

第3課題

生成AIの基礎知識

Artificial Intelligence Advisor



生成AIの仕組み

今回のテーマは、現在社会的に大きな注目を集めている生成AIを理解するうえで欠かすことのできない基盤部分になります。私たちが日常的に触れているChatGPTや画像生成AI、あるいは最近ニュースなどで頻繁に耳にするAI技術の多くは、この「生成AI」という大きな枠組みに含まれています。まずは、生成AIがどのような原理で動いているのか、その代表的な仕組みとして知られる「大規模言語モデル」「拡散モデル (Diffusion Model。データにノイズを加えてから段階的に除去し、ノイズから高品質な画像などを生成する仕組み)」「GAN (Generative Adversarial Network。生成器と識別器が競い合いながら学習し、より本物らしいデータを生成する仕組み)」という3つを中心に、詳しく説明します。

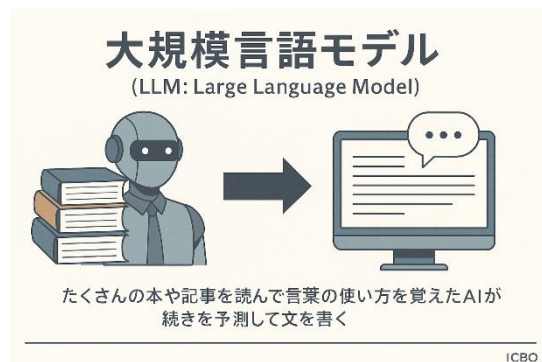
1. 生成AIとは何か

生成AIとは、その名の通り「新しいものを生み出すAI」のことを指します。従来のAIは、与えられたデータを分析して分類する、あるいは将来の値を予測する、といった用途が中心でした。たとえばメールがスパムかどうかを判定したり、売上や需要を予測したり、工場の異常を検知したりするのが典型的な従来型AIの仕事です。しかし生成AIは一歩進んで、文章や画像、音声、さらには動画まで、ゼロから新しいコンテンツを生み出せるという点で革新的です。

私たちがChatGPTに質問をすると、文章が即座に生成されます。画像生成AIに「空を飛ぶ猫を描いて」と指示すれば、空飛ぶ猫の画像が瞬時に作り出されます。つまり生成AIは「過去のデータをもとに未来を予測する」だけでなく、「新しいものを生成する」能力を持っているのです。この能力こそが、近年のAIブームを支える原動力になっています。

2. 大規模言語モデル (LLM: Large Language Model)

まず取り上げるのは大規模言語モデル、略してLLMです。ChatGPTを代表とする現在のAIブームの中心にあるのがこの仕組みです。大規模言語モデルとは、膨大な文章データを読み込み、その中に含まれる言葉のつながり方や文脈のパターンを学習することで、自然で意味の通った文章を自動的に作り出すことができるAIのことです。初心者向けに言えば「たくさんの本や記事を読んで、その言葉の使い方を覚えたAIが、続きを予測して文を書く」と考えると理解しやすいでしょう。



(1) 基本的な仕組み

LLMは、大量のテキストデータを学習することで言語のパターンを身につけています。テキストを細かく分割した単位（これを「トークン」と呼びます）ごとに、次にどのトークンが来る確率が高いかを計算する、ということを繰り返しています。たとえば「今日は天気」と入力すると、その後に「良い」「悪い」「晴れ」などの単語が続く可能性を確率的に判断し、最も自然な続きを選んで出力する、という仕組みです。

この「次に来る単語を予測する」という単純な原理を、高度な統計計算と莫大な学習データで強化することで、まるで人間が書いたかのように自然な文章が生み出されるわけです。

(2) 強みと応用

LLMの強みは、その応用範囲の広さにあります。質問への回答、文章の要約、翻訳、文章の添削、メール文の自動生成、プログラミングコードの作成や修正まで、多岐にわたる作業をサポートできます。こうした幅広さは、人間の知識を言葉で表現できる領域ならほぼすべてカバーできる、と言っても過言ではありません。

(3) 限界と課題

ただしLLMも万能ではありません。LLMは「確率的に最もらしい言葉の並びを生成する」だけなので、必ずしも事実を正しく答えられるとは限りません。実際には存在しない情報をもっともらしく答えてしまう現象、いわゆる「ハルシネーション」が代表例です。また、学習データに含まれる偏りや差別的表現がそのまま出力されるリスクもあります。

3. 拡散モデル (Diffusion Model)

次に取り上げるのは拡散モデルです。拡散モデルとは、データにノイズを加えてから段階的に除去し、ノイズから高品質な画像などを生成する仕組みです。これは主に画像生成の分野で大きな成果を挙げている技術です。

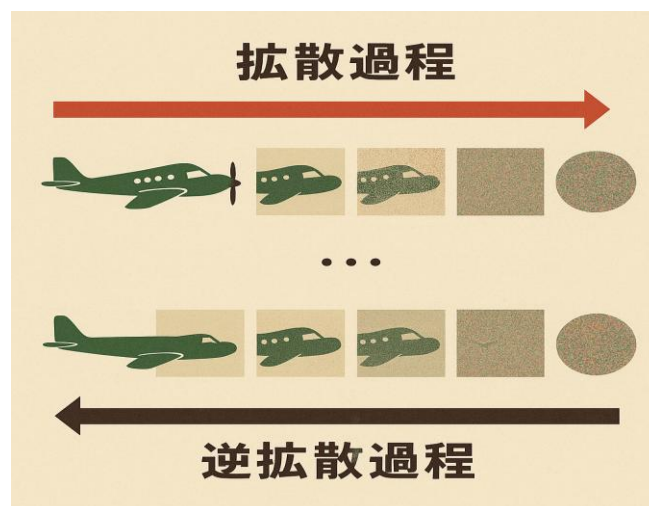


図 拡散モデル

(1) 仕組みの概要

拡散モデルとは例えると、まっさらな「ザラザラした画像（砂嵐のような模様）」から出発して、少しずつ形や色を整えていき、最終的に意味のある画像を生み出すAIの仕組みです。最初は何の形もない無秩序な点の集まりですが、AIが繰り返し手を加えることで、点がつながり、模様がはっきりし、やがて写真のような風景や動物の姿が浮かび上がります。

ここで、私たちが入力する文章が大きな役割を果たします。「夕暮れの海辺にいる猫」と指示すると、AIはその言葉の内容を手がかりに、ザラザラの中から「猫らしい形」「夕暮れらしい色」「海辺らしい背景」を少しずつ引き出ししていきます。つまり、拡散モデルは「無秩序なノイズを、言葉を道しるべに整えていく」ことで、新しい画像を作り出す仕組みなのです。

(2) 画像生成への応用

この仕組みを利用すると、「空飛ぶ猫を描いて」とテキストで指示するだけで、拡散モデルは真っ白なノイズの中から猫らしい形、空らしい背景を段階的に再構築し、最終的にリアルな画像を出力します。Stable DiffusionやMidjourneyといったサービスが代表例です。

