

<論 文>

ディープフェイクの政治風刺利用と
自由権規約第19条
— EU法・米国法を素材とした立法的含意 —

北 村 理依子*

Political Satire Deepfakes and Article 19 of the ICCPR:
Legislative Implications from EU and U.S. Law

Rieko Kitamura

Deepfakes are widely portrayed as a threat to democracy, yet not every deepfake is harmful. Political satire, long protected under freedom of expression doctrine, is increasingly produced with generative AI, and because synthetic media are now nearly indistinguishable from authentic recordings, the boundary of protected satire has blurred. This article asks what requirements a regulation of politically satirical deepfakes must meet to comply with Article 19 (3) of the ICCPR, and examines EU and U.S. (California) law as contrasting models. Article 19 (3) requires legality, a legitimate aim, and necessity and proportionality, criteria that blanket bans on deepfakes are unlikely to satisfy. The EU AI Act treats satire at two stages, excluding manifestly exaggerated content from the definition of deepfake and relaxing the manner of disclosure for satirical works, whereas California's Elections Code exempts satire and parody outright. Regulation of satirical deepfakes thus does not converge on a single model.

Keywords : deepfake, freedom of expression, ICCPR Article 19, EU AI Act, California law

キーワード : ディープフェイク、表現の自由、自由権規約第19条、EU AI法、カリフォルニア法

* 立命館大学国際関係学部助教

はじめに

2023 年 9 月、スロバキアの議会選挙直前、革新政党プログレッシブ・スロバキアの党首らの間で選挙不正を謀議したとされる音声ファイルが SNS 上で拡散した。これは後に生成 AI による合成音声であることが判明したが、選挙終盤まで拡散され、有権者の判断に影響を及ぼしたと報告されている。2024 年 1 月には、米国大統領予備選挙でも、投票棄権を呼びかけるバイデン大統領の声を模した自動音声通話が拡散した (Holland, 2024)。これらは、ディープフェイク (AI によって生成・編集された画像・音声・映像コンテンツで、まるで本物のように見せかけたもの) が民主主義の脅威となりうることを示している。Recorded Future の調査では、2023 年 7 月から 2024 年 7 月の 1 年の間に 38 カ国で 82 件の政治関連のディープフェイクが確認されており、金銭賠償、虚偽発言による印象操作、選挙工作、人格攻撃、性的画像流布による威嚇・名誉毀損という実害を及ぼしている (Recorded Future, 2024)。

しかし、ディープフェイクが常に悪だとは限らない。その例として政治風刺を目的としたディープフェイク (以下、政治風刺ディープフェイク) がある。そもそも政治風刺は、誇張・歪曲表現を用いて権力者の言動を批判するものとして表現の自由において手厚く保護されてきた表現類型である。その表現手法は風刺画から写真、動画へとその時々新しい技術を取り込みながら変容しているが、ディープフェイクもこのバリエーションの一つであると言える。問題は、AI 技術の発達により、情報の受け手が生成物を風刺と見抜けない事態が生じており、その結果保護対象となる風刺表現の範囲も曖昧になっていることである。

こうした状況において、政治風刺ディープフェイクをどのように保護または規制するのかについて先行研究では十分な検討が尽くされているとは言い難い。たとえば、Helm and Nasu (2021) は市民的及び政治的権利に関する国際規約 (自由権規約) 第 19 条 3 項に照らしてフェイクニュースへの規制を論じるが、その射程はテキスト型の偽情報にとどまり、ディープフェイクには及んでいない。ディープフェイクと人権の交錯を扱った研究としては Romero-Moreno (2024) と Kuźnicka-Błaszczowska and Kostyuk (2025) が挙げられる。前者は人権及び基本的自由の保護のための条約 (欧州人権条約) と欧州連合における人工知能に関する調和的規則 (EU AI 法) を準拠枠組とし、自由権規約は対象としていない。後者はウクライナ紛争の文脈でディープフェイクの国際法上の規制を論じるが、風刺的表現への規制の可否には踏み込んでいない。つまり、自由権規約第 19 条 3 項を軸に政治風刺ディープフェイク規制を分析した研究は見当たらず、本稿の狙いの一つはこの間隙を埋めることにある。

そこで本稿では、「自由権規約 19 条 3 項に適合的な政治風刺ディープフェイク規制は、いかなる要件を備えるべきか」を問いとして設定する。以下ではまず、ディープフェイクの技術的特性を概観し、本稿が問題とする政治風刺ディープフェイクについて整理したうえで、自由権規約第 19 条 3 項の枠組みを確認する。次に、政治風刺ディープフェイクの具体的な規制の例

として EU 法と米国（カリフォルニア）法を記述する。EU と米国（カリフォルニア）を対象とするのは、両者が政治風刺ディープフェイクの規制において対照的なモデルを提示しており格好の検討素材であるからである。最後に結論において、自由権規約第 19 条の枠組みを EU 法と米国（カリフォルニア）法の例に当てはめて、本稿の問いへの応答を試みる。

1 政治風刺ディープフェイクと自由権規約第 19 条の枠組み

(1) ディープフェイクの概要

ディープフェイクとは、深層学習（ディープラーニング）の技術を用いて人物の顔・声・動作を別の映像や音声に合成したコンテンツ、あるいはその合成を可能にする技術のことをいう。元々は 2017 年頃にオンラインコミュニティで使われ始めた俗語にすぎなかったが、いまでは学術・政策の場でも広く定着している（Chesney and Citron, 2019；総務省、2024）。

