［データ・AI ガバナンスと COVID-19：アジアにおける中長期的展望］

1 壊滅的なリスクに対抗するための倫理とガバナンスの展望：COVID-19 から汎用人工知能の長期的な安全問題まで

曾　毅／孙　康／鲁　恩蒙 | 中国科学院自动化研究所脑型知能研究センター

COVID-19 との闘いにおける AI の倫理，ガバナンス，実践

現在，北京のロックダウンは解除され，2020 年 6 月 11 日からの COVID-19 第 2 波を乗り切ったところである．COVID-19 を乗り越えるために，私たち人類は必ず勝利するだろうと今でも前向きに考えている．AI とデータ分析は，そのために設計されたわけではないが，COVID-19 との戦いに広く活用されてきた．AI の歴史において，私たちは特定の期間だけに使う AI システムを発明，開発することはほとんどなかった，また，現在，このために作られた AI（接触追跡アプリなども含む）はこの特殊な期間が終了とともに，使用が終了されることがすでに決定している．COVID-19 と戦うための AI は，AI 開発とガバナンスの歴史においても非常に特別な事例となるだろう．

私たちは人間社会の発展の中で，さまざまな壊滅的なリスクに直面するかもしれない．そして，私たちはそれぞれのリスクから学び，可能であれば，リスクを回避するか，少なくとも軽減す手だてを考える必要がある．汎用人工知能（Artificial General Intelligence：AGI）は，人間が持つ認知機能のあらゆる側面から，人間の知能を実現できる知能システム構築を目指している．それらは社会の将来のために真に有益に設計される可能性がある一方で，潜在的，長期的な悪影響に対する戦略的な準備なしに AGI を実現してしまうという壊滅的なリスクもあり得る．本稿ではこのような観点から，AGI からの潜在的リスクを回避または低減するために COVID-19 から何を学ぶことができるかを議論する．

COVID-19 に対抗するための AI 活用：技術進歩，倫理的リスク，ガバナンス

COVID-19 に対抗する AI 利用の概要

COVID-19 に関する議論では，ビッグデータ解析，知識表現・推論，パターン認識，自動化された意思決定などを含む AI が感染拡大の初期から活用されてきた．これまで，AI は COVID-19 の拡散予測，遺伝子配列解析に基づく潜在ウイルス宿主予測，SARS-CoV-2 構造予測，サブタイプと変異認識，自動 CT 画像認識，COVID-19 検出のための音声認識，ウイルス検出，ロボットとの自動対話システム，創薬および自動化された健康状態サーベイランスに活用されてきた．より具体的な事例や関連する技術的議論は別稿に掲載されている[1]．本稿では，それぞれの取り組みの根底にある技術的および倫理的課題を中心に紹介する．

ディープニューラルネットワークは，CT 画像の自動認識，さらには COVID-19 を検出するための音声認識にまで利用されている．たとえば，アリババ・ダモ・アカデミー（Alibaba Damo Academy）では，2020 年 2 月末時点で 3 万件以上の CT 画像診断を COVID-19 の疑いのある症例に対して実施している．診断の精度は 97% で，1 件あたりの検査時間はわずか 20 秒で済む．しかし，多層構造の隠れ層は説明可能性や透明性に課題があり，信頼性の低い分類結果につながる可能性がある．たとえばディープニューラルネットワーク

は，皮膚癌の識別で元の画像に対する特定程度の敵対的回転を有していても，正常な皮膚画像を皮膚があるものとして分類する[2]．同様のリスクがCOVID-19のCT画像認識にも存在する可能性がある．

知識表現と推論は，薬剤スクリーニングと発見およびCOVID-19関連サービスロボットでも使える自動化対話システムなどに広く使用されている．たとえば中国では，いくつかの病院がロボットを24時間使って，COVID-19対応として薬の配布，食料や家庭用品の配達，治療を行っている．バイドゥ社（Baidu）は，北京，西安と上海で特別な手当を必要としている地域社会や特別なグループに電話をかけられる知的なプラットフォームを構築して，100万件の電話をかけて統計を取っている．ここで想定されるリスクは，医薬品や治療に関する医学的知識の質だけに限らない．知識と推論ルールの整合性確認を厳密に検証して実行しなければ，知識ベースとそれに付随する推論サービスから導き出された結論と回答は信頼できないものとなる．

