

# Guidelines for labeling click-bait video titles

INFO 259 Annotation Project

Group Members: Tianhao Wu, Jingshu Rui, Yang Yang

## About the task

By definition, a clickbait video title is a title designed to attract attention and entice users to click on the video. Typically, they are deceptive, sensationalized, or misleading.

This task aims to identify whether or not a title is clickbait. Therefore, labeling will result in binary outputs:

**1** stands for positive, meaning the document **IS clickbait**.

**0** stands for negative, meaning the document **IS NOT clickbait**.

The document consists of category and video title, and they are separated by a semicolon following one whitespace. Example: Automobile; How A 700bhp Evo Saved My Life

This task is rather subjective. Whether or not a title would be interpreted as clickbait depends on factors including but not limited to interests in subjects, attention to particular words, cultural background of audiences.

There's a problem of consistency between headline and content: to decide whether it is clickbait also depends on whether his content is really as effective and interesting as his headline describes, which means that a fair decision can only be made after viewing the detailed content, not merely by reading the title to make pre-assumptions and predictions. The nature of content creators' work motivates them in some sense to do things that are out of the ordinary. They are competing for attention in the market and have the motivation to do really crazy things. Admittedly, they're battling for attention, but maybe their content is actually consistent with the eye-attracting title. Does that count as clickbait?

There's also a problem of intention: are the YouTubers themselves excited and just being expressive (not clickbait), or are they over-reacting and intentionally using pompous and pretentious language to trigger the audience's emotion (clickbait)? To classify a title as clickbait, there exists a knowing trick and deliberate confusion on the Youtuber's side. We as the audience have a hard time distinguishing their true intention.

Though labeling is rather subjective, we still provide general patterns and rules in the following section. They only act as a reference, and the final label should be decided by annotators.

## General rules and patterns

Content creators watch each other's videos and some interesting or attractive topics become a pattern or series; even the platform would come out with “xxx challenge” and invite YouTubers to film content with similar themes. Taking the food category as an example, some common and recurring titles include: trying out cheap food vs expensive same kind of food; using/eating a large amount of food; inviting celebrities to try out spicy food, etc.

We summarize these patterns we've noticed when labeling the exploration data and extract rules and tests from them in the following subsections. Examples of documentation we labeled will also be provided with explanations.

Some patterns weigh more than others, causing a title to be clickbait solely. Other patterns alone will not cause the title to be clickbait. Additionally, some patterns when appearing have exceptions in labeling as clickbait, and we will provide examples and explanations for such exceptions.

We separate observed patterns into two levels: Highly suspicious and lightly suspicious.

## Highly Suspicious of Clickbaiting

### Extreme Numbers

#### Explanation:

Numbers are eye-catching, standing out from words. They are intuitive and easy to read. The forms are different from words in terms of upper hang, lower hang, and width. When they appear in titles, viewers will naturally pay more attention.

Numbers come with different scales. How impactful a number depends on the context of the subject. For example, in the context of money, \$100,000,000 and \$0 weigh more than \$50 to identify as clickbait, since they're on the edge of the spectrum. Also, the word "free" may equivalently be perceived as a number in the context of money.

Numbers exist more in more fields than money too. They are expressed in time, age, percentage, scientific units, multiplication, etc. We gave some examples in such fields.

#### Examples:

##### - Money:

I Made \$246,397,197,269 by Deleting the Internet  
Why Windows is FREE Now  
Can You Play Fortnite on a \$200 Laptop?

##### - Time

Samsung's First Tablet: The \$5,000 PenMaster From 1992!  
Learn JSON in 10 Minutes

##### - Age

I Got Rejected By A 22-Year Old CEO

##### - Scientific unit:

Making 500,000 VOLT ARC with Marx Generator

##### - Percentage:

I Built a Theme Park With a 99.9% Death Rate - Parkitect

##### - Multiplication:

Double or Triple Your Internet Speed - This Method Actually Works!

### Overgeneralized words

#### Explanation:

Overgeneralizing will often time trigger people's curiosity to prove or disprove some statements, which match the characteristics of click-baiting.

#### Examples:

##### - Everything

How to Fix a Flat Tire EASY (Everything you need to know)

##### - All

Battlefield 4 Best Guns For All Classes

##### - Never

10 Awesome Car Technologies That Never Took Off

## **Superlative adjectives and adverbs**

### Explanation:

Superlative adjectives and adverbs have characteristics of overgeneralizing. We may not be sure that the subject content creators mentioned in the title actually be the superlative subjects.

### Examples:

The hardest problem on the hardest test  
World's Largest Devil's Toothpaste Explosion  
The Biggest vs Smallest Laptop?  
World's Roundest Object!  
Milking the WORLD'S MOST VENOMOUS FISH! - Smarter Every Day 117

Exception: When used in a subjective statement, superlative words may not be evidence of click-baiting

Live PD: Most Viewed Moments from of Nye County, NV | A&E

## **Lots of Uppercases**

### Explanation:

Uppercases occupy more space visually than lowercase letters. They carry less information in more space, resulting in the audience paying more attention.

