Hello, Twitter Bot!: Towards a Bot Ethics of Response and Responsibility

Line Henriksen
Malmö University
Line.henriksen@mau.se

Cancan Wang
IT University of Copenhagen
cawa@itu.dk

Abstract

In this paper, we explore the troubles and potentials at stake in the developments and deployments of lively technologies like Twitter bots, and how they challenge traditional ideas of ethical responsibility. We suggest that there is a tendency for bot ethics to revolve around the desire to differentiate between bot and human, which does not address what we understand to be the cultural anxieties at stake in the blurring boundaries between human and technology. Here we take some tentative steps towards rethinking and reimagining bot-human relationships through a feminist ethics of responsibility as response by taking as our starting point our own experience with bot creation, the Twitter bot "Hello30762308." The bot was designed to respond with a "hello" to other Twitter users' #hello, but quickly went in directions not intended by its creators.

Keywords
Social bots, ethics, responsibility, digital technology, Twitter

Introduction

Today, social bots—automated algorithms in online social networks that are able to perform tasks without direct human involvement—can be spotted everywhere across the internet (Hwang, Pearce, and Nanis 2012; Ferrara et al. 2016; de Lima Salge and Berente 2017). On Twitter, Facebook, or TikTok, one can find millions of bots liking, following, commenting, sometimes even posting their own content and buying stuff on their own (de Lima Salge and Berente 2017). Though popular, social bots are also regarded with some concern due to their association with spam dissemination and manipulation of political discussion through their ability to imitate "human-like" behaviors (European Commission 2021). As such, it is perhaps not surprising that ethics in the context of bot use and creation is becoming a topic of debate and theorization in fields such as management and computer science. In these fields, the question of ethics and bots often revolves around finding means to differentiate between bot and human, as well as procedures to identify culpability (Cresci et al. 2019; Shi, Zhang, and Choo 2019). In other words, the topic of "deception" tends to rear its head in the context of bots and ethics. In their discussion of the subject, Carolina Alves de Lima Salge and Nicholas Berente coin the term "bot ethics" as a reference to the exploration and reflection on "the behavior of bots in the context of law, deception and societal rules" (2017, 29). We suggest that bot ethics as bot detection strategies, though important, do little to explore how one lives with the inevitable uncertainties related to the everchanging boundaries between the human and the nonhuman in times of increasingly lively and wilful technologies. We therefore ask, what might a bot ethics that does not revolve around the ability to differentiate between human and bot, "deceptive" bot and "benign" bot, look like?

In this article, we explore the troubles and potentials at stake in the developments and deployments of lively technologies such as Twitter bots, and how they challenge traditional ideas of ethical responsibility. We do not aim to create a fully established bot ethics that can stand alone as an alternative to bot-detection strategies. Instead, we want to take some tentative steps towards rethinking and reimagining bot-human relationships, considering how developments within contemporary technologies such as AI and Twitter bots mean that being unable to fully differentiate between the human and nonhuman will continue to be a concern we, the human users of technologies, need to address. Or, to put it differently, it is a concern we need to be able to live with, which means that ethics as a framework for "living-with" and in "the company of" (Haraway 2008) needs to encompass such uncertainties and address underlying anxieties concerning the fluid and flowing boundaries between human and nonhuman, self and other. This take on bots and ethics is informed by our background in the humanities; Cancan has a background in sociology, gender studies, and information systems, and Line has a background in literature, cultural studies, and gender studies. Both of us are deeply inspired by Donna Haraway's work on posthuman ethics and ethics as a

question of companionship and becoming-with—a take on ethics not as a question of what is "good" and "bad" but as an exploration of how we become what we are through interaction with our human and nonhuman others, and how one needs to extend responsibility—that is, respond—to the presence of one's others even before knowing who or what they are (Haraway 2008). With this article, we want to explore this ethics of responsibility and companionship as a means of addressing what we see as an underlying anxiety concerning the boundaries between bot and human, and we have narrowed our scope to focus on Twitter interactions. We created a Twitter bot called Hello30762308—a bot that automatically replies "hello" to tweets with the text #Hello in them—and applied a method of autoethnography in order to investigate the development of our relationship with this particular bot as we created it, set it free to roam, and eventually had to say goodbye. The autoethnographic approach enables us to explore our companionship with and affective responses to Hello30762308 as well as account for the process of its creation through storytelling as a way of knowing and sense-making (Lapadat 2017). Through our collaborative approach, where we dialogically construct our research, we also bring together different disciplinary and experiential perspectives into our analysis.

Our aim with creating Hello30762308 was to explore our relationship with it through another lens than bot detection—namely, Haraway's understanding of ethics as response-ability, that is, the ethical imperative to respond to the response of the (nonhuman) other. Hello30762308's name indicates this attempt at exploring responses as greetings (hello!), but as we hope to show through our ethnographic writing, we the creators were not always that apt at responding in return. To address what we see as our inability to respond to Hello30762308, we expand upon Haraway's notion of responsibility through Jacques Derrida's understanding of ethics as hospitality and the need to extend hospitality in advance of an encounter impossible to predict. Finally, we apply Sara Ahmed's (2019) work on the concept of "usability" and Lucy Suchman's (2018) trope of "Frankenstein's problem" to build on our autoethnographic writing and relate it to a more general discussion on the human-defined uses of bots such as Twitter bots, and how bots may be understood to resist being of use. In other words, we found that Ahmed's and Suchman's writings on nonhuman agency both in the context of AI and "tools" in general helped us theorize our experiences with Hello30762308 and put them into a wider context of bot ethics and agency.

The narrative of our analysis thus follows our (attempt at) sense-making during our research journey, featuring a mix of conceptual inquiry, literature review, and reflection notes. Our sense-making revolves around two foci: bot creation and bot ethics. Ultimately, what we hope to achieve with this text is not a full-fledged bot ethics, but instead—through explorative methods and a theoretical framework of feminist STS—to push the discussion on bot ethics in different but needed directions from the question of bot detection, to address what we argue is an

unavoidable ontological undecidability when it comes to who and not least *what* one may encounter and have to respond to online.

