AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations

Jeffrey T. Hancock [1], Mor Naaman [2,3], & Karen Levy [2]

1 Department of Communication, Stanford University, Stanford, CA 94305, USA
2 Information Science Department, Cornell University, Ithaca, NY 14850, USA
3 Cornell Tech, Cornell University, New York, NY 10044, USA

We define Artificial Intelligence-Mediated Communication (AI-MC) as interpersonal communication in which an intelligent agent operates on behalf of a communicator by modifying, augmenting, or generating messages to accomplish communication goals. The recent advent of AI-MC raises new questions about how technology may shape human communication and requires re-evaluation – and potentially expansion – of many of Computer-Mediated Communication's (CMC) key theories, frameworks, and findings. A research agenda around AI-MC should consider the design of these technologies and the psychological, linguistic, relational, policy and ethical implications of introducing AI into human–human communication. This article aims to articulate such an agenda.

Keywords: Interpersonal Communication, Computer-Mediated Communication (CMC), Artificial Intelligence (AI), Artificial Intelligence-Mediated Communication (AI-MC), Language, Impression Formation, Relationships, Ethics

doi:10.1093/jcmc/zmz022

The advent of Computer-Mediated Communication (CMC) revolutionized interpersonal communication, providing individuals with a host of formats and channels to send messages and interact with others across time and space (Herring, 2002). In the classic social science understanding of CMC (e.g., Walther & Parks, 2002), the medium and its properties play important roles in modeling how actors use technology to accomplish interpersonal goals. Agency remains with the communicator: message production and impression management are broadly understood to manifest the communicator's goals. Similarly, the message receiver is assumed to understand and accept that agency.

The introduction of AI into interpersonal communication has the potential to once again transform how people communicate, upend assumptions around agency and mediation, and introduce new ethical questions. CMC is now expanding to include *Artificial Intelligence-Mediated Communication (AI-MC)*:

Corresponding author: Mor Naaman; e-mail: mor.naaman@cornell.edu
Editorial Record: First manuscript received on 1 November 2018; Revisions received on 16 March 2019 and 29 June 2019; Accepted by Dr. Mike Yao on 31 July 2019; Final manuscript received on 9 September 2019

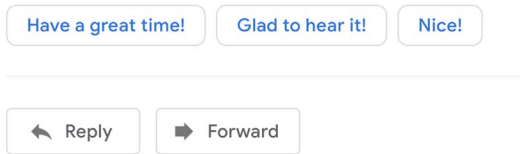

Figure 1 Gmail's AI-generated suggested responses often trend positive.

interpersonal communication that is not simply *transmitted* by technology, but *modified*, *augmented*, or even *generated* by a computational agent to achieve communication goals.

Conceptualizing AI-MC

Our definition of AI-MC builds on Russell and Norvig's description of AI as a computational "rational agent" that acts given inputs (percepts) to achieve the best expected outcome (2010, p. 4). This definition casts AI in terms of agent behavior, and is not concerned with how the agent reasons. Following this definition, we use AI to refer broadly to computational systems that involve algorithms, machine learning methods, natural language processing, and other techniques that operate on behalf of an individual to improve a communication outcome. The computational agent may analyze inputs including human-authored messages, communication history, personal information, or any other source of data. The agent may then suggest, augment, modify or produce messages to achieve an expected outcome.

CMC research can be defined as the study of social effects of communication that takes place between people using network-connected digital devices to exchange messages (e.g., email and text messaging, social network site interactions, videoconferencing) (Thurlow, Lengel, & Tomic, 2004). We follow Walther and Parks' (2002, p. 530) emphasis on social scientific approaches to the interpersonal dynamics of human-to-human communication via technology.

We integrate these AI and CMC conceptualizations to define AI-MC as: *mediated communication between people in which a computational agent operates on behalf of a communicator by modifying, augmenting, or generating messages to accomplish communication or interpersonal goals*.

We are already beginning to see examples of AI across many types of CMC. In verbal channels, the use of AI has advanced text-based communication from auto-correct, predictive text, and grammar correction (Grammarly, 2018) to smart replies, auto-completion, and auto-responses (e.g., in Gmail and mobile phones). For example, in Gmail's smart replies an email recipient can select one of several responses produced by AI (Figure 1). This trend is equally, if not more, advanced for nonverbal CMC, such as the auto-insertion of emoji. Commonly, the involvement of AI is not disclosed, with the partner presumably assuming that the message was produced by the sender.

