ビッグデータ倫理

ニール・M・リチャーズ[1]、ジョナサン・H・キング

凡例・出典

本稿は以下の論文の要約である。なお、原文中でのイタリックの強調は下線にて示した。
Neil M. Richards and Jonathan H. King, “Big Data Ethics,” in *Wake Forest Law Review*, 2014, pp.393-432.

論文の背景と目的

我々は、「ビッグデータ革命」の分岐点に立っている。通話履歴や購入履歴に始まり、顔認証のデータや位置情報にいたるまで、大量の個人情報が知らぬままに企業や政府に収集・共有されており、こうした傾向がこれからますます加速していくことはほとんど不可避的だろう。ビッグデータ分析は、我々の認識を拡大し疑いようのない利便性を提供する一方で、我々が気にかけている既存の価値を便宜のために犠牲にしかねない危険性を持っている。それゆえ、こうした価値とビッグデータ利用の間の釣り合いを取るビッグデータ倫理が構築されねばならないのである。

著者のニール・M・リチャーズとジョナサン・H・キングは、この論文において、ビッグデータ社会において守られるべき 4 つの価値――プライバシーprivacy・機密性 confidentiality・透明性 transparency・アイデンティティ identity――を提示し、それらを保障する具体的方法を論じている。

論文の構成

I. ビッグデータ革命 The Big Data Revolution……ビッグデータ革命の沿革とその利益、および想定される危害について。

   A. ビッグ・メタデータ・コンピュータ The Big Metadata Computer―――情報革命の起源と急速な成長が説明され、データをコンピューティング（計算）し、あらゆるメタデータに関わるビッグ・メタデータ・コンピュータを、我々の社会が構築してきたことが記述される。

   B. ビッグデータの採り入れ―――Big Data Adoption ビッグデータが政府や企業などあらゆる組織によって採り入れられ、我々のあらゆる活動がビッグデータ分析の対象となっていることが述べられる。

[1] 本論文の第一著者であるニール・M・リチャーズはワシントン大学の法学部教授であり、プライバシー法、情報法についての世界的権威である。

C. ビッグデータによる認識———Big Data Awareness ビッグデータ革命が根本的には我々の認識についてのものであると述べられ、ビッグデータ分析によってもたらされる我々の認識の拡大が具体例を挙げて説明される。

II. ビッグデータ倫理 Big Data Ethics……ビッグデータ時代に重要となる 4 つの価値について。

A. プライバシーPrivacy

1. 情報ルールとしてのプライバシーPrivacy as Information Rules———ビッグデータ時代にあっても「プライバシーの死」は訪れておらず、むしろプライバシーの重要性は増していると主張される。本論文で筆者らは、「本人しか知らない自身についての情報の量」といった単純なプライバシー理解を批判し、情報ルール一般に関わる広い概念としてプライバシーを理解することを提案している。

2. 共有された私的情報は内密でありうる Shared Private Information Can Remain Confidential ———ビッグデータ時代における情報の守秘について。筆者らによれば、一度他者に共有された情報はもはやプライベートではありえないというこれまでの二元的な binary プライバシーの観念は退けられるべきである。プライベートな情報はたとえ第三者に共有されたとしても内密 confidential でありえるのであり、情報ルールによる保護の対象となるべきである。

3. 透明性 Transparency ———データの二次利用と透明性について。ビッグデータ分析により多くの情報の二次利用がなされており、結果として企業や政府がデータを共有する動機を与えられていることを鑑み、組織の力の悪用防止と個人の安心のため透明性が重要であることが主張される。

B. アイデンティティ Identity ———ビッグデータ分析とアイデンティティについて。筆者らは、アイデンティティを自身のありようを自分で決定する能力と定義した上で、この意味でのアイデンティティがビッグデータ分析によって脅かされうることを主張している。具体的には、ビッグデータ予測は組織による個人の監視を可能にし、我々が決断する前に我々のありようが決定されてしまう危険性がある。こうしたリスクに備えるため、いくつかの有害なビッグデータ分析は禁止される必要がある。

III. ビッグデータ倫理の保障 Securing Big Data Ethics———ビッグデータ社会において上述の価値を守るための方策が述べられる。筆者らによれば、新たな情報法を制定することに加え、倫理的原理が確立され、広く実践されることが必要である。プライバシーが顧慮される情報社会においては、エンジニアからユーザーまで万人が対話と解決に参加する一員とならねばならない。

