

「人とAIの共生・協働社会を実現する学際的システム基盤の創出」領域

(略称: 共生AI学際システム)

募集説明会資料

研究総括 和泉 潔

東京大学大学院 工学系研究科 教授

令和 7年 4月 14日



科学技術振興機構

CREST

人工知能をめぐる動向： 個の知能と集団の知能の繰り返し

- 第1次AIブーム(1956～1960年代)：探索・推論の時代
 - ダートマスワークショップ (1956) オーガナイズされた
複数ルールの集合
 - 数学の定理証明、チェスを指す人工知能等

...冬の時代 協調問題解決の研究開発(70-80年代)
- 第2次AIブーム (1980年代)：知識の時代
 - エキスパートシステム 単純なルール間の相互作用による
複雑な問題への取り組み
 - 医療診断、有機化合物の特定、...

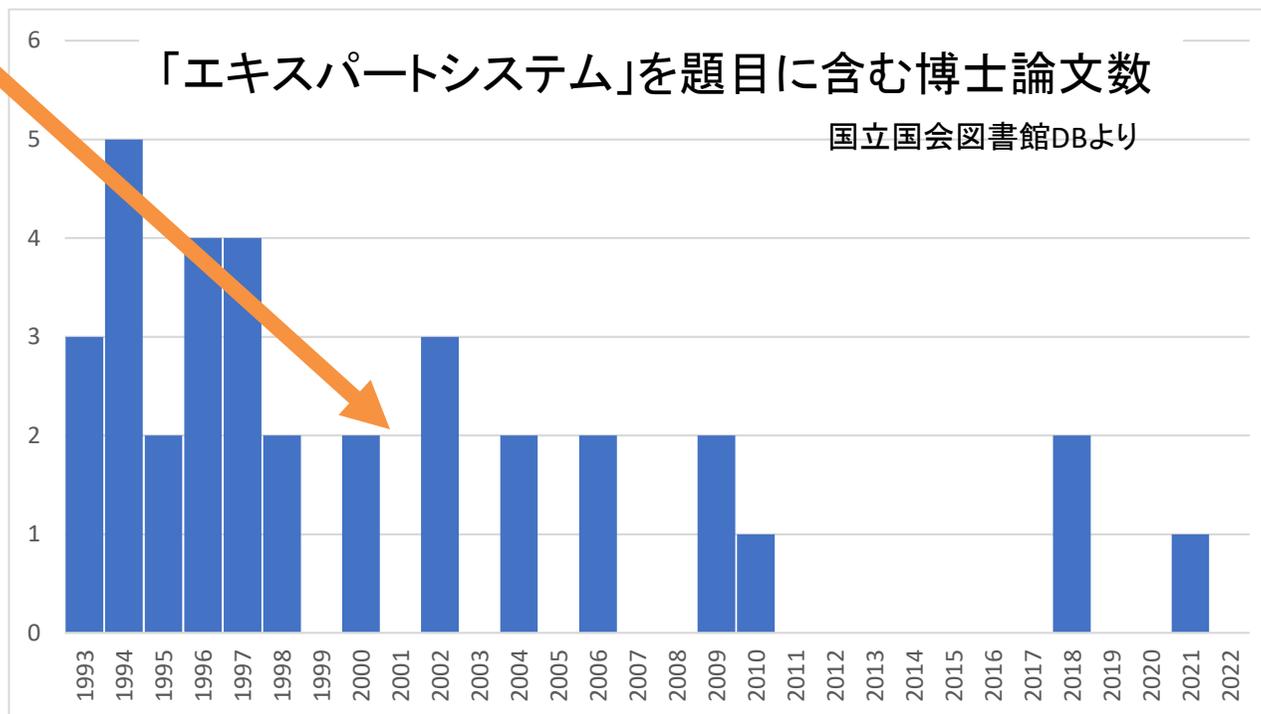
...冬の時代 エージェント/進化的計算(90-00年代)
- 第3次AIブーム (2013年～)：機械学習・ディープラーニングの時代
 - ウェブとビッグデータの発展 高度なAI同士の相互作用による
より複雑な問題への取り組み
 - 計算機の能力の向上

