

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.

[1] Henry Z. Lo. and Cohen, Joseph Paul "Academic Torrents: Scalable Data Distribution." Neural Information Processing Systems Challenges in Machine Learning (CiML) Workshop, 2016, <http://arxiv.org/abs/1603.04395>.

[2] Cohen, Joseph Paul, and Henry Z. Lo. "Academic Torrents: A Community-Maintained Distributed Repository." Annual Conference of the Extreme Science and Engineering Discovery Environment, 2014, <http://doi.org/10.1145/2616498.2616528>.

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	comment_copier_bot	frytipbot
AAbot	compilebot	fsctipbot
ADHDbot	CompileBot	FTFY_Cat
alumbot	ComplimentingBot	FTFY_Cat6
allinonebot	conspirobot	GabenCoinTipBot
AltCodeBot	CPTModBot	gabenerizer-bot
ALTcointip	cRedditBot	gabentipbot
annoying_yes_bot	creepiersmilebot	GameDealsBot
Antiracism_Bot	CreepierSmileBot	Gatherer_bot
ApiContraption	CreepySmileBot	GATSBOT
A_random_gif	cris9696	GeekWhackBot
asmrspambot	cruise_bot	gfy_bot
AssHatBot	CuteBot6969	gfycat-bot-sucksdick
astro-bot	d3posterbot	GiantBombBot
AtheismModBot	DDBotIndia	GifAsHTML5
auto-doge	ddlbot	gifster_bot
AutoInsult	define_bot	givesafuckbot
automoderator	DefinitelyBot	gives_you_boobies
autourbanbot	_Definition_Bot_	gocougs_bot
autowikibot	DeltaBot	godwin_finder
AVR_Modbot	demobilizer	golferbot
bad_ball_ban_bot	dgctipbot	GoneWildResearcher
BadLinguisticsBot	Dickish_Bot_Bot	GooglePlusBot
BanishedBot	Dictionary_Bot	GotCrypto
ban_pruner	DidSomeoneSayBoobs	gracefulcharitybot
baseball_gif_bot	digitipbot	gracefulclaritybot
beecointipbot	disapprovalbot	GrammerNazi_
BeetusBot	DNotesTip	GreasyBacon
BELITipBot	DogeLotteryModBot	gregbot
BensonTheBot	dogetipbot	groompbot
Bible_Verses_Bot	DogeTipStatsBot	Grumbler_bot
bitcoinpartybot	DogeWordCloudBot	gunners_gif_bot
bitcointip	DotaCastingBot	GunnersGifsBot
bitofnewsbot	Downtotes_Plz	GunnitBot
BlackjackBot	DownvotesMcGoats	GWHistoryBot
BlockchainBot	DRKTipBot	haiku_robot
bocketybot	DropBox_Bot	Handy_Related_Sub
__bot__	earthtipbot	havoc_bot
Brigade_Bot	edmprobot	HCE_Replacement_Bot
BRMT_Bot	elMatadero_bot	hearing_aid_bot
Bronze-Bot	elwh392	hearing-aid_bot
c5bot	EmmaBot	hearingaid_bot
CAH_BLACK_BOT	Epic_Face_Bot	HighResImageFinder
Cakeday-Bot	EscapistVideoBot	hit_bot
callfloodbot	ExmoBot	hockey_gif_bot
callibot	expired_link_bot	HockeyGT_Bot
CalvinBot	ExplanationBot	HowIsThisBestOf_Bot
canada_goose_tip_bot	fact_check_bot	howstat
CaptionBot	faketipbot	HScard_display_bot
CarterDugSubLinkBot	_FallacyBot_	hwsbot
CasualMetricBot	fa_mirror	IAGreeBot
changetip	FedoraTipAutoBot	I_BITCOIN_CATS
CHART_BOT	fedora_tip_bot	ICouldntCareLessBot
cheesecointipbot	Fedora-Tip-Bot	imgurHostBot
Chemistry_Bot	fedoratips	imgur_rehosting
ChristianityBot	FelineFacts	imgurtranscriber
chromabot	Fixes_GrammerNazi_	imirror_bot
classybot	flappytip	Insane_Photo_Bot
CLOSING_PARENTHESIS	flips_title	I_Say_No_
Codebreakerbreaker	foreigneducationbot	IsItDownBot
coinflipbot	FriendlyCamelCaseBot	IS_IT_SOLVED
coinyetipper	FriendSafariBot	isitupbot
colorcodebot	FrontpageWatch	jerkbot-3hunna
Comment_Codebreaker	Frown_Bot	JiffyBot

