

## 人工知能（AI）技術を活用する ESG 評価の現状と課題

富永 健司

### ■ 要 約 ■

1. 近年、人工知能（AI）に関連する技術（AI技術）の進展に伴い、資産運用における AI 技術の活用が進んでいる。昨今拡大が続いている環境・社会・ガバナンス（ESG）投資においても、企業の ESG に関連する取り組みを評価する ESG 評価の分野で AI 技術が活用されている。
2. AI 技術を活用する ESG 評価機関としては、スイスを本拠として ESG 評価・データを提供する RepRisk、米金融情報大手であるファクトセット（FactSet）傘下の Truvalue Labs、資産運用事業を中核としてサステナビリティ金融事業を推進するアラベスク・グループでドイツを本拠として ESG 評価事業を行うアラベスク S-Ray、等が挙げられる。
3. 伝統的な ESG 評価と比較した AI 技術を活用する ESG 評価の特徴として、①評価の担い手、②更新頻度、③主な情報ソース、等についての違いがある。AI 技術を活用する ESG 評価については、評価のプロセスにおいてアナリスト等の関与が総じて少なく、メディアニュース等の情報を比較的迅速に反映する特徴があることから、特に情報の信頼性・適切性の確保が重要となると考えられる。
4. 他方、ESG 評価における AI 技術の活用は、企業のサステナビリティに関する慣行及び行動について通常すぐには入手・分析しにくいような情報等を明らかにし、グリーンウォッシュ等のリスクを軽減することにつながるとの見方もある。ESG 分野における AI 技術の活用が、企業による ESG に係る取り組みの適切な把握につながり、ESG 投資の健全かつ持続的な発展に寄与するのか注目されよう。

### 野村資本市場研究所 関連論文等

- ・富永健司「ESG 投資の拡大と市場参加者に求められる ESG 評価の理解—ESG 評価の特徴と評価間のばらつき事例—」『野村サステナビリティクォーターリー』2021 年秋号。
- ・富永健司「重要性が高まる ESG 評価の現状と課題」『野村サステナビリティクォーターリー』2021 年冬号。
- ・佐藤広大「人工知能・ビッグデータを活用した資産運用への期待と課題」『野村資本市場クォーターリー』2017 年春号。

## I はじめに

近年、人工知能（AI）に関連する技術（AI 技術）の進展に伴い、資産運用における AI 技術の活用が進んでいる<sup>1</sup>。昨今拡大が続いている環境・社会・ガバナンス（ESG）投資においても、企業の ESG に関連する取り組みを評価する ESG 評価の分野で AI 技術が活用されている。

例えば、世界最大規模の年金基金である年金積立金管理運用独立行政法人（GPIF）は、スイスを本拠として ESG 評価・データを提供する RepRisk、米金融情報大手であるファクトセット（FactSet）傘下の Truvalue Labs、資産運用事業を中核としてサステナビリティ金融事業を推進するアラベスク・グループでドイツを本拠として ESG 評価事業を行うアラベスク S-Ray、といった AI 技術を活用する ESG 評価機関のデータに基づいた各種分析の実施等、資産運用・リサーチにおける AI 技術の活用を進めている<sup>2</sup>。

しかし、ESG 評価の分野における AI 技術の活用が、伝統的評価機関<sup>3</sup>による ESG 評価との間にどのような違いを生み、既存の ESG 評価の課題である ESG 評価間のばらつきにどのように影響するかといった点については必ずしも見方が定まっているわけではない<sup>4</sup>。

本稿では、AI 技術を活用する ESG 評価の特徴について述べ、評価機関の代表例として、RepRisk、Truvalue Labs、アラベスク S-Ray の評価手法を説明する。その上で、伝統的 ESG 評価機関と評価の枠組みが類似する<sup>5</sup>アラベスク S-Ray の ESG 評価を取り上げて、評価の特徴、伝統的 ESG 評価機関の評価とのばらつきの状況等を示し、AI 技術を活用する ESG 評価の課題と今後の注目点について論考する。

## II AI 技術を活用する ESG 評価機関の特徴

ESG 評価は、1980 年代に社会的責任投資（SRI）に関する調査・評価機関が設立されたことに端を発すると言われている<sup>6</sup>。その後、信用格付会社、指数提供機関、金融・経済情報会社等を中心に、ESG 評価機関の再編が進んだ<sup>7</sup>。このような中、AI 技術を活用する代表的な ESG 評価機関である RepRisk は 2006 年より、AI 技術をアナリストの知見と組み合わせ、リサーチ、分析、リスク指標の算出に活用し始めた<sup>8</sup>。そして、Truvalue Labs

<sup>1</sup> 詳しくは、佐藤広大「人工知能・ビッグデータを活用した資産運用への期待と課題」『野村資本市場クォーターリー』2017年春号、を参照。

<sup>2</sup> 年金積立金管理運用独立行政法人「平成 29 年度 ESG 活動報告」2018 年 8 月 13 日、27-28 頁、年金積立金管理運用独立行政法人「2018 年度 ESG 活動報告」2019 年 8 月 19 日、41-42 頁、アラベスク S-Ray ウェブサイト。

<sup>3</sup> 伝統的な ESG 評価機関として、ブルームバーグ、リフィニティブ、V.E、サステナリティクス、S&P グローバル、FTSE Russell、MSCI ESG Research、ISS ESG 等が挙げられる。

<sup>4</sup> 水口剛「『水口教授のヨーロッパ通信』AI と ESG 評価」QUICK ESG 研究所 2017 年 6 月 12 日。

<sup>5</sup> アラベスクについては、ESG 関連の総合スコア及び個別スコアを算出する等、伝統的 ESG 評価機関と ESG 評価の算出の枠組みが類似している。

<sup>6</sup> EIRIS, Ethical Investment Research Services, “The state of responsible business: Global corporate response to environmental, social and governance (ESG) challenges Executive summary version,” September, 2007, p.2.

<sup>7</sup> 詳しくは、富永健司「重要性が高まる ESG 評価の現状と課題」『野村サステナビリティクォーターリー』2021 年冬号、を参照。

<sup>8</sup> RepRisk, “RepRisk methodology overview.”

