CENTER on
NATIONAL
SECURITY
GEORGETOWN LAW

PIT UN Final Report

Social Media and Emerging Technology Project

June 2022

I. Goals, Objectives, & Results

Artificial Intelligence. Big Data. Third Generation Social Media. Each, on their own, presents challenges to our individual security. Combine these technologies and they have the potential to violently disrupt the rule of law and our collective security. Emerging technology allows us to create new life forms and killer robots. Modern social media companies capture what you read, what you believe, and what you do (and with whom). Using predictive algorithms, they can anticipate your friendships, your purchases, and which candidate you will support. Like their online retail counterparts, they have access to billions of records, creating the risk of mass surveillance, microtargeting, and identity theft. With the generation of new forms of data and communication, and our adversaries' ability to hack it, steal it, and use it, comes the opportunity to shape, disrupt, and cleave society, or segments thereof, threatening U.S. national security.

The lines between security, law, and technology have never been less clear, yet so important. The role of private industry in emerging technologies is dominant, but the negative effects on our democratic society are often an afterthought. Risks posed by social media are particularly pernicious for vulnerable groups, including alienation, addiction, false identities, deep fakes, and group/network manipulation.

Our goal with this project was to tackle these connected issues *as they are likely to manifest themselves in the future*. After a deep dive review of the hundreds of studies and reports looking at social media and the underlying technologies, we discovered that virtually all of them are backward looking, focused, for example, on disinformation/misinformation and the 2016 and 2020 elections. We took a different approach. Starting with deep research and over 60 interviews of very diverse stakeholders, we developed alternative future scenarios to push our Task Force - made up of a cross-section of decision makers, subject matter experts, and representatives of diverse sectors and disciplines - to "think forward and think differently" about how the risks and opportunities of social media will play out over the next 3-5 and even 10 years. We interviewed across racial, gender, age, geographic, able-bodiedness, and socioeconomic lines. Fundamentally, we wanted to ensure that every aspect of American online life was reflected and represented in our project.

Our specific objective was to allow our Task Force to frame the problems they saw as most important, informed by an agreed-upon set of principles and values, and generate creative governance solutions. To do so, we undertook an intensive, 20-phase process with our Task Force (discussed in Section IV), which not only produced creative governance solutions (see recommendations in Appendix A), but also created a team of high-level, well-networked change agents who are committed to taking forward our recommendations and expanding them internationally in phase II of our project (funded by a new grant for 2023-24).

We also wanted to pilot a new methodology for how to analyze and address complex national security challenges, a new way of thinking driven by diverse perspectives and design thinking tools. This approach is in stark contrast to the traditional, siloed - defense, intelligence, diplomacy - approaches in the national security field.

Over the course of 18 months, we have broken new ground in the field. Our specific lessons and recommendations are captured in significant detail in the attached appendices (described below). A few major takeaways, though, are worth calling out here.

- There is a significant gap in future-oriented thinking about the societal, legal, and security ramifications of our collective move toward an increasingly virtual world(s). Substantively, we learned that experts - even in tech and tech-adjacent fields - are not particularly attuned to *future* risks. The alternative future scenarios we developed were an extremely effective way at forcing the Task Force to confront and understand risks that are coming down the pike.

- After exploring third and fourth generation social media, the underlying technologies, and alternative future scenarios, we see social media as the "canary in the coal mine" for a much broader and complex set of risks attendant to the metaverse and future virtual worlds. Risks will likely metastasize because of the convergence of and interactions between technologies. Those risks will be exacerbated in their societal effects because the underlying technologies are being developed almost exclusively in the private sector, without public review and without checks and balances.

- The ramifications for our society are immense, leading to the Task Force's overwhelming emphasis on the need to build in democratic consultation and processes into the development, design, use, and oversight of emerging technologies. Throughout the process, we saw the need for the public to understand the implications of where the world is going. What happens when companies can identify and even manipulate our innermost thoughts through our online interactions? Will we have the ability to meaningfully live outside a virtual world? Do we need human rights in digital spaces, and what should they look like?

- In terms of methodology, our multi-faceted approach allowed us to take on a topic far too complex for traditional national security models. Our innovations allowed us to consider a broad range of perspectives, far-reaching and diverse technologies, and colliding threats rarely considered together. It enabled us to deal with complex systems, rather than an isolated, bite-size problem. Our stakeholder interviews produced some of the most creative ideas, and also brought to the fore indicators and signs of otherwise unforeseen connections. They helped us see around corners, understand the hidden glaciers, and consider viewpoints that national security practitioners would otherwise miss. The Task Force allowed us to stress test these creative ideas and keep focused on the big picture and the most important peaks.

- Going forward, we will likely adjust and refine our methodology to take advantage of the strengths of each approach, while streamlining and tailoring how we utilize the Task Force. This project should have had a budget of $300,000 over two years, and the strain on our small team was significant, even with significant volunteer help.

II. Challenges

Our challenges going forward include:

(1) Society must reconcile the lack of incentives tech companies have to change with the political gridlock in forcing accountability or transparency.

(2) Individual users exhibit a fundamental behavioral preference for short-term convenience and connection over long-term, broad-based security risks.

(3) There is a need for real resources to support backbone organizations that can galvanize the broad stakeholders needed for change. Current public interest technology resources are one-off, extremely limited, not flexible enough, and insufficient to be able to implement real change.

(4) There is a lack of talent willing to work in the public interest tech sector for the long term; the draw and pay of tech companies is just too strong.

III. Lessons Learned: Key Substantive Themes

In our Task Force meetings and interviews, we were able to forge connections across individual perspectives that brought new ideas to light. Among a wide variety of insights we plan to integrate into our final report and other messaging, four major themes emerged: 1) the potential hazards for our woefully underprepared cybersecurity infrastructure, 2) the need to incorporate ethics at the design phase of every emerging technology project, 3) the dangers of over-generalizing internet access and capacity (particularly in rural areas), and 4) the urgent need for further real-time study.

Though it was not an issue we initially anticipated exploring, cybersecurity was a key theme for an overwhelming number of stakeholders. There were extensive discussions on whether current private, personal, and government cybersecurity measures rise to the occasion given the rapid development of cyberattacks including Denial-of-Service attacks, malware, and ransomware. A number of private entities with limited cybersecurity capacity have access to tremendously sensitive data: we spoke with a high-end matchmaking service, an adult performer who offers personalized serviced with military clientele, and a well-respected industry insider who expressed his shock that Meta opts against a Chief Cybersecurity Officer (or did, until June 2022). The costs of protection, lack of technical capacity, and seeming inevitability of cyber breaches with no mechanisms for accountability all deterred various stakeholders from more actively engaging on cybersecurity, despite acknowledging the large societal consequences should they suffer a breach. Echoing our theme of ethical design, discussions also touched upon whether cybersecurity should be "baked into" the technology from the beginning and whether it has become too convoluted for users to employ. For example, we interviewed two international political activists, both noting that even though their safety frequently depends on good cybersecurity hygiene, it is too impractical and time consuming.

In addition, several stakeholders were wary of U.S. readiness when it came to critical infrastructure cyberattacks, particularly at the state and local government level, and their implications for U.S. national security. Such concerns were not surprising themselves, as the vulnerability of critical infrastructure has been raised as a national security concern for a very long time. Instead, we were met with the disbelief that despite this widespread knowledge, the same inertia that currently grips the private sector has seemingly stymied significant efforts to address the problem in critical public contexts as well, even with the benefits of knowledge of the stakes of the problem, government resources, and time.

Second, stakeholders sounded the alarm on rushed, profit-focused design of upcoming technologies. We discussed emerging technologies like the metaverse, Web 3.0, FinTech, general blockchain applications, AI/ML targeting, deepfake technology, and satellites. Multiple stakeholders highlighted the concept of ethical debt in technology, a current area of deep academic interest in the field, as one of the roots of our current tech challenges. These individuals saw promoting ethical design, through educational frameworks for coders and engineers, transparency of algorithms, altered investor incentives at the VC and angel levels, government regulation, and other channels, as the solution. Essentially, in the same vein that cybersecurity should be baked into the product from its inception, many stakeholders thought that engineers should be more mindful of the impact of their products and anticipate possible harms, especially when it comes to technologies that can have a substantial impact on our lives. One stakeholder framed it starkly: our generation's Golden Gate Bridge is being engineered by children–20 year olds often with no formal education!

The speed of innovation also raises unique challenges to ethical design. While in the physical world innovation has a life span that can take a generation to become obsolete (like moving from railroad tracks to buses and planes), in the digital world those cycles take between five to ten years. Unfortunately, the speed of adoption does not negate the importance of such infrastructure while in use. At the same time, the speed of innovation underlies a tension inherent in emerging technologies: sometimes they rapidly reinvent society, and sometimes they simply don't work. Plenty of stakeholders were quick to temper expectations as to how worried we should be about the risks and general applicability of tech like deepfakes and the metaverse. Whether skeptical of the timeline to commercial adoption, the actual capacity and reliability of such technology once deployed, or its likelihood of misuse, discerning hype from technological reality is an ongoing challenge–especially as technical experts themselves often hotly debate these issues. Regardless, all agree that ethical design principles are critical for emerging technology. The controversy arises over what that means in practice.

Third, we were consistently reminded that high-speed internet is far from ubiquitous in America, especially in rural areas. Large parts of the country cannot access the immense benefits its greatest proponents claim it confers. A local official in a rural state pointed out that there are too many barriers to internet access for rural areas, including high costs for subscribers, broadband leaders being restricted by regulations, and inadequate incentives for companies to upgrade their systems for remote residents. The lack of high speed internet was particularly felt during the Covid-19 pandemic, which resulted in a considerable reliance on telehealth. According to one stakeholder, more than 90% of rural hospitals use telehealth services like tele-radiology and tele-pharmacies. Internet access issues plague more than just rural areas, however. As one example, in the middle of the Capitol Hill neighborhood in Washington D.C., 9.1% of the County do not have internet access. While we explore the implications of an increasingly connected world, recalling those who aren't as well connected, and the digital divide that exacerbates many broader societal issues, is important to take into account.

The fourth consistent theme we heard was the lack of research and understanding of these issues, even as we know they are reshaping our world. We are almost completely in the dark on how our internet experiences work and what that means for democracy and society. This manifests through: lack of transparency into algorithms and how emerging technology works (including the worrying fact that ***no one including their creators*** actually understands how the most complex neural networks function); lack of transparency around the extent of individual surveillance and the ability to deanonymize individuals across datasets; lack of understanding of the long term physical, mental, and emotional health impacts of this technology; lack of qualified people to conduct this research (especially given the hiring shortages and scramble among tech companies themselves); and lack of funding to research any of these issues or explore solutions to them beyond cursory grants and internal company work. As the internet becomes an increasingly fundamental part of everyday life for most Americans, in ways direct and indirect, we need to understand it and guide it for society's benefit. Almost universally, we heard that the government must step in with research funding and reporting, and that tech companies must pay their fair share to support these efforts. An avalanche of change is coming, and we simply aren't prepared with our current research and understanding.

It is hard to underestimate the importance of the internet. As one stakeholder aptly phrased the current state of affairs: "the internet is that important, and if it breaks, we're all screwed and the fact that it's being held together with tape is lunacy." The above key themes are the most important insights we pulled from stakeholder interviews, but they are not the only ones. We hope that our work will provide a first look into the future, and a solid foundation for future efforts to guide our society's approach to emerging technology in an increasingly virtual age.

IV. Lessons Learned: Innovation Processes for National Security

Our Process:

Over the course of this 18 months, our process has yielded significant scholarship that we hope can drive thinking around the future national security risks of emerging technologies. Our project will have produced an innovative report with a comprehensive solutions strategy, built strong relationships across many diverse Task Force members, planted our flag for a new type of national security thinking, and cultivated our own network of interesting individuals across the space. For an overview of our full design process, see Appendix E: Design Methodology for Task Force. Looking at the individual steps more specifically, we began with extensive literature reviews and stakeholder interviews with approximately 60 diverse experts. We then developed high quality thought leadership through a 20 phase Task Force process:

1. Task Force pre-meeting reading materials (Introductory meeting);
2. Introductory Task Force meeting;
3. Two-on-one interviews with each member;
4. Task Force pre-meeting reading materials (Second meeting);
5. Second Task Force Meeting;
6. Network contact requests for additional stakeholder interviews;
7. Task Force pre-meeting reading materials and pre-work idea generation process (Third meeting)
8. Third Task Force meeting;
9. Task Force pre-meeting reading materials and ranking activity (Small group discussions);
10. Small group discussions on government and the private sector;
11. Task Force pre-meeting reading materials, including written feedback on initial idea lists, rights lists, and draft code of conduct (Fourth meeting);
12. Fourth Task Force meeting;
13. Additional written feedback on newly developed recommendations, idea lists, rights lists, and draft code of conduct;

14. Task Force pre-meeting reading materials (Rights discussion);
15. Small group rights discussion;
16. Written feedback on recommendations draft;
17. Written feedback on full report draft;
18. Final draft report comments;
19. Dissemination assistance through member networks; and
20. Official rollout event support and further report amplification.

Overall, this was an incredibly valuable pilot that took on a topic far too complex for traditional national security models. Our model, particularly in the national security space, is entirely novel, and it successfully considered a huge variety of perspectives as well as a huge variety of different technologies and colliding threads in the national security picture rarely considered together.

Insights for the Future:

In terms of process, we had many takeaways about the applicability of general design elements to a complex, evolving future issue in national security. We also came away with a number of insights around Task Force composition to tackle these unique, 360-degree types of challenges. We were lucky that our Task Force members developed strong rapport, and were highly committed to the process. We are excited to continue working with them through future phases of the report, and benefited tremendously from their work. Here, we offer insights from them, as well as our own observations, that we hope can guide our future engagement on these complex, thorny issues.

Looking at general design, we made the following broad observations. First, small group meetings and interviews are good for developing focus, generating new ideas, surfacing diverse perspectives, and identifying nuance. Second, large group meetings are good for building consensus, evaluating detailed proposals, and stress-testing scenarios. They require surgical precision around key questions/deliverables to offer value. Third, pre-meeting materials engaged Task Force members to varying degrees, and written work outside the meetings was not often taken on. The most useful pre-meeting materials were the shortest, and we should take care to limit pre-meeting materials to 15 minutes of reading/pre-work time. Fourth, 20 phases in a process, especially on top of stakeholder interviews, literature reviews, and other research channels, requires a significant investment in labor and materials. A team of 6 is the minimum necessary to accomplish a fully successful, comprehensive policy planning and ideation process.

Our unique Task Force composition also yielded innovative insights, and we hope to continue to build on that model in the future. Our Task Force involved three “types” of members, though many individuals blurred across categories: high-level individuals, subject matter experts, and non-traditional background experts. We found that high-level individuals provided key value through their own credibility/status conferring onto the report; sharing their network connections; and "war stories" from their individual experiences. Subject matter experts provided key value through in-depth knowledge of how ideas would fit into the existing landscape; quick, trustworthy analysis of areas that piqued Task Force curiosity; useful nuance and guidance to make the report relevant and compelling; and more subject matter engagement and creative ideas. Non-traditional background experts provided key value through creative ideas or considerations that the rest of the Task Force hadn't seen; emphasis in the report on concepts or ideas that other Task Force members didn't originally see as so important; and grounding in "reality" outside of expert bubbles to reorient discussion at various points away from elite-focused concerns. For example, raising questions like why we are so focused on virtual reality taking over the world when 15 percent of Maine households don't even have internet service?

This combination of Task Force member enabled robust discussion and new perspectives we might not otherwise have surfaced, and we want to continue to iterate on this model in future Task Forces. To best use their skills, we may separate subject matter experts from high level Task Force members on a future Task Force to give us the ability to drill down on their substantive contributions. For example, we found that positive interactions with our team and the process were more conducive to getting buy-in from high-level individuals than giving them full ownership over the substance and ideas of the report, while subject matter experts seemed to delight in jumping in and even offered to write large sections of the report themselves for the review of the rest of the group. We were also very happy with how much better our report was for its focus on diversity. One area we struggled with was that, on the Task Force itself, it's really hard to cover "every" nontraditional background–especially because minorities are not monolithic, and there can be wide divergence of opinion within any group. We addressed these blindspots with stakeholder interviews, but we want to continue to find ways to center underrepresented voices in these discussions beyond giving their expertise simple consideration–all such voices need ownership. Ultimately, we will continue to experiment with new ways to integrate non-traditional backgrounds into a Task Force's work that balance the power conferred by membership with the need to consider a wide variety of perspectives. We think that considering future "Expert Advisor" or other roles, in addition to traditional Task Force membership roles, is the best way to make use of subject matter and nontraditional experts in this critically important work.

For a future Task Force then, we are honing a process that focuses on our four key needs:

1. using Task Force expertise most effectively,
2. helping Task Force members buy into the report,
3. not asking too much commitment from Task Force members during the content development stage, and
4. making efficient use of our own limited resources.

Our current prototype for a future successful Task Force, like our efforts in the international space, is as follows:

1. Pre-meeting reading materials (30 minutes, backgrounders);
2. Task Force kickoff meeting, with introductions to each other and ample open discussion on key issues that will set the stage for breakout groups;
3. Pre-meeting reading materials (5 minutes, only basis for discussions);
4. Small group interviews/breakouts of ~4 TF members and a moderator on various topics prioritized at the first meeting;
5. Network contact requests for additional stakeholder interviews;
6. Pre-meeting reading materials (15 minutes, read through our initial documents with the option to give written feedback);
7. Big Task Force meeting to get feedback on initial document, surface thorny issues/areas of disagreement, and identify areas for further clarification/research;
8. Pre-breakout reading materials (5 minutes, only basis for discussions);
9. Optional interviews/breakouts of ~4 TF members and a moderator to dig more deeply into the new set of issues, only as TF members choose to participate;
10. Pre-meeting reading materials (20 minutes, circulate an updated, more fleshed out document to guide discussion and identify areas where we need final consensus);
11. Big Task Force meeting to reach final consensus, build support for next steps (report dissemination and promotion) as a group;
12. Written feedback on draft of report;
13. Final draft report comments;
14. Dissemination assistance through member networks; and
15. Official rollout event support and further report amplification.

V. Lessons Learned: Prototyping a National Security Virtual Simulation Tool

Georgetown's Center on National Security has over a decade of experience developing national security crisis simulations that help a range of stakeholders understand and confront future risks. Our PIT UN funding allowed us to apply that expertise to a new set of risks and a new set of stakeholders attendant to emerging technology. In collaboration with the MIT Game Lab, our two teams worked together to:

- Develop the parameters and affordances of a potential virtual social media game
- Review the open source applications that could serve as an initial prototype
- Develop a budget and timeline for the full game
- Develop ideas for storylines for a potential game

These ideas from our collaboration with MIT Game Lab are included in Appendix C.

We discovered a few key takeaways from our collaboration. First, we discovered that our primary audience or client - the Task Force and other stakeholders - understand with some degree of sophistication the ways in which risks attendant to today's technologies *currently* manifest themselves. For example, the Task Force and similarly situated policymakers are very familiar with disinformation and the ways in which bad actors can manipulate social media channels to produce certain outcomes. Second, we discovered that our target audience was far *less* familiar with the emerging technologies under development, and lacked an understanding of how those new technologies could be used in the future to undermine individual and societal-level security. And finally, we discovered that the virtual game we were contemplating with MIT would take far too long to build and was not flexible enough to keep pace with the changing technologies that would allow us to unpack future risks in semi-real time.

These discoveries allowed us to make an important shift in our planning. Namely, we were able to pivot to develop future scenarios that could be presented to the Task Force without significant time delays and development costs. This pivot - developing alternative future scenarios (including those in our forthcoming Task Force Report) - produced some of the richest insights from the Task Force and for our team. As a result, we used the remaining funds to do a deep dive into (1) various methodologies for developing alternative future scenarios by establishing an informal partnership with the Institute for the Future (Appendix C); and (2) a particularly important technology - blockchain - and its ramifications for future risks (e.g., cryptocurrencies especially for marketplaces in virtual reality platforms, Web3, data storage, supply chain monitoring, creative content industries, decentralized organizations) (Appendix D).

Our design methodology - using a relatively small budget to explore an initial prototype and make necessary adjustments to our approach - was important in allowing us to identify and explore the kinds of tools that are most effective in helping a set of diverse stakeholders understand and unpack future risks.

VI. Certification

In compliance with Section VII(B) of our Agreement, we certify that: All Georgetown University activities conducted with the Grant funds were and are consistent with charitable purposes as set forth in Section 501(c)(3) of the Internal Revenue Code, and Georgetown University complied with all provisions and restrictions contained in this Agreement, including, for example and without limitation, those provisions relating to lobbying and political activity.

VII. Appendices

A. Task Force Recommendations for Governing Social Media and Emerging Technologies

B. Prototyping A Virtual National Security Simulation: MIT's High-Level Design Document

C. Developing Alternative Future Scenarios

D. Prototyping National Security Simulations for Blockchain

E. Design Methodology for Task Force: Fluid Hive Innovation Workshop Summaries and Analysis

F. Financial Report

Appendix A: Task Force Recommendations for Governing Social Media and Emerging Technologies

Building a Healthy Digital World: A Roadmap

Vision: A Healthy Digital World			
Foundation for Future Actions: Democratic Principles & Criteria to Evaluate Novel Solutions			
Pillars	Effective Governance	Responsible Platforms	Empowered Public
Goals	Research; Innovation; Accountability	Trustworthiness; Ethics; Respect for Rights	Education; Mobilization; Agency
Task Force Recommendations*	Fund Research and Grants to Address New Harms	Codify a Digital Bill of Rights and Developers' Code of Conduct	Foster Civic Education and Engagement
Tools for Enactment: 1. Sample Digital Bill of Rights, 2. Sample Developers' Code of Conduct, 3. Sample Information Governance Principles			

**All members of the Task Force have participated in their personal capacities and not on behalf of any other organization or entity. The recommendations reflect the sense of the Task Force as a whole and are not attributable to individual members. They further reflect the diverse expertise of the members, who work on different aspects of the final recommendations.*

It is the strong sense of the Task Force that in our rapidly evolving digital world, Americans need to protect our 1) democratic processes and institutions, 2) freedom of speech and Constitutional rights, 3) individual health and well-being, 4) trust in community and accurate information, and 5) ecosystems that foster innovation. Balancing such complex and ever-evolving considerations is difficult. Accordingly, we offer a multi-faceted strategy that incorporates both structural guides for the inevitable future efforts that will address new challenges and concrete steps for today. We also propose tools to assist on both fronts.

After cataloging the most concerning risks, the Task Force sought to establish the principles most critical to protect for a healthy internet ecosystem. Consistently returning to concepts of democratic legitimacy and rights, the Task Force developed a framework for evaluating potential solutions with a set of criteria that put democratic principles at the forefront. Applying the framework to potential solutions to our current problems, the Task Force whittled approximately 80 ideas down to the current recommendations, organized by their promotion of effective governance, responsible platforms, or an empowered public. Those recommendations aim to achieve three types of goals for each of those three pillars of the online ecosystem, also detailed in the recommendation sections. Finally, members of the Task Force undertook a first pass at the key tools that the recommendations highlighted as necessary, creating sample documents that express the democratic principles in different forms and can serve as the jumping off point for future discussion in these arenas.

A healthy digital world is possible. With good leadership, the rapid evolution of technology brings immense promise to improve all our lives. The following vision can help lay the groundwork.

I. Foundation For Future Actions

Our world is changing, fast. Competition among the "great powers" is no longer the singular critical national security concern, and even great power competition is no longer measured by a single dimension as nations vie for dominance in the arenas of commerce, innovation, infrastructure, and more. In this increasingly complex world, new tools must be adopted to bolster conventional instruments of power, like defense and diplomacy that alone can longer cannot ensure our security against the new harms discussed in this report that are capable of causing damage that is more far-ranging and insidious than tools of old. The most concerning example of such harms is the widespread concern about threats to American democracy. Novel challenges magnify the ongoing struggle to preserve liberty and underscore the role of non-government actors in protecting against potential threats to the United States.

The Task Force, cognizant of the current and increasingly complex threat environment, worked to strike a balance, elevating above all the national security concern of safeguarding American democracy. It crafted solutions with democratic principles top of mind, creating a more traditional policy evaluation framework around the ultimate goal of fostering democracy and democratic processes. We therefore offer the frameworks to support future ideation efforts. While the national security problems will undoubtedly continue to evolve, the need to center democratic principles and involve diverse actors will remain constant.

A. Democratic Principles

The Task Force began by asking what principles would be needed to protect a healthy digital ecosystem; however, the discussion quickly shifted to the need to think more broadly than online communications. Task Force members of all political persuasions and areas of expertise raised concerns about the survival of American democracy itself. Fundamentally then, the Democratic Principles outlined by the Task Force answer the question **"What principles are needed to maintain and foster a healthy democracy?"** Framed around online participation, the principles, below, infuse the Task Force recommendations and express the core values of this report.

1. Free Expression: The ability to express oneself without interference by a public authority is a treasured American freedom and a critical protection against authoritarianism. What today are considered bad ideas may be just that, but in both physical and virtual public squares, some of those unpopular ideas of today may become the cornerstones of public values tomorrow.

When internet platforms adopt the role of public fora, safeguarding free expression becomes more complex. Social media companies do not have the same constitutional obligations as the government and so have the ability to limit expression on their platforms as they see fit without violating the First Amendment . This can be a good thing, as content moderation reduces hate speech and weeds out misinformation, a concept enshrined in 47 US Code Section 230, shielding internet platforms from civil liability over their hosted content. However, as individual platforms grow to encompass massive swaths of the virtual public square, their sheer size and opaque control of algorithms, removal of content, and increasing practice of deplatforming individuals and organizations–both alone and in conjunction with other online platforms–can have a profound impact on public debate. Free expression is a thorny concept, with its own internal inconsistencies, with which all potential solutions must grapple.

2. Information Access: The ability to access information, including the information of one's choice, is an important facet of free expression that takes on added significance in the digital world. The most basic iteration of access is literal access to digital public spaces in the first place via a internet connection, and as internet applications become more resource intensive, a fast internet connection is also quickly becoming a necessity. In the modern age, internet access is necessary to provide true economic, social, and political participation, whether used for researching political issues or communicating with politicians. Unfortunately, equitable and affordable internet service for rural, socioeconomically disadvantaged, and marginalized populations remains elusive as commercial considerations drive how Internet access is made across communities nationwide.