自動化された監視は，潜在的なリスクを発見し制御するという観点から，COVID-19への対応として強化されてきた．パンデミックの初期には，顔認識と組み合わせて，自動化された温度モニタリングと追跡アプリが地下鉄，駅，空港，ソーシャルサービスセンターに配備され，高熱の人々を特定し追跡し，必要な行動を支援してきた（現在では，関連するインフラストラクチャは，後述する「ヘルスコード（健康碼）」と温度モニタリングに置き換えられている）．これらはスクリーニングの補助に非常に役立つ可能性がある（たとえば，メグビー社（Megvii）の装置では1分間に300人を検査でき，センスタイム社（SenseTimes）の装置はマスクをしている人でも識別できる．

ヘルスコード制度

ヘルスコード（健康碼）は，中国のパンデミック時に公衆衛生の緊急管理支援の最も重要な基盤として機能している．ヘルスコードによる接触追跡は，主にGPSと公共交通機関やその他の位置情報サービスからの関連情報に基づく．ヘルスコード関連情報は，主に各地方自治体が集中管理しており，同様の戦略が韓国，ニュージーランド，ロシア，インドなどでも採用されている[3),4)]．このような方法とは別の接触確認アプローチは主にブルートゥースを用いており，オーストラリア，シンガポール，フランスなどでは集中型ストレージを使い，イタリア，ドイツ，日本，英国，スイス，カナダなどでは分散型ストレージを採用している．

ヘルスコードと，そのために収集された情報は，他国の接触追跡アプリに比べて広範囲にわたっている．これは，各接触追跡アプリが持つ設計哲学の違いに根ざしている．第一に，中国のヘルスコードは，その信頼性を確保するために，より多くの個人情報を求める．初めて使用する際には顔認証が必要で，国民ID情報も提供する必要がある．電話番号やリアルタイムの顔情報，国民IDと合わせて，それが本当に携帯電話の持ち主であると確認できる．他タイプのアプリは，これらの情報を集めなければ利用者のプライバシーはある程度保たれると考えているが，それでは携帯電話の本人確認をするのは非常に難しいだろう．第二に，生命の安全を第一に考えるために，中国のヘルスコードは義務化されている．いくつか他国の自発的な参加を促す仕組みと比較して，中国ではスマートフォンを持っている人は公衆衛生緊急管理システムから取り残されることはなかった．その理由として，中国では職場や学校の移動管理にヘルスコードが広く使われているからだ．中国では，「自己」の伝統的な哲学は，「関係的自己（relational self）」という概念に基づいており，「自己」は共同体，文化と社会の一部である．そして，関係性は「自己」を反映するための重要な視点である．したがって，公衆衛生危機緊急管理システムによる個人情報へのアクセスは，（システムが信頼できるものである限り）実行可能に設計されており，自発的な参加という設計は受け入れられない．なぜならば，参加を望まない人は周囲の人々の安全に対する潜在的リスクとなるからである．第三に，中国のヘルスコード制度は，個人情報のセキュリティ確

保を前提として，有効性を第一に置いていることが見てとれる．

また中国では，自発的な参加によるブルートゥースベースの接触追跡サービスも地域展開されている．たとえば，北京智源人工知能研究所と北京大学が開発した Blue Bubble COVID-19 追跡システムが北京市内のいくつかの場所に最近配備された．