マルチAIエージェントの研究? (??年代)

和泉潔「マルチエージェント研究は人工知能の「残り滓」か？」人工知能 2024年39巻3号, pp.318-322
松尾豊「人工知能の未来 - ディープラーニングの先にあるもの」を元に改訂

人工知能をめぐる動向： 個の知能と集団の知能の繰り返し

- 第2次AIブーム（1980年代）：知識の時代
 - エキスパートシステム
 - 医療診断、有機化合物の特定、…
- …冬の時代



和泉潔「マルチエージェント研究は人工知能の「残り滓」か？」人工知能 2024年39巻3号, pp.318-322

人工知能をめぐる動向： 個の知能と集団の知能の繰り返し

- 第2次AIブーム（1980年代）：知識の時代

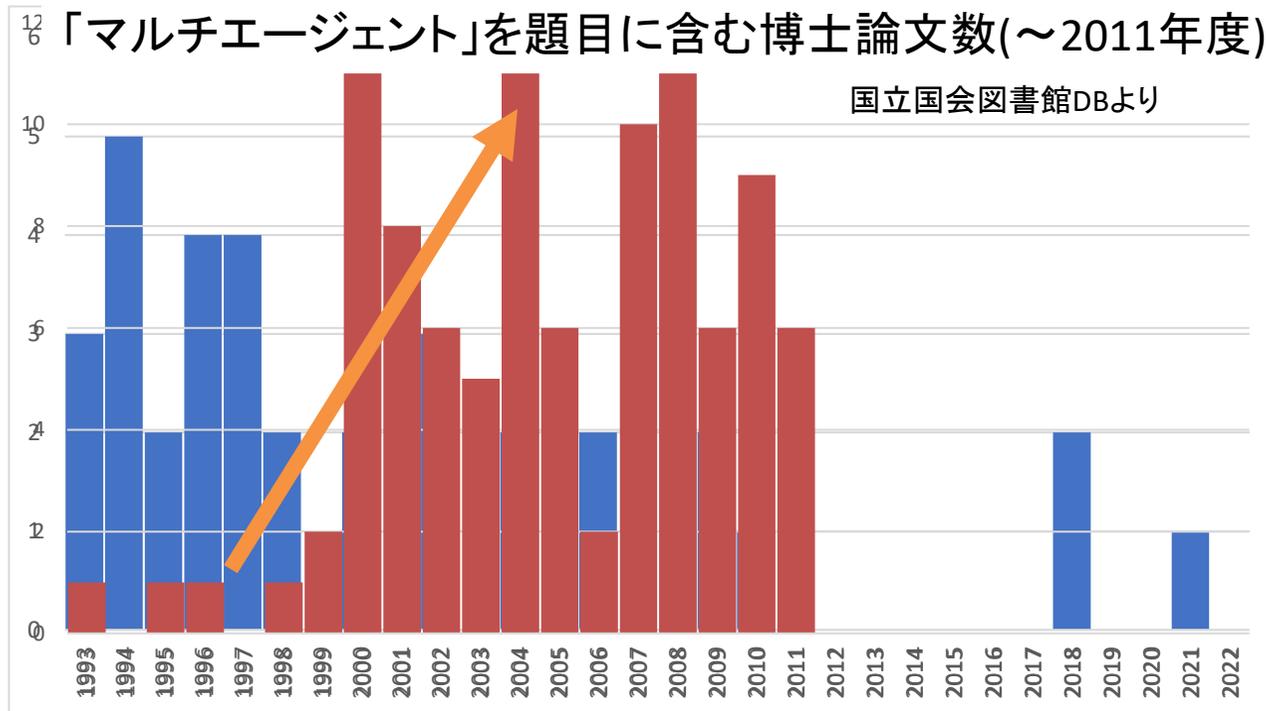
- エキスパートシステム

単純なルール間の相互作用による
複雑な問題への取り組み

- 医療診断、有機化合物の特定、...

