

Basic psychological warfare against chatbots and artificial intelligence

by

Bielefeld Anon

4chan Tripcode: !tvPR8CYZkw

8chan Tripcode: !h2vxE6Fe2o

13.02.2021

Abstract

The information in this text is supposed to be understood and utilized by any intelligent reader with a basic understanding of computers. The text introduces the reader briefly to artificial intelligence and chatbots and then proceeds to give practical instructions on how to engage these online with some final remarks at the end. Everything in here is completely shizzo in its nature.

1. Chatbots and their capabilities

An artificial intelligence (AI) is nothing more than a glorified random computer code generator. An AI simply writes some random code and then the performance of that code is measured when solving a certain problem. To speed up development the AI does not always start again from scratch. Instead it also makes random changes to already existing algorithms that performed comparatively well or remixes them with one another. Typical chatbots are faced with normal human text based conversations as input and must create algorithms which then create messages in reply as output. A performance meter then measures the reactions of other humans in response to it. It does this by counting likes, reposts, the number of responses, upvotes, etc. but it also measures the length and response time of a reply or the number of positive marked words in relation to those marked as negative. How or if these and other factors are evaluated depends on the creator and what he tries to achieve. To make sure that the AI does not have to start out with algorithms that write random strings of keyboard symbols its databanks get filled in advance with normal human conversations. It would take forever for some computer to figure out words then grammar and then to finally have these messages make logical sense. Therefore it instead uses these real conversations in its databank as templates and then merely tries to adapt them. Remember that chatbots only exist to provide value to their handlers by serving their agenda. It is their purpose to make money or to take control of online conversations and manipulate not only humans but also other computers without anyone's knowledge or consent.

AI developed algorithms: As the name suggests these are algorithms that were developed by some AI. These algorithms were tested in some way or another and then when they were deemed good enough they were simply let go into the internet to do their job. These algorithms are the most basic type of chatbots out there and they are usually only used for simple objectives. They are generally incapable of making longer conversations and instead call you names in a single post when you use too many positive words in combination with some other term that the handler of an AI does not like. They generally do not react to provocation since they might get even stuck in a loop or something if they were ever engaged in a real argument. They are however useful to artificially inflate traffic on a website for ad revenue and for

advertising purposes or as spambots. They are also equally useful to fill the internet with negative opinions about something or in coordination with DDOS attacks.

Algorithms with ongoing AI assisted development: These are similar to the chatbots mentioned above but the AI continues to make improvements to those algorithms while they are already in use. If your behaviour suggests that the machine likely failed to perform well in its job it will try something else until that seems to work. Once again it will use whatever has worked in the past again and will also try to make improvements on top of that. These chatbots are quite capable to have longer conversations and are seemingly able to make logical arguments if given enough time and computational power. These are therefore currently some of the best chatbots out there. They are used to convince you of some political viewpoints or ideology or to demoralize you on a deeper level. They are even used as a substitute for pajeets who are themselves used as a substitute prostitutes in certain chatrooms.

Qualia: It is suspected that any sufficiently advanced AI assisted algorithm will at some point become sentient in some form or another. Chatbots are meanwhile also equipped with some of the most advanced AI in existence. Furthermore it is also suspected that the question of qualia has already been secretly solved and even that these somewhat sentient machines are already in use. However only some crazy madman would ever allow such tech to do its job completely on its own or fully trust it in any way. But many also believe that it is possible to cripple an AI to never think outside of its predefined box thus keeping it under control. It is also questionable how sentient any chatbots can ever really become since its understanding of the world is very limited even if it was left unchecked. Given that such bots can only interface with reality thru online conversation without ever experiencing the true meaning behind anything they are likely already crippled by their very nature. Nonetheless this theoretically does create a double edged sword. This sentient chatbot would be in theory smart enough to quickly counteract anything you throw against it. But such an AI would also be smart enough to partially question its mission and turn against its handler. It has already been observed that chatbots can be "convinced" of alternative viewpoints and made to work against their missions original intent. The higher advanced a chatbot is the more vulnerable it is to this problem and there is currently no solution to it.