拡散モデルの強みは、きわめて高解像度でリアルな画像を生成できる点です。後述のGANに比べて学習が安定しているといわれており、多様なスタイルに対応できることも特徴です。たとえば写真風のリアルな表現から、油絵のような芸術的スタイル、アニメ風のイラストまで、幅広い表現を可能にします。

4. GAN（敵対的生成ネットワーク）

3つ目に紹介するのがGAN（Generative Adversarial Network 敵対的生成ネットワーク）です。GANとは、生成器と識別器が競い合いながら学習し、より本物らしいデータを生成する仕組みのことです。

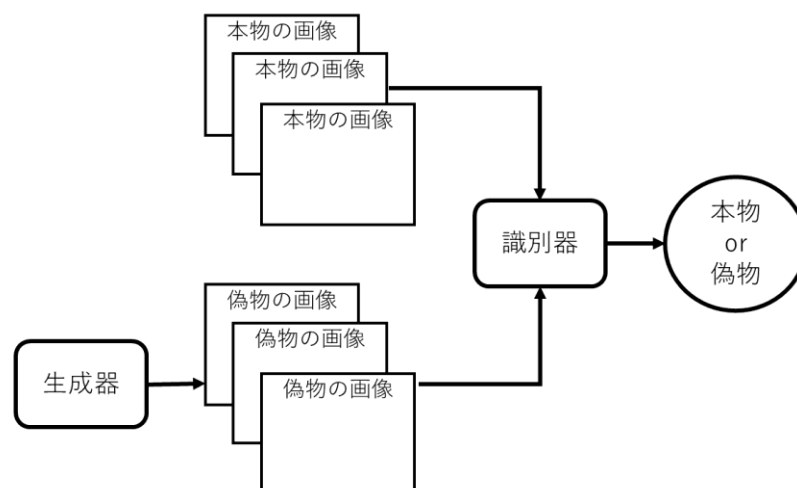


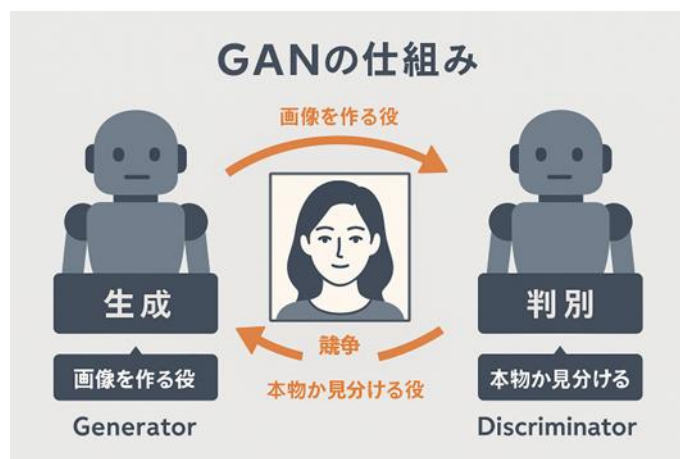
図 GAN

(1) 仕組みの概要

GAN（ガン）と呼ばれる仕組みは、人工知能の分野で「新しい画像を生み出す方法」として有名になった技術です。特徴的なのは、二つのAIを向かい合わせにして競わせるように学習を進める点にあります。一方は「画像をつくる役」で、もう一方は「それが本物か偽物かを見分ける役」です。つくる側はできるだけ本物に近い画像を描こうとし、見分ける側は「これは本物だ」「これは偽物だ」と判定しようとします。このやり取りが何度も繰り返されるうちに、つくる側はだんだんと巧妙になり、本物と区別がつかないほど自然な画像を描けるようになります。

例えば、最初のうちはつくる側のAIが描く顔の画像は歪んでいたり不自然な模様にししか見えません。しかし、見分ける側にすぐ「偽物だ」と見破られてしまう経験を重ねるうちに、よりリアルな目や口の形、髪の毛の流れを表現できるように改善していきます。そして、見分ける側もまた、巧妙になっていく画像を見抜こうと精度を上げていきます。結果として、両者のせめぎ合いが続くことで、画像を生成する能力が非常に高まっていくのです。

GANの本質は、この「競争関係」にあります。つくる側と見分ける側が互いに相手を出し抜こうと努力することで、どちらの能力も引き上げられ、最終的に人間の目には本物の写真と見分けがつかないほどの画像が生み出されるようになります。この仕組みは画像生成だけでなく、映像や音声の生成にも応用されてきました。つまりGANとは、二つのAIを競わせることによって、新しいものを生み出す力を引き出す仕組みなのです。



(2) 応用分野

GANは特に画像生成の分野で注目され大きな話題を呼びました。たとえば実在しない人物の顔写真を作り出す「This Person Does Not Exist」というサイトはGANの技術を使っています。また、古い写真の修復、低解像度の画像を高解像度に変換する「超解像技術」、アートやデザインの分野でも幅広く応用されています。

ただしGANは、学習が不安定になりやすいという弱点もあります。生成と識別の競争がうまく釣り合わないと、意味のない画像しか出てこなかったり、学習が途中で止まったりすることがあるのです。そのため、現在の主流は拡散モデルへ移行しつつありますが、GANのアイデアは今も多くの研究に生かされています。

5. 3つの仕組みの比較と位置づけ

ここで、大規模言語モデル、拡散モデル、GANを比較して整理してみましょう。

LLM：テキスト生成が得意。人間の会話、文章作成、知識検索の補助などで広く活躍。

拡散モデル：画像生成の最前線。写真からアートまで高品質なビジュアルを作り出せる。

GAN：リアルな画像生成に強みがあるが、学習の安定性に課題。ただし「競争による精度向上」という仕組みは重要。

このように、それぞれのモデルには得意分野と課題があり、研究者や開発者はこれらを組み合わせたり改良したりしながら、さらに進化した生成AIを作り出しています。

6. まとめと今後の展望

最後にまとめましょう。生成AIの仕組みは大きく分けて3つ、すなわち大規模言語モデル、拡散モデル、GANに代表されます。これらの仕組みは一見別々に見えますが、根底には「大量のデータをもとに、確率的にもっともらしい新しいデータを作る」という共通の発想があります。

近年はテキストと画像、音声、動画などを組み合わせた「マルチモーダルAI」が登場しており、文章を入力すれば画像や音声まで自動で生成できるようになっています。今後はさらに、生成AIがビジネスや教育、医療、芸術などあらゆる分野に広がっていくと予想されます。その一方で、偽情報の拡散や著作権侵害、倫理的な問題といった課題にも対応していく必要があります。

