

# Google’s Duplex and Deception through Power and Dignity

Nathaniel Grevatt

December 10, 2018

## 1 Introduction

In May of 2018, Google announced Google Duplex, “a new technology for conducting natural conversations to carry out ‘real world’ tasks over the phone” [6]. Specifically, Google created an artificially intelligent system capable of vocal communication with human beings which sounds “natural,” or alternatively, human. This focus on natural conversation seeks to “[allow] people to speak normally, like they would to another person, without having to adapt to a machine” [6].

While Duplex represents an amazing advance in the naturalness of artificial conversation, Duplex constitutes Artificial Narrow Intelligence (ANI), not Artificial General Intelligence, because Duplex can only carry out natural conversations in a few extensively trained domains [9]. The announcement revealed Duplex scheduling a hair salon appointment and Duplex calling a restaurant, and Google’s team described how focusing on these closed domains helped overcome the challenges of how to understand and adapt to natural language and how to model natural behavior. Yet despite the trivial nature of these example tasks, and the caveat of the narrowness of Duplex’s prowess, the media widely characterized Duplex as “scary” and questioned the possibility of an approaching AI “tipping point” [7].

Months later, Google announced “Call Screen,” a new role for Duplex as a means for Google users to screen phone calls, (marketed as only for use with calls from unknown numbers) [8]. Instead of using Duplex to converse with business employees to handle a user’s tasks, Call Screen enables users to have Duplex converse with unknown callers. This occurs under the user’s but not the caller’s

supervision, and the user can direct Duplex as needed or switch to an actual phone call as desired. In this new feature, Duplex identifies itself as a virtual assistant at the beginning of the call, perhaps influenced by California’s new law [1]. The media reaction to this new purpose for Duplex differed substantially from that to Google’s initial announcement, as Call Screen was characterized as useful rather than scary [8].

In this essay, I explore the idea of Duplex for “real world” tasks and its use as Call Screen to attempt to compare the two uses in a manner which explains the differing reactions of the media. To accomplish this, I consider the power dynamic in each scenario between caller and answerer and analyze the implication of Duplex’s role within this context. I argue that the public approval of Duplex’s use as a technical tool depends on which side of this power dynamic it falls on, yet there also exists a question of what it means for Duplex to converse so naturally. I attempt consider the implications of this reality for the human individuals with which it interacts, and I explore the effect of Duplex disclosing its artificiality and the extent to which this enables those interacting with it to consent. I also contextualize Duplex’s deceptive naturalness in the recent California bill, which mandates bots to disclose their artificiality, and consider how this relates to and characterizes Duplex.

Ultimately, Duplex’s technical success in emulating natural vocal conversation enables users to wield a powerful artificial assistant to reshape their interaction with other humans. And while Duplex remains only narrow artificial intelligence, its level of intelligence already enables users to exercise a (potentially imperceptible) power relation on other individuals through articulating a notion of the user’s time and presence as more valuable than that of others. And while such unequal valuations of one individual’s time in relation to another individual’s already exist (some affluent individuals hire personal assistants to perform the functions of Duplex), Duplex’s artificial nature increases the moral weight of this unequal valuation by challenging the dignity of those Duplex interacts with.

## **2 “Real World” Tasks**

Duplex purports to empower both the user placing the call (the caller) and the individual called (the callee), but while the technology improves the callee’s experience of interacting with a computer, it also creates the possibility of subjecting a callee to interacting with a computer, in a manner which disproportionately benefits the user. Google’s initial announcement of Duplex focused on Duplex’s ability

to make natural sounding calls to complete tasks for the user. The announcement contextualized the technical inspiration for Duplex within a narrative of humans seeking better and more natural communications with computers:

“Even with today’s state of the art systems, it is often frustrating having to talk to stilted computerized voices that don’t understand natural language. In particular, automated phone systems are still struggling to recognize simple words and commands. They don’t engage in a conversation flow and force the caller to adjust to the system instead of the system adjusting to the caller” [6].