Information access is also a question of free expression and information integrity, as individuals need to be able to reach the information they desire to inform their opinions unencumbered by the filters of an algorithm's value judgements on what to show them. As companies can increasingly dictate the information ecosystem of an individual without their conscious consent, democratic independence and truly free choice in public decision making are threatened.

3. **Information Integrity:** The ability for citizens to rely on news, academic, and other sources of their choice for credible, trustworthy information is foundational to democratic decision-making. Democratic engagement requires citizens to be able to make decisions based on available information. Regardless of whether it relates to societal, political, economic, scientific, or other matters, that information must be reliable, accurate, and complete, and citizens must be confident that it is so.

That reliance can become problematic when a citizen's informational bubble is curated by algorithms over which they have limited control or platforms which deliberately remove access to information and speakers outside of users' knowledge. Behavioral economics warns of "nudges," or the power of habits and subconscious cues that can direct individual behavior. Social media algorithms are analogous funnels to particular information, but they are dictated not by the individual's goals but by the company's, which may be more aligned with their advertisers than their users. As many critics point out, with many social media companies user behavior *is* the product. Platforms sell the ability to nudge that behavior. If citizens are relying on social media for news and information, the integrity of what is presented to them is potentially compromised by the goals of the companies doing the presenting. Even where a platform's intentions are entirely aligned with user goals, algorithms are products of our own unconscious biases and human error, a problem magnified when the teams creating and implementing these algorithms harbor preconceived notions or lack diversity. The siloing enabled by disparate algorithmic learning, moreover, may lead to contradictory information bubbles, with the result that users may have little faith in any information thus obtained–further laying the groundwork for authoritarian tendencies to grow. Information integrity can be achieved in a variety of ways. Citizens may benefit from understanding what is shown to them, why, and how they can change their individual information bubbles to fit their own goals, rather than those of the companies selling their behavior.

4. **Communal Trust:** Central to faith in democracy and its attendant political horse-trading is the idea that, at the end of the day, we are all on "Team USA," meaning we share a common set of values and beliefs that together form a national identity. The societal cohesion that develops from community formation is critical to maintaining public trust in the wisdom of the crowd, as well as, preventing slides into chaos, on the one hand, and authoritarianism on the other. From this broad perspective, developing new relationships, sharing interests and values, exploring new possibilities, and enjoying leisure time, are all positive ways of finding and reinforcing commonalities and communion that further democracy through connection.

5. Inclusion: One citizen, one vote applies to every adult, regardless of race, religion, gender, political views, sexual orientation, or ability. American democracy relies on collective action, and all are invited to participate. In the online context, inclusion can look different than in the physical world. Online threats and harassment, as well as the refusal to allow individuals to access online platforms or systemic economic forces that place such access out of reach, can stymie participation by marginalized populations, as can a lack of accessibility accommodations like screen-reader compatibility on webpages. As the digital world takes over more of our political, social, and economic lives, all people must be able to take part.

6. Institutional Trust: Protecting democratic institutions and processes must be a key goal for all democratic governments. The loss of public faith in those institutions and processes can have violent consequences, as illustrated by the January 6, 2021 U.S. Capitol attack. The internet has played a significant role in eroding American institutional trust, through the amplification of conspiracy theories, the strengthening of extremist recruitment, and the destruction of respected information sources. Institutional trust must be earned, and where broken, rebuilt, but it should not intentionally be undermined. Allowing for questioning and scrutiny while countering mis or dis-information and thwarting malicious actors is a difficult but vital task.

7. Security: Citizens need to feel safe in order to participate in voting processes, reflect on issues of import, and otherwise contribute to a democratic society. More than a democratic principle, the safeguarding of citizen health, be it physical, mental, or otherwise, is a critical component of any social contract between citizens and government. The enactment of legal protections tends to be reactive, not proactive. This is particularly true of digital users who increasingly need protection from both virtual and physical harm. Moreover, legal protection have not tended to be conceived of in a holistic manner, as evidenced by the patchwork of laws and law enforcement coordination around cyberstalking, revenge porn, cyberflashing, doxxing, phishing, hacking, and other internet threats. As these threats evolve and disproportionately thwart internet participation by certain populations, providing new forms of security to all becomes a critical democratic issue.

8. Privacy: Privacy in one's beliefs, thoughts, emotions, and sensations fosters personal development, self-reflection, and intellectual inquiry. It protects a sphere of intimate relationships and allows users to express themselves outside of the public eye. It also provides users with the space to learn about, debate, and decide how to approach matters that democracy requires its populace to address. Some users may be willing to divest themselves of certain matters related to privacy insofar as convenience and services can be better delivered when data is shared. And some invasions of privacy are sanctioned in the interest of security. But not all individuals are comfortable with these approaches. Just as privacy and other rights are constitutionally established to protect minorities, so too does the ability of users to engage online while still being able to mediate their bounds of intimacy and knowledge about their own behavior matter.

The stakes are high. Democracy cannot exist without dissent, and privacy creates the conditions for dissent to arise and thrive, within and among individuals as well as when people inject controversial ideas into the discourse. Dissidents of authoritarian regimes demonstrate this privacy imperative. Our discussions with international activists underscored how the ability to mask physical identity enables them to develop and spread messages through social media and the internet. One activist, concerned about their own unmasking on certain platforms, recounted how the Chinese government pressured a social media platform to remove posts critical of the Communist Party over complaints of anti-Han Chinese racism.

In the privacy sphere, social media platforms are faced with difficult choices around squashing fake accounts, sharing user data with advertisers, responding to government inquiries, integrating privacy into products without impacting user experience, and other concerns. Whether that status quo should continue is a matter of much debate.

9. Transparency: Public decision-making by those acting on the government's behalf facilitates democracy. Voters' and citizens' decisions, in turn, require information—not simply data, but whole data sets within the appropriate context. Transparency is critical to good governance and avoiding corruption.

As internet platforms play increasingly large roles in our lives and create new societal problems with which the public must grapple, private sector transparency becomes important as well. Algorithmic transparency, for instance, frequently arises as a critical emerging concern. It incorporates insight into how data is collected and analyzed, the contours of data sets employed to train algorithms to their tasks, and how the algorithm undertakes decisions and makes value judgements (the most technically difficult to achieve). And competing concerns present: as discussed above, the privacy and safety of dissidents must be weighed as a danger of transparency misuse. The degree and extent of both the current and requisite future private sector transparency is debatable, but there is wide agreement that more transparency is needed to inform public decisions on emerging technology and our future.

10. Accountability: The legal maxim that "rights warrant remedies" applies as a broad democratic principle. In a functioning democracy, bad actors and even well-meaning actors who cause bad consequences are held accountable for their impacts on society. As technology rapidly evolves, the law struggles to keep pace. That is no reason to sideline either the judicial concepts of fairness and equity, however, or the remedies that enforce those concepts. The question of who should enforce compliance, particularly in internet spaces untethered from the physical world, remains open.

B. Criteria to Evaluate Solutions

Recognizing the novelty of addressing national security concerns through broader societal contexts, the Task Force established the following criteria to guide strategic decision-making. The primary criterion of the Task Force to guide the national security focus of the solutions was to support the previously discussed Democratic Principles. The other criteria offer a broader emphasis on policy efficacy important to balance across a variety of stakeholder interests.

1. Supports Democratic Principles: The degree to which a solution bolsters the Democratic Principles outlined above. The Task Force considered this criteria of paramount importance.

2. Feasibility: The degree to which "we" can actually implement a solution. This may depend on:
    a. the number of actors and complexity of the processes involved in agreeing to and implementing the solution;
    b. whether those actors have conflicting interests or the processes require expending significant political capital;
    c. the cost of implementing the solution;
    d. ease of enforcement;
    e. how long it will take to actually implement the solution.

1. Opportunity for Impact: The degree to which a solution will meaningfully address the worst harms. "We can't nibble around the edges."

2. Likelihood of Adoption: The degree to which or likelihood end users will embrace the change (e.g., usability, user experience, convenience).

3. Quick Implementation: The degree to which a solution represents a quick win, including to help incentivize further governance actions. The general sentiment adopted was that we need to start somewhere: we can't dither while Rome burns.

4. Innovativeness: The degree to which a solution presents a new way of tackling a hard problem; on the other hand, innovative solutions are sometimes untested. We want to do more than reinvent the wheel.

5. Preserves benefits of social media: The degree to which a solution supports, fosters, or otherwise leaves intact social media's benefits. Don't throw the baby out with the bathwater.

6. Evergreen: The degree to which a solution is flexible and/or sustainable enough to be relevant in governing future technologies and attendant risks. Tech changes every 2-3 years. The most effective solutions will last beyond the latest innovation.

7. Efficacy: The degree to which a solution accomplishes its intended purpose.

8. Avoids collateral damage: The degree to which a solution minimizes unintended consequences.

9. Involves multiple stakeholders: The degree to which a solution is informed by and involves a range of stakeholders.

10. Abuse proof: The extent to which a solution cannot be manipulated or bypassed by bad actors.

11. Political will: The degree to which there is widespread consensus and agreement on the fundamental problem and the pressure on / willingness of key actors to address it. "We need enough, and the right, people at the table to make it happen."

12. Plays to strengths: The extent to which a solution plays to the strength of the Task Force.

13. Uses available resources: The extent to which a solution uses existing building blocks and resources (e.g., infrastructure, processes, frameworks, legal principles, and/or institutions).

II. Recommendations

Task Force Recommendations*

Effective Governance: Policymakers should identify and codify protections against harms that apply to the digital world, and provide funds for research and grant programs for investigating and responding to those challenges.

Responsible Platforms: Industry, civil society, academia, and the public should develop a users' digital bill of rights and a developers' code of conduct, and promote their adoption and adherence.

Empowered Public: All stakeholders, including government, platforms, community groups, academia, and civil society, have an obligation to educate and provide tools to online users so they are empowered to think critically, to advocate for their interests in the digital world, and to participate in democratic processes.

**As previously mentioned in this Report, all members of the Task Force have participated entirely in their personal capacities and not on behalf of any other organization or entity. The recommendations put forward are not attributable to any individual members. Not all members work directly on, or profess expertise in, all of the recommendations set forth below; nevertheless, this set of recommendations reflects the sense of the Task Force as a whole.*

Applying their own framework centered on Democratic Principles, the Task Force honed three primary recommendations. Each recommendation is keyed to one of the three pillars of the American digital ecosystem: (1) effective governance mechanisms, (2) responsible platforms, and (3) an empowered public.

Each pillar must possess three hallmark qualities to succeed in its digital world role, which are termed its goals. The Task Force-endorsed recommendations are intended to guide efforts within each pillar to meet its goals. The compendium of further concrete steps to reach those goals draws upon the Task Force's collective expertise. These recommendations cover myriad actors in society, reinforcing the need for creative collaborations, including many that exclude government altogether, in the name of national security and democracy.

A. Effective Governance

<table>
<tr><th colspan="2">Task Force Recommendation

Policymakers should identify and codify protections against harms that apply to the digital world, and provide funds for research and grant programs for investigating and responding to those challenges.</th></tr>
<tr><th>Goals</th><th>Possible Steps to Enactment</th></tr>
<tr><td><b>Research</b></td><td>Federal Research Consortium</td></tr>
<tr><td><b>Innovation</b></td><td>Early Stage Grants</td></tr>
<tr><td><b>Accountability</b></td><td>Disclosure and Reporting Requirements</td></tr>
</table>

The digital world fundamentally alters the relationship between government and society, rendering our current governance authorities inadequate. The internet creates new, loosely governed spaces and interactions that will create new security challenges that will only expand with the advent of 5G and 6G networks. Policymakers and regulators currently lack the technological sophistication, coordination, authorities, information, and resources to effectively safeguard a positive digital future. Digital Rights require protections our current structures do not fully offer.

Goals: We need governance that 1) fosters research, 2) encourages innovation, and 3) enhances accountability.

New technology exists at the cutting edge of knowledge. To properly understand how to best manage its consequences, government regulators need reliable research. Government must also adapt to this flexible environment, supporting private innovation and bringing some innovation in-house to bolster the public interest. Finally, good governance will bring accountability for the missteps and bad actors in the space, as good policy only succeeds where fairly and consistently enforced.

Recommendation: Policymakers should identify and codify protections against harms that apply to the digital world, and provide funds for research and grant programs for investigating and responding to those challenges.

As civil society works to prioritize and clarify the principles that are important to a healthy digital society, such as proposed in the Digital Bill of Rights, policymakers should watch and consider what policies, rules, and laws may be needed to realize the full public benefit of these principles. A bipartisan Congressional Commission, Federal Advisory Committee, or other executive branch body should be created and staffed by representatives from industry, government, academia, and civil society to conduct extensive research into emergent online harms and necessary regulatory protections. Its recommendations should cover issues including data transparency, information quality, security, data ownership, necessary platform disclosures, maintaining American competitiveness, accessibility and inclusion, and opportunities for public-private partnerships, with consideration of potential comprehensive legislation. In light of the rapid advancement of new and emerging virtual technologies blending the digital and physical worlds, the laws and policies proposed by the body must be technology neutral.

1) Establish Federal Research Colloquium

As part of their recommendations, the body should pay particular attention to identifying the most fruitful avenues for future research and grant programs, whether distributed through existing executive branch programs or through new organizations. A primary pathway to a coordinated funding program could be, via the NSF or NIH, a research consortium on digital harm. Congress should direct the FTC to clarify when and how platforms can share data with researchers from academia and civil service while protecting user privacy rights. Research should focus on establishing definitions and baselines for harms (including psychological harm); effective reporting mechanisms for user safety issues; developing an open research training data set for researchers at higher education institutions; convening discussions on standardizing data quality; understanding and publicly sharing the impact of automated decision-making; and experimenting with technology that supports democracy and transparency principles, rather than simply creating technology for government use; among other topics.

2) Distribute Early Stage Grants to Encourage Innovation

In addition, the body should specifically explore funding programs to incentivize companies or early stage investors to adopt practices that promote democratic and information governance principles (like those discussed in this report), Environmental, Social, and Governance principles (ESG) principles, and Digital Rights (*see* discussion, *infra*, for potential rights to be taken on board). In those areas of science and technology where innovations are still unproven or may not be immediately profitable, the government could fund companies or emerging venture capital funds that seek to promote the aforementioned principles. Through such research and grant programs, the government can explore the most promising avenues for new regulation.

3) Undertake Study and Establish Disclosure and Reporting Requirements for Enhanced Accountability

The first step towards accountability is to understand the current social media landscape, the baseline of regulatory authorities, and the regulatory needs of relevant State and Federal agencies, which the previous emphases on research and innovation aim to do. Building on that work, government agencies will need to define and promulgate rules for mandatory platform disclosures and reporting, including based upon standardized ESG principles (currently applicable in the investor realm) and information governance (as outlined earlier in this report) principles. To expedite implementation, relevant federal agencies like the SEC and FTC (in conjunction with private industry, research community, and national security leaders) could be asked to develop disclosure and reporting requirements under existing regulatory frameworks, public and nonpublic, for all digital media and platform companies. They should also explore whether new authorities requiring legislation are needed.

Topics that they should consider for disclosure and reporting requirements include transparency of algorithms and automated decision-making, data sharing with researchers, business practices monetizing user data, and evaluating current health and systemic risks, among other topics. For example, they should report on the incidence of “deepfaked” postings on their platforms, and if they cannot provide such information, they should explain in detail why not and the steps that would need to be taken to glean that. Such efforts must take into account that all government reporting required by the United States will pressure social media companies to share that same data with authoritarian adversaries. Given the demonstrated ease of de-anonymizing aggregated data sets that could be used to target dissidents, government agencies must take great care in what information they ultimately request through reporting. Although there is much flexibility in how, the government must take a more active role in ensuring a healthy digital world.

B. Responsible Platforms

<table>
<tr><th colspan="3">Task Force Recommendation
Industry, civil society, academia, and the public should develop a users’ Digital Bill of Rights and a developers’ Code of Conduct, and promote their adoption and adherence.</th></tr>
<tr><th>Goals</th><th colspan="2">Possible Steps for Enactment</th></tr>
<tr><td><b>Trustworthiness</b></td><td>Interagency Working Groups with Formal Channels for Public Input</td><td>Self-Regulatory Industry Collaborations around ESIG Standards</td></tr>
<tr><td><b>High Ethical Standards</b></td><td colspan="2">Support for Diverse Employee Perspectives</td></tr>
</table>

<table>
<tr><th colspan="3">Task Force Recommendation
Industry, civil society, academia, and the public should develop a users' Digital Bill of Rights and a developers' Code of Conduct, and promote their adoption and adherence.</th></tr>
<tr><th>Goals</th><th colspan="2">Possible Steps for Enactment</th></tr>
<tr><td><b>Respect for Digital Rights</b></td><td>Company Training</td><td>Formal Commitments to a Digital Bill of Rights and Developers' Code of Conduct</td></tr>
</table>

Emerging technology gives internet platforms an outsized role in societal interactions, for which they are not currently well-suited. The internet's role in our daily lives is growing faster than even some of its most sophisticated platforms can manage. The situation will become more pronounced as augmented and virtual reality come of age and platforms that are already well-established will have advantages that may help them achieve market dominance. Society expects more from companies with their increasing acquisition of power. Simultaneously, history has shown the importance of ensuring a greater role for consumers and consumer protections. As significant portions of the public square move into private hands, platforms must adapt to new public responsibilities.

Goal: We need platforms to exhibit 1) trustworthiness, 2) high ethical standards, and 3) respect for users' digital rights.

Platforms must gain the trust of consumers by judiciously and equitably enforcing their policies, aligning their incentives with user interests, and presenting trustworthy information. This does not mean that platforms must be the ultimate arbiters of truth; instead, transparency and empowering users to decipher credibility are two key factors in gaining the public's trust. Responsible platforms will also set high ethical standards internally, both for their employees and for their broader business decisions. As platforms receive tremendous societal power and space to innovate without onerous government regulation, they must also commit to ethical standards to curb abuse of that power and discretion. Finally, platforms must honor the digital rights of all their users—the foundation for their social contract with American society.

Recommendation: Industry, civil society, academia, and the public should develop a users' Digital Bill of Rights and a developers' Code of Conduct, and promote their adoption and adherence.

The users' Digital Bill of Rights and developers' Code of Conduct provide model norms and standards that could be adopted by the producers of digital society and used to inform their policies. They are intended to focus the attention of industry, civil society, academia, the public, and the government at every level in order to produce a better coordinated all-society strategy for a healthy digital universe. Critical topics to address include public reporting and data disclosures that could help facilitate transparency of algorithms and automated decision-making, algorithmic auditing, portability or interoperability standards, data sharing with researchers, ways to measure the credibility/accuracy in original content, accessibility and inclusion, ethical safeguards for business practices monetizing user data, and evaluations of current health and systemic risks.

1) All-Society Digital Strategy Should Be Adopted to Bolster Platform Trustworthiness

In order to be credible and, therefore, effective, such an all-society strategy for a healthy digital society will need to be drafted using a process that avoids giving any company an economic advantage. It must create space for new entrants and avoid a monopoly of power for any single large platform. This all-society digital strategy for a healthy digital society could grow from interagency working groups at the federal, state, and/or local levels, with representatives from the major online platforms, as well as representatives from smaller startup companies, civil society, and academia. Developing these norms and standards will require opening formal channels for public input, engaging community and advocacy groups in dialogue, and ensuring the perspectives of minorities and vulnerable communities are heard.

Additional efforts at norm or standard-setting could be based on successful efforts to develop the EU Code of Practice on Disinformation, which has brought together online platforms and the advertising industry to self-regulate. Industry undertook a similar effort in Australia, as this model seems to be gaining favor as an initial step. An all-society effort requires engagement in a multitude of ways at different levels of specificity to generate buy-in and operationalize the new norms and standards on the ground.

Building off current efforts in the finance industry, platforms could also commit to a model focusing on responsible Environmental, Social, and Governance principles (ESG), as well as the information governance principles discussed later in this report. Combined, stakeholders could develop a tech industry-focused set of ESIG standards. ESIG criteria could become part of defined guidance for companies, but also can help investors when evaluating companies for investment.[1] *The Task Force developed an illustrative set of Information Governance Principles for this purpose, included later in this report.*

[1] Sue Gordon, https://register.gotowebinar.com/recording/recordingView?webinarKey=1110374122902055179&registrantEmail=scif%40mattabrams.org

Much like the SEC enabled the private sector in the 1940s to develop its own practices that were eventually codified into law, the tech industry should be encouraged to set its own standards that promote the essence of ESIG criteria.[2] In fact, robust self-regulation is in the industry's best interest. As pointed out to us by the CEO of a major advertising enterprise, self-regulation can even persuade Congress against regulating an industry altogether, as in certain corners of direct selling and direct marketing.[3] ESIG would better equip investors to gauge risks posed by internet platforms and provide a common basis for the industry to codify appropriate behavior.

2) Adopt High Ethical Standards

Part of ethics is going beyond the ordinary call to ensure inclusivity, accessibility, diversity of thought, and fairness. Online platforms must adopt ethical postures, and a clear first step is through supporting employee resource groups and diverse recruitment initiatives at all levels of companies–from entry-level to senior management–with money and institutional clout. This is meant to provide alternative perspectives and to build out internal feedback channels for positive change. By bringing the views of diverse communities into the design process from day one, services and products will more effectively take into account differing perspectives on ways in which services and products are used, including input on how they may be used in harmful ways and how those harms may be mitigated.

3) Ensure Respect for Digital Rights

To integrate these concepts on the ground, some of these efforts must develop best practices and training for employees in online platforms that can serve as a guiding north star in ethical design, operation, administration, and governance. Such training should include a signed commitment from employees to follow a Developers' Code of Conduct and from companies to adhere to the Information Governance Principles, in addition to respecting users' Digital Rights. *For further detail on these concepts, please see the illustrative Digital Bill of Rights and Developers' Code of Conduct discussions later in this report.*

C. Empowered Public

[2] Sue Gordon, https://register.gotowebinar.com/recording/recordingView?webinarKey=1110374122902055179&registrantEmail=scif%40mattabrams.org

[3] https://www.ntia.doc.gov/legacy/ntiahome/privacy/mail/disk/DMA.htm

<table>
<tr><th colspan="3">Task Force Recommendation

All stakeholders, including government, platforms, community groups, academia, and civil society, have an obligation to educate and provide tools to online users so they are empowered to think critically, to advocate for their interests in the digital world, and to participate in democratic processes.</th></tr>
<tr><th>Goals</th><th colspan="2">Possible Steps for Enactment</th></tr>
<tr><td><b>Education</b></td><td>Locally-Focused Civic Education: K-12 and Adults</td><td>Support for Local Arts, Culture, and Journalism</td></tr>
<tr><td><b>Mobilization</b></td><td>Community Conversations to Set Online Norms</td><td>Engaging New Audiences with Workshops and Influencer Outreach</td></tr>
<tr><td><b>Agency</b></td><td>Interoperability</td><td>User Interface Adjustments</td></tr>
</table>

The blending of the physical and digital worlds shifts some corporate incentives around consumer interests in ways that could harm the public if left unchecked, and the public is not equipped to deal with the growing misalignment. A new digital age offers immense possibilities for individuals, but only if we are properly prepared to manage its complexity. Under the current internet model, individuals' attention is treated as a product to sell, with the consumer but a means to the ends. New technologies, like blockchain for example, could shift the onus of security squarely onto the shoulders of the individual, with consequences both good and bad. The public must prepare to grapple with the implications of new power dynamics in the digital landscape, so as not to be left vulnerable to future exploitation by narrow interests.

Goal: The public needs 1) education to understand new technologies and their consequences, 2) effective mobilization to advocate on issues of public importance, and 3) agency over personal data and choices.

Individuals must actively grapple with the complexity and implications of their online existence. This is only possible with some education about the technologies themselves, the way they influence individual choice, and the motives of platforms and content producers. To safeguard the rights and interests of consumers, individuals must mobilize, in grassroots campaigns and otherwise, and take active steps to advocate with both policymakers and corporate leaders. Finally, individuals need agency to fully participate in and shape their experiences online. Micro-targeting and other funneling techniques by online platforms absorb users' attention, benefitting advertisers at the expense of individuals' time and autonomy. While there are social goods that come out of these models, like free products, users need the freedom to choose these relationships, rather than being forced into acceptance by monopolistic realities.

Recommendation: All stakeholders, including government, platforms, community groups, academia, and civil society, have an obligation to educate and provide tools to online users so they are empowered to think critically, to advocate for their interests in the digital world, and to participate in democratic processes.

Democracy requires an engaged public to think critically about the rules and norms necessary for a healthy society, and this is true as much for digital society as it is for the real world. All stakeholders should have an opportunity to participate in specifying these rules and norms, including developing and using tools to realize them.

1) Develop Civic Tools of Education to Ensure Greater Societal Understanding of Risks and Opportunities Presented by Online Participation

In the first order, stakeholders should develop, promote, and distribute tools that support civic education that promotes a healthy digital society. Civic education for a healthy digital society is most effective where there is greatest trust, likely at the local level. Local level education efforts can be in dialogue with national, and sometimes international conversations, about how to establish rules and norms and what tools work best under what circumstances. To ensure quality and consistency however, local efforts should be ultimately guided by national-level frameworks. Moreover, civic education works best when it includes and empowers diverse communities, especially since marginalized communities are the largest targets for online abuse. Part of this diversity are advocacy groups and government agencies that protect consumers and support accessibility.

From a content perspective, civic education must offer politically neutral understanding and promote critical thought while furnishing a baseline-level understanding of technology. Educational efforts should cover, with ample input from all stakeholders including across the political spectrum: digital literacy, privacy and security consciousness, tech ethics, digital readiness, mitigating digital risks, and mental health management. Critical thought, in this context the constant questioning of the content, algorithms, and other systemic structures of digital society, is therefore an important element of public empowerment alongside general understanding of technology and its consequences. Such critical thinking regarding the digital world should be part of widespread civic education.

Educational content and materials should be provided at the state and local levels to students as young as kindergarten, and funding for these efforts may be supplemented through grants made available at the federal level. In addition, congressional funding and expanded mandates of government-funded entities with public education and information functions like PBS, universities, and state boards of education, can offer such education to the broader adult public. Other promising avenues for education include local government-sponsored seminars, wherein local leaders could discuss how to identify misinformation online, the harms caused by toxic and uncivil social media environments, and safety measures to protect children.

State and local governments, foundations, companies, and individuals must also invest in local arts and culture organizations to run exhibits that combat online misinformation through education, and other community dialogues to build consensus around social norms that apply on/offline. These groups must also come together to support local journalism to build a shared base of trusted information and local connection. The resulting community cohesion can foster understanding and spark collective advocacy efforts if community members decide they want particular changes.

