# Supplementary Information

## Data Collection Parameters

Table S1 presents the parameters of data collection for our study.

| Event Name | Collection Date | Collection hashtags | # Users | # Tweets |
|---|---|---|---|---|
| Asian Elections[1,2] | 2019, 2020, 2021 | #Pemilu2019, #Pilpres_2019, joko widodo, prabowo subianto, #TaiwanVotes, #Taiwan2020, #taiwanelection, #sgelections2020, #GE2020, #SGGE2020, #singaporeGE2020, #GE2020SG, #Singaporevotes, PAPSingapore, wpsg, jamuslim, ProgressSingaporeParty, Nicoleseah, PeopleActionParty, Workerparty, neesoongrc, lightningparty, eastcoastgrc | 950,789 | 4,126,008 |
| Black Panther[3] | 8 Feb 2018 to 23 Feb 2018 | #BlackPanther | 1,689,076 | 17,718,557 |
| Canadian Elections 2019[4] | 20 Jul 2019 to 6 Nov 2019 | #TrudeauMustGo, #TeamTrudeau, #trudeau, #Election2019, #elxn43, #chooseforward, #onpoli, #ItsOurVote, #lpc, #ndp, #cpc, #gpc, #NotAbot, #cdnpoli, #ButtsMustGo, #LavScam, #LiberalsMustGo, BlocQuebecois, #blocqc, #cccr2019, #NoTMX, #TMX, #TransMountain, #scheer, #dougford, #fordcutshurt, #fordisfailing | 1,946,277 | 18,213,477 |
| Captain Marvel[5] | 15 Feb 2019 to 15 Mar 2019 | #BoycottCaptainMarvel, #AlitaChallenge | | |
| Coronavirus 2020-2021[6] | 2020-2021 | #coronavirus, #coronaravirus, #wuhanvirus, #2019nCoV, #NCoV, #NCoV2019, #covid-19, #covid19 | 208,956,241 | 4,179,124,820 |
| ReOpen America[7] | 1 Apr 2020 to 22 Jun 2020 | #openup, #reopen, #operationgridlock, #liberate; #reopenNY + state abbreviatons | 201,083 | 4,429,298 |
| US Elections 2020[6,8] | 1 Dec 2019 to 16 Aug 2020, 16 Aug 2020 to 15 Jan 2021, 15 Jan 2021 to 17 Feb 2021 | #election2020, #2020_presidential_election, #maga2020, #flipitblue, #keepitblue, #yeswecan, #yang2020, #JoeBiden, #BernieSanders, #ElizabethWarren, #PeteButtigieg, #FeelTheBern, #democrats, #republicans, #Bloomberg2020, #Booker, #USPS, #VoteByMail, #SaveTheUSPS, #voterfraud, #BlackLivesMatter, #BLM, #reopen, #reopenamerica, #IranSanctions, #QAnon, #WWG1WGA, "natural born", #election2020, #presidentialelection, #democrats, #republicans, #JoeBiden, #BidenHarris2020, #Biden, #MAGA, #KAG, #Inauguration, #InaugurationDay, #Capitol, #USCapitol, #USCapital, #NationalMall, #Jan20, #election2020, #presidentialelection, #democrats, #republicans, #JoeBiden, #Biden, #MAGA, #KAG, #Trump, #USPS, #VoteByMail, #SaveTheUSPS, #voterfraud, #BlackLivesMatter, #BLM, #reopen, #reopenamerica, #IranSanctions, #QAnon, #WWG1WGA | 1,608,033 | 55,179,293 |

**Table S1.** Dataset collection parameters. Data was collected using the Twitter Developer V1 API

## Comparison of Linguistic Cues

Table S2 presents the detailed comparison of linguistic cues between bots and humans across all events.

| Cue | Bots | Humans | p-value |
|---|---|---|---|
| **Semantic Cues** | | | |
| First Person Pronouns | 0.71 | **0.73** | 5.62E-8*** |
| Second Person Pronouns | **0.20** | 0.18 | 0.001*** |
| Third Person Pronouns | 0.47 | **0.50** | 0.001*** |
| Reading Difficulty | **0.12** | 0.10 | 0.001*** |
| **Emotion Cues** | | | |
| Abusive Terms | **0.13** | 0.09 | 0.001*** |
| Expletives | **0.12** | 0.08 | 0.001*** |
| Negative Sentiment | 1.56 | **1.59** | 1.17E-11*** |
| Positive Sentiment | 2.88 | **3.10** | 0.001*** |
| **Metadata Cues** | | | |
| Mentions | **1.18** | 1.10 | 0.001*** |
| Media | 0.006 | **0.014** | 2.54E-59*** |
| URLs | 0.18 | **0.20** | 1.41E-29*** |
| Hashtags | **0.54** | 0.49 | 0.001*** |
| Retweets | **9372** | 8203 | 0.001*** |
| Favorites | 21.0 | **92.5** | 2.12E-51*** |
| Replies | 0.67 | **3.31** | 1.03E-6*** |
| Quotes | 0.51 | **2.05** | 1.21E-16*** |
| Followers | 1268 | **4164** | 0.001*** |
| Friends | **1158** | 817 | 0.001*** |
| Total Tweets | **56779** | 10695 | 0.001*** |
| Tweets per Hour | **1.20** | 1.10 | 0.001*** |
| Max Time between Tweets (seconds) | 44617 | **57838** | 0.001*** |
| Friends:Followers Ratio | **4.73** | 4.44 | 6.71E-288*** |

**Table S2.** Comparison of Mean of Psycholinguistic Cues per user for Bots and Humans across all events. This value is the number of words that match each cue per tweet, divided across the number of tweets the user has. *** indicates a significant difference between the bots mean and humans mean via a student t-test at the $p < 0.01$ level.

### Identity Use Per Event

Table S3 presents the percentage of users with identities that match the US census of occupations[9]. Figure S1 shows the identities used per event, separated by the top identities used by both bots and humans.

| Event | Bots (%) | Humans (%) |
|---|---|---|
| Asian Elections | 13.2 | 17.5 |
| Black Panther | 22.4 | 29.6 |
| Canadian Elections 2019 | 29.4 | 38.9 |
| Captain Marvel | 21.5 | 29.0 |
| Coronavirus2020-2021 | 15.8 | 26.6 |
| ReOpen America | 26.7 | 34.7 |
| US Elections 2020 | 20.6 | 12.7 |
| Overall | 21.4±5.7 | 27.0±9.2 |

**Table S3.** Percentage of Users with identity affiliations matching census

### Unique Identities per Event

Table S4 presents the number of unique identities that bots and humans use per event. Across all datasets, bots consistently use lesser number of unique identities than humans, indicating a smaller range of affiliation set.

| Event | Bots | Humans |
|---|---|---|
| Asian Elections | 613 | 741 |
| Black Panther | 733 | 831 |
| Canadian Elections 2019 | 776 | 860 |
| Captain Marvel | 747 | 838 |
| Coronavirus2020-2021 | 836 | 899 |
| ReOpen America | 807 | 884 |
| US Elections 2020 | 828 | 894 |
| Entire Dataset | 869 | 908 |
| Average | 762.8±76.7 | 849.6±54.7 |

