個人に対する誹謗中傷等を内容とする投稿への対応について

～ 対策強化に加え、深層学習を用いた自然言語処理モデル (AI) の技術提供や検討会を設置 ～

ヤフー株式会社（以下、Yahoo! JAPAN）は、個人に対する誹謗中傷等を内容とする投稿への対応に関し、サービスをユーザーにより安心して使っていただくため、現在「Yahoo!ニュースコメント」において導入している深層学習を用いた自然言語処理モデル（AI）のさらなる活用などの対策強化を進めてまいります。

Yahoo! JAPANでは、多数の投稿系サービスを提供しており、多くの皆さまに手軽にご利用いただける場を提供すると同時に、安心してご利用いただけるよう、サービスの健全化にも力を入れて取り組んでおります。誹謗中傷等の他人を傷つける表現行為は許されないものであり、Yahoo! JAPANでは、以前より誹謗中傷等の内容を含む投稿行為を禁止し、ユーザーのみなさまに遵守をお願いしています。

また、ユーザーのみなさまへの周知・啓発などにより当該投稿の事前抑止に努めるとともに、専門チームによるパトロールや、深層学習を用いた自然言語処理モデル（AI）を利用して不適切な投稿対策を行ってまいりました。「Yahoo!ニュース コメント」においては、深層学習を用いた自然言語処理モデル（AI）による検知を通して、1日平均約2万件の不適切な投稿（記事との関連性の低いコメントや誹謗中傷等の書き込みなど）の削除を行っています。今般、大変痛ましい事態についての報道があったことを踏まえ、Yahoo! JAPANはこれからも対策を強化して対応をしてまいります。

今後は、こうした深層学習を用いた自然言語処理モデル（AI）について、他の投稿系サービス事業者に対する技術提供を進めていきたいと考えています。

さらに、これらの問題への対処にあたっては、法的課題や実務的課題があると認識しており、これら課題をデジタル時代に即した共通規範に基づき解決すべく、議論を行う場である検討会を2020年6月中めに設置します。本検討会においては、法律家のみならず幅広い分野から有識者を募り検討を行うこととし、その結果を公表していくことといたします。