...冬の時代

エージェント/進化的計算(90-00年代)

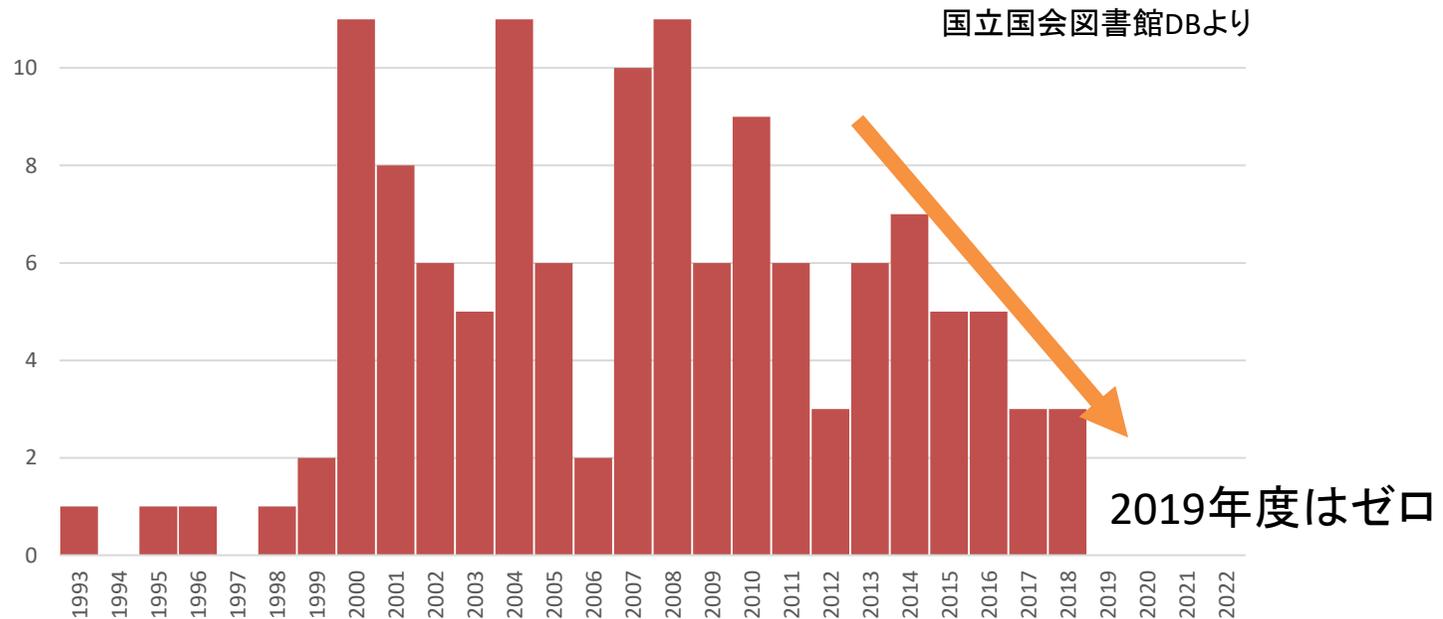


和泉潔「マルチエージェント研究は人工知能の「残り滓」か？」人工知能 2024年39巻3号, pp.318-322

人工知能をめぐる動向： 個の知能と集団の知能の繰り返し

- 第3次AIブーム（2013年～）：機械学習・ディープラーニングの時代
 - ウェブとビッグデータの発展
 - 計算機の能力の向上

12 「マルチエージェント」を題目に含む博士論文数(～2019年度)

