Written Public Comments Submitted to PCAST

May 22nd, 2023 to July 20th, 2023

As specified in the Federal Register Notice, because PCAST operates under the Federal Advisory Committee Act (FACA), all public comments and/or presentations will be treated as public documents and will be made available for public inspection, including being posted on the PCAST website.

Cyber-Physical Resilience

Written: 5/31/2023

Email: [redacted]

Mr. Horvitz, Mr. Venables, and PCAST Working Group,

This is Ben Ruddell, a Professor at Northern Arizona University and lead on the FEWSION academic supply chain data science program. I saw the PCAST working group press release on May 15th, and I am responding to provide information that your working group might find useful. Cyber-Physical Resilience requires, among other things, large datasets mapping the supply chain and critical infrastructure, with an emphasis on the cascading interdependencies between those networks. FEWSION is a world-class program to build and analyze these datasets and the leading R&D investment by the U.S. Department of Defense in the creation of a national capability serving this purpose. Dr. Evelyn Dahm is our FEWSION program manager with DoD and can be consulted on USG-side questions. We believe that the existing FEWSION capability is important for national security, resilience, supply chains, and policies beyond DoD applications, so we provide you here with a summary of the capability for your consideration. I have also made other notes at the bottom of this message on other capabilities that I think you should be aware of addressing your specific requests for information.

Good luck with this important effort, and please let me know if I can be of further assistance.

Ben

--

The FEWSION Supply Chain and Critical Infrastructure Data Science Program

The FEWSION program provides U.S. communities with real-life data to map the vulnerabilities and environmental footprints of supply chains as well as their resilience. FEWSION™ has built the first complete empirical description of the U.S. supply chains so that every citizen and policymaker in the U.S. can see where their commodities come from. FEWSION was founded in 2016 by a grant from the National Science Foundation and U.S. Department of Agriculture under the Innovations in Food Energy and Water Systems program (INFEWS). The FEWSION website is: >https://fewsion.us<

The FEW-View Platform

FEW-View is the publicly available supply chain and critical infrastructure visualization tool that has operated since 2019 as the primary public service of the FEWSION program. FEW-View is the world's first complete map and visualization system for supply chains and has been used to support each FEMA/DHS national preparedness exercise since 2019, along with response to hurricanes, the trade war with China, preparation for DoD logistics support to Ukraine, and (currently) planning for a conflict in east Asia. Check out FEW-View on the FEWSION website; it's free for your use, easy to use, and is offered to all US communities to empower their resilience. The FEWSION version 2 dataset is also available to researchers interested in mapping the nation's supply chains and critical infrastructures.

Department of Defense requirements and funding for FEWSION

In 2022 the FEWSION program was funded at $8 Million per year by U.S. Special Operations Command's Science and Technology program to develop a second-generation capability that is integrated with DoD's data systems and that proves the capability to provide data for supply chains and critical infrastructures for any given area of responsibility anywhere in the world from the "last-mile" level (facilities, addresses, power lines, pipelines) up to international networks. This is an unclassified program that we are able to discuss with PCAST. The second-generation FEWSION program's capabilities are not discussed on the project website for security and intellectual property reasons.

PCAST-specific requests for information

Recovery and survivability in the face of attacks and events. See points below. Survivability and recovery are part of the overall equation for critical infrastructure continuity and resilience. What should be emphasized for PCAST is that we need to build "all-hazards resilience" using methods that do not require the accurate prediction or characterization of a threat. This is the main weakness of most existing methods. Black Swans are real and unavoidable, and intelligent attackers will specifically circumvent defenses, so the focus of PCAST should be on recovery and survival from events and attacks from which we are not adequately protected. This is the main lesson we must learn from 9-11 and from the current cybersecurity crisis.

Approaches to assure continuity of operations in degraded states. I have authored a publication on "Continuity of the Economy" that is intended for the national emergency management curriculum and addresses a practical approach to achieving continuity of operations for critical infrastructures. The emphasis is placed on a whole-supply-chain approach to preparation that accelerates recovery, enables uncoordinated recovery actions by each supplier in the network, and ensures that adequate buffers and in place (e.g., inventory) to support supply chain recovery for critical functions. Reach out to me if you'd like a pre-publication copy, please.

Mechanisms to measure and assess modularity and limitations of scope or costliness of failures. Ted Lewis, Executive Director (retired) of the Naval Postgraduate School's Center for Homeland Defense and Security (CHDS), developed a branch of network mathematics specifically for the limitation of the scope and costliness of cascading failures on critical infrastructure networks. https://www.chds.us/c/chds-executive-director-ted-lewis-wins-literati-network-award/ This new branch of network theory and applications has been commercialized by Criticality Science Inc. of Alexandria VA, a small business headed by 9-11 Commission member Susan Ginsburg. As far as I am aware this capability is currently the state of the art for limiting scope and costliness of cascading failures on critical infrastructure networks. >https://www.critsci.com/<

Incentives to balance efficiency which can reduce resilience vs. the investment needed to maintain sufficient resilience. The financial incentives are already in place for balancing resilience against efficiency- but what is missing is authoritative and accurate methods for translating investments in resilience into reductions in risk- meaning reduction in dollars at risk. If financial risk is associated with resilience, the reduction in that risk can be directly included in financing, in bonding applications, in engineering optimization of cost, and in insurance rates. The capability developed by Criticality Sciences Inc. (see point above) has already demonstrated the ability to measure resilience benefits in terms of reduced dollar risk and should be considered as a tool for understanding the financial incentives that already exist (or lack thereof, as the case may be).

Out-of-band or systems-independent means of assuring physical control in the event of digital failures. No comment at this time.

Methodologies and standards to encourage resilient systems design and adoption. I believe that a role of the US Federal Government is to define and enforce minimum resilience standards for national critical functions. This implies the need for adoption of a set of science-based resilience measurement methods, and for the establishment of minimum performance standards for critical functions and infrastructures (including critical supply chains). There are several candidate methods and standards that PCAST should consider; the two best methods I am aware of are fairly new and are explained below.

1. The Lewis Score. Ted Lewis (read comment above) developed a benchmark that measures the tipping point in a critical infrastructure network, separating a network that is fragile and prone to "blow up" via cascading failure from a network that is resilience and tends to naturally suppress and contain failures keeping them small and localized. This is an easily understood, mathematically robust, binary metric for whether a critical infrastructure is "resilient enough". I believe that the measurement of and achievement of a minimally resilient Lewis Score should be required for all critical infrastructure networks and critical supply chain networks in the U.S. NIST is currently considering the Lewis Score for network resilience measurement.

2. Resilience to shocks using an intensity-frequency-duration (IDF) model. The FEWSION project's main scientific contribution to date is the publication of an intensity-duration-frequency model that established for the first time the probability of shocks impacting a community's supply chains and critical functions. The model is surprisingly simple and accurate, and it demonstrates that the resilience of a supply network to shocks due to any hazard or cause is a simple positive function of the supply network's diversity, measured as the Shannon Diversity. This model is a real breakthrough because it allows a city or country to establish policies that engineer and optimize a specifically controlled level of supply chain shock risk. I believe this model should be used for U.S. critical function and supply chain policy going forward. The model was published in *Nature* in 2021 for the nation's food supply chains, and an in-review paper demonstrates that the same model holds for all kinds of supply chains and critical infrastructures. The *Nature* article, Gomez et al. 2021, is attached for your convenience.

--

Benjamin L. Ruddell, Ph.D., P.E.

Professor, School of Informatics Computing and Cyber Systems (SICCS)

Director, FEWSION

Northern Arizona University, Flagstaff, AZ, 86011, USA

Email and Calendar [redacted]

Office Phone [redacted]

CSIL Lab https://csil.rc.nau.edu/

Time Zone MST/Arizona (UTC-7), which is the same as PDT in summer

Attachment:

Gomez_etal_2021.pdf

Cyber-Physical Resilience

Written: 6/14/2023

Email: [redacted]

The following actresses and actors:

Emma Watson, James Franco, Ezra Miller, Seth Rogan, Dave Franco, Danny McBride, Daniel Radcliffe, Zac Efron, Laurence Fishburne, Ben Affleck, George Clooney, Brad Pitt, Hugh Jackman, Bradley Cooper, Jesse Eisenberg, Edward Norton, Billy Bob Thornton, Paul Giamatti, Dwayne Johnson, Clark Duke, Lewis CK, Natalie Portman, Susan Sarandon, Woody Harrelson, Tom Hanks, Tom Hardy, Anthony Mackie, Tom Holland, Robert Downey jr., Scarlet Johansson, Chris Evans, Emma Thompson, Emma Roberts, Emma Stone, Kristen Stewart, Kristen Bell, Kristen Wiig, Elliot Page, Caitriona Balfe, Cillian Murphy, Mark Strong, Rachel McAdams, Zooey Deschanel, Zendaya, Sarah Jessica Parker, Martha Nussbaum (philosopher, University of Chicago), John Searle (philosopher, UC Berkeley)

need to be informed that there is a group (or groups) out there who have gotten ahold of new neurotechnologies and are using the voice and image (accompanied by sensations) of various actors and actresses against me (attacking me, tormenting me, molesting and raping me, and so on); either that, or these actresses and actors are directly involved in these abuses of neurotechnology (see end of message for some credible resources to get started regarding neurotechnology, neuroethics, neurorights, and the abuses of neurotechnology)… this is being done remotely. See: adamchristiannielsen.com for my work (poetry)...this might be its only chance at life. There is an autobiography in the letter attached, or a briefer one on my website.

Please pass along the following messages to the appropriate parties (all of this is newsworthy; a story that needs to be told, an issue which needs to be addressed):

1: An Urgent Appeal; Activism Against the Abuses of Neurotechnology
2: The Torture and Rape of Adam Christian Nielsen, Remotely, by New Neurotechnologies
3: The Coercion of Adam Christian Nielsen into A Relationship with Emma Watson (or imposter)

4: Counts of what was done to me during an eight week span of time
5: Credible Resources Regarding neurotechnology, neuroethics, neurorights, and the abuses of neurotechnology

1:
Activism against the abuses of neurotechnology

This is an appeal, an urgent plea for your immediate help... please, they are tormenting me even as I send this. I am being held captive, tormented, abused, tortured, and raped, remotely, by new neurotechnology... others are being targeted as well. This issue concerns us all, and needs to be addressed and resolved before it goes any further. We, as targeted individuals, need help your help in dealing with the abuses of these new neurotechnologies… we are scattered, disorganized, and often powerless and/or terrorized to the point of being virtually incapable of doing anything about it, and need credibility, especially with the media (which can be hard to come by insofar as it's easy to pass this off as schizophrenia or other mental health related disorders). Really throw your weight into it, now, or wait around until it's happening to you or a loved one I guess... there are people out there who are doing just about whatever they want to others with new neurotechnology. So let's try to coordinate, because I've been tired of this for a while now… here's what needs to be done:

-contact news media and get the word out (you can use my or other testimonies) (maybe use your own contacts, connections, and colleagues)

-adamchristiannielsen.com for my work (poetry)

-develop counter measures against the abuses of neurotechnology (ex. means of identifying or detecting its occurrence)

-contact your local elected representative about the abuses of neurotechnology
-enact laws against the abuses of neurotechnology
-equip law enforcement to deal with the abuses of neurotechnology
-investigate the situation in general as well as specific cases
-justice for targeted individuals
-keywords/phrases: neuroscience, neurotechnology, neurorights, neuroethics, etc... still working out the language here

Let's start there… I may have forgotten a few things, but it'll get us going. You can read the rest below, the point is that this is happening and we need to rally now.

2:

Adam Christian Nielsen

My name is Adam Christian Nielsen. I am a poet-philosopher from the northern California valley, and am fighting for my life and the lives of many others. I am a targeted individual of surveillance, torture, rape, voice-to-skull, neural monitoring (and so on) by new neurotechnologies (calling this, tentatively, the abuses of neurotechnology); this is being done remotely, it affects your body directly through your brain from however many miles away… human rights violations, civil rights violations, and crimes against humanity are taking place against myself and many others all over the world by new neurotechnologies. This started, in my case, in Fall 2016, with voice-to-skull and remote neural monitoring. By 2018 mind-games were common. In 2019 I was unable to complete a Masters degree in the Humanities through California State University, Northridge (which I was attaining so as to teach) after a year into my studies due to interference by these neurotechnologies. And 2020 - 2021, I was being verbally, emotionally, physically, and sexually abused, as well as harassed, molested, raped, and tortured on a daily basis. By February 2022 I was forced into mental health facilities under a false diagnosis, which I was in for ten months, until December 2022, in their effort to conceal what's going on. 2023 has so far been mostly just remote neural monitoring (probably for the purposes of developing this technology, as well as surveillance), voice-to-skull, mind games, image induction, and damage control… I have by now become accustomed to someone's being there in mind all the time. I suspect this is being done by corporations, universities, government agencies, and/or independent groups. And I have been an activist against the abuses of neurotechnology since 2020, when I first realized what was happening to me.

This, the truth, the whole truth, and nothing but the truth… I'll let you decide what to do with it.
I am speaking out on behalf of everyone who has been targeted, and on behalf of everyone who will be targeted if this issue goes unchecked. I am trying to raise awareness about the issue, reach out about my situation, give as accurate of an account of my experience of torment, torture, and rape via neurotechnology as possible, and get this matter resolved.
See: adamchristiannielsen.com for poetry I've written over the years... also, you can see below for more information about the abuses by way of neurotechnology, and other testimonies from other people of these things happening to them (especially the comments section at cyber-torture.com).
When Mark Zuckerburg says that Facebook will be able to be operated 'telepathically' in the future, or you hear talk of the 'brain-computer interface' or 'neural monitoring', these are instances of the kind of technology we are dealing with here. This stuff is new, it is being developed now, so it may be unfamiliar, and may come as a surprise… it was and did to me. Though I'm not exactly sure who all the players are yet, they no doubt have access to technology of this sort. In all likelihood there are multiple players involved (ex. technology companies linked up with entities of government linked up with private interests).
Manipulation of other persons without their knowing is fairly easy with this technology, I witnessed their doing this firsthand with my family and with medical staff; anyone can be targeted at this point for any reason, and any number of things can be done to them. You can do some seriously messed up things with this, especially if people don't know it's going on. Below, I have included a list of things they can do to you with this technology, or rather, some of the things they have done to me (and thus, are capable of doing to other people).
So there are two things going on, one positive and one negative. Negatively, they can see and hear everything that is going on in your mind (can "read" your mind) and can see and hear everything you see and hear (complete lack of privacy), positively, they can do whatever they want to you, including inducing images, feelings, ideas, and thoughts directly into your brain, it just goes to your body through your brain rather than to your brain through the rest of your body. They of course try to keep their crimes concealed, and so, for the most part, don't affect a very noticeable change in things unless it is in an extreme case like mine... but in extreme cases like mine they'll try to hide it by, say, calling my sanity into question, or trying to silence me. Sometimes there is no discernible sign, not even in mind, they are there just reading it and are watching what you do; invading the most intimate of your privacies all the while. Said again, this is being done remotely, and so is even more difficult to detect. I'm still considering the implications of this technology.
This will take the concerted effort of us all, but especially of law makers, government agencies, and news media at least; ambitious detectives and reporters and so on… so contact your local elected representative, call up the private investigators and investigative reporters, grill a couple of technology companies about the implications of their technology (and how it's being developed), and of course be informed about what is happening here, and be aware of the signs of its happening to you or others in your life if it does. But please, do it promptly, before I have accrued any long-term psychological damage, and before this issue goes any further than it's already gone. I am not alone in this… it concerns all of us. Do your part.
Laws will probably need to be made, legal action will probably need to be taken, law enforcement will probably need to be equipped to deal with its occurrences, but justice be done… even if this isn't illegal yet (on grounds that it's unprecedented of course), it is just wrong… this is happening to individuals who don't consent to it and ruinous things are being

done; it goes against literally everything the U.S. stands for. This should be made public knowledge, and the people responsible need to held accountable. This technology is here to stay… it needs to be kept in check.