ディープフェイクの技術的な基盤としては、当初は敵対的生成ネットワーク（GAN）が主流であった。GAN は、コンテンツを生成する生成器と、その精度を評価する識別器という 2 つのニューラルネットワークを反復的に学習させることで、人間による作業を大きく上回る速度と精度でサンプルを洗練させていく手法である。この技術の高度化とともに、ディープフェイクも精巧になっていった（Chesney and Citron, 2019, 1760）。2022 年以降はこれに代わって拡散モデルが急速に普及した。拡散モデルは GAN より安定した学習が可能で品質も高く、現在流通するディープフェイク生成ツールの多くはこの手法に基づく（Masood et al., 2023）。

法的な観点からとりわけ注目すべきは、ディープフェイク作成への参入障壁が急激に下がったことである。2017～18 年頃には、精巧なディープフェイクの作成に大量の GPU と相当の技術的知識が必要であった。ところが 2023 年以降、スマートフォンのアプリを用いて本物そっくりの合成映像・音声を誰もが簡単に作れるようになっている。この結果、ディープフェイクによる被害の規模は飛躍的に拡大した（Security Hero, 2023）。

(2) 政治風刺ディープフェイク

ディープフェイクは様々な目的で用いられ、選挙操作、性的コンテンツの合成、詐欺など、その害悪の性質は一様ではない。だが、その多くは情報の受け手に事実誤認を生じさせることで害悪が生じる、または受け手の認識のいかんにかかわらず表現の対象となった者の尊厳を侵害する（例として性的コンテンツ）というものであり、ディープフェイクへの規制の必要性についてはおおむね異論がない（Chesney & Citron, 2019）。しかし、政治風刺やパロディは、受け手が誇張・演出として了解したうえで受容することを前提としており、害悪があるとはいえない点で規制対象からは区別されるべきである¹⁾。

実際、伝統的な法の枠組みのもとでは、政治風刺は原則として手厚く保護されてきた。一般

的意見 34 は、国家元首・政府首脳を含むすべての公人が正当に批判と政治的対抗の対象となると述べており（Human Rights Committee, 2011, para 38）、権力者を標的とする風刺はとりわけ手厚い保護を受けると解される。

とはいえ、風刺が保護される根拠の一つは、受け手がそれを誇張表現であると看守できる点にあるのに、ディープフェイク技術の発達によって、一見しただけでは本物と偽物の区別がつかなくなっている。ディープフェイクかどうかを検出する技術の研究も進んではいるものの、生成技術の改善速度が検出技術を常に上回っており、状況はたちどころの様相を呈している（Chesney & Citron, 2019, 1787-1788）。

さらに、ディープフェイクが社会に広まると、目の前のコンテンツの真偽の区別がつかないだけでなく、認識論的な問題も生じる。すなわち、本物の映像・音声さえも偽物ではないかと疑われるようになり、風刺であるかどうかを裏づける証拠も同様の疑念を免れ得なくなるのである（Chesney & Citron, 2019; Rini, 2020）。その結果、風刺として保護されるべきディープフェイクとそうではないコンテンツの区別はさらに難しくなっていくという悪循環が生まれる。

このように、政治風刺ディープフェイクは表現の自由の保護対象となる一方で、その外延が曖昧になるという問題を抱えている。この点を踏まえた上で、以下では政治風刺ディープフェイクへの規制について自由権規約第19条がどのような分析枠組を提供しうるかを検討する。

(3) 自由権規約第19条の枠組み

自由権規約第19条は3つの項から成る。第1項は「意見を持つ自由」を、第2項は「表現の自由」を保障する。第3項は、一定の要件のもとで第2項の権利に制限を加えることができると定める。第2項が保護する「表現の自由」の範囲は広い。保護される表現の形態・公表手段は本、新聞、芸術作品などの伝統的な媒体にとどまらず、「音声・映像によるあらゆる表現形態、および電子的手段やインターネットを基盤とする表現形態にまで及」び（Human Rights Committee, 2011, para 12）、ディープフェイクを用いた表現もこの射程に含まれるといえる。第19条3項が表現の自由への制限を許容するための要件は、その制限が法律によって定められていること（法定性）、正当な目的のためになされること（目的の正当性）、必要かつ比例的であること（必要性・比例性）の3つである。以下、順に検討する。

i 法定性

委員会は、制限を定める「法律」について、個人が自らの行動を規律できる程度に十分な精確さをもって定式化されていなければならないと述べる（Human Rights Committee, 2011, para 25; Schabas, 2019, 565）。つまり、制限の内容が予見可能であり、当局に恣意的な裁量を与えないことが求

められる。

政治風刺ディープフェイクの文脈ではこれをディープフェイク規制の対象から除外する立法が考えられるが、前項で論じたように、問題となる表現が風刺的であるかどうかは生成技術が精巧になるほど判断がつかなくなっていく。この曖昧さが法律の文言に持ち込まれれば規制当局に広い裁量を与えることになり、法定性の要件が問われかねない²⁾。

ii 正当な目的

第19条3項が許容する制限の目的は、(a) 他者の権利または信用の保護、(b) 国の安全・公の秩序・公衆衛生・道徳の保護、の2カテゴリに限定される。この列挙は網羅的であり、それ以外の目的を根拠とした制限は第19条のもとでは許容されない (Human Rights Committee, 2011, para 22)。