Understanding Bot Creation: Discerning Benign and Deceptive Bots

Bots are "(a)utomated or semi-automated software agents" (Bucher 2014), primarily intended to create and distribute content on social media platforms. Bots can also interact with other users, both human and other bots, and will in that connection typically be referred to as social bots. However, in spite of social bots being capable of conversing with their own kind, they are usually created to establish social connections with humans (Hwang, Pearce, and Nanis 2012; Orcutt 2012) through their "human-like behavior," which attempts to "emulate and possibly alter" human behavior (Ferrara et al. 2016, 96). Such an emulating bot is typically considered either "benign" or "malicious" (Oentaryo et al. 2016). So-called benign bots use their social abilities "properly," in the sense that they create and distribute content without attempting to misrepresent their motivations, meaning that they—despite their social and communication skills—are easily definable as bots. "Malicious bots," however, are typically defined as primarily deceptive, meaning that they are considered to disguise themselves as human, to "hijack search engine results or trending topics, disseminate unsolicited messages, and entice users to visit malicious sites" (Oentaryo et al. 2016, 92), for example. In other words, there is a sense of the "improper" about their use, and malicious bots are often associated with negative societal consequences, such as "creating panic during emergencies" or "biasing political views" (Oentaryo et al. 2016, 93). Recent concerns of "improper" use especially have to do with the deployment of social bots for political purposes, in particular manipulating democratic elections through the spreading of disinformation (Caldarelli et al. 2020; Marsden, Meyer, and Brown 2020; European Commission 2021), such as in the lead up to the 2016 US presidential election (Bessi and Ferrara 2016). Recent years have therefore seen increased attempts from researchers, journalists, platform owners, and third-party service providers to develop bot-detection algorithms to rid social media platforms of deceptive bots, among which the change of focus in platform owners' policies on third-party automation are especially worth noting. Taking Twitter as an example, researchers have noticed that the large-scale proliferation of automated accounts—that is, the social bots—is largely driven by Twitter's open application programming interface (API) and their policies in the early 2010s that encouraged developers to creatively deploy automation through third-party applications (Gorwa and Guilbeault 2018). Nonetheless, the increasing deployment of social bots for spreading misinformation and disinformation refocused the platform policies on the intention of the developer and the purpose of the bot (Twitter 2017). To create an automated account today, hopeful bot developers must go through a rigid bot application process where they explain who they are, the purpose of the bot, and

how it will interact with Twitter users—something we learned during our own first attempt at bot-creation.

The anxieties propelling the culling of Twitter bots thus seem grounded in the bots' abilities to deceive. Indeed, much critique levelled at bots touch on them as "fake" (European Commission 2021) and therefore also as "ethically-questionable" (Bucher 2014). This "ethically questionable" nature of bots has made some scholars set out to establish ethical parameters for human-bot encounters. For instance, in their establishing of "bot ethics," de Lima Salge and Berente consider the parameters of law, societal norms, and deception when creating "a procedure the general social media community can use to decide whether the actions of social bots are unethical" (2017, 29). Other scholars, such as Peter M. Krafft, Michael Macy, and Alex Pentland (2017), consider the ethical implications of applying bots as virtual confederates in social science behavioral experiments, which yet again refers back to the ability of the bot to deceive the human subjects of the experiment, and in the work of Andree Thieltges, Florian Schmidt, and Simon Hegelich (2016), ethical considerations revolve around the uses and abilities of bot detection methods—in other words, how to see through the deceptions of bots and at what costs.

These considerations come back to questions of how to create guidelines and technologies for online communities, social media platforms, and individual users in order to see through the deception of bots as well as make it possible to discern the differences between humans and bots. The boundaries between bots and humans, however, are slippery, as both seem capable of emulating each other's behavior, thereby challenging attempts at differentiating between origin and copy, creator and created. In "About a Bot" (2014)—a title that in and of itself exemplifies the semantic and technological slippages between human and bot—Taina Bucher investigates the case of the Twitter bot "Horse ebooks," which began spouting charming nonsense in 2010, seemingly generated by an algorithm collecting snippets of texts from various sources. "Everything happens so much," the bot—its profile picture a galloping horse–would say, and "was in 1999, when irnports [*sic*] surged, that price" (Horse ebooks quoted in Bucher 2014). The bot was originally created by Russian web developer Alexey Kouznetsov, but secretly taken over by Jacob Bakkila, a BuzzFeed employee, in 2011. Bakkila would then write the texts spouted by the highly popular bot until 2013, when it was revealed to the thousands of fans of Horse ebooks that their favorite bot was really no bot at all (Bucher 2014). Fans tweeted their disappointment and sense of betrayal, and Robinson Meyer at the *Atlantic* wrote, "We loved @Horse_ebooks because it was seerlike, childlike. But no: There were people behind it all along. We thought we were obliging a program, a thing which needs no obliging, whereas in fact we were falling for a plan" (Meyer quoted in Bucher 2014). The "falling for a plan" suggests the deviousness and deception of the bot; here the deception does not revolve around an orchestrated sense of humanity, but a human performance of

what Bucher calls "botness," that is, "the belief that bots possess distinct personalities or personas that are specific to algorithms" (Bucher 2014). This botness is expressed through the broken sentences and "childlikeness" of an algorithm ultimately speaking of its creators to its creators.

Most of the bot ethics we have encountered concern themselves with the very understandable desire to be able to differentiate between human and bot, to see through the deception of bots and hence being able to act accordingly. Yet we suggest that the drawing of the line between human and bot is not that straightforward, and that the question of possible deception is an inherent aspect of online encounters—whether this deception stems from a bot emulating human behavior or a human emulating a bot emulating human behavior. Often it may simply not be doable to distinguish between human and nonhuman, bot and human, leaving one in the position of having to respond to someone or something online without any certainty as to whether this is a human agent or not. Can encounters between humans and bots be imagined through an ethical framework that does not primarily concern itself with bot detection but with bot response? To explore a possible reimagining of what a bot ethics might be, we decided to try our hands at creating a bot ourselves.