3:

The Seduction of Adam Christian Nielsen by Emma Watson (or imposter)

Add 'seduction' to the list of things they can do: I have fallen in love with Emma Watson (the actress)… she was presenced to me through this neurotechnology, and I proposed to be wed to her, she (they?) said yes, and I've been under the impression that we've been engaged since (2020). This is a confusing situation for me, so bear with me a moment if you would… Emma Watson has either been directly involved, or her image (and voice, a presencing of sorts) have been used against me; I'm not sure which, and aim to find out; it could be both.

I have attached a letter explaining all of this, a letter to Emma from myself, you can take a look if you're interested… there is a fairly detailed autobiography there you can check out for more information on my person. I could use your help informing her that this is happening; maybe getting us in contact with each other… she really should know what's going on. I'm sending you the letter with hopes it makes its way to her; she can decide for herself on me… just throwing it out there and am going to see what happens. It's truly a messed up thing though, messing with people's hearts.

Other people this situation of 'direct involvement and/or presencings being used against me' has happened with in my case so far include: primary suspects: Martha Nussbaum, Emma Watson, James Franco, Ezra Miller, and Jay Dodd... other possible suspects (having a brief presence at least throughout my being abused via neurotechnology) include: Seth Rogan, Dave Franco, Danny McBride, Daniel Radcliffe, Zac Efron, Laurence Fishburne, Ben Affleck, George Clooney, Brad Pitt, Hugh Jackman, Bradley Cooper, Jesse Eisenberg, Edward Norton, Billy Bob Thornton, Paul Giamatti, Dwayne Johnson, Clark Duke, Lewis CK, Natalie Portman, Susan Sarandon, Woody Harrelson, Tom Hanks, Tom Hardy, Anthony Mackie, Tom Holland, Robert Downey jr., Scarlet Johansson, Chris Evans, Emma Thompson, Emma Roberts, Emma Stone, Kristen Stewart, Kristen Bell, Kristen Wiig, Elliot Page, Caitriona Balfe, Cillian Murphy, Mark Strong, Rachel McAdams, Zooey Deschanel, Zendaya, Sarah Jessica Parker, John Searle, and whoever they are working with... Martha Nussbaum is a philosopher from the University of Chicago, John Searle was a philosopher at the University of California, Berkeley, the rest are actors or actresses.

Again, if these people are themselves not in fact involved, then their images (including voice and/or a physical presence of sorts) have been used against me throughout my being tortured and raped via neurotechnology in various ways at various times; if these persons have been involved, then the extent and kind of their involvement will, I hope, be worked out in detail in time (including any others involved not currently listed). All of the suspects on the list should be notified, and I could use some help getting word around.

I do not currently know the reasons they are doing this to me, and am still trying to understand it all; still am trying to find the right terms; still researching. You can contact me by the contact form on my website (adamchristiannielsen.com) if you have any questions, comments, concerns, etc., or have any ideas on a course of action which could be taken here. Again, anyone can be targeted and/or manipulated by neurotechnology anywhere at any time for any reason by now... be wary.

This is the best that I can presently do given the circumstances. Please, forward this to anyone who might be interested, concerned, or able to address it (e.g. family and friends, human rights groups, civil rights groups, law enforcement and other government agencies, elected representatives, news media, private investigators, coworkers, religious communities, scientists, poets, philosophers, and so on)… inform others that this is happening.

Means they have used and might continue to use to conceal it (this is not an exhaustive list): inconspicuousness and embeddedness in everyday life; routinedness; distraction; trying to get you to not think about it, or think about other things; memory erasure; questioning of mental health; defamation, slander, and discreditation; their putting up of a professional front; tampering with medical records and other documents; silencing of the opposition; relationship sabotage; human hijacking (of family, friends, medical professionals, law enforcement and other government personnel, so as to further the efforts at concealment).

4:
counts of what was done to me within an eight week span of time

Weeks spanned: 8 [1/27/2022 – updated and explicating over time, this will do for now][it will be indicated if it was not within those weeks preceding][this is not an entirely exhaustive list of what has happened to me]

forced vomiting ~x30
>20 times in half a night (February 2022)
choking me with my own hand ~x5
forced erection ~x90
forced unerection ~20
rape ~x170 (ex. forced orgasm)
ex: “I’m going to rape you over and over again in front of her” (they said)
seducing me into an engagement with Emma Watson x1 (May 2020)
stimulation of genitalia ~x100
molestation ~x170
tortured in some way or another (ex: sharp pains) ~x1000 or more
memory erasures and other mental interferences ~x1000 or more
staging a scene in public ~x10
head slammed to the ground (fortunately there was a pillow) ~x15
being crucified ~x15
being crucified while being raped ~x3
forced collapse ~x120
sore throat ~x15
sprained ankle, fractured heel ~x1
forced defecations / attempted herniations ~x25
forced defecation with forced vomiting at the same time x1
forced constipation ~x20
forced sleep ~x50
dizziness / weariness / drowsiness / fatigue ~x200
limbs forced to side (sometimes semblance sexual bondage) ~x35
fractured ribs (fall 2020) ~x2
utter agony ~15
heart convulsions / redirections of blood-flow ~x65
shortness of breath ~x100
disruption of circadian rhythm ~x50
sleep deprivation ~x2 (once 7 days, another 5 days)
mucus formation ~x75
clogged nasal passage ~x75
depreciation / disparagement / slander of my person ~x350
intellectual property theft
relationship and employment sabotage
control of bowels, bladder, brain waves, emotions, genitalia, heart rate and rhythm, sleep cycles,
ability to cause sensations throughout the body

ability to manipulate the light left in the eyes after you have closed them
tampering with accounts
mind reading (and neural monitoring)
mind games
threatening me and my loved ones
locking me up in behavioral health facilities for 10 months in their effort to cover up the abuses via a false diagnosis (February 2022 - December 2022)
etc…

5:

Resources

The Neurorights Foundation

Home (neuroethicssociety.org)

targeted evidence - Home

Information on Psychotronics

Frontiers | Human Brain/Cloud Interface (frontiersin.org)

Human Brain/Cloud Interface - PMC (nih.gov)

US investigating possible mysterious directed energy attack near White House | CNN Politics

U.S. probing suspected directed-energy attack on government personnel in Miami - POLITICO

US investigating possible ‘Havana syndrome’ attack near White House: CNN | The Hill

As mystery over 'Havana Syndrome' lingers, a new concern emerges (nbcnews.com)

Targeted Individuals: Now that we know it’s real, Will someone finally do something? – VT | Alternative Foreign Policy Media (veteranstoday.com)

Cyber-Torture – EU-Coallition Against Cybertorture

-see comments for more testimonies

-I have heard this also called "cybertorture"

UN warns of rise of 'cybertorture' to bypass physical ban | Torture | The Guardian

(PDF) Towards new human rights in the age of neuroscience and neurotechnology (researchgate.net)

(PDF) Cognitive liberty. A first step towards a human neuro-rights declaration | Paolo Sommaggio, Marco Mazzocca, Alessio Gerola, and Fulvio Ferro - Academia.edu

Patents (targetedmassachusetts.org)

V2K - Targeted Individuals 101 (google.com)

The Rise of Neurotechnology Calls for a Parallel Focus on Neurorights - Scientific American

TARGETED JUSTICE - Targeted Justice for Targeted Individuals

Mark Zuckerberg says Facebook of the future will be powered by telepathic thoughts | The Independent | The Independent

Patents for Mind Control Technology – Fighting Monarch

-"mind control" is simply a pop-culture sci-fi phrase for roughly the same thing; check out the patents and ignore the rest

Cyber-Physical Resilience

Written: 7/19/2023

Email: [redacted]

Apologies for the slow reply to this request but your email came to me late via several friends in DoD who only recently learned what we are doing and felt what we were doing might be highly relevant to your efforts.

Allow me to introduce myself and our company – AP Cyber – and to a solution we will be piloting through a contract with the Virginia Innovative Partnership Corporation and with the financial support and oversight of DHS.

My personal history includes over 30 years of working with the Government in a variety of capacities and through multiple companies – much of which was involved in supplying solutions to DoD and the IC for National Security. Several years ago, I started a company with Alan Wade, former CIO at CIA and ODNI with the goal of building a secure networking solution designed specifically for Zero Trust deployments to deliver "national security grade" protection to Critical Infrastructure. Our mission was to create a platform that was as secure as SCIF-based networking but easier to deploy, with less overhead and greater flexibility. That company is currently supplying solutions to various DoD customers and to the private sector and the platform will have an ATO issued in the next several weeks. I spent much of my time working with NIST, DHS, DoD (SECDEF) and MITRE toward helping to architect and configure Zero Trust solutions using that company's technology.

A year ago, I shifted my efforts toward driving Zero Trust. After seeing how expensive and time-consuming the process of testing, purchasing and implementing would be, it was clear to me that our national critical infrastructure was incredibly vulnerable. So I decided to expand my work to build a **Zero Trust Managed Service**, that would be compliant with/mapping to the ZTAs from CISA, DoD, NIST, MITRE and others leading the transition.

The goal was to offer enterprises the ability/option to "lift and shift" into a pre-configured and pre-integrated ZT environment quickly and less expensively than rebuilding/building from scratch. My approach followed along the lines of the work DISA was doing in their Thunderdome program, which focuses on 5 of the 7 pillars of a ZTA and allows customers to bring their applications and data since those are unique to each enterprise.

AT that same time I was approached by Available Power (AP) who was seeking a Zero Trust solution to protect energy assets and we formed a company called **AP Cyber**. Available Power is a developer of large-scale battery systems designed to provide resilience to the national grid. They have multiple projects with ERCOT and with the national cell tower providers. The platform we created, ZTMS,. was designed to protect those battery assets and ensure their availability and ensure they would never be subjected to cyber and cyber-physical attacks.

Our resulting solution set combined with AP's unique ability to deliver it using the existing cell tower infrastructure has enabled us to secure a pilot contract with the State of Virginia through the Virginia Innovation Partnership Corporation and with the support and participation of DHS S&T.

Our solution will be tested and deployed in the corridor between Richmond and Washington DC and the initial applications will include a drone detection system and wildfire and flood sensor systems. However, the platform is agnostic and can be used to operate and protect critical infrastructure and almost any cyber-physical system (CPS) and users. Once proven through this pilot, AP Cyber, though our unique contract relationships, can quickly scale our solution nationally to provide a safe haven for all kinds of CPS including power, water, transportation, and others identified in the DHS/CISA sectors.

In summary, AP Cyber will deliver a functional, Zero Trust environment to VIPC and DHS that offers highly secured communications delivered as a mesh network with integrated AI/ML-based continuous monitoring, Identity Management, Asset Management, endpoint protection and management from the cloud with on-premises backup. It will be capable of detecting and responding via humans, automated policies or autonomously to cyber threats and mitigating them in real time.

Our communications component is completely encrypted point-to-point communications with the ability to incorporate NIST approved quantum-resistant encryption technologies quickly. By employing the existing tower infrastructure through our unique contractual relationship with the two largest tower owners in the U.S., we have an opportunity to scale our solution nationally and allow it to be used independently and securely by public and private customers – each having their own ZT environment to run apps, assets, etc.. It should be noted that our solution does not expose any IP addresses or assets to the Internet and our ZTMS solution is agnostic and will work with existing applications and tools.

We would welcome the opportunity to discuss how we can support your efforts and to show you the work we have done to date. We believe that we can provide protection, resilience, and speed toward the protection of critical infrastructure without high costs up front – or long term.

Warm regards,

Glen Gulyas

President

AP Cyber

[redacted]

Cyber-Physical Resilience

Written: 7/20/2023

Email: [redacted]

Here is the feedback from the ISSA.org side on this request. I'd be happy to continue the consultation from a practitioner standpoint if need be.

Francesco Chiarini

Founder, ISSA.org Cyber Resilience Special Interest Group

List of recommended resources

- Recovery and survivability in the face of attacks and events.
  - Cyber Survivability https://www.mitre.org/sites/default/files/2022-09/pr-19-02172-10-cyber-resiliency-constructs-cyber-survivability.pdf
- Approaches to assure continuity of operations in degraded states.
  - Damage-Limiting Operations https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-172.pdf
  - Cyber Courses of Actions >https://apps.dtic.mil/sti/trecms/pdf/AD1107798.pdf< ; https://www.mitre.org/sites/default/files/publications/pr-15-1334-cyber-resiliency-engineering-aid-framework-update.pdf
- Mechanisms to measure and assess modularity and limitations of scope or costliness of failures.
  - Assess loss of functions >https://www.researchgate.net/publication/327009274_Implementing_Cyber_Res

ilient_Designs_through_Graph_Analytics_Assisted_Model_Based_Systems_Engineering<

- Incentives to balance efficiency which can reduce resilience vs. the investment needed to maintain sufficient resilience.
  - Nothing immediately available
- Out-of-band or systems-independent means of assuring physical control in the event of digital failures.
  - Nothing immediately available
- Methodologies and standards to encourage resilient systems design and adoption.
  - NIST 800-160 and MITRE CREF https://csrc.nist.gov/publications/detail/sp/800-160/vol-2-rev-1/final https://csrc.nist.gov/publications/detail/sp/800-160/vol-1/archive/2018-03-21
  - High Value Target >https://www.highvaluetarget.org/<

Think big, start small, act now. {Robin Sharma}

Francesco Chiarini – SABSA, TOGAF, CGEIT, ISO22301, EnCase, CEH

Head, Resilience & Architecture Risk Strategy

Chief Information Security Risk Office (CISRO)

Email [redacted]

Generative AI

Written: 5/22/2023

Email: [redacted]

Generative AI seems like a technological cover for old fashioned plagiarism: it thinly blends a few sources without reference, and the design of the technology makes it impossible to prove that it is plagiarism because even the creators don’t understand the internal representation. But if a human did it, we would just call it plaglarism. --

Carl Gold PhD

Data Scientist

Link|Book|Stream

Tube|Tweets|Blog

Generative AI

Written: 5/22/2023

Email: [redacted]

Regarding PCAST Generative AI questions 1-4, I wanted to contribute the following ideas.

The predecessor to the internet was designed initially to provide access in times of crisis to machines and users who were assumed to be authenticated. This is similar to a telephone from that era, where you did not need to unlock the phone to make a call. Physical access to the room with the phone was enough. This design may need to be reconsidered.

