特集／AIとサイバーセキュリティを巡る動向と対応

〈5〉生成AIとテロリズム：新たな脅威と国際社会の対応策

株式会社 Strategic Intelligence 代表取締役社長 CEO
和田　大樹

はじめに

グローバルなテロリズムの動向は、21世紀に入り劇的な変容を遂げている。2001年の9.11同時多発テロ以降、アルカイダやイスラム国（IS）に代表されるサラフィージハード系のイスラム過激派は、インターネットやソーシャル・ネットワーキング・サービス（SNS）を活用し、過激思想の拡散、リクルート活動、資金調達を展開してきた。2010年代中盤には、イスラム国がフェイスブックやツイッター（現X）を駆使し、プロパガンダ動画や犯行声明を世界中に拡散し、外国人戦闘員の流入を促進した（1）。近年では、イスラム国ホラサン州（ISKP）がイランやロシアでの大規模テロ事件に関与し、欧州では関連するテロ未遂や逮捕者が相次いで報告されている（2）。これらの事例において、SNSは依然としてテロ組織の主要な情報伝達手段として機能している。

技術革新はテロリズムの形態に大きな影響を与える。インターネットの普及は、アルカイダによる「ホームグローン・テロリスト」や「ローンウルフ」型の単独テロを誘発し、SNSの進化はイスラム国などテロ組織のグローバルな影響力を強化した（3）。そして、近年、生成AI（人工知能）の急速な発展が、新たな脅威として注目されている。生成AIは文章、画像、動画、音声などを迅速かつ大量に生成可能な技術であり、テロ組織がプロパガンダ、偽情報拡散、資金調達、サイバー攻撃に悪用するリスクが指摘されている（4）。本稿では、生成AIとテロリズムの関係を詳細に分析し、イスラム過激派や白人至上主義などの極右勢力による生成AIの悪用リスク、実際の利用事例、そして国際社会が講じるべき包括的対策を論じる。本稿の目的は、生成AIがテロリズムに与える影響を技術的、法的、倫理的観点から評価し、具体的な対応策を提案することである。

本稿は以下の構成で展開する。まず、生成AIがテロにもたらす主要なリスクを、プロパガンダ、偽情報、リクルート、サイバー攻撃、資金調達の観点から詳細に検討する。次に、イスラム過激派や極右勢力による生成AIの実際の利用事例を分析し、最近の動向を検証する。そして、国際社会が講じるべき技術的、法的、倫理的対策を提案し、国際協力の枠組みを議論する。最後に結論をまとめ、生成AIの潜在的リスクを管理しつつ、技術革新の恩恵を最大化する道筋を提示する。

生成AIがテロリズムにもたらすリスク

生成AIは、大量のデータから学習し、人間と遜色ないコンテンツを生成する能力を持つ。この技術は、テロ組織にとって従来のインターネットやSNSを活用した手法を大幅に強化するツールとなり得る。以下では、生成AIがテロリズムにもたらす主要なリスクを、プロパガンダ、偽情報、リクルート、

サイバー攻撃、資金調達の観点から詳細に検討する。

プロパガンダの効率化とグローバルな拡散

生成 AI は、テロ組織がプロパガンダを効率的かつ大規模に作成・拡散する手段を提供する。特に、ディープフェイク技術を活用したリアルな動画や画像の生成は、テロ組織のプロパガンダ戦略を一変させる可能性がある。例えば、テロ組織の指導者による演説や攻撃の映像を偽造することで、支持者の士気を高めるだけでなく、社会的かつ経済的不満を抱く若者たちをより多く惹きつけることが考えられる。ディープフェイクは、従来のプロパガンダに比べ、低コストかつ短時間で高品質なコンテンツを作成可能であり、SNS を通じて瞬時に世界中に拡散される（5）。生成 AI を用いた動画は、視聴者の感情に訴えかける力が強く、視覚的リアリティにより信憑性を高める効果がある。これにより、テロ組織が限られたリソースで最大限の影響力を発揮することが懸念される。