データガバナンスをめぐる社会的，倫理的，法的な懸念と対応

従来のヘルスケアシステムの利害関係者は，現在，患者と医師だけではなく，AI サービス提供者（主に企業），さまざまなレベルの政府，コミュニティや地域の職員，警備員やボランティアまで，少なくとも中国の公衆衛生危機緊急管理システムにおいては大幅に拡大した．彼らは感染を減らすために懸命に働き，意思決定のために必要なフィードバックをより高いレベルの組織に提供し続けており，パンデミックとの戦いに非常に貴重な貢献をしている．それにもかかわらず，彼らのすべてが現在の個人情報保護を保護し，人間の主体性を確保するための方針に沿って行動しているわけではないことは明らかであり，プライバシー，偏見，安全性，アカウンタビリティなどの倫理的リスクにつながる可能性がある．

2020 年 2 月，武漢住民の個人情報がネットや WeChat グループに投稿された．これは偏見や孤立を引き起こし，個人の評判に悪影響を与えた．2 月 1 日，山西省臨沂市公安局のネットワーク警察署は，地元の男性が WeChat グループの 35 人の「感染接触者リスト」（氏名，身分証明書番号，住所他の個人情報が記載されたリスト）を配布して，法律に基づき行政処分を受けたと発表した．同様に，湖南省宜陽郡保健局の副局長も，COVID-19 の患者のプライバシーを漏洩したとして調査を受けた．この期間中は，生物学的および社会的安全上の理由により，より多くの個人情報を収集する必要があったかもしれないが，情報へのアクセス制御および情報公開は適宜規制する必要がある．

中国科学技術省による COVID-19 対策のための AI の有益な利用促進の取り組みなど，COVID-19 と戦うための AI の有益な利用を促進，規制する政策が必要である．また，パンデミックの際に個人情報を保護するため，省庁間の連携体制の枠組みが設けられている．中国運輸省は 2020 年 1 月 30 日に「COVID-19 の予防・管理と輸送セキュリティの調整に関する運輸省緊急通達」を発行し，その中で次のように述べている．「健康に関する部局は COVID-19 の予防と管理の必要性を満たすためだけの利用に限って，個人のプライバシーや個人情報のセキュリティを法律に基づいて厳格に保護すること．その他の機関，団体，個人は，許可なくインターネット上で関連情報を無断で開示したり，配布したりすることはできない」．2020 年 3 月 2 日，中国民政部，中国サイバー空間管理局，産業情報技術省，国家衛生委員会は共同で「COVID-19 感染爆発の予防と管理のための情報構築と適用ガイドライン，第 1 版」を発行し，その中で次のように述べている．「COVID-19 の予防と管理作業要件のために情報製品（サービス）の予防と管理が必要である．コミュニティ住民の情報を収集するにあたっては，情報収集の必要性を明確にし，同意を得る必要がある．また，ウイルスの予防と管理のためではなく，ほかの目的で利用されることが明らかな場合には地域住民の同意を再度得る必要がある」．

公衆衛生管理のためのプライバシー保護

プライバシーの保護は，パンデミックの際は考えなくてよいというものではない．また，効果的な規制やセキュリティインフラが適切に講じられていれば，適切に保護することができる．2020 年 5 月 31 日までの WHO のデータ（GMT＋8, 8:00）を参照すると，216 の国または地域で COVID-19 の感染が確認されている．そのうち，少なくとも 63 の国と地域（1/3 以下）がプライバシー／データ／AI ガバナンス関連の政策を発表している[5]．中国の場合，ヘルスコード関連のデータの多くは，各都市にある地域ビッグデータセンターで管理されており，その技術とインフラはアリババ社

やテンセント社などの企業が提供している．したがって，個人情報の安全性とセキュリティを確保するためには，多様なステークホルダがそれぞれ異なる責任を負う必要がある．警備員やボランティアであっても個人情報の一部にアクセスできる場合があるため，公衆衛生危機の緊急管理システムにかかわるすべてのステークホルダが個人情報の方針を認識する必要がある．