And when describing Duplex’s innovation, Google employed a similar language emphasizing the benefit for the human interacting with Duplex:

“For such tasks, the system makes the conversational experience as natural as possible, allowing people to speak normally, like they would to another person, without having to adapt to a machine” [6].

These quotes describe the merits of Duplex for a human seeking to interact with a computer, yet they seem out of touch with the initially stated use for Duplex. Specifically, while Duplex improves the experience of the interacting individual, focusing on this improvement alone ignores the extent to which users empowered by Google can subject employees of salons and restaurants to calls which require their time and presence, but not the attention of the caller. This creates an unequal relationship in a scenario that would otherwise be reciprocal (at least, to the extent which monetary transactions in the service industry are reciprocal).

The media criticism of the Duplex’s launch took issue with this possible future reality of a well-meaning service employee unwittingly interacting with an AI instead of the person with which they would do business [10]. But the characterization of Duplex as “scary” stems more from the juxtaposition between the deceitful humanity of the AI and the knownness of the business and implied trust of a future transaction than the possibility of a new power dynamic. This distinction highlights the importance of disclosure in situations involving conversational artificial intelligence. After Duplex’s debut, a Google representative told *The Verge* that it believes it has a responsibility to inform individuals conversing with Duplex, though Google did not acknowledge this during the demonstration [10]. Had Duplex disclosed its artificial identity to the businesses it called, the business might have ended the calls prematurely, but this would have likely changed the public’s initial reaction to Duplex.

Anticipating other concerns, Google has suggested that businesses can join a Do Not Call List to opt-out of Duplex calls. And as a result, businesses will have to decide if and how Duplex alters the experience of their employees and if Duplex infringes upon their dignity, and weigh these costs against the increase in business Google promises. In the future, businesses might also enlist their own AI conversational assistants to interact with customers or use services to communicate directly with Duplex and other AI assistants to approximate application programming interfaces for making reservations. Yet if Duplex becomes available on a large scale, the carefree reservation-making Duplex enables could pose another problem: if everyone can automate calling restaurants to ask for a reservation, phone reservations would lose meaning as individuals speculate and make unnecessary reservations [10], and over time, making a reservation might become impossible due to the buildup of extraneous reservations. In response, restaurants could choose to charge for reservations or adopt a ticketed system, though consumers have previously met such change with resistance [5]. Google assured *The Verge* that it enforce a daily limit on how many calls a business can receive from Duplex and how many calls a given Duplex assistant can place to avoid use of Duplex for spam, but other providers of Duplex-like technology might not exercise such diligence.

### 3 Screen Call

Whereas some of the media condemned the initial announcement of Duplex, they responded positively to Google’s announcement of Duplex’s Screen Call functionality. A few contextual differences in this use of Duplex contribute to this different perception. Specifically, Call Screen reverses the nature of the power dynamic: the unknown caller does not have the trust of the user and users often assume unknown callers are malicious, and Duplex’s technology identifies itself and enables user oversight. These differences significantly change Duplex’s role in the interaction and the meaning of its use.

The different mediating role Duplex plays when screening calls increases the ethicality of using an AI conversational assistant. Many of the concerns about the use of Duplex to make reservations worried about how an AI conversational assistant would reinforce existing inequalities of power between the hypothesized affluent Duplex-using caller and the “most likely low-paid service workers” on the other end of the call [10]. *The Verge* wondered if Duplex would become another example of “tech privilege,” and *The Atlantic* questioned if “low-wage

workers [might] increasingly interact with bots and screens” in the near future. In these situations, the implication of a future economic transaction establishes a grounds of respect, and the undisclosed use of an AI conversational assistant infringes upon this trust and articulates the inequality of the relationship, as critiqued by the media. The context of Duplex as Call Screen differs fundamentally because the interaction does not begin with a relation of respect, as the individual called does not know the unknown caller, and there is no implied future economic transaction from which to establish unknowing respect. Additionally, the called individual lacks information and is thus disempowered relative to the unknown caller. Because using Duplex to screen the call does not constitute an exercise in reinforcing the power of the called user over the unknown caller, and serves as more of an attempt at leveling the playing field, this use of Duplex does not raise the same concerns as making reservations with Duplex.

