



We

ENGINEERING CULTURE

Are All

ONLINE GOVERNANCE

Tech

CHALLENGES OF SOCIAL CHANGE

Builders

By Sudhir Venkatesh

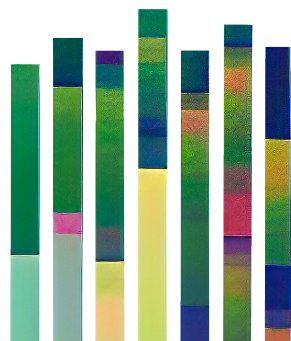
The contemporary technology sector creates an uneasy set of contradictions for the rest of society.

On the one hand, there's really no way to live without products that rely on digital technologies. On the other hand, no one likes relying on these products let alone suspecting that their producers are intentionally hiding aspects of their business, acting unethically, or playing fast and loose with the data they release.

Some of us make our livelihoods challenging and confronting this sector. Independent journalists, activists, and academics, to cite the most well-known examples, successfully extract significant goods from tech firms — money, fines, data, disclosure — and use these goods to improve our understanding (think detailed exposés of tech practices and whistle-blower reports to large-scale research studies, and the like). Despite the occasional victory, it is reasonable for the outsider to conclude that, at the end of the day, we are really just powerless in the face of Silicon Valley.

And, equally reasonable is the sense that this *must* change.

So, *how* can this change?



My point of view on this question arises from having spent ten years working in the tech industry, managing product teams, building academic advisory boards, releasing data to the public, and helping shape corporate policy. I've been both an employee and a consultant. This disclosure is critical because my job has included either explicitly safeguarding company data or finding ways to reconcile their needs with the asks of change makers—journalists, academics, activists. My teams have executed data release agreements, funded academic research, tested and launched disclosure reports, and supported independent journalists. Mine is an insider's view.

Outsiders underestimate the value of *imminent* critiques and how such standpoints might help them to leverage social change. So far, we've tilled the ground with pleas from the outside—pleas based on ethical

standards, human or civil rights, fair market competition, and other *externally-driven* standpoints of criticism (external because they are not grounded in the logic of the industry). Critiquing tech from an outside vantage point is valuable and necessary. In this essay, I am suggesting we need to add a perspective grounded in the lingua franca of the tech industry itself, namely, the logic of product development. Exposing the internal contradictions of product development in the tech sector will enable us to diagnose some of the challenges that arise, challenges such as adverse social impacts and negative imprints on well-being. Will this imminent posture be more efficacious to change efforts? Is this approach any better than the other options? I'm not entirely sure and leave those questions to others.

Let's unpack my point of view.

Governance & Product Culture: technology built by consumers

Most outsiders who seek goods from the tech industry spend little time understanding how tech works. I don't mean how a computer works. Instead, I mean how people in the tech sector labor together.¹ In fact, I'd argue that most of us look at the industry and think, "Looks like just another place where people make money selling me stuff. Doesn't seem all that different than shoes or baby food from the standpoint of business. The magic must be

in that damn computer."

There's some truth here. Tech is a lucrative business like many other businesses. But, there's a difference worth considering that has nothing to do with arcane technical knowledge or complex hardware. People who build many kinds of tech products do so in *collaboration* with you and me, which is unlike many other industries. It doesn't really look like it, I get it. Seems as though we run to the Apple Store to buy that nifty new iPhone *after* it is built. But, that's not really the case. For any such tool, let's call tech products "tools" for simplicity—

¹ This also complements studies, such as those of Sarah T. Roberts (2019), that critically examine the industry's complex labor arrangements with contractors, vendors, and other external parties.

what is *initially* available is only one-half of the story. What remains is the way consumers use the tool. An electric car, a bank website, an online hotel booking service, or a social media platform. The specific type of tool is irrelevant. A firm builds a version of the tool, but the company is waiting on consumers to put that tool to use. Thus, what is built on Day 1 is not what will exist on Day 9000. In fact, we should really think of the tool on Day 1 as an unfinished product—which is unlike other products, like shoes or baby food.

This is important for two reasons. First, this process is the principal factor generating tech’s social imprint, including notably the harmful effects on the wider society. It is what generates a wide range of problems, ranging from child exploitation to election interference to online bullying.

Second, this process is at the heart of the issue of power and control—or more to the point, outsiders’ feelings of powerlessness and their literal lack of control.

