2025 年度年次報告書

先端国際共同研究推進事業
【2025 年度採択　日英共同公募　AI・情報分野】

【研究課題名】
統計的セーフガーディング：責任あるデータ駆動型学習パラダイムに向けた日英共同研究

研究代表者：
石田 隆（東京大学　准教授）

主たる共同研究者：
松田 孟留（東京大学　准教授）
二見 太（大阪大学　准教授）

相手側研究代表者：
Wenkai Xu（University of Warwick, Assistant Professor）

1. 研究進捗状況と成果

本研究課題「統計的セーフガーディング」は 2026 年 1 月 1 日に研究を開始した。2025 年度（2026 年 1 月～3 月）は研究期間の最初の 3 か月であり、プロジェクトの立ち上げと研究基盤の構築に注力した。

研究構想に関しては、データ汚染下での AI 信頼性を保証する統計的セーフガーディングの枠組み構築に向け、各 work package の推進体制を整備した。研究者間で研究ディスカッションを開始し、プロジェクト管理のための GitHub チームアカウントおよびウェブサイトを構築・公開した。ポスドク研究者 1 名と事務補助員 1 名の採用手続きを完了した。

国際ネットワーク構築・拡大に関しては、2026 年 4 月に東京大学・理化学研究所 AIP との共催による Statistical Safeguarding ワークショップおよび東京大学での Gathering イベントの企画・準備を行った。5 名の英 Lead、Co-Lead が日本を訪問し、ワークショップおよび Gathering イベントにも参加予定である。

国際頭脳循環に資する若手研究者の人材育成に関しては、英国への若手研究者派遣計画の策定を開始し、修士課程学生が英 Lead とオンラインミーティングを実施して研究テーマのマッチングを進めた。

2. 代表的な成果事例

なし（本年度は 3 か月間のプロジェクト立ち上げ期間であり、論文発表等の研究成果は次年度以降に向けた準備段階にある）