

AI と人間の協働社会における労働の在り方と成果分配の方法

岡村 優希^{1,*} 池田 美穂¹

概要： AI による労働代替が進行しているが、セキュリティやアクチュエータの観点から、人間による労働の必要性が再認識されている側面がある。とりわけ、LLM を活用したマルチエージェントシステムの普及により、人間と AI が協働して成果を創出することが期待される。この場合には、人間を基軸とする労働の在り方が変容すると予想されるときともに、雇用主が労働者に対して賃金を支払いさえすれば労働者の実施した労働の成果を独占しえた従来の成果分配の在り方にも一定の疑問を投げかけている。本報告では、それらについての問題提起を現状の法制度に即して明確化したうえで、既存の法理論の改善の方向性を示すことを目指す。

キーワード： 人工知能、AI、労働代替、AI に学習されない権利

The Style of Labour and the Distribution of its Results in a Collaborative Society between AI and Humans

Yuki Okamura^{1,*} Miho Ikeda¹

Abstract: While AI-driven labour replacement is ongoing, there is a growing recognition of the necessity for the labour conducted by humans from the perspectives of security and actuators and so on. In particular, with the spread of AI multi-agent systems utilizing LLM, it is expected that humans and AI will collaborate to create output that has economic value that businesses pursue. In such cases, it is anticipated that the traditional approach which takes human labour as a basis will be called into question. This paper aims to clarify the problems the current legal system has and to indicate guidance for existing legal theory.

Keywords: Artificial Intelligence, AI, Labor substitution, Workers' Rights not to be learned for AI

1. はじめに

労働分野における AI の開発・社会実装の進展に伴い、これまで人間の行っていた仕事を AI が自動的に遂行することができるようになってきた。

このような現象には、人間の仕事が AI に奪われてしまうという側面があるが、他方で、人間によって行われるべき(=AI によっては代替が困難な)仕事があることを認識させるという側面もある。これらの両側面を併せて考慮するならば、「AI vs. 人間」という単純な二項対立的構図は妥当ではなく、AI と人間が協働するという方向性が同時に存在すると言わなければならない。

もっとも、人間のみで労働を行ってきた従来の社会から AI と人間が協働する新たな社会への移行プロセスにおいては、労働者(人間)の権利・利益の侵害が生じうするため、法的な利害調整の手段を検討する必要がある。このような移行期における利害調整を経ないまま強行的に AI の開発・社会実装が進むと、短期的な視座から効率化を行うことに傾倒するという企業側の機会主義的な行動により、長期的に見ると、却って技術発展が阻害されるなどの懸念がある。例えば、詳細は後述するが、熟練工の技能を AI に学習させ

ることにつき、元となった労働者に対する保護や利益保障が行われず、整理解雇等のリスクにさらされるのであれば、当該技能の習得に係るインセンティブが確保できず、結果として技能面での進展の機会が制約されてしまうなどの事態が懸念される。

そこで、本稿では、労働分野に応用可能な AI の技術水準に照らして、AI と人間の協働社会がどのようなものであるのかの具体像を想定した上で、そのような社会が実現する過程で生じうる法的課題を明らかにしつつ、それに対する一応の対応策についての試論を展開することとする。

なお、そのような検討にあたり、著者らがこれまでにやってきた一連の研究の成果を踏まえた検討を行う旨をお断りしておく。

2. 労働の在り方の変容

(1) AI マルチエージェントシステムによる構造的代替¹

現在の AI による自動化は、特に LLM を活用したマルチエージェント・システムにより、労働に関して、企業組織における営みが構造的に代替されうるという点で、従来から存在する技術的自動化とは異なるものである。これは、

¹ NTT 社会情報研究所
NTT Social Informatics Laboratories
* yu.okamura@ntt.com

¹ 本項目は、岡村優希＝池田美穂「労働法的見地からの“AI による自動

化”の影響」信学技報 125 巻 100 号(2025 年), pp. 64-71 を基礎とするものである。

とりわけ監督者型といわれる類型において顕著である。

まずもって、LLM を活用したマルチエージェントシステムの技術的特徴を列挙すると、下記の通りである。

- プロファイリング定義によって各エージェントの行動原理を決定することで、システム全体としての集合的な統制が可能である。
- 計画を策定する能力と当該契約を実行する能力が備わっており、例えば、複雑なタスクを管理容易なサブタスクに分割した上で、各サブタスクについて連鎖的に順序を決定するとともに、実行の過程でフィードバックを得て、そのような順序を適宜見直しながら目標の実現に向けた実行を行うことができる。
- 策定した計画を実行するに際して、外部ツールを使用するかどうか、使用するとしてどのようなツールを使用するかどうかの判断を行うとともに、外部環境とのインタラクションを行うことができる。
- 過去のタスク実行に関する短期記憶と長期記憶の双方の記憶メカニズムが備わっている。これらによって蓄えられているメモリーからの知識探索を行ったり、検索拡張生成 (Retrieval Augmented Generation: RAG) を合わせて活用することで、外部の知識からの検索結果を推論過程に組み込むことができる。

