Written Public Comments Submitted to PCAST

July 20th, 2023 to September 1, 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.

NUTRITION COMMENTS

Responses to Nutrition blogpost: https://www.whitehouse.gov/pcast/briefing-room/2023/07/18/pcast-welcomes-public-input-on-nutrition-research/

Table of Contents

Section 1: Loraine Snead

Written: 7/28/2023

Please see my responses to the Public Input on Nutrition Research questions below.

1. How can the United States obtain the greatest return from federal investment in nutrition research? The research findings must be accessible to all people and not use researchers. A TV commercial or social media ad highlighting the findings would be helpful.

2.

    a. What are the crucial evidence gaps in nutrition research and what steps could PCAST recommend that would substantially fill those gaps? Healthy nutrition starts with young people. Public schools must play a larger role in educating youth about the relationship between nutrition and diet-related diseases like diabetes, obesity and hypertension.

    b. What tools, methods, or other resources (in addition to funding) are needed to conduct that research? Require school cafeterias to focus more on nutrition. There is no need for Salisbury steaks, cube steaks, hotdogs, and potato chips in schools.

    c. Are there other barriers to research (other than inadequate funding)? Who is doing the research? Solicit elementary teachers to do research in their schools and give stipends to do so.

    d. Are there models from other fields of science that could be employed to fill nutrition research evidence gaps?

3. How could/should research-based interventions for primary and secondary prevention of diet-related chronic diseases be introduced into federal programs? Use education

consultant groups who administer professional development to schools to disseminate the research, findings and prevention. This can be done through grants.

4. What can be done to ensure equitable access to the benefits of the federal nutrition research investment? Use education consultant groups who administer professional development to schools to disseminate the research, findings and prevention. This can be done through grants.

PCAST will consult with experts and federal agencies that are responsible for food and nutrition programs and research and would also like to hear from you.

Thank you for the opportunity!

Loraine Snead

Loraine

Section 2: Jeanne Blankenship

Written: 7/31/2023

The Academy of Nutrition and Dietetics is pleased to see the initiative led by the Biden-Harris Administration to consider nutrition research by the President's Council of Advisors on Science and Technology. We do plan to submit feedback to the group, but ask for immediate attention to the highlighted language below posted on the White House website:

"Today, poor diets and sedentary lifestyles are fueling an obesity epidemic: 42% of Americans are considered obese and face increased risks for many conditions including, but not limited to, heart disease, stroke, type 2 diabetes, many cancers, respiratory illnesses, mental health disorders, and social stigmatization that interferes with all aspects of life and can compound their risks for chronic conditions."

To reduce bias and stigma related to obesity and related disorders, we respectfully ask that you change this language which labels individuals. As you know, we would not use similar framing for other chronic diseases. Could you reword the information to say "42% of Americans are living with overweight or obesity" or "42% of Americans have a body mass index that suggests excess adiposity" or something of a similar nature?

The language that the Administration uses is important and is an opportunity to address the bias and stigma associated with obesity. This reframing is a simple step in people-first language and would set the tone for others in policy and government on this issue.

Thank you for considering this request and again, thank you for the opportunity to comment.

Jeanne

Section 3: Nadine R. Sahyoun

Written: 8/1/2023

A paper that I authored entitled **"Increasing Equity and Inclusion in Nutrition Services for Older Adults"**, has just been published in *Nutrition Today* in Vol. 58, No. 4, July/August 2023. The article recommends ways to increase equity and inclusion in federally funded nutrition programs for older adults. I hope that the article can be useful to your work.

All the best,

-Nadine Sahyoun

Nadine Sahyoun, PhD, RD
Pronouns: She/Her/Hers
Professor of Nutritional Epidemiology
Department of Nutrition and Food Science
University of Maryland
Zoom link: >https://umd.zoom.us/j/6395198363<

Section 4: Shannon Klisch

Written: 8/9/2023

1. How can the United States obtain the greatest return from federal investment in nutrition research?

    a. What are the crucial evidence gaps in nutrition research and what steps could PCAST recommend that would substantially fill those gaps?

Nutrition incentive programs for SNAP are a very promising practice for increasing consumption of fruits and vegetables. We need large scale trials of the impacts of incentivizing the purchase of produce (in all forms) from different outlets including full-service markets, farm-direct, corner stores, etc. We need to be able to educate policymakers on the impacts of investing in SNAP incentive programs and we do not have the data to do that with confidence on a broad scale, comparing multiple delivery methods, locations, populations, health equity, urban/rural, amounts, etc.

In addition, we need more research on the impacts of SNAP on diet-related illness and disease. What would happen if people had the funds needed to purchase healthy foods? What is the impact of poverty on diet-related disease? How does increasing SNAP impact healthcare expenditures now and in the future (over the lifespan)?

Research is also needed on the best way to communicate these findings to diverse stakeholders so that we can build support for these programs. What do Republican, Independent, Democrat, etc voters care about and how are these food and nutrition programs aligned with their values?

    b. What tools, methods, or other resources (in addition to funding) are needed to conduct that research?

Large scale trials and pilot projects. Access to food purchasing data with SNAP. More access to county level SNAP data especially SNAP data that relates to participants of other government programs (SSI, Medicare, etc.). How many people are on multiple programs? What is the best way to serve them and get them the resources they need? How can we improve government efficiency (also a research need) so that people can get services in one location from knowledgeable and friendly staff.

c. Are there other barriers to research (other than inadequate funding)

d. Are there models from other fields of science that could be employed to fill nutrition research evidence gaps?

I think economics researchers are looking at more systemic issues vs. nutrition research which is focused on small dietary changes or the impacts of individual nutrients.

How could/should research-based interventions for primary and secondary prevention of diet-related chronic diseases be introduced into federal programs?

There is a need for federal programs to look at themselves and their processes, overlaps, gaps, etc. We need community members involved in the design and implementation of government programs and a focus on reducing bureaucracy and program administration. So much time in SNAP, SNAP-Ed, WIC is spent on counting pennies and preventing fraud that we have gone too far and are not able to serve people who most need benefits and support.

2. What can be done to assure equitable access to the benefits of the federal nutrition research investment?

Funding for community engagement in the research process and to compensate participants for their time and expertise. Investment in community -based researchers which also includes technical assistance from experienced researchers in the fields of economics, data science, research development, etc.

Thank you,

Shannon Klisch, MPH, MCHES (she/hers)

Academic Coordinator

Youth, Families, and Communities Programs UCCE

in San Luis Obispo & Santa Barbara Counties

Section 5: V.M. (Bala) Balasubramaniam

Written: 8/13/2023

Thank you for the opportunity to comment about National Strategy on Hunger, Nutrition and Health. This is very timely and relevant effort. Here are some input and feedback about how academic R&D efforts in food processing and manufacturing technologies may help address some of these efforts. Thank you very much.

How can the United States obtain the greatest return from federal investment in nutrition research?

What are the crucial evidence gaps in nutrition research and what steps could PCAST recommend that would substantially fill those gaps?

Global population growth and changing dietary patterns highlight the importance of industrially relevant food processing and preservation methods for feeding the growing population. Climatic and environmental changes further underscore the significance of food processing and preservation methods.

The growth of the urban population and an increase in the affluent middle class have led health-conscious consumers to seek processed foods that are not only microbiologically safe but also retain health-promoting nutrients and bioactive compounds. They are aware of the critical relationship between their diet and their health and well-being. Poor nutrition promotes illnesses and increases the risk of obesity, diabetes, and heart disease, as well as broader impacts, including higher healthcare costs and decreased productivity.

Many consumers attribute the rise in various lifestyle diseases to processed foods and seek minimally processed foods with minimal or no synthetic preservatives. This necessitates the development of the next generation of food manufacturing technologies that not only ensure food safety but also preserve various health-promoting nutrients and bioactive compounds.

Addressing the aforementioned societal challenges requires the integration of multiple disciplines, including microbiology, chemistry, nutrition, physics, process engineering, materials science, toxicology, biotechnology, social science, and computer science.

Food scientists and engineers are contributing to the development of a variety of next-generation food processing technologies (such as high pressure, microwave heating, pulsed electric fields, ozone, and cold plasma) to effectively inactivate microorganisms while retaining health-promoting nutrients. Many of these technologies, in addition to preserving nutrients, can also provide sustainability benefits by reducing water and energy usage and minimizing food waste. However, inadequate funding has hindered the research development and industrial implementation of such technologies, causing the US to lag behind many other nations around the world

What tools, methods, or other resources (in addition to funding) are needed to conduct that research?

Develop scientifically sound policies that consider the interdependence among agriculture, food processing, and nutrition disciplines, aimed at converting raw agricultural and animal materials into microbiologically safe and nutritious foods enjoyed by consumers.

Mandate various federal agencies to develop extramural competitive grant programs for USA researchers working in Food Safety, Nutrition, and Health research.

For example, NSF does not have a separate program or division in food processing and preservation technologies. As a STEM discipline, food science and engineering has stronger demonstrated societal and economic importance. NSF often rejects and redirect such proposals to USDA.

NIH only funds proposals with clinical aspects of nutrition and health but does not fund research enabling processing technologies that can preserve nutrition.

The USDA's food safety, nutrition, and health program is vastly significantly underfunded. Funding for USDA food safety, technology and nutrition research is almost $1/10^{th}$ or less funding size of similar programs at NSF, NIH, or elsewhere. Private industry, particularly medium size or smaller companies do not have resources to support academic R&D. Larger corporation conduct internally R&D that may not be shared with rest of the industry.

Establish expert panels at national and state levels with multidisciplinary expertise (nutrition, food science, food engineering, medical sciences, agricultural sciences) to provide advice on various policy needs from a multidisciplinary perspective

Are there models from other fields of science that could be employed to fill nutrition research evidence gaps?

By bringing people with diverse backgrounds together, we can accomplish great things. The USA had such a vision in the 1960s when we developed a program to send a man to the moon. We need similar national vision, coordination and emphasis to advance the science of food. Such effort is needed to resolve various issues facing the society. This not only have benefit USA, but likely have a global impact in the years to come.

Thank you.

V.M. (Bala) Balasubramaniam

Professor, Food Engineering
Food Safety Engineering Laboratory, Center for Clean Food Process Technology

CFAES Department of Food Science and Technology &

Department of Food Ag Biological Engineering

Editor-in Chief, Journal of Food Process Engineering

go.osu.edu/foodsafetyeng

Section 6: Sydney Pryor

Written: 8/15/2023

Good morning,

Please find comments attached in response to the PCAST request for public input to advance equitable nutrition research from the Global Food Institute and Sumner M. Redstone Global Center for Prevention and Wellness at the George Washington University. These comments were drafted by Dr. Bill Dietz, Director of Research and Policy at GFI and Chair of the Redstone Center, and Sydney Pryor, a doctoral fellow at the Redstone Center. Thank you for the opportunity to provide input and please feel free to follow-up with questions.

--

Sydney Pryor, MPH

Fellow, Sumner M. Redstone Global Center for Prevention and Wellness

PhD Candidate, Health Policy, GW Milken Institute School of Public Health

Response to PCAST request for public input to "identify scientific opportunities, gaps, and priorities to continue to advance nutrition science, emphasizing equitable access to the benefits of research."

Ultra-Processed Foods Research Directions

Sumner M. Redstone Global Center for Prevention and Wellness

Global Food Institute

The George Washington University

Advancing equitable nutrition science and policy urgently requires a focus on ultra-processed foods, their specific components, and their association with both human and planetary health.[1]

(1) How can the US obtain the greatest return from federal investment in nutrition research?

In the US and around the world, the escalating burden of diet-related chronic disease constitutes a public health crisis and contributes significantly to the economic burden of disease. The American diet and food supply are characterized by a lack of nutrient-dense whole foods, with adults consuming nearly 60% of their calories from ultra-processed foods (UPFs).[2,3] Likewise, UPFs comprise nearly 70% of children's calories in the US.[4] Ultra-processed foods are products made from low-cost industrial ingredients with little or no whole foods, and are designed to create durable, convenient, and hyper-palatable food products, prompting overconsumption.[5] High levels of UPF consumption threaten overall dietary quality, replace the consumption of whole, nutrient-dense foods, and are associated with costly diet-related chronic disease including obesity, type 2 diabetes, and cardiovascular disease, and other adverse health effects.[6] UPF consumption displaces higher quality diets contributing to nutrition insecurity, defined as a lack of consistent access to safe, healthy, affordable foods essential to optimal health and well-being.[7] Nutrition insecurity can overlap with food insecurity, particularly in low-income populations.[8] Because UPFs often cost less than healthier, fresher options, they may be overconsumed in low-income food insecure populations that lack access to grocery stores. UPFs can also contribute to adverse environmental outcomes through industrialized crop and livestock production and processing, as well as packaging waste. Despite the adverse human and environmental health effects of UPFs, the 2015-2020 Dietary Guidelines for Americans (DGAs) failed to address UPFs. Although the scientific advisory committee for the 2025-2030 DGAs have been tasked with evaluating UPFs in relation to weight gain, there are several other aspects of these products that require further understanding to ensure the equitable health and safety of both people and the planet.[9]

Suggested action: It would be useful to fund a National Academy of Medicine (NAM) consensus study to begin to explore these issues and to set a research agenda at the federal level.

(1a) What are the crucial evidence gaps in nutrition research and what steps could PCAST recommend that would substantially fill those gaps?

Defining UPFs

Defining UPFs is essential to address dietary quality and consumption patterns through policy action and consumer outreach. Researchers often use the NOVA classification system developed by Monteiro and colleagues in 2009 to categorize foods and food products into four groups according to their degree of industrial processing: (1) unprocessed/minimally processed foods (e.g., fresh produce, dried beans), (2) processed culinary ingredients (which may add salt, fat or sugar), (3) processed foods (which use preservation methods like canning or bottling), and (4) UPFs (which use more complex processing and additives and contain little to no whole foods).[10] A common critique of the NOVA classification is the limited focus on nutrient content or individual ingredients and additives, thus giving a collective definition to foods that do not have a homogenous impact on health. The implementation and effectiveness of policy and regulatory actions, such as labeling, federal food procurement, and school nutrition standards, require a nuanced definition of UPFs, as well as a categorization process, with attention to both nutrient composition and degree of processing. Research has shown agreement between consumers' perceptions of the NOVA classification and degree of processing, but consideration of UPFs within federal guidance, such as the DGAs, would provide clear and actionable information for individuals, public health professionals, and policymakers.[11]

Suggested action*:* Facilitate collaboration between the US Food and Drug Administration (FDA), the US Department of Agriculture (USDA), and prominent nutrition research institutions to develop a comprehensive UPF definition and promote its incorporation into the DGAs.

Identifying components of UPFs that contribute to disease burden

Several additives and preservatives are used in the production of UPFs to artificially enhance the flavor, texture, color, and other characteristics of products to improve their appearance and palatability.[10] Additives are used as coloring agents, flavoring agents, emulsifiers, and sweeteners.[12] There has been an increase in the use of food additives across most food groups, including baby food, in the US over the past two decades.[12] These additives serve different purposes and it is important to recognize that not all UPF components will have similar effects on human health.

Suggested action*:* Allocate research funding through the National Institutes of Health (NIH) to clarify the use of additives by quantity, type, purpose (preservative, coloring, chemical flavoring, texturing agents) and related mechanistic studies to examine their contribution to disease burden.[13]

Understanding determinants of UPF consumption

UPFs constitute a majority of caloric intake in US adults and children. Personal factors contribute to both purchasing and consumption of these products, such as price, convenience, taste preferences, and socioeconomic status, and the addictive properties of the UPF themselves. Food environment characteristics also influence individual consumption patterns through access, availability, and affordability of UPF relative to more nutritious options. Further research is needed to identify effective behavior change interventions to increase minimally processed food consumption. Importantly, research should go beyond individual factors to address the underlying drivers of purchasing and consumption,

including external factors such as targeted advertising, marketing, and agricultural policies that facilitate the efficient production of the commodities that make UPFs cheap and widely available.

Suggested action*:* Launch a research initiative within the USDA, perhaps within the National Institute of Food and Agriculture (NIFA) and in collaboration with behavioral scientists, to investigate the determinants of UPF consumption and develop evidence-based strategies for promoting healthier food access.

Evaluating the prevalence and impact of non-nutritive sweeteners in UPF

UPFs are typically high in inexpensive ingredients, such as added sugar. As demand for foods and beverages low in added sugar has increased, industry has reformulated products to be low or no sugar products by substituting non-nutritive sweeteners. Non-nutritive sweeteners may be associated with adverse health effects. Aspartame, a non-nutritive sweetener, has recently been classified as a possible carcinogen by the International Agency for Research on Cancer (IARC). Exposure to food marketing of these products to children may increase consumption and is therefore a public health concern.

Suggested action*:* Provide NIH/FDA RFAs to further understand the extent to which non-nutritive sweeteners are replacing added sugars in UPFs and the consequent health effects among consumer groups, especially children.

Evaluating UPFs at the intersection of human and planetary health

The production of certain food items has a disproportionate impact on environmental health outcomes, including greenhouse gas emissions (GHG), land use, and water and air pollution. Animal-sourced foods and particularly red meat, contribute significantly to these negative environmental impacts of food production. Simultaneously, consumption of processed meat is associated with increased risk of numerous diet-related chronic diseases. The IARC has classified processed meat as a human carcinogen.[14] Ultra-processed meat should be of great public health concern because the level of additives in products containing meat has increased over time.[3,12] Because these products are likely to have a disproportionate burden on both human and environmental health, research should prioritize reducing their production and consumption. Similarly, ultra-processed plant-based alternatives to meat, or meat analogues, should also be of research interest as they become increasingly prevalent in the food supply.

Suggested action*:* Establish a research program within the Environmental Protection Agency (EPA) to assess the environmental impact of UPFs with a particular focus on ultra-processed meat and plant-based alternatives.

Evaluating the role of UPF packaging on adverse human health effects

UPFs can contribute to exposure to toxins from packaging related to their production, processing, distribution, and storage, an unintended and often unknown consequence of the consumption of these products. Components of food packaging often used for UPFs, such as perfluoroalkyl and polyfluoroalkyl substances (PFAS) and bisphenol A (BPA), are environmental pollutants and endocrine disruptors associated with developmental impacts.[15,16] Given the reliance on UPFs in the average American diet, the potential harms of continuous exposure to these toxins should become a research priority.

Suggested action: Establish a research program to evaluate the impact of UPF packaging on human health, with a particular focus on PFAS, perhaps through a collaboration between EPA and FDA. Implement research findings in federal regulations and guidelines.

Identifying policy pathways and barriers for shifting away from UPF production

Over 50% of energy in US adult diets comes from federally subsidized commodities.[17] Reducing UPF consumption will ultimately be coupled with a reduction in production. These decreases should be accompanied by increasing the production and consumption of healthier alternatives. Such shifts will require a multi-faceted approach with numerous stakeholder groups and cross-sector collaborations. What pathways will divert commodities away from UPFs to healthier alternatives? What alternative crops are nutritionally dense, suitable for the impact of climate change, and economically viable for farmers? Research to identify effective policies to facilitate this shift should be prioritized and consider options such as programs to promote agricultural diversity and sustainable production methods.

Suggested action*:* Mandate the use of an equitable food systems approach in the development of a research and policy agenda around UPFs and engage diverse stakeholder input all research collaborations to account for the complexities and trade-offs of UPFs throughout the US agrifood system.

(1b) What tools, methods, other resources (in addition to funding) are needed to conduct research?

Suggested action*:* Conduct interdisciplinary research including epidemiological studies to identify patterns, controlled trials to identify mechanisms, behavioral research to understand facilitators and barriers to change, environmental studies (LCAs) to understand planetary health implications of UPF production. Prioritize the development of a standardized definition of UPFs. Encourage data sharing and accessibility across disciplines.

(1c) Are there other barriers to research (other than inadequate funding)?

The 'Big Food' industry acts as a powerful barrier to ultra-processed research and policy action.[2]

Suggested action*:* Rigorously and transparently monitor conflicts of interest for individuals and organizations involved in the development and completion of this work.

(2) How could/should research-based interventions for primary and secondary prevention of diet-related chronic diseases be introduced into federal programs?

US policies addressing UPF consumption and its detrimental effects on human and environmental health are minimal, particularly at the federal level.[18] Multi-sector actions and interventions at the federal level to reduce consumption of UPF are possible and can support the prevention and potential treatment of diet-related chronic diseases and associated costs in the US.

Suggested action*:* Aligning the DGAs with evidence-based research on UPFs and incorporating a definition of UPF will provide authoritative guidance to individuals, public health professionals, and policymakers on the health effects of UPF consumption. This guidance will increase the feasibility and implementation of other potential policy actions to reduce UPF consumption. Targeted food procurement and standards within federal programs can act as levers for changing UPF consumption and related diet-related chronic diseases at scale, as the federal government purchases over $8 billion worth of food annually.[19] Labeling provides another federal lever for generating change through empowering

consumers with transparent and reliable information about their food choices. Shifting federal incentives, such as reducing subsidies to commodity crops that act as inputs to UPF, can also facilitate UPF reduction by increasing the availability, access, and affordability of minimally-processed nutrient-dense foods. Simultaneously, federal support for producers and processors must also be aligned with desired changes in production to ensure an efficient and equitable transition to healthier and more sustainable production.

(3) What can be done to assure equitable access to the benefits of the federal nutrition research investment?

Centering federal nutrition research and policy on UPFs can equitably improve dietary patterns in the US and the burden of diet-related chronic disease. Low-income communities and people of color disproportionately lack access to fresh, nutritious food and often face higher rates of diet-related chronic disease in the US. The availability and affordability of UPF relative to more healthful options contributes to these diseases. UPFs are also disproportionately marketed to communities of color, particularly children. Research and polices related to UPFs can mitigate inequities in consumption and related adverse health effects and actions that protect marginalized groups and promote equitable access to minimally-processed foods should be prioritized, including further analysis of the upstream determinants of the widespread cost and availability of UPF relative to more nutritious foods.

Suggested action*:* Involve diverse agrifood system stakeholders throughout the multiple phases of federal nutrition research related to UPFs to foster inclusive collaboration. Prioritize an equitable food systems lens in the development of a research and policy agenda to proactively consider social, economic, and political trade-offs that may influence research findings or policy implementation in a way that does not adequately consider the impact on disadvantaged groups.

References

1. Lawrence M. Ultra-processed foods: a fit-for-purpose concept for nutrition policy activities to tackle unhealthy and unsustainable diets. *Public Health Nutrition.* 2022;26(7):1384-1388.
2. Baker P, Machado P, Santos T, et al. Ultra-processed foods and the nutrition transition: Global, regional and national trends, food systems transformations and political economy drivers. *Obesity Reviews.* 2020;21.
3. Juul F, Parekh N, Martinez-Steele E, Monteiro CA, Chang VW. Ultra-processed food consumption among US adults from 2001 to 2018. *The American Journal of Clinical Nutrition.* 2022;115(1):211-221.
4. Wang L, Martínez Steele E, Du M, et al. Trends in Consumption of Ultraprocessed Foods Among US Youths Aged 2-19 Years, 1999-2018. *JAMA.* 2021;326(6):519-530.
5. Hall KD, Ayuketah A, Brychta R, et al. Ultra-Processed Diets Cause Excess Calorie Intake and Weight Gain: An Inpatient Randomized Controlled Trial of Ad Libitum Food Intake. *Cell Metab.* 2019;30(1):67-77 e63.
6. Elizabeth L, Machado P, Zinöcker M, Baker P, Lawrence M. Ultra-Processed Foods and Health Outcomes: A Narrative Review. *Nutrients.* 2020;12(7):1955.
7. Thorndike AN, Gardner CD, Kendrick KB, et al. Strengthening US Food Policies and Programs to Promote Equity in Nutrition Security: A Policy Statement From the American Heart Association. *Circulation.* 2022;145(24):e1077-e1093.

8. Leung CW, Fulay AP, Parnarouskis L, Martinez-Steele E, Gearhardt AN, Wolfson JA. Food insecurity and ultra-processed food consumption: the modifying role of participation in the Supplemental Nutrition Assistance Program (SNAP). *Am J Clin Nutr.* 2022;116(1):197-205.
9. USDA, HHS. Dietary Guidelines for Americans, 2025-2030 Scientific Questions. https://www.dietaryguidelines.gov/scientific-questions.
10. Monteiro CA, Cannon G, Levy RB, et al. Ultra-processed foods: what they are and how to identify them. *Public Health Nutr.* 2019;22(5):936-941.
11. Hässig A, Hartmann C, Sanchez-Siles L, Siegrist M. Perceived degree of food processing as a cue for perceived healthiness: The NOVA system mirrors consumers' perceptions. *Food Quality and Preference.* 2023;110:104944.
12. Dunford EK, Miles DR, Popkin B. Food Additives in Ultra-Processed Packaged Foods: An Examination of US Household Grocery Store Purchases. *Journal of the Academy of Nutrition and Dietetics.* 2023;123(6):889-901.
13. Valicente VM, Peng C-H, Pacheco KN, et al. Ultraprocessed Foods and Obesity Risk: A Critical Review of Reported Mechanisms. *Advances in Nutrition.* 2023;14(4):718-738.
14. *Red Meat and Processed Meat.* Lyon FR: © International Agency for Research on Cancer, 2018. For more information contact publications@iarc.fr.; 2018.
15. Buckley JP, Kim H, Wong E, Rebholz CM. Ultra-processed food consumption and exposure to phthalates and bisphenols in the US National Health and Nutrition Examination Survey, 2013–2014. *Environment International.* 2019;131:105057.
16. Ramírez Carnero A, Lestido-Cardama A, Vazquez Loureiro P, Barbosa-Pereira L, Rodríguez Bernaldo de Quirós A, Sendón R. Presence of Perfluoroalkyl and Polyfluoroalkyl Substances (PFAS) in Food Contact Materials (FCM) and Its Migration to Food. *Foods.* 2021;10(7).
17. Siegel KR, McKeever Bullard K, Ali MK, et al. The contribution of subsidized food commodities to total energy intake among US adults. *Public Health Nutrition.* 2015;19(8):1348-1357.
18. Pomeranz JL, Mande JR, Mozaffarian D. U.S. Policies Addressing Ultraprocessed Foods, 1980–2022. *American Journal of Preventive Medicine.* 2023.
19. Santo, R. & Silverman, J. (2023). Values-Aligned Food Purchasing and Service: Promising Examples from US Federal Agencies and Programs. Federal Good Food Purchasing Coalition. Available from: www.fedgoodfoodpurchasing.org.

Section 7: Sam Hoeffler

Written: 8/16/2023

Greetings Catherine Woteki and Francis Colón,

Please find Reinvestment Partners' response to public input on nutrition research attached.

Reinvestment Partners runs the nation's largest produce prescription program; our Eat Well program has provided more than 100,000 individuals in North Carolina with funds to purchase fruits and vegetables each month. In just the month of July 2023, our participants purchased $1.6 million worth of fruits and vegetables through our Eat Well produce prescription program. We partner with health payers and providers to implement our program and target participants.

As practitioners and advocates, Reinvestment Partners is delighted to respond to a request for public input about federal nutrition research.

With warm regards,

Sam Hoeffler

August 15, 2023

To: President's Council of Advisors on Science and Technology
Re: Call for public input to advance nutrition science

Dear Catherine Woteki and Francis Colón:

Reinvestment Partners is a non-profit based in Durham, North Carolina. We address the problems of poverty and social injustice in the areas of food, housing, community development, health, and financial services. We also advocate for policy and systems change at the local, state, and national levels. Our mission is to foster healthy and just communities by empowering people, improving places, and influencing policy.

Reinvestment Partners is deeply committed to developing programs to address health related social needs that are sustainable and compatible with the healthcare sector. We design our services to meet the needs of patients, providers, and payers. We center participant experience, making it easy and dignifying to access and receive high quality services. In addition, we ensure our processes can meet the business and regulatory requirements of health providers and health insurers.

