AIはなぜウソをつく?

生成AIのしくみがわかる本

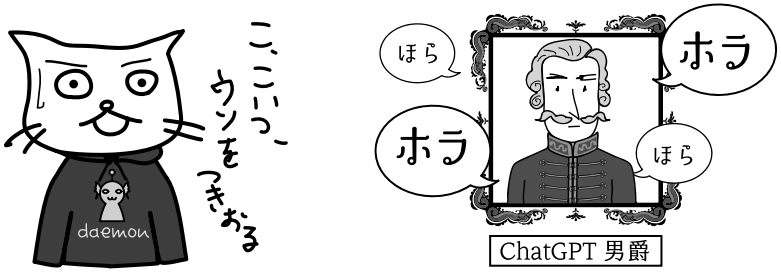
でも、あれだけ滑らかに自信満々に答えられてしまうと、信じてしまう気持ちもわかります。僕も、プレゼンや、謝罪の場では、いつも堂々としています。そうすると、皆さん信じてくれますから！

コラム

アカシックレコード／ラプラスの悪魔

アカシックレコード：神智学（しんちがく）などで語られる、宇宙で起きた出来事や人間の思考など、あらゆる情報が記録されているとされる概念。「宇宙の記録庫」のようなものです。

ラプラスの悪魔：宇宙のあらゆる物体の状態と自然法則を完全に知ることで、過去から未来まですべてを計算できると仮定された架空の知性。ラプラスの決定論を説明する際に使われる有名な思考実験です。



ほら吹き男爵と化すAI

生成AIには、**OpenAI（オープンエーアイ）**の**ChatGPT（チャットジーピーティー）**、Google（グーグル）のGemini（ジェミニ）、Anthropic（アンソロピック）のClaude（クロード）など、さまざまなサービスがあります。画像生成AIにも、**Stable Diffusion（ステーブル・ディフュージョン）**やMidjourney（ミッドジャーニー）、Adobe Firefly（アドビ・ファイアフライ）など、いくつもの種類があります。それぞれ得意なことや細かな仕組みには違いがありますが、まったく別物というわけではありません。

文章を扱う生成AIであればChatGPT、画像生成AIならStable Diffusionを知ることで、画像生成がどのような考え方で行われているのかを理解しやすくなります。

本書では、これらの仕組みについて説明しながら、生成AIとはどのようなものなのかを解説していきます。

1-1 機械学習とは

では、文章生成AIや画像生成AIの中身を見る前に、その土台になっている「機械学習」について押さえておきましょう。

みなさんが聞いたことあるようなChatGPTやStable Diffusionは、生成AIと呼ばれる**「なにかを作成するのが得意なAI」**です。

生成AI登場以前からあった、従来の機械学習と根本的な仕組みは同じですが、学習のさせ方やサービスのありかたが違っています。生成AIのしくみを理解するには、元となった機械学習から説明するとわかりやすいので、まずはそこから解説していきます。

「そんな初歩的なことは知ってるよ」という方は、この章を読み飛ばして、次の「ChatGPTの仕組み」にジャンプしてしまって大丈夫です。

機械学習は、AIを実現する手法の一つ

最初に言葉の定義をはっきりとさせておきましょう。僕は、定義付けが大好きなので。

AIの話になると、「人工知能」「機械学習」「深層学習」など様々な用語が出てきます。これらを整理することにしましょう。

まず、**人工知能（AI＝Artificial Intelligence）**というものがあります。AIの定義は難しく、専門家の先生達も色々言っているのですが、とりあえず、「知的な振る舞いをする人工物（システム）」と思って良いでしょう。「知的とは何か」は置いておくとして、「人間の知能っぽい行動を取る何か」ってことです。

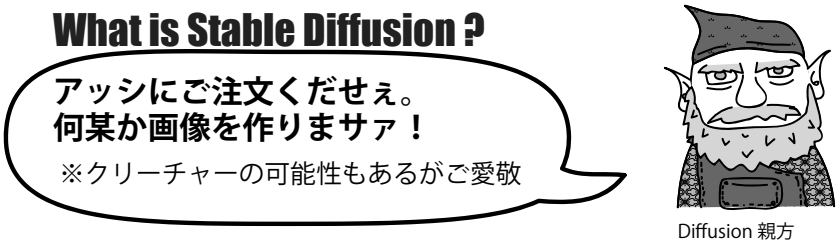
そして、**機械学習（Machine Learning）**とは、人工知能（AI）を実現するための手法（アルゴリズム）の一種です。事前にたくさんのデータを読み込ませて、その性質を覚えさせておき、新しいデータを与えたときに、その答えを得る仕組みです。特に、たくさんのデータを、自分で学習していくものを言います。

画像生成AIの仕組み

文章生成AIの仕組みがわかったところで、次は画像生成AIです。この章では、代表例として**Stable Diffusion（ステーブルディフュージョン）**を中心に、画像生成AIがどのように学習し、画像を作っているのかを見ていきます。

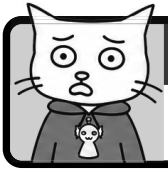
Stable Diffusionは、拡散モデルを使った画像生成AIの代表例の一つです。テキストで「こんな画像が欲しい」と指示すると、その条件に合う画像を生成できます。