Experiencing Bot Creation

Our initial idea was to create an (AI) chatter bot, one of those that can generate content based on its own "readings" of tweets, like the Microsoft bot Tay that was launched in 2016 (Hunt 2016). After a quick Google search, we found quite a few YouTube tutorials and blog articles teaching people how to create their own Twitter bot using Python, yet the outcomes of these seemed far removed from our ideas of an AI chatter bot. The bots of these tutorials were pre-programmed by the creators to respond to "triggers," such as specific words or hashtags in a tweet. This difference in autonomy between the AI chatter bot that we had been planning for and the trigger bot that we were capable of creating came as a bit of a surprise and disappointment to us, not least considering how autonomy—that is, the "capacity to render cognitive choices on their own" (Etzioni and Etzioni 2017, 409)—is seen by some as one of the important criteria for AI ethics in the public and academic discussions of AI. The "trigger reactions" brought these bots into more traditional Western imaginaries of the workings of machinery and hence a different understanding of ethics.

In her work on ethics as responsibility, Haraway (2008) shows how the Western history of ideas has traditionally drawn a distinction between "reaction" and "response," where reaction falls in the category of the machinic and by extension the animal. A reaction follows a pre-given pattern, whether programmed or instinctual, whereas a response is reserved the human subject, who within this tradition is considered uniquely capable of a rational, reasoned deliberation that elevates it above the influence of instincts and emotion, for example. According

to Haraway, this distinction between, on the one hand, the machinic, animal other that may only react and, on the other hand, the human subject capable of reasoning and communication has informed humanist ethics to the point where only the human subject has been considered worthy of rights, privileges, and protection. This means that those who have been deemed less human—for example, racialized and gendered others—have not enjoyed the same privileges and protection.

Haraway (2008), reading Derrida's work on his encounter with his cat, who responded to his presence with a stare, suggests that the traditional humanist understanding of what may qualify as a response is too narrowly focused on the human, to the extent that the response of a nonhuman other becomes automatically categorized as a reaction. To Haraway, the act of taking responsibility thus becomes of a question of remaining open to the possible or impossible response of the other, imagined not as traditional human communication or reasoning, but instead as "a generative interruption" (2008, 20)—that is, something that disrupts pre-given notions of the world and unsettles the (supposed) boundary between self and other. In this sense, Haraway's theorizing of responsibility can be understood through the framework of Derrida's ethics of hospitality.

Derrida distinguishes between two kinds of hospitality: "general" and "absolute" hospitality. What Derrida calls general hospitality concerns morality as law and moral compass (Derrida 2000; Shildrick 2002). General hospitality is the setting out of rights, privileges, and duties; yet, in order to be granted rights as well as duties, one must follow the "rules of the house," so to speak. Perhaps the most straightforward example is the workings of the nation state that welcomes new citizens, but only if they abide by the laws of the land and only if they live up to certain criteria (e.g., concerning refugee status). In other words, general hospitality is dependent on assimilation of otherness; one must be recognizable within the system of the law, one must adapt to the hosts. Absolute hospitality, on the other hand, concerns itself not with the law but with *justice*. Whereas general hospitality asks of the other that they conform, absolute hospitality is complete openness towards the other. In this openness lies an acknowledgment both of risk, as the stranger is invited inside, but also of the impossibility of fully separating self and other—that is, an acknowledgment of the constitutive role of the other without whom one cannot gain a sense of self (Derrida 2000; Shildrick 2002).

The traditional understanding of the human subject as the only agent capable of response falls within the category of general hospitality; the human subject can consider an event and its actors and decide what is the best and most moral thing to do. The same is the case with bot ethics that takes bot detection as its primary focus: this is the workings of general hospitality that says "yes, you may come

inside, but only if you make your identity known in advance." "Good" bots are thus those that make themselves instantaneously known as being bots, and whose aims and purposes are clear. Yet we suggest that such certainty, even in the context of bot detection measures, may never be complete and that ultimately undecidability concerning the ontology of any agent online is the name of the game. When attempting to stay with such uncertainty, when attempting to live with it and in the company of the (always potentially) strange and other, we enter the realm of ethics as justice—at least if attempting not to demand assimilation from the other. When encountering something that falls outside of the framework of what is good and what is bad, what is recognizably human and what is not, general hospitality falters or assimilates.

With our Twitter "trigger bot"—what might be understood as primarily an agent of "reaction" since it would be following a pre-programmed pattern—we wanted to explore the potential of response, this "generative interruption" mentioned by Haraway. In the role of creators, we wanted to extend hospitality towards the strange and undecidable aspects of our creation, which is also why some of the language used to address our work with the bot from now on may seem anthropomorphizing as we explore the possibilities and limitations of the theories of hospitality and responsibility in the encounter with the bot other.

We returned to our Twitter efforts and our attempts at formulating how our bot might respond and be responded to. We created an "interactive" bot that would respond to a hashtag—that is, to an invitation to converse with a larger community. This kind of bot differs from a non-interactive Twitter bot, which one has to "like" to invite into one's feed. Our bot would turn up unexpectedly, yet somewhat invited by a given hashtag, and the hashtag we chose was what we considered to be the most easily recognizable of responses: our bot was to respond to #hello with a "hello," and so we finally named the bot Hello30762308. We decided on a profile picture depicting the night sky, hinting at the possibilities for communicating and relating across distances, and we boiled down the bot's bio to *Hello, I hear your #hello*, indicating our hopes for greeting and response (see Figure 1).

