

## INTISARI

Penelitian ini bertujuan untuk menganalisis penerapan moderasi konten pada media sosial dengan menggunakan perspektif utilitarianisme John Stuart Mill. Penelitian ini dilatarbelakangi oleh keraguan atas konsep moderasi konten pada media sosial yang mampu melindungi penggunanya dari konten-konten merugikan, dengan tetap menjaga aspek kebebasan berekspresi di dalamnya. Oleh sebab itu, penelitian ini bertujuan, *pertama*, menjelaskan konsep moderasi konten pada media sosial, *kedua*, menjelaskan bagaimana analisis utilitarianisme John Stuart Mill meninjau penerapan moderasi konten pada media sosial.

Penelitian ini merupakan penelitian kepustakaan yang berusaha mengidentifikasi permasalahan aktual terkait moderasi konten pada media sosial melalui pandangan utilitarianisme. Penelitian ini menggunakan unsur metodis: interpretasi, koherensi intern, dan deskripsi.

Hasil penelitian ini adalah: *Pertama*, konsep moderasi konten pada media sosial merupakan proses peninjauan konten-konten membahayakan menggunakan metode manual, metode otomatis, dan bantuan pengguna dalam pendeteksiannya. Hasil dari peninjauan tersebut, kemudian ditindaklanjuti dengan penghapusan konten, pembatasan konten, pensorsan akun, atau penghapusan akun. *Kedua*, moderasi konten pada media sosial dapat diterima jika mampu memberikan manfaat untuk masyarakat secara luas dengan tetap mempertimbangkan aspek kebebasan di dalamnya. Hal tersebut dapat dicapai dengan mengadopsi prinsip utilitas sebagai tujuan, prinsip kebebasan sebagai batasan, dan *harm principle* sebagai dasar penerapan moderasi konten pada media sosial.

Kata kunci: utilitas, kebebasan, kerugian, keamanan, moderator

## ***ABSTRACT***

This research aims to analyze the implementation of content moderation on social media from the perspective of John Stuart Mill's utilitarianism. The study is motivated by uncertainties regarding the concept of content moderation on social media that can protect users from harmful content while preserving the aspect of freedom of expression. Therefore, this research aims, *first*, to explain the concept of content moderation on social media, and *second*, to elucidate how John Stuart Mill's utilitarianism analyzes the implementation of content moderation on social media.

This is a literature review that seeks to identify current issues related to content moderation on social media through the lens of utilitarianism. The study employs methodological elements such as interpretation, internal coherence, and description.

The results of this research are as follows: *First*, the concept of content moderation on social media involves the process of reviewing harmful content using manual methods, automatic methods, and user assistance in detection. The results of this review are then followed by the removal of content, content restrictions, account suspension, or account deletion. *Second*, content moderation on social media can be acceptable if it can provide benefits to society at large while still considering the aspect of freedom within it. This can be achieved by adopting the principle of utility as the goal, the principle of freedom as a limitation, and the harm principle as the basis for the implementation of content moderation on social media.

**Keywords:** utility, freedom, harm, safety, moderator