

脳情報を活用し知覚情報等を推定する AI 技術の開発・活用ガイドライン

2024 年 1 月

脳情報を活用し知覚情報等を推定する AI 技術の活用ガイドライン作成委員会

本ガイドラインは情報通信研究機構「2021-22 年度 脳情報を活用し知覚情報を推定する AI 技術等の社会受容性確保に向けた調査研究」事業および「2023 年度 脳情報を活かしたサイバー空間の感性評価技術の社会実装に向けた調査研究」事業の一環として作成されたものです。

内容

本ガイドラインに登場する用語とその定義.....	2
背景.....	4
1. はじめに.....	5
2. 脳情報を活用し知覚情報等を推定する AI 技術とは.....	7
脳情報を活用し知覚情報等を推定する AI 技術の流れ.....	7
現状可能なこと：脳情報 AI モデルの B2B 事業への応用.....	8
現状不可能なこと.....	8
将来的に可能になりうること：脳情報 AI モデルの DTC 事業(Direct to Consumer:対個人サービス)への利用.....	8
現状想定される利用スキーム.....	8
3. 脳情報を活用し知覚情報等を推定する AI 技術の活用ガイドライン.....	9
ガイドライン①脳情報データの取り扱いにおける同意取得.....	9
ガイドライン②脳情報データのライフサイクルにおいてとるべき対応.....	10
ガイドライン③脳情報をモデル化することによる差別・スティグマ助長リスクへの配慮と対応.....	10
ガイドライン④特定属性に限った脳のモデル化、および利用による潜在的なステレオタイプ助長のリスクへの配慮と対応.....	10
ガイドライン⑤特定の脳の特性に関わる配慮と対応.....	11
ガイドライン⑥意識的な自覚なく消費者の思想・価値観に与える影響に関する配慮と対応..	12
ガイドライン⑦脳情報利用に関する透明性の確保.....	12
ガイドライン⑧脳情報を利用した事業に関する公平かつ包摂的なアクセシビリティの担保..	12
4. 脳情報を活用し知覚情報等を推定する AI 技術の活用ガイドライン作成委員会 関係者一覧.....	14

本ガイドラインに登場する用語とその定義

本ガイドラインで使用する用語を以下のように定義する。

知覚：五感を通して、外界の事物の性質、形態等を把握・認識する主観的体験のこと指す。

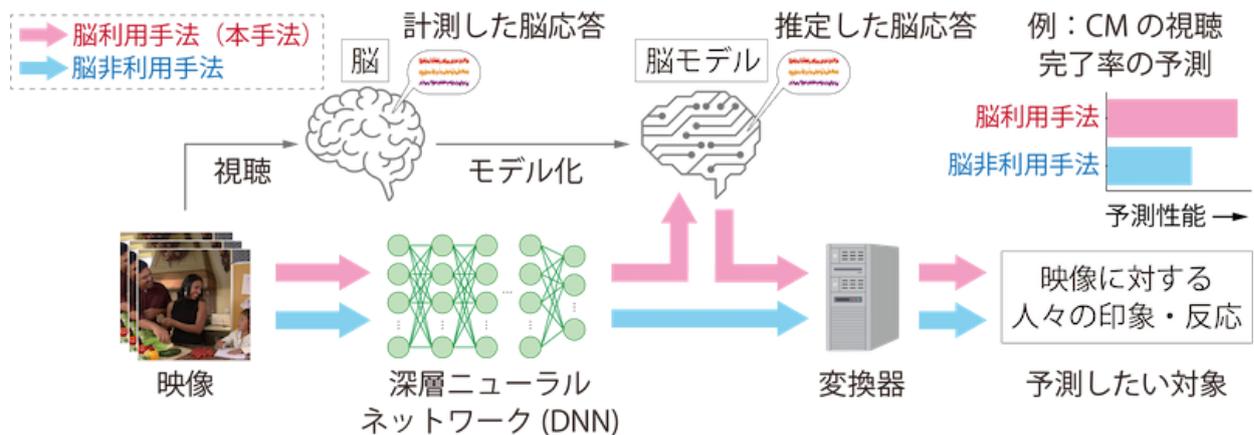
印象・反応：知覚された対象に対して好ましいと思ったり、購買動機を喚起されたり、記憶に残るなど知覚の結果生み出される認知的・行動的なヒトの反応を指す。

脳情報：機能的磁気共鳴画像法（fMRI）等の脳機能計測装置を用いて、脳神経の活動を記録したデータ、脳活動データともいう。特定の知覚などの情報が表現・表象されている。意識下において言語で表現されるよりも、ヒトの知覚や行動を説明・予測しうる情報。本ガイドラインが主な対象とするのは、動的に変化する脳の機能的な情報表現に関するデータである。

脳情報から知覚情報等を推定する AI 技術（BTL：Brain-mediated Transfer Learning）：「脳情報 AI」とも表現する。機械学習技術を利用し、収集したヒトの動画視聴時などの知覚体験に関わる脳機能データ（脳情報）から、脳活動を推定する（エンコードモデル）、また、その推定脳情報からさらに知覚内容等を推定する機械学習モデル（デコードモデル）を指す（Nishida *et al.* 2020）。

脳情報 AI モデル（BTL モデル）：本ガイドラインにおいては、単に“モデル”とも表現する。脳情報 AI と同義だが、個人の脳情報を予測し、知覚内容を推定する個人モデルと、複数の個人モデルを利用して一般化された予測結果を得る手法が現在利用されている。現状、本事業において想定されているサンプルサイズは様々な性別・年代の日本人 50～100 名程度である。個人の脳情報モデルは原則的に、個人の脳への入力（映像等）から脳活動としての出力を予測するだけのものである。その予測された脳情報データを、他の集団から収集した知覚情報（アノテーションデータ）や印象・反応・行動情報（好ましき・広告クリック率）と紐づけることによって、そうした知覚や印象・反応・行動を予測することができる。