I. ビッグデータ革命

著者らは、ビッグデータ革命 The Big Data Revolution を情報革命 Infomation Revolution の最新の段階と捉えており、情報技術一般の進歩の歴史を踏まえてビッグデータ革命の沿革を記述している。情報革命の第一幕がマイクロプロセッサと計算能力 the power to compute によって、第二幕がネットワークと接続能力 the power to connect によって定義されるとすると、第三幕はデータと予測能力 the power to predict によって定義される。

A. ビッグ・メタデータ・コンピュータ

ビッグデータの定義

はじめに、「ビッグデータ」という言葉の定義が行われている。技術的なものとしては「在来型のデータベースシステムの処理能力を超えるデータ」といった定義が用いられるほか、技術者たちはしばしば 3-V 定義――high-volume（大量の）, high-velocity（高速な）, and high-variety（多種多様な）データ資産――を用いている。こうした定義があくまでビッグデータ分析の技術的側面に焦点を当てた比較的狭い定義であるのに対して、筆者らはビッグデータがもたらすより広い社会的インパクトに注目しており、メイヤー・シェーンベルガーとキュキエによる「市場や諸機関、および市民と政府の関係などを変化させ、小規模なものでは不可能であった新たな洞察や価値の新形態の創造を可能にする大規模なデータ」[2]という定義を好意的に引用したうえで、「ビッグデータ」という語を多量のデータ資産の収集と保存、「ビッグデータ分析」という語を多量のデータ資産からなる推論と予測を指示するものとしている。

コンピュータの進歩

言うまでもなく、ビッグデータ革命に最も大きな影響を与えているのはコンピュータの進歩である。第二次大戦以降に起きたプロセッサの急速な進歩により、コンピュータの高性能化・低廉化が進んだ。「事務所の机ごとに一台、家庭ごとに一台のコンピュータを」というビル・ゲイツの標語の時代は既に通り過ぎ、スマートフォンとタブレットの時代が到来している。

政府や企業は急速にクラウド・コンピューティングを導入し始めており、同様に個人のレベルでも、コンピューティングはあらゆる物事に驚くべきスピードで広がっているのである。

アプリケーションとソフトウェア

こうしたコンピュータの力は、アプリケーションとソフトウェアの前代未聞の成長を焚き付けてもいる。2008 年のローンチ以来、Apple の App Store は総計 1000 万弱のアプリを抱えるまでに成長しているし、Android 向けの Google Play ストアも 2013 年 7 月に総計 100 万アプリという大

[2] Viktor Mayer- Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think,* 2013, p.6.

台を達成している。また、SaaS(Software as a Service)と呼ばれる、オンデマンドのまったく新たな形態のソフトウェア提供も現れている。

ネットワークの進歩

このように驚くほど広く分布しているコンピュータを接続し、増殖しつづける大量のアプリケーションを稼働させているのは、これもまた驚異的なほどのグローバルなコミュニケーションネットワークである。インターネットは急速に広まり、アカデミア、企業個人、そして今では我々の街や家の物理的デバイスを接続している。Cisco[3]が報告するところでは、グローバル IP トラフィックは直近 5 年で 4 倍に増加しているという。

電信からインターネットへの移行によって、グローバルコミュニケーションは現在海底の光ファイバーを駆け巡っており、グーグルやアマゾンなどの組織は、自社のコンテンツと経済状況をよりコントロールするため、自社用の光ファイバーネットワークを構築している。また、人工衛星や個人用の Wi-Fi など、無線技術も大きく発展している。

メタデータ

コンピュータやネットワークは主にデータを取り扱うわけだが、ビッグデータ社会を特徴づけるのは「メタデータ」、すなわちデータについての情報を記述するデータへの注目である。たとえば、アメリカ議会図書館はツイッター社と連携して 130 テラバイトものツイートをアーカイブしているが、彼らは、140 文字のツイートだけでなく、それに紐付けられた 31 ものメタデータ――日付や時間、位置情報など――をも収集している。

ビッグデータ革命以前にも、我々は長らくメタデータを用いてきた。たとえば、図書館は図書目録を作るために図書のメタデータを作成してきたし、あるいは、郵便局に手紙を配達してもらうために我々は宛先と差し出し人住所というメタデータを封筒に記入している。

しかし今日では、メタデータの活用事情が大きく異なる。上述したコンピュータ、通信技術の進歩、およびストレージの性能向上・低廉化により、莫大な労力を要していたメタデータの作成・保存を極めて低いコストで行うことが可能になり、我々のほとんどすべての行為についてのメタデータが収集・保存されているからである（たとえば、グーグル検索を利用したり、フェイスブックに投稿したりするだけで、我々はメタデータを生み出している）。更には、我々についてのメタデータはフェイスブックの顔認識システムのような商業アルゴリズムに組み込まれ、こうしたアルゴリズムをより力強いものにしているのである。

