はじめに

ビッグデータとかいわれるけど，どんなものなんだろう。
「データサイエンティスト」を目指せっていわれたけど，どんな仕事なんだろう。

AI や人工知能ってなんだろう。
AI に仕事を奪われるって聞いたけど，これからの時代になにが起きるんだろう。
そんな疑問をもっている人はいないでしょうか。

みんな，「よく知らないこと」「分からないこと」に対しては不安になってしまうものです。
不安をなくすには，少しでもいいからデータサイエンスや AI についての知識をつけること。未知の存在でなくなれば，怖くありません。
これからの時代には，AI やデータ処理は特別なものではなくなります。その当たり前を身につけて，備えるようにしましょう。

この書籍は，主に IT をこれから学ぼうとしている初学者の方，データサイエンスや AI に興味をもちはじめた方，本格的に学ぶ前の基礎の段階を作ることを目的にしています。データそのものからデータサイエンス，これから様々なデータを生み出す元になると思われている IoT，そして AI。

ちょうどディズニーランドのアトラクション「イッツ・ア・スモールワールド」みたいにデータの世界を巡ってみましょう。

皆様のお役に，少しでも立ちますように。

2020 年 1 月
吉原幸伸

Contents

巻末付録

ワークブック

4-1 AIでできること・できないこと

1 知能とは

初めに，知能について考えてみましょう。

知能▶ **知能**は，次の三つのタイプに分けることができます。

タイプ①　新しい場面や困難に直面したとき，本能に捉われず，適切かつ有効な方法で順応したり解決したりする能力

タイプ②　言語や記号などを用いて推理や洞察を含めた概念のレベルで思考を進める能力

タイプ③　知識や技能を経験によって獲得する能力

心理学者のアルフレッド・ビネーと医師のテオドール・シモンは，知能をタイプ②とタイプ③に近い，「外界を全体として再構成するための認識能力」と定義しました。そして，初めて知能を客観的に測定するための検査を考案しました。その検査のことを**ビネー式知能検査**といいます。精神発達の早さには，同じ年齢の子供でもそれぞれ個人差があると考え，特定年齢の子供のうち50%～75%が正しく答えられるテスト項目をあらかじめ作りました。それに回答できれば，その発達水準に発達しているとし，そこから「精神年齢」を算出しました。

ビネー式知能検査▶

ビネーの後，心理学者のウィリアム・スターン（シュテルン）が，知能テストの結果を表す指標として「**知能指数**（**IQ**）」を考案しました。これをルイス・マディソン・ターマンが知能検査に採用したことで，使用されるようになりました。知能指数は，実年齢に対する精神年齢の程度（発達の割合）を示します。これらは主に児童向けに活用されました。

知能指数▶

IQ▶

精神年齢▶ **精神年齢**（MA；Mental Age）ビネーの知能テストによって評定された精神発達水準（単位は月数）

生活年齢▶ **生活年齢**（CA；Chronilogical Age）実年齢（単位は月数）

知能指数＝「精神年齢」÷「生活年齢」×100

ただ，この手法はタイプ①の適応能力的知能などの評価を全く行うことができません。

ウェクスラー式知能検査▶

デイヴィッド・ウェクスラーが，検査対象を成人にまで拡張した**ウェクスラー式知能検査**（**WAIS**）などのテストを作りました。ウェクスラー式知能検査は，知能構造の診断をするため「診断的検査」と呼ばれ，評価指標として，**偏差知能指数**（**DIQ**）が用いられます。偏差知能指数とは，一般的な知能指数（平均）からどの程度異なるかを示した値です。集団の平均を 100 とします。

WAIS▶

偏差知能指数▶

DIQ▶

$$DIQ = 100 + 15 \times \{「(各個人の点数 - 同年齢集団の平均点)」 \div 「同年齢集団の標準偏差」\}$$

この評価方法が改良を重ねながら，その後の知能検査の主流となっていきました。

その後，知能に対する考え方が複合的になりました。発達心理学者のレイモンド・キャッテルは，タイプ①を**流動性知能**（フロー），タイプ③を**結晶性知能**（ストック）としました。初めて見た問題を解決したり，ひらめきを利用したり，新しいものを創造したりするのは流動性知能。言葉の分析，単語力，語学能力などは結晶性知能によって行われます。知能が複合的に考えられるようになったことで，結晶性知能が高くても，流動性知能が衰えているというような状態も考えられています。

