

ニューヨークのタイムズスクエアで行われたカウントダウンに今年も多くの人を訪れたようだ。タイムズスクエア周辺は世界的な観光地であると同時に大手金融機関の本社も多数近隣に位置するビジネス街区でもある。最寄りのタイムズスクエア駅は、実に12もの地下鉄の路線が四方八方から乗り入れる混雑駅でもある。少し気になってChatGPTに「同駅に最も多く入場客を訪れる時間帯はいつか」と問いかけてみたところ、「平日の朝と夕方がピークであり、イベント時は週末や祝日の昼間の時間帯も混雑」との平凡な答えが返ってきた。しかし、

ングという駅が最も多いことも立ちどころにわかる。

冒頭の質問に対する答えも「午後5時から午後6時の間」と明快である。さらに言えば、なぜか水曜日の入場者数が多く、とりわけ2024年4月3日午後5時台の入場者は18,396名と最多であった。

こうしたデータの公開について常に話題になるのがプライバシーの問題であるが、ニューヨーク市の警察は、駐車違反の情報を場所や時間だけでなく、違反車両のナンバーや車の色、年式まで公開している。分析機能を用いれば、にわかには信じ難いが1年間で1000件近い交

数 | 理 | の | 窓

AIがオープンデータとつながる日



この答えは近い将来、生成AIが直接、オープンデータのデータセットを深く学習する日がくれば、より洗練されたものになるに違いない。

そのオープンデータ、実はどんどん公開が進んでいる。2021年にニューヨーク州が制定した「MTA Open Data Act」により、ニューヨーク市の交通当局はオープンデータの継続的拡充を義務付けられることとなった。2023年に公開されたMTA Subway Hourly Ridershipでは、ニューヨーク地下鉄全駅の時間別、料金種別の入場者数を誰でも利用でき、分析もすべてウェブサイトで行える。例えば、2024年1月1日午前10時に最も入場者数が多かった駅はタイムズスクエア駅であったが、種別を高齢者・障がいのある人に絞ると、郊外のフラッシュ

通違反をする車もあることがわかる。その車は2024年2月20日に1日で7件もの駐車違反を繰り返していたようだ。どんな洗練された分析も現実世界を確実に描写することは困難だが、全データがあれば話は別である。

日本においても平成28年の官民データ活用基本推進法にて国及び地方公共団体はオープンデータへの取り組みが義務付けられた。ただ、ニューヨークのようなデータまで公開してよいか、という点は議論が必要であろう。

オープンデータの質が高ければ高いほどAIは様々な予測を行えるようになり、現在人間が様々な手法を用いて行っている予測を凌駕していくかもしれない。それを思うと、筆者もどこか首筋が冷たいような気がするのだ。

(神井 隆)