Future AI-MC systems can go even further, optimizing messages for interpersonal outcomes like conveying high status (Pavlick & Tetreault, 2016) or appearing trustworthy (Ma et al., 2017). Moreover, emerging technologies can optimize messages for a specified receiver: Crystal (Zeide, 2018) informs a sender about "how [the recipient] wants to be spoken to" based on the recipient's social media profile; Respondable (Boomerang, 2018) uses AI to advise email writers about how to "strike the right tone when emailing your boss." Future AI-MC systems will be able to do more than just suggest text. As algorithms for natural language generation improve (Graves, 2013), AI technologies will be able to wholly generate messages on behalf of a sender—including creating online profiles, or even generating messages in synchronous communications (Statt, 2018). With rising concerns around AI's involvement

Table 1 Dimensions of AI-MC

Dimension	Definition	Examples
magnitude	The extent of the changes that AI enacts on messages	Correcting spelling errors vs. generating entirely new messages
media type	The media in which AI operates (e.g., text, audio, video)	Suggesting text replies vs. modifying one's appearance in video
optimization goal	The goal for which AI is optimizing the messages	To appear attractive, trustworthy, humorous, dominant, etc.
autonomy	The degree to which AI can operate on messages without the sender's supervision	Sender chooses between AI suggested messages vs. AI engages in conversation with minimal input from the sender
role orientation	The role that the AI is operating on behalf of (e.g., sender vs. receiver)	Sender: offering messages to enhance reply efficiency vs. Receiver: assessing whether sender is potentially lying

in human communication, like deep fakes in which AI is used to create a misrepresentation of what a person says or does in audio or video (Suwajanakorn, Seitz, & Kemelmacher-Shlizerman, 2017; Thies, Zollhofer, Stamminger, Theobalt, & Niessner, 2016), the need to understand its effect on interpersonal communication is urgent.

Dimensions of AI-MC

The examples above illustrate several dimensions that could be useful in characterizing AI-MC interventions (see Table 1). The first is the *magnitude* of the involvement by the AI agent. These changes can range from minimal suggestions (e.g., change in wording) to full content generation. Another key dimension is the *media type* used in communication, such as text, audio and video. Currently available automatic editing tools and filters—for example, methods to detect and fix images where the subject is blinking (Emerson, 2018)—are first steps toward technologies that make one look more attractive (Leyvand, Cohen-Or, Dror, & Lischinski, 2008), more trustworthy, or more similar to the receiver (Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013). Audio can be manipulated to make a speaker sound more calm or authoritative (Klofstad, Anderson, & Peters, 2012), or synthesize entirely new speech for a given speaker (Vincent, 2017). Other tools will allow individuals to realistically portray deceptive signals in video—say, lifting heavy weights or dancing like a professional (Chan, Ginosar, Zhou, & Efros, 2018).

S*ynchronicity* is another key dimension of AI-MC systems. Synchronous forms of CMC will likely become AI-mediated, especially with recent advances in real-time audio and video manipulation (Suwajanakorn et al., 2017). Real-time video "filters" will allow individuals to change their appearances (Shah and Allen, 2019) and expressions even in live computer-mediated conversations (Thies et al., 2016).

AI-MC can also be categorized by the *optimization goal* for which it is deployed. The goals for interpersonal communication are manifold, and there are already several goal taxonomies to describe them (e.g., Chulef, Read, & Walsh, 2001). For example, self-presentation goals of appearing attractive, likable, or competent are important interpersonal functions that AI could conceivably be trained to optimize communication for.

The *autonomy* granted to the AI to operate on behalf of the communicator is another key dimension. This is similar to the principal–agent relationship (Sappington, 1991), in which the human communicator is the principal who delegates some authority and autonomy to the AI agent. For example, in smart replies, the principal retains substantial agency: they choose which suggested message to use or to ignore, and may also modify the message. Future AI systems may be granted much more autonomy to engage in communication tasks without supervision by the principal, from automated and personalized birthday wishes to automated scheduling (Statt, 2018) or online dating conversations.