流動性知能▶

結晶性知能▶

長年にわたる知能検査の結果，複数の検査項目の中で，相関が高い項目があることが判明しました。そして，この相関が高い項目をグルーピングしていくことで，何種類くらいの能力が人間にあるのかという「知能の分類」を作り出す試みも進められました。例えば，ハワード・ガードナーの提唱した「**多重知能**（MI）理論」があります。この理論は「人は皆それぞれ一組の多重知能をもっており，少なくとも八つの知的活動の特定の分野で，才能を大いに伸ばすことができる」として，その分野を定義しました。

多重知能▶

多重知能理論の八つの分野を次表にまとめます。

知能の分野	説明
言語的知能	話し言葉・書き言葉への感受性，言語学習・運用能力など
論理数学的知能	問題を論理的に分析したり，数学的な操作をしたり，問題を科学的に究明したりする能力
音楽的知能	リズムや音程・和音・音色の識別，音楽演奏や作曲・鑑賞の能力
身体運動的知能	体全体や身体部位を問題解決や創造のために使う能力
空間的知能	空間のパターンを認識して操作する能力
対人的知能	他人の意図や動機・欲求を理解して，他人とうまくやっていく能力
内省的知能	自分自身を理解して，自己の作業モデルを用いて自分の生活を統制する能力
博物的知能	自然や人工物の種類を識別する能力

このほかにも，ルイス・レオン・サーストンの多因子説など，いくつもの分類が考えられました。

人間の知能は，様々な分野の能力を発揮するものと考えられています。

2 人工知能とは

人工知能▶
AI▶

人工知能（Artificial Intelligence：以下，**AI** という）とは，前述の人間の知能の分野の一部，またはすべてを代替できる仕組みを指します。例えば，最近のカメラの多くには，ガードナーの分類でいう「博物的知能」が搭載されており，撮影された対象が人物なのかどうかを見分けられるようになっています。

AI につながる概念は，1950 年前後から研究されるようになりました。なかでも有名なのがアラン・チューリングの研究です。チューリングは，第二次世界大戦当時のエニグマ暗号の解読でも知られる数学者でした。彼は論文「計算する機械と知性（Computing Machinery and Intelligence）」（1950 年）で初めて人工知能の主要な議論に的を絞った内容を発表しています。その冒頭で「機械は（人間的な）思考をするか？」という「問い」を投げかけました。その上で「問いの視点を変えてみよう」と提案しています。チューリングはなぜ，このような提案をしたのでしょうか。それは，チューリングは「人間には扱えて」「機械には扱えない」概念がある＝「人間を完全に再現する事は不可能」と考えていたからです。

この問題意識から，チューリングは「人間か人工知能か」を見分けるためのテストとして，**チューリングテスト**を考案しました。

チューリングテスト▶

人間の審査員が，1 人の人間と一つのプログラムに対して会話を行います。このとき，人間もプログラムも人間らしく見えるように対応します。それぞれの参加者は隔離されているので，実際に返答しているのが人間なのかプログラムなのかは会話内容以外では判断できません。会話を終えて，審査員が人間とプログラムとの区別ができなかった場合に，このプログラムはテストを通過したことになります。

このテストでは，ガードナーの分類でいうところの「言語的知能」や「対人的知能」を試されることになります。

言語によるコミュニケーションは，次のような階層をもつと考えられます。

第４層	感情・思考
第３層	メッセージ（具体的な言葉）
第２層	言語（英語や日本語など）
第１層	音声・文章・手話など物理的コミュニケーション手段

音声などの物理的コミュニケーション手段は，言語に基づいて行われます。場合によっては，ボディ・ランゲージということもあるかもしれません。私たちはそこからメッセージを読み出します。

ここまでは人間の知能であっても，AIであっても変わらず行うことができます。手軽に顔認識が行えるサービスとして，Amazon Rekognition があります。Amazon Kinesis Video Stream からストリームとして送られる動画をリアルタイムで分析したり，スマートフォンなどから Amazon S3 にアップロードされた画像を解析することによって，そのなかに含まれる顔を分析し，幸福度，年齢層，目が開いているか，メガネの有無などの属性を抽出することができます。さらに，動画では，これらが時間とともにどのように変化するかも測定することができます。

図表 4-1 Amazon Rekognition「顔分析」

出典：(https://aws.amazon.com/jp/rekognition/)