Research into novel methods of authentication of users on a public, unclassified, U.S. government network (or virtual network) might be useful when machine-generated content becomes indistinguishable from human-generated content.

Ideally, content that was government generated would be digitally signed, and robust signature verification services could be part of a future network service or function, so that content if downloaded could be verified at a later date (such as when it is republished) to check its authenticity.

How to do this verificated at scale, and in a tamper-proof manner, is an open research problem as there would be a high cost to implementing the naive solution of deploying some kind of PKI for every U.S. person, including all the associated key management infrastructure. OMB will have information on the historical costs of DoD or other PKI programs.

A potentially fruitful area of research would be into applications of secure multiparty computation or zero knowledge proofs that might allow for a level of mutual authentication or agreement of identity engaged in a transaction such as the downloading of content from a government service, without the cost or complexity of traditional PKI.

If signatures of authenticated content could be verified in a distributed manner, robust to a variety of cyber effects (such as replay, spoofing, denial of service), and that verification could happen efficiently, functionality may be able to be built into the devices on which the public use to consume content.

The argument above is made assuming that generative content will eventually become indistinguishable from human generated text, speech, and images at common resolutions found in broadcast media. There may be thresholds at which generative methods could not keep up in real time, based on current and forecasted processing power (e.g., generating 50mp RAW images at broadcast framerates) but network bandwidth would then become a problem. That scaling problem seems to indicate that increasing file size (image resolution, sampling rate of speech, etc.) would not be viable alone.

The argument above is also network-centric, assuming a traditional notion of what a "file" is in modern Operating Systems. Research into authenticated file systems may also be of interest to address the challenges posed by generative AI.

For example, when a file is created in a POSIX file system, username information is currently associated with the file object, and can be changed.

Can novel file systems be designed that are performant and scalable, but also provide a form of authentication of their content and a history of changes to their metadata?

For example:

1. Information from TPM and a unique identifier associated to a user could be combined with the hash of the file contents of a file, only adding a few words of memory to the size of a file.
2. Many of the techniques originally developed for remote attestation and trusted computing might be adaptable to this problem of authenticating content so the device and user of the device can be associated to the data.

This type of technology could have benefits for reliability of government communications, provenance of data associated to scientific simulation or experiments (benefiting open science and public health), and also potentially help more quickly trace unauthorized disclosures of data to their source.

A balance would need to be struck between chain of custody of information and privacy. Private citizens would not want their content to be authenticated/signed necessarily, but if all government communications were authenticated/signed, then they could at least be verified as genuine.

Best,
Mark

Mark Raugas

Pacific Northwest National Laboratory

Please send your ideas to pcast@ostp.eop.gov with "Generative AI" in the subject line. We especially welcome public comments addressing the following questions (please indicate in your submission which questions you are addressing):

1. In an era in which convincing images, audio, and text can be generated with ease on a massive scale, how can we ensure reliable access to verifiable, trustworthy information? How can we be certain that a particular piece of media is genuinely from the claimed source?

2. How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens?

3. What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?

4. How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise?

5. How can we help everyone, including our scientific, political, industrial, and educational leaders, develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation?

Generative AI

Written: 5/22/2023

Email: [redacted]

Dear Members of the PCAST Working Group on Generative AI,

I hope this message finds you well. I am writing to share my thoughts on the potential threats posed by Generative AI and possible strategies for mitigating these risks. As an advocate for responsible AI with a background in urban design, planning, and urban sociology, I believe I can offer a unique and valuable perspective on this matter.

Attached to this email, you will find a detailed letter outlining my ideas and suggestions. I hope this will provide useful insights as you continue your important work on Generative AI. I am looking forward to the opportunity to contribute further to this crucial discussion.

Thank you for your time and consideration.

Sincerely,

Mehri M. Mohebbi, Ph.D. (She/Her)
Program Director - TE Certification Program
Equity in Transportation Initiative Lead
UFTI Research Faculty

[redacted]
512 Weil Hall, P.O. Box 116580 Gainesville, FL 32611

Attachment:

Dear Members of the PCAST Working Group on Generative AI,

I am writing to you as an advocate for responsible AI, with years of experience in the areas of inclusive public decision-making, transportation equity, regional planning, and urban sociology. My multidisciplinary background has provided me with a unique perspective on how emerging technologies, like Generative AI, can impact our communities and daily lives. Given the broad implications of Generative AI, having diverse voices and expertise at the decision-making table is crucial. This includes not only those with technical expertise in AI but also those with experience in various fields that will be impacted by this technology, such as urban planning and sociology. This breadth of perspectives can help ensure that the development and deployment of Generative AI are informed by a comprehensive understanding of its potential societal impacts and are guided by principles of fairness, inclusivity, and equity.

Generative AI holds immense potential for innovation and societal advancement, but it also poses serious challenges that need to be addressed thoughtfully and promptly. Here, I would like to share my thoughts and ideas on this significant matter:

Robust Legislation and Regulation: We must consider introducing comprehensive regulations that clearly outline the permissible and impermissible uses of Generative AI. Balancing the growth of technological innovation with protection against misuse will be crucial in this endeavor.

AI Ethics Guidelines: A set of ethical guidelines for the usage of Generative AI should be developed, promoting transparency, accountability, and respect for human autonomy.

Research into Threat Detection and Mitigation: Encouraging research into detecting AI-generated content and developing "watermarking" methods will be pivotal in mitigating potential threats.

Public Awareness and Education: By promoting public awareness and integrating AI literacy into education curriculums, we can ensure the general public is well-informed about the potential risks and signs of AI-generated content.

Collaboration with Tech Companies: It's crucial to work closely with tech companies to establish robust policies for handling AI-generated content on their platforms and to promote data sharing on detected AI threats.

International Cooperation: With the global nature of the internet, international cooperation will be essential in managing Generative AI risks, including creating international standards and sharing resources.

Privacy Protection Measures: Strict data handling and privacy protection measures should be implemented to prevent misuse of personal data.

Auditing and Transparency Measures: It's important to have auditing processes to ensure compliance and transparency measures that require companies to disclose how their AI systems work.

Red Teaming: Conducting "red teaming" exercises will help identify potential threats and improve security.

Socioeconomic Impact Assessment: Regular assessments of the impact of Generative AI on jobs and the economy will be needed to develop strategies to mitigate negative consequences.

Managing these risks requires ongoing effort and adaptation as the technology continues to evolve. I am confident that with thoughtful deliberation and action, we can guide our nation towards a future where the promise of Generative AI is realized, and its potential threats are effectively managed.

I appreciate your consideration of these ideas and look forward to contributing further to this important discussion. My unique blend of experiences and expertise positions me to be a valuable contributor to this pursuit, and I welcome the opportunity to help shape our collective approach to Generative AI.

Thank you for your time and commitment to this crucial matter.
Sincerely,
Mehri M. Mohebbi
Transportation Equity Program Director – UFTI
University of Florida Transportation Institute, FL.

Generative AI

Written: 5/23/2023

Email: [redacted]

Dear PCAST Representative:

Artificial Intelligence (AI), unlike any technology in our history, holds the power to revolutionize our world, yet it also poses significant potential threats. As AIs' reach expands and its capabilities increase, it opens doors for unprecedented opportunities and, unfortunately, unprecedented risks, including existential ones.

As a concerned global citizen, I urge you to prioritize the development of comprehensive, international AI regulation. This isn't about stifling innovation, but rather about ensuring the safe and responsible evolution of this transformative technology. AI warrants a unique level of governance, one that is discerning, proactive, and international in scope.

Effective AI regulation should promote accountability, transparency, and the safeguarding of human rights. It must act to prevent misuse, mitigate unintended consequences, and ensure AI development and deployment align with the broad interest of humanity. This includes the establishment of safeguards against "bad actors" who may seek to misuse this technology, leading to potentially catastrophic outcomes.

With the rapid pace of AI development, the window for effective action is narrowing. We are venturing into uncharted territory, and unlike many areas of policy, we may not get a second chance to correct our course.

I implore you, as our representative, to take decisive action. Your leadership can play a pivotal role in guiding the trajectory of AI, ensuring that its profound potential benefits humanity, and does not lead to our detriment.

Thank you for your attention to this pressing issue. I, along with the rest of humanity, eagerly await your thoughtful response and decisive action. We are counting on you and trust that you will take the necessary actions to safeguard our future.

Sincerely,
Mike Brooks, Ph.D.

Mike Brooks, PHD, PC
Psychologist, Author, Speaker

Website: >www.drmikebrooks.com<

Email: [redacted]

Twitter: [redacted]

Facebook: [redacted]

Author of Tech Generation: Raising Balanced Kids in a Hyper-Connected World

Address:

[redacted]

Generative AI

Written: 5/23/2023

Email: [redacted]

Response to question 3

I believe it is important to develop Neural Entanglement Technology, which would use quantum entanglement to connect a quantum AI system with humans' neurons, creating a dual consciousness (the system and mind operating in unison) in which the system would be able to utilize human consciousness and experience to learn while the human would be able to tap into the systems' computational and processing capabilities. This would successfully create a conscious symbiotic relationship between humans and AI.

Generative AI

Written: 5/25/2023

Email: [redacted]

1. In an era in which convincing images, audio, and text can be generated with ease on a massive scale, how can we ensure reliable access to verifiable, trustworthy information? How can we be certain that a particular piece of media is genuinely from the claimed source? First, I think we have to "tag" any AI production as artificial. It should be illegal for AI products or AI generated social media/news posts to not have to include a tag and disclaimer that the information or product was artificially produced. Think about a watermark. It would be in the same vein as a watermark but it would have to be legally tagged to any AI material. The next step would be to make sure that the AI company was also part of the tag so if there is unverifiable and untrustworthy information being spread, that AI company would be responsible and face legal ramifications.

2. How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens? In my opinion, we would be better served to focus on how to train citizens to be more discerning as well as more educated about the potential deception of AI. All public education curriculums should begin that training from kindergarten. Additionally, incentivize truth telling. We have monetized sensationalism to such a degree that it is now more profitable to stretch the truth into a lie, rather than tell the truth. We have really creative and brilliant people willing to counter false narratives. Encourage them even more through financial incentives and awards.

3. What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation? Educate and improve citizen's abilities to detect and counter AI-generated disinformation. Every AI company should have to provide mandatory financial support for massive education programs to assist citizens. Think in terms of Driving License classes or incoming college freshmen who have to pass a swim test. Whether you are applying for a driver's license, entering community college,

college, military, junior college, etc, you need to be able to pass an AI disinformation test.

4. How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise? Ban AI generated posts, personalities, accounts from social media and television space.

5. How can we help everyone, including our scientific, political, industrial, and educational leaders, develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation? If they can think things through, they already have the skills needed. Now, there needs to be more comprehensive training to apply their thinking skills. There is way too much ignorance around AI and that needs to be eliminated immediately. Don't treat AI like we did COVID because there will be ridiculous amounts of people succumbing to the ills of AI due to ignorance. Get entertainers and athletes involved. Advertise and buy advertising space on all major social media platforms. Flood the colleges and universities with training and incentivize educators to devote class time to teaching and improving competency about AI. One group that you have left out of this question is clergy. The religious world will have a lot of impact on AI-generated misinformation, impersonation, and manipulation. Don't forget about them.

--

Danielle Koonce

Doctoral Candidate *Sociology*

Lecturer, TA for *Department of Sociology*

Member *Student Affairs Committee* and *The Office of Graduate Diversity and Inclusion Advisory Board*

Generative AI

Written: 5/26/2023

Email: [redacted]

Hello,

Following the recommendations here I wanted to share what Vermont is doing in the AI Ethics space, specifically regarding how AI (including generative AI) is used within government processes to maintain trust in institutions. Our latest Policy Draft on use within government is here: https://legislature.vermont.gov/assets/Legislative-Reports/Council-Report-on-AI-Ethics-Policy.pdf

In general, we require that all AI-generated content be attributed to a human author, who is required to take responsibility for any decisions made by AI (including content produced by generative AI). Additionally, we are currently considering guidelines that require AI made decisions (and generated content) to be labelled, and we are actively discussing the design of a "nutrition facts" label for AI-generated content, with some standard information about the inputs and processing system.

Additionally, I'd love to be more connected with the work going on at the federal level on AI governance. If there's any way I can help or if there are resources you are working on that would be relevant to us, I'd love to talk in more detail.

Thanks,

Josiah Raiche | Director

Division of Artificial Intelligence

Vermont Agency of Digital Services

Dewey Bldg, 1 National Life Dr | Montpelier, VT 05620

[redacted]

Generative AI

Written: 5/26/2023

Email: [redacted]

Dear PCAST,

My name is Linden Li. I am an AI researcher at Stanford interested broadly in foundation models and generative artificial intelligence. I have worked on the following areas of research:

1. AI interpretability research. At Stanford, I worked with Professor Fei-Fei Li on a project to develop interpretable video understanding models. The goal was to develop a dataset to allow for understanding complex scenes, with applications to healthcare. This research was eventually published at NeurIPS, a top machine learning conference.
2. Efficient model training. During internships, I worked on software to efficiently train vision models at NVIDIA Research and built a library for users to train large language models on custom data at technology startup MosaicML.
3. Generative diffusion models. I've recently been working in a collaboration with the Stanford AI Lab and Graphics Labs on methods to enable artists to have greater control over image generation outputs.

I have attached my comment to this email. Let me know if you have any questions!

Thanks,

Linden

Attachment:

Dear PCAST,

My name is Linden Li. I am an AI researcher at Stanford interested broadly in foundation models and generative artificial intelligence. I have worked on the following areas of research:

1. **AI interpretability research.** At Stanford, I worked with Professor Fei-Fei Li on a project to develop interpretable video understanding models. The goal was to develop a dataset to allow for understanding complex scenes, with applications to healthcare. This research was eventually published at NeurIPS, a top machine learning conference.
2. **Efficient model training.** During internships, I worked on software to efficiently train vision models at NVIDIA Research and built a library for users to train large language models on custom data at technology startup MosaicML.
3. **Generative diffusion models.** I've recently been working in a collaboration with the Stanford AI Lab and Graphics Labs on methods to enable artists to have greater control over image generation outputs.

I'm also passionate about CS education through my involvement as a Section Leader for Stanford's introductory computer science classes. Below are my thoughts on generative AI and its influence on democracy.

Please reach out to me at: lindenli@stanford.edu with any questions about my response below.

In an era in which convincing images, audio, and text can be generated with ease on a massive scale, how can we ensure reliable access to verifiable, trustworthy information? How can we be certain that a particular piece of media is genuinely from the claimed source?

Broadly speaking, I think there are two aspects to ensuring access to trustworthy information:
1. There should be a method to distinguish AI-generated outputs from real ones.
2. AI systems should minimize generation of explicitly harmful content.

I will discuss each modality separately.