2) Mobilize Democratic Tools to Protect Users and Society from Potential Harms

An empowered public requires a democratic voice in digital society, which can be accomplished through means remarkably similar to those of the physical world. Grassroots campaigns can leverage the convening power of such organizations to pressure both platforms and government on issues like increasing user ownership capabilities on platforms, adopting a Bill of Digital Rights, improving data transparency for academics, accessibility and inclusion, and other key issues. Building campaigns in any domain requires a high degree of social cohesion, which could be achieved in part through in-person conversations among diverse community contacts. At the conversations, local communities would establish a "social contract" around behavior on local social media message boards and sites. The use of local social norming can reinforce civil online interactions and create new outlets for online dispute resolution outside of the platforms themselves.

In addition to promoting local engagement, it will be critical to bring together parties who don't normally communicate, and to convince a wide swath of society to participate and be heard. For example, fiction workshops by nonprofits in communities could invite youth to write about the types of new technology they want to see, and connect them with startups looking to innovate in rural settings, as a way to co-create business models around new technology. As well, a national foundation or advocacy group could convene a diverse cohort of social media literate young people to educate and mobilize the public to demand new, responsible solutions to clean up the privacy, security, and information ecosystems. To be successful, mobilization must be both local and inclusive.

3) Give Users Agency Over their Online Experiences

The final component to user empowerment is that they must have the opportunity and the ability to participate in shaping the contours of their personal online experiences. Ideally, they need to control their own data, make choices unencumbered by platform manipulation or subterfuge, have the freedom to move seamlessly between platforms, and access tools that give them more control over different aspects of their online expressions. The ability to "vote with your feet" by leaving platforms that do not meet users' needs or expectations is a critical component of shaping platform practices and policies; however, portability of information that may have been amassed on one platform and not transferable to another prevents users from exercising this option.

An interesting first step in this direction would be an embrace of interoperability standards, wherein consumer rights advocates and platforms would engage in discussions about priorities for users when developing interoperability standards and best practices for offering users more control over their profiles, content creation, and personal data.

At a platform specific level, users would benefit from a wider variety of tools to give them insights "behind the curtain" of their own internet experiences, including a more over ability to shape the content they see and why that content is presented to them. Enhanced user tools might include a button or flag to easily communicate to their networks their personal levels of assuredness about the credibility of the content they are sharing. In addition, community groups could dialogue and generate new best practices ideas for user interface hurdles to accessing questionable content. For example, they could discuss forcing multiple click-throughs to access longform articles or removing automatic hyperlinks for sites deemed to provide large volumes of misinformation; not automatically including article and headline previews with user posts; giving users the ability to remove "like/dislike" or "retweet" buttons; or requiring a checkmark that "I read this" before posting. While actual solutions may vary by platform, creating mechanisms for user participation in shaping their own experiences in digital society is fundamental to the thriving of democratic principles online.

III. Other Steps Towards a Healthy Digital Ecosystem

In addition to the Task Force's three primary recommendations, members analyzed over 80 additional ideas developed over the course of Task Force deliberations and stakeholder interviews to promote a healthy digital ecosystem. Here is a sampling of some of the ideas that received some, but not universal, support. While some remained controversial among Task Force members, we include them in this report with the hope that they will increase the breadth of *future dialogue and, in the course of further debate, perhaps generate new approaches to the current risks posed by next generation social media.*

1. Standardized Consumer Credit Scoring System

Currently, consumers have no easy way to discern credible sources from those that consistently spread thirdhand, false information, including from authoritarian state-supported and amplified propaganda. One way to tackle the misinformation deluge might be to create an "originality index" that prioritizes accounts in search algorithms and labels them based on the volume of dis/misinformation they create, promote, and share. From a structural perspective, the index could initially draw on the voluntary industry model used by the Codes of Practice in the EU and Australia. The index would serve two purposes, giving users the ability to see the quality of information they consume from others and providing a "check" for users before they repost from sources known to spread propaganda or falsehoods.

As the metaverse evolves and avatars become increasingly personal, the index could extend the existing American Credit Score system to standardize digital responsibility and ethical standards across platforms. There may also be a need to compare virtual with physical behaviors, cross-referencing metaverse behavior with airline "No Fly" lists, for instance. Given the metaverse's likely role as a new virtual public square of sorts, such systems should offer public accountability and appeals processes, potentially with private citizens invited to serve on screening panels under a type of "citizen jury" system. With careful attention to bias factors, some form of standardized credibility score might offer unique benefits as virtual technologies evolve.

2. Amending Existing Laws to Keep Pace

A key theme throughout the Task Force's discussions, and particularly future threat simulations, was the inadequacy of current legal frameworks to address some of the novel concerns that arise as these new technologies become widespread. In the near future, a federal work group could recommend amendments to consumer protection laws, criminal laws, and regulations at both the federal and state level to keep pace.

As a starting point for such exploration, consumer law could expand to cover algorithmic bias under anti-discrimination laws and psychological harms under consumer product safety issues. New laws could create private rights of action or be enforceable by existing or new regulatory agencies. With novel fact patterns slowly emerging, tweaks to criminal law definitions of threat and specific harms might improve applications to the VR context. Federal rulemaking to define what constitutes data abuse and exploitation in decentralized or virtual environments and to stipulate compliance and enforcement might be another preliminary executive branch step in this space. Updates to existing regulatory definitions and rules for executive agencies to address emerging threats might include broadening CFIUS purview, increasing export controls industry coverage, clarifying SEC disclosure rules on DeFi, instructing the FTC to publish Codes of Conduct for tech companies, or augmenting IRS authorities. The diversity of these examples demonstrate the breadth of implications a virtual world might have for our laws, and the need for a comprehensive legal strategy to prepare for them.

3. Taxation and Government Enforcement

The fundamental misalignment of incentives for some companies vis a vis consumer interests may require more direct action. There are both voluntary and mandatory ways to address this problem, as with the internally enforced industry standards and through taxation or government enforcement mechanisms mentioned under the Responsible Platforms recommendation above.

To take such an ESIG effort further, the government, at a later stage, may consider offering incentives to encourage investors to promote ESIG principles. Incentives might include directly funding anti-surveillance or other democracy-technology business models, tax breaks for individuals who invest in ESIG solutions, funds for state and local government that invest in ESIG solutions, or creating a new category of "accredited investors and qualified purchasers" that must meet ESIG principles to be recognized as investors in funds.

Financial support for such programs could come from a levy of 1% or more on targeted advertising that tracks, combines demographic and psychographic data to generate user profiles.[4] Based on ideas raised from a number of sources, including media activist organization Free Press and noted economist Paul Romer, a tariff at 1% alone would bring in between $1-2 billion annually to support the incentives outlined.[5] Aside from funding incentives, this tariff could also support the study of the effects of social media on individuals and society at large.[6] This is an important proposition given the addictive effects of social media and their role in increased political polarization.[7]

Additional taxation would be necessary to support this proposal at scale, and could also act as a deterrent for irresponsible investors and force public disclosure of investors whose investments or practices do not further ESIG principles. For example, a new tax on funds and fund managers might penalize those that either invest or receive money from non-ESIG committed countries

[4] https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure
[5] https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure
[6] https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure
[7] https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure

and limited partners. Similarly, for investors whose limited partners come from adversarial or non-democratic regimes, there could be an interest tax. Finally, there should be a restriction and reduction of federal funding for states that invest in business models that do not promote ESIG principles.[8] Finally, we need new sales tax breaks and penalties, as well as other incentive programs, to reward consumers who purchase products by companies that support ESIG principles. Healthy digital media ecosystems can only thrive where incentives are properly aligned to foster that. A concentrated emphasis on promoting ESIG principles is a potentially useful first step.

IV. Tools for Implementation

The following three tools correspond to the Responsible Platforms recommendations, and are meant as a jumping off point for the realization of those ideas.

A. Sample Digital Bill of Rights

As virtual technologies evolve, users will need protections from harms that don't fall neatly into existing frameworks. Accordingly, the Task Force highlighted the necessity of establishing new foundational digital rights to guide a healthy digital world. Here is a preview of potential rights that could be included to ensure that individuals can engage in the digital world. They are drawn from the considerations raised through the Task Force's lengthy discussions. In many ways they are a companion document to the Democratic Principles highlighted earlier, a distillation of the individual protections that flow naturally from those principles that must be considered by all stakeholders addressing the internet's new challenges.

1) Individual Rights

Right to Identity: Individuals have the right not to have their identity assumed for the purpose of engaging in fraudulent behavior or material misrepresentation.

Right to Bodily Autonomy & Integrity: Individuals have the right to protect themselves or otherwise be protected against unconsented interference with their body through external manipulation, such as haptic gaming suits, VR headsets, or sensors. Individuals have the right to not experience harm or unwanted touching of their physical body or their avatar.

Right to Control Data: Individuals have the right to control the collection, sale, transfer, and deletion of personal data, including:

[8] https://www.oregonlive.com/politics/2021/12/oregon-might-dump-controversial-spyware-investment.html

Data Transparency: Individuals have the right to know the truth about how user data-driven companies, platforms, and other private entities are using user-generated data. This includes the ability to obtain information about how the platforms are feeding information to users and handling users' own data, as well as a right to have publicly available and documented APIs that facilitate auditing, research, interoperability, and standardized access to digital platforms.

Biometric Data: Individuals have the right to keep the measurements of their physiological characteristics private; public authorities may only retain biometric data under certain circumstances. The data of individuals that is gathered through brain or body scans cannot be used against them in legal or administrative proceedings. Biometric data includes, inter alia, pupil dilation, sweat responses, heart rates, and other indicators of brain and bodily activity.

Data Portability: Individuals have the right to collect and transfer their personal data from one platform to another.

Right to Express Consent: Individuals have the right to give initial, express, consent before being monitored, surveilled, or engaged in interactions by other users. Individuals have the right to prevent corporate entities and others from tracking or surveilling their movements online.

2) Rights within the Public Square

Right to Free Association: Individuals have the freedom to associate with others in the digital realm.

Right to Verification: Individuals have the right to know with whom they are interacting in the digital realm, whether their identity is masqued, and whether the entity with which they are interacting is a person or not.

Right to Block: Private individuals have the right to foreclose metaverse interactions with another user for any reason, at any time.

Right to Due Process/Right Against Erasure: Individuals have the right to notice, third party review, an opportunity to be heard, an appellate process, and a reasonable alternative means of communication prior to removal from a platform.

Right to the Physical World: Individuals have the right not to be forced into virtual reality. They have the right to live and obtain essential goods and services in the physical world.

3) Participatory Rights

Right to Inclusion in Decision Making: When governments or platforms make major decisions that affect the direction of the internet and virtual worlds, the public has a right to be consulted and included in that decision making.

Right Against Discrimination: Companies will not make decisions that will discriminate against individuals on the basis of race, colour, religion, sex, national origin, disability, age, sexuality, or any other protected status.

Right to Protection of Vulnerable People/Communities: Guardians have the right to implement measures they consider imperative for the protection of and best interest of vulnerable people under their care.

Right to Accessibility: All individuals shall have access to the technology and platforms needed to fully participate in virtual worlds whether it is for education, entertainment or other purposes, regardless of socioeconomic or other status. To enable persons with disabilities to independently access and participate in all aspects of digital life, appropriate measures will be taken to ensure equal access to the Internet, communications technologies and systems.

Right to Communal Safety: Communities have a right to establish and enforce norms for appropriate behavior within their online forums, similar to how restaurants may refuse service to those acting in ways outside the scope of their accepted standards for decorum.

4) Algorithmic Inclusion and Transparency

Right to Algorithmic Transparency: Individuals, researchers, and others have the right to explanations of any algorithms governing data collection and distribution and the right to study and make public their findings on the logic, significance, and impact of algorithms and automated decision-making.

Right to Financial & Business Model Transparency: Individuals and the public have a right to obtain information about company pricing, revenue, and profit generated on private data across the digital sphere. Individuals have a right to information about personal data supply chain instantionations.

Right to Representation: In mass data sets used to train algorithms, people of all races, religions, disability status, genders, and other protected classes have a right to be represented to try to minimize algorithmic bias.

5) Tools to Navigate the Public Square Safely

Right to Digital Public Education: Individuals have the right to public education that will equip them to navigate the digital realm in a safe and secure way, as well as evaluate the consequences of adopting new technology or sharing their personal data online.

Right to Disconnect: All hardware will be built to offer individuals an immediate disconnection from electronic devices and online platforms, at will, in order to remove themselves from a harmful situation.

Right to Notification: Individuals have the right to be alerted if threats or actions targeting their virtual or physical presence are made in a particular virtual space.

Right to be Free from Deceptive Commerce: Individuals have the right to be protected from unfair, deceptive and fraudulent products and services. Companies or platforms who sell user data or otherwise make profit from user data must treat users fairly and honestly, putting user interests first.

6) Right to Enforcement

Right to Enforcement Information: The public has the right to information about steps companies are taking to adopt, adhere to, and enforce these Digital Rights. Public and private grant makers and investors have a right to require the adoption of these Digital Rights or information about the steps companies are taking to adopt and enforce them as a condition of investment.

B. Draft Developers' Code of Conduct

The developers' code of conduct, modeled on the Hippocratic Oath undertaken by medical professionals, is meant for online platforms and emerging technology companies to consider their broader societal obligations. Building out industry certifications, like those adopted by medical personnel, mechanical engineers, lawyers, and other professions will require significant further discussion, but companies, educational programs, and other stakeholders can adopt ethics trainings and codes of conduct to orient developers now towards the greater societal good. Such a code could contain the following concepts:

Emerging technologies derive from a culmination of humanity's collective quest for knowledge. Those who utilize such knowledge must bear the weight of the gift. We ask them to undertake this social contract:

I pledge to design projects with public safety, human rights, democracy, and the good of society in mind.

I pledge to refrain from intentionally causing harm to my enterprise, society, or others, in service of my own personal gain.

I pledge to make the fruits of my efforts accessible to all people, regardless of race, sex, disability, or other status.

I pledge to architect my products to provide data and functionality through publicly available and documented APIs and service interface calls that will be externalizable.

I pledge to develop products who's user experience provides less friction and better ease of use in every iteration.

When confronting a problem with uncertain impacts, I will seek input from others, including those who might face disproportionate impact, and consider the risks before proceeding and throughout.

I will respect the privacy and sanctity of individuals, taking care to ensure the security of their personal data and only using it with their approval.

I will aim for transparency and share my knowledge as much as possible, to help advance scientific knowledge and to open that information to all who come after me.

Most importantly, I will remember that technological innovation is meant to serve the good of humanity, and I will strive to contribute to that progress.

C. Information Governance Principles

The following Information Governance Principles could form the basis for an addition to ESG principles in the investing world and standards for industry self-regulation. They are meant as a guide to support more concrete metrics for companies to implement, and, above all else, to support democracy.

Information Integrity: Provide reliable, contextualized content to users.

Democratic Norms: Create channels to include the public in decision making with regard to major decisions that affect the direction of the internet and virtual worlds.

Intellectual Diversity: Allow a wide variety of perspectives to flourish, limit the ideological funneling of content to users without their express consent.

Privacy: Protect user data, including access history and sensitive personal information, from cyberattacks and deanonymization.

Universal Accessibility: Support access to internet services via a variety of mechanisms to accommodate for socioeconomic, racial, religious, disability, and other statuses.

Special Protections: Ensure that users with heightened risk profiles are considered and protected, whether they are minors, the developmentally disabled, political dissidents, or otherwise.

Data Ownership: Help users control and understand their data, including data collection and usage practices, and create pathways for them to monetize their own data or carry it across different platforms.

Openness: Share anonymized data with and support researchers, and educate the public about commercial practices that implicate their privacy, access to information, and other aspects of their online lives to inform public discourse.

Fairness: Apply Terms of Service and platform standards evenly across all users, enforce penalties for harassment and threats, and, where possible, offer services for victims to be made whole.

Beyond the information governance principles outlined here, a number of other organizations like have suggested potential ESIG metrics that individual Task Force members sought to highlight. Examples from 7Pillars Insights, one such organization, of some potential metric areas oriented at company evaluations include:[9]

[9] https://www.7pillarsglobal-insights.com/_files/ugd/24200f_4d90e4734a2346cdbd5de97b72162411.pdf

- Does the company host or support programs aimed at enhancing the ability of employees, particularly in digital spaces, to recognize hate speech, targeted disinformation and conspiracy theories?

- Does the company encourage and support independent local journalism?

- Does the company fund external civic education programs aimed at combatting disinformation?

- Does the company donate to politicians that have a track record of spreading disinformation?

- Does the company have a policy on advertising on networks or programs that perpetuate disinformation about the 2020 election or upcoming elections, that otherwise promote baseless claims about election fraud, or that incite attacks targeting state election officials?

- Does the company publicly endorse government efforts to address disinformation?

- Does the company have a track record of spreading disinformation?

- Does the company have a policy on advertising on networks or programs that perpetuate disinformation about the 2020 election or upcoming elections, that otherwise promote baseless claims about election fraud, or that incite attacks targeting state election officials?

- Does the company publicly endorse government efforts to address disinformation?

Appendix B: Prototyping A Virtual National Security Simulation: MIT's High-Level Design Document

Social Media Wars (SMW)

This appendix was developed by the MIT Game Lab, with input from the Center, to describe the parameters for a large-scale multiplayer alternate reality game that was conceived of by the Center on National Security. The Center's goal was to develop a concept for a game that will help players understand how social media can be used to both provide opportunities for new market economies and community engagement, but also contribute to serious security threats that can undermine our nation's fundamental democratic institutions and processes and individuals' safety and well-being. The appendix provides the key requirements and affordances, the potential open source applications that could be used, and a budget and timeline to develop such a game.

Vision Statement:
Create a social media simulation / game that will allow the players to understand, evaluate, use, explore, analyze, predict and attempt to counter bad actor (rogue government, hacker organization, etc) strategies. Such strategies include but are not limited to disinformation campaigns, crypto currency manipulation, DDOS attacks, widespread phishing attacks, as well as coordinated attempts to damage or destroy highly networked infrastructure (energy grids, banking systems, Internet of Things (eg, Amazon Ring / other Internet capable poorly secured household items.)

Genre / Requirements / Key Players
SMW is a massively multiplayer alternate reality (MMAR) game, run by a control team that releases information as the game goes on, monitors and responds to players' actions, advancing the plot and ensuring the game is responsive to the players, giving players the ability to genuinely affect the overall results of the game. Potentially, the people interacting with SMW consist of the control team (the game masters), the active players, and the inactive players.

Since one of the goals of the game is to monitor not only players' ability to respond and analyze social media manipulation, but also to engage in manipulation of social media as well, having some number of influenceable actors - players - would be ideal. While these actors could be represented by bots, controlled by influence algorithms monitored by the control team, a more realistic experience - which would also yield better information about the relative success and failure of different influence approaches - would be to have a number of people with social media accounts who can react to the manipulations more active players are doing.

This would give three types of people involved in the game: firstly, the **Control Team**, who control the storyline, initial information releases, and keep the game moving. Secondly, there are ''active players' - or **manipulators**, who react to the information released by the control team, and then may add additional information or messaging to react to the actions the control Team is taking. Finally, there are the background, or **inactive players**, who provide the background of people to be manipulated. Inactive players, as envisioned, are assumed to represent the many people who absorb social media and are influenced by it, but don't seek to control or analyze it.

The game is played on a set of privately maintained and curated social media platforms and private email servers, mimicking popular sites such as Facebook, Twitter, Twitch, Youtube, etc., in the 2D world. For a forward looking game, setting it entirely in the Metaverse[10] would be better, but it is not clear how that would be kept private (particularly since the Metaverse is still in development, and is not yet a usable platform for massively multiplayer endeavors.)

Key Mechanics/Gameplay : Active Players (Manipulators)
Most of the mechanics for the game are similar to those existent in social media: making posts, videos, memes and so on; liking them, sharing them, following them and, most importantly gaining followers. The way people gain power using social media is by convincing others to believe them and sharing those beliefs. So, in many ways, success/goals is measured in how widely spread players' thoughts and beliefs are as time goes onward.

By including cybercurrencies and NFTs in the game, the players acquire an interest in the digital economy, and success can also be measured by wealth gained and lost. Competing cybercurrencies and driving excitement/interest in NFTs gives players concrete reason to compete with each other for followers, fame and interest. The Metaverse also allows for ownership of digital 'physical' spaces, creating interest in digital assets, buildings, and gathering areas for social events.

Similarly, creating, managing, and encouraging social media gatherings - be they groups or forums in 2D media, active chat groups in a Discord-similar setting - or spaces/meetups in 3D media - can create interest/excitement goals, draw in followers, and increase the ability of players to influence each other.

Metaverse Gameplay Example
Imagine a team of players arranging and selling 'tickets' to a concert or game competition, in 3D space, requiring usage of a particular cybercurrency to purchase entrance - but only to the first

[10] Metaverse: Term for the currently imagined 3D virtual reality environment envisioned as being a continuous digital experience. For some reference, see here.
(https://www.wired.com/story/what-is-the-metaverse/)

limited (N) who sign up and play. There are doorprizes for a particular type of NFT, only available to attendees. Hype and excitement convince various players (nonactive) to come for the entertainment, driving up the price of the cybercurrency used to pay for tickets, and enriching the entertainment coordinators and the owner of the 3D space, while creating 'in-crowd' bonds for the attendees.

Key Goals: Active Players (Manipulators)
Active players are divided into teams, each with their own goals within the social media space. Team sizes depend on the overall size of the active player / Control team contingent - in a small game, teams might be one or two players; in a game with a hundred, teams might be five or six. It depends on the size of the team, and the potential complexity of the game description. Player goals - or how the players define winning vs losing the game - depends on their team's goals.

Note that teams can be potentially co-operative or primarily competitive - it depends on the nature of the game the Control Team wants to run. Also, there can easily be multiple teams of each style, seeking alternate but similar goals.

- Team Style A: On behalf of a government, distribute & promote information that diminishes trust in another government, seeking to either weaken its external influence and power or destabilize it by undermining confidence in it.
    - Could easily be multiple teams, working with separate information pushes against various governments / associations: e.g., QAnon influence, vaccine deniers, questioning election validity, encouraging and arranging demonstrations
    - Doesn't need to be just against governmental institutions: targets could include minority groups - POC, religious affiliations. Could also destabilize & encourage distrust against key infrastructure / instructions: financial, justice departments, judicial rulings, relief organization (Red Cross/Red Crescent), law enforcement, social outreach organizations.

- Team Style B: On behalf of a government or other major institution (non-profits, targeted corporations, etc.,) attempt to debunk, defuse, or otherwise minimize the effects of disinformation campaigns.
    - Could also have the goals of tracking down /identifying the particular alliances of individual
    - Could also represent the owners of the Social Media sites themselves, trying to maximize profits while minimzing disruptions and complaints from users, governmental institutions, black hat hackers, and others.

- Team Style C: Private individuals/organizations seeking to maximize profit out of the digital economy.

- Could be creating/pushing their own unique cybercurrencies/NFTs
- Could be seeking to profit by maximizing trading / encouraging buying/selling spikes in other peoples' cybercurrencies, or by generating and selling digital assets.
- Or, by attempting to steal and profit off of other peoples' currencies /digital assets.

Necessary Technologies:
The Control Team needs complete control over the multiple social media space the players are inhabiting, as well (ideally) the ability to oversee private (email / chat) conversations between the players. This means setting up and configuring multiple social media sites, and enabling the appropriate permissions.

In order to keep the servers privately controlled, an open source app hosted on private servers would make the best basis for the potentially multiple social media sites to be supported. This will require developer support to install, configure, and develop apps and analytics for the Control Team and players.

Potential open source applications:

- HumHub (https://www.humhub.com/en)*
- WallStant (https://github.com/wallstant/wallstant)
- Mastodon (https://joinmastodon.org)*
- GNU Social (https://gnusocial.network)*
- Pleroma (https://pleroma.social)*

There are multiple other open source social media software; there are also purchasable solutions. Most purchasable hosted solutions do not cover hosting modified versions of the software, and many do not provide the source code itself.

*Note that most open source social media platforms are licensed using the GNU Affero Public License which requires any modifications to the source code to be made available to its users. If the resultant social media platform is limited to a private group of users (the players), then only this private group would require access to the source code per the license requirements (a legal review would be necessary to confirm this). As this platform is designed as a game, the rules of the game could specify that while source code access is available, it should only be accessed after the game is completed. In the case of Mastodon and GNU Social, if the network was connected to real-world networks through the federation model, then the source code would need to be made publicly available because the user base would expand beyond just the game players.

Features:

- *All* communications and exchanges should be viewable/reviewable by the Control Team.
    - Control Team should be able to enable/disable various features and access to features to individual players.
- Customizable social media, maintained on private servers, mimicking multiple sites: Facebook, Twitter, YouTube, Snapchat, Tik-Tok, etc. ***Note that not all features are available to all players/teams: ie, Active Players may have access to more detailed features than Inactive Players; some Active Player teams may have access to better features than other teams, depending on their goals and setups.***
    - Creating websites, blogs, forums, posts
    - Uploading videos, pictures, etc.
    - Likes/dislikes
    - Emotes (customizable)
    - Email (private/internal)
    - Chats - group, one on one
    - Polls
    - Cookies
    - Following
    - Automated Posting - definitely for the Control Team, possibly for some players
    - Bots, BotNets
        - Could be used to represent the inactive players (as automated NPCs)
        - Could also be used by manipulators as a means to gain influence
    - Blocking
    - Logging of all activity so that it can be reviewed by the Control Team
    - Control over player news feed view
        - Moderation features for users given moderator powers
        - Controllable algorithms used to show/hide certain posts by Control Team
- Social Media Analytics - two levels
    - Control Team: Full analytics for posts, followers, keyword searches, advertisements, etc for all users/platforms.
    - Players: Control team should be able to decide what level of analytics are available to players - ideally, different teams may have access to different levels of analytics, representing private vs corporate level analytics.
        - Usage of cryptocurrency to allow players to upgrade analytics / other features
- Cryptocurrency earning/spending:
    - Earning - large numbers of viewers/followers for videos, blogs etc running advertisements.
    - Spending - paying for advertisements, NFT purchases, etc.
    - Ability to launch a Cryptocurrency or NFT in game.