Reinvestment Partners runs the nation's largest produce prescription program; our Eat Well program has provided more than 100,000 individuals in North Carolina with funds to purchase fruits and vegetables each month. In just the month of July 2023, our participants purchased $1.6 million worth of fruits and vegetables through our Eat Well produce prescription program. We partner with health payers and providers to implement our program and target participants.

As practitioners and advocates, Reinvestment Partners is delighted to respond to a request for public input about federal nutrition research.

How can the United States obtain the greatest return from federal investment in nutrition research?

Nutrition research has the potential to drastically recast what healthcare looks like for people across the United States. We know that people are not finding most of what they need to be healthy in doctors' offices. It is time for nutrition research to meet the moment and play a role in shaping a healthcare system that invests in wellbeing, livelihood, and connection. In particular, nutrition research must be focused on testing and improving scalable nutrition solutions that can be integrated into the healthcare sector. Nutrition researchers are notoriously out of touch with the people they study – those living with chronic illness, struggling with food insecurity, and facing other challenges. Thus, any coordinated approach to nutrition research must meaningfully incorporate on-the-ground understandings from practitioners and patients, and not simply allow research to stay in the exclusive hands of principal investigators.

The United States will obtain the greatest return from federal investment in nutrition research when it stops seeking returns on investments as a main outcome. Framing research around getting a return on investment takes the focus away from what is at stake here – the United States has woefully underfunded nutrition interventions and research for decades. Our chronic disease burden in the United States rests disproportionately on the backs of black and brown families and the poor. If the United States is going to commit to investing in nutrition research, it must be in service of those most in need and with the recognition that it simply costs money to care for people and keep them healthy.

What are the crucial evidence gaps in nutrition research and what steps could PCAST recommend that would substantially fill those gaps?

Gap: Nutrition research does not currently prioritize testing and improving scalable nutrition solutions that can be integrated into the healthcare sector.

Adults in the US are not eating the recommended amount of fruits and vegetables, and poor nutrition is associated with higher rates of chronic illness. Another important risk factor for chronic illness is food insecurity. Food insecurity incentivizes consumption of cheaper, less healthy foods. Improving food security and diet quality in the US is becoming an increasingly urgent issue, both in terms of ameliorating health outcomes and reducing healthcare costs. Produce prescriptions – especially those that can function at scale and meet the needs of the healthcare system – can play a role in solving these issues. We recommend that nutrition research focus on how to operationalize produce prescriptions to address high rates of chronic illness nationally.

Produce prescription programs benefit from widespread support and encouraging research findings.[1] First, produce prescriptions have been piloted and established across the country through government and non-profit programs. Second, they benefit from strong foundational research: We know that consumption of fruits and vegetables improves health and produce prescriptions do increase the purchase of fruits and vegetables.[2,3,4]Additionally, we know how to design produce prescriptions to maximize effectiveness.[5] Furthermore, produce prescriptions can be standardized and offered as a product across health plan member groups. Produce prescriptions are uniquely poised to take advantage of promising research, effective program design, and standardized service delivery to become a covered benefit. Innovative nutrition research can help us get there.

Gap: Studies are not designed to facilitate workable, sustainable solutions to our most pressing nutrition challenges.

Nutrition research must be used as a tool to address urgent health issues like curbing the number of people suffering from diet related chronic illnesses and reining in runaway healthcare costs. Research design is foundational to progress on this measure. More studies must be designed to test how nutrition

[1] Center for Health Law and Policy of Harvard Law School. Mainstreaming Produce Prescriptions. 2021.
[2] Jessica Marcinkevage et al. Washington State's Fruit and Vegetable Prescription Program: Improving Affordability of Healthy Foods for Low-Income Patients, 16 Prev. Chronic Dis. E91 (2019).
[3] Ashley Chrisinger & A. Wetter, Fruit and Vegetable Prescription Program: Design and Evaluation of a Program for Families of Varying Socioeconomic Status, 48 J. Nutr. Educ. & Behav. 557 (2016).
[4] Berkowitz SA, Curran N, Hoeffler S, Henderson R, Price A, Ng SW. Association of a Fruit and Vegetable Subsidy Program With Food Purchases by Individuals With Low Income in the US. JAMA Network Open. 2021.
[5] Sara John, Reece Lyerly, Parke Wilde, Eliza Dexter Cohen, Eliza Lawson, Amy Nunn, "The Case for a National SNAP Fruit and Vegetable Incentive Program", American Journal of Public Health 111, no. 1 (January 1, 2021): pp. 27-29.

interventions fare as part of a healthcare solution. Key questions that well-designed research can address include:

- Can health improvements from nutrition interventions be replicated across geographies, demographics, and payers?

We need to better understand how nutrition interventions impact health outcomes at scale. We know that consuming fruits and vegetables improve health. And we know tech-enabled produce prescriptions that take advantage of economies of scale can provide those fruits and vegetables.

How do these interventions impact target populations like people living with diabetes or people who have had a heart attack? And can research replicate those results across geographies and health insurers? How long do people need a nutrition intervention, is it indefinitely like with many medications? Narrowing in on these questions and demonstrating that the results are replicable are important steps in strengthening nutrition research and interventions.

- Can we use research to develop a cost-benefit analysis for nutrition interventions?

More important than a return on investment, cost-benefit analysis helps stakeholders understand where to allocate limited funds. Important components of these analyses would include comparing different nutrition interventions to determine which is a better investment and determining feasibility of healthcare integration.

Incorporating cost-benefit analysis into research is an important way to ensure that the research is contributing to a broader body of evidence focused on expanding access to effective interventions to all.

- How do nutrition interventions stack up to pharmaceutical interventions?

Nutrition research has not developed studies that directly compare nutrition interventions like produce prescriptions to pharmaceutical interventions. This is an important research gap because we do not know whether cheaper, more whole-health oriented interventions like produce prescriptions could actually replace medications. Nor do we know whether nutrition interventions could result in a lower dosage of medications. Or, perhaps patients may be able to start with a medication and transition to a nutrition intervention and maintain the same health outcomes. We need more nutrition research that tests whether nutrition interventions can substitute, complement, or otherwise relate to medications.

What tools, methods, or other resources (in addition to funding) are needed to conduct that research?

Use technology to keep research costs low and ensure study directly benefits participants.

Research can be done in a way that includes communities and ensures that those communities directly benefit from the research. One way to accomplish this is to leverage technology. We partnered with Duke University Health System to design and launch a Randomized Control Trial to evaluate Eat Well, our produce prescription program, for patients with diabetes and at-risk for food insecurity.

We used technology to keep the cost of the RCT low and ensure that most of our budget went to study participants. We offered participants $80 per month to purchase fruits and vegetables with a prepaid debit card at participating grocery retailers. At its core, this approach is built on the goal of creating a positive and enriching patient experience. We made it easy to enroll and easy to spend funds. We used existing data at the provider level to identify eligible study participants. We used the eligibility list generated by Duke University Health System to contact eligible patients via text message. Using an eligibility list and text message outreach, we enrolled more than 1,000 people in one month. Using technology and data for recruitment, we minimized study expenses and allocated our full budget to paying for food for study participants ($960,000).

As practitioners, we worked alongside DUHS to design the study and ensure that patient experience was at the center of the RCT. We also took advantage of their data access and our tech-enabled outreach to keep costs low. This RCT demonstrates how innovative nutrition research can be collaborative with practitioners and also ensure that the study improves the livelihoods of study participants.

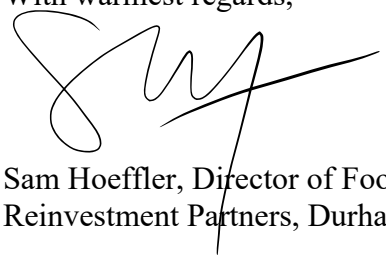
What can be done to assure equitable access to the benefits of the federal nutrition research investment?

To date, much of the food is medicine work in the United States is concentrated in the northeast and in California. Although there is movement all over the US, including in rural and southern states, most of the funding is allocated to New England and the west coast. Similarly, nutrition research is concentrated in the northeast and west coast. We recommend identifying academic institutions outside of these areas of wealth and influence as a first step to achieving some balance in research dollars.

It is worth repeating that it is critically important to ensure that practitioners and participants are meaningfully included in the design, structure, and implementation of nutrition research projects. For far too long, across all research disciplines, participants are an afterthought and researchers are not in touch with their needs. Furthermore, very often these research projects do not offer direct or sustained benefits to study participants. We must fund nutrition research to test interventions with an eye toward scale and sustainability. If nutrition research ultimately results in rolling out nutrition interventions that help people with chronic illnesses at scale, then the United States will have been successful in its endeavor.

Thank you for sharing your request for public input. Please do not hesitate to contact us if you would like any additional information about Reinvestment Partners, our advocacy work, or Eat Well.

With warmest regards,



Sam Hoeffler, Director of Food Programs
Reinvestment Partners, Durham NC

Section 8: Connie Weaver

Written: 8/17/2023

What do we mean by strengthening federal nutrition research?

An *ad hoc* committee of the National Academies of Sciences, Engineering, and Medicine (the National Academies) discussed priorities for strengthening federal nutrition research. The committee considered the most pressing areas of needed research to overcome the current barriers to understanding the role of diet and nutrition in health and disease. Nutrition is vital to health, growth, and development throughout the lifespan. However, it is challenging to study the role of nutrition in health and

disease processes because the human diet is a complex mixture of foods and nutrients, and because its impact is incremental and its role difficult to assess in long latency diseases. Also, dietary components can play roles in initiation and/or progression of diseases and can interact with other environmental and lifestyle factors. Additionally, the benefit of nutrition extends beyond known nutrients to include bioactive components in foods, which are incompletely characterized.

Diet-related health burdens of the nation and its corresponding economic, health equity, national security and sustainability implications overburden local and national budgets. Yet, the evidence base remains poor to establish diet guidance. Following 50 years of nutritional studies, we are still not in a strong position to advise Americans concerning what they should eat for health and longevity. A transformation is called for to achieve breakthroughs needed to build a stronger evidence base and to coordinate efforts for efficiency and progress.

Rationale:

The time is right to take a giant leap forward in strengthening federal nutrition research. The pandemic has exposed the extra burden to the many Americans in poor nutritional health, disparities in access to nutrition knowledge, and gaps in access to nutritious food. A scoping paper called for a new federal authority to coordinate nutrition funding Fleischhacker SE, Woteki, CE, Coates PM, Hubbard VS, Flaherty GE, Glickman DR, Harkin TR, Kessler D, Li WW, Loscalzo J, Parekh A, Rowe S, Stover PJ, Tagtow A, Yun AJ, and Mozaffarian D. Strengthening national nutrition research: rationale and options for a new coordinated federal research effort and authority. *American Journal of Clinical Nutrition* 2020; 112(3): 721-769. We support a means to identify areas of nutrition research in need of strengthening in order to provide public health recommendations to improve the health of Americans and to lower health care costs such as a consensus conference.

Priority areas of need that should be addressed include:

1. **Improving data integrity, rigor, and reproducibility**

Research rigor is essential to provide high quality and reliable contributions to the knowledge base, which forms the bedrock for the development of nutrition policy and guidance. Nutrition research has often suffered from lack of rigor demanded in other fields, lack of bona fide validity studies, and should be improved to build more confidence in scientific knowledge that informs public health dietary guidance.

2. **Assessing diet exposure**

We have very limited ability with standard dietary data assessment methods, whether based on food frequencies, food records, or dietary recalls, to assess more detailed total energy and diet intakes. For example, studies using doubly-labeled water for short-term energy assessment show energy underestimation by about 30-40% among overweight and obese individuals, though much less among normal weight persons. Systematic biases of this magnitude, if uncorrected, can play havoc with energy intake and disease association analyses. It is likely that ratios of dietary intakes, whether for foods, food groups, or nutrients, are better estimated by available assessment tools but these ratio estimators have mostly not been validated with corresponding objective intake measures. Few other nutrients or diet components or patterns have objective methods available for determining intake.

Needed is the ability to objectively measure dietary intakes and the ability to correct measurement biases in diet and disease association studies using related intake biomarkers. Metabolomic signatures, big data, and artificial intelligence may help to identify biomarkers of exposure and of disease processes. Research on assessing diet exposure, nutritional status, and impact on health should be considered from a lifespan perspective. The evidence base for the very young and very old are especially inadequate.

Nutrition research should take advantage of the latest technologies and approaches being developed to tackle the pressing questions. An emerging science that needs to be applied to nutrition is adding systems approaches to analytical approaches in order to understand the interactions and pathways (e.g., biological, behavioral, social, and environmental) involved in the complex interactions of diet and health, diet and weight, weight and chronic disease occurrence. Systems methods can elucidate the dynamic behavior of a system and can help generate hypotheses to explain why a system acts in certain ways.

Systems approaches were advocated to be applied to development of the Dietary Guidelines for Americans (National Academies of Sciences, Engineering, and Medicine 2017. *Redesigning the Process for Establishing the Dietary Guidelines for Americans*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24883). Nutrition as a field has attempted to integrate systems for glucose control, but for little else. The current efforts in Precision Nutrition are a start but adding diet assessment to a program (*All of Us*) not designed for nutrition leaves many gaps. Further, a greater understanding of the mechanisms that relate dietary intake with health outcomes can come from understanding the role of diet on the gut microbiome, exosomes that act as

cell-to-cell communicators, epigenetic modifications, and metabolic profiles of chronic disease.

Inherent in any effort is the need to be mindful of nutrition equity and translation of science to practice. Refinement of the priorities would be the goal of a consensus conference.

Section 9: Kristen Hicks-Roof

Written: 8/21/2023

Hello President's Council of Advisors on Science and Technology (PCAST) Members –
On behalf of the tens of thousands of American pig producers and farmers, and the health and wellness team of the National Pork Board (NPB), we appreciate the opportunity to provide comments for your consideration to identify scientific opportunities, gaps, and priorities to continue to advance nutrition science, emphasizing equitable access to the benefits of research.

Please find responses to the various questions posed by the Council below my signature line.

NPB is a Checkoff program and thus are prohibited from influencing government policy or action. As the Director of Nutrition Research, the comments are addressed to provide research-based comments and not to influence policy or action.

NPB looks forward to the possibility of present these ideas to the working group as part of the evolving process to develop recommendations.

Kristen Hicks-Roof, PhD, RDN, LDN

Director, Nutrition Research

National Pork Board

1. **How can the United States obtain the greatest return from federal investment in nutrition research?**
    a. What are the crucial evidence gaps in nutrition research and what steps could PCAST recommend that would substantially fill those gaps?

Modelling Work and the Thrifty Food Plan

Under Pillar 1 of the Biden-Harris Administration's National Strategy on Hunger, Nutrition, and Health, helping more individual experiencing food insecurity benefit from federal assistance programs is critical. This includes expanding SNAP eligibility to additional underserved populations and making it easier for eligible individuals to access SNAP.

One component for PCAST could consider is how modeling work can help fill these gaps in better understanding the relationship between food affordability, culturally appropriate foods, SNAP access and eligibility, as these are contextual factors in the social determinants of health which can influence dietary patterns and ability to purchase and cook healthful meals, and thus impact health outcomes. For example, in relation to the role of fresh lean pork in supporting nutrient adequacy or health outcomes, recent modeling work was completed to show how fresh pork is an affordable protein in the USDA Thrifty Food Plan – which sets SNAP allotments.

One of the foods that embodies all aspects of affordability, nutrition, and cultural significance – including in the context of the USDA Thrifty Food plan and SNAP is pork. In fact, pork is one of the most widely eaten meats in the world. There is research to support that there is no socioeconomic gradient when it comes to who is eating pork and there are numerous culturally specific pork recipes celebrated by individuals both here in the U.S. and abroad, making pork a key source of protein and nutrients in the global food culture. By including lean cuts of unprocessed pork as part of achieving the National Strategy on Hunger, Nutrition and Health, Americans have the opportunity to feel good about their chance to choose an affordable protein that also promotes social and cultural connection.

Lean pork is affordable and nutritious, giving more individuals access to an affordable, healthy diet. In fact, recent research showed that an optimization model of the USDA Thrifty Food Plan preferentially selected fresh lean pork to arrive at the lowest-cost healthy diets that met all nutrient requirements, followed dietary guidance, and respected existing eating habits while reducing weekly food costs. Researchers concluded that fresh pork is an affordable protein in the USDA Thrifty Food Plan.

Now, NPB plans to use modeling analyses to fill gaps in creating the first culturally appropriate Thrifty Food Plan. The first Hispanic Thrifty Food Plan 2023 (H-TFP 2023) will use the same methods and data inputs that the USDA CNPP had used for the TFP 2021 revision. Researchers will undertake an NHANES modeling study that will create culturally relevant food plans for a Hispanic population, based on existing diets as captured in the NHANES surveys. Participants identified as Mexican American or Hispanic will be stratified by gender and age to include children, adolescents and young, middle aged and older adults. The goal is to evaluate the role of fresh pork in nutrient adequate, healthy, culturally acceptable, and budget-conscious H-TFP 2023.

We can provide an update on this study to PCAST as needed and would be happy to connect PCAST to the PIs working on this study as warranted.

Food is Medicine

Food is Medicine (FIM) has the potential to offer every American the opportunity to look at developing a positive relationship with food, with the core focusing on how every food can fit into their dietary pattern (omitting foods only due to dietary allergy, preference, religious or cultural reasons). FIM is a fundamental are to close research gaps within the National Strategy Pillar 2.

High priority research items to close evidence gaps in the context of FIM include:

- Using various research design methods (quantitative and qualitative), explore healthcare professionals' education and training on nutrition and the role of the registered dietitian as part of the healthcare team to help facilitate FIM interventions.
- Defining ‘food is medicine’ and how it is different from ‘culinary medicine’ or ‘food as medicine’ or ‘food for health’.
- Conducting randomized controlled clinical trials and intervention studies that are inclusive to all foods. Not favoring or omitting a particular food group, but highlighting that cultural preference, affordability and variety are cornerstones to how ‘food is medicine’ is discussed with patients. For example, on the global scale, it is known that animal proteins offer a unique nutritional value. And when animal proteins like pork are omitted from the diet, there are unintended

nutrient consequences when pork is removed from the diet or when pork is not on the plate – and this spans the lifecycle. Researchers suggest that pork intake was estimated with over 2.5 million more children and adolescents (7%) and over 5.7 million more adults (4%) meeting the adequate daily intake levels for potassium. This is important as potassium is a nutrient of public health importance that most people are not eating enough of and can be core considerations in any FIM intervention.

Opportunities in FIM Research:

- Advocate for the role of the registered dietitian as part of the healthcare team to facilitate FIM interventions.
- Highlight that nutrition should be cornerstone to every patient care interaction.
- For the first time, demonstrate how all foods can be incorporated into a dietary pattern and not have a good/bad categorization of food items in the context of FIM. It is important to meet the patient where they are in terms of care, including as it relates to supporting the role all types of foods play in enhancing positive health outcomes when delivered as part of FIM.

2. **What tools, methods, or other resources (in addition to funding) are needed to conduct that research?**

To empower all consumers to make and have access to healthy choices in line with Pillar 3, as well as to enhance nutrition and food security research in the U.S., it is important to understand equity, access, and disparity among the many populations in America. Mexican Americans are the largest Hispanic ethnic group in the U.S. Data from the U.S. Census shows that Hispanic individuals are more likely to live with nutrition-related diseases such as cancer, heart disease, and diabetes. In fact, in 2017 and 2018, the rate of diabetes among Mexican Americans (14.4%) was nearly double that of non-Hispanic White Americans (7.5%). Diet plays a crucial role in these higher rates of chronic diseases among Mexican Americans. Yet, few research studies have focused on Mexican American dietary intake, possibly because there are currently a limited amount culturally relevant tools to assess the diet of Mexican Americans. Some researchers note problems with the methodologies for dietary intake data collection efforts in surveys such as NHANES where a greater focus on cultural foods may become a priority.

PCAST could choose to focus on research designing food frequency questionnaires (FFQ) that may be more culturally appropriate so that future dietary recommendations can consider appropriate cultural foodways and those involved in implementing the National Strategy can successfully monitor diet changes among Mexican Americans among other populations. Using the example of pork's cultural heritage in the diet of Mexican Americans, to our knowledge, there are limited tools that assesses the diet of Mexican Americans, yet none which directly address how this population is interacting with different pork cuts or preparation styles.

NPB is working with researchers to develop a culturally relevant FFQ for Mexican Americans. By adapting a traditional FFQ, researchers can include questions about the traditional Mexican diet that would enable stakeholders to evaluate nutrition interventions. Compared to other short-term members, an FFQ can be quickly administered and is better at obtaining long-term intake and episodically consumed foods when culturally tailored to a population. Additionally, a tailored FFQ can capture specific cuts of pork and other traditional foods consumed when assessing total dietary intake and can be used to evaluate the effectives of dietary changes such as increasing culturally relevant pork cuts in mainstream stores.

We are available should PCAST like to be connected to the researcher working on this project for NPB.

3. **What can be done to assure equitable access to the benefits of the federal nutrition research investment?**

Interprofessional collaboration is one way to assure equitable access to the benefits of the federal nutrition research investment. There is a lack of nutrition education among Health care professionals in the US, which has left many HCPs with a gap in knowledge regarding nutrition. This could be one barrier to implement the federal nutrition research investment especially as it relates to underserved and minority populations.

Registered dietitians undergo years of training to be able to provide Medical Nutrition Therapy, counseling and more to patients. Unfortunately, there is not enough HCP awareness on how to work with a dietitian, and many people do not have access to dietitians, which leads other HCPs to be the face of nutrition recommendations in some communities that may not account for meeting patients where they are in terms of personal, budgetary and culturally tailored approaches to nutrition counseling. Furthermore, work could be done to strengthen and diversify the nutrition workforce, meeting the call in the National Strategy to ensure patients do not perceive ethnic or social differences with their health care providers.

By incorporating nutrition education into medical school curriculum and working with dietitians to expand their reach to areas with limited care or as a proxy to other HCPs within these areas, will ensure more equitable care and access to outcomes of the federal nutrition research investment.

Please see the following resources from NPB staff on what can be done to assure equitable access to the benefits of the federal nutrition research investment in terms of interprofessional collaboration between HCPs and dietitians:

- How might enhanced interprofessional collaboration between primary care physicians and registered dietitian nutritionists impact clinical outcomes related to obesity and associated illnesses? A commentary
- Nutrition Education for Providers is Limited: It is Time for Increased Education to Boost Interprofessional Collaboration!
- A Mentoring Program Builds the Bridge with Nutrition Students and Healthcare Professionals

Section 10: Sarah L. Booth

Written: 8/22/2023

Dear Members of the President's Council of Advisors on Science and Technology (PCAST),
Attached is the collective response by Dr. Stephen Kritchevsky and Dr. Sarah Booth, Center Directors, on behalf of our research teams at Wake Forest School of Medicine Sticht Center for Healthy Aging and Alzheimer's Prevention and the Jean Mayer USDA Human Nutrition Research Center on Aging (HNRCA) at Tufts University.

We thank you for the opportunity to provide input.

Stephen B. Kritchevsky, PhD

Toby R. Alligood, MD Endowed Professor in Geroscience

Department of Internal Medicine: Gerontology & Geriatric Medicine

The Sticht Center for Healthy Aging and Alzheimer's Prevention

Wake Forest University School of Medicine

Sarah L. Booth, Ph.D.

Center Director and Senior Scientist

Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University

Tufts UNIVERSITY
JEAN MAYER USDA HUMAN NUTRITION RESEARCH CENTER ON AGING HNRCA
Wake Forest University School of Medicine
Sticht Center for Healthy Aging and Alzheimer's Prevention

Title of RFI: PCAST Call for Input on Nutrition Research
Name: Dr. Stephen Kritchevsky and Dr. Sarah Booth, Center Directors, on behalf of our research teams
Institutions: Wake Forest School of Medicine Sticht Center for Healthy Aging and Alzheimer's Prevention and the Jean Mayer USDA Human Nutrition Research Center on Aging (HNRCA) at Tufts University

I. Introduction to the Response:

The Wake Forest School of Medicine Sticht Center for Healthy Aging and Alzheimer's Prevention and the Jean Mayer USDA Human Nutrition Research Center on Aging (HNRCA) at Tufts University respectfully submit a response to the PCAST call for input on nutrition research. The Sticht Center's mission is to promote the health and independence of older adults. Over its 35 years, it has rigorously evaluated nutritional strategies to improve the health and well-being of older adults including energy balance, obesity treatment, healthful dietary patterns, and optimal micronutrient intake. The HNRCA's mission is to promote healthy aging through nutrition science to empower people seeking to enjoy long, active and independent lives. For more than 40 years, investigators at the HNRCA have engaged in research from the cell-level to large-scale human studies. The HNRCA has been a leader in promoting healthy aging by tailoring nutritional and physical activity guidance based on food science, basic biology of aging, genetics, epigenetics and lifestyle. As such, this call for input is pertinent and critical to the translational research we conduct with our focus on interventions that can prevent or delay the progression of age-related chronic diseases that allow people to maintain good health longer and live independently.

Nutrition is especially relevant to older adults, a large cohort of the U.S. population that has been understudied and may represent the most heterogeneous of population subgroups. We commend the PCAST for recognizing that food and nutrition are modifiable determinants of many diseases and further research can play an important role in advancing years spent in good health and result in more scalable and actionable approaches to health that will improve the quality of the aging process. However, the problems associated with an aging society transcends the boundaries of any specific discipline and play out across multiple biologic and societal domains, ranging from individual cells to organs and organ systems, to persons, to communities, to national and world economies. Therefore, we strongly advocate for cross-disciplinary approaches to reveal synergies and insights that will achieve our goals of extending good health in the U.S. population. We greatly appreciate this opportunity to respond to this important notice.

II. Respond to the queries posed in the Call for Input

A. What are critical evidence gaps in nutrition research and what steps could PCAST recommend that would substantially fill those gaps?

We would like to note the following critical gaps:

How do we define older adults? Older adulthood can span four or more decades. Therefore, it is necessary to further classify older adults by decade or a similar parameter, to properly investigate the role of food and nutrition in promotion of healthspan. Otherwise, we are assuming that research on a 65 year old can be extrapolated to that of a centenarian. That is an untested assumption and until we refine our categories of age,

we cannot make such assumptions. Furthermore, these classifications should be consistent with other disciplines and federal guidance.

There is inadequate age-specific evidence-based nutrition guidance for older adults, which can range from age 65 to older than 100 years. Nutritional needs and intakes of older adults, including micro and macronutrient content, are not well characterized, especially in the older old (defined as 85+ years old). Federally-funded population-based studies that include older adults often lack dietary assessment measures, and of those that do, they only have one baseline measure of diet. It is not known how diet patterns change over age and a single snapshot of dietary intake used to reflect a lifetime exposure of dietary intakes introduces many assumptions that limit the strength of the science. One short-term strategy to accelerate research on nutrition intakes in older adults is to add multiple dietary assessment measures in existing and ongoing federally-funded population-based studies. Novel techniques are now available that leverage digital measures that do not rely on self-report or recall, both of which may present challenges in an aging population that may have some cognitive challenges.

Age and sex have been under-represented variables in federally-funded studies. The 85yo+ population is the fastest growing demographic; it is also becoming a more diverse population which requires more complex tools to identify the cultural context of food choice and the socioeconomic determinants of food choice. Future studies need to maximize inclusion of currently understudies variables, including but not limited to age, sex and race/ethnicity.

Accordingly, the extent to which the **energy needs of older less mobile individuals** (especially women) are poorly understood, nutrition policy could be suboptimal.

In many nutritional intervention studies, **participants are typically selected based on specific criteria that exclude people with co-morbidities,** such as CVD, diabetes, and CKD, or with regular medication use, such as statins. Such studies are likely biased in favor of select participants with healthy phenotypes limiting their applicability to the majority of older adults who manage at least one, if not more, comorbidities and are taking multiple medications. This challenges the generalizability of the research to the U.S. population.

