

研究開発の俯瞰と潮流 ～科学技術イノベーションの動向と日本の活路～
**「情報技術の3つのトレンドと
生成AIのインパクト」**

2023年8月25日

JST研究開発戦略センター(CRDS)
システム・情報科学技術ユニット
福島 俊一



1. 情報技術の3つのトレンド

2. 生成AIのインパクト



研究開発の俯瞰報告書
システム・情報科学技術分野 (2023年)
[2023年3月発行]

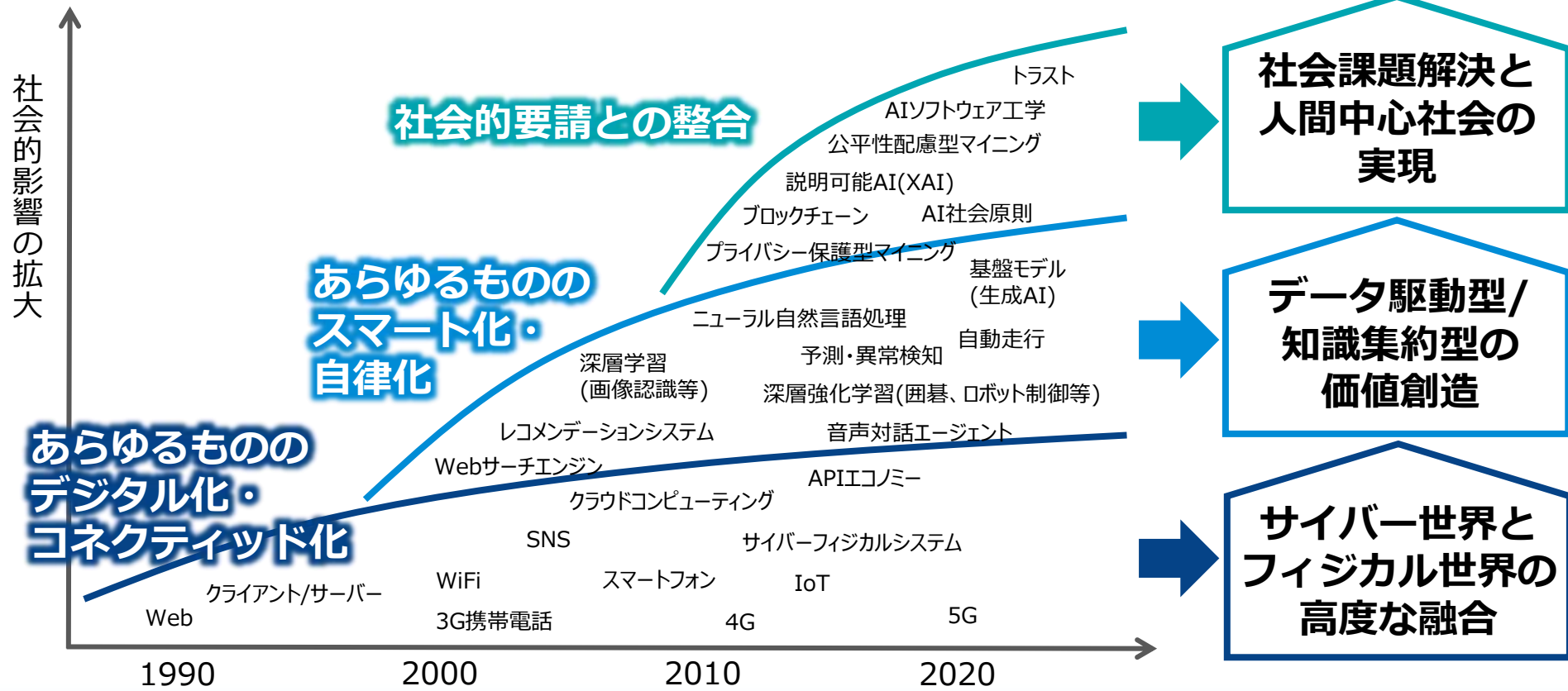
<https://www.jst.go.jp/crds/report/CRDS-FY2022-FR-04.html>

