*Episteme* (2025), **22**, 745–767
doi:10.1017/epi.2024.22

episteme

ARTICLE

# Building Epistemically Healthier Platforms

Dallas Amico-Korby, Maralee Harrell and David Danks

University of California, San Diego, La Jolla, CA 92093, USA
**Corresponding author:** Dallas Amico-Korby; Email: damico@ucsd.edu

(Received 11 September 2023; revised 9 February 2024; accepted 9 April 2024; first published online 04 June 2024)

**Abstract**

When thinking about designing social media platforms, we often focus on factors such as usability, functionality, aesthetics, ethics, and so forth. Epistemic considerations have rarely been given the same level of attention in design discussions. This paper aims to rectify this neglect. We begin by arguing that there are epistemic norms that govern environments, including social media environments. Next, we provide a framework for applying these norms to the question of platform design. We then apply this framework to the real-world case of long-form informational content platforms. We argue that many current long-form informational content platforms are epistemically unhealthy. The good news? We provide concrete advice on how to take steps toward improving their health! Specifically, we argue that they should change how they verify and authenticate content creators and how this information is displayed to content consumers. We conclude by connecting this guidance to broader issues about the epistemic health of platforms.

**Keywords:** Epistemology; social epistemology; expertise; design ethics; social media design

When you type a query like "back pain relief exercises" into TikTok, you'll be presented with a wide variety of results: people claiming to be doctors with miracle advice "Fix low back pain in SECONDS!" (moorewellness 2023), people claiming insight from their own experiences "His back was hurting, I saw this on tik tok and it worked!! See it till the end" (anitaduran34 2021), and people who claim to be medical professionals using medical jargon "Fix SCOLIOSIS naturally!! – functional scoliosis is when the spine appears to be curved, but the apparent curvature is actually the result of an irregularity elsewhere in the body" (physicaltherapysession 2023). For people with back pain and no medical training, sifting through these results to find quality information can be confusing and overwhelming. Is the Scoliosis video made by a physical therapist, like they claim? Are any of these videos made by hucksters or snake-oil sales people? Even if they are telling the truth about their identity and are sincerely trying to offer helpful advice, do they know what they're talking about? None of this is clear from scrolling through the list of videos, reading their descriptions, or even from watching them.

This familiar example reveals a significant flaw in TikTok's epistemic design: those searching for information regarding a certain query are often unable to reliably find

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

high-quality information regarding that query. Similarly, those *providing* high-quality information regarding that query are often drowned out by the noise of click-bait and miracle cures. The design of TikTok does not permit the kinds of ranking-by-expertise that is often epistemically required.

When thinking about designing social media platforms, we often focus on factors such as usability, functionality, aesthetics, ethics, and so forth. Epistemic considerations have rarely been given the same level of attention in design discussions.[1] This paper aims to rectify this neglect. We begin in section 1 by arguing that there are epistemic norms that govern environments, including social media environments. Next, in section 2, we provide a framework for applying these norms to the question of platform design. In section 3, we apply this framework to the real-world case of long-form informational content platforms – TikTok, YouTube, Apple Podcasts, Substack, etc.[2] We argue that many current long-form informational content platforms are epistemically unhealthy, and so we provide concrete advice on how to take steps toward improving their health. Specifically, we argue that they should change how they verify and authenticate content creators and how this information is displayed to content consumers. We conclude by connecting this guidance to broader issues about the epistemic health of platforms.

## 1. Environmental epistemology

Our aim is to epistemically evaluate a certain type of social media platform. Very roughly, social media platforms are online *environments* where people interact and communicate.[3] But how do we evaluate environments *epistemically*? On our view, there are two types of epistemic norms – what we will call "specific" and "general" epistemic environment norms (EENs) – that govern environments.[4] Further, we adopt a health-based model for evaluation. That is, environments that satisfy the norms that apply to them count as epistemically healthy. Those environments that fail to satisfy the norms that apply to them are, to varying degrees, epistemically unhealthy.[5]

Importantly, this is an account of how to evaluate the epistemic health of *environments*, not those who design, build, or maintain those environments.[6] To see how it is possible to evaluate environments independently of the people who created them, consider the fact that we can aesthetically evaluate natural environments (like a grove of trees), even though they do not have creators. Even human-created environments can be evaluated independently of their creators. Imagine, for example, that an engineer and construction team design and build an excellent home. Next, imagine that a lightning strike causes a fire that destroys the home's kitchen. The home would – after the strike – be a poor-quality home, as a necessary condition for a good home is a functioning kitchen. Nonetheless, this would not reflect poorly on the engineer, the construction

[1]This is not to say that no attention has been given to these questions. Simon (2010), Rini (2017), Miller and Record (2017), and Record and Miller (2022) are examples.

[2]It may be surprising that we include TikTok here, but TikTok allows video uploads of up to ten minutes.

[3]Here, and throughout, we rely on an intuitive sense of the term "environment" as referring to collections and arrangements of objects, norms, conventions, and/or people.

[4]We defend these norms in greater detail in Amico-Korby *et al.* (2024).

[5]We explain and defend this account in much greater detail in Amico-Korby *et al.* (2024).

[6]Although, as we'll argue, epistemic environmental norms do have implications for these latter evaluations.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

team, or anyone else. This is a feature of the home, and not of its creators.[7] Of course, we are not evaluating environments epistemically in either of these cases. Explaining how to do this is our next task.

First, consider specific EENs. These are norms that arise from the specific function of the environment. Consider the house analogy again. A house is an environment with a particular function: to provide a place for people to live. And given the needs of people to store, cook, and eat food, fulfillment of this overall function implies that a "good" house must have a kitchen. This is why a house without a kitchen is a poor house. The function of the environment (a house) generates norms for evaluating the quality of the environment.

Sometimes environments have epistemic functions. In these cases, the environments generate specific EENs that can be used to evaluate the quality of the environment, and the environment will count as epistemically healthy with respect to these norms to the extent that it satisfies them. Consider, for example, a newsroom. A newsroom is an environment with clear epistemic functions. Specifically, it has (at least) the functions of facilitating the discovery, vetting, and transmission of newsworthy information. Many other environments have epistemic functions as well. Schools function as places of learning. Libraries function as storehouses of information. Even grocery stores have epistemic functions, as part of the function of a store is to communicate what is for sale, and how much those items cost.[8]

In the same way that the function of a house generates norms for evaluating the environment, the epistemic functions of these environments generate epistemic norms for evaluating them. For example, a "newsroom" without transmission capabilities – perhaps lacking camera and sound equipment, or a connection to the world outside the room – would fail to fulfill its epistemic function and be a poor newsroom (if it counts as a newsroom at all). Similarly, a culinary school that did not ensure its students to have access to kitchens would be a very poor culinary school; while some aspects of cooking can be taught without practice in a kitchen, most require some element of practice.

Environments can be better or worse at meeting the epistemic standards that apply to them; they can fulfill their epistemic functions to varying degrees. So, for example, a school environment may be largely conducive to learning, but still have weaknesses. Perhaps an otherwise excellent school fosters an overly competitive environment that occasionally results in tension between students that inhibits their ability to learn. Such a school may be largely epistemically healthy with respect to the specific EENs that apply to it, but still less than fully healthy.

At this point one may have two questions. First, one may wish to ask about environments that have problematic epistemic functions. For example, what if the environment has the function of disseminating disinformation? For all we've said so far, this function may seem to generate problematic epistemic standards of evaluation for the environment: the environment will satisfy the specific EEN (i.e., be "healthy") just to the extent that it succeeds in causing people to have incorrect beliefs (due to disseminating disinformation). That is, the environment seems to be epistemically healthy only if

[7]A similar story can be told about property maintainers, such as a property management group. Of course, once the house has been damaged, they may have a responsibility to intervene. But the quality of the house at the time of the lightning strike may be poor and yet not reflect poorly on them.

[8]To see this, note that a grocery store that had no information about what products are for sale and how much they cost would be a very poor-quality grocery store!

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

the people in it are not. Second, one may wish to ask about how environments get epistemic functions in the first place.

Our answers to both questions involve an appeal to *general* EENs: epistemic norms that govern *all* environments, and that have normative priority over specific EENs. As we've argued, the epistemic function of an environment sets epistemic standards for evaluating the environment – specific EENs. Specific EENs are norms for determining whether the environment is a good environment of its kind. General EENs are norms for determining whether the kind (or function) itself is eligible to count as epistemically healthy. If we think about these norms in terms of their implications for action, we might understand general EENs as permissibility norms. General EENs determine which types of environments may permissibly be designed, built, and maintained. Specific EENs determine how to do well in designing, building, and maintaining permissible environments.

