研究開発構想（個別研究型）

人工知能（AI）が浸透するデータ駆動型の経済社会に必要な AI セキュリティ技術の確立

# 「SYNTHETIQ X：フェイク情報拡散の防御と予防を実現する研究基盤」

## 研究開発実施報告書（年次）

令和6（2024）年度

研究代表者

越前　功

国立情報学研究所 情報社会相関研究系・教授

## １．当該年度における研究開発の実施概要

### （１）研究開発概要

本研究開発では、画像・映像、音声、テキストといった単一のモダリティ（データの種別）によるユニモーダルメディアに加えて、複数のモダリティで構成されたマルチモーダルメディアを対象に、フェイク情報拡散の防御技術と予防技術を確立します。さらに、リアル情報とフェイク情報からなる大規模データセットと多数の生成・防御・予防モデルで構成された、フェイク情報の生成とその防御・予防という攻防の実践に適したフェイク情報研究基盤（SYNTHETIQ X）を構築します。

### （２）実施内容と成果の概要（研究開発開始から当該年度末まで）

令和 6（2024）年度

期間内の 4 つの研究目的である、[目的 1] フェイク情報の生成技術とフェイク情報を用いた攻撃技術の検討、[目的 2] フェイク情報の拡散を防御する技術の確立、[目的 3] フェイク情報の拡散を予防する技術の確立、および[目的 4] フェイク情報研究基盤（SYNTHETIQ X）の構築、において設定した研究開発項目に取り組みました。具体的には、[目的 1]では、3D Deepfake 生成と大規模データセット構築の検討、および 3D 顔と音声を対象としたマスター情報の生成手法の検討に着手し、基本的な技術仕様を検討しました。[目的 2]では、改ざん領域を推定するローカライゼーション手法の検討、および AI による多様な改ざん処理にロバストな来歴管理手法の検討に着手し、基本的な技術仕様を検討しました。[目的 3]では、学習データの自動収集を不能にする敵対的サンプル生成手法の検討、およびフェイク情報の生成を不能にする予防手法の検討に着手し、[目的 4]では、防御手法と予防手法を対象とした持続可能なモデル更新手法の検討に着手し、それぞれにおいて、基本的な技術仕様を検討しました。

## ２．主たる研究分担者一覧

本研究開発には、主たる研究分担者は参画していない。