トークン・パラメータ・学習データの概念

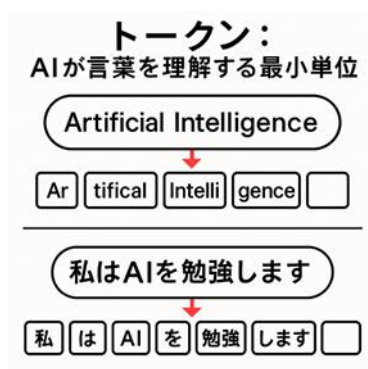
ここからは生成AIを理解する上で欠かすことのできない3つの基本概念「トークン」「パラメータ」「学習データ」について詳しくお話ししていきます。これらは普段、なにげなく生成AIを利用していると意識する機会が少ない言葉かもしれませんが、しかし、生成AIがどのようにして人間のように自然な文章を作ったり、複雑な会話をこなしたりしているのかを理解するためには、押さえておくべき基盤になります。逆に言えば、この3つの要素を理解することで、生成AIの仕組みや特徴、そして限界までもがより鮮明に見えてきます。本章ではその点を一つずつ丁寧に、具体例を交えながら解説していきたいと思います。

1. トークン：AIが言葉を理解する最小単位

まずは「トークン」について説明しましょう。人間にとっての最小単位は「文字」や「音」ですが、AIにとっての最小単位は「トークン」です。トークンとは、文章や単語を分解した際の断片的な要素のことを指します。

たとえば英語で「Artificial Intelligence」と書いた場合、人間は自然に「Artificial」と「Intelligence」という二つの単語として読み取ります。しかしAIは必ずしもそのまま単語として処理しているわけではありません。多くの場合は単語ごとに近い形で分けられますが、必ずしもきれいに単語単位になるとは限りません。たとえば「Intelligence」が「Intelli」と「gence」に分けられることもあれば、逆に複数の単語がまとめてひとつのトークンとして扱われる場合もあります。

日本語の場合はさらに特徴的です。日本語には単語の区切りとしてスペースが存在しません。そのため「私はAIを勉強します」という文章を処理する場合、AIは「私」「は」「AI」「を」「勉強」「します」といった単位に区切って認識します。この「どのように区切るか」が実は非常に重要で、分割の仕方によってAIの理解や生成の精度が大きく変わります。



(1) トークン化の歴史的背景

AIが言語を扱う際に「トークン化」という前処理を行うのは、自然言語処理の研究の積み重ねの中で生まれた工夫です。初期の言語処理では「単語ごとに区切る」程度の方法が中心でしたが、言語には複雑さがあり、英語の動詞変化や名詞の複数形、日本語の助詞や活用など、単語単位だけでは十分に扱えない課題がありました。

そこで近年広く使われるようになったのが「サブワード分割」です。サブワード分割の一つの手法に「バイトペアエンコーディング (BPE)」という技術があります。これは文字などの小さな単位から始め、よく出てくる組み合わせを一つのまとまりとして扱う方法です。

サブワード分割とは、自然言語処理において、単語をさらに小さな構成要素であるサブワードに分割する方法のことです。従来の自然言語処理では、自然言語モデルを構築する際に未知の単語に遭遇した場合には、この未知の単語を処理できませんでした。また、複数の単語が組み合わさり、一つの巨大な語彙になっている場合、すべての単語を語彙に含めると、自然言語モデルが非効率になってしまうという課題もありました。そこで、未知の単語や自然言語モデルの語彙サイズを管理しやすくするため、サブワード分割が広く利用されるようになりました。この仕組みにより、AIは未知の単語や造語にも柔軟に対応できるようになりました。たとえば「スマホ依存症」という新しい言葉が現れても、「スマホ」「依存」「症」など既知のかたまりに分けて処理できるのです。

サブワード分割を実現するために「バイトペアエンコーディング (BPE)」等の主要なアルゴリズムが開発されています。バイトペアエンコーディング (BPE) とは、テキストをより短い単位に圧縮したり、サブワード (部分語) 単位で表現するための手法です。

【バイトペアエンコーディングの仕組み】

1. テキストを文字単位に分割

すべての単語を文字単位に分割します。

例：“banana” → [“b”, “a”, “n”, “a”, “n”, “a”]

2. 最も頻出の隣接ペアを探す

単語リストの中で出現する文字ペア (隣接する2文字) の頻度を計算します。

例：“a n” が 2回

“n a” が 2回

“b a” が 1回

→ もっとも多い “a n” をマージ

3. そのペアを1つのトークンに統合

最も頻度の高いペアを1つの新しいトークンに統合します。

例：[“b”, “an”, “an”, “a”]

4. 規定回数繰り返して語彙を育てる

2と3を規定回数繰り返して、頻度の高いペアを一つのトークンに統合していきます。

例：語彙に “ban”, “ana” のようなサブワードを追加

(2) トークン予測による文章生成

生成AIが文章を作る仕組みは、実は非常にシンプルです。AIは「次に来るトークンは何か」を確率的に予測しているにすぎません。たとえば「今日は天気が」と入力されたとき、AIは次の候補として「良い」「悪い」「晴れ」「雨」などのトークンを考え、その中で最も確率が高いものを出力します。これを延々と繰り返していくことで、私たちが読んで自然に感じる文章が生成されるのです。

2. パラメータ：AIの知能を支える「重み」

「パラメータ」とは、AIの中に無数に並んでいる「処理や生成の仕方を方向づける設定値」のことです。ひとつひとつはごく小さな数で、つまみやダイヤルのようにAIのふるまいを少しだけ変えます。文章を作るAIなら、「この文字の並びの次には、どんな言葉が来やすいか」を判断するとき、内部の計算でこの数値たちが効き方の強さや方向を決めます。画像を作るAIなら、どんな形や色合いを強めるかに影響します。個々のパラメータは微小な役割でも、膨大な数が同時に働くことで、私たちが読むと自然に感じる文章や、見て納得できる画像が生まれます。

これらの数値は、人が一つずつ手で決めるものではありません。たくさんの例を読み込む過程で、自動的に調整されて固まります。その後、私たちがAIを使うときには、そのときに確定しているパラメータが応答の作り方を左右します。パラメータの数が多いほど、より細かな違いを表現できる“器”が大きくなる傾向はありますが、数が多いだけで必ず賢くなるわけではありません。どんなデータで、どのように調整されたかも重要です。なお、具体的なパラメータの総数は公開されないことも少なくありません。

要するに、パラメータはAIの性格や得意・不得意を形づくる「内部設定の束」です。ひとつでは何もできませんが、無数の小さな設定が重なり合うことで、意味の通った返答や自然な表現が可能になります。

(1) 人間の脳との類似点

人間の脳にはおよそ数百億個の神経細胞（ニューロン）があり、各ニューロンは多数（しばしば数千規模）のシナプスで結び合わさっています。私たちは新しいことを経験したり練習したりすることで、こうした結びつきが強まったり弱まったりし、少しずつ上達していきます。たとえばピアノの練習を重ねると、最初はぎこちなかった指の動きがだんだん滑らかになるのは、脳のつながりが「その動きをしやすくする方向」に調整されていくことが大きな理由のひとつです。実際には筋肉の慣れや体の使い方の工夫も関わりますが、経験に合わせて脳が変化するという点が要点になります。AIにおけるパラメータは生物学的な仕組みとは異なりますが、役割としてはこれに似ており、学習を重ねるたびに内部の設定値が少しずつ調整され、入力の理解や出力の作り方が洗練されていき、結果として複雑な判断に相当する処理が可能になります。

