

差分プライバシーや公平性などの 倫理的なデジタル社会構築に向けた先導的な研究への貢献

シンシア・ドワーク博士

1958年6月27日生まれ(67歳)

ハーバード大学コンピュータサイエンス 教授

デジタル社会の利便性と引き換えに 増加する危険

私たちの社会は、インターネットや AI、ビッグデータなどの技術革新によって、経済活動や日常生活のあらゆる場面でサイバー空間をより多く利用する「デジタル社会」へ急速に移行しています。その利便性や効率性の背後には、膨大なデータやその解析結果が存在します。この広く集められたデータの中に個人情報や機密情報が直接的あるいは間接的に含まれているため、漏洩や監視が懸念されます（図1）。

また、これらの膨大なデータから有用な情報を抽出するために行った操作により、「AI による差別的判断」や「アルゴリズムによる市場支配」などが起こり、倫理的・社会的な課題が深刻化しています。特に、プライバシーの侵害やサイバー空間の公益性が低下し、本来、社会全体が享受すべき利益を一部の企業や国家が独占することにより、政治的あるいは資本主義経済に構造的な歪みが生まれています。また近年、生成 AI に代表される革新的技術がこの問題に拍車をかけています。

デジタル社会において私たちは「情報の偏在」、「倫理的責任の所在」、「社会的公平性の担保」など、従来の法制度や倫理基準では容易に解決できない課題に直面しているのです。



図1：デジタル社会における懸念
サイバー空間を通して、私たちは気づかぬうちにAIや権力による支配や監視、差別を受けているかもしれない。

個人情報漏洩リスクを見積もる 「差分プライバシー」

デジタル社会における倫理的課題に対して、ドワーク博士は数学的に精緻で理論的な枠組みを構築し、まったく新しい学問領域を切り開きました。特に、2006年に提案した「差分プライバシー」は、ビッグデータ時代における個人情報保護の在り方を根本から変える内容でした。

差分プライバシーとは、データ分析の結果が、特定の個人のデータを含んでいてもいなくてもほとんど変わらないことを示すことで、データに含まれる個人情報が表に出ないことを保証する数理的定義です（図2）。重要なのは、統計データを公開した際に個人情報の漏洩リスクを数学的に評価できる点です。

さらに、統計処理結果に適度な「ゆらぎ（ノイズ）」を意図的に加えれば、統計処理の結果の有用性は失われず、同時に求められるプライバシー保護の水準を満たせることを示したことから、技術や製品への応用が進みました。すでに Apple、Google、Meta、Microsoft、NTTドコモなど世界の主要 IT 企

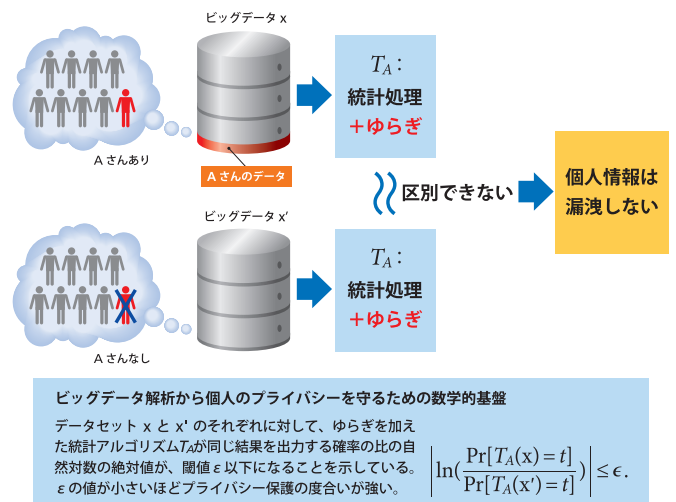


図2：個人情報を守る「差分プライバシー」
ビッグデータxとx'が1人分しか異なっていないとき、両者の統計処理結果を区別できなければ、異なる1人のデータが明らかにされることはない。すなわち個人情報は守られる。この発想から、統計処理結果に個人情報漏洩しないために加える適度な「ゆらぎ」が、数学的に定義される。

業のさまざまなサービスをはじめ、米国の 2020 年の国勢調査に採用されており、企業や行政は利用者のプライバシーを守りながら、社会に必要な統計情報を取得しています。

暗号資産の運用に採用された Proof of Work (PoW)

すでに 1992 年にドワーク博士は、将来、迷惑メールが氾濫することを予見し、それを防ぐ仕組みとして計算コスト負担を提案して注目されました。メール送信や取引記録の作成などの業務を実行する際に一定の計算作業を課して経済的コストを発生させることで、不用意な業務の実行を防ぎます(図 3)。このアイデアはその後、「Proof of Work (PoW)」として広く知られるようになり、2009 年に登場した暗号資産を運用するブロックチェーン（改ざん困難な分散型台帳技術）に採用されています。これは銀行のような特定の中央管理者を介さなくても、関係者が取引台帳を共有すれば信頼できる金融取引ができることを示したもので、まったく新しい民主的な情報共有システムが成立したのです。

