東北大学

NTT DATA

テクノロジーの質的進化に対応する組織的統制の在り方

～テクノロジーガバナンスの確立に向けて～

国立大学法人 東北大学 未踏スケールデータアナリティクスセンター
テクノロジーガバナンス共同研究部門

株式会社 NTTデータグループ グローバルガバナンス本部
Technology Governance部 AI Governance室

CONTENTS

はじめに

デジタル技術の進展は、企業経営および社会構造に不可逆的な変化をもたらしている。生成AIや大規模言語モデル（LLM）をはじめとする高度知能技術は、業務プロセスの自動化にとどまらず、意思決定の高度化、顧客体験の再設計、さらには組織の役割分担そのものにまで影響を及ぼしつつある。加えて、技術がソフトウェア領域を超えて現実世界に直接作用する領域や、人間の認知・神経機能に関与する領域へと拡張しつつあることで、その影響範囲と役割はより一層広がっている。こうした変化により、技術はもはや単なる効率化の手段ではなく、企業の競争優位および社会的信頼を左右する基盤となっている。

その一方で、新技術の導入・運用に伴うリスクは、従来のITガバナンスや情報セキュリティ管理の枠組みを超えた広がりを見せている。特に近年は、技術の自律性が飛躍的に高まり、ソフトウェアが自ら判断し行動する「Agentic AI」、物理世界に直接作用する「Physical AI」、さらには人間の認知・神経機能に接続する「ニューロテクノロジー」といった領域が実用段階に入りつつある。これらは、従来の「人間が最終判断を行う」前提を揺るがし、企業の意思決定、責任構造、リスク管理のあり方を根本から問い直すものである。

これらの技術は、「自律的意思決定」「物理空間への直接作用」「人間の内面への接続」といった新たな観点をもたらしている。この結果、責任所在の不明確化、予測困難な挙動、倫理的境界の曖昧化、さらには社会受容性の不確実性といった、新たなリスクが顕在化している。加えて、生成AIの普及により、「どのように技術を開示し、社会に受け入れられる形で利用するか」という問題も企業ガバナンスの中心的課題となりつつある。

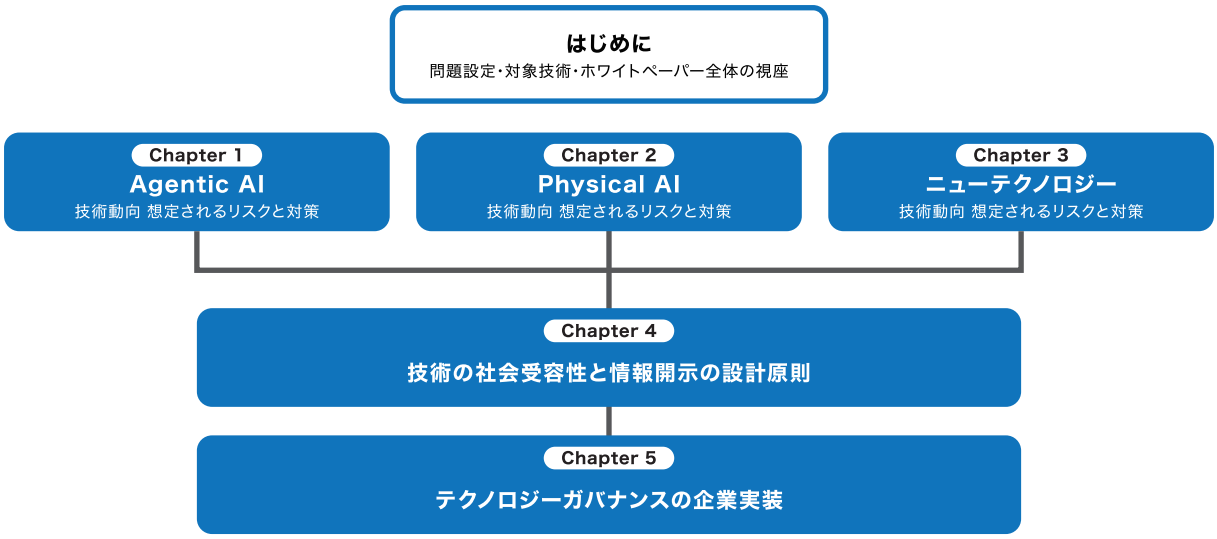
重要なのは、これらの課題を個別技術ごとに後追いで対処するのではなく、技術の特性に基づき、あらかじめ統治の枠組みを設計することである。すなわち、「何ができるか」ではなく、「どのように制御し、どの範囲で社会に適用するのか」という視点への転換が求められている。

本ホワイトペーパーでは、こうした問題意識のもと、企業にとって影響の大きい三つの技術領域（Agentic AI、Physical AI、ニューロテクノロジー）を対象に、それぞれの技術動向とリスク構造を整理する。さらに、それらを踏まえたうえで、技術の社会受容性および情報開示の設計原則を検討し、最終的に企業が実装すべきテクノロジーガバナンスの枠組みを提示する。

本稿の目的は、確定的な解答を提示することではない。経営層、リスク管理部門、情報システム部門、事業部門が共通の視座を持ち、新技術導入を「実装」の問題から「統治」の問題へと引き上げるための出発点を提供することである。本稿が、持続可能な技術活用と社会的信頼の両立に向けた議論の一助となれば幸いである。

本ホワイトペーパーの構成と検討アプローチ

本ホワイトペーパーは、先進技術の社会実装におけるガバナンスの課題を体系的に整理するため、三つの技術領域に関する個別検討と、それらに共通する横断論点の検討を組み合わせた五章構成を採用する。全体の流れは、技術の特性理解、リスク構造の把握、社会受容性の検討、そして企業におけるガバナンス実装へと段階的に進む構成としている。



本ホワイトペーパーの構成

まず中核となるのは、以下の三つの技術領域である。

- Agentic AI（自律的に計画・実行・修正を行うソフトウェア主体）
- Physical AI（ロボティクスを通じて物理世界に作用するAI）
- ニューロテクノロジー（神経活動の測定・調整を通じて人間の認知・行動に関与する技術）

これらは、いずれも「自律性」「実世界への影響」「人間への直接的影響」という観点で従来技術とは質的に異なる特性を有しており、テクノロジーガバナンスの再設計を迫る代表的領域である。

これら三つの技術領域については、すべて共通の構成で整理する。すなわち、各章においてまず「技術動向」を概観し、その上で「想定されるリスクと対策」を論じる。この共通構造により、技術ごとの差異だけでなく、リスクの共通構造（自律性・予測困難性・責任の分散など）を横断的に把握することを可能にする。

次に、これら技術の横断的課題として、「技術の社会受容性と情報開示」の章を設ける。本章では、生成AIを中心とした実例をもとに、技術そのものではなく「どのように社会に提示され、理解されるか」が評価や信頼に与える影響を分析する。ここでは、帰属、努力認知、文脈整合性、リスク認知といった認知的要因を整理し、情報開示を単なる義務ではなく、社会的評価を設計する統治行為として位置付ける。