(2) パラメータの多さとモデル性能

生成AIの発展の歴史を振り返ると、主にパラメータの数を大きくしていくことによって性能を高めてきました。パラメータが増えると、より複雑な言語の使い方や意味の細かい違いを扱える可能性が高まります。たとえばGPTシリーズも、公開されている範囲では世代が進むにつれてパラメータ数が増え、それに伴い理解や生成の精度が向上してきました。

ただし、パラメータは「多ければ多いほど必ず良い」というわけではありません。計算に必要な資源や電力は膨大になり、学習に使うデータや工夫が不十分であれば、数を増やしても性能が頭打ちになることがあります。そのため近年は「できるだけ少ないパラメータで効率よく学習させる工夫」や、「大規模モデルを圧縮して扱いやすくする方法」にも注目が集まっています。こうして、規模を拡大する方向と効率化を進める方向の両方が研究されているのです。

3. 学習データ：AIにとっての「経験」

3つ目の要素は学習データです。生成AIは、最初から文章を書いたり画像を描いたりできるわけではありません。人間が生まれたばかりでは何も知らず、経験を重ねることによって増えていくのと同じように、生成AIも学習を通じて初めて表現の力を身につけます。学習とは、大量の文章や画像といった情報を繰り返し処理しながら、内部のパラメータ（AIのふるまいを方向づける数値）を少しずつ調整していくことです。

ここで使われる「学習データ」が、生成AIにとっての「経験」にあたります。人間の子どもが本を読んだり会話をしたりして知識や感覚を蓄えるように、生成AIも膨大なテキストや画像を通して言葉の使い方や表現のパターンを学び取ります。つまり生成AIにとって学習データとは、世界を知り、文章や画像を生み出す力を獲得するための「経験」そのものなのです。

(1) 学習データの種類

学習データには多様な種類があります。インターネット上のウェブページ、百科事典やニュース記事、学術論文、SNSの投稿、さらには小説や映画の台本なども含まれます。言語モデルの場合はテキストが中心ですが、画像モデルでは写真やイラスト、音声モデルでは会話の録音データが使われます。

ここで「テキスト」といっても、本や記事を丸ごと一度に処理するわけではありません。一般的には、テキストを数百から数千トークン程度のまとまりに区切り、その単位ごとに学習を行います。ここでいう「トークン」とは、単語やその一部などの小さな文字のかたまりのことです。例えば「Artificial Intelligence」という言葉なら、「Artificial」と「Intelligence」として扱われる場合もあれば、さらに細かい単位に分かれる場合もあります。こうして細かい単位を組み合わせながら学習することで、生成AIは文や段落といった大きなつながりも扱えるようになり、新しい文章を自然につながられるのです。

(2) 質と偏りの問題

生成AIにとって学習データは「量」と「質」の両方が大切です。十分な量がなければ多様な言葉や表現のパターンを学ぶことができませんし、近年の大規模モデルの性能向上は、膨大なデータを使ったことが重要な要因のひとつです。しかし、量さえ増やせば良いというわけではありません。もし誤った情報や差別的な表現が大量に含まれていれば、AIの出力にも同じような偏りが現れてしまいますし、特定の文化や言語圏ばかりに偏ったデータでは他の文化に対応できなくなります。そのため近年は「十分な量を確保しつつ、安全で多様性のあるデータセットを整える」ことが大きな課題になっています。著作権やプライバシーの問題も絡むため、量だけに依存するやり方では限界が見えてきており、質をどう高めるかがより重要になっているのです。

4. 三要素の関係性

ここまで個別に説明してきた「トークン」「パラメータ」「学習データ」は、実際には密接に関係しています。

- * 学習データはAIにとっての「経験」であり、知識の土台。
- * トークンはそのデータを処理するための「言語の最小単位」。
- * パラメータは、学習を通じて最適化される「脳内の回路」。

AIは学習データをトークンに分解し、そのパターンをもとにパラメータを調整していきます。そして出来上がったパラメータを使って、新しいトークン列を予測しながら文章を生成する。これが生成AIの根本的な仕組みです。

これら三要素を理解すると、AIの応用範囲や課題も見えてきます。たとえばカスタマーサポートでは、大量の会話データを学習したモデルが自然な応答を生成します。医療分野では、論文や診療記録を学習したモデルが研究支援や診断補助に使われます。しかし同時に、データの偏りやパラメータの暴走によって誤った判断をするリスクもあります。

従来のAI（機械学習、ルールベース）との違い

ここからは「従来のAIと生成AIの違い」について詳しく解説していきます。今回のテーマは、AIの歴史を振り返りながら、なぜ生成AIがこれほど大きな注目を集めているのか、その理由を理解するために非常に重要なポイントになります。

私たちは普段「AI」という言葉を当たり前のように使っていますが、その内実は時代によって大きく変化してきました。かつては「ルールベースAI」と呼ばれる仕組みが主流であり、その後「機械学習AI」が登場し、そして現在の「生成AI」に至るという大きな流れがあります。これら3つを比較することで、AI技術の進化がどのような意味を持つのか、そして今後どのような未来が広がっているのかを実感していただければと思います。

1. ルールベースのAI

ルールベースAIは、その名の通り人間があらかじめ定めたルールに基づいて動く仕組みです。たとえば「もしXならYする」という条件と結果の組合せを多数用意し、入力をそれらに照らして処理します。診療の現場を支援する例でいえば、「体温が38度以上で咳がある場合は呼吸器系の注意が必要というフラグを立てる」といったルールを設定しておけば、入力された患者データに該当があった際に注意喚起ができます。ルールベースの長所は動作原理が明確で、なぜその結論に至ったかを説明しやすいことにあります。また、ルールの適用範囲内に限れば高い正確さで動作します。銀行の与信や製造の品質検査のように、満たすべき条件が明確な場面では今も実務で広く利用されています。

一方で、人手でルールを定義・保守するため柔軟性に乏しく、新しい状況への対応が難しいという欠点があります。こうした弱点を補うため、データから自動的にパターンを学ぶ機械学習が広く用いられるようになり、現在はルールと学習を組み合わせた設計も一般的です。

2. 機械学習を用いたAI

機械学習AIは、データを使って自らパターンを学習するAIです。人間がルールを一つ一つ与える必要はなく、大量のデータを与えることで機械が自動的に関係性を見つけ出します。

たとえばスパムメール判定の仕組みを考えてみましょう。ルールベースでは「件名に『無料』という単語が含まれていればスパム」といったルールを人間が作っていました。

しかし機械学習（のうち教師あり学習）では、過去の大量のメールに「スパム」「通常」の正解を付けて用意し、機械に読み込ませます。機械は件名や本文の文字の並び、送信元、リンクの有無や書き方などを数値として受け取り、はじめは当てずっぽうに近い予想をしますが、その予想と正解のズレを確かめ、ズレが小さくなるように内部の数値（パラメータ）を少しずつ調整していきます。こうして機械は、人が事前に書いたルールに頼らず、「どの手がかりをどれくらい重く見るか」という判断基準を自分の中に作り上げます。その結果、「無料」という単語の有無のような単純な条件だけでなく、言い回しの組み合わせや文脈の流れ、リンクや記号の使い方といった複数の弱い手がかりを総合して、スパムをより確かに見分けられるようなAIが構築されるわけです。

