

## データ・フェミニズム<sup>1)</sup>

キャサリン・ディグネイジオ

行政や企業のサービスにおけるデータ利用が進むにつれ、データ・サイエンティストにとって——また仕事でデータに頼る人々にとって——、その不平等な生産、応用の不均衡な方法、個人と集団への不平等な効果というものがますます無視できないものとなっている。だが、このような権力ゆえに次のような問いを立てる意味もある。すなわち誰による、誰のためのデータサイエンスなのか、誰の利益を念頭に置いたデータサイエンスなのか、という問いである。これらは筆者がデータ・フェミニズムと呼ぶもの、つまり過去数十年のインターセクショナルなフェミニズム運動と批判的思想に基きデータサイエンスとそのコミュニケーションについて考える方法から生まれた問いのいくつかである。本稿では、行動するデータ・フェミニズムを説明し、いかにして男／女の二分法への挑戦が他の垂直的な（かつ実証的に間違った）分類システムに挑む上で役に立つのかを示す。またいかにして感情の理解が効果的なデータ・ビジュアライゼーションについての考えを拡張することができるのか、いかにして見えざる労働の概念が自動化されたシステムが必要とする重要な人間の努力を露呈させることができるのか、なぜデータは決して「自明のもの」ではないのかについて論じる。本稿の目的は、データ・フェミニズムのプロジェクト同様、いかにして学術研究は行動に転換されるのか、つまり、より倫理的で平等なデータの実践を想像するためには、いかにしてフェミニズム的な思考は操作可能であるか、そのモデルを呈示することである。

キーワード：フェミニズム、データサイエンス、ジェンダー、アルゴリズム、バイアス

## はじめに

現在、筆者はマサチューセッツ工科大学（MIT）の都市研究・計画学科の助教を務めている。「データ・プラス・フェミニズム・ラボ」という研究室を主宰し、テクノロジーの構築やデータサイエンスに取り組んでいる。ラボの使命は、人種とジェンダーの平等を推進することだ。2020年にMIT Pressから『データ・フェミニズム (*Data Feminism*)』という本を出版したが、共著者のローレン・クライン（エモリー大学の英語と量的調査法の准教授）もまたデジタル・ヒューマニティーズのラボを主宰している。

本稿では、この本の中で私たちがどのような意味でフェミニズムという言葉を使っているのか、その定義に関する説明から始めたいと思う。なぜならフェミニズムには様々な種類があり、どのようなフェミニズムについて私たちが言及しているのかを明らかにしたいからだ。その上で、同書の中で取り上げた事例のいくつかを紹介し、私たちが提示したデータ・フェミニズムの原則を論じることとする。

人種差別的・性差別的・階級差別的なデータ製品が作られる昨今、データ・フェミニズムとは、企業や政府のアクターに説明責任を果たさせようとする、拡大しつつある取り組みの一つだといえる。例えば、有色の女性を認識できない顔認証システム、女子校卒業生にとって不利な判定を下す人事採用アルゴリズム、貧困層の親を罰する判定を下す児童虐待検出アルゴリズムなどがある。このような人種差別的・性差別的なデータサイエンスの製品の事例には枚挙にいとまがない。こうした差別の事例は日々のメディア報道でもしばしば取り上げられていた。それが、筆者がローレン・クラインと『データ・フェミニズム』という本を書こうと思ったそもそものきっかけである。というのも私たちはこう自問したからである。アルゴリズムが性差別をするなんて、データベースが人種差別的だなんて、と。技術的なシステム誤作動の原因を探るとともに、問題解決のためにはどうすれば良いのかも考えたかったのである。

一方、企業はどうか。ソーシャルメディア上のミーム（ネットで広まる情報）のような「データは、新たな石油である」ということを、私たちは批判的に捉えなければならない。企業はデータを未開発の天然資源のように見ているのだから。石油の比喩は、データを未開発の天然資源と同様にみているということである。これは、石油や水と同様、データは抽出・加工・精製されれば利益をもたらすということを意味している。

しかし、女性、特に有色の女性、先住民の人々や、移民、LGBTQの人々は、従来の抑圧同様にこのデータ抽出のプロセスを経験している。つまり「略奪」

が起きているのである。新しいテクノロジーの恩恵は全ての人に平等に分配されてはいない。データ駆動型のテクノロジーが実際に行なっていることは、昔からある、鬱屈した状況や不平等を加速させ、悪化させているのだ。このシステムは、背後で作動し、状況を更に悪化させているのである。