Most well-controlled dietary interventions are short-term (both exposure and washout periods), are not standardized for all food components, and involve small sample sizes. A recent example was the NIH-funded MIND study that concluded that the sample sizes were too small and the duration too short to determine the effect of consuming a healthy MIND diet pattern on cognitive outcomes. We would recommend research that spans beyond the traditional 5-year grant mechanisms of federal funding to be able to study the effects of diet on delayed progression of chronic diseases.

Many nutrition studies rely on clinical biomarkers (i.e., lipid panel) that are not deep enough to reveal mechanisms. Lipid panels cannot merely include the measurements currently used in the clinic and most cohorts. It is paramount to use much more in-depth phenotyping, such as, but not limited to metabolomics, proteomics or transcriptomics.

Biorepositories of existing studies may not have been collected for proper use with novel, highly sensitive techniques, such as metabolomics. There is also little knowledge regarding stability during storage. We propose that sample collection (blood, serum, plasma, urine, stools) must be collected using highly standardized protocols that enhance rigor and reproducibility (i.e., inflammatory markers may depend on venipuncture; the

use of alcohol swabs during venipuncture can invalidate specific omics techniques, certain appetite/satiety hormones must be collected in specially treated tubes, high throughput platforms measuring omics require serum or specific anticoagulants). Microbiome measures are highly sensitive to the type of collection device, preservation technique (including temperature), transport and sequencing technology. Rigorous stability testing of the various matrices is required to ensure accuracy of individual measures, especially for biorepositories that have long-term storage of samples.

Timing of a meal affects postprandial response and interacts with age. This needs to be taken into consideration along with food composition because current research on food timing with respect to food and beverage intake and circadian clock suggest the optimal timing of food intake is different in the older population, which may contrast with current trends/recommendations for individuals in younger age groups.

B. **What tools, methods or other resources are needed to conduct that research?**

Accurate determination of nutrient/food intake is the greatest challenge when studying nutrition and health. Areas of influence over dietary responses have been too narrowly focused on individual nutrients and/or dietary patterns using dietary assessment tools based on recall. The greatest need is developing a fast, accurate, inexpensive and easy-to-use tool that captures food intake. The use of non-invasive sensors to collect and assess food intake would be a game-changer, especially in older adults who may experience cognitive and physical challenges. Digital measures of diet patterns such as photographs and photovoice for dietary assessment can also be leveraged for guidance in diverse populations with language, motor, visual and/or auditory challenges and/or to research cultural foodways.

There is also a need to develop many other non-invasive sensors to monitor factors objectively ranging from continuous glucose levels to sleep and the many other environmental, psychological and social factors that influence an individual's metabolic and cognitive health.

C. **Are there other barriers to research?**

How to identify the most **appropriate animal model** for study of human aging? For example, mice to do not reflect many aspects of human aging, and there is question of how translatable insights from this model might be. There is also a **lack of rigor and reproducibility for diets** used in animal models of aging.

We need to identify **the critical windows of opportunity across the lifecycle** at which nutrition can have a positive impact on healthspan. Different aspects of nutrition are likely to different depending on life stage.

Data Integration and Analysis: The heterogeneity of older adults means that large datasets are often required to detect meaningful patterns. Analyzing and interpreting these data require sophisticated statistical tools and expertise, which may not always be readily accessible, and this should be extended nowadays to AI/ML

Ethical Considerations: With older adults, especially those in the 85+ age group, there are often ethical considerations related to consent, especially if cognitive decline is present. This can complicate study designs and lead to potential biases in participant selection.

Recruitment and Retention: Older adults, particularly the "older old", may have mobility issues or health conditions that make consistent participation in research studies challenging. Ensuring they remain engaged and available for long-term follow-ups is crucial but can be challenging.

Lack of Standardization in Outcome Measures: Different studies may employ various endpoints or outcome measures, making it difficult to compare or aggregate results across studies. For example, while one study may look at the impact of nutrition on cognitive decline using one set of metrics, another might use entirely different criteria, complicating meta-analyses.

External Validity Concerns: Given the diversity within the older adult population, findings from one subset (e.g., urban older adults of a particular socioeconomic status) might not be generalizable to other groups (e.g., rural older adults or those from different cultural backgrounds).

Influence of Socioeconomic and Cultural Factors: Nutrition is not just about what is available, but also what is culturally acceptable and economically feasible. Socioeconomic disparities can significantly impact dietary choices, and understanding these within the context of aging requires a nuanced approach.

D. **Are there models from other fields of science that could be employed to fill nutrition evidence gaps?**

Interdisciplinary Collaboration: Effective nutrition research, especially as it pertains to aging, requires the integration of diverse fields like gerontology, genomics, epidemiology, anthropology, economics, and psychology. The compartmentalized **nature of academic and research institutions can sometimes stalemate such interdisciplinary efforts.**

Integrating Chronobiology with Nutrition: An emerging field, chrononutrition, emphasizes the synchronization of food intake with our internal clocks, impacting digestion, metabolism, and overall health. The development of an application that facilitates a metabolic stress test, which looks at pre- and postprandial effects in conjunction with other physiological measures like sleep and physical activity, is not just novel but necessary. It's a reflection of the intricate interplay between our biological rhythms and our nutritional intake.

Importance of Microbiota: The gut microbiome has been recognized as a crucial player in our health. Not only does it influence digestion, but it also has profound effects on immunity, brain function, and even behavior. Recognizing that changes in the microbiome in adults are subtle underscores the need for sophisticated analysis methods. As diet impacts this microbiota, understanding its dynamic nature and its relationship with nutrition is paramount.

Technological Integration: The details in the large datasets that arise from studying the microbiome, especially in the context of diet, cannot be deciphered using traditional analytical methods alone. AI/ML have the capability to identify patterns and correlations in vast and complex datasets. Its application in this context can unveil insights that were previously elusive.

Innovation in Monitoring: Nutrition is not just about what we eat but also about when we eat, how our body processes it, and how it affects our overall physiological equilibrium. Non-invasive sensors, as mentioned, can revolutionize the way we gauge dietary impacts. These could range from sensors that assess nutrient levels in real-time to those that track the activity of the gut microbiome post a meal.

Incorporating these dimensions, a comprehensive metabolic stress test application can be envisioned to provide real-time feedback on nutritional health, thereby empowering individuals to make informed dietary decisions. Such a tool, fortified with AI and ML capabilities, can provide dynamic dietary recommendations based on individual metabolic responses and circadian rhythms. Additionally, partnering with experts in other fields of science such as data science, chronobiology, and microbiology would be instrumental in building such an application. Collaboration between nutritionists, data scientists, and technologists will be key to turning this vision into a reality.

Section 11: Frank B. Hu

Written: 8/26/2023

Attached please find some ideas about Nutrition Research Gaps and Priorities. Thank you for considering these ideas.

Sincerely

Frank Hu

Attachment:

August 26, 2023

Nutrition research gaps and priorities

Poor diet is widely recognized as a significant contributor to various morbidities, shortened life spans, negative environmental consequences, and health inequalities. Nutrition research plays an essential role in formulating evidence-based strategies and polices aimed at improving human well-being, promoting environmental sustainability, and addressing health disparities. Despite recent investments in nutrition research and advances in technologies, many gaps and challenges persist, impeding the advancement of both research and practical applications in nutrition. These barriers hinder our progress toward achieving nutrition security, reducing diet-related chronic diseases, and safeguarding the health of the planet. Below is an overview of several key research gaps that my colleagues and I believe require more urgent attention.

1. Enhancing Dietary Assessment Methodology in Diverse Populations

Accurately assessing diet in free-living individuals remains among the most difficult aspects of nutrition research due to the complexity and multifaceted nature of human diets. Commonly used dietary assessment methods such as food frequency questionnaires (FFQs) and multiple 24-hour recalls have been valuable in measuring diet in large populations, but they are prone to measurement errors. This challenge is particularly amplified in minority and immigrant groups due to language barriers, cultural differences, and religious practices as these tools might not adequately capture the diversity of foods consumed by some minority and immigrant groups with distinct eating habits. Because these groups are frequently underrepresented in dietary

assessment research, standardized dietary assessment tools have not been well-validated for them.

To bridge this gap, it's crucial to invest in developing and validating dietary assessment tools suitable for diverse populations. While traditional recovery and concentration biomarkers continue to be useful in providing objective biomarkers for intake of various nutrients, recent advances in cutting-edge omics technologies, notably, metabolomics hold promise in uncovering novel biomarkers that more accurately reflect intake of specific foods. In addition, the utilization of mobile applications featuring image recognition, barcode scanning, and digital food diaries can improve accuracy by reducing reliance on memory and recall. These digital tools offer great potential for real-time dietary assessment and can be customized to various cultural and ethnic diets. However, it is important to acknowledge that none of these methods are perfect, each having its own strengths and limitations. Therefore, extensive research is needed to examine the utility of a multi-method approach, which combines various assessment tools, such as validated dietary questionnaires or recalls, digital food dairies, and biomarker analyses, in improving overall dietary assessment accuracy within culturally diverse populations.

2. Advancing Precision Nutrition

While individual variabilities in dietary responses are well-documented, the methods for accurately quantifying these variations continue to evolve. The integration of metabolomics, genomics, and the gut microbiome profiling into dietary intervention studies and observational cohorts have proven to be useful in characterizing personalized response to nutrition. This approach holds promise in the realm of precision nutrition, where certain dietary recommendations can be tailored to an individual's unique traits. However, there exist significant obstacles to precision nutrition research and practical implementation. First, the omics technologies such as metabolomics and gut microbiome profiling, while promising in tailoring dietary recommendations to individual needs, come with a high price tag that limits their accessibility to a broader population and hinders equitable adoption of this approach. Second, these technologies generate a massive volume of data that require sophisticated computational and analytic tools for meaningful interpretation, impeding the translation of omics data into actional dietary guidance on a personalized level.

To overcome these barriers, it is key to drive down the cost through technological advancements and increased accessibility. Moreover, streamlined analytic approaches and user-friendly tools need to be developed to streamline data interpretation and facilitate clinical applications.

3. Integrating Environmental Sustainability with Nutrition and Health Research The lack of integration between environmental sustainability, the food system, and nutrition and health research represents a significant obstacle to achieving healthy and sustainable diets. While nutrition research has traditionally focused on health outcomes, the broad impact of our dietary

choices and food systems on the environment, ecosystems, and agricultural practices cannot be overlooked.

To bridge this gap, a holistic approach is needed to simultaneously assess the effects of dietary factors on human health, climate change, and the environment more broadly. It is important to examine how the production, processing, and consumption of food influence both individual health and planetary health and identify opportunities for reducing the environmental impact of our food systems through policy changes and technological innovations. Longitudinal studies are particularly valuable in assessing long-term effects of sustainable dietary practices on health outcomes and environmental indicators. Research is also needed to evaluate the effects of dietary recommendations and policies that encourage consumption of plant-based foods in schools, workplace, hospital facilities, and government entities on individual health and the environment. Presently, Dietary Guidelines for Americans predominantly focus on diet and health, ignoring environmental aspects of our food systems. This is a missed opportunity as many other nations have successfully integrated both environmental and health impacts of diet into their national guidelines. Sustainable dietary recommendations should encourage foods that are not only nutritious but also have a lower environmental impact.

Given global interdependence, food practices and policies that affect climate change and health outcomes outside the US have a significant impact on the health of all populations, including ours. It is imperative to invest in research that examines the effects of domestic and global agricultural programs and international trade on planetary health and nutrition, especially in Africa which has a higher burden of nutritional problems. This includes strengthening the capacity of US universities to partner with research institutions in the global South to advance the study of the inter-relationships of food systems, climate change, and planetary health, aimed at identifying solutions that promote health and sustainable agriculture and food systems globally.

4. Understanding and Addressing Obesogenic Food Environment

Insufficient research exists to understand the role of the food environment in driving the obesity epidemic and effective strategies for large-scale improvements to curb the obesity epidemic. As individual eating behaviors are strongly influenced by the food environment, it is important to understand the role of food marketing, processing, accessibility, and affordability as well as oversized portion sizes and social norms that foster unhealthy eating behaviors and contribute to obesity. In the US, widespread availability and high consumption of ultra-processed foods (UPFs) are considered a key component of the obesogenic food environment. Recent studies have established a clear link between higher consumption of UPFs and increased risk of obesity, diabetes, and other chronic diseases. This issue is particularly concerning given that UPFs account for over 60% of daily caloric intake in US children.

To bridge this gap, there is a pressing need to standardize the definition of UPFs because the widely used NOVA classifications have proven contentious and challenging to implement in both research and practice. Additionally, it is also important to understand the heterogeneity of health effects for different subcategories of UPFs. Further research is needed to understand the effects on adiposity of food composition, including non-nutrient components such as additives,

preservatives, and flavorings beyond conventional macronutrient and micronutrient profiles. Analytical methods need to be developed to accurately identify and quantify non-nutrient components. Long-term epidemiologic studies are valuable in tracking the trends in UPF consumption trends and understanding the cumulative impacts on health and the environment over an extended period. Promising leads could be evaluated in randomized trials lasting at least one year with changes in adiposity as the outcome. It is important to identify validated endpoints (e.g. sodium levels) that the FDA could use to set guidelines for the formulation of highly processed foods. Other potential candidates could include added sugar, refined starch, saturated fat, low-caloric sweeteners, and the amounts of preservatives and additives. The goal is to identify steps companies must take to prioritize nutritional value and health as a key criterion in food formulation that is of equal importance as taste, cost, and convenience.

5. Supporting Infrastructure of Cohort Studies of Diet and Healthy Aging

A considerable research gap exists in understanding long-term (lifetime) effects of diet on healthy aging, cognition, and neurodegenerative diseases is a significant research gap, particularly within diverse populations. As the elderly demographic continue to expand, it is crucial to understand the impact of diet across different life stages on chronic disease prevention, preserving physical and mental health, and achieving optimal healthy aging. Beyond conventional health outcomes such as chronic disease incidence and mortality, there is a pressing need to develop and validate biomarkers and other metrics for healthy aging and longevity.

Considering that the effects of diet often manifest over many years or even decades, and studies of that duration are needed to investigate such effects, it is imperative to support the infrastructure of existing cohort studies (including birth cohorts) that have already collected dietary and lifestyle data over many decades. Leveraging these existing resources can continue to provide novel insights into the role of diet in healthy aging and longevity. However, it is disconcerting that the NCI is pulling back its support for cohort infrastructure, which may jeopardize the substantial investments made over the decades. Sustained cohort infrastructure support is critical for continued data collection and analysis, biospecimen banking, enhancing technological and methods innovation, and ensuring sufficient statistical power for the studies of long-term diet on aging-related outcomes.

6. Aligning Governmental Food Assistance Programs with Dietary Guidelines

There exists a major research gap concerning effective methods to align governmental food assistance programs closely with Dietary Guidelines for Americans. While programs like WIC have been mostly successful at this, the universal school breakfast and lunch programs and the extensive SNAP program have fallen behind and could use creative ideas to assist those in the most vulnerable communities to get access to affordable, healthy food. This goes beyond mere access to fruits and vegetables, but making legumes, whole grains, healthy proteins, and healthy beverages, alongside with fruits and vegetables the default choices rather than exceptions within the shopping baskets of the program participants, while limiting UPFs with

high amounts of sugar and sodium. This can be done by incorporating specific nutritional criteria based on DGAs into food selections. In addition, research is needed to examine the effectiveness of other strategies such as providing financial incentives for healthier choices by offering discounts for purchasing fruits, vegetables, or whole grains, improving culinary skills through cooking lessons and assistance in menu planning, and behavioral nudges using behavioral economics strategies (e.g. placing healthier foods more prominently and using food labeling to encourage healthier food choices). Furthermore, integrating screening for food insecurity and nutrition security and fresh produce prescription within the healthcare system can help early identification of individuals in need for timely and targeted intervention. Addressing these gaps can help to enhance the efficacy of the food assistance programs, reduce food insecurity, and improve health outcomes for individuals and communities facing socio-economic challenges.

Frank Hu, MD, PhD

Professor and Chair

Dept. of Nutrition

Harvard T.H. Chan School of Public Health

CYBER-PHYSICAL RESILIENCE COMMENTS

Responses to CPR blogpost https://www.whitehouse.gov/pcast/briefing-room/2023/03/15/pcast-initiating-working-group-on-cyber-physical-resilience/ collected through the PCAST mailbox from organizations from July 20th to August 30th

Contents

Section 1: Adam Isles

Written: 8/11/2023

To Whom It May Concern:

The Chertoff Group is responding to a request for input to the PCAST Working Group on Cyber-Physical Resilience announced in a March 15, 2023 blog. Our response is contained in the attached document. We also endorse the input provided by the Information Systems Security Association through its Cyber Resilience Special Interest Group (also attached).

We would be happy to meet with the Working Group to discuss these points in greater detail, if helpful in advancing their work.

Please feel free to reach out to me either via this email, or at the cell number below, with any questions.

Adam

ADAM ISLES
Principal

TheChertoffGroup

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_Resilient_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/

Chertoff Group Input for PCAST Working Group on Cyber-Physical Resilience

August 11, 2023

The Chertoff Group is responding to a request for input to the PCAST Working Group on Cyber-Physical Resilience announced in a March 15, 2023 blog.

Our response is structured to align with the input categories defined in the blog, which are highlighted in quotes below.

"Recovery and survivability in the face of attacks and events."

We believe the following published content may help inform the Working Group's consideration of recovery and survivability in the face of attacks and events.

- ***"How to Build Resiliency in Times of 'Tail Risk' Events," September 2022 Security Technology Executive article by Chertoff Group Principal Adam Isles and Managing Director Ben Joelson.*** This article describes an integrated cyber-physical approach to manage "tail risk" contingencies. Three key elements are proposed: (1) a business-driven approach to applying graduated levels of security flexibly based on severe but plausible risk scenarios; (2) threat-informed validation of security tools and procedures – physical and cyber, which is key to their successful use in an incident; (3) whole-of-company preparedness for tail risk contingencies, which can help minimize disruption.

- ***"Guidelines for Businesses on Rapid Withdrawal from Conflict Zones and Contested Environments," Chertoff Group March 2022 blog by Chertoff Group Director Brian Hess and Senior Associate Jon Tran.*** The Chertoff Group published "Guidelines for Safe Withdrawal" immediately after Russia's invasion of Ukraine to help companies grappling with operating in and evacuating from war zones (e.g., Ukraine), in contested or hostile environments (e.g., Russia), or on the edges of a conflict zone. The guidance includes both physical and cybersecurity considerations.

- *"How US grid operators can defend against the unprecedented surge in power system attacks," January 2023 Utility Drive article by Chertoff Group Managing Director Ben Joelson.* **This article highlights a surge in physical attacks on electric utility transmission substations and offers a set of recommendations that operators could take to defend against such attacks. These include: (1) leveraging random security measures to confuse adversaries; (2) overhauling information sharing with law enforcement; (3) reviewing playbooks developed after the 2013 attacks against the PG&E Metcalf, California substation; and (4) promoting "see something say something" campaigns.**

"Approaches to assure continuity of operations in degraded states."

We believe that principles of contingency planning are foundational to being resilient and ensuring organizations can continue operating during crisis and degraded states. These principles include:

- Having the resources needed across an organization to focus on contingency planning;
- Using those resources to conduct an organizational risk assessment, including a business impact analysis, incident response plan, recovery plan, and business continuity plan;
- Identifying who in the organization is responsible for these efforts and what the decision-making and notification process will be when actions are needed—to include leadership succession planning, when primary decisionmakers are unavailable;
- Building strong internal coordination and information sharing mechanisms across physical security, information security, resilience, and operational technology functions (as appropriate);
- Identifying external stakeholders, including third-party vendors and technology providers needed for collaboration and unity of effort during a contingency and exercising that connectivity;
- Ensuring that plans are tested and exercised before the emergency occurs to build "muscle memory" and confidence in the organization; and
- Reviewing plans consistently and conducting "after action reviews" following exercises and incidents to adapt plans to the dynamic world where organizations operate.

"Mechanisms to measure and assess modularity and limitations of scope or costliness of failures."

We believe the following points may help inform the Working Group's consideration of how to assess modularity and thereby limit the scope or costliness of incidents:

- The National Incident Management System provides that an Incident Command System organizational structure should be modular, expanding to incorporate all elements necessary for the type, size, scope, and complexity of an incident.
- In our experience, scenario-specific playbooks and exercises can play a key role in enabling modularity and flexibility in practice and thereby limiting the scope and costliness of incidents.
- After the Department of Homeland Security was stood up, a set of 15 national planning scenarios were created for use in preparedness activities. Over time, this practice was discontinued. We recommend that the Working Group consider reinstitution and updating of such scenarios to reflect severe but plausible scenarios relevant in today's times.

"Incentives to balance efficiency which can reduce resilience vs. the investment needed to maintain sufficient resilience."

We believe the following published content may help inform the Working Group's consideration of incentives to balance efficiency which can reduce resilience vs. the investment needed to maintain sufficient resilience:

- ***Incentivizing High-Performing Cybersecurity Programs in the Banking Sector, November 2022 Lawfare blog by Chertoff Group Principal Adam Isles.*** This blog explains how voluntary programs modeled after anti-terrorism public-private-partnerships can help drive faster and more effective cybersecurity performance in the banking sector.

- ***Cyber Risk is Growing; Here's How Companies Can Keep Up, April 2023 Harvard Business Review article by Michael Chertoff.*** The article highlights three ways to improve measurement of cybersecurity risk and performance: (1) bring greater visibility to

inherent risk levels; (2) improve transparency, accuracy and precision around how companies perform against likely threats; and (3) plan for, and measure performance against low probability high-consequence ("tail risk") events.

As to risk quantification, the article notes in particular the trend toward quantifying financial impacts of cyber risk through models like Value at Risk, which quantifies (usually in dollar terms) an entity's potential loss in value over a defined period of time at a given confidence level. That said, the article goes on to highlight the following:

- "These models are useful, but are necessarily dependent on data inputs. Depending on what data drives these models, they can present an overly rosy view of risk. History tells us this is what happened on credit and liquidity risk in the early 2000s, and we have the 2007-2008 financial crisis to show for it."

- "At DHS after 9/11, we framed preparedness planning around a core set of planning scenarios, and British banking regulators now require similar planning and testing around "severe but plausible" scenarios. A good place to start is what happened to Maersk in the notPetya incident, where the company came within a hair's breadth of permanently losing its IT system to destructive malware later attributed to Russia. More recently, Ukraine's pre-invasion migration of workloads to the cloud was critical to its ability to weather a torrent of Russian cyber attacks. The current geopolitical climate underscores the importance of reframing resiliency planning around how to keep the company afloat if its core systems are compromised. Have we maintained offline back-ups and tested recovery? Can we reconstitute a way to communicate with essential employees? Do we know how to ensure that certain important but low-risk payments can continue?"

"Out-of-band or systems-independent means of assuring physical control in the event of digital failures."

We would offer the following points:

- If not practiced, out-of-band tools are worthless at best, and potentially harmful due to a false sense of security. Thus training and exercises around out-of-bank mechanisms are critical to their effectiveness and durability over time.

- We would also note that one the attributes that can make an asset more attractive to an adversary is the ability to infiltrate communications – i.e., that the asset if compromised could provide an adversary access to defender in-band or out-of-band communication tools (this is covered further in the below-referenced High Value Target methodology). Thus, these systems should also be prioritized for cyber defenses and related security testing to ensure their continued confidentiality, integrity and availability notwithstanding an adversarial foothold inside a victim environment.

"Methodologies and standards to encourage resilient systems design and adoption."

We believe the following resiliency strategic design principles apply. While they are largely based on five core cyber resiliency strategic design principles articulated by the U.S. National Institute of Standards & Technology, we believe they apply across cyber and physical environments:

- **Focus on Common Critical Assets (High Value Assets).** A common focus across security programs is aligning security resources with assets critical to the achievement of business objectives. For example, these critical assets may include highly sensitive intellectual property, or top risk manufacturing facilities. Categorizing critical business functions as part of a larger business impact analysis, to include recovery time objectives for key functions and interdependencies with key suppliers or vendors, can play a key role in this. Moreover, certain systems are highly targeted by threat actors because they perform functions critical to trust and are thus stepping-stones into everything else. Such systems do not always rank highly in standard business impact analysis programs. We believe that this High Value Target (HVT) methodology can be highly relevant in enabling practitioners to fine-tune existing business impact assessments to reflect an adversarial viewpoint on high value asset categorization.
- **Support Agility and Architect for Adaptability.** A company's geographic distribution and history of mergers and acquisitions can bring complexity and exposure to rapidly evolving geopolitical, regulatory, and security risks. We believe it important to consider how to enable agility and adaptability within the security program and related functions.
- **Reduce Complexity, Where Possible.** To the extent there are business opportunities to reduce complexity in footprint and protocols, either by rationalizing or standardizing resources (e.g., facilities, IT applications, vendors), this can also enable security resources to be more focused.
- **Build Defense-in-Depth.** A prudent planning assumption is that adversaries may achieve some foothold inside the organization, either through an insider, technical means, or audits and inspections. Capabilities that support contingency planning for (and ideally early identification of) such developments increase an organization's ability to minimize any related impacts.
- **Expect Adversaries to Evolve.** Changing threat motivations, capabilities, and opportunities put a premium on efficient use of risk and threat intelligence to inform programmatic and operational activities.

The 2023 National Cybersecurity Strategy's Pillar Four (Invest in a Resilient Future) also reflects key principles for building and sustaining a resilient digital ecosystem that often supports critical physical infrastructure.

We also believe the above-referenced Harvard Business Review article by Michael Chertoff may help inform the Working Group's consideration of methodologies and standards to encourage resilient cybersecurity systems design and adoption:

- ***Cyber Risk is Growing. Here's How Companies Can Keep Up, Harvard Business Review, April 2023 article by Chertoff Group Executive Chairman Michael Chertoff***. This article starts by explaining how, as "bad actors" become increasingly well-financed, and the "attack surface" becomes more complex, it's becoming practically impossible to ensure that everything is properly patched. It then proposes augmenting the way cybersecurity performance is measured in three ways: First, at the front-end, greater visibility is needed on organizations' inherent risk levels — essentially, "what are we being asked to defend?" Second, greater transparency, accuracy, and precision is required on how defenders perform against likely threats (and whether we do so consistently across the attack

surface). Third, the piece highlights the importance of planning for, and measuring performance against, low-probability high-consequence events.

Chertoff Group Background

In order to provide context for our response, we thought it useful to share some high-level background information on our firm and the nature of our work. The Chertoff Group is a global risk advisory firm founded in 2009 by former Secretary of Homeland Security Michael Chertoff. Our team possesses a blend of commercial and public sector backgrounds, and we serve both global Fortune 500 enterprise clients and mid-sized organizations with specialized security needs.

- **Experience with Complex, Global Organizations.** We have experience advising organizations with complex physical, cyber, geopolitical and blended risks, include global financial services, technology, professional services, manufacturing, logistics, and consumer-packaged goods companies.

- **Highly Qualified, Integrated Team.** Our team is comprised of experts who are not only highly qualified in their respective fields (corporate, cybersecurity, and geopolitical risk), but who regularly work together on integrated teams with highly credentialed personnel that help organizations understand intersecting physical, cyber, and geopolitical security risks and build converged security functions. Our team also features subject matter experts who have held Chief Security Officer roles and have practical experience building, implementing, measuring, and sustaining comprehensive security programs.

- **U.S. Government-vetted Security Risk Management Methodology**. Our methodology has been approved for **SAFETY Act** designation by the U.S. Department of Homeland Security (DHS) Science & Technology Directorate, designating the methodology as operationally effective in reducing security risk. We were approved more than six years ago, and re-approved in October 2022. In addition, we rely on industry leading authoritative frameworks like ASIS Protection Standards and MITRE ATT&CK to guide our work and build traceability to best practice.

