



Google AI 倫理チームをめぐる騒動

1 要旨

本文書はウェブ記事をもとに Google の AI 倫理チームをめぐる騒動についてまとめたものである。

はじめに騒動が発生した 2020 年 12 月から 2021 年 2 月までの出来事について確認する。続いてティムニット・ゲブル氏、マーガレット・ミッチェル氏の解雇を見ていくこととする。

2 キーワード

- AI 倫理
- 論文レビュー

3 タイムライン

- 2020 年 12 月 2 日¹
 - 黒人女性で AI 倫理研究者であるティムニット・ゲブル (Timnit Gebru) 氏が Google から解雇されたとツイッター上で発表した。
- 2020 年 12 月 3 日²
 - ジェフ・ディーン (Jeff Dean) 氏が Google 研究チームにゲブル氏の解雇に関して述べたメールを送信した。
- 2021 年 2 月 18 日^{3,4,5}
 - 倫理的 AI (Ethical AI) チームに代わるとみられる「責任ある AI に関する新専門知識センター (a new center of expertise on responsible AI)」の開設を

¹ 佐藤由紀子「Google を解雇された AI 倫理研究者が指摘していた「大規模言語モデル」の危険性」, *ITmedia NEWS*, 2020 年 12 月 7 日, <https://www.itmedia.co.jp/news/articles/2012/07/news090.html>。

² “About Google’s approach to research publication”, <https://docs.google.com/document/d/1f2kYWDXwhzYnq8ebVtuk9CqQqz7ScqxhSIxeYGrWjK0/edit#heading=h.aplcvu32myqt>

³ Richard Nieva (CNET 編集部訳)「グーグル、AI 倫理に取り組むチーム率いるもう 1 人の責任者を解雇」, *CNET Japan*, 2021 年 2 月 22 日, <https://japan.cnet.com/article/35166792/>。

⁴ 佐藤由紀子「Google、AI 倫理チームの共同リーダーを解雇 行動規範違反を理由に」, *ITmedia NEWS*, 2021 年 2 月 22 日, <https://www.itmedia.co.jp/news/articles/2102/22/news085.html>。

⁵ 佐藤由紀子「Google、倫理的 AI チームに代わる新チームとその黒人女性リーダーを発表」, *ITmedia NEWS*, 2021 年 2 月 19 日, <https://www.itmedia.co.jp/news/articles/2102/19/news098.html>。

発表し、Google のエンジニアリング担当バイスプレジデントのマリアン・クローク (Marian Croak) 氏が新チームの責任者に就任した。

- クローク氏は発表文で「AI に関する規範を標準化することについては多くの意見の相違や対立があり、時には二極化する可能性がある。私がやりたいのは、今よりも良い方法で会話し、この分野を前進させていくことだ」と述べ、また、「AI に関する規範を標準化することについては多くの意見の相違や対立があり、時には二極化する可能性がある。私がやりたいのは、今よりも良い方法で会話し、この分野を前進させていくことだ」とも述べた。
- マーガレット・ミッチェル (Margaret Mitchell) 氏の上司の上司にあたる AI 部門のトップ、ディーン副社長はゲブル氏退社に関する調査を終え、ポリシーの変更や論文発表のプロセスの合理化、従業員退社に関する新たな手順の制定などについて説明をした。
- これに対してゲブル氏はツイッターで「黒人女性をリーダーとする新組織の発表は、まるでわれわれが交換可能だと言っているようだ。私のチームに嫌がらせをし、怖がらせ、「ガスライティング」(映画『ガス燈』に由来する心理的虐待)しながら、危害を加えたことを全く認めず、修正もしないのは『ガスライティング』を超えている」と述べた。
- 2021年2月19日⁶
 - 倫理的 AI チームの共同リーダーを務めていたミッチェル氏が解雇された。
 - 解雇されて以来ツイッター上で Google の従業員に対する「ガスライティング」について訴えを続けていたゲブル氏は、ミッチェル氏によるクビ報告ツイートの数分後に次のようなツイートをしている。
 - ◆ 「Ethical AI チームのリーダー、メグ・ミッチェルが解雇された。社内メールシステムから 5 週間締め出しされていた彼女は自分のプライベートメールでその知らせを受け取った。これで言いたいことが言える。みんながやつらのうそっぱちに騙されていないことを知ってうれしい。Google の副社長たちはお気の毒。」
 - ミッチェル氏はツイッターで「これで知りました。信頼を再構築できてよかったですね。私は完全に消されて、私のチームは連れ去られたようだ」と述べた。