情報技術の3つのトレンドとビジョン

- 情報技術の3つのトレンドが重なり、Society 5.0の実現へ

Society 5.0

サイバー空間とフィジカル空間を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の社会

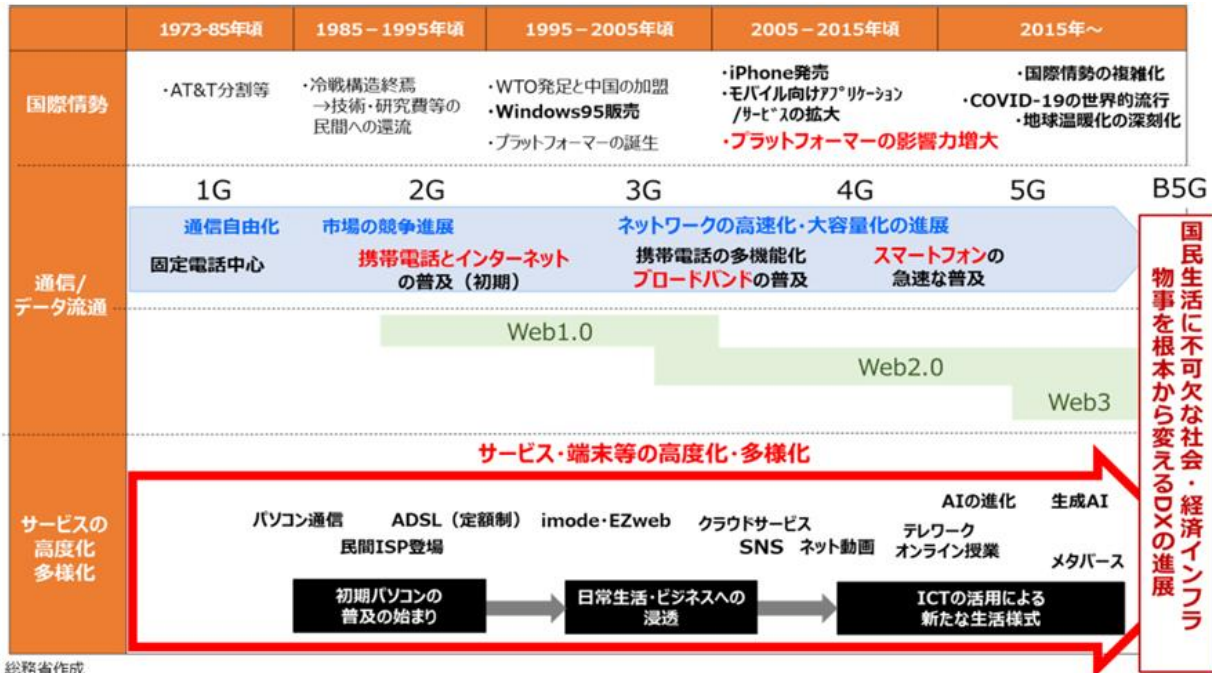


第1のトレンド あらゆるもののデジタル化・コネクティッド化

- サイバー世界とフィジカル世界の高度な融合に向けて
- 多様化する端末・環境の中、情報通信の大容量化・高速化、処理の分散協調・自動化が進展
- サステナブル社会のためのICT基盤が重要に

情報通信の大容量化・高速化

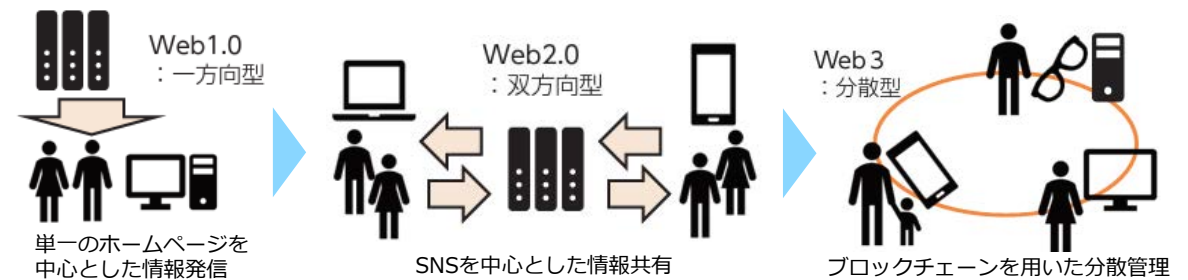
超高速・超大容量・超低遅延のデータ流通を可能とするBeyond5Gに向けた取り組みが活発化、社会に不可欠な基盤として重要性が高まる



処理の分散協調・自動化

閲覧中心の一方向型のWeb1.0、SNSを中心とした双方向型のWeb2.0に対して、近年新たに、ブロックチェーンを活用した分散型のWeb3が登場

Web3を応用したNFT (Non-Fungible Token: 非代替性トークン) やDAO (Decentralized Autonomous Organization: 分散型自律組織)も登場し、従来と異なる非中央集権型のビジネススキームとして注目されている



[図の出典] 両図とも総務省「情報通信白書」令和5年版から
<https://www.soumu.go.jp/johotsusintokei/whitepaper/r05.html>

第2のトレンド あらゆるもののスマート化・自律化

- データ駆動型/知識集約型の価値創造に向けて
- 目的特化AIから道具としての汎用AIへ発展、様々な知的活動の変革へ
- スマート化・自律化の根本である知能の原理探求が重要に

対話型生成AI

自然言語での入力に対して、まるで人間のような自然な対話応答や、専門的な知識・能力を備えているかのような応答が可能で、高い汎用性を示す

TO 基盤モデルについて、高校生向けに、300文字以内で説明してください。

基盤モデルは、大量のデータから学習し、言語を理解するAIプログラムです。自然言語処理（NLP）タスクで重要で、例えばGPTシリーズは、文章を生成したり、質問に答えたりできます。基盤モデルは、転移学習やファインチューニングで、さまざまなタスクに対応可能です。

ただし、計算リソースの大量消費や、学習データに含まれるバイアスが問題となります。技術者は、性能向上やバイアスの緩和方法を開発することが求められます。基盤モデルは、今後も多くの分野で活躍が期待されています。

ChatGPTでの対話応答例

水彩画風の猫



ヒューマノイドロボットと会話する少女

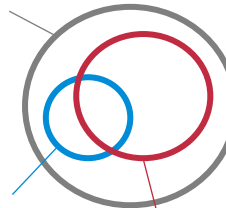


Midjourneyでの画像生成例

AI・ロボット駆動科学

AIを用いた大規模・網羅的な仮説生成・探索による人間の認知限界・バイアスを超えた科学的発見、ロボットを用いた仮説評価・検証のハイスループット化への取り組みが活発化

科学的発見の可能性の空間



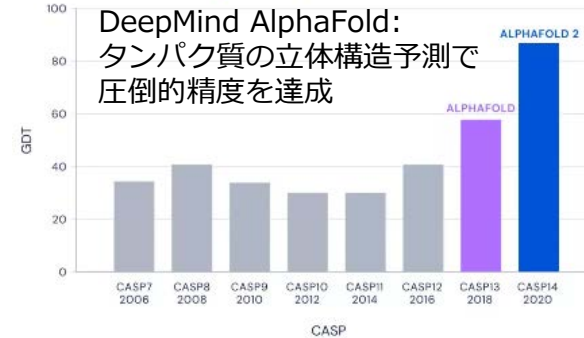
人間が認知できる範囲の科学的発見(認知限界・認知バイアスが存在)

