Ethical Tech Finalist Prompt

Keefer C. Rourke[*a], Alexander C. Parent[a], and Salem Abuammer[b]

[a]School of Computer Science, University of Guelph, 50 Stone Rd. E, Guelph, Ontario, N1G 2W1 Canada
[b]Illinois Institute of Technology, 10 West 35th Street Chicago, Illinois, 60616 USA

June 14, 2018

1 Forward

As one of the top 20 finalists for the Major League Hacking Ethical Tech Initiative, I have prepared responses to the provided writing prompts in the sections that follow. Please accept this document as a submission for the final task of this initiative.

I have a great passion for ethics and technology — making positive change through technology and remaining critical of innovation in a human context makes up a large part of my ethos. I have sincerely enjoyed preparing these responses.

Thank you,

Keefer Rourke

https://krourke.org

(on behalf of our team from *MHacks X*)

[*]Corresponding author.

2 Responses

> 1. Find one real world example of an issue that happened where the ethics of a technology were called into question.
>   (a) How did the tech sector react to this issue?
>   (b) How did the general public react to this issue?
>   (c) How was this response difference from the reaction from the tech sector, if at all, and what factors do you think lead to it being different?
>   (d) What is the significance of any disconnect between these points of view?

Through the course of the last few decades, technology has undergone a revolution like no other. In particular, the advancement of communications technologies has enabled connectivity on a scale that has never been seen before, enabling free distribution of information, ideas, independent content, and an explosion of technological innovation. Much of this technology has greatly benefited society, or at least aims to, but rather often it seems that innovation occurs for the sake of innovation, with little regard to societal impact or consequences. Many services are at fault for invading user privacy, directly misleading people, or improperly handling data security, however even with those issues aside, the ethics of certain practices, services, and products must be called into question.

To keep this discussion topical to recent events, consider the *May 2018 Google I/O* conference which introduced the world to a new AI-driven product called *Duplex*. A demonstration that Google provided showed *Duplex* automatically making a phone call to book an appointment for its user, where the receiver on the other end of the call was completely unaware that they were speaking to a robot. Notably, *Duplex* included natural pauses and interruptions in its speech[1], and was able to gracefully back out of calls when the context of the conversation was not properly understood. This software demonstrated unprecedented advancements in the domains of artificial intelligence, natural language processing, and speech synthesis, creating human-like, context-aware speech. The tech sector on a whole seems to be in awe of this capability, as the problem of speech synthesis has fascinated humankind as early as the late 1700s, beginning with acoustic sound resonators, evolving to electrical synthesizers in the early 1900s, eventually leading to the development of the first text-to-speech (TTS) program in 1968 [6]. Speech synthesis has improved greatly since the 60s, but it has not been until now that synthesized and natural speech have been indiscernible.

The potential impact of *Duplex* is profound. Seamlessly parallelizing and offloading administrative tasks to artificial intelligence sounds like Asimov's[2] science-fiction pipe dream. Human productivity could increase through naturally communicating with robot assistance. The revelation of this technology sparked immediate debate, with mixed responses to its capabilities, particularly focusing on the deceptive nature of the software. Proponents of the software insist the likeness to human voice is necessary for the software's goals, and to reduce friction while it interacts with clients. Scott Huffman, an executive on the Google Assistant team said, "People [would] probably hang up," if *Duplex* didn't sound human. The team at Google claims not to want to pretend to be human, but

[1] *Duplex*'s synthesized speech included sounds like "hmm," "um," and "ah."

[2] Isaac Asimov is the author of *I, Robot.* In 1942, he devised a set of fictitious rules for robots to follow in society, known as "The Three Laws of Robotics".

rather to ensure engagement for very specific tasks, such as booking appointments and restaurant reservations. Meeting resistance from the public with growing concerns about the capabilities of artificial intelligence, Google further assures that the AI is not up to the task of learning to do new things [1], but the pairing of realistic speech and AI automation is still worrying.

People are mostly uncomfortable with how natural the speech synthesis sounds, voicing the concern that when *Duplex* makes phone calls, Google has a responsibility to disclose that it is an AI that is speaking. Zeynep Tufekci, a professor at the University of North Carolina School of Information and Library Science, called this technology horrifying and ethically lost, emphasizing that, "As digital technologies become better at doing human things, the focus has to be on how to [...] delineate humans and machines," [7]. Hyper-realistic speech synthesis coupled with human-like AI raise questions of how easy it may be to impersonate someone, and what harm might follow as a result. Citing this concern, Shane Mac of *Assist*[3] suggested in an interview with NPR that new laws need to be created regarding disclosure and intent of machine intelligence [3]. It is clear that the general public is made uneasy by an artificial intelligence that can deceive a human, and doubly so when it is not disclosed as such.