A couple of quick caveats. There are some tech products that seem less pernicious, or look like other industry products, like shoes or baby food. Software as a Service (SaaS) is one example of an “off-the-shelf” tool; so too, can one point to enterprise products, like Adobe’s many creative tools. We can debate endlessly whether these too have negative social outcomes, but that’s beyond the scope of this essay. To keep things simple, let’s limit the scope of the observations below to tech products that depend on user-generated content.

Let me use a fictitious example to ground the discussion. Imagine two digital tools, created by two different companies, each built

to help parents motivate their children to read. Parents use Tool #1 for communication; most of the time, they share books their children love, and offer summaries of those selections alongside helpful reading strategies. Alternatively, parents use Tool #2 to discuss health concerns and review and locate pediatricians.

In two years, the company building Tool #1 is acquired by a large publishing house. They see Tool #1 as a catalyst for their children’s

book publishing business. They rebrand the tool as part of their overall online marketplace and focus on attracting the kinds of advertisers interested in this kind of tool. By contrast, the leaders at Tool #2 realize that they are essentially building a referral service where parents can share health care information. So, they rebrand themselves as an online health company and shift

their advertising, marketing, etc., to reflect this new direction.

Uncomfortable as it might sound, it is you and I that have helped build these two tools, for the respective companies. It was our activity—our active engagement, our willingness to share personal information, our time and energy that came to create Tool #1 and Tool #2. Put this way, we should start feeling that we are participating in a grand experiment—that we are taking a risk by using an untested and unproven tool in our lives. Unfortunately, the firms who launch experimental tech products do not feel the need to announce that they are in an experimental phase. Without any meaningful disclosure beyond their Terms of Service, they routinely carry out ongoing testing and refinement without our knowledge. (One might reasonably argue that product development for tools

Day 1 is not what will exist on Day 9000. In fact, we should really think of the tool on Day 1 as an unfinished product.

based on user-generated content *are always in experimental mode.*)²

That the firms developing Tool #1 and Tool #2 began in similar places, with similar goals, will likely be a forgotten element of the story. By the second year, they will be classified differently in the App Store or Google Play, and this classification will shape how the public perceives them for the foreseeable future. (Think of Twitter, reclassifying itself as a “News” app in 2016.) Nevertheless, they both began in the same way: by depending on consumers to hand over to them key aspects of their life—thoughts, book preferences, their child’s health data, eating habits, friendships, etc. In addition, each of the firms must gather and analyze the information that consumers are sharing with them. Only then can they adjust the tool so it fits what consumers are doing. If the firms fail to analyze the data and retrofit their tool, then consumers will stop using it and find (and help create) another tool that meets their needs.

Implications for Product Builders

That you and I, and the company, together build such digital tools is nothing new and has been the subject of much critical inquiry.³ Here, we want to explore some of the implications for changemakers. Let’s start by pointing to some of the complicated situations that arise when you and I help build products for companies.

² Here, I invite the reader to read the work of two scholars, Christin (2020) and Benjamin (2019).
³ For more on this, see Gillespie (2018).

For starters, as I mentioned, it is unlikely that the company disclosed that *you* would be part of an experiment. As a result, you might justifiably feel cheated, used, or deserving of compensation. We’ve got lots of rules in society about false advertising and about unethical research, and it is fair to ask whether tech companies are getting away with something in this regard. Second, you might feel trapped. You might feel that there’s not much

of a choice in the matter. Tech is everywhere. Who has the time to pause and ask, “Before I ride this plane, turn on this app, or do some online shopping, I *have* to get in touch with the company to talk about my role as one of their product builders!” Further, as we know from national elections and health epidemics, entire communities depend on digital tools for critical, sometimes lifesaving,

information. It’s hard to fight against those who developed tools that have become instrumental for living.

There’s a third way that this situation can be complicated and unsavory. To see this, we must continue to unpack our example of the two companies building online tools to support children’s reading.

Imagine that the firm building Tool #2—the one that helps parents exchange stories about children’s health—notices problems having to do with unwanted consumer use of their tool. They notice that users are harassing each other, engaging in hostile and hateful political debate over health practices such as vaccination or drug approval, and there are incidents of child “grooming” or early-stage exploitation. The firm did not anticipate these

problems. They were in “startup” mode, which means their focus was to bring as many users to the site as possible. This means they did not build a large internal safety team. They might be upset about facing such problems, but it is likely they neither have the experience or infrastructure to handle these issues, nor does their leadership team want to redirect resources away from what the industry calls “growth” prerogatives to “safety” needs. For them, all hands are on deck to increase the number of users. This metric, not safety indicators, enables them to secure investment and keep the lights on.