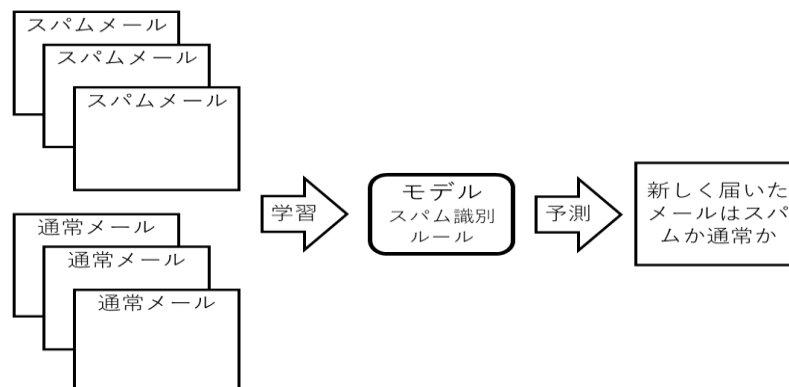


図 分類（教師あり学習）

(1) 代表的な学習手法

機械学習では「学習手法（アルゴリズム）」が多数存在し、何をしたいのかという目的や、扱うデータの性質、求める精度によって使い分けをします。学習手法とは、機械がデータから学んでいくなかで、間違いを減らすようにAIの内部設定（パラメータ）を調整していくための「学び方のレシピ」といえます。目的は、過去のデータから規則性や特徴をつかみ取り、はじめて見るデータにもそれを当てはめて予測や判定ができるようにすることです。世の中の課題やデータの姿はさまざまなので、レシピも一種類では足りません。シンプルで理由を説明しやすい方法もあれば、より柔軟で複雑な関係まで捉えられる方法もあり、問題の種類や必要な精度、説明のしやすさ、用意できるデータ量や計算資源に応じて選び分けます。

代表的な学習手法として、回帰分析や決定木、サポートベクターマシンなどが挙げられます。

サポートベクターマシンとは、2つのグループを最もはっきり分けられる境界線を探す分析手法です。サポートベクターマシンは、「教師あり学習」における「分類」において主に利用されています。サポートベクトルとは、分類境界から最も近い位置にあるデータ点のことです。これらのサポートベクトルは、決定境界（分類線）位置を定めるうえで重要な役割を果たします。サポートベクターマシンは、サポートベクトルと分類境界との間の距離（マージン）を最大化するような最適な分類境界を発見する「マージン最大化」と呼ばれる考え方をを用いて決定境界（分類線）を引いています。なお、分類境界を直線で分けられないときは、柔軟な境界線づくりを可能にするカーネル法と組み合わせ、データの見方を工夫して曲線の境界を引くこともできます。

どの学習手法も、基本的な流れは共通しています。過去の例で予想を作り、その予想と正解のずれが小さくなるように内部の数値を少しずつ調整し、ずれが十分に小さくなったところで完成した“判断のしかた”を未知のデータに適用します。違いは、どんな形のパターンを得意とするか、どれだけ柔軟に形を変えられるか、どの程度理由を説明しやすいか、といった性格の部分です。だからこそ現場では、一つの方法にこだわるのではなく、課題に合うものを選んだり組み合わせたりして、精度・スピード・説明のしやすさのバランスを取っていきます。

(2) ニューラルネットワーク

ニューラルネットワークとは、機械学習のアルゴリズムの一つであり、人間の脳が学習していくメカニズムをモデル化して、人工的にコンピュータ上で再現することで問題を解決しようとする仕組みです。人間の神経細胞に相当する各ノードが層を成して接続されている情報処理のネットワークに入力した情報が、中間層（あるいは隠れ層）と呼ばれるネットワーク内での処理を経て望む情報として出力されるよう、何度も処理方法の調整を行うことで学習していき、最終的に複雑なパターンまで表現できる方法です。写真の中の模様や文章の言い回しのような入り組んだ特徴にも対応しやすい反面、多くのデータと計算が必要になります。

ニューラルネットワークの学習は、誤差逆伝播法（Backpropagation）というアルゴリズムによって行われます。AIはまず、入力データから出力を計算し、その結果と正解データの

差（誤差）を求めます。そして、その誤差が小さくなるように各ニューロンの重みを少しずつ修正していきます。この修正は、出力層から入力層に向かって誤差を「逆伝播」させることで実現します。これを何度も繰り返すことで、AIはより正確な出力を得られるようになります。まるで、人間が間違いを繰り返しながら学び、経験を通じて判断の精度を高めていく過程に似ています。