4 ティムニット・ゲブル氏の解雇

⁶ Richard Nieva (CNET 編集部訳)・前掲注 3) 「グーグル、AI 倫理に取り組むチーム率いるもう 1 人の責任者を解雇」, *CNET Japan*, 2021 年 2 月 22 日, <https://japan.cnet.com/article/35166792/>

- 事件の流れ⁷
 - Google の検索エンジンで使われているシステムを含む、AI におけるバイアスの危険性を訴えた研究論文を出したことや、Google の複数の従業員あてに、同社の多様性や公平性関連の取り組みを批判するメールを送信していたことを理由に解雇された。
 - 2700 人近い Google 従業員がゲブル氏を支持する公開書簡に署名した。
 - ゲブル氏が所属していたチームの従業員らは、最高経営責任者（CEO）のサンダー・ピチャイ（Sundar Pichai）氏にゲブル氏の復職を求める書簡を送付した。
 - 2021 年 2 月 19 日に Google はゲブル氏に対する同社の処遇に関する社内調査を終了したことを明らかにした。社外の専門家と調査を実施したと述べたが、その調査結果を明らかにすることは控えた。また、広報担当者は、調査の実施方法に関する具体的な質問に答えなかった。
 - Google は調査を受け、人材とダイバーシティー（多様性）に関するポリシーを次のように変更すると述べた。
 - ◆ ダイバーシティーの目標をバイスプレジデント以上の幹部による業績評価と関連付ける
 - ◆ 従業員の維持に関わるチームを強化する
 - ◆ 議論を呼ぶ可能性のある従業員の退職に関する対応について、HR 専門家の意見を仰ぐ
 - このようなポリシーの変更に対して、ゲブル氏や同氏の支持者はツイッター上で次のような批判の声を挙げた。
 - ◆ ゲブル氏「私は物事を求めるメールを書いて、解雇された。その後の3か月にわたる調査の末に、彼らは、自らの行動に誰も責任を負うことなく、おそらく私が職を失う原因になった要求のいくつかを、実行しなければならないかもしれないと述べている。」
 - ◆ ゲブル氏「そこには何の責任もない。ゼロだ。」
- この事件に対して寄せられた意見⁸⁹

⁷ Richard Nieva (CNET 編集部訳)・前掲注 3) 「グーグル、AI 倫理に取り組むチーム率いるもう 1 人の責任者を解雇」, *CNET Japan*, 2021 年 2 月 22 日, <https://japan.cnet.com/article/35166792/>

⁸ Karen Hao, "We read the paper that forced Timnit Gebru out of Google. Here's what says.", *MIT Technology Review*, December 4, 2020, <https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/>.

⁹ 佐藤由紀子・前掲注 1) 「Google を解雇された AI 倫理研究者が指摘していた「大規模言語モデル」の危険性」, *ITmedia NEWS*, 2020 年 12 月 7 日,

- ゲブル氏
 - ◆ ゲブル氏は規則に従い、社内外の研究者と共同で書いた論文を公開前に社内レビューの申請を行った。
 - ◆ 2ヶ月間音沙汰がなく、感謝祭の休暇の1週間前になって上司の上司であるディーン氏から「論文は撤回して。これは決定事項だから」と告げられた。
 - ◆ これに対してゲブル氏は、「撤回してもいいがその理由と、誰がそうしろと言ったのかを教えて欲しい」と提案したが、拒否をされた。
 - ◆ ゲブル氏は休暇後に辞任に対する返事をすると言ったが、ディーン氏は勝手に判断をし「ゲブル氏の辞任を受け入れた」とゲブル氏の上司にメールを送信した。
- Google (ディーン氏)
 - ◆ Google は論文のレビューにかかる時間を2週間と設定しているが、ゲブル氏が共著となっている論文は評議会への提出1日前にレビューのために提出された。また、フィードバックを返す前にゲブル氏はその論文を提出した。
 - ◆ 様々な部門から構成されたチームが通常のプロセスの一環として論文のレビューを行い、ゲブル氏を含めた論文の著者たちに公開基準を満たしていないこととその理由（関連研究の言及が少ないなど）を述べたフィードバックを返した。
 - ◆ 既に論文を提出してしまっていたこともありメーガン・カコリア (Megan Kacholia) 氏やディーン氏の決定に不満を持ったゲブル氏は、今後 Google で研究を続けるための条件をメールで伝えてきた。その条件は、例えば、メーガン氏とディーン氏がレビューの実施やフィードバックの作成を行った時に相談をした全員の身元を明らかにすることなどである。そして提示した条件を満たさなかった場合、Google を辞めると書かれていた。
 - ◆ Google はその条件を満たすことをしなかったため、ゲブル氏の解雇を決定した。
- エミリー・M・ベンダー (Emily M. Bender) 氏 (ゲブル氏の共著者)
 - ◆ 自然言語処理に関する研究の現状を目に見える形にし、明らかに分かる言語モデルの良い点のみならずそのモデルが引き起こす可能性のある問題にどう対処するかを考えることを論文の目標として掲げている。