が 2013 年より、企業のリスク及び機会を明らかにすることを目的として、AI 技術を活用した金融サービスを開始した。その後、金融分野において分析ツール、データ、テクノロジーに関するソリューションを提供するファクトセットが 2020 年、顧客からの ESG に係る迅速な情報へのニーズが高まったこと等を背景に、Truvalue Labs を買収した<sup>9</sup>。

ESG データに関するニーズが高まる中、英金融大手のバークレイズから独立したアラベスク・グループにおいて、ESG リサーチの社内ツールとして開発された S-ray が 2018 年、アラベスク S-Ray として ESG 評価事業を開始した。

伝統的な ESG 評価機関においては、データの判別・分析・評価等の ESG 評価のプロセスにおいて、アナリストが中心的役割を担っている。他方、AI 技術を活用する ESG 評価機関においては、これらのプロセスにおいてデータに基づき構築されたモデルが中心となる（図表 1）。

具体的には、例えば、メディアニュースの内容について、AI 技術の代表的手法である自然言語処理技術（NLP）<sup>10</sup>を通じて、ESG との関連性が肯定的か否定的かといった判断や、不祥事情報の抽出が実施されることが多い。こうした点から、AI 技術を活用した ESG 評価は企業の ESG に係る取り組みに対する市場のセンチメントを分析するものとも考えられる<sup>11</sup>。

伝統的 ESG 評価機関の評価の更新頻度は、半年に 1 度又は年次であることが一般的である。重大なイベントが発生した場合については適宜更新が行われているものの、評価の更新についての迅速性の観点で課題が指摘されている<sup>12</sup>。

これに対して、AI 技術を活用する ESG 評価機関による評価の更新は、日次が一般的であり、相対的に高頻度となっている。情報ソースは、企業による開示情報の他、メディアニュース等を積極的に取り入れていることが特徴である。このことにより、AI 技術を活用する ESG 評価機関による ESG 評価は、迅速性及び正確性の課題を軽減することにつながるとの見方が示されている<sup>13</sup>。

図表 1 伝統的 ESG 評価機関と AI 技術を活用する ESG 評価機関の比較

	伝統的 ESG 評価機関	AI 技術を活用する ESG 評価機関
評価の担い手	アナリストが中心的役割	一定のアルゴリズムに基づくモデル及びアナリストによる評価
更新頻度	一般的には、半年に一度、年次 (重大なイベントが発生した場合は、追加的な更新が実施)	一般的には、日次
主な情報ソース	開示情報、アンケート調査、ニュース等	開示情報、ニュース等 (開示情報が含まれないケースもある)

(出所) 野村資本市場研究所作成

<sup>9</sup> FactSet, “FactSet enters into agreement to acquire Truvalue Labs,” October 20, 2020.

<sup>10</sup> 人間の言語の分析を行う技術であり、テキストデータの分析等に活用される。

<sup>11</sup> S&P Global, “How can AI help ESG Investing?,” February 25, 2020.

<sup>12</sup> 年金積立金管理運用独立行政法人「平成 29 年度 ESG 活動報告」2018 年 8 月 13 日、27-28 頁。

<sup>13</sup> London School of Economics, “Using AI to screen, search, and structure environmental, social, and governance data,” November 28, 2021.

### Ⅲ AI 技術を活用する ESG 評価機関の事例

AI 技術を活用する代表的な ESG 評価機関として、RepRisk、Truvalue Labs、アラベスク S-Ray が挙げられる（図表 2）。本書ではこれらの ESG 評価機関の評価手法と、伝統的評価機関と比較した際の AI 技術を活用する ESG 評価機関の評価の特徴である、①評価の担い手、②更新頻度、③情報ソース、等についての比較結果を示す。

図表 2 AI 技術を活用する ESG 評価機関の事例

評価名	RepRisk		Truvalue Labs				アラベスク S-Ray	
	RepRisk Index	RepRisk Rating	インサイト・スコア	パルス・スコア	モメンタム・スコア	ボリューム・スコア	GC スコア	ESG スコア
内容・目的	企業の ESG 関連のレピュテーション（風評）リスクエクスポージャー	RepRisk Index に国・セクターの ESG リスクを組み合わせて算出	ESG に係る長期の実績	短期的なパフォーマンスの変化	長期の ESG 行動の傾向	過去 1 年の情報量及び記事数	レピュテーションリスクを測る規範的な枠組み	長期的な財務パフォーマンスを測る指標
スコアの範囲	0～100	—	0～100	0～100	0～100	記事等の絶対数	0～100	0～100
定性評価	Low から extremely high の 5 段階	D から AAA の 10 段階	—	—	—	—	—	—
対象企業	世界の上場企業及び非上場企業 185,000 社超（上場企業の割合は約 15%）		世界の約 26,000 社				世界の約 8,000 社	
情報ソース	23 言語、10 万件超の公表ソース（企業による開示情報は考慮していない）。活字メディア、オンラインメディア、ソーシャルメディア、政府機関、規制当局、シンクタンク等		13 言語、10 万件超のソース（企業による開示情報は考慮していない）				・ 非財務情報開示に基づく 250 種類超の指標 ・ 公表ニュースを基にしたサステナビリティ関連の不祥事（170 か国超を対象とする 3 万件超のニュースソース） ・ NGO 活動	
AI 技術の活用とアナリスト等の関与	リサーチのデータのインプットは、機械学習及びアナリストの分析等により実施		NLP により ESG 関連課題の抽出、ESG との関連性を判別。評価においてアナリスト等の関与は原則行われない				公表ニュースよりサステナビリティ関連の不祥事を抽出。評価においてアナリスト等の関与は原則行われない	
更新頻度	新たなリスク情報が公表された場合に評価が更新。データは日次で更新		評価は日次で更新。データはイベントが発生した都度更新				日次で更新	

（出所）RepRisk, “RepRisk methodology overview”、FactSet, “At a glance: Truvalue Labs insight360 datafeed”、Arabesque S-Ray, “Arabesque S-Ray Methodology (v.2.6),” September 2020、各種資料より野村資本市場研究所作成

## 1. RepRisk

RepRiskは1998年、スイスを本拠として設立された。RepRiskは2006年より、AI技術・機械学習<sup>14</sup>とアナリストの知見を組み合わせることで公表情報を分析し、ESG 関連のリスク指標を提供している。RepRisk による主な企業に対する ESG 評価として、①RepRisk Index (RRI)、②RepRisk Rating (RRR)、が挙げられる。

RRIは、企業のESG関連のレピュテーション（風評）リスクエクスポージャーを示す指標である。RRIは、投資及びビジネスのESGリスクについての一次的評価としての役割がある。

RRIのスコアの範囲は0~100であり、スコアが0~24で低 (low) リスク、25~49で中 (medium) リスク、50~59で高 (high) リスク、60~74で超高 (very high) リスク、75~100で極高 (extremely high) リスク、と定性評価が行われる。

RRIは、①Current RRI、②Peak RRI、③RRI Change or Trend、に分類される。Current RRIは、メディア及びステークホルダーによる企業のESG課題に関連する現状の関心の水準を示す。Peak RRIは、過去2年間の最も高いRRIを示しており、当該数値はESG及びビジネスのコンダクトリスクに関連する全体的なレピュテーションリスク・エクスポージャーの指標となる。RRI Change or Trendは、過去30日間のRRIの増減を示す。

他方、RRRは、RRIに国・セクターのESGリスクを組み合わせることで算出される。具体的には、①企業特有のESGリスク (Peak RRI)、②国・セクターに係るESGリスク (本拠地及び主要セクターの所在国のESGリスクエクスポージャー、企業のESGリスクに関連する国及びセクターについての国際ESGリスクエクスポージャー、から算出される) を組み合わせることで計算される。