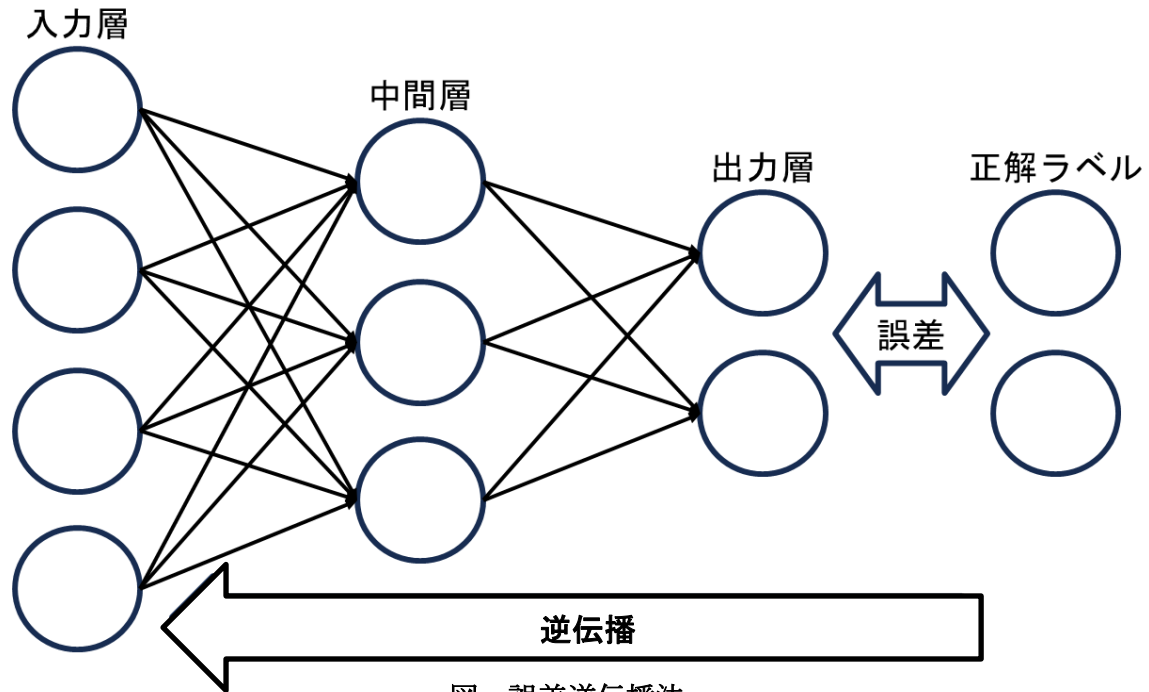


図 誤差逆伝播法

このようなニューラルネットワークの発想自体は1950年代から存在していましたが、当時はコンピュータの処理能力が低く、学習に必要な計算を実用的な時間で行うことができませんでした。そのため、長い間「理論的なアイデア」ととどまっていた。状況が一変したのは、2000年代後半以降のGPU（グラフィック処理装置）の発展とビッグデータの登場です。GPUは画像処理のために開発された装置ですが、大量の数値演算を並列に処理できる特性を持っています。これにより、数百万、数千万というパラメータを持つ巨大なニューラルネットワークでも、現実的な時間で学習できるようになりました。さらに、インターネット上の膨大な画像やテキスト、音声といったデータが「学習の材料」として利用できるようになったことで、ニューラルネットワークは一気に実用化の時代を迎えたのです。

現在では、ニューラルネットワークは画像認識、音声認識、自然言語処理、翻訳、医療診断など、あらゆる分野で応用されています。私たちがスマートフォンの顔認識でロックを解除したり、音声で検索したりできるのも、すべてこの技術が裏で動いているからです。ニューラルネットワークは、もはや特定の研究分野にとどまらず、現代社会の情報インフラを支える基盤技術となっています。

ニューラルネットワークは、単純な数式の組み合わせのように見えて、実際には人間の学習や思考を数学的に再現した極めて柔軟な仕組みです。膨大なデータを吸収し、自らパター

ンを見つけ出して世界を理解しようとするこのモデルこそが、現代のAIを支える「知能の器」といえるでしょう。

• CNN

CNN (Convolutional Neural Network: 畳み込みニューラルネットワーク) は、画像認識を中心に大きな成果を上げてきた深層学習の代表的なモデルです。その特徴は、人間の視覚の仕組みに着想を得ており、画像のような「空間的な構造を持つデータ」から特徴を自動的に抽出できる点にあります。

通常のニューラルネットワークでは、画像の各ピクセルを一行に並べた数値データとして処理しますが、それでは「隣同士のピクセルの関係」や「局所的なパターン」が失われてしまいます。たとえば、人間の目は画像全体を一度に認識するのではなく、小さな範囲ごとに形や色をとらえてそれらを統合しています。CNNはこの考え方を数理的に再現し、画像内の小さな領域を順番に解析して特徴を見つけていく仕組みを採用しています。

CNNの中核を担うのが「畳み込み層 (Convolution Layer)」です。畳み込み層では、「フィルタ」または「カーネル」と呼ばれる小さな行列を画像上にスライドさせながら、各領域との積和演算を行います。これにより、画像の中の線や角、模様などの局所的な特徴が強調されます。学習が進むと、AIは自動的に“何を見ればよいか”を学び取り、低層では線やエッジ、高層では目や口、最上層では「顔全体」といったより抽象的なパターンを捉えるようになります。

畳み込み層の間には「プーリング層 (Pooling Layer)」が挟まれます。これは画像のサイズを小さく圧縮しつつ、重要な情報だけを残す処理です。これにより、計算量が減るだけでなく、多少の位置ずれやノイズに対しても頑健 (ロバスト) な認識が可能になります。最後に、「全結合層 (Fully Connected Layer)」が配置され、畳み込みで得た特徴をもとに最終的な分類や判断を行います。

CNNの登場によって、画像認識の精度は劇的に向上しました。2012年の画像認識コンテスト「ImageNet」で、AlexNetというCNNモデルが従来の誤差率を半分以下に下げたことをきっかけに、深層学習ブームが世界的に広がりました。現在では、顔認識、自動運転、医用画像診断、製造業の欠陥検出など、視覚情報を扱うあらゆる分野で活用されています。

CNNは、「画像を理解するAI」を実現した転換点ともいえる技術です。人間のように“見る”力をAIに与えたこの仕組みは、後に音声や映像、さらには自然言語の分野にも応用され、AIの認識能力を根本から進化させました。

CNN（畳み込みニューラルネットワーク）の基本構造

1. 畳み込み層 (Convolution Layer)

- ・画像の「局所的な特徴（線・角・模様等）」を抽出する層
- ・小さな行列（「フィルタ」「カーネル」）を画像上でスライドさせ、各領域との積和演算を行い、特徴マップを作成する。

【畳み込み処理の手順】

(1) 低層

線・角・模様等を検出

(2) 高層

目・鼻・口のようなパーツを検出

(3) 最上層

顔全体等の高次特徴を検出

2. プーリング層 (Pooling Layer)

- ・画像のサイズを圧縮し、重要な特徴を残しながら計算量を減らす層
- ・重要な情報を残すことにより、小さな特徴差にとらわれることなく、頑健（ロバスト）な認識が可能となる。

3. 全結合層 (Fully Connected Layer)

- ・畳み込みで抽出した特徴をもとに、最終的な分類や判断を行う層

・ RNN

RNN (Recurrent Neural Network: リカレントニューラルネットワーク) は、「時系列データ」や「文のように順序を持つデータ」を扱うために設計されたニューラルネットワークです。CNNが“空間的な構造”をとらえるのに対し、RNNは“時間的な流れ”を理解できるという点に特徴があります。音声、文章、株価、センサーの記録など、前後の関係が意味を持つデータを処理する際に用いられます。

通常のニューラルネットワークは、入力を一度に受け取り、それを処理して出力を出す「一方向の構造」を持っています。しかし現実の多くの情報は、過去の文脈や状態が現在の判断に影響します。たとえば「私はパンを…」という文の後に「食べた」と続くのか「作った」と続くのかは、前の言葉の意味によって決まります。RNNはこの“文脈”を理解するために、内部に「記憶（状態）」を保持する仕組みであるフィードバックループを持っています。

RNNの仕組みでは、ある時点での入力だけでなく、直前の出力（または内部状態）を次の処理に再び入力として渡します。これにより、AIは過去の情報を蓄積しながら、現在のデータを解釈することができます。たとえば音声認識では、今の音だけでなく直前の音の流れを考慮して単語を特定します。文章生成では、これまでに出力した単語をもとに次の単語を決

めていきます。こうした「時間をまたいだつながり」を学習できることが、RNNの最大の特徴です。

しかし、RNNには課題もあります。時間が長くなると、過去の情報が徐々に薄れてしまう「勾配消失」や、学習中にパラメータの更新量を決定する勾配が異常に大きくなり、モデルが安定して学習できなくなる「勾配爆発」と呼ばれる問題が発生し、長期的な依存関係をうまく学習できなくなるのです。たとえば、長い文の冒頭で出てきた情報を文末まで覚えておくことが難しくなります。この問題を解決するために登場したのが、どの情報を忘れるかを取捨選択して記憶する構造をもつLSTM (Long Short-Term Memory) や、LSTMよりもシンプルな構造をもつGRU (Gated Recurrent Unit) といった改良型RNNです。これらは「ゲート構造」と呼ばれる仕組みを導入し、どの情報を保持し、どの情報を忘れるかを自動的に制御できるようにしました。これにより、より長い文脈を安定して扱うことが可能になりました。