Budget:
Developing a social media platform is not a small task, but hosting an existing social media platform is relatively easy and inexpensive. It is recommended to use as much of an existing social media network as possible for early prototyping and design of the game, to better understand exactly which features are required for the required gameplay. In this case, budget would be required for a designer and costs for hosting the platform (hosting costs if using an external provider that manages the platform, a system administrator and basic costs if hosting on a controlled server environment such as Amazon Web Services).

For more advanced features, a minimum budget should be found for a small three-person, full-time development team to modify an existing open source software package. In particular, analytics and modifying user stream algorithms would be more advanced features that would require time and skill from the development team. Which platform the team uses would be both a factor of the teams' skills in the development environment (such as Ruby, PHP, React) as well as the features required for the platform.

Anticipated costs include:
Three-person development team ($25,000 per month)
Design team ($10,000 per month)
Hosting costs for the platform ($100-1000 per month)

At these rates, a 12-month design and development project could run $432,000.

Yearly, on-going maintenance would be expected to run at 5-10% of the development budget, depending on whether the environment needs further modification or development post-launch.

These costs increase based on the number of social media platforms required for the game. A version of the game could include a primary modified social media network, and additional secondary networks with smaller feature sets (less/no modifications, etc…) unlocked and made available after the game's official start.

Note that costs increase dramatically if security of the platform is required, especially if players are providing any personal or private information into the game. If the network environment the platform is hosted on is not properly secured, this data can be stolen/become public (as happened to Parler). If the platform itself is not properly developed, it can be hacked or be an un-optimal experience (as is currently the case with the Truth Social launch). As the game is designed as a private experience and will not be publicized widely, these factors may not be as important as if this were a publicly released platform.

Appendix C: Developing Alternative Futures Scenarios

Georgetown Law, under the leadership of Professor Laura Donohue, has been at the leading edge of experiential learning, pioneering a large-scale, multi-player simulation introducing students to various decision-making roles, requiring them to respond to a range of national security crises. It is the only national security simulation of its kind, growing to over 120 participants from law schools around the country and the globe. (For more information, see a preview of our largest scale simulation and this American Inno article.) To develop this simulation, a small team of futurists come together to develop plausible and even likely scenarios of future risks. The methodology they have developed over time includes (1) bringing together diverse expertise; (2) undertaking deep research into the organizations, players, industries, economies, and other attendant issues at play in potential real-world security crises; and (3) iterating on the alternative future crises to develop the scenarios that are either the most likely or are unlikely but would have the worst consequences.

As we hone our Center's methodology for developing simulation storylines and seek to apply it to new sectors and problems, we wanted to explore other methodologies for building out alternative futures. To that end, the Center's Executive Director partnered with Institute for the Future to learn their methodology and reflect on how it could be applied for national security policymakers.

The IFTF methodology focuses on building out scenarios along four archetypes found across societies and eras:

- **Growth** scenarios, which show an *acceleration* of the present, with more of the status quo. It's often a story about *winners*, and their accumulation of power and goods.
- **Constraint** scenarios, which show a rebalancing of the system through regulation, *limitations, or restriction*. It's often a story of overcoming a *common threat* by reordering the haves and have nots.
- **Collapse** scenarios, which show struggles with a broken system. It's often a story of tragic failure, *irreconcilable differences, failure of human systems*, and breakdown of trust.
- **Transformation** scenarios, which show a reimagining of a different world. It's often a story of world-changing or *groundbreaking insight*, a *mutation or seed of an idea* or vision that becomes mainstream.

These scenarios are developed by:

- Ensuring a diverse team of futurists
- Identifying the key actors and their motives, values, and core dilemmas

- Building a body of research, including identifying the broad range of stakeholders and important issues, the relevant timeframe (often 10-20 years), geographic scope, trends and evidence of past/present examples of relevant change
- Charting out various horizons of change or story boards, with plot points, events or catalysts for change, and the potential choices and actions of key actors.

I found that the most useful part of this methodology was to begin to track data and graphics that will help you understand drivers of change (broad sociological, technological, economic, environmental or political patterns) and signals of innovation and disruption (specific, concrete examples from the present that will disrupt the status quo in some way). Analyzing those drivers and signals of innovation or disruption - with pictures, graphs, maps, stories, statistics - into the four categories (growth, constraint, collapse, and transformation) then helps you to begin to map out the various horizons of change. By instilling this habit in your team's natural discourse and rhythm, you can begin to develop and train your team to "think forward and think differently."

Appendix D: Prototyping National Security Simulations for Blockchain

Importance of Understanding Blockchain

Recognizing the potential importance of blockchain and digital currencies in the future digital world, we undertook an in-depth study of the current state of technologies, fundamentals of cryptocurrency economics, and future risks associated with increasing blockchain adoption across various sectors. A fundamental component of this research was participating in a Wharton School of Business Executive Education program on the Economic Fundamentals of Blockchain and Digital Assets. Although developing our own in-game cryptocurrency proved cost prohibitive and highly volatile, we identified straightforward ways to capture the effects of independent digital assets and blockchain in our future national security simulations based on how we see the technology evolving.

Use Case Understanding: Current State of Tech

Our first key goal was to develop sufficient understanding of the technology to understand its current capabilities and use cases. The economics of cryptocurrencies center on their use-cases, whether as a substitute value storage device like Bitcoin (analogous to digital gold), a currency for a closed ecosystem like Helium (where payments allow users to "rent" internet access to power their IoT devices away from home), or a public ledger to facilitate transactions like Ethereum (where users pay fees for strangers to conduct their transactions in a trustless manner). We analyzed the pros and cons of cryptocurrency systems effectively controlled by "whales," wherein single users with outsized holdings could "pump and dump" or otherwise manipulate the pricing of different cryptocurrencies to their own advantage in ways reminiscent to the 1980s Wall Street penny stock frauds, Ponzi schemes, and other financial chicanery. We also looked at the value of "governance" tokens in these environments, which give users direct input to the "rules" of any given system, and how those tokens can be (and have been) manipulated to perpetrate hacks and fraud against individual consumers.

At the same time, we explored alternative cases where blockchain is useful, including Web3 internet applications, data storage and dissemination, supply chain monitoring, art and creative content industries, decentralized organizations and project investment vehicles, and decentralizing backend financial system transactions. From a national security perspective, we considered how blockchain systems shift the onus of security from a central party controlling the system of transactions to the individuals at the "nodes" of each transaction, putting consumers and other, often less sophisticated parties, entirely in charge of their own security in ways they are not under traditional centralized systems. For example, a user who forgets their password or "key" can be completely out of luck to re-access their accounts in pure decentralized systems in ways that a centralized system almost always offers a "forgot your password" function; and a user who is duped into sharing their key has no "flagging fraud" recourse as is currently offered by major credit card companies and other payment systems. Transactions are final on blockchain without any sort of appeals, for better and for worse, which presents legal challenges as well as practical challenges for anyone engaging with the technology.

Future Scenario Building: Evaluating Systemic Risk

The course also offered a three part framework to evaluate the systemic risk of crypto and blockchain technologies that we found useful when considering future risks it poses:

1. The risk to a given blockchain network posed by failure at the level of a **single node or application.**

- A major under-discussed problem where a blockchain network is overly reliant on a small number of nodes/actors, as many current blockchain systems actually are. For example, even bitcoin has high concentration of miners; if the 5 biggest miners colluded they could theoretically commit a onetime attack destroying the network's value.

2. The risk to the blockchain ecosystem at large posed by the **failure of a single network**, protocol, or service provider.

- For example, 60% of total bitcoin purchases since 2019 occurred via Tether (a US-dollar pegged stablecoin), and there are questions about Tether's actual US dollar reserves. This could could lead to a rush, which could cause shocks across multiple digital assets. In fact, Tether was recently trading at $0.60 to the USD because of its volatility.

3. The risk to broader economies and financial systems posed by the **failure of the blockchain ecosystem itself.**

- Market cap across all blockchain networks right now is $2.4T, about 10% of US GDP; as corporations make inroads they become more susceptible to its fluctuations themselves and so a huge depression could spillover into traditional economy. As crypto is used more for payments, central banks can't control currency reserves to blunt the impacts of economic swings as easily and stop them from preventing recessions. These are mostly theoretical concerns right now.

For our purposes, the framework provided a helpful guide for the types of crypto-based scenarios we would need to put in front of future national security simulation participants. It has been, and will continue to be, key in helping us develop new scenarios that deal with financial instability and internet security breaches by both state and non-state actors.

Simulation Integration: How to Include Blockchain and Crypto

Ultimately, we saw many potential ways to integrate blockchain and cryptocurrencies into the simulations that would meet our pedagogical and innovative thinking goals. Originally, we hoped to create a simulation-specific decentralized currency that could run independently throughout a simulation for players. As our study progressed, we also considered 1) using a centralized database and algorithmic model to allow for maximum participant model cryptocurrency spends; 2) offering "choose your own adventure" decision points to in-game participants based around their own theoretical crypto asset holdings; 3) control team "drops" of information about imaginary third party use/misuse of existing crypto assets; and 4) independent mini-scenarios around theoretical hacks of existing or imaginary crypto assets owned by participants.

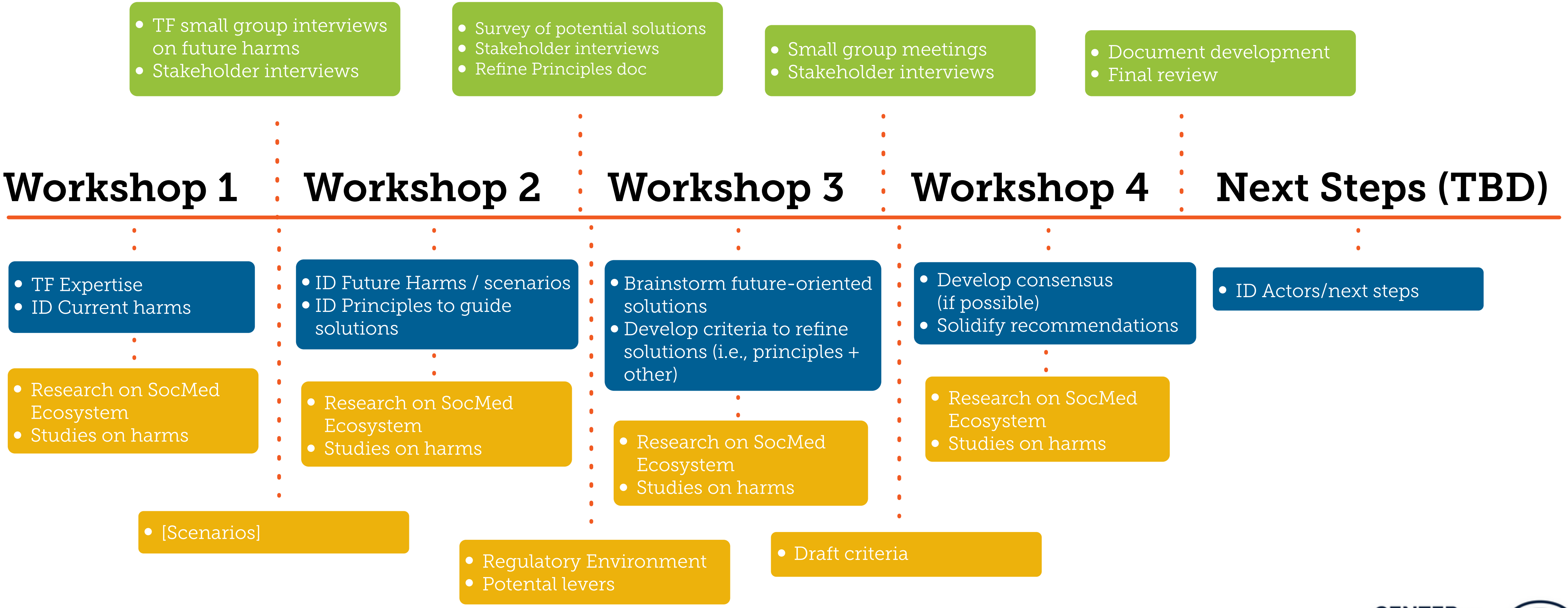
Based on our learnings, we see the best ways to incorporate blockchain technologies into our future simulations as through control team "drops" of third party information and independent mini-scenarios. Our game participants from the national security side are not the types of actors who will normally buy/spend cryptocurrencies—consumer fallout is highly complex and easier to make hypothetical via the control team. We see them needing to consider the second and third order consequences of such hacks, rather than the mechanics of the crypto and hacks themselves. Perhaps unfortunately, the wealth of past hacks and other system manipulations lays quite clear the consequences of such hacks, and the current thinking we lack is response and prevention. We believe we can use historical consumer and criminal behavior as simulation components we can pose to sim participants as they deal with the fallout of these technologies. Because so many of the concerns mimic the financial crisis tools of 1980s Wall Street, and the tech itself can be very difficult to understand beyond a broad theoretical basis (especially as it is so rapidly evolving and is therefore a moving target), we think that educating sim participants in financial system and other analogies will give them the background context necessary to play out decisions. We want blockchain and digital assets to enhance their overall sim understandings, especially how these elements collide with other new emerging technologies. We don't want them to be an overly complicated and technical distraction. Therefore, we are preparing to deploy this new knowledge via a more overarching framework with traditional sim injects.

Although we heavily studied it, we determined that our initial concept of an in-game digital currency was not feasible. We realized that gas fees are too expensive on the widely adopted Ethereum blockchain to make it useful for a simulation, and that they would make true gameplay cost prohibitive if our games happened to fall, for example, over the same weekend as a major NFT drop. Major NFT drops have been known to raise the fees of individual transactions into the hundreds of dollars, which is, of course, not reasonable for a simulation. We also considered commissioning a cryptocurrency build on Algorand, another network that does not have gas fees, but the need for players to have an advanced understanding of blockchain, plus their own digital wallets, meant that even building on Algorand might distract too much from other elements of a simulation.

Overall, we found the course very useful to 1) building our foundational understanding of blockchain, a highly complex and opaque new technology; 2) giving us new mental models for considering future risks of blockchain and digital assets; and 3) designing emerging technology simulations with the future of blockchain and digital assets as a key framing for participants to digest and analyze. We plan to incorporate the risks posed by blockchain into our future projects and simulations.

Appendix E: Design Methodology for Task Force: Fluid Hive Innovation Workshop Summaries and Analysis

How We Got Here



Workshop 1

- TF Expertise
- ID Current harms

- Research on SocMed Ecosystem
- Studies on harms

- TF small group interviews on future harms
- Stakeholder interviews

- [Scenarios]

Workshop 2

- ID Future Harms / scenarios
- ID Principles to guide solutions

- Research on SocMed Ecosystem
- Studies on harms

- Survey of potential solutions
- Stakeholder interviews
- Refine Principles doc

- Regulatory Environment
- Potental levers

Workshop 3

- Brainstorm future-oriented solutions
- Develop criteria to refine solutions (i.e., principles + other)

- Research on SocMed Ecosystem
- Studies on harms

- Small group meetings
- Stakeholder interviews

- Draft criteria

Workshop 4

- Develop consensus (if possible)
- Solidify recommendations

- Research on SocMed Ecosystem
- Studies on harms

- Document development
- Final review

Next Steps (TBD)

- ID Actors/next steps

🟢 Interim Actions 🔵 Group Action 🟡 Supporting Materials

CENTER on NATIONAL SECURITY
GEORGETOWN LAW

GEORGETOWN LAW

Social Media Task Force Meeting

Center on National Security and the Law

Workshop One

Dr. Laura Donohue started this inaugural task force workshop off by explaining why this group of people were chosen for this work, and how the Center on National Security at Georgetown Law (CNS) will be able to help the taskforce achieve the expected outcomes:

- creating the foundations for social media governance solutions
- changing pedagogy
- influencing people in DC
- mobilizing additional networks
- mobilizing the networks in the room

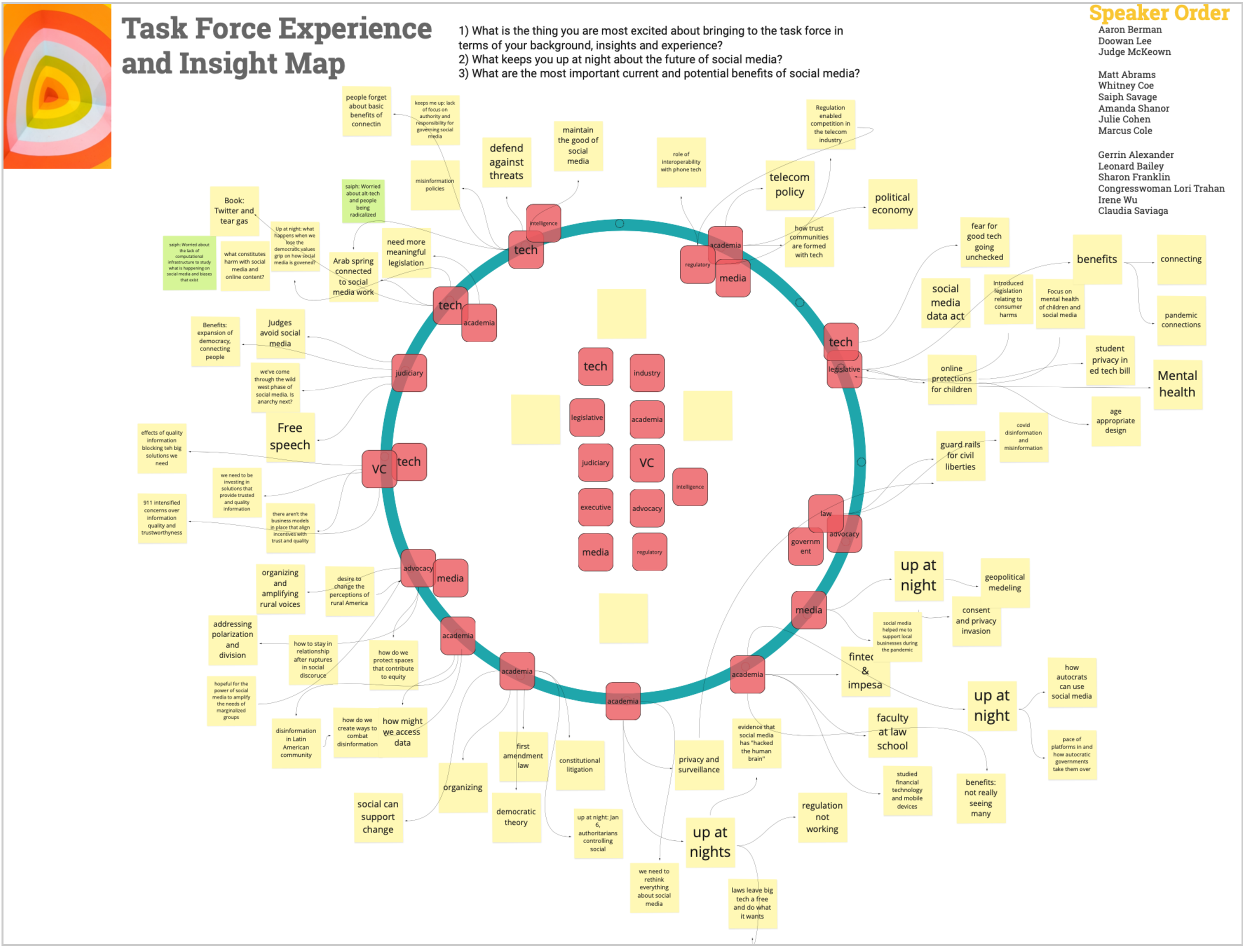
Dr. Dawan Stanford, the President of Fluid Hive, was introduced as the facilitator for the taskforce workshops and walked the task force through the ground rules for the workshop:

1. Operate with Chatham House rule
2. You may second something if you add/improve the contribution so that we are nudging people to dig into WHY rather than just +1 another idea
3. People can edit documents and/or Miro boards at any time to reflect changes in point of view.

Task Force Experience and Insight Map: The first exercise in this workshop was intended to quickly help the task force become a team by exposing everyone's expertise and letting everyone hear how other team members described not only what they did, but what they were thinking about in relation to social media and national security.

The task force attendees were then given three minutes each to introduce themselves, including the answers to the following questions:

1. What is the thing you are most excited about bringing to the task force in terms of your background, insights and experience?
2. What keeps you up at night about the future of social media?
3. What are the most important current and potential benefits of social media?


Task Force Experience and Insight Map

1) What is the thing you are most excited about bringing to the task force in terms of your background, insights and experience?
2) What keeps you up at night about the future of social media?
3) What are the most important current and potential benefits of social media?

Speaker Order

Aaron Berman
Doowan Lee
Judge McKeown

Matt Abrams
Whitney Coe
Saiph Savage
Amanda Shanor
Julie Cohen
Marcus Cole

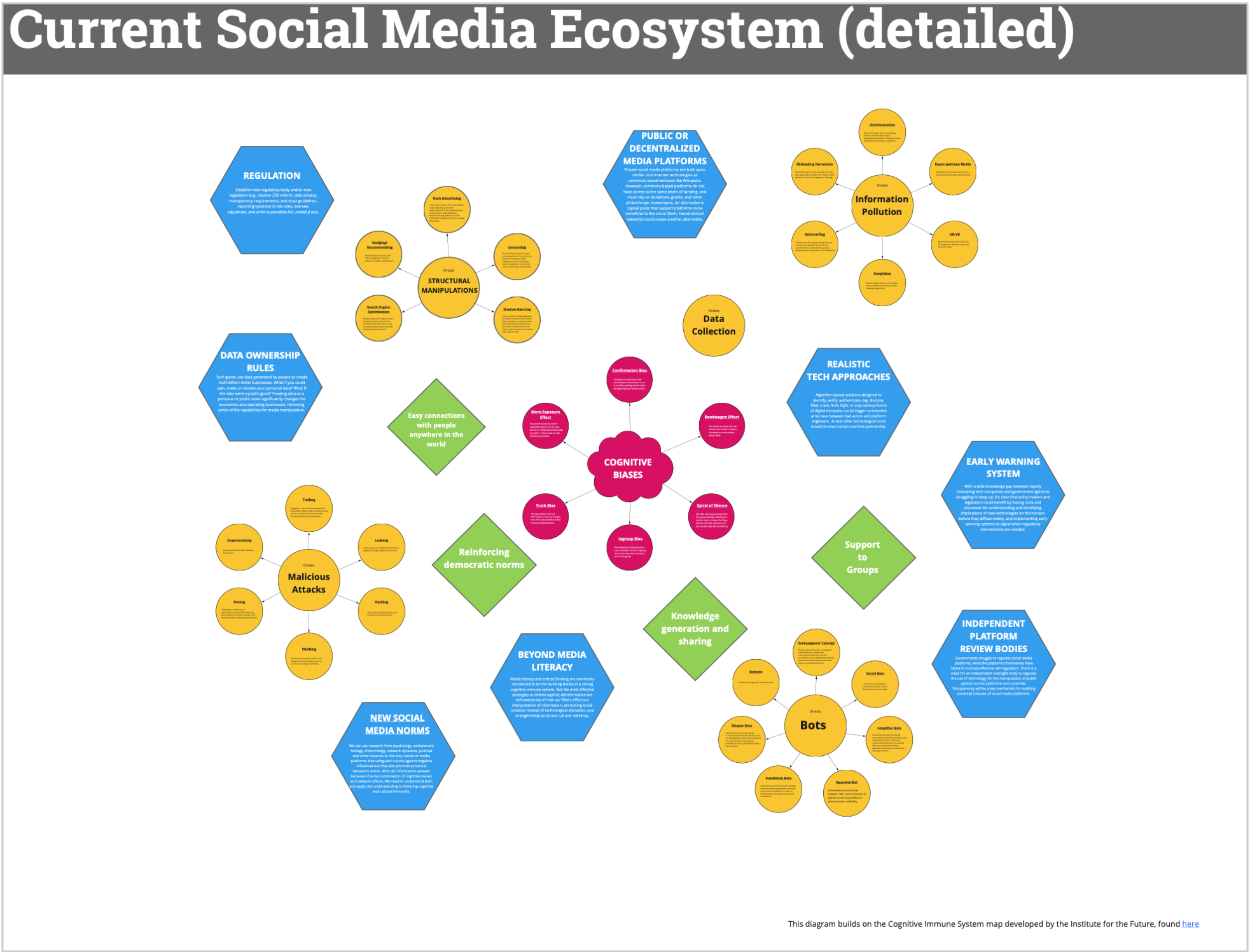
Gerrin Alexander
Leonard Bailey
Sharon Franklin
Congresswoman Lori Trahan
Irene Wu
Claudia Saviaga

people forget about basic benefits of connectin
keeps me up: lack of focus on authority and responsibility for governing social media
defend against threats
maintain the good of social media
misinformation policies
saiph: Worried about alt-tech and people being radicalized
Book: Twitter and tear gas
Up at night: what happens when we lose the democratic values grip on how social media is govened?
saiph: Worried about the lack of computational infrastructure to study what is happening on social media and biases that exist
what constitutes harm with social media and online content?
Arab spring connected to social media work
need more meaningful legislation
intelligence
tech
tech
academia
Benefits: expansion of democracy, connecting people
Judges avoid social media
judiciary
we've come through the wild west phase of social media. Is anarchy next?
Free speech
effects of quality information blocking teh big solutions we need
VC
tech
we need to be investing in solutions that provide trusted and quality information
911 intensified concerns over information quality and trustworthyness
there aren't the business models in place that align incentives with trust and quality
organizing and amplifying rural voices
desire to change the perceptions of rural America
advocacy
media
addressing polarization and division
how to stay in relationship after ruptures in social discoruce
how do we protect spaces that contribute to equity
academia
hopeful for the power of social media to amplify the needs of marginalized groups
disinformation in Latin American community
how do we create ways to combat disinformation
how might we access data
academia
organizing
social can support change
first amendment law
constitutional litigation
democratic theory
up at night: Jan 6, authoritarians controlling social
academia
we need to rethink everything about social media
privacy and surveillance
up at nights
laws leave big tech a free and do what it wants
regulation not working
evidence that social media has "hacked the human brain"
academia
fintec & impesa
faculty at law school
studied financial technology and mobile devices
benefits: not really seeing many
up at night
how autocrats can use social media
pace of platforms in and how autocratic governments take them over
media
up at night
geopolitical medeling
social media helped me to support local businesses during the pandemic
consent and privacy invasion
law
advocacy
government
guard rails for civil liberties
covid disinformation and misinformation
tech
legislative
online protections for children
student privacy in ed tech bill
Mental health
age appropriate design
social media data act
Introduced legislation relating to consumer harms
Focus on mental health of children and social media
fear for good tech going unchecked
benefits
connecting
pandemic connections
academia
regulatory
media
how trust communities are formed with tech
political economy
telecom policy
Regulation enabled competition in the telecom industry
role of interoperability with phone tech
tech
industry
legislative
academia
judiciary
VC
intelligence
executive
advocacy
media
regulatory

As each member of the task force who was present talked about those three questions and themselves, we created this visual map that everyone could see in real time. The map showed connections between their ideas and expertise, so that we could really build up a sense of who we are and what we can do. The idea here was that once people started to see this map, they would know that, "Oh, you know, I'm not an expert in free speech, but I know someone in the room is, so I can try to connect an idea that I have to something that might be related to free speech and that person can pick it up from there."

