



大学への国策 AI 推進圧力への疑問

松原 望 まつばら のぞむ
東京大学名誉教授(統計学)

国策 AI の議論のレベル劣化が止まらない。大学はいろいろな面で AI とは大きなかかわりがあり、この事情はさまざまな形で大学に及んでいる。このことを筆者の専門である統計学とその関連領域から述べてみよう。キーワードは「デジタル化」とその光と陰である。

目下の問題の根源は、「AI」、「データ・サイエンス」、「機械学習」、「ビッグ・データ」の領域で、訳もわからず流行の様相を呈している。もっとも、近い友人は最近では多少陰り気味で落ち着きそうだという。では、たとえば、私の分野に近いところで脚光を浴びている「データ・サイエンス」とは何であろうか。それは「統計学」とどう違うのか。統計学専門やその周囲のほとんどの人にもその意義や位置づけがわからない。統計とどう違うのか。ここの「データ」とは何であり何でないのか、議論は出来のよくない研究費申請のように効能ばかりで学問的議論がなく、わずかなデータ・サイエンスの国策予算をめがけて大学がコンペのように殺到している。

これらの問いに一つの答えはある。アメリカでよく売れている『機械学習—純然たる初心者に』*はわかりやすいが、グラディエント・ブーストなども扱いけっこうレベルは高い。はじめにこの周辺の定義の説明があるが、そこでは「計算

機科学」—「データ・サイエンス」—「AI」—「機械学習」と入れ子の(ロシアのマトリョーシカ的な)構造になっていて、したがって、当然ながら「データ」はどこまでもデジタル情報に限られる。そしてこのことが重要である。もちろん「デジタル化」による計算処理速度は途方もなく大きく、このメリットの上に AI や「ビッグ」な世界がある。しかしながら、その反面に人間価値にとって非常に重要で大きな制約もかかる。情報科学の始祖クロード・シャノン 以来すべての情報科学が成り立つ基本の前提条件は、人間的な「意味」をらち外として除外することにある。金属電子論で高名な物理学者 L. ブリルアンは優れた情報理論のテキスト『科学と情報理論』(みすず書房)で、深い本質洞察を残している。

—絶対的な情報の定義(情報のデジタル化・筆者)は、実際上大きな重要性をもっている。人間的要素を除外したことによって、多くの疑問に答える道がひらけた。電話方式を設計する技術者は、回線がうわさ話のおしゃべりに使われるのか、株式市況の伝送に使われるのか、あるいは外交上の秘密の通信に使われるかを配慮する必要がない。技術上の問題は常に同一である。—

「データ・サイエンス」でいう「データ」も、デジタル数字の並びであって、統計学でいうところの統計計算の部分にすぎず、エクセルで平均とか標準偏差とか相関係数はどこまでも常に同一の計算である。人間にとっての「意味」まで配慮されるわけではなく、範囲外であることはいうまでもない。「意味」とは極めてアナログ的な存在なのである。計算のあと、コンピュータ・ソフトのアウトプットの見方、統計量の意味や目的は統計

*一回帰分析にミスがある。