まあ、何か注文を唱えると、ドロロン！と画像を出してくれる工房の親方みたいなものだと思えば、最初のイメージとしてはそんなに間違っていません。



Stable Diffusionは工房の親方

文章生成AIも画像生成AIも、学習によって調整されたモデルを使って出力するという点では共通しています。ただし、扱うものが文章と画像では違うので、中で行っている計算や学習方法もかなり違います。



6-4 AIは「ツール」を使う

さっきは、AIサービスがWeb検索を使って外部から情報を取り込み、それを材料に回答を作る仕組みについて見てきました。実は、このWeb検索も、AIから見ると利用できる「道具」の一つです。

AIが利用できる道具は、Web検索だけではありません。正確な数値計算が必要なら計算機、大量のデータを処理するならPython（パイソン）などで書かれたプログラムを実行できる環境、会社の情報を調べるならデータベース、といった具合に、目的に応じてさまざまな機能を利用できます。画像を作る場合には、画像生成を担当する別のモデルや機能を利用することもあります。

このように、AIが必要に応じて呼び出して利用できる機能を「**ツール（tool）**」と呼びます。

人間が仕事をするときに道具を使うのと少し似ています。人間も、簡単な計算なら暗算しますが、複雑になれば電卓や表計算ソフトを使います。知らないことがあれば検索しますし、場所を調べるなら地図を使います。AIも、必要に応じて得意なツールへ仕事を任せることで、モデルだけでは難しい仕事まで扱えるようになります。

たとえば利用者から、「12345×6789を計算して」と頼まれたとします。このときAIサービスは、この依頼には計算が必要だと判断し、計算機へ「12345×6789」という式を渡すことができます。計算機から「83810205」という結果が返ってくれば、それを受け取ったモデルが、「計算結果は83,810,205です」と文章にして利用者へ返します。

ここでは、「何をする必要があるのかを判断する部分」と、「実際に計算する部分」が分担されています。モデル自身ですべてを計算するのではなく、計算を得意とする道具に仕事を任せ、その結果を受け取っているわけです。

同じことは、もっと複雑な仕事にも使えます。たとえば、「この売上データを月ごとに集計して、グラフを作って」と頼んだとします。モデルは依頼やデータの内容から必要な処理を考え、プログラムを実行できる環境へ処理を頼みます。そして、返ってきた集計結果やグラフを受け取り、その内容を文章で説明する、といった流れです。

ただし、ツールを使えるからといって、仕事全体が必ず正しくなるわけではありません。ここでも、「道具が正しく動くこと」と「道具を正しく使うこと」は分けて考える必要があります。

たとえば計算機そのものは正確でも、AIが間違った式を渡せば、目的とは違う答えが返ってきます。「税込価格を計算してください」という依頼に対して、本来10%として計算すべき商品を8%として式に入れてしまえば、計算機はその間違った式を正確に計算します。

プログラムを使う場合も同じです。読み込ませるデータを間違えたり、集計する列を取り違えたり、条件を勘違いしたりすれば、プログラム自体が正常に動いていても、正しい結果にはなりません。

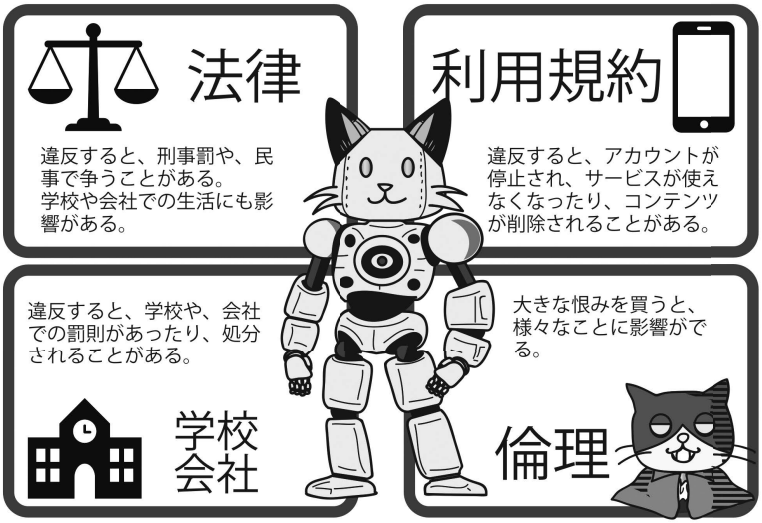
つまり、ツールを利用する場合には、「どのツールを使うか」「ツールへ何を渡すか」「返ってきた結果をどう解釈するか」という部分にも、AIによる判断が入ります。前の項で説明したWeb検索でも、「何を検索するか」「どの資料を使うか」「その内容をどう回答へ反映するか」に間違いが入り得ました。それと同じことが、計算機やプログラム、データベースなどでも起こるのです。