To see how this works, consider a disembodied brain in a vat (Harman 1973: 5; Putnam 1992). The function of the vat is to give the brain experiences that are qualitatively indistinguishable from the sorts of experiences people outside the vat have. Such a vat may be an excellent vat, fulfilling its function well. And yet this would not make the vat environment an epistemically healthy one. Intuitively, environments of this kind are epistemically unhealthy. Another plausible example is Goldman's (1976) Fake Barn County, wherein a county has lined their roads with barn facades designed to fool motorists into thinking that there are actual barns lining the roads. Further examples may include conspiracy websites or cults. In light of these examples, we propose the following as a candidate for a general EEN: any environment with the function of massively deceiving its (human) occupants cannot count as epistemically healthy.[9]

Return now to our two questions. First, general EENs explain why environments that fulfill their problematic epistemic functions well cannot count as epistemically healthy. Environments with problematic epistemic functions will violate general EENs, and thereby count as epistemically unhealthy.

Second, while general EENs do not provide a descriptive explanation for how environments get epistemic functions in the first place, they do explain which epistemic functions we may design, create, and maintain. In our view, environments may get their epistemic functions through a wide range of processes. In some cases, an environment may get its epistemic function from its creator, in other cases inhabitants of the environment may alter existing epistemic functions or create new ones (e.g., the hashtag was neither created nor initially welcomed by Twitter (Panko 2017)). Further, the function of an environment may change over time. Regardless of the precise details of how environments get their epistemic functions, general EENs tell us which functions may count as epistemically healthy, and that provides limits on what kinds of environments (or what epistemic functions) we (as designers, creators, maintainers, or inhabitants) may epistemically bring about. Specific EENs tell us how to make permissible environments fulfill their functions well.

General and specific EENs not only govern the health of environments, but also have implications for action. In particular, if one is designing, building, or maintaining an environment, one has – at least – a defeasible reason to ensure it is epistemically healthy. To see this, note that the epistemic health of an environment is closely tied

[9]Plausibly a wide range of other general EENs can be defended. Further, one might think that a sufficiently lower threshold of deception would count as sufficient for violating a general EEN. We do not provide a complete normative account of such norms here.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

to people's ability to be epistemically successful in that environment. And epistemic success is often necessary for people to flourish and successfully live their lives as they please. Consider an epistemically unhealthy learning environment. If a classroom has rave lights flashing, a live DJ, and a person banging pots together just off-beat, it will likely be very difficult for students to learn much of anything. However, epistemic success for the students may be integral to their ability to successfully live their lives as they please. Well-recognized moral obligations to not harm others thus imply that we have defeasible moral reason to not create epistemically unhealthy environments nor take actions that harm the epistemic health of existing environments.

There are likely further reasons grounded in what we owe each other as epistemic agents. If one's testimony is discounted simply because they are a woman, they are wronged. Intuitively, this wrong is epistemic (or at least epistemically flavored).[10]

We also have obligations to others in virtue of their epistemic agency that arguably provide reasons for us to work to ensure environments are epistemically healthy as well. To see this, imagine someone (perhaps inspired by Thaler and Sunstein 2009) designing procedures for how medical patients and their families make end-of-life decisions. They might consider creating an overly paternalistic environment that pushes patients toward certain decisions. Perhaps they think that euthanasia is the most reasonable option for people with a certain diagnosis, and so create an environment that pushes patients toward that decision by requiring doctors to only explicitly offer that option to patients. An environment like this one is unhealthy because it disrespects peoples' epistemic agency. To the extent that we have obligations to respect peoples' epistemic agency, we therefore have obligations to ensure environments are epistemically healthy.

Importantly, our obligations for action are not limited to creating or designing environments. For example, when a person attends a lecture, they play a role in determining the lecture environment, even if they had nothing to do with its design or creation. If they decide to interrupt the speaker, or make repeated loud noises, the epistemic health of the lecture environment will be degraded. And we plausibly have obligations to do our part to contribute to the epistemic health of the environments we occupy because of our obligations to not harm others and to respect others' epistemic agency (by avoiding such behavior).

These obligations are defeasible obligations in accordance with our degree of control over the relevant environment, as well as the weight of our other normative obligations. For starters, our EE obligations may conflict, particularly when we occupy multiple environments simultaneously. Further, our EE obligations may conflict with moral reasons. For example, there may be contexts where being disruptive violates one's EE-based reasons, but interrupting is sufficiently morally urgent that it may be what one all things considered ought to do. It is especially important to keep this in mind given that appeals to something resembling EE norms have often been used to oppress or silence in morally and epistemically problematic ways. For example, appeals to civility norms are often made in order to silence or hinder morally important protest. A complete normative account will include (at least) an account of traditional epistemic norms, EE norms, moral norms, and an account how to handle conflicts between these norms.[11] We cannot hope to give such an account here, but we do wish to emphasize

[10]Miranda Fricker (2007) would call this an "epistemic injustice." See also Basu (2019) for more on epistemic wrongs.

[11]For a range of approaches to resolving conflicts between epistemic norms and moral and pragmatic norms, see Feldman (2000), Basu (2019), Bolinger (2020), and Davia (2022).

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

the importance of being on the lookout for such conflicts and of not assuming that EE-based obligations automatically win the day. This is especially important in the context of the advice we offer to platform designers in what follows. Designing and building healthy social media environments, both epistemically and morally, should be our aim.

The account of general EENs we've provided is incomplete in many ways. While we've sketched their normative role and offered a plausible example of such a norm, we have not done the difficult work of defending a substantive account of general EENs. Such an account must ultimately be grounded in an account of the epistemic good and the right, but what that amounts to is a matter of contention amongst epistemologists.[12] But – as will become clear – this single general EEN will be sufficient for our purposes in this paper, and for many epistemic design questions.[13]

## 2. Epistemologically sensitive design

Typically, when thinking about designing an environment, people focus on usability, functionality, aesthetics, ethics, etc., but epistemic health has rarely been given the same level of consideration. In the previous section we introduced norms for epistemically evaluating environments. In this section, we translate this account into five actionable steps that one can take when designing an environment. In the following section, we apply these steps to real-world cases.

Step 1: *Determine the kind of environment one aims to build or alter, and any epistemic functions of that kind of environment.*

We earlier argued that the kind of environment one aims to build determines standards for evaluation of environments of that kind (i.e., specific EENs). In order to determine the specific standards, one must first determine the type of environment one aims to set up, and whether this type of environment has any epistemic functions. As it happens, nearly every environment has *some* epistemic functions, so in practice this step will often involve determining exactly what those functions are.

Step 2: *Determine whether the epistemic functions are consistent with the satisfaction of the general EENs.*

This step is perhaps the most difficult, because it requires one to do serious normative work. This is, of course, not special to the epistemic case. If we want to determine whether it is morally permissible to create a platform of the sort we want to create, we

[12]We suggest a more complete account in Amico-Korby *et al.* (2024).

[13]Our account is related to, and rooted in, the work of many other scholars. Harding (1978: 205–6) discusses ways that the communities we occupy can shape our values and influence our inquiries. Similarly, many scholars of race and gender have recognized the epistemic effects that our social environments have on us. Examples include Hooks (1990) and Haraway (1992). Our work provides a way of epistemically evaluating these environments. Nguyen (2020) provides an account of one way that environments might be hostile to epistemic agents like us. Interpreted through the lens of our view, we would say that Nguyen has identified an important way that some environments are in violation of the epistemic norms that govern them – i.e., epistemically unhealthy. What we have done, is provide a more general normative framework for epistemically evaluating *all* environments. We, thus, think of our projects as friendly companions exploring similar terrain. We suggest that similar things can be said about the relationship between our view and Mills (1997) and Fricker (2007).

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

will have to do analogous normative work. Nonetheless, because this step requires normative commitments (as well as specification of relevant EENs), we will be agnostic about the exact implementation of this step in the design process.