これらの技術的特徴により実用性が向上しているのが、図 1 のような構造をとる管理者型のエージェントモデルである。

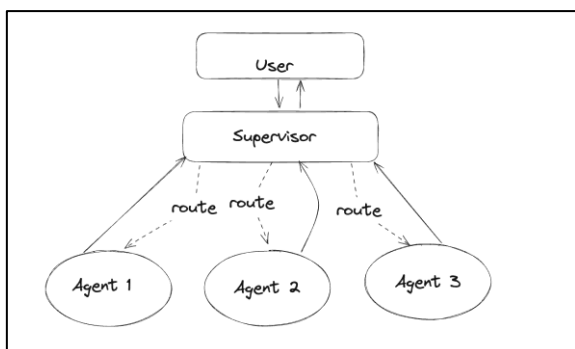


図 1：管理者型のマルチエージェントシステムの構造²

これは、中央に監督者(Supervisor)エージェントが配置され、この監督者エージェントのみが各エージェントと通

信する構造である。監督者エージェントはマルチエージェントシステムの全体の統御に特化した AI であり、各エージェントとの通信を独占し、どのタスクをどのエージェントに担当させるのかを決定する。各エージェントは、監督者エージェントからの指示を受けて、割り当てられた特定のタスクを実行する。

以上のような監督者型のマルチエージェントシステムによって、労働に関する企業の営みが構造的に置き換えられる可能性がある。

一般に、企業組織における業務執行プロセスとしては、取締役会が経営上の意思決定を行い、それを各部署に伝達し、それを受けた管理職がそれを実行するために具体的なタスクへの分割を行った上で、それぞれの実行を各従業員に対して指揮命令する、というものが想定される。従前の自動化技術は、これらのうち、特定のタスクの実行段階を代替するものとどまり、全体の統御はあくまでも人間が行っており、ここまでを自動化するものではなかった。これに対して、上記の監督者型のマルチエージェントシステムは、特定タスクの実行にとどまらず、タスク分割やタスクの割り振りも行えるので、業務執行を構造的に代替する可能性を有するものと言える。

(2) AI 作成のための労働者データの収集と法的保護：

「労働者の AI に学習されない権利」の保障範囲と実効性確保³

以上のようなエージェントシステムを構築するにあたっては、学習用データが必要となる。この点、労働分野における AI を作成するためには、労働者から職務に必要な技能や身体的特性に係るデータを、センシングデバイス、モーションキャプチャ等の技術的手段を用いて、取得することが必要となる。

もっとも、一度、労働者の技能や身体的特性をデータとして取得してしまえば、使用者は、それらの元となった労働者本人がいなくとも、事業活動を行うことができるため、それら労働者を解雇する可能性がある。また、それらのデータをもとに作成された AI が製品として流通する場合には、労働者は自らの技能等を活かした就業自体が一たとえ転職をする場合であっても一困難になるという不利益を被る懸念がある。そこで、労働者の技能等の情報が AI の学習のために利用されない権利（＝労働者の AI に学習されない権利）を保障すべきか、保障する場合にはどのような場合に保障すべきかが問題となる。

この問題を規律する専用の立法は存在しないため、既存の一般法理をもとにした規律が考えられるところ、使用者が技能情報の収集への協力を命じる場合には、そのような

² https://langchain-ai.github.io/langgraph/tutorials/multi_agent/agent_supervisor/

³ 本項目は、Yuki Okamura and Miho Ikeda, Workers' Rights not to be learned for AI: General Legal Framework Suggested from Japanese Labour Law Context (proceedings of the 24th Workshop on the Economics of Information Security) (available in [WEIS2025_paper_24.pdf](#))を基礎にしつつ、モデル・マージの技

術から生じるアイデンティティの保持に係る人格的利益の問題や、学習されない権利を侵害した場合の法的サンクションとして、原状回復義務が認められるか（認められるとすればどのようにして実現するのか）という側面からの検討を加えたものである。

指揮命令権の行使が権利濫用（労働契約法 3 条 5 項）に該当するか否が問題となる。ここで条文を見ると、労働契約法 3 条 5 項は、「労働者及び使用者は、労働契約に基づく権利の行使に当たっては、それを濫用することがあってはならない」と定めるのみで、いかなる事情があれば権利濫用性が肯定されるかは文言上明らかではない。そのため、権利濫用該当性の判断は、「労働者の有している技能等は誰に帰属すべきものであるのか」という実質的な考慮をもとに、個々の事案に即した判断が求められる。