Say you are one of the users who has had an undesirable or negative experience with Tool #2. It is likely that you are not alone. Depending on their rate of growth, tech firms have hundreds of thousands, perhaps millions, of people who face safety-related issues at any one time. If you approach the firm, you probably won’t be treated as a co-builder, that is, as an insider helping to create the tool with the firm. It is far more likely that you will be told, that according to the “Terms of Service”—the legal agreement that describes your rights—you have limited recourse. The firm’s response is very much a direct function of the product development process. How they treat not only consumers, but activists, academics, and journalists who wish to study these situations is based on this co-creation effort. To put it another way, their likely response would be that you should have realized your role was to help the company *grow*.

Defining Governance

The term for how firms manage this collaborative building process is *governance*. Governance is an old word in social science, and you may already be familiar with it in different contexts. Ergo, a quick caveat. I use it here not in the traditional sense—namely, the study of how a company’s leaders

fulfill the basic administrative, policy and financial functions of a firm. In techspeak, governance refers to the challenge of building products in which content is created by users, members, subscribers, and customers.

I contend that paying attention to how tech workers organize the collaborative building process with consumers—how they *govern*— will help us to get at those bigger goals of power, control, and accountability.

Reactive Policymaking

Most tech firms are marked by a separation between *product* and *policy* units. Product teams build and maintain the technology infrastructure—including the hardware and software, and the design of the experience. Policy teams are responsible for legal functions, and, importantly for us, they manage the relationships with the outside world via communications, contracting, crisis response, data release, and government engagement. When it comes to the consumer, both policy and product units are relevant, but they think about the consumer in different ways. Product teams help consumers use the tool. Policy teams help answer questions and address concerns about the tool.

Thus far, I’ve been making the point that there’s a difference between tech products and other products, like baby food or shoes. We can expressly see this by looking at the separate work of product and policy teams. In companies that build traditional products, like baby food or shoes, the policy team typically writes the rules and policies for the consumer coterminous *with* the building of the product itself. When the product is launched in the market, the rules are already in place. In tech, however, recall that the product is not complete until you and I use it. What this means is that policymaking is also half-finished. Only after people use the product, can the teams of policy associates observe the specific use cases,

and *then* develop and formalize the policies/ rules. Furthermore, as those uses change, the associates will rewrite the rules—including writing new ones that directly contradict earlier versions. As so often happens, at one time, you could do or say something with a tool, and then suddenly, the same speech or behavior is unlawful and subject to a fine, law enforcement investigation, etc. It is up to you to stay abreast of all the rules, especially whether it they’ve shifted to require a different user responsibility.

If this feels a bit unfair or worse, unlawful, that’s a legitimate reaction. The company is changing the rules to protect themselves as they find new and unanticipated consumer uses. It is reasonable to ask, *How can a tech company construct their policies reactively, and shouldn’t they be held liable for failing to understand what might go wrong—and for failing to prevent the problems from occurring (especially the harmful ones)?*

Yes and no. If asked this question, the firm would likely have a two-fold response. First, as noted, they would tell you, “We’ve done nothing wrong. Please read our Terms of Service (ToS) where we’ve explained our product and your rights.” Alas, in practice, consumers rarely review the ToS. They might also point you to S.230 of the Communications Decency Act that does not hold them liable for user-generated content on their tool. Neither of these are entirely satisfactory, so let’s drum up a more helpful response based on the point of view of this essay: namely, how tech works.

We can start by acknowledging that this reactive policymaking is *itself part of the product building process*, not an anomaly or

vestigial component. Tech firms release products that are often little more than hunches—the fancier, polite word for this is “prototype.” They don’t know what you’ll do with their tool, so they throw out a version, buy some advertising, and then watch as consumers put that tool to use.

Think about our example of the two hypothetical firms building online tools to support child reading

(i.e., Tool #1 and Tool #2 above). Neither firm knew what they had really built until two years of consumer use had passed. When they started, they simply released a product that had a huge promise attached to it (e.g., *We can help your kid to read!*). If a consumer has a negative experience or suffers harm when using the product,

they might demand that the company provide redress.

Consider this from the standpoint of the firm’s *product development* process. Would it be unreasonable for us to conclude that the negative experience was *necessary*? It sure looks like the firm needs to see that a consumer was harmed before they acted. And, as we will discuss in the next section, it might be the case that the firm needed to see the harm occur many, many times—because their internal detection mechanisms (human review, automated algorithmic review, etc.) might only detect the harm after it becomes an established pattern. Put in the language of product development, a harm occurring at a large volume creates a signature or fingerprint that enables detection and creates the conditions for future intervention.