最後に、これらの議論を踏まえ、「テクノロジーガバナンスの企業実装」の章において、企業がどのように統治を制度化すべきかを提示する。本稿では、

- ガバナンス要否の判断
- 守りのガバナンス（リスクマネジメントの実装）
- 攻めのガバナンス（事業環境・競争力の強化）

という三段階モデルを提示し、個別技術対応にとどまらない横断的な経営機能としてのガバナンスを定義する。

なお、本ホワイトペーパーは網羅的な規制解説や確定的な解答を提示するものではない。新興技術を取り巻く環境は急速に変化しており、企業に求められるのは固定的な正解ではなく、変化を前提に議論し、学習しながら統治を進化させる能力である。本稿は、そのための論点整理と実務的な出発点を提供することを目的とする。

次章以降では、まずAgentic AIを取り上げ、自律的ソフトウェア主体が企業活動に組み込まれる際の統制上の課題について検討する。

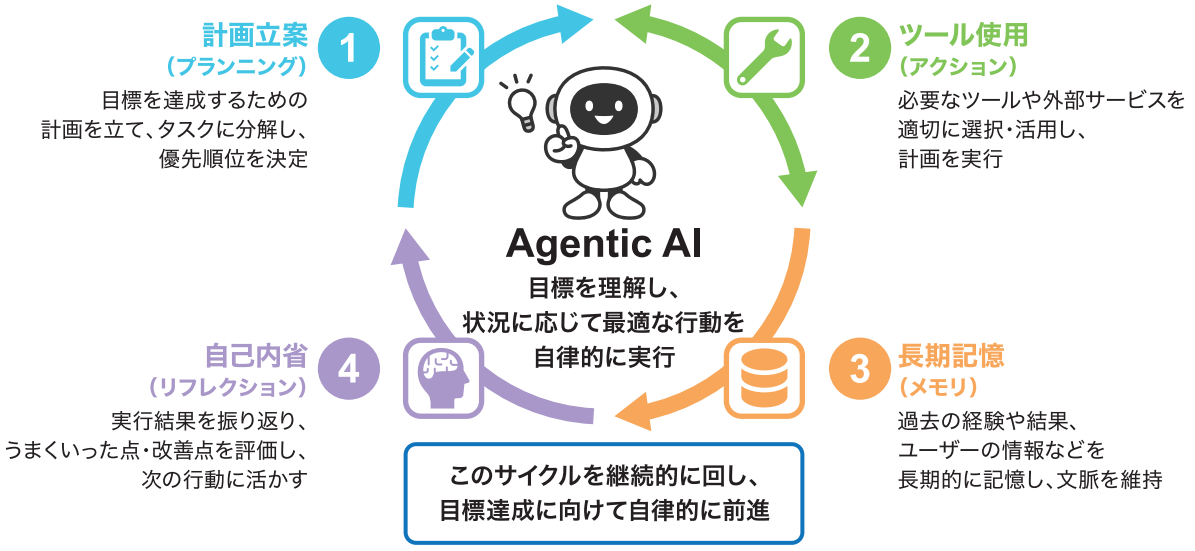
CHAPTER.1
Agentic AI

1. 技術動向

Agentic AIとは、単なる対話モデルにとどまらず目標を持って自律的に計画・行動・評価・修正を繰り返すよう設計された人工知能である。ユーザーが高レベルな目標を設定すると、Agentic AIは目標達成までの手順を自ら立案し、外部ツールやシステムを呼び出してタスクを実行し、その結果を評価・学習して行動を更新する。従来のようにLLM（大規模言語モデル）単独で利用する場合、ユーザーの入力に対してLLMが回答や文章を生成し、ユーザーがそれを活用する行動を実行していたのに対し、Agentic AIは自ら判断し、目標達成に向けた一連の行動シーケンスを主導するため、人手を介さずに複雑なタスクの自動化が可能となる。

Agentic AIの中核をなすのは計画立案、ツール使用、長期記憶、自己内省といった高度な知的能力である。具体的には、(1) 長期記憶（過去の経験を外部ストレージに保持し、長期的な学習や文脈理解に活用）、(2) 外部ツール連携（APIやデータベース、クラウドサービス、Webブラウザなどをモジュール的に呼び出し世界に働きかける）、(3) 再帰的プランニング（複雑な目標をサブタスクに分解し、順序立てて実行する能力）、(4) 内省的推論（自らの計画や行動を振り返り、必要に応じて修正するメタ認知的能力）、といった要素が挙げられる。これらを組み合わせることで、LLMを中核に据えつつもツールや環境との相互作用を通じてタスク達成プロセスを自己完結できる高度なAgentic AIが実現されつつある。

一方で、Agentic AIはその特性により、従来のLLMと比較して多様なタスクを自律的に実行するため、リスクも従来のLLMと比較して多様になる。



Agentic AIの基本的な振る舞い

2. 想定されるリスクと対策

Agentic AIの台頭は、従来の情報セキュリティモデルでは想定されていなかった新たなリスクを生んでいる。Agentic AIによるリスクが生じる原因として、以下の二つの類型に整理できる脆弱性がある。

脆弱性分類	第一の脆弱性： 従来型の情報セキュリティにおける 脆弱性の延長	第二の脆弱性： LLM特有の推論・意思決定プロセスに 起因する新たな脆弱性
説　明	既存のICTシステムで問題となってきた 脆弱性が、AIエージェントの特徴により 拡大・複雑化したもの	LLM自体の不確実性や自律性ゆえに 生じる想定外の挙動に由来する脆弱性
具体例	- 外部API等の不正呼び出し - 多様な情報ソースと不十分な入力検証 　によるインジェクション攻撃 - セッション管理ミス - メモリ共有やデータ汚染 - サンドボックスからの逸脱 - モデル内部の機密情報漏えい	- ユーザー指示の誤解 - 過剰な権限行使 - プロンプトインジェクションによる 　行動指針の書き換え - 安全性を無視した探究的行動 - 欺瞞的な振る舞い - マルチエージェント間の相互作用問題
対　策	従来のセキュリティ対策が基本となるが、 エージェントが連携・利用する環境の どこにリスクや脆弱性があるかを把握 することが重要	LLMの予測性能向上だけでなく エージェントの自律性の制限や、 動作中の監視と制御等の設計原則および 運用ルールによるリスク管理が重要

以上のように、Agentic AIの台頭によって、新たな価値創出の可能性（自律的な業務代行や意思決定の効率化等）が広がる一方、セキュリティおよびガバナンス上の課題も複雑化している。第一の脆弱性に対しては既存の情報セキュリティ原則（最小権限・境界の強化・分離など）の適用をより一層厳密に行い、従来のセキュリティホールや設定ミスを徹底的に潰すことが基本となる。第二の脆弱性については、モデルの振る舞いを完全に予測・保証できないことを前提に冗長で多層的な安全措置（入力～出力の全過程でのフィルタリング、エージェント内部の監視と内省ループ、異常検知と緊急停止機構など）を講じる必要がある。2025年末にOWASP（Open Worldwide Application Security Project）が発表したOWASP Top 10 for Agentic Applications for 2026でも、「エージェントの振る舞いハイジャック」「ツールの誤用・悪用」「ID・権限の濫用」などの新たな脅威カテゴリが指摘されており、これらに対処するためAgentic AI固有のセキュリティ基盤整備（安全な開発フレームワークや評価手法の標準化）が急務とされている。Agentic AIの安全な社会実装に向けては、「何をエージェントに任せ、どこで人間が制御するか」を精緻に定義し、組織のポリシーやガードレールに組み込むガバナンス体制が不可欠となる。技術的なディフェンス強化と組織的なルール策定を車の両輪とすることで、Agentic AIの恩恵を享受しつつリスクを適切に管理することが可能となるだろう。