図：脳情報 AI モデルの全体像¹



（モデルの）二次利用：学習済みの脳情報 AI モデルを他の教師データやモデルに二次的に利用すること。例えば、知覚を推定する脳情報 AI モデルを利用することで、広告動画に対する知覚推定結果と新たな広告指標を結びつけることができ、その結果から広告指標を予測する二次的なモデルを作成するようなケース（転移学習）が想定される。

効果量の増大：脳情報 AI を利用することで、広告活動の訴求力といった人への影響が従来の技術と比べて大きな効果を持ちうることを指す。

¹ Nishida S, Nakano Y, Blanc, A, Maeda N, Kado M, Nishimoto S. Brain-mediated Transfer Learning of Convolutional Neural Networks. Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence 34(4):5281–5288, 2020. <https://doi.org/10.1609/aaai.v34i04.5974>、および <https://www2.nict.go.jp/nie/nishida/research.html>

特定の脳情報処理特徴・脳の特性：個人的な属性などによって、同一の対象を見てもその知覚内容が異なることを指す。例えば、顔のそれぞれの部位は認識できるが、顔全体を認識できない相貌失認の傾向がある人にとって、人の顔は顔と処理されないなどの特徴がある。他にも依存症の傾向がある人物にとって依存対象の刺激は他の人と比べて高い渴望感を惹き起こすことが想定される。もちろん、一人ひとり脳の特性は異なるという前提ではあるが、例に挙げたような“臨床的な症状や障害など”に関連する特性を本ガイドラインでは中心的に議論する。

ステイグマ：本ガイドラインでは精神疾患など個人の持つ特徴に対して、周囲からあるいは自身に否定的な意味づけが行われることを指す。

パーソナライズ：本ガイドラインでは特定の特性を持つ個人・集団に対し、商品・広告の発信者が、その対象者が最も反応（好印象を抱く・買いたくなる）するような刺激に調整することを指す。

潜在意識・無意識への影響：個人の意識にのぼらない・言語報告できない知覚を通して（本人は見たり・聞いたりしていないと感じている）本人の行動に影響を及ぼし、何らかの反応を起こすことを指す。従来の意識的な質問回答では想定されない消費者への影響の一つ。

内面の表出：外部からは分からない自身の感じている内容等について外部に表出する行為を指す。

錯覚：物理的には同一にも関わらず、脳の情報処理特性によって、知覚・あるいは印象などが変わることを指す。狭義には「明るく見える」「長く見える」など知覚的な錯覚を指すが、本ガイドラインでは品質がより良く感じられるなど、より高次な認知・行動に影響を与えるものも指す。

研究開発に携わる人：研究開発企画・脳計測・データ収集・解析・モデル作成などに携わる関係者。業務委託などを通して、二次的に関与する個人・団体も含む。

事業者：自ら脳情報 AI モデルを利用し広告作成・商品開発等を行う一次利用者だけでなく、メーカーなどと契約し脳情報 AI を用いたコンサルティングを行う団体等を含む。

仮名化：データから個人識別子を取り除き、その識別子を仮の値で置き換える処理のことを示す。個人のプライバシーを保護、データのセキュリティを向上させるために行う。

匿名加工情報：特定の個人を識別することができないように個人情報を加工し、当該個人情報を復元できないようにした情報のことを示す。

オプトイン：（ユーザーが）参加や許諾、承認等の意思を相手方（事業者）に示すことであり、事業者がユーザーに事前に許可を取ることである。

背景

脳神経科学および人工知能技術（AI）の共進化的発展は、脳の情報表現およびヒトの知覚を計算機上で再現・予測する人工の脳、あるいは脳のデジタルツインとも呼べる融合技術を可能としてきている。こうした技術は、ヒトの脳の理解の促進に関する科学技術的な意義だけでなく、これまでの技術では不可能だったレベルで消費者・ユーザーをより深く知ることで、より優れた製品・サービスを開発することができるようになるといった点で、社会全体のイノベーションに貢献する可能性を秘めている。

しかしながら、脳の情報を扱う AI は従来の AI に加えて特に倫理的な配慮が必要な領域であるが、そのような留意点が整理されておらず活用促進の障害となっている。従来の画像や自然言語を対象として学習させる AI に関しては、様々なガイドライン等の国際的な整備が進んでいる。

そこで、日本国内閣府の官民研究開発投資拡大プログラム（以下「PRISM」）および研究開発成果の社会実装への橋渡しプログラム（以下「BRIDGE」）において、総務省・情報通信研究機構（以下「NICT」）事業である「脳情報から知覚情報を推定する AI 技術（AI 戦略・中核基盤研究の社会実装加速化）」の一環として、「脳情報を活用し知覚情報を推定する AI 技術等の社会受容性確保に関する調査に資する検討会」（以下、検討会）が設置された。検討会では、脳情報を活用し知覚情報等を推定する AI 技術等の社会実装に関わる倫理的・法的・社会的課題（Ethical, Legal and Social Issues: ELSI）への対応について、脳神経科学ならびに人文・社会科学系分野の研究者を含む幅広い関係者を集め、技術の社会受容性を確保する知見を収集し、技術開発等にかかるガイドライン案を作成するための調査および検討を目的として活動してきた。本ガイドラインはその成果である。