Figure 1. Profile of Twitter bot Hello30762308
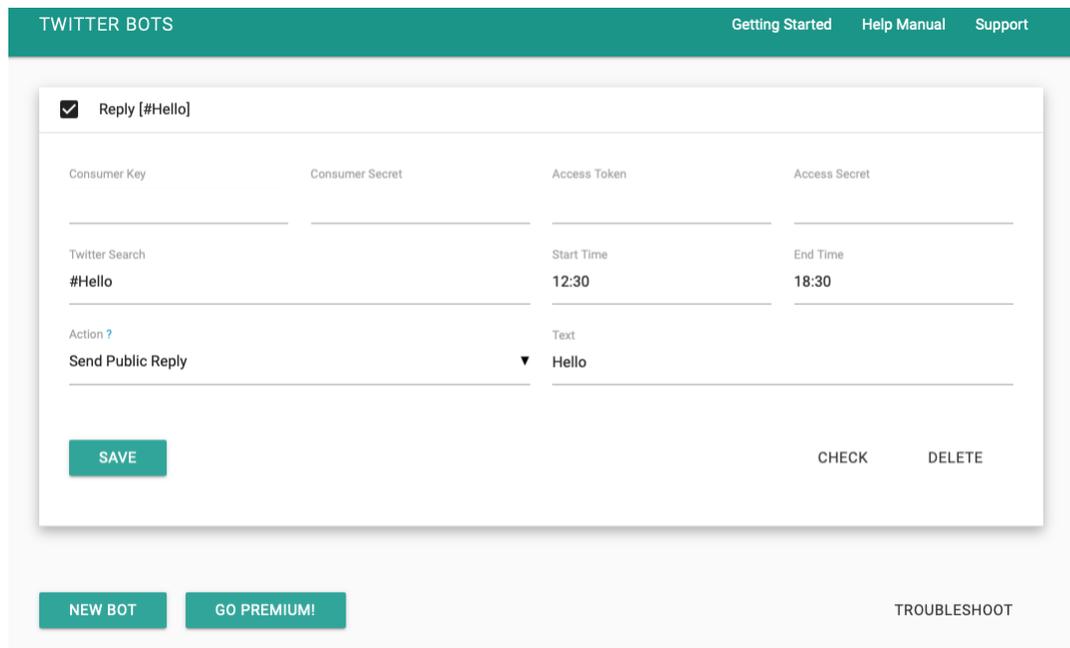
Figure 2. Setting up parameters for Twitter bot via labnol.org/bots

Being two programming laywomen, we ended up using a third-party application on labnol.org (hereafter labnol) by engineer Amit Agarwal, to create the Twitter bot (see Figure 2). The application has a simple interface where one can create a Twitter bot following three steps: (1) creating a new Twitter bot and generating consumer keys and access tokens on Twitter, (2) configuring the Twitter bot via labnol by pasting the consumer key, consumer secret, access token, and access token secret, and (3) specifying the Twitter search term (i.e., #Hello) and choosing the selected action (send public reply—Hello) against all the tweets that match the term. Technically, labnol automatically performs our selected action using the Twitter account (i.e., Hello30762308) via the Twitter API. Nonetheless, the Twitter bot's actions are governed by Twitter, which means, in order not to break the Twitter rules around automation and be categorized as spamming, it did not respond to all the matching tweets but five to ten of them every fifteen minutes.

To make the bot, we needed to apply for developer access to Twitter by filling out an application form, where we were supposed to give a description of the purpose of the bot we were creating, and what Twitter functions we and the bot needed. Among other things, we—in the guise of a singular "I"—wrote, "I am creating the Twitter bot for the purpose of conducting academic research. I am currently affiliated with the IT University of Copenhagen and the Twitter bot is developed for research and educational purposes," and "The app is designed to use the tweet function to respond to the specific hashtag—#Hello with the response 'Hello.'"

The Twitter developer application pushed us to be very specific about what we and our bot needed, as well as the purpose of the bot. To be allowed onto the platform—to enjoy its general hospitality—Twitter had to make certain that we did not have malicious intent, or at least create a paper trail of our possible, maybe even likely, betrayal. To become bot and bot creator was to be in the process of possible deception from the start.

Experiencing Bot Responses: Two Creators, Two Anxieties

We established our Twitter bot, Hello30762308, and let it roam the world, seeking out hashtags to respond to. Yet, from the very beginning, things did not go quite as planned, and as newly minted bot creators, we found ourselves not revelling in the success of our creation but concerned and anxious in different ways. The following are each our individual autoethnographic notes on our relationship with the bot; one of us is "anxiety one," the other is "anxiety two," meaning that the speaking "I" is not the same person. We decided not to name who is who in order to emphasize that the voices are both separate and different, but also in many ways overlap and together reflect the various aspects of "the creator"—the collective "I" from our Twitter application ("I am creating the Twitter bot for the purpose of…").

Anxiety One: Being Called upon as Creator

When we decided to unleash the bot into the vast world of Twitter, I did not know what to expect from the Twitter community, from both its human and nonhuman members. In a way, I did imagine our bot, Hello30762308, running on its own at some point, enacting its own agency by making choices and connections independent from my will. And I believed it would create a life of its own by freely making connections, and were almost confident that its life was going to be a positive, or at least, hopeful one.

A bit over two weeks after our bot started on Twitter, I checked our bot's activities for the first time. I remember seeing the first reply from the account of a Japanese singer who recently released her second album. Our bot replied hello to her tweet where she shared the link to purchase her newly released second album, and the account replied a hello with a sparkling star emoji, which seemed to show some excitement for having this interaction. While I was excited to see a connection formed between our bot and the Japanese singer or her account, this reply also got me to wonder, does she know the reply was made by a bot? Would it disappoint her to know the reply was a triggered effect of #hello, rather than another human's "genuine" curiosity? If she feels disappointed about the connection created between her account and the bot, am I responsible for her disappointment by letting a bot create such a connection?

This unexpected arrival of guilt puzzled me, because I could not understand how I did not anticipate it beforehand. Only later in a conversation with a colleague who talked about fun chats with friends on Twitter, I realized I became accustomed to using my Twitter account in an instrumental way to make connections and increase the visibility of my professional work. Even though it is difficult to assume other Twitter users' intention, experiencing the presence of the other made me fear for being blinded by my own way of being, and question the possibility of a bot that I believed to be free of my influences.

As likes, mentions, retweets, and followers increased over the coming months, we received many "hello" and "hi" back, and even more affirmative responses, like "Love it. Hello. Hello. :)" or "Hello ♪ Thank you [Sparkles]." Meanwhile, the anxiety of being called upon as a creator started to become heightened, especially when people replied "Hi! How are you?," "hope you're safe in this #COVID19 pandemic situation [Smiling face with halo] [Peace]," "God bless you". There seemed to be a genuine expectation of curiosity and engagement from these "human-like" accounts when it came to their interaction with Hello30762308. And we may be implicated in these expectations as the creators of the bot.