Images
In this section, I will focus on how to mitigate harms associated with imagery generated from AI. In a later section, I will elaborate on technologies that will help the public in determining if media comes from a real person. I think there are two classes of solutions that could ensure reliable access:
1. **Images that are generated by artificial intelligence should require a watermark.** This can be done in two ways. First, images could have an explicit visible watermark decided by the manufacturer on the bottom of the image. Getty Images already does this with their licensed images and OpenAI's DALLE-2 has a rainbow watermark that appears on the bottom right of the image. This allows users to visibly distinguish that a given image is synthetically generated. Second, images could have an invisible watermark. Recent work in this paper establishes a recipe for watermarking images. It allows for a diffusion model to be trained with a specified binary watermark string; any generated image from the diffusion model can have the binary watermark string recovered. A piece of regulation could involve AI image generation companies like OpenAI and Midjourney being assigned a government-registered watermark to be properly licensed. Browsers can be implemented to check for registered watermarks; if one doesn't exist, it can be flagged to let a user know. *Note that this is very similar to existing web security infrastructure.* To get rid of fraudulent imitation websites, browsers check to see if websites have recognized certificates given out by registered Certificate Authorities to verify legitimacy.
2. **Prevent the generation of likenesses.** Recent harms generated from AI-generated images have largely resulted from the generation of likenesses that resemble public figures (e.g., a trending image of former US President Donald Trump getting arrested). It could be easy to verify that an AI-generated image violates a policy by instituting a *safety checker*: these are simple computer vision models that could detect if a face is an a given image. AI-generated models could have their outputs vetted by a safety checker to filter out negative outputs (such as the generation of likenesses).

Audio

I have personally used open-source audio generation tools (e.g., Tortoise-TTS on GitHub) and experimented with APIs like ElevenLabs. Generating these fake audio messages was easy; it's a lot simpler to generate plausible audio samples compared to imagery. APIs right now allow you to upload an arbitrary voice recording from which it can generate a new audio sample very quickly. There should probably be some limitations to doing this – instead, API providers should just provide a set of voice presets that users can generate audio samples with. Allowing users to quickly get a model to generate whatever voice they want is harmful.

Text

There is currently technology to detect whether text was generated by an AI that I elaborate on in a later section. I believe the biggest harm from text is two-fold:

1. It is easy to disseminate information that resembles the linguistic style of a speaker.
2. Language models can hallucinate misinformation. If treated as a ground truth source of information, it can feed misinformation to a client – a good example could be a voter asking ChatGPT about a politician.

There are ways to reduce misinformation by grounding outputs in facts (data regulation and retrieval augmented generation mentioned in a later section).

What technologies, policies, and infrastructure can be developed to detect and counter AI generated disinformation?

Technology and infrastructure suggestions:

1. **Verifying that visual media comes from a trusted source.** It's important to distinguish if visual media comes from a *real camera* or not. A potential solution is to have a mechanism that allows for easy verification that a given image comes from a camera. There are cryptographic techniques to do this: Sony recently manufactured a camera that signs its image outputs, allowing someone to determine that the output came from a Sony camera. Camera manufacturers could send known signatures to some authority, where visual media can be verified as coming from a trusted source.
2. **Detectors that can determine if text is generated from a fake source.** There are existing systems like DetectGPT from Stanford that try to determine the probability that a chosen excerpt of text came from a model. These discriminators work with very good performance (higher than 0.95 AUROC). Since there are still errors (especially false positives), this should be used a tool to flag potentially suspicious content that may need to be verified. If a user sees a warning that some tweet sharing political news was potentially generated by a language model, they are more likely to be more skeptical.
3. **Reducing the risk of hallucination.** When people chat with a language model, they may not understand the risk that it could produce harmful information – even if there is a disclaimer indicating that this is possible. There are technologies being developed that ground GPT-generated outputs with a knowledge base known as *retrieval-augmented generation*: this allows an AI agent to access a search engine and synthesize the relevant information, giving it the ability to cite its sources. This would reduce the risk of misinformation.

Here are some additional policy recommendations:

1. **Data regulation and transparency.** In my experience, there's a big difference between image generation models like Adobe Firefly which have been trained on properly licensed images versus Midjourney, which doesn't release its training data. Similarly, it's hard to assess the capacity of a system like GPT-4 for misinformation and harm since we don't know what data it was trained on; comparatively, open-source datasets like the Pile that many open models are trained on have significantly larger data transparency, which allows for crowdsourcing analysis of the harms. If language models are trained on faulty data that contains misinformation (e.g., it was trained on human-generated text that was originally designed by malicious actors to misinform), then it's more likely to regurgitate this harmful content.
2. **Consider accountability for misinformation.** I'm undecided as to who should have culpability for misinformation, but instituting penalties on the provider side and/or the user side could provide disincentives for building systems that create misinformation. The response from providers could be to institute greater safety checks to counter misinformation and users would be to avoid using these tools to for malicious means.

How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise?

I get my political information primarily from two sources:

1. Official news channels like the New York Times and the BBC.
2. Social media, like Twitter and Facebook. This primarily comes from two sources: people's reported experiences and verified government accounts.

I want to note that the latter source (reported experiences on social media) is important to ensure that there are other sources of media outside of the government. However, these sources are more prone to manipulation.

1. We should continue to implement verification from government sources. This has been helpful for me to distinguish if something is truly coming from a government voice or not. There should be some official verified and irreplicable symbol that certify that something comes from them rather than media produced by AI to mock a government message. This will reduce the risk of AI-generated impersonation accounts.
2. Social media should try to rely on their large user bases to crowdsource whether information is true. If there's a video of an event that someone records on their phone, people could choose to flag it as suspicious if they were also at the event and did not verify that the event happened.

How can we help everyone, including our scientific, political, industrial, and educational leaders, develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation?

When talking to people, I think there are two misconceptions surrounding AI.

Underestimating the capabilities of AI

The message surrounding the capabilities of AI is largely inconsistent: some people focus on the failure cases of AI, which affects how people perceive the risks. People become more susceptible to misinformation when they don't believe that AI systems are capable enough of producing plausible outputs. A lot of people's information about generative AI capabilities—e.g., image generation technologies—might be from what they saw years ago; previous generative AI technologies like GANs, for example, were only able to do well at generating synthetic portraits of novel faces but could not generate plausible complicated scenes.

I think there are a couple of ways to help people identify AI-generated misinformation:

1. **Properly understanding capabilities:** training programs for leaders in various fields could expose people to outputs generated from state-of-the-art generative audio, text, and image models. The most important idea people should understand right now is that *AI now can generate highly realistic outputs*, meaning that assessing how *real* an image looks is no longer a good proxy to check if something is AI generated. By being aware of capabilities, identifying misinformation is a lot easier.
2. **Re-emphasis on trusted sources.** Malicious actors are more likely to exploit forums that are not highly audited – it's harder to manipulate an official news source than an online forum or a social media post. Training people to cross-reference what they see with a trusted source could be a helpful way to reduce misinformation.

Overestimation

The other class of misinformation has people overestimating the capabilities of AI. The abstraction presented by the media makes AI appear like a black box system capable of learning an arbitrary skill at an incomprehensible, superhuman capability (cf. reinforcement learning systems like AlphaGo). I think explaining the way that AI systems are built, without focusing on the low-level technical details, can reduce the risk of overestimating capabilities and allow people to become more capable in identifying AI-generated content:

1. **Understanding what data the models are trained on.** An understanding that language models like ChatGPT are trained on real political speeches, image generation models like Midjourney are trained on pictures of public figures, and audio models are trained on audio recordings of real people can allow people to become more skeptical of what they see online. Implicit in understanding this is some sort of data regulation: a lot of models are closed source right now, so it's difficult to assess what data was input into these models.
2. **Understanding the training procedure of the models:** the training objective of these models can also provide insight into why certain harms materialize. Understanding that large language models are built via modeling a statistical distribution of text illuminates why systems to return outputs that seem plausible, but aren't necessarily real or grounded in facts. This can lead to greater skepticism by understanding the limitation of current technology.

Generative AI

From: Simha Sethumadhavan

Written: 5/31/2023

Email: [redacted]

What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?

REGULATIONS FOR EMERGING SYSTEMS

Prof. Simha Sethumadhavan
Department of Computer Science, Columbia University
Email: [redacted]

Today, conversations about regulating various technology companies, including social media and AI firms, as well as software and hardware providers, are more pertinent than ever. This short note describes how we could establish practical regulatory guidelines for these rapidly evolving systems.

Indeed, regulations have proven beneficial across numerous domains, yet they also pose significant challenges for emerging domains. Firstly, creating regulations is a lengthy process that requires careful social consideration. With the fast-paced advancement of AI and rapidly evolving methods of spreading misinformation and disinformation, regulators face a tough decision: How frequently should rules be updated? There is a fundamental mismatch between rulemaking frequency and the rate of progress in these fields.

Secondly, these systems are intricately complex, with multiple layers of abstraction, and unexpected behaviors can arise when these layers interact. Therefore, regulators must determine at what level they should apply their rules. Complicating this issue further is that regulators may lack access to crucial proprietary information necessary for effective rulemaking. Typically, industry experts are invited to participate, but this can give an unfair advantage to large companies that can afford to invest time and resources in rulemaking. This could potentially lead to regulatory capture, unfairly disadvantaging smaller entities. Conversely, exempting smaller entities from regulation can be detrimental, as it is challenging to retrofit security measures and abuse prevention once a product or piece of media goes viral.

This note describes an innovative solution to the intricacies of regulating evolving technological systems. We propose an “open security mandate,” an approach that bypasses the rigid stipulations of conventional regulations and grants companies the flexibility to respond quickly to emerging threats. At its core, the concept is highly uncomplicated: All product vendors, be it a single-product company like TikTok or a multi-product vendor like Microsoft, must dedicate a portion of their resources to security and transparently disclose their security expenditures.

This proportion of resources, termed the "security budget," is broadly defined. It may include various company operating costs, such as employing human moderators to filter content, legal counsel, as well as system resources like the processing power and energy consumed by AI in data centers or on mobile devices. Crucially, under our open security mandate, decisions regarding how and where to allocate the security budget are made by the product vendors, not regulators.

Let's take the example of a mandate that mandates social media companies to dedicate a certain percentage of their operational budget to content moderation. In the context of control theory, this requirement can be considered part of a closed-loop system where the regulators act as the controller. If the system's output, which is the level of misinformation and disinformation, exceeds a preset threshold, the controller (regulators) can increase the input (mandated spending) to the system by changing the required level of spending. This action can be equated to a negative feedback mechanism that aims to maintain the system's stability by decreasing the error (prevalence of misinformation and disinformation). Conversely, the system could be seen as over-compensating if the spending is too high and the output (instances of abuse/misinformation/disinformation) is acceptably low. Here, the companies might request the regulators to reduce the mandate, a positive feedback mechanism to adjust the system's input (spending), and optimize performance. Since security spend is directly related to profit, companies are incentivized to spend the exact amount needed to keep mis and disinformation under the threshold. The feedback for the system can be obtained through netizens who can report this data, through simulated experiments carried out in controlled environments, free market companies that produce this information, or a combination of all of the above.

Moreover, this mechanism provides a unique opportunity for emerging small companies. Given the flexibility and control over their security budget, they can strategically allocate resources to create safer platforms, ultimately enhancing their market reputation and competitive advantage. Consequently, these smaller entities are incentivized to actively participate in this regulatory framework actively, aligning their growth objectives with a more secure ecosystem.

The same mandate can apply to a generative AI company. A security budget can be satisfied by work needed to train or fine-tune models to provide more truthful answers, work needed to filter out socially harmful behaviors, work required to insert "watermarks" to identify AI content, work needed to protect the AI model from illegal copying or stealing the model, or the cost of implementing access controls.

While the above mechanism grants significant flexibility to companies in executing their security measures, they will need to develop accounting mechanisms to track their security spending, preserve this data in an auditable form, and ultimately disclose this information in a manner that the public can meaningfully interpret. Although regulatory bodies cannot inspect these records regularly, they can conduct comprehensive “no holds barred” investigations in the event of an incident. Substantive penalties can be levied if an AI company or the social platform has neglected its responsibility to gather and maintain all necessary information accurately. A new federal AI agency or an existing agency like FCC can enforce these requirements.

Our prior work on applying Open Mandates to combat ransomware threats and implementing memory safety (https://arxiv.org/abs/2203.05015) demonstrates that many intricate technical questions can be addressed. It shows that such regulation can be advantageous.

The open mandate idea is not intended to replace solutions effective in other domains (e.g., FCC rules and regulations, medical, student, and financial privacy rules and regulations, etc.). Instead, it's a different concept potentially more beneficial to emerging technologies and complementary to existing regulatory frameworks.

BIO:

Simha Sethumadhavan is a Professor of Computer Science at Columbia University, specializing in developing practical solutions to cybersecurity and computer architecture challenges. He has been recognized for his contributions with several awards, including an Alfred P. Sloan Research Fellowship, NSF CAREER award, ISCA Hall of Fame nomination, and ten best paper awards at highly peer-reviewed venues.

His work has significantly improved the security of widely-used hardware and software products: from mobile phone processors to web browsers benefiting millions of users globally. Simha pioneered a type of hardware-based antivirus that is now productized in laptop processors to protect against ransomware. His contributions to hardware security have been used by standards organizations.

In addition to his academic work, Simha founded Chip Scan Inc., a hardware security company located in a climate-disadvantaged part of NYC, to bring DARPA and NSF-funded microelectronics security research from his lab to the market. The company, founded in 2014, is one of the few companies accredited as a DoD/DMEA trusted supplier for microelectronics trust and assurance services and provides cybersecurity for critical infrastructure and weapons systems.

Simha has also served as a technical advisor for US federal government research lab projects and on an FCC Technical Advisory Committee. He was the chair of the Cyber Security Center at the Data Science Institute at Columbia University.

Simha received his Ph.D. from UT Austin in 2007.

Generative AI

From: Zachary Elewitz

Written: 5/31/2023

Email: [redacted]

To whom this may concern,

My name is Zachary Elewitz and I know there has been significant justified governmental focus on AI recently. I am elated that this work has begun but it will need continued reform in order to stay current and effective. I am interested in helping out.

I have a decade of data science and artificial intelligence experience across numerous industries as well as holding a PhD in Mathematics and an MBA. I am currently the Director of Data Science at Wex and the Chief Data & Analytics office for Mystery on Main Street. I stay up to date on the ever-changing AI landscape and am able to communicate its nuance without relying on jargon.

I recently moved to DC and am looking for an area to make a difference. Is PCAST interested in advisory board members from industry to help with policy and strategy development? I am not looking for compensation - only to make a difference.

Please let me know how I can help.

Thank you!

Zachary Elewitz

Generative AI

From: Sean Koon

Written: 6/7/2023

Email: [redacted]

Dear Council,
Hello! Here's commentary in response to the invitation for input. The attached file contains the text below.

Sean Koon, MD, MS

RE: Commentary for PCAST Working Group on Generative AI

This comment is relevant to question #2 regarding the manipulation of the "beliefs and understanding of citizens" and perhaps other questions listed. However, this suggestion is qualitatively different from (but complementary to) to any efforts to validate information sources or prevent disinformation. This commentary emphasizes the opportunity combat misleading information by augmenting and supporting human reasoning itself using AI tools.