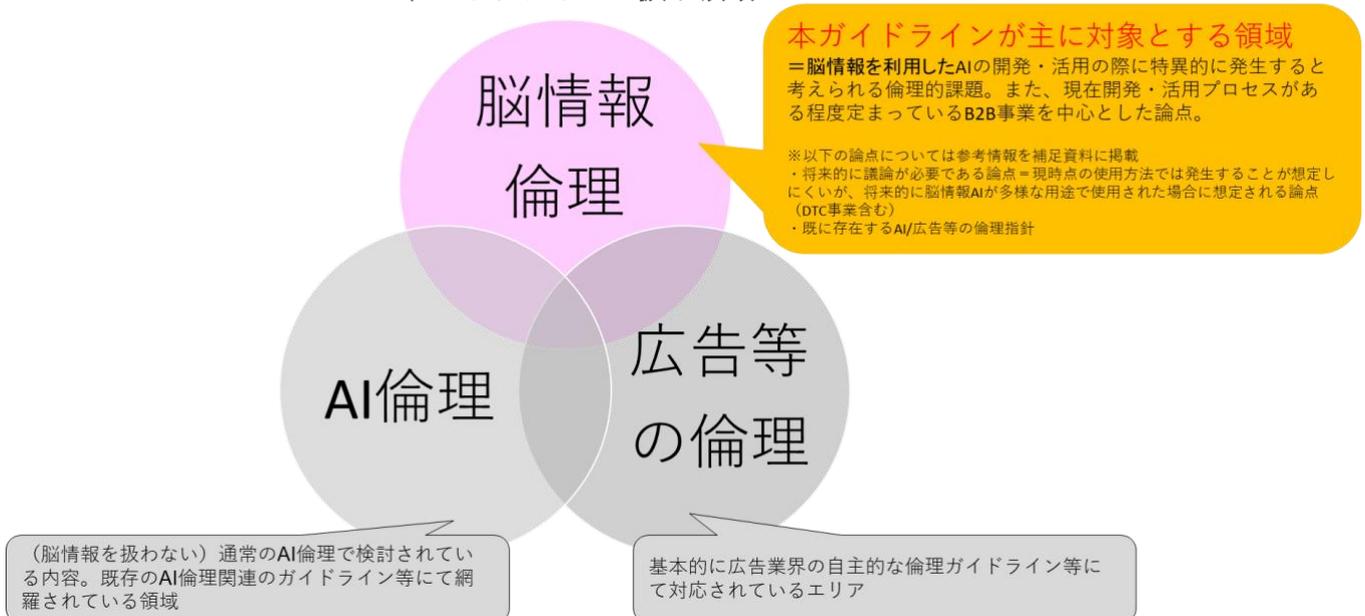
1. はじめに

本ガイドラインの目的：本ガイドラインは上述の背景の下、日本国・内閣府の科学技術開発プログラムである PRISM および BRIDGE 事業で開発している「脳情報から知覚情報等を推定する AI 技術」の ELSI に配慮した利用促進・指針の提供をすることを主たる目的としている。

想定読者・対象者：本ガイドラインは、主に「脳情報から知覚情報等を推定する AI 技術」を事業用途に研究開発する関係者、および実際に事業に活用する企業・団体を主な対象としている。

扱う範囲：本ガイドラインは、脳情報を利用することで特異的に発生する倫理的課題を主に対象としており、いわゆる通常の「AI 倫理」やこれまで広告業界等の民間事業主体が自主的に作成・運用している広告に関する倫理指針で扱える範囲は非優先としている（下図）。なお、本

本ガイドラインが扱う領域



ガイドラインは脳情報 AI の利用促進を主な目的としているが、利用における倫理的な課題は上流の開発プロセスとも大きく関わるため、利用だけでなく、「開発～利用」の全体的なプロセスの中で留意すべき課題と指針を取りまとめている（次ページ図）。

拘束力：本ガイドラインは法的な拘束力を有するものではなく、また当事者間における交渉や契約の自由を制約するものではない。

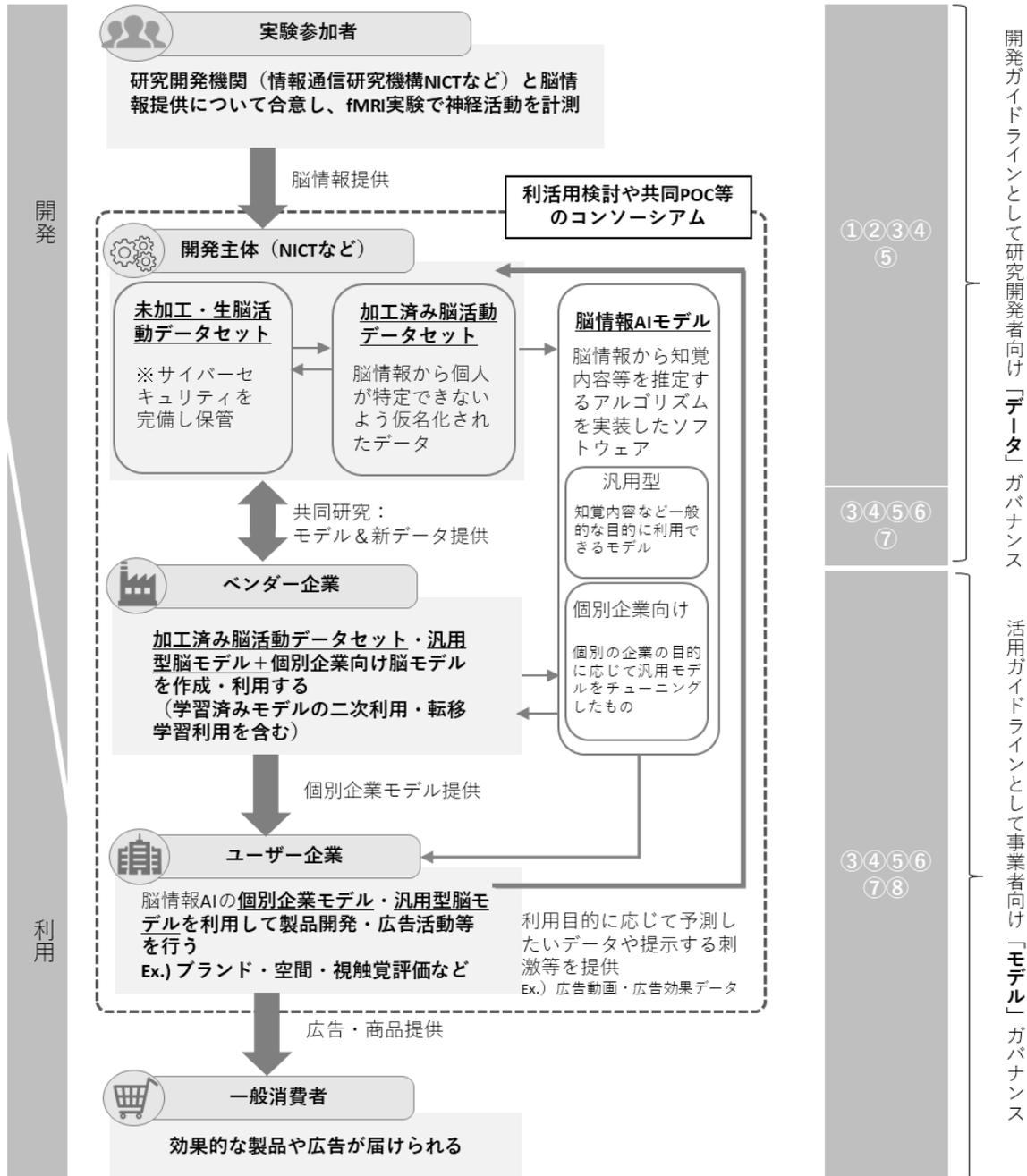
展望：本ガイドラインは、日本国の公共研究開発プロジェクトにおける脳情報の利用や提供を主たる想定として作成されたものであるものの、ここで提示される視点と論点は、NICT における脳情報の活用に限らず、今後の脳情報利用全般に関わる開発～利用の全体的なプロセスの中で留意すべき課題への視座を有すると考えられる。このガイドラインが主対象とする「脳情報から知覚情報等を推定する AI 技術」だけでなく、今後個人の脳情報を直接扱ったり、活用する事業者が出てきた場合に、考慮すべき ELSI の参考となり、適切な科学技術の発展および社会還元に貢献するものとなることを期待する。現に、脳情報と AI の新たな融合領域として、生成 AI との組み合わせや、リアルタイムニューロフィードバックへの利用などの事例が続々と誕生してきており、想定外の倫理的課題が生じる可能性があり、極めて慎重な対応が求められる。今後研究開発が進む新たな事業応用領域にも適用可能であることを目指している。

