



ESRI Research Note No.84

人工知能等の発展が 労働市場に及ぼす影響に関するサーベイ

黒木 裕鷹 久米 功一

March 2024



内閣府経済社会総合研究所
Economic and Social Research Institute
Cabinet Office
Tokyo, Japan

ESRI Research Note は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません（問い合わせ先：<https://form.cao.go.jp/esri/opinion-0002.html>）。

ESRI リサーチ・ノート・シリーズは、内閣府経済社会総合研究所内の議論の一端を公開するために取りまとめられた資料であり、学界、研究機関等の関係する方々から幅広くコメントを頂き、今後の研究に役立てることを意図して発表しております。

資料は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません。

The views expressed in “ESRI Research Note” are those of the authors and not those of the Economic and Social Research Institute, the Cabinet Office, or the Government of Japan.

人工知能等の発展が労働市場に及ぼす影響に関するサーベイ[◆]

黒木 裕鷹* 久米 功一**

要旨

人工知能等の発展が雇用にもたらす影響に関して、さまざまな議論が続いている。そのうち、本稿ではAI (Artificial Intelligence, 人工知能) exposureスコアに関連する先行研究を整理した。AI exposureスコアとは、AIへの暴露度（仕事の内容が技術要素と重複する度合い）の指標であり、AIによる仕事の代替・補完の可能性を評価する際に利用される。先行研究によると、AI exposureスコアの作成にあたり、タスクや技術に関する情報を数値化するための手法や、そのための詳細なデータベースが開発・活用されており、分析手法とデータが共進化していることが確認された。さらに、本稿では、この研究領域の今後の方向性や日本における発展可能性についても展望した。

◆ 本稿は、内閣府経済社会総合研究所（ESRI）科学技術ユニットのプロジェクト「デジタルエコノミーの進展が経済活動に与える影響」の成果の一部である。本稿の原案に対して、川口大司氏（東京大学）、小松恭子氏（労働政策研究・研修機構）、新谷元嗣氏（東京大学）、山本勲氏（慶應義塾大学）、村山裕氏（ESRI）、林田雅秀氏（ESRI）、信濃正範氏（ESRI）、篠崎敏明氏（ESRI）、宗像扶早子氏（ESRI）から有益なコメントをいただいた。記して感謝申し上げたい。なお、本稿の見解は筆者個人の責任で発表されるものであり、本稿における誤りのすべては筆者に帰するものである。

* 内閣府経済社会総合研究所研究協力者（Sansan株式会社）

** 内閣府経済社会総合研究所客員主任研究官（東洋大学経済学部）

目次

要旨	1
1. はじめに.....	3
2. 人工知能・機械学習技術の発展	5
2.1 機械学習	5
2.2 自然言語処理の発展と大規模言語モデル	6
3. データの開発と応用	9
3.1 職業とタスク (O*NET)	9
3.2 PIAAC調査	10
3.3 オルタナティブデータ	11
3.4 技術に関連するデータ (特許データなど)	12
3.5 クラウドソーシングの活用	13
4. 機械学習や人工知能が労働市場に与える影響についての先行研究	13
4.1 外観.....	13
4.2 情報技術の革新と職業の代替	14
4.3 AIexposureスコア(1) : 判断基準 (ルーブリック) の策定と人手のアノテーションによる算出	15
4.4 AIexposureスコア(2) : 自然言語処理による算出	17
4.5 AIexposureスコア(3) : 機械学習による算出.....	18
4.6 小括.....	19
4.7 AIexposureスコアのメタ分析	20
4.8 各国における実証分析	21
4.9 その他のテーマ.....	23
5. おわりに.....	26
5.1まとめと考察.....	26
5.2 日本における研究に向けて.....	28
参考文献.....	30

1. はじめに

この十数年間、人工知能 (artificial intelligence; AI) が人間の労働を代替するという議論が続けられてきた¹ (Frey and Osborne 2017; Brynjolfsson and McAfee 2011; 2014, Ford 2015)。その背景には、機械学習 (machine learning; ML) による労働市場への影響は、過去の技術革新による持続的な影響とは異なる性質を持つという認識がある。機械学習は、動作を規定するルールを人間が予め定めるのではなく、アルゴリズムがデータから自動獲得する点で従来の自動化技術と異なる。この特性により、これまで自動化の難しかった複雑な作業についても、入出力のデータから処理パターンの獲得にアプローチすることが可能となり、人間が得意とされていた予測・判断といった認識業務にまで自動化の領域が拡大している (Agrawal et al. 2018)^{2,3}。

この変化を踏まえて、人工知能等の発展が労働市場に及ぼす影響に関する研究が盛んに行われている。労働者の作業を代替し大量解雇や失業を引き起こす可能性があるか (Acemoglu and Autor, 2011; Frey and Osborne, 2017; Brynjolfsson et al., 2018; Agrawal et al., 2019; Webb, 2020; Felten et al., 2021)、新たな雇用を創出するか (Damioli et al., 2022; Yang, 2022)、労働者を補完し知的生産性を向上させるか (Noy and Zhang, 2023; Dell'Accia et al., 2023) などのテーマがあり、複数の議論が未だ途上で一貫した結論は出ていない。より最近では、従来よりも遥かに多い制御パラメータで膨大なデータセットを学習した言語モデルである大規模言語モデル (Large Language Models; LLMs) が雇用にもたらす影響についても分析対象となっている (Brynjolfsson et al., 2023; Eloundou et al., 2023; Felten et al., 2023)。

これらの研究には、大きく二つのポイントがある。一つは、人間の仕事をタスク・ベースで計測し指標化している点である。一つのジョブ (仕事) はさまざまなタスク (業

¹ 人工知能とは、人工的に知能を実現する技術を指す。本稿では、機械学習や自然言語処理などの人工知能の要素技術や、ソフトウェアやロボット等に応用・実装された人工知能の技術も、広義の意味で「人工知能等」という言葉で表している。また、機械学習は、人工知能の技術要素の一つであり、特定の事象に対してデータを解析して学習し、判断や予測を行うアルゴリズムとそれをを用いた技術である。その一つであるディープラーニング (深層学習) は、ニューラルネットワークを層化して特徴量を抽出する技術である。これを用いた自然言語処理の技術であるChatGPTは、ユーザーの指示 (プロンプトPrompt) に従って学習済のデータからコンテンツを生成するため、生成系AIと呼ばれる。

² AIの用途で分けると、識別系 (文字や画像、音声などから情報や物事を認識して特定する)、会話系 (チャットボットなどでの入力内容に対してアルゴリズムを介して適切な回答を出力する)、予測系 (過去データから未来予測や異常の検出を行う)、実行系 (機械の操作や制御を行う)、生成系 (創造的なコンテンツを新規に生成する) などがある。

³ AIの社会実装による影響に関する具体例については、本稿第3章を参照されたい。

務) からなり、タスクごとに技術の応用可能性が異なる⁴。このため、技術による代替・補完の可能性を検討する際には、タスクを定義し、測定して指標化する必要がある。労働者のタスクとAIの関連性(=exposure, 暴露度)をスコアリングし、職業別に集約することによって、どの程度代替の可能性に曝されているかを予測する手法の開発が進んでいる(Felten et al. 2021ほか)。

もう一つは、タスクの性質やAIとの関連性を指標化するためのデータがさまざまに開発・活用されている点である。例えば、職業をいくつかのタスクに分解して詳述しているO*NET (Occupation Information Network) やAI関連の技術を記載した特許データベースなどがあり、タスクに対する技術の応用可能性が検討されている(Felten et al. 2019; Webb 2020など)。テキスト情報を含むデータベースに対しては、自然言語処理の手法を用いた分析がなされている(Webb 2020)。

以上の二点から、人工知能等の発展が労働市場に及ぼす影響を把握するために、機械学習・自然言語処理などの分析手法の応用とO*NETや特許データベースのようなデータセットの開発・活用が同時に起きていることがわかる、つまり、この研究領域では、分析手法とデータの共進化によってさまざまな研究が蓄積されているのである。

このように人工知能AI/機械学習MLの発展は(1)労働を代替・補完する意味でも、(2)労働の代替・補完の尺度であるexposureスコアを計算する手法においても、また、その手法の利用に適した(3)データの開発・応用を促すという意味においても、さまざまな影響をもたらしている。そこで、本稿では、図1の通り、第2章で議論の全般にかかわる人工知能・機械学習技術の発展を整理し、第3章では、第2章で述べた技術の応用とともに開発・利用されているデータ等を紹介する。それらを踏まえて、第4章で労働の代替・補完の尺度であるexposureスコアを中心に、人工知能等と労働市場の関係を分析した先行研究について整理する。最後に、第5章で今後の研究の展望と日本における研究の可能性を考察する。

以上の通り、本稿は、人工知能等による人間の労働の代替・補完可能性に関する先行研究の成果を分析手法とデータの共進化の視点から整理して、技術による労働の代替・補完の問題に対して、機械学習・自然言語処理の手法を用いることの利点や限界、今後の研究の方向性を展望することを目的とする。AIが労働市場に与える影響に関する国内外の研究動向を整理した北原(2018)の続編にあたる。

⁴ 労働の総量が社会全体で一定であると仮定して、ある雇用を増やすことが他の雇用を奪うと主張したり、タスクの一部が自動化・機械化されることをもってその仕事全体が失われると解釈したりすることを労働塊の誤謬(lump of labour fallacy)という。

図1. 労働研究における技術とデータの共進化と本稿の構成



2. 人工知能・機械学習技術の発展

はじめに分析手法の発展として、機械学習と自然言語処理を取り上げる。とりわけ、自然言語処理は、その社会実装による労働市場への影響だけでなく、次章で紹介する先行研究における分析手法としても活用されていることから、本節でも焦点を当てる。続いて、データの開発として、O*NET (Occupation Information Network) やOECD (Organisation for Economic Cooperation and Development, 経済協力開発機構) のPIAAC (the Programme for the International Assessment of Adult Competencies, 国際成人力調査) などを取り上げる。機械学習技術の発展が労働市場への影響を考慮する上で、これらのデータがタスクなどの概念や研究手法の発展とともに開発・活用されてきた点を強調する。

2.1 機械学習

機械学習とは、与えられたデータから法則性を自動で獲得し、その結果をもとに分類や予測などを行う一連の技術体系であり、人工知能の重要な要素技術となっている。機械学習技術の発展では、入出力データから推論パターンを自動獲得するアルゴリズムの開発と、大規模な演算を可能にする計算資源の発展が中心的な役割を果たしている。機械学習の適用範囲は広く、基本的な表データの他に、画像、自然言語、信号（時系列）、グラフ（ネットワーク）といったデータの構造を明示的に考慮するアルゴリズムの開発も盛んであり、産業での応用も進んでいる。特に、自然言語処理（natural language processing; NLP）とコンピュータビジョン（computer vision; CV）の領域では、大規

模事前学習モデルの発展が目覚ましく、多くの革新的な成果が生み出されている。NLP分野ではChatGPT⁵ (OpenAI, 2023) やGemini (Gemini Team, 2024) , Claude 3⁶, CV分野ではDALL-Eシリーズ (Ramesh et al., 2022) やMidjourney⁷, Stable Diffusion (Rombach et al., 2022) といった生成モデルの活躍が著しく、複雑で高品質な出力を行うことができる。ChatGPTやGemini, Claude3は、チャット形式のインターフェースをもつアプリケーションとして提供されていることから、機械学習やプログラミングに精通していなくとも、誰もが簡単に馴染みのある自然言語で利用することが可能であり、広範な職業においてタスク遂行に利用する行動変容が起きている可能性がある。また、ウェブAPI (Application Programming Interface) などを利用して、これら大規模言語モデルを要素技術として組み込んだシステムやサービスの開発も進んでいる。

