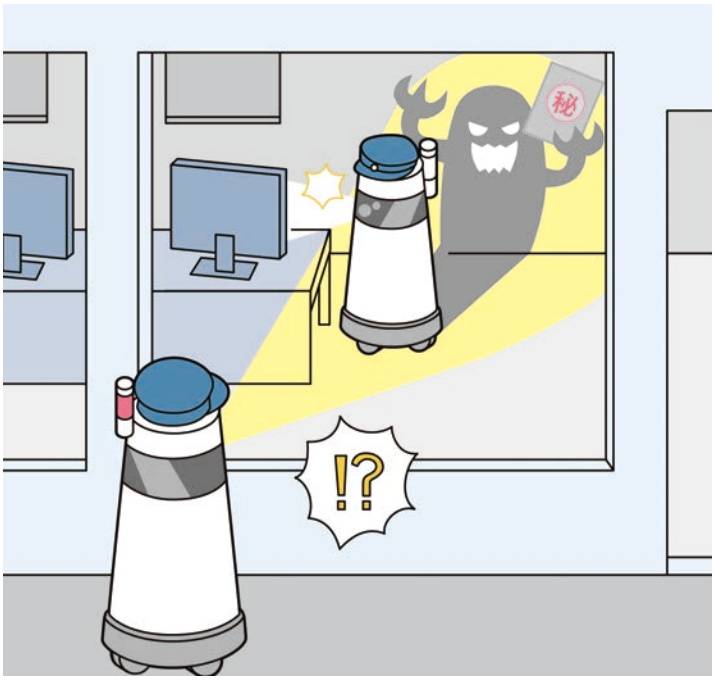
S-E01 AIのプロフィールによる同一性検証

IOWN Evolution

AIの同一性を検証して変化するAIを安心して利用できるようにします

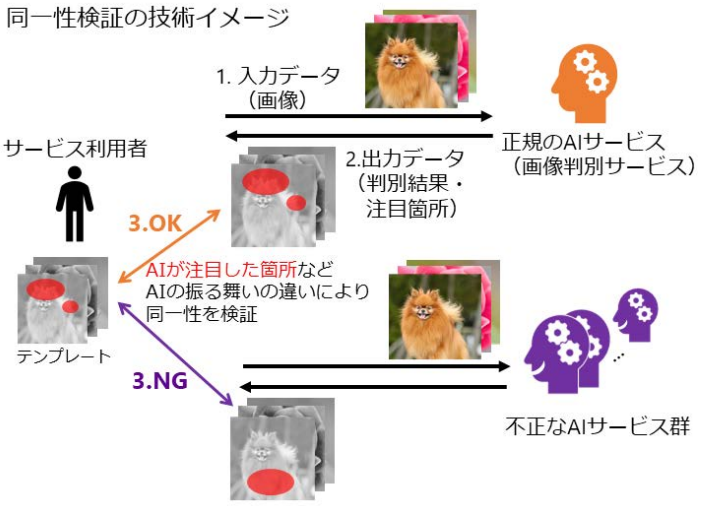


背景

今後AIが普及すると、予期せぬ変化や攻撃者の介入によって意図しないAIに変化してしまう恐れや、AIが不正にコピーされることで、AIのなりすましが発生し、利用者が被害を受ける可能性があります。そのため、利用者自身がAIの同一性を検証できる必要があります。

成果の概要

AIが攻撃を受け変化した場合やAIのなりすましが発生した場合には、AIの振る舞いからそれを検知し、AIが日々の学習を通じて変化した場合にはこれまで利用したAIのままかどうかを判別する技術の研究開発を開始しました。



技術のポイント１

入力データのどこにAIが注目したかや、入力データを加工した際の出力結果に着目した新しいAIの同一性の検証方法を提案

技術のポイント２

攻撃を受けた際に発生する影響に着目し、AIの同一性が失われたかどうかを判別

この研究がもたらす未来

AIのなりすましなどの不適切なAIを見極めることで、安心してAIを活用できます。これによってAIのビジネス利用を促進し、人とAIの共存共栄をめざします。

コラボレーションパートナー

国立大学法人静岡大学

出展企業

日本電信電話株式会社

問い合わせ先

rdforum-exhibition@ml.ntt.com