



松原 仁 Matsubara Hitoshi 人工知能研究者

京都橘大学工学部情報工学科教授。公立はこだて未来大学特命教授。元人工知能学会会長、前情報処理学会副会長。著書に『AIに心は宿るのか』（集英社インターナショナル、2018年）『やさしくわかる！文系のための東大の先生が教えるChatGPT』（ニュートンプレス、2024年）など

認識するAI(2)

今回も認識するAIについて考えてみましょう。AIによって作られたフェイクを見破るためにもAIが使われています。

ディープフェイクを見破る

AIの技術が急速に進歩して、さまざまな所でAIの成果が使われるようになってきたのはよいことなのですが、残念ながら悪いことにもAIの成果は使われています。最先端の科学技術はこれまでの歴史をみても、よいことに使われた半面で悪いことにも使われてきました。自分が発明したダイナマイトが戦争の兵器などに悪用されたので、発明で得た財産で世界の幸福に貢献する業績を挙げた人々を表彰することにしたのがノーベル賞です。ダイナマイトが悪用されたことに対する後悔からきた行為といえるでしょう。ほかの科学技術、例えば、原子力も同様にプラスの面とマイナスの面があります。AIも例外ではありません。

ディープラーニングを用いてフェイク（偽）の画像や音声を作ったものをディープフェイクといいます。いま世の中には多くのディープフェイクが氾濫（はんらん）しています。AIの負の面を象徴しています。政治家が実際には会っていない人に会っている写真であったり、言ってもいないことをしゃべっているインタビュー動画などは、選挙のときに対立候補をおとしめる目的で使われており、選挙民はだまされないように注意しないといけません。芸能人の裸などのわいせつなフェイクの静止画・動画もよくみられます。地震や水害が起こったときにも、実際の災害とは異なる動画がSNSに投稿されることがあります。

最近のディープフェイクは技術が進んできて簡単にはフェイクと見破ることができません。

その一方で、ディープフェイクを作るのは比較的簡単にできるようになっています。皆さんも静止画や動画を見たとき、あるいは音声を聞いたときに、常にこれは本物かと疑ってかかる必要があります。何でも疑ってかかるのはどうかとも思いますが、犯罪に巻き込まれたり詐欺にあったりしないためにはやむを得ない態度なのです。

ディープフェイクを見破るAIの開発も盛んに行われています。これも本物かフェイクかどうかを判定するという意味で、認識AIの一種といえるでしょう。ディープフェイクはディープラーニングを用いて作られますが、それを見破るAIもディープラーニングを用いて作られています。疑似的にフェイク画像を作ってそれらと本物の画像をたくさんAIに入力してその違いを学習させるということを行います。例えば、人の顔のフェイク画像は目に不自然さがしばしば出ることが指摘されています。本物の画像の目では瞳孔がちゃんとしたかたちをしているのに対して、フェイク画像の目では瞳孔のかたちが歪（ゆが）んでいることがあります。動画であれば、まばたきの回数やパターンもフェイクでは不自然さが出ることがあります。このような不自然さに注目することによってフェイク画像が判定できるのです。2022年に東京大学の山崎俊彦研究室が、動画の中の人物の顔が本物かフェイクかを高精度で判定する手法を開発して話題になりました。ディープフェイクを作る技術とディープフェイクを見破る技術は、いたちごっこで進歩していくものと考えられます。

AIによる管理社会

現在は多くの場所に監視カメラが設置されており、カメラの映像の分析にもAIが用いられています。そのおかげで犯罪者が逃走してもかなりの確率でカメラが捕捉して捕まえられるようになっています。店舗の中の監視カメラは万引

きも把握しています(認識AIが万引きとなる行為を判定することができます)。不審な動きも分かります(例えば、不審者が刃物を持っているのを見つけることができます)。夜間の防犯カメラもAIが見守っています。安全の確保に認識AIが貢献しているといえます。

クレジットカードをお持ちの読者が多いと思いますが、クレジットカードを作るときは一定の審査があります(クレジットカード会社によって審査が厳しい所とそれほどでもない所があるようです)。まず本人の属性です。年齢、配偶者の有無、家族構成、居住状況(持ち家か賃貸か)、職業、勤務年数、雇用形態(正規か非正規か)、年収などの情報がチェックされます。次に信用情報です。過去にクレジットカードやローンなどで支払いがちゃんとしていたか滞ったことがあるかを信用調査機関が把握しているので、それもチェックされます。従来はこういう審査は人間がやっていたのですが、今はかなりの部分をAIがやっています。これも一種の認識AIといえるでしょう。

さらに認識AIによる管理が進んでいる「ある国」の話をししましょう(どの国か想像がつく読者もいらっしゃるかもしれませんが)。その国ではすべての国民に政府によって点数が付けられているそうです。点数はその人の行動や言動に応じて付けられます。口頭でよろしくないこと(例えば、政府の悪口など)を話すとそれを聞いた人の密告によって政府に伝えられてマイナスの点が付きます(密告をするとプラスの点が付きます)。SNSに匿名でよろしくないことを書いても、誰が書いたかを特定されてマイナスの点が付きます(今のAIの技術を用いれば匿名で書いても誰が書いたのかほぼ特定できます。日本でもそうです。日本は犯罪に関係する場合しか特定しませんが)。その一方で、その国の政府が敵視している国の悪口を言ったりSNSに書いたりするとプラスの点が付きます。道端にゴミを捨てたらマイナス点ですし、道端のゴミを捨てたらプラス点が付きます(誰がゴミを捨てたか

捨てたかも監視カメラで捕捉されているということです)。電車の中でお年寄りに席を譲ったらプラス点です。赤信号の横断歩道を渡ったら当然マイナス点です。よからぬ所に行ってもマイナス点です。寄付をしたらプラス点です。罪を犯したら当然ながら大きなマイナス点です。一挙手一投足がAIによって見られているということです。

このようにして付いた点数でこの国ではすべてが決まります。クレジットカードの審査もこの点数で行われます。入学試験でも入社試験でも使われます。出世できるかどうかも点数次第です。結婚しようというときは当然相手の点数を気にします。点数が低いと外国へ行くパスポートを発行してもらえません。こうなると大多数の国民は高い点数をめざして活動するようになります。国民を管理したい政府にとって認識AIは優秀な道具なのです。

ユートピアかディストピアか

ジョージ・オーウェルというイギリスの作家が1949年に書いた「1984」という小説をご存じでしょうか。未来の全体主義国家の管理社会を描いたディストピア(不幸な未来)の小説です。前述の「ある国」は、オーウェルの「1984」の管理社会を既の実現したといえそうです。AIをずっと研究している筆者からすれば、このような管理社会を実現するための道具としてAIを研究してきたのでは決してありません。人間が健康に安心して暮らせるユートピア(幸福な未来)の世界に近づけたいと思って頑張ってきたつもりですし、これからもそうしていくつもりです。

AIはある特定の社会を実現したいなどという目的は持っていません。そもそもAI自体には「こうしたい」という目的がなく、言われたことをするだけなのです。AIを用いてユートピアをめざすのか、あるいはディストピアをめざすのかは、AIを使っている我々人間の側にかかっています。そのことをよくよく考える必要があります。