Testimony of Beth Simone Noveck

Professor and Director, The Governance Lab, New York University

U.S. House Committee on Financial Services

Subcommittee on Oversight and Investigations

Hearing – "Fake It till They Make It: How Bad Actors Use Astroturfing to Manipulate Regulators, Disenfranchise Consumers, and Subvert the Rulemaking Process"

February 6, 2020

Introduction

Chairwoman Waters, Ranking Member McHenry, thank you for the opportunity to participate in today's House Financial Services Committee Oversight and Investigations Subcommittee hearing: "Fake It till They Make It: How Bad Actors Use Astroturfing to Manipulate Regulators, Disenfranchise Consumers, and Subvert the Rulemaking Process."

I am a Professor of Technology, Culture and Society at New York University's Tandon School of Engineering, where I direct the Governance Lab, a nonprofit action research center focusing on the use of new technology to improve governance and strengthen democracy. At the Governance Lab, I direct our work on "CrowdLaw," where we collaborate with public sector partners to study and design use of new technology to improve the quality of law and policymaking. I previously served as Deputy Chief Technology Officer and Director of the Open Government Initiative under President Obama, where I led White House policy and projects on citizen engagement. I currently also serve as Chief Innovation Officer of the State of New Jersey and as a Member of Chancellor Angela Merkel's Digital Council.

In this submission, which reflects only my personal opinions, I set out the crucial importance of citizen and stakeholder engagement to increase the effectiveness and legitimacy of regulations, and to strengthen democracy and trust in policymakers when both are under severe challenge. I examine some difficulties attached to public commenting in rulemaking, and how they can be overcome using new tools and

technology. Finally, I showcase how jurisdictions around the world are turning to crowdlaw, the use of online public engagement in order to improve the *quality* of the law- and rulemaking process, and provide examples that the United States could draw on as it seeks to deepen the foundations of its democracy in uncertain times.

Using New Technology to Improve the Quality of Public Participation

The Administrative Procedure Act of 1946 provides the public with an opportunity to participate in the rulemaking process through the submission of data, views, or arguments, which a federal agency is then required to consider prior to promulgation.[1] The right to public participation is not intended to elicit popular opinion about the draft rule or to have people vote on the proposal. It is not an occasion for what constitutional law scholar Alexander Meiklejohn (1872-1964) described as "unregulated talkativeness." Instead, it is an important opportunity for the public to participate in politics, when "everything worth saying shall be said." In other words, the goal of public participation in rulemaking is to apprise the relevant agency of the best available information, in order to inform how it crafts the rule. As the Senate Permanent Subcommittee on Investigations found in its 2019 report on Abuses of the Rulemaking process, "agencies depend on relevant, substantive information from a wide variety of parties to assist them in developing and updating federal regulations."[2] Furthermore, the regulations.gov website states, "public participation is an essential function of good governance. Participation enhances the quality of law and its realization through regulations (e.g. Rules)."[3]

High quality participation in rulemaking is also vital for Congress in its oversight capacity. Although agencies often promulgate rules without significant oversight,[4] Congress still retains and uses its lawmaking authority after it delegates responsibility for implementing laws to regulatory agencies. In addition to oversight hearings, members frequently communicate with agencies during the rulemaking process through meetings,

---

[1] 5 U.S.C. §§ 553, Pub. L. No. 79-404, 60 Stat. 237 (1946).

[2] United States Senate Permanent Subcommittee on Investigations, "Abuses of the Federal Notice-and-Comment Rulemaking Process," *U.S. Senate Committee on Homeland Security & Governmental Affairs*, October 24, 2019, https://www.hsgac.senate.gov/imo/media/doc/2019-10-24 PSI Staff Report - Abuses of the Federal Notice-and-Comment Rulemaking Process.pdf?mod=article_inline.

[3] "Frequently Asked Questions." Regulations.gov Beta. General Services Administration, accessed January 28, 2020, https://beta.regulations.gov/faq.

[4] Sidney A. Shapiro, "Political Oversight and the Deterioration of Regulatory Policy." *Admin. L. Rev.* 46 (1994): 10 (oversight of rulemaking as "limited, infrequent, and ad hoc rather than systematic." Even as oversight has become more popular, "monitoring and reporting only reveals what an agency is doing; these activities do not automatically cause the agency to adhere to, or alter, a policy").

letters and calls.[5] For example, this Committee's Democratic members wrote to the Comptroller of the Currency to ensure that the upcoming Community Reinvestment Act regulatory processes would include meaningful engagement with the public and have suggested extending the period for public commenting from 60 to 120 days to facilitate more diverse participation.[6] Moreover, since the enactment of the Congressional Review Act, Congress has possessed and used its sweeping powers to review and overturn rules and policies within sixty days of submission to Congress (the 60 days *from submission* technicality is enabling Congress to overturn many rules and policies that have long been in effect).[7]

Thus, the process of public commenting provides a vital opportunity for agencies and Congress to obtain important and relevant information from diverse audiences that will help them to understand whether and how a regulation fulfills its legislative purpose.

However, new technology has also created challenges to successful public participation in rulemaking.[8] The shift from a predominantly paper-based to a digital process has made commenting easier but it has also inadvertently opened the floodgates to voluminous, duplicative and even "fake" comments – what I call notice-and-spam – thereby lessening the value of public participation.

As I predicted in an article in the *Emory Law Review* in 2004, shortly after the launch of regulations.gov: "Automating the comment process might make it easier for interest groups to participate by using bots—small software 'robots'—to generate instantly thousands of responses from stored membership lists. Moving from long standing agency traditions to a rationalized online system levels the playing field and lowers the bar to engagement. Suddenly, anyone (or anything) can participate from anywhere. And that is precisely the potential problem. Increased network effects may not improve the

---

[5] Melinda N. Ritchie, "Back-Channel Representation: A Study of the Strategic Communication of Senators with the US Department of Labor," *The Journal of Politics* 80, no. 1 (2018): 240-253. See also Rachel Augustine Potter, "Member Comments: The Other (Less Visible) Way Congress Oversees Rulemaking," *The Brookings Institution*, November 21, 2019, https://www.brookings.edu/research/member-comments-the-other-less-visible-way-congress-oversees-rulemaking/ and Nicole Kalaf-Hughes, Jason A. MacDonald, and Russell W. Mills, "Agency policy preferences, congressional letter-marking and the allocation of distributive policy benefits." *Journal of Public Policy* 36, no. 4 (2016): 547-571.
[6] "Waters Opens Investigation." *U.S. House Committee on Financial Services*, January 15, 2020, https://financialservices.house.gov/news/documentsingle.aspx?DocumentID=406070#1.
[7] Dylan Scott, "The New Republican Plan to Deregulate America, Explained," *Vox*, April 25, 2018, https://www.vox.com/policy-and-politics/2018/4/25/17275566/congressional-review-act-what-regulations-has-trump-cut.
[8] For comprehensive information about E-Rulemaking, see University of Pennsylvania Law School's E-Rulemaking.org program: https://www.law.upenn.edu/institutes/ppr/erulemaking/papers-reports.php.

legitimacy of public participation. For without the concomitant processes to coordinate participation, quality input will be lost; malicious, irrelevant material will rise to the surface, and information will not reach those who need it. In short, e-rulemaking will frustrate the goals of citizen participation."[9]

Although much current attention is focused on the problem of fake comments and astroturfing, where an interest group hides its identity and manufactures the appearance that comments come from the "ordinary public," the more salient and urgent concern for regulators and overseers, is not who signed the comment – if anyone – but the failure to invite and then to use high quality and diverse participation to inform the rulemaking process.

There is a remedy. In the almost two decades since participation moved online, data science tools and methods have evolved to deal with the problems of voluminous, duplicative and fake comments. Yet neither agencies nor the regulations.gov administrator are using them in a substantial way. The more agencies are deluged by voluminous, duplicative and fake "astroturf" comments, the more this race to the bottom reinforces a disturbing disregard for the potential value of public participation. We are failing to recognize the value of public commenting to enhance the quality of rules and, therefore, we have chosen not to solve the real problem at issue, which is not astroturfing, but taking the value of public commenting seriously. Failure to address the real challenge will only set us back further from the growing number of advanced nations that use new technology to tap the collective experience and expertise of their citizens.

I argue that the Committee should direct the agencies it oversees to use – and itself use – easily available tools to:

1) **Mine and summarize relevant comments for information**. As we shall explore, machine learning and natural language processing software, namely those subfields of artificial intelligence used for making sense of large quantities of text, have created unprecedented ways to manage information— to sort the informational wheat from the extraneous content chaff. These technologies could enable agencies to process and analyze public comments rapidly and effectively.

2) **Adopt complementary mechanisms for public commenting in addition to notice and comment.** The technologies of collective intelligence that enable people to

---

[9] Beth Simone Noveck, "The electronic revolution in rulemaking," *Emory LJ* 53 (2004): 433, https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/emlj53&section=20.

communicate and collaborate via the Internet, have led to new ways of soliciting information that are a substantial improvement on the traditional, open-ended submission process of notice-and-comment. Around the world, regulatory agencies and the legislative committees that oversee them are turning to "crowdlaw," namely the use of the Internet to create a meaningful and deliberative two-way conversation with the public, yielding more relevant, timely and diverse information. I explain how we could—how we must—adopt these practices in the United States and reimagine how agencies engage with citizens and stakeholders.

For additional information on the platforms and processes described herein, please see "CrowdLaw for Congress: Strategies for 21st Century Lawmaking," a report and short video training materials I authored, available at congress.crowd.law.