詐欺や性的被害等個人への害悪を生むディープフェイクは「他者の権利」を侵害し、選挙工作などの社会への害悪を生むディープフェイクは「公の秩序」を乱すとして第19条3項にいう正当な制限目的に対応させる。これに対し、政治風刺ディープフェイクの場合、当該表現が特定個人の名誉を毀損する場合であれば同要件を満たしうる (Schabas, 2019, 568) が、純粋に政治権力を標的とした風刺にとどまる場合、いずれの目的を正当化根拠とすればよいのか不明確である。

iii 必要性・比例性

第三の要件は、当該制限が正当な目的の達成に必要なかつ比例的でなければならない、というものである。この要件は政治的表現との関係でとりわけ重要な意味を持つ。上にも触れたように、一般的意見34は「市民、候補者、公選の代表者の間での公的・政治的事項についての情報と意見の自由な交流は不可欠である」と述べ (Human Rights Committee, 2011, paras 13, 20)、政治権力者への批判がとりわけ手厚い保護を受けることを確認している。さらに同意見は、「政治的言論の可能性に対するいかなる制限も、第3項との適合性を評価するために十分な審査を受けなければならない」と述べたうえで、「制限は保護のための機能を果たすために適したものでなければならない」、「最も制限的でないもので」、「保護される利益に対して比例的でなければならない」と述べ、「民主社会における公共の議論においては公的・政治的な人物に対する忌憚のない表現」にとりわけ高い価値がおかれると結んでいる (Human Rights Committee, 2011, para 34)。政治風刺ディープフェイクの多くは権力者の政治的言動に関するものであるから、その規制に際しては以上の基準をクリアしなければならない。

さらに委員会は、必要性要件について、締約国が規制の必要性を一般論としてではなく、個別の脅威との関係で具体的に示さなければならないと述べている (Human Rights Committee, 2011, para 35)。この点、政治風刺ディープフェイクを含む一切のディープフェイ

クを一律に禁止すると、政治風刺ディープフェイクが個別にどのような害悪をもたらすかを具体的に示すことなく、受け手が事後的に風刺と認識できたはずのものまで一括して排除してしまう点で、この個別の脅威との具体的な結びつきが証明できないおそれがある。

また、比例性の観点からは、より制限的でない代替手段の有無が問われる。後の第 2 章で詳論するが、EU AI 法が採用した開示義務（当該コンテンツが AI により生成されたコンテンツであることを開示させる義務）は、受け手にコンテンツの真偽を判断する材料を提供することで風刺的ディープフェイクの一律禁止を回避しており、より制限的でない代替手段として機能しうる。このような代替手段が存在する以上、ディープフェイクを一律に禁止する規制を必要最小限の手段と主張することは難しい。

ここまでの検討をまとめると、自由権規約第 19 条の要件は、政治風刺ディープフェイクの規制について次のような分析枠組を提示する。第一に、政治風刺ディープフェイクは政治権力に対する風刺的表現に該当するため手厚い保護が保障される。第二に、法定性の観点から、（政治）風刺ディープフェイクの概念の外延が曖昧になっている状況においてその規制の予見可能性の確保に課題がある。第三に、政治風刺ディープフェイクの規制目的は第 19 条 3 項の列挙目的に対応しない可能性が高い。第四に、規制の必要性は個別の脅威との関係で具体的に示されねばならず、さらに規制目的のために必要最小限の規制にとどめねばならないところ、ディープフェイクの一律規制は、必要性と比例性の要件を満たすとはいいがたい。次章では、この枠組みを手がかりに、EU 法と米国（カリフォルニア）法がそれぞれ政治風刺ディープフェイクの規制をどう設計しているかについて検討する。

2 EU 法・米国（カリフォルニア）法にみる政治風刺ディープフェイクの規制

(1) EU 法における規制

i EU AI 法の全体像

EU AI 法（Regulation (EU) 2024/1689）は 2021 年 4 月に欧州委員会により提案され、2024 年 5 月に成立、同年 8 月 1 日に施行された、AI に関する包括的な規則である。同法は、域内市場の機能を向上させ、人間中心で信頼できる AI の導入を促進し、AI システムがもたらす有害な影響から EU 基本権憲章が掲げる基本的権利の高水準の保護を確保し、技術革新を支援することを目的とし（第 1 条）、域内における AI 利用を規制する。同法は AI システムを「容認できないリスク」（原則禁止）、「ハイリスク」（厳格な適合性評価のもと利用可）、「限定的なリスク」（限定的な透明性義務のもと利用可）、「最小限のリスク」（自由利用可）の四段階に分類し、さらに汎用 AI モデルのプロバイダーには別途独自の義務体系を設けている（欧州連合日本政府代表部、2025；石川、2025；古川他、2025）。