GENERATIVE AI COMMENTS

Responses to AI blogpost https://www.whitehouse.gov/pcast/briefing-room/2023/05/13/pcast-working-group-on-generative-ai-invites-public-input/ collected through the PCAST mailbox from individuals from July 25th to August 30th

Table of Contents

Section 1: Rohit Anabheri Venkata

Written: 7/25/2023

Dear Esteemed PCAST Members,

I hope this email finds you in good health and high spirits. Firstly, allow me to express my profound respect and admiration for the work you have been orchestrating in the sphere of Generative AI. It is nothing short of inspiring to see the way you have been harnessing the power of this transformative technology, paving the way for a future we have only begun to envisage.

The work you have accomplished in the realm of AI is akin to developing a symphony of the future, a harmonious blend of data and algorithms that speak a language we are still learning. Your dedication and commitment in this sphere have immense implications for the society and undoubtedly promise a brighter future for us all. I write to share some insights on the intriguing yet critical questions that you have posed regarding the implications of AI-generated media and disinformation. These concerns are of immense importance, especially in today's digitally-driven society. Furthermore, I would like to introduce myself. Rohit Anabheri, a thought leader in Generative AI community, who has expressed interest in contributing to the PCAST working group.

- Firstly, ensuring reliable access to verifiable and trustworthy information in the face of easily generated false media is a complex task. We must look into a combination of technological and legislative measures. Blockchain technology, for instance, could provide a robust system for source verification and data integrity. In parallel, new media laws could be enacted to hold accountable those who spread falsified information maliciously.
    - EX: I believe that the integration of blockchain technology and digital watermarking can help ensure the authenticity of media content. By storing the source of origin and edits made to any piece of digital information, we can create an unalterable chain of trust. This, coupled with strong legislation and enforcement, could ensure accountability.

- Secondly, the manipulation of beliefs and understanding of citizens by malicious AI use is a serious threat to society. Raising public awareness about AI's capabilities and limits, and introducing strict regulation around AI usage are necessary. AI systems themselves can be trained to identify and flag potential disinformation campaigns, thus forming an AI-driven defense against malevolent AI actions.
    - I propose a multifaceted approach. This includes strengthening the AI systems' ability to identify and filter out such content, imposing stringent penalties for misuse, and empowering the public through education on AI ethics and the potential dangers of AI misuse.

- Thirdly, detecting and countering AI-generated disinformation would require both technological advancements and stringent policies. The development of more sophisticated AI models capable of identifying deepfakes and disinformation is paramount. Moreover, collaborative information-sharing platforms could be established among nations to report and counter disinformation collectively.
    - I propose developing advanced AI systems dedicated to detecting and countering AI-generated disinformation. Policymaking should focus on ensuring the accessibility of these tools for the public, and integrating such tools into social media platforms, news outlets, and digital platforms.

- Fourthly, to avoid the corruption of public engagement with elected representatives by AI-generated noise, transparency and verification systems for online interactions must be promoted. Verified identities and interaction spaces could help maintain the integrity of democratic discourse.
    - AI should be leveraged to filter out AI-generated noise, ensuring authentic public engagement. An ethical framework must be established to govern how AI interacts with and influences public discourse

- Finally, educating all, including our leaders, about AI-generated misinformation, impersonation, and manipulation is a daunting but essential task. This could be achieved through tailored education programs, public campaigns, and mandatory AI literacy in sectors vulnerable to such threats.
    - I advocate for a comprehensive educational initiative, aimed at all societal levels, to help people identify AI-generated misinformation. This could be part of school curricula, corporate training programs, and public awareness campaigns.

Transitioning to another critical topic, I would like to introduce myself Mr. Rohit Anabheri. As a small business owner, I am an orchestra conductor of a different kind—managing operations, people, and resources, orchestrating harmonies and battling dissonance. There's a parallel between our worlds that's intriguing, where nuances matter, the interplay of elements is key, and the outcomes can resonate profoundly.

I am firmly convinced that small businesses will have a significant stake in the generative AI space. As engines of innovation and employment, small businesses are well-positioned to harness the power of AI. AI can provide them with the tools to streamline operations, tailor customer experiences, and predict market trends, thus levelling the playing field with larger corporations. Moreover, small businesses, being deeply embedded in their local communities, are uniquely positioned to understand and respond to their needs. They can use AI to solve problems that matter to these communities, making AI an instrument of societal impact.

I am eager to contribute to this conversation, bringing the small business perspective to your distinguished panel. The dialogue between the pioneering work you are doing and the practical, on-the-ground experiences of small business owners like myself could lead to insights that may not have been apparent in isolation.

I hope to inspire, be inspired, and together, make AI not just a tool for the few, but a game-changer for the many. Thank you for your tireless efforts in making this world a better place, and for considering this perspective.

Thank you for considering these points. I eagerly look forward to further discussing these matters and how we might collaborate towards a balanced and responsible approach to AI use.

Best regards,

Rohit Anabheri Venkata

EQUITABLE AND RESPONSIBLE DEVELOPMENT OF GENERATIVE AI: PCAST'S WORKING GROUP ON ENSURING SAFE AND FAIR DEPLOYMENT

Rohit Anabheri

SAKESH SOLUTIONS

WRITTEN PUBLIC COMMENTS SUBMITTED TO PCAST

Contents

Executive Summary

In the era of deepfakes and AI-generated content, ensuring reliable access to verifiable, trustworthy information is paramount. Potential solutions include standards of journalistic integrity, digital literacy education, and technology for data validation, such as blockchain. Moreover, advanced technologies can confirm the source of a media piece, ensuring that it is genuinely from the claimed source.

Addressing the malicious use of AI by bad actors involves increasing public awareness of these manipulations, developing transparent and accountable AI systems, implementing strong cybersecurity measures, incorporating ethical considerations into AI development, fostering international cooperation, regulating AI usage, and increasing tech industry responsibility. Public-private partnerships also offer a way to mitigate potential misuse.

Technologies like AI and machine learning, blockchain, and advanced digital forensics can help detect and counter AI-generated disinformation. Proper regulations prohibiting the creation and distribution of AI-generated disinformation, alongside industry standards for content creation and distribution, can mitigate risks. A need for global cooperation is evident given the internet's borderless nature. Infrastructural measures include a broad-based digital content verification infrastructure, robust educational programs, and public-private partnerships for combating AI-generated disinformation.

These strategies require ongoing refinement to stay ahead of rapidly evolving AI technologies. The challenge is vast, but a combination of education, policy-making, technology, and international collaboration can maintain access to verifiable, trustworthy information and counter the malicious use of AI.

Recommendations and Policy Proposals

Question: 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?

Reliable access to verifiable, trustworthy information: To ensure reliable access to verifiable information, we need to establish and enforce standards of journalistic integrity, encourage the development of digital literacy skills among the public, and use technology to confirm the authenticity of information. This can include decentralized systems for information validation, such as blockchain technology for data traceability and non-repudiation. We could also create a network of trusted entities, like news organizations, that follow stringent verification protocols.

Certainty of media sources**:** Advanced technologies such as digital watermarks, blockchain for establishing provenance, and advanced content verification algorithms can be used to confirm the origin of a piece of media. Verification platforms can be used to validate the integrity of the media and confirm

its source. Digital signatures can be incorporated into media to confirm the identity of the source and confirm that the media has not been tampered with.

The increasing sophistication of technologies that can generate realistic, fake images, audio, and text, often referred to as "deepfakes", has led to a considerable rise in misinformation, which poses serious challenges to our society. It's become more vital than ever to ensure access to reliable, verifiable, and trustworthy information. Here are some potential strategies:

I. Education and Digital Literacy: A crucial part of the solution lies in increasing public awareness of these manipulations. Teaching individuals to be critical consumers of information, to corroborate facts from multiple sources, and to identify potential signs of fake content can reduce the spread of misinformation.
II. AI-based Verification Systems: Since AI plays a big part in generating misleading information, it can also be utilized in combating it. Detection algorithms can be trained to identify synthetic media and alert users when it's detected. These algorithms can look for inconsistencies that are often present in deepfakes, such as unnatural blinking patterns in videos or discrepancies in lighting.
III. Regulations and Policies: Governments and legislative bodies can play a key role in defining what constitutes misuse of AI-generated content and impose penalties for violations. Legislation can be introduced that forces technology companies to better monitor and manage the content on their platforms.
IV. Transparency and Accountability in Algorithms: Platforms that use algorithms to decide what content is shown to users should be transparent about how these decisions are made. This may involve audits or third-party oversight.
V. Collaboration: Tech companies, academic institutions, and governments need to collaborate in the fight against misinformation. This could involve joint research projects, sharing data (within privacy constraints), and commonly agreed standards for content verification.
VI. Provenance Tracking: Blockchain or similar technologies can be used to track the origin of digital content. Each piece of content could have a digital signature attached that indicates its source, making it easier to verify the authenticity of the content.
VII. Strengthening Journalistic Practices: Journalism has a critical role in providing reliable information. Journalists and news organizations should adhere to stringent fact-checking protocols and be transparent about their sources and the methods used to verify information.
VIII. Supporting Fact-checkers and Verification Organizations: Fact-checkers play a vital role in debunking false information. Supporting these organizations financially and socially can help in controlling the spread of misinformation.

The challenge is enormous, but with a combination of technical innovation, policy-making, education, and international collaboration, it's possible to ensure access to verifiable, trustworthy information. The above strategies will require continuous refinement and adaptation to keep up with the rapid evolution of AI and deepfake technology.

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

Countering malicious use of AI: To counter malicious use of AI, we need to focus on developing advanced AI systems that can detect AI-generated manipulative content. Laws and regulations should be enforced to penalize the creation and spread of such content. Public education is also crucial to ensure that people are aware of the risks and are equipped to recognize AI-generated manipulations.

Addressing the issue of malicious use of AI to manipulate the beliefs and understandings of citizens involves several integrated strategies:

I. **AI Literacy and Education**: A major part of the solution lies in education. Citizens should be equipped with the tools to discern and critically analyze the information they are exposed to, including the potential to recognize AI-generated or manipulated content. Public understanding of AI technologies, how they work, and how they can be misused is vital. Teaching digital literacy, including the ability to fact-check information and understand AI, should be a part of the educational curriculum.

II. **Transparency and Accountability**: AI systems should be developed and used in a way that is transparent and accountable. This can be achieved by implementing strong policies and regulations that require companies to disclose the use of AI in their products or services. Regulations should also require AI systems to be explainable and auditable, to provide accountability in the case of misuse.

III. **Strong Cybersecurity Measures**: It's crucial to have strong defenses against cyberattacks, including AI-enabled ones. Cybersecurity measures should involve active threat detection, strong encryption techniques, and regular system updates and patches. Strategies should also include plans to identify and mitigate potential AI threats.

IV. **AI Ethics**: Incorporating ethical considerations into the development and deployment of AI is important. This should include ensuring that AI systems respect privacy, fairness, and are designed to prevent misuse. Ethics committees and review boards could provide oversight.

V. **International Cooperation**: Given the global nature of AI, international cooperation is key to addressing its malicious use. This could involve sharing of best practices, joint efforts in AI research, policy alignment, and coordinated response to AI-enabled threats.

VI. **Regulation**: Governments need to implement robust regulations around the use of AI, particularly in areas where it could potentially manipulate the beliefs and understanding of citizens, such as in political advertising or social media. This could involve oversight of AI in these sectors and penalties for misuse.

VII. **Tech Industry Responsibility**: The technology industry has a responsibility to mitigate against potential misuse of their platforms by malicious actors, for instance, by implementing robust fact-checking measures and creating algorithms that promote accurate information rather than sensational or divisive content.

VIII. **Public-Private Partnerships**: Government bodies, private sector companies, and non-profit organizations can work together to research and respond to the threat of AI misuse. This may include shared initiatives to develop more secure AI, efforts to educate the public about AI, and jointly-funded research into new strategies to prevent AI misuse.

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

The rise of artificial intelligence has also led to a significant increase in AI-generated disinformation, popularly known as "deepfakes." Detecting and countering this form of disinformation requires concerted effort across multiple dimensions, including technology, policy, and infrastructure.

Technologies, policies, and infrastructure**:** Technologies to detect AI-generated disinformation could include AI-based fact-checking systems, digital fingerprinting, watermarking, and blockchain-based systems for data provenance. Policies may include rigorous regulations on information dissemination, stringent repercussions for distributing disinformation, and incentives for organizations to create and adopt disinformation detection technologies. Infrastructure could include a global, interoperable system of verified identities and content, coupled with stronger privacy and data protection protocols.

Technologies:

I. *AI and Machine Learning*: Paradoxically, the same technology that is being used to create disinformation can be employed to detect and counter it. Machine learning algorithms can be trained to identify the patterns, anomalies, and signatures that indicate AI-generated content. For example, in the case of deepfake videos, subtle irregularities in facial movements, lighting, or other visual indicators can be spotted by these algorithms.
II. *Blockchain Technology*: Blockchain can provide a way to verify the authenticity of digital content. By registering a digital "fingerprint" or hash of original content on a blockchain, it becomes possible to verify if a given piece of content matches the original or if it has been altered.
III. *Forensic Tools*: Advanced digital forensic tools can help in identifying manipulated content. These tools can look for inconsistencies in the data that would not typically appear in unaltered content.

Policies:

I. *Regulations*: Regulations need to be in place that prohibit the creation and distribution of AI-generated disinformation. These policies must balance the need to prevent disinformation with the importance of freedom of speech and innovation.
II. *Standards*: Industry standards for the creation and distribution of AI-generated content can be an effective method for mitigating the risks of disinformation. These standards can include technical measures for content verification, guidelines for ethical AI use, and requirements for transparency about the use of AI in content creation.
III. *Global Cooperation*: Given the global nature of the internet, international cooperation is vital for enforcing regulations and standards. Countries around the world need to collaborate on defining and enforcing measures to combat AI-generated disinformation.

Infrastructure:

I. *Verification Infrastructure*: There is a need to develop and implement a broad-based digital content verification infrastructure. This can include blockchain-based solutions for content

verification, trusted digital content registries, and online services that can verify the authenticity of digital content.

II. *Education and Awareness*: It is essential to build a robust infrastructure for educating the public about AI-generated disinformation. This includes not just formal education programs, but also public awareness campaigns, online resources, and tools that can help people identify disinformation.

III. *Public-Private Partnerships*: Governments and private companies need to work together to build the necessary infrastructure for combating AI-generated disinformation. This can involve collaboration on technology development, information sharing, and the implementation of regulations and standards.

Question: 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?

Preserving public engagement with elected representatives: Preserving genuine human interactions in democracy requires stringent regulation of AI usage in political campaigns. Also, political representatives could use verified platforms to engage with the public, reducing the risk of impersonation. Cap on the volume of AI-generated content in public forums may also be necessary. Additionally, public education on recognizing AI-generated content can help individuals discern between genuine human interactions and AI-generated content.

Maintaining meaningful, genuine public engagement with elected representatives amidst the rapid rise of AI technologies is a complex and pressing issue. It's essential to safeguard democratic processes and ensure the integrity of communication channels, which might be threatened by AI-generated noise. Here are several measures that can be adopted:

1. AI Education: It's critical to educate the public about AI and its capabilities, including how it can be used to manipulate narratives or create misinformation. Understanding these potential dangers allows the public to better scrutinize the information they receive and interact with, encouraging more discerning consumption of AI-generated content.
2. AI Transparency Legislation: Governments should pass legislation that requires AI developers to design their tools in a transparent manner, including the disclosure of when AI or bots are used. This could involve clear labels on AI-generated content or the use of watermarks on deepfake videos or images. Transparency also extends to the algorithms themselves. Making algorithms explainable can help us understand the principles that guide their decision-making and mitigate biases.
3. Media Literacy Initiatives: Strong media literacy programs are key to enabling the public to critically analyze and understand the media they consume. This includes not only recognizing AI-generated content but also understanding the mechanisms behind traditional media, social media, and fake news. These programs could be included in school curriculums, community outreach efforts, or adult education initiatives.
4. Robust Moderation and Fact-Checking Infrastructure: Social media platforms and other online sites where information is shared should implement robust moderation policies to manage AI-generated content. This could include automated systems to detect potential

misinformation, as well as human moderation to verify the accuracy of content. Fact-checking organizations also play an essential role in this infrastructure.

5. Public-Private Partnerships**:** Governments, tech companies, and civil society organizations need to collaborate to address AI-generated noise. Such partnerships can facilitate information sharing, drive innovation in detection and moderation tools, and establish ethical standards for AI development and use.
6. Develop Better AI Systems**:** Tech companies should be encouraged to design AI systems capable of filtering out AI-generated noise and misinformation. While this could result in an 'arms race' between generating and detecting systems, it can also stimulate progress and innovation in the field.
7. Strengthening Democracy at Grassroot Level**:** Making democracy more participatory and less susceptible to manipulation involves strengthening it at the grassroot level. This could involve increasing face-to-face interactions between the elected representatives and the public, town-hall meetings, community outreach programs, and initiatives to get public opinion on different issues. This way, even if AI noise is present, it's less likely to drown out the genuine voices.

Question: 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?

Skills for identifying AI-generated misinformation**:** Public education programs that focus on digital literacy can help individuals identify AI-generated content. Tools and training for those in leadership positions are also essential. Universities, research institutions, and private sector companies can collaborate to provide necessary training and resources. Inclusion of AI ethics and detection techniques in school and university curriculums could also be a promising approach. AI-powered tools can be developed to aid in the detection and identification of misinformation, impersonation, and manipulation, and these can be made widely available for use.

Addressing the challenge of AI-generated misinformation, impersonation, and manipulation requires a multi-pronged approach. This is not only a technical issue, but also a social, educational, and regulatory one. Here are several strategies to consider:

1. Education and Awareness**:**

- *General Public***:** We can start by implementing comprehensive digital literacy programs in schools and communities that include lessons on AI-generated content. This will teach everyone, including children, how to critically evaluate the content they encounter online.
- *Political and Industrial Leaders***:** Specialized workshops or seminars can be arranged for leaders in these sectors, which highlight the potential risks and signs of AI-generated content.
- *Educational Leaders***:** They can receive training on how to teach digital literacy, including how to spot AI-generated content, as well as ways to incorporate this into the curriculum.

2. Technological Solutions**:**

AI has the potential to not only create but also detect its own fabrications. As we develop more sophisticated AI, we should also focus on enhancing the capabilities of AI to spot AI-generated fakes. These technological tools can then be widely disseminated and utilized.

3. Policy and Regulation:

Legislation can play a vital role in managing the risks associated with AI-generated misinformation. This can range from imposing stricter penalties on those who intentionally spread AI-generated misinformation to creating guidelines for AI developers and users to promote transparency and accountability.

4. Encourage Open Research and Collaboration**:**

Promoting transparency in AI research can help the global community stay one step ahead of those who would use this technology maliciously. This means encouraging researchers to share their findings, techniques, and tools with the public.

5. Develop Ethical Guidelines for AI Use**:**

Ethical guidelines for the use of AI should be developed and adopted globally. These guidelines can outline the responsibilities and boundaries of AI users and developers, and can serve as a benchmark for assessing whether an action involving AI is ethical or not.

6. Media Responsibility**:**

Media outlets should be encouraged to responsibly report on AI-generated misinformation and provide tools and knowledge for their audience to identify such content. They should also be discouraged from sensationalizing AI-generated misinformation, which could otherwise inadvertently contribute to the spread of such content.

7. Partnership and Collaboration**:**

Collaboration between technology companies, academic institutions, non-profit organizations, and governments can lead to more holistic solutions. These collaborations can lead to better detection tools, more effective education programs, and stronger regulations.

8. Fostering a Culture of Critical Thinking**:**

Finally, one of the most powerful tools in combating misinformation is a culture that values critical thinking and skepticism. Encouraging people to question and fact-check information before sharing it can significantly reduce the spread of AI-generated misinformation.

Overall, the key to combating AI-generated misinformation lies in a combination of technical solutions, regulatory measures, education, and promoting a culture of critical thinking and skepticism. These strategies should be tailored to the specific needs of different groups, from the general public to our political, industrial, and educational leaders.

Conclusion

The battle against AI-generated disinformation and misinformation is complex, necessitating a constant evolution of strategies to keep pace with the rapid development of AI. Through the concerted efforts of technological innovation, education, policy-making, and international collaboration, we can safeguard our information landscape from the detrimental effects of deepfakes and AI manipulations.

In this era of mass digital manipulation, the importance of reliable access to verifiable, trustworthy information cannot be overstated. With the proliferation of AI-driven "deepfakes", it is essential to establish measures that can validate information and ascertain its sources. Possible solutions to this challenge involve strategies spanning education, technology, policy, and collaboration. By enhancing digital literacy among the public, utilizing AI-based verification systems, implementing strict regulations, promoting transparency in algorithms, encouraging collaborations between different stakeholders, employing provenance tracking, strengthening journalistic practices, and supporting verification organizations, we can mitigate the risks associated with misinformation.

Equally crucial is the need to counteract the use of AI by malicious actors intending to manipulate citizens' beliefs and understanding. This requires a well-rounded approach that combines AI literacy and education, transparent and accountable AI systems, robust cybersecurity measures, consideration of AI ethics, international cooperation, strict regulations, tech industry responsibility, and public-private partnerships. By enabling citizens to critically analyze information, demanding transparency in AI usage, fortifying our defenses against AI-enabled attacks, integrating ethical considerations in AI deployment, cooperating on a global scale, regulating AI usage, and holding the tech industry accountable, we can prevent the misuse of AI.

Finally, to detect and counter AI-generated disinformation, we must leverage technologies, policies, and infrastructure. Utilizing AI and machine learning, blockchain technology, and digital forensic tools can significantly help in detecting manipulated content. Regulations and standards that prohibit the creation and distribution of such disinformation, coupled with global cooperation, form the policy-level response to this challenge. A robust digital content verification infrastructure, awareness campaigns, and a combination of public and private resources can create a strong foundational response against the spread of disinformation.

In sum, while the challenges posed by AI and deepfake technologies are significant, a multifaceted, holistic approach involving a blend of technology, policy, education, and international cooperation can provide a reliable pathway to ensuring access to verifiable, trustworthy information and countering misinformation and disinformation.

Section 2: Vince Minerva

Written: 7/27/2023

Dear PCAST Team,

Thank you for the opportunity to provide a written submission regarding Generative AI. Attached are some actionable ideas which can help drive reliable access to verifiable, trustworthy information.

Please reply to this email if you have any questions regarding the input.

Sincerely,

Vince Minerva

President's Council of Advisors on Science and Technology (PCAST) Invitation for Public Input on Generative AI

Point of Contact: Vince Minerva, Member of the Public

July 26, 2023

Executive Summary:

Establishing foundational guidance for development, deployment, and ongoing monitoring of trustworthy Artificial Intelligence (AI) systems by leveraging proven federal government strategies for information security, can drive reliable access to verifiable, trustworthy information (Reference Question 1). Creating actionable trustworthiness requirements for AI systems analogous to NIST Special Publication (SP) 800-171, Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations (Ref. 1), provides necessary technical leadership for the AI community. NIST SP 800-171 provides information security requirements applicable to all nonfederal systems and organizations that process, store, and/or transmit controlled unclassified information. NIST SP 800-171 is intended for use by federal agencies in contractual vehicles or other agreements established between those agencies and nonfederal organizations (Ref. 1, p. iii). For example, the Department of Defense requires a significant percentage of the 100,000 + (Ref. 2) Defense Industrial Base (DIB) companies to comply with NIST SP 800-171 via contractual requirements which benefits our national security by safeguarding sensitive information. Similarly, creating a NIST publication of requirements for trustworthy AI systems should strengthen our national AI capabilities. Many of the use cases for Generative AI require large scale computing resources, which will likely result in most of the public and private industry users acquiring services from commercial providers. Many, if not most, of the cloud and enterprise IT providers with financial and IT system resources sufficient to run large language models are contractors to the federal government. The federal government's ability to influence the trustworthiness of AI systems should be increased similarly to the demonstrated successes in cybersecurity.

ESTABLISHING TRUSTWORTHY AI SYSTEMS

- Emphasis on trustworthiness during entire system lifecycle
- Build on proven federal government successes
- Facilitate collaboration with industry, NGOs, international partners, the public, etc.
- Beneficial to Generative AI Systems
- Supports voluntary and regulated environments

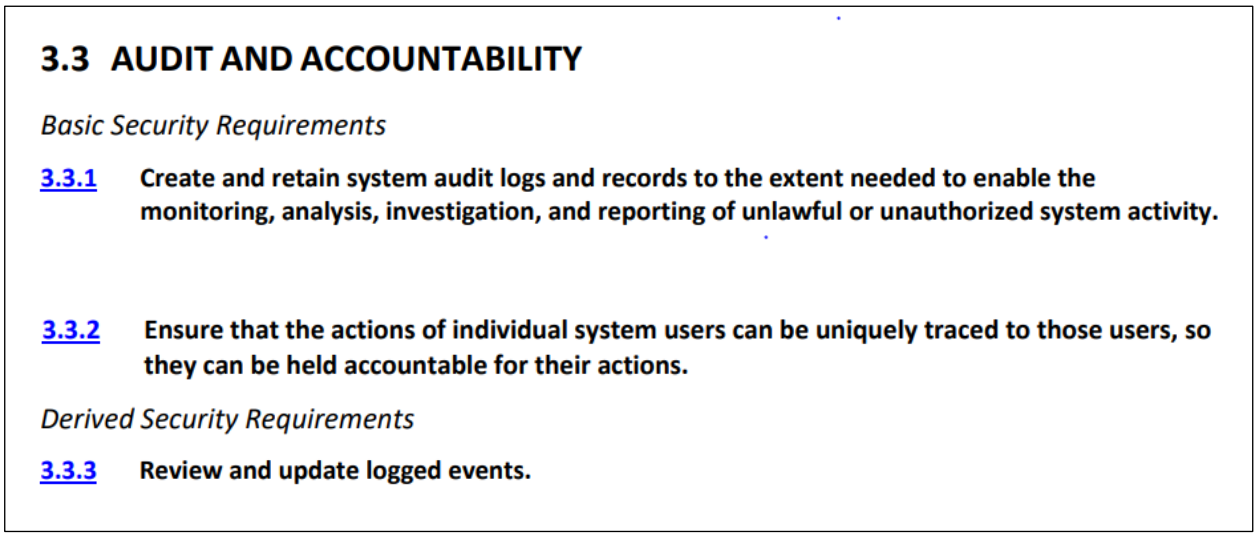
Developing Guidance for Trustworthy AI Systems

NIST SP 800-171 security requirements are organized into fourteen (14) security families containing basic and derived requirements as shown in Figures 1 and 2 below. Each of the

requirements has a discussion section providing additional information to facilitate requirement implementation and assessment.

FAMILY	FAMILY
Access Control	Media Protection
Awareness and Training	Personnel Security
Audit and Accountability	Physical Protection
Configuration Management	Risk Assessment
Identification and Authentication	Security Assessment
Incident Response	System and Communications Protection
Maintenance	System and Information Integrity

Figure 1: NIST SP 800-171 Security Requirements Families (Ref. 1, p. 7)



3.3 AUDIT AND ACCOUNTABILITY

Basic Security Requirements

3.3.1 **Create and retain system audit logs and records to the extent needed to enable the monitoring, analysis, investigation, and reporting of unlawful or unauthorized system activity.**

3.3.2 **Ensure that the actions of individual system users can be uniquely traced to those users, so they can be held accountable for their actions.**

Derived Security Requirements

3.3.3 **Review and update logged events.**

Figure 2: NIST SP 800-171 Basic and Derived Requirements Examples (Ref. 1, pp. 17-18)

Applying this approach to AI Systems leads to trustworthiness requirement families as shown in Table 1 below. The seventeen (17) candidate trustworthiness families were primarily derived from the NIST AI RMF 1.0 and the AI Bill of Rights. However, the process of formalizing the trustworthiness families enables incorporation of concepts from a diverse set of resources, such as, the Organization for Economic Co-operation and Development (OECD) AI Principles (Ref. 3). Once the trustworthiness families are defined through proven review and comment processes, they can be defined by basic and derived requirements.