こうした背景をもとに、機械学習は自動運転、医療補助、言語翻訳など、多様な分野での応用が進んでおり、人間の思考や判断を模倣するシステムの開発において、ますます重要な役割を担うようになってきている。

2.2 自然言語処理の発展と大規模言語モデル

機械学習の扱うデータ形式のうち自然言語については、前述の通りあらゆる労働者にとって身近で不可欠なものである。ChatGPTをはじめとしたアクセスしやすい高性能な言語モデルが登場したことは、労働市場に少なからず影響を与えると予想される。また、exposureスコアを提案するいくつかの先行研究においても、特許テキストや職業タスクの定義という形で扱われるほか、自然言語処理の技術をexposureスコア算出のツールとして利用する取り組みも存在する。そこで本節では、自然言語処理 (NLP) の概要とその発展について概説する。

NLP の黎明期では、ルールベースのアプローチが盛んに開発された。それらの手法は、言語の構造を分析し理解するために設計された手作りのルールに依存するものである。そのような初期のシステムにELIZA (Weizenbaum, 1966) やSHRDLU (Winograd, 1972) がある。ELIZAはセラピストとの会話をシミュレートするルールベースのチャットボットであり、パターンマッチングとキーワード認識を用いてユーザー入力に対する応答を生成した。また、SHRDLUはユーザーからの自然言語命令を理解し、仮想ブロック世界内で行動を実行する言語理解プログラムである。これらのシステムはNLPにおけるルールベースのシステムの可能性を示したが、一方でスケーラビリティや堅牢性の点で大きな制

⁵ <https://openai.com/blog/chatgpt>

⁶ <https://www.anthropic.com/news/claude-3-family>

⁷ <https://www.midjourney.com/home>

限に直面した。事前に策定されるルールは作成が困難である上に、自然言語の複雑さや変動性を網羅的に捉えきれない。その結果、統計モデルや機械学習技術など、新しいアプローチが模索されはじめた。

NLPにおける最初期の統計モデルの一つは、隠れマルコフモデル (HMM; Hidden Markov Model) である。HMMは、品詞や離散的な状態などの未知のカテゴリの系列をモデル化するために使用され、言語モデリングでは文内の前の単語から次の単語の確率を予測するために使用された。統計モデリングにおけるもう一つの初期の進歩は、n-gram言語モデルの開発である。n-gramはn個の連続するトークン (言語の分析単位) の系列であり、n-gramを用いた言語モデリングでは、前のn-1個の単語を考慮して単語の確率を予測する。これらのモデルは、機械翻訳などの多くのNLPタスクで使用されたシンプルで効果的な統計モデルである。

1990年代後半には、サポートベクターマシンやニューラルネットワークなど、NLPに限らない有力な機械学習アルゴリズムが開発された。自然言語のような系列データをニューラルネットワークで扱うために、LSTM (Long Short-Term Memory) などの再帰的構造を持つアーキテクチャが開発されたが、当時の計算資源では高次元かつ膨大な言語を扱うには計算資源が乏しく、NLPを推し進めるには至らなかった。一方で進められたのが、LDA (Latent Dirichlet Allocation, Blei et al., 2003) をはじめとする潜在的な意味を統計的に獲得するための解析である。NLPの分野における重要な進歩の一つは、2013年に単語の分散表現を学習するニューラルネットワークベースの言語モデル、Word2vec (Mikolov et al., 2013) の開発である。Word2vecは、周辺の単語との関係 (共起) を利用し、トークンの相対的な (意味) 関係のベクトル表現を学習するモデルである。このアプローチは、分類や回帰といった後段のタスクへの活用や文章単位の表現ベクトルの獲得などに発展し、手作りの機能やルールに依存する従来のNLPメソッドからの脱却と共に、多くのNLPタスクに大きな改善をもたらした。

近年では、膨大な言語データの蓄積と計算資源の発展を背景として、大きく深いニューラルネットワークである深層学習の活躍が目覚ましい。特に、アテンションメカニズムとそれを利用したTransformerアーキテクチャ (Vaswani et al., 2017) の開発により、モデルは予測を行う際に入力の特定の部分に選択的に焦点を当てることができるようになった。また、Transformerと共に後続の大規模言語モデル (LLMs) の発展を支えたのが、事前学習とfine-tuningである。fine-tuningは、事前学習モデルのパラメータを、より少数のデータで再更新する方法である。事前に広範な言語コーパスで学習された表現を活用して、他の方法で必要となるデータよりもはるかに少ないデータで新しいタスクに適用できる。NLPにおいて最も普及した事前学習モデルが、BERT (Devlin et

al., 2018) やGPTシリーズ (Brown et al., 2020; Radford et al., 2018; Radford et al., 2019; OpenAI, 2023) などがある。

GPTシリーズではGPT-1からGPT-4に至るまで、学習データとモデルサイズは大規模化の一途をたどっている。GPT-1で1.17億であったパラメータ数はGPT-2では15億、GPT-3では1,750億、GPT-3.5では3,550億となっている (GPT-4は非公開)。GPT-2以降は生成型のモデルであり、前のトークン列から次のトークンを予測するような形式になっている。この形式はChatGPTで利用されているGPT-3.5やGPT-4まで一貫しており、使用者は自然言語で指示となる文章 (プロンプト) を与えることで、それに続くトークン列を出力として得ることができる。その性能の向上、モデルの大規模化とともに、少数のデータでパラメータを再更新するfine-tuningではなく、プロンプトの中でさらに少ない正解例を与えるfew-shotプロンプティングが行われるようになっている。

こうしたNLPの発展は英語だけでなく日本語にも同様に適用することが可能であり、日本語に特化したLLMsも多数開発されている⁸。しかし、大量の計算機資源や予算を用意することの困難さに加えて、いくつか日本語特有の困難さが広く認識されている。一つの明確な違いは、英語のような多くの言語が単一の文字体系で構成されスペースによって単語を区切っているのに対し、日本語はひらがな、カタカナ、漢字という様々な文字体系が混在しており、スペースを用いて単語を区切ることが一般的ではないことである。これにより、英語では単語やトークンを容易に識別できるのに対し、日本語ではテキストを単語や意味の単位に分割するために形態素解析が必要となる。しかし、この形態素解析の過程は使用される形態素解析器⁹に大きく依存し、用いる辞書も固有表現などを追加する必要がある。加えて、日本語には同義語が豊富に存在し、さらに口語表現においては主語や目的語の省略が頻繁に見られることから、文脈や潜在的な意味を機械的に理解する上での難易度が高いことが挙げられる。Exposureスコアの算出に日本語テキストデータを用いる場合には、先行研究のアプローチをそのまま実施するのではなく、これらの課題に十分配慮した処理を綿密に設計する必要がある。

⁸ 日本語LLMsの多くは<https://github.com/llm-jp/awesome-japanese-llm>にまとまっている。

⁹ 代表的日本語の形態素解析器にはMeCab (Kudo, 2005) やSudachi (Takaoka et al., 2018) などがある

3 データの開発と応用

3.1 職業とタスク (O*NET)

国際労働機関 (ILO; international labour organization) は、職業に関する国際基準として ISCO (International Standard Classification of Occupations) を策定しており、これを通じて各国の職業別統計を突合せ比較、分析することができる。最も新しいバージョンである ISCO-08¹⁰での職業 (occupation) とは、主に類似性によって特徴付けられる職務 (job) の集合と定義されている。職務は、雇用主や自営業を含め一人で実行されることを意図されているタスク (task) と義務 (duty) のセットとされている。また、与えられた仕事のタスクや義務を遂行する能力としてスキル (skill) についても整備されている。スキルには職業の複雑さや範囲に基づくスキルレベル (skill level) と、職業内の特定の技能や知識に基づくスキル専門性 (skill specialization) という2つの次元があり、これらを用いて職業がグループ分けされる。日本においては、日本標準職業分類¹¹が定められている。

一方、米国では職業紹介用の分類として、細分類レベルの職業を補正したO*NET¹²が整備されている。O*NETは、職業ごとのさまざまな属性を提供しており、能力、興味、知識スキル (基本的なものとかロスファンクショナルなもの)、仕事の活動、仕事の文脈、仕事のスタイル、仕事の価値観などがテキストまたは数値情報として含まれる。各職業が行う具体的なタスクや作業により重点が置かれ、それらを職業の特徴として分類しているデータベースである。ISCOやO*NETのような、職業について分類、記述した公的なデータセットは、技術革新が労働市場に及ぼす影響を分析する上でも積極的に活用されている。日本においても、厚生労働省によって「jobtag職業情報提供サイト (日本版O-NET)¹³」が運営されている。約500の職種について、ジョブ (職業、仕事)、タスク (仕事の内容を細かく分解したもの、作業)、スキル (仕事をするのに必要な技術・技能) が紹介されている (図2)。小松・麦山 (2021) は、日本版O-NETの数値情報と国勢調査の集計データ (1990年~2015年) を用いてタスクの分布の現状及びそのトレンドを分析している。また、労働政策研究・研修機構 (近刊a) はタスクの分布とトレンドに関する日米比較を行い、具体的な職種レベルで考察することにより、日本の労働市場 (タスク分布) の特徴を明らかにしている。

¹⁰ <https://www.ilo.org/public/english/bureau/stat/isco/isco08/>

¹¹ https://www.soumu.go.jp/toukei_toukatsu/index/seido/shokgyou/index.htm

¹² <https://www.onetonline.org/>

¹³ <https://shigoto.mhlw.go.jp/>

図2. jobtag (日本版O-NET) の一例



出所) 職業情報提供サイト jobtag 飲食チェーン店員 職業詳細

<https://shigoto.mhlw.go.jp/User/Occupation/Detail/418>

3.2 PIAAC調査

O*NETは職業単位でタスクを把握するものであるのに対して、労働者単位でタスクを把握するものがある。具体的には、OECDのPIAAC調査 (De La Rica et al. 2020¹⁴, 表1) やプリンストン大学による PDII 調査 (Princeton Data Improvement Initiative survey, Autor and Handel, 2013) である。これらのデータから、各労働者が従事する仕事について、ルーティンタスク、抽象タスク、マニュアルタスクの3つの指標が得られ、定型業務集約度 (Routine Task Intensity, RTI) を試算することができる。Kobayashi and Yamamoto (2020) は、厚生労働省「賃金構造基本統計調査」の個票データとPIAAC調査の設問を取り入れた独自調査を接合して分析し、抽象タスクと賃金との間に正の相関があり、抽象タスクのスコアが1標準偏差上昇すると21.2%の賃金プレミアムが生じる一方、ルーティンタスクとマニュアルタスクによる賃金プレミアムは負であることを明らかにしている。Arntz et al. (2016) はPIAAC調査のスキル、タスク、コンピテンシーなどのデータから自動化リスクを計算している。Nedelkoska and Quintini (2018) は、Frey and Osborne (2013) をもとに、PIAAC調査に基づく職種別の自動化リスクを推計している。このほかにも、German Qualification and Career Survey

¹⁴ PIAACでは、スキルの保有と利用について、認知スキル (読解力と数的思考力) とスキル利用に分けて把握している。例えば、Kawaguchi and Toriyabe 2022は、項目反応理論を用いて認知スキルとスキル利用を指標化し、これらが男女間の賃金格差、移民とネイティブ間の格差を説明できることを示している。

(ドイツ資格キャリア調査) の設問を援用して、転職前後のタスクの距離と賃金変化の関係を分析したGathmann and Schönberg (2010), 勇上・牧坂 (2021) がある¹⁵。

表1. PIAAC調査の分析例

Task framework with PIAAC data.		
Task	PIAAC questionnaire item	item no.
Abstract	Face complex problems(<30 mins)	F_Q05b
	Use more advanced math or statistics such as calculus, complex algebra, trigonometry, or use regression techniques	G_Q03h
	Read articles in professional journals or scholarly publications	G_Q01d
	Planning the activities of others	F_Q03b
Routine	Persuading/influencing people	F_Q04a
	Planning your own activities(inverse)	F_Q03a
	Organising your own time(inverse)	F_Q03c
	Instructing, training or teaching people, individually or in groups(inverse)	F_Q02b
	Making speeches or giving presentations (inverse)	F_Q02c
Manual	Advising people(inverse)	F_Q2e
	Working physically for a long period	F_Q06b
	Using skill or accuracy with hands or fingers	F_Q06c

Note: To ensure the reliability of the statistical constructs, all questions provide the same time answers: (i) every day; (ii) at least once a week but not every day; (iii) less than once a week; (iv) less than once a month; (v) never.