本稿の対象であるディープフェイクは、「特定の AI システムのプロバイダーおよびデプロイヤーに対する透明性義務」（第 50 条）により規律されている。当該条項により規制される特定 AI システムは「限定的なリスク」に位置づけられ、「ハイリスク」AI システムのような重い適合性評価は要求されず、AI 生成コンテンツである旨の開示を中心とする比較的軽い規制に服するという特徴を持つ。以下では、本条の構造を整理する。

ii 第 50 条の構造

(a) 全体像——規律類型とプロバイダー／デプロイヤーの義務

第 50 条は、特定の AI を使用していることを開示する義務（透明性義務）を、対象となる AI システムの類型ごとに四つ定める。すなわち、①自然人と直接やりとりする AI システム（1 項）、②合成音声・画像・動画・テキストコンテンツを生成又は改変する AI システム（2 項）、③感情認識システム・生体分類システム（3 項）、④ディープフェイクを生成又は改変する AI システム及び公共の関心事について公衆に知らせる目的で公表されるテキストを生成又は操作する AI システム（4 項）である。義務の名宛人は類型によって異なり、1 項・2 項はプロバイダーに、3 項・4 項はデプロイヤーに、それぞれ義務を課するという構造を取る。本稿の主題であるディープフェイクは、このうち 4 項（ディープフェイクを使用したコンテンツであることを開示するデプロイヤーの義務）を中心に規律されるため、以下では同項を中心に検討するが、後述（d）のとおり、2 項（プロバイダーの義務）も別途重疊的に適用されうる点に注意を要する。

(b) 「ディープフェイク」該当性

デプロイヤーへの義務を検討する前提として、「ディープフェイク」の定義（3 条 60 号）を確認する必要がある。同号は、ディープフェイクを「既存の人物・物・場所・団体・出来事に類似し、かつ人にとって真正又は真実であるかのように誤って見える、AI が生成又は改変した画像・音声・動画コンテンツ」と定義する。EU AI 法第 50 条に基づく特定の AI システムに関する透明性義務の実施に関するガイドライン（以下、ガイドライン）はこれを、①類似性（既存の対象に類似していること）、②実在性（対象が実在すること）、③対象の種類（人物・物・場所・団体・出来事のいずれかであること）、④真正性への誤認可能性（真正又は真実であるかのように誤って見えること）という 4 つの要件に分解して詳述する（European Commission, 2026, paras 113-116）。

このうち政治風刺的表現との関係で特に重要なのが②と④である。まず②実在性について、ガイドラインは、自然法則や物理法則を無視した表現や生物学上ありえない生命体を描く表現は、そもそも「実在」の主体に類似するとはいえず、ディープフェイクの定義に該当しないとする（European Commission, 2026, para 113）。次に④真正性への誤認可能性について、当該評価は客観的に行われるものの、コンテンツが提示される文脈及び受け手の期待が考慮され、

受け手が真正性を期待していない文脈（風刺誌への掲載、ゲーム内コンテンツ、映画等）で用いられる場合には、たとえコンテンツ自体が非真正・非事実であっても、「虚偽に真正又は真実であるかのように見える」とは評価されないことがあるとされる（同, paras 114-115）。

したがって、荒唐無稽・誇張的な性質の強い風刺的コンテンツは、そもそもディープフェイクの定義自体に該当しないため、第50条4項の開示義務は最初から生じない。これに対し、たとえばSNS上に出回る実在の政治家の外観を精巧に再現した風刺画像は、ディープフェイクの定義に該当し、第50条4項の開示義務がいったん生じる。もっとも、この場合も、次に見る軽減規定により、その義務の履行方法が簡略化されうる。

(c) デプロイヤーの開示義務とその軽減

生成したコンテンツが上記(b)の定義に該当する場合、デプロイヤーは、当該コンテンツがAIによって生成・加工されたものであることを開示する義務を負う。この義務は、当該コンテンツが政治風刺目的であるか詐欺目的であるかを問わず、ディープフェイクの定義に該当する限り一律に及ぶ。

ただし、同項には例外が置かれており、当該コンテンツが「明らかに芸術的、創造的、風刺的、架空の、又は類似の作品若しくは番組の一部を構成する場合」には、当該義務は「作品の表示又は享受を妨げない適切な方法で」当該コンテンツの存在を開示することに限定される。つまり、風刺的コンテンツについては、開示義務そのものは免除されないものの、開示義務の方法が簡略化される。前項と合わせると、EU AI法は、AIにより生成された風刺的表現に対して、ディープフェイク該当性と開示の簡略化措置の二段階の特別措置を用意していることになる。

ガイドラインは、上記「明らかに」という要件についても検討を加えている。それによれば、開示義務の簡略化措置は、ディープフェイクへの制限が表現の自由と芸術・科学の自由に潜在的に影響することへの配慮によるものであることを指摘しつつ、あるディープフェイクが「明らかに」芸術的・創造的・風刺的・架空の又はこれらに類する作品若しくは番組の一部を構成するかどうかの判断について、①当該コンテンツが当該カテゴリーに特徴的な形式・様式を備えているか（風刺であれば皮肉や誇張、芸術であれば特定の作風や技法など）、②提示される文脈（プラットフォーム・媒体・場所が芸術的・創造的・風刺的・架空的な利用と結びついてるか）、③受け手の期待（映画・ゲーム環境・仮想現実空間など）という三つの考慮要素を示す。また、専ら情報提供的・商業的な性質を有するコンテンツ（たとえば報道）はそもそも範囲外に置かれるとし、情報提供的な性質と創造的な性質を併せ持つ場合には情報提供的な性質が常に優先し、通常の開示義務が適用されるとしている（European Commission, 2026, para 122）。このように、「明らかに」という要件はディープフェイクによる誤情報、操作、詐欺のリスクを減じるために厳格に解釈される。