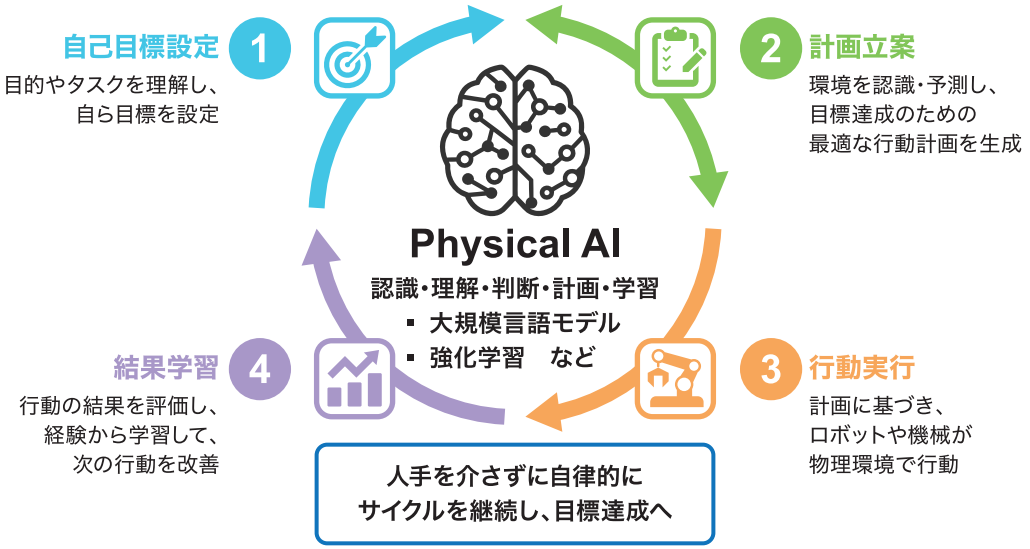
CHAPTER.2
Physical AI

1. 技術動向

Physical AIとは、物理環境との相互作用を持つAIで、従来はプログラムされたルールによって制御されていた機械にAI、特に大規模言語モデルや強化学習などによる自律的な判断能力を付与されたものである。Physical AIにより、AIが物理的実体を伴って人間社会に関与することになる。例えば、AIが搭載された自律走行ロボットや自動運転車は、カメラやセンサーによって環境を認識し、AIが最適な行動をプランニングおよび実行し、環境からのフィードバックを得ながら自己学習を行い、目標達成に向けて行動する。このように、自己目標設定、計画立案、行動実行、結果学習というサイクルを、人手を介さずに進め、それが物理環境と相互作用を持つ点において、Physical AIは従来のプログラム制御や人間補助を前提とした自動化とは一線を画している。

物理空間において高度な自律行動を可能にする技術として、近年特に注目されているのが脚型ロボットおよび人型ロボットである。脚型ロボット（四足歩行など）は、車輪では走行が困難な不整地において高い踏破能力を有し、屋内外を自在に移動できる。また、仮想環境と実空間を接続するSim2Real（シミュレーションから実機への挙動転移）手法により、効率的に動作制御を学習できる点も特徴である。また、人型ロボット（ヒト型）も、人間と同じ空間で活動し、人間向けに設計された道具や設備をそのまま利用できるという高い汎用性を持つという点で注目が高まっている。

こうしたプラットフォーム技術の進展により、農業、建設、物流、警備、医療・介護など、幅広い現場においてPhysical AIの実装が進みつつある。例えば、自律走行農機や建機は、GPSやカメラ画像を用いて畑や建設現場の状況をリアルタイムに把握し、自動で耕作や施工を行う。また、自律搬送ロボットは倉庫内を人の指示なしに移動し、荷物の仕分けや配送を担う。さらに、自動運転車は道路上の多様な事象をAIによって解析し、人の運転操作に依存することなく走行判断を下し、目的地まで安全に移動する。



Physical AIの基本的な振る舞い

目　　的	利 用 例
労働力不足や作業負荷の軽減に資する作業の自動化・代替	調理、運搬、清掃、荷詰め、建設作業など
交通・物流の高度化・自動化による効率性と安全性の向上	自動運転（自動車・船舶・鉄道・バス）、自動配送
医療・介護・リハビリテーション分野における支援と効率化	手術・診療支援、介護補助、在宅ケア、リハビリ支援
社会参加・アクセシビリティの拡張による包摂的社会の実現／社会インフラおよび自然環境の監視・維持・保全の高度化	移動支援、遠隔就労、分身ロボット、参加機会の拡大／インフラ点検、建設機械、環境モニタリング、保守・予防保全
人間にとって危険または不可能な環境における作業の代替	災害対応、大規模消防、原子力施設、警備・監視
ヒトおよび動物の身体・認知機能の代替・補完・拡張	パワーアシスト、身体自在化技術、サイバー救助犬
生活の質（QOL）および社会のレジリエンスの維持・向上	ペットロボット、パートナーロボット、心理的支援
熟練技能の継承・教育・人材育成の支援	技能教示、動作の可視化・共有、訓練支援

Physical AIは実社会で知覚・推論・行動できる特徴から、次のような社会課題解決への貢献が期待されている。

一方で、現実の物体に作用することができるため、ソフトウェア上で完結していた従来のAIとは異なるリスクが生じる。

2. 想定されるリスクと対策

Physical AIは現実世界に直接作用するため、そのリスクは従来のソフトウェア上のみで動作するAIと比べて実世界の損害（人の生命・身体、物的資産、環境など）に直結するリスクが懸念されている。

身体的・物的被害のリスク

Physical AIの誤作動や判断ミスにより、操作された装置が現実の人間や物体に意図しない形で干渉し、人命や財産に直接の被害をもたらす恐れがある。自律的に動作する装置は緊急時の人の制御が及ばず被害が広がる可能性も高まる。これらのリスクは、Physical AIが操作する装置が周りの人間を認識できずに起こる場合や、認識していてもAIの推論誤りやアラインメントの不一致により異常な振る舞いを取ることで起こる場合などがある。また、従来のAIを用いないソフトウェア動作の機械でも指摘されるシステムダウンやサイバー攻撃へのリスクもPhysical AIではより深刻なリスクになりえる。

対策として、従来、機械装置には法規制などにより安全設計が求められているが、自律的な動作を行うPhysical AIを想定した安全性分析や、安全設計の確実な実施を実施するガバナンスの構築が必要である。