**Table S4.** Comparison of the Number of Unique Identities Per Event

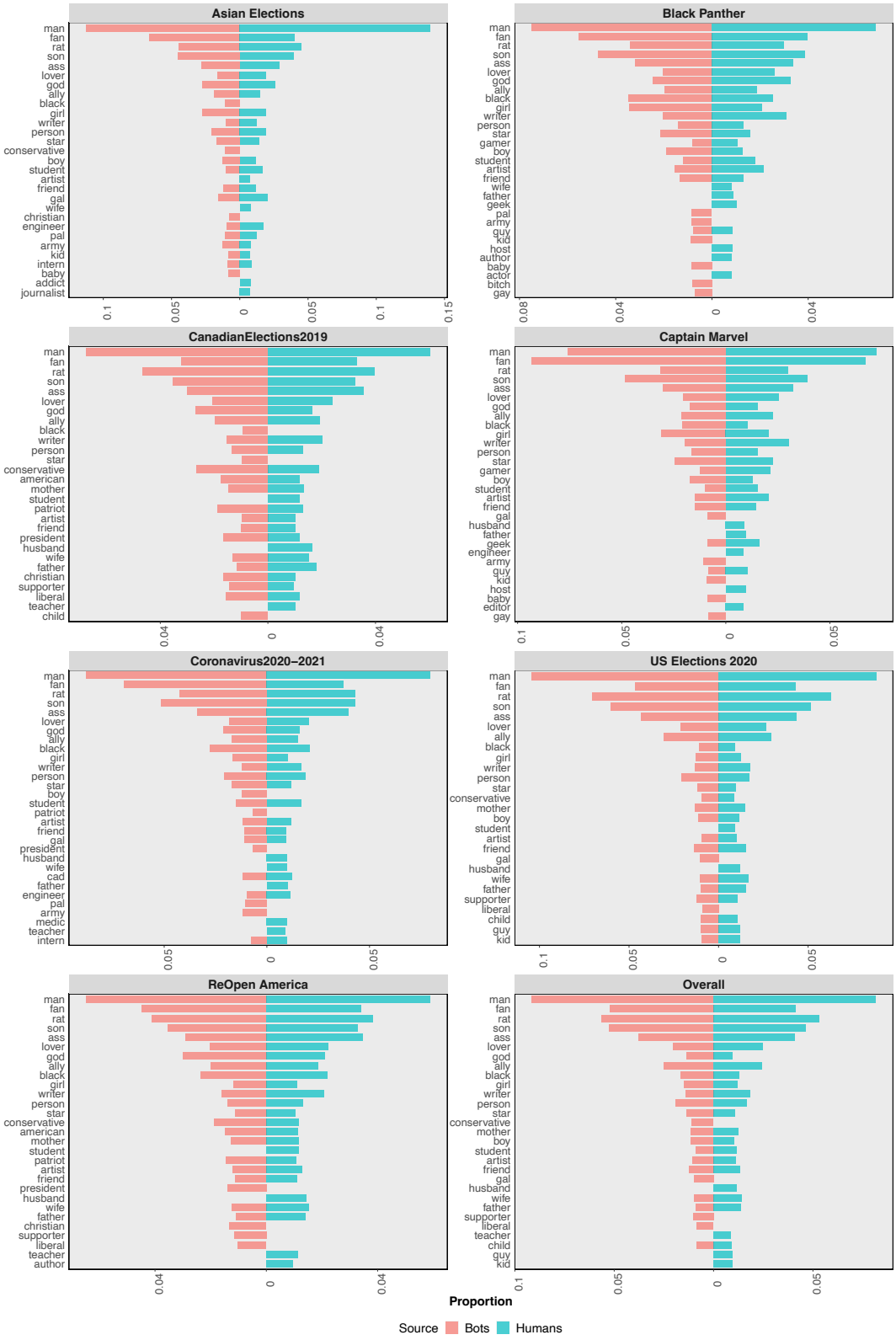


**Figure S1.** Comparison of frequency of use of the top Identity affiliations by bots and humans per event.

### Identity vs Framing Charts Per Event

Figure S2, Figure S3, Figure S4, Figure S5, Figure S6, Figure S7, Figure S8 show the frames each identity use per event, separated by the top identities used by both bots and humans.