出所) De La Rica et al. 2020 table 1.

3.3 オルタナティブデータ

公的に整備されたデータセットのほかに、オルタナティブデータの整備と活用が進んでおり、従来では難しかった需要側からの側面や、より詳細なスキル要件などについての洞察を得ることが可能になっている¹⁶。Burning Glass Technologies (BGT) は、ウェブ上に掲載された求人データをウェブクローリングとスクレイピングによって収集し、労働とスキルの需要に関するより詳細な情報をデータセットとして整備している。BGTデータは、米国における不況期 (2007年から2015年) におけるスキル要件の変化に関する研究 (Hershbein and Kahn 2018) や求人広告におけるスキル要件と給与や企業業績との関係の研究 (Deming and Kahn 2018) , STEM (科学, 技術, 工学, 数学) 職の賃金, 技術スキル, スキルの陳腐化の関係の研究 (Deming and Noray 2020) で使用されており, 米国労働統計局の求人データ (JOLTS, Job Openings and Labor Market Turnover, Davis, Faberman, and Haltiwanger 2012, 2013) を補完する役割を果たしている。

OECDのレポート (Cammeraat and Squicciarini, 2021) によると, BGTデータの収集は4万件以上の求人サイトや企業ページから行われ, オーストラリア, カナダ, ニュージーランド, シンガポール, 英国, 米国を対象とし, 米国では2010年から, その他の国

¹⁵ 日本版O-NETのタスク指標でホワイトカラー転職者の転職前後のタスク距離と賃金変化の関係を分析した研究として小松 (2022) がある。

¹⁶ 労働分野におけるオルタナティブデータの活用については, 森脇 (2022) を参照されたい。

では2012年からの期間をカバーしている。また、複数のウェブサイトに掲載された重複する求人を削除した後、テキスト解析によって地理情報や職種、業種、必要なスキル、教育レベル、系意見レベルなどの情報が70以上の変数として構造化されている。Cammaraat and Squicciarini (2021) は、BGTデータについて第一次産業のカバレッジが高くはないことなどサンプルの偏りを指摘しているものの、マネージャやプロフェッショナル、技術者や専門家のカテゴリについては比較的良いサンプルとなっていることを明らかにしている。特定の職種でスキル要件がどのように変化しているか、業界によってスキル要件がどのように異なるか、どのスキルが最も急速に成長しているか、どのスキルが最も雇用者に価値あるかなど、より詳細な分析を行うことが可能であり、技術革新と労働市場の変化を分析する上でも重要なデータソースとなる。たとえば Acemoglu et al. (2022) は特に2015年から2016年ごろにかけてAI関連の求人が増加していることを明らかにし、後述するexposureスコアと需要の関連を分析している。また Alekseeva et al. (2021) は同じ企業内でAI関連スキルを要求する求人では11%、同じ役職内では5%の賃金プレミアムをもつことを明らかにしている。BGTデータの収集・整備についてのより詳細なプロセスは Carnevale et al. (2014) を参照されたい。

3.4 技術に関連するデータ（特許データなど）

タスクが技術的に代替・補完可能であるかを検討するためには、タスクにおける技術浸透度を把握する必要がある。Felten et al. (2019) は、電子フロンティア財団 (Electric Frontier Foundation) のAI進捗測定プロジェクトの調査データ (EEF AI Progress Measurement) を用いている。また、自然言語処理を用いて、特許と職業記述の類似性スコアを算出する研究がいくつか存在する (Webb 2020, Kogan et al. 2021, Google Patents Public Data database, Meindl et al. 2021 the PATSTAT patent database)。タスクの説明と特許抄録のテキストを比較してタスクごとに関連性の高い特許を特定するものである。例えばWebb (2020) は、Google Patents Public Data databaseとO*NETから人工知能、ソフトウェア、産業用ロボットの特許とタスクを紐づけている。Autor et al. (2022), Mann and Püttmann (2023) は、技術によるタスク自動化を特許の登録数に代理し、その雇用への効果を分析している。

3.5 クラウドソーシングの活用

次章で詳述するexposureスコアの一つであるFelten et al. (2021) のAI Occupational Exposure (AIOE) は、電子フロンティア財団 (EFF) のAI進捗測定プロジェクトから収集した10のAIの応用について、Amazonのメカニカルターカー200人にアンケートを実施し、O*NETデータベースにある52の能力のそれぞれに関連しているかどうかの意見を聞いている。Amazon Mechanical Turk¹⁷ (アマゾンメカニカルターク) は、アマゾンウェブサービスの一つで、クラウドワーカーがタスク (HITs; Human Intelligence Tasks, 人間の知能が必要なタスク) を実行するものである。機械学習の正解データの作成などに活用されるほか、利便性の高さから社会科学分野での利用が増えており、個人の特性に一貫性がみられるという意見 (Johnson and Ryan 2020) もあれば、このプラットフォームに対する懐疑的な見方も存在する (Aguinis et al. 2021; Brodeur et al. 2022)。このほかにも、Brynjolfsson et al. (2018) は、O*NETデータベース (米国労働省が編集) で定義されている詳細な作業活動 (DWA, Detailed Work Activity) ¹⁸がどの程度機械学習に適しているかを推定するための23の質問を作成し、Crowdfunder (現Appen) というクラウドソーシングプラットフォームを使用して、各DWAの分野に詳しい7~10人のクラウドワーカーに採点させている。これをタスクレベル、さらにO*NETの重要度で重み付けされた米国の職業レベルへとExposureスコアを集計している。このように、一部の研究では、スコア作成という重要な場面でクラウドワーカーを活用している点もこの研究領域の新規性である^{19, 20}。

4. 機械学習や人工知能が労働市場に与える影響についての先行研究

4.1 概観

人工知能は消費者向けから産業用途まで、あらゆる分野で革新的な進歩をもたらしており、第四次産業革命の中核となっている。様々な産業と現場における人工知能の活用は、これまでの技術革新と同様に経済に多様な方法で影響を与え、労働市場の構造を大きく変える可能性がある。既存の労働を代替していると想定される一方で、機械学習エ

¹⁷ <https://www.mturk.com/>

¹⁸ Work Activityは一般的な記述からより特異的な記述に向かって、Generalized Work Activity (GWA), Intermediate Work Activity (IWA), Detailed Work Activity (DWA) の3段階に分類される。

¹⁹ インターネットを利用した経済実験については森 (2019) を参照されたい。

²⁰ さらに言えば、脚注29で言及するようにLLMsを用いてデータを生成する研究も増えている。

エンジニアやMLOpsエンジニア²¹といった新しい職種も生まれてきている。こうした背景を踏まえ、人工知能が労働市場に及ぼす影響はさまざまな先行研究によって論じられてきた。

これまでの研究では、主にタスク指向のアプローチが採用され、人工知能の能力と労働者のタスクとの類似性を評価するために様々な手法が提案されてきた。例えば、労働タスクの説明をAI関連の特許の説明とマッピングする手法 (Kogan et al., 2021; Webb, 2020) や、労働タスクが機械学習に適しているかのスコアリング (Brynjolfsson and Mitchell, 2017; Brynjolfsson et al., 2018) などである。次節では、まず人工知能以前の技術革新でどのような議論が行われてきたかを整理し、続く3節以降では、人工知能を対象とした先行研究について、主要なものを列挙し、そのアプローチをいくつかのパターンに整理する。

4.2 情報技術の革新と職業の代替²²

情報技術の革新が労働市場に与えた影響を考察する研究では、賃金格差や労働市場の二極化が重要なテーマとなっている。この文脈で、スキル偏向技術進歩の仮説が提示されている²³。この仮説は、ICTなどの技術進歩が高スキル労働者への需要を増加させるというものであり、賃金格差の拡大を高スキルと低スキル労働者間の需要の変化によって説明する。しかし、1990年代以降の労働市場の二極化、特に中スキル労働者の雇用減少については、この仮説では十分に説明できていないとされる²⁴。

スキル偏向技術進歩仮説に対する代替として、タスクアプローチが提案された。Autor et al. (2003), Acemoglu and Autor (2011) によって発展し、生産過程を構成するタスク (職務) が、ICTなどの技術進歩によって代替される可能性に注目している。タスクは定型的 (routine) と非定型的 (non-routine) に分けられ、さらに定型的タスクは手作業的 (manual) か認知的 (cognitive) に、非定型的タスクは分析的 (analytical), 相互的 (interactive), 手作業的 (manual) に分類される。ICTやロボットなどによる自動化技術は、特に中スキル労働者が担う定型的な職務に影響を及ぼし、労働市場の二極化を進めると解釈される。この視点から、人工知能以前の技術革新につい

²¹ Machine Learning Operationsの略語で、機械学習 (Machine Learning) と運用 (Operations) をかけた用語である。

²² より詳細な説明は北原 (2018) を参照されたい。

²³ Bound & Johnson (1992), Krueger (1993), Berman et al. (1994), Levy and Murnane (1996), Machin & Van Reenen (1998), 櫻井 (1999), 小原・大竹 (2001) などがある。

²⁴ Autor et al. (2003), Goos and Manning (2007), Spitz-Oener (2006), Goos et al. (2009) などがある。

ては、労働市場にもたらすタスクの変化に関して実証研究が進められてきた²⁵。Acemoglu and Restrepo (2018) は、タスクの複雑さに伴い労働の資本に対する比較優位性が増加するという条件のもと、労働者と資本があり、技術には自動化と新たなタスク創出の2通りがあるとき、技術制約>コスト制約の場合の自動化は資本分配率を高める一方、新しいタスクの創出の場合は必ず労働分配率が高まることを示した。そこでは、自動化による置換効果 (Displacement effect) だけでなく、生産性効果 (Productivity effect) や再配置・復職効果 (Reinstatement effect) が考慮されている。