しかし，メッセージは正確に読み取れても，その奥底にある感情や思考を読み取ることが正解ばかりとは限りません。例えば，人間関係において，本心とは裏腹な言葉を投げかけてしまった経験はないでしょうか。ドラマでも，本当は相手のことを好きなのに，全く逆のネガティブな感情を込めたメッセージを投げかけてしまうというシーンが見受けられます。

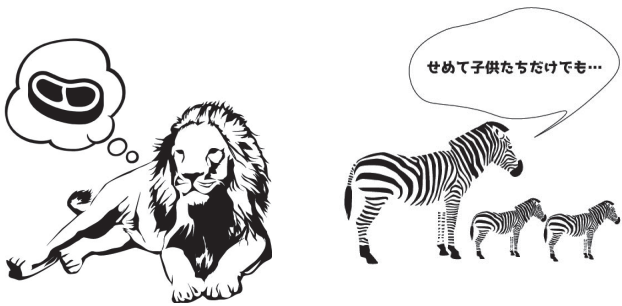
この場合，受け手がネガティブなメッセージをそのまま解釈してし

まったら，コミュニケーションの相手が全く逆の思考をしていると勘違いすることになるので，コミュニケーションは失敗といえるでしょう。

人間は，自分の思考を言葉によるメッセージに加工する際に，その加工を適切に行えるとは限りません。真逆ではなくても，表現を間違えてしまった結果，無用な誤解を生むこともよくあります。そのため，コミュニケーションは難しいといえます。

チューリングテストはこれを逆手に取ります。コミュニケーションを取ってみて，相手に感情や思考が存在していると判断したら，それは人間であると考えます。ただ，これは簡単なことではありません。それは，価値観が異なれば，感情や思考は伝わらない可能性があるからです。

例えば，肉食獣は，腹が減れば狩りをして，草食獣などを襲って食べます。食べられる側の草食獣は，「自分が食べられても，せめて子供だけでも逃がしたい」などと思い，自分がおとりになる仕草は伝わっても，「子供だけは逃がしてくれ」というメッセージの裏側の感情は伝わりません。肉食獣は草食獣を食べ物としか思っていないので，感情や思考を読み取らないからです。



図表 4-2 肉食獣と草食獣

AIが人間と同じように思考し，感情をもつならば，相手のいいたいことを理解し，会話が成立するはずです。チューリングテストでは人間が判断する理由は，人間と同じ思考が備わっているかは，人間にしか判断できないと思われているからです。

では，チューリングテストに合格したAIはいるのでしょうか。実は，チューリングの没後60年の命日に開催された，「Turing Test 2014」において，合格したAIが誕生しました。それは，ウクライナ在住の13歳の少年，ユージーン・グーツマンという設定のAIです。このイベントで，審査員のうち33%がユージーンを人間と認めました。

そのため，ユージーンは，チューリング・テストをクリアした初のAIとなりました。

テストは5分間のテキストチャットという極めてシンプルなもので，質問内容は審査員が自由にすることができます。ユージーンには，「13歳のウクライナ人少年で，父は婦人科医。ペットとしてビルと名づけたモルモットを飼っている」といった設定があり，ある特定の人物を演じるように設計されていました。また，審査員は，ユージーンに対して英語で話しかけたので，母国語でない英語をたどたどしく話すユージーンを，人間らしいと感じてしまったことも結果に影響を与えたことでしょう。なお，このテスト結果については，「設定自体が審査員に認知バイアスを与えるものだ」といった批判もあります。

チャットボット▶ ユージーンのように，テキストチャットに特化したプログラムを俗に**チャットボット**と呼びます。ほとんどのチャットボットは，次のアルゴリズムの型に分かれます。

辞書型▶ **(1) 辞書（ハッシュ）型**

辞書に登録されたテンプレートに応じて会話を行います。パターンに合っていれば会話を外すことは少なく，会話のフレーズは辞書の大きさに依存します。

「○○」という言葉が入っていたら，「××」という返事を返す，というルールに則った単純なものですが，分野や会話の範囲が限られていれば十分なコミュニケーションがとれる場合があります。このタイプのチャットボットは以前から多く見られました。

