ABSTRACT:

Words behind walls: Content moderation in social media and the impact of new regulatory initiatives. The “Everything in Moderation?” interdisciplinary project examines present-day and future challenges for moderation of online content, focusing primarily on end-to-end encrypted spaces. Starting from questions about phenomena like the spread of misinformation, the project investigates perceptions of moderation and of “walled” online participation across the spectrum of what is seen to be public or private communication. It aims to identify how challenges related to moderation of content in online settings (from public feeds to online communities to encrypted conversations) change across different spaces, and how content moves between such spaces. In this paper we will concentrate on work done within the regulatory aspect of the project. This translates the experience of content moderation practices, as mostly caught in the key dualities of (platform) norms and legal (co)regulation, and moderation internally applied by community moderators/administrators and moderation practices imposed by platforms. We pay particular attention to the role traditionally reserved for discussion spaces in democratic societies as facilitators for developing active and healthy citizenship, and whether emerging communication and moderation attitudes, nurtured in the confines of closed and encrypted spaces, have eroded this fundamental performance. In this respect, we will examine in parallel the anticipated impacts of the latest EU regulatory initiatives on platforms and platform content accountability, and how these are likely to shape moderation practices and communication on social media.