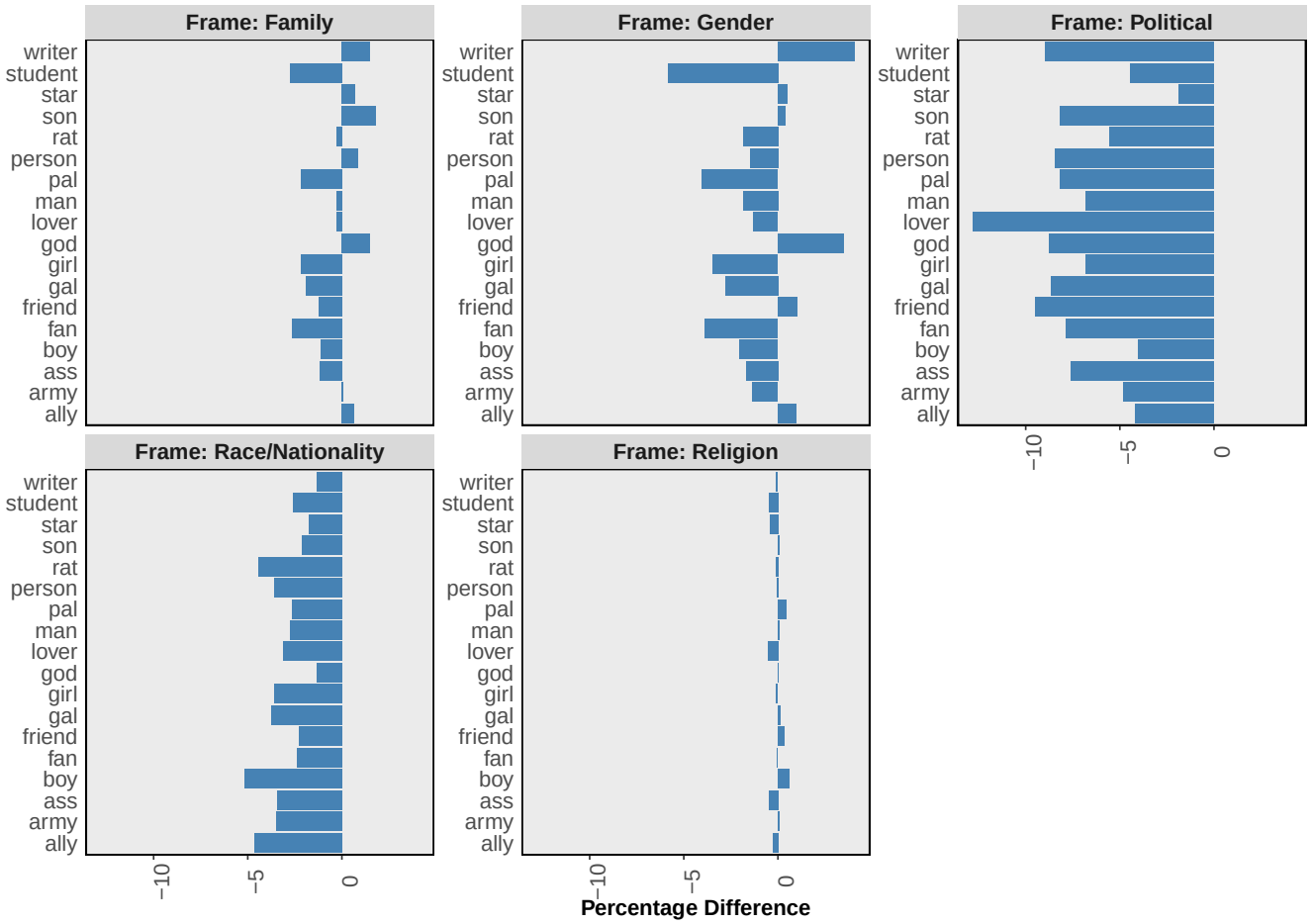


**Figure S2.** Percentage difference of narrative frames used for identities that are common between bots and humans in **Asian Elections** dataset

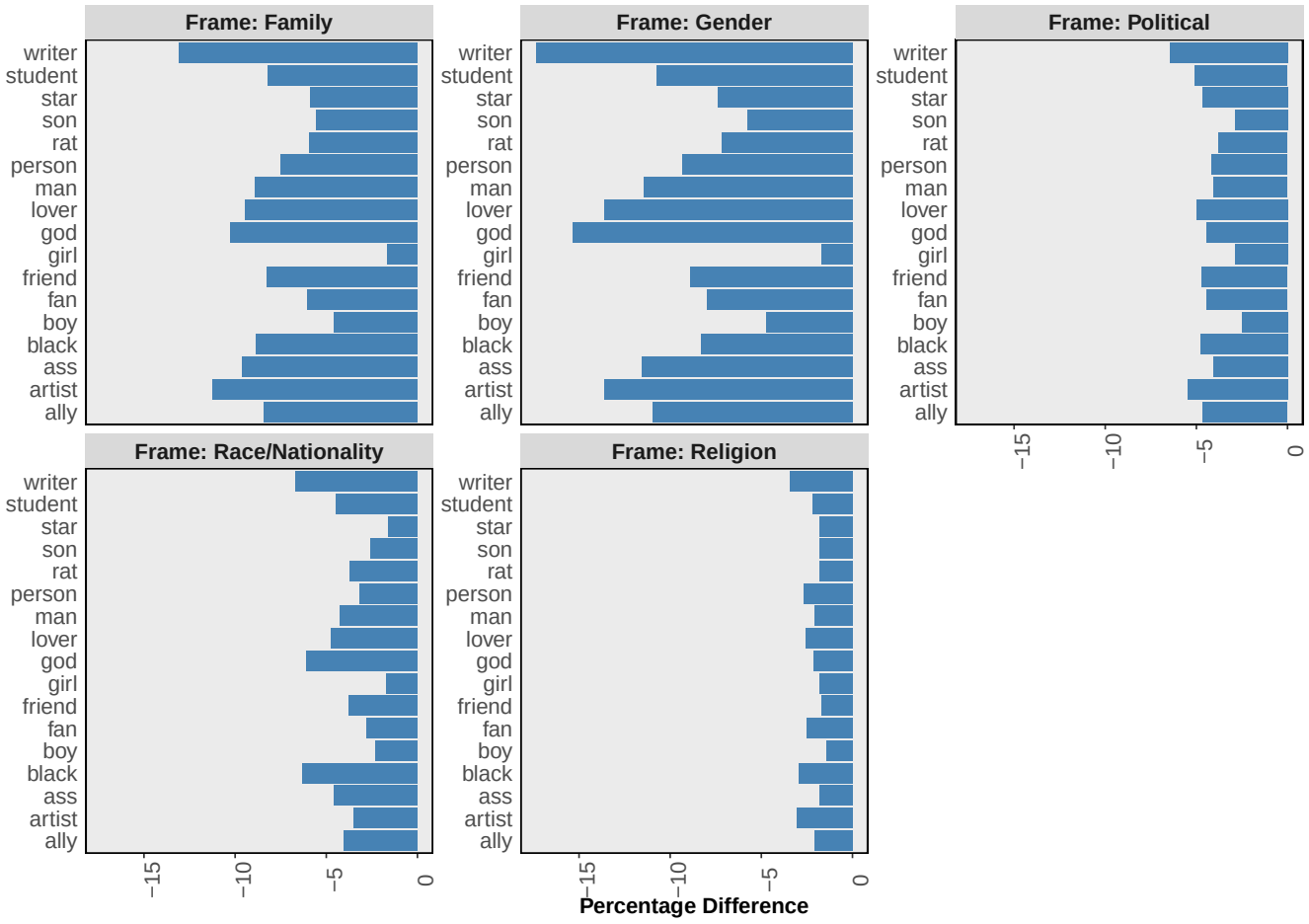


**Figure S3.** Percentage difference of narrative frames used for identities that are common between bots and humans in **Black Panther** dataset

