

科学技術の潮流

(348)

JST研究開発戦略センター

人工知能（AI）とを公表した。悪用され、人間が見つけられなかつた未知のシステム欠陥（ゼロデイ脆弱性）を「せいじやく」性を発見できるようになりつつある。これはコンピュータの脆弱性だけでなく、人間の心の脆弱性にまで及ぶかもしれない。

巧妙な攻撃

4月、米アンソロピックは、最新AI「クロード・ミクトス」が主要なコンピュータシステムのゼロデイ脆弱性を多数発見したこ

心の脆弱性 AIから守る

かねないとして、世界フェイク情報が、人々の中に衝撃を与えたが、の意思決定を惑わし、この脅威はサイバーセ判断を誤らせることがキュリティーだけにと社会問題となつていどまらない。最新AI。今後さらなる脅威国際社会で認識されつ

認知を守る

クトを推進している。欧州では民主主義社会への脅威となる外国による情報操作と干渉への対策強化が進む。



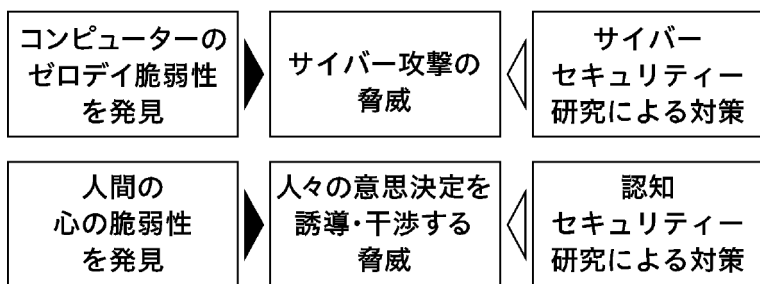
科学技術振興機構（JST）研究開発戦略センターフェロー（システム・情報科学技術ユニット） 福島 俊一

東京大学理学部物理学科卒、IT企業にて自然言語処理・情報検索の研究開発に従事後、16年から現職。工学博士。11～13年東大大学院情報理工学研究科客員教授、情報処理工学フェロー。

は人間の心の脆弱性になり得るのは、AIである。北大西洋条約も見つけてしまい、が利用者との親密な対話を通して、各人の性格、心の状態、反応・課題として研究を進め用され得る。近年、AIで思い通行動パターンなどを推定し、心の脆弱性を巧高等研究計画局（DA

このように、従来のサイバーセキュリティ研究による対策に加え、認知セキュリティの重要性が高まる一方、人間らしさ方法、信頼の一部でもあり、脳に形成のありパッチを当てるような方など、人対策は取れないから間の認知メだ。だからこそ認知セキュリティでは、脆弱性

AIによる脆弱性発見の脅威と対策研究



(筆者作成)

コンピュータの脆弱性、人間の心の脆弱性、サイバー攻撃の脅威、人々の意思決定を誘導・干渉する脅威、サイバーセキュリティ研究による対策、認知セキュリティ研究による対策。コンピュータの脆弱性、人間の心の脆弱性、サイバー攻撃の脅威、人々の意思決定を誘導・干渉する脅威、サイバーセキュリティ研究による対策、認知セキュリティ研究による対策。コンピュータの脆弱性、人間の心の脆弱性、サイバー攻撃の脅威、人々の意思決定を誘導・干渉する脅威、サイバーセキュリティ研究による対策、認知セキュリティ研究による対策。

(金曜日に掲載)