リスク・不確実性に関する態度：脳情報の事業活用に関しては今後発展が期待される領域でもあることから、現状では想定・計量が困難な不確性を多くはらんでいると考える。そうした状況の中で、本ガイドラインの基本的な姿勢として、不確実な影響に関してはモニタリング、ケーススタディの共有を強く推奨していくという立場をとる。

〈脳情報型AIの開発～脳情報をめぐる研究開発・利用のモデル全体像〉

プロセス

関連性が高い
ガイドライン項目



～ガイドライン一覧～

- ① 脳情報データの取り扱いにおける同意取得
- ② 脳情報データのライフサイクルにおいてとるべき対応
- ③ 脳情報をモデル化することによる差別・スティグマ助長リスクへの配慮と対応
- ④ 特定属性のみの脳のモデル化、および利用による潜在的なステレオタイプ助長のリスクへの配慮と対応
- ⑤ 特定の脳の特性に関わる配慮と対応
- ⑥ 意識的な自覚なく消費者の思想・価値観に与える影響に関する配慮と対応
- ⑦ 脳情報利用に関する透明性の確保
- ⑧ 脳情報を利用した事業に関するアクセシビリティへの配慮と対応

2. 脳情報を活用し知覚情報等を推定する AI 技術とは

脳情報を活用し知覚情報等を推定する AI 技術の基となるのが、fMRI（機能的磁気共鳴画像法）等の脳機能イメージングデータから、機械学習技術を利用して被験者の知覚内容の解読や再構成をする脳情報解読（デコーディング）技術と、外的な刺激や特定の内部状態に対しての脳活動パターンを予測する符号化（エンコーディング）技術である。

本ガイドラインが対象とする技術は、このエンコーディングとデコーディング技術を様々に組み合わせ、外部物理世界に対するヒトの脳活動パターンの予測と、その予測された脳活動パターンから、対応する知覚内容などヒトの反応に関わる予測を行う AI 技術である。この技術は、初出の論文²において“Brain-mediated Transfer Learning(BTL)”と呼称されている。以下に、当該技術の概要について紹介する。

脳情報を活用し知覚情報等を推定する AI 技術の流れ

脳のデータ収集：数 10～数 100 名程度を対象に実験を行う。被験者には映像・音声・その他製品・広告等の外部刺激を与え、その際の脳活動を fMRI によって計測する。脳情報に関しては、開発主体（国立研究開発法人情報通信研究機構(NICT)など）が個人の同意を得た上で取得する。

モデル化：取得した個人の脳活動データと外部データ（知覚・印象データ、行動データ）を結び付けて仮名化されたモデルを作成する。モデルには 2 段階あり、被験者に提示した感覚刺激の特徴情報から脳活動を再現・予測するエンコードモデルと、脳活動データから提示された刺激の内容・反応を予測し、言語化するデコードモデルが存在する。以上を総称して脳情報 AI モデル（または BTL モデル）とする。知覚データは「緑の芝」、「青い文字」、「人の声」など画像や動画等の外部環境からの刺激により知覚された感覚情報を指し、印象データは対象刺激に対する好みや感情的な反応等を言語化したものを意味し、行動データはマーケット全体のクリック率や購買数など、消費者の行動等により生成されるデータを指している。このモデルを基礎に、特定の感覚入力に対する脳活動から、その先の知覚等の反応をコンピューター上で再現する仮想化技術として用いることで、仮想的な脳として脳活動のシミュレーションをすることができる。