After the introductions were complete, the taskforce took a short break, and were asked to consider the following question: "It's 5 years from today. What would make you proud that we accomplished together as a Task Force?"

Current Social Media Ecosystem: We created a map of the current social media ecosystem to help the taskforce members start to gain a sense of the landscape. Everyone was coming in from different areas with different backgrounds, and we wanted to help everyone see the good things (the green), the types of threats that are out there that we know about in yellow, and the red is about the way human brains function and the cognitive biases therein. The blue hexagons are some of the levers that we can pull in terms of things that can help us get more of the benefits and fewer of the harms from social media when it comes to national security.



Current Social Media Ecosystem (detailed)

REGULATION
Establish new regulatory body and/or new legislation (e.g., Section 230 reform, data privacy, transparency requirements, anti-trust guidelines, reporting systems) to set rules, oversee, adjudicate, and enforce penalties for unlawful acts.
PUBLIC OR DECENTRALIZED MEDIA PLATFORMS
Private social media platforms are built upon similar core Internet technologies as commons-based ventures like Wikipedia. However, commons-based platforms do not have access to the same levels of funding, and must rely on donations, grants, and other philanthropic investments. An alternative is capital pools that support platforms more beneficial to the social fabric. Decentralized networks could create another alternative.
threats
STRUCTURAL MANIPULATIONS
Dark Advertising
Nudging/ Recommending
Censorship
Search Engine Optimization
Shadow Banning
threats
Information Pollution
Disinformation
Misleading Narratives
Hyper-partisan Media
Astroturfing
AR/VR
Deepfakes
threats
Data Collection
DATA OWNERSHIP RULES
Tech giants use data generated by people to create multi-billion dollar businesses. What if you could own, trade, or donate your personal data? What if the data were a public good? Treating data as a personal or public asset significantly changes the economics and operating businesses, removing some of the capabilities for media manipulation.
REALISTIC TECH APPROACHES
Algorithm-based solutions designed to identify, verify, authenticate, tag, disclose, filter, track, limit, fight, or stop various forms of digital deception could trigger unintended arms race between bad actors and platform engineers. AI and other technological tools should involve human-machine partnership.
Easy connections with people anywhere in the world
COGNITIVE BIASES
Confirmation Bias
Mere-Exposure Effect
Bandwagon Effect
Truth Bias
Spiral of Silence
Ingroup Bias
EARLY WARNING SYSTEM
With a wide knowledge gap between rapidly innovating tech companies and government agencies struggling to keep up, it's clear that policy makers and legislators could benefit by having tools and processes for understanding and identifying implications of new technologies on the horizon before they diffuse widely, and implementing early warning systems to signal when regulatory interventions are needed.
threats
Malicious Attacks
Trolling
Impersonating
Leaking
Doxing
Hacking
Phishing
Reinforcing democratic norms
Support to Groups
Knowledge generation and sharing
BEYOND MEDIA LITERACY
Media literacy and critical thinking are commonly considered to be the building blocks of a strong cognitive immune system. But the most effective strategies to defend against disinformation are self-awareness of how our filters affect our interpretation of information, promoting social cohesion instead of technological alienation, and strengthening social and cultural resiliency.
INDEPENDENT PLATFORM REVIEW BODIES
Governments struggle to regulate social media platforms, while the platforms themselves have failed to institute effective self-regulation. There is a need for an independent oversight body to regulate the use of technology for the manipulation of public opinion across platforms and countries. Transparency will be a key mechanism for auditing potential misuses of social media platforms.
NEW SOCIAL MEDIA NORMS
We can use research from psychology, evolutionary biology, immunology, network dynamics, political and other sciences to not only construct media platforms that safeguard culture against negative influences but that also promote prosocial behaviors online. After all, information spreads because of some combination of cognitive biases and network effects. We need to understand both and apply this understanding to fostering cognitive and cultural immunity.
threats
Bots
Sockpuppets/ Cyborgs
Botnets
Social Bots
Sleeper Bots
Amplifier Bots
Roadblock Bots
Approval Bot

This diagram builds on the Cognitive Immune System map developed by the Institute for the Future, found here

We set this up so that the taskforce could play with it, and then we broke ourselves into three smaller groups so that people could have deeper conversations about what's on the map, as well as what's *not* on the map, from their perspectives. Each group had a facilitator helping them think about the map, what's missing on the map, additional ideas, new connections between what's on the map, and more. Facilitators also pushed for additional harms we should seek to avoid, opportunities to capture, and benefits to protect.

Specifically, the breakout groups were supposed to address three questions:

1. What are the blind spots that no one is addressing?
2. What are some of the emerging challenges people are not thinking about?
3. What are emerging benefits and governance opportunities that no one is seeing?

At the end of the breakout sessions, the facilitators each had three minutes to share findings from their small group.

Dr. Stanford then posed the following questions to the taskforce:

1. Are there any areas where the task force needs more information or we could do more research?
2. CNS is about to do some stakeholder interviews. We did not include all of the actors in our task force, so who are the people we should talk to and what should we ask them?
3. If there are any thoughts that you did not share in your small group or did not get on the board, please send them to us.

Dr. Donohue then closed out the workshop by talking about the next workshop and next steps for the taskforce and the Center.

Social Media Task Force Meeting

Center on National Security and the Law

Workshop Two

Activity 1: Harms

Participants analyzed the harms from the Insights received prior to the workshop and identified those that were high priority and those that were missing. The harms were aligned with principles also received prior to the meeting.

Prioritized Harms (highest # of votes to lowest # of votes)

7 votes: **Information chaos.** Everyone is or can be a content creator and publisher, with minimal barriers to entry and no verification filters, leading to information pollution.

5 votes: **Psychosocial harm.** Studies reflect that the use of social media is associated with an increase in addiction, depression, stress, anxiety, isolation, including as a result of harassment, stalking, and bullying.

5 votes: **Ungoverned gray areas between protected and unprotected speech** which results in further corruption of truth, with platforms making profound decisions with national security and public interest implications, including decisions related to content causing acute harm, particularly against vulnerable communities.

5 votes: **Users have limited control over their data.** As a result, consumers/users (1) are susceptible to micro-targeting, exploitation, and social/political manipulation; (2) have limited control over their privacy and data security (and not all tech platforms are incentivized to provide security); and (3) cannot transfer their data between and to other platforms.

4 votes: **Potential loss of innovation.** Inadequate incentive structures and immigration processes limit the ability to attract and retain a diverse workforce, while overregulation will disincentivize the talent pool and investors.

2 votes: **Negative Externalities**, from the extra energy necessary to sustain these systems to the loss of entire professions fundamental to the exercise of our democratic ideals, like journalism.

1 vote: **Decentralization/democratization of the ability to conduct widespread attacks.** The ability to cause significant physical and monetary harm is now available to non-state actors.

Harms

Psychosocial harm: Studies reflect that the use of social media is associated with an increase in addiction, depression, stress, anxiety, isolation, including as a result of harassment, stalking, and bullying.

Information chaos. Everyone is or can be a content creator and publisher, with minimal barriers to entry and no verification filters, leading to information pollution.

Ungoverned gray areas between protected and unprotected speech which results in further corruption of truth, with platforms making profound decisions with national security and public interest implications, including decisions related to content causing acute harm, particularly against vulnerable communities.

Users have limited control over their data. As a result, consumers/users (1) are susceptible to micro-targeting, exploitation, and social/political manipulation; (2) have limited control over their privacy and data security (and not all tech platforms are incentivized to provide security); and (3) cannot transfer their data between and to other platforms.

Potential loss of innovation. Inadequate incentive structures and immigration processes limit the ability to attract and retain a diverse workforce, while overregulation, will disincentivize the talent pool and investors.

Decentralization/democratization of the ability to conduct widespread attacks. The ability to cause significant physical and monetary harm is now available to non-state actors.

Negative Externalities, from the extra energy necessary to sustain these systems to the loss of entire professions fundamental to the exercise of our democratic ideals, like journalism.

Missing Harms

- Technoauthoritarianism
- Erosion of societal norms, particularly related to civility; societal norms from physical world have not translated to digital world
- Monopoly power. Companies becoming more powerful than governments, monopoly power captured- just the idea that we have companies bigger/more powerful than governments (democracy does not survive with monopolies)
- Treating the digital world as “different” from the physical world
- Risk to veracity in the digital world
- Threats to democracy, ability of government to address pandemic/public health, from spread of misinformationP
- Political and social instability
- Pollution/energy usage
- Broad cleavage of society
- 1st Amendment rights
- Reinforcing power structures
- Everyone thinks of themselves as an expert, ignoring actual experts
- Lack of social norms in the digital world destabilizes real-world society
- Digital identity (protecting anonymity but also needing to verify ownership, citizenship, etc)- also civility, people tend to be bigger jerks when anonymous
- Threats to democracy, ability of government to address pandemic/public health, from spread of misinformation
- The detrimental environmental impact of all this technology, and about how easy it is for folks who don’t understand the tech to just not understand what they’re getting into and the potential risks of situations and thus not being able to make informed decisions.
- Emergence of Alt tech and more polarization and radicalization of people
- Educational content on disinformation
- Digital identity verification to verify ownership, citizenship
- Fewer close relationships and social and community ties
- People having fewer children

Idea: We might use the harms as a scoring rubric for companies or services.

Activity 2: Principles

Principles were aligned with harms, including those generated during this meeting and then discussed and prioritized by the group. The scenario was introduced in phases to facilitate even more discussion about harms and principles that appear to be common among society's most crucial issues regarding social media.

Prioritized Principles

11 votes Protect democratic norms. We value the preservation of democratic norms, institutions, and processes.

7 votes Protect digital users from virtual and physical harm.

6 votes Hold space for airing of grievances and expression of views.

6 votes Offer consumers data control and protection.

5 votes Create an ability to verify and weigh information you receive or share.

5 votes Promote transparency and access to data.

5 votes Lead domestic and global social media policymaking as a matter of US national security.

4 votes Realign current incentive structures.

1 vote Ensure positive community formation.

1 vote Provide universal access to digital services.

1 vote Maintain a right to experience as a First Amendment principle.

1 vote Safeguard privacy.

1 vote Explore decentralized networks and competition.

0 votes Carve out special protections for tech innovation.

Principles

Ensure positive community formation.

Provide universal access to digital services.

Create an ability to verify and weigh information you receive or share.

Maintain a right to experience as a First Amendment principle.

Hold space for airing of grievances and expression of views.

Protect democratic norms. We value the preservation of democratic norms, institutions, and processes.

Promote transparency and access to data.

Safeguard privacy.

Offer consumers data control and protection.

Carve out special protections for tech innovation.

Explore decentralized networks and competition.

Realign current incentive structures.

Lead domestic and global social media policymaking as a matter of US national security.

Protect digital users from virtual and physical harm.

Missing Principles

- Sustainability and access to information, capital, etc.
- Protection for workers (content moderators as well)
- Egalitarian principles: analyze who are the elites (people in power) that benefit from the Silicon Valley model and who is in less power? I am thinking about the rural communities that exist and have been left out of the innovation process.
- Provide universal access to digital skills
- Standardized reporting and information / data portability
- Right to fix and propose new directions to technology
- Protection to digital workers
- Effective sanctions
- No right without a remedy
- Keeping standards democratic (not overrun by monopolies)
- Content creator participation in data rights design
- Monetization of user created wealth/(Re)distribution and allocation of profits
- (Re)distribution and allocation of profits/monetization of user-created wealth
- Provide workers with more transparency about what happens in their digital workplaces
- Right to own workplace data
- A principle on limiting principles
- Tying regulation to potential harm
- Channel innovation to positive outcomes
- Accountability - we don't have that today in the digital world

Idea: how do we create a process for keeping the governance structure nimble. FDA uses "incentivized guidance" - not perfect at all but is even more flexible

What about an emergent exception process?

We need to rethink the way we approach things like community guidelines and user education. I'm thinking specifically in VR spaces that have been built on small communities of enthusiasts that established their own norms that make sense for their small group. The problem is that's not going to scale. So how do we build in-community guidelines to make sure people understand they work from a person-to-person interaction?

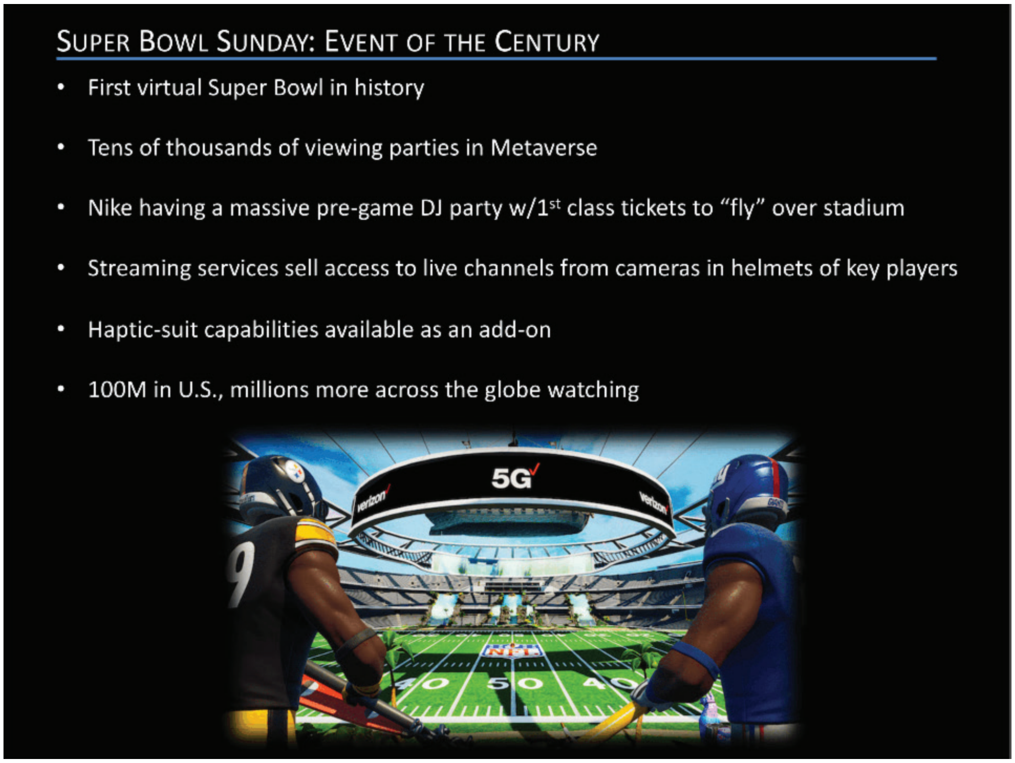
To me almost all of these principles speak to democratic values

I feel the Silicon Valley model of "moving fast and breaking things" can also hurt certain populations. It is important to be aware of also how it might not be the best model for everyone.

Scenario Introduction

The year is 2028, and we are approaching the Super Bowl. In 2028, there are quite a few things different from our current world. Everyone of all ages has VR headsets. A new internet architecture based on blockchain, called Web 3.0, has taken off in some corners and competes with traditional internet players like Amazon Web Services and other hosting services. Private crypto currencies periodically take off, along with NFT auctions associated with various real-life events or brands. Some things are the same: climate change has only increased erratic and extreme weather patterns, and most entities still lack robust cybersecurity.

For the first time in history, the Super Bowl is virtual. Attendance is expected to include 100,000 attending live, with millions more on television and connected in virtual reality. The virtual components of the Super Bowl are held on Web 3.0 VR platforms, and various companies offer adjacent festival booths or Super Bowl services, like AI-based betting competitions, the ability to "fly" over the stadium, and live streams from a player's helmet. This one-time, nationwide event takes a LOT of electrical power. Our generally neglected, aging power grids have been reinforced, but there isn't much room for error. Finally, February forecasts multiple snow storms, and is the coldest month on record.



SUPER BOWL SUNDAY: EVENT OF THE CENTURY

- First virtual Super Bowl in history
- Tens of thousands of viewing parties in Metaverse
- Nike having a massive pre-game DJ party w/1st class tickets to "fly" over stadium
- Streaming services sell access to live channels from cameras in helmets of key players
- Haptic-suit capabilities available as an add-on
- 100M in U.S., millions more across the globe watching

Scenario Introduction Concerns

- Inequities — things cost extra money
- Risk of significant economic damages to companies/economies from fraud or disruption
- Distraction from events happening in reality
- Citizens could forget their "citizenship" and the rules/laws that govern their acts.
- There's lots of other things you could be distracted from, too. A huge storm. An attack.
- Risk of targeted attacks by foreign actors who use large global events to distract. Hard to identify who is behind the distraction.
- Exacerbating how online harms cause real world harms — if people can "feel" injuries as players feel them.
- Loss of privacy (thinking of fly-over events) that compromises city/state security because bad actors can do surveillance of spaces/places previously off limits.
- One can use behavioral and psychological profiles to drive a specific profile of users to a private room or VIP room and introduce trigger stimulation to induce an online or physical attack.

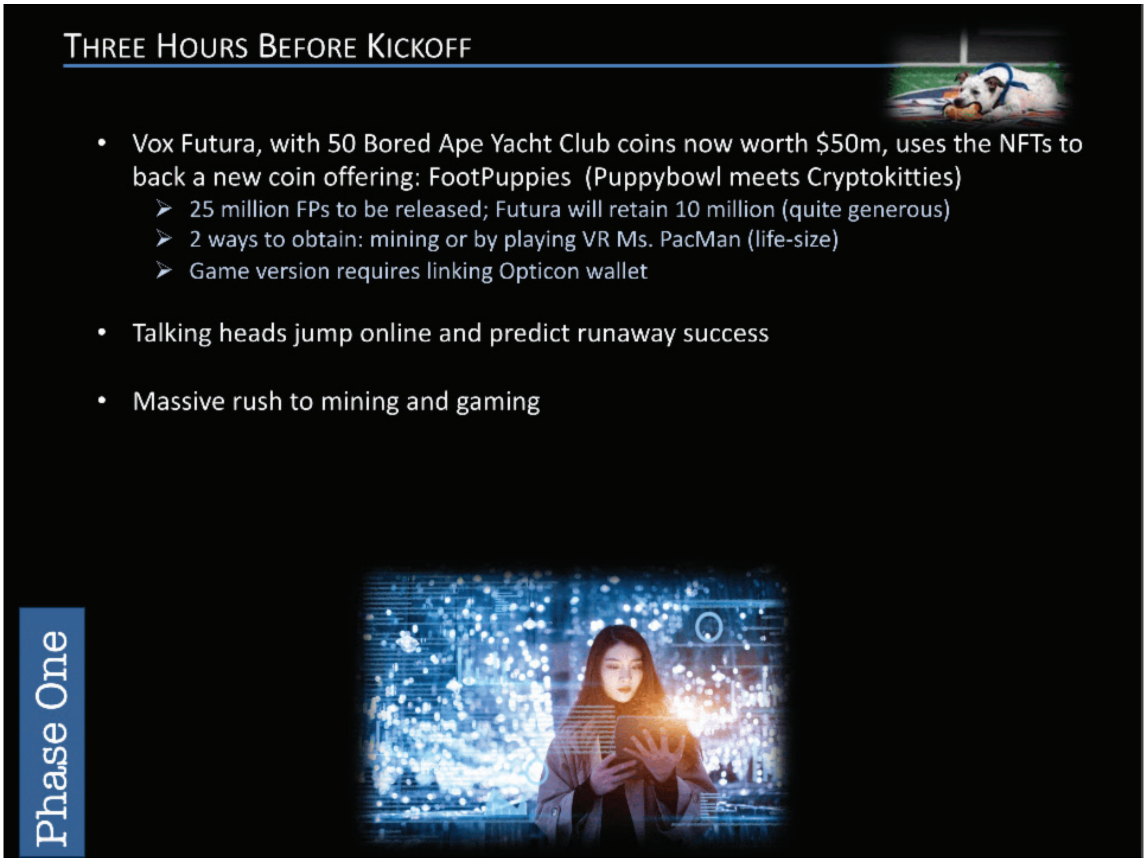
Scenario risks/ principles/ ideas

3 hours before Kickoff

Scenario Stage 1 Summary: Crypto Coin Drop

Three hours before the Super Bowl, Vox Futura announces a new commemorative coin offering. Backed by her personal NFT collection worth $50M of Bored Ape Yacht Club NFTs, one of the original NFT collections, she releases 25 million new coins called FOOTpuppies. Vox Futura keeps 10M coins, leaving 15M for the general public to harvest. FOOTpuppies are unique images of puppies in football gear, and are successively generated by mining or gameplay. The game is a takeoff on Ms. Pac Man, using a FOOTpuppy in a blue jersey eating footballs and running away from FOOTpuppies in red jerseys. The announcement prompts a massive public rush.

Unfortunately, the game requires linking to your Opticon wallet and a criminal enterprise exploits a software vulnerability in Opticon. Using AI algorithms to match wallets with publicly available personal information, the enterprise penetrates individuals' wallets and cleans out all Opticon users. As such, those who mine the coin get rich, but those who played the game lose everything. It is nearly impossible to track all the losses, which are distributed across thousands of burner wallets.



THREE HOURS BEFORE KICKOFF

- Vox Futura, with 50 Bored Ape Yacht Club coins now worth $50m, uses the NFTs to back a new coin offering: FootPuppies (Puppybowl meets Cryptokitties)
  - ➢ 25 million FPs to be released; Futura will retain 10 million (quite generous)
  - ➢ 2 ways to obtain: mining or by playing VR Ms. PacMan (life-size)
  - ➢ Game version requires linking Opticon wallet
- Talking heads jump online and predict runaway success
- Massive rush to mining and gaming

Phase One

Risks

- Crime at scale. Ransomware at scale. Malware at scale.
- Infrastructure fail
- Malware takes over your suit and hurts you or freezes you so you can't get out
- Hacking wallets and stealing both $$ and data
- Risks to power grid which exacerbates climate change/energy use
- scammers using this situation for money laundering
- Big quantum computing break though decrypts the entire network
- Personal "bankruptcy"/catastrophic financial loss - loss of entire digital wallet
- hard to build trust when you can't identify
- Emergence of counterfeits
- Identify spoofing
- can it also facilitate losing track of what really matters?
- The 1A for code as it pertains to blockchain (smart contract protocols) is really scary because smart contracts are governance - that would be a really interesting court case
- Web 3 can also have oligarchic exchanges much like Web2
- Computer code as "speech" is a nightmare with regards to 1st Amendment
- Increased psychological harms from visceral experience

Principles

- systems should allow interaction testing at scale
- Creating identity management
- "There is no way mere coders or mere website operators can register with FINRA, track the identities and trades of AMM systems occurring on decentralized autonomous blockchain systems or otherwise comply with the ATS / exchange reporting and registration regime, and therefore, if applied to such persons, this new rule would be banning a vast swathe of technologies and free speech regarding those technologies, which is beyond the SEC's authority and would constitute an unconstitutional violation of our civil and human rights."
- https://www.evernym.com/blog/w3c-vision-of-decentralization/
- Like with like regulations in the metaverse. For instance contracts are not "speech" and protected by 1A
- Leveling the playing field via regulation

Ideas/Questions

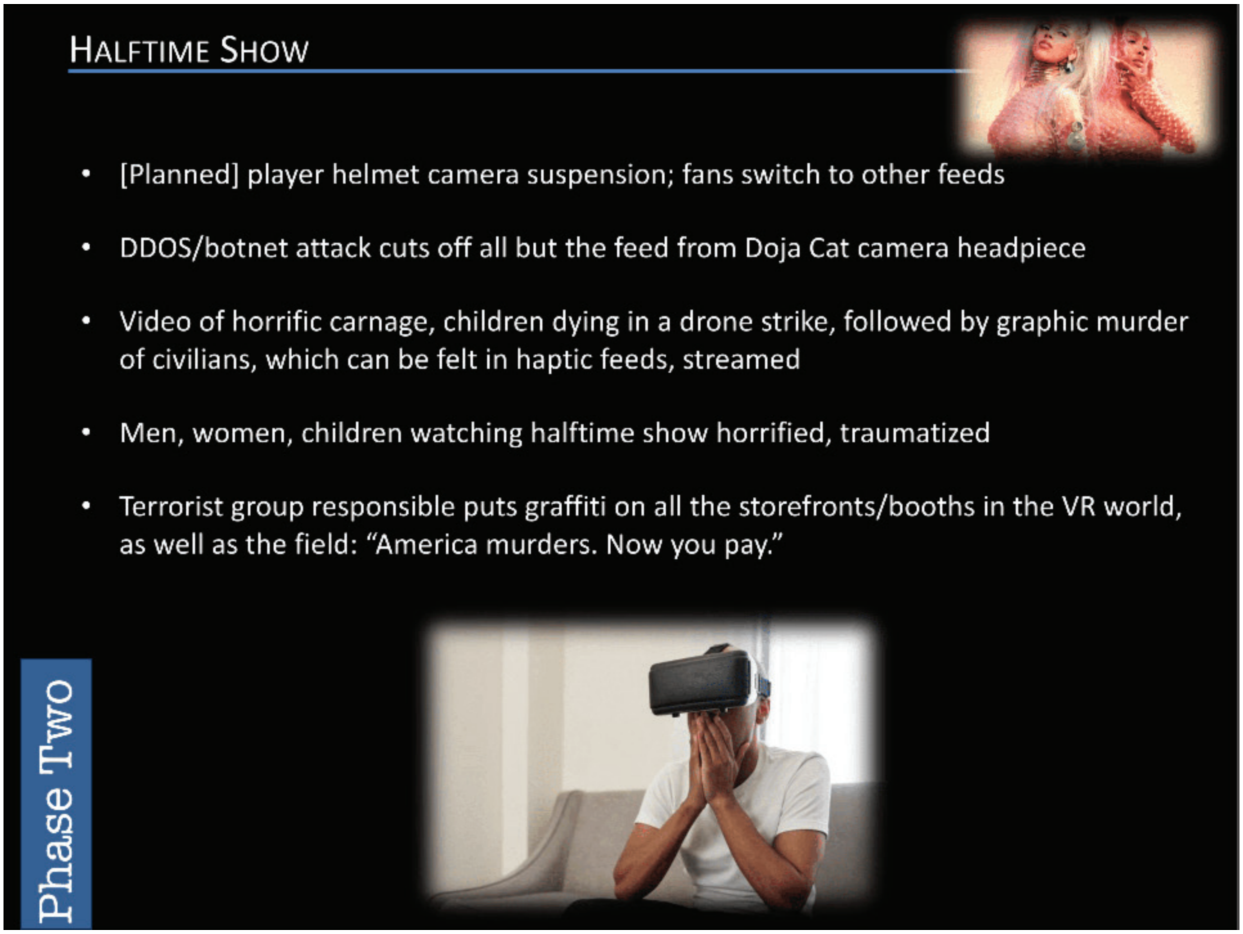
- I hear the SEC proposed this week to make some forms of software "exchanges" for regulatory purposes, freaking a lot of crypto guys out
- Is there a "kill switch" for the sensory part of the meta verse?
- The 2021 Infrastructure bill—signed into law by President Biden on November 15, 2021—now requires that transactions involving digital assets exceeding $10,000 be reported to the IRS.
- Do we need different principles for dominant companies (market power distinction currently works)?
- This is what my colleague said, which perhaps will make more sense to you all than it did to me: The SEC just proposed changes to the definition of Alternative Trading Systems (aka exchanges). This is in response to the market manipulation and fraud in the recent frenzy around "meme stocks" like Gamestop. Folks in the crypto community are going berserk, arguing that the language would cover decentralized finance (DeFi) protocols. These are basically software codes on the blockchain that automatically process trades.
- That a work of authorship for copyright purposes does not mean that something can't be regulated as e.g. a contract
- The 1A for code as it pertains to blockchain (smart contract protocols) is really scary because smart contracts are governance - that would be a really interesting court case
- can regulate platform or the thing sold on the platform (or a third "thing")
- "content neutral processes" should have different 1A scrutiny (I think in web 2 and web 3)
- purpose, context and social function of code matters (same as speech in real life)

Halftime Show

Scenario Stage Two Summary: Terror Attack

Just before halftime, there are massive DDOS/botnet attacks on all VR live streams and festival booths (which isn't hard, because these were developed for just a one-time event). The halftime performer, Doja Cat, hosts a camera feed allowing VR audiences to view the stadium from her perspective. The feed, into which most viewers have been funneled both because it is the "coolest" and because others have been taken down by the attacks, is penetrated. Terrorists stream horrifying footage of a drone strike, followed by two feeds into viewers' haptic suits to control their movements and physical sensations. Half the users are transported into the feed of a terrorist murdering a child, and half the users are transported into the feed of that victim. In this way, everyone tuned in experiences the murder as though it were real, and they were either perpetrating it or the victim. The stream ends with a recorded message: "America murders. Now you pay."

