Testimony of Raffi Krikorian
Chief Technology Officer, Emerson Collective
Before the
Committee of Energy and Commerce
Subcommittee on Innovation, Data, and Commerce
of the
U.S. House of Representatives
"Safeguarding Data and Innovation: Building the Foundation for the Use of Artificial Intelligence"
October 18, 2023

Subcommittee Chair Bilirakis, Subcommittee Ranking Member Schakowsky, Chair Rodgers, Ranking Member Pallone, and members of the Subcommittee, my name is Raffi Krikorian, and I am the Chief Technical Officer at Emerson Collective. On behalf of the Collective, I welcome the opportunity to testify today.

I very much appreciate the Subcommittee's ongoing interest in protecting the digital privacy of Americans, as well as its efforts to address the intersection of digital privacy and artificial intelligence systems. I welcome the opportunity to discuss the American Data Privacy and Protection Act, exploring what the foundation of data privacy could look like in this new era of artificial intelligence. The Committee has successfully put a legislative stake in the ground with the ADPPA, and I am eager to explore how we can build on this good work and to discuss the challenges and opportunities surrounding data privacy and artificial intelligence.

I've been fortunate to work in the tech industry, as well as with the practical applications of artificial intelligence, for over twenty years. At Twitter, I was a Vice President of Engineering, in charge of a 500-person global engineering team, after which I was the Director in charge of the self-driving car efforts at Uber. There I led over a thousand engineers as we put the first self-driving fleet on public roads. I now have the pleasure of working at Emerson Collective, an organization that recognizes that complex societal problems require innovative solutions. We use a unique combination of philanthropy, venture investing, and art to spur measurable, lasting change in a number of disciplines, including technology. As an organization focused on ensuring that all people have access

and opportunity, we want to strive for the development of people-centric technologies that allow for an equitable deployment of artificial intelligence technologies to improve lives.

Understanding the Present

To set the frame, I would like to start with the simple fact that we live in an age of rapidly increasing digital surveillance. Very few of the users who are impacted have a clear understanding of the implicit tradeoff they are making for the conveniences they crave from the commercial applications on their phones and the web[1]. When it comes to the General Data Protection Regulation cookie consent banners — probably the most familiar implementation of user data consent agreements — there is significant evidence that a lot of users have negative feelings toward these disclaimers, and that these banners do not have a significant impact on users' decisions with regards to the website[2].

Notice and consent are failing us[3]. Not only are these technologies doing other than what users expect, they're evolving at an unprecedented speed. To move forward, we first need to step back and look at what's at the heart of the problem:

- The "data economy" is becoming incredibly complicated, and it is increasingly difficult to explain to everyday consumers how their data is being used.

[1] "Are data privacy concerns driving consumer behavior? Not yet." https://www2.deloitte.com/us/en/insights/industry/technology/protecting-consumer-data html
[2] "Has the GDPR hype affected users' reaction to cookie disclaimers?" https://academic.oup.com/cybersecurity/article/6/1/tyaa022/6046452
[3] "How "Notice and Consent" Fails to Protect Our Privacy" https://www.newamerica.org/oti/blog/how-notice-and-consent-fails-to-protect-our-privacy/

- Data minimization, or the desire for entities to reduce the collection and use of data to what is "reasonable" and "proportionate," may be in direct conflict with artificial intelligence and algorithmic desires for more data.
- There are technological trends that go beyond a regime of commercial entities capturing personal data via applications and the web, bringing us into new territory.

I want to offer an anecdote from my personal experience as an engineer and builder. When I worked at Twitter (2009-2014), we created a "Twitter button." This software allowed publishers to embed a Twitter-like experience onto their own site and app. On any given site that had a Twitter button, a user could simply click and share that link across the social network. However, what most users did not understand was that the button was also designed to track them across the web. As the Twitter button became more prevalent, the more visibility Twitter got into your browsing history – what sites you visited, what applications you used, and therefore what your interests were. Every time you visited a site that had a Twitter button installed, Twitter would receive notice that you were visiting this site. Now imagine that at the scale of Twitter, with its incredible reach and user base. Twitter could discern a lot of information about our users from this browsing history, harness it to power our algorithms, and sell access to users' interests and preferences to advertisers.