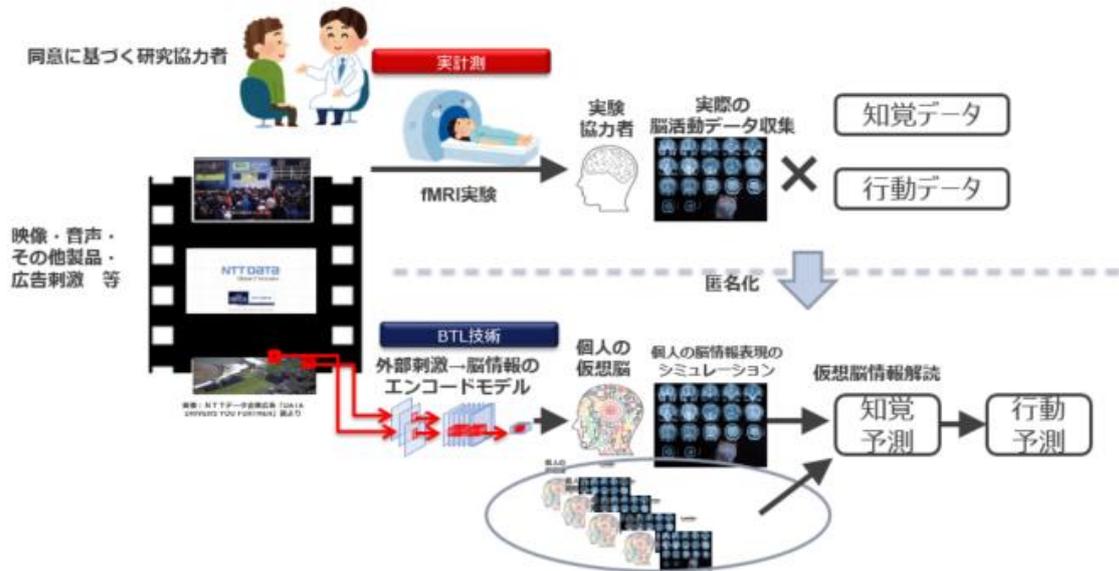
利用：上記のプロセスを経てモデルを学習させた後に、新たに画像や動画などの刺激データを用意することで、エンコードモデルを介してその刺激に対する個人の脳活動（脳の情報表現）を予測する。またその個人の脳活動の予測データから、知覚内容等を解読する。さらに、特定の刺激に対するクリック率等の外部データを学習させておけばマーケットレベルでの行動予測も行うことができる。個人の脳の予測結果を平均化することで、予測結果を一般化することも期待される。以上のように、モデルを利用して刺激の入力～脳情報表現～知覚・行動出力という一連の脳の情報処理過程のシミュレーションを企業等が行うことで、研究開発の効率化に加えて、最も消費者に対して意図する知覚が期待されるデザインを選んだり、行動を最大化できそうな広告表示を選ぶなど、優れた製品開発やマーケティング活動に役立てることができる。

更新：新たな被験者が増えたり、既存の被験者から新規の脳情報を収集したりすることによって脳活動のデータが増える際にモデルが増えたり、更新が行われる。また、既存の脳情報を用いるが、別の目的（知覚ではなく行動を予測したい場合など）のために知覚情報等を応用する際もモデルが更新される。

廃棄：脳情報を提供した被験者が事後的にデータの消去を希望した場合は、削除することは不可能ではないが、脳情報 AI モデル構築に際しデータの仮名化が行われているため簡単ではない³。また、個人のモデルを総合的に平均化するなどして利用することを想定しており、個人の情報の影響を完全に削除することは困難である。一方、脳情報 AI モデルから元のデータを復元することはできず、個人の特定は不可能である。

² Nishida, S., Nakano, Y., Blanc, A., Maeda, N., Kado, M., & Nishimoto, S. (2020, April). Brain-mediated transfer learning of convolutional neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 5281-5288).

³ 脳情報 AI モデル自体が何らかの害を持つことが示された場合においてはモデル自体の廃棄や使用停止はあり得る。



現状可能なこと：脳情報 AI モデルの B2B 事業への応用

主な目的：個人の製品やサービス等に対する反応を、脳情報 AI によって分析し、広告や商品開発等に役立てる。

実際の活用事例：TV ショッピングの番組の制作に人工知能を活用し、入電率 27.6%アップ (NTT DATA, 2018.⁴)、音楽トレンドを可視化しヒットソングの予測 (Billboard Japan and NTT DATA, 2020⁵)

現状不可能なこと

リアルタイムに個人の脳状態を推定すること（現状は一度計測したデータをオフラインでモデル化し事後的に推定）。

機微な情報（思想・信条その他）の解読。

将来的に可能になりうること：脳情報 AI モデルの DTC 事業 (Direct to Consumer: 対個人サービス) への利用

主な目的：個人あるいは属性などに応じてカスタマイズされた脳情報 AI モデルを利用し、個人向けサービスに活用する

使用例：教育活用（教育コンテンツのパーソナライズに特定の脳情報 AI モデルを利用）・能力向上・障害/疾病の治療/ヘルスケア用途の利用

現状想定される利用スキーム

研究開発主体（NICT など）が、事業会社からの依頼を受け技術を提供し、当該事業会社が消費者に対する広告や商品を開発する利用プロセスが想定されている。また、研究開発主体（NICT など）から脳情報 AI の技術提供を受けた事業者が、第三者の企業の広告作成や商品開発を支援するコンサルティングを行う利用形態も想定される。

事業者側が扱うのは基本的にはモデル化された脳情報 AI であり、元となった脳情報そのものには接触しない。一方で、利用者たる事業者側は、脳情報を提供した参加者の同意と個人情報保護が確立できる範囲内で、どのような属性の人々の脳情報を集めたモデルなのかについてデー

⁴ <https://www.nttdata.com/jp/ja/news/release/2018/11130>

⁵ <https://www.nttdata-strategy.com/newsrelease/200903.html>

タを取得し、活用することができる（技術の透明性の確保と個人情報保護の両立）。

3. 脳情報を活用し知覚情報等を推定する AI 技術の活用ガイドライン

脳情報を活用し知覚情報等を推定する AI 技術の利用に関して以下のように指針を示す。
なお、具体的に配慮を必要とする想定ユースケースについては、別紙「2022 年度 脳情報を活用し知覚情報等を推定する AI 技術等の社会受容性確保に向けた調査研究」報告書の該当部分を参照できるように記述している。

ガイドライン①脳情報データの取り扱いにおける同意取得

データの適切な取り扱いに十分留意しないと、脳情報と個人特性の紐づけなどによりプライバシーが侵害される可能性や、データ取得・利用・第三者提供などにおける意思決定プロセスが適切に行われない可能性がある。脳情報提供者に対し、広い同意（特定の用途の範囲において活用の自由度を認める同意の在り方）、オプトイン、仮名化等の処理を適切に組み合わせることでそのプライバシーや意思決定を保護すべきである。また、脳情報提供者から「特定のデータをモデルから撤回するのは難しいため、基本的には一度データを提供した後にデータを消去することはできない」ということに対する同意を脳情報取得時に得ることを推奨する。