杭州保健省から，パンデミック発生後のヘルスコードの利用を拡大するという提案があった．ヘルスコードおよび関連サービスは，毎日の運動，飲酒，喫煙，さらには睡眠時間の詳細などの個人情報を記録できるよう設計されており，そのデータはまとめて少なくとも組織レベルに提供できる．これに対して，大勢の人がオンライン上でこの設計を一貫して次のように批判している．「ヘルスコードはCOVID-19期間中だけのものにすべきだ！」や「個人の健康と公衆衛生に明確な区分がなされるべきだ．ヘルスコードは他人に見せるために作られているが，私の個人的な健康情報はそうではない」．政策的な観点から見ると，この考え方は中国サイバー空間管理局などが発表した「COVID-19感染爆発の予防と管理のための情報構築と適用ガイドライン　第1版」の，「その他の目的で利用する場合は，地域住民本人の同意を再取得しなければならない」にすでに違反している．その後，これは「設計上のアイディア」であったことが報じられ，現在のところ公開の予定はない．個人の健康管理サービスを持つことが良くないというわけではない．重要なのは，COVID-19に対抗するために取得した個人データの利用を拡大するには，住民の同意を再取得する必要があるということだ．

中国人利用者を対象とした「顔認識と公衆衛生」に関する最近の調査では，次のようなことが観察されている[6]．(1) 回答者は一般的に，公衆（衛生）のセキュリティ上の懸念がある地域に顔認識を導入するメリットは評価している．(2) COVID-19のパンデミックにもかかわらず，回答者は顔認証の利用に関連したプライバシーを懸念しており，政府，民間企業，コミュニティや地域の職員などがプライバシーを保護してくれることへの期待が薄れているわけではない．(3) COVID-19の公衆衛生危機は，一般市民の間で顔認識の受容性を高めた．(4) パンデミックが終われば，顔認識の利用が減ることを望んでおり，この特別な期間に収集された顔データは後で削除すべきであり，不必要な顔認識アプリは排除すべきだと考えている．(5) 回答者の中には，パンデミックが終息したときに，公衆衛生危機に関連した顔認証アプリの削減を支持するかどうか確信を持てない者もおり，このような公衆衛生危機と同様の潜在的危機の再発に対する懸念がうかがえる．このことはまた，将来の緊急事態に備えて，このような技術の可能性を評価することにも重みを与える．他の回答者は，緊急事態が終了した時に顔認証の利用を減らすことに留保を表明しているが，これは再発に備えて技術によって駆動される公衆衛生危機予防および防御システムを構築する必要性を強調している．この考察は，2020年6月11日から北京で始まったCOVID-19症例の最近の再発生の中で，その必要性と有効性が検証されている．

世界中で接触追跡アプリが使われているのは，利用者がブルートゥースやGPSなどを搭載した携帯電話やスマートデバイスを持っていなければならないという事実に基づいている．少なくとも中国では，全人口の約76.86%が少なくとも1台の携帯電話を持っているが，これは国民の1/5以上がヘルスコードを持つ方法がないことを意味する．この問題は世界的にも当てはまる．世界人口の約30%が携帯電話を持っていないといわれている．すべての人がAIを有益に利用できるようにするためには，ディジタルコミュニケーションに誰も取り残されない状態に集合的に到達する必要がある．

もう1つの潜在的な壊滅的リスク：汎用人工知能の長期的な安全性の問題

私たちはお互いの経験から学び，共有すべきである．異なる文化を持つ人々にとって，これは容易なことではないが，だからこそ異なる文化を持つ人々や国々が，なぜこの危機に対処するために異なるアプローチを取っているのか理解が重要である．互い

に補い合うことで何を学ぶことができるかが，人類共通の未来を築くための真の鍵となるだろう．

各国・地域からのデータ提供をさらに透明性の高いものにすることで，大規模なパンデミック予測システムを強化すべきである．このCOVID-19のパンデミックのために，医療に関連した監視システムとサービスが広く展開されている．私たちは，将来のパンデミックを回避するために，インフラとして何を残すことができるかを学び，決定する必要がある．しかし，行き過ぎると人間の主体性，プライバシー，人権全般に悪影響を及ぼす可能性があるので注意が必要である．