That harm is tolerated, nay even encouraged, by a process that tech leaders have

The company is changing the rules to protect themselves as they find new and unanticipated consumer uses.

adopted for some of their products should not feel acceptable. The situation becomes even more complicated and distressing as we play out these dynamics.

Automation by Myth

In most tech companies, the scale of consumer use of the tool blunts the effectiveness of policy as a safety tool. The consumer use patterns might be manifesting in real-time and at such a large scale—that is, at such a fast pace and at such a high volume—that policy associates are simply unable to develop systematic rules and enforcement protocols for the problems that the product teams send them. Think of products that are being used by millions of users around the world. In most cases, the policy associates will not have the background to adroitly predict all the cases (we address this below). Policy associates are dependent on the product teams for making it a priority to detect and report incidents of consumer problems. At best, companies typically advertise a consumer help line or email address, and most do not always have access to user reports of harm, if indeed such reporting options are even in place. For this reason, we must be careful to draw assumptions from the largest companies, like Meta or Google or Twitter/X, where policy teams are sufficiently large and well-resourced. The vast majority of companies feature policy teams that are small and have limited capacity to investigate consumer experiences.

In some cases, the action may be to proactively remove that content so a user never sees it.

The Two Myths of Online Governance & Safety

To create safe products when the scale of content created by users is high, other techniques must be adopted. Namely, the governance workers must turn to computer algorithms that can review and act upon content at a fraction of a second. These computational programs are primarily deployed to identify repeated instances within a large sample of violating content. That is, each piece of content moves through a process of algorithmic review. Based on a simple binary decision-making model, the output is a value assigned to that content. Either it requires action, or it can continue to be featured in the consumer's overall experience. In some cases, the action may be to *proactively* remove that content so a user never sees it. Alternatively, the action might be to *reactively hide that content from further exposure* (i.e., ensure that future users do not see that content anymore).

But, how to instruct the computer on which pieces of content to remove proactively or hide after-the-fact? To answer this question, we need to introduce the concept of *myth*.

In any tech firm, there will be beliefs in place about the best way to manage with the social and behavioral challenges that arise from using their tool. The word for the set of beliefs is *myth*. In organizational analysis, *myth* is a common analytic to explain why members of an organization make decisions in a particular way. A myth is a feature of collective social life. It is not a falsehood. Myths are durable, deeply ingrained sets of beliefs and notions that motivate behavior. And, they are features of groups that emerge over time. Myths are

commonly associated with large, abstract groups, such as societies, subcultures, and nations, but they are also particularly useful in examining interactions in bounded organizational settings—at school, among workplace groups, and so on.

In some contexts, a myth can have the connotation of obviousness—something unremarkable. Consider the well-worn phrase, *if you work hard, you can get ahead*. Whether true or false is beside the point for those who live according to this myth. It is a convenient means for individuals to express views, reinforce collective bonds, and transmit values and expectations across generations. In many tech firms, a common myth is “Be your Authentic Self.” You might see this on posters or within online employee discussion forums. It is partly a means of handling diversity in a global workforce—where dress might take different forms for different social and cultural groups. Over time, a myth may end up making things feel natural or timeless. It may be impossible to identify the precise origins of any particular myth. Here, we invoke the writings of Thomas Kuhn on *paradigms*—which is a close cousin to the concept of myth, as it is used here. Kuhn (2012) writes, “[C]onsiderable time elapses between the first consciousness of breakdown and the emergence of a new paradigm. When that occurs, the historian may capture at least a few hints of what extraordinary science is like”. Following Kuhn, the best we can do in this essay is to “hint” at some of the conditions under which the myths shaping governance labor in a tech firm arose.


We can point to two overarching myths that animate the governance work of tech firms. Taken together, these two myths provide a benchmark for determining *who* will perform governance-related labor and *how* governance work should be accomplished. In other words, myths have a normative element by virtue of setting expectations for acceptable or proper conduct. The first myth is *governance is an engineering problem*. This myth teaches that

managing users at scale requires prioritizing engineering-based approaches. The second—the *myth of self-sufficiency*—tells tech workers that the governance team should work on its own, and ideally should have limited, if any, engagement with outside parties or experts. Taken together, then, employees of a tech firm are expected to understand that, above all else, those who direct the engineering functions of the firm have the greatest say in prioritizing resources and making decisions. And the governance team should rely on their own engineering and product-driven expertise to solve the consumer problems with the tool.