Duplex’s disclosure of its own artificiality at the beginning of a Call Screen constitutes an essential difference between this use and its original use for making reservations. And this small detail likely appeased the concerns about the asymmetrical degrees of information and power present in use of Duplex to making reservations. Specifically, if Duplex notifies the individual that they are conversing with an Artificial Intelligence at the start of the interaction, no deception occurs, and the individual can decide for themselves whether the interaction violates their dignity and proceed with the call as they see fit. (Society currently deems it reasonable to hang up on, or block entirely, spam phone calls [12], and Duplex would likely fall into this category unless it attains acceptance). Of course, employers could require their employees to treat Duplex with the respect given to a human assistant, but at least the employees would suffer no deception in the case of disclosure.

## **4 Duplex and the California Senate Bill No. 1001**

The media interpretations of Duplex make a compelling case for disclosure for public acceptability, and California’s recent legislation on bots requires some forms of bots to disclose their artificiality. Specifically, *California Senate Bill No. 1001* declares:

“It shall be unlawful for any person to use a bot to communicate or interact with another person in California online, with the intent to mislead the other person about its artificial identity for the purpose of

knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election. A person using a bot shall not be liable under this section if the person discloses that it is a bot” [1].

The language of the bill does not explicitly require Duplex to disclose its artificiality, yet the reservation making application of Duplex parallels the behavior of the bots which the bill addresses. Duplex does not interact with individuals online, but its phone-based means of interaction leads to scenarios in which identifying the artificial identity of Duplex is even more challenging than identifying online bots, due to its believable conversational skills. A critical reading of the bill could also interpret Duplex’s advanced conversational capabilities as a deliberate attempt to “mislead the other person about its artificial identity.” Duplex does not strictly attempt to “incentivize” the individuals it interacts with to accept its proposed reservations. Yet in posing the reservation and offering the possibility of a future economic transaction to employees in a believably human manner, Duplex provides little room for exercising agency, because the employees cannot discriminate and deny service. While prosecuting an individual for using Duplex to make a reservation at a restaurant might not hold up well, the developers of Duplex could eliminate the possibility of this occurrence by building in mandatory disclosure.

#### **4.1 History of the Bill and the Influence of the Electronic Frontier Foundation**

In order to understand the rationale for the language of the bill, which does not restrict the use of Duplex to make reservations without disclosure, it is important to consider how and why the bill arrived at its current language. The first draft of the bill, proposed by California State Senator Robert Hertzberg, did not limit disclosure requirements to only the types of bots mentioned in the final bill [2] and focused primarily on social media bots [4]. Additionally, the initial draft included detailed requirements for disclosure over “audio communications,” which would mandate disclosure by all existing forms of Duplex.

In April of 2018, the Electronic Frontier Foundation (EFF) sent an open letter to California State Senator Jerry Hill, expressing concern over the bot bill. The letter primarily critiqued the measure in the bill which required social media platforms to implement content moderation features for reporting bots. Specifically,

the EFF discussed the history of misuse of content moderation features by users to maliciously censor non-offending users and warned that malicious users could misuse the tools described in the bill to harm those the bill sought to protect [13]. The stipulation for features to facilitate reporting bots no longer exists in the current bill, and the EFF claims they worked with Senator Hill to accomplish this “crucial victory” for online anonymity and lawful human speech [14].