RRRについては、対象となる企業はRepRisk Ratingにより、DからAAAの10段階で評価される。その上で、Dは超高 (very high) リスク、CCC・CC・Cは高 (high) リスク、BBB・BB・Bは中 (moderate) リスク、AAA・AA・Aは低 (low) リスク、と定性評価が行われる。

同評価は、国及びセクターのリスクエクスポージャーと共に、企業自身のリスクエクスポージャーを考慮して算出される。RepRisk Rating の評価対象となる企業は、上場及び非上場企業185,000社超である。このうち、上場企業は約15%である。

RepRiskは、ESG評価の算出にあたり、ESG関連のリスク事象 (risk incident) の特定を目的として、AI技術を活用して日次で23言語、10万件超の公表ソースに対してスクリーニングを行う。具体的な情報ソースは、活字メディア、オンラインメディア、ソーシャルメディア、政府機関、規制当局、シンクタンク等である。

情報ソースについては、RepRiskは、企業による開示情報を考慮していない点の特徴である<sup>15</sup>。同社はこの理由として、企業による開示は個社毎に開示状況が異なることに加え

<sup>14</sup> データを用いて、予測、分析、判別等を行うアルゴリズム (問題解決のための方法や手順) の総称。

<sup>15</sup> アウトサイド・イン手法 (outside-in approach) と呼ばれることがある。

開示の比較可能性が確保されていないことなどを挙げ、特にリスクの観点からは、開示情報を基にした ESG 評価については信頼性に疑問符がつくとの見方を示している<sup>16</sup>。

RepRisk の主な調査対象となるのは 28 の ESG 課題である（図表 3）。全ての ESG 関連のリスク事象は少なくとも一つの ESG 課題と紐づけられる。さらに、複数の ESG 課題に結び付けられた 73 種類の関連トピック（topic tag）についても評価の対象となる。

ESG 課題は、世界銀行グループの環境・健康・安全に関するガイドライン、国際金融公社（IFC）パフォーマンス・スタンダード<sup>17</sup>、エクエーター原則（赤道原則）<sup>18</sup>、経済協力開発機構（OECD）多国籍企業行動指針、国際労働条約に沿って選定され、定義されている。

ESG 関連のリスク事象は機械学習のモデルによる予測に基づき、企業、プロジェクト、セクター、国、ESG 課題等にタグ付けされる。あるリスク事象が複数の情報ソースに含まれる場合は、最も影響力のあるソースのみが活用される。

スクリーニングされた結果は、RepRisk のアナリストチームにより、①深刻度（Severity）、②情報ソースの到達範囲（Reach of the information source）、③新規性（Novelty）、について分析が実施される。深刻度については、リスク事象の結果、インパクトの範囲、発生理由が考慮され、低・中・高の水準に分類される。情報ソースの到達範囲については、現地メディア・小規模な非政府組織（NGO）・地方政府・ソーシャルメディア等については限定的（limited）、全国メディア・地域メディア・国際 NGO・州政府／中央政府／国際政府機関は中間（Medium）、少数のグローバルメディアが広範囲（High）に分類される。

ESG 関連の各リスク事象は内容の精査とシニアアナリストによる承認を経て、RepRisk のデータとして公表される。

図表 3 RepRisk の ESG 評価項目

	環境分野	社会分野		ガバナンス分野
	環境フットプリント	地域社会との関係	従業員関係	コーポレートガバナンス
個別分野 の項目	<ul style="list-style-type: none"> <li>気候変動、温室効果ガス、グローバル汚染</li> <li>地域汚染</li> <li>景観、エコシステム、生物多様性への影響</li> <li>資源の乱用及び浪費</li> <li>廃棄物問題</li> <li>動物虐待</li> </ul>	<ul style="list-style-type: none"> <li>人権侵害及び共謀</li> <li>社会への影響</li> <li>地域参画問題</li> <li>社会的差別</li> </ul>	<ul style="list-style-type: none"> <li>強制労働</li> <li>児童労働</li> <li>結社の自由及び団結権</li> <li>差別的雇用</li> <li>職場の健康及び安全</li> <li>劣悪な雇用条件</li> </ul>	<ul style="list-style-type: none"> <li>汚職、贈収賄、恐喝、マネー・ローンダリング</li> <li>役員報酬問題</li> <li>誤解を招くコミュニケーション</li> <li>詐欺</li> <li>脱税</li> <li>税務コスト最適化</li> <li>反競争的慣行</li> </ul>
分野横断 項目	<ul style="list-style-type: none"> <li>プロダクト（健康及び環境問題）</li> <li>法令違反</li> <li>製品及びサービスに関する不祥事</li> <li>サプライチェーン問題</li> <li>国際基準の違反</li> </ul>			

（出所）RepRisk, “RepRisk methodology overview.”、より野村資本市場研究所作成

<sup>16</sup> RepRisk, “Responding to criticisms of ESG ratings,” September, 2019.

<sup>17</sup> IFC の支援先企業が取り組むべき環境及び社会的リスク管理に関する責任を定めた基準。

<sup>18</sup> インフラ建設プロジェクトなど大規模プロジェクトへの融資を行う際に、環境・社会リスクを評価管理する金融業界の自主的ガイドライン。

## 2. Truvalue Labs

Truvalue Labs は、2013 年に米サンフランシスコを本拠として設立された。同社は AI 技術を活用し ESG 情報を提供した最も初期頃の企業である。同社は 2020 年 11 月に、金融情報大手であるファクトセットの傘下となった。

Truvalue Labs が提供する ESG 評価には、①企業の ESG に係る長期の実績 (track record) を示すインサイト・スコア (Insight Score)、②機会と不祥事に焦点を当てた短期的なパフォーマンスの変化を測るパルス・スコア (Pulse Score)、③長期における企業の ESG に係る行動の傾向を示すモメンタム・スコア (Momentum Score)、④過去 12 か月間の情報量又は記事数を示すボリューム・スコア (Volume Score)、の 4 種類がある。

Truvalue Labs の評価対象の企業は世界の 26,000 社超であり、13 か国語の 10 万件超のソースより情報が収集される。Truvalue Labs による ESG 評価には、①ニュース、②NGO 及びその他の監視グループ、③ブログ及び業界公表物、④ソーシャルメディア、がある。ESG 評価は、イベントが発生した都度更新される。Truvalue Labs の ESG 評価は、RepRisk と同様、企業による開示情報を考慮していない。また、評価においてアナリスト等の関与は原則行われない。

Truvalue Labs は、ニュースの判別にあたり、NLP を用いて ESG 関連課題を抽出する。抽出については単純なキーワード検索ではなく、NLP により文脈を予測し、ESG との関連性を判別する。より具体的には、関連記事を 0~100 の範囲に定量化し、50 を上回る場合をポジティブ (肯定的)、50 を下回る場合をネガティブ (否定的) と定義し、当該定義に従って記事の判別を行う。