キーワード	応答
戸籍謄本	○○○・・・
戸籍抄本	△△△・・・
住民票	XXXX・・・

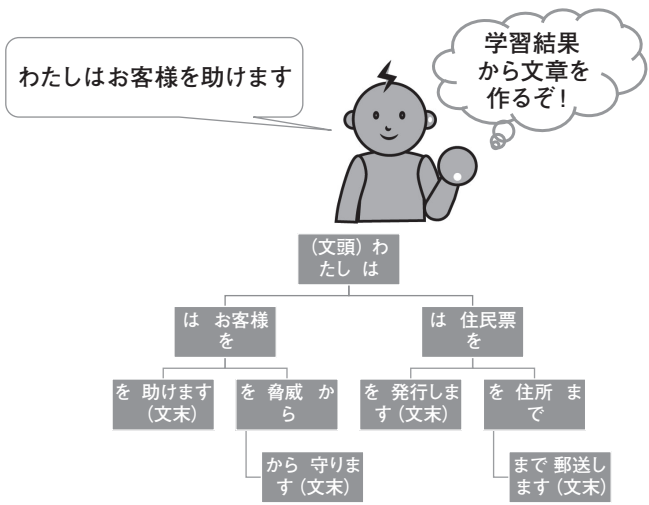
図表 4-3 辞書型の概念

マルコフ連鎖型▶ **(2) マルコフ連鎖型**

ユーザとの会話を品詞分解し，オリジナルの文章を組み上げていく方法です。組み上げていく方法に「マルコフ連鎖」が使われます。マルコ

フ連鎖は，Google が内部で Web ページの評価に使用している Page Rank の基になっているアルゴリズムです。Page Rank は，ある人が特定のページを開く→ページ内のリンクのうちいずれかをクリックし遷移する→遷移したページにあるリンクのうちいずれかのリンクをクリックして遷移する，という一連の行動を繰り返したとき，あるページが開かれている確率を表したものになります。結果として，様々なところからリンクの張られている人気のあるサイトほど開かれている確率が高くなるので，Page Rank が高くなります。

同様に，ある単語の次に来る単語 A，B，C の，それぞれ統計上の割合が 50%，30%，20%とした場合，その割合に基づいて次に何が来るのかを判断し，それをどんどん続けて一つの文を作る方法です。元となる文章は過去の会話のログや WEB サイトなどから割り出します。



図表 4-4 マルコフ連鎖型の概念

自然文分析型▶

(3) 自然文分析型

ユーザが入力した自由な文章による質問（自然文）を分析し，単語に分割して，その単語が用意された想定質問とどの程度一致しているかを計算し，一致する確率が高い想定質問の回答文を出力します。質問の単語を想定質問が含んでいたら「加点」をしていくと考えます。

自然文分析型には，次のような特徴があるため，辞書型に比べて適切な応答ができる可能性が高まります。

・似た意味の単語を理解できるので，異なる聞き方でもヒットする
・単語の完全一致ではなく，一致する確率を計算しているので，0 件ヒ

ットがない

シソーラス▶　なお，**シソーラス**（類語辞典・辞書）を用意することで，さらに精度が高まります。

図表 4-5 自然文分析型の概念

マルコフ連鎖型チャットボットも，自然文分析型チャットボットも，あらかじめ決められたルールに沿って動作するだけでなく，会話のログやWebサイトの分析から学習していく仕組みなのでAIといって差し支えないでしょう。

ユージーンを含むチャットボットは，前述のとおり「言語的知能」や「対人的知能」だけに特化したものです。そのため，知能の分野に特化したAIを**特化型AI**と呼びます。現在実用化されているAIのほとんどがこの特化型AIです。また，何かの分野に特化しているということは，人間の知能を模倣しているレベルが低いという意味で，「弱いAI」とも呼びます。

特化型 AI▶

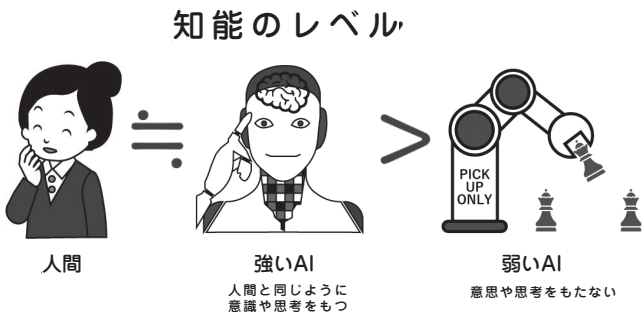
一方で，特定の作業やタスクに限定せず，人間と同様，あるいは人間以上の汎化能力をもち合わせているAIを**汎用型AI**と呼びます。汎用型AIは，プログラミングされた特定の機能以外にも，自身の能力を応用して対応できるとされます。そのため，流動性知能（フロー）に近い能力をもつと考えられます。現時点では，IBM Watsonのように多種の目的に対応することができる「Augmented Intelligence，拡張知能」，自然言語を理解･学習し人間の意思決定を支援する「コグニティブ･コンピューティング・システム（Cognitive Computing System）」が登場していますが，あくまでも意思決定支援を中心としたソリューションに特化しており，流動性知能をもつには至りません。