The same message appears graffitied across the virtual festival booths–they were hacked too. The footage is widely covered by news outlets and social media feeds. With rampant secondary trauma or retraumatization from viewing others' recordings of event, panic ensues. Stunned reporters liken the attack to a virtual 9/11.



Halftime Show

- [Planned] player helmet camera suspension; fans switch to other feeds
- DDOS/botnet attack cuts off all but the feed from Doja Cat camera headpiece
- Video of horrific carnage, children dying in a drone strike, followed by graphic murder of civilians, which can be felt in haptic feeds, streamed
- Men, women, children watching halftime show horrified, traumatized
- Terrorist group responsible puts graffiti on all the storefronts/booths in the VR world, as well as the field: "America murders. Now you pay."

Phase Two

Risks

- Targeting of people who may have a mental health issue who don't even know it and wreaking havoc on their mental health
- Bringing to scale the targeting of certain vulnerable people to commit crimes and be the fall guy
- targeting people with greater potential to react violently in the real world to content
- Haptic suits could be used for large scale sexual assault
- Online harms can create serious harms in the physical world
- Risk to 9-1-1 emergency assistance capabilities (ability to announce warnings, ability of harmed individuals to call for help)
- Has Customs and Border Protection considered using avatars for security interviews? I'd imagine a terrorist would want to get their hands on this kind of technology
- Terrorists get their hands on Customs and Border Protection avatars
- Grooming facilitation
- How do you find someone engaging in criminal activity when they could be anywhere in the world? Enforcement seems to be a serious issue.
- The ease with which false identities can exist in this space

Principles

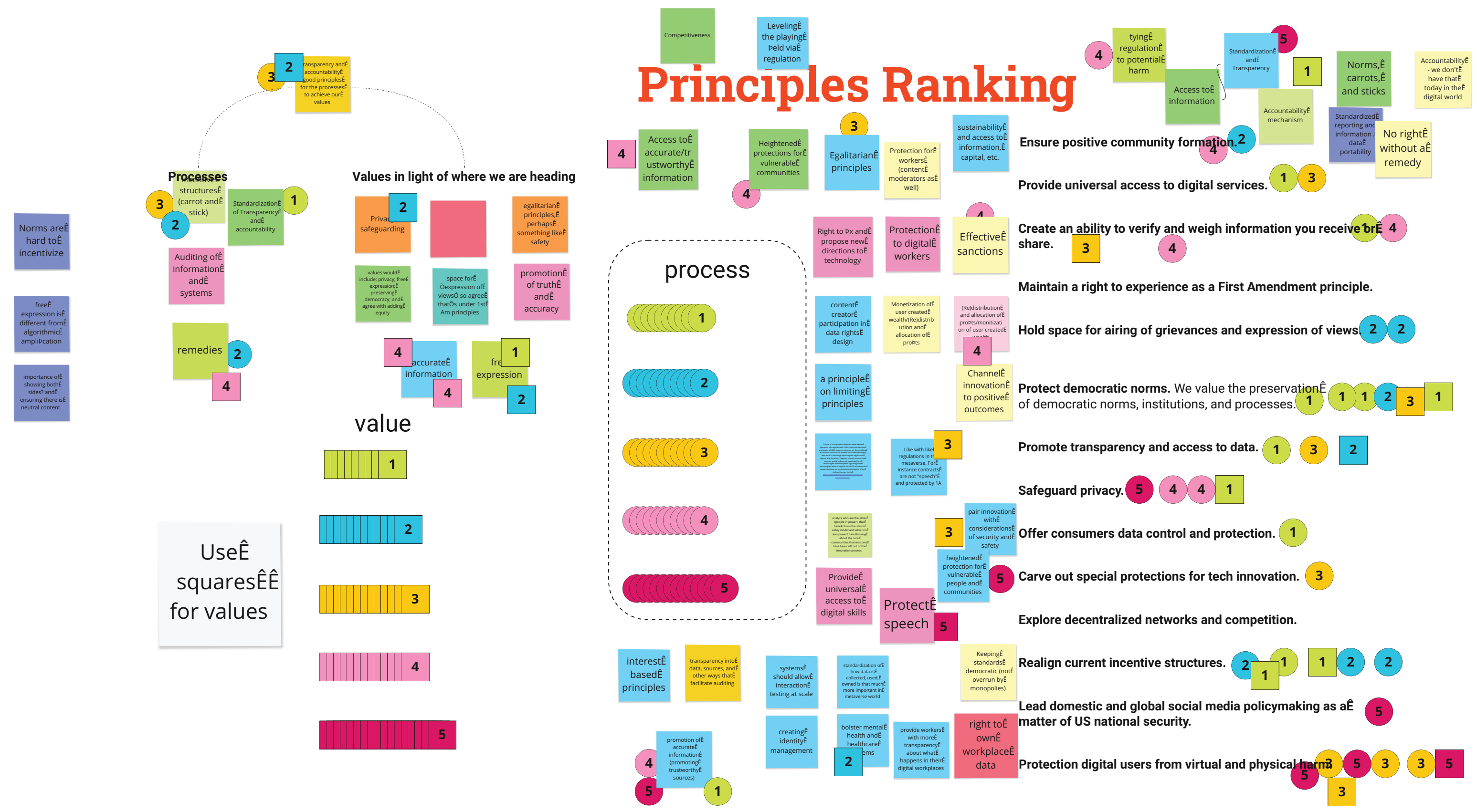
- standardization of how data is collected, used, owned is that much more important in metaverse world
- bolster mental health and healthcare systems
- pair innovation with considerations of security and safety
- heightened protection for vulnerable people and communities

Ideas/Questions

- this scenario also highlights the importance of keeping other channels of communication, because if the meta verse is taken over, it could not be a big deal if you know you have other channels and people have autonomy to know they can just turn off their device.
- how will we work with anticipatory harm?
- Why isn't code treated like medicine, i.e. drug/drug development, etc?

Activity 3: Highest Priorities

Participants marked highest priority principles and explored the difference between process principles and value principles.


ransparency andÊ accountabilityÊ good principlesÊ for the processesÊ to achieve ourÊ values
Processes
IncentiveÊ structuresÊ (carrot andÊ stick)
Standardization of TransparencyÊ andÊ accountability
Auditing ofÊ informationÊ andÊ systems
remedies
Norms areÊ hard toÊ incentivize
freeÊ expression isÊ different fromÊ algorithmicÊ ampliÞcation
importance ofÊ showing bothÊ sides? andÊ ensuring there isÊ neutral content.
Values in light of where we are heading
Privacy safeguarding
egalitarianÊ principles,Ê perhapsÊ something likeÊ safety
values wouldÊ include: privacy; freeÊ expression;Ê preservingÊ democracy; andÊ agree with addingÊ equity
space forÊ Òexpression ofÊ viewsÓ so agreeÊ thatÕs under 1stÊ Am principles
promotionÊ of truthÊ andÊ accuracy
accurateÊ information
free expression
value
UseÊ squaresÊÊ for values
process
Competitiveness
LevelingÊ the playingÊ Þeld viaÊ regulation
Principles Ranking
Access toÊ accurate/trustworthyÊ information
HeightenedÊ protections forÊ vulnerableÊ communities
EgalitarianÊ principles
Protection forÊ workersÊ (contentÊ moderators asÊ well)
sustainabilityÊ and access toÊ information,Ê capital, etc.
Right to Þx andÊ propose newÊ directions toÊ technology
ProtectionÊ to digitalÊ workers
EffectiveÊ sanctions
contentÊ creatorÊ participation inÊ data rightsÊ design
Monetization ofÊ user createdÊ wealth/(Re)distribution andÊ allocation ofÊ proÞts
(Re)distributionÊ and allocation ofÊ proÞts/monitization of user createdÊ wealth
ChannelÊ innovationÊ to positiveÊ outcomes
a principleÊ on limitingÊ principles
Like with likeÊ regulations in the metaverse. ForÊ instance contractsÊ are not "speech"Ê and protected by 1A
pair innovationÊ withÊ considerationsÊ of security andÊ safety
heightenedÊ protection forÊ vulnerableÊ people andÊ communities
ProvideÊ universalÊ access toÊ digital skills
ProtectÊ speech
KeepingÊ standardsÊ democratic (notÊ overrun byÊ monopolies)
interestÊ basedÊ principles
transparency intoÊ data, sources, andÊ other ways thatÊ facilitate auditing
systemsÊ should allowÊ interactionÊ testing at scale
standardization ofÊ how data isÊ collected, used,Ê owned is that muchÊ more important inÊ metaverse world
creatingÊ identityÊ management
bolster mentalÊ health andÊ healthcareÊ systems
provide workersÊ with moreÊ transparencyÊ about whatÊ happens in theirÊ digital workplaces
right toÊ ownÊ workplaceÊ data
promotion ofÊ accurateÊ informationÊ (promotingÊ trustworthyÊ sources)
tyingÊ regulationÊ to potentialÊ harm
Access toÊ information
StandardizationÊ andÊ Transparency
AccountabilityÊ mechanism
Norms,Ê carrots,Ê and sticks
StandardizedÊ reporting and information / dataÊ portability
No rightÊ without aÊ remedy
AccountabilityÊ - we don'tÊ have thatÊ today in theÊ digital world
1
2
3
4
5

Ensure positive community formation. 4 2

Provide universal access to digital services. 1 3

Create an ability to verify and weigh information you receive orÊ share. 1 4 3 4

Maintain a right to experience as a First Amendment principle.

Hold space for airing of grievances and expression of views. 2 2

Protect democratic norms. We value the preservationÊ of democratic norms, institutions, and processes. 1 1 1 2 3 1

Promote transparency and access to data. 1 3 2

Safeguard privacy. 5 4 4 1

Offer consumers data control and protection. 1

Carve out special protections for tech innovation. 3

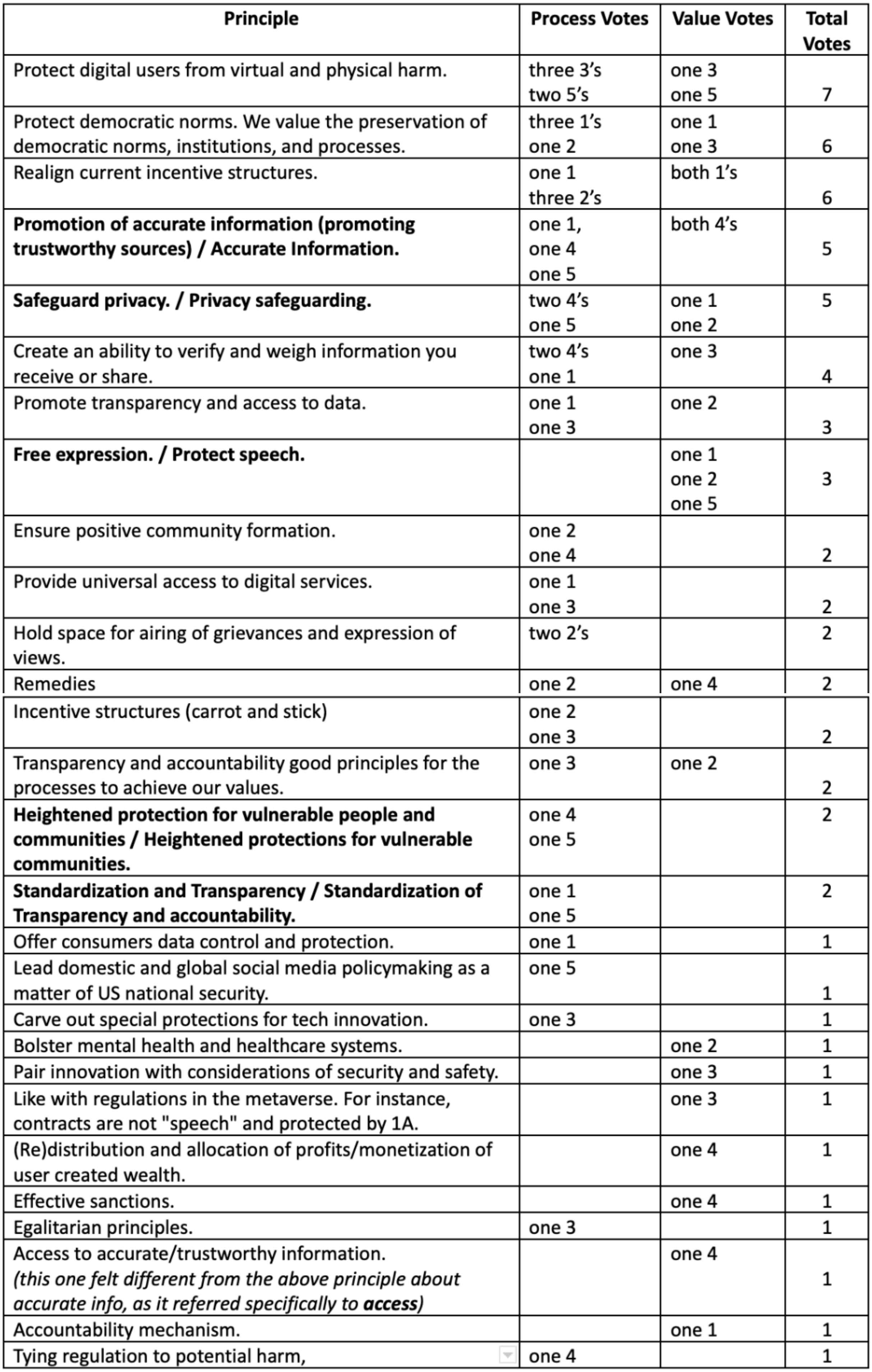
Explore decentralized networks and competition.

Realign current incentive structures. 2 1 1 1 2 2

Lead domestic and global social media policymaking as aÊ matter of US national security. 5

Protection digital users from virtual and physical harm 5 3 5 3 3 3 5

Value/Process Principle Ranking 1-5
P= process, V=value



Principle	Process Votes	Value Votes	Total Votes
Protect digital users from virtual and physical harm.	three 3's two 5's	one 3 one 5	7
Protect democratic norms. We value the preservation of democratic norms, institutions, and processes.	three 1's one 2	one 1 one 3	6
Realign current incentive structures.	one 1 three 2's	both 1's	6
Promotion of accurate information (promoting trustworthy sources) / Accurate Information.	one 1, one 4 one 5	both 4's	5
Safeguard privacy. / Privacy safeguarding.	two 4's one 5	one 1 one 2	5
Create an ability to verify and weigh information you receive or share.	two 4's one 1	one 3	4
Promote transparency and access to data.	one 1 one 3	one 2	3
Free expression. / Protect speech.		one 1 one 2 one 5	3
Ensure positive community formation.	one 2 one 4		2
Provide universal access to digital services.	one 1 one 3		2
Hold space for airing of grievances and expression of views.	two 2's		2
Remedies	one 2	one 4	2
Incentive structures (carrot and stick)	one 2 one 3		2
Transparency and accountability good principles for the processes to achieve our values.	one 3	one 2	2
Heightened protection for vulnerable people and communities / Heightened protections for vulnerable communities.	one 4 one 5		2
Standardization and Transparency / Standardization of Transparency and accountability.	one 1 one 5		2
Offer consumers data control and protection.	one 1		1
Lead domestic and global social media policymaking as a matter of US national security.	one 5		1
Carve out special protections for tech innovation.	one 3		1
Bolster mental health and healthcare systems.		one 2	1
Pair innovation with considerations of security and safety.		one 3	1
Like with regulations in the metaverse. For instance, contracts are not "speech" and protected by 1A.		one 3	1
(Re)distribution and allocation of profits/monetization of user created wealth.		one 4	1
Effective sanctions.		one 4	1
Egalitarian principles.	one 3		1
Access to accurate/trustworthy information. *(this one felt different from the above principle about accurate info, as it referred specifically to **access**)*		one 4	1
Accountability mechanism.		one 1	1
Tying regulation to potential harm,	one 4		1

Received No Votes

- Auditing of information and systems
- values would include: privacy; free expression; preserving democracy; and agree with adding equity
- space for "expression of views" so agree that's under 1st Amendment principles
- egalitarian principles, perhaps something like safety
- promotion of truth and accuracy
- Competitiveness
- Leveling the playing field via regulation
- Protection for workers (content moderators as well)
- sustainability and access to information, capital, etc.
- Right to fix and propose new directions to technology
- Protection to digital workers
- content creator participation in data rights design
- Monetization of user created wealth/(Re)distribution and allocation of profits
- a principle on limiting principles
- "There is no way mere coders or mere website operators can register with FINRA, track the identities and trades of AMM systems occurring on decentralized autonomous blockchain systems or otherwise comply with the ATS/ exchange reporting and registration regime, and therefore, if applied to such persons, this new rule would be banning a vast swathe of technologies and free speech regarding those technologies, which is beyond the SEC's authority and would constitute an unconstitutional violation of our civil and human rights." https://www.evernym.com/blog/w3c-vision-of-decentralization/
- analyze who are the elites (people in power) that benefit from the Silicon Valley model and who is in less power? I am thinking about the rural communities that exist and have been left out of the innovation process.
- Provide universal access to digital skills
- interest based principles
- transparency into data, sources, and other ways that facilitate auditing
- systems should allow interaction testing at scale
- standardization of how data is collected, used, owned is that much more important in metaverse world
- creating identity management
- provide workers with more transparency about what happens in their digital workplaces
- right to own workplace data
- Keeping standards democratic (not overrun by monopolies)
- Channel innovation to positive outcomes
- Access to information
- Norms, carrots, and sticks
- Standardized reporting and information / data portability
- Accountability - we don't have that today in the digital world
- No right without a remedy

Ideas

Norms are hard to incentivize
free expression is different from algorithmic amplification
importance of showing both sides? and ensuring there is neutral content.

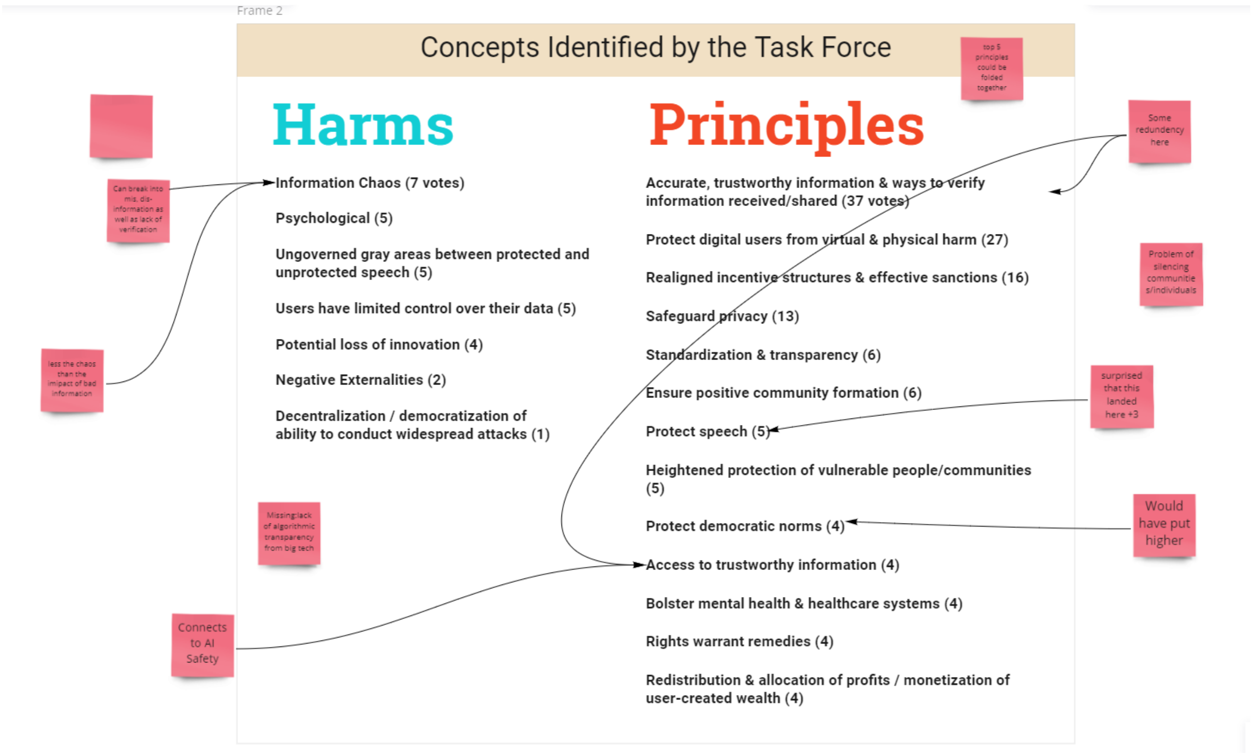
GEORGETOWN LAW

Social Media Task Force Meeting

Center on National Security and the Law

Workshop Three

Top Harms / Principles Review



Frame 2

Concepts Identified by the Task Force

Harms

Information Chaos (7 votes)

Psychological (5)

Ungoverned gray areas between protected and unprotected speech (5)

Users have limited control over their data (5)

Potential loss of innovation (4)

Negative Externalities (2)

Decentralization / democratization of ability to conduct widespread attacks (1)

Principles

Accurate, trustworthy information & ways to verify information received/shared (37 votes)

Protect digital users from virtual & physical harm (27)

Realigned incentive structures & effective sanctions (16)

Safeguard privacy (13)

Standardization & transparency (6)

Ensure positive community formation (6)

Protect speech (5)

Heightened protection of vulnerable people/communities (5)

Protect democratic norms (4)

Access to trustworthy information (4)

Bolster mental health & healthcare systems (4)

Rights warrant remedies (4)

Redistribution & allocation of profits / monetization of user-created wealth (4)

top 5 principles could be folded together

Some redundency here

Problem of silencing communitie s/individuals

surprised that this landed here +3

Would have put higher

Can break into mis, dis-information as well as lack of verification

less the chaos than the imipact of bad information

Missing:lack of algorithmic transparency from big tech

Connects to AI Safety

Under Information Chaos, two stickies:

- Can break into mis, dis-information as well as lack of verification
- less the chaos than the impact of bad information

Missing Harms

- lack of algorithmic transparency from big tech
- problem of silencing communities/individuals