Step 3: *Determine the specific EENs that govern this kind of environment.*

One might hope that the function(s) would transparently imply the specific EENs, but in practice, this step will typically be more complicated. Many epistemic functions readily imply the general types of norms, but not necessarily the exact norms. For example, if one aims to set up a school, the epistemic function of the environment may be (a) the acquisition and development of understanding, skills, capabilities, and practices (hopefully leading to expertise), and (b) the credentialing or certification of that development.[14] These functions suggest specific norms around, say, the efficacy of skill acquisition, but fail to provide much concrete guidance. Instead, we suggest that specific EENs should be elicited by thinking in terms of ideals or prototypes. Given the function, one should next determine what sort of environment(s) is ideal for fulfilling that function (where multiple environments may be equally ideal).[15] This ideal can then serve as the standard for measuring other environments. The ideal environment will be conducive to achieving the epistemic purposes to a certain degree. The epistemically healthiest environments will be conducive to a similar degree as the ideal environment, and the further away from that level of conductivity that an environment is the less epistemically healthy it is.

Importantly, this ideal should be ideal for the agents that are likely to occupy the relevant environment. For example, one function of an elementary school may be the development of appropriate skills (e.g., basic literacy) by elementary school children. The ideal elementary school, then, should not be an ideal environment for the education of ideal Bayesians, or ideal adults, as these are not the people who will occupy this environment. Rather, we should be interested in the ideal environment for children with the characteristics of those who the environment is intended for.

It is also important to note that what is ideal may change as circumstances change. For example, over time, a school's demographics may change from primarily children whose first language is the language of instruction, to most of its students being in the process of learning the language of instruction. The ideal way to fulfill the educational function is likely to change with these demographic changes. For example, ensuring that certain translation supports or spaces for learning the language of instruction are available and accessible is likely to be an important part of the ideal in the latter circumstances but not in the former.

Step 4: *Determine the design and functionality constraints that will (likely) foster epistemic health for the environment, where it is possible that no system could simultaneously satisfy all of the epistemic functions.*

Step 3 helps one determine what the environment needs to be like to fulfill its function, as well as how to (approximately) measure the epistemic health, relative to the relevant specific EENs. This knowledge must then be translated into particular design choices for the environment. This step is thus concerned with determining how to

[14]There are likely many more, but we keep it to two for simplicity's sake.

[15]With multiple purposes, it may also be that no environment is ideal for all purposes.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

build an environment that will fulfill the core functions, rather than analyses of that environment. As a result, relatively standard design thinking approaches (e.g., those advocated by the Interaction Design Foundation, or value-centered design, or many others) will be appropriate for this step, as design methodologies guide people toward creative solutions that respect many distinct constraints (including from specific EENs).

Step 5: *Build and empirically test the environment.*

Finally, one should build their environment, and empirically test it to determine whether it is epistemically healthy by being as conducive to fulfilling its epistemic functions as it ought to be. We emphasize the importance of this final step, as many seemingly well-designed systems have turned out to be less-than-ideal when deployed broadly in the world. Further, there is the risk that we end up maximizing whatever we have decided to measure at the outset, rather than the complex things we actually care about.[16]

## 3. Case study: a long-form informational platform

Step 1: *Determine the kind of environment one aims to build or alter, and any epistemic functions of that kind of environment.*

In this section, we explore a case study of environment design that specifically focuses on online platforms that provide users with the tools to create, publish, share, consume, and, in some cases, react to and comment on various types of long-form content. For example, YouTube provides a platform for long-form videos; Substack provides a platform for long-form written content; and Spotify provides a platform for long-form audio content.[17] Importantly, this function is not accidental for these companies; they *aim* to provide a platform that does these things.

Long-form content can involve two very different types of content: informative and performative, where the former *essentially* involves the sharing of information while the latter does not.[18] Examples of informative content on these platforms include non-fiction prose, personal newsletters, DIY videos, news, biography, etc. Examples of performative content include fiction, comedy, poetry, concert performances, podcast serials, theater, etc.

Of course, the same piece of content can play both roles, which is why we add the qualifier "essentially" in the previous paragraph. Performative content often conveys information, such as when an individual performance has epistemic aims. For example, a performance may have the goal of enlarging the audience's sense of possibilities – "I wasn't aware that you could play Bach on a set of glass bowls!"; or a performance may have the goal of convincing the audience of some moral claim, such as a play that aims to convince you to care about a moral issue.

[16]We're grateful for an anonymous reviewer for encouraging us to include this final thought. We also point the reader toward further discussion of related issues in O'Neil (2016) and Nguyen (2021*a*).

[17]Of course, many of these platforms have other goals and other functions as well. We focus on this one because we think it is a particularly central epistemic function of social media platforms. We are hopeful, however, that our proposed steps and our proposal provide people with sufficient resources to perform similar analyses on platforms' other epistemic functions.

[18]There are certainly other ways of carving up the space of content. We choose this way because of its fruitfulness for this discussion.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

But while performative content sometimes conveys information and sometimes has epistemic aims, these things are not essential to what it is to be a performance, nor do these things typically set standards or constraints for performances. To see this, note that a juggling performance, rock show, or comedy video could be excellent – and fully achieve its ends – while conveying no information at all. Complaining "Metallica played incredibly tonight and the energy in the arena was electric, but I didn't learn anything (I already knew all the songs!) so that concert failed as a performance," misunderstands the essential function of performances.

On the other hand, informative content essentially involves the sharing of information. If a lecture or a DIY instructional video was to consist primarily of juggling while conveying little or no information, it would fail to be a very good lecture or DIY video (it may not even *count* as a lecture or DIY video)! Good informational content essentially involves the transmission of information. The function of informational content is to inform.

Of course, that informative content essentially involves the sharing of information does not imply that informational content never involves performance. Enrapturing lectures often involve many of the techniques utilized in theatre performance, and natural history museums, documentaries, and DIY YouTube videos are best when they're entertaining and engaging.[19] Further, some of these types of content plausibly have both functions.

In this section, we will focus on online environments that aim to (and do) provide a platform for content that has at least one epistemic function – the function of informing.[20] We next consider the EENs connected with such environments.

Step 2: *Determine whether the epistemic functions are consistent with the satisfaction of the general EENs.*

We've now identified two broad functions for these types of environments: (1) to provide a platform for long-form performative content, and (2) to provide a platform for long-form informative content. In step 3, we argue that function (2) gives rise to epistemic norms that govern the platform and that allow us to evaluate the epistemic health of the platform. But before making that argument, it's worth considering whether these epistemic functions are at odds with any general EENs.

So, what are the general epistemic norms that govern social environments, and do these platforms run counter to those norms? Rather than trying to answer this immense and difficult question, we argue instead that, on any reasonable view of the general EENs, they allow the building of social environments with the epistemic functions that we consider here.

[19]For examples of entertainment and education coming together in natural history museums and nature documentaries, see Griffiths (2008) and Louson (2018). As those authors point out, informational functions and performative functions in these contexts are often entangled in various ways. On our view, when a piece of content has both functions, standards for both will apply to it. That said, our focus in what follows will be on creating platforms suitable for fulfilling the informational function.

[20]There may be more specific epistemic functions that specific types of long-form informational content have as well. For example, what it is to be a DIY video may essentially involve attempting to enable people who watched it to do (roughly) what the person was instructing them to do (fix a leaky faucet or play a piece of music, e.g.). To be a streamed lecture on a topic x, on the other hand, it is not required that the lecturer aims to enable the audience to give lectures on their own on x at the end of it. These functions may play a role in determining further conditions on epistemic health. But for our purposes, we'll just focus on the broad function of information transmission.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

More specifically, we contend that no plausible epistemic norms would rule out setting up environments that platform informative long-form content. Of course, an environment might be bad at providing such content, but there is no general EEN that *forbids* such an environment. Many paradigm cases of fruitful epistemic interactions involve exactly these types of content, such as streamed lectures and audio books that – when done well – often provide great epistemic benefits for the watcher or listener.

Of course, environments close to those we discuss here might violate general EENs if they provide a platform for *mis*informative content. For example, platforms that have the function of hosting and spreading misleading content may run afoul of general EENs (imagine a flat earth podcast network which requires hosts to affirm and spread flat earth content). And individual design and functionality choices on a platform may result in a platform that runs afoul of them as well. For example, it may be that a platform has a stated function that does not violate any general EENs, but in a designer's attempt to create a platform that executes the relevant function they may create something that – in practice – has a function that is in violation of a general EEN. This latter concern should give platform designers reason to keep reflecting on general EENs even for the types of environments we consider here.

Step 3: *Determine the specific EENs that govern this kind of environment.*

We are focused on environments that provide people with a platform for informational content, and this function straightforwardly generates specific standards for epistemic evaluation.