This disconnect, a gap in trust, is significant. Google has the motivation to promote its product and increase its user-base, while maintaining its established position at the forefront of ground-breaking technology. The ethical dilemma arises when Google, an organization with immense power over many aspects of people's lives, insists that a technology that makes people uneasy (such as a realistic speech synthesis with natural pauses and interjections) is needed in order to enable basic functionality of their service: an artificially intelligent personal assistant. An AI assistant doesn't require realistic speech synthesis, it simply requires a method by which it can relay information to humans. It should be considered that if the public would object to such a service, even without deceptive features, then that should be an indication of a poor service. Undercutting trust to justify a use case for a technology, or for the sake of profit and innovation, without evaluating the risks that it may pose with concern to social wellness is harmful. Moreover, this is not the first time that technology has tested public trust. A similar attempt to introduce *Google Glass* into the mainstream failed due to similar concerns about privacy[4], and innovators need to take note of these mistakes so that efforts can be directed to creating useful services that don't breach basic societal contracts like trust. If society is opposed to a technology, the technology should not be made implicitly more acceptable by making it less noticeable.

> 2. Imagine the project you made is now a near ubiquitous service/technology utilized by a significant number of people. Identify 2–3 ways (separate from data privacy or security) in which this technology might propose a risk to people and things we care about — including but not limited to users, society, democracy, economic security, civil discourse, public safety our collective mental health, etc. For each potential risk, think of how it might affect the user, people unrelated to the technology, and society as a whole.

The project that we created, tentatively named *stegamsg* for "steganographic messaging," is a platform for secret messaging, unconventionally hidden within innocuous data-streams like PNG

[3] https://assi.st

[4] *Google Glass* had a discrete camera capable or recording video, which raised many concerns regarding unannounced video recording in private venues.

formatted images. The software allows for users to communicate securely[5], privately, and discreetly. In the case that data streams may be intercepted, the eavesdropper will only see images (usually memes) and not be privy to the actual information exchange. *Stegamsg* was created to enable secret communication between silenced individuals and groups. A world which might use it on a grand scale would probably see many independent service providers all running the same base software distribution backed by our project. As with any online communications platform (data privacy and security aside) there are concerns around the purposes for which it will be used, matters regarding freedom of speech and complying with legislation that directly pertains to different forms of speech, as well as cultural impact, sensitivity, and reaction.

Stegamsg is a platform which, for the purposes of anonymity, does not have user accounts. Users may identify themselves in a channel, but the software itself does not provide a means for authenticating any claim to an identity by users. This platform was originally described[6] under the use-case where minorities would have the need to secretly organize themselves under the watchful eye of groups aiming to persecute them, such as in the case of the violent persecutions of LGBTQ+ persons in Chechnya, Russia in the spring of 2017 [5]. If one is to consider an environment where secret messages are being relied upon to organize communications between marginalized groups, one must also consider the possibility that malicious actors may claim false identities in an attempt to disrupt conversation, obtain unknown information, and actively seek to do harm against the groups.

Due to the abstract nature of software, if one is to consider a more general-purpose usage of our software platform, anonymous communication remains a double-edged sword in our design. While a user's identity is protected, people may still be misled by misuse of the platform. This might accidentally allow for the spread of misinformation intended to harm specific individuals or groups. Harmful side effects of this may range from the dissemination of fake news[7], to actual impersonation, misrepresentation, and/or defamation of individuals. Due to the abstract nature of messaging, and the unlimited number of contexts under which humans communicate, the ramifications of this could be severe. Adversarial messaging could disrupt and hinder civil discourse, and unmonitored, potentially abusive messaging within groups could pose a risk to users' mental health.

Beyond issues of trust and authenticity, one must also consider format and transmission protocol for messaging software. *Stegamsg* was designed with a bias towards Western culture and infrastructure. It uses images as containers for delivering messages, a trait of the software that has several limitations and consequences. Considering most of the data in each transmission is effectively useless (we only care about particular bits in an image data-stream), *stegamsg* is extremely inefficient, and effective usage relies on high volume, high bandwidth connections. In areas of the world where bandwidth is restricted and connectivity is sparse, networks with insufficient throughput would very likely disable usage of our software. This, as a result, alienates users from more than half the world, as almost 4 billion people do not have internet access, and bandwidth is extremely limited in many developing countries [4]. If internet usage is already being monitored in low-bandwidth geographic areas, then comparatively high bandwidth consumption from our software may have unintended consequences. Monopolizing bandwidth may deny other individuals access to the internet, thus indirectly harming others. In some cases, high volumes of incoming and outgoing images may raise red flags and attract unwanted attention, which is the opposite of the intended effect.

[5] Though the cryptographic and steganographic software libraries that we used are likely not as strong as industry standard encryption schemes. Our team is not composed of experts in the domain of information security.

[6] Please refer to the abstract which was submitted in late 2017 as an application to the *Ethical Tech Initiative*.