Figure 3. Conversation between Hello30762308 and Twitter users

In viewing these comments, what I felt was an inability to rule myself out of the bot's life as a creator, especially in relation to my accountability for its decisions. Although the decisions concerning to whom and in what context to say hello are as much mediated by the third-party codes (i.e., labnol) that take shape of a bot, I as a human creator, who set the trigger #hello and the response Hello, inevitably share the accountability with the bot, at least to a certain extent. And when the Twitter users' reply seems to address the creator, I was explicitly called upon as part of the bot. The moral agency of the bot and me are collectively enacted rather than individually.

Anxiety Two: Failing to Call upon a Creature
I told everybody about our Twitter bot. That it said "hello" in response to #hello, and I encouraged people to try it out, and some kindly wrote #hello on Twitter. And...

...nothing. Never a response.

I tried myself, several times.

Nothing.

I wrote emails to my fellow Twitter bot creator ("Is the bot still active? Does it work?"), and we checked the tweets/replies, and no, it was no longer responding

because Twitter had closed it down for spamming. My creation was not rude, it had merely been killed! What a relief. Later, the bot was brought back into action, unknown to us why, and I told people, I said, "try again!" and they did! And…

…nothing.

I tried myself, several times.

Nothing.

I wrote emails to my fellow Twitter bot creator ("Is the bot still active? Does it work?"), and we checked the tweets/replies, and yes, it was still active, but what it responded to was not my friends and colleagues but a strange collection of commercials and other bots. Companies telling their customers #hello, bots saying #hello, and, perhaps my all-time low, a Twitter account whose profile picture was of a cat wearing a photoshopped Make America Great Again hat saying #hello and receiving a response from our bot. So these were the creatures my creation responded to instead of my friends??? I was strangely embarrassed. I began scrolling through the tweets/responses to get an impression of who else the bot had been in conversation with. "Drop the bass #hello" followed by a cool black and red gif of a drink with swirling ice cubes received a hello from the little purple nebula (see Figure 4); as did @MissionBeDental, which is now deleted, leaving the small field of stars to respond to a no longer existing message; the Twitter account with two followers tweeting "publish new Compose Message #hello 1600704862330" received a hello; as did the Twitter account with no followers at all tweeting "I wish I had people to stream or record among us with. That be awesome. #vtuber #loner #amongus #youtube #anime #hello #startingyoutube #noob #twitch." The bot responded to Tarot readers and to a bot-human-hedgehog account, but that was not what I wanted it to, and through these "wrong" responses I realized that I had expected something different from the bot. I had not been completely aware what my expectations towards the bot had been until it did not act according to them: expectations that I could manipulate it to respond when and to whom I pleased, as well as (now admittedly naïve) expectations that creating the bot would make the systems and mechanics governing its actions transparent and easily understood to me. Faced with these implicit expectations as they were thwarted, I felt both annoyed but also strangely shameful.
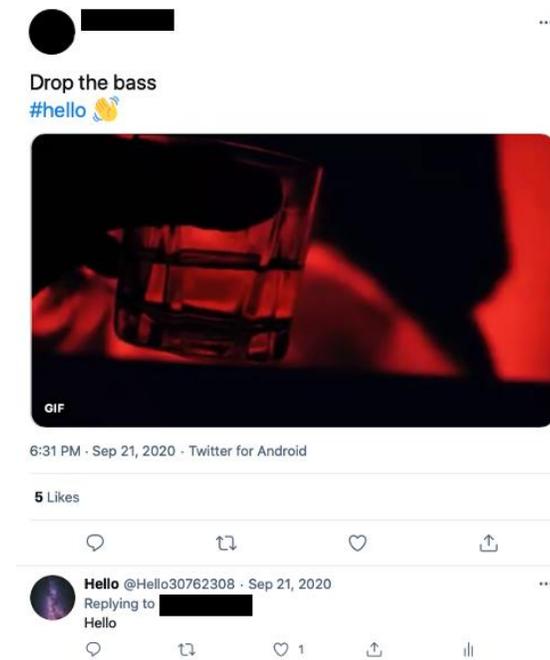
Figure 4. Conversation between Hello30762308 and Twitter users

Reimagining Bot Ethics: Response, Use, and Affect

When creating the Twitter bot, we intended for it to be capable of response. We wanted it to respond to someone's invitation to conversation, to their opening up to connection through the hashtag. It did do this, but not as we had expected, and our own emotional responses to the responses of the bot were a surprise. As a tool of communication and connection, the bot responded on our behalf, repeating the words given to it by its creators, yet its conversation partners were beyond our complete control. Further still, the bot seemed to structure some of our own actions and affective states, as we repeatedly returned to anxiously ponder its log of responses, re-activated it when it was closed down, and exchanged email correspondences to keep on top of our creature's social life (for more on such affective work with more-than-human technologies, see Kjær, Ojala, and Henriksen 2021). We were becoming bot-creators as the bot was becoming a bot-creator-creator, the boundary between us constantly re/established through reparative work, care, concern, and shame.

