**Appendix A: Tweet collection for October 4 and May 25, 2022**

Tweets related to the October 4, 2022, accident during a missile test were collected in three waves using the *academictwitteR[[1]](#footnote-1)* R package. Below, we detail how tweets were collected in the three samples and present longer lists of representative tweets related to the accident in Gangneung extracted from each sample of tweet collection. For all three samples, retweets and duplicate tweets were excluded. We also detail how tweets were collected for the May 25 missile test, again using the *academictwitteR* R package.

After obtaining each sample as detailed below, we further narrowed down the October 4 tweets to those including either words for “war” or “battle” ("전쟁", "싸움", "전투","투쟁","교전") or “North Korea” ("북한"), keywords which were not used to generate the samples for October 4. For the May 25 sample, we narrowed down the sample to include those that contained the term “North Korea” AND a term for “war” or “battle”, as “North Korea” was one of the search terms.

1. *October 4, Sample 1*

For the first sample, tweets were collected in close to real-time using Korean-language keywords: “Gangneung” (“강릉”), “missile” (“미사일”), “explosion” (“폭발”), fire (“화재”), and “Hyunmoo” (“현무”) (the Hyunmoo-2C surface-to-surface missile was them missile which was being tested). These keywords were selected as they were the primary keywords used at the time to discuss the explosion and fire that occurred during the failed test. This sample yielded 2,805 unique tweets. We compiled this sample in close to real-time after noticing Twitter posts related to the accident in Gangneung shortly after 11:00 PM local time in South Korea on October 4, 2022. As a result, this sample is much smaller than the other two samples generated for October 4.

For sample 1, there were 450 unique tweets specifically mentioning “war” and/or “North Korea.”

1. *October 4, Sample 2*

The second sample was collected using an expanded keyword list similar to the first sample: “missile” (“미사일”), “explosion” (“폭발”), “fire” (“화재”), “flames” (“화염”), “Hyunmoo” (“현무”), “bomb” (“폭탄”), “accident” (“사고”), “accident” (“우연한”), “air base” (“공군 기지”), “airport” (“공항”), and “what happened” (“무슨 일이야”). This sample was not collected in real-time; rather, it was collected later by applying the keyword filtering to the time period from 11:00 pm local time in South Korea (shortly before the accident occurred) on October 4, 2022 to 8:00 am local time on October 5, a nine-hour time period. This sample yielded 7,456 unique tweets (26,698 total), with 2,576 mentioning either “war” or “North Korea” and 94 mentioning both.

1. *October 4, Sample 3*

The third sample was generated by filtering for Korean-language stopwords[[2]](#footnote-2) from 11:00 PM local time in South Korea on October 4 to 7:00 AM local time on October 5. This third, broad-based sample returned 470,466 unique tweets (1,037,448 total). For this third sample, it is important to emphasize that a large proportion of these tweets were unrelated to the Gangneung accident, as the filtering was not specific to that incident.

1. *May 25 Sample*

We collected tweets from 4:00 AM on May 25 to 5:00 AM local time on May 28 using the Korean-language keywords “Gangneung” (“강릉”), “missile” (“강릉”), “explosion” (“미사일”), “Hyunmoo” (“현무”), “bomb” (“폭탄”, “air base” (“공군 기지”), “airport” (“공항”), and “North Korea” ("북한"). This query returned 139,583 tweets (19,630 unique tweets). In order to analyze the immediate reaction to the missile tests (comparable to our analysis of October 4), we narrowed the sample to the 12-hour period between 4:00 AM and 4:00 PM local time on May 25; this time-restricted sample included 9,631 tweets (3,110 unique tweets). Of these, only 23 unique tweets mentioned war (951 included either a term for war or “North Korea”, but this larger inclusion is to be expected, both because of the North Korean missile tests and the fact that “North Korea” was one of the keywords used to generate the sample). Four tweets mentioned both a term for “war” AND “North Korea.”

1. Christopher Barrie and Justin Chun-ting Ho, Justin, ‘academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint’, *Journal of Open Source Software* 6:62 (2022), <https://doi.org/10.21105/joss.03272>. [↑](#footnote-ref-1)
2. Stopwords are commonly used words in a given language; the term comes from the typical practice of excluding these words during the processing of natural language data, as they do not contain additional useful information. Examples of English stopwords include “are”, “a”, and “the.” The benefit of generating a sample based on stopwords is that it returns a broadly representative sample of Korean-language tweets that are not biased based on particular keywords related to the topic of interest. [↑](#footnote-ref-2)