一歩下がって見つめてみると、ネットワーク接続されたデバイスと多種多様なデータを生み出すアプリケーションを駆動する、広く分布したコンピューティングは、ある種のビッグ・メタデータ・コンピュータになり始めているのである。個人や企業、政府はみなビッグ・メタデータ・コンピュータと常に触れ合っており、ハードウェアやソフトウェア、プロトコルの急速な改善により、ビッグ・メタデータ・コンピュータはよりよいメタデータを生み出しそれをより簡便に共

[3] アメリカのネットワーク機器開発会社。

有することを可能にしている。メタデータの利用は我々に新たな洞察や推論を提供しており、その利益は明らかである。それゆえ、我々が繁栄のためにメタデータを必要とし、望んでいることははっきりと認めなければならない。しかし一方で、他の新たなツールと同じく、メタデータは我々に解決すべき難問を投げかけてもいる。

B. ビッグデータの採り入れ

データ分析の初期の時代には、企業は社内で生成したデータをデータ倉庫に供給するのに莫大な時間を費やしていた。この状況を変革したのは、グーグルやヤフーが開発した新たなオープンソースソフトウェア（Hadoop など）である。こうしたソフトウェアによって、企業はこれまでよりも大量のデータを保存・分析することできるようになり、さらには社内のデータだけでなく社外のデータを利用することまでも可能になった。ビッグデータ革命の本質は、まさに扱われるデータ量の増大にあるということだ。

現代は、ビッグデータ分析がシリコンバレーを超えてあらゆる法人、政府組織に利用される第三の時代である。企業や政府は、将来的な二次利用のため、今のところはどうでもいいデータまで収集し保存している。さらにはデータブローカーも現れ始め、政府関連団体や企業、非営利団体、私人を相手取り数千億ドルの収入を生み出す一大産業となっている。

以上のように、ビッグデータはますます多くの分野に採り入れられ始め、人間のほとんどあらゆる活動———デートから投票、テロリストの特定に至るまで———に影響をおよぼすまでになっている。ビッグデータが採用されたことで、既に市民と政府、企業間の関係に影響が及ぶほどになっているが、あまりにも変化が急激なために、人々のほとんどはその規模にも速度にも気づかないままでいるのである。

C. ビッグデータによる認識

ビッグデータ革命は、根本的には認識 awareness に関わるものである。ビッグデータ分析によって、我々は世界についてよりよく知り、予測を立て、問題を解決できるようになる。

たとえば、2012 年に行われた MIT と UC バークレーの研究者による共同研究では、携帯電話の信号ログの分析によって、これまでにないほど詳細な人々の往来パターンマップを作り出し、隠れたパターン *previously hidden patterns* を明るみに出した。

テロリズム対策の分野でも、ビッグデータは、状況認識 situational awareness——多様な情報源から集められた、事件管理と意志決定の基礎を形作りうる情報——を大幅に拡大することによって、テロ攻撃に対する特効薬を提供している。こうした状況認識の拡大のために政府が必要とするのは、ただあらかじめあらゆる情報を収集しておいて、必要なときに必要なものを検索できるようにすることだけである。事件発生後には、捜査官はビッグ・メタデータ・コンピュータにアクセスすることで嫌疑をかけられたテロリストを特定することができる。実際、ボストンマラソ

ン爆弾テロ事件の際には、米司法省がボストンの通信ログにアクセスし、監視カメラと目撃写真とのクロスチェックによって犯人を特定しようとした。

加えて、将来的には、国家が通信業者のログにアクセスしたり、電話番号のデータを収集することによって、テロ攻撃の隠れたパターンが明るみに出され、テロが起きる時点を予測したり、犯人を特定することが可能になるかもしれない。このために、国家権力が国内外のあらゆる通信業者からメタデータを収集することが許される可能性がある。

また、サイバーセキュリティや日常的な治安維持業務においては、犯罪が起きる確率が高い地点を示したり、強盗に遭う可能性のある家を予測したりする、犯罪予測アルゴリズムが既に多く利用されている。

企業が商品改善のために、また政府がテロリズム・サイバー攻撃からの防衛のためにビッグデータを活用していることは驚くべきことではない。しかし一方で、人々がビッグデータによるプライバシー侵害のおそれに疑問を付すのもまた当然のことである。情報が第三者の手に渡る危険性は拭い去れず（機密性への懐疑）、そもそもビッグデータの運用方式自体が法的・商業的内密として秘匿されており、ビッグデータ予測の影響を評価することもできない（透明性の欠如）からである。