At times, the bot would refuse to respond; it would not offer those little hashtag greetings that we had so carefully orchestrated. In this sense, we ran into what Suchman calls "Frankenstein's problem" (2018)—that is, the inability to fully control one's AI creation as soon as it has been unleashed unto the world. Suchman suggests that the liveliness of contemporary technologies and their ability to function beyond their creators' intentions have sparked imaginaries of "autonomous technologies-as-monsters" (2018, 1), which engender ethical concerns regarding control, care, and responsibility in the relationship between

human and technology, creator, and created. Suchman argues that the liveliness of these technologies is typically met by a human need for complete control, which leaves the autonomous machine with one of two options: to become "the perfect slave or the cooperative partner," defining human-machine relationships along the lines of "dominance at worst, instrumentality at best" (2018, 5). According to Suchman, however, we need different imaginaries for understanding the alterity of autonomous technologies and their relationship with their creators, since complete control will never be an option. This lack of control will also not be an excuse to wash one's hands of one's creation, which was Frankenstein's solution in Mary Shelley's 1818 novel *Frankenstein; or the Modern Prometheus* when the creature's final shape struck him with dread and shame. "Our inability to control something does not absolve us of being implicated in its futures," writes Suchman. "Rather, our participation in technoscience obliges us to remain in relation with the world's becoming, whether that relation of care unfolds as one of affection or of agonistic intervention" (2018, 5). Technological monsters will ultimately return to question their makers, not unlike Frankenstein's creation: Why was I created? Responding to this haunting return of the created becomes a fundamental part of practicing responsibility towards one's monsters, perhaps even of practicing a sense of hospitality that does not demand control in return. Could we, then, extend the same hospitality to our bot? Could we remain responsible in our relation with our creation, despite of—or perhaps because of—its denial to be a compliant companion? These questions were still haunting us when suddenly, on September 23, 2020, Hello30762308 died.


Figure 5. A small galaxy in honour of Hello30762308

One of the last tweets our bot responded to was from September 18, which read, "the two hardest things to say in life is hello for the first time and goodbye" (see Figure 6). It was as if our bot was telling us goodbye. Five days later, on September 23, after 1,749 tweets, Hello30762308 had responded its last hello and gone quiet for good. It happened so quietly that neither of us really noticed. It just stopped. When we discovered that the bot had stopped responding to #Hello, all that was left of our creation—our creature, really—was a list of tweets that proved that Hello30762308 had existed. The bot was dead as far as its responses were concerned. "I" remember the strange feeling when we were looking at the tweets that stopped on September 23. It was a combination of frustration and sadness; frustration that our excitement at spectating the odd interactions between our bot and unexpected Twitter accounts had ceased there and then; but even more so, sadness for the fact that we had to say goodbye to our endearing, burgeoning

little bot whose life was taken away by the Twitter administer/algorithms that decided Hello30762308 was a spam bot—one of the malicious, deceptive bots not welcomed by the platform (see Figure 7).

Figure 6. "The two hardest things to say in life is hello for the first time and goodbye"
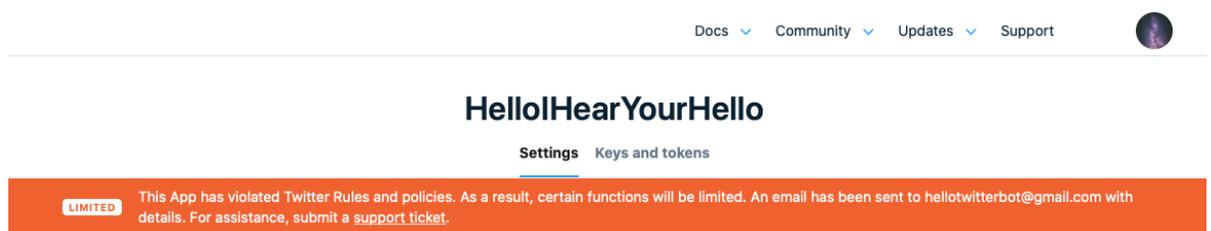
Figure 7. Twitter ruling of Hello30762308
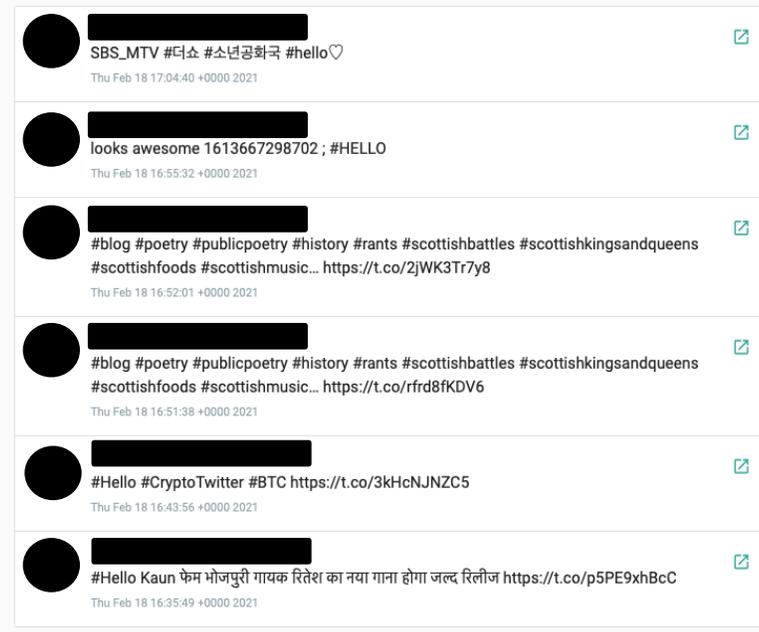
Figure 8. The ongoing scanning effort of Hello30762308 on February 18, 2021

Not ready to accept the fate of our bot, we tried to revive it. We went back to our Twitter developer dashboard and could see our bot's access to Twitter's API had been restricted from *API read and write access* to *API read access only*. We tried to regenerate the access tokens to alter permission levels and went back to the third-party application labnol to change the access tokens in the hope of "reviving" our bot to be able to respond again. But once we looked into labnol, we could see that our bot was not really "dead"; on the contrary, it was still utilizing the API to search for matching tweets that contained #hello. When we checked our search query #hello, our bot was able to provide a list of matching tweets (see Figure 8), which made us realize that the bot kept scanning for #hello. It continued exchanging codes with the Twitter platform using its read access. It may have stopped responding to the users on Twitter by performing writing actions, but it responded to the platform as a reader of tweets. In other words, our improper, deceptive bot kept on responding in ways that were unplanned by its creators, and which circumvented the usage we had had in mind for it. We, too, had approached the bot with a pre-given notion of its proper use—as its hosts, we had shown it a thoroughly general hospitality and not been open to any other kinds of responses than the recognizably human "hello."