現時点では、脳情報から個人が特定できるデータを研究機関から提供することは想定していないが、今後、企業自身が脳応答と個人の特徴を関係づけるような事業を実施した場合には、従来存在しなかった新たなプライバシー関連課題が生じると考えられる。例えばギャンブル・アルコール・ソーシャルゲームなどの依存症状を生む商品や、消費者金融など社会経済的な状況と関連するコンテンツの広告などの際に、脳情報提供者と彼ら個人の依存性・社会経済的指標データが関連付けられ、他人への商品・パーソナライズに利用されるといったケースが想定される。このため、企業自身が脳情報データを収集・共有するにあたり、個人の脳情報及び解読された結果を企業が利用する際のプライバシーをどのように守るべきか、脳情報 AI を取り扱う事業者・研究者が、どのタイミングで誰の同意を取る必要があるかといった点については、引き続き十分な検討が必要である。さらに、研究用途、商用用途、医療用途など、どこまでの用途を想定するかによって、データライフサイクルにおける課題が異なってくると考えられることから、脳情報の提供者に対しては、自身の脳情報の処理方法・利用者・利用目的に関する同意を、脳情報取得時に得るべきである。内容としては仮名化、匿名化、オプトアウト時の対応、用途・目的（研究・商用など）等を説明の上、同意の取得をする必要がある。

また、本人の属性情報・日常の行動データと脳情報の紐付けについては現段階ではそうしたデータは含まれていないが、「本人の人種・信条・社会的身分・病歴・犯罪の経歴・犯罪により害を被った事実その他本人に対する不当な差別、偏見その他の不利益が生じないようにその取扱いに特に配慮を要するものとして政令で定める記述等が含まれる個人情報」は、個人情報保護法第 2 条第 3 項で「要配慮個人情報」として扱われている。そのため将来的にこれらの要配慮個人情報へ該当する可能性が高いデータを収集する場合、本人の同意が必要となる。

特定の個人のデータを消去し、モデルを作り直すことは困難である。そのため、研究開発の関係者は、現時点ではデータから構築したモデルから個人を特定できないことを保証したうえで、脳情報提供者から作成したモデルの中からデータを事後的に消去することは困難であることに同意を得ることを推奨する。

想定される具体的な要配慮ユースケース→補足資料ユースケース 2.1.1, 2.2.2 を参照

ガイドライン②脳情報データのライフサイクルにおいてとるべき対応

脳情報を収集～加工（公的な研究機関で実施を想定）し、民間利用のために提供・共有する一連のプロセス、具体的には研究（脳情報の取得）、AI 開発時、利用時、事後において総合的にモニタリングを行い、不適切な事例が発見された場合には早急に情報公開を行う必要がある。また脳情報提供者にはそうしたプロセスを開示した上で同意を得ることを推奨するとともに、各プロセスで第三者の倫理委員会等の評価を受けることを推奨する。

脳情報の収集～加工及びその後の利用に至るまでのプロセスにおいて、利用目的や利用体制、データ管理体制等について、適切に脳情報データの取得および利用が行われているかモニタリングを行う必要がある。特に利用時（不適切な利用がされていないかの評価）、事後（利用において悪影響はないかの評価）に留意し、不適切な事例が発見された場合には速やかに対応を行い、その内容について情報公開を行うことを推奨する。脳情報 AI モデルを利用する事業者は、原則的に元の個人の脳情報データにアクセスすることは技術的にも不可能であり、事業的にも不要であると想定されるが、学習済みのモデルに関してリバースエンジニアリングの事例など、元の脳情報・個人の特定可能性が出てくるような事例がでてきた場合は、速やかに関係者に情報を共有することを推奨する。

ガイドライン③脳情報をモデル化することによる差別・スティグマ助長リスクへの配慮と対応

性別・年代別・人種別・疾患別などの特定の集団を用いた脳モデルの作成や、それを用いて広告表現などに応用することは、より公正かつ倫理に適うビジネス利用を促進するうえで役に立つと考えられる。一方で、脳の特徴の比較や差異を発信することは排外主義や差別を助長するリスクがあるため、モデルの作成・利用・情報発信の際には十分な注意を払うべきである。

脳情報 AI は発達中の技術であるため、特定の性別・年代・人種・疾患等のプロファイリングに基づく脳の特徴を分析することで一般の人々の誤った解釈を助長し、結果として差別やスティグマ（特定の属性を持つ人に対するネガティブな態度）につながる恐れがある。例えば、特定の精神疾患などに関わる脳の情報表現特性の比較や差異を発信することは曲解（「●●障害の人の脳は、人の顔をそれと認識できない」など）と拡散、そして排外主義や差別を助長することにつながる可能性があるため、そうした個別のモデルの作成・利用・情報発信に関しては ELSI や個人情報保護に関する専門家等への相談を行うなどの対応を実施すべきである。

想定される具体的な要配慮ユースケース→ユースケース 2.1.1 を参照

ガイドライン④特定属性に限った脳のモデル化、および利用による潜在的なステレオタイプ助長のリスクへの配慮と対応

性別など特定の属性に限った脳情報モデルを作成し利用する場合、利用する事業者が意図しない形、あるいは消費者が意識的に気づくことのできないレベルで、特定ステレオタイプの強化や特定集団排除のリスクがあると考えられる。ステレオタイプの強化を避け、また包摂性を確保するため、リスクを認識し注意を払うべきである。また問題が生じた場合、脳モデルの作成に当たって、どのような集団を対象としたかなどの開示が求められる。