Intuitively, informational content is bad when it fails to inform, as seen by reflection on what counts as failure for a creator or a consumer of informative content. Start with the former. Imagine that a medical doctor in their official capacity attempts to inform a patient about an illness they have but misremembers the details and so misinforms their patient. In this scenario, they have failed to achieve their ends. When one aims to inform, to misinform is to fail.[21]

Similarly, people consume informational content because they want to figure out the truth regarding some question(s) so that they may either fulfill their curiosity or meet their (non-epistemic) ends. One might read Substack articles and watch YouTube videos on medical treatments for an ailment *because* one wants to figure out how to treat that ailment. But one will gain very little from that content if it's filled with inaccurate information – if it misinforms. Fulfilling one's aims in consuming informational content requires that the content *actually* informs.[22]

Thus, there is a standard by which we can evaluate informative content: namely, whether it informs. Moreover, there are many ways of failing to meet this standard; it is not vacuously satisfied. To see this, consider an ideal case of information sharing: a person (or persons) with knowledge of what they're sharing accurately conveys what they know to a person who has reason to trust them and who comes to believe the

[21]Those creating informative (or informative-like) content may not always aim to inform. They may aim to deceive or to entertain, for example. But to the extent that they succeed in deceiving, they will fail to inform.

[22]Sometimes consumers will have other aims when consuming informative content. For example, one might consume informative content for enjoyment. And misinformation can certainly entertain! But when one consumes the content for the sake of the information, misinforming will foil one's goals.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

shared information. There are (at least) five ways a case of information sharing can fail to meet this ideal: (1) the sharer doesn't know the relevant information, (2) the sharer doesn't accurately share what they know, (3) the consumer doesn't have reason to trust the sharer, (4) the consumer has reason to trust the sharer but nevertheless fails to come to believe what's been shared, or (5) the consumer doesn't understand what's being shared. We now must convert this ideal (and ways to fall short) into specific EENs for these environments.[23]

In the simplest cases of information sharing, there are three loci of evaluation: the sharer, the consumer, and the environment in which sharing takes place.[24] It is thus worth noting that failure to satisfy the ideal might be due to one of the first two loci, rather than an unhealthy environment. For analogy, consider a school. There are many ways that students might fail to learn at school, and only some of those will be because of an epistemically unhealthy school environment. For example, if a school doesn't credential its teachers or blasts Sweet Home Alabama through the intercom all day, the learning environment will be sub-par. But many ways that students might fail are not properly attributable to the environment. Even in ideal environments, students will daydream, doodle, and goof off, and credentialed teachers will have bad days. These things may prevent students from learning, but they are not attributable to the environment. Instead, what matters when evaluating the school environment is whether it is conducive to the transmission of the relevant information and skills.

Similarly, when epistemically evaluating an information-sharing platform, what matters is whether the *platform* is conducive to meeting the ideal of information sharing, even if the sharers or consumers fail in other ways.[25] This is *a* specific EEN that governs information-sharing platforms. And, thus, the more conducive to the ideal such platforms are, the epistemically healthier they will be. Further, as we argued at the end of section 1, EENs generate obligations for people to act. Thus, those that design, build, and maintain social media platforms plausibly have obligations to create and maintain an environment that is conducive to meeting that ideal (even if information is not necessarily successfully shared on every occasion).[26]

In the same way that there are many permissible ways to set up a school, there are surely many permissible ways for platforms to satisfy these obligations. Further, these obligations may be mitigated based on the difficulty of satisfying them.[27] And finally,

[23]One may wonder why we reference ideals in such applied contexts. We have offered a partial defense of this in our introduction to this step. We are also sympathetic to Carr's (2022) discussions of the use and importance of ideal epistemic theorizing – even for non-ideal contexts. Though also see McKenna (2023) for concerns with some types of ideal theorizing.

[24]Compare with Frost-Arnold (2014: 65) who claims "any problem of deception can be approached in two ways: (i) by focusing on the speaker and attempting to make her more honest, or (ii) by focusing on the hearer and attempting to shield her from dishonesty […] or increase her abilities to detect and reject falsehoods" (Frost-Arnold 2014).

[25]Of course, one must be a bit careful since the environment might attract sharers or consumers who are less likely to act appropriately. For simplicity, we leave aside those second-order effects.

[26]We view ourselves as developing the normative architecture for evaluating platforms epistemically. Determining how to ensure platforms meet those obligations is an extremely important question – for example, whether should we leave them unenforced, leave enforcement to the public to enforce by choosing to stay or leave platforms (or other forms of crowd pressure), or by attempting to institute some legal mechanisms for enforcement. We leave these important questions unanswered here.

[27]There are a few reasons this might hold. First, one might think that such requirements simply can't be too demanding. For analogy, in moral theory, some authors think that moral requirements can't be too demanding (this is commonly levied as an objection to certain versions of maximizing consequentialism

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

there may be competing obligations that platforms must satisfy. That said, we contend that most existing platforms have clearly not done enough to make their environments conducive to meeting the ideal of information sharing.[28] In particular, we identify two key features of epistemically healthy information-sharing environments that many information-sharing platforms lack (to varying degrees). In step 4, we argue that there are relatively simple ways to implement these features.

First, a key feature of healthy information-sharing environments is that content consumers and creators understand the goal/function of the environment. Consider the situation from the content consumer's perspective. If one is scrolling The Onion (a popular satirical news site) or a website they know to be a fake news site, then they will not be disposed to believe what they read. Perhaps they read The Onion to laugh, or the fake news site to learn about current conspiracy theories. But they don't consume these sources in order to learn about what the authors purport to be reporting on. From this, we can conclude that if content consumers believe a platform is a fake news or satirical site, then they won't come there to learn. But current environments make this difficult to determine.

Currently, most long-form content platforms include a mix of informative and performative content, real and fake news, authentic teachers and snake oil salespeople. And there is little that distinguishes these distinct types of content for consumers except the costless attestation of content creators.[29] Anyone can set up a YouTube, Substack, or Apple Podcast account, claim to be a lawyer, and begin offering legal advice. And outside of a few special circumstances, platforms typically sit back and leave to the consumer the very important work of determining: (1) which content is intended to be informative (rather than performative), and (2) which content is actually likely to inform. Neither of these tasks is easy to accomplish.

From the content consumer's perspective, this is a difficult position. There is valuable information on these platforms that can improve one's life. But there's also an enormous quantity of junk that can hinder one's attempts to improve one's life, or even make it worse (e.g., if you come to believe medical misinformation or attempt a dangerous DIY project on the basis of bad advice). And, as we'll argue next, consumers often aren't in a good position to answer questions (1) and (2) on their own.

Of course, in some cases, answering (1) is straightforward. Content consumers aren't likely to confuse a SpongeBob styled remix of a Dua Lipa song as informative content, for example. But things aren't always so easy.[30] Specifically, things are more difficult when it comes to content that is intended to appear (to varying degrees) like

(Kagan 1984; Wolf 1982)). One might think something similar applies to these kinds of norms. Second, one might appeal to a sort of consequentialist reasoning. If meeting the constraints is particularly difficult, it will take extensive resources on the part of the platform. This will necessarily divert finite resources from other ends. And, if particularly demanding, may not be possible to meet while remaining solvent. In these scenarios, one may find other moral or epistemic ends more pressing. Further, one might conclude that the existence of the platform is valuable enough (even as it is) to justify it, even if the only way for it to exist is in an epistemically unhealthy form.

[28]In fact, in some cases, they have intentionally made them epistemically unhealthy. See, for example, the list of common deceptive web designs provided by Brignull *et al.* (n.d.).

[29]One interesting exception is how YouTube has handled medical content in response to the spread of Covid-19 misinformation on the platform (Currin 2021; Google 2023).

[30]Consumers of testimony also face problems with determining the content of a testifier's testimony, for more on this see Goldberg (2007), Peet (2016), and Davies (2019).

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

informative content: e.g., deceptive content (Rini 2017), satirical content, and bullshit (Frankfurt 2005). Consider, for example, deceptive content.[31]

The most straightforward way to distinguish between informative content and deceptive content is to know the answers. For example, if you come across an article purporting to be news about a major earthquake in your hometown, but you know no such event occurred, you'll be able to deduce that the author is either aiming to deceive or made a serious mistake. But the problem is that we typically consume informational content precisely because we do *not* know the corresponding information. Content consumers are in a position to determine authentic informative content only when they have no (informational) need to consume the content.[32]

Other methods of determining authenticity are also typically unavailable to content consumers. For example, personal relationships with content creators would (to some extent) allow consumers to build trust, determine motivations, and otherwise assess the reliability of the sharer. But it's rare that content consumers have personal relationships with content creators, whether direct or indirect.