Trustworthiness Requirement Family	**Summary Objective**
Accountability	"This is the fundamental need: to ensure that machines remain subject to effective oversight by people and the people who design and operate machines remain accountable to everyone else. In short, we must always ensure that AI remains under human control." Ref. 4, p. 4
Accuracy	"Accuracy is defined by ISO/IEC TS 5723:2022 as "closeness of results of observations, computations, or estimates to the true values or the values accepted as being true." Measures of accuracy should consider computational-centric measures (e.g., false positive and false negative rates), human-AI teaming, and demonstrate external validity (generalizable beyond the training conditions)." Ref. 5, p. 14; AI systems should provide a confidence level for predictions.
Explainability and Interpretability	*"Explainability* refers to a representation of the mechanisms underlying AI systems' operation, whereas *interpretability* refers to the meaning of AI systems' output in the context of their designed functional purposes." Ref. 5, p. 16
Fairness	Fairness in AI includes concerns for equality and equity by addressing issues such as harmful bias and discrimination. Ref. 5, p. 17
Ongoing Monitoring	Automated systems should have ongoing monitoring procedures, including recalibration procedures, in place to ensure that their performance does not fall below an acceptable level over time, based on changing real-world conditions or deployment contexts, post-deployment modification, or unexpected conditions." Ref. 6
Planning	Develop, document, and disseminate policies, plans, and procedures necessary to implement trustworthiness requirements.
Privacy	"Privacy refers generally to the norms and practices that help to safeguard human autonomy, identity, and dignity." Ref. 5, p. 17; "Automated systems should be designed and built with privacy protected by default". Ref. 7
Reliable	"Reliability is defined in the same standard as the "ability of an item to perform as required, without failure, for a given time interval, under given conditions" (Source: ISO/IEC TS 5723:2022)." Ref. 5, p.13
Resiliency	"AI systems, as well as the ecosystems in which they are deployed, may be said to be resilient if they can withstand unexpected adverse events or unexpected changes in their environment or use – or if they can maintain their functions and structure in the face of internal and external change and degrade safely and gracefully when this is necessary (Adapted from: ISO/IEC TS 5723:2022)." Ref. 5, p. 15
Risk Management	"AI risk management offers a path to minimize potential negative impacts of AI systems, such as threats to civil liberties and rights,

	while also providing opportunities to maximize positive impacts. Addressing, documenting, and managing AI risks and potential negative impacts effectively can lead to more trustworthy AI systems." Ref. 5, p. 4
Robustness	"Robustness or generalizability is defined as the "ability of a system to maintain its level of performance under a variety of circumstances" (Source: ISO/IEC TS 5723:2022). Robustness is a goal for appropriate system functionality in a broad set of conditions and circumstances, including uses of AI systems not initially anticipated." Ref. 5, p. 14
Safety	"AI systems should "not under defined conditions, lead to a state in which human life, health, property, or the environment is endangered" (Source: ISO/IEC TS 5723:2022)." Ref. 5, p. 14
Security	"…concerns related to the confidentiality, integrity, and availability of the system and its training and output data…" Ref. 5, p. 8
Supply Chain Risk Management	"Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues." Ref. 5, p. 24
Test, Evaluation, Verification, and Validation (TEVV)	Tasks are performed throughout the AI lifecycle that are carried out by AI actors who examine the AI system or its components or detect and remediate problems. Ref. 5, p. 35
Transparency	"Transparency reflects the extent to which information about an AI system and its outputs is available to individuals interacting with such a system – regardless of whether they are even aware that they are doing so." Ref. 5, p. 15
Valid	"Validation is the "confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled" (Source: ISO 9000:2015)." Ref. 5, P. 13

Table 1: Candidate Trustworthiness Requirement Families

Additional support can be provided to the AI Systems community by creating a document that provides procedures for self-attestation or third-party audits of trustworthiness requirements similar to NIST SP 800-171A, Assessing Security Requirements for Controlled Unclassified Information (Ref. 8). NIST SP 800-171A enables organizations to generate evidence to support assertion of requirements satisfaction.

Summary

- Establishing foundational guidance documents for developing and using trustworthy AI systems strengthens federal government leadership in the AI community.
- Leveraging federal government demonstrated successes with information security provides confidence in a high value outcome.
- Processes used to develop trustworthiness requirements facilitate collaboration amongst a diverse set of stakeholders.

- The voluntary commitments to manage risk posed by AI secured from leading AI companies by the Biden-Harris Administration (Ref. 9) are easily incorporated into the trustworthiness requirements.
- Trustworthiness requirements support development of Generative AI systems aligned with federal government priorities such as the AI Bill of Rights.
- NIST-led effort ensures that the necessary cybersecurity principles for secure AI systems are incorporated into trustworthiness requirements.

References:

1. NIST Special Publication (SP) 800-171 Revision 2, Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations
2. https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors/defense-industrial-base-sector
3. https://oecd.ai/en/ai-principles
4. Governing AI: A Blueprint for the Future, Microsoft
5. NIST AI 100-1, Artificial Intelligence Risk Management Framework (AI RMF 1.0)
6. https://www.whitehouse.gov/ostp/ai-bill-of-rights/safe-and-effective-systems-3/
7. https://www.whitehouse.gov/ostp/ai-bill-of-rights/data-privacy-2/
8. NIST Special Publication (SP) 800-171A, Assessing Security Requirements for Controlled Unclassified Information
9. FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI | The White House

President's Council of Advisors on Science and Technology (PCAST) Invitation for Public Input on Generative AI

Point of Contact: Vince Minerva, Member of the Public

July 26, 2023

Executive Summary:

Establishing foundational guidance for development, deployment, and ongoing monitoring of trustworthy Artificial Intelligence (AI) systems by leveraging proven federal government strategies for information security, can drive reliable access to verifiable, trustworthy information (Reference Question 1). Creating actionable trustworthiness requirements for AI systems analogous to NIST Special Publication (SP) 800-171, Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations (Ref. 1), provides necessary technical leadership for the AI community. NIST SP 800-171 provides information security requirements applicable to all nonfederal systems and organizations that process, store, and/or transmit controlled unclassified information. NIST SP 800-171 is intended for use by federal agencies in contractual vehicles or other agreements established between those agencies and nonfederal organizations (Ref. 1, p. iii). For example, the Department of Defense requires a significant percentage of the 100,000 + (Ref. 2) Defense Industrial Base (DIB) companies to comply with NIST SP 800-171 via contractual requirements which benefits our national security by safeguarding sensitive information. Similarly, creating a NIST publication of requirements for trustworthy AI systems should strengthen our national AI capabilities. Many of the use cases for Generative AI require large scale computing resources, which will likely result in most of the public and private industry users acquiring services from commercial providers. Many, if not most, of the cloud and enterprise IT providers with financial and IT system resources sufficient to run large language models are contractors to the federal government. The federal government's ability to influence the trustworthiness of AI systems should be increased similarly to the demonstrated successes in cybersecurity.

ESTABLISHING TRUSTWORTHY AI SYSTEMS

- Emphasis on trustworthiness during entire system lifecycle
- Build on proven federal government successes
- Facilitate collaboration with industry, NGOs, international partners, the public, etc.
- Beneficial to Generative AI Systems
- Supports voluntary and regulated environments

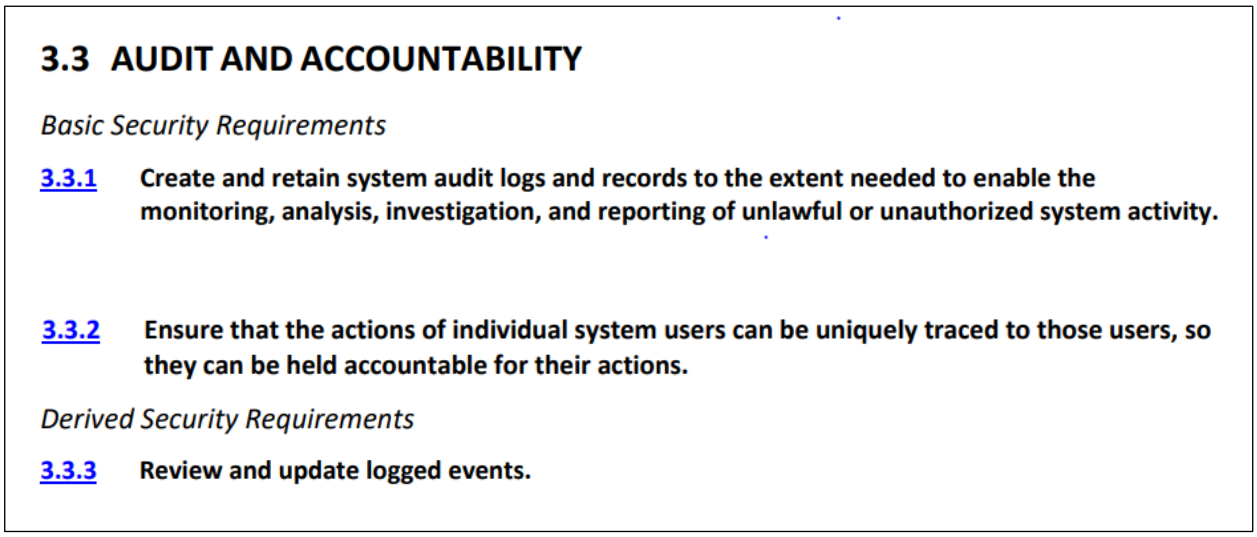
Developing Guidance for Trustworthy AI Systems

NIST SP 800-171 security requirements are organized into fourteen (14) security families containing basic and derived requirements as shown in Figures 1 and 2 below. Each of the

requirements has a discussion section providing additional information to facilitate requirement implementation and assessment.

FAMILY	FAMILY
Access Control	Media Protection
Awareness and Training	Personnel Security
Audit and Accountability	Physical Protection
Configuration Management	Risk Assessment
Identification and Authentication	Security Assessment
Incident Response	System and Communications Protection
Maintenance	System and Information Integrity

Figure 1: NIST SP 800-171 Security Requirements Families (Ref. 1, p. 7)



3.3 AUDIT AND ACCOUNTABILITY

Basic Security Requirements

3.3.1 **Create and retain system audit logs and records to the extent needed to enable the monitoring, analysis, investigation, and reporting of unlawful or unauthorized system activity.**

3.3.2 **Ensure that the actions of individual system users can be uniquely traced to those users, so they can be held accountable for their actions.**

Derived Security Requirements

3.3.3 **Review and update logged events.**

Figure 2: NIST SP 800-171 Basic and Derived Requirements Examples (Ref. 1, pp. 17-18)

Applying this approach to AI Systems leads to trustworthiness requirement families as shown in Table 1 below. The seventeen (17) candidate trustworthiness families were primarily derived from the NIST AI RMF 1.0 and the AI Bill of Rights. However, the process of formalizing the trustworthiness families enables incorporation of concepts from a diverse set of resources, such as, the Organization for Economic Co-operation and Development (OECD) AI Principles (Ref. 3). Once the trustworthiness families are defined through proven review and comment processes, they can be defined by basic and derived requirements.

Trustworthiness Requirement Family	**Summary Objective**
Accountability	"This is the fundamental need: to ensure that machines remain subject to effective oversight by people and the people who design and operate machines remain accountable to everyone else. In short, we must always ensure that AI remains under human control." Ref. 4, p. 4
Accuracy	"Accuracy is defined by ISO/IEC TS 5723:2022 as "closeness of results of observations, computations, or estimates to the true values or the values accepted as being true." Measures of accuracy should consider computational-centric measures (e.g., false positive and false negative rates), human-AI teaming, and demonstrate external validity (generalizable beyond the training conditions)." Ref. 5, p. 14; AI systems should provide a confidence level for predictions.
Explainability and Interpretability	*"Explainability* refers to a representation of the mechanisms underlying AI systems' operation, whereas *interpretability* refers to the meaning of AI systems' output in the context of their designed functional purposes." Ref. 5, p. 16
Fairness	Fairness in AI includes concerns for equality and equity by addressing issues such as harmful bias and discrimination. Ref. 5, p. 17
Ongoing Monitoring	Automated systems should have ongoing monitoring procedures, including recalibration procedures, in place to ensure that their performance does not fall below an acceptable level over time, based on changing real-world conditions or deployment contexts, post-deployment modification, or unexpected conditions." Ref. 6
Planning	Develop, document, and disseminate policies, plans, and procedures necessary to implement trustworthiness requirements.
Privacy	"Privacy refers generally to the norms and practices that help to safeguard human autonomy, identity, and dignity." Ref. 5, p. 17; "Automated systems should be designed and built with privacy protected by default". Ref. 7
Reliable	"Reliability is defined in the same standard as the "ability of an item to perform as required, without failure, for a given time interval, under given conditions" (Source: ISO/IEC TS 5723:2022)." Ref. 5, p.13
Resiliency	"AI systems, as well as the ecosystems in which they are deployed, may be said to be resilient if they can withstand unexpected adverse events or unexpected changes in their environment or use – or if they can maintain their functions and structure in the face of internal and external change and degrade safely and gracefully when this is necessary (Adapted from: ISO/IEC TS 5723:2022)." Ref. 5, p. 15
Risk Management	"AI risk management offers a path to minimize potential negative impacts of AI systems, such as threats to civil liberties and rights,

	while also providing opportunities to maximize positive impacts. Addressing, documenting, and managing AI risks and potential negative impacts effectively can lead to more trustworthy AI systems." Ref. 5, p. 4
Robustness	"Robustness or generalizability is defined as the "ability of a system to maintain its level of performance under a variety of circumstances" (Source: ISO/IEC TS 5723:2022). Robustness is a goal for appropriate system functionality in a broad set of conditions and circumstances, including uses of AI systems not initially anticipated." Ref. 5, p. 14
Safety	"AI systems should "not under defined conditions, lead to a state in which human life, health, property, or the environment is endangered" (Source: ISO/IEC TS 5723:2022)." Ref. 5, p. 14
Security	"…concerns related to the confidentiality, integrity, and availability of the system and its training and output data…" Ref. 5, p. 8
Supply Chain Risk Management	"Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues." Ref. 5, p. 24
Test, Evaluation, Verification, and Validation (TEVV)	Tasks are performed throughout the AI lifecycle that are carried out by AI actors who examine the AI system or its components or detect and remediate problems. Ref. 5, p. 35
Transparency	"Transparency reflects the extent to which information about an AI system and its outputs is available to individuals interacting with such a system – regardless of whether they are even aware that they are doing so." Ref. 5, p. 15
Valid	"Validation is the "confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled" (Source: ISO 9000:2015)." Ref. 5, P. 13

Table 1: Candidate Trustworthiness Requirement Families

Additional support can be provided to the AI Systems community by creating a document that provides procedures for self-attestation or third-party audits of trustworthiness requirements similar to NIST SP 800-171A, Assessing Security Requirements for Controlled Unclassified Information (Ref. 8). NIST SP 800-171A enables organizations to generate evidence to support assertion of requirements satisfaction.

Summary

- Establishing foundational guidance documents for developing and using trustworthy AI systems strengthens federal government leadership in the AI community.
- Leveraging federal government demonstrated successes with information security provides confidence in a high value outcome.
- Processes used to develop trustworthiness requirements facilitate collaboration amongst a diverse set of stakeholders.

- The voluntary commitments to manage risk posed by AI secured from leading AI companies by the Biden-Harris Administration (Ref. 9) are easily incorporated into the trustworthiness requirements.
- Trustworthiness requirements support development of Generative AI systems aligned with federal government priorities such as the AI Bill of Rights.
- NIST-led effort ensures that the necessary cybersecurity principles for secure AI systems are incorporated into trustworthiness requirements.

References:

1. NIST Special Publication (SP) 800-171 Revision 2, Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations
2. https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors/defense-industrial-base-sector
3. https://oecd.ai/en/ai-principles
4. Governing AI: A Blueprint for the Future, Microsoft
5. NIST AI 100-1, Artificial Intelligence Risk Management Framework (AI RMF 1.0)
6. https://www.whitehouse.gov/ostp/ai-bill-of-rights/safe-and-effective-systems-3/
7. https://www.whitehouse.gov/ostp/ai-bill-of-rights/data-privacy-2/
8. NIST Special Publication (SP) 800-171A, Assessing Security Requirements for Controlled Unclassified Information
9. FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI | The White House

Section 3: Dr. Francesca Tripodi & Dr. Justin Reich

Written: 7/31/2023

To the members of the PCAST Working Group on Generative AI –

Attached please find our ideas and comments addressing questions 3 and 5 per the open call for comments. We hope you find our strategies useful and that we may have an opportunity to formally present these ideas to the working group in the future.

Sincerely –

Dr. Francesca Tripodi & Dr. Justin Reich

Submission to PCAST Generative AI Working Group

Dear colleagues,

In this message, we respond to Questions #3 and #5:

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

Regardless of partisanship, inaccurate or misleading information receives higher levels of engagement than trustworthy content (Edelson et al., 2021). Given the propensity with which AI can generate duplicitous information, technologies that can detect these manipulation strategies, and interventions that can enhance human decision-making are critical to reducing the potential harm AI misinformation can have on public health, financial markets, and civic participation. Existing interventions are often too late to be effective, disrupting less than 1% of that engagement (Goldstein et al., 2023). Thus, the ability to combat threats to information integrity is dependent on better tools to empower civic actors to adequately respond to constant threats and tools to help people fact-check for themselves.

Ultimately, we argue that because information is generated by people, interpreted by people, and shared onwards by people– it is fundamentally a human, social problem. While technological approaches to labeling and removing this kind of content may be possible, our research indicates that a robust and comprehensive AI strategy must include detection mechanisms, public participation, and education.

Based on our research, we recommend two key interventions to detect and counter AI-generated attacks on information integrity: 1) early identification of "data voids" to help civic actors develop quality information to fill these gaps, and 2) new approaches to information literacy based on existing interventions that we have already developed, tested, and disseminated to tens of thousands of people.

Early Detection of "data voids"

Propagandists, conspiracy theorists, and hostile foreign governments have a sophisticated understanding of how tags and metadata work. They regularly suggest that users verify information independently, but only after seeding the Internet with problematic content and tagging it with keywords designed to surface and amplify those ideas (Tripodi, 2019; Tripodi, 2022; Williams & Carley, 2023). This strategy all but ensures that the information created by the producers in this network will be the first results returned—turning search engines into self-fulfilling information prophecies. In the retail context, this strategy is known as the "IKEA effect," where marketers encourage their consumers to assemble their products, since this involvement in construction builds affinity (Norton et al., 2012). This "IKEA effect of misinformation" makes audiences feel like they are drawing their own conclusions, thereby valuing them more (Tripodi, 2022). AI has the potential to further this risk since large language models can more readily create websites and digital-first content that mimics authentic information outlets.

"Data voids" are topics associated with little to no reliable and informative content online (Golebiewski & boyd, 2018). Bad actors organize around these information absences to cause confusion, sow mistrust, and undermine national security (Ballantyne & Dunning, 2022; Murthy, 2021; Noble 2018). Research indicates that those trying to sway information integrity use data voids to artificially manipulate the search engine results page (SERP) and guide search traffic to nefarious websites (Ghosh & Scott, 2018; Tripodi, 2022; Williams & Carley, 2023). Since limited or insufficient information about a data void exists, it creates an opportunity for conspiracy theorists and propagandists to manipulate that absence and "fill the void" with corrupt information (Flores-Saviaga et al., 2022; Marwick & Partin, 2022; Starbird et al., 2019; Tripodi, 2020; Tripodi, 2022).

If researchers could identify a list of words or phrases associated with corrupt content before it becomes widespread, programmers, journalists, politicians, educators, and Wikipedians could fill the void(s) with trustworthy information and resources (Urman et al., 2022; Williams & Carley, 2023). New research indicates that AI can play a role in helping detect these voids and enable journalists to fill those gaps quickly (Flores-Saviaga et al., 2022).

Civic Online Reasoning

To create a more secure and trustworthy internet, the NSF (Lyu et al., 2022-2023; Wagner et al., 2021-2023) and industry (Rosen, 2020) have allocated considerable resources to retroactive solutions like fact-checking and content labels. While these efforts can help mitigate the effects of inaccurate information, new research indicates the need for proactive measures to try and "prebunk" strategies used to erode information integrity (Nassetta & Gross, 2020; Porter & Wood, 202l; Roozenbeek et al., 2022). Researchers have found that highly educated people still struggle with basic search literacy tasks and that most search literacy strategies currently conveyed in educational settings are outdated and ineffective (Breakstone et al., 2021; Caufield, 2018; Wineburg & McGrew, 2017). Over the last eight years, researchers at the Stanford History Education Group have teamed up with Dr. Reich's Teaching Systems Lab at MIT to develop curricular resources to effectively teach research-backed strategies for effective search practices. This approach has been rigorously tested in middle school classrooms (Kohnen et al., 2020; Pavlounis et al., 2021), in high school classrooms (McGrew, 2020; Nygren et al., 2021; Wineburg et al., 2022), in college classrooms (Breakstone, Smith, Connors, et al., 2021; Brodsky et al., 2021a; Brodsky et al., 2021b; McGrew et al., 2019), with adults (Panizza et al., 2022), in the U.S. (e.g., Wineburg et al., 2022), in Canada (Pavlounis et al., 2021), in Sweden (Nygren et al., 2021), and in India (Anand & Srivastava, 2022). In each case, results show that it is possible to improve individuals' digital savvy through focused educational interventions.

These tools and resources are not only proven to work in field experiments, but they are also widely adopted across the United States and Canada. The freely available Civic Online Reasoning curriculum has registered more than a quarter of a million downloads since its launch 30 months ago and UNESCO awarded the curriculum a Global Media and Information Literacy Award in 2020, and the American Association of School Librarians named the curriculum a 2022 Best Digital Tool. Digital literacy videos developed in conjunction with the Civic Online Reasoning curriculum have been viewed over three million times on YouTube (Crash Course, 2019).

As Co-PIs of the NSF Convergence Accelerator Track F: *Adapting and Scaling Existing Educational Programs to Combat Inauthenticity and Instill Trust in Information* Federal Award #2137530, Dr. Tripodi and Dr. Reich worked alongside their colleagues, Drs. Breakstone,

Caulfield, and Wineburg, identifying five key innovation areas for expanding search literacy. The broader impacts of this research include two short-form educational videos, a mobile-first online course designed for librarians, and a gamified platform where players tried to prevent their clients from posting bad information on their social media accounts by searching for information about unfamiliar claims (see Exhibits One, Two, and Three). The engagement videos were produced by Retro Report, a non-profit journalism organization that creates content for classrooms to foster engagement and critical thinking skills. The first video "Can You Spot Misinformation?" was viewed by close to 70,000 people in less than a year. The second video, "Where's That Photo From?" was cross-posted on PBS Learning Media and specifically designed to reach students in middle and high school. The team rigorously tested the effectiveness of these tools, asking 822 people to watch "Can You Spot Misinformation" and complete a pretest and posttest that assessed essential search literacy skills. After watching the 90-second Retro Report video, performance between the pre and post-test improved significantly ($p<.001$). Such insights indicate that short, engaging, interventions can have big impacts. Participants also responded positively to the video, with over 76% indicating that the video was useful and engaging. Insights from this research were published in a special issue on Information Integrity in *Library Quarterly,* a peer-reviewed academic journal covering leading issues in information and library science (Tripodi et al., 2023).

Exhibit 1. *Search Lit Strategies from 'Secrets of Effective Search on Mobile Devices' Course*