<https://www.itmedia.co.jp/news/articles/2012/07/news090.html>。

- ◆ 関連研究への言及の少なさを指摘されているが、様々な学問領域からこの問題を捉える必要があると考えたため、この論文の共著者の学識は幅広く、128 個の文献を引用している。また、さらに参考文献を加えることも考えていることを明らかにしている。
- ◆ 論文公表の前に提出する評議会でも厳しいレビューが行われるにもかかわらず、自社でそれほど厳密なレビューを行う必要はあるのか。
- ◆ 今回のゲブル氏の解雇を知った AI 倫理研究者が委縮してしまう可能性がある。最前線の専門家たちの多くは資金が潤沢なテクノロジーを扱う大企業で働いている。確かに大企業で働くことは多くのメリットがあるが、今回の事件のように世界の科学の進歩にとって最善とは言えない事象が発生してしまっている。
- ニコラス・レ・ルークス (Nicolas Le Roux) 氏 (モンリオールの Google AI 研究者)
 - ◆ 彼は自身がこれまでに公表してきた論文のレビューでは機密資料の開示の確認に留まっており、ゲブル氏の論文が指摘された参考文献の質や量は対象になっていなかったと語り、ゲブル氏の論文に対して行われたレビューの不可解な点を指摘している。
- ゲブル氏が撤回を求められた論文¹⁰
 - ゲブル氏が撤回を求められた論文の公開前に、共著者の 1 人が MIT Technology に提供し、同社が公開した概要記事。
 - ゲブル氏解雇に対する抗議文に 1400 人以上のグーグル職員と 1900 人以上のサポーターが署名した。
 - 論文の概要
 - ◆ ここ 3 年で進化を遂げてきた言語モデル (大量のテキストデータでトレーニングされた AI) は正しい条件下で利用された場合は素晴らしい働きをするが、リスクもはらんでいる。この論文では「開発する一方で潜在リスクに対して十分な処置を施しているのか。また、そのリスクを抑える方法を考えているのか」を議論している。
 - 論文が指摘する言語モデルの 4 つの課題
 - ① 環境・財務コスト

¹⁰ Hao, *supra* note 8.

- ◆ 大規模な AI モデルのトレーニングでは大量の電力が消費される。エマ・ストルベル (Emma Strubell) と共著者が 2019 年に発表した大規模言語モデルの炭素排出量と経済コストに関する論文によれば、言語モデルに与えられるデータが増加したことを受け、2017 年以降のエネルギー消費とカーボンフットプリントは爆発的に増加している。下図は各言語モデルの環境・財務コストを表したものである。なお、言語モデルは研究や開発のために何度もトレーニングや保持が行われるため、上記のコストは最低水準のものを表している。

	Date of original paper	Energy consumption (kWh)	Carbon footprint (lbs of CO2e)	Cloud compute cost (USD)
Transformer (65M parameters)	Jun, 2017	27	26	\$41-\$140
Transformer (213M parameters)	Jun, 2017	201	192	\$289-\$981
ELMo	Feb, 2018	275	262	\$433-\$1,472
BERT (110M parameters)	Oct, 2018	1,507	1,438	\$3,751-\$12,571
Transformer (213M parameters) w/ neural architecture search	Jan, 2019	656,347	626,155	\$942,973-\$3,201,722
GPT-2	Feb, 2019	-	-	\$12,902-\$43,008