[7] Fake news promoted on Facebook in Sri Lanka this year spread racist propaganda, and lead to an outbreak of racially-charged violence [2].

> 3. For the potential risks you listed above, what could you do to offset each of them? How might you prevent or mitigate those impacts? These can be purely technological solutions, but consider any initiatives or investments you could also make with your company's resources.

The biggest danger of *stegamsg* as implemented, is the lack of authenticated communications within the platform. *Stegamsg* is anonymous by design, and deliberately does not have user accounts which are traceable to the user's real-life identity. However, without provisions for verifying an individual's identity, users of the platform are made responsible for trusting the members of a channel. If people require that their messages are only sent to their intended recipients, then there should be capabilities within the platform to guarantee this.

A few technological solutions to this issue may be constructed as follows:

1. provision roles within encrypted group channels;
2. allow a creator of a channel to close the channel, preventing new members from joining (or requiring approval for new members to join);
3. expose a user ID or authentication code which can be externally verified against an identity, without directly registering that identity with the system.

These solutions: 1) allow for finer grained control over group conversations, enabling creators or administrators to monitor the channel for abuse; 2) restrict access to private communications to only authenticated individuals; and 3) provide some level of trust and safety where users may exchange authentication codes via some third-party. These enhancements to the *stegamsg* platform would mitigate (to an extent) efforts of trolls or malicious parties, without compromising on the anonymous design of the application.

Considerations of internet access and bandwidth availability are more complex issues, which are deeply rooted in the expensive nature of the traditional infrastructure required for internet connectivity. However, there are a number of high-quality solutions that are in development (e.g. RightMesh[8], Project Loon[9], etc.) that aim to reduce the prevalence of connectivity issues. Given sufficient resources, promotion and integration with these platforms, as well as other outreach initiatives to establish infrastucture it is possible to mitigate some of the ethical challenges *stegamsg* would pose in developing areas. Further investment into research, development, and integration of efficient data compression algorithms would help alleviate challenges related to the technical limitations of using images as containers for messages.

[8] RightMesh AG, Switzerland. `http://rightmesh.io`.

[9] Alphabet Inc. X Company. USA. `https://x.company/loon`.

3 Acknowledgements

I would like to extend special thanks to Major League Hacking for creating the *Ethical Tech* initiative. As someone who is passionate about ethics, technology, and human rights, I feel that this initiative is extremely important. Thank you for your consideration of this submission.

I would also like to specially thank Alexander Parent and Salem Abuammer for their collaboration and efforts in creating the prototype for `stegamsg`, the hack which I wrote about in my original abstract.

Special thanks also go to Thomas Kosciuch, for his unwavering strength and support in all my endeavours.

Finally, I wish to express gratitude for the excellent staff of instructors and supervisors at the University of Guelph's School of Computer Science. I acknowledge that the University of Guelph resides on the ancestral lands of the Attawandaron people and the treaty lands and territory of the Mississaugas of the Credit. I recognize the significance of the Dish with One Spoon Covenant to this land, and I offer respect to our Anishinaabe, Haudenosaunee and Métis neighbours as we strive to strengthen our relationships with them.

References

[1] Bergen, Mark. "Google Grapples With 'Horrifying' Reaction to Uncanny AI Tech." *Bloomberg*, 10 May 2018, www.bloomberg.com/news/articles/2018-05-10/google-grapples-with-horrifying-reaction-to-uncanny-ai-tech.

[2] Goel, Vindu, et al. "In Sri Lanka, Facebook Contends With Shutdown After Mob Violence." *The New York Times*, 8 Mar. 2018, www.nytimes.com/2018/03/08/technology/sri-lanka-facebook-shutdown.html.

[3] Kelly, Marie Louise, and Shane Mac. "Google's 'Duplex' Raises Ethical Questions." *NPR*, 14 May 2018,
www.npr.org/2018/05/14/611097647/googles-duplex-raises-ethical-questions.

[4] Kemp, Simon. "The Global State of the Internet in April 2017." *The Next Web*, 9 May 2017, thenextweb.com/contributors/2017/04/11/current-global-state-internet/.

[5] Kramer, Andrew E. "Chechen Authorities Arresting and Killing Gay Men, Russian Paper Says." *The New York Times*, 1 Apr. 2017, www.nytimes.com/2017/04/01/world/europe/chechen-authorities-arresting-and-killing-gay-men-russian-paper-says.html.

[6] Lemmetty, Sami. "Review of Speech Synthesis Technology." *Helsinki University of Technology Department of Electrical and Communications Engineering*, 1999, research.spa.aalto.fi/publications/theses/lemmetty_mst/thesis.pdf.

[7] Tufekci, Zeynep. (`@zeynep`) *Twitter*, 09 May 2018,
twitter.com/zeynep/status/994233568359575552.