(d) プロバイダーの義務との関係

(c) で見た第4項の開示簡略化措置は、デプロイヤーの開示義務にのみ関わるものであり、プロバイダーの義務には及ばない。第50条2項は、音声・画像・動画・テキストという4つの様式のいずれかにおいて合成コンテンツを生成または改変するAIシステムのプロバイダーに対し、当該コンテンツを機械可読な形式でマーキングし、AIによる生成または改変であることを検知可能にすることを義務付ける。ディープフェイクは、その定義が「AIが生成又は改変した画像・音声・動画コンテンツ」であることを前提とするため（第3条60号）、ディープフェイクに該当するコンテンツは必然的に第50条2項が対象とする合成コンテンツの範疇にも含まれる。ガイドラインは、この二つの義務が同一のコンテンツに累積的に適用されうることを明示し、あるAIシステムがディープフェイクを構成しうる画像を生成する場合、プロバイダーは第50条2項を、デプロイヤーは追加的に第50条4項を、それぞれ遵守しなければならないとする（European Commission, 2026, paras 8, 111）。しかも、第50条2項の例外は「標準的な編集にとどまる場合」「入力データを実質的に改変しない場合」「法執行目的での使用」の三つに限られ（同, paras 89-93）、風刺・創作等を理由とする例外規定は存在しない。

したがって、あるディープフェイクが（政治）風刺ディープフェイクとして第50条4項の軽減措置の対象となり、開示の態様が簡略化されたとしても、当該コンテンツには第50条2項に基づく機械可読な痕跡（電子透かし・メタデータ等）が付される。

iii 小括——EU AI 法における風刺表現の位置づけ

以上のとおり、EU AI 法は（政治）風刺ディープフェイクに対して二段階の特別措置を設けている。荒唐無稽・誇張的な性質の強い表現はディープフェイク該当性の段階で規制対象から外れる一方、外形上は真正に見える（政治）風刺コンテンツはディープフェイクに該当したうえで、簡略化規定によって生成AIによる表現であるとの開示の態様が簡略化される。この簡略化措置はデプロイヤーの開示義務に限られ、プロバイダーによる機械可読なマーキング義務（第50条2項）は（政治）風刺コンテンツにも及び続ける。すなわち EU AI 法は、（政治）風刺表現を完全に規制対象外としているわけではなく、作品の画面上に表示されるラベルのように、受け手の目に見える形で開示を求める場面、つまり表現の自由に最も直接的に影響する部分に限って規制を緩和している。さらに、機械可読の電子透かしのように作品の見え方を変えない技術的手段によって欺罔や誤情報のリスクにも備えている。このように、EU AI 法は（政治）風刺表現の保護とディープフェイクによる欺罔・誤情報という実害への対処のバランスを法的に担保しようとしているといえる。

他方で、法定性の要件との関連では課題も残る。ディープフェイク該当性と開示義務の簡略化措置の対象となるかどうかの判断が当該要件を満たすには、ディープフェイク該当性に関わる真正性への誤認可能性と、「明らかに」風刺的等であることの該当性がいずれも十分に明確

でなければならない。EU AI 法において、その具体化は主として非拘束的なガイドラインに委ねられており、権威的な解釈権限は最終的には欧州連合司法裁判所（CJEU）に留保されている（European Commission, 2026, para 5）。同ガイドラインにより、規制対象となるディープフェイクの様式・技法の特徴、それが提示される文脈、受け手の期待といった考慮要素は相当程度要件化されているが、現時点ではそれが拘束力ある規範のレベルで確立されているわけではないため、法定性については一定の曖昧さが残る。

（2）米国法における規制

i 連邦レベルの規制

米国では、EUのような包括的なAI規制法は連邦レベルでは存在しない。連邦議会は数多くのAI関連法案を審議してきたがどれも成立には至っておらず、本稿執筆時点（2026年8月）でも連邦による規制は大統領令や個別分野の立法にとどまる（DLA Piper, 2026; The White House, 2026）。ディープフェイクについても、非同意の性的画像のオンライン投稿を禁止し通知後48時間以内の削除を義務付ける2025年成立のTAKE IT DOWN Act (SEC. 3 (a) (3)) や、被害者に民事上の損害賠償請求権を認め2026年1月に上院を通過したDEFIANCE Act（本稿執筆時点では下院で審議中）（U.S. Senate, 2025）のように、欺罔の意図が明確で性的被害をもたらすディープフェイクを対象とする立法が実現の途上にあるにとどまる。本稿の検討対象である政治風刺目的のものを含むディープフェイクを一般的に規制する連邦法は現時点で存在せず、AI規制は各州に委ねられている（DLA Piper, 2026）。

この状況を踏まえ、本稿はカリフォルニア州における政治風刺ディープフェイクの規制について検討する。というのも、同州は政治的な場面におけるディープフェイクの規制について最も早くから継続的に立法を重ねてきたためである。同州は以下にみるような選挙におけるディープフェイクを規制する立法を進めるだけでなく、California AI Transparency Act (SB 942; AB 853) や Transparency in Frontier Artificial Intelligence Act (SB 53) 等、全米で最も広範な関連立法を展開する州の一つでもある。

ii カリフォルニア州法

（a）California Elections Code におけるディープフェイク関連の規定

カリフォルニア州は2019年、California Elections Code に、選挙候補者に関する重大な誤解を生じさせる画像・音声・映像の頒布を選挙の60日前から禁止する規定（AB 730、§ 20010）を新設した。当該禁止は、風刺・パロディに該当するコンテンツを一律に適用除外とし（§ 20010 (d) (5)）、同コンテンツがディープフェイクであることの開示も求めない³⁾。