Finally, the *role orientation* of the AI agent is important. Most current instantiations of AI in communication tools are sender-oriented, but we imagine receivers will increasingly use AI tools. Google Translate allows both sender and receiver to converse, using AI as mediator. One may imagine other tools that will claim to help receivers, for example by offering to extract social cues (Razavi, Ali, Smith, Schubert, & Hoque, 2016), or detect emotion, deception and lies (Sen, Hasan, Tran, Levin, Yang, & Hoque, 2018) from real-time speech. As AI-MC research continues we expect this initial set of dimensions to evolve.

Boundaries of AI-MC

For clarity, it is useful to outline some examples of what we do *not* consider to be AI-MC as we have defined it here. Most closely related is the growing field of AI–human interaction, or the study of human interactions with bots and other smart virtual agents who *do not* represent other individuals, such as Apple's Siri or Amazon's Alexa. This area, sometimes called Human–Machine Communication, overlaps with the scope of AI-MC, though in our formulation the interesting questions involve the introduction of AI that operates on *communication between people*. For example, AI-MC bots could be used to multiply how many people an individual can talk to interpersonally (e.g., running for political office and using a bot to talk to many possible voters).

AI-MC could be more broadly conceptualized to include all algorithms that mediate human communication, like the Facebook "Newsfeed" and other content ranking, recommendation and classification algorithms (e.g., email filters, friend suggestions) that use algorithms to support human communication. This type of conceptualization is too large to serve as a useful research framework and we therefore exclude these kinds of algorithmic tools that indirectly support human communication.

Research agenda

The introduction of AI-MC has the potential to upend and extend existing CMC knowledge, theories, and scholarship. The use of AI in interpersonal communication challenges CMC assumptions of agency and mediation and may subvert existing social heuristics. Here, we aim to lay the foundation for studying the emerging field of AI-MC, providing a research agenda that revisits core CMC topics, from linguistic and communication processes to relational and interpersonal dynamics to ethical, cultural, and policy implications. We offer research themes that progress from lower to higher levels of analysis: from the design of AI-MC and its immediate impact on individuals interacting with these systems, to relational aspects of AI-MC, and to its potential longer-term societal and cultural impact. Finally, as AI-MC systems expand into more interpersonal domains, their social and ethical implications become increasingly salient. We outline these areas below.

Design of AI-MC

A first set of research questions concerns how people *interact with and understand* AI-MC. For senders, how do design choices influence their use AI-MC suggestions like smart replies (Hohenstein & Jung,

2018) or take advantage of AI-MC's help in crafting a profile? For receivers, are they aware of the fact that AI-MC is involved in communication, and if so, what are the outcomes of such awareness? Recent work suggests that mixing AI- and human-generated profiles can negatively impact trustworthiness ratings of profiles *suspected or labeled* as AI (Jakesch, French, Ma, Hancock, & Naaman, 2019). Future research could study the type and efficacy of disclosure of AI's involvement, while considering the factors that make people suspect AI has been involved in communication.

Design choices may also influence the interpretation of agency with AI-MC, especially given that people accept algorithmic input differently than human input (Dietvorst, Simmons & Massey, 2015; Waddell, 2018). Under which design conditions will AI involvement be perceived as augmenting a sender's agency, and when will it be viewed as usurping it? An early study shows that when task-oriented conversations were not successful, participants assigned significantly *less* responsibility to their partners when AI mediation (smart reply) was used (Hohenstein & Jung, 2019). Research is required to understand when AI-MC is viewed as *a filter* on human representation, as an independent social agent, or somewhere in between.

The question of agency will be complicated by the context of use. For example, AI-mediation is widely accepted when used to improve clarity, like auto-correct or machine translation between languages in text-based communication (Xu, Gao, Fussell, & Cosley, 2014). It is less clear how AI's role in optimizing self-presentation goals will be accepted. Consequently, it is essential to examine the contexts in which AI-MC is used, the kinds of goals it is used to accomplish, and how it is perceived in each context by senders and receivers.