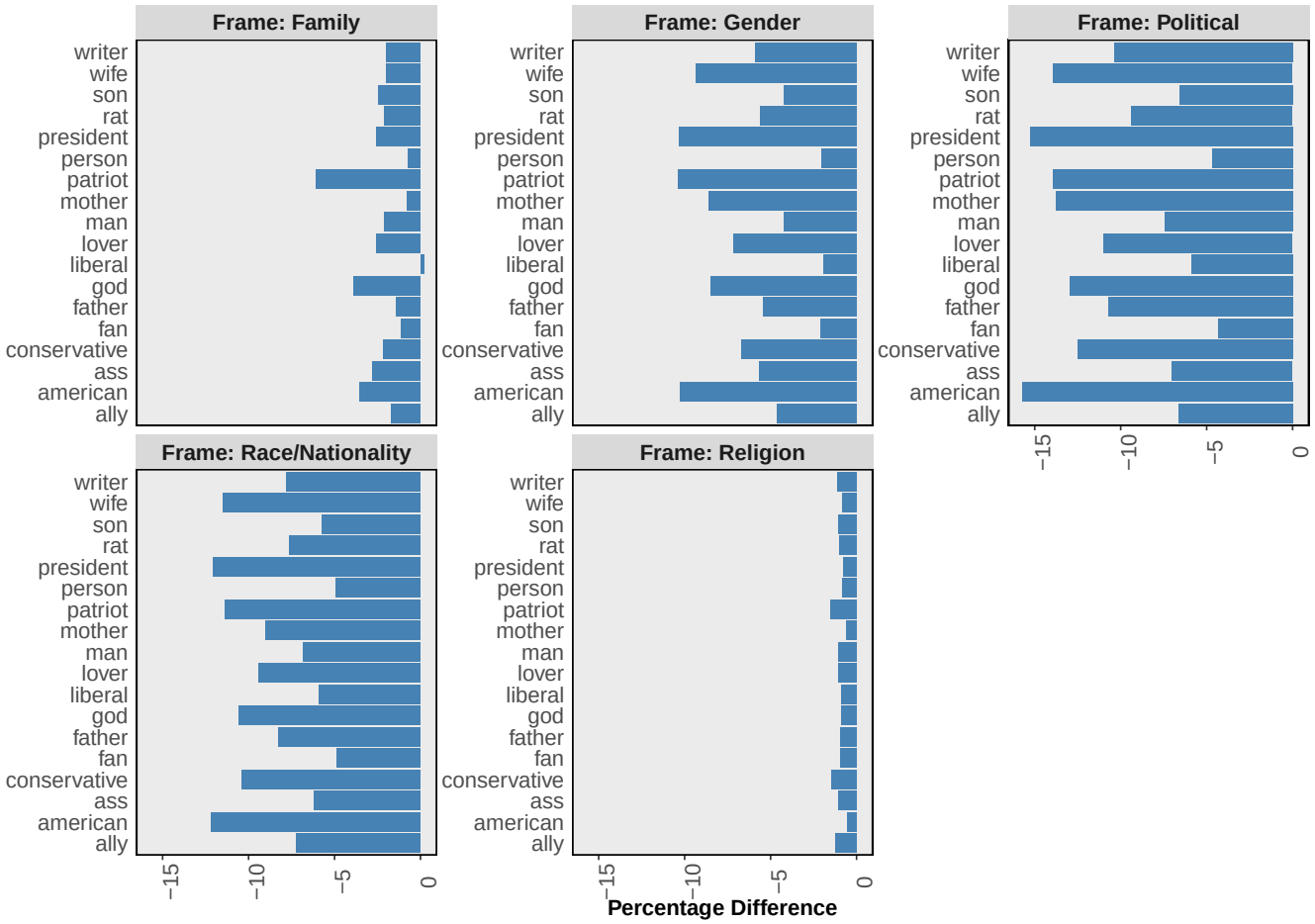


**Figure S4.** Percentage difference of narrative frames used for identities that are common between bots and humans in **Canadian** dataset

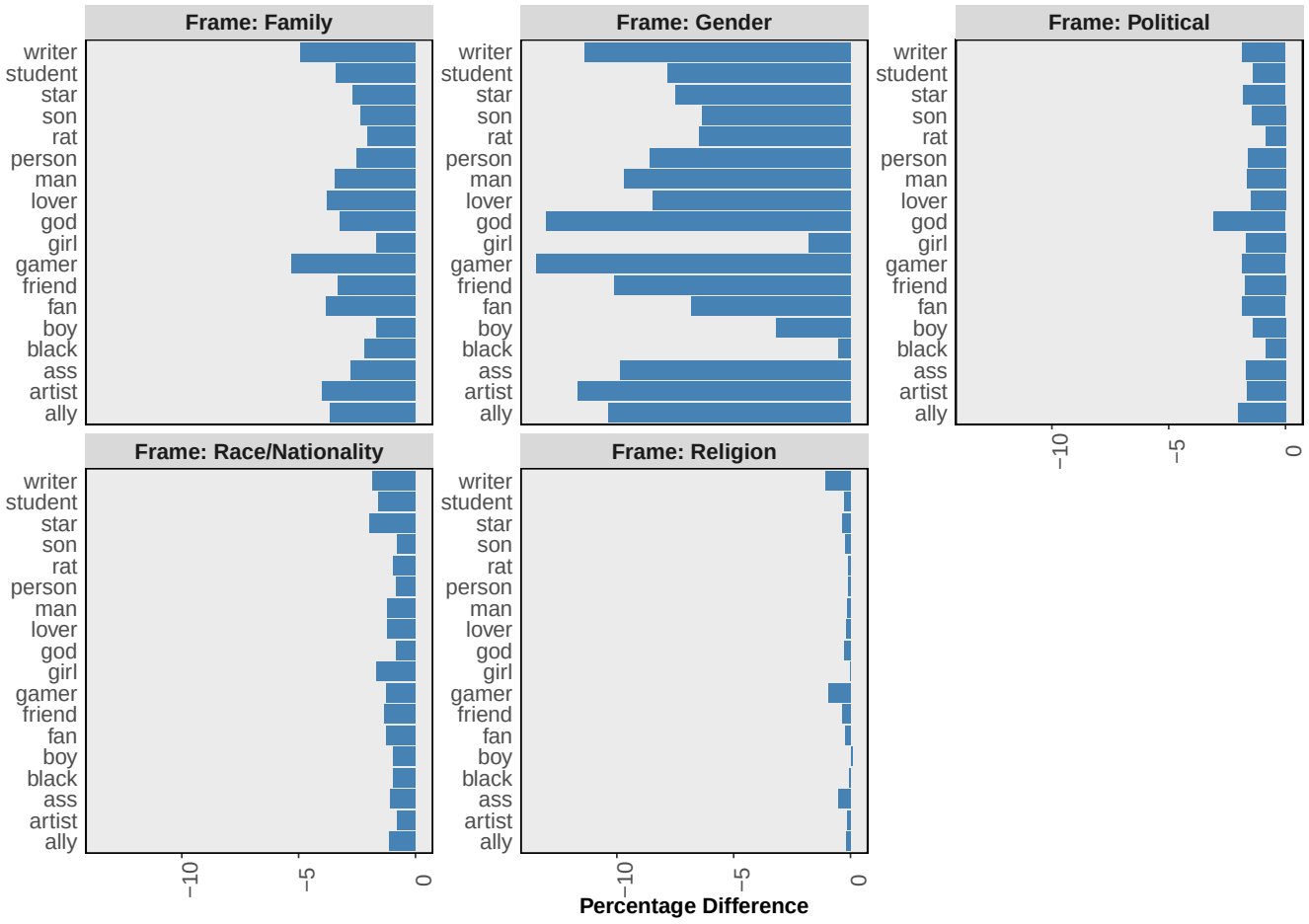


**Figure S5.** Percentage difference of narrative frames used for identities that are common between bots and humans in **Captain Marvel** dataset