There are, of course, other methods of determining authenticity, but they tend to be time intensive for an individual to accomplish – the sort of thing investigative reporters get paid to do. One must determine that: credentials are authentic; verifiable claims are verified; the sharer has a track record of being reliable and honest; they lack connections to institutions or networks known to produce or traffic deceptive content; and many other related tasks. While these steps are all effective in identifying actual informative content, they are also quite costly for individual content consumers, particularly given the large quantity of informative content that people consume from a wide range of sources. Even if a single person could do this work in individual cases, no one has the time to do this work for each source of informative content they encounter each day.

One might suggest that individuals look for surface-level indicators of authenticity, like claims to credentials, professional or institutional seals, citations of other work, etc. But these markers are easy to mimic, and even the best of us often fall for the mimicry. For example, Wineburg and McGrew (2019) asked Stanford undergraduates and Stanford professors of history to distinguish between posts written by a legitimate medical organization and those written by a hate group, and both groups "fell victim to easily manipulated features of websites, such as official-looking logos and domain names," and ultimately struggled to successfully distinguish between the two posts (4).

Thus, if platforms do nothing, content consumers are left in a difficult position. Since they are typically unable to do the work of determining what content is deceptive from what content is not, they may be forced to treat content with one-size-fits-all responses. For example, they may: (A) trust all content, (B) reject all content, or (C) be agnostic about all content.

Each of these blanket responses leads to problems: (A) results in frequently believing deceptive content; (B) and (C) will result in failing to believe important information that could improve one's life.

[31]But problems with bullshitters provide similar challenges (Robinson 2022).

[32]This is partly inspired by arguments Cholbi (2007) has made about morality, although whether his argument succeed depends on different considerations. Specifically, he has argued that moral experts (if they exist) would not need to recognize moral experts but would also be the only people suitably positioned to do so.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

Anecdotally, these seem to be common responses to information gathered from social media. For example, people are often dismissive if one claims to have learned something on social media: "you can't trust anything you see on there" (B); or people often throw up their hands about important issues: "to be honest, I don't know who to believe!" (C).

We contend that these failures are not typically the responsibility of the individual consumer, as individuals do not have the time or resources to do the relevant work. And while some of the fault surely lies with those who create and spread deceptive content, we further contend that the environments (i.e., the platforms) also bear some epistemic responsibility.[33] To the extent that the above problems arise from the platform design (and not only malicious sharers), the platform is failing to create an environment conducive to the ideal of information sharing.[34]

Concretely, we contend that these types of environments are subject to a specific EEN to ensure that informational content is clearly demarcated. Creators need to know where and when it's appropriate to engage in satire or use deceptive tools (which are sometimes legitimate), and where and when it's inappropriate. And consumers need to know what content they can safely take to be non-deceptive. There are many ways of realizing this (see step 4), but this feature must be included to have an epistemically healthy information-sharing platform.[35]

Even if content consumers can determine that a content creator is genuinely attempting to inform, they must also determine whether the person is likely to know what they're talking about (i.e., problem (2) from earlier). And this may be an even more difficult challenge.

The proliferation of knowledge and the fact of hyperspecialization mean that individual people can't be experts in every domain that matters to their lives (see, e.g., Millgram 2015). One cannot be all of: computer scientist, geologist, bridge engineer, urologist, mechanic, plumber, chef, cardiologist, economist, etc. But our plight is worse than this common observation. On our own (without institutional help such as credentials), *we're not even in a position to recognize who the experts are* in most of the domains that matter.[36] The advantages of division of cognitive labor depend on people being able to learn from the *right* people when they have to make a decision outside of their own expertise.[37]

[33]See Buzzell and Rini (2023) for a closely related argument.

[34]We do not mean to imply that people cannot do better. We are very encouraged by the work of Caulfield and Wineburg (2023) (amongst other researchers) to create and test *Lateral Reading* methods. Methods that involve leaving platforms and following simple heuristics to evaluate information credibility. But that people *can* do better in current information environments does not imply that platforms do not also have obligations to improve. Further, in current environments people often do extremely poorly (Breakstone *et al.* 2021). Solving these problems effectively will likely take interventions at multiple levels.

[35]For those concerned about whether it's always possible to demarcate performative and informative content, this suggestion would let content creators and consumers know where and when fulfilling the function of informative content for content they produced took priority. On the other parts of the platform (if they exist), creators may have more leeway and freedom to find their preferred balance between entertainment and informational functions.

[36]Very roughly, we understand expertise as a success term. To be an expert in X is to have knowledge, skill, or ability in X. Experts in plumbing know how to fix plumbing issues, and can reliably do so. Experts in category theory know the truths of category theory (and avoid believing false propositions in this field). There's much more to be said here, but we hope this minimal account is sufficient for our purposes.

[37]For more discussion of novices dependence on experts and their (in many cases poor) likelihood of being able to recognize them, see Hardwig (1985), Goldman (2001), Anderson (2011), and McKenna (2023).

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

Experts are often capable of recognizing whether a putative expert knows what they're talking about by comparing what the putative expert says to what they (the expert) know. But non-experts do not have this luxury. By definition, they don't have expert knowledge.[38] So how can they recognize experts in online settings?

A natural place to start is to check for credentials, but social media platforms for long-form informational content typically do not provide such information, at least not in an easily verifiable form. Consumers will encounter a wide range of content sources, including credentialed experts who claim they're credentialed experts; credentialed experts who don't claim to be credentialed experts; those who claim to be credentialed experts but actually lack credentials (or whose credentials are in a different domain than the one they're creating informational content about); those who don't have any credentials but are nonetheless experts; and many other types. Since most long-form content platforms don't verify credentials (for most topics), content consumers are left to perform the required investigation. This is certainly an achievable task, but it's often time consuming, clunky, and – pretty clearly – not something most long-form informational content consumers actually take the time to do.

Alternatively, non-experts can recognize experts by examining their track records. Consider this example. You and your friend are watching Wimbledon. During the tournament, you might disagree over whether various shots landed in. One effective way to solve your dispute would be to take a close look at the slow-motion replay. If you do this frequently enough, you may discover that whenever you have a disagreement, your friend is correct; your friend has a better track record than you do. And this discovery would make it reasonable for you to conclude that your friend was an expert with respect to you when it came to judging tennis shots in real time (since they have a higher propensity for forming accurate credences when faced with novel problems in the domain).[39]

But notice that assessing track records like this will often be challenging for non-experts. First, in online contexts, people purporting to be experts sharing information often do not have easily accessible track records. They may be publishing information from a new account (so there is no online track record), or they may delete old posts (distorting their available online track record), or they may post at multiple sites, or they may have recently changed what site they post at.

Second, even if there is an available track record, it will often be time consuming for non-experts to gather and assess. Imagine, for example, that you wish to evaluate the track record of a person who has been making NBA draft projections over the last 10 years. This will entail watching multiple videos, reading multiple articles, and/or listening to multiple podcasts just to determine what their previous claims were. Then, one must actually assess their projections. But projections of basketball players (like many projections) are probabilistic. The purported draft expert may claim that a player has a 5% likelihood of superstardom, 40% likelihood of stardom, a 40% likelihood of being a role player, and a 15% player of failing to make it in the league. The non-expert then will have to evaluate (1) how good the player ended up being (a task that's often

[38]Nguyen (2021*b*) discusses the ways that experts must simplify and distort their work in certain cases in order to make it assessable by the public.

[39]This type of evidence is commonly referred to as "Independent track record evidence." Because it allows a non-expert to evaluate the track record of an expert by using judgments they (the non-expert) make in an independent domain. In this case, the non-expert uses their judgments of slow-motion objects, to evaluate an expert's judgments of objects in real-time. For more on this, see McGrath (2009) and Nguyen (2020).

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

difficult and time consuming), and (2) how to credit the purported expert given that projection.[40]

Third, in many domains, non-experts will not be able to assess the claims on their own. For example, if a non-expert is trying to determine if an account that is posting information about Covid-19 is reliable, they're unlikely to be in a position to evaluate those claims. Instead, they'll have to rely on the word of other putative experts to evaluate those claims.

