

インフォメーション・ヘルス AWARD2024

今回私が考えた案は、ネット上の誹謗中傷への対策についてだ。

インターネットが普及している現代で、SNS での誹謗中傷は大きな問題の 1 つと言える。

そんな誹謗中傷の対策として私が提案するのは、

人工知能を用いた「誹謗中傷コメントの自動削除システム」だ。

この案について、今から大きく 3 つのテーマに分けて説明していこうと思う。

1. 誹謗中傷の現状

YouTube や TikTok、Instagram や X(旧 Twitter)などのアプリは、今やスマートフォン所持者の大半が利用しているのではないだろうか。世界中の人たちとネット上で繋がることができるからこそ生まれる問題が、「誹謗中傷」だ。

誹謗中傷とは、人格や容姿・存在を否定するものや、相手の根拠のない悪口やデマを言いふらすことを指す。これらによって不安な思いをするだけでなく、精神的に追い詰められ自殺者が出る事件も起こっている。この現状は、SNS における「言葉」は時に人を死に至らせる凶器になるという意見に繋がる。

SNS が私たちの生活と深い関わりがあるということは誹謗中傷も同様に身近なものであり、誰もが被害者・加害者になる可能性があるということだ。

2. 匿名性について

誹謗中傷がなくなる大きな原因の 1 つに“匿名性”がある。

自分の本名を伏せ、ペンネームなどを使用することだ。お互いの顔や名前がわからないことを利用して誹謗中傷が起こるのだ。以前ニュースで、「匿名は包丁だ」という言葉を耳にしたことがある。食べ物を切るといふ面では包丁はとても便利なものであるように、匿名性も正しく使えば便利なものだが、使い方を間違えると人の命を脅かすことがある。私はこの言葉を広くたくさんの人に知って欲しいと思う。

相手の表情が見えないからこそ自分の言葉によって相手はどのような気持ちになるかを想像し、一度立ち止まって考えるという行為がとても大切だからだ。

しかし、それで誹謗中傷がなくなるのならば簡単な話だ。

これは私個人の意見だが、誹謗中傷を完全になくすには「匿名性」の廃止は絶対条件だと思う。

一人ひとりの顔や名前がわかる状態ならば、自分が発信する言葉にもっと責任を持つようになり、無責任な誹謗中傷はなくなるだろう。

しかし、匿名性であることで生まれる自由やメリットもある。匿名は自分を守る盾でもあるからだ。顔や本名が誰でも閲覧できる場所に記載されていると、悪用されることもあるだろう。

そういったことを考えると、匿名性を廃止するという方法は非現実的な部分が多い。

そこで私は、まずは誹謗中傷を「減らす」ということを第一に考えたしくみを提案する。

3. AI を用いた誹謗中傷コメントの自動削除システム

このシステムの構造は3段階に分かれる。

① NGワードを決める

② AIによりNGワードが含まれる文(コメントなど)を探索・発見

③ NGワードが含まれるものは自動削除

① で決める NGワードとは、誹謗中傷に該当する容姿や人格などを否定する言葉、存在を否定する言葉などだ。

それらのワードをNGワードとしてAIに記憶させ、反応できるように準備をする。

そしてAIが自らコメント欄などで探し、発見し、即座に自動削除をする。

このシステムがあれば、誹謗中傷が本人に届くことがなくなるため「誹謗中傷」自体を減らしていき、最後にはなくすことができるかもしれない。AIがコメント欄などネット上のパトロール的な役割をすることで誹謗中傷によって傷つく被害者はもちろん、新たな加害者を生まずに済むのだ。

現代社会を生きていく上で、インターネットから得る情報の量は膨大で、私たちの生活を豊かにしてくれる便利なものであり、必要不可欠だと思う。同時に、正しく使わなければ命すらも脅かす危険なものでもある。だからこそこれから先、インターネットを利用する中で傷つく人や命を奪われる人が減っていけばいいと思う。