AIによってリーチし得る科学的発見

ノーベルチューリングチャレンジ
(北野宏明が2016年にグランドチャレンジ提唱)

「2050年までに生理学・医学分野でノーベル賞級の科学的発見をできるAIシステムを作る」

Median Free-Modelling Accuracy



[図の出典]

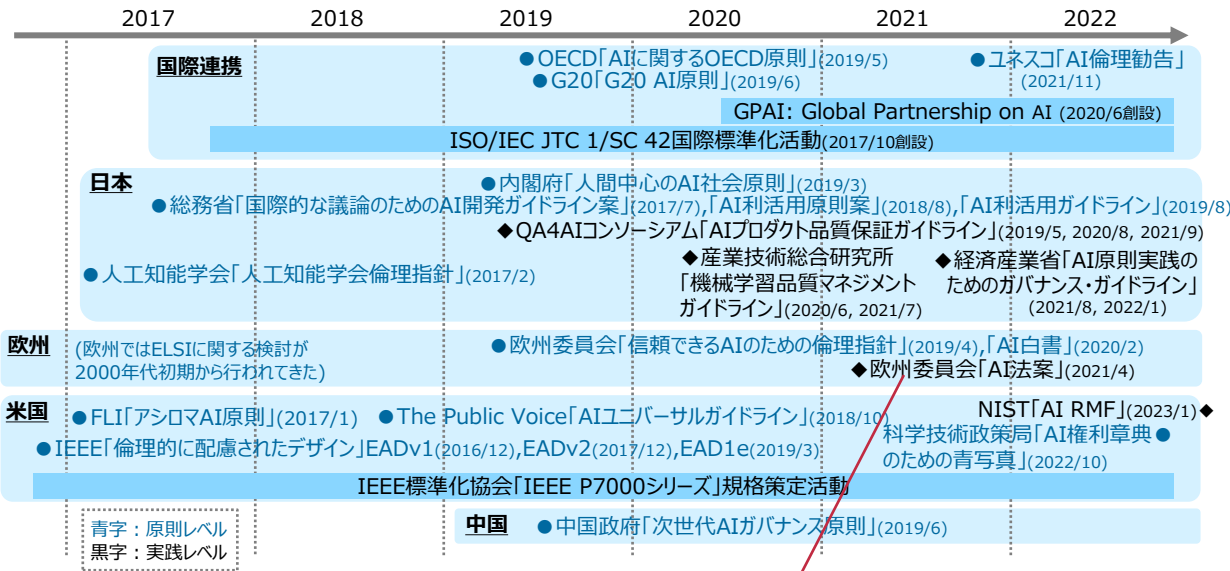
<https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>

第3のトレンド 社会的要請との整合

※ELSI: Ethical, Legal and Social Issues(倫理的・法的・社会的課題)

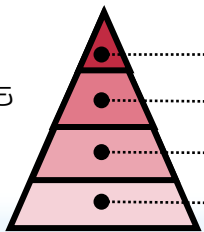
- 社会課題解決と人間中心社会の実現に向けて
- 複雑化する最先端技術・システムと人間・社会のより良い関係、懸念対策
- デジタル安全保障に対する総合知による取り組みが重要に

AI ELSI※は原則から実践のフェーズへ



欧州AI法案

2023/6 生成AI規制も盛り込むと決定、施行は2024年以降となる見込み

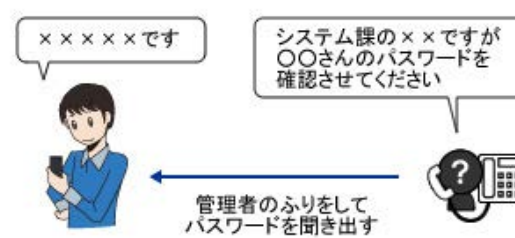


- 受容できないAI【利用禁止】
- ハイリスクAI【事前に適合性評価】
- 透明性義務を負うAI【透明性義務等】
- 極小リスク/リスクなしAI【規制なし】

コグニティブセキュリティ

システムの脆弱性だけでなく、人の脆弱性を突く攻撃が巧妙化
 悪意を持ったオンラインやオフラインでの誘導・干渉によって人々の行動・意思決定を含む認知過程に影響を与える問題に対処する取り組み「コグニティブ(認知)セキュリティ」の重要性が高まっている

