
Introducing the Journal of Online Trust and Safety

Elena Cryst, Shelby Grossman, Jeff Hancock,
Alex Stamos, David Thiel

We launched the Journal of Online Trust and Safety to bring together rigorous trust and safety research, which is currently spread across many disciplines, and to spur new research in this field.

Trust and safety is the study of how people abuse the internet to cause real human harm, often using products the way they are designed to work. If someone uses a peer-to-peer messaging app to send a message that threatens to hurt the recipient's family, the product is being used as intended, namely, to send a message. However, the message content itself is causing harm.

eBay was a prominent early user of the phrase "trust and safety." In a 1999 press release they introduced their "SafeHarbor trust and safety program," which included easy access to escrow services and customer support.¹ In 2002, eBay appointed a former federal prosecutor to head a newly formed "Rules, Trust and Safety" team that aimed to identify fraudulent activity before the transfer of funds between buyer and seller.² Early academic uses of the term are also tied to eBay. A 2002 academic article in the Journal of Computer-Mediated Communication analyzed "how eBay has rhetorically constructed trust and safety on its site."³ "Trust" was used to reference trust among eBay users and between eBay users and eBay itself; "safety" was used to reference keeping platform users safe.

Priority areas for the Journal of Online Trust and Safety include child exploitation and non-consensual intimate imagery; suicide and self-harm; incitement and terrorism; hate speech and harassment; spam and fraud; and misinformation and disinformation. We are interested in these abuses on social media platforms, end-to-end encrypted platforms, blockchain technologies, and more.

Currently, some of these topics have one or more disciplinary homes. Others do not. Even for topics with disciplinary homes, accessing this research is not always easy for individuals doing trust and safety work at smaller online platforms. Journal access can be prohibitively expensive for those outside of academia. Even with journal access, non-academics may struggle to identify and access cutting-edge research in these fields. With the Journal of Online Trust and Safety, we strive to reduce friction to accessing relevant research that could support people working to make online platforms safer.

While we expect a majority of articles in this journal will come from academics at universities, we strongly encourage submissions from practitioners, particularly researchers at civil society groups or online platforms. Large online platforms have teams of trust and

1. <https://web.archive.org/web/20000815064513/http://pages.ebay.com/community/aboutebay/releases/9901.html>

2. <https://www.fastcompany.com/46858/how-ebay-fights-fraud>

3. <https://academic.oup.com/jcmc/article/7/3/JCMC736/4584230>

safety professionals conducting multimethod research into online harm, but these findings are rarely made public. We hope this journal becomes a space for these platforms to share this research so it may benefit smaller platforms.

The Journal of Online Trust and Safety will emulate successful practices developed by the Journal of Quantitative Description: Digital Media,⁴ and develop some new practices of our own. We implement several protocols to use author, editor, and reviewer time efficiently. First, we encourage the submission of an optional letter of inquiry, which allows the editors to assess whether the paper is a good fit for the journal. Authors may submit full manuscripts without submitting a letter of inquiry.

Second, we clearly set submission deadlines and publication dates for each issue, with the intent that this bounds the timeline for the review process and allows authors to know when their accepted article will be published.

Every issue will be launched with a livestream briefing where the editors will interview authors about the key takeaways from their research. We invite media and the public to these briefings and will work to publicize the briefings among the tech community as well.

We are excited about this first issue, which includes research on the relationship between hate speech and COVID-19 misinformation, an experimental study evaluating the effectiveness and scalability of crowdsourcing fact-checking, a review of perceptual hashing techniques for limiting the spread of harmful audio and visual content, a paper on the promises and limitations of online platform transparency regimes, and a study evaluating what content search engines surface for suicide-related queries in English and Spanish. Authors include a computer scientist, a legal scholar, political scientists, psychologists, and others.

Please reach out to trustandsafetyjournal@stanford.edu if you would like to be involved in the journal. Peer review and non-peer review submissions can be made at tsjournal.org.

Authors

Elena Cryst is the Assistant Director of the Stanford Internet Observatory.

Shelby Grossman is a Research Scholar at the Stanford Internet Observatory.

Jeff Hancock is the Norman Chandler Professor of Communication at Stanford University and the Founding Director of the Stanford Social Media Lab.

Alex Stamos is the Director of the Stanford Internet Observatory.

David Thiel is the Chief Technical Officer of the Stanford Internet Observatory.

Acknowledgements

We thank Alex Feerst for feedback.

4. <https://journalqd.org/>