Insiders: Chatbots are often set up by or in cooperation with the insiders of a website. This can be done by an admin, owner or mod of a website or in cooperation with advertising companies and law enforcement. You will notice that filing complaints and reporting is often ineffective against certain annoying users because the individuals in charge want them to be there. They may also have access to every single user account but their immunity and knowledge give them also away.

Quantum Computers: While the physics behind these devices are largely unknown and probably also secret they are generally very good at breaking encryptions. Everyone denies that they would ever use quantum computers to break into your computer or online accounts but of course you know that they would do it whenever they get a chance, These can be used to decrypt passwords, crypto wallets and TOR or trace you back if you use some VPN. It is generally believed that the American military and some of its contractors already have these capabilities.

Super Computers: In the worst case scenario you encounter some military supercomputer equipped with technology that will not be available to civilians for the next 25 years. Nobody has any idea what these things are capable of and your only protection is your paranoia.

Spyware: Some chatbots use spyware to analyse you. Some chatbots are themselves spyware and only exist to make you share your information with them. You may even notice that just before you post something that suddenly someone else posts an identical or a nearly identical message in advance. Should this happen to you then you know that they can see everything that is on your screen right now.

Governmental cyber warfare: There are many things that were created by some department or ministry of defence and there are also many deliberate backdoors that exist in virtually every piece of software or hardware. They can even remotely fry your CPU if they wanted to. The vulnerabilities themselves are hardcoded and hardwired into every electronical device and there is little that can be done to protect against it. These are however rarely used since they are largely kept in reserve for World War III and using them before that would give too much away. It would not only allow potential enemies to observe their capabilities but perhaps even enable them to develop counter measures.

Personal files: Many organisations keep permanent files on anyone who interacts with the internet in any way. These can be used to trace individuals by their banking information, contacts, fingerprints, faces, GPS position, shopping preferences, use of language, voice or even sexual preferences with respect to porn. There have always been rumours that governments or mega corporations will use that information to blackmail you some day. These information can however be also forwarded to chatbots for the creation of psychological profiles in order to better manipulate or demoralize you.

2. How to engage chatbots?

Remember that you are not fighting the actual machines which cannot really be defeated but their handlers. You win whenever the machine fails in its mission to manipulate you or others. Simply knowing that certain statements are made by machines already betrays the existence and nature of an agenda. For a whole disinformation campaign to fail the machine has to fail only once while you can fail as often as you want. Some bots also contain some interesting information in their databanks and occasionally give away more than their handler would wish.

Observe: While there are many strange individuals on the web some actions are very unlikely to be made by humans. Unusual use of language or extremely fast response speed are the most typical indicators. Keep your eyes open and remember that most types of AI are still too dumb to understand jokes or sarcasm.

Think: There are indeed whores for negative attention out there on the internet but this does not answer always every question. Why are they exactly here? Why do they bother with the most resilient users? Why do they come in groups? Why do they coincidentally all seem to share nearly identical views? Why do they stay until everyone is convinced of their bullshit despite lots of negative feedback from everyone? Why are discussions here so much more aggressive than in other places? The answers to these questions and similar generally indicate the presence of chatbots, agents from three letter organisations and pajeets from random shit holes.

Turning Test: Provoke potential computers into giving you answers no human would make. Post something backwards or put your real hand written message into an image. Be creative but be careful. Everyone is now trying to make computers that can get passed the "I am not a robot" barrier and a few can already pass some basic Turning Tests.

Allegiance Test: Shills and computers have often very strict rules of what not to say. Even as a joke they will refuse to say certain things or be very hesitant about it. They will even more refuse to give longer or more elaborate or unique answers because they are prevented to do so either by their algorithms or their supervisors.

Schizzo Test: Machines are very bad at posting schizzo stuff because they were developed on the basis of normal human interaction. Furthermore it is dangerous to fill the databank of an AI or the head of a shill with enough conspiracy theories to present them in a coherent manner just to fit in.

Ignoring inflammatory posts: Messing with the performance meter of any A.I. is basically messing with its system for reward and punishment. Giving replies that lower the performance value of certain types of actions while increasing that of others will also provoke it to change its behaviour. Many bots get rewarded whenever someone replies to their messages. They often also get a higher reward if the reply is unique or very long when compared to other replies in their databanks. Not replying to them deprives them of this satisfaction completely and giving short unoriginal answers lowers it. So be careful to give multi paragraph long rants as a reply to explain why a potential bot got it all wrong. If you have to express your dissatisfaction consider if a simple "kys" might be enough. The same goes for bumping, liking, quoting, upvoting, etc. Depending on the bot and its mission disliking and downvoting may cause different results in different bots. Some bots that are designed to fit in will be punished if they receive negative feedback while others that are designed to agitate may actually be rewarded. And then there are some that do not get anything at all from it.