Impact on language

From the outset of CMC research, questions about how technology affects the way people write and interpret messages have been a core concern (Herring, 2008). AI-MC raises these questions anew, and introduces a host of new inquiries.

One important issue is AI-MC's potential to shape human language and thought. According to the interactive alignment model (Pickering & Garrod, 2013), language production and comprehension are tightly interwoven in a process that can help create linguistic alignment between speakers. Thus, when AI-generated text is inserted into dialogue it is likely to influence this alignment, with the potential to modify not only the speaker's word choices but also the partners. For example, Gmail's smart reply function (Figure 1) provides three options for a response, priming syntactic and semantic content whether the sender chooses to use them or not.

A study of "smart reply" suggestions in text messaging revealed that they were overly positive ("sounds great!") (Hohenstein & Jung, 2018). This excess of positive language could cause the sender and receiver to also use more positive language in subsequent messages. These "smart replies," now part of Google's popular Gmail service, and "smart compose," in which Gmail suggests automatic completions for words and sentences, have introduced AI-MC to millions of conversations worldwide, yet we know little about their impact on language use dynamics at the individual or societal level. Given its scale, Gmail's overly-positive language suggestions have the potential to shift language norms and expectations even when communicators are not using these tools, and produce long-term language change over time.

Further impact of AI-MC on language may be in how people adapt to one another in interpersonal communication. Interpersonal adaptation, or modifying one's behaviors to adjust to a communication partner, is fundamental to social interaction (Toma, 2014). This process may be disrupted or intensified when AI-MC is involved. What happens when AI-MC plays a role in suggesting or generating messages by one of the communicators? Research is needed to investigate the impact of these suggestions and

other AI-MC language trends, from pragmatic phenomena (e.g., politeness and rudeness) to register phenomena (gender styles, regional dialects, and ingroup languages used in online communities), all of which could be impacted by AI.

The implementation of AI-MC systems can further balkanize or homogenize language and linguistic practices by allowing—or concealing—personal or group-based variations. In organizational contexts AI-MC could alter the diffusion of shared social semantics. For example, in the process of *constrained searching* authors use concepts from a restricted (e.g., political or organizational) semantic population (Margolin & Monge, 2013). If AI's suggestions are drawn from this population, this tendency will be exaggerated, resulting in terminology and concept balkanization. Such subtle decisions, activated here almost invisibly through simple AI-MC mechanisms like suggested text, may have a profound impact on our common language and terminology.

Interpersonal dynamics

CMC research has often examined how communicating through different technologies may change our messages, how they are crafted, and how they are perceived. AI-MC raises important questions about these processes. Consider, for example, the key aspects of the Hyperpersonal Model of Communication (Walther, 2011): the sender, receiver, channel, and feedback. How will the sender's selective self-presentation, afforded by channel attributes, be transformed when AI tools can be used to modify a sender's messages to achieve desired impressions? Will receivers change how they scrutinize messages if they know the message was optimized by a machine?

Social Information Processing (Walther, 2011) suggests that language content and style characteristics are primary conduits of interpersonal information in CMC, but what happens to these cues when messages are not only transmitted through technology, but also modified or generated by AI? One possibility is that these cues will lose their diagnosticity for interpersonal information as receivers come to understand that AI is modifying language content and style. Another possibility, however, is that perceived agency will remain with the sender, just like receivers continued to attribute responsibility for *spelling* errors to the sender and not to failures by the spellcheck system (Figueredo & Varnhagen, 2005).

Self-presentation, impression formation and trust

AI-MC is likely to impact both how people present themselves online and how they evaluate others. We expect that theories and frameworks concerned with signaling and trust, like Profile as Promise (Ellison, Hancock, & Toma, 2012), Warranting (DeAndrea, 2014) and Signaling Theory (Donath, 2007) will need to be updated for AI-MC, where the use of AI tools can undermine—or enhance—the reliability of profiles and messages. For example, in CMC, evaluators must interpret signals presented online to infer characteristics about the individual, and the awareness that a communication partner has used AI may affect perceptions of online profiles. According to Warranting Theory, AI-generated information could be perceived as more warranted (DeAndrea, 2014) given that messages and profiles can seem more objective when generated by a computer (Sundar, 2008).