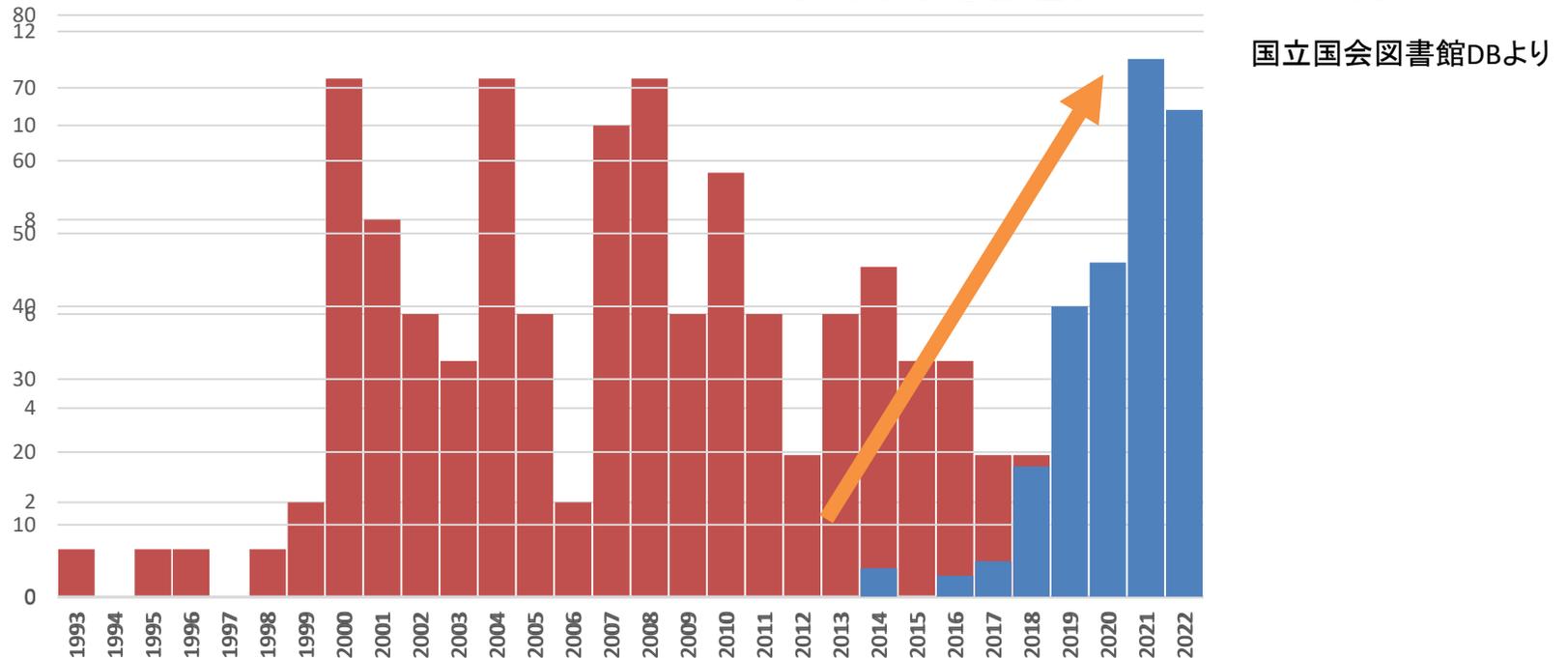


和泉潔「マルチエージェント研究は人工知能の「残り滓」か？」人工知能 2024年39巻3号, pp.318-322

人工知能をめぐる動向： 個の知能と集団の知能の繰り返し

- 第3次AIブーム（2013年～）：機械学習・ディープラーニングの時代
 - ウェブとビッグデータの発展
 - 計算機の能力の向上

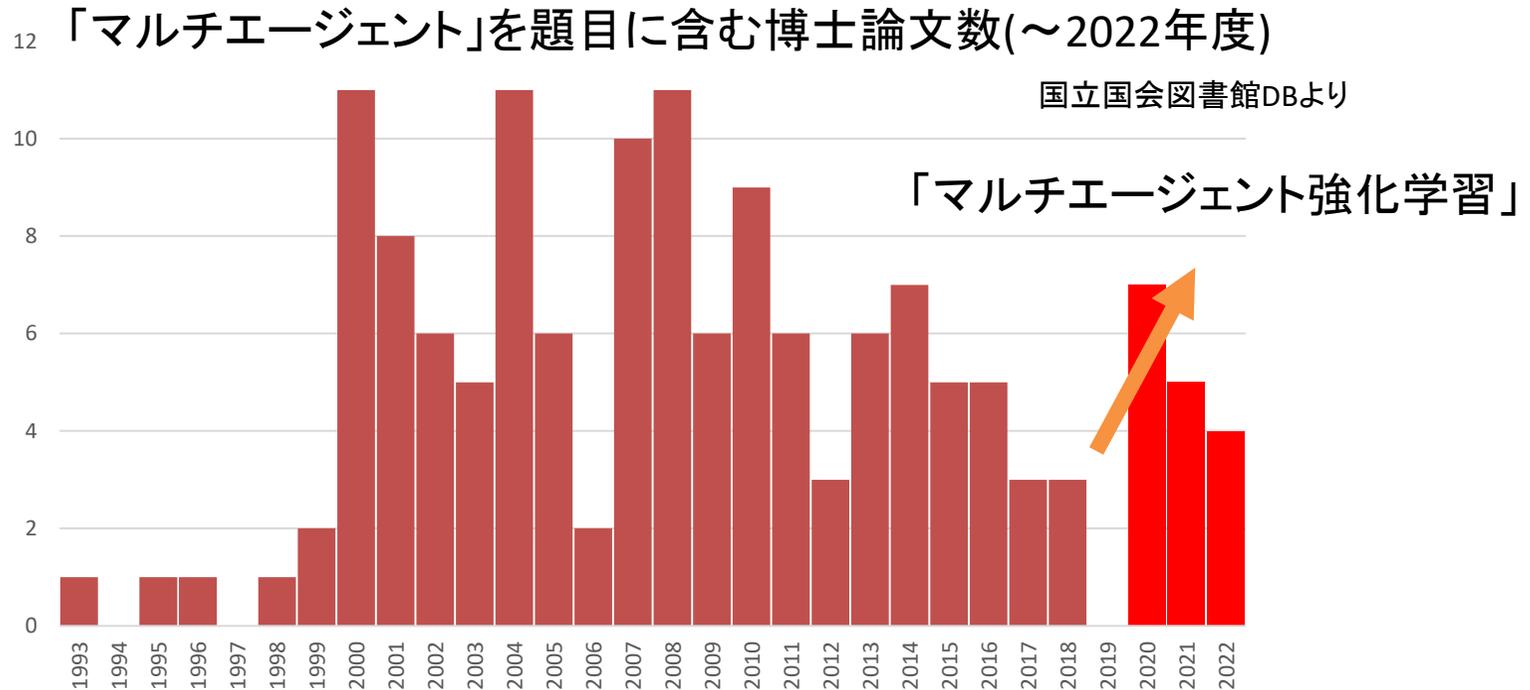
「深層学習」を題目に含む博士論文数



和泉潔「マルチエージェント研究は人工知能の「残り滓」か？」人工知能 2024年39巻3号, pp.318-322

人工知能をめぐる動向： 個の知能と集団の知能の繰り返し

- 第3次AIブーム（2013年～）：機械学習・ディープラーニングの時代
 - ウェブとビッグデータの発展
 - 計算機の能力の向上



和泉潔「マルチエージェント研究は人工知能の「残り滓」か？」人工知能 2024年39巻3号, pp.318-322

AIツールから マルチAIエージェントへ

Generated by ImageFX



戦略目標



安全かつ快適な“人とAIの共生・協働社会”の実現

アライメントに配慮した、多様な人と多様なAIの連携の推進

趣旨

- AIが急速に社会に普及する中、**進化したAIエージェントによる、他のAIや人間との協働への期待**が高まっている。
- 他方、AIが高度化・多様化することで、**人間がAIをコントロールできなくなる等の懸念**がある。



多様なエージェントが連携し、リスクを抑制しつつ価値創出を最大化するための研究開発を促進

達成目標

信頼性・公平性・安全性などを考慮しながら、以下の実現を目指す。

- 1 **人とAIの共生**：知識や意図の共有技術等
- 2 **多様なAIの連携**：相互運用プラットフォーム等
- 3 **複数の人と複数のAI協働**：社会デザイン等