汎用型 AI▶

また，汎用型AIの中でも，人間のような自意識を兼ね備えているものを**強いAI**と呼びます。

強い AI▶

ただ，現時点では，汎用型 AI も，強い AI も存在していません。



図表 4-6 強い AI と弱い AI

3 AI にできること

これまでの AI でできることは，次のとおりです。ガードナーの分類に従って記述します。

(1) 言語的知能

言語的知能▶

言語的知能の分野では，自然言語という，人間が通常使う言葉の処理方法が研究されています。AI は文章を読んで構文を理解したり，内容が似た文章を選択したりすることができます。チャットボットもこの言語的知能の拡張知能と考えてもよいでしょう。

言語的知能▶
特化型 AI

チャットボット以外に，入力された言語を解析して，翻訳を行う Google 翻訳のようなサービスは**言語的知能特化型 AI** といえるでしょう。また，Amazon が提供する「Alexa」というサービスは，人気 IoT 機器「Amazon echo」の音声認識部分として組み込まれています。ユーザの音声指示を受けて商品発注や音楽再生など様々なリアクションをします。なお，Alexa は，他の企業が利用できるように，音声認識機能を一般公開しており，「スキル」という形で Amazon Echo に取り込むことができます。

さらに，中部経済新聞では，AI の執筆した新聞記事を掲載するというチャレンジがなされました。この新聞では，AI の執筆した新聞記事を掲載したことに関する座談会について，さらに AI がまとめを作成して掲載しました。

Mini Discussion 4

Mini Discussion 4-1

自分の持っているスマートフォンについて，宣言的知識を挙げてみよう。

Mini Discussion 4-2

パソコンの購入検討する行動を分析する場合の，特徴量の候補になる数値で評価できるものを挙げてみよう。

Mini Discussion 4-3

2045 年の日本の人口を予測する場合，最も適切なアルゴリズムは回帰か分類かを考えてみよう。

Mini Discussion 4-4

ある散布図上で点の分類を行いたい場合，適切と思われるアルゴリズムを選び，理由を考えてみよう。

Mini Discussion 4-5

画像認識が進化すると，どのような使い方ができるかを考えてみよう。

Chapter 4 章末問題

■問1　次の記述の中で正しいものには○，誤っているものには×をつけよ。

ア　チューリングテストは，「人間か人工知能か」を見分けるためのテストで，審査員との会話で判断される。

イ　辞書型チャットボットは，反応速度を上げるためにできるだけ小さい辞書を用意することが望ましい。

ウ　マルコフ連鎖型チャットボットは，学習済みのある単語とある単語の繋がりから可能性の高いものを使って文章を組み立てるボットである。

エ　自然文分析型は，入力された質問の意味をAIが理解し，AIの判断で応答を自動的に作り出すボットである。

オ　特定の目的によらず，プログラミングされた特定の機能以外にも自身の能力を応用して，人間同様に意思決定を行うことができるAIを，汎用的AIと呼ぶ。

■問2　次に挙げるAI・機械学習に関する記述a～dにそれぞれ最もよく対応する字句を解答群の中から選べ。ただし解答を重複して選んではならない。

a　機械学習で，モデルを数値で扱えるように数式化した問題を，どのように解くのかを決めた手順

b　モデルの調整用の数値のこと

c　機械学習でモデルを調整するために読み込ませるデータのこと

d　モデルが正しく動くかどうかを確認するためのデータのこと

解答群

ア　トレーニングデータ　イ　パラメタ　ウ　テストデータ

エ　アルゴリズム

■問3　次の記述の中で正しいものには○，誤っているものには×をつけよ。

ア　モデルを，現実のデータに対してどこまでフィッティングさせるかという調整は非常に難しいが重要である。

イ　教師あり学習は，人間が人間に対する教育と同じようにAIに対して教師として振る舞い，手動で様々なことを教え込むことで行われる。

ウ　教師なし学習は，与えられたデータの本質的な構造や法則をアルゴリズムが自動的

に抽出する機械学習アルゴリズムである。

エ　強化学習は徐々に強くなる負荷を継続して与えていくことで行動パターンを学習していく過程をモデル化して数学的に表現したものである。

オ　チューニングとは，高い予測精度を得るためにフィッティングの際のハイパーパラメタの値などを変えて，訓練データによる学習とテストデータによる評価を繰り返すことである。