収集された情報については、サステナビリティ会計基準審議会 (SASB)<sup>19</sup>の枠組みにより、企業価値に重要な影響を与えるものと特定される 26 種類の項目 (カテゴリー) に分類される (図表 4)。項目毎にスコアが算出された後、スコアが合算される。

図表 4 Truvalue Labs が活用する SASB の 26 種類の重要カテゴリー

	リーダーシップ 及びガバナンス	環境	人的資本	社会資本	ビジネスモデル 及びイノベーション
個別分野 の項目 (カテゴリー)	<ul style="list-style-type: none"> <li>・ビジネス倫理</li> <li>・競争行動</li> <li>・法的・規制環境の管理</li> <li>・重大事象のリスク管理</li> <li>・事故及び安全管理</li> </ul>	<ul style="list-style-type: none"> <li>・温室効果ガスの排出</li> <li>・大気の水質</li> <li>・エネルギー管理</li> <li>・水及び廃水管理</li> <li>・廃棄物及び危険物管理</li> <li>・生態学的影響</li> </ul>	<ul style="list-style-type: none"> <li>・従業員エンゲージメント、多様性及び包摂</li> <li>・労働者の健康及び安全</li> <li>・労働慣行</li> </ul>	<ul style="list-style-type: none"> <li>・人権及び地域社会との関係</li> <li>・顧客プライバシー</li> <li>・データ保護</li> <li>・アクセス及び入手可能性</li> <li>・製品の質及び安全性</li> <li>・顧客利益</li> <li>・販売慣行及び製品表示</li> </ul>	<ul style="list-style-type: none"> <li>・製品設計及びライフサイクル管理</li> <li>・ビジネスモデル耐性</li> <li>・サプライチェーン管理</li> <li>・原材料の調達・効率性</li> <li>・気候変動の物理的影響</li> </ul>

(出所) ファクトセットウェブサイト、より野村資本市場研究所作成

<sup>19</sup> SASB は 2021 年 6 月、国際統合報告評議会 (IIRC) と合併し Value Reporting Foundation (VRF) が設立された (Value Reporting Foundation, “IIRC and SASB from the Value Reporting Foundation, providing comprehensive suite of tools to assess, manage and communicate value,” June 9, 2021.)。

### 3. アラベスク S-Ray

アラベスク S-Ray は、独フランクフルトを本拠とする ESG 評価機関である。アラベスク・グループは 2013 年に、英金融大手のパークレイズからマネジメント・バイアウト (MBO) により独立し、資産運用事業を中核としてサステナビリティ金融事業を推進してきた。その後、ESG リサーチの社内ツールとして開発された S-ray が 2018 年、ESG 評価事業としてアラベスク・アセット・マネジメントよりスピノフされ、アラベスク S-Ray が設立された。

アラベスク S-Ray が提供する ESG 評価として、①GC スコア (グローバル・コンパクトスコア)、②ESG スコア、が挙げられる。GC スコアは、国際連合グローバル・コンパクト (UNGC) 原則に基づいて企業行動規範を評価するものであり、レピュテーションリスクを測る規範的な枠組みである。これに対して、ESG スコアは、企業による長期的な財務パフォーマンスを測る指標である。両スコアとも 0~100 の範囲で評価される。アラベスク S-Ray による ESG 評価の対象企業は世界の約 8,000 社である。また、評価においてアナリスト等の関与は原則行われない。

アラベスク S-Ray による GC スコア及び ESG スコアの算出は、①インプット層、②フィーチャー層、③スコア層、の 3 層の段階に分かれる。

インプット層においては、①非財務情報開示に基づく指標、②公表ニュースを基にしたサステナビリティ関連の不祥事、③NGO 活動、についての情報が日次で取得される。非財務情報開示に基づく指標については、サステナビリティ又は統合報告書等から 250 種類超の指標が対象となる。公表ニュースについては、170 か国超、3,000 件超のニュースソースが対象となっており、これらのニュースを対象にサステナビリティ関連の不祥事を抽出する目的で NLP が活用されている。

取得されたデータは、クオリティ・チェック等が行われ、取捨選択された後、内部の分類基準 (Internal Taxonomy) に従いラベル付けされる。ラベルが付与される際には、取得された情報に関して、①フォーカス、②トピック、の 2 つの観点で評価が実施される。フォーカスは、①整備 (Preparation)、②成果 (Outcome)、③ビジネス関与 (Business Involvement)、④ニュース、⑤NGO キャンペーン、の観点が含まれている<sup>20</sup>。トピックについては、フィーチャー (Feature) と呼ばれる、22 種類のサステナビリティ項目が定義される。この結果、例えば非開示情報開示に基づく指標内で、「腐敗防止規定」にはトピックとして「ビジネス倫理」が、フォーカスとして「整備」がラベル付けされる。

インプット層からのデータについては重複が数多く発生することから、フィーチャー層においては、インプット層における重複データの削減等を行い、フィーチャーに沿った整理が行われる (図表 5)。

<sup>20</sup> フォーカスの具体的な例として、整備については「人権に関する規程の有無」、成果については「現場で作業がストップする重大事故の発生回数」、ビジネス関与については「煙草関連商品の売上」、ニュースについては「個人情報の流出に関する不祥事」、NGO キャンペーンについて「海洋プラスチック汚染問題に関するキャンペーン等」。トピックの具体的な例として、「職場の多様性、エネルギー利用、取締役会の独立性」がある。



図表 5 アラベスク S-Ray による ESG 評価のサステナビリティ項目（フィーチャー）

	環境分野	社会分野	ガバナンス分野
個別分野のサステナビリティ項目（フィーチャー）	<ul style="list-style-type: none"> <li>・ 排出</li> <li>・ 環境に関するステュワードシップ</li> <li>・ 資源利用</li> <li>・ 環境ソリューション</li> <li>・ 廃棄物</li> <li>・ 水</li> <li>・ 環境管理</li> </ul>	<ul style="list-style-type: none"> <li>・ 多様性</li> <li>・ 職場の健康及び安全</li> <li>・ トレーニング及び開発</li> <li>・ 商品アクセス</li> <li>・ 地域社会との関係</li> <li>・ 商品の質及び安全性</li> <li>・ 人権</li> <li>・ 労働者の権利</li> <li>・ 報酬</li> <li>・ 雇用の質</li> </ul>	<ul style="list-style-type: none"> <li>・ ビジネス倫理</li> <li>・ コーポレートガバナンス</li> <li>・ 透明性</li> <li>・ 法廷会計学</li> <li>・ 資本構成</li> </ul>

（出所） Arabesque S-Ray, “Arabesque S-Ray Methodology (v.2.6),” September 2020、より野村資本市場研究所作成