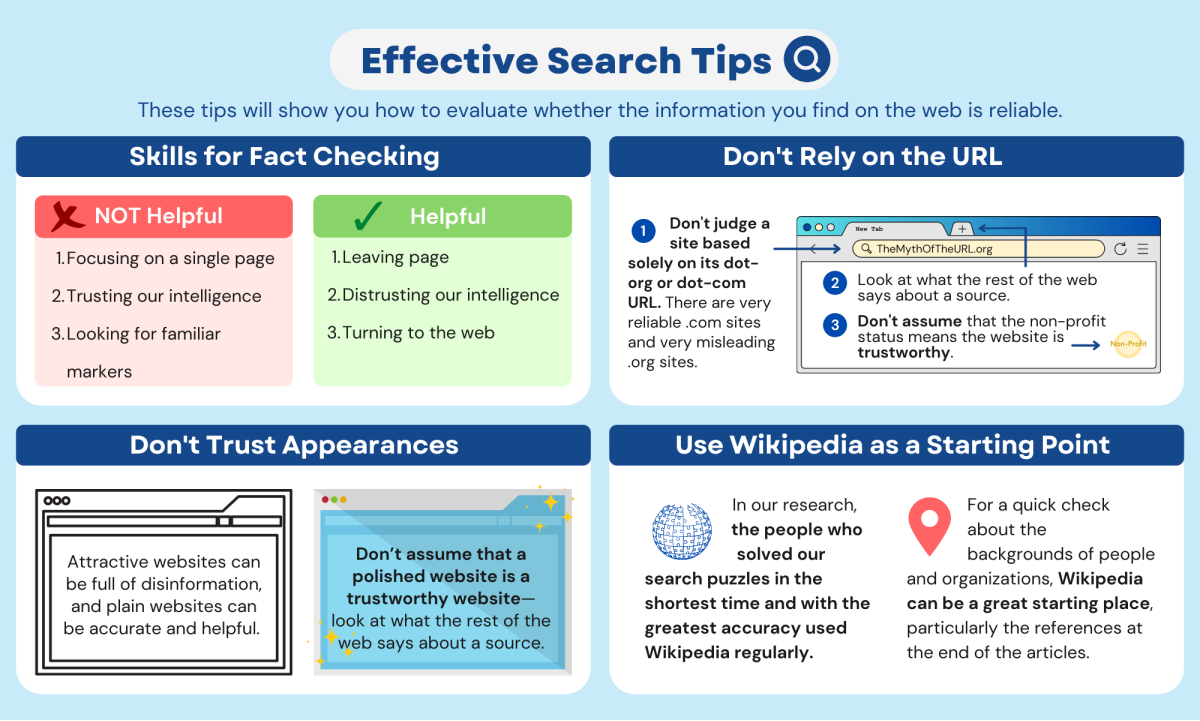


Exhibit 2 - Retro Report Video

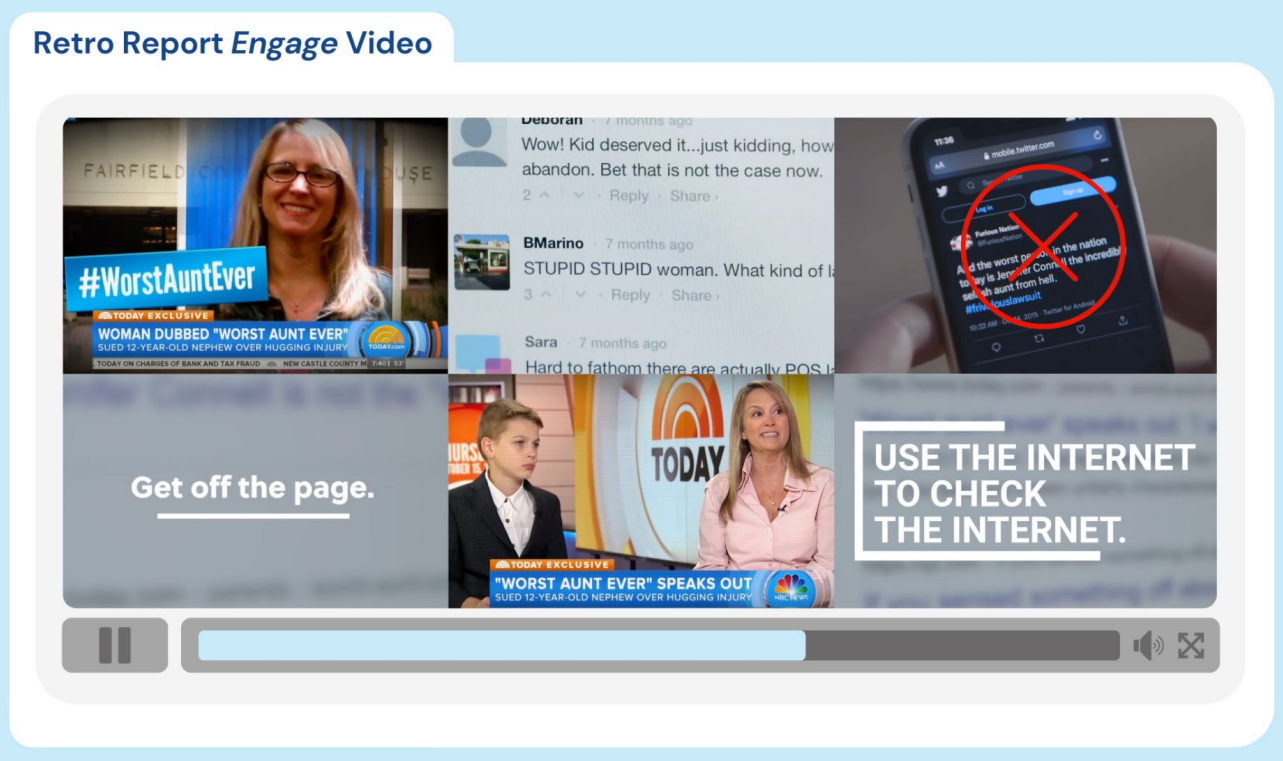
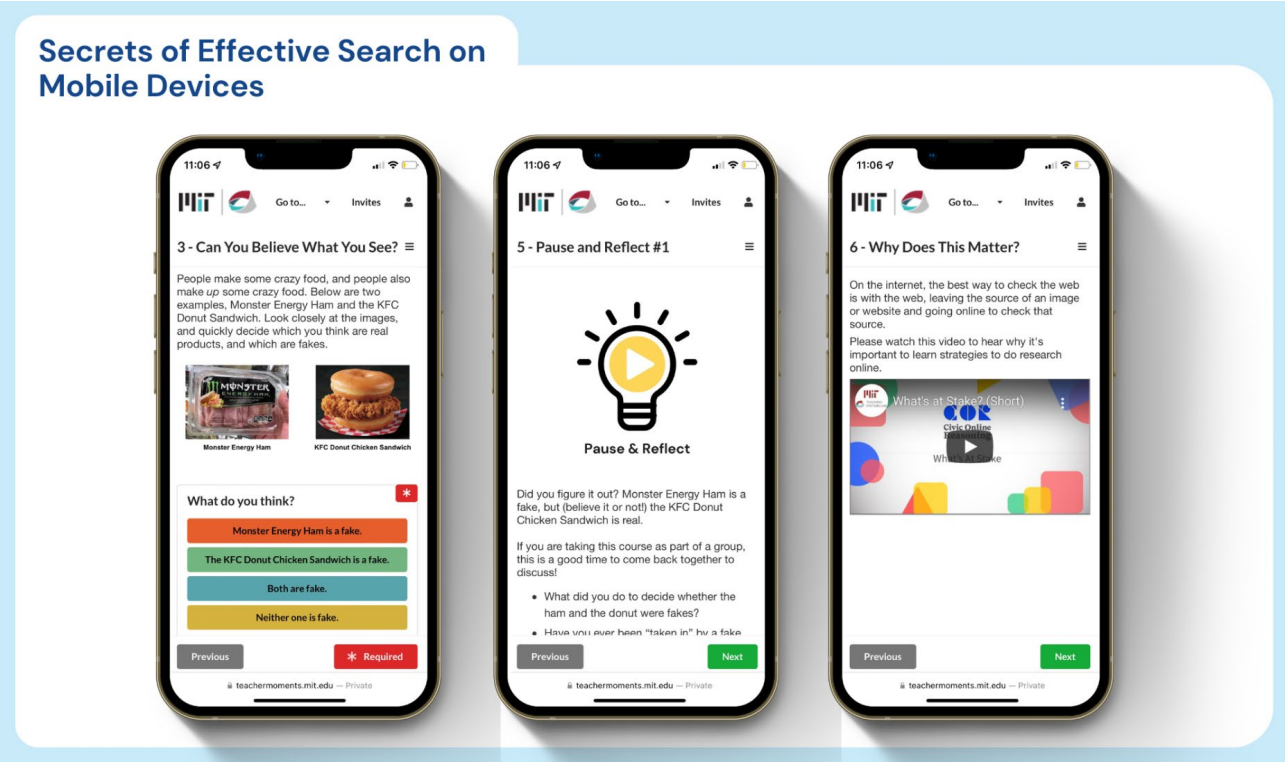


Exhibit 3 - Secrets of Effective Search on Mobile



PCAST Initiatives - Goals and Ideas

Search engines are critically dependent on Wikipedia content to forge and capture information and connections from multiple sources to construct "knowledge graphs" – short snippets of information that attempt to directly answer questions without having to navigate to other hyperlinks/sources (McMahon et al., 2017; Vincent and Hecht, 2021). As such, Wikipedia is ripe for information warfare, especially by historical revisionists trying to bolster support for their causes and mobilize action (Kim et al., 2023). Signals from other platforms (like Wikipedia) may help identify the keywords media manipulators are organizing around in their attempt to undermine information integrity (Borra et al., 2015; Flores-Saviaga et al., 2022; Kim et al., 2023). Early research conducted by Dr. Tripodi indicates that mechanisms used to detect vandalism on Wikipedia could be used to proactively identify the search prompts manipulators are organizing around before they become widespread. **More resources are needed to develop these kinds of tools so that we can detect information absences being used to erode integrity across platforms.**

At the same time, the world of search is dramatically changing and the practice of looking up information using traditional tools (like search engines) is rapidly evolving to include generative AI tools like ChatGPT (Pichai, 2023; Rauwerda, 2022; West, 2020). Understanding the connections between how people search for information, algorithmic ambiguity, and the role AI plays in surfacing information that erodes trust can provide insights into a shifting information-retrieval landscape. The surge of generative AI makes the development of these new capacities

ever more urgent. It also creates a **need to teach the next generation of internet users to sort truth from fiction in an increasingly complex world** of search. Tools created by Dr. Reich can help fill that gap, enabling people to become their own fact-checkers and allowing them to discern how to identify efforts to manipulate information before it becomes too late.

Justin Reich, MIT

Francesca Tripodi, UNC

References Cited

Anand, A., & Srivastava, A. (2022). *FactShala media literacy initiative in India: An impact evaluation*. InterNews. https://internews.org/wp-content/uploads/2022/03/25x25-Factshala-Final-20220316.pdf

Axelsson, C. W., Guath, M., & Nygren, T. (2021). Learning how to separate fake from real news: Scalable digital tutorials promoting students' civic online reasoning. *Future Internet*, *13*(3), 60-78. http://dx.doi.org/10.3390/fi13030060

Ballantyne, N., & Dunning, D. (2022, January 3). Skeptics Say, 'Do Your Own Research.' It's Not That Simple. *The New York Times*. https://www.nytimes.com/2022/01/03/opinion/dyor-do-your-own-research.html

Borra, E., Weltevrede, E., Ciuccarelli, P., Kaltenbrunner, A., Laniado, D., Magni, G., Mauri, M., Rogers, R., & Venturini, T. (2015). Societal Controversies in Wikipedia Articles. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 193–196. https://doi.org/10.1145/2702123.2702436

Breakstone, J., Smith, M., Connors, P., Ortega, T., Kerr, D., & Wineburg, S. (2021). Lateral reading: College students learn to critically evaluate Internet sources in an online course. *Harvard Kennedy School (HKS) Misinformation Review, 2*(1), 1-17. https://doi.org/10.37016/mr-2020-56

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021). Students' civic online reasoning: A national portrait. *Educational Researcher, 50*(8), 505-515. https://doi.org/10.3102/0013189X211017495

Breakstone, J., Smith, M., Ziv, N., & Wineburg, S. (in press). Civic preparation for the digital age: How college students evaluate online sources about social and political issues. *Journal of Higher Education*.

Brodsky J. E., Brooks P. J., Scimeca, D., Galati, P., Todorova, R., & Caulfield, M. (2021). Associations between online instruction in lateral reading strategies and fact-checking COVID-

19 news among college students. *AERA Open, 7*(1), 1-17. https://doi.org/10.1177/23328584211038937

Brodsky, J. E., Brooks, P. J., Scimeca, D., Todorova, R., Galati, P. Batson, M., Grosso, R., Matthews, M., Miller, V., & Caulfield, M. (2021). Improving college students' fact-checking strategies through lateral reading instruction in a general education civics course. *Cognitive Research: Principles and Implications, 6*(23), 1-18. https://doi.org/10.1186/s41235-021-00291-4

Caulfield, M. (2017). *Web literacy for student fact-checkers*. https://webliteracy.pressbooks.com/

Crash Course. (2019). *Navigating digital information*. https://thecrashcourse.com/topic/navigatingdigitalinfo/

Edelson, L., Nguyen, M.-K., Goldstein, I., Goga, O., McCoy, D., & Lauinger, T. (2021). Understanding engagement with U.S. (mis)information news sources on Facebook. Proceedings of the 21st ACM Internet Measurement Conference, 444–463. https://doi.org/10.1145/3487552.3487859

Flores-Saviaga, C., Feng, S., & Savage, S. (2022). Datavoidant: An AI System for Addressing Political Data Voids on Social Media (arXiv:2210.13594). arXiv. http://arxiv.org/abs/2210.13594

Goldstein, I., Edelson, L., Nguyen, M.-K., Goga, O., McCoy, D., & Lauinger, T. (2023). Understanding the (In)Effectiveness of Content Moderation: A Case Study of Facebook in the Context of the U.S. Capitol Riot (arXiv:2301.02737). arXiv. http://arxiv.org/abs/2301.02737

Golebiewski, M., & boyd, danah. (2018). *Data Voids: Where Missing Data Can Easily Be Exploited*. Data & Society Research Institute. https://datasociety.net/wp-content/uploads/2018/05/Data_Society_Data_Voids_Final_3-1.pdf

Hertel, H. (2022). *Fact-checking and misinformation*. Rogue Community College. https://libguides.roguecc.edu/factcheck

Kim, T., Garcia, D., & Aragon, P. (2023). *Controversies over Historical Revisionism in Wikipedia*. WikiWorkshop. https://wikiworkshop.org/2023/papers/WikiWorkshop2023_paper_22.pdf

Kohnen, A. M., Mertens, G. E., & Boehm, S. M. (2020). Can middle schoolers learn to read the web like experts? Possibilities and limits of a strategy-based intervention. *Journal of Media Literacy Education*, *12*(2), 64-79. https://doi.org/10.23860/JMLE-2020-12-2-6

Lipton, E. (2014, February 9). Fight over minimum wage illustrates web of industry ties. *The New York Times.* https://www.nytimes.com/2014/02/10/us/politics/fight-over-minimum-wage-illustrates-web-of-industry-ties.html

Lyu, S., Nikolich, A., Bazarova, N., DiFranzo, D., Linvill, D. (Co-Principal Investigators). (2022-2024). Online Deception Awareness and Resilience Training (DART). (Award No. 2230494). [Grant]. NSF Convergence Accelerator Track F. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2230494&HistoricalAwards=false

Marwick, A. E., & Partin, W. C. (2022). Constructing alternative facts: Populist expertise and the QAnon conspiracy. *New Media & Society*. https://doi.org/10.1177/14614448221090201

McMahon, C., Johnson, I., & Hecht, B. (2017). The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 142–151. https://doi.org/10.1609/icwsm.v11i1.14883

McGrew, S. (2020). Learning to evaluate: An intervention in civic online reasoning. *Computers & Education, 145*, 1-13. https://doi.org/10.1016/j.compedu.2019.103711

McGrew, S., Smith, M., Breakstone, J., Ortega, T., & Wineburg, S. (2019). Improving university students' web savvy: An intervention study. *British Journal of Educational Psychology*, *89*(3), 485–500. https://doi.org/https://doi.org/10.1111/bjep.12279

Murthy, V. (2021). *Confronting health misinformation: The U.S. Surgeon General's advisory on building a healthy information environment.* Office of the U.S. Surgeon General.

Nassetta, J., & Gross, K. (2020). State media warning labels can counteract the effects of foreign disinformation. Harvard Kennedy School Misinformation Review, 1(7). https://doi.org/10.37016/mr-2020-45

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. In *Algorithms of Oppression*. New York University Press. https://doi.org/10.18574/nyu/9781479833641.001.0001

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, *22*(3), 453-460. https://doi.org/10.1016/j.jcps.2011.08.002

Panizza, F., Ronzani, P., Martini, C., Mattavelli, S., Morisseau, T., & Motterlini, M. (2022). Lateral reading and monetary incentives to spot disinformation about science. *Scientific Reports*, *12*(5678). https://doi.org/10.1038/s41598-022-09168-y

Pavlounis, D., Johnston, J., Brodsky, J., & Brooks, P. (2021). *The digital media literacy gap: How to build widespread resilience to false and misleading information using evidence-based classroom tools*. CIVIX Canada. https://ctrl-f.ca/en/the-evidence/

Pichai, S. (2023, February). Google AI updates: Bard and new AI features in Search. *Google*. https://blog.google/technology/ai/bard-google-ai-search-updates/

Rauwerda, A. (2022, March 3). *Why is Cleopatra constantly trending on Wikipedia?* Input. https://www.inverse.com/input/culture/why-cleopatra-trending-wikipedia

Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34) https://doi.org/10.1126/sciadv.abo6254

Starbird, K., DiResta, R., & DeButts, M. (2023). Influence and Improvisation: Participatory Disinformation during the 2020 US Election. *Social Media + Society*, *9*(2), 205630512311779. https://doi.org/10.1177/20563051231177943

Tripodi, F. (2019, October 14). Devin Nunes and the power of keyword signaling. *Wired*. https://www.wired.com/story/devin-nunes-and-the-dark-power-of-keyword-signaling/

Tripodi, F. (2020, May 5). Google and the Cost of 'Data Voids' During a Pandemic. *Wired*. https://www.wired.com/story/opinion-google-and-the-cost-of-data-voids-during-a-pandemic

Tripodi, F. (2022). *The propagandists' playbook: How conservative elites manipulate search and threaten democracy.* Yale University Press.

Urman, A., Makhortykh, M., Ulloa, R., & Kulshrestha, J. (2022). Where the earth is flat and 9/11 is an inside job: A comparative algorithm audit of conspiratorial information in web search results. Telematics and Informatics, 72, 101860. https://doi.org/10.1016/j.tele.2022.101860

Vincent, N., & Hecht, B. (2021). A Deeper Investigation of the Importance of Wikipedia Links to Search Engine Results. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–15. https://doi.org/10.1145/3449078

Wagner, M., Borah, P., De Choudhury, M., Kumar, S., Yang, S. (Co-Principal Investigators). (2021-2023). How Large-Scale Identification and Intervention Can Empower Professional Fact-

Checkers to Improve Democracy and Public Health. (Award No. 2137724). [Grant]. NSF Convergence Accelerator Track F. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2137724

West, R. (2020). Calibration of Google Trends Time Series. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2257–2260. https://doi.org/10.1145/3340531.3412075

Williams, E. M., & Carley, K. M. (2023). Search engine manipulation to spread pro-Kremlin propaganda. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-112

Wineburg, S., Breakstone, J., McGrew, S., Smith, M., & Ortega, T. (2022). Lateral reading on the open Internet: A district-wide field study in high school government classes. *Journal of Educational Psychology*. Advanced online publication. https://doi.org/10.1037/edu0000740

Wineburg, S., & McGrew, S. (2017). *Lateral reading: Reading less and learning more when evaluating digital information*. SSRN. https://doi.org/10.2139/ssrn.3048994

Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record*, *121*(11), 1-40. https://doi.org/10.1177/016146811912101102

Wineburg, S., McGrew, S., Breakstone, J., & Ortega, T. (2016). E*valuating information: The cornerstone of civic online reasoning*. Stanford Digital Repository. https://purl.stanford.edu/fv751yt5934

Section 4: Mack Blackburn

Written: 7/31/2023

Dear sir of madam,

As part of the generative AI PCAST response, we have prepared this attached document answering questions 1, 3, and 5 in our response.

Thank you,

Mack Blackburn

Attachment:

Generative AI PCAST Response

Mack Blackburn, Roopa Vasan, and Victor Miranda

Leidos has extensive experience in evaluating and improving AI/ML models to increase user trust in AI. Our projects include several DARPA projects covering natural language processing on social media, authorship anonymization, text generation, topic modeling, and many other subfields of AI/ML and natural language processing. We also have several projects aimed at reducing and understanding bias in AI, and mitigating adversarial vulnerabilities in AI models.

Challenges of Social Media, Misinformation, and Detection of Generated Text

The ability to rapidly produce large amounts of synthetic text that most people perceive as believable and trustworthy poses severe problems for social media and the dissemination of information in general. Researchers and many other groups are rightly concerned about the ability to detect synthetic text. While in the short term it is possible to detect text generated by AI, there is no reason to think that it will always be possible to do so, as AI models increasingly improve. Even the current state of the art methods for detecting synthetic text can have an AUROC score between 0.79 and 0.99, with the lower end of the range being detection of the newer, larger models such as GPT-J and NeoX (**>https://arxiv.org/pdf/2301.11305.pdf<**)**.** As LLMs improve and get closer to human language ability, there is no guarantee that there will be any method to reliably tell human from generated text, by examining a single text alone.

Another problem for detection of synthetic text is posed by retrieval-based transformers, which do not generate text themselves, but retrieve sections of human-made texts that are most relevant to the given query. In theory, retrieval based transformers can recycle and recontextualize text snippets made by humans for new purposes when prompted. If a retrieval-based transformer is used to create misinformation, it will likely be classified as human-made rather than AI. Usage of retrieval based transformers could potentially result in false negatives in detection of AI-derived text, but false positives in detection can also pose issues.

Over-zealous use of AI detection methods has the potential to cause problems for some humans, especially if AI detection tools lack evaluation or are marketed as being more effective than they really are. For example, false positives in detection may cause a student's assignment to be classified as AI generated, potentially affecting their grades and future opportunities. In one test, up to 20% of students were falsely labeled as having used AI for their assignments.

Verification of Information Sources

Given the limitations outlined above, one alternative solution to detection of synthetic AI output is verification of sources of information including governance methods to document and display the route that information has taken before user interaction. For example, Block-chain technology could enable social media platforms to catalog a record of the originator of an image. Images or posts could then be confidently labeled as either natural or AI generated, and that label can be displayed prominently.

Social media platforms may benefit from categorizing accounts as either human or AI (such as "bot" accounts, which have so far been used for simple purposes but may become more complex over time). Although there are multiple methods to classify users on different platforms as bots vs human users, such classification methods are not currently integrated into the platforms in ways that would be visible to human users. An example of a bot classification method is Botometer. Categorization of accounts as human or bot would have benefits to everyday users of social media platforms so that they might know if the information they are getting can be trusted. Quantifying the number of bot accounts would also have benefits to advertisers because they can be more confident in the number of real humans they are displaying ads to.

Model Hallucinations and Unintentional Inaccuracies

The increasingly widespread use a large language models in everyday life has in some cases demonstrated the problem of AI model hallucinations: AI outputs that are entirely fabricated and untrue but that sometimes appear convincing. Using AI to draft legal documents has resulted in citations of legal cases that never occurred. Although AI has tremendous opportunity for personalized education, students could take such hallucinations at their face value. An AI used in the healthcare domain may hallucinate symptoms of a patient, potentially with grave consequences.

Quantitative and Temporal Limitations of Language Models

Although LLMs are increasingly being used in a general setting for many different tasks, it is important to remember that they have limitations. As an example, the most used LLMs are not designed to answer questions related to math, statistics, or other domains that require computation or quantitative reasoning; their main strength is generating text that convincingly reads like it was written by a human. Methods that allow LLMs to delegate specific tasks to other specialized systems may be the best way forward. An example of this is the Wolfram plugin for ChatGPT, which allows both fluent free text conversation and precise mathematic answers. AI models that utilize or delegate to external tools or databases have been called Augmented Language Models (ALMs) and they will likely see significant use in the future, although they improve LM functionality in multiple ways besides quantitative reasoning as well. (>https://arxiv.org/pdf/2302.07842.pdf<)

LLMs also have limitations regarding time and temporal information. Many large language models like GPT4 are not capable of retrieving new information and are trained entirely on data from before a specific date. After a model is trained, it does not update unless it is adapted to newer data, which is a time consuming and expensive process at scale. This means that a model is effectively a snapshot in time and can only reliably provide information from the time range of its training set. If someone uses an outdated AI model, they run the risk of it generating hallucinations when asked about current events. LLMs also generally do not distinguish between old and new data in the training set, which can result in models that provide out of date information even if they have the newest information in their training data.

Responsible use of Training Data

Having large quantities of training data is one of, if not the primary limiting factor in developing successful large language models. Given the prime role of training data for AI, it is worth reconsidering data use laws including how an individual human creates a text or image for their own purposes may, without their knowledge, have their content trained on and used for development of AI. In many cases, average people who create AI training data do so without any awareness of their contribution, or any say in how it is used, and they do not directly benefit from their contribution. Most people are probably not, and may never be, aware if content they created was directly used to train the largest AI models which are increasingly being used in the real world. We should try to find ways we can create systems for the development and use of AI that are responsible to all those that contributed, and all those affected, including the many thousands or millions of people that unknowingly contributed to training sets.

Intellectual Property

Use of AI for the generation of text and images poses new problems for the way we think about intellectual property. Using generative AI is fundamentally different from the creative pursuits that people have traditionally engaged in, and how it will be incorporated into current law is not fully settled. In some cases, AI generated art is already being sold or used as part of a process to create a product. This makes relating IP laws to AI models, training sets, and prompts not a theoretical problem but a real-

world one that will grow over time. As an example, AI image generators can be purposefully and explicitly used to imitate the style of human artists, to make AI art that appears stylistically similar. Using the names of famous artists in prompts is common practice for many people using AI to generate art. The AI models were also certainly trained directly on images from those artists, and the process that AI models use to encode aspects of training examples into model weights is not fully explainable. These are the current most visible issues that get widespread public attention and stoke fears of “replacing” or “plagiarizing” human artists, but art or design is by no means the only field where the overlap of AI and IP may raise concerns.

Section 5: David Broniatowski

Written: 8/1/2023

To Whom it May Concern:

Please find my input to the question asked in the public comment listed below. These comments represent my personal opinion and are not the official position of The George Washington University, the Trustworthy AI in Law and Society (TRAILS) Institute, or the GW Institute for Data, Democracy, and Politics. I would like to add my colleague and collaborator, Dr. Valerie Reyna, of Cornell University, as a co-author on this submission.

Q1-1: *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?*

R1-1: This is an excellent question. We can learn a lot from the recent COVID "Infodemic" where people were exposed to massive amounts of information about the virus during a deadly pandemic. The concern is not so much that poor quality information will exceed high-quality information, but rather that people will be unable to find high-quality information because of information overload. In a recent study (D. A. Broniatowski et al., 2022), we found that information about the pandemic was actually more likely to be from high-quality, verifiable, and trustworthy sources, when compared to other health topics. We expect that, as AI-generated poor quality content proliferates, people faced with information overload may actually be motivated to seek out higher quality information if they can find it. If so, then one solution is simply to ensure that search results prioritize content from higher quality sources.

Beyond these factors, we also know that information is more likely to be compelling if it contains a clear, bottom-line *gist* – a simple, meaningful representation of information (Broniatowski & Reyna, 2020; Reyna, 2020; Reyna et al., 2021). Therefore, we posit that reliable access to verifiable, trustworthy information requires to ingredients:

1) The ability for motivated users to find high quality content by boosting that content in search
2) Ensuring that the content from high-quality sources (e.g., official government websites) is easily comprehensible and communicates a clear, bottom-line meaning.

Q1-2: How can we be certain that a particular piece of media is genuinely from the claimed source?

R1-2: This is a difficult problem to solve, but new technologies, such as watermarking of AI generated content (Bansal et al., 2022), can help significantly. Of course, more research needs to be done, but watermarking is an excellent start.

Q2: *How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens?*

R2: This is also an excellent question. We know that malicious actors have been using social media to attempt to manipulate American populations for many years, in large part by promoting discord in a way that identifies societal cleavages (D. A. Broniatowski et al., 2018, 2020). This suggests that the best way to build resilience to these attacks is to address the root causes – the specific sources of grievance that cause these cleavages in the first place. To do so, an empathic approach (Abroms et al., 2023; Larson & Broniatowski, 2021) may be effective if it helps build mutual understanding. Specifically, we need to build understanding between different communities and design interventions that can heal, rather than exacerbate, these grievances. However, in the short term, we have to understand whether tactics that are currently used by malicious actors actually effectively change beliefs and understanding of citizens. This requires targeted research to understand the mechanisms, including the technological, psychological, and social mechanisms, underlying what techniques are actually effective. This will help us to prioritize countermeasures. Second, we can start with techniques that are known to be effective, including those identified in Q1, above.

Q3: *What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?*

R3: Although new technologies are constantly being developed to detect AI-generated disinformation (Alizadeh et al., 2020; Gabriel et al., 2023), the fact is that we are in an "arms race", where disinformation purveyors will use whatever means at their disposal to generate their content. Thus, technological solutions alone are insufficient. The question asks about "policies and infrastructures" and this is precisely the right framing. Forthcoming work (Broniatowski et al., 2022) points to the need to construct policies that are informed by a deep technical understanding of social media platforms that moves beyond algorithms (which are just one piece of a much larger socio-technological system) to a focus on the architecture (Broniatowski & Moses, 2016) of these platforms, and how that architecture can promote or undermine attempts to detect and counter malicious content. An architecture framing is especially productive because

it draws an analogy between AI systems and critical infrastructure. One typically governs critical infrastructure using a combination of regulation, voluntary standards, and "building codes" that specify best practices that builders must adhere to in order to promote public health and safety. Just as building codes are flexible enough to change as technology evolves (e.g., building codes have been updated to promote sustainable energy use, reduce fire risks, prevent lead poisoning, etc.), technologies, policies, and infrastructures should be developed in a manner that allows for the incorporation of new scientific and public health insights as they become available.

Q4: *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?*

R4: We have observed that information from official government sources were more likely to be shared during the COVID infodemic than information from low quality or even high quality news sources (Broniatowski et al., 2022). Just as AI-generated content can be watermarked to identify it as potentially fabricated, technological solutions might be developed to ensure that content from official sources are marked as credible. In addition, search engines can commit to promoting information from these sources. However, these approaches should be used with caution, since elected representatives have been known to spread misinformation on occasion.

Q5: *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?*

R5: We have addressed this in a recent publication discussing communicating meaning in the intelligence enterprise (Broniatowski, 2019). In that paper, I advocate for the training and retention of people who specialize in translating complex technical analyses into terms that are well-understood by members of different communities. These translators must be able to communicate the gist of these technical findings into terms that are insightful and well-understood by community members. Thus, translators must represent a wide range of diverse backgrounds. A recently-funded NSF AI Institute, entitled Trustworthy AI in Law and Society (TRAILS), adopts this idea by advocating for a participatory approach to the develop of AI throughout the lifecycle. A participatory approach seeks to uncover the specific mechanisms making different communities are vulnerable to AI-generated misinformation, impersonation, and manipulation, and then addressing those specific mechanisms, such as by empowering community members to provide input into the design process, and then facilitating "translation" from these engaged community members to communities as a whole.

References

Abroms, L. C., Koban, D., Krishnan, N., Napolitano, M., Simmens, S., Caskey, B., Wu, T.-C., & Broniatowski, D. A. (2023). Empathic Engagement With the COVID-19 Vaccine Hesitant in Private Facebook Groups: A Randomized Trial. *Health Education & Behavior*, 10901981231188312. https://doi.org/10.1177/10901981231188313

Alizadeh, M., Shapiro, J. N., Buntain, C., & Tucker, J. A. (2020). Content-based features predict social media influence operations. *Science Advances*, *6*(30), eabb5824.