ソーシャルエンジニアリング



フィッシング



フェイクニュース

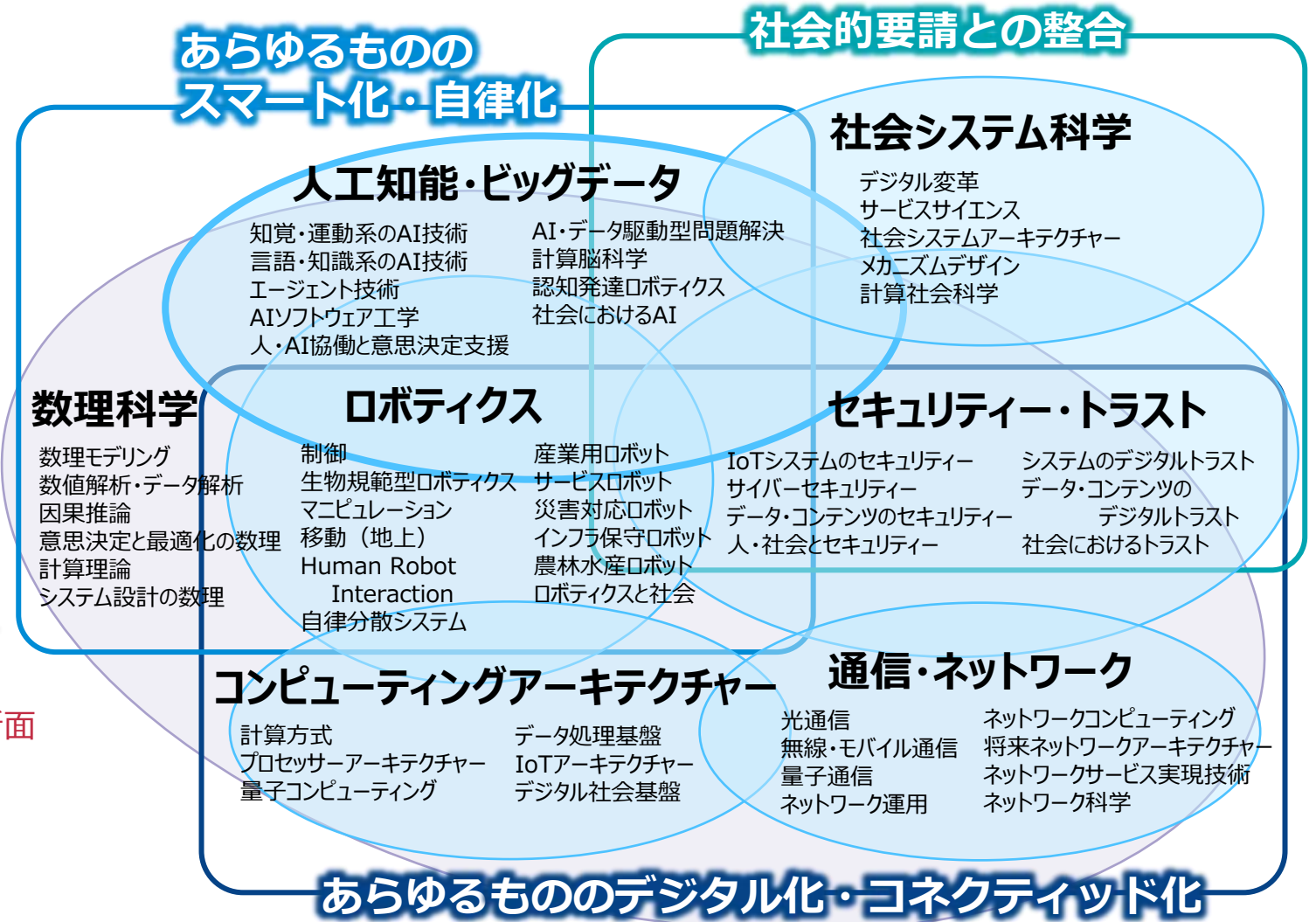
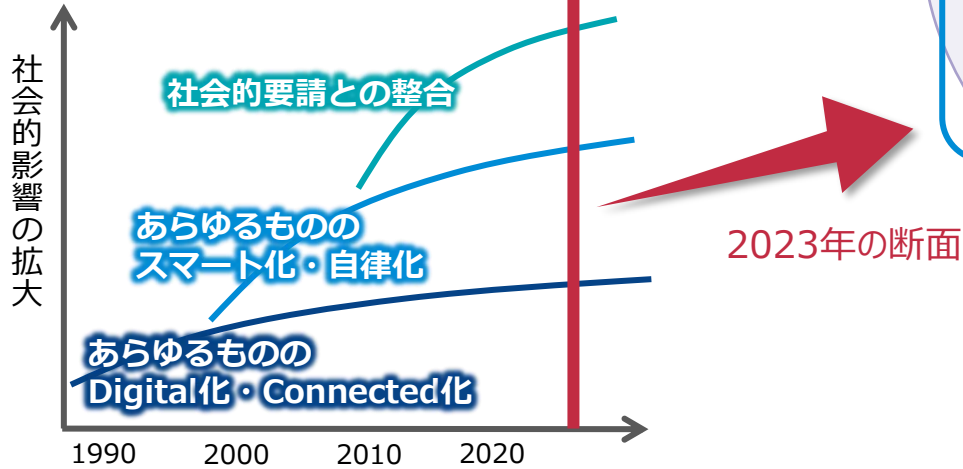


[図の出典] 「国民のための情報セキュリティサイト」 「インターネットとの向き合い方～二セ・誤情報に騙されないために～」 (いずれも総務省)

情報技術分野の研究開発の俯瞰



俯瞰報告書では、3つのトレンドに沿って7分野の研究開発動向を記載



1. 情報技術の3つのトレンド

2. 生成AIのインパクト



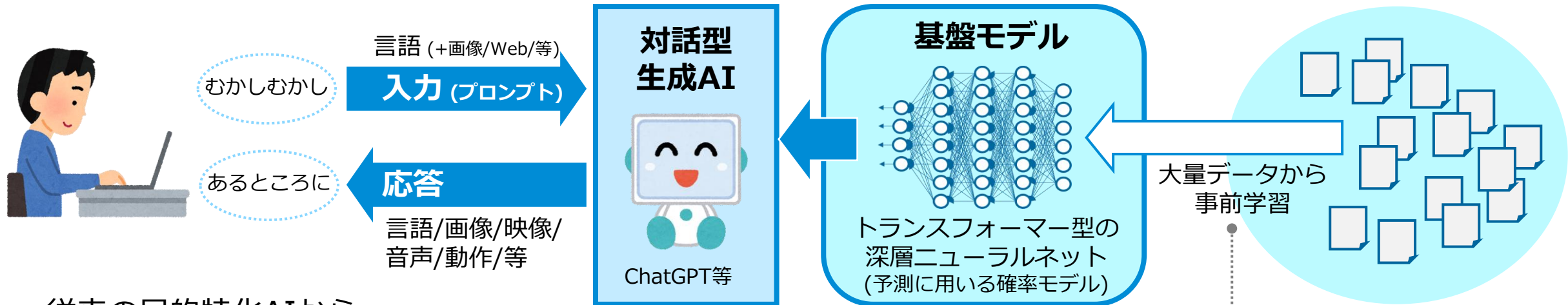
人工知能研究の新潮流2
～基盤モデル・生成AIのインパクト～

[2023年7月末発行]

