

○戦略目標「安全かつ快適な“人と AI の共生・協働社会”の実現」の下の研究領域

「人とAIの共生・協働社会を構成する要素研究と基盤技術の創出」

研究総括:山下 直美（京都大学 大学院情報学研究科 教授）

研究領域の概要

本研究領域では、信頼性・公平性・安全性等を考慮しながら、AI と人の共生や多様な AI の連携を可能とする技術の発展及びそれらを基にした複数の人と複数の AI による協働の実現を目指します。

AI と人が共生・協働する社会の実現に向けては、単なるシステムの精度向上にとどまらず、人や社会のウェルビーイングに資する AI やデザインの在り方を検討することが不可欠です。また、AI による社会課題の解決だけでなく、その普及がもたらす社会課題の予兆を捉え、未然に防ぐ手法も求められます。本研究領域では、このような観点に留意しながら、多様な AI と人間が共存し、調和した共生・協働社会の実現に向けた AI の開発やその要素技術の研究を推進します。また、AI が人や社会に与える影響を明らかにすることも重要な課題とし、システム開発に関する研究だけでなく、人文・社会科学の視点や、医療・教育・法律など AI が活用される分野ごとの専門的な視点を取り入れた研究にも取り組みます。

なお、本研究領域は文部科学省の人工知能/ビッグデータ/IoT/サイバーセキュリティ統合プロジェクト（AIP プロジェクト）に参画します。

募集・選考・領域運営にあたっての研究総括の方針

1. 背景

人工知能(AI)関連技術の急速な進化と社会への普及に伴い、高機能な AI エージェントが人間と協力し、複雑な問題の解決や業務の効率化、創造性の向上に貢献しています。一方で、多様な AI が乱立し、人間や AI エージェント同士が自律的に相互作用することで、人間による制御が困難になるというリスクも指摘されています。

AI と人が共生・協働する社会の実現に向けて、単なるシステムの精度向上にとどまらず、人や社会のウェルビーイングに資する AI やデザインの在り方を検討することが不可欠です。また、AI による社会課題の解決だけでなく、その普及がもたらす社会課題の予兆を捉え、未然に防ぐ手法の解決も求められます。

本研究領域では、多様な AI と人間が共存し、調和した共生・協働社会の実現に向けた AI の開発やその要素技術の研究を推進します。また、AI が人や社会に与える影響を明らかにすることも重要な課題とし、システム開発に関する研究だけでなく、人文・社会科学の視点や、医療・教育・

法律など AI が活用される分野ごとの専門的な視点を取り入れた研究も積極的に歓迎します。

2. 募集・選考の方針

(1) 基本方針

上記の背景の下、本研究領域は、多様な AI と人間が共存・協働する社会の実現に向けて、複雑で複合的な社会課題を解決し得る AI、あるいは AI の普及がもたらす社会課題の予兆をとらえて未然に防ぐ手法の開発などの要素技術を開発することを主眼とします。

なお、AI が人や社会に与える影響を明らかにすることも重要な課題とし、システム開発に関する研究だけではなく、人文・社会系研究の視点や、AI が活用される様々な分野、例えば医学、教育、法学、経済学などの専門的な視点を取り入れた研究も積極的に歓迎します。

具体的には以下のような研究に取り組めますが、必ずしもこれらに限定するものではなく、より自由で挑戦的な提案を期待します。

(2) 想定する研究内容

(ア) 人と AI の共生

人と AI が安全かつ快適に共生し、ともに成長するために必要となる技術の開発といったシステム開発に関する要素技術の開発の他、AI が人に与える影響に関する研究、AI 依存による人間の思考低下リスクの回避方法に関する研究。また、遠隔医療や、法律相談、オンライン授業のように、AI が人間の相談事項に対して回答する場合に、回答の信頼性を担保するための技術開発や、AI が人間の不安に寄り添うことが可能となるような技術、およびその理論的基盤など。

(イ) 多様な AI 間の連携

複数のエージェントが存在する場合に、互いが協調する仕組みの開発。また、エージェント同士の交渉に際して、依頼者のプライバシーやセンシティブなデータを保護し、大規模なエージェント群として管理あるいは制御する手法の開発。例えば自動運転車が相互に連携して交通ルールを守りつつ、個々に最適な行動をとるための連携技術の開発。また、AI 同士が勝手に話を進め、不適切な事態を招かないための予防措置など。

(ウ) 複数の人と複数の AI による協働

複数の人と複数の AI (エージェント) が混在する中で、合意形成を行うための理論やモデルの構築。例えば大規模言語モデル (LLM) を用いたエージェントシミュレーション。人と AI の相互学習、相互監視によりシステム全体の信頼性、安全性を保全する仕組みの開発。あるいはこうしたシステムにおいて人間が安心・共感する機構の研究など。

研究テーマ (ア) ～ (ウ) は CREST と共通の内容ですが、さがかけでは、そのうち1つ以上のテーマを選定し、より基礎的・要素技術的な内容を期待します。本領域は、若手研究者の育成も重要

視しており、挑戦的な研究テーマによる研究提案を期待します。

なお、研究開始後に開催される領域会議等を通じ、さきがけ研究者同士による連携テーマを設定し、連携テーマによる研究を推奨することも検討しています。

3. 研究費と研究期間

研究期間は3年半以内とします。研究費(直接経費)は提案内容の達成に必要な額を申請することとし、上限は4千万円とします。ただし、総括による精査の結果、採択にあたっては研究費の調整を行う場合がありますので、予めご了承ください。

4. 領域運営の方針

本研究領域では、情報科学技術を中心とした自然科学系の研究者による個人研究を対象とします。研究テーマは、2. で示した(ア)～(ウ)の課題のうち、どれか単独、あるいは複数の観点からの研究提案でも構いません。また、AI の社会への影響について十分に考慮するため、これら自然科学系研究者と、認知科学、行動社会科学、経済学、法学など多様な分野の人文・社会系研究者との意見交換・協力も推奨します。

なお、本研究領域は文部科学省の人工知能/ビッグデータ/IoT/サイバーセキュリティ統合プロジェクト(AIP プロジェクト)を構成する「AIP ネットワークラボ」の1研究領域として、理化学研究所革新知能統合研究センターをはじめとした関係研究機関等と連携した取り組みなどにも貢献していきます。