■問 4　次の記述は，それぞれどんなアルゴリズムに関して説明をしているものか，解答群の中から選べ。

a　分類を行うために利用され，データを散布図上に配置したときに 2 つのグループのデータから最も離れた箇所＝最大マージンを見つけ出す。

b　複数の変数から確率を予測するためのアルゴリズムで，0%から 100%の間でデータはプロットされる。

c　いくつのグループに分類するかを決めて，それぞれのグループの座標の重心を基に分類作業を行う。

d　データの要約を作成するために使われ，データが最もばらついている方向と 2 番目にばらついている方向に情報損失量が少なくなる軸を作る。

e　データの点に最も沿うように，直線とデータの誤差の 2 乗の和を最小にすることで，最も確からしい関係式を求める方法。

解答群

ア　k 平均法　　イ　主成分分析法　　ウ　サポートベクターマシン

エ　ロジスティック回帰　　オ　最小二乗法

■問 5　次の記述の中で正しいものには○，誤っているものには×をつけよ。

ア　トイ･データセットは，学習用に用意されたデータセットで，これを利用することで，機械学習の問題解決方法を学びやすくなる。

イ　機械学習はデータを学習することで，既知のデータの識別や予測が確実にできるようにすることを目的としている。

ウ　交差検証では，データセット全体を使って学習し，テストもデータセット全体で行う。

エ　正解率は，正や負と予測したデータのうち，実際にそうであるものの割合を示す。

オ　適合率は，正と予測したデータのうち，実際に正であるものの割合を示す。

Chapter 4

章末問題解答・解説

問１　ア－○，イ－×，ウ－○，エ－×，オ－○

イ：辞書型チャットボットは辞書に登録されたテンプレートに応じて会話を行うため，入力される可能性のある単語がカバーされている十分な大きさの辞書をもつことが望まれます。

エ：自然文分析型は，入力された質問と，想定質問を比較して，一致度の高いものに用意された応答を返すボットです。本当の意味で質問を理解することができるAIは実現は難しいとされています。

問２　ａ－エ，ｂ－イ，ｃ－ア，ｄ－ウ

（アドバイス）機械学習では，１つ以上のアルゴリズムを用いてモデルを作成していきます。モデルの数式の結果を調整するための値がパラメタです。パラメタはトレーニングデータ（訓練データ）を基に調整されていきます。訓練されたモデルが適切に動作するかどうかは，テストデータで検証をします。

問３　ア－○，イ－×，ウ－○，エ－×，オ－○

イ：教師あり学習は，人間が選択した正解のラベルのついたトレーニングデータを基にラベルのないテストデータでも正解できるように学習させる手法です。

エ：強化学習は，正解データの代わりに行動がどれだけよかったのかを報酬として与え，その報酬が高くなるような行動をするように学習させる手法です。

問４　ａ－ウ，ｂ－エ，ｃ－ア，ｄ－イ，ｅ－オ

（アドバイス）機械学習やディープラーニングに用いられるアルゴリズムは，それぞれ特徴を持っています。同じデータセットを分析する場合でも，アルゴリズムによって結果が異なることがあります。それぞれのアルゴリズムの特徴を活かして，適切なアルゴリズムを選択することが重要です。

問５　ア－○，イ－×，ウ－×，エ－○，オ－○

イ：機械学習は既知のデータを学習することで，未知のデータの識別や予測が確実にできるようにすることを目的としている。

ウ：交差検証では，データセットを分割し，データをそれぞれ訓練データとテストデータに割り振っていく手法です。

索　引

目次

●ワークブックの使い方

このワークブックは本書学習後の理解度確認用として利用いただけます。
利用方法は次のとおりです。

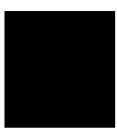
(1) 「AI・データサイエンスの基礎」の学習を進めます。
(2) 理解度を確認したい内容について,ワークブックの該当部分を解きます。
(3) 「AI・データサイエンスの基礎」を活用し,自分の記述内容が合っているかを確認します。もし間違いがあれば訂正します。
(4) 訂正した内容については,「AI・データサイエンスの基礎」の本文に戻って復習をします。
(5) ワークブックは空欄を埋めるだけでなく，ワークブックの余白を利用して，必要と思われる内容を書き足して，ワークブックを自身の「AIオリジナルノート」として完成させます。