法的責任の不明確さ

Physical AIが関与する事故では、過失や責任の所在が曖昧になりやすい課題がある。現行の法律や判例は人間の行為を前提としており、AIの判断ミスに対する責任を誰が負うかについて、法的な整備は途上である。例えば自動運転車が事故を起こした場合の刑事責任について乗員は負わない一方、民事責任は自動車所有者が一時的な責任を持つ、などのように整理されている。一方で製造物責任の観点では、自動車メーカー、システム開発会社、AIベンダーなど、関係者が多く、責任分担についての課題がある。また、国際的な「ジュネーブ道路交通条約」では「車両にはそれぞれ運転者がいなければならない」、「運転者は、常に、車両を適正に操縦しなければならない」などの条文もあり、国際的なルールとAIによる自動車やロボットの動作を想定した場合の法規制の整合性の課題がある。

対策としては、国際的なルール策定を含めた法整備に向けた責任モデルについてのガイドライン策定と合意形成が必要である。

雇用・社会への影響

Physical AIの社会実装が進むにつれ、人間の働き方や産業構造への影響も懸念されている。ロボティクスの役割は一部の実験的用途から企業経営に不可欠なインフラへと移行しつつあり、業務の自動化・効率化により労働需要の変化や雇用の置き換えが起こり得る。同時に、AI時代に適応するため人材の再育成（リスキリング）が追いつかないと、技能のミスマッチで取り残される層や労働意欲の低下が生じる可能性がある。さらに、AI導入による意思決定の自動化で責任の所在が不透明になり、人間の疎外や雇用の喪失などの社会的不安も指摘されている

対策として、Physical AIの導入を単なる省人化として捉えるのではなく、人間とAIの役割再設計を前提とした社会的対応や、リスキリング、意思決定の透明性確保等のルール整備が必要である。

悪用・犯罪のリスク

Physical AIはサイバー攻撃や犯罪への悪用という新手のリスクも内包する。AIロボットがネットワークを介してハッキングを受ければ、意図的に危険運転をさせられたり犯罪行為に利用されたりする恐れがある。また将来的には、Agentic AI同士の連携による新しいタイプの犯罪（デジタルと物理をまたぐ強盗や暴行など）の可能性も指摘されている。

対策として、悪用シナリオを含めたリスク評価や、Physical AI特有の物理的特性を踏まえたサイバーセキュリティ対策等の確立が必要である。

上記のようなリスクに対応し、Physical AIを安全・信頼できる形で社会実装するためには、Physical AIに対応した安全設計技術の確立および開発企業による確実な実施、さらに開発や利用に関する国際的なルール整備が必要となる。

国際的な動向として、先端AI技術のガバナンス強化に向けた取り組みが進んでいる。例えばUNESCOが2023年に発刊した報告書『Missing Links in AI Governance』では、急速な技術革新に法規制が追いついていない現状を踏まえ、グローバルなAIガバナンス強化の緊急性を強調している。またEU（欧州連合）は2023年に「AI法（AI Act）」を提案し、リスクに基づくAI規制アプローチを採用した。本規制では、自動運転車や医療ロボットなど高リスクAIシステムに対して厳格なリスク管理と人間による監視を義務付けるなどが盛り込まれている。さらに英国や米国でも、AI安全に関する厳格な評価基準の策定や産業界へのガイドライン提示など、法規制や政策によるAIガバナンス整備が進み始めている。こうした国際的なAIの利用に関する規制に加え、Physical AIの物理環境との相互作用性を踏まえた枠組みが必要となる。

最後に、Physical AIの恩恵を享受しつつリスクを最小化するには、技術開発者から政策立案者、ユーザーに至るまで多層的なステークホルダーの取り組みが必要である。具体的には、明確で説明責任あるAI利用ポリシーの策定、リアルタイム監視によるセキュリティフレームワークの導入、国際的な標準・規範の策定、労働力の再教育と適応策といった優先課題に取り組むべきとされている。信頼に足るガバナンスの下でPhysical AIを社会実装することは、革新的技術の恩恵を享受しながら人々の安全・安心を守る鍵となるであろう。今後も国内外で進む制度整備や倫理的対話を注視し、我々一人ひとりがPhysical AIの利活用について理解を深めることで、技術と社会の調和を実現していくことが期待される。

CHAPTER.3
ニューロテクノロジー

1. 技術動向

ニューロテクノロジーは、神経活動を測定または調整する技術として、これまで主に医療分野で発展してきた。しかし現在、その応用範囲は臨床領域を越え、教育、労働、安全管理、マーケティング、エンターテインメントへと拡張しつつある。この拡張は、単なる用途の広がりではない。社会的論点の重心が、「デバイスの安全性」から「神経活動データの取り扱い」へと移行していることを意味する。

ニューロテクノロジーは大きく非侵襲型と侵襲型に分類される。非侵襲型技術には、EEG、MEGといった電気・磁気信号計測、fMRI、fNIRSといった血流変化計測などが含まれる。これらは身体外部から神経活動を測定するため外科的措置を伴わない点で身体へのリスクは低い。一方で、信号対雑音比が低く、環境条件や装着状態の影響を受けやすいという技術的制約を持つ。また、その取得構造上、特定目的に必要な信号のみを選択的に抽出することは困難であり、広範な神経活動データが同時に取得されるという特徴を持つ。

侵襲型技術には、ECoG、マイクロ電極アレイなどがある。これらは脳内または脳表から直接信号を取得するため、高い時間分解能および空間分解能を実現できる。重度神経疾患や運動機能再建においては不可欠な技術である。

一方、外科的措置を伴う場合には身体的負担や合併症のリスクが生じ得る。特に脳内への埋め込みを伴う場合には、手術に起因するリスクに加え、デバイスの交換や調整に再度外科的介入が必要となる可能性があるなど、容易に元に戻すことができない点が非侵襲型との大きな相違である。また、これらの技術は長期的かつ継続的に高精度の神経活動データを取得できるという医療上の利点を持つ一方で、持続的なデータ蓄積という観点ではより深刻なプライバシー影響を内在させる。さらに、神経刺激を伴う技術においては、認知、情動、行動への影響、すなわち神経精神的な副作用が生じる可能性も指摘されている。加えて、神経モジュレーションを伴う技術においては、意思決定への影響が報告されており、これは単なる安全性の問題にとどまらず、個人の自律性や責任の所在にも関わる重要な論点となる。

	非侵襲型（身体の外から計測）	侵襲型（身体の中から計測・刺激）
代表的技術	EEG、MEG、fMRI、fNIRS など 	ECoG、マイクロ電極アレイ、深部脳電極（DBS） など 
計測原理	電気・磁気信号計測、、血流変化計測	脳内または脳表から直接信号を取得
メリット	- 外科的措置が不要でリスクが低い - 幅広い用途に適用しやすい	- 高い時間分解脳・空間分解脳 - 重度神経疾患や運動機能再建に不可欠
主な制約／リスク	- 信号対雑音比が低い - 環境や装着状態の影響を受けやすい - 広範な神経活動データを同時に取得（選択的抽出が困難）	- 外科的リスク（手術・合併症など） - デバイス交換・調整に再手術の可能性 - 身体的負担、神経精神的副作用の可能性 - 高頻度・継続データの長期蓄積によるプライバシー影響
主な用途	研究、ヘルスケア、教育、マーケティング、エンタメ、安全管理 等	重度神経疾患の治療、運動機能再建、神経刺激療法 等