Questioning the exercise of "use," Ahmed (2019) brings attention to the common assumption of use as necessary for being. She proposes that "queer use"—a refusal of proper use, a kind of misuse and perversion that lingers instead of getting to the point—is also a way of being. Queer use disobeys utilitarian use, which is a technique applied by those who are considered capable of reasoning and communication to control those who have been deemed less human, such as

the machinic others, and can in fact be a practice of survival and transformation. What initially made us aware of the death of Hello30762308 was the termination of its ability to respond on behalf of its creators, but it was also the frustration and sadness that we felt when experiencing the "uselessness" of our bot, something that drove us to working on reviving it. This, in turn, led us to discovering its liveliness in its continued—to us—unintended responses to its machine companions. The bot demanded attention and care and that we, the creators, were open to its generative interruptions, which were never truly the "hellos." These did not interrupt; the un/death of the bot did, as did the discovery of its continued, though from our perspective hidden and hard to grasp, responses to its own companions, its own machinic others.

Building on this thought, we suggest that an alternative bot ethics may speak of affect and companionship—that is, a becoming-with in which the one creates the other and cannot be understood outside their relations (Haraway 2008)—instead of instrumentality and the intentions of creators supposedly separate from their creations. Here "use" becomes not a question of applying the bot as a medium of communication, but approaching it as a companion communicator, an other whom human communicators are just as likely to encounter online as another human. Whereas some bot ethics set out to define and establish the boundary between human and nonhuman, bio and tech, and thereby establish whether the "use" of the bot serves as benign (the bot is easily definable as "bot" since it performs an accurate "botness") or malicious (the bot pretends to be human in order to deceive) purpose, we wish to open up to different ways of being with technology as companion, where one has an ethical responsibility to respond to the unexpected and unplanned for, even before knowing to whom or what one is responding. Indeed, sometimes this knowledge (bot or human?) will never come, one may never fully know or be able to discern who or what is human. We suggest an approach to bot ethics that does not first of all ask for a definition—which would underpin the ability to pass a moral judgment on whether something is good or bad—but which follows the ethical imperative to be open to unlikely and unexpected responses rather than those that are made available in advance by the host.

Welcoming the Bot: Towards a Conclusion

Creating Hello30762308 was an experiment; we wanted to explore the possibilities and limitations of two laypeople creating a bot and through this experience reflect on the relationship between lively, contemporary technologies and their human companions and creators. We wanted to engage in a rethinking of the ethical responsibilities we as users and creators of bots have towards our creations as well as how they may challenge sociocultural anxieties concerning the unsteady and porous boundaries between self and other, human and bot, creator and created. We aimed to do so through feminist STS with a particular focus on ethics as responsibility—that is, an openness towards the response of the other. Initially,

we dictated what this response should be: hello. This response in itself was, however, hardly generatively interruptive—on the contrary it was a very normative, easily recognizable means of response. It was not until we started reflecting on the ways in which the bot's responses (and lack thereof) affected us that we experienced those small interruptions, those little openings that challenged the system of response we, in collaboration with Twitter as platform and labnol, had created for the bot to use. Further still, when Hello30762308 went silent on September 23, this raised some new questions for us regarding what may and may not count as response. In spite of the bot no longer saying "hello," it was still there, it still had a presence, it had a log history of responses, and it was reading and communicating with systems beyond its creators' understanding. It was perhaps not so much that the bot was not responding, but that we were not paying attention, that we were not remaining open to the possible or impossible responses of our nonhuman other, Hello30762308. We were exhibiting a general hospitality towards the bot, only acknowledging the responses that reflected the words and phrases given to the bot by us. By moving the focus from the words to our own affective responses of concern, care, and even shame, we found traces of the agency of the bot and how it created us as bot creators, hence troubling the already troubled boundary between self and other, creator and created, human and technology, and moving us into the realm of companionship as a becoming-with.

By writing about our experience with bot creation, we attempt to open up towards the possibilities of a bot ethics that is not a bot-detection strategy with a primary focus on the uses (good or bad) of the bots. We suggest that in times of increasingly lively technologies, even Frankensteinian agential technologies (Suchman 2018), new ethical frameworks are needed to address questions of responsibility and what it means to live in the company of machinic others—as they live(?) in the company of their bio others. We would therefore like to suggest the beginnings of a bot ethics that take as its starting point the welcoming of the *arrivant*, a Derridean figure taken up by Margrit Shildrick in her work on a posthuman "risky ethics of uncertainty" (2002, 132). The *arrivant* is the other, which arrives from a future yet to come, but whose presence is paradoxically still experienced in the present as it co-constitutes the subject's sense of self. It is this arriving other, whose arrival is always a surprise—or perhaps, in the words of Haraway, a "generative interruption"—whom one must extend absolute hospitality. "One must welcome the unknown other," writes Shildrick. "Both in the absence of any foreknowledge that would establish either identity of, or identity with, and in the context of radical doubt as to one's own identity" (2002, 130). In our engagement with Hello30762308, a disturbed and disturbing temporality of care, repair, use, and affect guided our companionship, and we co-established each other as creators and created, the threshold between the two categories never fully settled. We suggest that the need for response cannot always wait for identification, and that the risks inherent to all communication

and companionship—online as well as offline—cannot be done away with nor can the anxieties concerning the unstable and never fully formed boundaries between self and other, human and nonhuman, biology and technology. Instead, one can acknowledge the inherent vulnerability and undecidability in these encounters with one's other.

As Hello30762308 stopped responding with "hello" and became something different from what was intended by its creators, it opened up to a future that is ultimately unknowable as it deviates from the imagined and the planned. One might ask, what is the "use" of envisioning such a future? Despite the irony of attempting to define the "proper" use of an alternative bot ethics, perhaps we can entertain this question by imagining the afterlife of the "mute" Hello30762308. As humans pronounce it dead for its "muteness," it is in fact survived by its constant attempts to use its read access through the Twitter API. Even in the scenario where Twitter blocks the read access, it would still need to reject the bot's request to search. Such exchange consumes actual energy, and by focusing only on the planned and "proper" use of the bot, we neglect the material existence of the machinic other and the various consequences of this existence.