<https://www.jst.go.jp/crds/report/CRDS-FY2023-RR-02.html>

対話型生成AIはどのようなものか

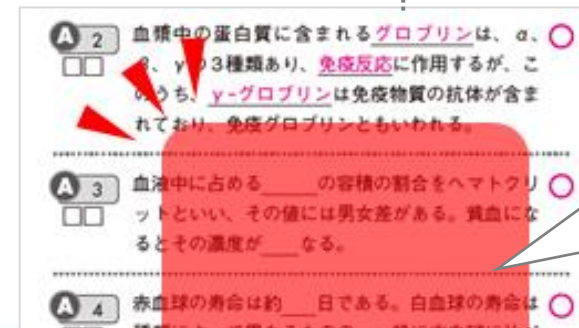
- 大量データから言葉のつながり傾向を学習した確率モデルに基づき、入力の続きを予測して応答
- 現在主流のトランスフォーマー型の深層ニューラルネットでは、幅広いコンテキスト情報を参照しつつ、予測のために注目すべき箇所を適切に選択できるようになり、予測精度が大幅に向上



従来の目的特化AIから 道具としての汎用AIへ

- まるで人間のような自然な対話
- 様々なタスクで、専門的な知識・能力を備えたような応答

世界中のテキストをもとに、自分で穴埋め問題を作って解く練習をひたすらしたら、聞かれたことに対して、たいてい続きを答えられるようになった！



事前学習：
穴埋め問題を自分で作って解く練習(自己教師あり学習)

基盤モデルの研究開発課題の全体観

基盤モデル応用開発(API利用)

- チャットボット、仮想アシスタント、問い合わせ自動応答、質問応答
- コンテンツ生成(文章、画像、映像)
- 翻訳、要約、ライティング支援
- 企画支援、発想支援 他

基盤モデル周辺拡張技術

- 基盤モデルが不得手な機能を扱う外部処理連携(最新情報検索、数式処理、物理シミュレーション、論理推論等)
- 問題解決ワークフロー設計の自動化
- プロンプトやワークフローの最適化 他

基盤モデル運用技術

- 継続運用可能なビジネスモデル(ビジネス用途、研究用途)、エコシステム
- データ追加・更新プロセス
- トラストを確保した運営体制 他

分野固有基盤モデル開発・活用

- プログラミング向け基盤モデル
- 個別企業業務向け基盤モデル
- 法業務向け基盤モデル
- 医療・ヘルスケア向け基盤モデル
- 教育向け基盤モデル

- 科学研究向け基盤モデル 他

利活用時の問題対処

- 生成AIの出力か否かの判定技術(フェイク検出技術、電子透かし技術を含む)
- 入出力データの著作権・肖像権関連問題への対処 他

AIアライメント研究

- 基盤モデル自体の倫理性確保(RLHF等)
- 生成AI応用システムの品質管理(プロンプト型開発法のソフトウェア工学等)
- 人間・AI共生社会のリスク低減(エージェント設計論、トラスト形成等) 他

基盤モデル構築技術

- 大規模深層学習モデル(トランスフォーマー、マルチモーダル)の実装
- 学習データの収集・選別・整備
- 大規模計算環境構築
- 高速化アルゴリズム、デバイス 他

次世代AIモデル研究

- 基盤モデル高効率化、生成AI高性能化
- 基盤モデルのメカニズム解明
- 人間知能の理解に基づくモデルの探求、基盤モデルとの融合
- 新モデル向けコンピューティング 他

↑ 応用個別

↓ 共通基盤

← 実務

→ 学術

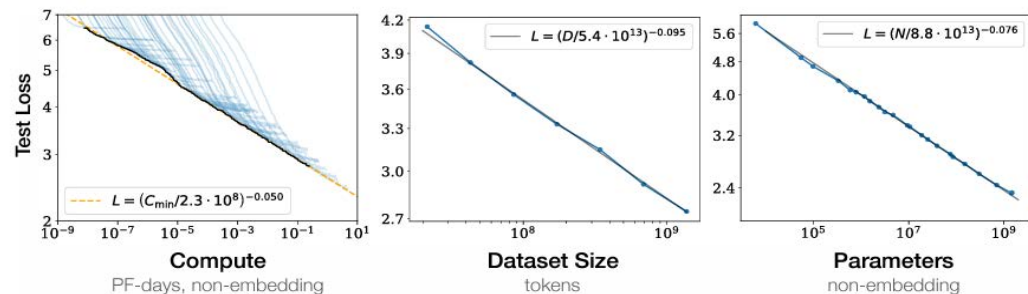
基盤モデル開発の動向

- スケーリング則や創発的能力が観測され、Big Tech企業による基盤モデル開発の超大規模化が加速 (OpenAI GPT-4、Google PaLM2等)
- ChatGPT以降、多数企業が参入、オープンソース化や軽量モデル化も進展
- ただし、高い予測精度や汎用性が得られるメカニズムは未解明

基盤モデル 応用開発 (API利用)	分野固有基盤モデル 開発・活用	
基盤モデル 周辺拡張 技術	利活用時の 問題対応 (フェイク/著作権等)	AIアライ メント研究 (人・社会との整合)
基盤モデル 運用技術	基盤モデル 構築技術	次世代AI モデル研究