The GovLab's CrowdLaw for Congress website with cases and examples of how parliaments around the world are using technology to engage with citizens and stakeholders. Available online at congress.crowd.law

Non-Endorsement: *The technologies referenced in this document are discussed as examples of platforms supporting public participation practices in lawmaking in legislatures around the world. Their mention does not constitute an endorsement of the companies behind these technologies. I derive no financial benefit from these firms.*

Summary of Recommendations

In order to address the challenge of voluminous, duplicative and fake comments:

1. Agencies should use machine learning to *summarize voluminous comments.*

2. Agencies should use deduplication software to *remove identical comments*.

3. Agencies should use filtering software to *sift out the real and the relevant.*

4. Lawmakers and agencies *should use complementary crowdlaw platforms and processes* used by other governments and organizations to enable better citizen and stakeholder engagement.

5. Like Brazil and New Jersey, agencies and committees should use *Wiki Surveys* to reduce volume and duplication.

6. Agencies and committees should use *Collaborative Drafting and Annotation* as Germany did to engage more experts in review of rules.

7. Committees should set up UK-style *Evidence Checks* to crowdsource review of comments and evidence.

8. Committees should democratize oversight and pilot the use of *Citizen Juries* as they do in Belgium.

Why Improving Public Commenting is Urgent: Declining Trust in the Effectiveness and Legitimacy of Government

The need to improve how we make rules has never been more pressing. For democracy to thrive, it has to work. Yet there is widespread public perception that the government's capacity to tackle the problems of our age is declining. Cynicism is up and rates of trust are down.[10] In particular, trust in Congress is at an all-time low. In a 2019 Gallup poll just 4 percent of Americans trusted Congress a great deal.[11] The executive branch does not

---

[10] Morris Fiorina, "Parties, Participation, and Representation in America: Old Theories Face New Realities," *Annual Meeting of the American Political Science Association*, August 2000, https://web.stanford.edu/~mfiorina/Fiorina%20Web%20Files/Fiorina%20SOD.pdf.
[11] Megan Brenan, "Americans' Trust in Government to Handle Problems at New Low," *Gallup* 31 (2019), https://news.gallup.com/poll/246371/americans-trust-government-handle-problems-new-low.aspx.

fare much better. Across the board, only 17 percent of Americans today say they can trust the government in Washington to do what is right "just about always" (3 percent) or most of the time" (14 percent).[12] Globally, the 2020 Edelman Trust Barometer finds that 66 percent of people do not trust their country's leaders to address the country's challenges.[13]

In part, the decline in trust stems from a crisis of effectiveness. Voters typically see their government as a "chronically clumsy, ineffectual, bloated giant."[14] Both Republicans and Democrats hold this view.[15] In *Why Government Fails So Often*, Yale Law professor Peter Schuck concludes that voters rate the government so poorly because it performs poorly. Similarly, political scientist Paul Light asserts that federal failures have become de rigueur. He writes: "the question is no longer if the government will fail every few months, but where. And the answer is 'anywhere at all.'"[16]

There is also a related crisis of legitimacy. The public feels disenfranchised. One study concludes that "the preferences of the average American appear to have only a minuscule, near-zero, statistically non-significant impact upon public policy."[17] This "implosion of trust" is compounded by a widening legitimacy gap – the sense that those who govern do not speak for us. Law professors Bryan Caplan and Ilya Somin (following in the tradition of Anthony Downs in his 1957 classic, *Economic Theory of Democracy*) see voting, that basic form of democratic participation, as irrational and irrelevant.[18]

With legislation and regulation often developed by a small group of elected or appointed officials working behind closed doors, often with the aid of lobbyists, it is no wonder that rates of trust are at historic lows.

---

[12] "Public Trust in Government: 1958-2019," *Pew Research Center*, April 11, 2019, https://www.people-press.org/2019/04/11/public-trust-in-government-1958-2019/.
[13] "Edelman Trust Barometer 2020 Global Report," *Edelman*, January 19, 2020, https://www.edelman.com/sites/g/files/aatuss191/files/2020-01/2020%20Edelman%20Trust%20Barometer%20Global%20Report_LIVE.pdf.
[14] Peter H Schuck, *Why government fails so often: And how it can do better* (Princeton University Press, 2014): 3.
[15] John Bridgeland and Peter Orszag, "Can government play moneyball? How a new era of fiscal scarcity could make Washington work better?" *The Atlantic*, July/August, 2013, https://www.theatlantic.com/magazine/archive/2013/07/can-government-play-moneyball/309389/.
[16] Paul C. Light, "A cascade of failures: Why government fails, and how to stop it." *Center for Effective Public Management. Washington, DC: Brookings Institution* (2014), https://www.brookings.edu/research/a-cascade-of-failures-why-government-fails-and-how-to-stop-it/.
[17] Martin Gilens and Benjamin I. Page, "Testing theories of American politics: Elites, interest groups, and average citizens," *Perspectives on politics* 12, no. 3 (2014): 564-581.
[18] Anthony Downs, *An Economic Theory of Democracy* (University of Chicago Press, 1957): 260-276; See also Guido Pincione and Fernando R. Tesón, *Rational choice and democratic deliberation: A theory of discourse failure* (Cambridge University Press, 2006).

That is one reason why the rulemaking process is so important. It provides an opportunity to implement Congress' broad-scale policy prescriptions and to involve the American public in doing so. It creates an opportunity to reverse the lack of engagement and improve rates of trust, as well as to strengthen later compliance by giving everyone equal chance to be part of the process. To overcome the twin crises of legitimacy and of effectiveness, it is increasingly urgent to create meaningful ways for people to participate.

The Purpose of Public Participation in Rulemaking

Section 553 of the 1946 Administrative Procedure Act enshrines the public's right to participate by codifying a longstanding practice of soliciting participation in rulemaking. Agencies must ensure that the right is real, not just theoretical, by giving the public ample opportunity to review and comment. Therefore, they must give notice of the rule and, under the 1993 Executive Order 12,866 (reiterated in 2011 in Exec. Order 13,563), keep the draft open for comment for no less than sixty days, after which agencies must respond to significant comments.[19] The D.C. Circuit has held that "there must be an exchange of views, information, and criticism between interested persons and the agency," allowing for a deliberative and two-way conversation between the public and the agency.[20]

Thousands of rules are enacted every year.[21] They touch every aspect of our lives. The purpose of participation is to advance both the legitimacy and the quality of these rules. Participation allows agencies to obtain information that will enable them both to improve rules and to anticipate their likely impact. This input—bringing to bear the collective intelligence of a wider audience of stakeholders, interest groups, businesses, nonprofits, academics and interested individuals—infuses the process with information that comes from participants' professional and lived experience. Committees also need this information in order to provide effective oversight of the executive, one of their core legislative responsibilities. Research shows that committees do not access enough information from diverse enough sources.[22] This impedes their capacity to conduct

---

[19] *Perez v. Mortg. Bankers Ass'n*, 135 S. Ct. 1199 (2015) Also Exec. Order 13,563, 76 Fed. Reg. 3821 (Jan. 18, 2011).

[20] Home Box Off. v. F.C.C., 567 F.2d 9, 35 (D.C. Cir. 1977).

[21] Maeve P Carey, "Counting regulations: An overview of rulemaking, types of federal regulations, and pages in the federal register," *Congressional Research Service* (Updated September 3, 2019), https://fas.org/sgp/crs/misc/R43056.pdf.

[22] Jonathan Lewallen et al., "Congressional Dysfunction: An Information Processing Perspective", *Regulation and Governance,* 10, no. 2 (2016): 179-190.

oversight. Access to more information from a greater diversity of perspectives will bolster committee oversight of the executive.

Participation is also designed to ensure legitimacy. People who get the chance to participate in a deliberative rulemaking process, especially one in which the agency responds to and addresses their concerns, will be more willing to comply with the rule. Participation also provides a public check on the rulemaking process, helping to ensure public oversight and scrutiny. It can also help to facilitate congressional oversight and judicial review by allowing for comments that, as Penn law professor Cary Coglianese writes: "assess whether agency decisions are in fact being made on the grounds asserted for them and not on other, potentially improper, grounds."[23]

The Challenges of Online Participation: High Quantity, Low Quality

While it represents an improvement on the paper-based processes of yesteryear, online participation in notice-and-comment rulemaking falls far short of the goal of effectively informing the regulatory process. The limitations can be grouped into three categories: 1) voluminous comments, 2) duplicative comments, and 3) fake comments. I examine each in turn.

Voluminous Comments

Although many of the 3-4,000 rulemakings agencies publish receive only a handful of comments, some receive voluminous responses, thanks to the ease of digital commenting. In 2017, when the Federal Communications Commission sought to repeal an earlier Obama-era rule requiring Internet Service Providers to observe Net Neutrality by transmitting content at the same speeds and not discriminating in favor of one content provider, the agency received 22 million comments in response.[24] According to the General Services Administration, which administers regulations.gov, the Social Security Administration's proposed rule on disability reviews attracted 91,720 comments between November 2019 and January 2020. A Justice Department proposal to expand the collection of DNA samples from people put under arrest and from immigrant detainees garnered over 24,000 comments over a three week period in October 2019.[25] The Fish

---

[23] See Cary Coglianese, Heather Kilmartin, and Evan Mendelson, "Transparency and public participation in the federal rulemaking process: Recommendations for the new administration," *Geo. Wash. L. Rev.* 77 (2008): 924, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1292911.