II. ビッグデータ倫理

我々の社会には既に個人情報の流出を統御するルールがあるが、これらはビッグデータ時代の新たなデータ流出、使用、そしてデータに基づく決定を統御するに足るものではない。我々は、ビッグデータ分析の持つ決定的な利得を犠牲にすることなく、この新たなツールの社会的コストを統御する新たなルールを必要としているのである。著者らは、こうしたルールを下支えする 4 つの価値としてプライバシー、機密性、透明性、アイデンティティを提唱する。

A. プライバシー

1. 情報ルールとしてのプライバシー

プライバシーの死？

1999 年の 1 月、サン・マイクロシステムズの CEO であるスコット・マクニールは、「なんにせよプライバシーなどまったくない、もう忘れよう」と宣言した。近年にも、「インターネットの父」の一人であるヴィント・サーフが「プライバシーは歴史的例外だったのかもしれない」と述べている。彼ら技術者たちは、我々はプライバシーへの期待を諦めるべきだと考えているのである。

しかし依然として、NSA（アメリカ国家安全保障局）による監視とプライバシーについての国際的議論が活発に行われている。これは、プライバシーが死んでいないことのなによりの証拠である。実際、プライバシーは決して死んではいないのである。

プライバシーの定義

プライバシーの死が訪れたかいなかは、当然プライバシーの意味するところに依存する。たしかに、単純に秘密にしておける情報の量としてプライバシーを理解するのであれば、大量の個人情報が収集される現代において我々のプライバシーは縮減している。しかし、個人情報を統御すべき規則とはなにかという問いとしてプライバシーについて考えるのであれば、プライバシーは死んだどころかむしろ我々の社会が直面する不可欠な論点である。

実際我々はどのようにプライバシーを定義しているのだろうか。公共的な議論でもっとも頻繁に用いられる定義は、「誰にも知られていない自身についての情報」といったものである。しかし、法律家はこの語をより洗練された仕方で理解してきた。たとえば、プライバシーの侵害は、(1) 保護された領域、関係、決定の侵害、(2) 情報の収集、(3) 情報の使用、(4) 情報の暴露、の 4 つを意味すると考えられている。

米国における既存のプライバシー関連ルールとしては、のぞきや写真の勝手な商業利用を禁じる不法行為法をはじめ様々なものが挙げられるが、重要な点は、プライバシーがいかに定義されようとも、それが情報に関わるものになる、ということである。プライバシーは単に秘密になっている情報の量として理解されるべきではなく、むしろ情報の暴露と同じく情報の利用をも統御するルールに関連するものと考えられねばならない。実際我々は、情報ルール一般について語るために「プライバシー」という語を用いているのであり、この用法は英語話者に根付いているのである。

個人情報の収集とプライバシー

「プライバシーの死」と言われるとき、実際に指されているであろう現象が２つある。第一に、大量の個人情報が収集されていることである。しかし、先に述べたとおりプライバシーが単なる情報収集からの保護以上のものを意味しているとすると、この事実からプライバシーが死んだと言うことはできない。むしろ、プライバシー関連規則の必要性は増しているのである。

プライバシー・セルフマネジメント

第二に、自身についての情報のやり取り・使用をコントロールする実践的能力が脅かされていることである。この能力はダニエル・ソローヴ[4]によって「プライバシー・セルフマネジメント」と名付けられたアプローチに基づくもので、既存のプライバシー関連法が焦点を当てているのもこのアイデアである。具体的には、データ処理業者が個人データの処理内容を明示する「通知 notice」と、人々が自分の気に入らないデータの使用を取り下げることができる「選択 choice」が最も重要な原理とされている。

[4] Daniel J. Solove, *Understanding Privacy*, Harvard University Press, 2008.

プライバシー・セルフマネジメントは、理念的には微妙な次元でのプライバシー保護を約束しているが、実際的には企業は擬制的な通知しか提供しておらず、選択についても個人は 0 か 100 かの合意を求められてしまっている。こうした中で、いたずらに個人にいままで以上のプライバシー・セルフマネジメントを求め大変な労力を負わせる方策か、あるいは個人の合意の効力を制限するパターナリスティックなアプローチばかりが提唱されるというジレンマが発生しているのが現状である。問題になっているのは、プライバシーが死んだということではなく、個人情報の流出を管理する既存のシステムの再考と追加の統御原理がより一層必要とされている、ということなのである。