最終的には個別判断になるが、一定の傾向を見出すことは可能であり、例えば、使用者による技能習得に対する投資が少ない一方で、労働者による技能習得への貢献度合いが高い場合には、権利濫用性が肯定されやすいと考えられる。他方で、使用者が技能習得に対して多額の投資をする一方で、技能習得に対しての労働者の貢献度合いが低い場合には、権利濫用性は否定されやすい（したがって、労働者は技能等のデータ収集に協力することが労働契約によって義務付けられ、これに違反した場合には解雇や懲戒処分の対象となりうる）ものと考えられる（図 2）。

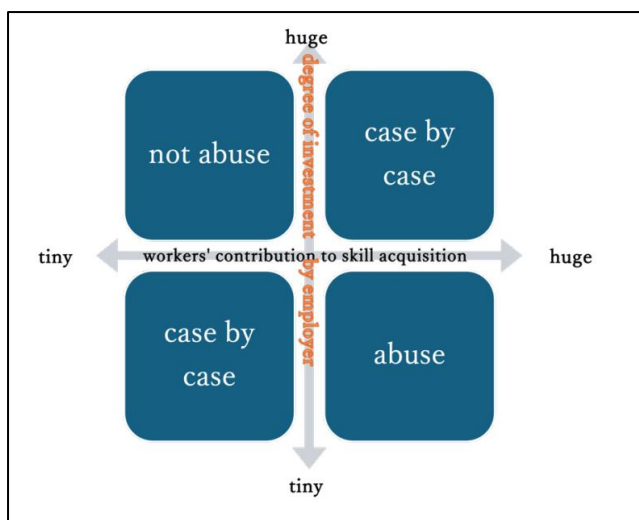


図 2 権利濫用性の判断傾向の分析枠組みの一例

ここでは、さらに、技能等にかかるデータが収集され利用することで生じる労働者の不利益が問題となる。というのも、データ収集の目的が、データ収集を命じる旨の指揮命令権の行使自体の権利濫用性の評価にも影響すると考えられるからである。

この点、解雇や労働条件の引下げ等が行われる場合には、労働者の経済的利益が問題になるが、これに加えて、自らの技能を活かして就労する機会を奪われること自体から精神的な損害が生じたりもする。この場合には、労働者の人

格的利益が問題となる。

このような人格的利益の問題は、近年、技術開発の進んでいるモデル・マージ (Model Merge) という手法によって、多様化する可能性がある。すなわち、例えば、熟練工やデザイナー等が自らの技能を他者のものと区別することで自己同一性を保持している場合には、自らのデータを用いてファインチューニングされたモデルに、他者のデータを用いてファインチューニングされたモデルをマージすると、上記の自己同一性が損なわれることによって、人格的利益の問題が生じうるのである。

以上のように、権利濫用性を評価する上では、経済的利益と人格的利益の双方を適切に考慮することが必要である。

その上で問題となるのが、仮にデータ収集への協力を要求する旨の指揮命令権の行使が権利濫用に該当する場合、いかなるサクションが（労働者の側から見ると、いかなる救済が）認められるのかという点である。

この点、民法 121 条の 2 第 1 項は、「無効な行為に基づく債務の履行として給付を受けた者は、相手方を原状に復させる義務を負う」と規定しているため、使用者には、当該技能等のデータを削除すること（＝データ取得前の状況に戻すこと）が求められる。もっとも、民法 412 条の 2 第 1 項では、「債務の履行が契約その他の債務の発生原因及び取引上の社会通念に照らして不能であるときは、債権者は、その債務の履行を請求することができない」と定められており、原状回復義務が生じない可能性を認めている（この場合には、債務不履行を理由とする損害賠償請求が認められうる（民法 412 条の 2 第 2 項、民法 415 条）。

とりわけ、LLM についてみると、学習データのアンラーニング (unlearning)⁴の実現には困難が伴うため、原状回復義務が否定されやすいとも考えられる。すなわち、アンラーニングされた状態を、当該データをデータセットから視て学習したモデルの状態として定義すると、再学習を要求することになるが、膨大な学習データのうち、特定のデータを除くために莫大なコストを要求する点で、「社会通念に照らして不能である」との評価を受ける可能性がある。もっとも、モデル自体の修正を行わずに、モデルの出力結果から当該データの影響を除くという方向性での研究も行われている。その技術的發展の程度によっては、出力結果から自身のデータの影響を除くことをもって原状回復がなされたと評価する余地があり、それが過度な負担を強いるものではないほどまでに普及する場合に、そのような技術の利用をもって原状回復がなされたと評価する方向性も想定される。