4.3 AI exposureスコア (1) : 判断基準 (ルーブリック) の策定と人手のアノテーションによる算出

どのようなタスクや作業がAI/MLにより代用可能か、もしくは相性が良いか、その判断基準となるルーブリックを定性的議論から定め、それに基づいてexposureをスコアリングするアプローチが存在する。ここでいうルーブリックとは、AI技術の適用可能性に関する評価基準を、評価の観点と尺度からなる表として示したものをいう²⁶。機械学習に適しているかの判断は通常複雑なものになるが、そのプロセスを細分し明文化することで、柔軟にスコアリングを実施することができる。一方、その柔軟さのためにルーブリックを厳密に遵守しながらのスコアリングはコストが高くなることが想定される。クラウドソーシングを活用する取り組みもあるが (Brynjolfsson et al. 2018, Felten et al. 2019) , 作業者によって機械学習や対象となる職業やタスクへの精通の程度が異なるため、精度の担保のためにも作業工程の綿密な設計に注意が必要である。また、最近ではChatGPTのような高性能な言語モデルにルーブリックとタスクのテキストを与え、スコアリングを行うような取り組みも存在するが (Eloundou et al. 2023) , その精度評価や推論の偏りにも注意が必要である。以下に、代表的なexposureスコアの研究例であるFelten et al. (2018), Felten et al. (2019) , Felten et al. (2021) を説明した後、関連する研究を示す。

²⁵ 米国 Autor et al. 2006, 英国 Goos and Manning, 2007, ドイツ Spitz-Oener 2006; Dustmann et al. 2009, 北欧 Asplund and Barth 2011, OECD諸国 Michaels et al. 2013, 欧州 Goos et al. 2009, 2014, 日本 池永 2009, 2011, Ikenaga and Kambayashi 2016, サーベイ論文として, Montobbio et al. 2023; Mondolo 2022; Lu and Zhou 2021がある。

²⁶ 例えば, Brynjolfsson and Mitchell (2017) のルーブリックは, 以下を参照されたい。
https://www.science.org/doi/suppl/10.1126/science.aap8062/suppl_file/aap8062-brynjolfsson-sm.pdf

Feltenら : Feltenらは、AIの主要な応用タスクとO*NETのabilitiesの関連度合いを人手でスコアリング、職業ごとに集計することでAI Occupational Exposure (AIOE) を算出している (Felten et al., 2018; Felten et al., 2019; Felten et al., 2021)²⁷。AIの主要な応用タスクとしては、Electronic Frontier Foundation (EFF) の AI progress measurement datasetを集約することで特定している。また、Felten et al. (2021) では職業だけでなく業界や地域単位で集計することによるAI Industry Exposure (AIIE) やAI Geometric Exposure (AIGE) も合わせて提案している。AIOEの特徴としては、クラウドソーシングのプラットフォームであるAmazon Mechanical Turkを利用し、人手のスコアリングでありながら、一定の客観性をもたせていることにある。また、近年ではChatGPTをはじめとする大規模な言語モデルの登場を受け、自然言語処理の応用タスクにより重みをおいた Language Model AIOE (LM AIOE) を提案している (Felten et al., 2023)。

その他の研究 : Manyika et al. (2017) は、米国O*NETの800以上の職種、2000以上の仕事に対して、技術による自動化の潜在的な可能性を求めている。それぞれの仕事に対して、18の能力 (capacities, 主に5つに分けられる : sensory perception : 感覚・知覚, cognitive capabilities : 認識能力, natural language processing : 自然言語処理, social and emotional capabilities : 社会的情動能力, physical capabilities : 身体的能力) が適用可能かについて、学術研究、社内や業界の専門家が4段階で評価してスコア化したものを45か国に展開し、世界の労働市場の80%をカバーしている。

Brynjolfsson and Mitchell (2017) は、様々な分野における実例をもとに、MLに適したタスクかどうかのルーブリックSuitability for Machine Learning (SML) を提唱している (脚注27を参照)。また、Brynjolfsson et al. (2018) はこれらの基準をもとにO*NETの各タスクを専門家の手によってスコアリングし、職業によってスコアの傾向が異なることを示している。職業によっては、SMLが大きく異なる複数のタスクから構成されており、そのような職業ではMLの能力をより活用するために職務の再設計や再編成の機会があることを示唆している。

Eloundou et al. (2023) は、「そのタスクを完遂するのに要する時間が、少なくとも50%以上削減される」ことをexposureスコアとして定義し、LLMsの技術に特化したスコアを算出している。スコアの算出は人手のアノテーションによる推論とGPT-4による

²⁷ Felten et al. (2021) は、Felten et al. (2019) で開発したArtificial Intelligence Occupational Impact (AIOI) 指標について、職業レベルでの曝露をより適切に測定するために、10のAI技術の応用を考慮してAI Occupational Exposure (AIOE) と呼ばれる指標に拡大・改良したものである。

推論によって行い、両者の比較やそれぞれについての分析を行っている²⁸。結果として、80%の職業でそのタスクの10%が曝露されており、平均賃金や教育レベルの高い職業ほど exposure スコアの値が高いことを報告している。

Tolan et al. (2021) は、AIの主要タスクと労働タスクを直接結びつけるのではなく、認知能力という中間次元をもたせることでより柔軟な理解を可能にしている。exposure スコアの算出はシンプルな行列演算によって行われ、職業にとってのタスクの強度を要素にもつベクトル、認知能力の強度ベクトル、AIの主要タスクの強度ベクトルが、それらの関係を表す係数行列で結びつけられる。ここで、係数行列は、人手のアノテーションによって与えられる。得られた係数行列より、あるベンチマークが進捗した場合にどの労働タスクが影響を受けるかを分析することができるほか、注目する職業とその一連のタスクへの潜在的な影響について、どのようなベンチマークを注視すべきかを測る逆向きの分析も可能となる。

4.4 AI exposure スコア (2) : 自然言語処理による算出

0*NETの職業テキストやタスクテキスト、特許のテキストなどを活用し、自然言語処理により exposure のスコアリングを行う手法が提案されてきている。入力には職業やタスクを表すテキストと、MLの技術を表すテキストの対であり、両者の関連性をMLを含む自然言語処理で定量化するようなアプローチである。機械的な処理によってスコアが付与されるため、誰もが同様の手順を実施すれば同じ結果を再現することが可能である。一方で、入力となるテキストが職業やMLの内容を完全に表しているとは考えにくいことや、MLを用いる推論ではその中身がブラックボックスであり解釈性が低くなってしまう欠点がある。代表例であるWebb (2020) とその関連研究を以下に示す。

Webb : Webb (2020) は、AI関連の特許と0*NETのタスクのタイトルに含まれる動詞と名詞のペアの共起によって、exposureの直接的な尺度を提案している。結果として、ロボットやソフトウェアに基づくこれまでの技術革新よりも、AIが熟練労働者や高齢な労働者に影響を与える可能性が高いことを見出している。

その他の研究 : Kogan et al. (2021) は、特許テキストと職業テキストそれぞれの表現ベクトルのコサイン類似度によって exposure スコアを算出している。特に、Kelly et al. (2021) の方法論によって画期的なイノベーションとなる特許を特定し、それらの特

²⁸ LLMsは分析手法としても実証的なデータソースとしても社会科学の研究での利用が高まっており、その妥当性の検証として、人間による評価との比較が行われている。例えば、Gmyrek et al. (2024) は、職業上の威信 (occupational prestige) と職業上の社会的価値 (occupational social value) に関する人間の評価とGPT-4による評価を比較している。

許と職業間のexposureスコアを算出している。表現ベクトルは、まず単語レベルのベクトルをWord2vecと同じくトークンレベルでの表現ベクトルを学習するGloVe (Pennington et al., 2014) によって獲得し、それをTF-IDF重みづけ²⁹で平均することによって文書レベルのベクトルを得ている。Kogan et al. (2021) の手法は、他のいくつかの実証研究で用いられている (Autor et al., 2020; Meindl et al., 2021)。Autor et al. (2020) Census Alphabetical Index of Occupations (CAI) において新しく出現する職種カテゴリの分析を行っている。Meindl et al. (2021) では、第4次産業革命 (4IR) の特許テキストコーパスとO*NETのDWAセクション (脚注19を参照) をTF-IDFスコアと単語類似性マトリックスを用いてマッチングさせている。結果として、金融業と専門職は非4IRの特許に比べ、4IRの特許に多く曝されている傾向が示されている。Montobbio et al. (2023) は、ロボットによる省力化技術へのexposureスコアを特許データから求めて、O*NETのタスク記述とコサイン類似度でマッチングし、職業コード別に集計している。米国の地理的な分布を示したうえで、1999年から2019年までの雇用と賃金の変化に対して、コサイン類似度が負に相関することを示している。

4.5 AI exposureスコア (3) : 機械学習による算出

職業やタスクの内容とMLに可能な作業の記述の対を考えるのではなく、直接「MLにより代替されるか」を目的変数とし、職業やタスクの特徴量との関係を直接モデリングするアプローチがある。単純なMLの予測タスクに帰着することができるが、確からしい目的変数を設定することや、バックテストを行う際にデータ取得時点ではなく過去時点の特徴量を入手することが困難である。失業率など過去の定量的な実績を設定することもできるが、その結果獲得されるモデルは過去の傾向を反映するものにすぎず、本質的に興味のある「MLによる労働市場への影響」は外挿である点に注意が必要である。以下では、Frey and Osborne (2013, 2017) とその関連研究を紹介する。

Frey and Osborne : Frey and Osborne (2013, 2017) は、「職業のタスクがビッグデータの活用と最先端のコンピュータによって実行できるようになるか」を専門家に対するアンケートで収集し、その回答をガウス過程によって学習・分類するモデルを提案している。説明変数には、テクノロジーによる自動化が難しい領域を表す「ボトルネック

²⁹ 文書中のトークンが、「文書をどの程度特徴づけるものか」を表すための尺度の一つであり、当該トークンの当該文書における出現頻度 (term-frequency) と、トークンに出現する文書の逆頻度 (inverse document-frequency) を用いて算出される。

変数」として、O*NETから9つの指標を利用して、職種別の自動化確率（probability of computerization, computerization probability, CP）を求めている。

その他の研究：Frank et al. (2023) は, Webb (2020) や Felten et al. (2021) , Frey and Osbornme (2017) のexposureスコアを説明変数に加え, 失業率を直接予測するlassoモデルを構築している。様々な定義とアプローチによるexposureスコアに加え, 地域や職種の固定効果を考慮に入れた上での予測確率が出力されるため, その値を「代替の危機に曝されているか」に直接的に関連するexposureスコアとして得ることができる。

4.6 小括

4.3～4.5でみたexposureスコアを整理すると, 表2 (Meindl et al.2021, Table 4を一部改変) の通りとなる。Exposureの記述において, AIに限定するものと, ロボティクスや第4次産業革命までを含めるものがある。タスクの多くは, O*NETに接合しているが, 指標のベースが, オンライン調査, 専門家の評価, 特許の記述などの違いがある。そのため, これらの指標間での整合性の有無ははっきりしない。この点に関して, 次節の研究紹介をみてみよう。