より具体的には、フィーチャー層においては、①長期トレンドの考慮、②短期調整、が実施される。長期トレンドについては、インプット層からの非財務情報開示に基づく計測指標について、①フォーカス、②次元（類似したインプット情報の統合）、③エキスパート・インプット（特定項目についての NGO 及び学術的な知見を適宜考慮）、が考慮され合算される。

短期調整については、長期トレンドが相対的に変更の頻度が低いことを踏まえ、より高頻度で発生する、①公表ニュースを基にしたサステナビリティ関連の不祥事、② NGO 活動、に基づく指標が考慮される。

その後、スコア層において、22 種類の長期トレンドスコア（0～100）と短期調整の値を掛け合わせた後、GC スコア及び ESG スコア等の最終スコアが算出される。

GC スコアの算出においては、まず関連するフィーチャーが 4 種類の GC カテゴリーにマッピングされる（図表 6）。この際、よりネガティブな側面に焦点が当たるフィーチャーと、よりポジティブな側面に焦点が当たるフィーチャーについては区別され、パフォーマンスが低水準である際には前者が優先される。各フィーチャーの合算によりカテゴリー毎のスコアが計算された後、低水準のスコアのウェイトを上昇させる調整が実施され、最終スコアが算出される。

図表 6 アラベスク S-Ray の GC カテゴリー及びサステナビリティ項目（フィーチャー）

カテゴリー	人権	労働者の権利	環境	反腐败
サステナビリティ項目（フィーチャー）	<ul style="list-style-type: none"> <li>・ 人権</li> <li>・ 労働者の権利</li> <li>・ 職場の健康及び安全</li> <li>・ 雇用の質</li> <li>・ 多様性</li> <li>・ 商品の質及び安全性</li> <li>・ 商品アクセス</li> <li>・ 地域社会との関係</li> </ul>	<ul style="list-style-type: none"> <li>・ 労働者の権利</li> <li>・ 職場の健康及び安全</li> <li>・ 多様性</li> <li>・ 報酬</li> <li>・ トレーニング及び開発</li> <li>・ 雇用の質</li> </ul>	<ul style="list-style-type: none"> <li>・ 排出</li> <li>・ 廃棄物</li> <li>・ 環境に関するステュワードシップ</li> <li>・ 環境管理</li> <li>・ 資源利用</li> <li>・ 水</li> <li>・ 環境ソリューション</li> </ul>	<ul style="list-style-type: none"> <li>・ ビジネス倫理</li> <li>・ コーポレートガバナンス</li> <li>・ 透明性</li> </ul>

（出所） Arabesque S-ray, “Arabesque S-Ray Methodology (v.2.6),” September, 2020、より野村資本市場研究所作成

ESGスコアの算出においては、フィーチャー層のスコアは、財務的重要性（マテリアリティ）に基づくウェイト付けがされる。マテリアリティは、全ての企業について、時価総額加重及び単純平均の月次指標の過去 1、3、5 年のリターン（収益率）によって、セクター（Sector）及び業種（Industry）毎に四半期に一度評価が行われる。これにより、各企業に 12 種類のポートフォリオの指数リターンが割り当てられる。

各ポートフォリオにおいて、マテリアリティは、①静的マテリアリティ、②データベースのマテリアリティ調整、③トータル・スコア、の 3 段階で評価される。静的マテリアリティは、各ポートフォリオのベースとなるマテリアリティであり、第三者の調査・分析、業界レポート等に基づき、各ポートフォリオのフィーチャー毎に付与される。次に、各フィーチャーによって、リターンの変動が説明できる程度が考慮され、ウェイトが割り当てられる。当該数値によってウェイト付されたフィーチャースコアの合算により最終スコアが算出される。

AI 技術を活用する ESG 評価の特徴である、①評価の担い手、②更新頻度、③情報ソース、に着目すると、評価の担い手については、RepRisk がリサーチのデータのインプットを機械学習及びアナリストの分析等により実施しているのに対して、他の 2 社については、評価においてアナリスト等の関与は原則行われていないとの違いがある。更新頻度については、いずれも日次でデータ又は評価が更新される。情報ソースについては、RepRisk 及び Truvalue Labs がメディアニュース等を使用し、企業による開示情報は考慮していないのに対して、アラベスク S-Ray については、企業による開示情報を含めて評価を行っている。また、評価の枠組みについては、RepRisk が RRI 及び RRR、Truvalue Labs がインサイト・スコア、パルス・スコア等、といった ESG 情報を合算した評価を提供している一方、アラベスク S-Ray は ESG スコアについて総合スコア及び個別分野のスコアを提供しており、伝統的 ESG 評価機関と評価の枠組みが類似しているとの特徴がある。

## IV AI 技術を活用する ESG 評価の事例分析

本章では、伝統的 ESG 評価機関と評価の枠組みが類似するアラベスク S-Ray による、日経 500 種平均株価の構成銘柄の中で数値が取得可能な日本企業を対象とした ESG 評価について、①業種別評価状況、②ESG 分野毎の業種内平均値、③伝統的 ESG 評価機関による評価とのばらつきの状況<sup>21</sup>、④アラベスク S-Ray の ESG スコアと株価リターンとの関係、についての分析を実施する。

