Neely Design Code for Social Media

Social media needs to be designed explicitly for the people and societies that it serves. Social interactions are essential to human flourishing, and relying solely on profit-focused business mechanisms to shape those interactions has led to societal harm. Content moderation alone, absent design changes that improve all interactions and not just those deemed to violate policies, will not be able to fix this harm.

USC Marshall's Neely Center for Ethical Leadership and Decision Making has consulted with numerous experts and practitioners to create a list of specific steps we can take to design social media systems to better serve society. We, who have signed this document, support the following nine minimum standards for companies that host online social interactions. These standards are meant to be specific and require no policy definitions, to complement other efforts that may state more general principles or may focus on policy defined harms. Broadly, they enable greater explicit user control, protect children through better defaults, improve incentives, and prevent small groups of users from manipulating and harming others. This is especially important given upcoming global elections set to take place in a world increasingly powerful AI.

These design codes are a minimum start that builds upon existing evidence. Most of us have other codes that we would like platform designers to abide by as well, beyond these minimum standards, such that these design codes do not represent every complaint that every signer has with social media platforms. Further, this document takes no position as to whether these steps should be adopted voluntarily or via app store policy, government policy, or consumer pressure. Some may be more important and readily implementable by larger platforms. Some may already be in place at some platforms. However, we all agree that these would be important positive steps for social media platforms to take and ask them to do so, to the best of their ability.

Neely Design Code for Social Media

	Design Code	Details	Supporting Argument	Goal
1	Allow users to easily and explicitly indicate content they do or do not want, and respect users' explicit preferences even if contradicted by users' engagement.	***Easily*** means that indicating unwanted content should be a top level action that requires no clicks to access. ***Explicitly*** means that the interface should clearly indicate that this is a subjective action with no required connection to formal platform policies. ***Respecting*** this preference means that users would not be shown content they indicate not wanting unless they explicitly seek it out (e.g. via search).	Platforms have identified many kinds of content (e.g. clickbait) that users engage with even though they do not value that content. Objectively defining that content has proven impossible, with "borderline content" that approximates gaining wide distribution. Lightweight ways for users to indicate their disapproval of such content can help reduce negative experiences on platforms.	Allow greater explicit user agency to improve subjective user experience.
2	Especially for conversations that users deem important or sensitive (e.g. about politics, religion, race, crime, health, finance, sexuality, etc.), replace engagement optimizations (e.g. view time, comments, shares, more ad distribution) with optimizations for user-perceived quality.	***Engagement*** refers to any behavior that does not indicate an explicit user preference, since users may spend time on, comment on, or share content that they actually dislike or find to be low quality. ***Optimizations*** refers to any increased algorithmic visibility or distribution due to the use of a particular signal, including in advertisements. ***Quality*** refers to user perceptions of quality, which can be measured in a variety of ways. ***Sensitive*** conversations refers to domains where quality information is especially important for user and societal well-being, including, but not limited to discussions of politics, religion, race, crime, health, finance, sexuality, and more.	Publishers and politicians have repeatedly indicated that they are incentivized toward misleading and divisive content due to engagement based algorithms. External studies have corroborated that observation. Removing that incentive has shown positive effects on content quality, misinformation, bullying, and graphic content. Greater content quality and less toxic content should also produce better norms that may lead to less production of low quality, toxic content.	Reduce bad experiences for users (e.g. sensational & divisive political discussions) and improve ecosystem of incentives for publishers (e.g. the financial incentive to spread divisive misinformation).

Neely Design Code for Social Media

	Design Code	Details	Supporting Argument	Goal
3	Prioritize public amplification to actors that varied users explicitly know and trust.	***Prioritize*** means that these actors would get greater relative distribution. ***Public amplification*** refers to increasing the visibility or distribution of messages to unknown others, often via algorithms. ***Varied*** refers to a set of users who have minimally different behaviors and histories online. For the purposes of this code, "varied" is meant as a low bar involving individuals who exhibit some degree of differential behavior, rather than requiring universal trust, which would unnecessarily stifle dissent. The goal is to make information spaces more robust to manipulation by extremely narrow groups. ***Trust*** refers to the subjective judgment of users, where platforms would have to be transparent about what signals they use to infer trust (e.g. a history of explicit positive interactions).	Many online harms have occurred as a result of motivated actors reaching large sets of strangers. Platforms have intermittently acknowledged this risk by removing capabilities for untrusted users during times of stress, but such definitions are often ad-hoc. In contrast, signals of reputation play an important part in reducing risk across domains, including in internal social media research. A safer ecosystem would reduce the risk of untrusted actors, requiring a positive explicit feedback history (reputation) from a broad set of other users before allowing users to affect people they do not already know. Signals of diverse approval have proven useful across domains in identifying actors that can be trusted.	Improve user experience since diverse approval is a good signal of quality. Reduce ecosystem vulnerability to manipulation by narrow groups. Improve ecosystem incentives.
4	Allow users to accessibly opt-out of revenue maximizing design features (e.g. optimizing for time spent, infinite scroll, auto-play) that encourage greater usage. Make this the default for minors. Offer all users tools to limit their platform usage.	***Revenue maximizing design features*** refers to optimizations that relate more to revenue and usage (e.g. time spent optimization, infinite scroll, auto-play) than to the well-being or experience of users. ***Minors*** refers to any user who a platform reasonably believes to be under 18. Identification of minors should be proportionate to potential risk and should be done in privacy safe ways (e.g. device based parental controls as described in #8).	Enables users to control design features that incentivize greater usage than what users want. This is especially important for children who are still developing their inhibition systems and so are more vulnerable to design features that target reward systems. Platforms have been implicated in myriad harms toward minors, including sexualization and reduced mental health. This is consistent with current Age Appropriate Design code, which requires risk assessments for design features that are likely to impact minors. A 2022 Pew study showed that 36% of teens say they use too much social media and that 54% say that it is hard to give up.	Enable greater user control and agency. Protect children, who are especially vulnerable to design features that target impulsive choices.