We have yet to solve how to explain these complexities to users in an upfront way. There is evidence that users already face anxiety when seeing application permission dialogs[4]. It has proven to be very challenging to determine how to obtain consent without creating a horrendous user experience, even with the best intentions that Twitter had in the aforementioned situation, during which it acted admirably, hosting a series of internal conversations tuned to make the product better and preserve users' trust.

[4] "Mobile users' information privacy concerns and the role of app permission requests" https://www.sciencedirect.com/science/article/abs/pii/S0268401218307965

Emerson Collective

Meanwhile, this is all rapidly colliding with technical possibilities and the desire to build better and more engaging experiences for users. Machine learning, artificial intelligence, and deep learning technologies pride themselves on finding patterns of which humans would never conceive. A person can find obvious correlations given a relatively small amount of data. However, computers have proven over and over that if given more and more data, they will find correlations and patterns that no humans could. Target, famously, has been able to identify newly expecting parents through the analysis of the couple's purchase history combined with other demographic information that it has obtained[5]. Retailers like Target purchase demographic information all the time – a user's age, marital status, which part of town they live in, their estimated salary – and the list goes on. And it provides retailers like Target a distinct advantage when it comes to recommending goods for purchase – providing for more sales for Target, and better experiences for their buyers.

Another example is that traditional booksellers typically only receive information about book purchases. Amazon, however, knows all the books that customer has viewed, as well as how long they dwelled on a specific page on their Kindle, as well as their searches across all of Amazon's retailers[6]. This leads to an immense advantage by Amazon when it comes to recommending the right book for readers. And, finally, this type of big data analysis is especially germane to the medical fields, where organizations such as the NHS in the UK have an inherent advantage in how they operate due to large-scale and longitudinal access to data[7], in ways that the US healthcare system does not allow. The NHS can detect and treat diseases differently because of access to all that information.

[5] "How Companies Learn Your Secrets" https://www nytimes.com/2012/02/19/magazine/shopping-habits html
[6] "Big Data: The Management Revolution" https://hbr.org/2012/10/big-data-the-management-revolution
[7] "Big data in digital healthcare: lessons learned and recommendations for general practice" https://www.nature.com/articles/s41437-020-0303-2

All these features are incredibly valuable to the data-hungry algorithms that companies are developing to create more engaging products. Drawing a clear line to how a particular data point helps an algorithm make a decision is nearly impossible, given the myriad of data that is now being fed to these systems. Companies would love both more data from their users, as well as to purchase additional data to link to their systems because it allows them to build better services for their users.

Finally, technological trends are occurring for which simply requiring users to minimize their data collection is not enough to protect them. One trend is voluntary data surrender: social media, along with the prevalence of cameras on mobile devices, has given us all new ways to connect with the world, but this new publishing mechanism has also caused an explosion of data to be put on the internet[8]. Users are voluntarily publishing information into public spaces, and are enabling tools like Clearview.AI to violate their ability to be anonymous. One may argue that there is no expectation of privacy in a public space[9], whether online or not. However, one may also argue that these 21st-century technologies are crashing into 20th-century norms.

These AI tools have downloaded imagery and other features from these public spaces to train models to identify people from surveillance cameras and be able to identify anybody anywhere. And these tools may not simply be designed to identify people, but to mimic them as well. The actor Tom Hanks has warned people that his likeness has been used to create advertising without his consent[10]. But this issue is not isolated to celebrities: This hearing alone will generate enough recorded samples of my voice so that someone could make a convincing synthetic replica of it — and not solely for entertainment purposes. Due to the rise of generative AI, all our society's identity

[8] "Social Media And The Big Data Explosion" https://www.forbes.com/sites/onmarketing/2012/06/28/social-media-and-the-big-data-explosion/
[9] "Bounds of Privacy in Public Locations -- What Is Legal?" https://www hg.org/legal-articles/bounds-of-privacy-in-public-locations-what-is-legal-35724
[10] "Tom Hanks disavows AI clone amid Hollywood's robot reckoning" https://www.latimes.com/entertainment-arts/business/story/2023-10-02/tom-hanks-pushes-back-on-ai-clone-amid-hollywoods-robot-reckoning

verification mechanisms are at risk if not updated to keep pace in this arms race. It may be entirely possible for an attacker to simulate my voice and use it to convince family members that it is myself conversing with them, asking them to do things for "me" (e.g. financial transfers)[11].

