

84. Content Moderation

Content moderation emerged as a prominent theme during Elon Musk's first week at Twitter, especially following the controversial actions and statements of the musician and fashion designer Ye, formerly known as Kanye West. Ye's "White Lives Matter" T-shirts and an abrasive tweet about Jewish people had sparked a social media firestorm, resulting in his ban from Twitter. Musk's subsequent interaction with Ye highlighted the complexities of free speech and the challenges of impulsive content moderation decisions. Musk proposed creating a content moderation council to address these issues, emphasizing the need for diverse global perspectives, but his commitment wavered as the concept's execution became less of a priority for him.

This chapter also introduces Yoel Roth, Twitter's thirty-five-year-old who suddenly became responsible for content moderation in the wake of legal officer Vijaya Gadde's firing. Roth, despite his previous criticisms of conservatives and firm Democratic leanings, hoped to find a middle ground with Musk. Their initial interactions shed light on Musk's fears of internal sabotage and his intention to consolidate access to Twitter's security tools.

The narrative unfolds with Musk's impulsive decisions to reinstate controversial accounts like the Babylon Bee and Jordan Peterson, highlighting the tension between Musk's vision for unbridled free speech and the reality of managing a social platform. Roth's strategy to mitigate this included non-removal policy interventions, aiming for a balanced approach to content moderation without outright bans.

Key figures such as David Sacks and Jason Calacanis emerge as influential advisors to Musk, bringing their own perspectives on free speech and moderation. Their involvement underscores the chaotic transition phase at Twitter, marked by Roth's efforts to address a spike in hateful content following Musk's takeover. The chapter reflects on the complexity of content moderation and the continuous struggle between ideologically driven decisions and the practical necessity for coherent, consistent policies to govern online speech.