(3) Human in the Loop の考え方

もっとも、上記のようにして AI エージェントを作成し

4 アンラーニング技術については、Jiahui Geng et al., A Comprehensive Survey of Machine Unlearning Techniques for Large Language

Models(<https://arxiv.org/abs/2503.01854>)等を参照。

たとしても、人間による労働が直ちに不要になるわけではない。例えば、ソフトウェア上の出力を現実世界で再現するためのアクチュエータの技術水準が十分でないため、アクチュエータとしての人間が必要となる場面がある。また、セキュリティの関係で、人間による操作を意図的に介在させるべき場合や、コンプライアンス関係など、出力結果に対する信頼性の要求が高い場合にも、人間をマルチエージェントシステムの一部として組み込むという発想が必要となる。

このような形で、管理者の AI エージェントの指揮監督の下で、人間が単独のエージェントとして、又は、各 AI エージェントをサポートする者として、機能することが想定される。これが、現状の技術的水準を前提とした場合に想定される AI と人間の協働社会の在り方である。

その実現の過程で、上述の通り、労働者の権利利益の侵害が生じうるという課題がある。この点について、本稿で検討したような内容で一定の対処を行わなければ、長期的な視点では、却って技術や技能の発展が阻害される懸念がある。例えば、労働者の学習されない権利に対する配慮がないまま、データの取得を無制約に認め、元となったオリジナルの労働者を整理解雇等のリスクにさらすのであれば、これから熟練工になるべき労働者に対して、当該技能の習得に係るインセンティブが確保できず、結果として技能面での進展の機会が制約されてしまう事態が想定される。

このような移行期における配慮は、AI と人間の協働社会を実現する上で必須のものである。また、Human in the Loop の考え方のもと、一部の仕事が人間に留保されるとはいえ、現在の人間の仕事が構造的に代替されうるとなると、AI による労働が事業活動の中核に据えられる可能性がある。その場合でも、過去のデータからの学習を行うことで作成される AI によって、新たに生じうる問題への対処が行えるのかについて、一定の疑問が存在する以上、人間の労働を通じた技術的発展の可能性を残すことが重要であり、そのためのインセンティブを確保するという観点からも、労働者の AI に学習されない権利についての検討が要請されるものと解される。

3. 労働の成果分配の在り方の変容

(1) 現状の分配の方法：一般雇用原則を基軸とする方法

労働を通じた成果創出に関する対価については、賃金でこれが賄われているとする考え方を一般雇用原則⁵という。この原則が妥当する限りにおいて、使用者は、労働者に賃金を支払ってさえいれば、労働者の創出する成果を適法に保持することができるということになる。

他方で、この原則の妥当する範囲を超える例外的な場面においては、賃金支払いをもって、労働の成果を適法に保持することはできない。その典型例が職務発明であり、特許法 35 条は、特許を受ける権利の法人への原始的帰属を容認しつつも、その場合には労働者に対して「相当の利益」を与えるべき旨定めている。この利益は賃金とは異なるものであるので、特許を受ける権利は一般雇用原則の枠外のものとして扱われているということが含意されている。

(2) 現状の分配方法の限界と将来展望

この点、労働者の技能データを用いた事業活動から生じる利益はどうか。これは、労働者の労働を代替的に行わせるものであり、当該労働者の労働から生じる利益自体は引き続き使用者が享受しつつも、労働者本人を不要とする点で、労働義務の履行そのものとは厳然たる相違点を有するものである。

そして、技能等のデータ収集やそれをもとに作成された AI の開発・利用から生じる不利益への対処は、上述の通り、労働者の AI に学習されない権利によって行われるものであり、その権利を侵害した場合のサンクションとして、現状回復が不能である場合には、損害賠償請求が認められる。このような損害賠償請求を通じて、利益に対する金銭的な補償を行うという方向性も想定されなくはない。

しかし、我が国における損害賠償制度は、いわゆる填補賠償を旨としているため、上記のようなスキルデータに対する利害調整システムとして十分に機能しないことも想定される。その場合には、特許法 35 条のような立法が要請されることも考えられる。

労働分野における AI の開発・利活用が進展する中、AI と人間が共存する社会に向けて、移行期にある現在において検討すべき法的課題は山積している。本稿で検討した内容について、引き続き理論的な検討を進めるとともに、技術によって実現される協働社会の在り方についても、技術発展の水準を見ながら、さらに検討を深めていきたい。

参考文献

脚注に掲げたもの。

⁵ 一般雇用原則については、横山久芳「職務上作成される創作物の権利の帰属について」日本工業所有権法学会誌 39 号(1978 年)185 頁以下、土田道夫「職務発明・職務著作と労働法の規律」日本労働法学会誌 132 号 (2019

年) 52 頁以下参照。