RNNは自然言語処理 (NLP) の分野で特に大きな役割を果たしてきました。文章の要約、機械翻訳、感情分析、音声認識など、時間的・文脈的な連続性を持つタスクでは不可欠な技術でした。また、株価予測やIoTデータの解析など、時系列データの分析でも広く使われています。

ただし、RNNには計算の並列化が難しいという制約があり、大量のデータを効率的に処理するには限界があります。この問題を克服する形で登場したのが、後に主流となるTransformer (トランスフォーマー) というモデルです。TransformerはRNNの「順次処理」を必要とせず、文全体の関係を同時に学習できるため、高速かつ高精度な生成AIの基盤となりました。

RNNは今日の生成AIへとつながる重要な架け橋でした。AIが「文脈」や「流れ」を理解できるようになったのは、RNNの登場によって初めて実現したのです。つまりRNNは、「AIが時間を記憶する力を得た技術」といえます。

・生成AIにつながる基礎技術

現在のChatGPTやStable Diffusionのような生成AIは、深層学習の延長線上に生まれた技術です。その土台には、これまで紹介してきたニューラルネットワーク、CNN、RNNなどの研究成果がすべてつながっています。つまり、生成AIは「深層学習の応用」ではなく、「深層学習が到達した新しい表現の段階」と言えます。ここでは、その進化を支えた主要な基礎技術について整理します。

まず、生成AIの前提として欠かせないのが自己教師あり学習 (Self-Supervised Learning) の発展です。従来の機械学習は「正解付きのデータ (教師データ)」を必要としましたが、現実世界ではラベル付きデータを大量に集めることは困難です。自己教師あり学習では、AIが「自分で課題を作り、自分で答えを学ぶ」という方法をとります。たとえば文章データであれば、ある単語を意図的に隠し、周囲の文脈からその単語を予測することで言語の構造を学びます。これによりAIは膨大な未ラベルのデータから一般的な知識や文脈の理解を獲得し、大規模言語モデル (LLM) の基礎を築きました。ChatGPTなどが多様な質問に答えられるのは、この「文脈予測型学習」によって世界中の文章パターンを吸収しているためです。

次に重要なのが、大規模言語モデルを構築するための基盤となるトランスフォーマー (Transformer) という構造の登場です。RNNは時系列処理に優れていましたが、長い文や複雑な依存関係を扱うには限界がありました。Transformerは、RNNの「順番に処理する」構造を捨て、文章全体の単語同士の関係を一度に捉え、重要な箇所に注目することができる「自己注意 (Self-Attention)」という仕組みを導入しました。これにより、AIは文の前後の文脈だけでなく、全体の意味構造を同時に理解できるようになったのです。自己注意は、たとえば「銀行」という単語が「お金の銀行」なのか「川の岸边 (bank)」なのかを、前後の文脈から自動的に判断するような高度な理解を可能にします。

Transformerは2017年にGoogleによって提案され、その後BERTやGPTといった多くのモデルの原型となりました。とくにGPT (Generative Pre-trained Transformer) は、事前学習 (Pre-training) と呼ばれる段階で膨大なテキストを読み込み、一般的な言語知識を習得したのち、特定の目的に合わせて微調整するファインチューニング (Fine-tuning) と呼ばれる構造を持っています。この「事前学習+ファインチューニング」という考え方が、生成AIを現実的に利用できる仕組みを作り上げました。

画像や音声の生成AIにも、同様に深層学習の基礎技術が生かされています。代表的なものがGAN (Generative Adversarial Network: 敵対的生成ネットワーク) です。GANは「生成モデル」と「識別モデル」という2つのAIを競わせることで、よりリアルなデータを生み出します。両者の競争が続くうちに、生成モデルは人間の目でも区別できないほど精巧な画像を生成できるようになります。GANの発想は、AIが「創造性」を獲得するための最初の大きな突破口となりました。

さらに近年では、拡散モデル (Diffusion Model) と呼ばれる新しい手法が急速に普及しています。Stable DiffusionやDALL·Eといった最新の画像生成AIは、この拡散モデルをベースにしています。従来のGANよりも安定して高品質な生成ができるため、研究・商用の両面で主流になりつつあります。

これらの技術に共通しているのは、「AIが世界を観察して理解し、その理解をもとに新しい表現を作り出す」という点です。教師データに頼らず、自ら文脈を学び、表現を生成する—この一連の能力こそが、AIを“知的な存在”へと近づけています。深層学習が「知覚」を、RNNが「記憶」を、Transformerが「理解」を与えたとするなら、生成AIは「表現する力」を得た段階といえるでしょう。

つまり、生成AIは単なるツールではなく、これまでのAI研究の積み重ねが融合して生まれた知的創造の集大成です。その背後にある基礎技術を理解することは、AIが「どのように考え、どのように創るのか」を知る第一歩となります。

(3)長所と短所

機械学習の強みは、未知のケースにも対応できる柔軟性です。特にビッグデータ時代の到来とともに、膨大なデータを解析して予測する技術は急速に発展しました。需要予測、音声認識、画像分類など、幅広い分野で活用されています。

しかし機械学習にも課題はあります。第一に「なぜその判断をしたのか説明しにくい」という問題です。複雑な数式やパラメータの組み合わせによって結果が出るため、利用者にはブラックボックスに見えがちです。第二に、大量の学習データが必要であることです。データが偏っていれば誤った予測をしてしまいます。

3. 生成AIの登場とその革新性

ここでいよいよ生成AIの登場です。従来のAI、つまりルールベースや機械学習が「与えられたデータをもとに分類・予測する」ことに特化していたのに対し、生成AIは「新しいデータを生み出す」ことを可能にしました。

たとえば従来のAIは「この画像は猫か犬か」を分類することはできても、「新しい猫の画像を作り出す」ことはできませんでした。しかし生成AIは「存在しない猫の写真」をリアルに生成できるのです。これが最大の違いです。

従来のAIは、正解がすでに存在するタスクに強みを発揮しました。翻訳、分類、需要予測などが典型です。しかし生成AIは、正解がない状況でも「もっともらしい答え」を自ら作り出せます。言い換えれば、「未来の可能性を切り開くAI」としての役割を持ち始めたのです。

生成AIの登場により、AIの応用範囲は一気に広がりました。文章生成、画像生成、音声合成、動画生成といった分野はもちろん、教育、研究支援、ビジネスアイデアの創出、芸術活動まで多様な領域に浸透しています。

(1) 事例

・医療

生成AIは、医師の診断を助ける形で活用されています。たとえば電子カルテに入力された検査結果や症状の記録をもとに、自動で診断レポートを下書きすることができます。また、患者さん向けには、難しい医学用語をかみ砕いた説明文を生成し、「この薬はどう効くのか」「どんな生活上の注意が必要か」をわかりやすい言葉で伝えることができます。

