

## データ数を増やしても残るAI診断のバイアス

### ◆AIによるfMRIデータの解釈には、データ数を増やしても人種的バイアスが残る

2022年3月、ドイツのハインリヒ・ハイネ大学、シンガポール国立大学などの研究グループは、脳のfMRI（functional Magnetic Resonance Imaging、機能的磁気共鳴画像法）のデータからAI（Artificial Intelligence、人工知能）を用いて行動様式を予測する際の人種的なバイアスを解析した結果を報告した。

白人米国人（WA）のデータを多く含むデータセットを用いて作成したAIモデルでは、アフリカに起源をもつ米国人（AA）に対する予測の正解率が低くなる。これは、一般に、データセットの偏りに起因するバイアスとして知られている結果と一致する。しかし、WAとAAのデータ数を等しくして構築したAIモデルでも、WAに対する予測よりもAAに対する予測の正解率が低かった。さらに、WAとAAのデータセットを分けてそれぞれのAIモデルを作成しても、正解率は向上するものの、依然として、WAに対する予測精度がAAに対する予測精度を上回っていた。

### ◆米国のNISTがAI応用の際のバイアスに関する報告書を発表

22年3月、米国のNIST（National Institute of Standards and Technology、国立標準技術研究所）は、AIを応用する際のバイアスに関する報告書「Towards a Standard for Identifying and Managing Bias in Artificial Intelligence」を発表した。

AI応用の際の、人種、性別、年齢などによるバイアスは、上記の医学的な研究に限った問題ではなく、あらゆる分野で取り上げられている。NISTの報告書では、具体的な例を紹介した後、バイアスを、歴史や社会など体系的なもの、計算手法や統計学的なもの、認識や判断など人間的なものに大きく分類し、挑戦すべき主な課題として、データセット、検査と評価、人間的な要因を挙げている。

AIの応用が産み出すバイアスの原因として、統計学的な手法やデータセットの偏りが挙げられることが多いが、それだけであれば、むしろ、対処は容易である。しかし、真の原因が歴史的、社会的、人間的なところにもあるとすると、この問題は想像以上に根深いものである。

【戸潤一孔】