これとは別に、ニューサム州知事は2024年9月、候補者個人に限定されない一般的な選挙関連メッセージを対象として、重大な誤解を生じさせるコンテンツの頒布を選挙120日前から

60日後まで禁止する AB 2839 (§ 20012 を新設) に署名し、同日、大規模プラットフォームに選挙関連の欺罔的コンテンツの削除・報告体制整備を義務づける AB 2655 にも署名した。当該禁止は、風刺・パロディに該当するコンテンツについては所定の書式による定型の開示を付すことを条件として適用除外としている (§ 20012 (b) (3))。

§ 20010 と § 20012 が併存することにより、同じ州の選挙法体系の中で風刺・パロディの扱いについて禁止規定からの「無条件除外」と「開示付き除外」が併存し矛盾しているようにも見えるが、両者は対象範囲を異にする。§ 20010 は規制対象を候補者本人に関するコンテンツに限定するのに対し、§ 20012 は候補者個人に限られない選挙関連メッセージ一般を対象としている。条文上、両者の適用関係を調整する明文の規定は存在せず、対象が重なる場面でどちらが優先するかは、特別法(候補者個人に関する § 20010) が一般法 (§ 20012) に優先するという解釈上の整理に委ねられると考えられる。

(b) Kohls v. Bonta 事件

上述の規定のうち、§ 20012 に関わる AB 2839 は、カリフォルニア州東部地区連邦地方裁判所(以下、裁判所)で争われた Kohls v. Bonta 事件(Kohls v. Bonta, 2025a)において恒久的差止が命じられた。裁判所は、同法案が内容に基づく規制にあたるものとして、最も厳格な違憲審査基準である厳格審査を適用した。裁判所は、§ 20012 に定めるコンテンツ規制が合憲であるためには、同規制が「やむにやまれぬ利益」のためのものであり、利用可能な手段のうち最も制限的であると証明できない限り正当化できないと判示した(Kohls v. Bonta, 2025a, 8-9) うえて、被告が主張する選挙の公正性の確保および有権者の混乱防止という利益はそうした利益にあたるものの、被告は代替可能な手段が存在しなかったことを示す証拠を提示しなかったとして、AB 2839 の合衆国憲法修正第1条違反を認めた(Kohls v. Bonta, 2025a, 9-13)。

さらに裁判所は、風刺・パロディ向けの開示付き例外規定についても、上記の憲法上の瑕疵を治癒するものではないと判断した。義務的な開示要件は、表現者に国家が定めた文言の表明を強制するものであり、とりわけ映像コンテンツでは当該表示が画面の相当部分を占めることで原告らのメッセージをかき消すものであるためである(Kohls v. Bonta, 2025a, 15-18)。

本件についてカリフォルニア州は2025年9月に第9巡回区控訴裁判所へ控訴しており(Case No. 25-6138)、同事件は本稿執筆時点でも係属中である。他方、AB 2655 も恒久的に差し止められている(Kohls v. Bonta, 2025b) が、§ 20010 は2027年1月1日の失効まで現行法として効力を維持している。

iii 小括

カリフォルニア州法上、選挙時の風刺的ディープフェイクへの制限について、2024年の時点では § 20010 という「無条件除外型」と、§ 20012 というディープフェイクであることの開示

を条件に表現することが許される「開示付き除外型」の 2 つの制限の在り方が存在したが、Kohls v. Bonta 事件により後者は現時点で無効化されており、前者の無条件除外型のみが有効となっている。

§ 20010 は前述のとおり 2027 年 1 月 1 日に自動失効する予定であったが、これを 2031 年まで再延長する AB 686 法案（2025-2026 会期）が現在審議されている。同法案は 2025 年 5 月 27 日に州下院で可決された後、本稿執筆時点において上院委員会で審議中である（LegiScan, n.d.）。Kohls v. Bonta 事件で § 20012 が差し止められている一方で、§ 20010 という無条件除外型の規律が存続・強化されようとしている点は、アメリカにおける表現の自由の価値の特殊性が表れているといえよう。

自由権規約第 19 条 3 項との適合性については、米国法は現時点で規制対象を法文上明確なカテゴリーに限定するという点で同項が求める要請に適合的な構造を持っている。ただし、外形的には政治風刺である欺罔的ディープフェイクにより生じる害悪について規制がかけられていない点において、同条約の別の条文（第 25 条など）との適合性の問題が生じる可能性があることを指摘しておきたい。

結論

本稿は、「自由権規約第 19 条 3 項に適合的な政治風刺ディープフェイク規制は、いかなる要件を備えるべきか」を問いとして設定し、政治風刺ディープフェイクの規制に対して自由権規約第 19 条 3 項の解釈から一般的に得られる示唆を提示し、さらにこれを検証するための具体例として EU 法と米国（カリフォルニア）法を取り上げた。