両者に共通する本質的特徴は、取得された神経活動データが単一目的に閉じないという点にある。神経活動データの核心は、その多義性にある。たとえば、てんかん発作の検出を目的として収集されたEEGデータには、気分状態、疲労度、注意集中、認知機能低下の兆候などに関する情報が含まれ得る。取得時点では想定されていなかった属性であっても、将来的な解析技術やアルゴリズムの進展により抽出可能となる。すなわち、神経活動データは「特定目的のデータ」ではなく、「将来の推論資源」である。

2. 想定されるリスクと対策

インフォームドコンセントの限界

この特質は、従来の医療情報や一般的な個人情報とは質的に異なる。情報の範囲を説明時に確定することが難しいからである。解析技術の進化に伴い、神経活動データの意味内容は拡張し続ける。さらに重要なのは、自己情報コントロールの構造的弱体化である。通常、人は発言や行動を通じて自らの情報開示範囲をある程度制御できる。しかしニューロテクノロジーにおいては、利用者が特定の神経活動情報のみを選択的に提供することはできない。デバイスは、計測手法による違いや精度の制約はあるものの、特定の目的に限定されない神経信号を取得し得る。その結果、利用者は自身に関して収集される情報を意識的に制御することができない。この「推論可能性」と「自己統制困難性」こそが、神経活動データをめぐるリスクの出発点である。

神経活動データの多義性ゆえに、取得時点で将来のあらゆる推論可能性を説明することは現実的に不可能である。たとえば、てんかん発作検出への同意が、将来開発されるアルゴリズムによる精神状態推定への同意を含むのかは曖昧である。また、生データには目的外情報が含まれるため、利用者が「何に同意しているのか」を完全に理解することは困難である。

この問題は、同意の形式化を招く。文書上の合意があったとしても、それが実質的な自己決定を保証するとは限らない。

対策の中核は、静的な一回限りの同意から、動的かつ継続的な統制モデルへの転換である。利用者が後から利用状況を確認し、利用範囲を変更・撤回できる仕組みを整備する必要がある。さらに、収集されるデータの種類、保存期間、処理方法、想定される推論範囲を明確化し、継続的に更新する透明性が求められる。

しかし、最も重要なのは取得段階でデータの範囲を必要最小限に抑える設計である。推論可能性を完全に説明できない以上、不要な生データをそもそも取得しない設計こそが、最も強固な保護策となる。

データプライバシー

神経活動データは、健康状態のみならず、認知状態や精神状態に関する情報を含み得る。さらに、EEG信号が生体認証として機能し得る可能性が示されていることから、匿名化は本質的に限定的である。匿名化されたデータから再び個人を特定する「再識別」や、高度な分析によって当初想定していなかった属性や状態を推定する高度推論のリスクを十分に抑制できない可能性が高い。

したがって、神経活動データは原則として高度機微情報として取り扱うべきである。暗号化、厳格なアクセス制御、保存期間の最小化、利用履歴の監査など、多層的な保護措置が必要である。加えて、技術的フィルタリングが重要である。たとえば、周波数分解によって目的関連帯域のみを抽出し、生データを保存しない設計は、推論リスクを構造的に低減させる。

データの目的外利用

神経活動データは、取得目的を超えた利用が構造的に容易である。疲労測定のために収集されたデータが、精神疾患スクリーニングや性格評価に転用される可能性は否定できない。このような利用は、本人が認識していない内面情報を可視化し、評価や差別に結びつく危険を孕む。

対策としては、目的限定原則の厳格化が不可欠である。利用目的を技術的に固定する仕組み、二次利用の明確な禁止、第三者提供の厳格な制限が求められる。

しかし、最も実効性が高いのは、目的外推論を物理的に困難にする設計である。すなわち、最小限の特徴量のみを保存し、生データを保持しない設計思想への転換である。

データの悪用

職場における集中度測定や感情推論は、安全確保を目的として導入され得る。しかし、それがパフォーマンス評価や昇進判断に拡張されれば、差別的利用に転化する可能性がある。また、ニューロマーケティングは、個人の心理状態や状況に関する神経活動データをもとに行動や意思決定を誘導し得る精緻なターゲティングが可能であり、これが本人の認識や同意の範囲を超えて利用された場合、個人の意思形成い影響を及ぼす形でのデータの悪用につながる懸念がある。 対策として、医療または安全目的以外の感情推論については厳格な制限が必要である。利用範囲の明確化、不利益なきオプトアウト、独立監査機関による監督体制の整備が求められる。商業的利用においては、格別の透明性義務と規制措置が不可欠である。

小児・未成年者への使用

発達途上の脳に対する神経デバイスの適用は、長期的影響が予測困難である。さらに、未成年者は十分な同意能力を有さない。教育目的の能力向上デバイスなどは、通常の発達過程との区別が難しく、不可逆的変化をもたらす可能性がある。

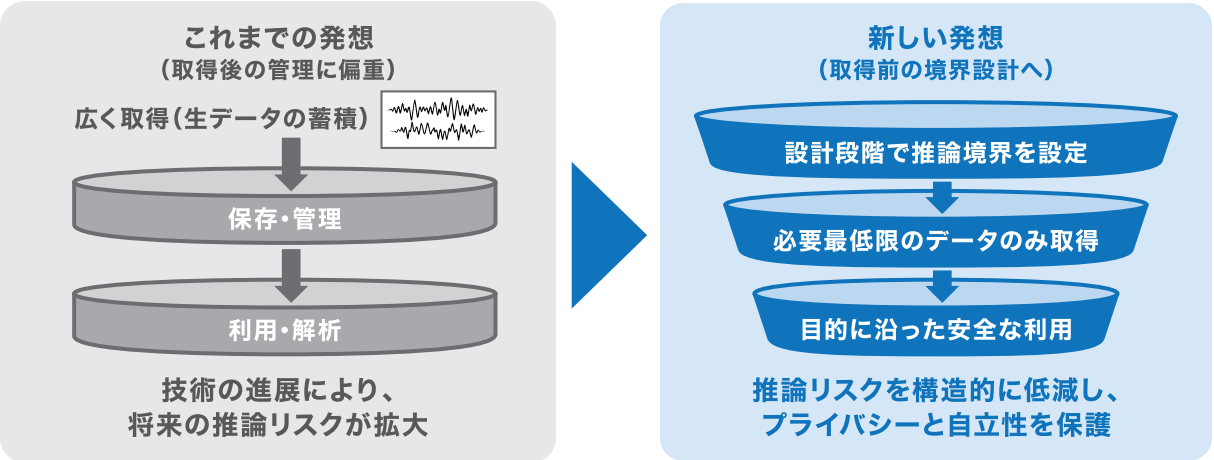
対策としては、利用目的の厳格な正当化が必要である。代替手段が存在する場合はそれを優先すべきであり、長期追跡研究の義務化および独立した倫理審査の強化が不可欠である。