Neely Design Code for Social Media

	Design Code	Details	Supporting Argument	Goal
5	Provide transparent, sensible rate limits for new, untrusted users who access functionality that can be used to target or influence others.	**Rate limits** for **functionality that can be used to target others** refers to functionality that may be reasonable in small amounts (e.g. friend requests, messaging strangers, inviting others to groups/pages, viewing other people's profiles, commenting on public posts), but which create systemic risk when a small group of people use it a lot (e.g. to manipulate others, misuse other people's images, or to solicit others for bad ends). **Sensible** refers to what most people would use in everyday usage. **Transparent** disclosure of these rate limits should provide accountability as to what limits conform to public expectations.	Small groups of motivated actors have been instrumental in promoting content with societal risk, targeting others, and misusing other people's images and information. Most people have no need to reach the rate limits that platforms have for such functionality, and so limiting high usage to those who are trusted and have a demonstrated need creates a safer ecosystem for both individuals and society (e.g. in this break glass measure). Such functionality mirrors the on-ramping process that occurs in offline communities, where newcomers have to build trust to be entrusted with more responsibility.	Prevent bad actors from accessing information and potentially harming others. Improve the robustness of information ecosystem.
6	Enable privacy by default for any situation involving minors or when a significant number of users expect their information to be inaccessible.	**Default** refers to the product use case before a user takes any explicit action. **Significant number of users** can be defined transparently by platforms, but should include any significant minority (e.g. >10%). Businesses have an incentive to make information accessible as content to other users, even when users may have chosen otherwise. Users and their content should be presumed to be private, unless users explicitly desire their content to be public. In situations where the expectation is ambiguous, platforms should default visibility to private.	Users care about privacy, but often do not understand enough about their choices to provide consent, which is why it is important to anchor on what users would expect, rather than what users are willing to agree to. Allowing strangers scalable access to minors has proven dangerous online. Rather than allowing strangers to reach minors and excluding those that violate policies, platforms should accept the limited ability of policies to catch harmful behavior and err instead on the side of caution, where minors' information is private by default.	Prevent bad actors from accessing information and potentially harming others, especially children.

Neely Design Code for Social Media

	Design Code	Details	Supporting Argument	Goal
7	Prohibit the public distribution of sexually explicit content depicting individuals who do not explicitly provide permission.	***Sexually explicit content*** clearly includes non-consensual nudity, but can include any content that conforms to standard legal definitions of sexual explicitness that do not require nudity. Broadly, if a private person considers something sexually explicit and non-consensual, any claim within reason should be honored. This is not meant to include historical or educational images that are not primarily sexual in nature. ***Permission*** should be proactive and ***explicit***, such that the default should be that such content will not be publicly accessible unless a platform can verify consent was given in a clear manner by the person being portrayed.	Sexualized content has proven to be a persistent barrier to the full participation of women and vulnerable groups in online life. The growing capabilities of AI systems means that it is even easier to generate sexual content involving non-consenting individuals. Given advances in facial identification, it should be possible for platforms to stop any non-consensual public distribution of sexualized content by limiting such distribution to individuals who proactively consented to have their sexualized images made public.	Prevent non-consensual sexual imagery from distribution.
8	Enable device-based parental controls with appropriate defaults for minors for any online activity that a broad group of parents considers risky.	***Device-based parental control*** can be done in a privacy safe way, by designating specific devices as belonging to minors, building on OS providers, who have experience linking permissions via Google and Apple family accounts. No identifying information about a user needs to be provided nor does the user experience for non-designated devices need to change. **Appropriate defaults** means setting the default setting for these controls according to broad parental preference, rather than business preference. This would ideally focus on design choices and include the safest settings for privacy, engagement optimization, and revenue maximization. **Risky online activities** would cover any online service that has been credibly connected to significant risk for minors, such that a broad group of parents, across political divisions, are seeking to mitigate that risk.	Better design will reduce risk, but not eliminate it. Parental controls are an optional tool that can work with other design changes to further mitigate the impact of social media on children. Parents traditionally protect children from risky behaviors, and given the identified risks of social media for kids, it should be no different. Enabling parental control will not ban minors from the internet, but will add an opportunity for parents and children to have explicit conversations about the risks and benefits of engaging in the online world. While a risk exists that some parents may abuse this control, most parents have their children's best interests at heart.	Protect children and empower parents. Incentivize conversations about digital literacy.