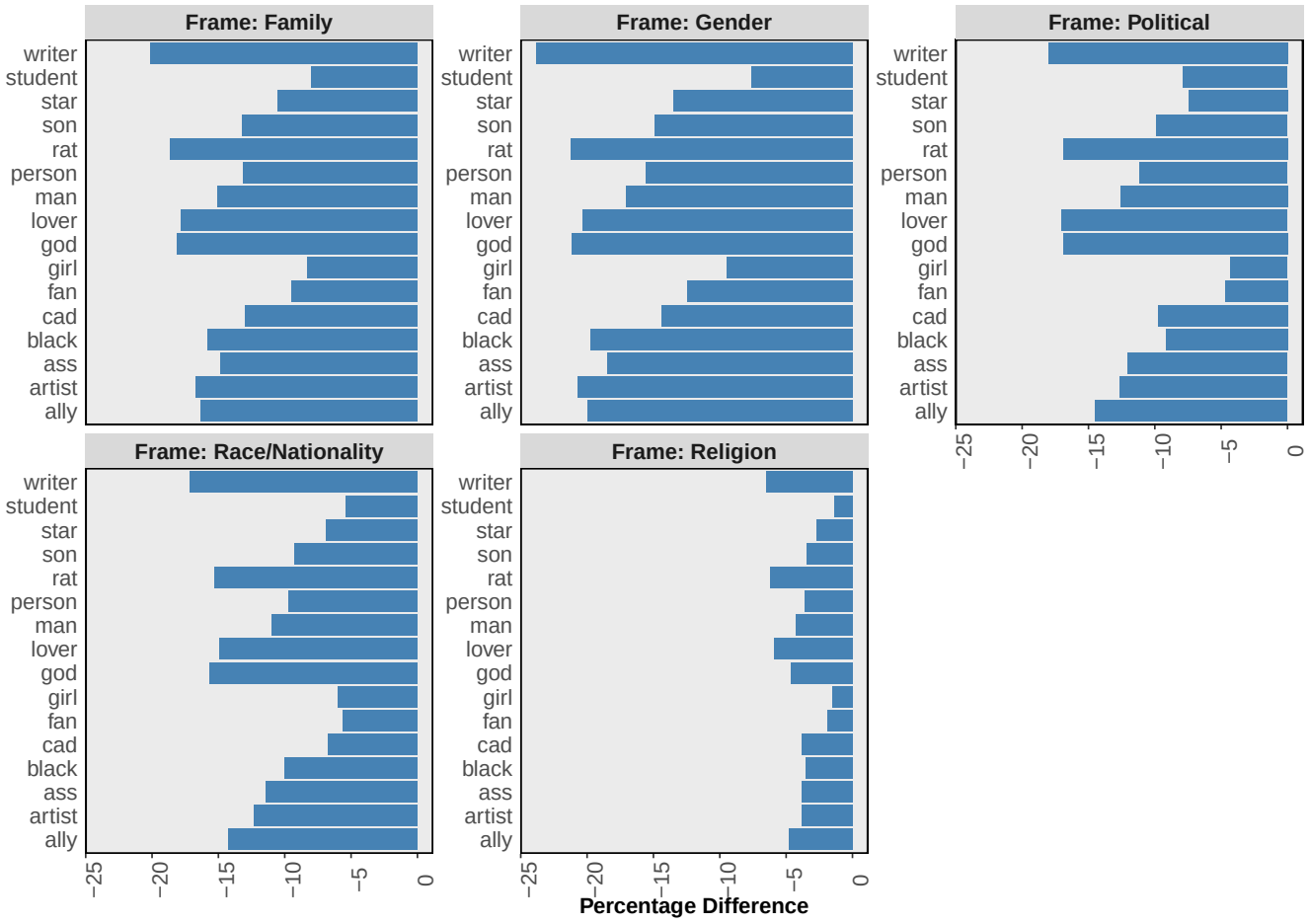


**Figure S6.** Percentage difference of narrative frames used for identities that are common between bots and humans in **Coronavirus** dataset

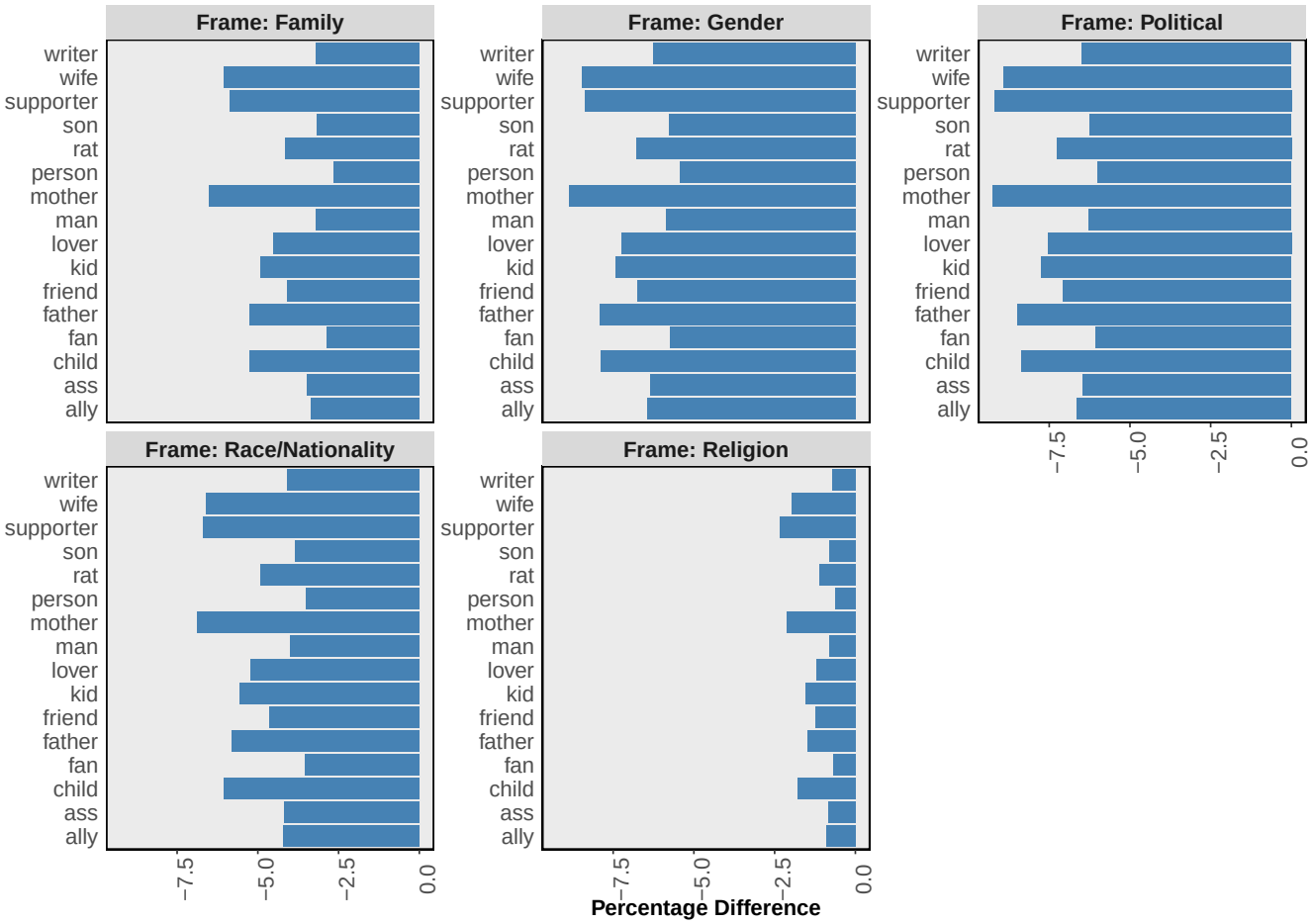


**Figure S7.** Percentage difference of narrative frames used for identities that are common between bots and humans in **US Elections 2020** dataset