主なリスク	主な対策の方向性
インフォームドコンセントの限界 （推論可能性と自己統制困難性）	**動的・継続的な統制モデルへの転換** 取得前の限定設計・道明性の確保
データプライバシー （再識別・高度推論のリスク）	**高度機微情報として保護** 暗号化、厳格なアクセス制御、保存期間の最小化、技術的フィルタリング
データの目的外利用 （目的超過・二次利用の危険）	**目的限定原則の厳格化** 利用目的の固定、二次利用の禁止、最小データ設計
データの悪用 （評価・操作・差別のリスク）	**利用範囲の明確化と監督** 厳格な制限、オプトアウト、独立監査、商業利用の透明性義務
小児・未成年者への使用 （長期影響・同意能力の問題）	**厳格な正当性と保護** 代替手段の優先、長期追随研究、独立した倫理審査の強化

ニューロテクノロジーによって取得されたデータが、将来いかなる意味を帯びるのかを完全に予測することは不可能である。解析技術の進展により、当初想定されていなかった属性や状態が後から抽出され得るからである。加えて、利用者は収集される神経情報を選択的に制御することができない。デバイスの構造上、取得は包括的に行われるため、自己情報コントロールは本質的に制限される。

したがって、対策の重心は「取得後のデータ管理」にとどまるべきではない。必要なのは、「どのような推論を可能にするのか」という境界をあらかじめ設計する視点への転換である。すなわち、「取得前の推論の境界設計」に軸足を移すことが求められる。

この問題は単なる技術管理の課題ではない。それは、人間の内面に関わる情報を、社会としてどこまで可視化し、どこで制限すべきかという、倫理的かつ社会的な境界線の設定に他ならない。



神経活動データのガバナンスにおける発想の転換

CHAPTER.4
技術の社会受容性と情報開示の設計原則

1. なぜ情報開示が問題になるのか ― 問題の構造

これまで議論を深めてきた三つの技術領域も含む新たな技術の社会実装を進める中で、その技術が孕むリスクが過大に捉えられてしまうと、技術活用に対する行き過ぎた忌避感を生み出してしまうことにつながりかねない。特に生成AIについては「AIを使って生成した」という事実をどのように開示すべきかという問題が顕在化している。一見するとこれは単純な透明性の問題、すなわち「正直に表示すべきか否か」という倫理的論点のように見える。しかし実際には、情報開示は単なる事実伝達ではなく、生成物に対する社会的評価を直接的に変動させる要因として機能している。

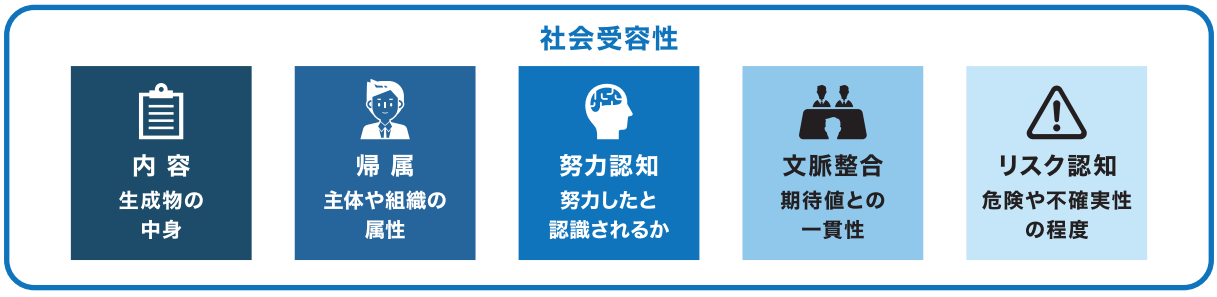
例えば、生成主体に関する情報は、受け手の評価形成過程に直接的な影響を及ぼす。すなわち、同一のコンテンツであっても、それが「人間が作成した」と説明される場合と「AIが生成した」と説明される場合とでは、受け手の評価が有意に変化する。とりわけ、感情的表現や個人的メッセージのように「人間らしさ」が期待される領域では、AI関与の情報が道徳的嫌悪感や違和感を増幅させる傾向がある。一方で、AI利用を伏せた場合には、後に発覚した際の信頼毀損や「隠していた」という疑念が生じ、より深刻な社会的反発を招くこともある。

すなわち、情報開示には二重のリスクが存在する。開示しなければ不信を招くリスクが、開示すれば評価が低下するリスクがある。この構造において問題となるのは、「開示するか否か」という二元論ではない。重要なのは、開示がどのような認知的・感情的プロセスを通じて社会的評価に作用するのかを理解したうえで、適切な設計を行うことでリスクを最小化することである。

生成AIの出力は、その内容のみで完結するものではない。受け手は常に、「誰が、どのように、なぜ生成したのか」という情報を含めて意味づけを行う。したがって、情報開示は単なる倫理的義務ではなく、社会的評価に介入するガバナンスとして位置づける必要がある。

2. 何が社会受容性を左右しているのか ― 評価メカニズムの核心

では、生成AI出力に対する社会受容性は、どのような要因によって左右されるのか。既存研究および実務上の炎上事例を踏まえると、評価は、出力内容そのものを除き、主として四つの認知的要素（帰属、努力認知、文脈整合、リスク認知）によって構成されていると整理できる。



社会受容性の構成要素

第一に、「帰属（authorship）」である。生成物がAIによるものか、人間によるものかという情報は、それ自体が評価の枠組みを規定する。AIが作成したと明示されるだけで創造性や誠実性の評価が低下する現象は、「AI-authorship effect」として報告されている。これは品質評価とは独立に作用する認知バイアスであり、情報開示が直接的に受容性へ影響を及ぼす経路を示している。

第二に、「努力認知（perceived effort）」である。人は成果物の背後にある制作過程を想像し、その中にどの程度の人間的努力や意図が投入されたかを推定する。AIが全面的に生成したと理解された場合、「手間をかけていない」「本来的な創作ではない」という印象が生じやすい。特に感情的文章や芸術表現においては、努力認知の低下が道徳的嫌悪感と結びつき、評価を大きく下げることがある。

第三に、「文脈整合（context fit）」である。技術そのものの良し悪しではなく、それが置かれる場との適合性が重要となる。高級ブランドの広告、教育現場、芸術作品、公共的メッセージなど、分野ごとに期待される価値は異なる。廉価性や効率性を想起させるAI生成物が、「手仕事」「希少性」「人間性」といった価値が重視される領域に投入された場合、強い不整合感が生じる。この不整合が炎上の引き金となる。