Conversely, AI mediation in self-descriptions can violate expectations and activate concerns about deception and manipulation. Recent work has demonstrated a robust effect on the evaluation of trustworthiness when AI is involved with self-presentation in the context of Airbnb (Jakesch et al., 2019), suggesting that in a system that mixes human- and AI-written profiles the trustworthiness of profiles *perceived* to be written by AI will decline. This effect, dubbed the "Replicant Effect," highlights the complicating role AI-MC may have for theories regarding self-presentations online.

AI-MC is likely to also affect how individuals describe themselves and disclose information online, potentially requiring reconsideration of CMC frameworks and theories concerned with self-disclosure. For example, a central tenet of the Profile as Promise framework (Ellison et al., 2012) is that people expect online profiles to be only approximate reflections of their offline identity given the constraints of presenting oneself online—but that may change with the presence of AI, resulting in different self-presentation and self-disclosure dynamics. For both receivers and senders, these questions are not limited to text alone, but extend to presentation via images and videos.

Feedback and relationships

AI-MC's effects may go beyond short-term communication and first impressions to introduce novel effects on the self, partners, and relationships. Identity shift, for example, suggests that selective self-presentation in CMC leads to corresponding changes in how we perceive our real-world selves (Gonzales & Hancock, 2008). It is unclear, however, whether or how identity shift will happen when a machine modifies one's communication behavior. If AI modifies a sender's messages to be more positive, more funny, or extroverted, will the sender's self-perception shift towards being more positive, funny, or extroverted? There are clearly behaviors that might have more worrisome consequences, such as using AI-MC to manipulate or deceive others leading to identity shifts towards considering oneself manipulative or deceptive.

AI-MC could also impact how we maintain and develop meaningful relationships with others, following CMC's effects on intimacy, attraction, and relationship maintenance (Tong & Walther, 2011). The amount of effort that partners dedicate to a relationship plays a crucial role in relationship maintenance. With the introduction of AI into communication processes, such as automated birthday wishes on social networking sites, will the relational value of expressions of intimacy or gratitude be undermined when AI plays a role? Research efforts can evaluate when AI-MC undermines the perception of effort in a relationship, and whether this effect carries over to perceived intimacy.

Policy, culture, and ethics

As AI-MC systems are developed, it is critical to assess their social and ethical implications. AI-MC systems are already deployed at scale, and may have widespread social impacts that disproportionately impact vulnerable populations. Research must attend to how AI-MC systems balance normative values, what factors policymakers should consider in governing AI-MC, and what practices designers can employ to mitigate these concerns. We describe some major areas of concern here.

Bias and fairness

AI-MC systems are trained, at least initially, on existing human communication; extant biases in training data are likely to be replicated and amplified by AI. The effect of such systems at scale may be to ossify social power structures as manifested in communication norms. When AI-MC modifies communications to be more socially efficacious (e.g., by sounding more authoritative), we might question the implications of normalizing certain styles of interaction while subsuming others. For example, the Chrome extension *Just Not Sorry* alerts users when they use qualifying words like "sorry," and "I think," that such words can undermine the writer's message (Cauterucci, 2015); the extension is aimed at women, and in encouraging them to make such changes, normalizes brasher language as the "right" way of speaking.

In other contexts AI-MC may help *mitigate* interpersonal biases. If AI-MC imparts signals of trustworthiness between peer-based social exchange partners, these countervailing cues may neutralize stereotypes that would otherwise impede the transaction (Levy & Barocas, 2018). Alternatively, AI-MC

may directly encourage users to modify communications that bear the mark of prejudice, e.g., using a warning dialog when an individual attempts to post negative comments.

Transparency: Must AI-MC reveal itself?

Another concern is whether, how, and when AI-MC systems ought to disclose their existence and functionality to users (senders and receivers). Sometimes, AI-MC may be so commonplace or non-controversial that disclosure is unnecessary: it seems absurd, for example, to require that a message bear a declaration that it has been auto-corrected. In other contexts we might require disclosure to aid in proper interpretation of the message and the sender's intent. For example, text that has been automatically translated using Google Translate often bears a disclaimer acknowledging as much; this disclosure might help a reader better interpret error-ridden text as the result of imperfect AI, not as a writer's intent.