2. 共有された私的情報は内密でありうる

プライバシーと機密性

さきほど退けた単純なプライバシー理解は、二元的 binary なものであった。つまり、情報は完全に私的であるか完全に公的であるか、すなわち自分にだけ知られているか世界中に発信されているかの二者択一だと扱われているわけである。しかし、我々の情報はほとんど二極の間の中間的な状態であるはずで、この理解は明らかにナンセンスである。我々は、信頼に基づき、他者に共有した情報を内密にしておいてもらうことを期待するのであり、機密性とは、関係性の文脈において信頼と約束への信頼にもとづくある種のプライバシーなのである。一度共有された私的情報でも内密でありつづけることは可能なのであり、我々が信頼する第三者に共有したプライベートな電子情報もプライバシー法によって統御されうる。

二元的なプライバシーの観念は、有益さのために情報が必然的に共有される我々の時代にはとりわけ危険である。たとえばデート相手を見つけるためにマッチングサイトを利用するときのように、我々はよろこんでビッグデータアルゴリズムに情報を共有する。しかし、我々の情報は残り続けるのであり、ビッグデータによって、我々自身はほとんど知り得ない我々の調書 dossiers が作成されうるようになっている。加えて、以前とは異なり、我々は自分の個人データが期待通りに用いられているかほとんど評価できておらず、情報提供に同意する際にその利得とコストを正確に比較考量できていない。プライバシー・セルフマネジメントの崩壊と技術的進歩の両方によって、セルフマネジメントの失敗によってもたらされる危害は輪をかけて大きくなっているのである。

組織の機密性と信頼

機密性に関して、意図されない帰結が生じるのは個人のプライバシーに限ったことではない。信頼の喪失という形で組織に危害がもたらされる可能性もある。エドワード・スノーデン[5]による NSA のスパイ行為の告発が出版されて以来、アメリカ政府は信頼喪失の問題を計り知れない

[5] アメリカ中央情報局の元職員で、本論文で説明されている通り NSA のスパイ行為をメディアに告発した。

コストをかけてなんとか切り抜けてきた。アップルやグーグルなどの企業も、この事件による顧客からの信頼喪失を恐れており、2013 年にアメリカ政府に監視体制の改革を求める公開文書を発表したほどである。

既存の法規定

機密性に関する法律は、特定の種類の共有された情報をプライベートに保っておくため数世紀前から存在している。たとえば、弁護士や医師はクライアント・患者から共有された情報を保護する義務を追うし、また FTC（連邦取引委員会）は、90 年代後半から、プライバシーポリシー上の違約は連邦取引委員会法に反すると述べている。こうした法制度が明らかにしているのは、我々が長らく一度共有された情報でも法という道具によってプライベートに保つことができると考えてきたという事実である。

メタデータと機密性

こうした既存の法規定の存在にも関わらず、メタデータについての法的保護は追いついていないことがしばしばであった。しかし、米国のいくつかの州裁判所と立法府はメタデータ収集がプライバシーに対して持つ含意を認識しはじめている。*Klayman v. Obama*（「Klayman 判決」）で、リチャード・レオン判事は、米国政府の通話記録メタデータの収集は合衆国憲法修正第 4 条に違反するとし、仮差し止め命令と原告のデータの破棄を求めた。これに対し、政府側は 1979 年の *Smith v. Maryland*（「Smith 判決」）に基づき、「通信会社がビジネスの記録として保持している電話メタデータについては、誰しもがプライバシーの期待を失うし、合理的な期待については言うまでもない」と論じた。レオン判事は情報技術の発展を踏まえ、30 年以上前の判決を単純に現代の事例にそのまま適用することを批判したが、Klayman 判決の二週間後には連邦地方裁判所のウィリアム・J・ポーレー判事が政府のメタデータ収集は修正第 4 条に違反しないとの判決を下すなど、法廷におけるメタデータとプライバシーの問題は解決とは程遠い。

こうした論争は、意図的に第三者に公開した情報について情報主はプライバシーへの合理的期待を失う、という Smith 判決以来の「第三者ドクトリン third party doctrine」の妥当性についてのものであり、根本的にはプライバシーの定義についての論争である。政府側は一旦情報が共有されたならばもはや保護されえないと主張しているが、この白々しい主張は情報時代の必要とも常識とも矛盾をきたすだろう。機密性という既存の法原理が十分に示しているとおり、我々は共有された個人情報を保護することができるし、また保護すべきなのである。

3. 透明性

透明性の重要性

透明性は、他者に説明責任を負わせることによって、機密性と同様、信頼を育みうる。透明性はよりよい社会を作り、市民の権利を保護するために不可欠であり、EU をはじめ、政府や企業は既にその重要さに気づいている。

