

「AI のリスクアセスメントガイドライン～生成 AI 編～」第 1 版公開のお知らせ

先進技術リスク検討ユニットでは、かねてより先進技術の中でも「AI」を題材としてそのリスクアセスメントの方法論について議論を重ねてきました。

AI をはじめとした先進技術のリスクの最大の特徴は、その「未知性」にあります。社会にとって新規性が高い先進技術のリスクは、従来の安全工学の発展の主要な領域を担ってきた原子力安全やプロセス安全分野の文脈でのリスクとは異なり、「そもそも何がリスクなのか」が未知であることが多く、これまでのリスクアセスメントの考え方を適用して対応するだけでは不足が生じる可能性がある、という課題を生み出してきました。

「AI のリスクアセスメントガイドライン～生成 AI 編～」は、こうした課題に対して、昨今の発展著しい「生成 AI」を題材として、

- ・「そもそも何がリスクなのか」を問い直す「リスク特定」
- ・それらリスクの性質や特徴を理解する「リスク分析」
- ・その理解を元にリスクへの対応方針を判断する「リスク評価」

のステップを含む、一連のリスクアセスメントの具体的な方法論を提案する試みです。

生成 AI は個人・組織・社会を大きく変えるポテンシャルを有しており、その活用はすでに好ましい影響を生み出しつつある一方で、常に同時に好ましくない影響をももたらす可能性がある以上、そのリスクの大きさや特徴を十分に認識した上で活用していく必要があります。本ガイドラインが、多様な目的に向かって AI を活用しようとする人々にとって、そのリスクを事前に把握し適切な対応を取るための一助として活用いただけるものになれば幸いです。

また、本ガイドラインはあくまで暫定版であり、今後の生成 AI の発展と共に新たな必要事項や対応項目が生まれてくる場合もあります。本ガイドラインをご一読いただいた皆様におかれましては、生成 AI をはじめとした未知性の高い先進技術のリスクへの向き合い方を考えるためのきっかけとしても本ガイドラインをご活用いただき、その方向性を模索していくための今後の議論にぜひご参画いただきたく、忌憚なきご意見・ご質問、その他コメントなどのご連絡をお待ちしております。

ガイドライン作成発行

2026 年 8 月 3 日

リスク共生社会推進センター 先進技術リスク検討ユニット

ユニットリーダー 鈴木智也