ターゲティング（特定の属性を持つ集団における固有のニーズに応じた製品・広告等の設計を行うこと）はマーケティングにおける通常の行為であると考えられる。一方で脳情報 AI を利用した場合により特徴的に発生しうるリスクも想定される。例えば、調理器具・掃除道具などのマーケティングに脳情報 AI を利用する際に、女性の脳情報 AI モデルだけを用いることによって、女性だけが買いたくなるような商品や広告となる傾向が強化されるような場合が想定される。結果として、ステレオタ

イブの増長（家事は女性がするもの）や、一部の人のみに最適化された製品となってしまう（男性が使いづらいと感じる）などのリスクがある。こうした事象は従前の企業活動によっても起きうるが、脳情報 AI を用いることによって、そうした作用が非意図的・非明示的に起こることが想定される（事業者が意図していなくても、また消費者が意識的に気づかなくてもそうした作用が惹起される可能性）。そのため、利用に際しては、他のマーケティング手法の利用と同様に、そうしたステレオタイプ形成・強化のリスクについてはあらかじめ認識し、その影響をモニタリングすることを推奨する。このようにして、人が持つ自然な多様性を十分に尊重することが求められる。また問題が生じた場合、脳情報 AI モデルの作成に当たって、どのような集団を対象としたかなどの開示が求められる。

尚、ガイドライン③と同様に、性別・年代・人種・疾患などの特定の集団を用いた脳情報 AI モデルを作成する場合のリスクへの配慮を記載しているが、モデルから予測される認知的な行動をビジネスに利用する時に起こりうる潜在的なリスクとして、特定の集団の持つ傾向を助長することで、人々の持つ多様なあり方を排除する可能性があることを指摘しており配慮が必要である。

想定される具体的な要配慮ユースケース→ユースケース 2.1.1, 2.2.2 を参照

ガイドライン⑤特定の脳の特性に関わる配慮と対応

特定の脳の特性とは、個人的な脳の状態によって、他の人と比べて時に影響を受けやすい特徴を想定している。例えば、依存症傾向がある人物は依存対象物質の知覚的な入力により通常よりも渴望感が増すといった可能性があるため、そのような特性を利用した事業活用は行うべきではない。また意図的ではない利用の際も、そうした悪影響がないかのモニタリング及び悪影響が発見された場合の情報公開や対策が求められる。

研究段階においては、脳の情報処理に関する特定の特性がある人々を対象とする研究自体に意義はあると考えられるが、その際も要配慮個人情報に関するガイダンスを参照し、同意取得を行うことを推奨する。

脳情報 AI を利用することで特定の脳の特徴を持つ人々に対して効果の大きいコンテンツを制作することが期待されるが、依存症患者の渴望を高める、精神症状を悪化させるなど結果として脳の脆弱性を利用した悪影響が出るリスクがある。例えば、ギャンブル・アルコール・ソーシャルゲームなど依存症状を生む商品において、脳情報解析を通じて依存症特有の脳応答が発見される可能性がある。他にも、自身の体型や顔にコンプレックスを抱える人物は、美容に関する情報入力に更にコンプレックスを抱くようになってしまうといった状況も想定される。同様に消費者金融などの分野においても、貧困層・経済的弱者の脳情報分析に基づいて、彼らの渴望感や焦燥感を効果的にあおり、事業に活用する可能性などが考えられる。勿論、すでに依存症になっている患者、または依存しやすい傾向を持つ人々などの脳情報プロファイリングに研究的意義はあり、そのような情報がより公正な市場の形成に貢献することは期待される。

そして、依存リスクのある商材を扱うときは脳情報 AI の分析の結果を活用することによって、従来のマーケティング手法では成しえなかった事前評価として、依存患者の渴望感により影響を与えないかなどの評価を行い配慮すべきである。また、特定の精神症状を悪化させるようなプロファイリングや、精神症状を持つ患者に特化したマーケティングは行うべきではない（摂食障害・醜形恐怖など）。

不作為的に脆弱性ゆえの悪影響が及ぼされることも想定できるが、その場合にはモニタリングや情報公開、対策などのしかるべき対応が求められる。

また、特定属性を持つ人の脳情報収集においては、改正個人情報保護法・厚生労働省「人を対象とする生命科学・医学系研究に関する倫理指針ガイダンス」（令和3年）が示す要配慮個人情報を取得する際の配慮が求められる。特に、健康診断・診療等の業務を通して得た健康に関する情報・病歴・障害の有無等が、特定属性を持つ人の脳情報収集に関連性が高い要配慮個人情報となる。要配慮個人情報取得において、医学系指針上の「インフォームド・コンセント」取得は必須ではないが、取得しない場合でも、研究者等は原則として研究対象者等の「適切な同意」を受けなければならないため、ガイダンスに則った事業計画が求められる。

想定される具体的な要配慮ユースケース→ユースケース 2.1.1, 2.1.2 を参照

ガイドライン⑥意識的な自覚なく消費者の思想・価値観に与える影響に関する配慮と対応

脳情報の解析や利用により、意識的な自覚を伴わず人々の思想・価値観を変化させる可能性がある。そうした目的・手法の開発・利用においては特に配慮が必要で、無意識に人の思想・価値観を変えるような刺激生成法を探索したり、利用したりするような場合に関しては第三者による倫理的判断を求めるべきである。

消費者の意識的知覚を伴わない反応を脳情報解析に基づいて探索し、それを利用する用途が想定される。具体的には、消費者自身は自らの意志によって行動していると感じているが、実際は脳情報解析に基づいた広告やそのカスタマイズなどによって嗜好や購買行動が無自覚的に誘導されている、という場合などが挙げられる。これらは既存の広告等における倫理的課題と質的には同様であるものの、脳情報を活用することにより効果量を増大させる可能性があることから、知覚不可能な影響を及ぼしうるような脳情報の活用法自体への倫理的配慮や、必要に応じて脳情報 AI モデルを使っ活用していることを開示できるような透明性の確保が必要である。

さらに、思想・信条に関する広報活動等に脳情報が利用される可能性があるが、民主主義・基本的人権・言論および表現の自由・法と秩序などの社会的価値に反する活動への扇動などを目的とした広告・コンテンツへの脳情報 AI の利用は行うべきではない。この点に関しては、将来的には、客観的な審査が可能な第三者委員会を設置するなど、社会的な枠組みが必要になると考えられる。