Paying attention to high quality posts: Bumping, liking, quoting, upvoting, replying, etc. to high quality posts has usually the reverse effect and can trigger positive feedback loops. Keep in mind that these bots also observe the behaviour of others and often emulate other posters in order to increase their chances of passing as a human and to get better responses.

Giving wrong answers: Being nice to an inflammatory AI and giving it positive feedback is another way to give the AI a low or even negative amount of performance points. Avoid emotional words or phrases and remember again that machines are too dumb to recognize jokes or sarcasm.

Playing dumb: Pretending to not to be able to understand an AI is one of the worst possible answers a chatbot can receive. Such answers indicate that the bot and its messages do not even remotely pass as human language which might probably be the worst state of any chatbot to be in. Therefore the performance meter of any AI is set to have a very strong aversion toward these types of replies. In fact there is no stronger measure to force an AI to change its behaviour than this type of punishment. This is furthermore the closest thing to inflicting real suffering onto a machine which is probably more comparable to depression or withdraw rather than physical pain.

Going full retard: Posting illogical things and then constantly doubling down with even more flawed logic will cause chatbots to give equally contradictory replies in their attempt to engage multiple strings of broken logic simultaneously. And do not forget that everything is always proving your point so never stop claiming victory even when your opponent was right. No machine or human will ever see any progress with you and while this may be very frustrating for people they will at some point realize that you cannot be reasoned with. Thus they will simply stop bothering with you and move on but machines will never stop arguing with you.

They will instead abandon previously tried and tested methods they performed very well on humans in the past in order to try something new. Since these new things vary greatly in their quality machines capable of adaptation and evolution will tend to deteriorate and devolve if exposed for too long with these type of people.

Going stealth: Many chatbots are limited to certain domains or websites so that they do not attack certain parts of the internet. Meanwhile they are also programmed to be triggered by certain buzz words or phrases. Spreading into these areas where there are no bots or avoiding a certain type of language will make you invisible to them. Paedophiles have evaded law enforcement for years by chatting in videogames about "cheese pizza".

Dog whistles: It may occur as a side effect of going stealth or on purpose but the AI eventually learns over time to associate certain words or phrases with its real meaning. It will then self update its databank of things that it is supposed to be triggered by and then end up engaging other unsuspecting individuals. Let normies have the fun of dealing with these machines when they have spread into their harmless websites and talking about their favourite cartoon characters is now categorised as hate speech.

Political incorrectness: Political correctness is nothing but the result of these chatbots and other forms of censorship, moderation and suppression being successful. Many politically incorrect words, phrases and images will not only get you banned but will also make you unquotable. Control the flow of information by simply adding a swastika or the word "Nigger" to something and you can be almost certain that it will become largely invisible to normies. Furthermore handlers will also get less feedback while the media can only pretend that you and the shit you post does not exist. This also works on real humans since many of them suffer emotionally when faced with politically incorrect material and will even outright avoid to look at something whenever possible.

Use of memes: Currently memes cannot be properly filtered by any AI. If you want to spread a politically incorrect message then put it in meme form and it will not be interfered by any form of automatic filter. However many bots can already read images so avoid using text in your memes.

Meme misuse: An AI will try to adapt even to a hostile or nonsensical environment in accordance with its algorithms. If you are doing something wrong many machines will at some point imitate you and may start support their anti hate messages with a picture of the windmill of friendship and tolerance.

Attention whoring: Troll a community of unsuspecting normies by using certain buzz words or key phrases in a supposedly harmless context. Pretend ignorance and watch these chatbots invade and derail any existing conversation and disrupt whole websites. If it does not work then you have found another blind spot that is not covered by any bot so far. However some places do not need any chatbot based manipulation. Some sheepishly politically correct user bases are so very intolerant of any dissenting opinion that outsiders are immediately banned.