表2. Meindl et al. (2021, Table 4) による各exposureスコアの整理

	4IR exposure (Meindl et al. 2021)	AI Occupational Exposure (AIOE) (Felten et al., 2021)	SML (Brynjolfsson & Mitchell, 2017)	probability of computerization (Frey & Osborne, 2017)	Automation potential (Manyika et al., 2017)	AI exposure (Webb, 2019)
Description	Exposure to technologies of the fourth industrial revolution (most of which enable the introduction of AI)	Exposure to artificial intelligence technology	Suitability for machine learning technologies	Risk of being automated through AI and robotics	Share of activities potentially being automated through automation technology	Exposure to AI patents
Level of analysis	Tasks conducted by occupations (O*Net)	Abilities required by occupations (O*Net)	Tasks conducted by occupations (O*Net)	Occupation level, scaled through abilities (O*Net)	18 performance capabilities (based on O*Net abilities)	Tasks conducted by occupations (O*Net)
Index basis	Relevant patents per task	Scientific AI advances, as described by the EFF, per ability	Online survey of SML characteristics per task	Expert evaluation of selected occupations automation potential	Expert evaluation of technology performance per capability	Relevant patents per task
Measure interpretation	Existing technology 4IR capabilities based on patents (current diffusion)	Existing theoretical AI capabilities based on scientific articles (possibly near future diffusion)	Potential future capability of ML (possible future diffusion)	Potential future capabilities of technology (possible future diffusion)	Potential future capabilities of technology (possible future diffusion)	Existing technology AI capabilities based on patents (current diffusion)

Source: Meindl et al. 2021. Table 4: Overview of automation, AI, and 4IR indicatorsの一部を改変した。

4.7 AI exposureスコアのメタ分析

これまで提案されてきたexposureスコアがどのような性質をもつか、メタ分析を行っている研究がいくつか存在する。今後本邦でもexposureスコアの継続的な算出と分析を行っていくうえでも、各手法の特徴を掴むうえでこれらの議論は大いに参考になる。賃金や雇用を被説明変数とし、異なるexposureスコアを説明変数とする回帰分析を行い、exposureスコア間の差異を把握するもの (Acemoglu et al. 2022, Frank et al. 2023), もう一つは、そもそもexposureスコアの違いを考慮した仮説を立ててそれを検証するもの (Fossen and Sorgner 2022) である。以下、それぞれの論文をみてみよう。

Acemoglu et al. (2022) は、主要なexposureスコアであるWebbのexposureスコア、AIOE, SML の各指標を用い、AIが労働市場にどのような影響を与えているか、オンライン求人データに由来する様々な側面からの分析を行っている。3つの指標のうちSMLはAI関連の求人数との相関が弱く、特にWebbのexposureスコアとAIOEを用いた結果を報告している。結果としては、AIのスキルに適したタスクを多く持つ企業では、AI関連の求人や新しいスキルを求める求人を増やし、一方で他の求人を減らしていることを示している。また、企業が労働者の雇用を全体的に減らしている傾向も示しているが、職業レベルや産業レベルでの雇用や賃金への影響は検出できず、AIとその普及の影響が労働市場全体ではまだ比較的小さい可能性があると考えしている。

Frank et al. (2023) は、Webb (2020) や Felten et al. (2021), Frey and Osborne (2017) などのexposureスコアを説明変数に加えた失業率を予測する機械学習モデルを構築している。結果として、個々のexposureスコアを単体で用いたときの予測精度は低いですが、地域や時間の固定効果などを考慮するベースモデルにexposureスコアを加えたとき、大きく精度向上がみられることを報告している。

Eloundou et al. (2023) はexposureスコアの提案だけでなく、提案スコアを既存のexposureスコアで回帰する比較も行っている。Brynjolfsson and Mitchell (2017), Webb (2020) との相関が高い一方で、Felten et al. (2021), Frey and Osborne (2013, 2017) との相関はなかった。Albanesi et al. (2023) は、Felten et al. (2021) と Webb (2020) のexposureスコアを使い、2011年から2019年の欧州16カ国では、AIにより多く曝される職業ほど、雇用シェアが増加しており、スキル偏向技術進歩仮説と整合的であった。Fossen and Sorgner (2022) は、2011年から2018年の米国の個人レベルの賃金・雇用に対して、Frey and Osborne (2013, 2017), Brynjolfsson and Mitchell (2017), AIOI (AI Occupational Impact, Felten et al., 2019), MLの職種内標準偏差を回帰させた。大卒の賃金成長率に対して、置換効果を表すFrey a

nd Osborne (2013, 2017) と Brynjolfsson and Mitchell (2017) は有意に負であるのに対して、復職効果を表すAI0Iは正であった。

Exposureスコアの違いはどう考えればよいだろうか。Brynjolfsson and Mitchell (2017), Felten et al. (2019), Webb (2020) では、AIに最も曝される職業として高技能職が挙げられていたのに対して、Frey and Osborne (2013, 2017) や Nedelkoska and Quintini (2018) は低技能職が自動化技術にさらされていた。この点に関して、Lane and Saint-Martin (2021) は、二つの理由を述べている。一つは、後者の研究が実施された2013年からの10年間の技術革新の進捗が結果の違いをもたらした可能性である。もうひとつは、Frey and Osborne (2013, 2017) はより広い技術革新を含むが、Brynjolfsson and Mitchell (2017), Felten et al. (2019), Webb (2020) は機械学習の技術的能力に焦点を絞っていることである。このことが、通常は、自動化技術は定形的な業務を実行する (Autor et al. 2003) と考えられてきたが、問題解決、論理的推論、知覚におけるAIの向上は、FeltenやWebbの研究では非定型な認識業務に関わるようになり、高度な技能を持つ技術者がAIに強くさらされているように見えているのではないかと指摘している。

また、多くの研究は、Frey and Osborne (2013, 2017) 流のコンピュータ化リスクが必ずしも雇用に影響していないことを示している。Arntz et al. (2016) は、3つの理由を挙げている。一つは、AIの利用には、経済的・法的・社会的なハードルがあるため、技術的代替は遅々として進まないこと、二つ目に、AIが導入されたとしても、労働者は仕事を切り替えて技術的失業を防げること、三つめは、AIは生産性向上と需要拡大を招き、追加的雇用を生む可能性があることである。その意味では、AIの導入に備えた、低スキル労働者に対するリスクリングが望まれる (Arntz et al. 2016)

4.8 各国における実証分析

exposureスコアの開発と実証は主に米国の労働市場を対象に行われてきたが、追随する形で米国以外の国での実証分析も進みつつある。

Stapleton et al. (2021) はインドの求人データセットを用い、インドにおけるAIの導入がサービス業界の企業の雇用と賃金に与える影響を調査している。米国におけるAcemoglu et al. (2022) と同様に、AI関連の求人が高い賃金プレミアムをもつこと、AI関連の労働需要が非AI関連の労働需要と負の相関にあることを報告している。その分析過程では、AIに関係なく労働需要と事業所ごとのアウトカムの間に関連が想定されるため、Webb (2020) のexposureスコアをインドの労働市場に適用し、操作変数として用いている。

中国においては、Chen et al. (2023) が Eloundou et al. (2023) のexposureスコアを適用し、業界単位と職業単位の集約をもとに LLMs による潜在的な影響を調査している。結果として、以前のコンピュータベースの技術革新とは異なり、非ルーチン的な認知タスクを遂行する職業で大きい影響を受ける可能性があるとしている。一方 Lou et al., (2023) はルーブリックの策定に基づいたexposureスコア算出のアプローチを取り、同様に LLMs による中国の労働市場への影響を調査している。exposureスコアは職業、業種、地域別に集約され、影響の程度は米国よりも小さいものの、より高い教育レベルや賃金をもつ職業、情報セクターや金融セクター、直轄市や省都などのプレミアム都市の労働力に大きな影響を与える可能性があるとしている。

国横断的な分析も行われ始めている。Engberg et al. (2023) は AI0E (Felten et al., 2021) を動的に拡張し、デンマーク、ポルトガル、スウェーデンの労働市場に適用している。この3カ国は労働市場の硬直性や産業構造、AIを採用する程度で異なり、この異質性に関する考察もなされている。結果として、よりexposureスコアの高い企業はブルーカラー労働者の雇用を減らし、高スキルのホワイトカラー労働者の雇用を増やすとしている。また、画像処理と言語処理の進歩が高スキルのホワイトカラーの労働需要を高める一方で、ブルーカラー労働者にはネガティブな影響を与えることを示している。Pizzinelli et al. (2023) は、同様にAI0Eを元にしたexposureスコアを、2つの先進国（米国、英国）と4つの新興国（ブラジル、コロンビア、インド、南アフリカ）について算出し、分析している。未調整のexposureスコアは国ごとに大きなばらつきがあり、先進国では管理職や専門職の割合が大きいため、新興国よりも高いリスクに晒されているとした。また、先進国と新興国に共通する傾向として、女性の労働者と高い教育レベルをもつ労働者は高いexposureスコアをもつことを指摘している。Carbonero et al. (2023) は、ラオスとベトナムについて、世界銀行のSTEP調査 (Skills Measurement Program, STEP) によって提供された機械学習のための作業活動の適合性の尺度を用いている。ラオスと比較して、ベトナムの都市部の労働者にAIの影響を受ける職業が集中していることがわかった。SBERT³⁰を用いた意味的なテキストの類似性に基づく研究方法を用いている。

exposureスコアの算出を通じた将来の洞察を得るのではなく、AIが労働市場にどのような影響を及ぼしたかを分析する実証研究も、同様に各国で行われている。Yang (2022) は、2022年から2018年にかけての台湾の電子業界の特許データと生産性のパネルデータを分析し、企業がAI関連の技術開発を実施することで、労働需要は約3.5%増加したこと

³⁰ Sentence BERT (SBERT) は2つの文章が与えられたときに両者の関係性を学習するニューラルネットワークである。

を示している。Hilstob and Massie (2022) は、カナダの労働組合が発行したオンライン資料をレビューし、AIによる技術革新と労働の自動化に反対を表明した労働組合はないものの、多くの労働組合がどのようにAIが企業で活用されるか、社会的および地政学的な影響に疑問を抱いているとしている。

4.9 その他のテーマ

ここまでexposureスコアの開発とその適用について先行研究をレビューしてきたが、ここでは、MLやAIの労働市場への影響を分析するその他のテーマについても簡単にまとめる。具体的には、MLやAIの導入に関するケーススタディ、導入による生産性効果の計測、労働需要の喪失・創出の3つのテーマである。

①MLやAIの導入に関するケーススタディ

近年では、あらゆる産業セクターでMLやAIの導入が進められており、労働市場への影響を測る上では、どのような応用が典型的であるのか、さらには代替などの現象が実際に起こっているのかを把握することが重要である。ここでは、Webb (2020) や Felten et al. (2021) でexposureスコアの上位に挙げられている職業が含まれることから、医療分野と金融分野における先行研究を紹介する³¹。

医療分野では、MLを駆使した医療画像解析がCOVID-19の検出のような事例で高精度な診断を達成し、放射線科医の役割を変えつつある (Shi et al., 2020; Nayak et al., 2021)。また、歯科医学においても、歯科インプラントの必要性を予測するMLアルゴリズムにより診断過程を簡略化する取り組み (Alharbi and Almutiq, 2022) などがある。その他広範な応用事例は Shebhab et al. (2022) や (Khanagar et al., 2021) にまとまっている。また、LLMsの応用としては、Lee et al. (2023) が医療用チャットボットのメリットとデメリットについて整理しており、特に活用の余地があるタスクとして、音声からのカルテ自動生成や簡易な医療相談、既存の医療知識の提供を挙げている。