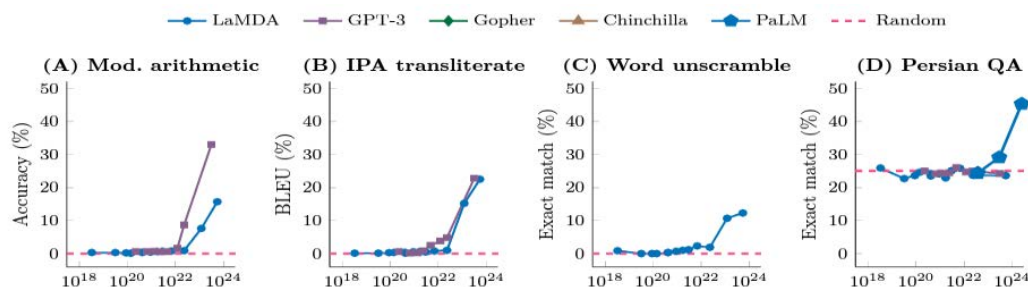
スケーリング則

基盤モデルは巨大化するほど性能が向上
(3変数のべき乗則に従う)

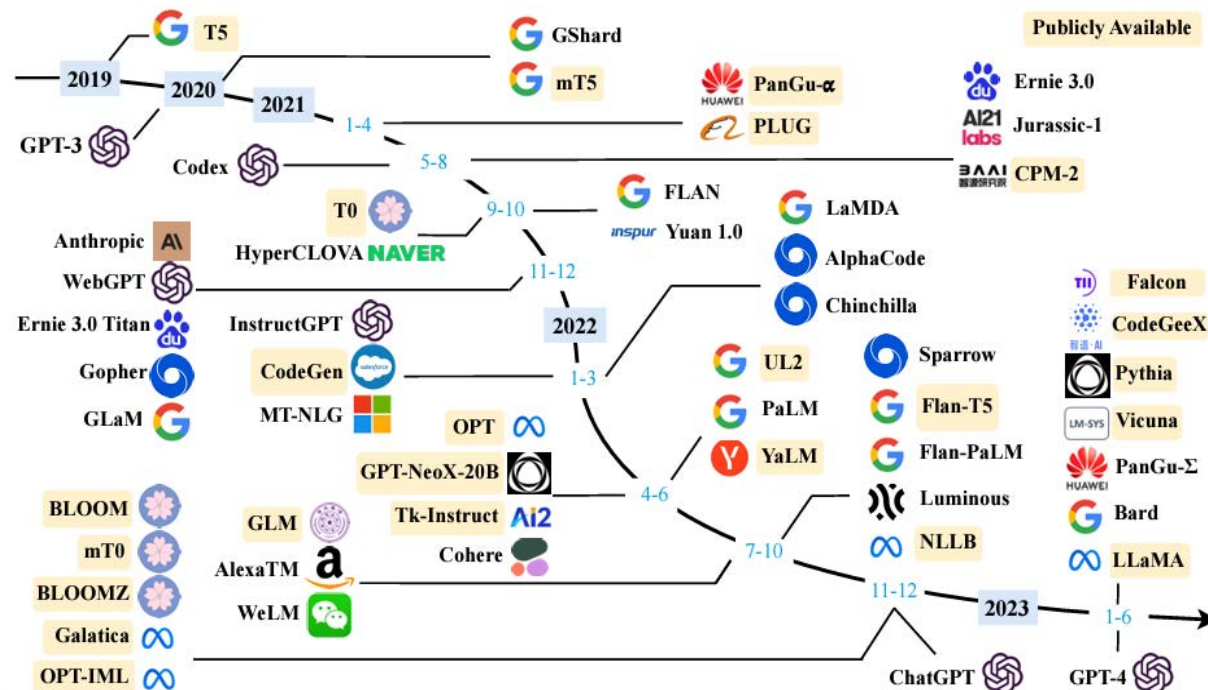


創発的能力

モデルの規模がある点を超えると性能が劇的に向上



大規模な言語モデル・基盤モデル開発の活発化



基盤モデル・生成AIの応用

例えばgpt3demo.comに応用800件以上

- 言葉で指示でき、汎用性が高く、**様々な応用が世界中で開発**されている
- 生成AIに意思・感情や真偽判別はなく、確率モデルに基づいて続きを予測するため、**論理推論・実世界操作等は不得意** (外部ツール連携で一部補完)

基盤モデル 応用開発 (API利用)	分野固有基盤モデル 開発・活用	
基盤モデル 周辺拡張 技術	利活用時の 問題対処 (フェイク/著作権等)	AIアライ メント研究 (人・社会との整合)
基盤モデル 運用技術	基盤モデル 構築技術	次世代AI モデル研究

プロンプト エンジニアリング

生成AIから望ましい応答を得るための
の**プロンプト(指示)の最適化手法**

- プロンプトでなるべく**詳しく的確に**
コンテキストを与える
- 目的・役割等の前提条件や口調・振
る舞い方等も指定して対話(**なりきり**
チャット)
- 文字数・難易度・出力形式等の応答
の制約条件の明示も有効
- “Let’s think step by step.” 手順の分
解によって精度が上がることもある
(**Chain-of-Thought Prompting**)



	そのまま利用する	分野に適応させる	独自モデルを開発する
モデル	訓練済みモデルをそのまま利用 (パラメータ固定)	モデルを修正(層の追加・削減、 再学習等でパラメータ更新)	独自モデルをスクラッチから 事前学習し、再学習もする
適応策	プロンプトエンジニアリング、 外部ツール連携	再学習、転移学習、ファインチューニング、蒸留、 人間からのフィードバックによる強化学習(RLHF)等	
構成	<p>プロンプト ↓ 訓練済み 基盤モデル ← 外部ツール 検索、記号処理、 四則演算 等</p>	<p>分野のデータ(ラベル付き)</p> <p>再学習、転移学習 (パラメータ更新) ↓ 独自に調整した 基盤モデル</p>	<p>分野固有 コーパス (文献、 知識ベース等)</p> <p>↓ 独自の 基盤モデル</p>
応用 事例	ChatGPT、Bard、 Microsoft 365 Copilot、 Walfram plugin、AutoGPT等	GitHub Copilot、 Galactica、PMC-LLaMA、 MatSciBERT等	タンパク質LLM (ESMFold他)、 化学・材料分野の分子LLM (MoLFormer他)等