1 では、まずディープフェイクの技術的特性を確認した。生成技術の高度化と作成コストの劇的な低下により、情報の受け手が一見してコンテンツの真偽を判別することは困難となり、検出技術も生成技術の進展に追いつかない結果、政治風刺として保護されるべきディープフェイクの外延は曖昧になっている。そのような政治風刺ディープフェイクの保護・規制について自由権規約第 19 条 3 項は以下のような分析枠組を提示する。すなわち、①政治権力を標的とする風刺は手厚い保護を受けること、②風刺概念の外延の曖昧さは規制当局に広い裁量を与えかねず法定性に課題を残すこと、③純粋に政治権力を標的とする風刺の規制は第 19 条 3 項が列挙する正当な目的に対応させにくいこと、④必要性は個別の脅威との関係で具体的に示されねばならず、より制限的でない代替手段が存在する以上、政治風刺を含むディープフェイクの一律禁止は必要性・比例性の要件を満たしがたいこと、である。

続く 2 では、この枠組みに沿って対照的な二つのモデルを検討した。EU AI 法は、ディープフェイクをディープフェイク該当性と開示の簡略化措置の二段階で扱い、風刺的表現を全面的に規制対象から除外するのではなく、文脈や受け手の期待を考慮しつつ開示義務の態様を調整

する。これは、政治的な風刺への配慮を行った上で（上記①）、より制限的でない代替手段を用いることで比例性（④）にも適合的であるといえる。他方で、風刺と非風刺の境界を文脈に依拠して判断するため、法定性（②）については一定の曖昧さが残る。③正当な目的については、第50条4項に明示はないが、EU AI法の目的に挙げられる（第1条）「民主主義（～）を含む欧州連合基本権憲章に記されている基本的権利」の保障が目的であると解すれば、自由権規約第19条3項「他者の権利」という正当な目的に少なくとも合致すると考えられる。

これに対し、カリフォルニア法 § 20010 は、風刺・パロディを明確に適用除外とし、規制対象を選挙候補者に関する重大に欺罔的な合成メディアに限定する。さらに、Kohls v. Bonta 判決で見られたように、開示義務も政治的言論への萎縮効果をもたらすことから政治風刺表現への規制は極めて限定的であるべきという姿勢が明確に現れている。こうした米国法の規制のあり方は自由権規約第19条3項が提示した枠組み①～④から逸脱しないという意味において同項に適合的であると言える。ただし、真正な選挙の権利（自由権規約第25条）など、別の条文との関連で問題が生じる可能性は残される。これについては本稿の射程からは対象外であるため、別の機会に検討することとしたい。

以上の検討から、自由権規約第19条3項に適合的な政治風刺ディープフェイク規制のあり方として、以下のような示唆が得られる。まず、法定性の要請に応えるために、政治風刺の保護範囲を法文上明確なカテゴリーとして画定することが必要である。風刺該当性の判断を個別の文脈に委ねる限り、規制当局に広範な裁量が生じ、萎縮効果を避けがたい。カリフォルニア法のように、規制対象を候補者に関し重大な誤認を生じさせる性質の有無という要件で限定し、風刺・パロディを明文で適用除外とする手法は、この点で参照に値する。しかし、このような狭い規制を採用すると、風刺の外形を装った欺罔的なコンテンツを規制し損ねるというコストが生まれる。そこで、EU法に見られるように、規制の手段に段階を設けて調整可能なものとし、簡略的な開示義務というより制限的でない手段を通じて必要性・比例性の要請に配慮した規制が考えられる。以上を総合すれば、自由権規約第19条3項の枠組みからは、風刺該当性を法文上明確なカテゴリーによって画定することで法定性を確保する規制のあり方と、規制の手段を状況に応じて調整することで必要性・比例性の要請に応える規制のあり方という、少なくとも二つの適合的な規制類型が導かれる。もっとも、法定性の確保と欺罔的なコンテンツを規制対象とすることはトレードオフの関係にあり、いずれの規制類型も他方が抱える限界を完全には免れない。自由権規約第19条3項がこの二つの規制タイプのいずれをより適合的と解するか、あるいは双方を等しく許容するとみるかについては、条文解釈から一義的に導くことは難しく、自由権規約委員会の今後の実行の蓄積が待たれる。

注

- 1) これに関連して、国連の特別報告者は、オンラインプラットフォームにおける表現の自由の規制一般の文脈で、政治風刺を含む正当なコンテンツをも排除してしまう自動コンテンツモデレーションに懸念を表明している (UN General Assembly, 2025)。
- 2) Helm and Nasu も、フェイクニュースの定義の曖昧さが予見可能性を害すると指摘しており (Helm and Nasu, 2021, 309-310)、ディープフェイク規制でも同様の問題が生じうる。
- 3) 同条には失効条項 (サンセット条項) が付されており、2019 年制定時は 2023 年に失効する規定であった (§ 20010 (g)) が、AB 972 (2022 年) により 2027 年 1 月 1 日まで延長されている。