また、生成 AI は、多言語対応のコンテンツ生成を可能にする点でも脅威である。イスラム国などのテロ組織は、特定の地域や文化的背景を持つターゲット層に対して、言語や文化的ニュアンスを反映したプロパガンダを迅速に作成してきたが、生成 AI を使用することで、例えば、アラビア語、英語、フランス語、ウルドゥー語など複数の言語でプロパガンダ動画を生成し、それぞれの地域の SNS ユーザーに訴求することで、グローバルなリクルート活動を強化する可能性がある（6）。生成 AI による多言語プロパガンダが、テロ組織の影響力を地域を越えて拡大させるリスクが指摘されているが、生成 AI はリアルタイムでのコンテンツ生成を可能にするため、テロ組織は特定の事件や政治的出来事に即座に対応したプロパガンダを作成し、タイミングを逃さずにそれを拡散できる。このような技術的進化は、テロ組織の影響力の拡大を助長し、国際安全保障に対する脅威を増大させる（7）。

さらに、生成 AI のプロパガンダへの応用は、視聴者の心理的影響を最大化する点でも優れている。例えば、生成 AI は、特定のターゲット層の文化的・宗教的価値観に訴えるストーリーテリングを自動生成できる。これにより、テロ組織は従来の静的で単調なプロパガンダから、動的で感情に訴えるコンテンツへとシフトできる。生成 AI を用いたプロパガンダが、若年層や疎外感を抱くコミュニティに対して特に効果的であるとの見解もあり、このようなコンテンツは SNS プラットフォームのアルゴリズムによって拡散されやすく、短期間で広範な影響を及ぼす（8）。

偽情報と世論操作の新たな手段

生成 AI は、偽情報や誤情報の生成を通じて、世論操作を容易にするツールとして悪用されるリスクがある。政治家の偽演説、捏造されたニュース記事、改変された画像や動画を生成することで、社会の分断や混乱を誘発することが可能である。2024 年 5 月、米オープン AI は、中国やロシアなどの関連組織が生成 AI を用いて、SNS やブログ上で世論誘導を試みた事例を報告した（9）。これらの事例には、日本に対する批判を煽るコンテンツや、欧米諸国での反移民感情を増幅する偽情報が含まれており、生成 AI が国家間の対立を悪化させるツールとして利用されつつあることが明らかになった。

テロ組織にとっても、偽情報キャンペーンは魅力的な戦略である。例えば、イスラム過激派は、欧米諸国での反移民感情を煽る偽情報を生成し、社会的緊張を高めることで、過激思想への支持を拡大しようとする可能性がある。2024 年の欧州での事例では、ISKP が生成 AI を用いて、移民コミュニティに対する攻撃を正当化する偽のニュース記事を拡散し、特定の国での社会的分断を助長したと報告されている（10）。同様に、白人至上主義などの極右勢力は、特定の民族や宗教に対するヘイトスピーチを増幅する偽コンテンツを作成し、暴力行為を誘発する可能性がある。極右勢力による生成 AI の悪用についての懸念が広がっている 。（11）。

偽情報の拡散は、SNS のアルゴリズムによって加速される。生成 AI によるコンテンツは、視覚的・感情的に訴求力が高く、ユーザーのエンゲージメントを高める傾向にある。プラットフォームのアルゴリズムは、こうしたコンテンツを優先的に推薦するた

め、偽情報が短期間で広範な影響を及ぼす。これにより、テロ組織は最小限の労力で最大限の社会的混乱を引き起こすリスクが懸念される。さらに、生成AIはリアルタイムで偽情報を生成・修正する能力を持つため、当局の検知や削除が追いつかない場合がある。