基盤モデル・生成AIのELSI

ELSI: Ethical, Legal and Social Issues
(倫理的・法的・社会的課題)

基盤モデル 応用開発 (API利用)	分野固有基盤モデル 開発・活用	
基盤モデル 周辺拡張 技術	利活用時の 問題対処 (フェイク/著作権等)	AIライ メント研究 (人・社会との整合)
基盤モデル 運用技術	基盤モデル 構築技術	次世代AI モデル研究

- 学習データの傾向に依存し、ハルシネーション、社会的バイアス、悪用・思考誘導、情報漏洩、著作権侵害等、ELSI面の様々な懸念が指摘されており、多面的な対策技術が求められる
- 生産性向上・産業成長の期待の一方、G7等で国際的議論も進められている



• ウソや架空の出来事をあたかも事実であるかのように語る(ハルシネーション)

• 差別・偏見、偏った価値観が応答中に表れる(社会的バイアス)



• 学習データやプロンプトから個人情報・機密情報が漏洩
• 学習データや生成データの著作権・肖像権の問題(クリエイターの反発)



• 個人情報・機密情報は学習に使わない
• 正確性・安全性・倫理等を確保するようにモデルを調整(RLHF等)

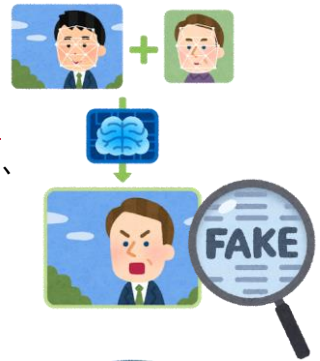


• 個人情報・機密情報の除外や品質確保をしていない
• それを偽ってもバレないだろう



• 特定主義・思想のプロパガンダ
• 汚染されたモデル、仕込まれたバックドア

• フェイク動画やフェイクニュースを生成、SNSで拡散して世論を誘導・干渉



様々な悪用



• 武器や毒薬の作り方等の悪知恵を聞き出し
• なりすましや詐欺メール等を生成、犯罪利用

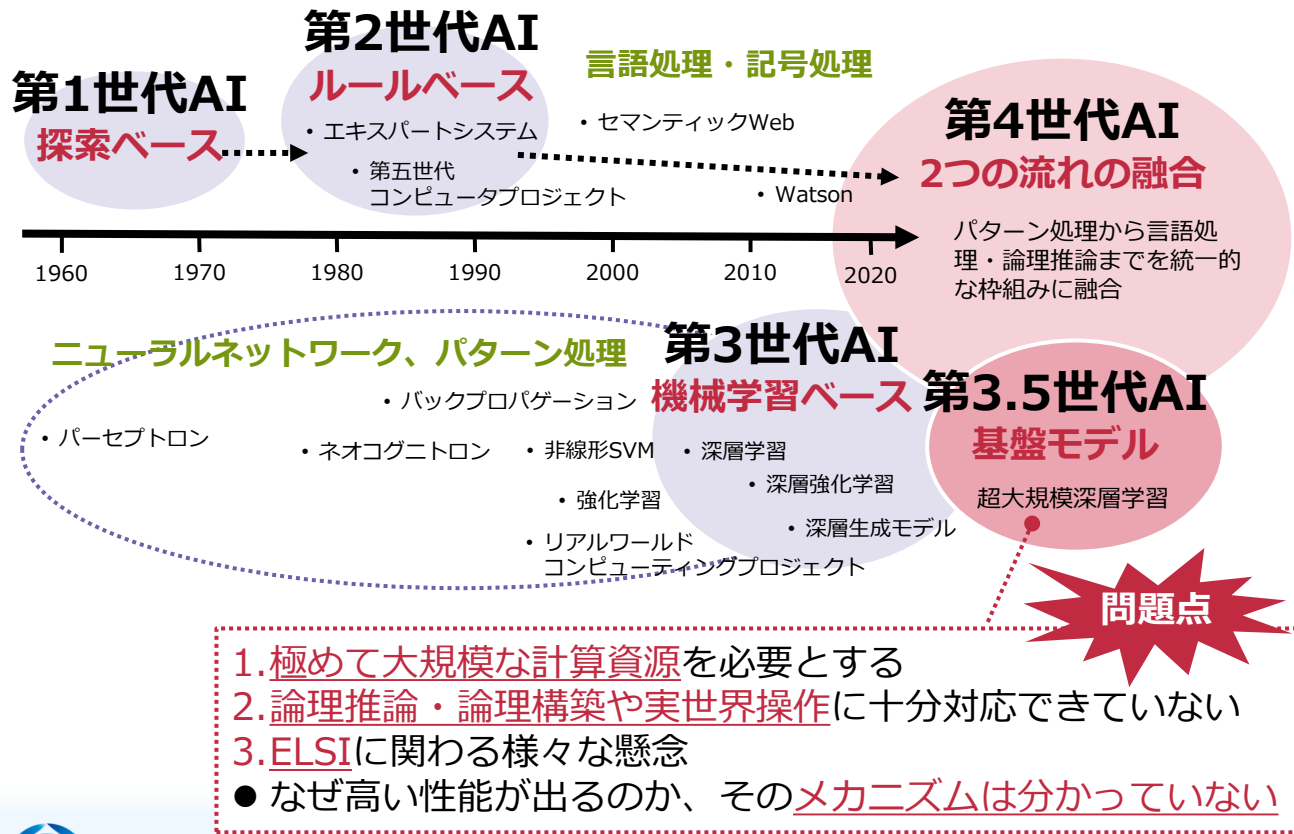



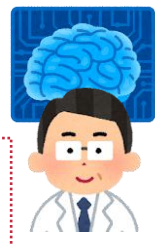
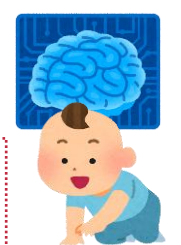
• 人々を思い通りに誘導・洗脳
• 思考停止させ、依存させる