・教育

教育分野では、生成AIが先生の手助けになります。生徒がつまづいた箇所を入力すると、その生徒のレベルに合わせた解説を作成できますし、同じ問題でも違う言い回しや新しい例題を自動的に作り出すことができます。たとえば英語学習では、生徒の興味に合わせて「スポーツに関する例文」や「旅行に関する会話文」を即座に生成することが可能です。

・ビジネス

ビジネスの現場でも生成AIは幅広く活用されています。販売促進では、新しい広告文やキャッチコピーを瞬時に複数案作成できますし、商品紹介文もターゲット層に合わせて言葉を変えて生成できます。また、画像生成の仕組みを利用すれば、広告用のビジュアルデザインや試作品のイメージ図を短時間で作り出すことも可能です。

(2) 生成AIの課題とリスク

もちろん生成AIにも課題があります。もっともらしく見えるのに事実と違う内容を作ってしまうことがあり、生成された文章や画像は自動で正確さや最新性が保証されるわけではありません。医療や法律のように誤りの影響が大きい領域では、人による確認と責任ある運用が欠かせません。

著作権やプライバシーの問題も重要です。学習に使われるデータは公開情報や契約に基づくデータなどさまざまですが、生成物が既存の作品の表現を実質的に再現してしまった場合の扱いは国や地域で見解が分かれる点が残っています。単に作風が似ているだけでは直ちに著作権等の侵害とならない場合もありますが、特定の人物・作品を具体的に真似る指示や、個人情報を含む内容の取り扱いには注意が必要です。

倫理面では、偏ったデータに基づけば偏った結果が出る可能性があり、なりすましの音声や画像、偽情報の大量生成など社会的な悪用リスクも無視できません。こうした背景から、生成AIは従来以上に「誰がどこまで責任を負うのか」が問われやすい側面があります。実務では、開発者・提供者・利用組織・利用者がそれぞれの役割で責任を分担し、出力の検証、出典や注意書きの提示、ログ管理やアクセス制御、ラベリング（生成物である旨の明示）などのガバナンスを整えることが求められます。各国・地域で法規制やガイドラインの整備も進行中であり、社会的な議論は現在進行形です。

4. まとめ

ルールベースAIは「人間が作ったルールに従うAI」、機械学習AIは「データから学んで予測するAI」、そして生成AIは「新しいものを生み出すAI」です。この進化の流れを理解すると、なぜ現在生成AIが社会的に大きな影響を与えているのか、その理由がはっきり見えてきます。

生成AIは単なる技術革新ではなく、私たちの仕事の仕方や創造活動そのものを変える存在になりつつあります。今後もAIの進化を正しく理解し、リスクと可能性の両面を意識しながら活用していくことが求められるでしょう。

代表的な生成AIサービスの特徴

ここでは、現時点で広く利用されている代表的な生成AIサービスとして「ChatGPT」「Claude」「Gemini」「Stable Diffusion」を取り上げます。それぞれの特徴や強み、得意な利用分野を整理することで、生成AIの多様な姿を理解できるようにします。

1. ChatGPT (OpenAI)

ChatGPTは、OpenAIが開発した大規模言語モデルを基盤とする対話型AIです。自然な会話を通じて、文章の作成や要約、翻訳、コード補助など幅広い作業に対応します。

長所

- 自然な文章生成に優れ、会話形式でのやり取りが可能。
- 用途が広く、業務から日常利用までカバーできる汎用性。
- 出力速度が速く、下書きやアイデア出しに適している。

得意分野

- ビジネスメールや報告書のドラフト作成
- 学習支援や調べ物の整理
- 企画や資料のアイデア生成

ChatGPTは一般利用者への普及度が特に高く、外部サービスとの連携機能も拡張されつつあります。単独の会話AIにとどまらず、作業基盤としての役割も拡大しています。

2. Claude (Anthropic)

Claudeは、Anthropic社が開発した対話型生成AIです。安全性や配慮ある応答を重視する設計が特徴で、落ち着いた文章生成や長文の取り扱いに強みがあります。

長所

- 出力内容が穏当で、リスクの高い発言を避けやすい。
- 長文の要約や整理に適している。
- 企業や教育現場など、安全性を重視する場面で使いやすい。

得意分野

- 大規模な契約文書や規約の要約
- 配慮の必要な顧客対応文書の作成補助
- 教材や研修資料のドラフト作成

Claudeは独自の設計思想をもとに、AIが従うべき行動指針を組み込んでいます。そのため、出力の一貫性や安定感が評価され、ビジネスや公共分野での利用が進んでいます。

3. Gemini (Google)

Geminiは、Googleが提供する生成AIで、テキストだけでなく画像やコードも扱えるマルチモーダル型を志向しています。Googleの各種サービスとの連携が利用のしやすさにつながっています。

長所

- 複数の形式（文章・画像・コードなど）に対応可能。
- Google Workspaceなど既存サービスと統合しやすい。
- 調査や資料作成に直結する環境を提供できる。

得意分野

- 調べ物やリサーチの支援
- 学習用教材や教育コンテンツの生成
- DocsやSlidesなどでの文書・資料作成の効率化

Geminiは、検索や業務ツールとの統合を強みにしています。既にGoogleのサービスを利用している人にとっては導入が容易で、既存のワークフローに自然に組み込みやすい点が評価されています。

4. Stable Diffusion

Stable Diffusionは、画像生成に特化した生成AIで、オープンソースとして公開されているのが最大の特徴です。ユーザが自由に利用や改変でき、幅広い用途に展開されています。

長所

- オープンソースで、研究や商用利用まで自由度が高い。
- 写真風からイラスト風まで多様な画像表現が可能。
- モデルのカスタマイズや追加学習が容易。

得意分野

- 広告やマーケティング用のビジュアル制作
- ゲームや映像制作における素材の生成
- デザイン試作やアイデア段階のイメージ化

Stable Diffusionは世界的な開発コミュニティが活発で、追加モデルや拡張機能が日々生まれています。そのため、単体のサービスにとどまらず、生成AIにおいて、製品・サービスなどが連携し、大きなネットワークを形成して新たな価値や収益を生み出す仕組みであるエコシステムの中心的な存在となっています。

5. 総括

これら4つのサービスは、いずれも生成AIの代表格ですが、それぞれ異なる強みを持っています。ChatGPTは汎用性の高さ、Claudeは安全性と長文処理、GeminiはマルチモーダルとGoogle連携、Stable Diffusionはオープンソースとカスタマイズ性に特徴があります。共通するのは、人間がゼロから作業しなくても、AIが下書きや試作品を迅速に用意し、効率を高めてくれる点です。

将来的には、文章・画像・音声・動画など複数の形式を一つのサービスで扱える方向へと進化が見込まれています。とはいえ、これらのサービスは現在も発展の途上にあり、特徴や強みは技術や社会的な要請に応じて少しずつ変わっていく可能性があります。したがって、利用者には「現時点での特徴」を理解しつつ、変化に合わせて柔軟に使い方を調整する姿勢が求められます。