金融分野においても、銀行、保険、資産運用、ヘッジファンドなどの各業態でMLへの投資と活用が進んでおり、国際金融の監督機関であるFSBもその影響を注視している (FSB, 2017)。また、He et al. (2018) は雇用創出、雇用削減、生産性向上の可能性があるとの見通しを示している。実際に、資産の価格過程のモデリング (Dixon, et al., 2020) やテキストからの情報抽出 (Loughran and McDonald, 2020) など、従来クオン

³¹ なお、聞き取り調査としてOECDの国際比較研究がある (Milanez (2023))。この一環として、日本では、金融業と製造業が調査されており (労働政策研究・研修機構2022, 2023)、その比較考察の結果が労働政策研究・研修機構 (近刊b) で公表予定である。

ツやアナリストが担ってきたタスクをMLで行うための技術が開発されている。また、様々なタスクでLLMsを活用しようとする取り組みが始まりつつあり、Dong et al. (2023)はそのサーベイを行なっている。本邦においても、経営者の決算説明からの主観的表現の抽出 (Kuroki et al., 2023) や、サステナビリティ情報の開示についての充足判定 (土井ら, 2024) などの事例が出てきている。

②生産性効果

AIやMLを用いることによる労働者レベルの生産性向上については、特に最近では、チャット形式という一般の労働者に馴染み良いインターフェースをもつLLMsであるChatGPTなどのツールの登場を背景に、ツールの使用をランダムに割り当てる実験が行われている。Brynjolfsson et al. (2023) は、フィリピンに拠点を置く米国のカスタマーサービス会社の担当者約5,000人について、ランダムに生成AI (ChatGPT) にアクセスさせて、1時間あたりの問題解決数を1年間にわたって観察した。その結果、生成AIを使用した担当者の生産性は平均13.8%向上した。とくに、初心者レベルや低スキルの労働者で生産性向上が大きかった。

Noy and Zhang (2023) は、大学教育を受けた444人の被験者を対象に、ChatGPTの使用がタスクの遂行に与える影響を検証した。被験者はマーケティング、ライティング、コンサルティングなどの職業固有の作文タスクが割り当てられ、そのアウトプットは同職種のエキスパートによって評価された。結果として、ChatGPTを使用したグループは作業時間を平均10分短縮し、評価も上昇させた。さらに、作業の内訳や被験者の属性との分析により、ChatGPTの使用により作業の焦点が作文からアイデア生成や編集に移行したことや、特にもともと能力の高くない労働者にとってChatGPTの恩恵が大きいことなども示唆している。

Dell'Accia et al. (2023) もコンサルタントに複雑な知識集約的なタスクを実施させ、同様にChatGPTの使用が質の向上と時間の節約に効果があったことを報告している。また、作業内容の観察と定性的な分析より、人によってAIとの役割分担があることを指摘し、AIの得意なタスクを切り分け役割分担を行うケンタウロス行動と、サブタスクのレベルでさえAIと使用者が密接に統合されているサイボーグ行動をとるグループに分けられると論じている。

Kanazawa et al. (2022) は、横浜の520人のデータを用いて、タクシードライバーへのAIナビ導入の空車時間削減効果を推定している。このAIナビは、乗客がいる確率が高い地域にタクシードライバーを誘導する需要予測AIである。分析の結果、AIナビをオンにすると、平均的に空車時間が5%削減され、低技能のドライバーの方が削減幅(7%)

が大きかった。これらの研究は、低スキルの労働者ほどAIの便益を得ることで格差が是正される可能性があることを示唆している。

③労働需要の喪失・創出

AIやMLが労働需要の喪失・創出にどうかかわるかを検証する研究では、アンケート調査であるMorikawa (2017) は、約3000社の企業アンケートから、高スキルの労働者や海外展開している企業ほど、AI等への取り組みに前向きであり、AI・ロボットが自社の雇用に及ぼす影響は、雇用抑制的と考える企業 (21.8%)の方が雇用増加につながると考える企業 (3.7%)より多かったことを示している。Bessen et al. (2018) はAIスタートアップ179社を調査し、75%のAIスタートアップが、彼らの製品はクライアントの予測、データマネジメントを改善したことを示す一方、50%のスタートアップが顧客の定型業務や労働費用の削減に貢献したと答えている。Kitahara and Shinozaki (2019) は、独自のアンケート調査により、受付、経理、品質保証、営業、人事の5つの職種について、タスク別にノン・ルーティンタスク集約度 (Non-Routine Task Intensity) を求めて、AI導入前後の変化を推定して、AI導入後は、労働時間が減少しノン・ルーティンタスクが増加したことを示した。労働政策研究・研修機構 (2022) は、約27,000の個人に対して、Frey and Osborne (2017) やAutor and Handel (2013) に準拠したアンケート調査を行っている。また、Nedelkoska and Quintini (2018) に準拠した雇用の代替可能性を表す自動化確率を算出して、雇用の代替可能性 (自動化確率) が70%以上の高リスク層が占める比率は全体の10.0%であることを示している。

実証分析では、Grennan and Michaely (2019) はAIのアルゴリズムが証券アナリストに与える影響を分析し、タスクの代替を見つけている。アナリストは、豊富なデータが入手可能な銘柄を担当する場合、離職する可能性が高かった。Alekseeva et al. (2021) は、2010年から2019年にかけて米国経済ではほとんどの産業と職種においてAIスキルの需要が大幅に増加しており、IT職業で最も高く、建築・工学、科学、管理職がそれに続くことを示している。Damioli et al. (2022) はAIやMLを活用する企業に着目するため、2000年から2016年の間にAI関連の発明の特許化した3,500社以上の全世界的なパネルデータを分析し、AI関連特許の取得数が労働需要の増加に有意に関連していることを示している。台湾での実証研究である Yang (2022) もAI関連特許はその他の特許と同様に、企業の生産性や雇用と正の関係にあるとしているが、従業員構成を変化させることも示している。

そのほか、White et al., (2023) は様々なタスクで LLMs からより効果的に性能を引き出すためのプロンプト構成のパターンを整理し示しており、生成AIの出現は質の

高いプロンプトとアウトプットを作成するプロンプト・エンジニアなど、新たな職務を生み出す可能性がある。また、求職者にとってのメリットもあると考えられる。直接的なメリットとしては履歴書の作成補助などが考えられるが、間接的にも生成AIによる言語能力による障壁の解消 (Jiao et al., 2023) やプログラミング能力による障壁の解消 (Qadir, 2023) では既に効果的であるとされており、求職者はスキル向上の効率化と費用の節約を見込める可能性がある。

5. おわりに

本章では、これまでのまとめと考察を行い、今後の研究の方向性を展望した後、日本における研究の可能性についても言及する。

5.1 まとめと考察

ここまで、exposureスコアの算出に関する既存研究をおおまかなアプローチごとにグルーピングし、周辺の文献と共に整理してきた。各アプローチにはそれぞれメリットとデメリットがあったが、デメリットには他のアプローチにより解決可能とかがえられるものがあり、今後はこれらの潮流の統合が進むと考えられる。

まず、ルーブリックの策定とそれを元にしたスコアリング (Brynjolfsson et al. 2018, Felten et al. 2018) では、GPT-4の活用 (Eloundou et al. 2023, Felten et al. 2023) を含む人間による複雑な推論が可能であることが大きな利点であった。これにより、設定されたexposureの概念を自然に反映したスコアリングが可能である。しかし、このプロセスでは一つ一つの判断が複雑であるため、特にタスクなどの単位でスコアリングするにはリソースの制約を受ける可能性がある。より一層クラウドソーシングの活用が進むことも考えられるが、その品質の客観性を完全に担保することが難しく、コストの制約も大きい。

職業やタスクを定義するテキストと、機械学習に可能な作業を記述するテキストの関連性をテキストマイニングするアプローチ (Webb 2020; Kogan et al. 2021) では、客観的なデータに基づいた推論が可能であり、同じデータとロジックを用いれば再現性が高いというメリットがある。ただし、入力であるテキストが職業や特許の複雑な情報を十分に考慮できていない可能性があり、機械学習を用いる場合には、推論過程の解釈性の低さが残りうる。ただし、今後LLMsがより長いトークンを受け入れられるようになれば、その性能の多面的な検証と共にこれらの問題を解決するだろう。

機械学習による失業確率の予測 (Frey and Osborne 2017; Frank et al., 2021) は、exposureの趣旨を直接的に表現できるという点で有益であった。しかし、この方法

には正解データが「代替されるか」という抽象的な質問への回答に過ぎないという問題や、モデルの精度評価や妥当性を測るバックテストにおいては、過去時点で知りうるデータのみが必要であるという重大な課題が残る。たとえばWebページに由来する情報や他のexposureスコアを変数として用いる場合には、過去時点で知り得た情報のみ利用することに細心の注意が必要である。

これらのexposureスコアは、それぞれの趣旨や扱うAIの範囲が異なり、結果としてそれぞれの相関が高くないことも報告されている (Acemoglu et al., 2022; Eloundou et al., 2023; Frank et al., 2023)。分析の目的に合わせてどのようなexposureの趣旨が望ましいかは異なることから、今後は、その細分化と適切な使い分けが進んでいくと考えられる。さらに、それぞれの性質を比較し継続的に分析する上でも、それぞれのexposureスコアは容易に閲覧でき、突合できる状態にあることが重要である。メタ分析を行う研究が出始めていることから、利用データやアルゴリズムを統一したデータベースの整備が進むと考えられる。また、それぞれのスコアのより詳細な性質や性能の理解が進むことが期待される。

テキスト情報を用いたexposureスコアの算出では、GPT-4などの先進的なモデルの利用が進むと考えられる。これはコスト的なメリットが大きだけでなく、高性能なLLMsの登場により、人手によるスコアリングの大きなメリットであった複雑な判断が可能になりつつある。一方で、推論過程のブラックボックス性と解釈の困難さについては解決の見通しが立っていない。説明可能なAIの研究が進んでいるものの、現状では完全な解決には至っていないため、解釈性や再現性に優れるアプローチが今後も重要であると考えられる。例えば、複数の研究において、算出したexposureスコアが大きく異なる場合、それがアルゴリズムの違いで生じたものか、データの特徴量の違いから生じるものかを識別するためには、同じデータで異なる手法を用いて分析して比較するなど、漸進的なアプローチも不可欠である。

最後に、人工知能が労働市場にもたらすインパクトに関する研究の起点には、それらが格差拡大をもたらすという懸念があったことに鑑みると、所得や仕事の質などのさまざまな格差の動向に関する研究が今後も期待される。低スキルの労働者ほど人工知能技術に裨益するという事例 (本稿4.9参照) も出てきている一方で、生成AIが高所得者に補完的に働いたり (Cazzaniga et al. 2024)、生成AIによって事務職や女性ほど自動化の影響を受けたりする場合 (Gmyrek et al. 2023) には格差拡大が懸念される。また、MLやAIによる生産性効果や再配置効果により人口構成の変化が起きた後の賃金や雇用の評価、いわゆる一般均衡効果に関する分析も望まれる。Acemoglu and Restrepo (2022) は、自動化の一般均衡効果を評価する方法として、産業構成の変化や異なるグルー