[24] Restoring Internet Freedom (82 Fed. Reg. 25,568 (June 2, 2017) and (83 Fed. Reg. 7,852 (Feb, 22, 2018)).

[25] "Site data," *Regulations.gov*. United States Government, https://www.regulations.gov/siteData.

and Wildlife Services received more than 640,000 e-mail comments on whether to list the polar bear as a threatened species in 2007.[26] When the National Parks Service proposed restricting snowmobile access, it received 360,000 comments. It is good, in principle, that the public comments on rules, but large volumes of both electronic and mailed-in comments make it hard for agencies to read and understand the material (and can present other problems that I examine in the next section).

Without the right tools and methods, the volume of comments makes it impossible for agencies to process comments, and renders public participation all but useless, frustrating the needs of regulators, overseers and the public.

Duplicative Comments

Moreover, comments are often both voluminous and duplicative. Interest groups have learned the tactic of getting people to submit identical or nearly identical comments -- so-called postcard comments -- by mail or electronically. One research study looked at 1,000 e-mails sent via the grassroots activism site MoveOn.org to the Environmental Protection Agency (EPA) in connection with its 2004 mercury rulemaking. The study concluded that "only a tiny portion of these public comments constitute potentially relevant new information for the EPA to consider. The vast majority of MoveOn comments are either exact duplicates of a two sentence form letter, or they are variants of a small number of broad claims about the inadequacy of the proposed rule."[27] The Pew Research Center found that only 6 percent of online comments in the 2017 Net Neutrality rulemaking used unique text. In fact, Pew found that the top five comments were each repeated over 800,000 times.[28]

Interest groups of every political stripe encourage this kind of clicktivism, in which people click a button to submit duplicative postcard comments. Political science

---

[26] The U.S. Fishing and Wildlife Service, "News Release: Polar Bear Range States Meet to Exchange Information." https://www.fws.gov/news/ShowNews.cfm?newsId=64E14600-FE16-C275-509DA6F8C494B903.

[27] Stuart W Shulman, "The case against mass emails: Perverse incentives and low quality public participation in US federal rulemaking," *Policy & Internet* 1, no. 1 (2009): 23-53, https://onlinelibrary.wiley.com/doi/abs/10.2202/1944-2866.1010?casa_token=kL6DVaO-CyQAAAAA:syPJG_lztTwFmGqwef9fIWfs9fnXZU8ZpZAbMA62KJUZkuxPymuodcPHOivbUkrDhKJ1F8Kx6xTO7LY.

[28] Jack Karsten and Darrell West, "Net Neutrality Debate Exposes Weaknesses of Public Comment System," *The Brookings Institution TechTank,* January 18 2018, https://www.brookings.edu/blog/techtank/2018/01/18/net-neutrality-debate-exposes-weaknesses-of-public-comment-system/.

professor Eitan Hirsch calls this process "political hobbyism;"[29] Digital technology is making it easier for some groups to engage in notice-and-spam clicktivism, to sign electronic petitions, to forward political messages on Facebook or to shout into the wind on Twitter.

In the main, these groups are not seeking to enhance the level of knowledge in the rulemaking. They do not place much stock in the ability of their members and would-be members to contribute substantively to the discussion. Even though the public commenting process is meant to air new ideas and identify novel issues, interest groups use it simply as an opportunity to recruit new members and solicit personal information for subsequent donation solicitations. Interest groups use the forum to signal popular approval or disapproval. These campaigns are intended to sway the outcome, and agencies should never base their decisions on the number of campaigners asserting for or against, but on the substance of the comments.[30]

With large numbers of duplicative comments, and without the right tools, agencies are not obtaining all ideas worth hearing because they cannot extract information of value. Moreover, the identical comments, even if genuine, do not add meaningfully to the discussion or even convey the sense that the individual feels strongly about the point of view. This further undermines respect for and belief in the value of public commenting.

Without the ability to visualize and understand the comments as a whole -- to see how many are identical and what the unique content is across the dataset of comments -- it also becomes too easy for regulators and others to cherry-pick those comments that support their point of view and claim that it reflects the public's submissions.

Fake Comments

Finally, there is the problem of astroturfing and fake comments. A 2017 Wall Street Journal investigation found that 41 percent of comments they investigated on several federal agency websites were from "fake people." Comments had been signed by people

---

[29] Eitan D. Hersh, "The Problem With Participatory Democracy Is the Participants," *The New York Times*, June 29, 2017, https://www.nytimes.com/2017/06/29/opinion/sunday/the-problem-with-participatory-democracy-is-the-participants.html.

[30] Stuart W Shulman, "The case against mass e-mails: Perverse incentives and low quality public participation in US federal rulemaking," *Policy & Internet* 1, no. 1 (2009): 23-53, https://onlinelibrary.wiley.com/doi/abs/10.2202/1944-2866.1010?casa_token=kL6DVaO-CyQAAAAA:syPJG_lztTwFmGqwef9fIWfs9fnXZU8ZpZAbMA62KJUZkuxPymuodcPHOivbUkrDhKJ1F8Kx6xTO7LY.

who either did not exist or had died, or whose names were used without their knowledge.[31] Nearly eight million of the FCC Net Neutrality comments came from email addresses associated with fakemailgenerator.com.[32] The ability to comment online has made it easy to submit fabricated comments that skew the informational inputs and further emasculate the public commenting process.[33]

Software is making it easier to auto-generate, duplicate and submit such comments. While some agencies have required fields for "name" and "address" in their rulemakings, any string of characters typed in the box is sufficient.[34] Although anonymous commenting is legal, interested parties are still choosing to falsify names, perhaps in an effort to lend more weight to comments and knowing that agencies lack the tools and the ability to investigate and prosecute perpetrators of identity fraud.

For example, the Department of Housing and Urban Development (HUD) reported to the Senate Subcommittee that:

> The Department has no way of determining whether a commenter has filed a comment under someone else's identity . . . HUD has received comments from commenters that identify themselves as "Mickey Mouse," "Donald Duck," and "John Q. Public." These comments have not been so numerous as to adversely affect the Department's efforts to review and summarize public comments. Generally, these comments are not substantive and are given appropriate weight.[35]

---

[31] James V. Grimaldi and Paul Overberg, "Millions of People Post Comments on Federal Regulations. Many Are Fake." *Wall Street Journal*, Updated December 12, 2017, https://www.wsj.com/articles/millions-of-people-post-comments-on-federal-regulations-many-are-fake-1513099188.

[32] In the Matters of Nicholas Confessore and Jeremy Singer-Vine, FOIA Control No. 2017-764; FOIA Control No. 2018-204, 33 FCC Rcd. 11808 (18) (Dec. 3, 2018).

[33] Jack Karsten and Darrell M. West, "Net Neutrality Debate Exposes Weaknesses of Public Comment System," *The Brookings Institution*, January 18, 2018, https://www.brookings.edu/blog/techtank/2018/01/18/net-neutrality-debate-exposes-weaknesses-of-public-comment-system/.

[34] United States Senate Permanent Subcommittee on Investigations. "Abuses of the Federal Notice-and-Comment Rulemaking Process." U.*S. Senate Committee on Homeland Security & Governmental Affairs* (October 24, 2019): 16, https://www.hsgac.senate.gov/imo/media/doc/2019-10-24 PSI Staff Report - Abuses of the Federal Notice-and-Comment Rulemaking Process.pdf?mod=article_inline.

[35] United States Senate Permanent Subcommittee on Investigations. "Abuses of the Federal Notice-and-Comment Rulemaking Process." U.*S. Senate Committee on Homeland Security & Governmental Affairs* (October 24, 2019): p. 17 citing Letter from Len Wolfson, Ass't Sec'y for Congressional & Intergovernmental Relations, U.S. Dep't of Housing & Urban Dev., to the Hon. Rob Portman, Chairman, U.S. Senate Permanent Subcomm. on Investigations (Mar. 5, 2018).

As the Chairwoman of this Committee, the Honorable Representative Maxine Waters, has written: such fraudulent comments "undermine legitimate debate on proposed rules by creating the false appearance that a position has widespread, grassroots support. Such misrepresentations have been increasing in frequency and complexity in recent years."[36]

Tackling the Problems: **Use Data Analytical Tools to Make Sense of Comments**

Companies, governments and researchers are keen to make sense of the increasing volume of information people post online. The good news is that the tools and methods already exist to be able to address the problem of voluminous, duplicative and fake comments, while preserving the right to comment anonymously.

Society is awash in information. IBM is fond of saying that 90 percent of the world's data has been created in the last two years alone. There are countless projects—both academic and commercial—to help us make sense of such overload. Whether there are too many comments, too many of the same comments, or fake comments, agencies need to extract meaning to understand the substance of what is submitted. The good news is that "there's an app for that," as the saying goes. The Center for Democracy and Technology writes: "Automated content filtering is not new. Many tools have been developed over the years to identify and filter content, including keyword filters, spam detection tools, and hash matching algorithms."[37]

Thankfully, with the proliferation of big data, technologists have evolved the means to make sense of large quantities of information, much of it far more complex than the text-based comments submitted to regulations.gov. To address the challenges discussed above, agencies need three kinds of tools.

First, they need help to make sense of large quantities of relevant comments. A large volume of messages can be hard to parse in a short time frame, especially when they are thoughtful, helpful and on topic. I examine each of these tools in turn.

Second, they need to be able to de-duplicate identical comments.

---

[36] "Waters Opens Investigation," *U.S. House Committee on Financial Services*, January 15, 2020, https://financialservices.house.gov/news/documentsingle.aspx?DocumentID=406070#1.

