mode

Due to multiple violations of our Community Guidelines, your account is restricted to view-only mode until 2021-09-30 18:48:44

The next violation could result in being permanently banned. Review our **Community Guidelines**

content violation

Content can't be restored

It was determined that your content violates applicable laws and cannot be restored.

Video details
Posted on 07-31 at 14:40

content violation

Adult nudity and sexual activity

We do not allow sexually explicit or gratifying content on TikTok, including animated content of this nature. Sexualized content carries many risks, such as

What can we do about it?

Teaching a machine model to do better

- **Collected 1000 videos**
  - Sex Education
  - Sexually Suggestive
  - Neither

- The annotator agreement - Cohen’s kappa on the class label is 0.92

To check for visual bias

- Encoded perceived gender expression NOT to be used a feature
- The annotator agreement - Cohen’s kappa on gender expression is 0.89.

Label	Fem	Masc	Non conforming	More than one person with different gender expressions	None (ie no humans)
Suggestive	115	84	0	1	0
Educative	85	84	6	8	17
Others	164	170	12	113	141
Total	364	338	18	122	158

Experiment Setups

- Transcription Only
- Video Only

Group	Acc	Micro			Macro		
		P	R	F1	P	R	F1
Majority	0.65	0.00	0.00	0.00	0.22	0.33	0.26
All Text	0.76 ±0.02	0.82 ±0.10	0.40 ±0.02	0.54 ±0.01	0.52 ±0.04	0.59 ±0.01	0.55 ±0.01
Non-empty Text	**0.83** ±0.06	**0.92** ±0.07	0.43 ±0.04	0.58 ±0.03	0.57 ±0.02	0.59 ±0.02	0.57 ±0.01
Video	0.77 ±0.02	0.72 ±0.05	**0.50** ±0.04	**0.59** ±0.05	**0.75** ±0.04	**0.64** ±0.03	**0.67** ±0.03

Table 3: We present the average and standard deviation of results from three different runs of our experiments. We use accuracy, micro-precision, recall, and F1 (with "Others" as a negative class, not included in the scores) and macro-precision, recall, and F1 as metrics. Text-based classification has higher micro-precision but lower overall metrics as it cannot detect suggestive content well. Video classification has lower micro-precision but higher macro-precision as it detects suggestive content better.

Group	Suggestive	Educative	Others
Majority	0.00	0.00	0.79
All Text	0.00 ±0.00	0.82 ±0.04	0.84 ±0.02
Non-empty Text	0.00 ±0.00	**0.84** ±0.04	**0.87** ±0.01
Video	**0.59** ±0.07	0.59 ±0.03	0.84 ±0.01

Table 4: We present the overall F1 of each class label with the average and standard deviation of three random runs. Text-based classification gives a higher F1 for educative content, but suggestive content is detected only in videos where educative content is misclassified higher.

What Next?

- Multimodal (current work in progress)
- Paper under review at ACL

Credits:

Animated Illustrations : Julia from Ouch!

Screenshots and main intro image: from Mashable

Image of phone from tiktok.com

Cartoons from truthdig.com/cartoons

Icons from flaticon.com