筆者らが同書で検討したのは、フェミニズム、特にインターセクショナル・フェミニズムが、長い間、様々な抑圧とそれを引き起こした権力の力を解体しようとしてきたことだ。このようなインターセクショナル・フェミニズムから、今の時代のデータサイエンスは多くを得ることができる。

## データサイエンスにフェミニズムを取り入れる

2014年のMTV ミュージック・アワードで、歌手のビヨンセが舞台上に登場した時、その背後に「フェミニスト」の文字が投影されていた。彼女が歌った「フローレス」という曲ではフェミニズムについて言及している節がある。アメリカの代表的な辞典の一つである『アメリカン・ヘリテージ・ディクショナリー (*The American Heritage Dictionary of the English Language*)』の定義を引用して、フェミニストとは、男女同権、そしてノンバイナリーの人の平等な権利を信じる人であると歌っているのだ。

これに倣うならば、フェミニズムとは平等を信じることだといえる。しかし、もしあなたが平等を信じ、本当に平等という考えを真剣に受けとめるのであれば、私たちが生きている世界がまだ平等ではないことに気づいているはずだ。ゆえにフェミニズムには二つ目の定義が必要となる。辞書の定義にもある「組織的な活動」というものである。平等を信じるならば、私たちは女性やノンバイナリーの人々の権利と利益のために、平等な世界を実現するための行動を起こさなければならない。つまりフェミニズムとは、信念であり政治的行動なのである。

そして第三の定義もある。一連の理論とアイデアとしてのフェミニズムだ。この定義は特に刺激的である。なぜならフェミニズムが知的遺産であることを意味するからである。私たちは、研究者や活動家、作家など、先達の多くの素晴らしい人々からフェミニズムを受け継いできた。ゼロから発明する必要はないのだ。

フェミニズム理論は、セックスとジェンダーに関する不平等の問題を考えることから始まった。しかし過去40年間の研究と現在の政治的現実により、人種、階級、セクシュアリティ、障害などを含め、不平等のより多くの次元が議論されるようになり、インターセクショナル・フェミニズムの考えにたどり着いた。

そして現在、フェミニズムはインターセクショナルなものとして理解されなければならない。

インターセクショナリティ(交差性)とは、法学者のキンバリー・クレンショーが作った言葉である。彼女は、この言葉を使い、社会的不平等がジェンダーのような一面的な差異だけでは説明できないことを述べた。人種差別は階級差別、植民地主義と組み合わせられているのだから、不平等や抑圧について論じる際には、それを生み出す多くの要因や力の交差について検討しなければならないというのだ。

インターセクショナリティについては、重要であるにもかかわらず見過ごされがちな点がある。それは、この用語が「私は白人です」「私はシスジェンダーです」「私は女性です」といった個人のアイデンティティの特徴を単に記述するだけでなく、性差別、家父長制、人種差別などの権力の構造的な力と、不平等な効果を生み出すそれらの交差を説明しているということだ。こうした議論は、有色の女性のフェミニスト、特に黒人のフェミニストの作業によるものである。

インターセクショナル・フェミニズムは、女性やジェンダーだけの話ではない。インターセクショナリティの議論では、誰が権力を持ち、誰が権力からシステムティックに排除されているかということを考える。今日の世界では、データは権力である。データは新しい石油であり、データはすでにある抑圧であると言われる。筆者らは、インターセクショナル・フェミニズムは、データサイエンスに適用することで、不平等な権力に挑みそれを変える上で役立つと考える。不当な権力や不均衡の転覆を望むのであれば、データサイエンスにはフェミニズム、特にインターセクショナル・フェミニズムが必要なのだと。

筆者と共著者のクラインは『データ・フェミニズム』の執筆にあたり、これまでの経験から学んだこと、特に教育と学術、様々なコミュニティ活動に参加したことで得たフェミニズム的な思考について考えた。文献調査を行って、フェミニズムがこれまでどのように応用されてきたのかについて調べた。例えば、法律や制作においては、フェミニスト・テクノロジー、フェミニスト・マップ(下記参照)など、実に色々な形で取り入れられてきたことがわかった。そして調査の結果、データに関するフェミニズムの最も重要な側面が7つの原則に集約できることがわかった。すなわち(1)権力を検討する、(2)権力に挑む、(3)バイナリーとヒエラルヒーを再考する、(4)感情と身体を高める、(5)多元性を大切にする、(6)文脈を考える、(7)労働を可視化する、の7つである。『データ・フェミニズム』が7つの章から構成されているのはそのためである。章ごとに一つの原則を取り上げているのだ。