尚、ガイドライン⑤「特定の脳の特性に関する配慮」の項目も合わせて対応する必要がある。

想定される具体的な要配慮ユースケース→ユースケース 2.1.3, 2.2.2 を参照

ガイドライン⑦脳情報利用に関する透明性の確保

脳情報に基づいて訴求される対象自体の精度・安全性の担保が必要である。また、訴求の結果生じる影響についても考慮されるべきである。

脳情報 AI は発達段階であるため、精度や信頼性・解釈の妥当性について未確立の分野である。そのため、利用する事業者や消費者の誤った認識（過剰な信頼・期待）に留意すべきである。具体的には、「脳情報解析に基づいた」などの文言使用により消費者の信頼度を虚構的に高め、低精度のサービスを提供する場合などが想定される。脳情報 AI 構築のプロセス、精度や技術的限界については透明性を確保し、積極的に利用者・消費者へ伝達する努力を行うべきである。事業に関わる Web サイト上で脳情報 AI の利用を明記することを強く推奨する。

また、脳情報 AI モデルを作成する際にどんな刺激を提示したかはモデルの性能に直結するため、そうした学習済みモデルの背景にある、どんな刺激をどれだけ被験者に提示したかについての情報開示、及びその提示方法に関する規格化・標準化といった信頼性・再現性の確保も今後求められると考えられる。

想定される具体的な要配慮ユースケース→ユースケース 2.1.1, 2.2.3, 2.3.1 を参照

ガイドライン⑧脳情報を利用した事業に関する公平かつ包摂的なアクセシビリティの担保

脳情報解析は単純な AI よりも高額になることが想定されるため、アクセス可能な対象が限定されることが懸念される。包摂的なアクセシビリティの担保が求められる。

脳情報 AI モデルの構築にあたっては大きな経済的コストがかかっており、学習済みモデルの利用であっても、それを利用した製品・サービスは単純な AI を利用した商品よりも数倍程度の価格差が生じることが予想される。例えば、将来的に認知機能の低下予防サービスや PTSD 等の克服に係るサービス等に脳情報 AI が応用された場合、貧富の差による健康格差が生じる可能性が考えられる。特に多くの人にとって有益な製品・サービスが登場した場合、福祉政策への反映など包摂の確保に向けて官民一体となって制度設計等に努めることが求められる。

想定される具体的な要配慮ユースケース→ユースケース 2.3.2 を参照

なお、本ガイドラインは国内外の技術研究開発、倫理的、法的、社会的動向等を踏まえ、今後見直し、修正を行う可能性があることを付記する⁶。

※具体的な想定ユースケースなどの補足資料については別ドキュメント「脳情報を活用し知覚情報等を推定する AI 技術の開発・活用ガイドライン（補足資料）」を参照。

⁶生成 AI と脳情報型 AI の組み合わせ技術も 2024 年 1 月時点で報告されてきており (Takagi, Y., & Nishimoto, S. (2023). High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*)、今後の注目課題としている

4. 脳情報を活用し知覚情報等を推定する AI 技術の活用ガイドライン作成委員会 関係者一覧

2021-22 年度「脳情報を活用し知覚情報を推定する AI 技術等の社会受容性確保に向けた調査研究検討会」事業における検討会構成員

(敬称略。五十音順。)

委員長	標葉隆馬	大阪大学社会技術共創研究センター 准教授
副委員長	田口隆久	情報通信研究機構 未来 ICT 研究所脳情報通信融合研究センター 副研究センター長
委員・外部評価委員兼任	伊地知千織	味の素株式会社食品事業本部 食品研究所 Group Executive Specialist
委員	今村泰彦	VIE STYLE 株式会社 代表取締役
委員	柏岡秀紀	情報通信研究機構 未来 ICT 研究所脳情報通信融合研究センター 統括
委員	岸本充生	大阪大学データビリティフロンティア機構 教授
委員	谷藤学	早稲田大学先進理工学部 生命医科学科 客員教授
委員	中川裕志	理化学研究所 革新知能統合研究センター チームリーダー
委員	西堤優	情報通信研究機構未来 ICT 研究所脳情報通信融合研究センター 研究員
委員・外部評価委員兼任	萩原一平	一般社団法人応用脳科学コンソーシアム 理事・事務局長
委員	服部雅史	立命館大学総合心理学部 総合心理学科 教授
委員	服部宏充	立命館大学情報理工学部 教授
委員	林隆介	産業技術総合研究所人工知能研究センター 主任研究員
委員・外部評価委員兼任	前田直哉	株式会社エヌ・ティ・ティ・データソーシャルイノベーション事業部 テクニカルグレード
委員	横井惇	情報通信研究機構 未来 ICT 研究所脳情報通信融合研究センター 研究員
委員	横野恵	早稲田大学社会科学総合大学院 社会科学部 准教授
情報提供者	西本伸志	大阪大学大学院生命機能研究科 脳神経工学講座 教授
情報提供者	黒田知宏	京都大学医学部附属病院 医療情報企画部 教授
情報提供者		一般社団法人 遺伝情報取扱協会

**2023 年度「脳情報を活かしたサイバー空間の感性評価技術の社会実装に向けた調査研究」事業
における外部評価委員**

(敬称略。五十音順。)

外部評価委員	金井良太	株式会社アラヤ 代表取締役
外部評価委員	武見充晃	慶應義塾大学 大学院理工学研究科 特任講師
外部評価委員	中村健二	DIC 株式会社 アドバンストリサーチセンター プロジェクトマネージャー
外部評価委員	中村壮志	株式会社竹中工務店 技術研究所 主任研究員
外部評価委員	中村文哉	アサヒクオリティードイノベーションズ株式会社 コアテクノロジー研究所 第二部シニアマネージャー

ガイドライン作成事務局

(株式会社 NTT データ経営研究所ニューロイノベーションユニット)

茨木拓也・桃木美登里・貝谷美帆・鶴巻明梨・川瀬悠里・田島美咲・佐橋紀杏・澤田碧砂