The members of a democracy must have the opportunity to think critically and to reason about issues important to them. This ability has become challenged by an increase in targeted misinformation from media and social media sources. Misinformation achieves a level of concern now with the advent of Generative AI such as the Large Language Models (LLM's) of Chat-GPT, image or multimedia-generating tools, etc. Many necessary approaches will emerge, with strategies to discern between human and computer-created content, regulation of LLM's, clarifying the provenance of the information or explainability, verifying the validity of the training datasets, etc.

A complementary and equally necessary approach involves supporting readers in being able to critically evaluate the information that they face, whether it is from generative AI or not. Misinformation often (perhaps most often) has less to do with the correctness of a particular "fact" and more to do with a persuasive but misleading presentation. Often some statements of fact are made, but with crucial context or opposing arguments left out. Exaggeration or understatements are made, persuasive rhetoric is used in place of facts, and false logical structures are employed that can lead to compelling, if incorrect, conclusions, etc. Generative AI can easily replicate these misleading approaches and fact-checking does not resolve these issues for at least two reasons. First, an article can be mostly factual but highly misleading. Secondly, we have recently seen that our established and authoritative sources can be rejected by citizens if someone is able to frame them as politically affiliated or biased. It is far more difficult to create an accepted and authoritative source than it is to cast doubt upon it through social media postings. Those who have grown up with the internet may not believe anything they read. Neither excessive trust nor skepticism supports the needs of a democracy. Citizens require an ability to reason about whether information is true, whether it is complete, and whether its presentation is appropriate to implications or conclusions made.

Reasoning Support

Reasoning support leverages innate reasoning capabilities and enhances them. It should help people not only to reason about problems or questions, but also to reason about the information sources themselves.

The jury system is an interesting example of this, in that people who may not have the technical understanding of the topic and may have low literacy or education in things like the law, statistics, science, medicine, etc. Yet the judicial system has created a structure that employs their innate reasoning capabilities and human values even though they face novel, complex, and potentially incorrect information. While this is certainly not a model to apply directly to the challenges of generative AI, it is proof of concept that the plurality of a democracy can be powerfully leveraged through the use of informational tools and processes.

In terms of technology tools to support reasoning, these are possibilities:

3. Offering plain-text summaries that can take a large, persuasive article and summarize its basic assertions in a short paragraph or set of bullets, with rhetoric removed
4. Highlighting potential logical fallacies: “This sentence makes a strong claim[x] with no supporting source”, “this makes a broad generalization from a single anecdote”, “this implies that [x] is causing [y] because they occurred together, but doesn’t give causal evidence”, “this text attempts to refute a statement, but only does so by making an attack on the character of the person stating it (*ad hominem)*, “this paragraph uses arguments that are “circular”—they are only valid if you assume their conclusion to be true” (i.e. tautology/begging the question),
5. Stating the single most concerning flaw of a text
6. LLM-based tools to rewrite text without bias
7. Exploring plausibility:

a. Pointing out internal inconsistencies within a text

b. Pointing out inconsistencies with universally accepted facts, visible realities, or common sense, i.e. “What else would we also have to believe in order to accept this assertion?”

c. Expanding our consideration of downstream impacts. An example might be historical revision. If a historical event is asserted to be different than documented, then whatever event actually occurred should have many logical downstream impacts. Basically, what would we also reasonably expect to be different in the world we agree with this assertion about a historical event?

8. On-demand automated commentary or argument about a topic
9. Offer a score that measures a ratio of emotionally powerful content (persuasive rhetoric) vs. supportive facts
10. Clarify exaggerations or understatements. “There has been a surge in ‘X’” might be annotated or footnoted with AI-generated, “‘X’ has increased 2% in the last month but is 20% lower than it was 2 years ago”. Or offer a graph of the trend.

Along with reasoning support, there can be tools that analyze the factuality of information. It is critical that these be separate resources, given that assertions statements of truth or falsehood may attract attacks on the tool itself and the source of that tool. A major rationale for reasoning support is to place the agency in the user and avoid some of the controversies around who has the authority to decide what is true.

Information Support

Efforts to validate information or to regulate or prevent misinformation may be hampered by challenges to authority: who has the right to say what is true? That being said, good information is obviously essential to good reasoning. Some ideas for potential tools may be as follows:

11. Tools that facilitate search and summary of factual sources
12. Assigning a "level of evidence" to assertions
13. Highlighting statements that are considered controversial based on verified sources
14. Showing which affinity groups support a controversial assertion. While this bears a similar risk to the problem of "filter bubbles", it may also serve to demonstrate to a user that they have been reasoning within a filter bubble with a specific group identity. It may also clarify the other groups that the reader is aligning with when adopting a controversial belief.
15. Providing supportive or non-supportive information from trusted sources.
16. Promote independent sources. For example, the Congressional Research Service (CRS) creates non-partisan information on important issues. However, it is typically not reviewed directly by citizens but rather filtered through congresspersons who access the resource, posing an obvious risk of bias. A "consumer reports" of information sources or information itself may be of some value.

Fitting the Needs of User Groups

Depending on the intended user group or circumstance, these tools might give feedback which can be reviewed in seconds, minutes, or in great depth over time. A reader may need a tool that scans an article in real time while they are reading it, highlighting key reasoning considerations. Writers may use tools integrated within their word processor, like grammar or spell-checking. Professional reviewers or "watchdogs" might use tools that analyze large volumes of content to support an overall rating of a publisher or source, etc. Generally, an ability to provide a minimalist insight, accompanied by an ability to "drill down" into the reasoning issues or factual basis of the content.

Feasibility

The basic building blocks of these tools exist or are emerging. Certainly, large language models have incredible capabilities that may be leveraged towards reasoning support. Traditional statistical methods and visualizations can give balancing context to assertions made. Many devices of rhetoric and logical fallacies are well described, available in any "Critical Thinking" text. AI classification algorithms may assist in analyzing new text, perhaps from a website, and in identifying/classifying sections where such devices or logical fallacies are employed. AI tools built to search and summarize internet resources have an essential role in both reasoning and fact-checking tools. Also, tools are emerging in sentiment analysis, intent analysis, and related areas to evaluate text. The Genesis Group at MIT CSAIL has an evolved research program with capabilities that may contribute to reasoning support tools. AI tools that interpret the content of pictures or videos can also be leveraged so that the interpreted content may be evaluated for the illogical inferences they may present.

Conclusion

Misinformation and disinformation have always posed a challenge to citizen participation, but the threat is perhaps greater now, or at least there are new and powerful threats. There is an urgent need to support citizen reasoners in real time, giving them technology-based tools that help them reason about the informational sources that they experience.

Sean Koon, MD, MS

Generative AI

Written: 6/14/2023

Email: [redacted]

The following actresses and actors:

Emma Watson, James Franco, Ezra Miller, Seth Rogan, Dave Franco, Danny McBride, Daniel Radcliffe, Zac Efron, Laurence Fishburne, Ben Affleck, George Clooney, Brad Pitt, Hugh Jackman, Bradley Cooper, Jesse Eisenberg, Edward Norton, Billy Bob Thornton, Paul Giamatti, Dwayne Johnson, Clark Duke, Lewis CK, Natalie Portman, Susan Sarandon, Woody Harrelson, Tom Hanks, Tom Hardy, Anthony Mackie, Tom Holland, Robert Downey jr., Scarlet Johansson, Chris Evans, Emma Thompson, Emma Roberts, Emma Stone, Kristen Stewart, Kristen Bell, Kristen Wiig, Elliot Page, Caitriona Balfe, Cillian Murphy, Mark Strong, Rachel McAdams, Zooey Deschanel, Zendaya, Sarah Jessica Parker, Martha Nussbaum (philosopher, University of Chicago), John Searle (philosopher, UC Berkeley)

need to be informed that there is a group (or groups) out there who have gotten ahold of new neurotechnologies and are using the voice and image (accompanied by sensations) of various actors and actresses against me (attacking me, tormenting me, molesting and raping me, and so on); either that, or these actresses and actors are directly involved in these abuses of neurotechnology (see end of message for some credible resources to get started regarding neurotechnology, neuroethics, neurorights, and the abuses of neurotechnology)… this is being done remotely. See: adamchristiannielsen.com for my work (poetry)...this might be its only chance at life. There is an autobiography in the letter attached, or a briefer one on my website.

Please pass along the following messages to the appropriate parties (all of this is newsworthy; a story that needs to be told, an issue which needs to be addressed):

1: An Urgent Appeal; Activism Against the Abuses of Neurotechnology
2: The Torture and Rape of Adam Christian Nielsen, Remotely, by New Neurotechnologies
3: The Coercion of Adam Christian Nielsen into A Relationship with Emma Watson (or imposter)

4: Counts of what was done to me during an eight week span of time
5: Credible Resources Regarding neurotechnology, neuroethics, neurorights, and the abuses of neurotechnology

1:
Activism against the abuses of neurotechnology

This is an appeal, an urgent plea for your immediate help... please, they are tormenting me even as I send this. I am being held captive, tormented, abused, tortured, and raped, remotely, by new neurotechnology... others are being targeted as well. This issue concerns us all, and needs to be addressed and resolved before it goes any further. We, as targeted individuals, need help your help in dealing with the abuses of these new neurotechnologies… we are scattered, disorganized, and often powerless and/or terrorized to the point of being virtually incapable of doing anything about it, and need credibility, especially with the media (which can be hard to come by insofar as it's easy to pass this off as schizophrenia or other mental health related disorders). Really throw your weight into it, now, or wait around until it's happening to you or a loved one I guess... there are people out there who are doing just about whatever they want to others with new neurotechnology. So let's try to coordinate, because I've been tired of this for a while now… here's what needs to be done:

-contact news media and get the word out (you can use my or other testimonies) (maybe use your own contacts, connections, and colleagues)

-adamchristiannielsen.com for my work (poetry)

-develop counter measures against the abuses of neurotechnology (ex. means of identifying or detecting its occurrence)

-contact your local elected representative about the abuses of neurotechnology
-enact laws against the abuses of neurotechnology
-equip law enforcement to deal with the abuses of neurotechnology
-investigate the situation in general as well as specific cases
-justice for targeted individuals
-keywords/phrases: neuroscience, neurotechnology, neurorights, neuroethics, etc... still working out the language here

Let's start there… I may have forgotten a few things, but it'll get us going. You can read the rest below, the point is that this is happening and we need to rally now.

2:

Adam Christian Nielsen

My name is Adam Christian Nielsen. I am a poet-philosopher from the northern California valley, and am fighting for my life and the lives of many others. I am a targeted individual of surveillance, torture, rape, voice-to-skull, neural monitoring (and so on) by new neurotechnologies (calling this, tentatively, the abuses of neurotechnology); this is being done remotely, it affects your body directly through your brain from however many miles away… human rights violations, civil rights violations, and crimes against humanity are taking place against myself and many others all over the world by new neurotechnologies. This started, in my case, in Fall 2016, with voice-to-skull and remote neural monitoring. By 2018 mind-games were common. In 2019 I was unable to complete a Masters degree in the Humanities through California State University, Northridge (which I was attaining so as to teach) after a year into my studies due to interference by these neurotechnologies. And 2020 - 2021, I was being verbally, emotionally, physically, and sexually abused, as well as harassed, molested, raped, and tortured on a daily basis. By February 2022 I was forced into mental health facilities under a false diagnosis, which I was in for ten months, until December 2022, in their effort to conceal what's going on. 2023 has so far been mostly just remote neural monitoring (probably for the purposes of developing this technology, as well as surveillance), voice-to-skull, mind games, image induction, and damage control… I have by now become accustomed to someone's being there in mind all the time. I suspect this is being done by corporations, universities, government agencies, and/or independent groups. And I have been an activist against the abuses of neurotechnology since 2020, when I first realized what was happening to me.

This, the truth, the whole truth, and nothing but the truth… I'll let you decide what to do with it.

I am speaking out on behalf of everyone who has been targeted, and on behalf of everyone who will be targeted if this issue goes unchecked. I am trying to raise awareness about the issue, reach out about my situation, give as accurate of an account of my experience of torment, torture, and rape via neurotechnology as possible, and get this matter resolved.

See: adamchristiannielsen.com for poetry I've written over the years... also, you can see below for more information about the abuses by way of neurotechnology, and other testimonies from other people of these things happening to them (especially the comments section at cyber-torture.com).

When Mark Zuckerburg says that Facebook will be able to be operated 'telepathically' in the future, or you hear talk of the 'brain-computer interface' or 'neural monitoring', these are instances of the kind of technology we are dealing with here. This stuff is new, it is being developed now, so it may be unfamiliar, and may come as a surprise… it was and did to me.

Though I'm not exactly sure who all the players are yet, they no doubt have access to technology of this sort. In all likelihood there are multiple players involved (ex. technology companies linked up with entities of government linked up with private interests).

Manipulation of other persons without their knowing is fairly easy with this technology, I witnessed their doing this firsthand with my family and with medical staff; anyone can be targeted at this point for any reason, and any number of things can be done to them. You can do some seriously messed up things with this, especially if people don't know it's going on.

Below, I have included a list of things they can do to you with this technology, or rather, some of the things they have done to me (and thus, are capable of doing to other people).

So there are two things going on, one positive and one negative. Negatively, they can see and hear everything that is going on in your mind (can "read" your mind) and can see and hear everything you see and hear (complete lack of privacy), positively, they can do whatever they want to you, including inducing images, feelings, ideas, and thoughts directly into your brain, it just goes to your body through your brain rather than to your brain through the rest of your body.

They of course try to keep their crimes concealed, and so, for the most part, don't affect a very noticeable change in things unless it is in an extreme case like mine... but in extreme cases like mine they'll try to hide it by, say, calling my sanity into question, or trying to silence me.

Sometimes there is no discernible sign, not even in mind, they are there just reading it and are watching what you do; invading the most intimate of your privacies all the while. Said again, this is being done remotely, and so is even more difficult to detect. I'm still considering the implications of this technology.

This will take the concerted effort of us all, but especially of law makers, government agencies, and news media at least; ambitious detectives and reporters and so on… so contact your local elected representative, call up the private investigators and investigative reporters, grill a couple of technology companies about the implications of their technology (and how it's being developed), and of course be informed about what is happening here, and be aware of the signs of its happening to you or others in your life if it does. But please, do it promptly, before I have accrued any long-term psychological damage, and before this issue goes any further than it's already gone. I am not alone in this… it concerns all of us. Do your part.

Laws will probably need to be made, legal action will probably need to be taken, law enforcement will probably need to be equipped to deal with its occurrences, but justice be done… even if this isn't illegal yet (on grounds that it's unprecedented of course), it is just wrong… this is happening to individuals who don't consent to it and ruinous things are being done; it goes against literally everything the U.S. stands for. This should be made public knowledge, and the people responsible need to held accountable. This technology is here to stay… it needs to be kept in check.

3:

The Seduction of Adam Christian Nielsen by Emma Watson (or imposter)

Add 'seduction' to the list of things they can do: I have fallen in love with Emma Watson (the actress)… she was presenced to me through this neurotechnology, and I proposed to be wed to her, she (they?) said yes, and I've been under the impression that we've been engaged since (2020). This is a confusing situation for me, so bear with me a moment if you would… Emma Watson has either been directly involved, or her image (and voice, a presencing of sorts) have been used against me; I'm not sure which, and aim to find out; it could be both.