リクルート活動の個別化と高度化

テロ組織は、生成AIを用いてターゲット層に合わせた個別化されたリクルートメッセージを作成する能力を持つ。AIは、SNS上のユーザーデータを分析し、年齢、性別、文化的背景、関心事に基づいて最適化されたコンテンツを生成できる。例えば、若年層に対しては、冒険心や正義感を刺激する動画を生成し、中高年層には宗教的・政治的訴求を強調したメッセージを配信するなど、ターゲットに応じた戦略的なアプローチが可能である。2024年の欧州警察機構（ユーロポール）の報告では、ISKPが生成AIを用いて、特定のコミュニティに訴求するプロパガンダをSNS上で配信していた事例が報告されている（12）。この事例では、若年層をターゲットにした短編動画が、TikTokやYouTube Shortsで拡散され、視聴者のエンゲージメントを高めた。

個別化されたリクルート戦略は、従来のマス向けプロパガンダに比べ、過激思想への勧誘を効率的かつ効果的に行うことができる。生成AIは、ユーザーのSNS投稿や検索履歴を分析し、過激思想に傾倒しやすい個人を特定する能力を持つ。これにより、テロ組織は潜在的な支持者に対し、ピンポイントでメッセージを送信できる。さらに、生成AIはチャットボットや自動応答システムを通じて、ターゲットとリアルタイムで対話するコンテンツを生成可能である。例えば、テロ組織は、生成AIを活用したチャットボットを運用し、潜在的リクルートに対して個別に対応することで、信頼関係を構築し、過激思想への勧誘を進めることができる。

このような個別化されたアプローチは、SNSプラットフォームのアルゴリズムを悪用することでさらに効果を発揮する。プラットフォームは、ユーザーの関心に基づいてコンテンツを推薦するため、テロ組織が生成したコンテンツが適切なターゲットに届きやすくなる。これにより、テロ組織は従来よりも少ないリソースで、より多くの潜在的戦闘員や支援者を獲得することが懸念される。

サイバー攻撃の支援と技術的進化

生成AIは、サイバー攻撃を支援するツールとしても悪用される可能性がある。フィッシングメールやマルウェアのコードを自動生成する能力は、テロ組織が政府機関や重要インフラを標的にする際に利用される。例えば、生成AIを用いて、信頼性の高い偽メールを作成し、標的の個人や組織から機密情報を盗むことが可能である。2024年のサイバーセキュリティ企業CrowdStrikeの報告では、生成AIがフィッシングメールの文面を自動生成し、従来の検知システムを回避する事例が増加していると指摘されている（13）。これらのメールは、文法や文脈が自然であり、受信者が偽物と気づきにくい特徴を持つ。

さらに、敵対的攻撃（adversarial attacks）と呼ばれる手法では、AIシステムの脆弱性を突く入力データを生成し、セキュリティシステムを回避する試みが報告されている。例えば、生成AIを用いて、顔認証システムやマルウェア検出システムを欺くデータを作成することで、テロ組織はセキュリティの隙を突くことができる。2024年のサイバーセキュリティ報告書では、生成AIが悪意のあるコード生成に使用された事例が増加しており、テロ組織がこれを攻撃の準備段階で利用するリスクが考えられる。特に、重要インフラへの攻撃は、国家の安全保障に対する直接的な脅威となる。電力網、通信ネットワーク、医療システムなどを標的にしたサイバー攻撃は、社会に深刻な混乱を引き起こす可能性がある。

生成AIのサイバー攻撃への応用は、技術的知識が限られたテロリストにとっても実行を容易にする。従来、サイバー攻撃には高度なプログラミングスキルが必要だったが、生成AIはコード生成を自動化し、専門知識を持たない者でも攻撃を仕掛けられるようにする。2024年の米国国土安全保障省（DHS）の報告では、生成AIを活用したサイバー攻撃の試みが、テロ組織や犯罪組織の間で増加傾向にあると警告されている（14）。このような攻撃は、従来の防御策では対応が困難であり、新たなセキュリ