**Project goal**: Determine what proportion of leading magazines covering Tesla’s ‘autopilot’ system view the label as misleading/a form of mismarketing, and what types of words they use to convey this.

**Deadline**: End of March (we need it for a conference submission, with a deadline on April 6).

1. From the top 5 journals of the United States- Wall Street Journal, New York Times, Washington Post, USA Today, New York Post
2. Construct a search query to look for articles that talk about accidents involving Tesla’s Autopilot cars. For instance, you can use search terms like "Tesla Autopilot” + “accidents" + “failures”, etc. Don’t explicitly include words like “mismarketing” in the title, since our hypothesis is that the articles will naturally suggest mismarketing.
3. *Randomly* sample 30 news articles from each journal on Tesla’ Autopilots accidents from the last 3-4 years. So this makes 30 articles x 5 journals = 150 articles total.
4. For each article, code whether the author explicitly or implicitly conveys that the word ‘autopilot’ misleading. Words that may suggest this include, but are not limited to, *overstated, defrauded, misuses, overconfidence, inattention, false sense of security, deceptive*, *mismarketing.* Code an article as ‘1’ if the author thinks autopilot is mismarketing, and ‘2’ otherwise. Later, we will use these codes to determine what proportion of articles on autopilot spontaneously suggest that the label is misleading.
5. For each article, concatenate and save all sentences that indicate mismarketing. Later, we will use these sentences to (i) plot word clouds, and (ii) quantify the most to last frequent words.
6. We need to ensure inter-coder reliability for the ‘1’ and ‘2’ codes only. To do this, please use the following procedure. Two people should code the same 10 articles, then measure what percentage of articles they agreed on. If they agree on less than 80% of the articles, they should resolve the discrepancies and then code another 10 articles. They should then re-check their agreement and repeat the procedure so long as they agree on less than 80% of articles. Once they agree on 80% or more, both coders can proceed with coding all remaining articles independently.
7. Format of final data frame: article title, link, journal, publish date, author, categorization for coder 1, categorization for coder 2, and concatenated sentences indicating mismarketing (in the event that the article is coded a ‘1’). If possible, please save the full text of each article as well in an easy-to-access format, e.g., JSON.