I have attached a letter explaining all of this, a letter to Emma from myself, you can take a look if you're interested… there is a fairly detailed autobiography there you can check out for more information on my person. I could use your help informing her that this is happening; maybe getting us in contact with each other… she really should know what's going on. I'm sending you the letter with hopes it makes its way to her; she can decide for herself on me… just throwing it out there and am going to see what happens. It's truly a messed up thing though, messing with people's hearts.

Other people this situation of 'direct involvement and/or presencings being used against me' has happened with in my case so far include: primary suspects: Martha Nussbaum, Emma Watson, James Franco, Ezra Miller, and Jay Dodd... other possible suspects (having a brief presence at least throughout my being abused via neurotechnology) include: Seth Rogan, Dave Franco, Danny McBride, Daniel Radcliffe, Zac Efron, Laurence Fishburne, Ben Affleck, George Clooney, Brad Pitt, Hugh Jackman, Bradley Cooper, Jesse Eisenberg, Edward Norton, Billy Bob Thornton, Paul Giamatti, Dwayne Johnson, Clark Duke, Lewis CK, Natalie Portman, Susan Sarandon, Woody Harrelson, Tom Hanks, Tom Hardy, Anthony Mackie, Tom Holland, Robert Downey jr., Scarlet Johansson, Chris Evans, Emma Thompson, Emma Roberts, Emma Stone, Kristen Stewart, Kristen Bell, Kristen Wiig, Elliot Page, Caitriona Balfe, Cillian Murphy, Mark Strong, Rachel McAdams, Zooey Deschanel, Zendaya, Sarah Jessica Parker, John Searle, and whoever they are working with... Martha Nussbaum is a philosopher from the University of Chicago, John Searle was a philosopher at the University of California, Berkeley, the rest are actors or actresses.

Again, if these people are themselves not in fact involved, then their images (including voice and/or a physical presence of sorts) have been used against me throughout my being tortured and raped via neurotechnology in various ways at various times; if these persons have been involved, then the extent and kind of their involvement will, I hope, be worked out in detail in time (including any others involved not currently listed). All of the suspects on the list should be notified, and I could use some help getting word around.

I do not currently know the reasons they are doing this to me, and am still trying to understand it all; still am trying to find the right terms; still researching. You can contact me by the contact form on my website (adamchristiannielsen.com) if you have any questions, comments, concerns, etc., or have any ideas on a course of action which could be taken here. Again, anyone can be targeted and/or manipulated by neurotechnology anywhere at any time for any reason by now... be wary.

This is the best that I can presently do given the circumstances. Please, forward this to anyone who might be interested, concerned, or able to address it (e.g. family and friends, human rights groups, civil rights groups, law enforcement and other government agencies, elected representatives, news media, private investigators, coworkers, religious communities, scientists, poets, philosophers, and so on)… inform others that this is happening.

Means they have used and might continue to use to conceal it (this is not an exhaustive list): inconspicuousness and embeddedness in everyday life; routinedness; distraction; trying to get you to not think about it, or think about other things; memory erasure; questioning of mental health; defamation, slander, and discreditation; their putting up of a professional front; tampering with medical records and other documents; silencing of the opposition; relationship sabotage; human hijacking (of family, friends, medical professionals, law enforcement and other government personnel, so as to further the efforts at concealment).

4:
counts of what was done to me within an eight week span of time

Weeks spanned: 8 [1/27/2022 – updated and explicating over time, this will do for now][it will be indicated if it was not within those weeks preceding][this is not an entirely exhaustive list of what has happened to me]

forced vomiting ~x30
>20 times in half a night (February 2022)
choking me with my own hand ~x5
forced erection ~x90
forced unerection ~20
rape ~x170 (ex. forced orgasm)
ex: “I’m going to rape you over and over again in front of her” (they said)
seducing me into an engagement with Emma Watson x1 (May 2020)
stimulation of genitalia ~x100
molestation ~x170
tortured in some way or another (ex: sharp pains) ~x1000 or more
memory erasures and other mental interferences ~x1000 or more
staging a scene in public ~x10
head slammed to the ground (fortunately there was a pillow) ~x15
being crucified ~x15
being crucified while being raped ~x3
forced collapse ~x120
sore throat ~x15
sprained ankle, fractured heel ~x1
forced defecations / attempted herniations ~x25
forced defecation with forced vomiting at the same time x1
forced constipation ~x20
forced sleep ~x50
dizziness / weariness / drowsiness / fatigue ~x200
limbs forced to side (sometimes semblance sexual bondage) ~x35
fractured ribs (fall 2020) ~x2
utter agony ~15
heart convulsions / redirections of blood-flow ~x65
shortness of breath ~x100
disruption of circadian rhythm ~x50
sleep deprivation ~x2 (once 7 days, another 5 days)
mucus formation ~x75
clogged nasal passage ~x75
depreciation / disparagement / slander of my person ~x350
intellectual property theft
relationship and employment sabotage
control of bowels, bladder, brain waves, emotions, genitalia, heart rate and rhythm, sleep cycles,
ability to cause sensations throughout the body

ability to manipulate the light left in the eyes after you have closed them
tampering with accounts
mind reading (and neural monitoring)
mind games
threatening me and my loved ones
locking me up in behavioral health facilities for 10 months in their effort to cover up the abuses via a false diagnosis (February 2022 - December 2022)
etc…

5:

Resources

The Neurorights Foundation

Home (neuroethicssociety.org)

targeted evidence - Home

Information on Psychotronics

Frontiers | Human Brain/Cloud Interface (frontiersin.org)

Human Brain/Cloud Interface - PMC (nih.gov)

US investigating possible mysterious directed energy attack near White House | CNN Politics

U.S. probing suspected directed-energy attack on government personnel in Miami - POLITICO

US investigating possible 'Havana syndrome' attack near White House: CNN | The Hill

As mystery over 'Havana Syndrome' lingers, a new concern emerges (nbcnews.com)

Targeted Individuals: Now that we know it’s real, Will someone finally do something? – VT | Alternative Foreign Policy Media (veteranstoday.com)

Cyber-Torture – EU-Coallition Against Cybertorture

-see comments for more testimonies

-I have heard this also called "cybertorture"

UN warns of rise of 'cybertorture' to bypass physical ban | Torture | The Guardian

(PDF) Towards new human rights in the age of neuroscience and neurotechnology (researchgate.net)

(PDF) Cognitive liberty. A first step towards a human neuro-rights declaration | Paolo Sommaggio, Marco Mazzocca, Alessio Gerola, and Fulvio Ferro - Academia.edu

Patents (targetedmassachusetts.org)

V2K - Targeted Individuals 101 (google.com)

The Rise of Neurotechnology Calls for a Parallel Focus on Neurorights - Scientific American

TARGETED JUSTICE - Targeted Justice for Targeted Individuals

Mark Zuckerberg says Facebook of the future will be powered by telepathic thoughts | The Independent | The Independent

Patents for Mind Control Technology – Fighting Monarch

-"mind control" is simply a pop-culture sci-fi phrase for roughly the same thing; check out the patents and ignore the rest

Generative AI

From: Cindy Chen

Written: 6/17/2023

Email: [redacted]

Hello,

I am responding to the call for public input on Generative AI:

1. In an era in which convincing images, audio, and text can be generated with ease on a massive scale, how can we ensure reliable access to verifiable, trustworthy information? How can we be certain that a particular piece of media is genuinely from the claimed source**.**

I do think part of it is certainly public education that we must be much more wary of what information we consume and how we can determine if it's from a legitimate source. It would also be ideal if there were a way to see the original source for something (like somehow using blockchain but there's got to be an easy, accessible way for everyone to trace the origin of something). Perhaps as part of metadata or something.

2. How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens.

To be frank, a question we don't seem to be asking is why we're allowing this technology to run rampant and why we take such a reactive response to this technology rather than an offensive. Maybe we shouldn't even allow this technology to begin with since it serves more harm than good. Sure, it's "fun" using generative-AI to enhance photos or videos or it makes it easier to write text, but at what cost? Do the trivial benefits really outweigh the significant harms/threats that this technology serves? Is it something worth banning general public use for in terms of stopping companies and entities from using this technology for 'general use'? The innocuous applications of generative AI don't seem that important/beneficial given the malicious applications it can also do.

Companies that disseminate content like Twitter, Facebook, TikTok, and Instagram (including adult video websites like Pornhub, etc.) should have a responsibility in ensuring that this type of content is abruptly removed from their website or have safeguards in place to detect manipulated images to do so proactively so they're not even posted or shared. Honestly, I think a blanket ban would be best, because even if someone is using it for non-nefarious purposes, they're still using copyrighted material or something where someone did not give them their consent to have their likeness manipulated. It might be best to encourage large tech firms to come together as a research coalition to develop algorithms for quickly detecting and combating generative AI.

3. What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?

I also wonder if policies to limit the use of this technology would be a much more fruitful approach than trying to "live" with it. Sure, we wouldn't limit how places abroad implement it and there would still be the possibility that other people will bypass those rules and still use it, but I really think there are benefits to making it MUCH less accessible for the general public to create certain things like deepfakes and fake images.

I think we should encourage open-source software as it promotes innovation and creativity, and I don't think that is the root of the problem. Someone had . Also, I think there should be more responsible practices around training data for generative AI; like you shouldn't be able to take copyrighted or someone's art/images/videos from the Internet and then use them in your model without their permission.

4. How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise?

Perhaps there should be more in-person events or mail-related avenues for public participation. In terms of mail-related avenues, I just think of the household census where people get something in the mail with a unique code and then they can fill out a survey. Or you could solicit online participation but require a verifiable residential address on the form, and then a sample of the submissions receive something in the mail (like a postcard) that summarizes the person's participation in a certain discourse, and they must mail it back (pre-paid) with a check mark on whether they recall participating in a certain call for public opinion or if not as a way to confirm true authentic participation (as part of a vetting process) and to help understand the percentage of illegitimate input/responses. Public libraries might also be a way to drown out AI-generated noise if people who don't have a fixed address can go to submit their input and they'd have to do it in person so we know there is a real human on US soil who made those comments.

Frankly, AI has made 'old' mediums for authentic public engagement like physical mail more secure than electronic communication. Of course, that would also mean that elected representatives or their emissaries need to make the effort to host, attend, and promote events that are widely accessible and potentially localized, so that people can give their feedback in-person, as to sift through AI-generated noise.

I recognize that requiring physical interaction or attendance at things can severely limit participation, but I believe there's a tradeoff here: you either get high participation and high risk of AI-generated noise, or lower participation with lower risk of AI-generated noise.

5. How can we help everyone, including our scientific, political, industrial, and educational leaders, develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation?

The media landscape (how we've traditionally consumed news/info) has evolved where many people nowadays get their news from social media (a new AP News survey found exactly that). Accordingly, it's important that the disseminators of this information around identifying AI-generated content should include "influencers" or trusted content creators on apps like YouTube, Twitter, or TikTok. I genuinely think recruiting people who represent the 'modern' ways that the public consume news and information, such as having these content creators be part of the public education effort on identifying AI-generated misinformation, impersonation, and manipulation.

I wonder if something as simple as workshops could help people understand how to identify AI-generated malicious content. For example, part of it is understanding where the content came from, who is sharing it, and if it's a trusted source. Of course, even a methodology like that has to go through several layers, since it's also possible that a trust source accidentally fell victim to misinformation and is now disseminating false information that they earnestly thought was true.

But especially with generative AI for voice and video, where it's difficult for people to even know if they're talking to the real person, it's perhaps giving people guidelines that they should be much more hesitant about sharing or disseminating information/news that they immediately hear, for fear that it's actually false and was planted for nefarious purposes, or from divulging / sharing confidential information with anything they come across as there is a material risk that it is generative AI.

Thank you.

Cindy Chen

Generative AI

From: Yonah Welker

Written: 6/21/2023

Email: [redacted]

Dear Sir / Madam,

Addressing your request for the AI working group, we are sending our group's public letters, materials and call-to-actions, addressing challenges of Generative AI and disability-centered algorithms.

- Our recent letter addressing AI Act and risks categorization - >http://lnkd.in/dSTZGniE<
- OECD publication - http://lnkd.in/dy2XvKvh.
- Our WEF publication - Agenda

Below - an overview of our suggestions.

Abstract

*The proposed paper / suggestion covers existing issues of the AI Acts, the challenges of high and unacceptable risks systems through the lens of individuals with facial asymmetry, different gestures, gesticulation, communication styles, behavior and action pattern. In particular, people with disabilities, cognitive and sensory impairments, autism spectrum disorders. It also covers statistics addressing misuse and silos including categories of algorithms, policing and city systems, proposed actions and criteria - 6 for facilitating assistive technology and disability-centered AI systems and 8 for safety and preventing misuse, as well as audit and compliance frameworks. (*Following public letter signed by 150 EU organizations including EU Disability Forum)*

Similar to how AI systems may discriminate against people of a particular origin or skin tone, systems such as computer vision, facial recognition, speech recognition, and hiring or medical platforms may discriminate against individuals with disabilities. Facial differences or asymmetry, different gestures, gesticulation, speech impairment, or different communication styles may lead to inaccurate identification or discrimination.

For instance, Workday's AI system was alleged by an older black man with a disability who mentioned that the algorithm potentially hinders his job search. It was also reported that people with disabilities face specific and disproportionate risks from police or security systems since autonomous systems may not correctly recognize assistive devices or target individuals with mental health conditions. Other examples include speech recognition systems that can be less accurate for individuals with speech impairments, leading to misinterpretation, or automated decision-making systems used in education that may not account for the diverse learning styles and needs of students with disabilities or neurodivergent individuals.

These challenges lead us to the necessity of “disability-centered” or “neurodiversity-centered” research, development, and audit frameworks that ensure fairness, transparency, and explainability, human-centeredness, and privacy and security for these groups.

AI & Social Exclusion

Data, algorithms, machine learning, and AI systems mirror the society that created them. Historically, individuals with disabilities were excluded from the workplace, educational system, and sufficient medical support. For instance, around 50-80% of the population with disabilities are not employed full time, 50% of children with disabilities in low- and middle-income countries are still not enrolled in school, public spaces meet only 41.28% to 95% of the expectations of people with disabilities, and only 10% of the population have access to assistive technologies. More importantly, disability is not a monolith, but a spectrum, existing through the lens of intersectionality, underlying conditions, and socioeconomic criteria. For instance, between 25 and 40% of people with learning disabilities also experience mental health problems. Girls are diagnosed at a substantially lower rate (4:1 ratio) and misdiagnosed due to the different manifesting criteria or historical exclusion from the research process. Particular ethnic and social groups were historically excluded from the research and available data. In Georgia State University’s study, it was reported that Caucasian parents of autistic children were 2.61 times more likely to report any social concerns to their child’s pediatrician than African-American parents, making them predominantly excluded from the research.