透明性にまつわる諸問題

透明性に伴うひとつの問題は、公開性と秘密性の間の緊張関係である。一定の情報公開が信頼のために不可欠である一方で、あまりにも多くの情報を公開すれば組織の重大な利益が損なわれ、場合によっては他者のプライバシーを侵害し、信頼を損なう恐れもある。

もうひとつの問題は、著者が以前論じた[6]「透明性のパラドックス」である。すなわち、個人についてのありとあらゆるデータが NSA やフェイスブックなどの組織によって収集されている一方で、同時にこれらの組織は法的・商業的秘密性のもとに完全なプライバシーを実現しているという不均衡である。さらなるねじれとして、近年はアップルやグーグルなどの企業が政府に透明性を要求している一方、彼ら企業は完全な暗号化によって国家に対してプライバシーを保持しており、国家・企業間の均衡も問題となる。

データの二次利用と透明性

ビッグデータの効力の大部分はデータの二次利用によるあらたな予測に由来するから、ビッグデータ社会において透明性はとりわけである。組織は我々の知らないままに、同意もなく我々についてのデータを収集しており、望まれない二次利用のためにそれらを使用・共有している。このために、特にデータブローカーについては、既に透明性の欠如が批判されているところである。

重要なのは、個人のプライバシーと組織のプライバシーの間のバランスを取ることであり、ビッグデータ社会においては、データを収集し、利用するものはより透明性をもち、したがってより説明責任を負っていなければならない。企業が政府に透明性を要求するだけの力があれば、我々はより政府を信頼することができるし、加えて企業側が自身のデータ利用について透明性を持っていれば、我々は両者に対してさらなる信頼を持つことができるだろう。

4. アイデンティティ

根本的権利としてのアイデンティティ

アイデンティティもまた、ビッグデータ社会において守られなければならない価値である。我々が自分のあり方を決定し、自分自身の決断をするためにはプライバシー、とりわけ知的なプライバシーが必要であるが、ビッグデータ分析は組織的な個人の監視を可能にし、我々自身の決断以前に我々のありようを決定してしまうことさえありうるかもしれない。

ここでいうアイデンティティとは、哲学的な同一性などではなく、自身のあり方――自分が何者であるか、自分がなにを行うか、自分がなにを好みなにを嫌うか――を自分で定める根本的な権利を意味している。修正第 1 条を始めとする多くの米国の憲法の規定は、この意味でのアイデンティティの保護を目的としていると理解することができるだろう。

[6] Neil M. Richards and Jonathan H. King, “Three Paradoxes of Big Data,” in *Stanford Law Review Online*, Vol.41, 2013, pp.41-46.

ビッグデータがアイデンティティにもたらす影響

より詳しく言うと、ビッグデータがアイデンティティを損ないうるとはどういうことなのだろうか。マーシャル・マクルーハンは、「メディアはメッセージである」と述べ[7]、メディアや技術がメッセージはもちろん人間の思考や表現の構造そのものを変容させることを明らかにした。ビッグデータ分析は、我々の表現方法だけでなく自身についての決定すらも変容させる可能性を持っているのである。

また、政府は、我々をテロリズムやサイバー攻撃から守るために我々を絶えず監視している。こうした監視は、市民として自分自身の考えを持つ知的なプライバシーを抑圧し、我々を差別・抑圧する政府の力を強化することで、アイデンティティを損なうかもしれない。加えて、こうした規模のビッグデータによる監視が意味するのは、サービス提供者に共有された情報が内密になっていない社会に個人が生きているということである。機密性の欠如と監視の不安からなる累積的な効果は、自由な社会における個人のアイデンティティを危険に晒すものに他ならないだろう。

消費者としても、我々のアイデンティティはますますビッグデータ予測とそれをコントロールする企業によって形作られるようになっている。企業は大量の我々のデータをしばしば我々の知らぬままに、同意なく収集しており、それらは我々が予測もしない望ましからぬ二次利用にもたらされる可能性がある。

ビッグデータの力は、データ資産の二次利用によってきわめて多様な洞察と予測を生み出すことに由来するから、ユーザーがサービスを利用すればするほど、ビッグデータを活用する組織は、我々の預かり知らない二次利用によってアイデンティティに影響するデータ利用手段を得ることになる。GAFA (Google, Apple, Facebook, Amazon) をはじめとする企業は、消費者が用いるインターフェイス（スマートフォンや Kindle デバイスなど）をコントロールしており、我々のあらゆるインタラクションを詳細に記録している。そしてこれらをビッグデータ分析の対象にすることで、企業はインタラクションを個人化 personalize（個人向けにカスタマイズ）しており、これによって消費者のアイデンティティが形成されてしまう可能性がある。