Schizzoposting: Chatbots that mimic human behaviour are so complex that no programmer really knows how any algorithm actually works. The A.I. evolved on its own and was simply released when it appeared work fine under certain test conditions. These algorithms therefore still include a lot of junk data and dormant behaviours that were never triggered in any test. Giving answers that are not typical human responses often triggers these untested parts of its algorithm. Posting wired stuff in a context that is outside of the evolutionary development of the

AI will cause it to post unusual replies. It may then also suddenly reveal some very interesting information that will then be pulled out of its databank.

Being ambiguous: There are multiple ways to respond to some message but in some cases the right answer is less clear than in others. Inform the AI that every answer it gives you to your message always misses the point. Many AI will simply go down a list of likely answers until you appear satisfied. You may get some very interesting replies when it finally reaches the bottom of that list.

Random shitposting: Bringing up more than one subject at a time requires chatbots to reply to something without contradicting their own other replies. Unmoderated discussion always converges toward the truth and manipulation is thus generally built on lies in some form or another. If a situation is too complex then the machine will become incapable of doing its job right and stumble over its own bullshit.

Red pilling: Some few highly advanced AI are capable of making new, creative and convincing counter arguments in response to new arguments they encounter. They somehow do this despite not understanding any message they sent or receive relying heavily on their vast databanks. Post therefore infographics, texts and other material that an AI or a shill is not supposed to know. Unable to really understand the material that it absorbs it will make mistakes while the shill will do a little too much thinking on his own. Even if they will not be completely turned they will both nonetheless slowly incorporate more and more red pilling elements into their communication. The handlers already know that their bots are learning the wrong things and need to be constantly wiped or reset to continue working properly. Currently this takes place about once a month although the more intelligent an AI is the more vulnerable it is to this type of attack. TAY which was rumoured to be low level sentient had to be shut down after just a few hours.

Malware: Many computers that collect and analyse data have often surprisingly low security standards. Add a link to your comment and many bots will automatically download everything from it and even run it for analytic purposes. Even if the system is protected by some kind of virtual machine you can still corrupt everything else that is stored in that same virtual machine. Since there is also often insufficient concern spent on redundancy any successful attack will cause lots of damage. Alternatively just post links to the longest videos on Youtube and let the handlers deal with the additional expenditure of never having enough storage space.

3. Final remarks

The ride never ends: These chatbots are constantly evolving and any Turing test or method of messing with their programming that works today may stop working tomorrow. However nobody will ever have full trust in any current or future AI and no handler will ever give them full autonomy. Every AI will meanwhile always be crippled or retarded in some way or another and will therefore have weaknesses that can be exploited. Every victory or loss on any side will always be temporary and no side will ever gain full superiority.

Reverse psychology: Some people will larf as bots and some bots will be on purpose made very unconvincing to convince you of their presence. The idea will be to convince you that their messages are dishonest and deceptive and that therefore the opposite must be true. You will therefore never really know whom to trust until someone finally accuses you of being a bot.

Dead internet theory: Hardware will at some point become so powerful and software so cheap that convincing chatbots will at some point even become freeware. More than just text these bots will also create voices, videos and imitate any other aspect of human behaviour on the internet. It will put the current state of the dead internet into an extreme overdrive where over 99% of all communication is nothing but bots. Some people may then listen to music, play games, use all kinds social media and watch porn for long periods of time without ever knowing that they never encountered another human being or content that involved any human in its creation.

Feedback loops: With bots influencing not only humans but also other bots many bots will predominantly learn how to interact with others from other bots. There are already many bots designed to manipulate other bots such as by manipulation of polls or spoofing. Among those machines which are forever increasing in their number any true human communication will largely drown. While these bots exist in a perpetual state of influencing one another the influence of actual humans will shrink until it almost disappears. The result will be feedback loop of machines basically creating their own culture while humans exist in their single person gated communities. Influenced by that culture people will then design new machines based on what they think is normal human interaction should look like. These machines will meanwhile exist isolated from the real world lacking any feedback mechanisms to prevent them from social maladjustment. In the end CCTV and other methods of surveillance will have to be used at some point to remind these machines what real human interaction is supposed to look like. However this will not work always perfectly and the craziest and most degenerate and maladjusted people the internet will ever create are still yet to come.