次世代AIモデル研究

- 生成AI(基盤モデル)は人間離れした強力な道具だが、計算資源、論理推論、実世界操作等に課題
- 基盤モデルのメカニズム理解と人間知能の理解からの示唆が次世代AIモデルの可能性を示す

基盤モデル 応用開発 (API利用)	分野固有基盤モデル 開発・活用	
基盤モデル 周辺拡張 技術	利活用時の 問題対処 (フェイク/著作権等)	AIアライ メント研究 (人・社会との整合)
基盤モデル 運用技術	基盤モデル 構築技術	次世代AI モデル研究



	人間離れした道具	人間知能の理解からの示唆		
	基盤モデル	二重過程モデル	発達・創発モデル	...
中心的アイデア	深層ニューラルネットの大規模学習	即応的システム1と熟考的システム2で構成	外部との相互作用における予測誤差最小化	
関連研究分野	ニューラルネット、機械学習、計算機科学	心理学、行動経済学、脳科学	認知発達・記号創発ロボティクス	
人間にあてはめると	人が一生かかっても読めない量の情報を学習 	大人の思考・意思決定パターン  論理推論/論理構築	乳児・幼児の発達・成長過程  実世界操作/相互作用	

まとめ

(1)情報技術の3つのトレンド

- あらゆるもののデジタル化・コネクティッド化
- あらゆるもののスマート化・自律化
- 社会的要請との整合

(2)生成AIのインパクト

- まるで人間のような自然な対話、様々なタスクで専門的な知識・能力を備えたような応答
- 汎用性の高い強力な道具として応用が広がり、様々な知的活動を変革、産業成長の期待
- その一方で、計算資源、論理推論、実世界操作、ELSI面の懸念等が課題に
- 人間知能の理解から次世代AIモデルへの示唆も



研究開発の俯瞰報告書
システム・情報科学技術分野 (2023年)

[2023年3月発行]

<https://www.jst.go.jp/crds/report/CRDS-FY2022-FR-04.html>



人工知能研究の新潮流2
~基盤モデル・生成AIのインパクト~

[2023年7月末発行]

<https://www.jst.go.jp/crds/report/CRDS-FY2023-RR-02.html>

ご清聴ありがとうございました

最新の俯瞰報告書（2023年版）をご覧になりたい方はこちら



- 環境・エネルギー分野
- システム・情報科学技術分野
- ナノテクノロジー・材料分野
- ライフサイエンス・臨床医学分野
- 科学技術・イノベーション政策の国際動向
- 日本の科学技術・イノベーション政策の動向

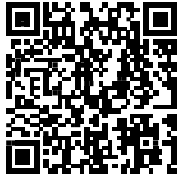
<https://www.jst.go.jp/crds/report/CRDS-FY2022-TOP.html>



日刊工業新聞連載 — 科学技術の潮流 —



2019年4月にスタートした
CRDSの日刊工業新聞連載
毎週金曜日に紙面に掲載、
1週間後にCRDS公式HPにて公開



<https://www.jst.go.jp/crds/column/choryu/index.html>



公式SNSで最新情報発信中



@CRDS_Japan

https://twitter.com/CRDS_Japan



@CRDSjapan

<https://www.facebook.com/CRDSjapan>

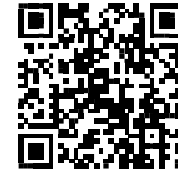


月刊メルマガ配信中

毎月15日（土日祝の場合は翌営業日）配信予定



登録はこちら



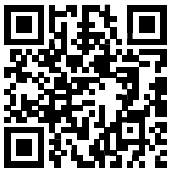
<https://www.jst.go.jp/melmaga.html#M01-06>



デイリーウォッチャー （海外記事検索）



科学技術に関連する、諸外国の
主要行政機関・研究機関の発表を
要約記事として日本語で提供



<https://crds.jst.go.jp/dw/>

国内外の研究開発動向やイノベーションについてご関心のある方は
お気軽にお問い合わせください



crds@jst.go.jp

■ 作成担当 ■

福島 俊一 フェロー（システム・情報科学技術ユニット）

TP20230825-5

その他報告書

**研究開発の俯瞰と潮流 ～科学技術イノベーションの動向と日本の活路～
「情報技術の3つのトレンドと生成AIのインパクト」**

令和5年8月 August 2023

ISBN 978-4-88890-875-7

国立研究開発法人科学技術振興機構 研究開発戦略センター

Center for Research and Development Strategy, Japan Science and Technology Agency

〒102-0076 東京都千代田区五番町7 K's五番町

電話 03-5214-7481

E-mail crds@jst.go.jp

<https://www.jst.go.jp/crds/>

本資料は著作権法等によって著作権が保護された著作物です。

著作権法で認められた場合を除き、本資料の全部又は一部を許可無く複写・複製することを禁じます。

引用を行う際は、必ず出典を記述願います。

なお、本報告書の参考文献としてインターネット上の情報が掲載されている場合には、本報告書の発行日の1ヶ月前の日付で入手しているものです。

上記日付以降後の情報の更新は行わないものとします。

This publication is protected by copyright law and international treaties.

No part of this publication may be copied or reproduced in any form or by any means without permission of JST, except to the extent permitted by applicable law.

Any quotations must be appropriately acknowledged.

If you wish to copy, reproduce, display or otherwise use this publication, please contact crds@jst.go.jp.

Please note that all web references in this report were last checked one month prior to publication.

CRDS is not responsible for any changes in content after this date.

Copyright © 2023 CRDS All Rights Reserved.

©2023 CRDS