AI Act and Disability

Suggestions to facilitate assistive technology and disability-centered AI systems:

- Legal status, stakeholders and caregivers - the majority of assistive devices are used by not one end-user, but caregivers
- Spectrums and comorbidity - such conditions as autism disorders present not monoliths, but spectrums. It means the necessity to reflect these spectrums and specific cases
- Accessible vocabulary - disability area actively involve language addressing terms of individuals or groups, intersectionality, medical terms
- Knowledge frameworks, adoption and curriculums - with more complexity of adoption cycle, the area of disability-system involves knowledge frameworks related to bioethics, medicine, children rights, intersectionality, educational and medical institutiions + area specific cases
- Feedback loop and assessment - type of framework addressing how developers are connected with end-users
- Technical fixes - with typical automation for bug identification and fixing, it’s important to address how potential loopholes are addressed with human-involvement

Suggestions to facilitate safety and prevent misuse

- High and unacceptable risk systems - police and autonomous security systems may falsely recognize your assistive device as a weapon or dangerous object. There is an importance of disability-specific assessment
- Low-risk systems and emotion recognition - even though emotion recognition is widely involved in the assistive technology, it’s important to stay aligned with the Convention of the disability rights
- Silos, echo chambers and social distortion - with disabled people constantly surronded with social robotics, it’s important to avoid silos and adjust human involvement

- Misuse scenarios and abuse - with social networks or platforms which may be used for the abuse or harassment. It’s an important to collect it.
- Emissions and accountability - not only actions, but non-actionsAutonomy, automation and decision-making - the mechanism should involve “double-check” principle and human involvement
- Data collection, creation and ownership - not only data input, but the creative process may happen with the use of AI, where the end used should be the rights
- Disability-specific audit and conformity assessment - addressing disability-specific aspects of fairness, explainability, and transparency.

Disability AI Bias and Audit

Algorithms do not create biases themselves but may perpetuate societal inequities and cultural prejudices. Bias can enter different development and deployment stages, including data sets, algorithms, and systems. The reasons may vary, including lack of access to data for target populations, unconscious and conscious bias from the developing team, organizational structure, and practices. As a result, algorithms may provide inaccurate predictions and outputs for certain subsets of the population or discriminate against particular groups.

The audit aims to include diverse perspectives when setting an algorithm’s purpose, evaluate disability bias in a dataset and determine how to address it, and establish disability equity-sensitive metrics and key performance indicators.

Criteria

The exact audit criteria list may vary depending on the type of target group, demography, intersectionality and socioeconomic parameters. Examples of the criteria may include:

- Representation, accessible vocabulary & accessibility frameworks. The team should involve the target population as a part of the research and resource group. Documents and communication should operate with vocabulary that properly incorporates accessibility terminology and is available for all stakeholders. The team should rely on the recent accessibility and ethics frameworks issued by specialized organizations and global documents issued by the UN, Unesco, Unicef etc
- Diverse input and stakeholders. The audit process should take into account sensory diversity (sensibility, physical, tactile and visual differences), parameters of cognition, communication, learning, and memory as well as diverse stakeholders, including individuals themselves, families, parents, caregivers, counselors and educators
- Accountability. We should identify accountable agents, actions - doing and omissions - not doing. We also should make sure that we established end-to-end answerability and auditability, where answerability stands for “who is accountable”, and auditability - “how is accountable”.
- Transparency & explainability serves the goal to make sure we avoid the “black box”, when we are not able to explain particular actions and decisions made by the system. For this purpose, we should make sure that we clearly identify the logic, semantics, social understanding, outcomes and moral justification for every action or non-action.
- Fairness allows to establish the principle of discriminatory non-harm. It means the elimination of any kind of primary or secondary discriminatory influences at the levels of the data, design, outcome and implementation.
- Feedback loop and impact metrics. The team should establish a constant feedback loop with all involved stakeholders to avoid isolated silos. Since technology may be used in different ways, we also should create and analyze scenarios of potential misuse to exclude potential harm or negative influence along with impact metrics.
- Autonomy and human decision-making. The objective of assistive solutions is to empower social integration and communication, but not replace them. For instance,

technologies such as social robotics or adaptive learning platforms are typically developed along with a curriculum that identifies parameters of interaction and learning for the child, but also involves the caregiver and educator. This means that we should identify the level of human involvement and the "double-check principle".

About speaker:

I'm a neuro disabled explorer, public evaluator and board member with a focus on social AI and robotics for disability, accessibility and neurodiversity technology. I oversee both public (EU Commission, Horizon 2020 etc) and private portfolios of technologies (MIT, Masschallenge, venture funds and programs) in the field, including my own portfolio of projects and accessibility accelerator, co-create frameworks for AI ethics, research and evaluation, future of work, organize hackathons and city projects, curate working groups, spaces and experiential projects. I"ve spent over 60 world appearances to bring awareness to the neuroexclusion crisis, as well as curated the world's largest AI Summit for the good of humanity, related space and neuromuseum.

Recent appearances:

AI Summit New York, Responsible AI, ML Con, Horasis Summit AI US, Dublin Tech Summit, RightsCon, World Humanitarian Forum, Times Higher Education, Women in AI, Women in IT, Wonder Women Tech, Women In Tech, She Loves Tech, IWIB, MozFest UK, Spark Fest Australia, ML Innovation, Datalift, Pondering AI, Digital AI Conclave, Disability Tech UK, Way Davos, Diversity In Tech Awards, AWE US, FWD 50, London Tech Advocates, Human Rights Copenhagen, Formwelt Institute etc (North America, Europe, APAC, Middle East, Australia)

--

Yours sincerely,

Yonah Welker

Explorer, Public Evaluator, Board Member

Future Of Algorithms, Research & Policy

Yonah.ai /.org, EU Commission, Women in AI

Generative AI

From: iProov Group

Written: 7/6/2023

Email [redacted]

Please find attached iProov's response to the Working Group's call for inputs on Generative AI.

If we can provide additional information in support of the Working Group's activities in this area please do not hesitate to contact me.

Kind regards,

Campbell Cowie

PCAST Working Group on Generative AI

Response to Request for Public Input

iProov 3rd July 2023

General Comment

iProov welcomes the opportunity to provide input in response to the request from the Working Group. The timing of the initiative is highly relevant with the rapid development of the public and political discourse on AI. Governments, regulators, consumer groups, and industry are each looking for frameworks within which AI can be safely harnessed for social and public purposes, as well as enabling productivity and economic growth. Addressing harms from widespread misinformation and the use of fake images, as well as other content, is increasingly a challenge for the function of an effective civil society. Whilst citizens require critical thinking skills when faced with the increasing variety of political media targeted at them, the sophistication of tools readily available to malicious actors means that technology solutions are increasingly crucial as intermediaries. In particular, consent-based digital identity verification solutions offer a route to help citizens and voters to identify often critically important content from legitimately elected representatives and genuine media. Although tackling the effects of misinformation and fake content is not straightforward, we would welcome the opportunity to contribute further to the thinking of the Working Group regarding the role of technology solutions.

Background on iProov

iProov provides mission-critical biometric face verification to protect governments and organizations from fraud. iProov ensures the highest level of identity assurance by confirming that the individual is the right person, a real person, who is authenticating in real time. Our products and processes are regularly subject to external auditing and accreditation. iProov products are certified and compliant to:

- ISO 27001 Information Security Management
- Systems and Organizations Control (SOC) 2 Type II
- European GDPR (EU) 2016/679 and the UK Data Protection Act 2018
- WCAG 2.1 AA and Section 508 (US)
- Infosec Registered Assessors Program (iRAP) - Identity Proofing Level 3
- eIDAS EN 319 401- Level of Assurance High & Quality Trust Service Level
- iBeta Level 1 and Level 2
- ISO/IEC 30107-3
- ISO/IEC 19795-1:2006

iProov is a technology innovator and market leader and is established as a trusted service provider in major markets globally. Founded in the UK in 2012, iProov is a research-led innovator with 24 registered patents. Not only is iProov a recognized world leader in identity verification, but we are also a highly successful exporter of services to major markets. iProov customers include the Australian Taxation Office, GovTech Singapore, the UK Home Office, the UK National Health Service (NHS), the U.S. Department of Homeland Security, and others. iProov was recognized by industry analyst KuppingerCole, in the Leadership Compass, Providers of Verified Identity 2022 report. In 2023 iProov was included by industry analyst, Gartner, in the Gartner Buyer's Guide for Identity Proofing report.
As the only biometric verification vendor with in-production data and intelligence through our cloud Security Operations Center (iSOC), we have been tracking the behaviors of threat actors over the last decade and have analyzed how their methodologies have evolved. We are pleased to share key learnings through this brief response and would be happy to engage in greater depth should that please the Working Group.

Response to Questions

1. In an era in which convincing images, audio, and text can be generated with ease on a massive scale, how can we ensure reliable access to verifiable, trustworthy information? How can we be certain that a particular piece of media is genuinely from the claimed source?
With the volume of AI-generated content increasing exponentially, any sort of content accreditation system may be quickly overwhelmed (or indeed gamed by AI, such that false accreditations will quickly materialize), so expectations on the likely success of such systems should be carefully managed. That does not mean that there should not be experimentation with such approaches. It may be advisable to identify and prioritize those context(s) within which false images and content may be most damaging and where the identification of false images is most technically and economically feasible. In such cases, it is likely that an automated biometric identity verification approach, based on liveness and

one-time biometrics, would be most appropriate.

With regards to determining the validity of a media presenter, as opposed to a piece of audio or text, the crucial first step is to be aware that deepfakes and digital injection attacks (whereby synthetic media is injected into the data stream, bypassing trusted device authentication) are both real and increasingly common. US Government agencies have shown leadership in their awareness of these threats.[1] The availability of generative AI through Crime-as-a-Service marketplaces means that the technical know-how barrier is evaporating, allowing less tech-savvy attackers access to online tools to create false images, including 3D face swaps, audio and text, which have also become more sophisticated to the point where they are almost impossible to discriminate with the naked eye (near-human fidelity) or ear. Publicity and media coverage further serves to drive curiosity and interest in testing the capabilities of AI both by malicious actors and by the moral majority – those who would never set out intentionally to act unlawfully.

The second challenge is detection. Once these tools are successful across certain platforms, they are then sold on to organized groups. With reduced accessibility barriers, increased tool sophistication, and a perception that tools to create deepfakes can be "fun" applications, the sheer volume of faked images and content can prove overwhelming for those trying to identify and isolate the work of malicious actors. Common attack vectors like phishing have data to understand the risk that they pose, as well as to identify an attack, but there is no equivalent for face biometrics. It's alarmingly apparent that these threats are challenging even for state-of-the-art machine-learning computer vision systems. We now need to include other complementary, multimodal approaches as imagery is becoming increasingly veridical – we can no longer solely rely on the trained human eye or even computer vision. Furthermore, privacy features on desktop and mobile devices are making it challenging to verify device authenticity, allowing attackers to conceal their identity and method of attack. Detection of a presentation attack detection (PAD), whereby a mask, photo, or video playing on a screen is presented to the camera in an attempt to create a faked image, is relatively straightforward and well-understood. Detection of injection-based attack vectors is more challenging, and more sophisticated tools are required, such as looking at the metadata or other information that comes from the device or detecting that the imagery has been synthesized or modified in some way. However, this is not straightforward, and many biometric systems are not equipped to defend against this threat vector. The most effective solutions must incorporate these sorts of liveness tools, whereby it can be determined whether the image presented is that of a real person and that the person is actually present at the time the biometric information is being captured. What's more, unlike PAD, there are no globally accepted standards pertaining to digital injection attack detection. This has given malicious actors an advantage in the biometric arms race as defenses blindly fail to keep up with novel attacks.

[1] Public-Private Analysis Exchange Program, "Increasing Threat of Deepfake Identities," https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf , 2021.

2. How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens?

Any approach will need to be multifaceted and dynamic. Policymakers and regulators will need to have a robust understanding of the economics of the threat landscape, including how barriers to participation faced by both malicious actors and the moral majority are changing. Visibility, threat monitoring, and sharing threat intelligence is essential for understanding how the market is evolving. Enforcement agencies will need to use this market intelligence to continually evolve their ability to identify and prioritize the malicious actors whose activities pose the greatest risk of harm. As awareness and understanding matures, there will be an increasing role for a standardized methodology for identifying injection attacks, such as that being developed by CEN-Cenelec.[2] Consumer education to ensure widespread awareness of fake information and false images will be a core element for promoting critical thinking amongst consumers. However, approaches that place the burden of responsibility on the consumer, although important, will only ever have a limited impact. Even to the trained eye, good synthetic media is impossible to tell apart from genuine. Policymakers must also consider the role of automated technical solutions designed to address the complex nature of threats, particularly biometric identity verification, which incorporates robust liveness, real time checks, as well as active threat monitoring . Such tools can support the ability of enforcement agencies to identify false images and other content, to be fully aware of the threats (via cloud Security Operations Center), and have visibility over the techniques used by malicious actors. Evidence and intelligence gathered using such tools can form part of the evidence chain for any actions against the malicious actors, whether that be notice-and-take down, site blocking, or prosecution. This can only be achieved through agile cloud monitoring and responsive defense deployment. For Generative AI, one-time biometrics with a SOC is required to secure mission-critical use cases. Consumer-centric tools, similarly automated, could play a role and may ultimately have to become as common as antivirus tools are today.

3. What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?

Biometric face verification technology offers a highly-convenient, secure, and inclusive approach to identifying sophisticated false images. However, as with any other verification technology, employing biometrics without insight into the evolving threat landscape may create a dangerous blind spot and amplify exposure to risk. Understanding the anatomy of a generative AI attack and the economics of how fake and false images are deployed is essential for policymakers and regulators to make decisions on what constitutes an action by a malicious actor for nefarious means. This requires the development of tools for gathering base real-world threat intelligence at speed. Threat intelligence and market awareness help to ensure that mitigation measures, including technology solutions, meet the expected level of assurance. Defending against biometric threats is no different than any other cybersecurity protocol. Organizations with high-risk use cases, especially governments and financial institutions, must deploy appropriate technologies according to the actions of individual users and current threats. As face verification technology gains traction and organizations continue to recognize its value, threat actors will utilize generative AI to create and deploy progressively more advanced attacks to circumvent the systems.

Another challenge to be addressed is the regulatory and policy framework. The framework within which any schemes are implemented is critical for securing the confidence and trust of consumers. There are many commentators who call for an ethical approach to AI and propose that this approach will address concerns about consumer harm, discrimination and exclusion.[3] Whilst a focus on ethics is advantageous for the interests of consumers and citizens, as a discipline ethics can often be too theoretical and imprecise for organizations to implement in a context specific manner with measurable impact. We are supportive of the pursuit of an ethical underpinning to behaviors, processes and product development, but prefer to view the approach as representing a north star for our approach. In practice, however, organizations need to put the needs of the end user (i.e. consumer or citizen) at the heart of what they do in developing and implementing AI solutions. Even in the absence of statutory regulation, organizations can follow some basic principles which, if properly adhered to, provide practical safeguards for end users and address barriers to inclusion and accessibility. Whilst the specific principles will best be driven by the context for each use case, we set out in the table below the key principles to which we adhere in our development and delivery of our remote digital identity verification service, based on advanced biometrics and deep AI.