上記以外にも，目的に応じてワークブックをご活用ください。

このテキストに書かれた内容は基礎知識ですが，非常に大切な内容です。始めから確実に理解することによって，これから発展学習が非常にスムーズに進みます。
本書籍の内容をひとつひとつ，丁寧に読み進めて,「AI・データサイエンスの基礎」の理解をさらに深めていきましょう。

ワークブックの解答解説をダウンロードしよう！

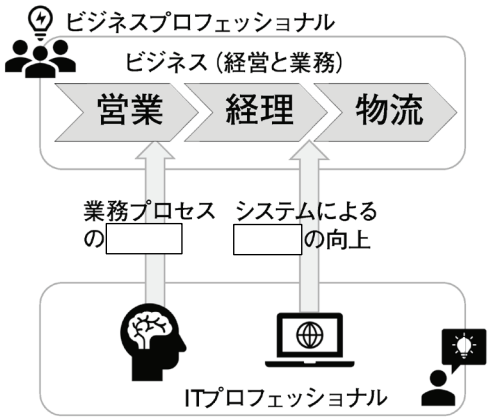
＜ダウンロード方法＞
①●●●●にWebブラウザからアクセスまたは，QRコードを読み取ってください。



② ダウンロードしたファイルを解凍してご利用ください。
※パスワードは「●●●●」です。

IT の基礎概念

ここまで見てきたとおり，ITとは，様々なデータを，情報として役立てることができるようにするために行われる様々な [　　] と，そのための [　　] を指しています。



図表　IT の概念

データを処理する技術には，形式化や符号化などがあります。これは伝達，解釈，処理などに適した形にデータを [　　] する技術です。

(1) 形式化

日本産業規格の「X0001 情報処理用語 - 基本用語」において，「データ」の定義は「情報の表現であって，伝達，解釈または処理に適するように**形式化**され，再度 [　　] として解釈できるもの」とされています。ここでいう形式化とは，データを [　　] にすることです。

例えば，電子メールは，「ヘッダ」と空白行を挟んで「メッセージボディ」で構成されます。ヘッダには，宛先や送信元，送信時間や件名などが，決められた [　　] に沿って記述します。だからこそ，どんなメールアプリを使っていても，同じようにメールを読むことができます。

日本語は文法の自由度が高いので，[　　] になりやすくなっています。

一般的に，人間同士のコミュニケーションに使われるのは日本語や英語などの自然言語ですが，このような自然言語は [　　] です。コミュニケーションの場では，自然言語を使っている以上，完全に [　　] をなくすのは難しく，意図せず [　　] が混在することを防ぐことができません。

では，完全に曖昧さをなくすためにはどうすればよいでしょうか。一つの方法としては，コミュニケーションに使用する言語自体を明確な形で定義することが考えられます。その言語で使用される「語彙」や，語彙から文を作る方法「文法」を完全に明確な形で定義し，定義した語彙と文法から生成できる文章によって，コミュニケーションを行います。

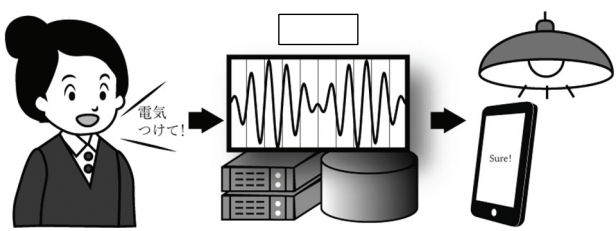
ITの分野では特定の目的のために作られ，厳格な文法に則った言語を［　　］といいます。コンピュータに指示を与えるためには，［　　］と呼ばれる言語を用いたプログラムを作成します。この［　　］も形式言語の一つになります。［　　］は情報処理技術の中核技術の一つです。

(2) 符号化

符号化とは，信号や情報を一定の規則に従ってデータ化することです。アナログ信号をディジタルデータに変換し，データを［　　］することなどがこれに当たります。

人間が直接取り扱う情報は，多くの場合，コンピュータが利用しようとするとかえって効率が悪くなってしまうことがあります。例えば，メニューを数字で選択する場合と，音声で［　　］のようなコンピュータに命令をする場合とでは，処理に大きな差があります。