There may be other ethical and social reasons for transparency. In May 2018, Google demonstrated its lifelike AI assistant Duplex at a conference by having it make an appointment over the phone. Duplex's oral presentation was replete with "um's" and "mm-hmm's," and did not reveal itself to be a robot. Google espoused Duplex's ability to interact naturally with humans, but critics lambasted the technology for deceiving the receiver about its true nature (Statt, 2018).

But disclosure is complicated. A recent California legislative proposal, the B.O.T. Act of 2018, requires "any person [using] a social bot to communicate or interact with natural persons online [to disclose] that the bot is not a natural person" (Williams, 2018). The measure seems to rely on the assumption that a disclaimer will better equip receivers to evaluate a message's intent and veracity. In addition to the empirical uncertainty of this assumption (Jakesch et al., 2019; Waddell, 2018), legal scholars have noted that broad AI-MC disclosure laws might have undesirable policy consequences: threatening First Amendment rights to anonymous speech, for example, or overburdening senders who have accessibility needs (e.g., text-to-speech users).

Even when transparency is warranted, there are important questions about what transparency requires and what goals it serves in different contexts. Is it sufficient for the use of an AI-MC system to be disclosed, or should more specific information about its objective function be made clear? How granular must explanation be? Do the same transparency requirements attach to both senders and receivers? These and other questions are likely to emerge in public discourse and policy around AI-MC.

Misrepresentation and manipulation

To an extent, all self-presentation techniques can be understood as strategically designed to impress upon others a particular understanding of the communicating parties and the situation (e.g., Ellison et al., 2012). But certain techniques strike us as crossing the line from representation to *mis*representation, from persuasion to manipulation. Manipulative AI-MC might have the aim of materially deceiving parties, perhaps by exploiting cognitive vulnerabilities or inducing false beliefs (Susser, Roessler, & Nissenbaum, 2018).

The appropriate lines between permissible misrepresentation and unethical manipulation are not firmly fixed, and dependent on the vulnerabilities of the parties, the intent of the communication, and the consequences of the action. Facebook researchers are training machine learning systems to recognize photos where the subject is blinking, and to replace closed with open eyes (Emerson, 2018). Though doing so technically misrepresents reality, many would consider it socially and ethically acceptable. In fact, *failing* to retouch an unflattering photo of someone might be considered impolite.

At the other end of the spectrum sit *deep fakes*, AI-synthesized videos meant to deceive or manipulate. Sundar's MAIN model (Sundar, 2008) includes the "realism heuristic," predicting that people are

more likely to trust the audiovisual modality because its content has a higher resemblance to the real world. Such manipulations may have widespread impact, for example on political communications, given that traits like attractiveness can predict electoral outcomes (Mattes et al., 2010) and changes in vocal pitch can improve leadership perception (Klofstad et al., 2012). With AI capabilities, candidate faces can be morphed, in photographs or in real-time, to match voters' preferences.

Between these two poles, social and ethical implications are muddier and contextually dependent. We might not object to the use of AI-MC in business exchanges, for example, where we might expect self-presentation to be highly strategic or even adversarial; but we might consider it inappropriately deceptive in intimate personal exchanges marked by greater expectations of authenticity (e.g., between romantic partners). Norms of acceptability are likely to evolve in response to the availability of new technologies.

Conclusion

AI-MC is advancing rapidly, with potentially critical impact in areas from interpersonal relationships to political decision-making. In the worst case, the effects may be quite bleak. If manipulative and false messaging is easier to generate and harder to detect, AI might render almost all CMC unreliable, undermining trust in any interaction besides those that occur face-to-face. But AI also has the potential to improve human communication by augmenting our natural ability to communicate with one another and improving the affordances of such interactions in CMC channels.