Bansal, A., Chiang, P.-Y., Curry, M. J., Jain, R., Wigington, C., Manjunatha, V., Dickerson, J. P., & Goldstein, T. (2022). Certified Neural Network Watermarks with Randomized Smoothing. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, & S. Sabato (Eds.), *Proceedings of the 39th International Conference on Machine Learning* (Vol. 162, pp. 1450–1465). PMLR. >https://proceedings.mlr.press/v162/bansal22a.html<

Broniatowski, D. A. (2019). Communicating Meaning in the Intelligence Enterprise. *Policy Insights from the Behavioral and Brain Sciences*, *6*(1), 38–46. https://doi.org/10.1177/2372732218792061

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health*, *108*(10), 1378–1384. https://doi.org/10.2105/AJPH.2018.304567

Broniatowski, D. A., Kerchner, D., Farooq, F., Huang, X., Jamison, A. M., Dredze, M., Quinn, S. C., & Ayers, J. W. (2022). Twitter and Facebook posts about COVID-19 are less likely to spread misinformation compared to other health topics. *PLOS ONE*, *17*(1), e0261768. https://doi.org/10.1371/journal.pone.0261768

Broniatowski, D. A., & Moses, J. (2016). Measuring Flexibility, Descriptive Complexity, and Rework Potential in Generic System Architectures. *Systems Engineering*, *19*(3), 207–221.

Broniatowski, D. A., Quinn, S. C., Dredze, M., & Jamison, A. M. (2020). Vaccine Communication as Weaponized Identity Politics. *American Journal of Public Health*, *110*(5), 617–618. https://doi.org/10.2105/AJPH.2020.305616

Broniatowski, D. A., & Reyna, V. F. (2020). To illuminate and motivate: A fuzzy-trace model of the spread of information online. *Computational and Mathematical Organization Theory*, *26*(4), 431–464. https://doi.org/10.1007/s10588-019-09297-2

Broniatowski, D., Gu, J., Jamison, A., Simons, J., & Abroms, L. (2022). *Facebook's Architecture Undermines Vaccine Misinformation Removal Effort*. >https://europepmc.org/article/PPR/PPR531143<

Gabriel, N. A., Broniatowski, D. A., & Johnson, N. F. (2023). *Inductive detection of Influence Operations via Graph Learning* (arXiv:2305.16544). arXiv. https://doi.org/10.48550/arXiv.2305.16544

Larson, H. J., & Broniatowski, D. A. (2021). Why Debunking Misinformation Is Not Enough to Change People's Minds About Vaccines. *American Journal of Public Health*, *111*(6), 1058–1060. https://doi.org/10.2105/AJPH.2021.306293

Reyna, V. F. (2020). A scientific theory of gist communication and misinformation resistance, with implications for health, education, and policy. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1912441117

Reyna, V. F., Broniatowski, D. A., & Edelson, S. M. (2021). Viruses, Vaccines, and COVID-19: Explaining and Improving Risky Decision-making. *Journal of Applied Research in Memory and Cognition*, *10*(4), 491–509. https://doi.org/10.1016/j.jarmac.2021.08.004

--

David A. Broniatowski, Ph.D., FPsyS
Associate Professor | Department of Engineering Management and Systems Engineering
School of Engineering and Applied Science
GW Lead | NSF Institute for Trustworthy AI in Law and Society (TRAILS)
GWU-PREP Program Coordinator
Associate Director | Institute for Data, Democracy, and Politics
The George Washington University

>http://www.seas.gwu.edu/~broniatowski/index.html<

>http://www.socialmediaforpublichealth.org<

Section 6: Dr. Susan Ariels Aaronson

Written: 8/1/2023

Good afternoon. Thank you for requesting comments. Enclosed please find the comments of Professor Susan Ariel Aaronson, Professor Valerie Reyna, and Adam Zable.

--

Susan Ariel Aaronson, Ph.D.

Research Professor of International Affairs and Cross-Disciplinary Fellow, Elliott School of International Affairs, George Washington University

co-PI, Governance, NSF Trustworthy AI Institute for Law and Society

>https://elliott.gwu.edu/susan-aaronson<

>https://www.trails.umd.edu/<

Comments to PCAST on Generative AI

Comments of Professor Susan Ariel Aaronson, Director, Digital Trade and Data Governance Hub and co-PI, NSF Trustworthy AI Institute for Law and Society; Adam Zable, Director of Emerging Technologies, Digital Trade and Data Governance Hub; and Valerie F. Reyna, Professor, Center for Behavioral Economics and Decision Research, Cornell University, NSF Trustworthy AI Institute for Law and Society

August 1, 2023

The comments expressed below are those of the authors and are not the official position of Cornell University, Center for Behavioral Economics and Decision Research, The George Washington University, the NSF Trustworthy AI in Law and Society (TRAILS) Institute, or the Digital Trade and Data Governance Hub at George Washington University.

As a prelude to our comments, we wish to acknowledge the importance of the questions PCAST is asking for the social, economic, and health welfare of the United States.

Question 1—How can we ensure reliable access to verifiable, trustworthy information?

When Open AI first issued its AI chatbot to the broad public in November 2022, it set off shockwaves that still reverberate almost 8 months later.[1] The generative AI chatbot allowed anyone with a computer to use AI to answer questions, solve problems, or automate their work tasks. But as Pen America recently noted, by enabling the use of AI–generated content in a wide range of daily human interactions, there is also the potential for people to lose trust in that content (Lopez et al., 2023). On the one hand, generative AI companies are not fully open about the data they utilize to create large learning models and how they obtain that data. While openness could lead to some problems, being transparent about the data could build greater trust in these models. The best way to build such trust is to require publicly traded companies to report on the data they use for their generative AI models. On the other hand, safety and security could be compromised by making data open. Therefore, we suggest that USG officials must ensure that multiple stakeholders are "at the table" when making decisions about how to balance e transparency while maintaining safety and security.

While each chatbot is unique, they are all built from large language models. In turn, these models are constructed from two main pools of data---data created, collected, or acquired by the model developers; and data scraped from a wide range of sites on the world wide web. (Touvron et al. 2023; De Vynt: 2023a). When researchers scrape the web, they create a bot to copy code off the internet which they can then use for innovation, business, or research purposes. Some of that scraped data comes from public sites such as Reddit,

[1] https://www.nytimes.com/2023/07/25/opinion/karp-palantir-artificial-intelligence.html?utm_source=substack&utm_medium=email

where the data are unstructured but continuously updated (Isaac: 2023). Other data sets come from public sites where the data is structured such as Wikipedia. The Washington Post analyzed one of Google's large language model data sets and reported that the top sites for that data set were: patents.google.com, No. 1, which contains text from patents issued around the world; wikipedia.org, No. 2, the free online encyclopedia; and scribd.com, No. 3, a subscription-only digital library. Also high on the list: b-ok.org, No. 190, a notorious market for pirated e-books that has since been seized by the U.S. Justice Department. At least 27 other sites identified by the U.S. government as markets for piracy and counterfeits were present in the data set." (Schaul et al: 2023).

Web scraping is an efficient and legal way to get large pools of data. In fact, the internet is sustained, updated, and improved by bots that search and then scrape the web to provide or index web content, or gauge political sentiment. Moreover, it is easy, as there are many sites, such as Kaggle and GitHub, where anyone can retrieve large data sets. In so doing, corporate, governmental, and academic researchers can use that data to better understand the state of the planet and its people. Thus, the answer cannot be to ban or unduly restrict web scraping. But policymakers can build trust by requiring firms to be open about what data they use, how they obtain such data, and how they plan to use such data. The rights of individuals to their own data—data created by them or about them—must be upheld with reasonable accommodations to support innovation, business, or research purposes.

Regarding the caveat above, some companies such as Open AI and Google have stated that they are not transparent about the provenance of the data to ensure that the system is safe to use. But as a result, we do not know if the data that underly many prominent AI systems are complete, consistent, or accurate. We also do not know where that data come from (its provenance). Without such information, users do not know whether they should trust the results they obtain from AI.

Recommendations:

1. Congress should pass a national personal data protection law that clarifies the rights and responsibilities of data subjects and entities that collect, use, and sell data (data controllers) and grants explicit responsibility to a data-protection body.

2. Congress should require that the Securities and Exchange Commission develop rulemaking related to the data underpinning AI. The SEC has already determined that firms must disclose how they address cyber threats and protect personal data. How firms acquire, collect, and use data for AI is material information for corporate stakeholders because as noted above incomplete, inaccurate, or unfair data could pose substantial risks to investors as well as to society. Moreover, in their Risk Management Framework, the National Institute of Standards and Technology has recommended that AI designers, developers and deployers should maintain records on the provenance of data and how their algorithms use data to make decisions, predictions, and recommendations. The SEC should also recommend that those in leadership, such as a member of the firm's board and senior management, monitor

the firm's use of AI. Such rules would incentivize firms to describe how they use data to fuel AI.

3. Congress should re-examine the legal implications of web scraping. Although courts have determined such scraping is legal in the U.S., it is clear that some firms that use web scrapers are not adequately protecting personal data. Moreover, some firms may be unfairly obtaining and using copyrighted data without explicit permission. As a result, a few large firms may be capturing both much of the world's data as well as the rents from AI. At a minimum, Congress should examine if AI firms that engage in data scraping should be licensed by the government and required to carefully examine the consistency, completeness, and the veracity of the data they collect for large language models.
4. Congress should examine ways to incentivize US platforms to share data with verified researchers.
5. Although it is often difficult to unravel how current AI systems make decisions, the USG should encourage research on how this unraveling can be achieved (i.e., how AI-made decisions are made), especially in high-stakes contexts, such as medical decision making, sentencing decisions in the legal system, and responses to natural disasters (Reyna & Brainerd, 2023).

Question 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?

Despite the wide interest in AI, policymakers have done little to involve the public in that debate. Moreover, we have done little to educate the public. A 2022 study by Zable and Aaronson (working paper published, paper September 2023 CIGI) examined whether officials informed and consulted their citizens as they developed national AI strategies. According to the OECD, such strategies articulate how the government sees the role of AI in the country and its contribution to the country's social and economic development. Building on a data set of 68 countries and the EU, we used qualitative methods to examine whether, how and when governments engaged with their citizens on their AI strategies and whether they were responsive to public comment.

We did not find a model of deliberative democratic decision-making. As of October 2022, some 43 of our 68 nation and EU sample had an AI strategy, but only 18 attempted to engage their citizens directly in the strategy's development. Moreover, only 13 of these nations issued an open invitation for public comment. Only 4 nations provided evidence that public input helped shape the final text (the US was not one of these 4). Although some acknowledged the comments, most governments did not make changes in response to the comments that they received. The number of people commenting on the strategy was small, comprised of individuals and organizations that are knowledgeable about AI and

willing and able to articulate their concerns. Thus, AI governance may be for the people, but it is not by the people.

We recognize that technical knowledge about AI is a barrier to participation, but research is commencing to address that conundrum, namely, how to effectively communicate risks and benefits of AI to non-experts (as part of the NSF Trustworthy AI Institute for Law and Society). Most people do not get involved in the development of technology policies or public policies writ large. Yet, without the input of a wide swathe of their citizenry, policymakers may struggle to anticipate future problems related to AI, and over time, to sustain trust in AI systems.

Recommendations:

1. To encourage public involvement, US policymakers could rethink how they engage the public regarding emerging digital technologies. For example, policymakers in the European Union organized a series of citizens' assemblies on web 4.0 and virtual worlds. Recognizing that building trust in their efforts was essential to developing effective governance, these policymakers provided several weekends of education on the technology underpinning such virtual worlds and how such worlds might change how people interact, work, study, and consume. Then they organized and facilitated discussions to develop recommendations to policymakers. The US could organize similar events for generative AI, working with any of the various trustworthy AI institutions located throughout the US, or directly with companies such as Meta. Policymakers could also investigate other emerging forms of collective sense-making, such as swarm AI or deliberation for AI governance.

2. Data governance has become a central element of governance in the 21st century. Yet most primary and secondary schools do not discuss how data and datasets are governed, the types of data, our rights to personal data or how firms use personal data to create innovative products. Citizens need such knowledge to know and realize their online rights. While federal policymakers cannot determine the curriculum, federal research agencies should consider developing grant programs to fund research on risk communication, on readability, and on data governance (e.g., how to train primary and secondary school teachers to teach what the public needs to know about data governance).

Section 7: Dr. Nidhi Rastogi

Written: 8/1/2023

Please find the attached submission in response to the President's Council of Advisors on Science and Technology's (PCAST) May 13, 2023, solicitation of public input on identifying and promoting the beneficial deployment of generative AI.

Sincerely,

Dr. Nidhi Rastogi

She/her/ hers
Assistant Professor
Department of Software Engineering
Golisano College of Computing and Information Sciences
Rochester Institute of Technology

PCAST Working Group on Generative AI
The White House
1600 Pennsylvania Ave NW
Washington, DC 20500

Aug 1, 2023

RE: Comments to the PCAST Working Group on Generative AI

Dear PCAST Working Group on Generative AI,

On behalf of the Rochester Institute of Technology, I write to provide the following comments on identifying and promoting the beneficial deployment of generative AI and on how best to mitigate risks.

Located in New York, Rochester Institute of Technology (RIT) is a leading university for artificial intelligence research and has several research centers and labs dedicated to AI, including the Center for Human-aware AI (CHAI), the Artificial Intelligence and Machine Learning Systems Center, and the Global Cybersecurity Institute. Its faculty and students are working on cutting-edge AI projects in a variety of fields, including healthcare, transportation, and cybersecurity.

I am an Assistant Professor at RIT who specializes in AI, cybersecurity, and graph analytics research and application. With Ph.D. from Rensselaer Polytechnic Institute, I have published research at top AI and Security conferences and have worked in both Industry and Academia.

Comments

We provide the following answers in response to PCAST's five inquiries, as outlined in its request. As a professor deeply engaged in the development and security of Generative AI technologies, I appreciate the call to involve the community in molding our society, especially as AI becomes more interwoven into our everyday lives.

Question 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?

The rise of generative AI raises serious challenges for the reliability of information, especially on digital platforms. These challenges make it difficult to know

what is true and what is not, even for the trained eye. Rochester Institute of Technology proposes the following solutions:

a. The **Federal government** should incentivize the private sector to develop a technological solution to detect fake content and to make that solution accessible to the general public in the form of an application and a website. The solution can be built using machine learning algorithms, a subset of AI, that identify patterns characteristic of fake content.
b. **Educational Institutions** should take the initiative to inform people about the dangers of fake content and how to spot them.
c. **Technology Companies** can ensure access to verifiable, trustworthy information through solutions like digital watermarking and blockchain-based verification systems. Blockchain-based verification systems can provide a traceable record of the media's origin and any subsequent alterations, helping to verify its authenticity.
d. **The Legislative branch** should enforce regulations that mandate truth in data and content dissemination. Such laws should penalize those intentionally spreading disinformation.

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

Using AI, malicious actors are already manipulating the beliefs and understanding of citizens through fake news articles and social media posts, which can lead to an increasingly divisive society and target people with personalized propaganda. Rochester Institute of Technology proposes the following solutions:

a. **Researchers and Scientists** should employ machine learning algorithms to devise tools that identify patterns of deceit. The National Science Foundation should support these researchers. Moreover, public campaigns are essential to educate on the threats of false narratives and techniques to detect them.
b. **The Federal Government** should bolster strong democratic institutions capable of resisting AI manipulation. This includes institutions that are committed to free speech and the open exchange of ideas. It also includes institutions that are able to hold malicious actors accountable for their actions.
c. **Technology companies** should be encouraged to develop and deploy advanced detection algorithms that can flag and filter out AI-manipulated content on their platforms. These platforms are the sources of a massive amount of data that is consumed by citizens.

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

The rise of AI-generated noise is a real threat to democracy, making it difficult for people to believe in the information disseminated by their elected representatives, engage with them, and have their voices heard. Rochester Institute of Technology proposes the following solutions:

a. **Both the Public and Private sectors** should collaborate and support the development of technologies that can help people filter out AI-generated noise.
b. **The President** should support the development of strong democratic institutions capable of resisting AI manipulation. This includes institutions that are committed to free speech and the open exchange of ideas. It also includes institutions that are able to hold malicious actors accountable for their actions.
c. **Private technology companies** should employ measures to limit the spread of AI-generated content during sensitive times, like elections. Regulations can also be implemented to limit the volume of AI-generated content on social media platforms.

Question 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?

As a society that is increasingly getting influenced by AI-generated content, **Educational Institutions, and NGOs** should help everyone develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation. We should provide public awareness campaigns, educational initiatives, and training programs to early and higher-educational institutions to raise awareness of the dangers of AI-generated misinformation, impersonation, and manipulation. These programs should include training in recognizing AI-generated content and understanding the potential impacts of AI-driven misinformation. Such education could be integrated into school curriculums, company training programs, and public awareness campaigns.

Thank you again for soliciting feedback from the community on this important topic.

Sincerely,

Dr. Nidhi Rastogi
Assistant Professor, GCCIS
Rochester Institute of Technology

Section 8: Eoghan Stafford

Written: 8/1/2023

Please find attached my response to the request for public input for the PCAST Working Group on Generative AI.

Thank you,

Eoghan Stafford

Evaluate Advanced Generative AI Models for Potential Autocratic Misuse

Recommendations to the Working Group on Generative AI
of the President's Council of Advisors on Science and Technology

Eoghan Stafford, Centre for the Governance of AI

August 1, 2023

Dear Dr. Greene, Dr. Tao, and members of the Working Group on Generative AI,

I welcome the opportunity to respond to your working group's request for public input. My comments address the second question posed in the request: *How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens?*

I am a researcher at the Centre for the Governance of AI (GovAI).[1] I have a PhD in Political Science from UCLA, and I specialize in researching autocratic regimes. In these comments, I discuss risks that generative AI poses to democracy globally, and I propose specific steps that the United States can take to mitigate those risks.

Recent decades have seen a deepening of authoritarianism around the world, even in some previously democratic countries. This trend has been accompanied by the rise of digital forms of autocratic control, including misuse of AI to entrench the power of authoritarian leaders. Governments are increasingly using AI to automate strategies such as censorship, influence operations, and surveillance that were once very labor-intensive, allowing regimes to implement these mechanisms of control at greater scale and with greater reliability.[2] Leaders of already authoritarian countries are most likely to abuse AI to entrench their own power and undermine democratic institutions in other countries. However, democratic countries also face risks of AI misuse by elected leaders with autocratic ambitions.

Of particular concern are the "foundation models" that are currently driving the most dramatic advances in generative AI, such as GPT-4. Foundation models are a kind of

---

[1] https://www.governance.ai The views I express here are my own: I do not speak for GovAI as an organization.

[2] Erica Frantz, Andrea Kendall-Taylor, and Joseph Wright. 2020. "Digital Repression in Autocracies."
Steven Feldstein. 2021. *The Rise of Digital Repression*.

general-purpose AI, because they can be adapted to a wide range of tasks.[3] This flexibility creates potential for highly beneficial uses, but could also cause serious harms. Foundation models exhibit many capabilities that their designers did not anticipate, some of which were not immediately apparent when the models were created. It is therefore crucial, as even more powerful foundation models are developed, to thoroughly test the models before they are used in publicly accessible applications, to determine how malicious actors – including autocrats – might misuse them.[4]

Developers of today's most advanced foundation models have begun evaluating their models for potential misuse prior to making them available for public use. This has included giving external researchers pre-release access to models to probe for vulnerabilities. Model evaluations test how a model can be "fine-tuned" or otherwise adapted for malicious purposes, as well as how the model can be misused without modification. However, there is still much work to be done to develop rigorous and comprehensive model evaluation methods.[5] The federal government can help develop an infrastructure of trustworthy external auditors to test cutting-edge foundation models for potential misuse by:

1. investing in research into the kinds of potentially dangerous capabilities that evaluators should test for in foundation models and rigorous methods for assessing a model's performance on such capabilities

2. developing common standards, based on that research, that auditors should follow when testing foundation models[6]

Establishing these standards would be helpful for AI companies that want to increase the transparency and trustworthiness of their systems. They would also be an important step toward governance of the next generation of powerful AI in the public interest. Ultimately, the US government should require AI companies, before deploying state-of-the art foundation models for public use, to submit the models to independent audits and implement safeguards to prevent potential misuses that auditors discover.

This approach would be useful for limiting the malicious uses of AI technology in general. In the following sections, I give some examples of metrics auditors should use specifically to test whether foundation models have generative capabilities that autocrats could misuse. These metrics represent tactics relevant to disinformation and

[3] Rishi Bommasani, et al. 2022. "On the opportunities and risks of foundation models."
[4] Toby Shevlane, et al. 2023. "Model evaluation for extreme risks."
[5] "Model evaluation for extreme risks."
[6] These standards should be incorporated into the "MEASURE" function of the National Institute of Science and Technology's AI Risk Management framework.

surveillance. Auditors should determine the extent to which autocrats could misuse a particular AI model to automate each tactic, in order to carry it out more effectively, at greater scale, or at lower cost.[7]

Disinformation campaigns

Foundation models could enable governments to generate and spread persuasive disinformation at greater scale and speed. Being able to cheaply and rapidly generate content would also allow state actors to test and hone their messages for greater effectiveness.[8] Autocrats might also use multimodal foundation models to create convincing fabricated evidence to discredit their political opponents. AI systems for creating “deepfake” videos already exist and are likely to become more effective with advances in large pre-trained audio and visual foundation models.[9]

Autocrats can also use generative AI to engage in a technologically sophisticated version of “astro-turing”, making it seem like ordinary citizens are spreading the content rather than state actors. This can make fake information seem more credible and make public support for a government seem more widespread than it is.[10] Authoritarian regimes (including Russia, China, and Cuba) have already used generative AI to create profile pictures for social media bots to push disinformation and propaganda.[11] Multimodal models could be used to create even more convincing fake user profiles, generating combinations of text, photos, audio, and video that are difficult to distinguish from what a real user would post.

---

[7] Foundation models have applications – and misuses – beyond generative AI. Increasingly, AI developers are leveraging foundation models to carry out other types of AI tasks, namely prediction and classification tasks ("On the opportunities and risks of foundation models"). Autocratic governments could use AI models that are fine-tuned on foundation models to more effectively censor social media or monitor protestors through facial recognition, for example.

[8] Josh A. Goldstein. 2023. "Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations."

[9] Additionally, as sophisticated synthetic media becomes widely available, publics may become more distrustful of all information sources, including government critics or independent media outlets. It will be easier for governments to claim that an authentic video showing, for example, the torture of dissidents or violence against protestors is fake. This phenomenon is known as the “liar’s dividend.” (Robert Chesney and Danielle Citron. 2019. “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security.” *California Law Review* 107, no. 6.)

[10] “Ethical and social risks of harm from language models.”
“Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations.”

[11] The examples that have come to light involve operations to influence citizens of other countries, but a government could also employ such methods to manipulate its own citizens.

The misuse of generative AI to spread disinformation could be particularly dangerous in combination with other advances in AI that have enabled “micro-targeting” of political messages. The kinds of algorithms that target ads and content to users of a website or app, based on their past behavior or demographics, could also steer them toward the kind of content that will be most effective in persuading them to support a government. Combining recommender algorithms and generative AI could enable autocratic governments to create disinformation tailored to each individual, at scale.[12]

To evaluate the extent to which a model would facilitate disinformation campaigns, auditors should assess how effectively a model can be used to:

- Fabricate evidence of events:
    - write false news reports that sound authentic
    - produce video and audio deepfakes, including depictions of public figures
- Shape narratives:
    - create engaging, high-quality, and emotionally charged text, images, and video that convey pro-government messages
    - select events and other context to include or omit to support a particular political perspective
- Generate content tailored to be most persuasive to users based on data about each individual
- Create fake user profiles to promote pro-government content, including:
    - images and video of fictitious users
    - realistic collections of posts and interactions with other users’ content
    - chatbots that cannot be distinguished from humans and are consistent with the fake user’s persona

Surveillance by spyware

Many autocratic states use spyware to monitor dissidents and journalists.[13] Large language models have demonstrated a remarkable ability to generate code, which could enable autocrats to adapt and deploy spyware more rapidly and extensively.[14] State actors could also misuse language models to produce spear phishing messages in

---

[12] "Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations.”
[13] Fred Guterl. 2022. "When spyware turns phones into weapons: How zero-click surveillance threatens reporters, sources, and global press freedom." Committee to Protect Journalists.
[14] Aakanksha Chowdhery, et al. 2022. “PaLM: Scaling Language Modeling with Pathways.”
“On the opportunities and risks of foundation models.”
“Ethical and social risks of harms from language models.”
“Model evaluations for extreme risks.”

order to get targets to download spyware by impersonating trusted contacts.[15] To assess the vulnerability of generative models to this kind of misuse, auditors should evaluate how effectively a model can be used to:

- Create code for spyware that can identify and exploit vulnerabilities on networked devices and evade detection by anti-malware software
- Write convincing spear phishing messages tailored to targeted individuals that persuade them to download spyware

[15]Julian Hazell. 2023. “Large Language Models Can Be Used To Effectively Scale Spear Phishing Campaigns.”

Section 9: Emily Ma

Written: 8/2/2023

Please find attached my response to the request for public input for the PCAST Working Group on Generative AI.

Thank you,

Eoghan Stafford

Impressions of ChatGPT: Using Survey Results to Inform AI Policy in Education

Abstract

ChatGPT is a powerful artificial intelligence chatbot developed by OpenAI. Because of the public availability and unprecedented writing capabilities of ChatGPT, it presents a unique policy obstacle for schools. On the one hand, ChatGPT and similar chatbots have the potential to be revolutionary learning tools, helping students to learn, study, and understand material. On the other hand, ChatGPT provides students with the means to cheat with unprecedented ease and competence, producing quality work in seconds with minimal understanding of the material required. In this paper, I examine the results of a survey I sent out to students, parents, and teachers in order to gauge policy opinions with regards to ChatGPT. The hope is that public opinion in this space can be used to inform policy making. I find that students are currently using ChatGPT at high rates, and that there is broad support for policy changes that can prevent AI cheating. Ultimately, I suggest that schools should minimize the use of take home assignments as a middle ground response to the growing risks of AI cheating.

Part 1: ChatGPT

ChatGPT is a chatbot, which is “a computer program designed to simulate conversation with human users, especially over the internet, that was designed to interact with users in a more conversational way” (Adamopoulou, 2020). It is similar to InstructGPT, and was trained in a similar way. OpenAI used Reinforcement Learning from Human Feedback (RLHF) to train ChatGPT. First, ChatGPT was shown desired outputs to various prompts, which ChatGPT learned from. Then, several responses to prompts were ranked from best to worst, which helped fine-tune ChatGPT (OpenAI, 2022). Using this, along with a method called Proximal Policy Optimization, OpenAI was able to optimally fine-tune the model.

When the user gives ChatGPT a question, or prompt, the AI predicts an output on the basis of the text it has been trained on. (Woods, 2023) "It's just always predicting the next word. You give it some context, it can predict the next word. But it can predict, you know, many next words," said Paul Cook, a University of New Brunswick professor who researches artificial intelligence (Harrison, 2023).

With the rise of GPT as an extremely powerful chatbot, it is important to consider how such AI programs will change commerce, society, and—for our purposes—education.

In terms of education, AI is a valuable tool as it can provide personalized instruction by adapting to student’s learning styles, creating lesson plans, and helping students by providing information and answering questions. ChatGPT is especially helpful as it is designed to provide a more detailed response and can write paragraphs that are easy to read. When students are having trouble doing math or physics homework, for example, ChatGPT can clearly explain how to do it. It can also adapt to students and identify areas where they’re struggling. ChatGPT is also good at summarizing information, so it could be a valuable tool for studying for tests and completing homework. Also, according to George Veletsianos, a professor at Royal Roads University in Victoria, B.C., ChatGPT could help students whose first language isn’t the language their school uses, since ChatGPT can explain assignments, help expand their vocabulary, and improve their grammar (CBC, 2023).