参考文献

- Chesney, Robert and Danielle Keats Citron (2019) "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *California Law Review*, Vol. 107, pp. 1753-1820.
- European Commission (2026) *Annex to the Communication to the Commission: Approval of the content of the draft Communication from the Commission - Guidelines on the implementation of the transparency obligations for certain AI systems under Article 50 of Regulation (EU) 2024/1689 (the 'AI Act')*, C (2026) 5054 final, Brussels, 20 July 2026.
- DLA Piper (2026) *AI Laws of the World: United States*, DLA PIPER INTELLIGENCE (last modified July 29, 2026), <https://intelligence.dlapiper.com/artificial-intelligence/?t=01-law&c=US>.
- Helm, Rebecca K. and Hitoshi Nasu (2021) "Regulatory Responses to 'Fake News' and Freedom of Expression: Normative and Empirical Evaluation," *Human Rights Law Review*, Vol. 21, No. 2, pp. 302-328.
- Holland, Steve (2024) "Fake 'Biden' Robocall Tells New Hampshire Democrats to Stay Home," *Reuters*, 23 January.
- Human Rights Committee (2011) *General Comment No. 34: Article 19 -- Freedoms of Opinion and Expression*, CCPR/C/GC/34, 12 September 2011.
- Kuźnicka-Błaszowska, Dominika and Nadiya Kostyuk (2025) "Emerging Need to Regulate Deepfakes in International Law: The Russo-Ukrainian War as an Example," *Journal of Cybersecurity*, Vol. 11, No. 1, article tyaf008.
- LegiScan. (n.d.) *California AB 686 (2025)*. <https://legiscan.com/CA/bill/AB686/2025> (last visited Aug. 21, 2026).
- Masood, Momina et al. (2023) "Deepfakes Generation and Detection: State-of-the-Art, Open Challenges, Countermeasures, and Way Forward," *Applied Intelligence*, Vol. 53, No. 4, pp. 3974-4026.
- Schabas, William A. (2019) *Nowak's CCPR Commentary: U.N. International Covenant on Civil and Political Rights*. 3rd rev. ed. Kehl: N.P. Engel.
- Recorded Future (2024) *Targets, Objectives and Emerging Tactics of Political Deepfakes: 2024 Report*.
- Rini, Regina (2020) "Deepfakes and the Epistemic Backstop," *Philosophers' Imprint*, Vol. 20, No. 24.
- Romero-Moreno, Felipe (2024) "Generative AI and Deepfakes: A Human Rights Approach to Tackling Harmful Content," *International Review of Law, Computers & Technology*, Vol. 38, No. 3, pp. 297-320.
- Security Hero (2023) "State of Deepfakes: Realities, Threats, and Impact," Security Hero, <https://www.securityhero.io/state-of-deepfakes/> (last visited Aug. 21, 2026).
- The White House (2026) *Legislative Recommendations: National Policy Framework for Artificial Intelligence*, March 2026.
- UN General Assembly (2025) *Threats to freedom of expression online in turbulent times: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*. A/80/341.
- U.S. Senate. (2025) *S. 1837 - DEFIANCE Act of 2025, 119th Cong*, <https://www.congress.gov/>

bill/119th-congress/senate-bill/1837/all-actions (last visited Aug. 21, 2026).
石川智也 (2025) 「EU の AI 法の実務対応と今後の展望」『ジュリスト』第 1616 号、29 頁。
欧州連合日本政府代表部 (2025) 『EU AI 法の概要』。
総務省 (2024) 『令和 6 年版情報通信白書』。
古川直裕他 (編) (2025) 『EU AI 法概説』(別冊 NBL No.192) 商事法務。

法令一覧

EU 法

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), OJ L, 2024/1689, 12 July 2024.

米国連邦法

S. 146, Pub. L. No. 119-12, 139 Stat. 55 (2025), the TAKE IT DOWN Act.
S. 1837, 119th Cong. (passed by the Senate Jan. 13, 2026; pending in the House as of Aug. 19, 2026), the DEFIANCE Act of 2025.

カリフォルニア州法

AB 730, 2019 Cal. Stat. ch. 493 (approved Oct. 3, 2019), codified at Cal. Elec. Code § 20010 [Elections: Deceptive Audio or Visual Media].
AB 972, 2022 Cal. Stat. ch. 745 (approved Sept. 29, 2022), amending Cal. Elec. Code § 20010 to extend its sunset to Jan. 1, 2027 [Elections: Deceptive Audio or Visual Media].
AB 686, 2025–2026 Reg. Sess. (Cal.) (passed the Assembly May 27, 2025; pending in the Senate as of Aug. 19, 2026) [Elections: Deceptive Audio or Visual Media].
AB 2839, 2024 Cal. Stat. ch. 262 (approved Sept. 17, 2024) [Elections: Deceptive Media in Advertisements].
AB 2655, 2024 Cal. Stat. ch. 261 (approved Sept. 17, 2024), the Defending Democracy from Deepfake Deception Act of 2024.
SB 942, 2024 Cal. Stat. ch. 291 (approved Sept. 19, 2024), the California AI Transparency Act.
AB 853, 2025 Cal. Stat. ch. 674 (approved Oct. 13, 2025), amending the California AI Transparency Act (S.B. 942).
SB 53, 2025 Cal. Stat. ch. 138 (approved Sept. 29, 2025), the Transparency in Frontier Artificial Intelligence Act.

判例

米国連邦裁判所

Kohls v. Bonta (2025a), 797 F. Supp. 3d 1177 (E.D. Cal. 2025). Case 2:24-cv-02527-JAM-CKD Document 101.
Kohls v. Bonta (2025b), 797 F. Supp. 3d 1177 (E.D. Cal. 2025). Case 2:24-cv-02527-JAM-CKD Document 98.