Many open questions remain about how AI tools will be used to optimize communications and achieve interpersonal goals. New hardware, software, laws, and policies will shape the role AI has in communication in the coming years. With the proper basis of evidence, these efforts might effectively address the new challenges raised by AI-MC. To design, implement, and regulate these systems responsibly, we must develop a foundational empirical understanding of their impact on a wide variety of behaviors, including impression management, trust, deception, language use, relationships and other key factors. We ask the community to join this effort and to expand consideration of AI-MC as an important topic of research.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CHS 1901151/1901329. We would like to thank Xiao Ma, whose work on online trust was the seed for the AI-MC framing; Xiao also came up with the phrase AI-MC. Other colleagues including Maurice Jakesch, Malte Jung, Megan French, Jess Hohenstein, Sunny Liu and the Stanford Social Media Lab provided ideas and feedback on prior versions.

References

Boomerang. (2018). Respondable: Write Better Email. Retrieved from https://www.boomeranggmail.com/respondable/.

Cauterucci, C. (2015, Dec. 29). New Chrome app helps women stop saying "just" and "sorry" in emails. Slate. Retrieved from http://www.slate.com/blogs/xx_factor/2015/12/29/new_chrome_app_helps_women_stop_saying_just_and_sorry_in_emails.html.

Chan, C., Ginosar, S., Zhou, T., & Efros, A.A. (2018). Everybody dance now. *arXiv preprint*. arXiv:1808.07371.

Chulef, A. S., Read, S. J., & Walsh, D. A. (2001). A hierarchical taxonomy of human goals. *Motivation and Emotion*, 25(3), 191–232.

DeAndrea, D. C. (2014). Advancing warranting theory. *Communication Theory*, 24(2), 186–204.
Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114.
Donath, J. (2007). Signals in social supernets. *Journal of Computer-Mediated Communication*, 13(1), 231–251.
Ellison, N. B., Hancock, J. T., & Toma, C. L. (2012). Profile as promise: A framework for conceptualizing veracity in online dating self-presentations. *New Media & Society*, 14(1), 45–62.
Emerson, S. (2018, June18). Facebook Wants to Use AI to Replace Your Eyeballs in Photos Where You Blinked. *Motherboard*. Retrieved from https://motherboard.vice.com/en_us/article/a3a8n8/facebook-wants-to-use-ai-to-replace-your-eyeballs-in-photos-where-you-blinked-artificial-intelligence.
Figueredo, L., & Varnhagen, C. K. (2005). Didn't you run the spell checker? Effects of type of spelling error and use of a spell checker on perceptions of the author. *Reading Psychology*, 26(4–5), 441–458.
Gonzales, A. L., & Hancock, J. T. (2008). Identity shift in computer-mediated environments. *Media Psychology*, 11, 167–185.
Grammarly (2018). Free grammar checker - Grammarly. Retrieved from https://www.grammarly.com/.
Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint*. arXiv: 1308.0850.
Herring, S. C. (2002). Computer-mediated communication on the Internet. *Annual Review of Information Science and Technology*, 36(1), 109–168.
Herring, S.C. (2008). Language and the internet. In W. Donsbach (Ed.), *The international encyclopedia of communication* (pp. 2640–2645). doi:10.1002/9781405186407.wbiecl005
Hohenstein, J. & Jung, M. (2018). AI-supported messaging: An investigation of human-human text conversation with AI support. *CHI EA '18: Extended abstracts of the 2018 CHI conference on human factors in computing systems* (pp. LBW089:1–LBW089:6). New York: ACM.
Hohenstein, J., & Jung, M. (2019). AI as a moral crumple zone: The effects of AI-mediated communication on attribution of responsibility and perception of trust. *Computers in Human Behavior*, in press, doi: 10.1016/j.chb.2019.106190.
Jakesch, M., French, M., Ma, X., Hancock, J.T., & Naaman, M. (2019). AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. In *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, 279(1738), 2698–2704.
Levy, K., & Barocas, S. (2018). Designing against discrimination in online markets. *Berkeley Technology Law Journal*, 32(2), 1–57.
Leyvand, T., Cohen-Or, D., Dror, G., & Lischinski, D. (2008). Data-driven enhancement of facial attractiveness. In *ACM SIGGRAPH 2008 Papers, SIGGRAPH'08* (pp. 381–389). New York: ACM.
Ma, X., Hancock, J.T., Lim Mingjie, K., & Naaman, M. (2017). Self-disclosure and perceived trustworthiness of Airbnb host profiles. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW'17* (pp. 2397–2409). New York: ACM.
Margolin, D. B., & Monge, P. (2013). Conceptual retention in epistemic communities: Examining the role of social structure. In P. Moy (Ed.), *Communication and community* (pp. 1–24). New York: Hampton Press.

