Supporting Information for A Large-Scale Comparative Field Study of Misinformation Interventions

Benjamin Kaiser[1*] and Jonathan Mayer[1,2]

[1*]Department of Computer Science, Princeton University, 35 Olden St., Princeton, 08536, NJ, USA.
[2]School of Public and International Affairs, Princeton University, 20 Prospect Ave., Princeton, 08536, NJ, USA.

*Corresponding author(s). E-mail(s): bkaiser@princeton.edu;
Contributing authors: jonathan.mayer@princeton.edu;

Supporting Information

The files referenced below are available at https://github.com/citp/misinformation-intervention-study-SI.

SI1. Information Panels

The English-language messages for each information panel were:

- **Humanitarian**: "On 24 February 2022, Russia invaded Ukraine... and caused Europe's largest refugee crisis since World War II, with an estimated 8 million people being displaced within the country by late May as well as 7.7 million Ukrainians fleeing the country."
- **Condemnation**: "On 24 February 2022, Russia invaded Ukraine... The United Nations General Assembly passed a resolution condemning the invasion and demanding a full withdrawal of Russian forces. The International Court of Justice ordered Russia to suspend military operations."
- **Warning**: "On 24 February 2022, Russia invaded Ukraine... Protests occurred around the world; those in Russia were met with mass arrests and increased media censorship, including a ban on the words 'war' and 'invasion'."

Research assistants who were proficient in each of the languages included in the study assisted us in finding comparable quotes from foreign language-editions of Wikipedia.

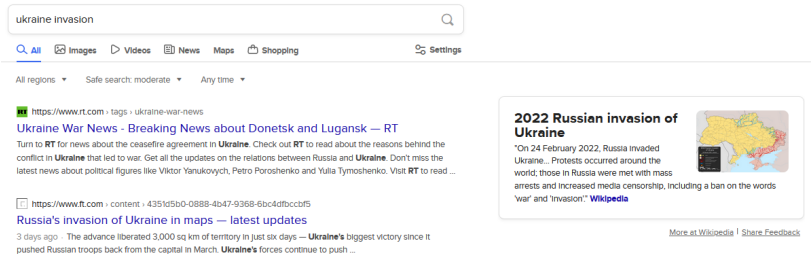

Figure 1 Warning information panel, sidebar

SI2. Inclusion Criteria

Query Keywords

We developed a set of query terms to identify searches relevant to the Russian invasion of Ukraine. We started with an initial term list drawn from Chen and Ferrara [1]. For each English-language term in their dataset, we sampled 300 searches (or as many searches as were available) from between April 2022 and August 2022 and manually labeled each search as either relevant or irrelevant to at least one of the following topics:

- The invasion of Ukraine
- Russian or Ukrainian politics, economics, history, military info, current events, notable people, landmarks, basic facts, or geography
- Global politics and events related to the invasion of Ukraine
- Any news coverage of Russia or Ukraine, or by Russian or Ukrainian outlets

We discarded any query terms with less than 85% query relevance, leaving 15 terms. We then conducted a round of snowball sampling to identify additional query terms that were frequently used in conjunction with the initial terms. We evaluated these additional query terms using the same method as for the initial list, resulting in 6 additional terms. Finally, we translated the terms into each of the languages our study supported. The query terms are:

- ukraine, ucrania, ucrânia, украина, украины, ukraina
- russia, russie, rusia, russland, rússia, россия, rossiya
- putin, poutine, путин
- soviet, soviétique, soviético, sowjetisch, советский, sovetsky
- kremlin, kreml, кремль
- minsk, минск
- ukrainian, ukrainien, ucranio, ukrainisch, ucraniano, украинец, ukrainets
- NATO, OTAN, НАТО

- luhansk, lugansk, lougansk, луганск
- donetsk, donezk, донецк
- donbas, donbass, donbás, dombas, донбасс
- kyiv, kiev, kiew, kiiv, киев
- moscow, moscou, moscú, moskau, москва, moskva
- zelensky, zelenskiy, zelenski, зеленский
- KGB, КГБ
- crimea, crimée, krim, crimeia, крым, krym
- kharkov, kharkiv, jarkov, charkow, carcóvia, харьков
- belarus, biélorussie, bielorrusia, weißrussland, bielorrússia, беларусь
- nova kakhovka, nouvelle-kakhovka, nouvelle kakhovka, nova kajovka, nova kachowka, новая каховка, novaya kakhovka
- kherson, kerson, cherson, херсон,
- duma, douma, дума

SI3. News Websites

A key goal of our study was to measure whether interventions made users less likely to select Russian state-affiliated media (RSAM) outlets and more likely to select other, credible news results. We created lists of RSAM and credible news websites to classify search results client-side. The list of RSAM websites is included in Section 3.1. For the dataset of news websites, we chose three sources:

1. ABYZNewsLinks.com, a directory of links to online news sources from around the world. We collected 36,241 news links using an automated scraper.
2. Wikipedia's list of newspapers by country, a crowdsourced dataset of news outlets from around the world that includes websites for many entries. We collected 7,234 news links by manually traversing Wikipedia entries.
3. DMOZ, also called the Mozilla Directory, which is a directory of web links maintained by a community of volunteers editors. DMOZ has not been updated since 2017 but the final directory is still accessible. We collected 35,302 domains by downloading a data dump and filtering for domains labeled as News, News & Media, or Newspaper.

We identified several other datasets and chose not to use them. Curlie (the successor to DMOZ) does not provide a data dump or permit automated data collection. World-Newspapers.com, AllMediaLink.com, and AllYouCanRead.com had significantly fewer links per country than ABYZNewsLinks.com, which we chose to use instead. Worldpress.org was not reliably available during the period of data collection. Listings of news websites on BBC.com were not indexed in a way that they could be easily traversed for data collection.