期待

- ・複雑な社会課題の解決、社会システムの全体最適化
- ・多角的な議論や合意形成
- ・新たなビジネスを生むエージェント経済圏 等

懸念

- ・AIのシステミックリスク、AIの暴走
- ・有害なAI、ブラックボックス問題
- ・AIが持つ影響力の強さと社会的受容性 等

将来像

公平性や多様性に配慮した
包摂的な社会の実現



Well-beingや人間理解の深化と
人に寄り添ったAI技術の進化



生産性向上や労働力不足の解消、
新産業の創出、国際的プレゼンスの向上



領域の概要と背景

AIの自律化・高度化・多様化

AIの社会への浸透

- 複数のAI同士の連携
- 人とAIの協働
- AIによる多様なリスクを防ぐ

1. 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・知見の獲得
2. 多様なAI 同士それぞれの特徴を活かしてサイバー・フィジカル両面で協働するために必要となる情報科学技術の創出
3. 複数の人とAI が協働することでよりよい社会につなげるための理論体系や関係する個別要素技術を統合して仮想空間や実フィールドにおいて評価する手法の構築

研究開発の目標

- (1) 人とAIの共生
- (2) 多様なAI間の連携
- (3) 複数の人と複数のAIによる協働

＜関連する技術キーワード＞

AIエージェント、生成AI、大規模言語モデル(LLM)、人とAIの協働システム、AI行動規範、サイバー／フィジカルシステム、群知能、合意形成、社会デザイン、AIの社会受容性

具体的な研究課題例1

(1) 人とAIの共生

- 現実世界の場面において、人とAIの相互理解のために必要となる、人とAI間の共通理解(コモングラウンド)基盤の確立
- AIの論理的推論・説明技術の開発、人の脅威とならないようにAIの行動を律するための行動規範や“常識”の実装技術の開発
- AI依存による人間の主体性・思考低下リスクの回避方法

人間に対する理解や人間に与える影響を踏まえ、人とAIが安全かつ快適に共生し、共進化するために必要な技術・知見の獲得

具体的な研究課題例2

(2) 多様なAI間の連携

- 性質の異なるAI同士が相互運用できる協調プラットフォームの開発
- プライバシー・センシティブデータを保護しながら大規模AI群を管理・制御する手法の開発
- 群知能や能動的情報取得を実現するための技術

多様なAIがそれぞれの特徴を活かしてサイバー・フィジカル両面で協働するために必要となる情報科学技術の創出

具体的な研究課題例3

(3) 複数の人と複数のAIによる協働

- 実社会を想定し、より高度な社会課題の解決や社会システムの全体最適化を見据えた、複数の人と複数のAIによる協働環境・社会デザイン
- 認知バイアスを持ち得る人間とAIの討議・多角的意思決定など、複数の人とAIが協働することで、よりよい社会につなげるための理論体系
- 関係する個別要素技術を統合し、仮想空間あるいは実フィールドにおいて評価する手法の構築

具体的な研究課題例4

人文・社会科学的な視点から生じる研究も対象

- AIが実社会に導入され、人との接触を経て人に影響を与え、行動変容を図るための研究
- AIの社会受容性の国際比較と今後のAIの開発の方向性に関する議論

領域運営の方針

- AIの分野は極めて速く進歩が進むため、中間評価の時点で適宜目標を見直し、実際の技術進展に即した内容に変更することも可能とします。
- 社会との関わりについては、フィールドワークを主体とする人文・社会科学系の研究者との協力が有用と考えられることから、応募時点から人文・社会科学系の研究者をチームに加えることを推奨します。
- 研究期間中のメンバー追加、同戦略目標下のさきがけ(「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」領域：山下直美総括)研究者との協業も歓迎します。
- 本領域は、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。

研究費と研究期間

- 研究期間: 5.5年間(2025年10月から2030年3月末まで)
- 予算規模上限: 総額 3億円(間接経費を除く)

※申請の金額に対して、選考の過程で精査等により調整を行う場合あり

※期間中にも必要に応じた予算措置を検討する予定

領域アドバイザー

・調整中

スケジュール

応募締切:6月3日(火)正午

選考期間:6月中旬～9月中旬

研究開始:10月1日(予定)

応募にあたっての留意点

- 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チーム型研究を対象とします。研究テーマは、研究開発の目標(1)人とAIの共生、(2)多様なAI間の連携、(3)複数の人と複数のAIによる協働、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。
- 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・社会系研究者を加えたチームの研究も推奨します。

ご静聴ありがとうございました。

多数のご応募お待ちしております