Myth #1: Governance is an engineering problem

Our first myth—namely *governance is an engineering problem*—arose as part of the overall transformation digital technology. Some of the earliest tools looked more like baby food or shoes in terms of product development. Companies built them, charged a fee for their use, sent them to users on floppy disks or other portable objects, and then consumers used them but without necessarily sharing their use patterns in real-time with the company. This changed, some have argued, as advertising models replaced single purchases of stand-alone products, and as technology enabled firms to surveil users as they used products and services in real time. It was possible to watch hundreds of millions of users sharing their information, and then adjust the product to keep those users interested and engaged—in most cases, a firm simply gave away the product for free and made their money on advertising.

In today's model, the reliance on rules and enforcement to stimulate healthy and safe consumer behavior will be minimally effective, so this story goes, given the volume of content that appears on most platforms is generated so rapidly and is so large. It would be a waste



of time to devote extensive resources to proactive policy development for the reasons mentioned above. Spending time predicting what rules will be needed, based on what consumers will do with the tool, is too slow and cumbersome a process. Nor is it worthwhile to educate users by providing them materials up front that set expectations for healthy behavior. The prevailing belief in tech firms is that most of the problems on the platform are likely caused by adversarial actors who would not respond to education; regardless, there are simply too many users and not enough time to educate them all. Instead, the firm is better off relying on their engineers to do their magic, namely harnessing automation and advanced computational processes (e.g., machine learning-based decision making, algorithmic-based recommendation systems, and the like) to handle governance needs. In practical terms, as I note below, this way of thinking enables engineering teams and their leadership to minimize other units inside the company that might challenge their authority.

The myth of *governance is an engineering problem* is a powerful force inside firms *not* because leaders have consciously tested and verified their beliefs against other beliefs or points of view that are available. Instead, as the scholar Tarleton Gillespie, notes, the unquestioned belief in the efficacy of product-based approaches has become a self-fulfilling ideal for the tech industry that no longer needs to be scrutinized: “This link between platforms, moderation, and AI is quickly becoming self-fulfilling: platforms have reached a scale where only AI solutions seem viable; AI solutions allow platforms to grow further” (Gillespie, 2020, p. 2).

The Engineering is a Governance Problem Myth in Action

A simple hypothetical scenario, one that is common to nearly all firms managing user content, will highlight the ways that the views and beliefs—*myths*—of team members shape their governance work. Let’s return to our example of the two companies building reading products for children—Tool #1 and Tool #2. Say each company faces a rising incidence of uncivil, harmful user content. Each company is concerned about user safety as well as a flurry of negative media attention.

Within one of the companies, Tool #1, the staff looks closely at user posts. They find problematic behavior occurring in the posts, including hateful speech and threats to harm other users. In the second company, Tool #2, the staff notice unwanted and harmful user behavior in the comment field. Each company rolls into action. Those in Tool #1 label the user posts as “harassment,” whereas the team in Tool #2 classifies the unwanted comments as “toxic” behavior.

An immediate task for the respective teams within each company is to develop a perspective on the unwanted behavior in question—why the meanness or incivility is occurring, who is responsible, what are the effects, and of course, what should be done. Let us say that in each company, a specific team—call it a Governance Team—is charged with developing such a perspective. They will be creating an operational point of view that enables each Governance Team to act on the respective problem.

As we noted, tech products typically have thousands, if not millions, of users. Which means an even larger number of posts, comments, emotional reactions, etc. It is simply too difficult to review every post or comment by hand in real time. This means that the governance team in each of the two companies will

be using automation (i.e., a computer-driven or computational process) to anchor their approach.

Ultimately, each team must be able to identify and segregate a creator’s harmful content so they can be reviewed in advance before it reaches other users. This way, the impact on the community is minimal.

Tool #1

Say that the governance team for Tool #1 decides that the user’s gender, age, and political persuasion are the most useful variables for predicting a potentially “harassing” post. In their reasoning, the propensity that any user decides to share harassing comments can be well predicted by knowing these three personal traits about the user. So, they build a computer model to segregate all posts in which the creator of that post has a particular gender, age, and political makeup. Once these are segregated off, the team labels them as “Potentially Harassing Posts.” This strong perspective, or point of view, motivates the team to select posts based on these three determinants or *signals*. Everything else is *noise*. A metaphor might be that they have used a large net to capture a large number of fish with three specific traits. Before we look at what they do with the captured fish, let’s turn to the company building Tool #2.