- ☆ ニューラルアーキテクチャ検索を使用した 1 つの言語モデルのトレーニングが放出する二酸化炭素は、アメリカ産の車 4 台が廃車になるまでに放出する量とほぼ等しい (626,155 ポンド)。
- ☆ Google の検索エンジンを支える言語プログラムである BERT のトレーニングは 1438 ポンドの二酸化炭素が放出されるが、これはニューヨークとサンフランシスコの往復便が消費するものに近しい。

② 大量のデータ・不可解なモデル

- ◆ 言語モデルのトレーニングに使用するために大量のデータをインターネットから入手するが、その時に人種差別や性差別などのデータが取り込まれ、AI がそういったデータを「普通」であると認識してしまう危険性がある。例えば人種差別のデータが取り込まれた場合、次の

3つのような問題が生じる。

1. 社会情勢の変化により意味が変わった単語（MeToo や Black Lives Matter など）を正しく解釈できない。
2. インターネットに触れる機会が少ない国や人の情報を収集することが出来ず、先進国やその国に住む人の情報しか反映されない。
3. 言語モデルのトレーニングに使用するデータが膨大なあまりデータの監査をすることができず、潜在的なバイアスを取り除くことができない。

③ 研究に費やす労力

- ◆ 言語モデルは言語を理解することはできないが、収益を上げることが出来るという理由から企業は投資を続けている。
- ◆ 言語理解ができる可能性があり、取捨選択したデータセットを使用してより良い結果を得ることができる AI モデルの開発は言語モデルよりも労力を抑えることができる。

④ 意味の錯誤

- ◆ 言語モデルは人の模倣に優れているため偽情報の生成や誤訳が発生する可能性がある。実際、2017年にフェイスブックでパレスチナ人の男性がアラビア語で「おはよう」と投稿したところ、AIがヘブライ語で「攻撃する」と男が投稿したと判断してしまい、男が逮捕されてしまうという事態が起こった。

● Googleの研究レビュープロセス¹¹

- 幅広い分野をカバーできるようにリサーチ・アンド・グーグルの研究者や社会学者、倫理学者、ポリシー・プライバシーアドバイザー、人権の専門家が構成するグループがレビューを行っている。このグループは、Googleが発表する論文は十分な説明が記載されており、最新の研究を考慮しているか、そしてGoogleのAI原則を遵守しているかを確認する。

5 マーガレット・ミッチェル氏の解雇事件の流れ¹²

- ミッチェル氏はゲブル氏が解雇された後、同氏に対する社内での差別や嫌が

¹¹ 前掲注2)“About Google’s approach to research publication”, <https://docs.google.com/document/d/1f2kYWDXwhzYnq8ebVtuk9CqQqz7ScqhxSIxeYGrWjK0/edit#heading=h.aplcvu32myqt>

¹² Richard Nieva (CNET 編集部訳)・前掲注3)「グーグル、AI倫理に取り組むチーム率いるもう1人の責任者を解雇」, *CNET Japan*, 2021年2月22日, <https://japan.cnet.com/article/35166792/>