プ間でのタスクの再配置による波及効果を組み込んだモデルを考案している。Shinozaki et al. (2022) では、Webb (2020) のexposureスコアを外挿し、32部門のCGEモデルを用いて、AI資本の増加の影響を分析している。その結果、AIが労働に補完的でない場合や、AIが高所得産業で代替的で低所得産業に補完的である場合には、賃金格差が縮小し得ることを示している。今後もシミュレーションに利用できるモデルやパラメータの開発が期待される。

5.2 日本における研究に向けて

本邦ではどのように研究が進展するであろうか。本邦と米国では労働市場についても蓄積されているデータについても性質が異なること、その他の国においてもexposureスコアの分析で米国市場との異質性が確認されていることから、本邦データでの追試とアレンジ、結果の解釈を継続的に行っていくことが重要である。既に千葉・福田 (2023) が Frey and Osborne (2017) を日本版O-NETへと拡張しているが、これまで提案されてきたexposureスコアはその性質が様々であることより、目的に沿った手法を選定しつつも、複数の指標を継続的に算出し観測することが重要である。クラウドソーシングのコストを鑑みると、着手しやすいのは Webb (2020) や Kogan et al. (2021) をはじめとするテキストマイニングを用いるアプローチや、LLMsを用いるアプローチであろう。ただし、係り受けや体言止めなど日本語特有の表現に対応した適切な自然言語処理手法を選択し、LLMsを用いる場合には英語テキストを扱う場合との性能差に注意する必要がある。

職業テキストデータの観点では、本邦においても日本版O-NETや厚生労働省編の職業分類³²では職業の説明テキストが存在することから、これらの拡充と利用が進むと考えられる。すでに金融分野では、金融文書 (鈴木他2021) や景況診断 (五島他2022; 新谷2023) で自然言語処理の手法が応用されつつある。ただし、日本版O-NETから特徴量を抽出する分析に際しては、日本の労働市場や雇用慣行の特徴を考慮する必要がある。日本ではいわゆるメンバーシップ型雇用 (濱口2021) がメインであり、職務記述書などで職務範囲が明確に示されていないことから、O-NETでジョブをタスクレベルに分解して記述する際にはさまざまな困難を伴う。また、国際比較する際には、exposureスコアだけでなく、タスクの記述内容の違いにも留意すべきである。タスクの定義と測定、国際比較を通して、日本でAIをいかに活用するか、自動化のための促進・抑制する条件 (雇用慣行や商慣習) は何かを見極めて実務面で応用していくことが期待される。

³² <https://www.jil.go.jp/institute/seika/shokugyo/bunrui/index.html>

AIやMLに可能な作業を表すテキストとしては、特許テキストなどが利用できる。特許文書を機械学習手法で調査する試みがすでになされているが（安藤2017；安藤2020；西尾・安藤2020），特許情報プラットフォームJ-PlatPat³³からいかに対象特許を抽出するかの問題が残る。この分野の研究の発展のためには、より妥当なクエリとさらなるフィルタの精査，コーパス化が急務であると考えられる。

本稿で述べたように、MLやAIが労働にもたらす影響の分析は、分析手法とデータが同時に発展しながら進められている。日本においても、求人情報、職業・スキル情報、特許情報などの数値・テキスト情報がどのような分析に利用可能であるか、新しい分析手法を想定したデータの開発と共有を同時に進めていく必要がある。また、日本語テキストを機械的に処理するにあたっては、2.3節で述べたように、前処理に日本語に特有の困難さが存在する。前処理によって分析結果も異なることが想定されるため、研究の再現性を高めるためには前処理を含めた分析過程の共有が必要である。そのためには、GitHub³⁴などのプラットフォーム上でオープンにしておくことで、後続の研究がより簡単に追試と改善を試みることができるようになるほか、課題や改善案を広く受け付けることが可能となる。また同時に、形態素解析と辞書などのライブラリに関してどのバージョンを利用したかを明記しておくことが重要である。

³³ <https://www.j-platpat.inpit.go.jp/>

³⁴ <https://github.com/>

参考文献

- 安藤俊幸（2017）機械学習を用いた効率的な特許調査方法—ニューラルネットワークの特許調査への応用，Japio year book, 230-232.
- 安藤俊幸（2020）機械学習を用いた効率的な特許調査方法—単語・文書のベクトル化方法と特許調査への応用，Japio year book, 222-234.
- 池永肇恵（2009）労働市場の二極化—ITの導入と業務内容の変化について，日本労働研究雑誌，No.584, pp.73-90.
- 池永肇恵（2011）日本における労働市場の二極化と非定型・低スキル就業の需要について，日本労働研究雑誌，No.608, pp.71-87.
- 北原聖子（2017）AI等の技術が労働市場に与える影響に関する内外の研究動向について．ESRI research note, (43).
- 五島圭一・新谷元嗣・高村大也（2022）景気単語極性辞書の構築とその応用．自然言語処理, 29(4), 1233-1253.
- 小原美紀・大竹文雄（2001）コンピュータ使用が賃金に与える影響，日本労働研究雑誌, 第494巻, pp.16-30.
- 小松恭子（2022）転職行動の男女差：転職前後のタスク距離と賃金変化に着目して，ディスカッションペーパー 22-05, 労働政策研究・研修機構.
- 小松恭子・麦山亮太（2021）日本版 0-NET の数値情報を使用した応用研究の可能性：タスクのトレンド 分析を一例として．JILPT Discussion Paper, 21(11), 1-58.
- 櫻井宏二郎（1999）偏向的技術進歩と雇用—日本の製造業のケース—，循環と成長のマクロ経済学, 東京大学出版会, 165-187.
- 新谷元嗣（2023）テキスト情報と機械学習を用いた景気動向分析．経済分析, 208, 128-145.
- 鈴木雅弘・坂地泰紀・平野正徳・和泉潔（2021）金融文書を用いた事前学習言語モデルの構築と検証．人工知能学会研究会資料 SIG-FIN-027.
- 千葉茂樹・福田節也（2023）職業の自動化確率についての日米比較 Frey&Osborneモデルの再現と日本版0-NETデータへの拡張．JILPT Discussion Paper 23. S-01.
- 土井惟成・小田悠介・中久保菜穂・杉本淳（2024）大規模言語モデルを用いたゼロショットテキスト分類によるTCFD推奨開示項目の自動判定．JPXワーキング・ペーパー, 43.
- 西尾潤・安藤俊幸（2020）AI系基盤技術と，オープンソースを用いた機械学習による特許文書解析, (298), 25-37.
- 濱口桂一郎（2021）ジョブ型雇用社会とは何か 正社員体制の矛盾と転機．岩波書店.

- 森知晴（2019）インターネットを利用した「経済実験」の動向と展望．日本労働研究雑誌．（705），2-7.
- 森脇大輔（2022（））労働分野におけるオルタナティブデータの活用．日本労働研究雑誌，（741）．61-64.
- 勇上和史・牧坂亮佑（2021）転職が不利にならない条件，労働政策研究報告書，No.210，pp.75-94.
- 労働政策研究・研修機構（2022）自動化技術の普及による雇用の代替可能性に関する個人調査，調査シリーズNo.225
- 労働政策研究・研修機構（2022）金融業におけるAI技術の活用が職場に与える影響—OECD共同研究—，JILPT資料シリーズNo.253
- 労働政策研究・研修機構（2023）製造業におけるAI技術の活用が職場に与える影響—OECD共同研究—，JILPT資料シリーズNo.262
- 労働政策研究・研修機構（近刊a）タスクの日米比較からみた日本の労働市場の特徴と変化—日本版0-NETと国勢調査（1980～2020年）を使用した分析から得られた示唆—」JILPT資料シリーズ No.280.
- 労働政策研究・研修機構（近刊b）職場におけるAI技術の活用と従業員への影響—OECDとの国際比較研究に基づく日本の位置づけ，労働政策研究報告書No.228.
- Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics*, 4, 1043-1171.
- Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: evidence from online vacancies. *Journal of Labor Economics*, 40(S1), S293-S340.
- Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American economic review*, 108(6), 1488-1542.
- Acemoglu, D. & Restrepo, P (2022). Tasks, Automation and the Rise in U.S. Wage Inequality, *Econometrica*, Vol. 90, No. 5 (September, 2022), 1973-2016.
- Agrawal, A., Gans, J. & Goldfarb, A. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence, *Harvard Business Review Press; Illustrated edition.*

- Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, 33(2), 31-50.
- Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management*, 47(4), 823-837.
- Albanesi, S., da Silva, A. D., Jimeno, J. F., Lamo, A., & Wabitsch, A. (2023). *New Technologies and Jobs in Europe* (No. w31357). National Bureau of Economic Research.
- Alekseeva, L., Azar, J., Gine, M., Samila, S., & Taska, B. (2021). The demand for AI skills in the labor market. *Labour economics*, 71, 102002.
- Arntz, M., Gregory, T., & Zierahn, U. (2016). The risk of automation for jobs in OECD countries: A comparative analysis. *OECD Social, Employment and Migration Working Papers* No. 189
- Asplund, R., Barth, E., Lundborg, P., & Nilsen, K. M. (2011). Polarization of the Nordic Labour Markets. *Finnish Economic Papers*, 24(2), pp.87-110.
- Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4), 1279-1333.
- Autor, D. H., & Handel, M. J. (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics*, 31(S1), S59-S96.
- Autor, D. H., Katz, L. F., & Kearney, M. S. (2006). The Polarization of the U.S. Labor Market. *American Economic Review*, 96(2), pp.189-194.
- Berman, E., J. Bound, and Z. Griliches, (1994). Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from Annual Survey of Manufacturing, *Quarterly Journal of Economics*, Vol. 109, pp. 367-397.
- Bessen, J., Impink, S. M., Reichensperger, L., & Seamans, R. (2018). The business of AI startups. *Boston University School of Law, Law and Economics Research Paper*.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

- Bound, John and George Johnson. “Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations.” *American Economic Review* 82 (June 1992): 371-392.
- Brodeur, A., Cook, N., & Heyes, A. (2022). We Need to Talk about Mechanical Turk: What 22,989 Hypothesis Tests Tell Us about Publication Bias and p-Hacking in Online Experiments. *available at SSRN*, 4188289.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901
- Brynjolfsson, E., & McAfee, A. (2011). Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy. *Lexington: Digital Frontier Press*.
- Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. *WW Norton & Company*.
- Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, 358(6370), 1530-1534.
Supplementary Materials for What can machine learning do? Workforce implications
https://www.science.org/doi/suppl/10.1126/science.aap8062/suppl_file/aap8062-brynjolfsson-sm.pdf 2024年3月13日最終アクセス
- Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy?. *AEA papers and proceedings*, 108, 43-47.
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. *NBER Working Paper Series*, 31161.