[37] Natasha Duarte, Emma Llanso, and Anna Loup, "Mixed Messages? The Limits of Automated Social Media Content Analysis," In *FAT*, p. 106. 2018, https://cdt.org/wp-content/uploads/2017/11/2017-11-13-Mixed-Messages-Paper.pdf.

Third, they need to separate the real from the fake and, as we shall discuss, separate out relevant from irrelevant comments.

Solution 1: Use Machine Learning to Summarize Voluminous Comments

While still a challenging task, researchers have developed tools for extracting meaning and summarizing text. The processes often combine automation with human, collective intelligence to make quick work of large data stores. Journalists took advantage of such tools, for example, when they needed to rapidly sift through the 13.4 million documents that comprised the Paradise Papers.[38] In short, researchers have cracked problems far more challenging than making sense of rulemaking data. So far rulemakers, legislators and agencies have paid little attention to them.

That should change, especially since both **Google** and **Microsoft** announced in 2019 that they had built systems that could summarize articles spanning news, science, stories, instructions, emails, patents, and even now legislative bills.[39] [40]
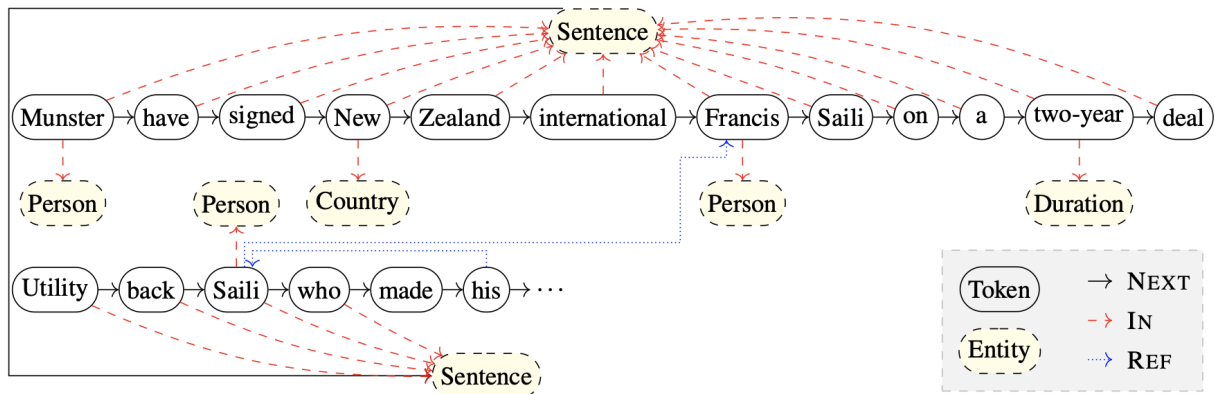
A graphic explaining how Microsoft's AI breaks down a sentence and interprets its various parts in order to develop a summary. Source: see Patrick Fernandes et al, "Structured Neural Summarization," https://arxiv.org/pdf/1811.01824.pdf

---

[38] Fabiola Torres López, "How They Did It: Methods and Tools Used to Investigate the Paradise Papers," Global Investigative Journalism Network (Dec. 4, 2017).
[39] Patrick Fernandes, Miltiadis Allamanis, Marc Brockschmidt, "Structured Neural Summarization", International Conference on Learning Representations (ICLR 2019), February 2019, https://arxiv.org/pdf/1811.01824.pdf
[40] Google AI Blog, "Text Summarization using TensorFlow", August 2016, https://ai.googleblog.com/2016/08/text-summarization-with-tensorflow.html.

Indian news aggregation app **InShorts** recently debuted its AI-based news summarization feature. The app creates 60-word summaries of full-length news articles using natural language processing techniques.[41]

Similar techniques have become commonplace for images. Both Android and iOS operating systems use machine learning to "summarize photos" -- that is to identify objects present in a photograph. This enables people to search for photos that contain dogs or cars, for example.

Professor Deb Roy directs the "Social Machines" Lab at MIT, which does research on large-scale Twitter data sets. Its **Electome** project extracts meaning from the entire corpus of Twitter data—billions of tweets—in order to summarize the core political messages of the day and help drive election coverage.[42]
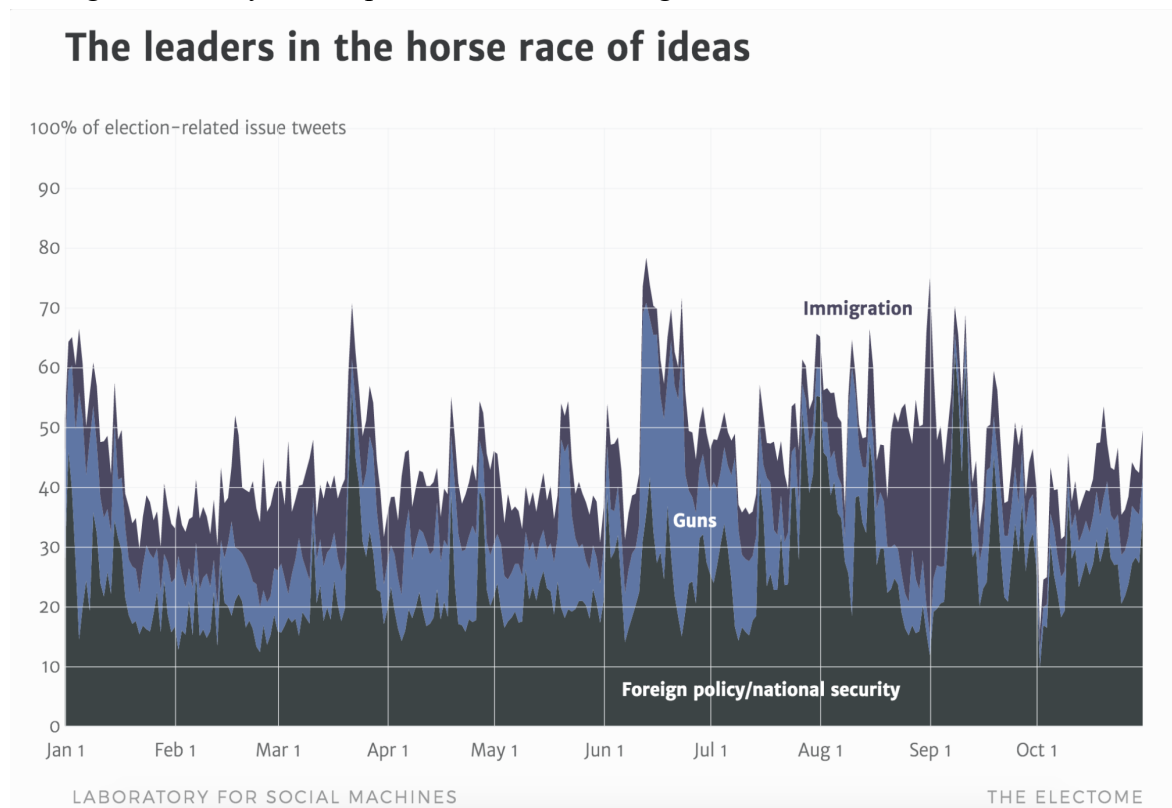
An example of a data visualization created by the MIT Electome project: This graphic shows foreign policy and national security issues were the dominant topics on Twitter between January and November 2016.

---

[41] Economic Times Tech, "Inshorts debuts AI-based news summarization on its app", June 2018, https://tech.economictimes.indiatimes.com/news/startups/inshorts-debuts-ai-based-news-summarization-on-its-app/64531038.
[42] The Electome, http://www.electome.org/.

Public institutions are also using data analytical techniques to make sense of social media data for public good. To help UNICEF and other actors craft more effective pro-immunization messaging programs, researchers set out to monitor social media networks, including blogging platforms, forums, Facebook, Twitter, Tumblr, and YouTube. They sought to analyze prevalent conversation themes according to volume, types of engagement, and demographic; to identify influencers across languages and platforms; and to develop specific recommendations for improving messaging strategies across languages, platforms and conversation themes. The research methodology involved scraping text conversations from social media platforms in English, Russian, Polish and Romanian, in order to be able to identify key patterns.[43]

In 2019, in order to make inputs more usable for civil servants, **CitizenLab**, a Belgian software company that designs software for citizen engagement in use in twenty countries, incorporated natural language processing (NLP) and machine learning to categorize and cluster the text submitted by citizens. As Wietse van Ransbeeck, the CEO, writes: "Analyzing the high volumes of citizen input collected on these platforms is extremely time-consuming and requires skills that administrations often do not have, which prevents governments from uncovering valuable learnings. Setting up a digital participation platform is therefore not enough: it is also necessary to make data analysis more accessible so that civil servants can tap into collective intelligence and make better-informed decisions."

CitizenLab's algorithms identify the main topics and group similar ideas together using an approach known as topic modeling. It works by grouping content that shares similar words, both in meaning and in form, i.e. the words "trees" and "forest" are similar in meaning, therefore two ideas with these words are more likely to be grouped together. With regard to word form, for example, "bicycle" and "cyclist" are also considered similar.