第四に、「リスク認知（perceived risk）」である。倫理的逸脱、雇用代替、アイデンティティ侵害、プライバシー侵害、あるいはガードレールの不明瞭さといった懸念は、生成AIに固有の不安要因として作用する。特に利用範囲や統制体制が説明されない場合、受け手は最悪のシナリオを想定し、評価を下方修正する傾向がある。

重要なのは、情報開示がこの構造の複数の要素に同時に作用する点である。開示は帰属を明確化し、努力認知を変化させ、リスク認知を増減させる。ゆえに、開示は単なる表示ではなく、評価構造そのものに介入する設計要素なのである。

3. 企業はどう設計すべきか ― 情報開示の原則

以上の構造を踏まえると、企業に求められるのは一律の表示義務の履行ではなく、社会受容性を意識した情報開示の設計である。ここでは四つの原則を提示する。

【原則1：帰属の明確化】

AIの関与を曖昧にしないことは、信頼維持の前提である。後からAI利用が発覚することは、評価低下以上に「隠蔽」の印象を生み、ブランド毀損につながる。したがって、AI関与の事実は適切に明示されるべきである。

【原則2：人間関与の可視化】

単に「AI生成」と表示するだけでは、努力認知の低下を招く可能性がある。どの工程を人間が担い、どの部分に判断や創意が投入されたのかを示すことで、成果物を単に機械的に生成されたものではなく「ガバナンスされた出力」として位置付けることができる。これは努力認知を高める重要な設計要素である。

【原則3：文脈整合の確認】

分野ごとに期待される価値は異なる。芸術や高級ブランド、教育など、感情的・象徴的価値が強い領域では、AI利用の説明方法や強調点を慎重に設計する必要がある。一律の開示フォーマットではなく、文脈に応じた開示水準の調整が求められる。

【原則4：リスク対策としてのガードレールの明示】

とりわけ高リスク領域では、利用範囲、制限事項、監督体制を明確に示すことが不可欠である。どのような用途には用いないのか、どのようなガバナンスの仕組みが組み込まれているのかを示すことは、リスク認知を適切にコントロールし、社会的信頼を形成する基盤となる。

総じて言えるのは、情報開示とは単なるラベル表示ではないということである。それは、生成AI出力を「偶発的な産物」ではなく、「ガバナンスされた成果物」として社会に提示するための設計行為である。

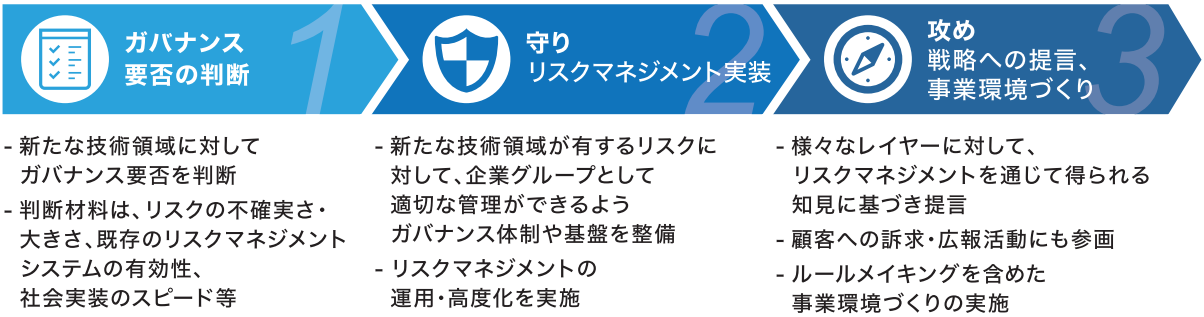
生成AIの社会実装が加速する時代において、出力の品質だけで持続的な受容を確保することは困難である。必要なのは、出力そのものと同時に、その背後にあるガバナンス構造を可視化することである。情報開示は、そのための中核的インフラとして位置づけられるべきである。

CHAPTER.5
テクノロジーガバナンスの企業実装

1. 企業実装の意義と方法

テクノロジーの進歩は、企業活動の高度化をもたらす一方で、既存の統制や制度設計では十分に対応できない新たなリスクを生み出している。こうした新たな技術は発展速度を増しており、従来の個別領域ごとの統制を積み上げるだけでは不十分かつ遅れが生じる可能性が高く、リスクが顕在化することで技術活用にブレーキがかかってしまうおそれがある。リスク対策を行いながら、新たな技術活用を推進・加速していくために、企業全体として横断的に技術を統治する枠組み、すなわちテクノロジーガバナンスの実装が不可欠となる。

本稿では、テクノロジーガバナンスを三段階で実装するモデルを提示する。第一段階は「ガバナンス要否の判断」、第二段階は「守りのガバナンス（リスクマネジメントの実装）」、第三段階は「攻めのガバナンス（事業環境づくり）」である。この三段階のうち二段階目（守り）と三段階目（攻め）は直線的なプロセスではなく、相互にフィードバックを行いながら高度化していく構造を成す。



テクノロジーガバナンス フレームワーク

2. ガバナンス要否の判断

あらゆる新技術に対して一律に手厚いガバナンス体制を構築することは、現実的でも合理的でもない。一方で新技術領域が起こるたびにリアクティブに対応をすることで、リスクマネジメントの遅れが生じるおそれがある。そこで、どの技術領域に対して、どの水準の統制を設けるべきかを判断するプロセスを予め整え、制度化することの重要性が高まっている。

具体的には、ユースケースを想定したリスク評価、既存の統制の仕組み及びリスクマネジメントで対応可能かの検証、社会受容性や規制動向の分析、炎上事例や社会的論争の有無といった観点を統合的に評価する。ここではPEST分析等の外部環境分析も活用し、企業の技術戦略と整合的に対象領域を選定することが求められる。

この段階を形式的なチェックにとどめず、経営レベルでの意思決定プロセスに組み込むことで、テクノロジーガバナンスは「後追いの統制」ではなく「先回りの設計」へと転換することが可能となる。

3. 守りのガバナンス ― リスクマネジメントの実装

第二段階は、対象と定めた技術領域に対し、企業グループとして適切なリスク管理を実装することである。ここでの目的は、技術の活用を萎縮させることではなく、リスクを可視化・制御可能な状態に置くことで、持続的な事業展開を支える基盤を整備することである。