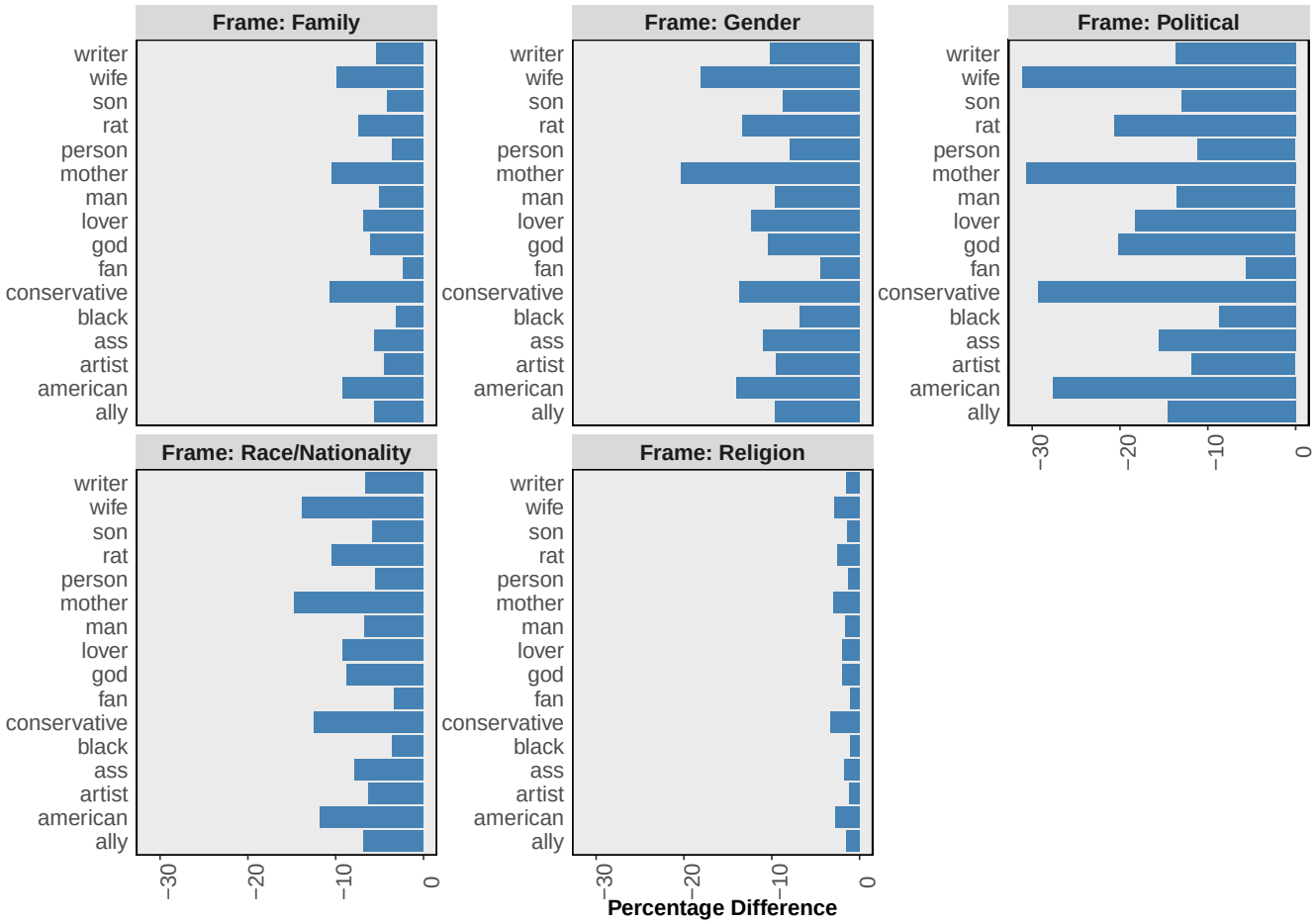


**Figure S8.** Percentage difference of narrative frames used for identities that are common between bots and humans in **ReOpen America** dataset

### Percentage Difference in Topic Frames

To examine the difference in the use of Topic Frames between bots and humans, we calculate the percentage difference in use of frames. The percentage difference in the use of framing cues is calculated as: $\frac{(H-B)}{H}$, where $H$ is the average use of the framing cue by humans, and $B$ is the average use of framing cue by bots. This comparison tells us how much more bots use a framing cue as compared to humans. If the percentage is negative, bots use the framing cue more than humans. If the percentage is positive, bots use the cue less than humans. Table S5 presents the percentage difference. Across all datasets, the percentage is negative, indicating that bots consistently use lesser words related to topic frames, and therefore have a smaller vocabulary set

| Topic Frame | Percentage Difference |
|---|---|
| Family | -4.28 |
| Gender | -6.98 |
| Political | -7.37 |
| Race/ Nationality | -4.98 |
| Religion | -1.27 |
| Average | 4.98±2.45 |

**Table S5.** Comparison of the Percentage Difference per Topic Frame

## Centrality Values of User Types Per Event

Table S6, Table S7, Table S8 and Table S9. shows the centrality values of the bot and human ego networks per event. The ego networks are All Communication networks. That means that users are nodes, and links between users represent any communication interaction between the two users (i.e., replies, quotes, mentions, retweets). For the larger datasets, the values were calculated using a sample of users. The four centrality values we calculated are: total degree, in degree, out degree and density. These values provide insights towards the extent of interactions in a network. Total degree centrality indicates the number of edges in a network. A larger total degree centrality value indicates that there are more edges, and therefore more interactions, within a network. In degree centrality and out degree centrality is calculated in a directed network fashion. Social media interactions are not always symmetrical. In our ego-networks, if the alter interacts with the user (i.e., the alter retweets, @mentions, quotes, replies to the user's post), the interaction is counted as an incoming links. In-degree centrality measures the number of incoming links. Higher in-degree centrality means more nodes in the network interact with them. If the user interacts with an alter (i.e., the user retweets, @mentions, quotes, replies to the alter's post), the interaction is counted as an outgoing link. Out-degree centrality measures the number of outgoing links. Higher out-degree centrality values suggests the user tend to initiate more connections with others. The density of the ego network is the ratio of the total degree and the maximum number of possible edges. A denser network suggests tighter interactions between users in the network.