Such clustering, according to van Ransbeeck, happens in real-time and takes between 5-15 seconds. This makes it easier for those running the consultation to see what the comments are about and understand priorities. If organizers require people to login then the comments can also be sorted by demographic groups and location, making it possible to cluster topics, for example, by location as well. For example, an engagement on youth

---

[43] Stefaan G. Verhulst and Andrew Young, "The Potential of Social Media Intelligence to Improve People's Lives," *The GovLab, Online:* http://www. thegovlab. org/static/files/publications/social-media-data (2017).

climate action in 2019 elicited 1700 contributions, which CitizenLab grouped into 15 concrete proposals. This helps decision makers to make sense of the content gathered through citizen participation and better understand the priorities and ideas of the public.
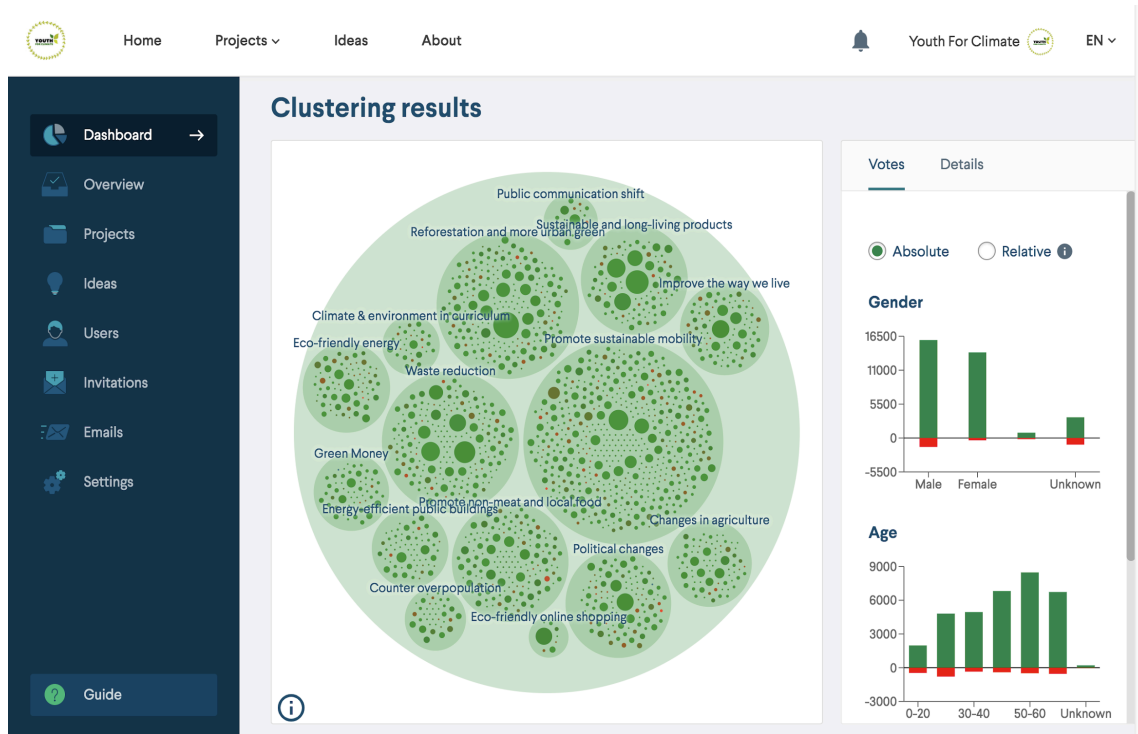
A screenshot of summarized and clustered public comments from a CitizenLab project on climate change policy

In addition, **Remesh** is an American engagement platform that also uses artificial intelligence to enable clustering of topics but for a real-time conversation, rather than asynchronous submissions. Remesh specializes in real-time online discussions with large numbers of participants, usually a thousand or more.

A recent **State Department project** offers a simple illustration for how agencies could take a more effective approach to making sense of rulemaking comments using a combination of artificial intelligence (AI) from machines and collective intelligence (CI) from humans. In 2016, the State Department sought to improve its passport application and renewal process in anticipation of an increase in the number of passport application and renewal forms. After consulting with the General Services Administration (GSA) and USA.gov, it ran an online public engagement process to ask people what improvements they wanted. It received almost 1,000 comments. In order to make rapid sense of those submissions, it engaged **Insights.us**, an Israeli American third-party software company

that helped the agency use two methods to whittle down answers to their most essential parts.

First, commenters were asked to highlight the 200 most important characters containing the key points of their answers. For users who declined to do so, the platform encouraged other users to highlight what they felt to be the other users' core ideas. Then the company applied a text-mining algorithm that scanned the highlighted text for responses containing similar keywords in order to create summaries, or what the company calls "highlights." The public was invited to proof and make suggestions for how to improve those highlights, adding accountability but in a way that is efficient.

The combination of human and machine intelligence made it faster and easier to summarize content than using an algorithm alone. Finally, the Insights.US team grouped the AI and CI summaries into nine insights. Inevitably, most individuals wanted a "much easier and more convenient" online process. Others wanted the Department to use simpler language on forms and web pages, make physical passport application facilities easier to access, and provide on-demand user support through an online web chat or other system. Indeed, research has also identified increased accessibility as a key way to improve participatory rulemaking.[44]After the Department of State reviewed these insights, it made them available on the tell-us.usa.gov site. Although the cost of this engagement is not public, Insights.us says services for cities and government agencies generally cost between USD 18,000 and USD 36,000.

Solution 2: Use Deduplication Software to Address Identical Comments

The large volume of submissions often results from duplicative "postcard" comments. Data mining technology to deduplicate public comments has existed for well over a decade. Deduplicating records in a database is a common process. **Dedupe.io** is an example of a software service which helps researchers and data scientists to identify and delete similar records in their database. Once similar entries are identified, duplicate records can be deleted if they are an exact match or consolidate similar entries.[45] Jeff Jonas, a world renowned data scientist, has spent his career developing tools like **Senzing**, an "entity resolution system" -- software that detects duplicates in large scale databases like voter registration systems. His work for the Electronic Registration Information Center has made it possible to identify 26 million people who are eligible but

[44] Cary Coglianese,"Enhancing Public Access to Online Rulemaking Information", *Michigan Journal of Environment & Administrative Law* 2 (2012): 1.
[45] "Dedupe - how it works." *Dedupe.io*, https://dedupe.io/documentation/how-it-works.html.

unregistered to vote, as well as 10 million registered voters who have moved, appear on more than one list or have died.[46]

	A	C	D	E	F
1	cluster_id	site_name	address	zip	phone
2	00448ef8-970d-4afc-aac4-799bff92edb8	Ada S. McKinley Community Services	725 S Wells Street	60607	3852000
3	005473f8-81b5-4ee6-8c90-2165f76cc14d	Armstrong	2110 W. Greenleaf	60645	5342150
4	005473f8-81b5-4ee6-8c90-2165f76cc14d	Armstrong , G.	2110 W. Greenleaf		5342150
5	013ddb28-e581-4bbf-9821-ade35c19ac0c	Chicago Public Schools Mason, Roswell B.	4216 W. 19th St	60623	5341530
6	013ddb28-e581-4bbf-9821-ade35c19ac0c	Mason	4216 W. 19th St.	60623	5341530
7	013ddb28-e581-4bbf-9821-ade35c19ac0c	CHICAGO PUBLIC SCHOOLS MASON, ROSWELL B.	4216 W 19TH ST	60623	5341530
8	0142b9e7-785c-40b1-8c6f-6f20c1e60dc0	Tarkington	3330 W. 71st St.		5354700
9	0160962a-02d5-4a9f-af2e-72db596190f9	HENRY BOOTH HOUSE ALLISON'S INFANT TODDLER	5522 S RACINE	60636	4363193
10	0160962a-02d5-4a9f-af2e-72db596190f9	Henry Booth House - Allison's Infant Toddler	5522 S Racine Ave		4363193
11	0160962a-02d5-4a9f-af2e-72db596190f9	Henry Booth House Allison's Infant Toddler	5522 S Racine	60636	4363193
12	0160962a-02d5-4a9f-af2e-72db596190f9	Allison's Infant and Toddler Center	5522 S. Racine		4363193
13	01da0507-fc62-4d30-a450-a3802399b9c9	Ashburn School	8300 S. St. Louis	60652	5357860
14	01da0507-fc62-4d30-a450-a3802399b9c9	Ashburn School	8300 S. St Louis Ave.		5357860
15	01edfddc-a559-4e2a-a587-829e2c6c66a5	ERIE NEIGHBORHOOD HOUSE FCCH-TANIA PEREZ	4502 N CHRISTIANA AVE	60625	5041813
16	024a0af2-03f4-4c71-88a7-d734a6902531	Bass	1140 W. 66th St.		5353275
17	024a0af2-03f4-4c71-88a7-d734a6902531	Bass	1140 W. 66th St.	60621	5353275
18	0285fac6-6b0b-4819-a247-223fd16caf54	Catholic Charities-St Mark	1041 N. Campbell	60622	7726606
19	0285fac6-6b0b-4819-a247-223fd16caf54	Catholic Charities Chicago - St. Mark	1041 N Campbell Avenue	60622	7726606
20	02ceb7fb-4237-4090-82a2-55ba1a2cc5bb	Easter Seals Gilchrist Marchman	1001 West Roosevelt Rd.	60608	4927402
21	02ceb7fb-4237-4090-82a2-55ba1a2cc5bb	Easter Seals Society of Metropolitan Chicago - Gilchrist-Marchman	1001 W Roosevelt Rd		9395115
22	02ceb7fb-4237-4090-82a2-55ba1a2cc5bb	Easter Seals Society of Metropolitan Chicago - Gilchrist-Marchman	1001 W Roosevelt Rd		9395115

Results downloaded from Dedupe: After analyzing the database, the tool adds a "cluster_id" to each entry and assigns a single cluster id to entries which are similar or are duplicates.