The EFF’s open letter also described how a blanket requirement that all bots disclose their artificiality would limit non-malicious forms of “ordinary speech activities” [13]. Specifically, the EFF suggested that “Speech generated by bots is often simply speech of natural persons processed through a computer program” [13], and the EFF cited “poetry” and “political speech” as examples of expression which a general law on bot disclosure would restrict. From these examples, it seems that the EFF’s concern lies not with the extent to which the initially proposed legislation would limit the availability of expression through bots, but rather, the degree to which the legislation would reduce the quality of expression available to natural persons through deceptive bots. To this extent, the EFF suggests that deception by bots is not always negative, though their argument focuses on the individuals choosing the deception rather than those interacting with it.

The amended bill preserves the right to the sorts of anonymous expression advocated for by the EFF and creates some protections for individuals against certain types of deceptive bots, yet the sentiment expressed by the EFF challenges the initial goals of the bill. Specifically, the argument for anonymous expression, and other critiques which invoke the first amendment [11], focus on the impact of the bill for individuals who use bots, whereas the bill initially sought to protect individuals from exposure to misleading and potentially malicious expression through bots. Furthermore, arguments for preserving anonymous expression through bots do not consider the potential impact of such expression on other individuals, such as false or hateful information [4]. However, the EFF’s critique makes sense when considering the bill as a precedent: from an online liberties perspective, a continuing problem of malicious users expressing themselves anonymously through bots is likely preferable to not having the option of expression through bots. The EFF correctly challenges a blanket requirement for disclosure, but determining the permissibility of deception for a type of bot through the extent to which it empowers its creator proves incompatible with the stated goals of the initial bill [4].

## 4.2 Beneficial Believability in Bots

The EFF argues against a comprehensive stipulation for bot disclosure with the claim that not all deception by bots is harmful. Yet there exist more compelling arguments to permit deception in which a seemingly human bot brings benefit to an individual because of its believability. For example, a family caring for an older relative in an end of life scenario might choose to enlist a virtual caregiver or companion for their relative. Whether the bot reveals its artificiality or obscures it to appear human would likely impact the experience of the individual interacting with it. And a believably human bot might enable the individual to have a more meaningful connection and thus benefit more from its care.

Of course, this decision comes with many considerations for the family and the individual in question. A virtual assistant would likely be able to provide constant care for the individual to supplement that which would usually be provided by a human caregiver, but the ideal qualities of such a virtual caregiver are not as certain. And while alluring, if the family chooses to have the bot not disclose its identity, their deception might harm the relative and challenge their relationship.

This example falls within the specific context of a single family and in a situation involving end of life care, both of which differ from the sorts of social interaction targeted by the bill. Yet it seems plausible that a believably human bot which does not disclose its artificiality can act not only non-maliciously but towards the benefit of an individual. And regardless of whether transparency or deception is more moral in this situation, individuals should have the option to make the choice. This argument could also rationalize the design choice to make Duplex's conversational patterns so natural-seeming.

## 5 Conclusion

Google's Duplex represents an amazing technological accomplishment and presents a compelling option for users if developed in a manner which makes its use socially and legally acceptable. Duplex could empower users through increasing the accessibility and ease of existing services by enabling anyone to place or receive a call in a believable manner. Enforcing disclosure in Duplex would ameliorate the concerns raised by the media about the ethics of Duplex's use, and disclosure might also satisfy potential legal problems with using Duplex under future iterations of the California bot bill. Due to the limited set of current uses of Duplex, developing a rule about when and in what roles using Duplex is acceptable

presents a challenge, but the general trend seems to disapprove of uses of Duplex which worsen existing inequalities. Ultimately, because of the substantial difference disclosure makes in the efficacy of Duplex's use, legislation which requires artificial intelligence to disclose its artificiality seems advisable, at least until interacting with AI becomes more socially acceptable.

If deceptive bots can improve the experience of individuals, mandating disclosure for all bots seems unwise. Yet the test for determining the acceptability of deception invoked in the previous example: "Does deception improve the experience of the individual the bot interacts with?" falls short. For example, false information can affirm an individual's perspective and improve their experience yet generally does not serve their best interest. Perhaps the impact of deception by bots is best understood through motivation: "Why deception?" or "Deception, to what end?" These sorts of questions could lead legislators to a more open and powerful legal understanding of deception in AI and could help the developers of AI approach the ethical dimensions of deception in AI.