また各章でフェミニズムの思想家を取り上げている。例えば、性別と人種の交差について論じたキンバリー・クレンショーや支配のマトリクスについて論じたパトリシア・ヒル・コリンズである。フェミニズムを知らない人達にもわかりやすくすることを心がけると同時にデータサイエンスの知識も前提にはしなかった。また、今この時代にここで実際に取り組んでいる人々の具体例を見せることも心がけた。データ・フェミニズムの原則は実践的なものであり、ジェンダーと人種の正義と解放に向けて、データサイエンスの実践をどのように再構築するかということである。その目的は、データサイエンスにフェミニズムを取り入れ、既にデータを扱っている人々の作業の指針となるようなモデルを提供することである。執筆にあたっては、データを扱いたい人だけでなくそれを拒絶したい人をも頭頭に置いた。なぜならフェミニズム的な拒絶は時には最も重要なデータサイエンスの行為となるからである。

## 見当たらないデータセットの問題

深刻な社会問題に関するデータをみると、そこにいるはずの人が存在しないことがある。死傷したトランスジェンダーの人々やヘイトクライムの事案、犯罪歴があるために公営住宅から排除された人々などがそうだ。政治的な意思が統一されていないため、どの当局もそうした情報の収集を行っていない。したがって、そのようなデータセットは見当たらず存在しないのである。

「見当たらないデータセットの図書館」(The Library of Missing Data Sets)というアート作品がある<sup>2)</sup>。ミミ・オヌオハ氏の作品だ。展示場所にはファイルキャビネットが置かれている。その中には多くのファイルフォルダが入っており、フォルダにつけられたタイトルは、見当たらないデータセットを指す。鑑賞者は実際にキャビネットを開けてフォルダ内を閲覧できるが、データセットが見当たらないのでフォルダは当然、空である。

そのフォルダに入るべきデータがないのはなぜだろうか。例えば、なぜアメリカの妊婦の死亡率の情報は、連邦レベルで包括的に収集されていないのだろうか。それは、今日の世界ではデータ収集において深刻な権力の不均衡があるためである。このような権力の不均衡により、どのデータが収集され、どのデータが収集されないかが決まるのであるが、政府はそのような権力の保持者なのである。

企業や金融機関もそうであるが、往々にして女性を含むマイノリティ集団にはそうした権力がない。数えられ、測定されるものは重要になるが、そうされないものは重要とはみなされなくなってしまうのだ。

データセットやアルゴリズムのバイアスの問題に取り組むことは不平等を暴くうえで非常に重要な作業であるが、その手前の段階で、すなわちデータセットの存在・不在以前に、何を収集するかしないかの選択のレベルで、既にバイアスがある。データサイエンスに対するフェミニズムのアプローチが権力の分析から始まるのはそのためである。

データの欠如のもう一つの例に、メキシコにおけるフェミニサイドがある。フェミニサイドとは、ジェンダーに関連した、シスジェンダーやトランスジェンダーを含む女性や少女の殺戮のことである。アメリカでは女性に対する暴力は親密なパートナーによる暴力と呼ばれることが多いが、実際にはそれだけではない。女性や女性のアイデンティティを持つ多くの人々がジェンダー関連の理由で殺されており、ドメスティック・バイオレンスや親密なパートナーによる暴力に留まらないのだ。

現在、フェミニサイドは多くの国で犯罪と定義されており、中南米18カ国ではフェミニサイドに関する法律がある。これは中南米のフェミニズム運動の力による部分が大きいが、実際には国家はフェミニサイドのデータを組織的に収集してはいない。法律があっても、データ収集がきちんと行われていないのであり、また法律に則ってきちんと手続きがなされているのかどうか、その確認もできないのである。

## データサイエンスを使って抵抗する

データセットの欠如という問題に関しては、データないしデータサイエンスを使って不均一な権力構造に抵抗するという事例もある。

Maria Salguero氏のメキシコでの取り組みについては、『デジタル・フェミニズム』でも取り上げている。Salguero氏は、ニュース番組を基に1日2～4時間かけてフェミニサイドの事件をグーグルマップに記録し始めた人物である。フェミニサイドに関するデータを独力で集めるということをかこれ5年も続けた結果、彼女の作ったデータはメキシコのフェミニサイドの最も広範な公的資料となった<sup>3)</sup>。彼女はこのデータベースを用い、大勢の人が愛する家族を探し出す手助けをするとともに、ジャーナリストやNGOにデータを提供し、メキシコ議会で何度も証言した。