らせの実態を調査するために社内通信を検索していたが、この企業データの取り扱いをめぐる Google の調査を受けていた。

- 調査を受けている間、ミッチェル氏は職務用アカウントを数週間ロックされていた。
- ミッチェル氏と Google の反応¹³¹⁴
 - ミッチェル氏
 - ◆ 2021 年 2 月初めに Google のプレスチームにメールを送った
 - ◆ 2021 年 2 月 19 日に「解雇された」とツイートした。
 - ◆ 2021 年 2 月 22 日に 4 年間一緒に働いてきた仲間に感謝を述べるとともに、Google の人種・性差別の問題に対する懸念を表明し、ゲブル氏の解雇について話そうとした結果であるとツイッターで述べた。
 - ◆ 2021 年 2 月 28 日に「9 日前に失職し、明日にはヘルスケアの資格を失う。15 年以上キャリアを築き、成功を重ねてきた AI 研究者の現状はこれです。」とツイートした。
 - Google
 - ◆ 解雇を認め、「このマネージャーの行為を調査した結果、ビジネス上の機密文書と他の従業員の個人データの不正取得など、当社の行動規範とセキュリティポリシーに対する複数の違反があったことを確認した」とコメントを出した。
- 2021 年 2 月初めに送信されたミッチェル氏のメール¹⁵
 - AI の倫理的開発の運用化に取り組む際に「フラクタル問題」や「インフィニティ・オニオン問題」に直面するが、この問題はさらに複雑な問題への入り口である。この状況を打破することは難しいが、あるポイントで一度立ち止まり、現状を詳しく説明することで状況を改善することが出来る。
 - ゲブル氏の解雇について不可解な点と今回のような問題が今後も発生することになる理由について以下のように述べている。
 - ① 結論
 - ◆ ゲブル氏の解雇には人種・性差別が関わっているが、これは AI システムにも同じことが言える。また、ゲブル氏の解雇の方法や解雇を許

¹³ Richard Nieva (CNET 編集部訳)・前掲注 3)「グーグル、AI 倫理に取り組むチーム率いるもう 1 人の責任者を解雇」, *CNET Japan*, 2021 年 2 月 22 日, <https://japan.cnet.com/article/35166792/>

¹⁴ "On the Firing of Dr. Timnit Gebru", Margaret Mitchell", <https://docs.google.com/document/d/1ERi2crDToYhYjEjxRoOzOuOUeLgdoLPfmx1JOErg2w/edit>.

¹⁵ *See id.*

してしまった環境にも問題がある。

- ◆ ディーン氏やカコリア氏は責任を取ろうとしないが、これはゲブル氏の解雇は正当であると会社全体で認めていることを意味している。
- ◆ 以上のことからゲブル氏の解雇は不当でその対応には侮辱が含まれているため、Google はゲブル氏に対して謝罪をすることが妥当であると言える。

② 背景：テクノロジーの開発に対する倫理的な AI のアプローチ

- ◆ 研究課題は人間の価値観や多様な経験の包含、複数の事象や社会運動からの学習に基づいている可能性があるため、ミッチェル氏は過去と未来から学ぶという「大局的な目線」を取り入れた新しい研究のアプローチ（倫理 AI）を考えている。
- ◆ 従来のアプローチとの違い
 - ◇ 従来のアプローチ
 - 「斬新的なこと」の実践や課題の解決
 - ◇ ミッチェル氏のアプローチ
 - 理想的な未来について考えてそこから逆算をし、今何をすべきかを考える
 - よりよい社会のために AI が何をできるかということを考える

③ 倫理 AI チーム

- ◆ AI の利用により公平にチャンスや有益な結果を分け与えることができることは容易に想像できるが、AI がそういったメリットを台無しにすることも想像に難くない。
- ◆ ミッチェル氏のチームは、研究に対して不適切なアプローチを取ることによって生じるネガティブな対人政治から脱却し、重要な事柄（未来や AI、社会）に集中することに取り組んできた。
- ◆ 2 年前、チームが素晴らしい活動をしている中でゲブル氏を共同リーダーとしてチームに招待することができたことを大変うれしく感じた。

④ 共同リーダー

- ◆ ゲブル氏はミッチェル氏と考えを共有し、リーダーとして素晴らしい働きをしてきていた。例えば、不当な処遇（人種・性差別など）を受けている人を見つけ、その問題の解決に取り組むなどである。
- ◆ AI に関わる問題の被害を受ける危険性が高い人々を助けるためには、様々な問題に取り組んできたゲブル氏の力が必要となる。
- ◆ ゲブル氏とミッチェル氏はスタッフ研究員に昇格したが、これはある

程度の雇用保障を意味していると捉えていた。

⑤ これまでに起こったこと

- ◆ 高い階級にいる人ほど権限を持つことになるが、あきらかに間違っている考えに固執してしまう事態を引き起こしてしまう。今回のゲブル氏の解雇やそれを「辞任」と呼んでいることも、このような体制から発生した問題である。

⑥ いま何が起きているか

- ◆ ゲブル氏は研究者の誠実さや研究者自身を軽視する現行の体制に反発した。これに倣い、ミッチェル氏たちは研究者を軽視するような体制を打破すると述べた。

文責：三國 陸真