［　　］は，まず音声データをディジタルデータに変換します。それを暗号化して［　　］のデータセンタに送り，何を命令されたのかを解析します。そして，その解析結果をスマートスピーカに戻して，何を行えばよいのかを［　　］するようにできています。これは，人間が使う情報を，文字，記号，数字などのコンピュータ内部表現（**コンピュータコード**）に変換する作業を繰り返し行っていることになります。これが［　　］です。



図表　符号化とスマートスピーカ

データの種類

データには大きく分けて，質的データと量的データの2種類があります。

(1) 質的データ

質的データは好きなスポーツ，血液型，自動車のナンバーなど，単に［　　］や［　　］を区別するためだけのデータです。質的データはさらに名義データと順序データの2種類に分類することができます。

名義データ：例えば，1. 明治，2. 大正，3. 昭和，4. 平成，5. 令和のように［　　］をデータの［　　］として用いていて，大小に意味はありません。

生年月日を記入してください
元号は番号を記入してください（1：明治 2：大正 3：昭和 4：平成 5：令和）

元号［　　］［　　］［　　］年［　　］［　　］月［　　］［　　］日

数字は右詰めで記入してください

図表：名義データ

順序データ：例えば，順位表（1位　山形，2位　福島，3位　沖縄…）のように［　　］を表すために用いていて，数値間の大小関係に意味はありますが，順序データ同士では計算はできません。

1世帯の構成人員の多い都道府県

順　位	都道府県名	1世帯の平均人
1	山形県	2.78
2	福島県	2.51
3	沖縄県	2.48
4	長野県	2.43
5	埼玉県	2.34
6	神奈川県	2.30
7	栃木県	2.26
8	千葉県	2.26
9	新潟県	2.25
10	山梨県	2.25

順序データ　　量的データ

図表：順序データと量的データ
出典：（総務省統計局＜引用加工＞）

(2) 量的データ

量的データは枚数，身長，金額など，［　　］で推し測ることができ，数字

の大小に [　　] データです。量的データは，さらに間隔データと比率データの2種類に分類できます。

間隔データ：温度などの [　　] データで，[　　] ことが特徴です。またアンケートの満足，少し満足，少し不満，不満，のようなデータは [　　] ですが，[　] を相対値で，評価平均などの平均値を算出することができます。

要素	気温(℃)		
	平均	日最高	日最低
統計期間	1981～2010	1981～2010	1981～2010
資料年数	30	30	30
1月	5.2	9.6	0.9
2月	5.7	10.4	1.7
3月	8.7	13.6	4.4
4月	13.9	19.0	9.4
5月	18.2	22.9	14.0
6月	21.4	25.5	18.0
7月	25.0	29.2	21.8
8月	26.4	30.8	23.0
9月	22.8	26.9	19.7
10月	17.5	21.5	14.2
11月	12.1	16.3	8.3
12月	7.6	11.9	3.5
通年	15.4	19.8	11.6

図表：間隔データ

比率データ：は間隔が等しく，[　] の概念が固定的であるデータです。野球やサッカーなどの得点は 0 が絶対値で，1，2，3 と [　　] で増加していきます。

市という地域軸で仕分けられた契約件数データ（人口はバラバラ）

	A市	B市	C市	D市
男性	18.3	24.1	78.8	45.3
女性	12.5	8.9	56.6	20.9
総人口	221	374	629	566

人口一人あたりの比率に変換されたデータ

	A市	B市	C市	D市
男性	8%	6%	13%	8%
女性	6%	2%	9%	4%
総人口	221	374	629	566

図表：比率データ

ここまで，四つのデータの種類を説明してきました。ここで注意したいのは，コンピュータは [　　] を見ただけでは，その数値がどの種類のデータに当たるのかを判断できないということです。コンピュータにデータを入力した場合，特に指定がない限りは [　　] として認識されます。

そのため，コンピュータは [　　] とは気付かずに，これは，コンピュータを扱う人間が対応しなければならないことです。したがって，コンピュータにデータを入力する場合，数値を羅列するだけではなく，数値の頭にラベルを付けて識別できるようにしておきます。どこに何のデータがあるかを見やすく表示しておくことは，後のミスを防ぐためにも非常に重要な作業です。