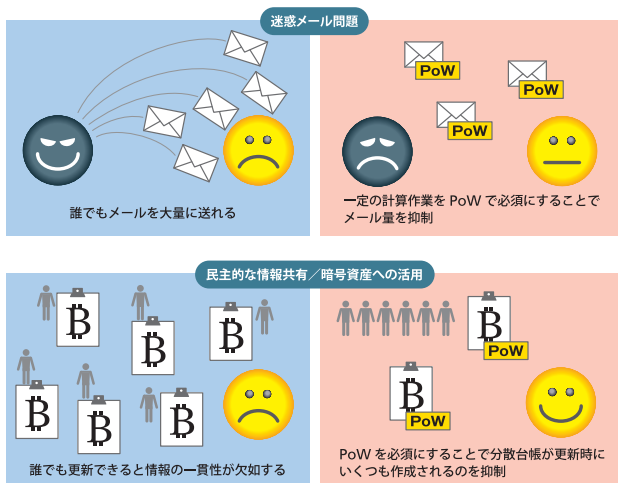


図 3：Proof of Work (PoW) の果たす役割
メール送信（上）や取引記録の作成（下）などの業務を実行する際に一定の計算作業を必須にすることで、悪意のあるなしに関わらず不用意な業務の実行を防ぐ。こうしてデジタル社会の秩序を守る。

安全で公平なデジタル社会を目指して

AI が人種・性別・年齢などの属性によって社会的に妥当ではない判断を下す可能性を危惧し、2011 年からドワーク博士は「アルゴリズムの公平性 (Fairness through Awareness)」という視点で、公平性を数理的に定義し保証する枠組みづくりを始めています。さらに、暗号通信の安全性をより厳密に保証する枠組みも確立しました。

いずれの研究も、IT 技術の発展によって成長するサイバー空間での経済活動やアルゴリズムの運用に潜む社会的・倫理的なリスクを先取りし、社会に深刻な影響が及ぶ前に、数理科学に基づく理論的解決策を提示したものです。こうした理論的成果は抽象的な思索にとどまらず、市民一人ひとりのプライバシーと安全を守りながら、世界的な情報基盤や経済システムの信頼性を支える中核理論としてすでに広く機能しています(図 4)。

シンシア・ドワーク博士は、真に世の中に役立つ研究を行うため広範な共同研究をしており、そこでは多くの人材が育っています。そして今後も、この研究者たちとともに複雑化するデジタル社会の秩序を守る先導的な力になって貢献してくれることでしょう。

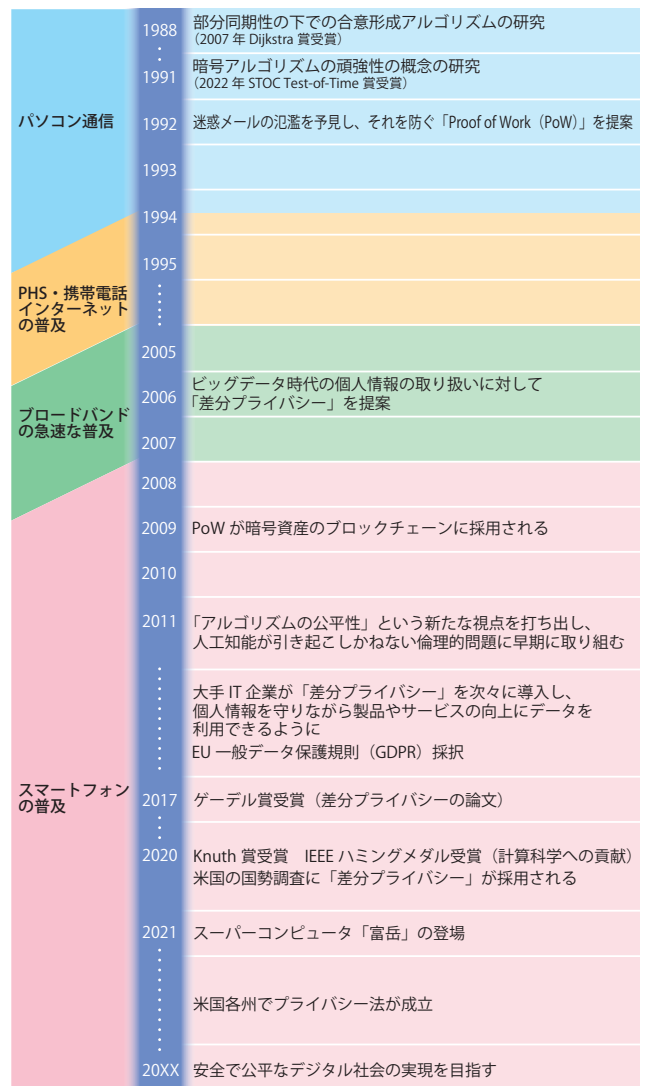


図 4：デジタル社会の進展とドワーク博士の業績
デジタル社会が形成されつつあった 1990 年代のはじめ、ドワーク博士は近い将来に起こるであろう倫理的な課題を予見し、それを防ぐための厳密な数理モデルを提案するなど、安全のための理論的な基盤を築いてきた。現在は、本格化する AI 社会における倫理問題に取り組んでいる。