組織がビッグデータを採用するにつれて、我々のアイデンティティはますます組織による予測と推論によって形作られていくだろう。多くの点で我々はこの事態を望んでおり、必要としてもいる。たとえば、グーグルなどの個人化されたサービスによって我々の生活は活気づけられるだろう。しかし、企業や政府は、ビッグ・メタデータ・コンピュータの大部分にアクセスしており、またビッグデータ分析を運用する手段とノウハウを持っているから、組織の力は個人のアイデンティティを犠牲にして増加しており、そのあり様を我々は十分に理解できていない。ビッグデータは法的・商業的秘密のもとに運用されており、レイプ被害者の特定や思春期妊娠の予測といった、問題含みのビッグデータの使用が及ぶ範囲とその本性は、明らかになりはじめたばかりであって、いうまでもなく理解などされていない。こうした理解の欠如を鑑みると、我々は一定のビ

[7] Marshall McLuhan, *Understanding Media: The Extensions of Man*, The MIT Press, 2013, p.19.

ッグデータ予測については境界をもうけ、またほかの一定の予測については利用を却下することを望むだろう。

III. ビッグデータ倫理の保障

前章で述べられた 4 つの価値をビッグデータ社会において実際に守っていくための法的・倫理的方策が述べられている。

法的方策

プライバシーについて

前述の 4 つの価値とビッグデータの利益とのバランスを取るため、第一には新たな法規則の制定が必要である。データ処理に関連する既存の法は十分とはいえないが、個人情報の運用についての原則である FIPs（Fair Information Practices）連邦情報処理標準はいまだ有意義である。FIPs に基づく法廷制度としては公正信用報告法 the Fair Credit Reporting Act があり、これは、消費者の財務情報を本当に必要としている限られた者にだけ受け取られるように保障することで、消費者の財務情報を保護するものであり、また、消費者が自身の財務情報を収めているデータベースにアクセスしそれを正す有意味な機会を持つように保障するものである。プライバシー保護を強化するためのありうる方策のひとつは、公正信用報告法の射程を拡大することだろう。

加えて、ビッグデータによる新たなリスクを処理するために、個人にデータについてのさらなるコントロールを付与することも多く提案されている。たとえば、2012 年 2 月には、連邦取引委員会が合衆国議会に対し、消費者がデータブローカーによって保持されているデータをよりコントロールできるように求める報告書を刊行した[8]。しかし、ビッグデータに対処するためには、プライバシーコントロールを強化するだけでは十分ではない。

機密性について

機密性については、ウッドロウ・ハーツォグ教授が提案している連鎖機密性 chain-link confidentiality のアプローチが注目される[9]。これは、個人情報を開示する際、契約上で情報を受取る側に情報保護を義務づけ、さらにその情報が第三者に渡る場合にはその第三者にも同様の義務を付与することで、機密性を連鎖させることを狙ったものである。しかし、こうした管理体制は制限が強くなりすぎないように注意深く設計しなければならないし、加えて、その実現可能性にも疑問符が付くかもしれない。

[8] *Federal Trade Commission, Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers,* 2012.
https://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers

[9] Woodrow Hartzog,” Chain-Link Confidentiality,” in *Georgia Law Review,* Vol.46, No.3, pp.719-741.

透明性について

透明性の適用にはさまざまな困難が伴うが、しかしその重要性からして我々は困難にひるまず法的整備を進めるべきである。ひとつのありうるアプローチは、FTC が既存のプライバシーの枠組みに加えて透明性のポリシーを加えるようプライバシー責任者に求めることであろう。透明性ポリシーが採用されれば、企業はより自由に経営できるようになり、また消費者も保護されるはずである。

メタデータと法

以上のいかなる方策についても、メタデータの保護を視野に入れたものでなければならないのはもちろんのことである。メタデータは企業の監視を容易にしてきたにも関わらず、これまで比較的プライバシー関連の制限が少なかった。メタデータが可能にするビッグデータ分析の効力を考えれば、法はメタデータの蔓延によるプライバシーの危機に合わせて発展する必要があるのである。

ビッグデータ分析の実質的制限の必要性

更に、ビッグデータ分析の予測能力、我々の行動に影響をあたえる能力を考えれば、コンプライアンス規則を制定する以上のこともなされうる。すなわち、我々はビッグデータ分析に限界を設け、また一定のビッグデータ予測については単に却下することを望むだろう。こうした分野のひとつの例は投票である。2012 年のバラク・オバマの選挙戦では大規模なビッグデータの活用がなされたほか、同年の韓国の大統領選では、大韓民国国家情報院が選挙情勢を変えるため 120 万ものツイッターメッセージを送信していたことが明らかになった。投票の重要性からして、投票について企業、国家、選挙勢力がビッグデータを用いてなしうる限界を考察する必要があるだろう。