Another trend is the rapid rise of "sousveillance."[12] Surveillance, or "watching over," refers to monitoring by an authority figure, such as law enforcement. Sousveillance, or "watching from below," refers to informal networks of everyday people who capture video or other data on each other. The technology that enables this is being developed and deployed right now. Take exciting and new software services such as Rewind[13] – that software provides a search engine for one's digital life. Once installed, Rewind allows users, like me, to ask questions like "Did I place that order on Amazon?" Or, "What was that hotel I was last browsing for my upcoming vacation?". Rewind's artificial intelligence software reads any text that appears on my screen, correlates it to which applications are running on the computer, and then allows me, as the user, to ask those very questions. It augments my memory. However, it does also record every single video conference that I participate in, and does so silently in the background.

The wearable computing community has also run into similar sousveillance issues historically and specifically in the real world. For example, devices such as Google Glass did not provide visual cues to others when it was recording[14] nor the Narrative Clip[15]. More recent devices, such as the Meta Smart Glasses[16] do have a small light as an indicator to others – however, this only highlights the lack of standards and guidelines here, and design decisions

[11] "How generative AI is reshaping the identity verification landscape" https://www.helpnetsecurity.com/2023/05/22/generative-ai-identity-verification-video/
[12] "Sousveillance" https://en.wikipedia.org/wiki/Sousveillance
[13] Rewind https://www rewind.ai/
[14] "How do you know if someone's recording with Google Glass?" https://www.techradar.com/news/portable-devices/other-devices/how-do-you-know-if-someone-s-recording-with-google-glass-1163374
[15] Narrative http://getnarrative.com/
[16] "The next generation of smart glasses" https://www meta.com/smart-glasses/

around privacy being left up to individual designers and manufacturers. Others participating in the sousveillance space could run up against the same issues.

And finally, as more commercial entities deploy devices and robots into the real world, such as smart home devices or self-driving vehicles, more and more data is being captured in passive ways. While an owner may have consented to, say, Amazon's terms of use when he or she installs Alexa in his or her home[17] (or equivalently with Google and their Google Assistant, or Apple with Siri), a visitor to the home almost certainly has not. The guests are most likely unaware that the device is in use and collecting data.

Self-driving and autonomous automobiles face a similar challenge, as they are equipped with a myriad of cameras, microphones, and other sensors. As they rove the streets, they are collecting personally identifiable information from pedestrians and fellow drivers without any notion of consent. Again, the expectation may be that there is no such thing as privacy in a public space, but one could also argue that nobody expected that every car driving next to you is collecting data about you. Imagine a world where a fleet of self-driving, data-collecting vehicles were capturing every single license plate, allowing them to be tracked from where they are going to what stores they go to the addresses of other homes they are visiting.

As the digital and physical worlds collide, this increasing ability for us to self-publish, coupled with this increasing need for companies to collect data, will prove to be a problem with unforeseen consequences for American citizens.

[17] "Amazon Services Terms of Use" https://us.amazon.com/gp/help/customer/display.html?nodeId=202140280

Emerson Collective

Building Upon the ADPPA

It is in companies' interest to hoover up as much information as possible, because it has the potential to unlock a lot of innovations and find differentiated ways to engage with users. However, we need to approach this encroachment on user privacy from all angles, as there is no silver bullet, and no single effort is enough. We should consider the following requirements:

- Increased efforts to promote and expand digital literacy, as well as continued pushes on the design patterns needed to transparently explain to users, up front, what they are consenting to.
- Allowing people to access the full life cycle of their user data - from creation to usage to sales and swaps to deletion.
- Offering mechanisms to still engage with applications without data collection activated, albeit perhaps in a limited way.

Data literacy is a fundamental part of addressing this privacy dilemma[18][19]. In addition, there needs to be increased incentives for application developers to push the boundaries on explaining to users, up front, what they are consenting to and how their data will be used. Historically, there have been prize challenges offered by private organizations[20], as well as organizations such as NIST[21] and the White House[22], to spur these designs – and better