<sup>21</sup> 比較対象として、伝統的な ESG 評価機関の代表例であるブルームバーグ、S&P グローバル、サステナリティクスを取り上げる。

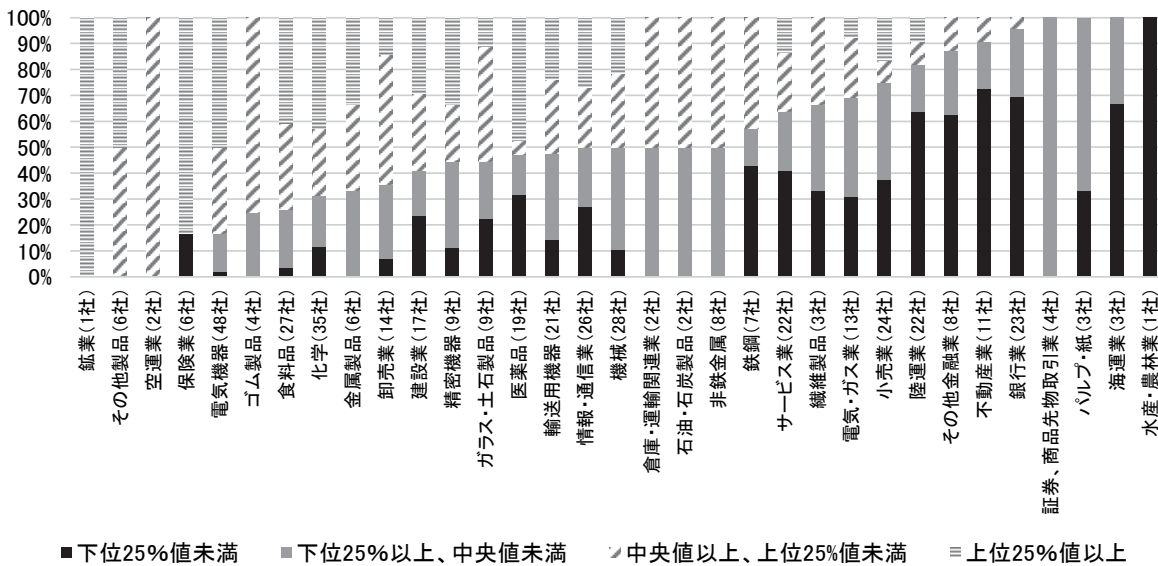
## 1. 業種別評価状況及び ESG 分野毎の業種内平均値

アラバスク S-Ray の ESG スコアの業種別評価状況について、高評価の企業が相対的に高い割合を占める業種として、その他製品<sup>22</sup>、保険業、電気機器等が挙げられる<sup>23</sup>（図表 7）。最も割合が高かったその他製品においては、ESG 分野毎の業種平均値について、全ての分野のスコアが相対的に高水準となっている（図表 8）。他方、低評価の企業が相対的に高い割合を占める業種として、海運業、パルプ・紙、証券・商品先物取引業が挙げられる。

ESG 評価が相対的に高い企業が多くを占める業種と ESG 評価が相対的に低い企業が多くを占める業種とを比べると、ガバナンスのスコアで比較的大きな水準の差が見られている。

これに加えて、ESG 分野毎の業種内平均値を見ると、環境及び社会のスコアが相対的に高水準となっている。このことから、ESG 分野の評価項目の設定等により、ガバナンスのスコアは、環境及び社会のスコアと比べて低い水準となりやすい構造となっていることが示唆される。

図表 7 アラバスク S-Ray による ESG 評価の業種別評価状況

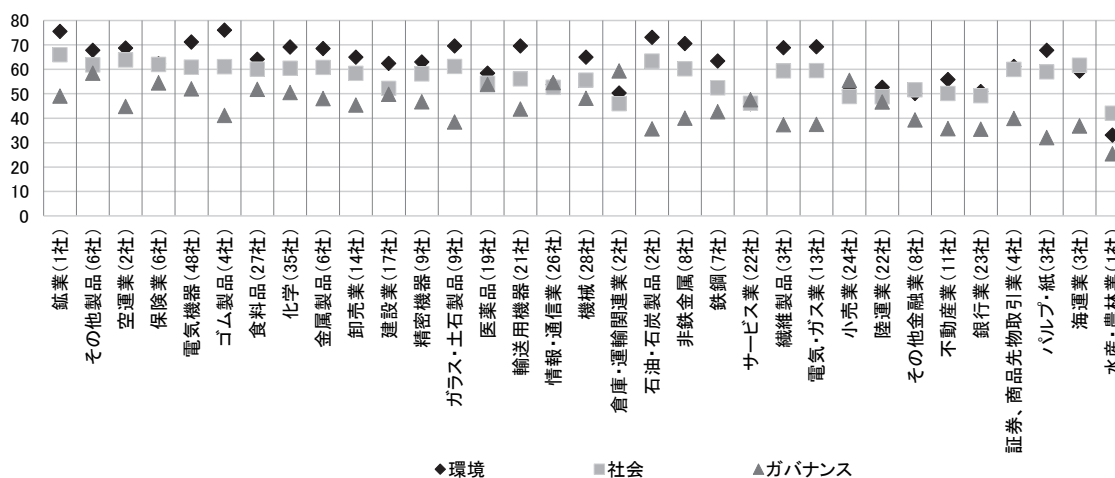


(注) 1. データは 2021 年末時点で取得可能なデータで最新だった 2021 年 2 月時点。  
対象企業は日経 500 種平均株価の構成銘柄の中で数値が取得可能な企業。括弧内は企業数。  
業種は東証 33 業種。  
2. ESG スコアについて、①下位 25%値未満、②下位 25%値以上、中央値未満、③中央値以上、  
上位 25%値未満、④上位 25%値以上、の水準における企業数の割合を示した。  
(出所) ブルームバーグより野村資本市場研究所作成

<sup>22</sup> 本分析で「その他製品」に含まれる企業は、パナソニック、凸版印刷、大日本印刷、アシックス、ヤマハ、任天堂。

<sup>23</sup> 業種分類は東証 33 業種。業種別評価状況及び ESG 分野毎の業種内平均値のグラフにおいて企業数が 3 社以上の業種を示した。

図表 8 アラベスク S-Ray による ESG 評価の ESG 分野毎の業種平均



- (注) 1. データは 2021 年末時点で取得可能なデータで最新だった 2021 年 2 月時点。  
対象企業は日経 500 種平均株価の構成銘柄の中で数値が取得可能な企業。括弧内は企業数。  
業種は東証 33 業種。  
2. 数値が取得可能な企業の平均値。スコアの範囲は 0~100。  
(出所) ブルームバーグより野村資本市場研究所作成

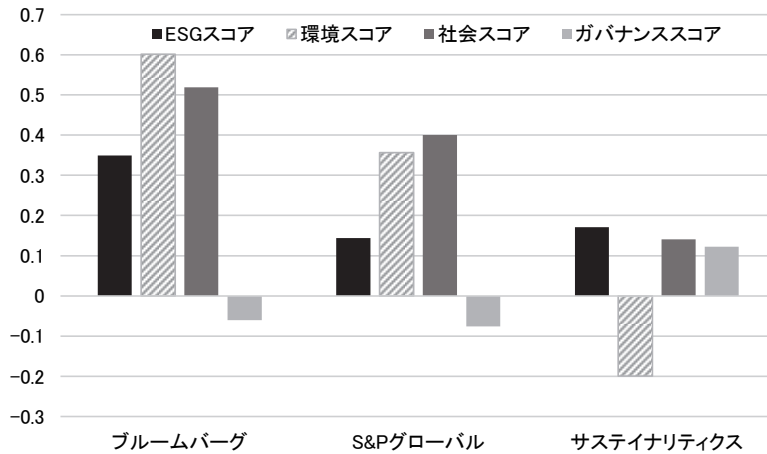
## 2. 伝統的 ESG 評価機関による評価とのばらつきの状況

アラベスク S-Ray の ESG スコアと、伝統的かつ代表的な評価機関であるブルームバーグの ESG 開示スコア、S&P グローバルの SAME ESG スコア、サステイナリティクスによる ESG Risk Rating との間のばらつきに関して相関係数を用いて確認すると、いずれの ESG 評価間についても相関は弱く、ばらつきが大きい(図表 9)。ただし、アラベスク S-Ray の ESG 各分野の個別スコアについて、環境及び社会のスコアについてはブルームバーグの開示スコアとの弱~中程度の正の相関が見られる。これらのスコアについては、サステイナリティクスとの相関が最も弱かった。これは、①サステイナリティクスが業種の環境及び社会に関わるリスクを積極的に評価に取り入れていること、②評価の際には ESG リスクの対応状況等が考慮されていることなどが関係性の低さにつながっている可能性がある<sup>24</sup>。