Note in particular the priority given to inclusion, through both the provision of a meaningful user choice over device and by ensuring that services are easily accessible. We would also argue that it is important to minimize the burden of responsibility placed on end users. For example, end users should not be required to undertake complex upgrades or accept liability for security patches in order to engage safely with an AI solution.

[2] *Biometric Data Injection Attack Detection*, CEN/TC-224. Draft dated 2022 - 01.

[3] *A Practical Guide to Building Ethical AI*, Reid Blackman, Harvard Business Review, October 15th, 2020, available from https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai (accessed 03/07/2023).

4. How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise?

It should be feasible to build an identity verification system for elected representatives and public officials based upon verifiable credentials when working with reporters and media outlets. It should be possible for the government to verify their elected representatives online with a similar ‘closed tool’ structure. This approach could potentially help ensure the public are receiving genuine messages from their representatives. Additionally, a process of building awareness amongst elected officials of the ease of use and security benefits of biometric solutions would encourage engagement and adoption, and therefore familiarity and confidence. By way of example, the use of biometrics during secure onboarding into role and in ensuring secure access to offices and other facilities would raise awareness and understanding, helping to demystify the technology.

5. How can we help everyone, including our scientific, political, industrial, and educational leaders, develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation? (Difficult of spotting Deepfake/Gen AI tech)

Identifying deepfakes was already challenging, with many tools able to create images that would easily fool humans and many of the available technical detection solutions. The rapid growth in awareness and availability of generative AI tools, as well as their sophistication, has served to accelerate the pace at which human cognition will become unable to identify most synthetic media . The training that will be required for even the most naturally able to identify a fake will make a human solution very scarce and increasingly costly, such that it will rapidly become infeasible for more situations. Technical biometric solutions incorporating liveness, one-time biometrics and active threat intelligence are going to be the primary means by which synthetic media is identified. However, humans are not totally redundant in this environment. Whilst the right algorithm will outperform most humans at identifying a fake image, a skilled human is superior at researching, active threat monitoring, and identifying potential future threats. The best solutions will be those which build on the skills and capabilities of humans and algorithms.

More information can be found in our Biometric Threat Intelligence Report
iProov Contact: Campbell Cowie, Head of Policy. Campbell.Cowie@iproov.com
5

--

Campbell Cowie
Head of Policy, Standards & Regulatory Affairs
[redacted]
iproov.com

Generative AI

Written: 7/18/2023

Email: [redacted]

Hello,

I write to provide the attached comments from the American Statistical Association in response to PCAST's May 13 call for public input on Generative AI. Thank you for the opportunity.

Sincerely,

Steve

Steve Pierson, Ph.D.

Director of Science Policy

American Statistical Association
Promoting the Practice and Profession of Statistics®
[redacted]

www.amstat.org/policy

Attachment:

Response to the White House's Council of Advisors on Science and Technology's (PCAST) Invitation for Input for Its Working Group on Generative AI

July 18, 2023

Prepared with the expertise and guidance of these ASA entities:*
Section on Text Analysis, Section on Statistical Learning and Data Science, and *Committee on Data Science and AI*

The American Statistical Association (ASA) appreciates this opportunity to provide the President's Council of Advisors on Science and Technology (PCAST) our input for its Generative AI Working Group regarding threats posed by use of large language models to spread disinformation. The call poses five related questions:

1. In an era in which convincing images, audio, and text can be generated with ease on a massive scale, how can we ensure reliable access to verifiable, trustworthy information? How can we be certain that a particular piece of media is genuinely from the claimed source?

2. How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens?

3. What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?

4. How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise?

5. How can we help everyone, including our scientific, political, industrial, and educational leaders, develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation?

Before addressing the questions separately, this response treats the problem in general while respecting the page limit in the call. Responses specific to the questions are given at the end.

Note that in this response, the definition of “large language models” (“LLMs”) extends the current technical definition from “a computerized language model, embodied by an artificial neural network using an enormous amount of parameters … resulting in a tokenized vocabulary with a probability distribution”, while separate from techniques using neural networks with generative pre-trained transformers (GPT), includes any Deep Learning model that uses a Generative Pre-Trained Transformer (GPT) framework that results in a tokenized vocabulary with a probability distribution and is defined by statistical methods.

Ideas That May Help

There is no technology on the horizon that will stop malicious actors from telling lies, nor some people from believing those lies. Nonetheless, there are ways to mitigate the damage, and many of them involve statistics and data science. Some of the proposed approaches apply to fake news in general, not just that produced by AI. It is worth noting that “fake news” has been generated manually by human beings for many years. In this response, we are also examining two new scenarios: human beings who now use AI to create and distribute fake news, as well as fake news purely generated by AI without human intervention.

One approach to mitigating the damage caused by fake news is the creation of an integrity score for any digital artifact, such as an article, broadcast, or image. An illustrative example of an integrity score might be where articles in the New York Times or Wall Street Journal have a probability of 0.95 of being true, articles in The Times (in London) have a probability of 0.98 of being true, and articles in Politifact have a probability of 0.93 of being true (these probabilities could be created using a Bayesian method or a frequentist (quantitative) historical accuracy method). When someone wanted to quantify the accuracy of a news article, representing the extent to which the article agreed with coverage in selected benchmarks, the integrity score would be one of multiple signals that would indicate trustworthiness. Other signals might include be a score based on an algorithm similar to an Erdős number, that is, a collaboration network that may the "collaborative distance" from the source of the artifact to a large and broad number of peers that have historically high integrity scores (Newman, 2001).
One possible criticism of the integrity score approach is the creation of an “echo chamber”, that is, where the integrity score reinforces existing views and alternative ideas have a lower probability. The statistical techniques discussed above can control for this issue, using detection of originality, for example, that would mitigate an echo chamber effect (Campbell, 2020). Less directly, it seems clear that there has been a recent breakdown of comity and trust in political and other spheres. Statistical research on the causes and exacerbating factors is ongoing (e.g., Bail et al., 2018) as well as the impact of using an integrity score within such an environment, but the phenomenon demands more study and interdisciplinary research.

A second approach is taking advantage of a digital record trail that cannot be counterfeited. It would not be needed for all disseminated information, but if a consequential claim is being made, it should have a verified source. By “consequential”, we mean one that has direct and significant impact on society and has a measurable level of specificity in its assertions. By “verified”, we mean that the original source can be traced and identified with a high degree of confidence. Blockchain networks are the obvious tool for tracing content through the Internet (Xiao et al., 2020). A related approach is a digital signature whose authenticity is ensured by a hash code (Kuznetsov et al., 2018).

Outlier detection is much studied in statistics (Ben-Gal, 2005). If a news item is flagged as an outlier, its accuracy may be questionable. Techniques have already been developed that apply to digital text (Kannan et al., 2017) and to images (Marchette and Solka, 2003), although some additional work would surely be needed to adapt those methods of identifying outliers to false information sourced or disseminated by generative AI applications, and disinformation more generally.

Cluster analysis, that is, statistically grouping similar items together, would also be helpful (Jaeger and Banks, 2022). Disinformation is usually tailored to further a specific agenda. Automatic identification of groups of media posts that share a common theme enables the public to recognize coordinated efforts at deception. Some clusters will correspond to accurate news, but others will correspond to fake news. Cui and Potok (2005) developed methods for clustering documents, and Verma, Verma and Tiwari (2021) explored methods for clustering images. Little work has been done on clustering videos (Asano et al., 2020). Again, research would need to be done to tailor such methods to disinformation detection.
Adversarial risk analysis (ARA) is a research area that may be relevant to countering the spread of disinformation. ARA enables one to build a model for the decision-making of a strategic opponent, then place a subjective (Bayesian) distribution over the unknowns, and using this information, choose the action that maximizes expected utility (Rios et al., 2009). Using ARA to identify disinformation would require some knowledge of the goals of the opponent (e.g., to manipulate an election) and a subjective distribution over the opponent's capabilities (priors). Sensitivity to the assumed subjective distribution is easily explored in this setting.

There is ongoing discussion of the use of AI to recognize deepfakes created by other AIs (Salazar, 2020). Deepfakes are digitally manipulated media where the manipulation cannot be identified by human beings without technological assistance. Statisticians and computer scientists have developed methodologies that are useful for assessing the performance of such systems. To improve the classification power, a wide array of approaches exist. Statistically-based machine learning techniques such as boosting, stacking, and ensemble methods are all strategies for improving the accuracy of classifiers that distinguish deepfakes (Hastie, Tibshirani and Friedman, 2009).
Likewise, there is ongoing research of the use of AI to recognize AI-produced content, whether or not the content is considered a deep fake (Liang and Tadesse, 2022). The use of recognizing AI-produced content would allow for a standard (not a regulation) of "tagging" the content as AI-produced, signaling to the consumer additional information about the potential accuracy of the content. In addition, the tag could be considered as statistical input to the generation of an integrity score, even if the tag is not shown to the end consumer.

Approaches That Probably Will Not Work

There have been some discussions of the use of regulation requiring enforcing copyrights and placing watermarks on AI generated content, laws to prohibit deliberate introduction of AI material into public discourse, and the formation of agreements to slow the pace of AI development. We believe these are at best temporary patches. Generative AI research is an international enterprise: foreign actors will not be impeded by such measures, and domestic disruptors will find loopholes and evasions. Statistical methods such as the ones discussed above are more durable, transcend language barriers and cultures, and may evolve to keep pace of AI development.
Finally, we accept many false convictions as part of our everyday life, such as an over-emphasis on the likelihood of the statistical improbability of rare events such as airline crashes or winning the lottery

(reference Thayer, Kahneman, etc.). We may want to accept or measure the level of risk generated by innocuous, inconsequential disinformation and solve instead for disinformation that disrupts individual or collective lives.

Responses to Specific Questions

1. How can we ensure reliable access to verifiable, trustworthy information? How can we be certain that a particular piece of media is genuinely from the claimed source?
To ensure access to trustworthy information, one can assign integrity scores to news outlets, or to news anchors, or to politicians. It would reward careful digital content and warn of purveyors of disinformation.
To ensure that a particular piece of media is from the claimed source, one needs a return address that cannot be counterfeited and will work in a network of transactions. Blockchain and other systems could work.

2. How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens?
We can provide new tools, often statistical, that make it easy for citizens to assess the accuracy of a statement. Politifact offers one such method today, rooted in fact-checking using humans to perform some of the automated techniques discussed above. If properly and transparently constructed, the public may very well buy into such a tool. Other techniques, such as semantic search used with generative AI, might provide a very powerful way to implement such a tool.
For example, building a system that could query for semantically similar articles/paragraphs and then generating a referenced summary of supporting/refuting resources.

3. What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?
Machine learning methodology can be used to detect AI-generated misinformation, and perhaps human generated misinformation. But it is an arms race and the classification will not be perfect. In addition, we can build tools to enhance rather than replace a human's ability to investigate validity of claims directly - AI Augmented Human Judgement - using building blocks such as semantic search and generative AI to provide immediate access to authoritative sources to support/refute claims, making it easier to judge real from fake news.
Policies that impact international actors or clever and resourced domestic ones will be difficult to regulate. Attempting to slow the pace of research gives an advantage to potential opponents who will ignore any roadblocks. Policy development in a technologically advanced and rapidly developing environment will be an ongoing challenge.

4. How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise?
The volume of AI generated noise seems less of a problem, considering the need for curation of trusted information. Such curation entails a combination of assessing accuracy and assessing significance---a minor error of fact and malicious disinformation are both wrong, but the latter is more consequential. Elected representatives should lead the way in establishing a system for quantifying the trustworthiness of media reports, but they will need the support of statisticians, sociologists, computer scientists, and others.

5. How can we help everyone develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation?
Ironically, flooding social media, news or the public domains with disinformation may drive people to believe with greater caution what they are told. We have learned not to answer emails from “catfish”, nor to share passwords online. However, to increase skills in discriminating truth from clever AI fakes, we must create easy-to-use mechanisms that allow fact checking.

Questions can be sent to ASA Director of Science Policy, Steve Pierson: pierson@amstat.org.

References

Asano, Y., Patrick, M., Rupprecht, C., & Vedaldi, A. (2020). Labelling unlabelled videos from scratch with multi-modal self-supervision. *Advances in Neural Information Processing Systems*, *33*, 4660-4671.
Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. "Exposure to opposing views on social media can increase political polarization." *Proceedings of the National Academy of Sciences* 115, no. 37 (2018): 9216-9221.
Ben-Gal, I. (2005). Outlier detection. *Data mining and knowledge discovery handbook*, 131-146.
Berkhout, J. (2016, May). Google's PageRank algorithm for ranking nodes in general networks. In *2016 13th International Workshop on Discrete Event Systems (WODES)* (pp. 153-158). IEEE. Colin Campbell, Kirk Plangger, Sean Sands & Jan Kietzmann (2022) Preparing for an Era of Deepfakes and AI-Generated Ads: A Framework for Understanding Responses to Manipulated Advertising, Journal of Advertising, 51:1, 22-38.
Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE signal processing magazine*, *35*(1), 53-65.
Cui, X., & Potok, T. E. (2005). Document clustering analysis based on hybrid PSO+ K-means algorithm. *Journal of Computer Sciences (special issue)*, *27*, 33.
Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

Jaeger, A., & Banks, D. (2022). Cluster analysis: A modern statistical review. *Wiley Interdisciplinary Reviews: Computational Statistics*, e1597.

Kannan, R., Woo, H., Aggarwal, C. C., & Park, H. (2017, June). Outlier detection for text data. In *Proceedings of the 2017 siam international conference on data mining* (pp. 489-497). Society for Industrial and Applied Mathematics.

Kuznetsov, A., Pushkar'ov, A., Kiyan, N., & Kuznetsova, T. (2018, May). Code-based electronic digital signature. In *2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT)* (pp. 331-336). IEEE.

Liang, W., Tadesse, G.A., Ho, D. et al. Advances, challenges and opportunities in creating data for trustworthy AI. Nat Mach Intell 4, 669–677 (2022).

Marchette, D. J., & Solka, J. L. (2003). Using data images for outlier detection. *Computational Statistics & Data Analysis*, *43*(4), 541-552.

Newman ME. The structure of scientific collaboration networks. Proc Natl Acad Sci U S A. 2001 Jan 16;98(2):404-9. doi: 10.1073/pnas.98.2.404. Epub 2001 Jan 9. PMID: 11149952; PMCID: PMC14598.

Rios Insua, D., Rios, J., & Banks, D. (2009). Adversarial risk analysis. *Journal of the American Statistical Association*, *104*(486), 841-854.

Salazar, A. P. (2020). AI tools on fake news detection: An overview and comparative study. *Researchgate*.

Verma, H., Verma, D., & Tiwari, P. K. (2021). A population-based hybrid FCM-PSO algorithm for clustering analysis and segmentation of brain image. *Expert systems with applications*, *167*, 114121.

Xiao, Y., Zhang, N., Lou, W., & Hou, Y. T. (2020). A survey of distributed consensus protocols for blockchain networks. *IEEE Communications Surveys & Tutorials*, *22*(2), 1432-1465.