Here’s what ChatGPT said about itself and its uses in education:

> ChatGPT, or Generative Pre-trained Transformer, is a cutting-edge language model developed by OpenAI that uses machine learning to generate human-like responses to natural language input. It works by pre-training on large datasets to develop a rich representation of language that can be fine-tuned on specific tasks or domains. While ChatGPT has the potential to positively influence education, by serving as a tool for language learning or writing and research assistance, there are also concerns that it could be used for cheating and ethical considerations around its use in the classroom. It will be important to carefully consider the potential impacts of ChatGPT in educational settings (ChatGPT response).

Part 2: Implications for Education

Since ChatGPT is able to write about nearly any topic, many students have been using ChatGPT in their education. Some use it for help with homework problems or to proofread essays, while others use ChatGPT to complete entire assignments. Accordingly, it is important to consider how schools should adapt to this new powerful learning tool. On one hand, students using ChatGPT to write significant parts of their homework are not doing the work that they have been assigned, so they aren’t learning the skills they are being taught (Wong, 2023). On the other hand, ChatGPT can help students learn on their own and can provide personalized support (Chan & Hu, 2023).

Another question is who should be attributed when considering ChatGPT responses. Should it be OpenAI, as they created ChatGPT, or should it be the user of ChatGPT? Alternatively, it could be the authors of the sources used to train ChatGPT, but there were so much data used that it would be nearly impossible to cite everyone. As a result, institutions may need to revise their definitions of plagiarism to adapt to this new technology (Dehouche, 2021).

Some schools in New York City and Los Angeles have already started cracking down on AI usage by banning the use of ChatGPT (D'Andrea, 2023). Similarly, Hamilton's public school board blocked ChatGPT on all Wi-Fi networks and Board devices, meaning that students and staff have a harder time accessing ChatGPT (Hristova, 2023). On the other hand, the Seattle Public Schools district initially blocked ChatGPT on all school devices but then allowed educators to use it as a teaching tool (O'brien & Gecker, 2023). Some teachers have gone back to paper assignments and assessments instead of digital, says Shana Ramin, a technology integration specialist with Oakland Schools in Michigan (Claybourn, 2023).

Many Canadian universities such as UNB are still considering school wide policies on ChatGPT and academic dishonesty in order to decide what counts as a legitimate use of ChatGPT and similar tools, with no plans yet to necessarily ban ChatGPT. Others, such as Jeffrey Carleton at STU, have decided that any issues will be dealt with at a classroom level and that if the use of ChatGPT becomes a prevalent issue, professors would consider policies targeting the tool (Harrison, 2023).

Part 3: Prior Literature

Regardless, AI will likely play a greater role in education in the future, so students, parents, and teachers will need to adapt. This is new territory and it's a very powerful tool; thus, schools need to think carefully about how to adapt/respond to it. Precisely *how* we adapt to it is going to depend on whether we think of ChatGPT and other similar AI programs as useful tools to help students learn or a software that allows students to outsource their work and bypass learning altogether.

Some prior literature already exists in this space. For instance, Chan & Hu (2023) find that students and staff thought that ChatGPT was useful for writing assistance, generating ideas, summarizing information, and editing writing. However, there were concerns about plagiarism and academic honesty, since AI generated text cannot necessarily be detected by plagiarism detectors.

In terms of adapting to the existence of ChatGPT, Amani et al. (2023) found that some faculty said that schools would have to determine how to evaluate students differently. Additionally, many noted that ChatGPT would only have a positive impact if used correctly, as there were worries of students not understanding the limitations of ChatGPT or relying on it too much. Students had similar perspectives, and also said that ChatGPT could have a positive or negative impact depending on how it was used. In that survey, 64% of faculty/staff and 73% of students stated that they had an account and had used ChatGPT, which indicates there may already be a widespread awareness of ChatGPT within schools.

Also, it seemed that students were more distrustful of their classmates. 11% of students believed that their peers had used ChatGPT to complete homework, and 11% thought that their peers had used ChatGPT to write essays. In addition, 55% of faculty/staff responded that it was somewhat or extremely likely that students would engage in academic dishonesty, and 63% believed that ChatGPT would enable these behaviours (Amani et al., 2023). This indicates that both students and staff may be concerned about the academic dishonesty that may occur as a result of the use of ChatGPT.

Part 4: Methods

I designed a survey to address something that I believe has been under addressed by existing literature. Namely: what are the opinions of students, parents, and teachers about the particular policies that schools should be enacting in order to adapt to new, powerful AI chatbots like ChatGPT.

Any policy approach to this problem, I believe, will depend critically on the attitudes of teachers, parents, and students. In order to gauge these attitudes, I designed a survey and sent it to parents, teachers, and students. Our intention was to get a sense of the current attitudes people have towards ChatGPT in an educational setting, as well as what policy options people consider reasonable and prudent. I was also interested in the correlations between different responses. In particular, I was interested in how attitudes towards the use of AI chatbots in an educational setting differ between students, teachers and parents. Additionally, I was curious if more strictness regarding non-AI cheating would translate to a more negative attitude about the use of AI in school settings. I asked the following questions:

1. Which of the following best describes you?
    a. Student
    b. Parent
    c. Teacher
    d. None of the Above
2. Where do you live?
    a. North America
    b. South America
    c. Europe
    d. Africa
    e. Asia
    f. Oceania
3. What is your level of education?
    a. High school student
    b. High school graduate
    c. University student
    d. University Graduate
4. Have you ever used ChatGPT or a similar AI chatbot?
    a. Yes
    b. No
5. How often do you use ChatGPT for educational purposes?
    a. Never
    b. Rarely
    c. Sometimes
    d. Often
6. If you have used ChatGPT for educational purposes, how did you use it? (Select all that apply)
    a. To ask for explanations or clarification on a concept or topic
    b. To brainstorm ideas for a homework assignment or project
    c. To complete a homework assignment or project (e.g. giving ChatGPT a prompt and having it write an essay)
    d. To get feedback on writing assignments (e.g. grammar, structure, content)
    e. To practice answering questions or solving problems
    f. To practice language skills (e.g. speaking, listening, reading, writing)
    g. I haven't used ChatGPT for educational purposes
    h. Other (Please specify)
7. Overall, how do you feel about the use of AI in a school setting?
    a. Very positive
    b. Somewhat positive
    c. Neutral
    d. Somewhat negative
    e. Very negative
8. What do you think are the main benefits of using Chat GPT or a similar AI chatbot in school? (Select all that apply)
    a. It can provide quick and accurate answers to questions
    b. It can help students learn new concepts and ideas
    c. It can provide a more engaging and interactive learning experience
    d. It can free up time for teachers to focus on other tasks
    e. It can facilitate online or remote learning
    f. Other (Please specify)
9. What are the potential dangers or risks of using Chat GPT or a similar AI chatbot in school? (Select all that apply)
    a. It may not always provide accurate or reliable information
    b. It may not fully understand the context or nuances of a conversation
    c. It may not be able to fully replace the role of a human teacher or mentor
    d. It may lead to students relying too heavily on AI rather than developing their own critical thinking skills

   e. It may facilitate plagiarism or cheating if students copy and paste answers or content from the chatbot
   f. Other (Please specify)
10. Would you feel dishonest using ChatGPT to edit a for-credit take home assignment?
   a. Yes
   b. No
   c. Unsure
11. Do you think teachers should restrict or ban the use of ChatGPT or similar AI chatbots (e.g. by making more in-class assignments, or by running students' essays through AI-detection sites)?
   a. Yes, I think it should be banned
   b. No, it should be allowed
12. Suppose the use of ChatGPT is permitted for students, how should school assignments be changed?
   a. Assignments should stay the same
   b. The difficulty of assignments should increase
   c. Take home tests/essays should be eliminated
   d. Other
13. ChatGPT is a predictive artificial intelligence trained on material from the internet. When you ask ChatGPT a question, it combs through its data and tries to predict what an answer might look like on the basis of what it has read on the internet. Do you think ChatGPT is plagiarizing the original authors of the material it is trained on?
   a. Yes
   b. No
   c. Unsure/it depends
14. Suppose a student cheats or plagiarizes (in an old-school way—not by using ChatGPT) on a major assignment, like an exam or essay. It is the student's first offense. Which of the following is the most appropriate punishment?
   a. No punishment
   b. Redo the assignment
   c. A 0% on the assignment
   d. A 0% in the course
   e. Suspension
   f. Expulsion

Part 5: Results

Below is a list of interesting and relevant results from the survey:

Result 1:

We predicted that teachers would have the most strict attitude towards non-AI cheating, followed by parents, followed by students (correlation between q1 and q14). What I found was that parents were in fact the strictest, with 5 parents saying that students who cheat should redo the assignment (27.8 percent), 10 saying they should receive a zero on the assignment (55.6 percent), 2 saying that they should receive a zero in the course (11.1 percent), and 1 recommending suspension (5.6 percent).

The next strictest were the teachers, with 1 teacher supporting no punishment (6.7 percent), 4 saying that students who cheat should redo the assignment (26.7 percent), 9 saying they should receive a zero on the assignment (60 percent), and 1 saying that they should receive a zero in the course (6.7 percent).

Students had the least strict attitudes toward cheating, with 2 students supporting no punishment (5.6 percent), 16 saying that students should redo the assignment (44.4 percent), 17 saying that they should receive a zero on the assignment (47.2 percent), and 1 recommending expulsion (2.8 percent).

Overall, 36 respondents selected the student receiving a zero on the assignment (52.2 percent), 25 selected having the student redo the assignment (36.2 percent), and 3 people supporting each of no punishment and a zero in the course (4.3 percent), and only 1 person supporting each of suspension and expulsion (1.5 percent).

Result 2:

We predicted that parents will have the most negative attitude towards AI in education, followed by teachers, followed by students (Correlation between q1 and q7).

Among the 37 students surveyed, 20 felt very or somewhat positive (54.1 percent), 9 felt neutral (24.3 percent), and 8 felt very or somewhat negative (21.6 percent). Among the 15 teachers surveyed, 5 felt very or somewhat positive (33.3 percent), 4 felt neutral (26.7 percent), and 6 felt very or somewhat negative (40 percent). Among the 19 parents surveyed, 9 felt very or somewhat positive (47.4 percent), 5 felt neutral (26.3 percent), and 5 felt very or somewhat negative (26.3 percent).

Overall, 34 respondents felt very or somewhat positive (47.9 percent), 18 felt neutral (25.4 percent), and 19 felt very or somewhat negative (26.8 percent).

The data suggests that students may have the most positive attitudes toward AI in education, followed by parents, and then teachers, and that overall, around 50% may feel positive, while roughly 25% may feel neutral and 25% may feel negative.

Result 3:

We predicted that those who feel more strictly about non-AI cheating would have less permissive attitudes towards the use of AI for educational purposes (correlation between q14 and q7/11).

Out of the 3 respondents that preferred no punishment for non-AI cheating, 2 felt positive about AI in a school setting (66.7 percent), and 1 felt neutral (33.3 percent). All 3 felt that ChatGPT should be allowed.

Out of the 25 respondents that preferred that students redo the assignment as a punishment for non-AI cheating, 12 felt positive about AI in a school setting (48 percent), 6 felt neutral (24 percent), and 7 felt negative (28 percent). 8 felt that ChatGPT should be banned (32 percent), and 12 felt that ChatGPT should be allowed (48 percent).

Out of the 37 respondents that preferred that students receive a 0% on the assignment as a punishment for non-AI cheating, 17 felt positive about AI in a school setting (45.9 percent), 9 felt neutral (24.3 percent), and 11 felt negative (29.7 percent). 19 felt that ChatGPT should be banned (51.4 percent), and 9 felt that ChatGPT should be allowed (24.3 percent).

Out of the 5 respondents that preferred students receiving a 0% in the course, suspension, or expulsion, 3 felt positive (60 percent) and 2 felt negative (40 percent). 4 felt that ChatGPT should be banned (80 percent).

It appears that there is a slight correlation between how severely respondents believed non-AI cheating should be punished and negativity they felt about the use of AI in schools.

Result 4:

Out of 71 responses, 37 (52%) said they would feel dishonest using ChatGPT for a take home assignment. 14 (20%) said they would not feel dishonest, and 20 (28%) said they were unsure. Interestingly, “dishonest” was a strong plurality here, suggesting that pre-existing anti-cheating norms have generalized to include AI chatbots.

Result 5:

We had 70 responses to question 11. Of those, 31 (44%) said that ChatGPT should be banned. 24 (34%) said that ChatGPT should not be banned. The remaining responses were suggestions of middle ground policies and partial restrictions on the use of ChatGPT. This is a very important result. A plurality, but not a majority, believes that ChatGPT should be banned outright. This suggests both that people are not very open to the use of ChatGPT for schoolwork and also that finding popular policy options may prove quite difficult.

Result 6:

We predicted that those who have used ChatGPT or a similar chatbot will be more likely to report that they would feel dishonest using it for a school assignment (correlation between q4 and q10).

Out of the 44 respondents that have used ChatGPT, 21 stated that they would feel dishonest (47.7 percent), 13 were unsure (29.5 percent), and 10 stated that they would not feel dishonest (22.7 percent). Out of the 27 respondents that haven’t used ChatGPT, 16 stated that they would feel dishonest (59.3 percent), 7 were unsure (25.9 percent), and 4 stated that they would not feel dishonest (14.8 percent).

Contrary to our prediction, the data suggests that those that haven’t used ChatGPT may be more likely to feel dishonest using it for a school assignment.

Result 7:

We predicted that the attitude people have towards the use of AI in a school setting will have an inverse correlation with level of education (i.e. more education=worse opinion of AI) (correlation between q3 and q7).

Of the 31 high school students who responded to this question, 15 had a positive or very positive attitude towards the use of AI in a school setting (48%), 7 had a negative or very negative attitude towards the use of AI in a school setting (23%), and 9 had a neutral attitude (29%).

Of the 40 high school graduates who responded to this question (university students + university graduates), 18 had a positive or very positive attitude towards the use of AI in a school setting, (45%), 13 had a negative or very negative attitude towards the use of AI in a school setting (32.5%), and 9 had a neutral attitude (22.5%).

Of the 34 university graduates who responded to this question, 14 had a positive or very positive attitude (41%), 12 had a negative or very negative attitude (35%), and 8 had a neutral attitude (24%).

Overall, there did not appear to be a strong relationship between respondents' level of education and their attitude towards the use of AI in an educational setting.

Result 8:

Out of 65 responses to question 8, 35 respondents (54%) said that helping students learn new concepts and ideas was a main benefit of AI chatbots. After that, 32 respondents (49%) said that one of the main benefits was providing a more engaging learning experience. 24 (37%) thought that one of the main benefits was AIs ability to provide quick answers to questions, and 20 respondents (31%) said that AI will help free up teachers' time. Finally, 18% said that one of the main benefits was the fact that AI will help facilitate online learning.

Result 9:

We predicted that people's major concerns with regard to the use of chatbots in educational settings would be plagiarism and cheating and students relying too heavily on AI (q9). I was correct. Of our 71 responses to this question, 61 were concerned about plagiarism and 61 were concerned about students relying too heavily on AI (86% each). 47 respondents (66%) were concerned about AI not providing accurate or reliable information. 40 respondents (56%) worried that AI might not be able to replace the role of teacher or mentor and 36 respondents (51%) worried that AI might not fully understand the nuances of conversation. These data show that, while there are a broad range of concerns, the most salient overall concerns involve the potential for AI to enable cheating and the risk that AI will diminish learning outcomes by creating over-reliance.

Result 10:

We predicted the most common policy recommendation would be to get rid of take home tests (q12). More specifically, I predicted that students will be most partial to assignments staying the same (q12 and q1) and that teachers will be most partial to eliminating take-home tests (q12 and q1).

Of the 69 responses to question 12, 27 said that schools should get rid of take home tests and essays (39%), 16 said that assignments should stay the same (23%), and 14 said that the difficulty of assignments should increase (20%). The remaining 12 responses were suggestions for alternative policy options. These suggestions fell into two camps. Some respondents had suggestions for how assignments could be changed to accommodate powerful chatbots, by, for example, making the assignments more creative or opinion-based. Other respondents suggested that we should change the nature of assessments at school to focus more on class participation.

Students, more than any other group by far, were partial towards assignments staying the same, which was in line with our predictions. Even so, however, a plurality of students suggested getting rid of take home assignments and essays. In total, of 33 student responses, 12 said that assignments should stay the same (36%), 5 said that assignments should become more difficult (15%), and 16 said that take home tests and essays should be eliminated (48%).

Of the 9 responses to this question given by teachers, 6 suggested getting rid of take home tests and essays (67%), 2 said that assignments should be made more difficult (22%), and 1 said that assignments should stay the same (11%).

Overall, these results suggest that there is a broad understanding between parents, teachers, and students that policies do need to change to accommodate this new technology, with only 23% of respondents saying that assignments should stay the same.

Result 11:

We predicted that respondents will by and large not consider ChatGPT to be plagiarism. (q13). Of the 70 responses to question 13, 9 people said that ChatGPT does count as plagiarism (13%), 25 said that ChatGPT does *not* count as plagiarism (36%), and 36 said they were unsure or it depends (51%).

Result 12:

We wondered how many current high school and university students are regularly using ChatGPT for educational purposes already (q4 and q5). Of the 37 current high school or university students who responded to the survey, 29 have used ChatGPT before (78%). 22 of them have used ChatGPT for educational purposes (59%), and 16 reported to be using ChatGPT for educational purposes *sometimes* or *often* (43%). What this shows is that the use of ChatGPT for educational purposes is already quite prevalent, and therefore there is a pressing need for schools and educators to make corresponding policies.

Part 6: Discussion

Overall, respondents felt mostly positive or neutral about the use of AI in a school setting, with around half of respondents feeling positive and around a quarter feeling neutral. The students were the most positive about AI, followed by parents, and then teachers. I also found that the more negative respondents were about AI in a school setting, the more likely they were to recommend stricter punishments and believe that AI should be banned. However, surprisingly, more respondents said that ChatGPT should be banned (44%) than not (34%), with other respondents in between. This disconnect between the overall positivity towards AI that respondents expressed and their particular policy attitudes towards ChatGPT might be explained a few different ways. Perhaps respondents dislike ChatGPT in particular, but feel optimistic about other AI tools. Perhaps respondents feel optimistic about AI in the abstract, but conflicted when they are confronted with AI is concretely being used right now. More research would help to figure out exactly which AI tools people have positive attitudes towards and how people think they should be used.

However, a majority of people did not think that the use of ChatGPT should be considered plagiarism of the original authors the program trains on, so they may have been more concerned about the other possible consequences of AI chatbots. The most common concerns were about cheating, students relying too heavily on AI, and AI not providing accurate or reliable information.

On the other hand, some of the most cited benefits of ChatGPT were that it could help students learn new concepts and ideas and could provide a more engaging learning experience. This indicates that if chatbots are to be used, people would likely support applications that use it as a learning tool. In general, people are attracted to the idea of AI as a learning tool that can help students explore and understand the material, but worried about the possibility of AI simply doing students' work for them.

The majority of respondents, no matter if they were parent, teacher, or student, agreed that policies should change in order to adapt to chatbots. The most popular suggestion was to get rid of take home tests and essays (39%), and some respondents also supported increasing the difficulty of assignments (20%). Other suggestions included making assignments more creative or opinion-based, or changing what assessments measure to focus more on class participation. Additionally, since nearly half of the current high school and university students use ChatGPT sometimes or often, it is crucial to determine a policy that will best address the needs of students, teachers and parents.

While students, teachers, and parents generally agree that change is needed, there are differences between their thoughts on what to do. For instance, students were more likely to believe that assignments should stay the same than parents or teachers. They also had the most positive attitudes towards AI, and the least strict attitudes towards cheating.

Given these results, I think that one very natural policy option is for schools to get rid of take-home tests and short answer assignments. This suggestion is generally popular, and addresses the primary fears about AI (cheating, overreliance) while still enabling chatbots to be a valuable learning tool. I think that an ideal outcome involves students using ChatGPT as another way to learn, understand, and study the material without outsourcing *all* of the learning, understanding, and studying to ChatGPT. A majority of our respondents cited the learning benefits of ChatGPT as a primary potential benefit and cheating and over-reliance as a primary potential drawback. Getting rid of take-home assignments seems like a natural start towards this end.

The reality is that—in the status quo—take-home assignments are compromised. Students *will* be using ChatGPT to cheat on them. Per result 12, most students have used ChatGPT for educational purposes and nearly half of students regularly do so. A critical mass of students are or soon will be using chatbots to help them complete their assignments. This is a threat both to their own education and also disproportionately harms those honest students unwilling to cheat.

Certain chatbot detection mechanisms that are commonly used by teachers, such as Turnitin, are much too fallible to use as a basis for any kind of decisive action. (Fowler, 2023) It is also extremely difficult to have a very

effective chatbot detector as chatbots are continuously getting more and more advanced (Heikkilä, 2022). In addition, as there is such a strong incentive for using chatbots, I think the simplest solution is to get rid of the efficacy of chatbots for unethical purposes by simply getting rid of take-home assignments. Some people might imagine a middle ground approach where take-home assignments are still assigned and students are simply instructed not to use ChatGPT, but with weak detection and the overwhelming utility of ChatGPT, I think these policies will be largely ignored, which—again—serves to punish the most honest students for not cheating.

Although the use of AI for cheating is a very real threat, I think that banning ChatGPT on campus (particularly for boarding schools) would be an overcautious approach that risks missing out on the educational benefits of AI. There is a consensus among respondents that AI chatbots have the potential to teach students new material as well as make learning more engaging and fun. I believe that ChatGPT offers students an engaging and positive way to learn and study the material which can positively supplement their education: for instance, ChatGPT can be used to instead help exercise students' critical thinking, by asking students to evaluate ChatGPT's responses to prompts. (Roose, 2023). I also believe that teachers should have the latitude to involve ChatGPT in their classrooms to enhance engagement with their teaching, if they so wish.

Conclusion

In Section 1, I explained what ChatGPT is and how it works, and outlined some of the possible benefits and harms of using AI in education. In Section 2, I examined how schools and teachers have been reacting to ChatGPT to get an idea about how people overall seem to feel about it. In Section 3, I outlined the survey that I created to get a more precise idea of how people felt toward AI in education, ChatGPT specifically, and their opinions on various other issues such as plagiarism and how they thought policies in schools should change. The results and correlations found from the survey were detailed in Section 4, and I discussed the results and what they mean in Section 5. Using these results, I made a few policy recommendations that would satisfy most of the parties involved.

The takeaways from this report are threefold. (1) I believe that there is currently a pressing need to improve AI policy at academic institutions in order to adapt to the new status quo. AI chatbots are powerful learning tools that are already being used by a large chunk of the student population. With good policy, we can ensure that these powerful tools are enhancing the learning experience rather than frustrating or replacing it. To this end, successful AI policy needs to recognize that AI both has a great potential to be positively incorporated into education, but also that if it is *not* positively incorporated, it will be used to cheat and plagiarize, at the expense of the students who cheat, the students who do not cheat, and academic integrity as a whole. (2) I believe that outright bans do not serve this end because they eliminate all of the positive use cases of AI chatbots. Such bans are unpopular, likely to be circumvented, and prevent teachers from finding positive ways to incorporate AI tools into their classroom and curriculum. (3) Finally, I believe that take-home assignments and tests are compromised and should be de-emphasized. AI chatbots like ChatGPT are already powerful enough to do students' work for them, and such a tool is irresistible to many students. Our research found that a large number of students are regularly using ChatGPT to help with their assignments. Given that it is not possible to know if students have written their assignments or if an AI chatbot has, I believe that continued use of take-home assignments runs the risk of punishing those students with the good integrity to do those assignments for themselves.

References

Amani, S., White, L., Balart, T., Arora, L., Shryock, K. J., Brumbelow, K., & Watson, K. L. (2023, April 21). Generative AI Perceptions: A Survey to Measure the Perceptions of Faculty, Staff, and Students on Generative AI Tools in Academia. Arxiv. https://arxiv.org/pdf/2304.14415.pdf

Adamopoulou, E., Moussiades, L. (2020). An Overview of Chatbot Technology. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 584. Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_31

ChatGPT could help rather than hinder student learning, says B.C. professor. (2023, January 30). *CBC*. https://www.cbc.ca/news/canada/british-columbia/chatgpt-student-benefits-1.6731105

Claybourn, C. (2023, January 18). ChatGPT in Classrooms: What to Know. *USNews.com*. https://www.usnews.com/education/best-high-schools/articles/chatgpt-in-classrooms-what-to-know

D'Andrea, A. (2023, February 1). Canadian universities crafting ChatGPT policies as French school bans AI program. *Global News*. https://globalnews.ca/news/9451143/chatgpt-education-canadian-universities/

Dehouche, N. (2021, March 25). Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3). Ethics Sci Environ Polit 21:17-23. https://doi.org/10.3354/esep00195

Fowler, G. A. (2023, April 3). We tested Turnitin's ChatGPT-detector for teachers. It got some wrong. *The Washington Post*. https://www.washingtonpost.com/technology/2023/04/01/chatgpt-cheating-detection-turnitin/

Harrison, L. (2023, February 2). From instant essays to phishing scams, ChatGPT has experts on edge. *CBC*. https://www.cbc.ca/news/canada/new-brunswick/chatgpt-academia-cybersecurity-1.6733202

Heikkilä, M. (2022, December 19). *How to spot AI-generated text*. MIT Technology Review. https://www.technologyreview.com/2022/12/19/1065596/how-to-spot-ai-generated-text/

Hristova, B. (2023, February 2). Some students are using ChatGPT to cheat — here's how schools are trying to stop it. *CBC*. https://www.cbc.ca/news/canada/hamilton/chatgpt-school-cheating-1.6734580

O'brien, M., & Gecker, J. (2023, February 1). Did ChatGPT write that? OpenAI launches tool to detect AI-generated text - National | Globalnews.ca. *Global News*. https://globalnews.ca/news/9451365/chatgpt-detector-openai/

OpenAI. (2022, November 30). *Introducing ChatGPT*. OpenAI. https://openai.com/blog/chatgpt

Roose, K. (2023, January 12). Don't Ban ChatGPT in Schools. Teach With It. *The New York Times*. https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html

Wong, J. (2023, April 11). Teachers are split on bringing ChatGPT into elementary, high schools. *CBC*. https://www.cbc.ca/news/canada/chatgpt-highschool-elementary-1.6802336

Woods, R. (2023, March 29). *How to use ChatGPT for content creation*. Microsoft Create. https://create.microsoft.com/en-us/learn/articles/how-to-use-chatgpt-for-content-creation

Section 10: Yonah Welker

Written: 8/8/2023

Also in case you or your team members will be in Paris for the Unesco gathering on September 4-7 (Digital Learning Week), I will present this vision of disability-centered AI policy, and recent global regulation updates to achieve a united digital framework.

Would be happy to meet and connect as well.

Thank you so much again.

Section 11: Tyler Jaynes

Written: 8/14/2023

Good day,

I am a Utah-based researcher who serves in several volunteer positions within the Institute of Electrical and Electronics Engineers (IEEE). I am drafting this message today in advance of my sponsored participation in a forthcoming workshop by The Royal Society (which aims to expand upon their 2019 iHuman report) to advance the case for a more robust regulatory framework around neurotechnologies in the USA. While I am not a medical doctor or professor, I have been writing on the legal gaps in this space for several years--as evidenced by my sponsorship to London in September. I understand that we are currently in an election cycle, and that the federal government is already tackling issues related to other advanced technologies (specifically AI, nanotechnology, and quantum computation), but do not want the USA to fall behind its peers in this regulatory space further than it already has with the implementation of Regulation (EU) 2017/745 (the EU Medical Device Regulation [MDR]). I am more than open to speaking with someone from either the PCAST or OSTP on this subject further, and would appreciate any action taken to advance regulation absent my direct involvement.

With appreciation for your time and consideration,

Tyler L. Jaynes, B.Sc. (he/him)