

人類のデータを使って学習する人工主体の公共的役割を考える

濱田 太陽 (Hiroaki Hamada)

株式会社アラヤ

2022年に text2image と呼ばれるテキストコマンドによって画像を生成する自然言語処理のアルゴリズムや複数のサービスが突如現れ議論を巻き起こした。生成される画像の精度が高まり人間が生成する画像に近いレベルで生成することができるようになった。また、画像生成を精度良く行うための技術やテキストコマンド自体が売り買いされることになり、画像生成に付随したサービスも展開され始めた。しかし、このアルゴリズムに必要となる学習データはインターネット上から収集された著作権がある画像も大量に利用されている。著作権者への承諾なしにサービスを展開し有料で画像生成するサービスには多くの批判もなされた。この批判の背景には、人間の習熟が必要なスキルによって生み出されたデータを無制限に吸収した AI によって、画像が生成され販売されることにより人間のスキルに対する還元がなされていないことと、彼らの職が奪われる可能性に原因がある。

この状況に対して、利用されたデータを生成した人間への還元は行うことはできないが、技術の進展は止めることはできないので、アルゴリズムをオープンソースとして公開することで有料のサービスが出現することを妨げ、公共財として利用しようという動きが出てきている。この動きは、アルゴリズム自体の商業的価値をなくすことで、データを生成した人間への還元を放棄したものである。今後このような人間の生成したスキルをもとに無制限に学習する人工主体が、商業的利益をうむことが出てくると考えられる。この場合、同意や契約なしに行われる学習をどこまで認めるのか、また学習データを提供した人間にどのような還元をすべきなのかは現在必ずしも明らかではない。

本講演では、人間のスキルが同意や契約なしに学習データとして使われる人工主体について text2image などのアルゴリズムから着想を得ながら、状況の整理を行いつつ、人工主体が担うもしくは担うべき公共性についての議論の端緒としたい。