Mattes, K., Spezio, M., Kim, H., Todorov, A., Adolphs, R., & Alvarez, R. M. (2010). Predicting election outcomes from positive and negative trait assessments of candidate images. *Political Psychology*, 31(1), 41–58.

Pavlick, E., & Tetreault, J. (2016). An empirical analysis of formality in online communication. *Transactions of the Association for Computational Linguistics*, 4, 61–74.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347.

Razavi, S. Z., Ali, M. R., Smith, T. H., Schubert, L. K., & Hoque, M. E. (2016). The LISSA virtual human and ASD teens: An overview of initial experiments. In *International Conference on Intelligent Virtual Agents* (pp. 460–463). Cham: Springer.

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: a modern approach*. Pearson: Third Ed.

Sappington, D. E. (1991). Incentives in principal-agent relationships. *Journal of Economic Perspectives*, 5(2), 45–66.

Sen, T., Hasan, M. K., Tran, M., Levin, M., Yang, Y., & Hoque, M.E. (2018). Say CHEESE: Common Human Emotional Expression Set Encoder and its Application to Analyze Deceptive Communication. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 357–364). IEEE.

Shah, D., & Allen, K. (2019, July 30). Chinese vlogger who used filter to look younger caught in live-stream glitch. *BBC News*. Retrieved from https://www.bbc.com/news/blogs-trending-49151042.

Statt, N. (2018, May 10). Google now says controversial AI voice calling system will identify itself to humans. *The Verge*. Retrieved from https://www.theverge.com/2018/5/10/17342414/google-duplex-ai-assistant-voice-calling-identify-itself-update.

Sundar, S. S. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital Media, Youth, and Credibility. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning* (Vol. 2008, pp. 73–100). Cambridge, MA: The MIT Press.

Susser, D., Roessler, B., & Nissenbaum, H. (2018). Online Manipulation. Working paper presented at Privacy Law Scholars Conference, 2018.

Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning lip sync from audio. *ACM Transactions on Graphics (TOG)*, 36, 95.

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Niessner, M. (2016). Face2Face: Real- time face capture and reenactment of RGB videos. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2387–2395.

Thurlow, C., Lengel, L., and Tomic, A. (2004). *Computer mediated communication: Social interaction and the internet.* Thousand Oaks, CA: Sage.

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, 13, 724.

Toma, C. L. (2014). Towards conceptual convergence: An examination of interpersonal adaptation. *Communication Quarterly*, 62, 155–178.

Tong, S., & Walther, J. B. (2011). Relational maintenance and CMC. In K. B. Wright & L. M. Webb (Eds.), *Computer-Mediated Communication in Personal Relationships* (pp. 98–118). Bern, Switzerland: Peter Lang US.

Vincent, J. (2017, Apr. 24). Lyrebird claims it can recreate any voice using just one minute of sample audio. *The Verge*. Retrieved from https://www.theverge.com/2017/4/24/15406882/ai-voice-synthesis-copy-human-speech-lyrebird.

Waddell, T. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism*, 6(2), 236–255.

Walther, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. In M. L. Knapp & J. A. Daly (Eds.), *The Sage Handbook of Interpersonal Communication*. (pp. 443–479). Sage Publications: Thousand Oaks, CA.

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in. *Handbook of interpersonal communication*, 529–563.

Williams, J. (2018). Should AI always identify itself? It's more complicated than you might think. *Electronic Frontier Foundation*. Retrieved from https://www.eff.org/deeplinks/2018/05/should-ai-always-identify-itself-its-more-complicated-you-might-think.

Xu, B., Gao, G., Fussell, S. R., & Cosley, D. (2014). Improving machine translation by showing two outputs. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI'14*, 3743–3746. New York: ACM.

Zeide, E. (2018). Algorithms can be lousy fortune tellers. *Slate*. Retrieved from http://www.slate.com/articles/technology/future_tense/2015/05/crystal_app_algorithmic_fortunetelling_for_employers_and_potential_customers.html.