Neely Design Code for Social Media

9	Design Code	Details	Supporting Argument	Goal
	Publicly provide product experimentation results on outcomes of societal interest for any meaningful product design decision.	**Product experimentation results** refers to the data that companies collect to understand the experimental impact of their products. **Design decisions** refer to any meaningful product change that potentially has societal or user impact. **Outcomes of societal interest** include (but are not limited to) measures of exposure to online harm, actions taken by users that indicate negative experiences, and surveys of user impact.	The codes listed above reflect causal estimates from numerous product experiments already done. Since products and technology are continually evolving, we need this level of transparency in order to meaningfully understand the causal impact of future product decisions. Publicly providing these results voluntarily would be a service to the world, but any compelled disclosure would need to be thoughtfully designed to avoid unintended consequences (see this article for one example of how this could be done).	Inform future design codes.

Current Signatories

- Ravi Iyer, Managing Director, USC Neely Center for Ethical Leadership & Decision Making; Co-chair, Council on Technology & Social Cohesion
- Juliana Schroeder, Professor in Management of Organizations, UC Berkeley Haas; Co-Director, Psychology of Technology Institute
- Aviv Ovadya, Thoughtful Technology Project; Affiliate, Berkman Klein Center, Harvard
- Center for Humane Technology
- Zach Rausch, Lead Researcher for The Anxious Generation
- Luke Thorburn, King's College London

- Nathanael Fast, Director, USC Marshall's Neely Center for Ethical Leadership & Decision Making; Co-Director, Psychology of Technology Institute
- Larry D. Rosen, Ph.D. Professor Emeritus
- Gloria Mark, UC-Irvine, Author of Attention Span
- Charles Montgomery, Author of Happy City
- David Evan Harris, University of California, Berkeley
- Jonathan Haidt, NYU-Stern School of Business
- Helena Puig Larrauri, Build Up
- Matt Motyl
- Frances Haugen
- New_Public

USC Marshall
Neely Center for Ethical Leadership and Decision Making

Neely Design Code for Social Media

Current Signatories

- Sarah Oh, Pebble.is
- Glenn Ellingson, Visiting Fellow, Integrity Institute
- Brandon Silverman, Founder, CrowdTangle
- Lisa Schirch, University of Notre Dame
- Wendy Wood, University of Southern California
- Matt Abrams, Founding Partner, Democracy Capital
- Emillie de Keulenaar, OILab and University of Groningen
- Peter Ondish
- Nick Chang, Civicue
- Matthew Katsaros, Yale Law School
- Steven Sloman, Brown University
- Nathaniel Lubin, Former Director of the Office of Digital Strategy, White House
- Clark Benson, CEO, Ranker
- Meetali Jain, Tech Justice Law Project
- Vardon Hamdiu, Sparkable
- Kostadin Kushlev, Georgetown University
- Prosocial Design Network
- Betty Ray, Pandora's Way and Center for Ritual Design

- Lena Slachmuijlder, Executive Director, Digital Peacebuilding, Search for Common Ground
- Nicolle Zapien
- Susan Benesch, The Dangerous Speech Project
- John Hunter, Chapman University
- Paul Bouchaud, Complex Systems Institute of Paris
- Kevin Frazier, Assistant Professor, St. Thomas University College of Law
- Filippo Menczer, Director, Observatory on Social Media, Indiana University
- Matt Pulley
- Dr. Michelle Ferrier, Executive Director, Media Innovation Collaboratory & Founder of TrollBusters
- Steve Downs & Thomas Getz, Founders of Building H
- Kristen Hansen, Executive Director, Civic Health Project
- Angela Lee
- Jeff Hancock, Stanford Social Media Lab
- Joo Ho Yeo
- Caitlin Ner, PsyMed Ventures
- Brett Frischmann, Villanova University
- Issue One

If you would like to add your public support for these design codes, either as an individual or as an institution, please email **raviiyer@marshall.usc.edu** or **fill out this form**.

Follow our work
NeelyCenter@Marshall.USC.edu
USC Marshall School of Business | 3660 Trousdale Pkwy, Suite 216 ACC | Los Angeles, CA 90089