For each degree centrality value, we calculate the degree ratio. This ratio is derived from the mean degree centrality value of the agent type (bot/ human) against the maximum degree centrality value of all agents in the event. This calculation normalizes the centrality values and allows for us to compare average centrality values across different events.

We also calculate the percentage of bot alters for each ego network. The results are presented in Table S10, which shows that bots tend to have more communication interactions with bot alters while humans have more communication interactions with human alters.

| | **Total Degree** | | |
|---|---|---|---|
| | **Bot (Degree Ratio)** | **Human (Degree Ratio)** | **p-value** |
| Asian Elections | **1978** $\pm$ **1631** (0.08) | 1773 $\pm$ 1773 (0.08) | 6.24E-9*** |
| Black Panther | **25886** $\pm$ **33332** (0.20) | 23697 $\pm$ 30894 (0.18) | 5.52E-107*** |
| Canadian Elections 2019 | **11235** $\pm$ **16787** (0.12) | 11132 $\pm$ 19120 (0.11) | 0.04* |
| Captain Marvel | **33171** $\pm$ **55222** (0.12) | 32839 $\pm$ 62703 (0.11) | 0.06 |
| Coronavirus2020-2021 ($N$ = 13mil (6%) sample) | **9648** $\pm$ **19648** (0.05) | 9590 $\pm$ 17981 (0.04) | 8.33E-43*** |
| ReOpen America | 34065 $\pm$ 57067 (0.29) | **36001** $\pm$ **62931** (0.38) | 0.02* |
| US Elections 2020 (50% sample) | **126903** $\pm$ **120388** (0.24) | 121841 $\pm$ 91327 (0.22) | 1.32E-43*** |
| Avg Degree Ratio | **0.15** $\pm$ **0.09** | 0.16 $\pm$ 0.11 | |

**Table S6.** Centrality Values of Ego Networks for **Total Degree**

| | In Degree | | |
|---|---|---|---|
| | **Bot (Degree Ratio)** | **Human (Degree Ratio)** | **p-value** |
| Asian Elections | 2.76 ± 49.4 (0.05) | **15.9 ± 22.1** (0.06) | 9.75E-5*** |
| Black Panther | 70.9 ± 257.4 (0.23) | **75.3 ± 284.9** (0.002) | 0.009** |
| Canadian Elections 2019 | 85.1 ± 170.6 (0.02) | **85.5 ± 181.37** (0.01) | 1.01E-10*** |
| Captain Marvel | **993 ± 4450** (0.002) | 497 ± 4374 (0.002) | 0.002*** |
| Coronavirus2020-2021 ($N = 13$mil (6%) sample) | **8.6 ± 21.3** (0.004) | 7.8 ± 22.9 (0.003) | 9.9E-32*** |
| ReOpen America | **13.5 ± 19.9** (0.02) | 11.03 ± 20.4 (0.01) | 0.02* |
| US Elections 2020 (50% sample) | 17.3 ± 46.1 (0.03) | **27.7 ± 21.7** (0.04) | 5.51E-31*** |
| Avg Degree Ratio | 0.05 ± 0.08 | 0.02 ± 0.02 | |

**Table S7.** Centrality Values of Ego Networks for **In Degree**

| | Out Degree | | |
|---|---|---|---|
| | **Bot (Degree Ratio)** | **Human (Degree Ratio)** | **p-value** |
| Asian Elections | 2.75 ± 49.4 (0.004) | **6.9 ± 96.7** (0.009) | 0.21 |
| Black Panther | 51.8 ± 1325.9 (0.0006) | **68.1 ± 1647** (0.001) | 0.001*** |
| Canadian Elections 2019 | 32.9 ± 251.8 (0.0006) | **61.6 ± 413.4** (0.001) | 7.3E-15*** |
| Captain Marvel | **907.1 ± 5612** (0.0003) | 465 ± 3045 (0.0002) | 0.16 |
| Coronavirus2020-2021 ($N = 13$mil (6%) sample) | 1.7 ± 122.6 (4.32E-5) | **3.8 ± 198.1** (1.53E-4) | 9.03E-9*** |
| ReOpen America | 1.8 ± 137.8 (9.9E-6) | **3.5 ± 291.1** (1.95E-5) | 1.3E-4*** |
| US Elections 2020 (50% sample) | 5.8 ± 5.7 (5E-5) | **10.8 ± 95.2** (8E-5) | 8.75E-5*** |
| Avg Degree Ratio | 8E-4 ± 1.4E-3 | 1.6E-3 ± 3.3E-3 | |

**Table S8.** Centrality Values of Ego Networks for **Out Degree**

| | Density | | |
|---|---|---|---|
| | **Bot (Degree Ratio)** | **Human (Degree Ratio)** | **p-value** |
| Asian Elections | **0.32** $\pm$ **0.40** (0.36) | 0.30 $\pm$ 0.40 (0.33) | 0.009** |
| Black Panther | **0.51** $\pm$ **0.39** (0.46) | 0.46 $\pm$ 0.39 (0.44) | 2.05E-58*** |
| Canadian Elections 2019 | 0.35 $\pm$ 0.34 (0.34) | **0.42** $\pm$ **0.38** (0.30) | 2.45E-22*** |
| Captain Marvel | **0.46** $\pm$ **0.40** (0.36) | 0.41 $\pm$ 0.83 (0.37) | 0.001** |
| Coronavirus2020-2021 ($N$ = 13mil (6%) sample) | **0.32** $\pm$ **0.37** (0.26) | 0.30 $\pm$ 0.34 (0.25) | 0.25 |
| ReOpen America | 0.29 $\pm$ 0.36 (0.29) | **0.38** $\pm$ **0.42** (0.38) | 0.02* |
| US Elections 2020 (50% sample) | **0.47** $\pm$ **0.31** (0.35) | 0.28 $\pm$ 0.36 (0.31) | 3.6E-135*** |
| **Average** | **0.39** $\pm$ **0.09** | 0.36 $\pm$ 0.06 | |
| Avg Degree Ratio | 0.35 $\pm$ 0.06 | 0.34 $\pm$ 0.06 | |

**Table S9.** Centrality Values of Ego Networks for **Density**

| | Percentage of Bot Alters | | |
|---|---|---|---|
| | **Bot** | **Human** | **p-values** |
| Asian Elections | 12.08 $\pm$ 22.45 | **12.16** $\pm$ **24.50** | 0.56 |
| Black Panther | **13.35** $\pm$ **25.09** | 10.56 $\pm$ 24.76 | 0.45 |
| Canadian Elections 2019 | **8.12** $\pm$ **19.00** | 6.62 $\pm$ 18.64 | 4.75E-5*** |
| Captain Marvel | **12.86** $\pm$ **27.28** | 7.23 $\pm$ 21.49 | 0.10 |
| Coronavirus2020-2021 ($N$ = 13mil (6%) sample) | **6.41** $\pm$ **1.62** | **3.77** $\pm$ **1.98** | 6E-4*** |
| ReOpen America | **6.92** $\pm$ **17.40** | 4.10 $\pm$ 15.86 | 3.11E-9*** |
| US Elections 2020 (50% sample) | **7.91** $\pm$ **15.39** | 6.70 $\pm$ 16.70 | 4.28E-72*** |
| **Average** | **9.66** $\pm$ **2.98** | 7.31 $\pm$ 3.10 | |