- Cammeraat, E., & Squicciarini, M. (2021). Burning Glass Technologies' data use in policy-relevant analysis: An occupation-level assessment. *OECD Science, Technology and Industry Working Papers* 2021/05
- Carbonero, F., Davies, J., Ernst, E., Fossen, F. M., Samaan, D. & Sorgner, A. (2023). The impact of artificial intelligence on labor markets in developing countries: a new method with an illustration for Lao PDR and urban Viet Nam. *Journal of Evolutionary Economics* 33:707-736.
- Carnevale, A. P., Jayasundera, T., & Repnikov, D. (2014). Understanding online job ads data: a technical report. *Georgetown University, McCourt School on Public Policy, Center on Education and the Workforce*.
- Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A. J., Pizzinelli, C., Rockall, E J., & Tavares, M (2024). Gen-AI: Artificial Intelligence and the Future of Work. IMF Staff Discussion Note SDN2024/001, International Monetary Fund, Washington, DC.
- Chen, Q., Ge, J., Xie, H., Xu, X., & Yang, Y. (2023). Large Language Models at Work in China's Labor Market. *arXiv preprint arXiv:2308.08776*.
- Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda* (pp. 115-146). University of Chicago Press.
- Damioli, G., Van Roy, V., Vertesy, D., & Vivarelli, M. (2022). AI technologies and employment: micro evidence from the supply side. *Applied Economics Letters*, 30(6), 816-821.
- Davis, S., Faberman, R. J., & Haltiwanger, J. C. (2012). Recruiting Intensity during and after the Great Recession: National and Industry Evidence. *American Economic Review*, 102 (3): 584-88.
- Davis, S., Faberman, R. J., & Haltiwanger, J. C. (2013). The Establishment-Level Behavior of Vacancies and Hiring. *Quarterly Journal of Economics*, vol. 128, issue 2, 581-622.
- De La Rica, S., Gortazar, L., & Lewandowski, P. (2020). Job tasks and wages in developed countries: Evidence from pиаac. *Labour Economics*, 65, 101845.

- Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraymer, L., Candelon, F., & Lakhani, K. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, 24-013.
- Deming, D. J. & Noray, K. L. (2019). STEM Careers and the Changing Skill Requirements of Work, *NBER Working Paper Series*, 25065.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Machine learning in finance. *Springer International Publishing*.
- Dustmann, C., Ludsteck, J., & Schönberg, U. (2009). Revisiting the German Wage Structure. *Quarterly Journal of Economics*, 124(2), pp.843-881.
- Dong, M. M., Stratopoulos, T. C., & Wang, V. X. (2023). A Scoping Review of ChatGPT Research in Accounting and Finance. *available at SSRN 4680203*.
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.
- Engberg, E., Görg, H., Lodefalk, M., Hellsten, M., Javed, F., Långkvist, M., Monteiro, N., Kyvik-Nordås, H., Pulito, G., Schroeder, S., Tang, A., & Tydén, O. (2023). AI Unboxed and Jobs: A Novel Measure and Firm-Level Evidence from Three Countries. *Available at SSRN 4691172*.
- Felten, E. W., Raj, M., & Seamans, R. (2018). A method to link advances in artificial intelligence to occupational abilities. *AEA Papers and Proceedings*, 108, 54-57.
- Felten, E. W., Raj, M., & Seamans, R. (2019). The occupational impact of artificial intelligence: Labor, skills, and polarization. *NYU Stern School of Business*.

- Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12), 2195-2217.
- Felten, E., Raj, M., & Seamans, R. (2023). How will Language Modelers like ChatGPT Affect Occupations and Industries?. *arXiv preprint arXiv:2303.01157*.
- Ford, M. (2015). *Rise of the Robots: Technology and the Threat of a Jobs Future*. Basic Books.
- Fossen, F. M., & Sorgner, A. (2022). New digital technologies and heterogeneous wage and employment dynamics in the United States: Evidence from individual-level data. *Technological Forecasting and Social Change*, 175, 121381.
- Frey, C. B., & Osborne, M. A. (2013) The Future of Employment: How Susceptible Are Jobs to Computerisation? Working Paper, University of Oxford, Oxford.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological forecasting and social change*, 114, 254-280.
- Gathmann, C. & Schönberg, U. (2010). How General Is Human Capital? A Task-Based Approach, *Journal of Labor Economics*, 28, 1, pp.1-49.
- Gemini Team, Anil, R., Borgeaud, S., Wu, Y., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., et al. (2023). Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.
- Goos, M., & Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in Britain. *Review of Economics and Statistics*, 89(1), pp. 118-133.
- Goos, M., Manning, A., & Salomons, A. (2009). Job Polarization in Europe. *American Economic Review*, 99(2), pp.58-63.
- Goos, M., Manning, A., & Salomons, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review*, 104(8), pp. 2509-2526.

- Grennan, J. & Michaely, R. (2020). Artificial Intelligence and High-Skilled Work: Evidence from Analysts. *Swiss Finance Institute Research Paper Series*. 20-84.
- Gmyrek, P., Berg, J., & Bescond, D. (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality,
- Gmyrek, P., Lutz, C., Newlands, G. (2024). A Technological Construction of Society: Comparing GPT-4 and Human Respondents for Occupational Evaluation in the UK, ILO Working Paper 102 (Geneva, ILO). <https://doi.org/10.54394/UQ0Q5153>
- Hilstob, K., & Massie, A. (2022). Artificial Intelligence and Labour: Perspectives from Organized Labour in Canada. *Labour*, 90, 223-253.
- Ikenaga, T. & Kambayashi, R. (2016). Task Polarization in the Japanese Labor Market: Evidence of a Long-Term Trend, *Industrial Relations*, Vol.55, pp.267-293.
- Jiao, W., Wang, W., Huang, J. T., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? A preliminary study. *arXiv preprint arXiv:2301.08745*.
- Johnson, D., & Ryan, J. B. (2020). Amazon Mechanical Turk workers can provide consistent and economically meaningful data. *Southern Economic Journal*, 87(1), 369-385.
- Kanazawa, K., Kawaguchi, D., Shigeoka, H., & Watanabe, Y. (2022). AI, Skill, and Productivity: The Case of Taxi Drivers, *NBER Working Paper Series*, 30612.
- Katz, L. & Murphy, K. M. (1992). Changes in the Wage Structure, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics*, Vol. 107, No. 1 (February), pp.35-78.
- Kawaguchi, D., & Toriyabe, T. (2022). Measurements of skill and skill-use using PIAAC. *Labour Economics*, 78, 102197.
- Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3), 303-320.
- Kitahara, S. & Shinozaki, T. (2019). Do Digital Technologies Complement or Substitute for Human Labor? *ESRI Discussion Paper Series No. 351*.

- Kobayashi, T & Yamamoto, I. (2020). Job tasks and wages in the Japanese labor market: Evidence from wage functions, *Journal of the Japanese and International Economies*, Volume 58. 101110.
- Kogan, L., Papanikolaou, D., Schmidt, L. D., & Seegmiller, B. (2021). Technology-skill complementarity and labor displacement: Evidence from linking two centuries of patents with occupations. *NBER Working Paper Series*, 29552.
- Krueger, A. B. (1993). How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989, *The Quarterly Journal of Economics*, Vol. 108, 34-60.
- Kudo, T. (2005). Mecab: Yet another part-of-speech and morphological analyzer. <http://mecab.sourceforge.net/>.
- Kuroki, Y., Manabe, T., & Nakagawa, K. (2023). Fact or Opinion? - Essential Value for Financial Results Briefing. *14th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. 375-380.
- Lane, M., & Saint-Martin, A. (2021). The impact of artificial intelligence on the labour market: What do we know so far? *OECD Social, Employment and Migration Working Papers*, 256.
- Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *New England Journal of Medicine*, 388 (13), 1233-1239.
- Levy, F. and R.J. Murnane, (1996). With What Skills Are Computers a Complement? *American Economic Review*, Vol. 86, pp. 258-262.
- Lou, B., Sun, H., & Sun, T. (2023). Gpts and labor markets in the developing economy: Evidence from China. *Available at SSRN 4426461*.
- Loughran, T., & McDonald, B. (2020). Textual analysis in finance. *Annual Review of Financial Economics*, 12, 357-375.
- Lu, Y. & Zhou, Y. (2021). A review on the economics of artificial intelligence, *Journal of Economic Surveys*, Volume35, Issue4, September 2021, p. 1045-1072.
- Machin, S. & Van Reenen, J. (1998). Technology and Changes in Skill Structure: Evidence From Seven OECD Countries, *The Quarterly Journal of Economics*, Vol. 113, No. 4, pp. 1215-1244, November

- Mann, K., & Püttmann, L. (2023). Benign effects of automation: New evidence from patent texts. *Review of Economics and Statistics*, 105(3), 562-579.
- Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). *A future that works: AI, automation, employment, and productivity*. Technical Report McKinsey Global Institute.
- Meindl, B., Frank, M. R., & Mendonça, J. (2021). Exposure of occupations to technologies of the fourth industrial revolution. *arXiv preprint arXiv:2110.13317*.
- Michaels, G., Natraj, A. & Van Reenen, J. (2014). Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years. *Review of Economics and Statistics*, 96(1) , pp.60-77.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Milanez, A. (2023) "The Impact of AI on the Workplace Evidence from OECD Case Studies of AI Implementation," OECD Social, Employment and Migration Working Papers, OECD Publishing, Paris, Vol. 289.
- Mondolo, J. (2022). The composite link between technological change and employment: A survey of the literature, *Journal of Economic Surveys*, Volume 36, Issue 4, September 2022, pp.1027-1068.
- Montobbio, F., Staccioli, J., Virgillito, M. E., & Vivarelli, M. (2023). The empirics of technology, employment and occupations: Lessons learned and challenges ahead, *Journal of Economic Surveys*, <https://doi.org/10.1111/joes.12601>, Early View Online Version of Record before inclusion in an issue.
- Montobbio, F., Staccioli, J., Virgillito, M. E., & Vivarelli, M. (2023). Labour-saving automation: A direct measure of occupational exposure, *The World Economy*, Volume 47, Issue 1, pp.332-361.
- Morikawa, M. (2017). Firms' Expectations about the impact of AI and Robotics: Evidence from a survey. *Economic Inquiry*. Vol. 55, No. 2, April 2017, 1054-1063.

- Nedelkoska, L., Quintini, G. (2018). Automation, skills use and training. *OECD Social, Employment and Migration Working Papers*, (202).
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Available at SSRN 4375283*.
- OpenAI (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv preprint arXiv:2302.06590*.
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532-1543.
- Pizzinelli, C., Panton, A., Mendes Tavares, M., Cazzaniga, M., & Li, L. (2023). Labor Market Exposure to AI: Cross-Country Differences and Distributional Implications.
- Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)*, 1-9.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2), 3.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684-10695
- Shinozaki, T., Ihara, S., Saito, N., & Takeda, S. (2022). The Effect of AI on Wages in Japan Using Computable General Equilibrium Model. *New ESRI Working Paper No. 65*.

- Spitz-Oener, A. (2006). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics*, 24(2), pp.235-270.
- Stapleton, K., Copestake, A., & Pople, A. (2021). AI, firms and wages: Evidence from India. *Available at SSRN 3957858*.
- Takaoka, K., Hisamoto, S., Kawahara, N., Sakamoto, M., Uchida, Y., & Matsumoto, Y. (2018). Sudachi: A Japanese tokenizer for business. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, 167, 599-606.
- Webb, M. (2020). The impact of artificial intelligence on the labor market. *Available at SSRN 3482150*.
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. & Schmidt, D.C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *arXiv preprint arXiv:2302.11382*.
- Winograd, T. (1972). Understanding natural language. *Cognitive psychology*, 3(1), 1-191.
- Yang, C. H. (2022). How artificial intelligence technology affects productivity and employment: firm-level evidence from taiwan. *Research Policy*, 51(6), 104536.
- Yilmaz, E. D., Naumovska, I., & Aggarwal, V. A. (2023). AI-Driven Labor Substitution: Evidence from Google Translate and ChatGPT. *Available at SSRN 4400516*.