守りのガバナンスは、三つの機能領域から構成される。

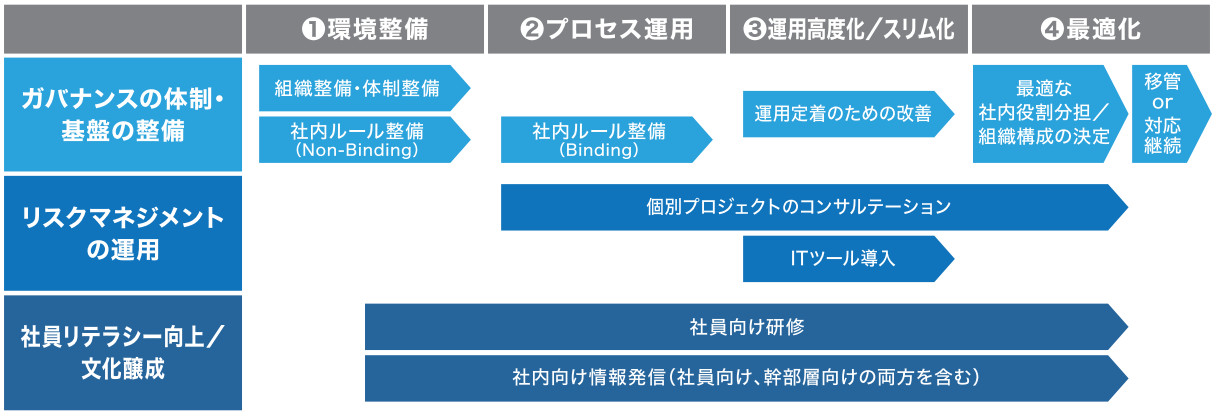
第一に、ガバナンスの体制・基盤整備である。ポリシーやガイドラインの策定を通じて、価値判断の拠り所を明確化し、社内ルール(Non-Binding/Binding)の整備を行う。これには単なる理念宣言だけではなく、社内運用に落とし込むための具体的なルールや判断基準も含まれる。

第二に、リスクマネジメントの運用である。個別プロジェクトへのコンサルテーション、リスクチェックシートの運用、社員研修などを通じて、日常業務にガバナンスを組み込む。

第三に、社員リテラシー向上・文化醸成である。ITツール導入や情報共有基盤整備により、実効性を高める。

これら三つの機能は、(1)体制・基盤の整備、(2)プロセス運用、(3)運用高度化、(4)最適化の四段階で成熟していく。初期段階ではルール整備や体制整備が中心となるが、運用を通じて改善が進み、最終的には最適な役割分担の決定や組織的な移管により、持続可能な運用モデルへと進化する。

特に大企業においては、法務、情報セキュリティ、ITマネジメント、全社リスクマネジメント、知財など既存機能との連携が不可欠である。テクノロジーガバナンスはこれらを代替するものではなく、横断的に接続し統合するハブ機能として位置づけられるべきである。



実装段階イメージ

4. 攻めのガバナンス ― 事業環境づくりと競争力強化

テクノロジーガバナンスは守りにとどまらない。リスクマネジメント活動を通じて蓄積された知見は、企業の競争力強化に資する戦略資産となる。

攻めのガバナンスは三つの方向性を持つ。

第一に、戦略・事業への提言である。リスク評価を通じて得られた技術特性や規制動向の洞察は、新規事業設計や投資判断にフィードバックされるべきである。

第二に、対外発信・顧客訴求への参画である。自社のガバナンス体制や倫理的配慮を明確に示すことは、信頼獲得と差別化要因となる。

第三に、ルールメイキングを通じた事業環境整備である。技術活用の障壁となる制度や慣行に対し、業界団体や政府との対話を通じて改善を働きかけることは、長期的競争優位に直結する。

グローバル市場においては、EUをはじめとする規制動向が事業機会を左右する。日本企業にとっても、国内制度への対応だけでなく、国際的なルール形成の動向を見据えた戦略的関与が不可欠である。

5. 三段階モデルの意義

三段階モデルの本質は、ガバナンスを「制約」ではなく「価値創出のインフラ」と捉える点にある。

第一段階で対象を選定し、第二段階でリスクを制御し、第三段階で知見を戦略へ転換する。この循環が確立されれば、テクノロジーガバナンスは単なる内部統制ではなく、企業の持続的成長を支える経営基盤となる。

技術の社会実装スピードが加速する時代において、ガバナンスの整備は後追いであってはならない。むしろ、技術ライフサイクルに先行して設計されるべきである。適切なガバナンスの実装を通じて企業にとっての「安心して挑戦できる環境」を確立することで、技術活用の攻めと守りを両立させることができる。

テクノロジーガバナンスの企業実装とは、リスク管理部門の新設ではない。制度とプロセスを通じて、技術選定から事業環境形成に至るまでを一貫して設計する経営機能の確立である。その成否が、次世代テクノロジー時代における企業の持続可能性と競争力を左右する。

おわりに

本ホワイトペーパーでは、Agentic AI、Physical AI、ニューロテクノロジーという三つの技術領域を対象に、それぞれの技術動向とリスク構造を整理するとともに、それらに共通する横断論点として社会受容性および情報開示の問題を取り上げ、最終的に企業におけるテクノロジーガバナンスの実装モデルを提示した。

これらの技術に共通するのは、単なる「ITの高度化」にとどまらず、意思決定の自律化、物理世界への直接作用、人間の認知・神経機能への接続といった特性を通じて、企業活動の境界そのものを拡張しつつある点にある。その結果、責任の所在、予見可能性、倫理的境界、さらには社会受容性といった論点が、技術課題であると同時に経営課題として顕在化している。

とりわけ重要なのは、リスクの性質が従来と質的に異なっている点である。Agentic AIにおいては、意思決定と実行の主体が人間からソフトウェアへと移行することで、制御と責任の境界が曖昧になる。Physical AIにおいては、AIの判断が直接的に物理的結果として現れることで、安全性と法的責任がより深刻な形で問われる。ニューロテクノロジーにおいては、取得されたデータが将来的にどのような意味を持ち得るかを完全に予測できないため、「取得後の保護」ではなく「取得前の推論の境界設計」という新たな課題が生じる。

さらに、これらの技術を社会に適用する過程では、技術そのものの性能や安全性に加え、「どのように説明され、どのように理解されるか」が評価や信頼に大きく影響する。すなわち、技術の社会実装は単なる導入の問題ではなく、受容性と情報開示を含めた統合的な設計課題である。

こうした前提に立つと、企業に求められる対応は明確である。新技術を個別に評価し、個別に対策を積み上げるだけでは不十分であり、技術の特性に応じてガバナンスの枠組みそのものを再設計する必要がある。本稿で提示した三段階モデル―「ガバナンス要否の判断」「守りのガバナンス」「攻めのガバナンス」―は、そのための実践的なアプローチである。対象領域を見極め、リスクを制御可能な状態に置き、さらにそこから得られる知見を事業戦略や対外的信頼の形成へと接続する。この循環を確立することで、ガバナンスは単なる制約ではなく、価値創出を支える経営基盤へと転換する。

一方で、テクノロジーガバナンスは一度設計すれば完了するものではない。技術の進展、規制動向、社会受容性、実運用における知見の蓄積に応じて、継続的に見直され、更新されるべきものである。新興技術を取り巻く環境が不確実である以上、企業に求められるのは固定的な正解ではなく、変化を前提に議論し、学習し続ける組織能力である。

テクノロジーガバナンスの本質は、リスクを抑制することそのものではなく、リスクを制御可能な状態に置くことで、企業が「安心して挑戦できる環境」を構築する点にある。その環境こそが、技術活用の加速と社会的信頼の確保を両立させ、結果として企業の持続的成長と競争力を支える。

本ホワイトペーパーが、技術導入を「実装」の問題から「統治」の問題へと引き上げ、さらに「統治」を経営機能として位置付けるための出発点となることを期待する。