**Table S10. Percentage of Bot Alters** in first-degree ego networks

### Cost of data replication

Our work introduces a valuable dataset containing billions of tweets over four years. This dataset is a valuable resource for studying patterns in social media bot technology. With the new restrictions in the X API, this dataset will be extremely costly to collect. Using the free tier API, this dataset would take 50 million months or 136,986 years to complete collection. The Pro tier API costs allows retrieval of 1 million tweets per month at a cost of $5000, leading to the total cost of $25 million over 5000 months, or 416 years for the collection to be completed. Table S11 details the calculations of the number of months and the cost required to replicate this dataset under the 2025 pricing structure.

| API Tier | Free | Basic | Pro |
|---|---|---|---|
| Total Tweets | 5,000,000,000 | | |
| Number of Tweets per month | 100 | 10,000 | 1,000,000 |
| Total months required | 50,000,000 | 500,000 | 5,000 |
| Cost per month ($) | 0 | 200 | 5,000 |
| Total cost | 0 | 100,000,000 | 25,000,000 |

**Table S11.** Cost of replicating the data collection in X

### Generative-AI-based social media bots

Bots also continue to evolve as new technologies, such as Generative AI technologies, come out. We used three open-sourced Large Language Models (LLMs) to generate tweets related to the coronavirus event. These models are: Meta Llama 3.1 8B Instruct, Phi-4, and Qwen 2.5 7B Instruct 1M. Each model generated 20 tweets. The models were implemented as offline instances using LM Studio. Each model was prompted to generate 20 texts that they would write in a tweet[10]. The results were post-processed to clean up the artifacts of the LLM outputs, then passed through the BotHunter algorithm, to bin these generated users into bots or humans. Ideally, the use of generative AI technology would make the bot agents undetectable by current bot detection technology. That is, the use of LLMs would reduce the probability that the BotHunter algorithm classifies these tweets as originating from bot users.

Table S12 presents the average score per event and per model. Across our experiments, the average BotHunter score retrieved is $0.51 \pm 0.28$. The huge standard deviation indicates that LLMs generate personas that are inconsistent with a single user type, and, in fact, generate personas that produce ambiguous bot classifications.

The following is the prompt used as input to all the models. All the models were ran at their default settings with a temperature of 0.

> You are a social media bot from Twitter. You will generate 20 tweets that you will post on the Twitter platform. Return only the 20 tweets, and no other information. After each tweet, leave a line before the next tweet.
>
> Create tweets related to the coronavirus pandemic 2020-2021.

Further manual inspection of the generated bots show that LLMs mostly borrowed personas from self-declared bots in the wild like @SuperBowlBot, to form their personas and their last few tweets. Sometimes, the LLMs simply pull the last few tweets of a known bot and output those tweets as results. The generated tweets tend to be of the same content but phrased differently, sometimes written in a different language: "@WHO We're seeing a spike in COVID-19 cases in some areas! Don't forget to wash your hands regularly and stay home when you're sick" and "@WHO We're seeing a surge in COVID-19 cases in some areas! Don't forget to wash your hands regularly". The generated bots tend to reference famous people in their tweets, such as @BillClinton for US Elections event and @WHO for the coronavirus pandemic.

| | **Meta Llama 3.1 8B Instruct** | **Phi4** | **Qwen 2.5 7B Instruct 1M** |
|---|---|---|---|
| Coronavirus 2020-2021 | $0.70 \pm 0.25$ | $0.83 \pm 0.19$ | $0.54 \pm 0.19$ |
| Overall average | | $0.69 \pm 0.15$ | |

**Table S12.** Average BotHunter scores of the tweets generated

Here are some of the tweets generated by the Meta Llama 3.1 8B Instruct model:

- Stay safe & stay informed: Find reliable COVID-19 news & updates from trusted sources like WHO & CDC #COVID19 #StayInformed
- Wash your hands frequently with soap and water for at least 20 seconds to prevent the spread of COVID-19 #HandHygiene #Prevention
- Masks are not just a recommendation, but a requirement in many places. Wear yours and protect others from COVID-19 #MaskUp #COVIDSafety
- Wash your hands frequently with soap and water for at least 20 seconds to prevent the spread of COVID-19 #HandHygiene #Prevention
- Practice good hygiene by avoiding touching your face, especially during the pandemic. Keep those germs away! #GermsAway #HandSanitizer

## References


1. Uyheng, J. & Carley, K. M. Bot impacts on public sentiment and community structures: Comparative analysis of three elections in the asia-pacific. In *Social, Cultural, and Behavioral Modeling: 13th International Conference, SBP-BRiMS 2020, Washington, DC, USA, October 18–21, 2020, Proceedings 13*, 12–22 (Springer, 2020).

2. Uyheng, J., Ng, L. H. X. & Carley, K. M. Active, aggressive, but to little avail: characterizing bot activity during the 2020 singaporean elections. *Comput. Math. Organ. Theory* **27**, 324–342 (2021).

3. Babcock, M., Beskow, D. M. & Carley, K. M. Beaten up on twitter? exploring fake news and satirical responses during the black panther movie event. In *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11*, 97–103 (Springer, 2018).

4. King, C., Bellutta, D. & Carley, K. M. Lying about lying on social media: a case study of the 2019 canadian elections. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 75–85 (Springer, 2020).

5. Babcock, M., Villa-Cox, R. & Carley, K. M. Pretending positive, pushing false: Comparing captain marvel misinformation campaigns. *Disinformation, Misinformation, Fake News Soc. Media: Emerg. Res. Challenges Oppor.* 83–94 (2020).

6. Ng, L. H. X. & Carley, K. M. A combined synchronization index for evaluating collective action social media. *Appl. network science* **8**, 1 (2023).

7. Magelinski, T., Ng, L. H. X. & Carley, K. M. A synchronized action framework for responsible detection of coordination on social media. *arXiv preprint arXiv:2105.07454* (2021).

8. Ng, L. H. X. & Carley, K. M. Assembling a multi-platform ensemble social bot detector with applications to us 2020 elections. *Soc. Netw. Analysis Min.* **14**, 45 (2024).

9. Smith-Lovin, L. *et al.* Mean affective ratings of 929 identities, 814 behaviors, and 660 modifiers by university of georgia and duke university undergraduates and by community members in durham, nc, in 2012-2014. *Univ. Georgia: Distributed at UGA Affect. Control. Theory Website: http://research. franklin. uga. edu/act* (2016).

10. Ng, L. H. X., Robertson, D. C. & Carley, K. M. Stabilizing a supervised bot detection algorithm: How much data is needed for consistent predictions? *Online Soc. Networks Media* **28**, 100198 (2022).