Tool #2

The governance team in the company building Tool #2 takes a different approach in line with their own unique proposition. Recall that they called the mean-spirited behavior on their

platform—occurring in the comment field—“toxic.” (Not “harassment.”) They believe that the user’s *history of rule violation* will be the key predictor of whether any comment is likely to be “toxic.” They do not prioritize gender, age, and political persuasion as relevant for prediction. These variables do not end up in their own computational model, which takes into account only one factor—namely, whether a user has violated rules in the past. Put another way, their

net captures a lot of fish based on a different approach, and they probably have caught all different kinds of fish.

Keep playing out this process across the industry. One can imagine a third company—creating Tool #3—that is offering another online resource to improve children’s reading. And a fourth and so on. Each will probably devise a unique

approach to battling unwanted comments based on how their internal governance teams understand human motivation. The final choices that a governance team will make reflect their own proprietary data, and their myths and beliefs about humankind and society.

It is useful to consider such points of view on human behavior because conventional discussions of technology workflows often describe the labor of tech firms as highly technical. Sure, there are some very arcane tasks like building a machine classifier or writing a software program. However, in reality, these technical efforts should be thought of as truly social—that is, they depend very much on the particular people and what they believe about the world—in this case, what they believe about people who break rules.⁴

They find problematic behavior occurring in the posts, including hateful speech and threats to harm other users.

⁴ For further reading see Seaver (2017) and Christin (2020b).

Developing a POV on Human Behavior

For this reason, the first task for a governance team's seeking to automate the management of large volumes of content is to develop a point of view on human behavior. In this initial stage of problem scoping, well before any algorithm is created, a team must develop such a proposition about very particular behavioral phenomena—mean-spirited posting, uncivil comments, terrorism, child exploitation, and the like. They must agree with one another about *why* people acted in one way or another. After coming to a consensus, they can then set goals for the team and select corresponding measures of success for their product work.

I have offered an admittedly highly simplified account in this example to highlight that the creation of an algorithm is the *culmination of a social process*, not the independent generator of a governance team's actions. As I said, I draw this distinction because conventional discourse often anthropomorphizes algorithms as animate objects capable of driving action independent of the people who build and utilize them.

But what qualifies this Governance Team to solve a behavioral problem like harassment, incivility, hate speech, etc.? What is their expertise, and do they come to their work with deep training in these behaviors? It is to this aspect of tech firms that I now turn—which we've already introduced, in fact, with our notion of *the myth of self-sufficiency*.

Myth #2: Self-sufficiency and the Tyranny of Design Thinking

Silicon Valley in particular, and the tech industry in general, loves design-thinking approaches to building their products and solving their problems. The design-thinking approach is a decades-old framework for supporting company teams that need to weigh options, make decisions, and create a common

path forward. Nearly every type of tech company, building nearly every kind of tool, will use this approach. Design-thinking processes have been shown to be particularly useful for figuring out how to enhance the consumer experience, such as shifting a color scheme or reducing friction for new consumers when adopting a tool. But this approach can also face challenges and have

less utility when tech firms need to address social and behavioral problems. Let us understand why.

The ultimate goal of design-thinking within tech firms is to support the *governance is an engineering problem* myth. That is, design-thinking processes are brought in because the team must quickly establish a perspective or point of view on a behavioral challenge, like toxic speech or harassment. The product manager instructs the entire team to move quickly so that the engineers can begin creating a computational model. This immediately puts team members in the position of making a tradeoff: they cannot afford to spend excessive amounts of time in a discovery and learning mode, lest the money allocated to the highly-compensated engineers becomes wasted. Alternatively, move too fast and they

They work on challenges, such as child predation, bullying, gender violence, and terrorism, that can cause deep emotional distress.

may develop an inadequate understanding of the problem. In this context, the initial challenge for the team is to dedicate a defined period of time to sort out an approach to combating unwanted behavior. Design-thinking becomes a way to legitimate their work and reach their desired end state quickly and with minimal cost.

Once again, we turn to our example of building an online tool to help children read. To keep things simple, the focus is on a single company, the one in which parents use their tool to share health-related information.