[18] "The Digital Literacy Imperative" https://www.csis.org/analysis/digital-literacy-imperative
[19] "Privacy Literacy Training" https://dataprivacyproject.org/initiatives/privacy-literacy-training/
[20] "Privacy by Design Awards" https://cybercx.com.au/privacybydesignawards/
[21] "PETs Prize Challenge: Advancing Privacy-Preserving Federated Learning" https://www.drivendata.org/competitions/98/nist-federated-learning-1/
[22] "U.S. and U.K. Launch Innovation Prize Challenges in Privacy-Enhancing Technologies to Tackle Financial Crime and Public Health Emergencies" https://www.whitehouse.gov/ostp/news-updates/2022/07/20/u-s-and-u-k-launch-innovation-prize-challenges-in-privacy-enhancing-technologies-to-tackle-financial-crime-and-public-health-emergencies

designs, better products, and better infrastructure is possible. Privacy-first products, such as Signal[23], have started to appear and become popularized. More exploration of this space should be facilitated and designers, organizations, and companies should be encouraged to explore here.

After the initial consent process, users should then have the ability to see how their data is flowing through these systems long after they've closed an app or a service, and retain agency over it. Users should be given the tools to understand and navigate which parts of a company's business model relies on their data, as well as which specific features have incorporated it. They should be given clear ways to understand the tradeoff of the data they have consented to be collected, and how it benefits — or harms — them or their community. This should not stop at a single company or application, but needs to be between companies and applications through data sales and swaps. If desired, the user should be able to both revoke consent and delete their data from the application. And users also need to be able to understand the use of their data that is acquired from other means – whether it is being scraped from public spaces, or acquired through commercial vendors or data swaps.

Above all, users should be given enough information to be able to make a reasonable choice. Companies such as Meta and Google already provide some transparency into what user data they have collected. On Facebook, a user can see the number of times that his or her face has been used to train a recognition algorithm[24], while on Google, a user can see all the location tracking history Google knows about him or her[25]. In addition, both Facebook[26] and Google[27] provide mechanisms to delete a user's data. These efforts are certainly a start toward what a user

[23] Signal https://signal.org/
[24] "What to Look for in Your Facebook Data—and How to Find It" https://www.wired.com/story/download-facebook-data-how-to-read/
[25] Google Maps Timeline https://maps.google.com/locationhistory
[26] "Deleting Facebook? Follow These Steps Carefully" https://www.cnet.com/tech/services-and-software/deleting-facebook-follow-these-steps-carefully/
[27] "How to See What Data Google Has on You (and Delete It)" https://www howtogeek.com/709263/how-to-see-what-data-google-has-on-you-and-delete-it/

interface could look like. These are the beginnings of user interface models that companies can adopt in order to provide transparency to end users.

Finally, for those who want to opt out of these trades of data for convenience altogether, there should always be the opportunity to use an app without personalization, even if that means that financial costs to the user increase. There are sometimes no reasonable options for users; putting that burden on application providers is a reasonable tradeoff, as they sometimes already have frameworks to allow users to opt out of advertising (not necessarily data collection) by simply paying more for subscription services.

Conclusion

I sincerely praise the work of this committee in the authoring of the American Data Privacy and Protection Act, but I submit that this should be treated as the foundation for more work going forward. These technologies are all racing ahead — from the machine learning and artificial intelligence algorithms to the unique ways to gather data to the product designers dreaming up different ways to engage people with their apps — which is good for innovation and the economy, but they are colliding with user privacy in potentially irrevocably dangerous ways.

The problems that we can identify today are just that: the problems of today. There will almost certainly be new issues to tackle as technology continues to evolve. Setting up legislation that can be quickly adapted and updated as issues appear in specific areas or in different communities is vital. It is incredibly important that the significant and impactful work of this committee, and that of the ADPPA, is the “floor” that the rest of our society should be building upon, and not necessarily the “ceiling.” Because, as we all know, this technology is changing every day.

Emerson Collective

To summarize:

- There are no silver bullets: We need to invest in digital literacy and user education, improve the clarity of the consent process; create mechanisms for users to understand and participate in the "data economy" in which their data is being gathered, used, and traded; and provide them with ways to opt out of data collection (including giving users options within applications).
- We need to be mindful of the impact of emerging technology trends, including the data hunger of artificial intelligence, the explicit desire for users to publicly share their data, and the increase of surveillance or sousveillance mechanisms.
- We should treat the ADPPA as the beginning and not the end, explicitly encouraging and inviting others to innovate on what protections and guidelines are needed for their communities.

I thank you for this opportunity to share the perspective of Emerson Collective. We look forward to continuing to work with you to address these issues with the attention and urgency they deserve.