ただし、ガバナンスのスコアについては、ブルームバーグの ESG 開示スコアを含めていずれの評価とも関係性は見られなかった。このことから、アラベスク S-Ray のガバナンスのスコアについては、伝統的評価機関の ESG 評価と比べて独自性のある評価結果となっているとみられる。

<sup>24</sup> 伝統的金融機関の ESG 評価間のばらつき等については、富永健司「ESG 投資の拡大と市場参加者に求められる ESG 評価の理解—ESG 評価の特徴と評価間のばらつきの事例—」『野村サステナビリティクォーターリー』2021 年秋号、を参照。

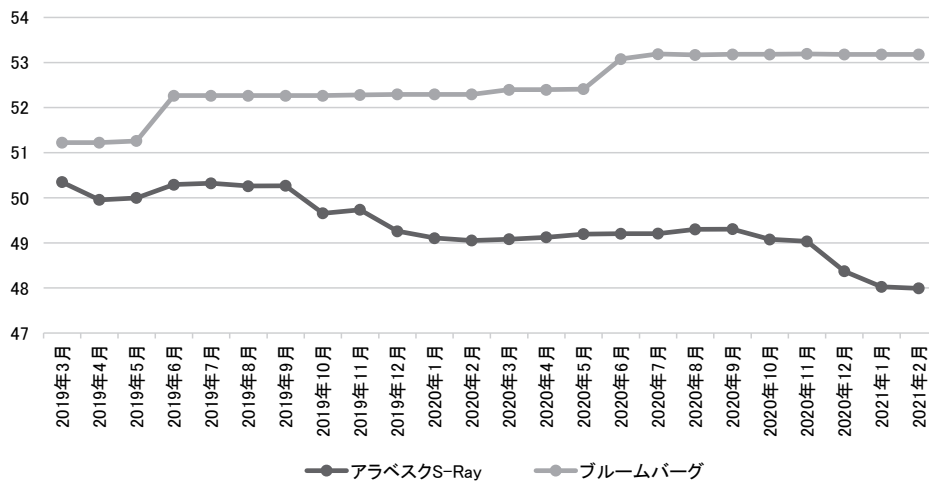
図表 9 アラベスク S-Ray と伝統的 ESG 評価機関 3 社の ESG 評価間のばらつき



(注) 1. 数値は相関係数を示す。日経 500 種平均株価の構成銘柄について、評価の水準毎の  
 ちらばりが異なること等を考慮して、ESG 評価が最も高かった企業を最大値、最も  
 低かった企業を最小値とした上で算出。  
 2. アラベスク S-Ray の ESG スコアは 2021 年 2 月時点。S&P グローバル及びサステナ  
 リティクスは 2021 年 7 月のデータ取得時点。ブルームバーグは 2021 年 5 月時点。  
 (出所) ブルームバーグ、各社ウェブサイトより野村資本市場研究所作成

アラベスク S-Ray のガバナンススコアについて、2019 年 3 月から 2021 年 2 月までの平  
 均値をブルームバーグと比較すると、ブルームバーグのガバナンススコアが上昇傾向で推  
 移しているのに対して、アラベスク S-Ray の同スコアは低下傾向にあることがわかる (図  
 表 10)。アラベスク S-Ray のガバナンススコアがブルームバーグの同スコアと相対的に異  
 なる動きをしている背景の一つとして、アラベスク S-Ray の ESG 評価における開示情報  
 以外の短期調整の項目である、公表ニュースを基にしたサステナビリティ関連の不祥事等  
 が影響している可能性がある。

図表 10 アラベスク S-Ray とブルームバーグのガバナンススコアについての比較



(注) 日経 500 種平均株価の構成銘柄の中で、アラベスク S-Ray 及びブルームバーグの  
 ESG スコアが取得可能だった企業の月次データの平均値。  
 (出所) ブルームバーグ、より野村資本市場研究所作成

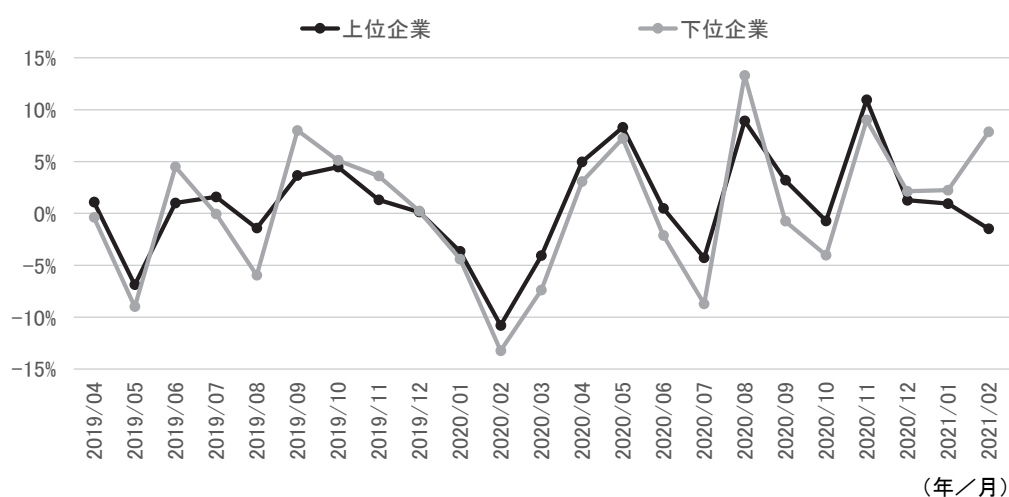
### 3. アラベスク S-Ray のガバナンススコアと株価リターンとの関係

ここでは、伝統的 ESG 評価機関と比べて相対的に独自性があり、短期調整の項目を比較的迅速に反映していると見られる、アラベスク S-Ray のガバナンススコアについて、スコアの水準と株価リターンとの関係性について簡易的な分析を実施した。具体的には、2019年3月時点においてガバナンススコアの水準が上位10%の企業と下位10%の企業について、2019年4月～2021年2月における月次平均リターンを比較した（図表11）。

その結果、ガバナンススコアが上位の企業の株価リターンは、下位の企業と比べて、株価が低下する時期と上昇する時期において、総じて変動が抑制される傾向が見られた。この点からはガバナンススコアの水準が株価リターンの変動と関係している可能性があることが示唆される。ただし、本分析の結果は、ガバナンススコアの上位企業及び下位企業の単純比較に基づくものであるため、スコアと株価リターンとの関係性についてはより詳細な分析が求められる。