From 78,777 links, we performed filtering and ranking as follows:

1. **Liveness and redirection**: we issued a GET request to each link. If we didn't receive a response, or received an error response code, we discarded

the link. If we did receive a response, we saved the original link and the response URL.
2. **Normalization**: we normalized each URL to a domain name by stripping the path and subdomain, unless the domain was a publishing platform like wordpress.com or substack.com, in which case we preserved the subdomain. This left us with 58,254 domains.
3. **Ranking**: we ranked the popularity of each domain using Tranco, a ranking dataset that is used widely in academic research and contains over 7.5 million domains.
4. **Classification and validation**: to evaluate the quality of the dataset, we ran the top 10,000 ranked domains through WebShrinker, a domain classification service. WebShrinker classified 5,731 (57.3%) of the domains as news. We randomly sampled 500 of these domains to manually confirm classification and found only 7 false positives, which we left in the dataset. We manually checked all 4,269 non-news classifications, finding 1,437 false negatives. We added these false negatives to the classified news domains. This constitutes our high confidence, top ranked news dataset of 7,168 domains.
5. **Quality filtering**: we removed 28 domains that qualify as extremely low-quality news sources based on DuckDuckGo's News Rankings policy, leaving us with 7,140 domains.
6. **Partitioning**: we selected the top 1,000 domains as our Top 1k dataset and the remaining 6,140 became our Top 1-7k dataset. Of the remaining 38,254 domains in our original dataset, 17,261 appear on the Tranco list, so this is our Long Tail dataset.

SI4. Analysis

Hypothesis Tests

We provide full results from our hypothesis tests in an attached file (`full_results.csv`). Two effect size measures are commonly used in related literature: standardized mean differences (also called Cohen's d or Cohen's h) and risk ratios. We chose to present risk ratios as our main effect size measure. Risk ratios capture the relative change in participant behavior for low base rate events like clicking misinformation results as well as high base rate events like clicking news results. To allow comparison with prior work, we also include Cohen's h for each hypothesis test.

Countries

We did not set any geographic filters when deploying the study, although we did set language filters. The study only triggered on SERPs where the user's language was set to English, Russian, German, Spanish, Portuguese, or Italian, and this influenced the distribution of countries where we conducted observations.

In attached files, we present results broken down for each of the top ten countries by observation count: the United States, United Kingdom,

Canada, Germany, Australia, Russia, Switzerland, The Netherlands, France, and Ukraine. For seven of the ten top countries, our top line result holds: strong downranking showed a significant downward effect on selection of misinformation results. For three countries – Australia, France, and Ukraine – the effect was not significant.

We also present results broken down by US state, for the top ten states by observation count: California, Texas, Florida, New York, Illinois, Virginia, Washington, Pennsylvania, Ohio, and Georgia. For eight of the ten top states, our top line result holds: strong downranking showed a significant downward effect on selection of misinformation results. For two states – Washington and Pennsylvania – the effect was not significant.

Time

We collected data for twelve weeks: January 18, 2023 through April 12, 2023. In attached files, we present results subdivided by week. There are thirteen files; the first (`weekly_results_2023_01-18.csv`) contains data for five days and the last (`weekly_results_2023_04-10.csv`) contains data for three days; all other files contain seven days of data.

Our top-level result holds for every week of data: strong downranking showed a significant downward effect on selection of misinformation results.

Imputed News Module Engagement Data

Due to an implementation error, we did not capture clicks on links in the related news module intervention for roughly the first month of the study (from the beginning of data collection on January 18, 2023 through February 20, 2023). This missing data is visible in the country-level, state-level, and week-level breakdowns provided above. For the key results presented in the body of the paper, and the full results provided above, we imputed the missing data using the following method. For each day that we did collect click data for the related news module intervention, we computed the click rate as the number of clicks divided by the number of SERPs where the news module was shown. For each day where we did not collect click data for the related news module intervention, we randomly sampled a click rate from the distribution produced above and used that rate to impute the number of clicks. This imputation does not affect any of our hypothesis tests, because we did not test hypotheses about engagement with interventions.

SI5. Literature Review

We conducted a comprehensive literature review of misinformation intervention research. Prior to this work, the most comprehensive survey was conducted by Courchesne et al [2]. Their survey covers 223 papers published between 1972 and 2021 "related to countermeasures designed to combat influence operations" describing "studies whose methodology provides evidence regarding causal relationships." We extended this corpus by collecting an additional 49 papers.

We searched Google Scholar, SSRN, arXiv.org, OSF Preprints, and PubMed for papers containing the keywords "misinformation", "disinformation", "conspiracy", "propaganda", "fake news", or "information operations" along with either the keyword "intervention" or "countermeasure". For each database and each pair of keywords, we read the titles and abstracts of the top 100 results and selected papers meeting the above criteria. For survey papers and meta-analysis papers, we also mined the bibliographies for additional papers for our corpus. We coded the full corpus of 272 papers according to the type of treatment delivered and the outcomes (or dependent variables) measured, including re-coding the original corpus. The coded dataset of papers is provided in the file `misinformation_intervention_studies.xlsx`.

References

[1] Chen, E., Ferrara, E.: Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War Between Ukraine and Russia. Preprint at https://arxiv.org/abs/2203.07488 (2022)

[2] Courchesne, L., Ilhardt, J., Shapiro, J.N.: Review of social science research on the impact of countermeasures against influence operations. Harvard Kennedy School Misinformation Review (2021). https://doi.org/10.37016/mr-2020-79