The current language of the California bill seeks to safeguard individuals from deception through this lens, identifying harmful deception and in bots which exhibit deliberate deception to malicious ends. The bill will hopefully protect individuals from the types of bots it specifies when it comes into effect on July 1, 2019. Though, because of its particular language, it seems unlikely to result in the "Botageddon" *Quartz* predicts. [3].

## References

- [1] *Bill Text - SB-1001 Bots: disclosure*. [https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill\\_id=201720180SB1001](https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001). Sept. 2018.
- [2] *Bill Text - SB-1001 Bots: disclosure*. [https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill\\_id=201720180SB1001](https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180SB1001). Mar. 2018.
- [3] Dave Gershgorn. *A new law means California's bots have to disclose they're not human*. <https://qz.com/1409350/a-new-law-means-californias-bots-have-to-disclose-theyre-not-human/>. Oct. 2018.
- [4] *Hertzberg Announces Legislation to Encourage Social Media Transparency*. <https://sd18.senate.ca.gov/news/212018-hertzberg-announces-legislation-encourage-social-media-transparency>. Feb. 2018.
- [5] Maura Judkis. *Would you pay to make a reservation at a hot new restaurant? You might have to*. [https://www.washingtonpost.com/lifestyle/food/why-your-next-hot-restaurant-reservation-may-come-with-a-price/2016/04/04/c7919d86-d5bb-11e5-9823-02b905009f99\\_story.html?utm\\_term=.e450e7836bdc](https://www.washingtonpost.com/lifestyle/food/why-your-next-hot-restaurant-reservation-may-come-with-a-price/2016/04/04/c7919d86-d5bb-11e5-9823-02b905009f99_story.html?utm_term=.e450e7836bdc). Apr. 2016.
- [6] Yaniv Leviathan and Yossi Matias. *Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone*. May 2018. URL: <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>.
- [7] Molly Roberts. *Could Google's creepy new AI push us to a tipping point?* May 2018. URL: <https://www.washingtonpost.com/blogs/post-partisan/wp/2018/05/10/could-googles-creepy-new-ai-push-us-to-a-tipping-point/>.
- [8] Tom Simonite. *Google's Human-Sounding Phone Bot Comes to the Pixel*. Oct. 2018. URL: <https://www.wired.com/story/google-duplex-pixel-smartphone/>.

- [9] Tim Urban. *The AI Revolution: The Road to Superintelligence: Part 1*. Jan. 2015. URL: <https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html>.
- [10] James Vincent. *Google's AI sounds like a human on the phone — should we be worried?* <https://www.theverge.com/2018/5/9/17334658/google-ai-phone-call-assistant-duplex-ethical-social-implications>. May 2018.
- [11] John Weaver. *Everything Is Not Terminator: We Need The California Bot Bill, But We Need It To Be Better*. <http://www.mondaq.com/unitedstates/x/739116/new+technology/The+content+of+this+article+is+intended+to+provide+a+general+guide>. Sept. 2018.
- [12] Chris Welch. *Spam calls: how to stop the robots from calling your iPhone or Android*. <https://www.theverge.com/2018/3/6/17071478/spam-calls-how-to-stop-block-robocalls-robots-scam-iphone-android>. May 2018.
- [13] Jamie Williams. *EFF Letter Opposing California Bot Disclosure Bill, SB 1001*. <https://www.eff.org/document/eff-letter-opposing-california-bot-disclosure-bill-sb-1001>. May 2018.
- [14] Jamie Williams and Jeremy Gillula. *Victory! Dangerous Elements Removed From California's Bot-Labeling Bill*. <https://www.eff.org/deeplinks/2018/10/victory-dangerous-elements-removed-californias-bot-labeling-bill>. Oct. 2018.