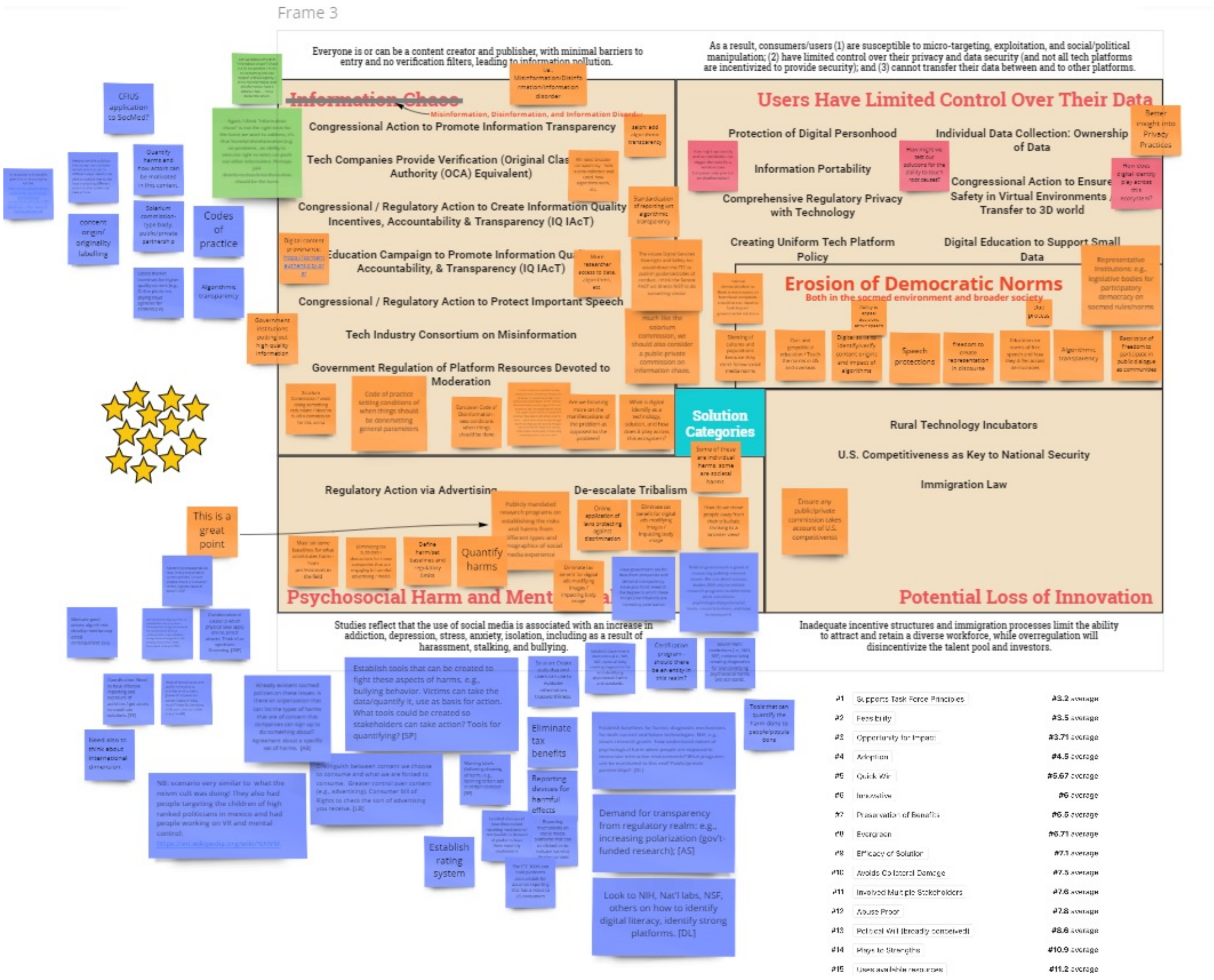
For both "Accurate, trustworthy information & ways to verify information received/shared" and "Access to trustworthy information", the following comments were made

- Some redundancy here
- Connects to AI safety

For Protect Speech and Protect democratic norms (note: these also seem to be redundant?) two comments expressed surprise at how far down the list they appeared:

- Surprised this landed here (with three +1s)
- Would have put higher

Solutions Review by Section


Frame 3
Everyone is or can be a content creator and publisher, with minimal barriers to entry and no verification filters, leading to information pollution.
~~Information Chaos~~
vs. Misinformation/Disinformation/Information disorder
Misinformation, Disinformation, and Information Disorder
Congressional Action to Promote Information Transparency
Tech Companies Provide Verification (Original Class Authority (OCA) Equivalent)
Congressional / Regulatory Action to Create Information Quality Incentives, Accountability & Transparency (IQ IAcT)
Education Campaign to Promote Information Quality, Accountability, & Transparency (IQ IAcT)
Congressional / Regulatory Action to Protect Important Speech
Tech Industry Consortium on Misinformation
Government Regulation of Platform Resources Devoted to Moderation
As a result, consumers/users (1) are susceptible to micro-targeting, exploitation, and social/political manipulation; (2) have limited control over their privacy and data security (and not all tech platforms are incentivized to provide security); and (3) cannot transfer their data between and to other platforms.
Users Have Limited Control Over Their Data
Protection of Digital Personhood
Individual Data Collection: Ownership of Data
Information Portability
Comprehensive Regulatory Privacy with Technology
Congressional Action to Ensure Safety in Virtual Environments Transfer to 3D world
Creating Uniform Tech Platform Policy
Digital Education to Support Small Data
Better insight into Privacy Practices
Erosion of Democratic Norms
Both in the socmed environment and broader society
Speech protections
Algorithmic transparency
Solution Categories
Some of these are individual harms, some are societal harms
Rural Technology Incubators
U.S. Competitiveness as Key to National Security
Immigration Law
Potential Loss of Innovation
Inadequate incentive structures and immigration processes limit the ability to attract and retain a diverse workforce, while overregulation will disincentivize the talent pool and investors.
Regulatory Action via Advertising
De-escalate Tribalism
Quantify harms
This is a great point
Psychosocial Harm and Ment
Studies reflect that the use of social media is associated with an increase in addiction, depression, stress, anxiety, isolation, and bullying, harassment, stalking, and bullying.
CFIUS application to SocMed?
Codes of practice
Algorithmic transparency
Eliminate tax benefits
Establish rating system
Demand for transparency from regulatory realm: e.g., increasing polarization (gov't-funded research). [AS]
Look to NIH, Nat'l labs, NSF, others on how to identify digital literacy, identify strong platforms. [DL]

#1	Supports Task Force Principles	#3.2 average
#2	Feasibility	#3.5 average
#3	Opportunity for Impact	#3.71 average
#4	Adoption	#4.5 average
#5	Quick Win	#5.67 average
#6	Innovative	#6 average
#7	Preservation of Benefits	#6.5 average
#8	Evergreen	#6.71 average
#9	Efficacy of Solution	#7.1 average
#10	Avoids Collateral Damage	#7.5 average
#11	Involved Multiple Stakeholders	#7.6 average
#12	Abuse Proof	#7.8 average
#13	Political Will (broadly conceived)	#8.6 average
#14	Plays to Strengths	#10.9 average
#15	Uses available resources	#11.2 average

General Comments

- What is digital identify as a technology, solution, and how does it play across this ecosystem?
- Some of these are individual harms, some are societal harms.
- Are we focusing more on the manifestations of the problem as opposed to the problem?

Information Chaos / Misinformation, Disinformation, and Information Disorder
Note: discussion around this section led to a change in name:

- Can we discuss the term "information chaos"? Chaos is a bit too passive, I think. It's something that can happen without agency. I think misinformation and disinformation have a different feel ... more deliberate action.
- Again, I think "information chaos" is not the right term for the harm we want to address; it's that harmful disinformation (e.g. on pandemic, on ability to exercise right to vote) can push out other information. Perhaps just disinformation/misinformation should be the harm

Pre-Criteria Discussion Solutions

- add algorithmic transparency
- Standardization of reporting with regards to algorithmic transparency
- We need broader transparency - how is data collected and used, how algorithms work, etc.
- The House Digital Services Oversight and Safety Act would direct the FTC to publish guidance/codes of conduct, I think the Senate PACT act directs NIST to do something similar

- More researcher access to data, algorithms, etc.
- Digital content provenance: https://contentauthenticity.org/
- Government institutions putting out high quality information
- much like the Solarium Commission, we should also consider a public-private commission on information chaos.
- Solarium Commission / avoid doing something redundant / Need to build a commission for this arena
- Code of practice setting conditions of when things should be done/setting general parameters
- European Code of Disinformation - sets conditions when things should be done
- I think it can be important to also measure harm in terms of money lost. because it is an amount that more people can understand. So for instance if Facebook bans certain pages from selling products, we can relate that to money the pages would lose, or amount of revenue they would not obtain due to the harm. harm also involves psychological harm but that can be hard to measure and quantify for people to rapidly understand. hence why i think using monetary terms can be useful.

Post-Criteria Discussion Solutions

- CFIUS application to SocMed?
- In response to Amanda's question on encouraging suicide, https://www.law.georgetown.edu/american-criminal-law-review/wp-content/uploads/sites/15/2019/01/56-1-The-Puzzle-of-Inciting-Suicide.pdf
- Need to be able to follow the money: can motivate certain actors to act in different ways. Need to be able to analyze that to see how motivating different actors so that others can take actions.
- Quantify harms and how actors can be motivated in this context.
- content origin/ originality labeling
- Solarium commission-type body, public/private partnership
- Codes of practice
- Create market incentives for higher quality content (e.g., Online platforms paying news agencies for content) [IW]
- Algorithmic transparency

Users Have Limited Control Over Their Data

Pre-Criteria Discussion Solutions

- Better insight into Privacy Practices

General Comments:

- How might we identify and set conditions that trigger the need for a solution? (see European code practice on disinformation)
- How might we test our solutions for the ability to touch root causes?
- How does digital identity play across this ecosystem?

Erosion of Democratic Norms

Pre-Criteria Discussion Solutions

- Internal democratization so there is more voices on how these companies should be run. Need to look beyond governmental solutions
- Silencing of cultures and populations because they don't follow social media norms
- Civic and geopolitical education / Teach the norms in US and overseas
- Ability to appeal decisions about speech
- Digital skills to identify/verify content origins and impact of algorithms
- Speech protections
- freedom to create representation in discourse
- Due process
- Education on norms of free speech and how they differ across democracies
- Algorithmic transparency
- Representative Institutions: e.g., legislative bodies for participatory democracy on socmed rules/norms

- Restriction of Freedom to participate in public dialogue as communities

Psychosocial Harm and Mental Health

Pre-Criteria Discussion Solutions

- Publicly mandated research programs on establishing the risks and harms from different types and demographics of social media experience
- Must set some baselines for what constitutes harm - from professionals in the field
- Eliminating tax subsidies / deductions for those companies that are engaging in harmful advertising / media
- Define harm/set baselines and regulatory limits
- Quantify harms
- Eliminate tax benefit for digital ads modifying images / impacting body image
- Online application of laws protecting against discrimination
- Eliminate tax benefit for digital ads modifying images / impacting body image
- How do we move people away from their tribalistic thinking to a broader view?

Post-Criteria Discussion Solutions

- Advertising/propaganda/political lines are blurred by current policies. Unsure whether there is a solution to this. Express cause of action? IIED?
- Motivate good actors: algorithmic development/unexpected consequences. [SS]
- identification of legal barriers to coordination on grey areas. Development of legal framework for coordination without antitrust/other vulnerabilities. And problem of migration. NB: ID Migratory Harms. Have framework, instead of [DL]
- Consideration of extent to which physical laws apply online. Direct attacks. Think of as spectrum. Grooming. [SBF]
- Coordination: Need to have effective reporting and exposure of activities / get actors to coordinate solutions. [SS]
- Issue of recruiting on one platform/moving to another environment. [Level of cooperation across industry? How much? Have for terrorism,, child porn, but not other scenarios AB]
- Need also to think about international dimension.
- Already existent socmed policies on these issues. is there an organization that can list the types of harms that are of concern that companies can sign up to do something about? Agreement about a specific set of harms. [AB]
- NB: scenario very similar to what the nxivm cult was doing! They also had people targeting the children of high ranked politicians in Mexico and had people working on VR and mental control: https://en.wikipedia.org/wiki/NXIVM
- Establish tools that can be created to fight these aspects of harms. e.g., bullying behavior. Victims can take the data/quantify it, use as basis for action. What tools could be created so stakeholders can take action? Tools for quantifying? [SP]
- Distinguish between content we choose to consume and what we are forced to consume. Greater control over content (e.g., advertising). Consumer bill of Rights to check the sort of advertising you receive. [LB]
- Warning labels (following showing of harm). e.g., banning certain ads in certain contexts [IW]
- Establish rating system
- Have government ask for data from companies and demand transparency. Have govt fund research the degree to which these things (like tribalism) are increasing polarization
- Federal government is good at resourcing publicly relevant issues. We can direct various bodies (NIH, etc) to initiate research programs to determine what constitutes psychological/psychosocial harm, create baselines, and how to measure it.
- Solution: Create tools that end users can use to evaluate information trustworthiness.
- Solution: Government institutions (i.e., NIH, NSF, national labs) creating diagnostics for and identifying psychosocial harms and standards.
- Certification program - should there be an entity in this realm?
- Government institutions (i.e., NIH, NSF, national labs) creating diagnostics for and identifying psychosocial harms and standards.

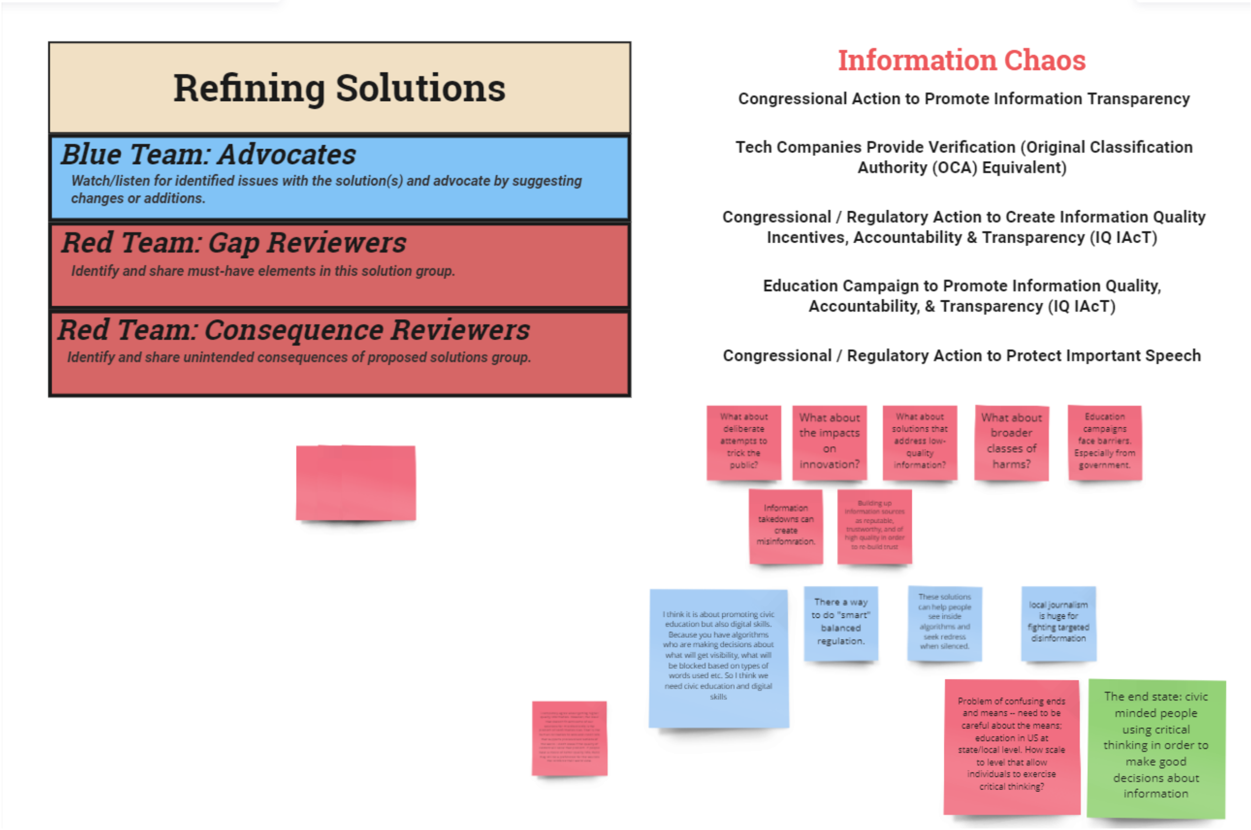
- Eliminate tax benefits
- Reporting devices for harmful effects
- Tools that can quantify the harm done to people/populations
- Certified sites would have these robust reporting mechanisms? Not feasible to demand all platforms have these reporting mechanisms
- Reporting mechanisms on social media platforms that can be clicked on to indicate harmful, abusive content
- The FTC 100% can hold platforms accountable for accurate reporting that has a nexus to US consumers
- Establish baselines for harms: diagnostic mechanisms for both current and future technologies. NIH, e.g., issues research grants--how understand extent of psychological harm when people are exposed to immersive interactive environments? What programs can be marshaled to this end? Public/private partnerships? [DL]
- Demand for transparency from regulatory realm: e.g., increasing polarization (gov't-funded research); [AS]
- Look to NIH, Nat'l labs, NSF, others on how to identify digital literacy, identify strong platforms. [DL]

<u>Potential Loss of Innovation</u>

Pre-Criteria Discussion Solutions

- Ensure any public/private commission takes account of U.S. competitiveness

Refining Solutions Red Team / Blue Team Discussion around Information Chaos/ Misinformation, Disinformation, and Information Disorder



Refining Solutions

Blue Team: Advocates
Watch/listen for identified issues with the solution(s) and advocate by suggesting changes or additions.

Red Team: Gap Reviewers
Identify and share must-have elements in this solution group.

Red Team: Consequence Reviewers
Identify and share unintended consequences of proposed solutions group.

Information Chaos

Congressional Action to Promote Information Transparency

Tech Companies Provide Verification (Original Classification Authority (OCA) Equivalent)

Congressional / Regulatory Action to Create Information Quality Incentives, Accountability & Transparency (IQ IAcT)

Education Campaign to Promote Information Quality, Accountability, & Transparency (IQ IAcT)

Congressional / Regulatory Action to Protect Important Speech

What about deliberate attempts to trick the public?

What about the impacts on innovation?

What about solutions that address low-quality information?

What about broader classes of harms?

Education campaigns face barriers. Especially from government.

Information takedowns can create misinfomration.

Building up information sources as reputable, trustworthy, and of high quality in order to re-build trust

I think it is about promoting civic education but also digital skills. Because you have algorithms who are making decisions about what will get visibility, what will be blocked based on types of words used etc. So I think we need civic education and digital skills

There a way to do "smart" balanced regulation.

These solutions can help people see inside algorithms and seek redress when silenced.

local journalism is huge for fighting targeted disinformation

Problem of confusing ends and means -- need to be careful about the means; education in US at state/local level. How scale to level that allow individuals to exercise critical thinking?

The end state: civic minded people using critical thinking in order to make good decisions about information

Red Team

- I completely agree about getting higher quality information. However, the issue that doesn't fit with some of our solutions for misinfo/disinfo is the problem of confirmation bias. That is, the human inclination to seek and credit info that supports preconceived notions of the world. I don't know if the quality of content will solve that problem. If people have a choice of better quality info, there may still be a preference for the sources the reinforce their world view.
- What about deliberate attempts to trick the public?
- What about the impacts on innovation?
- What about solutions that address low-quality information?
- What about broader classes of harms?
- Education campaigns face barriers. Especially from government.
- Information takedowns can create misinformation.
- Building up information sources as reputable, trustworthy, and of high quality in order to rebuild trust
- Problem of confusing ends and means -- need to be careful about the means; education in US at state/local level. How scale to level that allow individuals to exercise critical thinking?

Blue Team

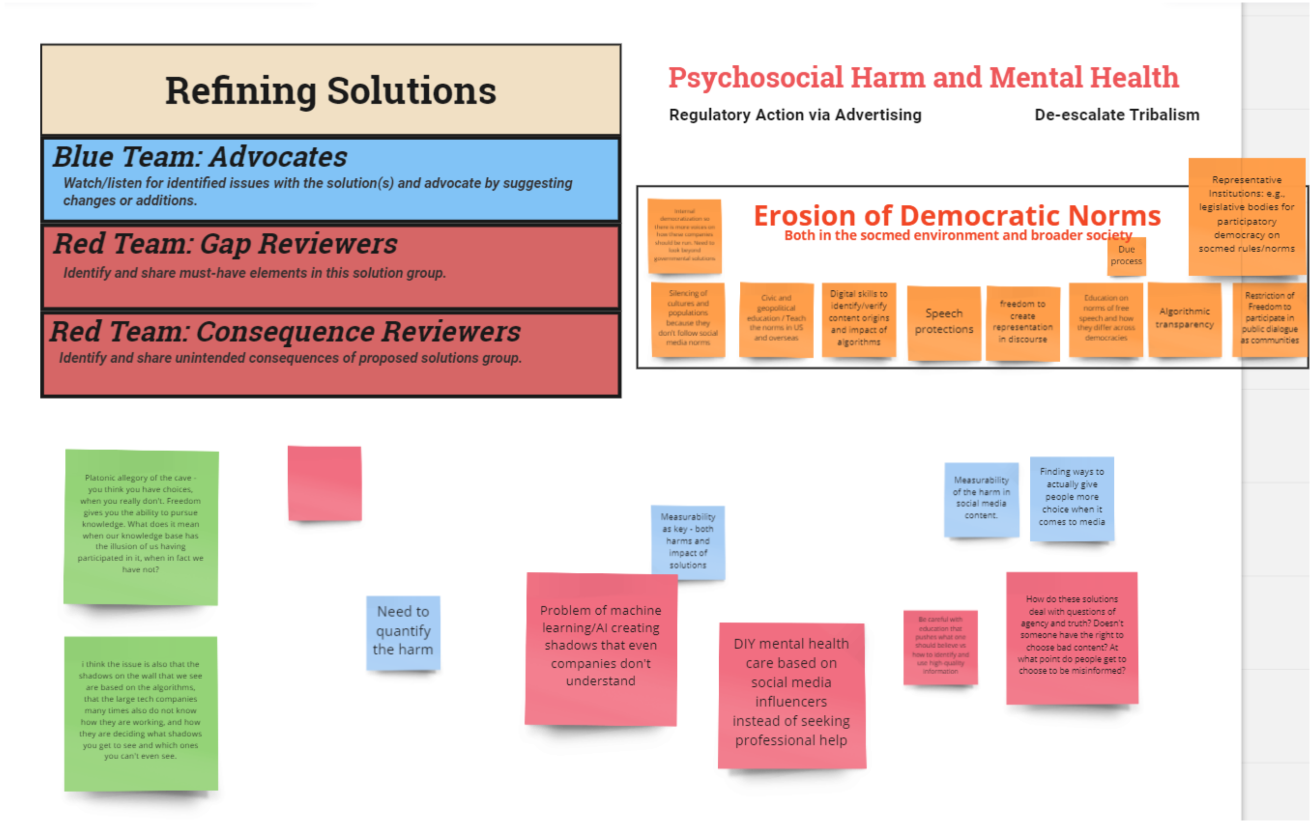
- I think it is about promoting civic education but also digital skills. Because you have algorithms who are making decisions about what will get visibility, what will be blocked based on types of words used etc. So I think we need civic education and digital skills

- There a way to do “smart” balanced regulation.
- These solutions can help people see inside algorithms and seek redress when silenced.
- local journalism is huge for fighting targeted disinformation

Additional Comment

- The end state: civic minded people using critical thinking in order to make good decisions about information

Refining Solutions Red Team / Blue Team Discussion around Psychosocial Harm and Mental Health and Erosion of Democratic Norms



Refining Solutions

Blue Team: Advocates

Watch/listen for identified issues with the solution(s) and advocate by suggesting changes or additions.

Red Team: Gap Reviewers

Identify and share must-have elements in this solution group.

Red Team: Consequence Reviewers

Identify and share unintended consequences of proposed solutions group.

Psychosocial Harm and Mental Health

Regulatory Action via Advertising

De-escalate Tribalism

Erosion of Democratic Norms

Both in the socmed environment and broader society

Internal democratization so there is more voices on how these companies should be run. Need to look beyond governmental solutions

Due process

Representative Institutions: e.g., legislative bodies for participatory democracy on socmed rules/norms

Silencing of cultures and populations because they don't follow social media norms

Civic and geopolitical education / Teach the norms in US and overseas

Digital skills to identify/verify content origins and impact of algorithms

Speech protections

freedom to create representation in discourse

Education on norms of free speech and how they differ across democracies

Algorithmic transparency

Restriction of Freedom to participate in public dialogue as communities

Platonic allegory of the cave - you think you have choices, when you really don't. Freedom gives you the ability to pursue knowledge. What does it mean when our knowledge base has the illusion of us having participated in it, when in fact we have not?

i think the issue is also that the shadows on the wall that we see are based on the algorithms, that the large tech companies many times also do not know how they are working, and how they are deciding what shadows you get to see and which ones you can't even see.

Need to quantify the harm

Measurability as key - both harms and impact of solutions

Problem of machine learning/AI creating shadows that even companies don't understand

DIY mental health care based on social media influencers instead of seeking professional help

Measurability of the harm in social media content.

Finding ways to actually give people more choice when it comes to media

Be careful with education that pushes what one should believe vs how to identify and use high-quality information

How do these solutions deal with questions of agency and truth? Doesn't someone have the right to choose bad content? At what point do people get to choose to be misinformed?

Red Team

- Problem of machine learning/AI creating shadows that even companies don't understand
- DIY mental health care based on social media influencers instead of seeking professional help
- Be careful with education that pushes what one should believe vs how to identify and use high-quality information
- How do these solutions deal with questions of agency and truth? Doesn't someone have the right to choose bad content? At what point do people get to choose to be misinformed?