While Dedupe and Senzing are designed to remove duplicates of names and addresses from databases, American technology firm **Texifer** was designed specifically to work on the problem of regulatory rulemaking comments. The company was a spinout from the research of Dr. Stuart Shulman, who has written many seminal articles about administrative rulemaking. In 2007, he created a text mining tool called DiscoverText through research funded by the National Science Foundation.

DiscoverText helps agencies and rulemaking researchers to quickly deduplicate comments.[47] For example, Shulman deduplicated the public comment dataset from a 2013 school nutrition rulemaking in order to be able to quickly reveal the substance of the comments and analyze them.[48] Shulman deduplicated the polar bear rulemaking in 2007 (660,000 comments) and the national monuments rulemaking in 2017 (3.3 million comments). According to the DiscoverText website, the company has worked with the Department of Homeland Security, Department of the Treasury, Federal Communications

---

[46] Lohr, Steve Lohr,, "Another Use for AI: Finding Millions of Unregistered Voters, *The New York Times* , November 5 2018. https://www.nytimes.com/2018/11/05/technology/unregistered-voter-rolls.html.
[47] Stuart Shulman, "Scholarly Citations of the Coding Analysis Toolkit," *DiscoverText*, March 31, 2018. https://discovertext.com/2018/03/31/scholarly-citations-of-the-coding-analysis-toolkit/.
[48] Lauren M. Dinour and Antoinette Pole, "Potato chips, cookies, and candy oh my! Public commentary on proposed rules regulating competitive foods," *Health Education & Behavior* 44, no. 6 (2017): 867-875, https://journals.sagepub.com/doi/abs/10.1177/1090198117699509.

Commission, National Labor Relations Board, Naval Postgraduate School, National Oceanic and Atmospheric Administration, Office of the Comptroller of the Currency, U.S. Department of Agriculture, U.S. Fish & Wildlife Service, U.S. Small Business Administration and U.S. Department of the Interior.
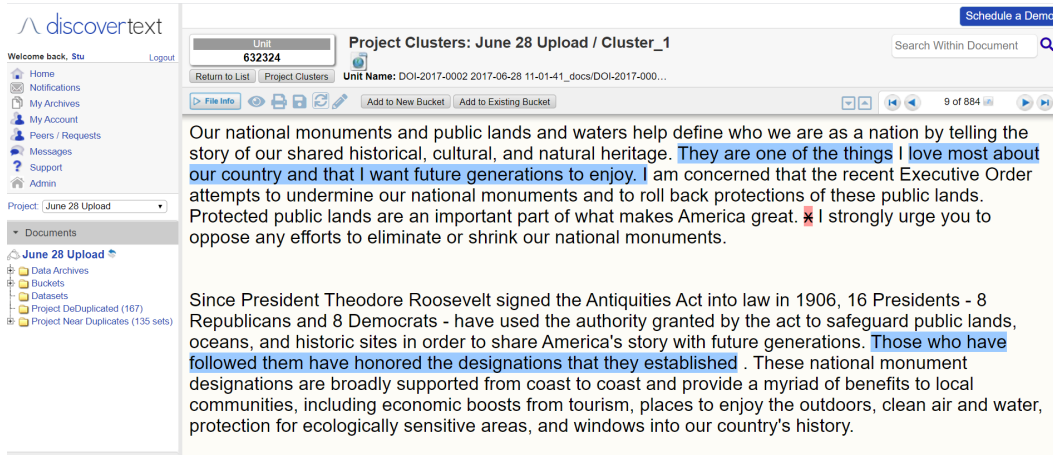

Interface of Discovertext with unique text in a paragraph highlighted

Another way to avoid some of the problems of duplicates is to process a user's submission and alert her if a similar entry already exists since the goal of rulemaking is not to hear every person but to hear every idea. This process happens now with **chatbots** or question-and-answer platforms like Stack Overflow, a popular site for hobbyists and technologists, where people ask and get answers to questions like "how often to water my begonias."
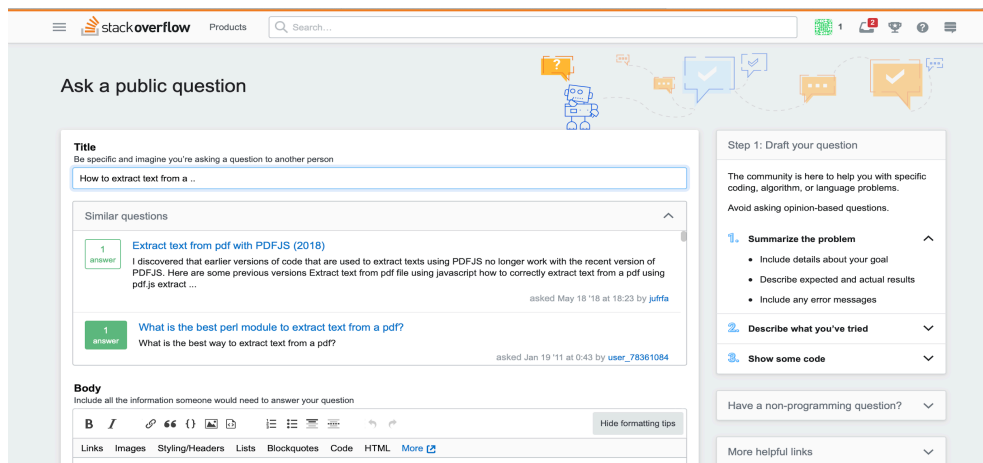
Screenshot of the question form on Stack Overflow (A Q&A platform for computer engineers): When a user attempts to ask a question, the tool automatically proposes similar questions which have previously been asked and answered.

Many websites, including government sites, use chatbots to respond to customer service queries. After typing in a question, the software shows the user similar questions that have already been posted (and answered). Smartphones and email clients like Google use a similar technique for the "auto-complete" feature that analyzes a user's input in real time. For example, the State of **New Jersey's Career Network** website (https://njcareers.org/), which provides automated coaching for job seekers, has such a chatbot. Once a question is asked, the software remembers it along with the answer so it can easily provide such information next time.
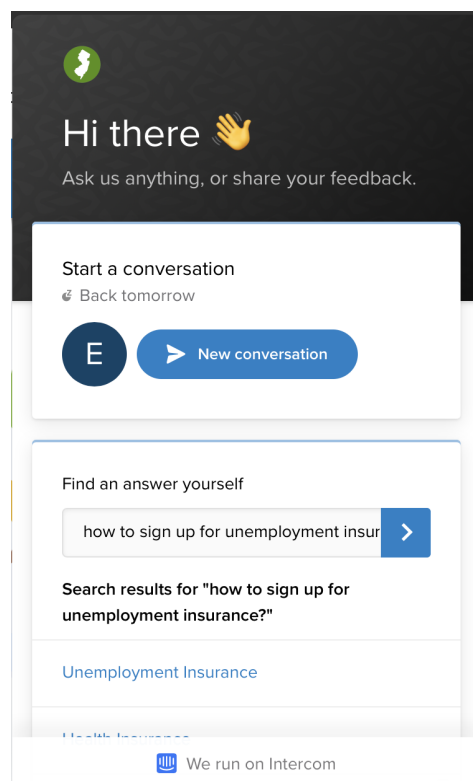
Screenshot of the chatbot on njcareer.org which provides users with the links to relevant answers based on the question it is asked.

Solution 3: Use Filtering Software to Sift Out the Real and the Relevant

A third challenge -- related to deduplication -- is to separate topical from irrelevant comments, whether they are posted maliciously or not. In other words, some comments are fake and posted by bots and others may be posted by humans but are irrelevant. These twin challenges have different solutions

Many people have already called for the use of captcha technology designed to filter humans from bots. **Captcha** is an example of a "Turing Test" - a technique developed by Alan Turing to tell humans and robots apart - and stands for "Completely Automated Public Turing Test to tell Computers and Humans Apart." With captcha, users are presented an image of a set of visually distorted letters and numbers and asked to enter the same characters into a textbox. When captcha was invented nearly two decades ago, it was believed that machines would not be able to complete this task since only humans would be able to interpret what the distorted characters were. With advances in computing power this is no longer true and techniques to defeat captcha have been created but captcha has also been reinvented to protect against these attacks. In its latest form, called **reCaptcha** (a free service, owned by Google), "bot activity" is identified using artificial intelligence.[49] "NoCaptcha ReCaptcha" simplified the user experience of Captcha by simply requiring users to click a box which says "I am not a Robot."[50] By analyzing several parameters on a website including mouse movements and button clicks, the service can differentiate between humans and robots. In 2018, Google announced "**reCaptcha v. 3**" which eliminated the need for any human interaction with Captcha at all. By using risk analysis algorithms that assign a"risk score" to every person browsing a website using the tool, the software alerts administrators if fraudulent activity is detected.[51]

But, in addition to sorting out the real from the fake, there is still a need to make it easier for federal agencies to sort the relevant from the irrelevant as part of the process of making sense of large quantities of content. Again, content-based filtering techniques that combine human intelligence and machine learning can help to sort irrelevant or off-topic comments. Such techniques are used, for example, with spam detection. Software analyzes content to determine whether it meets certain "rules." The simplest of these methods measures the occurrence of certain words and phrases – telltale signs of spam. More complex techniques involve identifying common patterns in submissions. Many blogging platforms such as Wordpress use such techniques to filter abusive/spam/advertorial comments in the comments section.