In the end, the most promising strategy is probably the one we discussed first – checking for credentials. Empirical evidence suggests its value (see Caulfield and Wineburg 2023 for discussion), and it is significantly less time consuming than the other methods. That being said, it does require leaving the relevant social media platform and doing research to determine that the relevant account is appropriately credentialed, and people (currently) struggle at these tasks.[41]

Needless to say, recognizing expertise in even singular cases is often a demanding task for non-experts. But people don't just consume one-off bits of informational content online. Many of us consume large quantities of information from a wide variety of sources every day. Even if it's strictly speaking possible for non-experts (which most of us are in most domains) to recognize experts in each individual context, it is unreasonable to expect people to do this work for the wide variety of putative experts they rely on. If a platform chooses to rely on its users to do this work, they're bound to have an information-sharing platform where information consumers frequently trust those they shouldn't. For this reason, we maintain that a second key feature of healthy information-sharing environments is a mechanism that helps content consumers recognize experts.[42]

The absence of such a mechanism in most online environments stands in sharp contrast with the many offline institutions that train, credential, and monitor experts in professions such as medicine, cosmetology, auto mechanics, engineering, plumbing, accounting, law, truck driving, aviation, etc. Everywhere we turn in the brick-and-mortar world, there are sophisticated institutions that help us recognize experts. While I might be skeptical if someone tells me they heard a piece of medical or legal advice on YouTube, I'm not skeptical if they tell me they heard it at their doctor's office or from their lawyer. Platforms have an epistemic obligation to help close this gap. Our proposal is to bring these institutional markers that already exist offline, online.

There are two worries one might have about our arguments. First, one might point to cases where institutions (or those credentialed by institutions) have failed. Are we

[40]For many of the problems discussed here, the degree of difficulty will be on a sliding scale, with problems being easier as one has more expertise. For example, the more one knows about basketball, the easier it will be to make evaluations about the quality of players.

[41]Between 2019 and 2020 a group of researchers at Stanford surveyed 3,446 high school students who they asked to evaluate a series of websites. Asked to evaluate a site claiming to "disseminate factual reports" about climate science, 96% never discovered the site's connection to the fossil fuel industry (Breakstone *et al.* 2021). See also Lucassen and Schraagen (2011) for more evidence that non-experts tend to focus on surface-level features when evaluating information reliability.

[42]These problems might be somewhat mitigated if people slowed their content consumption. But there are clear downsides to this. First, it's unlikely. Second, it would reduce the quantity of accurate and helpful information that people consume. And, third, social media platforms don't typically want us to reduce our content consumption, as this tends to be at odds with their business models. This is (at least in part) why platforms are often hesitant to add friction to problematic content, even when they know their platform will be epistemically healthier if they do.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

really safe relying on institutions? Second, one might wonder whether there are other possible routes we might take. Do we really need to trust institutions? Two cases might make these worries particularly salient. First, Wikipedia is extraordinarily reliable, but its authors are often laypeople and non-experts. Second, expert fact checkers cannot be experts in all the areas in which they check information, but they are nonetheless extremely accurate.

We agree that many credentialing institutions have had failures (in some cases very high-profile ones). But we also contend that reliance on (healthy) institutions is still the best way of proceeding, and the cases of Wikipedia and professional fact checkers actually help make this point.

In our view, what explains Wikipedia's reliability is that laypeople (and in some cases experts) gather and convey information that has been produced, discovered, or curated by institutions or by experts who have been vetted by institutions.[43] Wikipedia in fact provides guidance to its authors along these lines. For example, Wikipedia tells its authors that "If available, academic and peer-reviewed publications are usually the most reliable sources on topics such as history, medicine, and science" (Wikipedia: Verifiability 2023). Of course, academic journals are a type of institution, and the vast majority of authors who publish in such journals have been trained and credentialed by universities, another institution. So, the success of Wikipedia in these domains requires and relies on the success of those institutions. More generally, Wikipedia tells its authors that "The best sources have a professional structure for checking or analyzing facts, legal issues, evidence, and arguments. The greater the degree of scrutiny given to these issues, the more reliable the source" (Ibid). Finally, Wikipedia tells its authors that they should not engage in any original research (Wikipedia: No Original Research 2024). In other words, Wikipedia tells its authors that they should not rely on their own judgment but that they should source their information.[44] And, further, that the best sources of information are those that are produced and vetted by epistemically healthy institutions. Similar principles explain the reliability of professional fact checkers (Caulfield and Wineburg 2023; Wineburg and McGrew 2019).

As we see things, Wikipedia is a reliable epistemic institution in its own right. And people do well by trusting it, because it does so well at ensuring its authors and information curators defer to other reliable institutions that do the work of discovery and knowledge production. So, rather than being a counterexample to the importance of institutions, it emphasizes their importance.

One way of summing our view is as follows. Institutional markers make it possible for individual non-experts (like those who often write Wikipedia, professionally fact check, or even just scroll YouTube) to determine what information to trust. Thus, we ought to build those institutional features into platforms in order to reduce the time and labor for the average person to search out and find that information. This doesn't imply that institutions never fail. But our response to such failures should be to strengthen and improve them, because so much of our epistemic success relies on their healthy functioning.

One final worry that one might have with our proposal is that it draws too sharp a distinction between informative and performative content. To make this worry clear,

[43]For more general discussion of the epistemology of Wikipedia, see Fallis (2008) and Tollefsen (2012).

[44]Of course, this does not mean Wikipedia always lives up to these principles or is without other flaws. See, for example, Ford (2022) for a particularly important discussion of the ways Wikipedia does more than merely reflect what experts have discovered, and for a discussion of other issues such as bias on Wikipedia.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

one might consider a satirical news content creator along the lines of Stephen Colbert on The Colbert Report. How should platforms treat such content?

In response, we note two things. First, we think the most important step for platforms to take is to clearly demarcate the highest quality information. Thus, while people may produce and consult whatever they please for information and entertainment, information that has been vetted and produced by those who are in a position to know about the relevant topic should be in some way marked. Second, there are many approaches platforms can take to improving along the two dimensions we've identified. And individual platform decisions will likely depend on the overall functions of the platforms. We will discuss this in greater detail in the next section, but an educational platform, for example, may have reason to create an environment where only high-quality information is shared on the platform. Content like satirical news would not typically meet these standards. On the other hand, general interest platforms like YouTube or TikTok may mark high-quality information while allowing a range of other types of content to be shared on their platforms. Satirical news would be perfectly acceptable, but should not be marked as meeting the high standards reserved for information intended to be informational and vetted and produced by experts.

We've now argued that social media companies have a responsibility to ensure that information-sharing environments have two key features for epistemic health: (1) they clearly demarcate informative content from non-informative content, and (2) they create and maintain mechanisms that help information consumers recognize experts.

The lack of these two mechanisms helps to explain two facts that, when taken together, seem puzzling: (1) there is a large quantity of valuable long-form informative content online that (if believed) could make a significant difference to peoples' lives, and (2) people are often skeptical and dismissive when someone claims to have learned something online. We suggest that people aren't skeptical that there is high-quality informative content online, but rather are skeptical about others' (and, in some cases, their own) ability to recognize that content either directly or indirectly (by recognizing expertise).

Step 4: *Determine the design and functionality constraints that will (likely) foster epistemic health for the environment, where it is possible that no system could simultaneously satisfy all of the epistemic functions.*

Present-day social media platforms capitalize on the fact that many of the same features and functionalities that are conducive for performative content are conducive for informative content. YouTube includes DIY videos and movie trailers in the same feed. A similar overlap is admittedly found in some offline environments. Consider the functional overlap between theaters and lecture halls, for example. Both ideally allow those on stage to be clearly seen and heard by the audience by making the stage the focal point with the way the room is designed (seats facing the stage, lightning is brighter on the stage, voices on the stage are amplified while others are not, etc.). Both tend to discourage interaction between audience members during the performance or lecture (the audience's seats don't face each other, and there are strong norms against speaking during a performance or lecture). And so on.

But while the base design and functionality constraints are the same, the end products should not be. As we've argued, performative and informative content have different primary functions and thus have different primary norms for evaluation. To that end, it will be important for developers to treat performative and (at least some)

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

informative content differently, even if they coexist on the same platform. Specifically, we've argued that platforms should (1) clearly demarcate informative content from non-informative content, and (2) create and maintain a mechanism that helps would-be information consumers recognize experts. And we think it is especially important that this be done for the highest quality content.

The most straightforward way of accomplishing (1) is to: (a) have content creators indicate what type of content they intend on making, and (b) mark this for content consumers. Regarding (a), providing a box for content creators to check whether they intend their content to be primarily informational before uploading should suffice. Regarding (b), building into the user interface some way of marking this selection should suffice. There are many number of ways of achieving this: having separate sections on the platform for the two types of content, color coding the two types of content, etc.[45]

Marking experts may initially seem more complex. After all, recognizing experts (as we've argued) is time consuming and demanding work, and so social media companies have good reason to avoid this effort. Further, platforms often want to avoid wading into tricky politically charged debates, be accused of taking sides, or risk picking the wrong experts. Similarly, credentialed experts sometimes get things wrong, and content consumers may reasonably want to pursue the advice of those who aren't credentialed experts at times. Even when it's unreasonable, one might think that people should be permitted to pursue the advice of whoever they'd like.