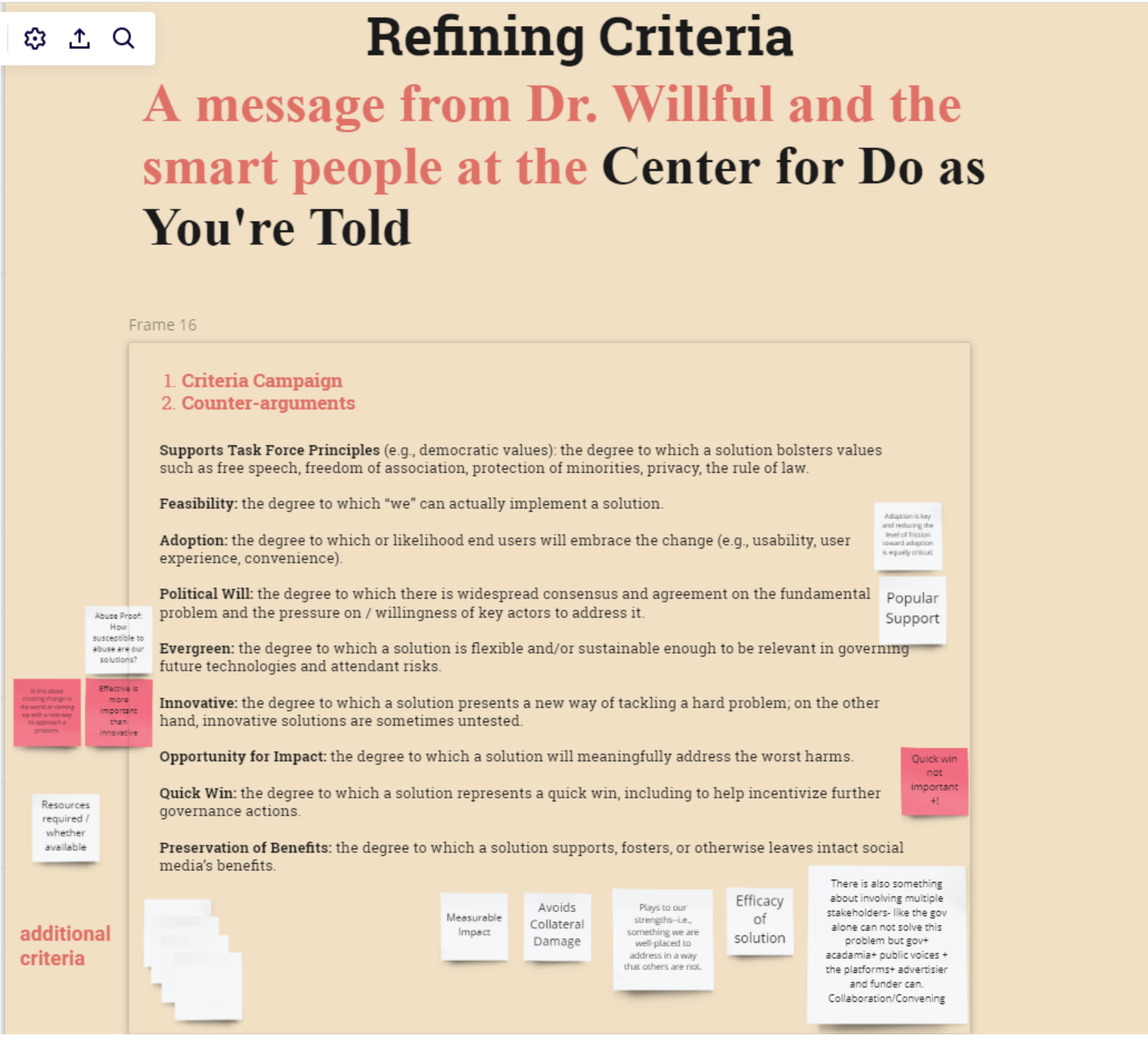
Blue Team

- Need to quantify the harm
- Measurability as key - both harms and impact of solutions
- Measurability of the harm in social media content.
- Finding ways to actually give people more choice when it comes to media

Additional Comments

- Platonic allegory of the cave - you think you have choices, when you really don't. Freedom gives you the ability to pursue knowledge. What does it mean when our knowledge base has the illusion of us having participated in it, when in fact we have not?
- I think the issue is also that the shadows on the wall that we see are based on the algorithms, that the large tech companies many times also do not know how they are working, and how they are deciding what shadows you get to see and which ones you can't even see.

Refining Criteria Discussion



Refining Criteria

A message from Dr. Willful and the smart people at the Center for Do as You're Told

Frame 16

1. Criteria Campaign
2. Counter-arguments

Supports Task Force Principles (e.g., democratic values): the degree to which a solution bolsters values such as free speech, freedom of association, protection of minorities, privacy, the rule of law.

Feasibility: the degree to which "we" can actually implement a solution.

Adoption: the degree to which or likelihood end users will embrace the change (e.g., usability, user experience, convenience).

Political Will: the degree to which there is widespread consensus and agreement on the fundamental problem and the pressure on / willingness of key actors to address it.

Evergreen: the degree to which a solution is flexible and/or sustainable enough to be relevant in governing future technologies and attendant risks.

Innovative: the degree to which a solution presents a new way of tackling a hard problem; on the other hand, innovative solutions are sometimes untested.

Opportunity for Impact: the degree to which a solution will meaningfully address the worst harms.

Quick Win: the degree to which a solution represents a quick win, including to help incentivize further governance actions.

Preservation of Benefits: the degree to which a solution supports, fosters, or otherwise leaves intact social media's benefits.

Adoption is key and reducing the level of friction toward adoption is equally critical.

Popular Support

Abuse Proof: How susceptible to abuse are our solutions?

Is this about creating change in the world or coming up with a new way to approach a problem.

Effective is more important than innovative

Quick win not important +1

Resources required / whether available

additional criteria

Measurable Impact

Avoids Collateral Damage

Plays to our strengths--i.e., something we are well-placed to address in a way that others are not.

Efficacy of solution

There is also something about involving multiple stakeholders- like the gov alone can not solve this problem but gov+ acadamia+ public voices + the platforms+ advertisier and funder can. Collaboration/Convening

Comment under Evergreen

- Abuse Proof: How susceptible to abuse are our solutions?

Comments under Innovative

- Is this about creating change in the world or coming up with a new way to approach a problem.
- Effective is more important than innovative

Comment under Political Will

- Popular Support

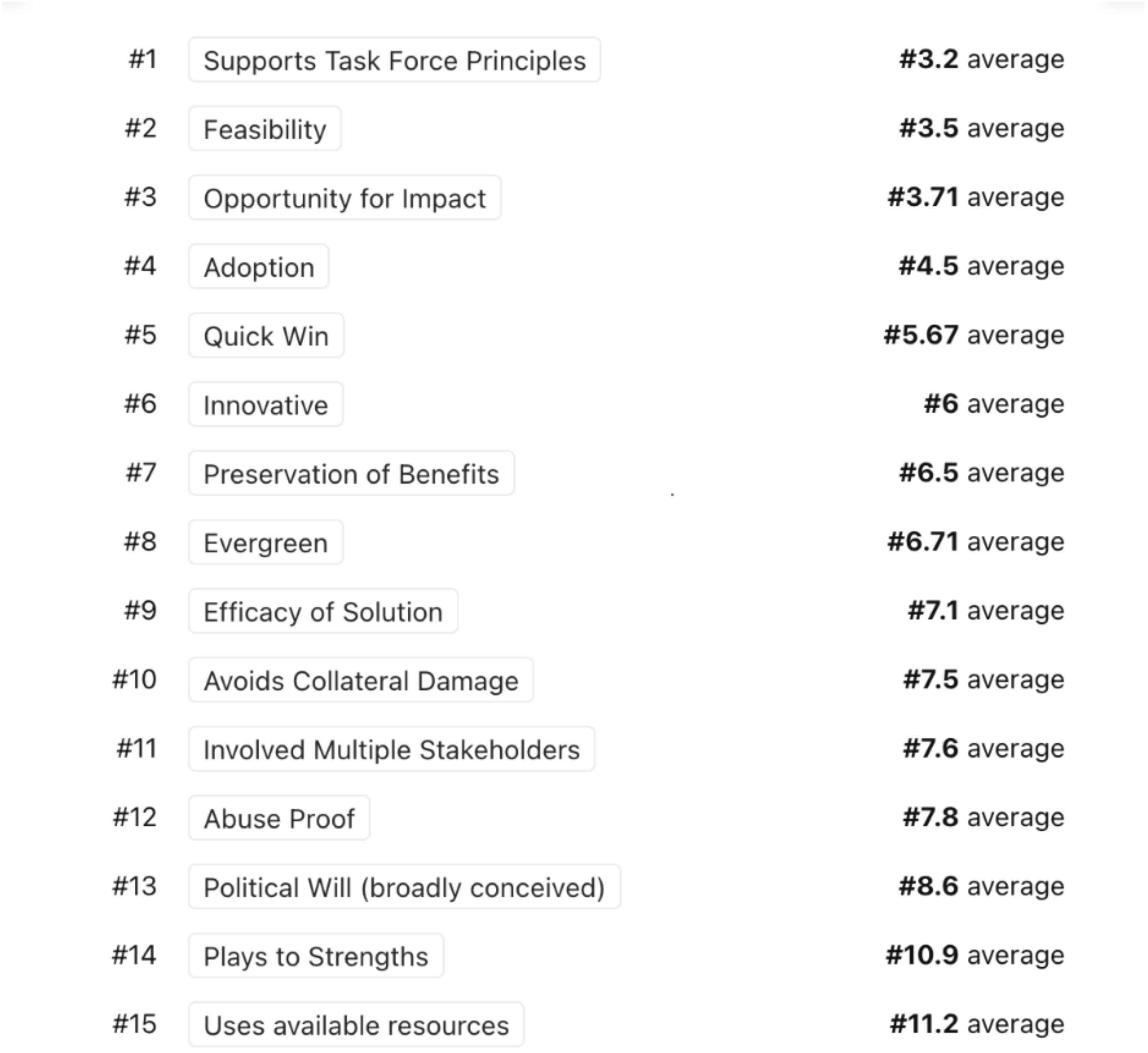
Comment under Quick Win

- Quick win not important (with one +1)

Additional Criteria

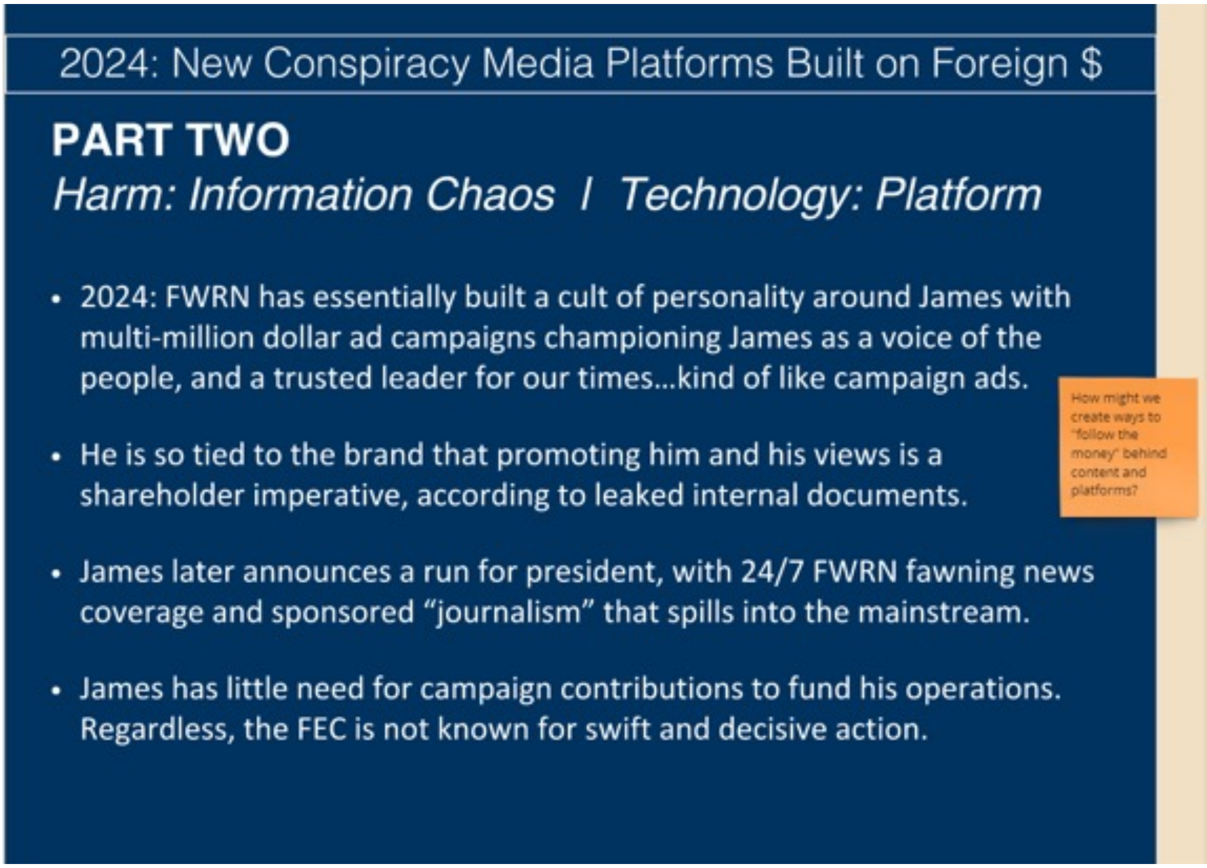
- Measurable Impact
- Avoids Collateral Damage
- Plays to our strengths--i.e., something we are well-placed to address in a way that others are not.
- Efficacy of solution
- There is also something about involving multiple stakeholders- like the gov alone can not solve this problem but gov+ academia + public voices + the platforms + advertiser and funder can. Collaboration/Convening

Criteria, including additional criteria, Card-Sorted Rank



#1	Supports Task Force Principles	**#3.2** average
#2	Feasibility	**#3.5** average
#3	Opportunity for Impact	**#3.71** average
#4	Adoption	**#4.5** average
#5	Quick Win	**#5.67** average
#6	Innovative	**#6** average
#7	Preservation of Benefits	**#6.5** average
#8	Evergreen	**#6.71** average
#9	Efficacy of Solution	**#7.1** average
#10	Avoids Collateral Damage	**#7.5** average
#11	Involved Multiple Stakeholders	**#7.6** average
#12	Abuse Proof	**#7.8** average
#13	Political Will (broadly conceived)	**#8.6** average
#14	Plays to Strengths	**#10.9** average
#15	Uses available resources	**#11.2** average

Scenario Comments



2024: New Conspiracy Media Platforms Built on Foreign $

PART TWO

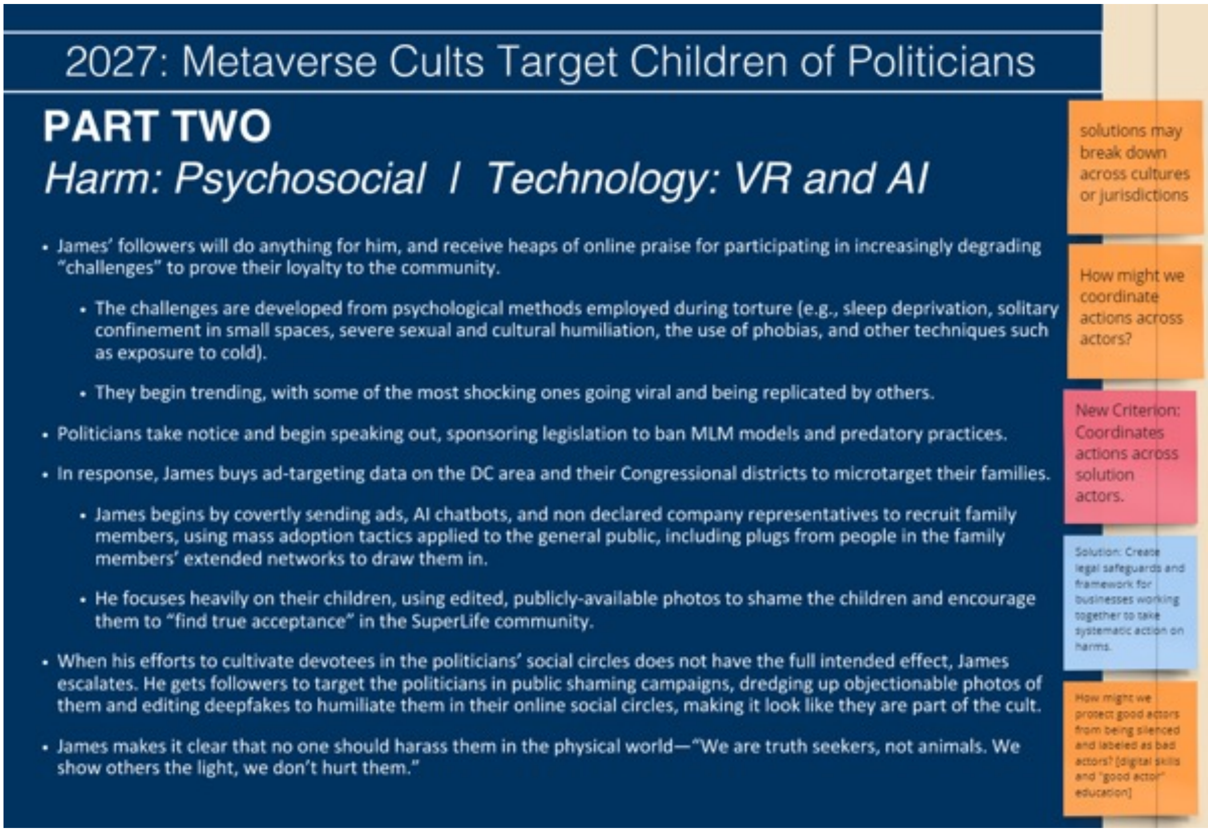
Harm: Information Chaos | Technology: Platform

- 2024: FWRN has essentially built a cult of personality around James with multi-million dollar ad campaigns championing James as a voice of the people, and a trusted leader for our times…kind of like campaign ads.
- He is so tied to the brand that promoting him and his views is a shareholder imperative, according to leaked internal documents.
- James later announces a run for president, with 24/7 FWRN fawning news coverage and sponsored "journalism" that spills into the mainstream.
- James has little need for campaign contributions to fund his operations. Regardless, the FEC is not known for swift and decisive action.

How might we create ways to "follow the money" behind content and platforms?

Comment:

- How might we create ways to "follow the money" behind content and platforms?



2027: Metaverse Cults Target Children of Politicians

PART TWO

Harm: Psychosocial I Technology: VR and AI

- James' followers will do anything for him, and receive heaps of online praise for participating in increasingly degrading "challenges" to prove their loyalty to the community.
  - The challenges are developed from psychological methods employed during torture (e.g., sleep deprivation, solitary confinement in small spaces, severe sexual and cultural humiliation, the use of phobias, and other techniques such as exposure to cold).
  - They begin trending, with some of the most shocking ones going viral and being replicated by others.
- Politicians take notice and begin speaking out, sponsoring legislation to ban MLM models and predatory practices.
- In response, James buys ad-targeting data on the DC area and their Congressional districts to microtarget their families.
  - James begins by covertly sending ads, AI chatbots, and non declared company representatives to recruit family members, using mass adoption tactics applied to the general public, including plugs from people in the family members' extended networks to draw them in.
  - He focuses heavily on their children, using edited, publicly-available photos to shame the children and encourage them to "find true acceptance" in the SuperLife community.
- When his efforts to cultivate devotees in the politicians' social circles does not have the full intended effect, James escalates. He gets followers to target the politicians in public shaming campaigns, dredging up objectionable photos of them and editing deepfakes to humiliate them in their online social circles, making it look like they are part of the cult.
- James makes it clear that no one should harass them in the physical world—"We are truth seekers, not animals. We show others the light, we don't hurt them."

solutions may break down across cultures or jurisdictions

How might we coordinate actions across actors?

New Criterion: Coordinates actions across solution actors.

Solution: Create legal safeguards and framework for businesses working together to take systematic action on harms.

How might we protect good actors from being silenced and labeled as bad actors? [digital skills and "good actor" education]

Comments:

- solutions may break down across cultures or jurisdictions
- How might we coordinate actions across actors?
- New Criterion: Coordinates actions across solution actors.
- Solution: Create legal safeguards and framework for businesses working together to take systematic action on harms.
- How might we protect good actors from being silenced and labeled as bad actors? [digital skills and "good actor" education]

GEORGETOWN LAW

Social Media Task Force Meeting

Center on National Security and the Law

Workshop Four

Dr. Laura Donohue began the fourth taskforce workshop by thanking the members for their hard work, and then gave a quick review of the work the taskforce had done over the course of the previous three workshops.

The goal of this workshop was to do a section-by-section review of the drafted report in order to:

- Surface disagreement and reach consensus on a broad approach.
- Adopt at least one forward-looking, innovative recommendation in each area.
- Push into subsidiary areas as much as possible.

Taskforce members were reminded that after the meeting, there were going to be additional time periods where they would be allowed to offer comments, ideas, and alternative verbiage for the report. The goal was for the report to be finalized and published on July 15, 2022.

Jenny Reich then led the group through a deep discussion of each of the main sections of the report:

1. Public Empowerment
2. Responsible Platforms
3. Effective Governance

Each resolution was given approximately 30 minutes for discussion, with the task force members commenting about everything from the overall idea to the nuances of the text itself, up to and including suggesting both edits to the existing text as well as entirely new verbiage.

At the end of the meeting, it was decided that a fifth meeting was needed in order to have more time to go over the remaining two sections of the report: the Digital Bill of Rights and the Digital Hippocratic Oath.

Dr. Donohue then presented a look into the future,including suggestions on how the taskforce members could continue to stay involved:

Report launch

1. Post on social media
2. Share with your networks

Implementation talks

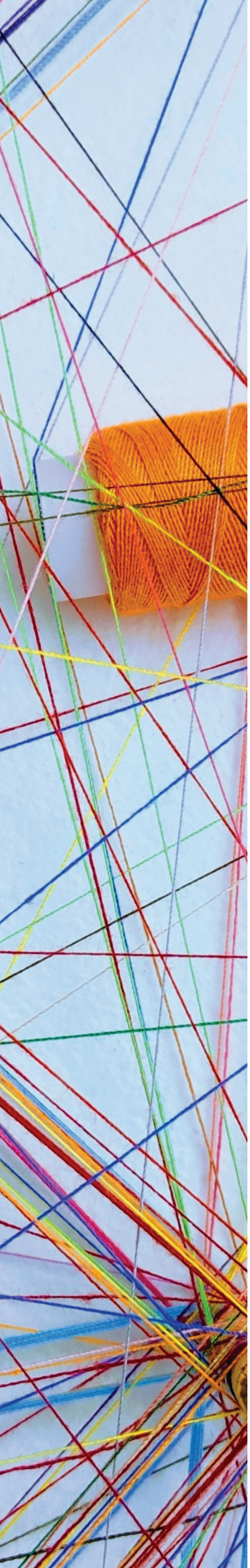
1. Public event/panels
2. Legislators
3. Tech platforms
4. Other stakeholders

International Expansion

1. Expert participation
2. New connections

Center Involvement

1. Meet our students
2. Attend our events
3. Keep these relationships going



Social Media Task Force Meeting

Center on National Security and the Law

Task Force Retrospective

Recommendations for Future Approaches to Task Force Projects

These recommendations present opportunities for constituting and facilitating the work of future task forces related to national security and social media. The comments below are based on Fluid Hive's experience with the Center on National Security at Georgetown Law's (CNS) social media and national security taskforce and our expertise in helping people think and solve like a designer.

Task Force Composition

CNS convened an exceptional group of thinkers with perspectives on multiple areas connected to social media and national security. Each participant was able to share insights into how their expertise connected to the questions and challenges we wrestled with.

Future task forces should consider three additional voices:

Social Work — A team member with a social work background would help the task force connect impacts, opportunities, and solutions to how communities function. This voice would also offer insights into community-level unintended consequences others might miss.

Behavioral Science — Many of the challenges and solutions in the social media and national security space involve understanding why people are behaving the way they are now and how we might influence those behaviors in the future. Someone with experience applying behavioral science to service design could open pathways into new kinds of solutions and experiments. They can also help make hidden emerging threats visible.

User Experience Design — Someone with experience designing aspects of the social media landscape will bring a designer's sensibilities to the task force work. This voice will help task force members understand how decisions are made when designing social media platforms and how we might influence those decisions. They can also provide insight into what it might mean to redesign aspects of a platform in response to a governance solution.

Task Force Approach

To deepen the impact of human-centered design and design thinking we recommend:

1. **Drip Communications:** Compiling background information and subject matter expert insights then sending them out in small weekly packages to make the information more digestible during the early days of the task force.

2. **Quick Launch:** A 1-hour to 2-hour task force launch event that builds relationships and reveals expertise. Something similar to the expertise mapping we did in Workshop 1 is enough.

3. **2/3-synthesize-all Pattern:** After the launch, follow a pattern that begins with a facilitated conversation between small groups of 2 or 3 task force members, then gives CNS time to synthesize, then brings the whole taskforce together to build on that synthesis. The small group work develops relationships while people surface problems, opportunities and challenges. CNS gathers all of the small-group data and makes inferential leaps to suggest frameworks, models, structures, or a collection of issues the synthesized data might contain. CNS would in effect make prototypes. Task force members would be presented with CNS prototypes as things to critique and develop during workshops while pursuing a set of criteria for effective solutions.

4. **Criteria Lead:** Focus on what an effective solution should accomplish (criteria) before attempting to create possible solutions. Developing the set of criteria for effective solutions, and using it when it is time to generate solutions, will focus what the task force creates on solution ideas connected to those criteria. Also, the solution criteria connected to a particular topic area can serve a broad range of people interested in that area and serve as a tool people can continue to develop over time.

5. **Criteria Focused Scenario Play:** Use scenarios after developing a set of criteria to push those criteria into the future. Instead of focusing on grasping a future world, future problems, future behaviors, and future consequences, task force members would focus on a set of criteria to evaluate and improve against imagined futures.

6. **Toolkit:** Create a simple toolkit to help people use the task force's solution criteria to develop their own solutions based on how they are positioned to act and create change.

Appendix F: Financial Report

GEORGETOWN UNIVERSITY
Office of the Chief Financial Officer
Sponsored Projects Financial Operations

	Date:	June 28, 2022
New Venture Fund	**Financial Report:**	Final FSR
1201 Connecticut Ave NW Suite 300	GU Reference #:	AWD-7774504
Washington, DC 20036	Project Period:	01/01/2021-05/31/2022
	GU PI:	Cave, Anna

360 Tech: Innovation, Security & Governance
Reporting Period: 01/01/2021 - 5/31/2022

Expense Items	Approved Budget	Current Period	Cumulative Amount through Current Period	Remaining Balance
Personnel	$ 76,187.00	$ 74,027.67	$ 74,027.67	$ 2,159.33
Fringe	18,813.00	$ 21,017.39	$ 21,017.39	(2,204.39)
Supplies	-	$ -	$ -	-
Services	55,000.00	$ 52,104.94	$ 52,104.94	2,895.06
Sub Award Costs	-	$ -	$ -	-
Travel and Training	$ -	$ 2,850.00	$ 2,850.00	$ (2,850.00)
Total Direct Costs	$ 150,000.00	$ 150,000.00	$ 150,000.00	$ -
Indirect Costs	30,000.00	30,000.00	30,000.00	-
Total Expenses	$ 180,000.00	$ 180,000.00	$ 180,000.00	$ -

Certification: "By signing this report, I certify to the best of my knowledge and belief that the report is true, complete, and accurate, and the expeditures, disbursements and cash receipts are for the purposes and objectives set forth in the terms and conditions of the Federal award. I am aware that any false, fictious, or fraudulent information or the omission of any material fact, may subject me to criminal, civil or administrative penalties for fraud, false statements, false claimes or otherwise. (U.S. Code Title 18, Section 1001 and Title 31, Sections 3729-3730 and 3801-3812)."

Luis Mancilla
Manager
Sponsored Programs Financial Operations

For billing questions contact: Nate Robinson
nr657@georgetown.edu