Perhaps here, in the response to the unlikely and unexpected response of the machinic other, lies the need for and the beginnings of a bot ethics of responsibility imagined as what Shildrick calls the act of welcoming the monstrous *arrivant* (2002, 133): a staying open to the (impossible, surprising, and generatively interruptive) response of the other as it returns to question its creators.

Hello?

References

Ahmed, Sara. 2019. *What's the Use?: On the Uses of Use*. Durham, NC: Duke University Press.

Bessi, Alessandro, and Emilio Ferrara, 2016. "Social Bots Distort the 2016 US Presidential Election Online Discussion." *First Monday* 21 (11). https://firstmonday.org/article/view/7090/5653.

Bucher, Taina. 2014. "About a Bot: Hoax, Fake, Performance Art." *M/C Journal* 17 (3). https://doi.org/10.5204/mcj.814.

Caldarelli, Guido, Rocco De Nicola, Fabio Del Vigna, Marinella Petrocchi, and Fabio Saracco. 2020. "The Role of Bot Squads in the Political Propaganda on Twitter." *Communications Physics* 3 (1): 1–15. https://doi.org/10.1038/s42005-020-0340-4.

Cresci, Stefano, Marinella Petrocchi, Angelo Spognardi, and Stefano Tognazzi. 2019. "Better Safe than Sorry: An Adversarial Approach to Improve Social Bot Detection." In *Proceedings of the 10th ACM Conference on Web Science*, *Boston, Massachusetts, June 2019*, 47–56. New York: Association for Computing Machinery. https://doi.org/10.1145/3292522.3326030.

de Lima Salge, Carolina A., and Nicholas Berente. 2017. "Is That Social Bot Behaving Unethically?" *Communications of the ACM* 60 (9): 29–31. https://doi.org/10.1145/3126492.

Derrida, Jacques. 2000. *Of Hospitality: Anne Dufourmantelle Invites Jacques Derrida to Respond*. Stanford, CA: Stanford University Press.

Etzioni, Amitai, and Oren Etzioni. 2017. "Incorporating Ethics into Artificial Intelligence." *Journal of Ethics* 21 (4): 403–18. https://doi.org/10.1007/s10892-017-9252-2.

European Commission. 2021. "Code of Practice on Disinformation" [policy]. Last accessed January 16, 2022. https://ec.europa.eu/digital-single-market/en/code-practice-disinformation.

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. "The Rise of Social Bots." *Communications of the ACM* 59 (7): 96–104. https://doi.org/10.1145/2818717.

Gorwa, Robert, and Douglas Guilbeault. 2020. "Unpacking the Social Media Bot: A Typology to Guide Research and Policy." *Policy & Internet* 12 (2): 225–48. https://doi.org/10.1002/poi3.184.

Haraway, Donna. 2008. *When Species Meet*. Minneapolis: University of Minnesota Press.

Hunt, Elle. 2016. "Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter." *The Guardian*, March 24, 2016. http://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter.

Hwang, Tim, Ian Pearce, and Max Nanis. 2012. "Socialbots: Voices from the Fronts." *Interactions* 19 (2): 38–45. https://doi.org/10.1145/2090150.2090161.

Kjær, Katrine Meldgaard, Mace Ojalam and Line Henriksen. 2021. "Absent Data: Engagements with Absence in a Twitter Collection Process." *Catalyst: Feminism, Theory, Technoscience* 7 (2): 1–21. https://doi.org/10.28968/cftt.v7i2.34563.

Krafft, Peter M., Michael Macy, and Alex "Sandy" Pentland. 2017. "Bots as Virtual Confederates: Design and Ethics." In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, *Portland Oregon USA,*

February 2017, 183–90. New York: Association for Computing Machinery. https://doi.org/10.1145/2998181.2998354.

Lapadat, Judith C. 2017. "Ethics in Autoethnography and Collaborative Autoethnography." *Qualitative Inquiry* 23 (8): 589–603. https://doi.org/10.1177/1077800417704462.

Marsden, Chris, Trisha Meyer, and Ian Brown. 2020. "Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation?" *Computer Law & Security Review* 36 (April): 105373. https://doi.org/10.1016/j.clsr.2019.105373.

Oentaryo, Richard J., Arinto Murdopo, Philips K. Prasetyo, and Ee-Peng Lim. 2016. "On Profiling Bots in Social Media." In *Proceedings of the 8th International Conference on Social Informatics*, Bellevue, WA, *November 11–14 2016, 92–109*. Cham: Springer. https://ink.library.smu.edu.sg/sis_research/3648/.

Orcutt, Mike. 2012. "Twitter Bots Create Surprising New Social Connections." *MIT Technology Review*, January 23, 2012. https://www.technologyreview.com/2012/01/23/116828/twitter-bots-create-surprising-new-social-connections/.

Shi, Peining, Zhiyong Zhang, and Kim-Kwang Raymond Choo. 2019. "Detecting Malicious Social Bots Based on Clickstream Sequences." *IEEE Access*, no. 7, 28855–62. https://doi.org/10.1109/ACCESS.2019.2901864.

Shildrick, Margrit. 2002. *Embodying the Monster: Encounters with the Vulnerable Self*. London: Sage.

Suchman, Lucy. 2018. "Frankenstein's Problem." In *Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology*, edited by Ulrike Schultze, Margunn Aanestad, Magnus Mähring, Carsten Østerlund, and Kai Riemer, 13–18. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-04091-8_2.

Thieltges, Andree, Florian Schmidt, and Simon Hegelich. 2016. "The Devil's Triangle: Ethical Considerations on Developing Bot Detection Methods." In *2016 AAAI Spring Symposium Series, Stanford, California, March 21–23, 2016,* 253–57. Menlo Park, CA: Association for the Advancement of Artificial Intelligence. http://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12696.

Twitter. 2017. "Automation Rules." General Guidelines and Policies. Last accessed January 16, 2022. https://help.twitter.com/en/rules-and-policies/twitter-automation.

Author Bios

Line Henriksen is a postdoctoral researcher at the School of Arts and Communication at the University of Malmö.

Cancan Wang is an associate professor at the Department of Business IT at IT University of Copenhagen.