Despite these concerns, there are a range of solutions that platforms can implement – many of which do not run afoul of *any* of these concerns. The key to all of the solutions is for platforms to offload the work of recognizing experts onto credentialing institutions that already exist. There's no need for platforms to reinvent the wheel here. There are reliable, trustworthy, and high-functioning credentialing institutions in a wide variety of domains. Thus, platforms do not need to do the work of recognizing experts, they only need to verify who possesses relevant credentials.

Once they've done this, there are a range of mechanisms platforms could create to help would-be information consumers recognize experts. On one extreme, there's what we might call the "university model." Platforms following this model may only offer a platform to credentialed experts. This might be ideal for educational platforms. On the other extreme, platforms could create some way of marking what informative content was produced by experts – for example, by using a display feature similar to Twitter's check mark. This sort of platform would help users easily recognize experts but give them the freedom to consult those who aren't credentialed if they wish. This model would put users in a situation similar to one they're often in offline. If my car breaks down, I can consult my mechanic who I have good reason to believe knows what they're talking about. But I can also choose to consult my neighbor, in which case it's up to me to determine whether they know what they're talking about, and whose advice I take at my own risk.

There are, of course, a wide variety of models in between these two that platforms could adopt. For example, a platform could prioritize verified experts in search, a platform could have different comment and reaction policies for credentialed experts, a

[45]We emphasize that an indication that some content is informative does not thereby imply that it is true, accurate, or correct. False content can nonetheless be intended to be informative. As noted in step 3, the key is that consumers know the relevant success conditions for the content (i.e., whether truth or accuracy is relevant).

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

platform could restrict sharing of non-expert content, and so on. Regardless of the exact implementation, it is also important for a platform to ensure that sharers are marked as experts only for content that is actually connected to their expertise. For instance, medical advice from a licensed physician ought to be marked as from an expert, but legal advice from the same individual ought not be marked in this way.[46]

All of these features should be relatively straightforward to implement for social media platforms. Adding a check mark to work produced by credentialed content creators, for example, is very simple. Similarly, determining who is credentialed will be relatively straightforward precisely because the necessary institutions, structures, and processes already exist in the offline world. Social media platforms can adopt a policy similar to Twitter's authenticity check to determine if the putative expert is in fact credentialed with the organization they claim to be credentialed with.

## 4. Big picture thoughts and conclusions

When the internet was younger, there was considerable optimism that it would democratize knowledge. Before the internet, information was primarily housed in libraries, in minds, and in institutions. But these are spread out across the world and often difficult to access for the average person. The internet, however, could easily contain all of this information and be accessible to the average person.[47] The supposed payoffs of this increased access would be improved decision making, greater intellectual autonomy, and a more informed public.

In many ways, those aims have succeeded. There are more than 5 billion internet users worldwide, and "anything that brain of yours can think of can be found" there.[48] But as with so many statements about the internet, the optimistic version is paired with a pessimistic side: the democratization of knowledge has not been an unalloyed good. While people increasingly rely on the internet – and especially social media – for news and information, they often do not trust the information they're consuming.[49] If someone asks you why you believe something, and you respond that you saw it on TikTok or YouTube, they're likely to scoff. Instead of becoming trustworthy sources of information, like the old more inaccessible institutions, they have become places where people (reasonably!) are slow to believe.

This somewhat gloomy reality is reflected in wider discussions of the internet and social media more specifically. Instead of cheery celebrations of a more informed public enabled by the power of the internet, authors are more likely to bemoan the spread of misinformation and conspiracy theories, the rise of polarization, decreasing attention spans, the harms of echo chambers, and so on.

So, what went wrong? Why has the promising vision of the early internet failed to materialize? Surely there's no single or simple answer to these questions. But in our view, a significant part of the story is the fact that, in democratizing the internet, we left behind many of the institutional features of the traditional but inaccessible institutions that made them trustworthy in the first place. Doing so made information accessible, but at the cost of making it difficult to determine what content was intended to be

[46]Considering Frost-Arnold's (2014) reasonable concerns about the loss of anonymity, it is worth noting that many of these models permit large swaths of anonymous content (from those who are not providing information in their capacity as experts).

[47]Consider, for example, Google's early efforts to upload and cross reference *every* book (Google n.d.).

[48]From Bo Burnham (2022).

[49]See, for example, Ray (2021) and Shearer and Mitchell (2021).

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

informational and what content was trustworthy. If we are to make our social media platforms – and the internet more widely – epistemically healthy, we must begin to reintegrate some of these lost institutional features.

In this paper, we've provided a framework for evaluating and making epistemic design decisions for social media platforms. Further, we applied this framework to a set of social media platforms – those that platform long-form informational content – and argued that many of these platforms are epistemically unhealthy. We argued that this implies that they have epistemic obligations to alter their platforms, and we provided concrete suggestions for implementing these changes.

We think the normative reasons should be sufficient for platforms to act on our suggestions. But there is, of course, the further question about whether those in charge of the relevant platforms will find this convincing. While we'd like to think the answer is yes, it's far from clear. One thing we will say, however, is that even if a platform was to reject the deeper normative implications of our view, it is hard to deny the pull of the specific EENs we've been discussing. After all, YouTube, for example, does desire to create and maintain a platform for the sharing of long-form informational content. There are clearly better and worse ways of satisfying their aims. Thus, even if they reject the more fundamental picture, we hope they would recognize that they have instrumental reason to do what they've set out to do *well*.[50]

## Statements and declarations

All authors contributed to all aspects of this paper. This research was partially supported by a grant to D. D. from the Office of Naval Research. The authors have no financial or proprietary interests in any material discussed in this article.

## References

**Amico-Korby D., Harrell M. and Danks D.** (2024). 'Environmental Epistemology.' *Synthese* **203**(81), 1–25.

**Anderson E.** (2011). 'Democracy, Public Policy, and Lay Assessments of Scientific Testimony.' *Episteme* **8** (2), 144–64.

**anitaduran34**. (2021, July 28). *#fyp#backpainrelief #tiktokremedies #vaporub #remedies #follow #homeremedies-forback*. Retrieved from TikTok: https://www.tiktok.com/@anitaduran34/video/6990203144651246854?q=what%20helped%20me%20relieve%20back%20pain&t=1707513247235

**Basu R.** (2019). 'Radical Moral Encroachment: The Moral Stakes of Racist Beliefs.' *Philosophical Issues* **29** (1), 9–23.

**Bolinger R.** (2020). 'Varieties of Moral Encroachment.' *Philosophical Perspectives* **34**(1), 5–26.

**Breakstone J., Smith M., Wineburg S., Rapaport A., Carle J., Garland M. and Anna S.** (2021). 'Students' Civic Online Reasoning: A National Portrait.' *Educational Researcher*, **50**(8), 505–15.

**Brignull H., Leiser M., Santos C. and Doshi K.** (n.d.). *Deceptive Patterns*. Retrieved from https://www.deceptive.design/types

**Burnham, B.** (2022). *Welcome to the Internet (from "Inside")*. Retrieved from YouTube: https://www.youtube.com/watch?v=k1BneeJTDcU

**Buzzell A. and Rini R.** (2023). 'Doing Your Own Research and Other Acts of Epistemic Superheroism.' *Philosophical Psychology* **36**(5), 906–30.

**Carr J.R.** (2022). 'Why Ideal Epistemology?' *Mind,* **131**(524), 1131–62.

**Caulfield M. and Wineburg S.** (2023). *Verified: How to Think Straight, Get Duped Less, and Make Better Decisions about What to Believe Online*. Chicago and London: The University of Chicago Press.