Say that this company detects an excessive amount of hate speech and there is a disagreement among members of the company's governance team regarding the causes. Some team members view the mean-spirited behavior as being the product of a user being inexperienced on the platform. Other team members say it is probably a function of the political leanings of users. Since time and resources are limited, decisions will need to be made quickly regarding their preferred cause. Recall that they must build a computational model—an algorithm—that predicts the behavior and segregates the potentially hateful comments before they reach the entire community. Do they build an algorithm that captures the content of all inexperienced users and reviews it for hate speech? Or, do they build a model that sets aside the content posted by those belonging to a political group—regardless of their experience with the tool? A computational model will look much different if the predictive variables include politics (or not) or experience (or not). So, whose perspective is right—or right *enough* to provide direction to engineers to start building a computational model? Time, resources, the viability of the business and the quality of the consumer experience all depend on the right decision.

To manage this uncertainty, a team will undertake a period of internal review to establish their point of view and identify key factors (*signal*) and discard others (*noise*). In

this case, they may be deciding between many potential variables, including but not limited to inexperience and politics. Typically, over the course of a week or two, a sub-group—ten to twenty company employees—gather to understand the problem and identify approaches and potential solutions. Those who come together can include different functional roles—researchers, designers, perhaps an engineer or two. Subject matter experts that might have valuable knowledge on relevant topics, such as hate speech, are typically not included. At most, they may be invited to share their knowledge for a few minutes, but the prevailing view is that non-employees don't really understand the tool so their value is limited.

It is worth noting that the design-thinking approach occurs as governance teams operate under conditions of multiple, unenviable stressors. They work on substantive challenges, such as child predation, bullying, gender violence, and terrorism, that can cause deep emotional distress. Governance teams are not composed of social workers, counselors, therapists, and probation officers trained to handle distressing issues. Moreover, tech firms rarely provide access to mental, psychological, and health supports for these teams. Their executive leadership is also likely to be impatient. Consumers, the media, and/or possibly government officials are continuously scrutinizing problems with their tools. Employing the well-worn tech sector mantra, "move fast," executives will demand that the team identify a viable approach. Viable could mean many things, including creating a meaningful distraction until the press moves to another news topic.

With the possible exception of larger firms, it is unlikely for a company to place individuals with any significant knowledge of human behavior relevant to safety matters on a governance team. Rarely does the recruiting team responsible for governance team positions connect with schools of social work,

policy, law, or criminal justice. In addition, people move freely inside the company, landing on governance one year, only to cycle off to sales or marketing or another division soon thereafter. This process makes it unlikely that a governance team will have a majority of its members with training in a relevant field.

The net effect on the work process is significant. With time pressures and knowledge limitations, most governance teams are forced to draw on their existing knowledge for the design-thinking approach. You can imagine the dangers if the team is largely drawing on untested and underexamined stereotypes regarding human behavior. A common stereotype that pervades every firm is the belief that the world is made up of two kinds of people: good and bad. Translated into product language, there are good and bad users, the former who play nice, follow rules, act civil, etc. and the latter who are not redeemable and should not be allowed to use the tool.

This belief instructs that bad users should be punished, harder and harder, until they behave or leave. It is an escalating set of punishments—removal of privileges, frozen accounts, banning, referral to law enforcement, etc.—that is the best way to safeguard the *good* consumers from bad ones. The *Good & Bad User* notion does not derive from an official training program or a set of manuals that instruct team members how to understand anti-social behavior. It is a view that pervades the broader society, and so its prevalence among teams working on tech company governance is simply a fact they are human.

Destroying the Myths & Building Better Governance

In a tech firm dominated by the product development gestalt, the twin myths of *self-sufficiency* and *governance is an engineering problem* puts into place distinct conditions of working. For changemakers challenging such firms, several ideas should be considered.

1. Fight Common Sense

Tech workers tend to think that an understanding of human behavior is fundamentally an extension of common sense—not a form of specialized expertise. The initial declaration of self-sufficiency inures team members to the notion that their own point of view will be materially enhanced with any consultation of, or engagement with, experts. Recall that the opposite is, in fact, occurring. The prevailing belief is that outsiders—even those who work at other online companies!—will not really understand the inner workings of *their* product or tool or service.