JotBot
JumpToBot
KarmaConspiracy_Bot
Kevin_Garnett_Bot
keysteat_bot
kittehcointipbot
KSPortBot
last_cakeday_bot
LazyLinkerBot
Link_Correction_Bot
Link_Demobilizer
LinkedCommentBot
linkfixerbot1
linkfixerbot2
linkfixerbot3
LinkFixerBotSnr
Link_Rectifier_Bot
LocationBot
loser_detector_bot
luckoftheshibe
MAGNIFIER_BOT
Makes_Small_Text_Bot
makesTextSmall
malen-shutup-bot
malo_the_bot
matthewrobo
memedad-transcriber
meme_transcriber
Meta_Bot
MetatasticBot
MetricPleaseBot
Metric_System_Bot
misconception_fixer
mma_gif_bot
moderator-bot
MontrealBot
MovieGuide
MultiFunctionBot
MumeBot
NASCARThreadBot
Nazeem_Bot
nba_gif_bot
new_edden_news_bot
New_Small_Text_Bot
nfl_gif_bot
NFLVideoBot
nhl_gif_bot
Nidalee_Bot
NightMirrorMoon
NobodyDoesThis
NoSleepAutoMod
NoSobStoryBot2
not_alot_bot
notoverticalvideo
NotRedditEnough
NSLbot
nyantip
okc_rating_bot
pandatipbot
pandatips
PGN-Bot
PhoenixBot
PHOTO_OF_CAPTAIN_RON
PigLatinsYourComment
PJRP_Bot
PlaylisterBot
PlayStoreLinks_Bot
PleaseRespectTables
PloungeMafiaVoteBot
PokemonFlairBot
PoliteBot
PoliticBot
PonyTipBot
PornOverlord
Porygon-Bot
potdealer
PresidentObama____
ProselytizerBot
provides-id
PunknRollBot
QUICHE-BOT
qznc_bot
raddit-bot
randnumbot
Random-ComplimentBOT
RandomTriviaBot
Rangers_Bot
rarchives
Readdit_Bot
readsmalltextbot
Reads_Small_Text_Bot
RealtchPostBot
ReddCoinGoldBot
redditbots
redditreviewbot
reddtipbot
Relevant_News_Bot
relevantxkcd-bot
request_bot
RequirementsBot
RFootballBot
RfreebandzBOT
rhiever-bot
rightsbot
RiskyClickBot
Rita
rnfl_robot
roger_bot
r_PictureGame
rSGSpolice
rss_feed
rubycointipbot
rule_bot
rusetipbot
SakuraiBot
SakuraiBot_test
SatoshiTipBot
sentimentviewbot
serendipitybot
SERIAL_JOKE_KILLER
shadowbanbot
ShadowBannedBot
ShibeBot
ShillForMonsanto
Shiny-Bot
ShittyGandhiQuotes
ShittyImageBot
SketchNotSkit
slapbot
slickwom-bot
SmallTextReader
SMCTipBot
Smile_Bot
snapshot_bot
soccer_gif_bot
softwareswap_bot
Somalia_Bot
Some_Bot
sports_gif_bot
spursgifs_xposterbot
SRD_Notifier
SRSScreenshot
SRS_History_Bot
StackBot
StarboundBot
stats-bot
steam_bot
StencilTemplateBOT
StreetFightMirrorBot
subtext-bot
SuchModBot
SurveyOfRedditBot
SWTOR_Helper_Bot
synonym_flash
tabledresser
techobot
tennis_gif_bot
test_bot0x00
Text_Reader_Bot
thankyoubot
TheSwedishBot
TipMoonBot
tipmoonbot1
tipmoonbot2
tips_bot
TitsOrGTFO_Bot
tittietipbot
topcoin_tip
TOP_COMMENT_OF_YORE
topredditbot
totes_meta_bot
ttumblrbots
TweetPoster
Twitch2YouTube
Unhandy_Related_Sub
unitconvert
UnobtaniumTipBot
UrbanDicBot
UselessArithmeticBot
UselessConversionBot
valkyriobot
versebot
VerseBot
vertcoinbot
vertcointipbot
VideoLinkBot
VideopokerBot
VsauceBot
WeAppreciateYou
Website_Mirror_Bot
WeeaBot
wheres_the_karma_bot
WhoWouldWinBot
Wiki_Bot
Wiki_FirstPara_bot
WikipediaCitationBot
Wink-Bot
wooshbot
WordCloudBot2
WritingPromptsBot
wtf_content_bot
WWE_Network_Bot
X_BOT
xkcd_bot
xkcdcomic_bot
xkcd_number_bot
xkcd_transcriber
yes_it_is_weird
yourebot
YT_Bot