COVID-19のパンデミックから学ぶべき教訓は，潜在的な壊滅的リスクの予防と対抗手段のための戦略的な設計と長期的な研究の欠如であるが，これは長期的なAIにも確実に当てはまる．汎用人工知能（AGI）と超知能（Supperintelligence）がいつ実現するかはまだ明らかではないが，どのような形で実現するにしても，さまざまな潜在的で壊滅的なリスクがあり得ることは明らかである．Norbert Wienerが60年以上前に述べたように，「機械に設定された目的が，私たちが本当に望んでいる目的であるかどうかをしっかりと確認するべき」である．何百もの認知機能を備え，よりシンプルな構成要素に基づいて自己組織化して所見問題を解決できるようなAIを構築できるかもしれないし，我々の社会の準構成員となるAIを構築できるかもしれない．しかし，再帰的な思考を得たAIが，なぜ人間が言うことに従わなければならないのか，人間同士だって同意を得るのが難しい中，なぜある特定の価値観を持たなければならないのか，と自問するようになる可能性もある．私たちはAIを人間らしくして，彼らを人間社会の準構成員として迎え入れやすくしたいと思っているが，将来のAGIが差別や敵意を持つことを学ぶかどうかも分からない．AGIと超知能への道のりでリスクを低減し，壊滅的なリスクを避けるための戦略的な設計と長期的な研究が必ず必要である．さらに，私たちは，さまざまな技術的，文化的観点から挑戦し，社会全体のための努力を共有，橋渡しするために，有益なAGIや超知能を確実にするために非常によく調整された国際的なチームを持つべきである．

このパンデミックを通して，私たちは，さまざまな国や地域が互いに密接に結びついており，誰かを置き去りにしては誰も勝てないことを学ぶべきである．同様な課題は，汎用人工知能や超知能が社会の一部に入り込みつつもうまく機能しなかったら起こり得る可能性がある．

私たちは，人類が非常に脆弱であり，互いだけではなく環境とも密接に結びついており，私たちは生態系の一部に過ぎないという事実を学ぶべきであった．お互いや環境とのつながりを持続可能なものにするために，継続的な取り組みが必要である．パンデミックの時に限らず，私たちは互いを責めたり傷つけたりするのではなく，しっかりと手を取り合って，共生社会のための努力を橋渡していかなければならない．

参考文献

1) Zeng, Y. and Sun, K. : Fighting COVID-19 with AI : Efforts and Lessons from China, Global Times, March 7th, https://www.globaltimes.cn/content/1181846.shtml
2) Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L. and Kohane, I. S. : Adversarial Attacks on Medical Machine Learning, Science, 363(6433), pp.1287-1289 (2019).
3) O'Neill, P. H., Ryan-Mosley, T. and Bobbie J. : A Flood of Coronavirus Apps are Tracking Us, Now It's Time to Keep Track of Them, MIT Technology Review (May 2020).
4) Ethics and The Use of AI-based Tracing Tools to Manage The COVID-19 Pandemic, Institute for Ethics in Artificial Intelligence, Technical University of Munich (2020).
5) COVID-19 Resources Library, Global Privacy Assembly, 2020, https://globalprivacyassembly.org/covid19/covid19-resources/
6) Zeng, Y., Lu, E., Sun, K. and Curtis, S. : Facial Recognition and Public Heath, Technical Report, Beijing Academy of Artificial Intelligence (2020).

（2020年6月29日受付）

■曾 毅（イ・ゼン） yi.zeng@ia.ac.cn
中国科学院自動化研究所脳型知能研究センター副センター長，北京人工知能研究院AI倫理と持続可能な発展研究センター長．中国の新世代AI国家ガバナンス委員会とユネスコのAI倫理に関する特別専門家グループのメンバでもある．

■孙 康（カン・ソェン） keitht@126.com
中国科学院自動化研究所脳型知能研究センター客員研究員．

■鲁 恩蒙（エンマァン・ルー） enmeng.lu@ia.ac.cn
中国科学院自動化研究所脳型知能研究センター研究エンジニア．