AIサービスが外部ツールを使う場合の流れ

ここまでの例では、AIは必要なツールを使い、その結果を受け取って回答を作れば仕事は終わりました。しかし、現実の仕事には、一度ツールを使うだけでは終わらないものもあります。途中の結果を見て、次に何をするべきかを考え、さらに検索したり、別のツールを使ったりしなければならない仕事です。

この「ツールを使う」という考え方を、さらに一歩進めたところにあるのが「**AIエージェント**」です。

規約、学校や職場のルール、そして相手への配慮や社会的な常識です。法律はもちろん守らなければなりませんが、それ以外も「法律ではないから無視してよい」という話ではありません。



AIを使うときに気を付けること

「法律じゃないなら、利用規約や学校・職場のルール、倫理まで守らなくてもいいのでは？」と思うかもしれません。でも、それぞれ守る理由があります。

法律、利用規約、学校や職場のルール、倫理は、誰が決めたのか、何のためにあるのか、破ったときに何が起こるのかが違います。だから、「法律に違反していないから大丈夫」と一つの基準だけで判断しないことが大切です。

利用規約は、サービスを提供する会社と利用者との約束です。違反したからといって、それだけで直ちに犯罪になるわけではありませんが、投稿の削除、機能制限、アカウント停止などの対象になることがあります。

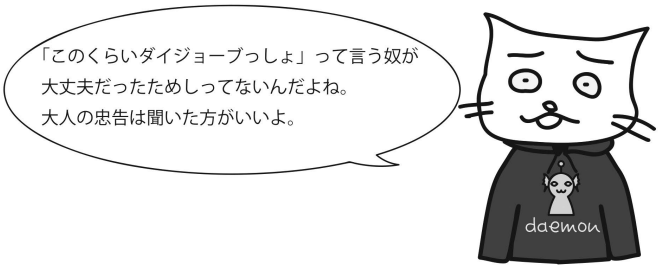
学校や会社にも、それぞれ**ルール**があります。課題でAI利用を禁止されているのに使えば学校のルール違反ですし、社外AIへの機密情報入力を禁止されているのに入力すれば会社のルールに反します。法律とは別の基準でも、処分や大きなトラブルにつながることがあります。

さらに線引きが難しいのが、相手への配慮や社会的な常識、**倫理**です。条文のような明確な境界がなくても、他人の写真を勝手に加工して笑いものにしたり、本人が言っていないことをAIで言わせたり、誰かをだますために偽物を作ったりするのが、よい使い方とはいえないでしょう。

法律や利用規約に、世の中の迷惑行為がすべて先回りして書かれているわけではありません。「禁止と書いていないからやっていい」ではなく、その行為で誰かが困らないか、だまされないか、傷つかないかも考える必要があります。

生成AIでは、この判断がとくに重要です。以前なら技術や時間が必要だったような偽物でも、短時間に大量に作り、広げられることがあります。AIは、作る力だけでなく、量とスピードまで増幅します。

本物と区別しにくい偽画像や偽音声が一つ広がるだけでも、信じる人が出れば影響は小さくありません。数十、数百と増えれば、訂正や削除だけでは追いつかないこともあります。



気軽にできる反面、リスクもある

しかも、いったん広がったものは簡単には元に戻せません。投稿を削除しても転載や保存までは消せませんし、法律上の問題が解決しても、傷ついた相手との関係や失った信用まで元どおりになるとは限りません。

軽い冗談や「ちょっと困らせるだけ」のつもりでも、AIで作ったものが広がれば、相手を社会的・精神的に追い詰めることがあります。「こんな大ごとになるとは思わなかった」で済まないこともあります。

■著者略歴

小笠原　種高（おがさわら　しげたか）

愛称はニャゴロウ陛下。技術著述家。
システム開発のかたわら、雑誌や書籍などで、データベースやサーバ、マネジメントについて執筆。図を多く用いた易しい解説に定評がある。綿入れ半纏愛好家。

［主な著書］

「図解即戦力　Amazon Web Servicesのしくみと技術が これ1冊でしっかりわかる教科書」　技術評論社
「仕組みと使い方がわかる Docker&Kubernetesのきほんのきほん」　マイナビ出版
「256(ニャゴロー)将軍と学ぶWebサーバ」　工学社　他

［イラスト］

小笠原種高

［Website］

モウフカブール　http://www.mofukabur.com

［X］

@shigetaka256

本書の内容に関するご質問は、
①返信用の切手を同封した手紙
②往復はがき
③E-mail　editors@kohgakusha.co.jp
のいずれかで、工学社編集部あてにお願いします。
なお、電話によるお問い合わせはご遠慮ください。

サポートページは下記にあります。

［工学社サイト］
https://www.kohgakusha.co.jp/

I/O BOOKS

AIはなぜウソをつく?生成AIのしくみがわかる本

2026年9月30日　初版発行　ⓒ2026

著　者　小笠原　種高
発行人　星　陽平
発行所　株式会社工学社
〒160-0011　東京都新宿区若葉1-6-2 あかつきビル201
電話　(03)5269-2041(代)［営業］
(03)5269-6041(代)［編集］
振替口座　00150-6-22510

※定価はカバーに表示してあります。

印刷：(株)エーヴィスシステムズ

ISBN978-4-7775-2350-4