In the face of this point of view, outsiders should consider a range of strategies that might be available. So far, we've relied on political organizing, leveraging government oversight powers, requests for data sharing, and other externally driven efforts that appeal to general rules, standards, and norms in society at large. To this, we should consider ways to dismantle the pervasive naïveté that proliferates across tech firms in regard to social and human behavior. The general view inside the tech sector, which is strengthened by their use of design-thinking approaches, is that a smart and capable group can crowd-source a solution to any problem. Fighting the absurdity of this proposition is paramount. To date, this view seems sensible in tech because most behavioral issues are repositioned in simplistic terms—why can't people just follow the law? If I can behave, why can't they? And so on. But, I doubt that groups working in other industrial sectors would plan a bridge construction effort or provide a medical diagnosis simply by virtue of their intellect and teamwork skills alone. At some point, the specific knowledge of transportation engineers and physicians would be required. Nevertheless, in tech firms, the twin myths of *self-sufficiency* and *governance is an engineering problem* makes it difficult for the firm to solicit help—and the individual employees to feel comfortable asking.

2. Exploit Potential Alignments in Product Development

There are some notable examples in which external parties have worked in an *imminent* fashion, using the product development process as a leverage to create change. Consider the adoption of transparency reporting for governance-related issues. In established firms and smaller entities, what began as voluntary disclosure of government requests for user data have now become comprehensive public reporting on a much wider range of governance and safety issues. We all now benefit from the industry norm that creates expectations for firms with user-generated content to disclose incidence, prevalence, and content management metrics for governance issues. The consultation with experts ended up as a powerful force that eventually transformed how the company measured and disclosed issues—inevitably leading to a new Transparency Report for Community Standards. Transparent reporting also created new pathways for external experts to advise the company on building safety products that could more effectively reduce harm, and eventually other firms followed suit. Consider that today, the extraordinarily impactful human rights, social activist, and governmental oversight work that can be carried out is a direct beneficiary of these reports.

At Facebook, this reporting did not arise because activists and the firm's policy team worked harmoniously. In fact, it was the product teams who were critical to the release of this information. The development of such reporting for governance issues was spear-

headed by the engagement of external subject matter experts who worked directly with the product teams responsible for keeping surfaces such as Groups, Pages, and Newsfeed safe for Facebook users. Various external parties—including academics, activists, and journalists—realized it was critical to partner with product teams to shape how the company measured problems, collected relevant data, and prepared public releases. As noted in this essay,

A common stereotype that pervades every firm is the belief that the world is made up of two kinds of people: good and bad.

the process of product development rests on accurate measurement to support the development of usable, safe products. So, the product team was incentivized to work with these external experts. In effect, these experts bypassed the policy directors whose responsibilities include *shielding* the external expert from involvement in product processes.

We can contrast this example with the more highly publicized Facebook Oversight Board, whose impact has been minimal in terms of truly reaching a large number of Facebook users. Ironically, the Oversight Board initiative began as a series of dialogues between Facebook's product leaders and academics who urged the adoption of an independent council for building "ground truth" into scalable enforcement practices. This was a sensible idea, and at first, the product teams were thirsty for such support and believed such ground truth mechanisms could make the product better and thereby create safety across the globe. But over time, the activists, lawyers, and academics who were recruited to build the initiative decided that it would be more influential to shape corporate policymaking rather than the product itself. The company's executive had no reason to resist since this

meant they could limit their need to have outsiders shape the core business. The net effect of the move into policy implementation, and away from the product development process, was to limit the overall impact of the Oversight Board. Today, most users are not affected by the work of Board, which reviews only a limited number of cases each year and has minimal insight into how Facebook's (now Meta's) products are built.

At the end of the day, to move the tech industry forward in a more responsible direction, we need a range of approaches, including adversarial activism, government oversight, and academic-driven data disclosure. To this, we should add a focus on understanding and leveraging opportunities within the product-development process.

Sudhir Venkatesh is William B. Ransford Professor of Sociology, and the Committee on Global Thought, at Columbia University in the City of New York.

References

- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code* (1 edition). Polity.
- Christin, A. (2020a). *Metrics at Work: Journalism and the Contested Meaning of Algorithms*. Princeton University Press.
- Christin, A. (2020b). The Ethnographer and the Algorithm: Beyond the Black Box. *Theory and Society*, 49(5), 897–918. <https://doi.org/10.1007/s11186-020-09411-3>
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.
- Gillespie, T. (2020). Content Moderation, AI, and the Question of Scale. *Big Data & Society*, 7(2), 2053951720943234. <https://doi.org/10.1177/2053951720943234>
- Kuhn, T. S. (2012). *The Structure of Scientific Revolutions: 50th Anniversary Edition* (I. Hacking, Ed.). University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/S/bo13179781.html>
- Roberts, S. T. (2019). *Behind the Screen: Content Moderation in the Shadows of Social Media* (Illustrated edition). Yale University Press.
- Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 2053951717738104. <https://doi.org/10.1177/2053951717738104>