本章の分析では、伝統的 ESG 評価機関による評価とのばらつきの状況について、いずれの ESG 評価間についてもばらつきが大きいことが示された。個別分野のスコアを見てみると、アラベスク S-Ray の ESG 各分野の個別スコアについて、環境及び社会スコアについてはブルームバーグ開示スコアと弱～中程度の正の相関が見られた。他方、ガバナンススコアについては、いずれの ESG 評価とも関係性は見られなかった。特に、アラベスク S-Ray のガバナンススコアについては、伝統的 ESG 評価機関と比べて、独自性のある結果であることが示唆された。このことに関連して、2019年3月から2021年2月までのスコアの平均値の推移を見ると、ブルームバーグのガバナンススコアの動きと異なる傾向

図表11 アラベスク S-Ray のガバナンススコアと株価リターン



(注) 日経500種平均株価の構成銘柄の中で、2019年3月時点のアラベスク S-Ray のガバナンススコアが上位10%と下位10%の企業の月次平均リターン。  
Y軸は、株価の前月比。

(出所) ブルームバーグ、より野村資本市場研究所作成

が見られた。さらに、アラベスクの当該ガバナンススコアを株価リターンと比較したところ、ガバナンススコアの水準が株価リターンの変動と関係している可能性があることが示唆された。ただし、この点についてはより詳細な分析が求められる。

以上の結果から、AI 技術を活用した ESG 評価機関においては、ESG に関する情報について、特に不祥事情報を含むメディアニュース等を比較的迅速に反映する指標としての有用性がうかがえる。ただし、AI 技術を活用した ESG 評価については、メディアニュースを迅速に反映するが故に、情報の信頼性・適切性をいかに確保していくかということが重要な論点となっている<sup>25</sup>。

## V ESG 評価における AI 技術の活用の進展とその課題

本稿では、AI 技術を活用する ESG 評価機関として、RepRisk、Truvalue Labs、アラベスク S-Ray の評価手法を取り上げて、伝統的な ESG 評価機関との比較、評価の特徴等を示した。AI 技術を活用する ESG 評価と伝統的な ESG 評価との相違点として、①評価の担い手、②更新頻度、③情報ソース、等が挙げられる。AI 技術を活用する ESG 評価は、企業の ESG に係る取り組みに対する市場のセンチメントを分析するものであるとも考えられる。

AI 技術を活用する ESG 評価手法の特徴や、アラベスク S-Ray の評価の各種分析結果で示唆されたように、AI 技術を活用する ESG 評価の課題としては、評価のプロセスにおいてアナリスト等の関与が総じて少なく、情報ソースにおいてはメディアニュース等の情報を比較的迅速に反映することから、特に情報の信頼性・適切性の確保が重要となると考えられる。これらの点を踏まえた上での具体的な課題として、①ESG データに関するバイアス<sup>26</sup>への対処、②評価結果についての理解度向上、が挙げられる。

ESG 評価は、従来から、①企業の規模と予算の大きさ、②マーケティング及び内容、等との関係性が深いと考えられている<sup>27</sup>。特に小規模企業の評価については、メディアニュース等についての情報に制約があることが指摘されている<sup>28</sup>。さらに、このような ESG データに関するバイアスは、データに基づいて構築されるモデルに影響をもたらす可能性もある。例えば、ESG データに関するものではないが、参考になり得る事例として、米インターネット通販最大手のアマゾン・ドット・コムが、求人活動を行う際に活用したモデルが、男性のみが働いていた過去データを使用して構築されたものであったことから、女性に対してのバイアスが含まれるものになってしまったとのケースが挙げられる<sup>29</sup>。本事例では、女子大学出身の候補者が排除されていたことが発覚したが、データのバイアスを把握し、的確に対処することが不可欠であることを示唆していると言える。

<sup>25</sup> RepRisk, “Responding to criticisms of ESG ratings,” September, 2019.

<sup>26</sup> 内容等の偏りとそのことで生じる偏った見方又は分析結果。

<sup>27</sup> Deutsche Bank Research, “Big data shakes up ESG investing,” October, 2018, p16.

<sup>28</sup> 年金積立金管理運用独立行政法人「平成 29 年度 ESG 活動報告」2018 年 8 月 13 日、28 頁。

<sup>29</sup> RepRisk, “Avoiding bias in AI: how a representative workforce empowers due diligence-grade data for financial decision-making,” July, 2021.

これらに加えて、AI 技術を活用したモデルに関連する評価結果の解釈についても課題として指摘されている。GPIF は、AI 分析における当面の課題として、スコアは AI が複雑な過程を経て算出しているものであり、その水準の解釈は簡単ではない、との見方を示している<sup>30</sup>。

GPIF は、近い将来に AI による ESG 評価・分析結果がアナリストに完全にとって代わると考えるのは難しいとの見方を示すと共に、より網羅的な ESG 評価を行うために、長期的な水準感には人間による評価を使用し、短期的な変動について AI 技術を活用した分析によって補う組み合わせが現実的である、との意見を紹介している<sup>31</sup>。AI 技術を活用する ESG 評価の課題を踏まえると、AI 技術を活用する ESG 評価について、① ESG データに関するバイアスへの対処、② 評価結果についての理解度向上、等を進めながら、伝統的評価機関による ESG 評価を補完し、企業の ESG に係るリスク及び機会を評価・可視化していくことが望まれる。

他方、ESG データに関するバイアスの問題について、AI 技術の活用は、人間の主観を排除し、むしろ伝統的な分析において生じる認知バイアスを抑制する可能性がある等との見方もある。すなわち、AI 技術を ESG 評価に活用することが、企業のサステナビリティに関する慣行及び行動について通常すぐには入手・分析しにくいような情報等を明らかにし、グリーンウォッシュ<sup>32</sup>のリスクを軽減することにつながるとの考え方である<sup>33</sup>。

AI 技術を活用する ESG 評価の課題への取り組みが進むことで、ESG 分野におけるグリーンウォッシュ等のリスクが軽減されれば、企業の ESG に係るより実効性のある取り組みを評価していくことにつながる可能性もある。ESG 分野における AI 技術の活用が、企業による ESG に係る取り組みの適切な把握につながり、ESG 投資の健全かつ持続的な発展に寄与するのか注目されよう。

<sup>30</sup> 年金積立金管理運用独立行政法人「平成 29 年度 ESG 活動報告」2018 年 8 月 13 日、28 頁。

<sup>31</sup> 前掲脚注 30 参照。

<sup>32</sup> 実際は環境改善効果がない、又は、調達資金が適正に環境事業に充当されていないにもかかわらず、環境面で改善効果があると称すること（環境省「グリーンボンドガイドライン グリーンローン及びサステナビリティ・リンク・ローンガイドライン 2020 年版」2020 年 3 月 10 日、53 頁）。

<sup>33</sup> OECD, “Artificial Intelligence, Machine Learning and Big Data in Finance Opportunities, Challenges and Implications for Policy Makers,” August, 2021, p.36.、SustainAbility, an ERM Group company, “Rate the Raters 2020: Investor Survey and Interview Results,” March 2020, p.39.