[50]We are grateful to Jennifer Rose Carr for her many insightful comments on a previous draft of this paper. We are also grateful to David Brink, Craig Callendar, and an anonymous reviewer for their thoughtful feedback.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

**Cholbi M.** (2007). 'Moral Expertise and the Credentials Problem.' *Ethical Theory and Moral Practice* **10**(4), 3232–334.
**Currin G.** (2021). 'YouTube's Plan to Showcase Credible Health Information Is Flawed, Experts Warn.' *Scientific American*.
**Davia C.** (2022). 'Moral Encroachment and the Ideal of Unified Agency.' *Philosophical Explorations*, **26**(2), 179–96.
**Davies A.** (2019). 'Testimonial Knowledge and Context-Sensitivity: A New Diagnosis of the Threat.' *Acta Analytica* **1**(34), 53–69.
**Fallis D.** (2008). 'Towards and Epistemology of Wikipedia.' *Journal of the American Society for Information Science and Technology* **59**(10), 1662–74.
**Feldman R.** (2000). 'The Ethics of Belief.' *Philosophy and Phenomenological Research* **60**(3), 667–95.
**Ford H.** (2022). *Writing the Revolution: Wikipedia and the Survival of Facts in the Digital Age*. Cambridge: MIT.
**Frankfurt H.** (2005). *On Bullshit*. Princeton: Princeton University Press.
**Fricker M.** (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
**Frost-Arnold K.** (2014). 'Trustworthiness and Truth: The Epistemic Pitfalls of Internet Accountability.' *Episteme* **11**(1), 63–81.
**Goldberg S.** (2007). *Anti-Individualism: Mind and Language, Knowledge and Justification*. Cambridge: Cambridge University Press.
**Goldman A.I.** (1976). 'Discrimination and Perceptual Knowledge.' *The Journal of Philosophy*, **73**(20), 771–91.
**Goldman A.I.** (2001). 'Experts: Which Ones Should You Trust?' *Philosophy and Phenomenological Research* **63**(1), 85–110.
**Google**. (2023). *Get Info on Health-Related Content*. Retrieved May 2023, from YouTube Help: https://support.google.com/youtube/answer/9795167?hl=en
**Google**. (n.d.). *Google Books History*. Retrieved from Google Books: https://web.archive.org/web/20160206043510/http://books.google.com/googlebooks/about/history.html
**Griffiths A.** (2008). *Shivers Down Your Spine: Cinema, Museums, and the Immersive View*. New York, NY: Columbia University Press.
**Haraway D.** (1992). 'Otherworldly Conversations; Terran Topics; Local Terms.' *Science as Culture* **3**(1), 64–98.
**Harding S.G.** (1978). 'Four Contributions Values Can Make to the Objectivity of Social Science.' *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, **60**(4), 199–209.
**Hardwig J.** (1985). 'Epistemic Dependence.' *The Journal of Philosophy* **82**(7), 335–49.
**Harman G.** (1973). *Thought*. Princeton: Princeton University Press.
**Hooks B.** (1990). *Yearning: Race, Gender and Cultural Politics*. Boston: South End Press.
**Kagan S.** (1984). 'Does Consequentialism Demand Too Much? Recent Work on the Limits of Obligation.' *Philosophy & Public Affairs* **13**(3), 239–54.
**Louson E.** (2018). 'Taking Spectacle Seriously: Wildlife Film and the Legacy of Natural History Display.' *Science in Context*, **31**(1), 15–38.
**Lucassen T. and Schraagen J.M.** (2011). 'Factual Accuracy and Trust in Information: The Role of Expertise.' *Journal of the American Society for Information Science and Technology* **62**(7), 1232–42.
**McGrath S.** (2009). 'The Puzzle of Pure Moral Deference.' *Philosophical Perspectives* **23**(1), 321–44.
**McKenna R.** (2023). *Non-Ideal Epistemology*. Oxford: Oxford University Press.
**Miller B. and Record I.** (2017). 'Responsible Epistemic Technologies: A Social Epistemological Analysis of Autocompleted Web Search.' *New Media & Society* **19**(12), 1945–63.
**Millgram E.** (2015). *The Great Endarkenment: Philosophy for an Age of Specialization*. Oxford: Oxford University Press.
**Mills C.** (1997). *The Racial Contract*. Ithaca, NY: Cornell University Press.
**moorewellness**. (2023, April 18). https://www.tiktok.com/@physicaltherapysession/video/7273203962516491522?q=scoliosis%20spine%20curvature&t=1707513394993. Retrieved from TikTok: https://www.tiktok.com/@moorewellness/video/7223512668043332910?_r=1&_t=8jkTVQzj2yo
**Nguyen C.T.** (2020). 'Cognitive Islands and Runaway Echo Chambers: Problems for Epistemic Dependence on Experts.' *Synthese*, **197**, 2803–21.
**Nguyen C.T.** (2021*a*). 'The Seductions of Clarity.' *Royal Institute of Philosophy Supplement* **89**, 227–55.

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press

**Nguyen C.T.** (2021*b*). 'Transparency is Surveillance.' *Philosophy and Phenomenological Research* **105**(2), 331–61.
**O'Neil C.** (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books.
**Panko B.** (2017, August 23). *A Decade Ago, the Hashtag Reshaped the Internet*. Retrieved from Smithsonianmag.com: https://www.smithsonianmag.com/smart-news/decade-ago-hashtag-reshaped-internet-180964605/
**Peet A.** (2016). 'Testimony and the Epistemic Uncertainty of Interpretation.' *Philosophical Studies* **173**(2), 395–416.
**physicaltherapysession**. (2023, August 30). *Fix SCOLIOSIS Naturally!! – Functional Scoliosis is When the Spine Appears to be Curved, but the Apparent Curvature is Actually the Result of an Irregularity Elsewhere in the Body #scoliosis #scoliosischeck #scoliosisawareness #physicaltherapy #physicalth*. Retrieved from TikTok: https://www.tiktok.com/@physicaltherapysession/video/7273203962516491522?q=scoliosis%20spine%20curvature&t=1707513394993
**Putnam H.** (1992). 'Brains in a Vat.' In K. DeRose, T. Warfield, K. DeRose and T. Warfield (eds), *Skepticism: A Contemporary Reader*. Oxford: Oxford University Press.
**Ray J.** (2021, November 18). *Young People Rely on Social Media, but don't Trust it*. Retrieved from Gallup Blog: https://news.gallup.com/opinion/gallup/357446/young-people-rely-social-media-don-trust.aspx
**Record I. and Miller B.** (2022). 'People, Posts, and Platforms: Reducing the Spread of Online Toxicity by Contextualizing Content and Setting Norms.' *Asian Journal of Philosophy* **1**(2), 1–19.
**Rini R.** (2017). 'Fake News and Partisan Epistemology.' *Kennedy Institute of Ethics* **27**(2), 43–64.
**Robinson N.J.** (2022, December 29). 'We Live in the Age of the Bullshitter.' *Current Affairs*.
**Shearer E. and Mitchell A.** (2021, January 12). *News Use across Social Media Platforms in 2020*. Retrieved from Pew Research Center: https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/
**Simon J.** (2010). 'The Entanglement of Trust and Knowledge on the Web.' *Ethics and Information Technology* **12**, 343–55.
**Thaler R.H. and Sunstein R.C.** (2009). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York: Penguin Books.
**Tollefsen D.P.** (2012). 'Wikipedia and the Epistemology of Testimony.' *Episteme* **6**(1), 8–24.
**Wikipedia: No Original Research**. (2024, February 7). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Wikipedia:No_original_research#:~:text=However%2C%20Wikipedia%20editors%20must%20base,no%20reliable%2C%20published%20source%20exists
**Wikipedia: Verifiability**. (2023, December 19). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Wikipedia:Verifiability#Reliable_sources
**Wineburg S. and McGrew S.** (2019). 'Information, Lateral Reading and the Nature of Expertise: Reading Less and Learning More When Evaluating Digital.' *The Teachers College Record* **121**(11), 1–40.
**Wolf S.** (1982). 'Moral Saints.' *The Journal of Philosophy* **79**(8), 419–34.

**Dallas Amico-Korby** is a PhD candidate at UC San Diego. He works in epistemology, the philosophy of technology, and metaethics. Email: damico@ucsd.edu

**Maralee Harrell** studies how students can effectively learn critical thinking and writing skills. She is Teaching Professor of Philosophy and Public Health at UC San Diego, and the Editor-in-Chief of the journal *Teaching Philosophy*. Email: mharrell@ucsd.edu

**David Danks** both develops novel AI methods and examines the ethical and societal impacts of AI technologies. He is a Professor of Data Science & Philosophy at UC San Diego, and serves on multiple advisory boards, including the National AI Advisory Committee. Email: ddanks@ucsd.edu

**Cite this article:** Amico-Korby D. Harrell M. and Danks D (2025). "Building Epistemically Healthier Platforms." *Episteme* **22**, 745–767. https://doi.org/10.1017/epi.2024.22

https://doi.org/10.1017/epi.2024.22 Published online by Cambridge University Press