Creating the data to train an algorithm for the relevant subject matter domain has become faster and cheaper using crowdsourced human labor. As the Center for Technology and

[49] Matt Burgess, "Captcha is dying. This is how it's being reinvented," *Wired Magazine* (Oct 2017), https://www.wired.co.uk/article/captcha-automation-broken-history-fix.
[50] Vinay Shet, "Are you a robot? Introducing "No CAPTCHA reCAPTCHA," Google Security Blog (December 3,2014), https://security.googleblog.com/2014/12/are-you-robot-introducing-no-captcha.html.
[51] "reCAPTCHA: Easy on Humans, Hard on Bots" https://www.google.com/recaptcha/intro/v3.html.

Democracy explains, training algorithms that teach a computer to classify content are sometimes developed "using crowdsourcing services such as CrowdFlower or Amazon Mechanical Turk. Researchers or developers typically provide definitions for the targeted content (e.g. hate speech, spam, "toxic" comments) or other instructions for annotating the text. While there are significant ethical challenges with these particular crowdwork platforms, which are notorious for underpayment of workers, the point to take away is that the combination of machine algorithm and manually coded training data to teach that algorithm is a potential avenue for sorting comments and extracting meaning from them to make them easier for regulators to read.[52]

Informing Rulemaking and Oversight with CrowdLaw

This testimony has examined how agencies can improve the outputs of the e-rulemaking process in order to enhance the quality of information received during the public comment period, pointing out the widespread availability of data analytical tools for making sense of comments. These solutions demonstrate that, with the right data mining tools, astroturfing need not be an impediment to effective and meaningful participation.

However, the need to improve the quality and diversity of participation remains urgent and unresolved. People are hungry for meaningful opportunities to participate. Half of respondents surveyed by Pew said they had participated in a civic activity in the past year.[53] But more want to do so and about three-quarters of those surveyed by the Public Agenda in 2019 said they would participate under two circumstances that are currently not present in notice-and-comment rulemaking: namely, if they knew that participation was relevant and if they could contribute their skills and experiences.[54]

Not only do people want to get involved in the life of their democracy, their involvement will benefit our institutions by bringing more diverse voices into the process. There is a need, as we have seen with the Community Reinvestment Act rulemaking, to create an opportunity for more people with lived experience of the statute to participate. As noted earlier, last week's hearing underscored how bank closures disproportionately affect low

---

[52] Natasha Duarte, Emma Llanso, and Anna Loup, "Mixed Messages? The Limits of Automated Social Media Content Analysis," In *FAT*, p. 106. 2018, https://cdt.org/wp-content/uploads/2017/11/2017-11-13-Mixed-Messages-Paper.pdf.

[53] Aaron Smith, "Part 1: Online and offline civic engagement in America." *Pew Research Center*, April 25, 2013, https://www.pewresearch.org/internet/2013/04/25/part-1-online-and-offline-civic-engagement-in-america/.

[54] "Strengthening Democracy: What do Americans Think?" *Public Agenda and the Kettering Foundation*, 2019, https://www.publicagenda.org/wp-content/uploads/2019/08/Strengthening_Democracy_WhatDoAmericansThinkFINAL.pdf.

income communities. Lawmakers have urged the OCC to establish a notice and comment period that allows for a larger and more diverse group of stakeholders to comment on the proposal.

Yet rulemaking -- as civic participation generally -- does not attract diverse perspectives. Cornell law professor Cynthia Fiorina explains that regulated entities tend to be more represented in rulemakings than regulatory beneficiaries. Studies by a variety of academics find that business groups dominate the commenting process.[55] While there is still limited empirical research on electronic rulemaking, it appears that individuals all too rarely submit substantive comments -- in the same way that freedom of information requests come far less often from investigative reporters or civic groups than from businesses.[56] Surveys undertaken by Pew in both 2008 and 2012 found that civic engagement is overwhelmingly the province of the wealthy, white and educated.[57]

The notice-and-commenting process is not attracting the balanced and deliberative discussion that it was intended to attract. The design of the current notice-and-comment process exacerbates armchair activism and amplifies some voices at the expense of others with relevant expertise and experience to share. Structural challenges compound the issue: the public is given the opportunity to participate only very late in the process; people are provided with limited information other than the draft rule or notice of proposed rulemaking; and agency officials are prohibited from responding to public comments during the process in order to foster more deliberation and elicit more information.[58]

---

[55] Jason Webb Yackee and Susan Webb Yackee, "A bias towards business? Assessing interest group influence on the US bureaucracy," *The Journal of Politics* 68, no. 1 (2006): 128-139.

[56] David E Pozen, "Freedom of information beyond the Freedom of Information Act," *U. Pa. L. Rev.* 165 (2016): 1097, https://scholarship.law.columbia.edu/faculty_scholarship/2022.

[57] Aaron Smith, "Part 1: Online and offline civic engagement in America." *Pew Research Center*, April 25, 2013, https://www.pewresearch.org/internet/2013/04/25/part-1-online-and-offline-civic-engagement-in-america/ ("a key finding of our 2008 research was that Americans with high levels of income and educational attainment are much more likely than the less educated and less well-off to take part in groups or events organized around advancing political or social issues. That tendency is as true today as it was four years ago, as this type of political involvement remains heavily associated with both household income and educational attainment.").

[58] Cary Coglianese, Heather Kilmartin, and Evan Mendelson, "Transparency and public participation in the federal rulemaking process: Recommendations for the new administration," *Geo. Wash. L. Rev.* 77 (2008): 924, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1292911.

Solution 4: Use crowdlaw platforms to increase diversity and quality of commenting

Thus, in addition to improving the commenting process ex post using machine learning tools, lawmakers should use alternative and complementary platforms and processes -- ones used by other governments and organizations -- in order to eliminate the challenge of voluminous, duplicative and fake comments from the outset. I examine these below.

Crowdlaw describes the use of the Internet to enable citizens to volunteer, or be selected to participate, in law and policymaking. In many different jurisdictions around the world, people are collectively taking part in every stage of lawmaking. They identify problems or solutions, they draft proposals, and they debate, vote on, implement or evaluate policies, regulations and laws. Innovative technologies, including collective intelligence, artificial intelligence and machine learning, are enabling these new forms of democratic experimentation. Parliaments and city councils have sought to reverse the decline in democratic trust by adopting more of these online participatory practices. To give just one example, Taiwan has enacted 26 statutes informed by online and offline deliberation by 250,000 people through a process known as vTaiwan.[59]

Innovative crowdlaw approaches should be undertaken by American regulatory agencies and the committees that oversee them to improve the quality of public comments. The innovative practices described below are explained in detail on congress.crowd.law, a website designed to educate public officials about crowdlaw processes and platforms.

---

[59] Audrey Tang, "A Strong Democracy is a Digital Democracy," *The New York Times*, October 15, 2019, https://www.nytimes.com/2019/10/15/opinion/taiwan-digital-democracy.html.

Solution 5: Like Brazil and New Jersey, Use Wiki Surveys to Reduce Volume and Duplication

Landing page for the New Jersey Future of Work Task Force public consultation using All Our Ideas

Later this month, the State of New Jersey's Future of Work Task Force will start multiple online public engagement exercises using a wiki survey tool called **All Our Ideas** in order to engage workers in defining the challenges associated with the impact of technology on the future of worker rights, health and learning. All Our Ideas is a free, open source tool developed by Princeton professor Matt Salganik, who explains that All Our Ideas combines "the scale, speed, and quantification of a survey while still allowing new information to 'bubble up' from respondents." A wiki survey presents respondents with a question and then a random series of two answer choices. People select the response they prefer (or "I can't decide" as a third answer) or they may submit their own response. As people are repeatedly selecting between two randomly generated options, it is a faster and easier mechanism for responding to a series of questions. People can answer as many or as few questions as they choose and, with enough people participating, the resulting list is a rank ordered list of the answer choices. Because people respond to questions and can add their own responses, it is known as a **Wiki Survey**.

The user interface for voting in the AllOurIdeas tool. Users can pick between the two options or add their own in the text box.

The All Our Ideas system automatically tabulates and visualizes the results and can also select "View Results" to see how other participants are responding.

The All Our Ideas tool results page in New York City (2010): The results are updated in real time to show a rank ordered list. All option choices are assigned an equal score (50) when voting begins. This score changes based on the users' votes.

Although participants select between two options (or add an idea), New Jersey is not seeking a referendum on the future of work policy. People are not voting in the conventional sense. Rather, the Task Force is using the tool to create a more manageable and structured commenting process.

The Task Force has pre-populated the wiki survey tool with dozens of issues that are already known to it, enabling people to select one of them, thereby reducing proliferation of duplicative comments. In addition, people can also submit new ideas, which, if different, are then added to the roster of options; this function enables the public to contribute new information to the policymaking process. For example, when the NYC Mayor's office used All Our Ideas to collect feedback on the City's sustainability plan in 2007, eight of the top ten ideas selected for the plan were respondent contributions. For

instance, the top-rated idea, "Keep NYC's drinking water clean by banning fracking in NYC's watershed," was submitted by a participant.[60]

Similarly, in 2011 in the State of Rio Grande Do Sul in Brazil, then-Governor Tarso Genro used All Our Ideas to create the "Governor Asks" program, by some measures the largest citizen consultation in the country's history. In its first iteration, the Governor's Office collected more than 1,300 new proposals and 122,000 votes on public health care policy within a period of 30 days. In 2012, Rio Grande do Sul collected more than 2,000 submissions and then used the tool to help identify 10 priorities in promoting traffic safety. In order to encourage participation, the government partnered with civil society organizations and sent two "voting vans" equipped with tablets to collect votes across the state. This broadened access to those without access to the internet at home. This innovative technique helped ensure that a broad and diverse group of people, representing the opinions of ordinary citizens across the country, was able to participate.