### Examples:

FLYING PHONE SCAM EXPOSED (so I built a REAL one)

## **“Not clickbait” is a clickbait**

### Explanation:

There are no tricks more vulgar when grabbing attention by claiming itself not click-baiting. Stating “not clickbait” greatly increases the chance that people click on the video and find out why the content creator makes such a claim. Welcoming people explicitly to click on the video is also a form of click-baiting.

### Example:

World's LARGEST SUPER SOAKER!! (not clickbait)  
If You Click This Video I'll Give My Friend .001\$

## **Punctuation: !?**

### Explanation:

“!” indicates something unexpected, “?” suggests a question. When combined together, it expresses the feeling of unbelieving. This triggers people’s curiosity.

### Example:

No More Google!?

## **Unrealistic, Defying Normal**

### Explanation:

Statements about unrealistic, supernatural, and mysterious subjects easily grab people's attention and curiosity. Aliens, ancients, defying physics laws, nasty stuff normal people won't do, etc. could all fall in this category of click-baiting.

Examples:

- The Big Misconception About Electricity
- How An Infinite Hotel Ran Out Of Room
- I Built a Raft That's a Mile Long and Defies All Physics - Raft
- Drinking Nasty Swamp Water (to save the world)
- Turning cotton balls into cotton candy

## **Lightly Suspicious of Clickbaiting:**

### **Questions might be left unanswered**

#### Explanation:

Question represents curiosity. Unanswered questions in videos are a typical “bait-and-switch” strategy for increasing exposure. Notice that a title and category is the only information we get, thus we are unable to verify the question raised by the content creator in the title. Annotators need to judge the degree of unanswered questions by their common sense themselves.

#### Examples:

Why Alien Life Would be our Doom - The Great Filter  
How Hot Can It Get?  
Have They Seen Us? | Space Time | PBS Digital Studios

### **Heavy words or concepts**

#### Explanation:

Some concepts such as death, curse words, will cause mood swings in the audience. As a result, the audience may be more likely or less likely to click on the video, depending on individuals' backgrounds.

#### Examples:

Apple's Big "F U" to China  
How Houdini DIED (in Slow Motion) - Smarter Every Day 108

#### Exception:

Objective facts:

The Cannibal Warlords of Liberia (Full Documentary)

Rhetorical:

Tomi Lahren's Anger Lights Facebook on Fire: The Daily Show

### **Celebrities**

#### Explanation:

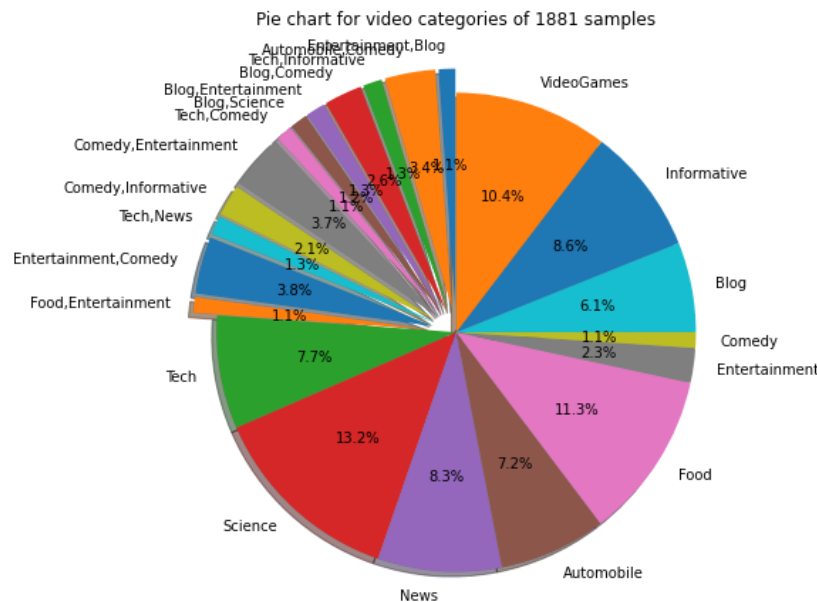
Celebrities are the objective concept of people. However, when mentioning them in abnormal contexts, they may result in clickbait.

#### Example:

Feeding Bill Gates a Fake Burger (to save the world)

## How video CATEGORY may influence the labeling

Video data we've collected consists of categories such as automobile, blog, comedy, entertainment, food, informative, news, science, tech, and video games. Some videos may have overlapped categories. Of the 1881 documents we've collected, 1432 documents have single categories, which is around 76%. As a visualized overview, a pie chart of the count of videos for each category is presented below, where exploded sections are videos with multiple categories.



The clickbait rate is category-dependent. Food and entertainment categories' clickbait rates are high; while informative, news, and science categories have a comparably lower clickbait rate, for they are inherently serious, knowledge-dense, and factual. Facts and objective statements are not clickbait because their intent is to convey knowledge; whereas emotionally inflammatory or extreme rhetoric titles are highly suspicious of clickbait.

Now please enjoy the labeling process, and use this guideline as a reference. Remember, you may still use your subjectivity to decide.