S1 Appendix: Dataset availability and supporting details

Data Availability

The reddit comment dataset was collected by Reddit user ‘Stuck In The Matrix’ using the official Reddit API. The post describing the dataset and its collection is available (as of 4 Nov 2019) at: https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/

The dataset has been widely used for research purposes, and is available via academic torrents [1,2] at: http://academictorrents.com/details/7690f71ea949b868080401c749e878f98de34d3d/tech

Please note that other versions of the dataset (such as those hosted by pushshift.io) may differ from the archived version used in our paper. Our ethics clearance does not permit hosting more specific intermediate versions of the dataset which might be able to identify particular users. Should academictorrents become unavailable in the future, please contact the authors and we will make reasonable arrangements to transmit the dataset.


Reddit Demographics:
Demographic information was taken from the 2016 Reddit Demographics survey (the most recent available at the time of writing). A post describing this information can be found at: https://www.reddit.com/r/dataisbeautiful/comments/5700sj/octhe_results_of_the_reddit_demographics_survey/

The following demographic information is a description of the graphics associated with the above post, should they become unavailable in the future:

Age:
- 18-25  56.7%
- 26-35  22.7%
- <17    14.4%
- Other  6.2%

Gender:
- Male:  63.1%
- Female: 33.3%
- Other:  3.6%

Ethnicity:
- White: 81.8%
- Other: 18.2%
Bot Removal:

The list of known bots was downloaded from:

https://www.reddit.com/r/botwatch/comments/1xojwh/list_of_320_reddit_bots/

And combined with a list of users whose forum diversity (described in the text) was 15 Standard Deviations above the mean. The list of 466 usernames culled as bots is as follows:

- A858DE45F56D9BC9
- AAbot
- ADDBot
- albumbot
- allinonebot
- AICodeBot
- ALTcointip
- annoying_yes_bot
- Antiracism_Bot
- ApiContraption
- A_random.gif
- asmrsambot
- AssHatBot
- astro-bot
- AthelismModBot
- auto-doge
- AutoInsult
- automodbot
- autowikibot
- bad_ball_ban_bot
- BadLinguisticsBot
- BanishedBot
- ban_pruner
- baseball_gif_bot
- beecointipbot
- BeetusBot
- BELITipBot
- BensonTheBot
- Bible_Verses_Bot
- bitcoinpartybot
- bitcointip
- bitofnewsbot
- BlackjackBot
- BlockchainBot
- bocketbot
- Brigade_Bot
- bRMT_Bot
- Bronze-Bot
- c5bot
- CAH_BLACK_BOT
- Cakeday-Bot
- callfloodbot
- callibot
- CalvinBot
- canada_goose_tip_bot
- CaptionBot
- CarterDugSubLinkBot
- CasualMetricBot
- changetip
- CHART_BOT
- cheesecakebot
- Chemistry_Bot
- ChristianityBot
- chromobot
- classbot
- CLOSING_PARENTHESIS
- Codebreakerbreaker
- coinflipbot
- coinflytippertime
- colorcodebot
- Comment_Codebreaker
- comment_copier_bot
- compilebot
- ComplimentingBot
- conspirobot
- CPTModBot
- cRedditBot
- creepier_smilebot
- CreepierSmileBot
- CreepySmileBot
- CuteBot6969
- DBotIndia
- ddibot
- define_bot
- DefinitelyBot
- _Definition_Bot_
- demobilizer
- dgctipbot
- Dickish_Bot_Bot
- Dictionary_Bot
- DidSomeoneSayBoobs
- digitipbot
- disapprovalbot
- DNotesTip
- DogeLotteryModBot
- dogetipbot
- DogeTipStatsBot
- DogeWordCloudBot
- DotsaCastingBot
- Downnotes_Plz
- DownvotesMcGoats
- DRKTipBot
- Dropbox_Bot
- earthtipbot
- edmprobot
- _ElMatadero_bot
- elwh392
- EmmaBot
- Epic_Face_Bot
- EscapistVideoBot
- _FakeryBot_
- faketipbot
- ExplanationBot
- fact_check_bot
- feature_bot
- fedora_tipbot
- fedora TipBotAutoBot
- fedora_tip_bot
- fedyrid
- FernandezNazi
- FGT_Bot
- frawbot
- frawbot
- fa_mirror
- fedoratips
- FelineFacts
- Fixes_GrammerNazi
- flask
- flips_title
- foreigneducationbot
- FriendlyCamelCaseBot
- FriendSafariBot
- FrontpageWatch
- Frown_Bot
- frytipbot
- factotipbot
- FTTY_CAT
- GabenCoinTipBot
- gabenizer_bot
- gabentipbot
- GameDealsBot
- gausfakbot
- gards_you_boobies
- GrammerNazi
- GreasyBacon
- gregbot
- grompbot
- Grumblor_bot
- gunners gift_bot
- GunnersGifsBot
- GunnitBot
- GWHistoryBot
- haiku_robot
- Handy_Related_Sub
- havock_bot
- HCE_Replacement_Bot
- hearing aid_bot
- hearing aid_bot
- hearingaid_bot
- HighResImageFinder
- hit_bot
- hockey gif_bot
- HockeyGT_Bot
- HowIsThisBestOf_Bot
- howstat
- HScard display_bot
- hwabot
- I_BTCI_N CATS
- ICouldntCareLessBot
- IAgreeBot
- imgurHostBot
- imgur_rehosting
- imgr_transcriber
- imr efficiently
- Insane_Photo_Bot
- I_Say_No
- IsItDownBot
- IS_IT_SOLVED
- isitupbot
- jerkbot-3hunna
- JiffyBot