それに加えて、ビッグデータの効力には、個人を特定し、カテゴライズし、ナッジするという側面もある。近年データブローカーがレイプ被害者のリストを販売用に作っていたことが明らかになったが、こうしたビッグデータの利用は当然許されるべきではなく、検討から外して排除すべきである。あるいは、エンジニアを性差別や人種差別に向かわせるような利用も絶対に許されない。

倫理的方策

倫理的方策の必要性

ビッグデータ倫理を保障するためには、法だけで十分であるとは言えない。(1)法はビッグデータの進歩についていけるほど素早いシステムではないし、(2)法が意図されない結果をもたらしビッグデータ革命に重荷を課してしまうかもしれないからである。実際に機能している法と最先端の技術の間には、不可避的にギャップが存在してしまうかもしれない。

それゆえ、このギャップを埋め、ビッグデータ革命を望ましいものとして保証するための最重要の方法は、情報技術周りの倫理的感受性を陶冶することである。この方策はさまざまな形態をとりうる。たとえば、情報責任者やプライバシーを専門とする弁護士、またデータセキュリティコンサルタントなどの情報分野の専門家たちは既に活躍し始めているし、加えて、ほかの分野の専門家の活躍も期待される。たとえばグーグルは社内で哲学者を雇っており、ほかにもデータサイエンティストやエンジニアも活動を始めている。

ユーザーの役割

ユーザーもまたこれからの世界に対して責任を負っている。過去にはユーザーの個人的な選択ばかりが注目されたが、これからは望ましくない結果が生じた際に積極的なフィードバックを行うという形で役立つことができる。組織が、プライバシーポリシーといった透明性のポリシーを持っていれば、ユーザーはどこに自身の懸念していることを送信するべきか知ることができるし、翻って組織はユーザーの苦情に即座に反応し、持続的なビッグデータ利用を改善することができる。しかしながら、ユーザー自身では作ることができず、反対に、ますますユーザーがそれへの依存を高めている技術やビジネスについて、ユーザーだけではその責任を取ることはできない。

技術者の役割

技術者達は、急速に変化するこの時代のパイオニアであり、しばしば他者よりも早くビッグデータによるプライバシーのギャップに気づくだろう。彼らは、「プライバシーの死」神話を反駁し、ビッグデータ倫理を前進させることで、このギャップを埋める先陣を切ることのできる存在である。この動きは実際に始まっており、一連の 7 つの情報原理と最良の実践からなる、「プライバシー・バイ・デザイン」が、法学者やテクノロジーの主導者によって支持されている。プライバシー・バイ・デザインの基本的な考えは、プライバシーは政府による監督のみによっては保障されえず、効果的なプライバシー保護のためには、企業も個人のプライバシーを尊重するということが必要とされるということである。

技術者たちはまた、新たな技術やビジネスモデル、そして最良の実践を生み出すこともできる。プライバシー関連のスタートアップ企業は投資の対象となりつつあり、たとえば Personal.com は、個人データを保護するためのロッカー（ストレージ）を提供しており、個人データを個人の利益を目指した上で収益化している。

ビッグデータ実験の倫理

最後に、ビッグデータはその本性からして実験を必要とするが、実験は十分な説明に基づいた協力的なものでなければならない。カロ教授は[10]、ビッグデータを扱う実験を制御するために、ヒト対象研究の倫理的問題の解決を目指すヒト対象審査委員会（human-subjects review boards）において既に確立されてきた原則をモデルとした、消費者審査委員会（consumer review boards）の

[10] M. Ryan Calo,” Against Notice Skepticism in Privacy (and Elsewhere),” in *Notre Dame Law Review*, No.87, Vol.3, pp.1027-1072.

設立を提案している。カロ教授が提案するこのような審査委員会は、実験の結果生じうる危害を最小化したり、不公正を生み出したりしないための有効な施策であるだろう。

結論

我々はいま、人類史上最もクリエイティブでエキサイティングな時代に生きていると言えるかもしれない。他の新たな情報技術と同じく、ビッグデータはイノベーションと莫大な利益をもたらすが、一方でプライバシーや機密性、アイデンティティにとっての脅威であることも確かである。こうした事態を統御するためには、手続き的な制限に加えて実質的な制限も必要になるだろう。ビッグデータのもたらしうる危害に屈することなくその利得を享受するためには、ビッグデータ倫理の発展が不可欠である。

（鈴木英仁）