これは「フェミニズムの対抗的データ (feminist counter-data)」だといえる。すなわち、国家や機関が国民の基本的安全の確保に組織的に失敗した際に介入するアクティビズム的なデータ収集である。この事例は、データやデータサイエンスを使って権力に挑む方法があることを示している。

同時に私たちは次の点を忘れてはならない。すなわちデータの欠如や不平等などの問題に関する対抗的データ (counter-data) を収集する際に、より多くのデータを収集するだけでは必ずしも解決にはならないことである。脆弱な立場にある人々の場合、国家機関の目に留まることで、彼らに危害が及ぶ可能性もある。アメリカ合衆国の場合、不法移民にとって自らの身元確認情報とともにデータベースに入れられることは有益ではない。定量的にデータを使って問題を可視化したいと思っても、それが人々を危険に晒す結果になることもあるのだ。データサイエンスを使ってマッピングを行う際には、このことも理解しておく必要がある。

## データ・ビジュアライゼーションを使った新たな試み

筆者らは「バイナリーやヒエラルキーの再考」について論じてきた (著書第4章参照)。それはフェミニズム理論がバイナリー、特にジェンダー・バイナリー (性別二元制) について活発な議論を展開してきたことによる。

アメリカでは一般的には性別には男女の二つがあるとされているが、フェミニズム理論はこの考え方に反対してきた。もしこれが偽のバイナリーであれば、つまり性別は二つではなく複数あるとすれば、性別は二つしかないという考え方自体がそもそも実証的にも間違っている。また、二つの性別の間のヒエラルキーの問題もある。これはフェミニズム思想にある種の青写真を提供し、時間をかけて真剣に検討されてきた。つまりフェミニズム理論は、ジェンダー・バイナリーを批判し、自然／文化、主体／客体、次の例で取り上げるような理性／感情等、さまざまな人為的な二分法の問題を論じてきたのだ。

英米・西洋の文脈では、とかく理性は感情に優ると教えられてきた。データサイエンスやデータ・ビジュアライゼーション、コミュニケーションでもそうである。データ・ビジュアライゼーションのベスト・プラクティスには、クリーンなデザインで最小の美意識かつ「単なる事実」のみといったアプローチがしばしば関わっているのだ。しかしなぜそれらがベスト・プラクティスなのか。それは感情的にもレトリックの面でも、いわゆる中立的・美的選択と解釈されており、「グラフ」というものは実際より信頼できると思われがちである。だが、感情を後押しすることを意図されたビジュアライゼーションには問題があるのだろうか。

このことを考える上で興味深い事例がある。とあるデータ・ビジュアライゼーションの動画である。アメリカ合衆国における銃による年間死亡数のデータを使い、その年に銃で殺された人ひとりひとりを一つの弧で表したものであ

る（2022年5月30日現在、<https://guns.periscope.com>で閲覧できる）。オレンジとグリーンがひとりひとりの命を表している。その線はゆっくりと動き出し弧を描いており、スクリーン上に痕跡を残している。その動きが次第に加速されていく様子に、みる者は圧倒される。心が揺さぶられ、耐え難いほどである。

このビジュアライゼーションは、アメリカにおいてあまりに多くの人が銃で殺されていること、またそれがあたかも伝染病であるかのように全米で広く起きていることを表している。このビジュアライゼーションは、方法論的には他の研究に劣らず正統な統計学的手法によって作られたものである。しかし、このビジュアライゼーションが公開されたとき、ビジュアライゼーションの分野では疑惑をもって受け止められた。なぜならそれが私たちに「感じ」させるからである。

このビジュアライゼーションでは、スクリーン右上に、命を奪われた人々が生きていたとしたら過ごしたであろう人生の年数が表記されている。コンピュータによる多変量解析によって、亡くなった人々が愛する人々と過ごしたかもしれない年数が示されているのだ。もしも命を失わなかったら、銃による暴力で人生が途中で終わることがなかったら、という悲嘆や喪失感といった感情が、このデータ・ビジュアライゼーションによってもたらされているのである。

フェミニズムのアプローチでは、感情を生じさせることは全く問題ではない。このビジュアライゼーションは感情を揺さぶる動画であるが、それは理性と感情の両方が混じり合う体験を引き起こすからである。感情と理性を再調整することにより、我々はデータ・コミュニケーションのツールボックスを開き、デザインプロセスにおいて重要なこと——すなわち文脈を重視すること、経験に耳を傾けること、私たちが直面する不均衡な権力の問題に挑戦する行動を起こすこと——に焦点を当てることができる。