The wiki survey method of showing people two ideas and having them choose between them or submit a new idea has several practical benefits. It makes it harder to manipulate or game results as one can with open-ended commenting of the kind found on regulations.gov. Respondents cannot manipulate which answer choices they will see. Second, because respondents must select one of two discrete answer choices from each pair (or add their own), this reduces the impulse to add new ideas unless there is something new to be said. New submissions can also be reviewed prior to posting to reduce duplication. Also, the need to pick one of two submissions does help with prioritization of ideas. While rulemaking is not a popularity contest, this prioritization can help ensure that the agency sees those ideas that are most novel. In any event, all submissions are still shown and readily available. This feature is particularly valuable in policy contexts in which finite resources make it helpful for agency officials to have some assistance extracting the most unique comments.

Solution 6: Obtain Better Exercise: Use Collaborative Drafting and Annotation as Germany Did to Engage More Experts in the Review of Rules

Other countries are turning to online collaborative drafting processes and platforms to develop policies, rules and laws with the public, especially with experts. Instead of a hearing with a handful of experts, **online collaborative annotation** could make it possible to hear from a broader and deeper range of experts. To enhance the level of

---

[60] Matthew J. Salganik and Karen E.C. Levy. "Wiki Surveys: Open and Quantifiable Social Data Collection." *PLoS ONE* 10, no. 5 (2015): 12. https://doi.org/10.1371/journal.pone.0123483.

expertise in rulemaking, an agency or a committee could use collaborative annotation to invite experts to annotate and comment on a draft rule. They could, as the German federal government did in 2018, invite a select subset of experts to discuss the draft as a complement to the notice-and-comment e-rulemaking process.

In 2018, the German government used an annotation platform to "expert source" feedback on its draft artificial intelligence policy.[61] By putting the draft on **Hypothes.is**, a free and open source annotation tool, the German Chancellor's Office, working in collaboration with Harvard University's Berkman Center for Internet and Society and the New York University Governance Lab, was able to solicit the input of global legal, technology and policy experts. Using an annotation platform also made it possible for people to see one another's feedback and create a robust dialogue, instead of a series of disconnected comments.

The Hypothes.is tool overlaid on the webpage of the draft artificial intelligence policy (2018)

Hypothes.is can be used on any webpage. It offers the ability to highlight, mark up or respond to other people's comments, and it offers both public and private annotations on the same page. Comments can be tagged for ease of filtering. Adding hypothes.is to a webpage does not change how the site looks or works; when the tool is active, it "adds a layer" on top of the website to enable annotation without affecting the underlying text. Users can turn on and off this layer as they want.

---

[61] Beth Simone Noveck, Rose Harvey and Anirudh Dinesh, "The Open Policymaking Playbook," *The GovLab*, April, 2019, https://www.thegovlab.org/static/files/publications/openpolicymaking-april29.pdf.

While collaborative annotation and drafting demand a higher level of commitment and greater knowledge of the subject matter than clicktivist commenting, when designed right, public participation in drafting offers key advantages. It provides an opportunity to obtain meaningful expert review of draft rules. It raises issues policymakers don't know about and ensures that drafts more effectively reflect the concerns of the people impacted by them. It is much faster and easier to organize online expert review using an annotation platform, making it efficient to organize, while providing the means to get balanced and thoughtful reactions to draft rules. An oversight committee could organize such an online annotation process alongside notice-and-comment rulemaking.

Solution 7: Set up a UK-Style Evidence Checks Process to Crowdsource Review of Comments and Evidence

The UK Parliament uses online **Evidence Checks** to invite members of the public to comment on the rigor of evidence on which a policy is based. This process allows a large and diverse group of people with relevant experience and expertise to identify gaps in evidence that require further review and aids in oversight.[62] If Congress would like to ensure diverse citizen input into the rulemaking process, the Committee should innovate the process of citizen engagement and **pilot the use of Evidence Checks.**

---

[62] "Evidence Checks UK," CrowdLaw for Congress, *The GovLab* (Feb 2020), https://congress.crowd.law/files/evidence-checks-uk-briefing-note.pdf.

In the UK House of Commons, where the role of committees is similar to that in the United States, there is a Select Committee conducting oversight for each government department, examining spending, policies and administration. In an Evidence Check, government departments and agencies supply information to the Committee about an issue. Committee staff publish that information at parliament.uk and share the task of scrutinizing it with a wider pool of experts, stakeholders, and members of the public. Typically, the Committee uploads the government statement as a publicly viewable PDF and frames the request with specific questions and problems that they would like participants to address. The process comprises three steps:

First, the Committee asks a government department to supply information about a policy, and the evidence on which the policy is based. Second, the Committee publishes the departmental submission and adds a page to their website to collect comments over a period of about a month, inviting academics, stakeholders, practitioners and members of the public affected by the policy to comment on the departmental advice. They might comment on the strength of the evidence provided by the department, highlighting contrasting evidence, selection biases and gaps. The web forum is public, but committee staff may choose to review comments before and after users post them to ensure that they are not defamatory, abusive, or otherwise inappropriate.

Finally, the Committee assesses comments and uses them to guide further investigation of the policy and/or integrates the commentary into its final report, which is supplied to the relevant government Minister for response.

Within this broad approach, Commons Select Committees have implemented evidence checks in varying ways. In 2014-15 the Education Select Committee used the process to help develop its work program. Initially, the Committee requested a two-page statement on each of nine topics from the Department of Education, inviting public comment via web forums on each topic, as well as general comments on the Department's approach to the use of evidence. Comments on the web forums then helped the Committee to decide what areas to focus on and what areas to hold oral evidence sessions for.

In 2016, the Science and Technology Select Committee published seven government statements on policy areas, including driverless cars, smart cities, digital government, smart meters and flexible working arrangements. It sought comments that aligned with a

framework that the Institute for Government developed in partnership with the Alliance for Useful Evidence and Sense About Science. The framework covered diagnosis of the issue, evidence-based action by government, implementation method, value for money, and testing and evaluation.

Targeted outreach, including social media, guest blogs on civil society organization websites, and leveraging the networks of organizations with expertise in the related policy topic, is crucial for obtaining high quality participation on an array of policy topics.

Evidence Checks help committees to more efficiently and effectively hold government to account by leveraging the collective intelligence of a broader expert audience. In 2016, an Evidence Check conducted by the Women and Equalities Committee into sexual harassment in schools (dubbed a "Fact Check") generated contributions from students with lived experience of harassment and led to a revised (upwards) estimate of the incidence of harassment. Information from contributors was incorporated into the subsequent Ministerial Briefing on the issue.

Solution 8: Democratize Oversight: Pilot the Use of Citizen Juries as in Belgium

If Congress would genuinely like to ensure diverse citizen input into the rulemaking process, the Committee should innovate the process of citizen engagement and **pilot the use of a citizen jury** to democratize and enhance oversight. Citizen juries have long been in use for civic deliberation in the United States but, as of December 2019, the legislature of the Brussels Region of Belgium formally introduced the use of citizen juries into the work of its legislative committees.[63]

The Regional legislature established that every committee shall include a citizen jury, comprising a random sample of 45 residents aged 16 and above, to participate in advisory committees alongside elected legislators. Citizen juries are attached to a specific standing committee, where they deliberate with selected parliamentarians on a given topic and formulate recommendations after four days of discussion and committee hearings of experts.

---

[63] For more on citizen juries, see Beth Simone Noveck, "The electronic revolution in rulemaking," *Emory LJ* 53 (2004): 499ff, https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/emlj53&section=20.

Inspired by ancient Greek democracy where citizens chosen by lot served in a wide range of governing roles, and by one-off sortition (aka selection by lot) experiments in Australia, Ireland, Spain and in the United States, the Brussels region is seeking to institutionalize the benefits of citizen engagement in formal lawmaking processes by adopting the use of citizen juries.

Under this arrangement, Regional Parliament standing committees each comprise 15 parliamentarians and 45 randomly chosen citizens. The two groups work together to draft recommendations on any given issue. The citizen participants are chosen as follows:

1) a first round, in which people are chosen by lot from among the population
2) a second draw among those who have expressed interest in participating. A random sampling method is used to ensure diversity of gender, age, geography, level of education and, important for Belgium, language spoken. Citizens serve for one year.

Political science Professor Min Reuchamps writes: "in this new deliberative process, the power of the citizens is significant. Nevertheless, the institutional and legal framework in Belgium does not allow non-elected citizens to officially vote in parliament. Accordingly, the recommendations adopted by both the randomly selected citizens and the parliamentarians will be voted upon separately."[64] In other words, the agenda and the final vote remain firmly in the hands of elected legislators. But citizens offer information, ideas and reactions as part of the deliberative process. Moreover, the fact that the jury comprises a random sample of citizens ensures a diverse spectrum of different citizen perspectives.

The House Financial Services Committee should consider a pilot project to incorporate the participation of a deliberative citizen jury if it wants to ensure both more diversity and legitimacy in its work.

I hope that these recommendations have been of value and I am happy to answer any questions now or in the future.

Thank you.

---

[64] Min Reuchamp, "Belgium's experiment in permanent forms of deliberative democracy," ConstitutionNet (January 17, 2019), available at http://constitutionnet.org/news/belgiums-experiment-permanent-forms-deliberative-democracy.