## おわりに

『データ・フェミニズム』において、筆者と共著者のクライン氏は、データ・ビジュアライゼーションやデータサイエンスの定義の拡大を主張している。これまでのデータサイエンスの対象範囲は狭く、またデータサイエンスとは、その作業をする人々の資格で定義されるものでもない。狭い定義により、女性や有色の人々は長い間データサイエンスの分野から排除されてきた。またデータサイエンスにおいては、社会技術的な作業よりも技術的な作業の方が評価されてきた。しかし、データサイエンスの定義を拡大し、かつ正義のためのデータ

サイエンスを意図すると、データサイエンスと正義の交差にある最も刺激的な作業の一部が見えてくる。従来データサイエンスが行なってきたような取り組みもあるが、それ以外にも、既にアーティスト、ジャーナリスト、ヒューマニスト、コミュニティ・オーガナイザー、司書、活動家らが着手しているが、彫刻やギャラリー、あるいはデータジャーナリズム、すなわちデジタルおよびマルチメディアを使った新たな取り組みもある。例えば、驚異的なデータ壁画（ミュラル）のような、グループデータ・セラピーを作品にしたものがある（具体的な事例については、Catherine D'Ignazio and Lauren Klein, 2020, *Data Feminism*, MIT Press参照）。

今日、データが非常に多くの問題の根源となっている。筆者は、このことを認識する一方で、解決の一端を担う可能性もあると信じている。そんな思いから、アーティストでもある筆者は、データ・フェミニズムの原則を表すインフォグラフィックスの作品を制作した<sup>4)</sup>。この作品では、複数言語にデータ・フェミニズムの原則を翻訳し、それをデータとして使っている。

筆者らのデータ・フェミニズムの研究が、日本でどのように受け止められるのか、楽しみである。今後、議論が更に展開することを期待したい。

（きゃさりん でいぐねいじお マサチューセッツ工科大学）

## [注]

- 1) 本稿は、国際ジェンダー学会2021年大会基調講演の内容を再構成したものである。日本語原稿の作成にあたっては、本基調講演の企画者・司会であり、大会資料の日本語訳（講演原稿及び発表スライドの和訳）に関わった田中洋美さん（明治大学）の協力を得た。なお、本稿で論じているデータ・フェミニズムに関する詳細は、MIT Pressから出版された筆者とローレン・クライン氏の共著『データ・フェミニズム (*Data Feminism*)』（2020）を参照のこと。同書は以下のリンクで閲覧できる。<https://data-feminism.mitpress.mit.edu/>（2022年5月30日閲覧）
- 2) 作品の詳細については、ミミ・オヌオハ氏のホームページ（<https://mimionuoha.com/>）内の作品専用ページで知ることができる。  
The Library of Missing Datasets: <https://mimionuoha.com/the-library-of-missing-datasets>（2022年5月30日閲覧）
- 3) Maria Salguero氏が作成したGoogleMapは以下のリンクで確認できる。  
<http://mapafemicidios.blogspot.com/p/inicio.html>（2022年5月30日閲覧）
- 4) 2022年5月30日現在、以下のリンクで閲覧できる。  
<http://datafeminism.io/blog/book/data-feminism-infographic/>

# Data Feminism

Catherine D'Ignazio

(Massachusetts Institute of Technology, USA)

As data are increasingly mobilized in the service of governments and corporations, their unequal conditions of production, their asymmetrical methods of application, and their unequal effects on both individuals and groups have become increasingly difficult for data scientists--and others who rely on data in their work--to ignore. But it is precisely this power that makes it worth asking: “Data science by whom? Data science for whom? Data science with whose interests in mind?” These are some of the questions that emerge from what we call data feminism, a way of thinking about data science and its communication that is informed by the past several decades of intersectional feminist activism and critical thought. Illustrating data feminism in action, this paper will show how challenges to the male/female binary can help to challenge other hierarchical (and empirically wrong) classification systems; it will explain how an understanding of emotion can expand our ideas about effective data visualization; how the concept of invisible labor can expose the significant human efforts required by our automated systems; and why the data never, ever “speak for themselves.” The goal of this paper, as with the project of data feminism, is to model how scholarship can be transformed into action: how feminist thinking can be operationalized in order to imagine more ethical and equitable data practices.

**Keywords:** feminism, data science, gender, algorithms, bias