

【DRニュース・035】：データサイエンティストのブームとスキル・感覚・エンジニアとの共創

2017年08月25日発信

最近になって耳にする機会が増えた、データサイエンティストという職業。

調べて見ると、アメリカで「ワークバランスが良い職業」で、2016年1月：ナンバーワンとなっています。
 (No1：データサイエンティスト、No2：グラフィックマネージャー、No3：小学校教師、No4：ソーシャルメディア・マネージャー、No5：ソフトウェア・ディベロッパー、No6：ヘアリスト、No7：ウェブ・デザイナー・・・ しかもNo1は、扱っているのは数字なので、基本的にはどこでも仕事ができます)

どのような職業で、スキル・能力・経験が必要か～現在の“技術的なブレークスルー”を追ってみました。

1. データサイエンティストとは

データサイエンス(data science)とは、データに関する研究を行う学問である。

2007年に「データサイエンス」という言葉が新しく生まれ、そしてデータサイエンスを行う者として
 ... 「データサイエンティスト」は、ビジネスにおいて最も需要の多い職種の1つになりました。

アメリカのビジネス雑誌『ハーバード・ビジネス・レビュー』(Harvard Business Review)
2012年10月号は、データサイエンティストを「21世紀で最もセクシーな職業」と形容しています。

ハードウェア・ソフトウェアの進歩に伴い、大規模で複雑な「ビッグデータ」の蓄積・分析が可能となり、ビジネス分野でのビッグデータの活用が進められています。

... このビッグデータを分析する人材が「データサイエンティスト」です。

(1) データサイエンティストとは、どのような職業なのか？

① データの収集と貴重な情報

Facebookのアカウントにはユーザーの貴重な情報がつまっていることや、Googleが個人についてのあらゆる情報を把握したいと意図していることは知られていますが、昨今では小規模なスタートアップ起業であっても ... 事業戦略に転用するために、さまざまなデータを収集しています。

しかし、多くの場合、企業が収集している生データは非常に乱雑なものです。不完全で整理されておらず、ちぐはぐに分類されていたり、明らかに間違っものが紛れ込んでいることも多いのです。
 ... それでも、そこには多くの貴重な情報があります。

② データサイエンティストの仕事

データサイエンティストの仕事は、統計学、コンピュータサイエンス、データ分析を駆使して、膨大なデータを構造化しながら整理して、企業がデータを活用したアクションを起こすために
..... 必要な情報となるように、解析結果を導き出すというものです。

データサイエンティストは、テラバイト単位のデータを解析している最中は、自分たちが何を探すべきなのかをまだ把握していないことがあります。しかし、ひとたび興味深いデータを見つけた時には、それがわかります。

.... このような発見には、直感やひらめきが必要であるため、仕事として成立している。

データサイエンティストは、発見した情報の意味を適切な表現で他人に伝える必要があります。通常の管理職は、データサイエンティストほどデータ分析の言葉に精通していないからです。

..... 要するに、データサイエンティストは、大量のデータを分析し、それらのデータを実行可能な事業戦略に変換するのが仕事です。

データサイエンティストはビッグデータの分析及び分析結果をもとに、問題の解決や状況改善のための施策立案を行います。大量のデータ、すなわちビッグデータから

..... ビジネスに活用できる情報を引き出す専門技術者のことです。

例えば、大量に蓄積された顧客データの分析によって、新商品開発や現行商品の改善に直結するヒントやアイデアを出したり、顧客の消費行動の変化を踏まえて、

..... 効果的な販売戦略を打ち出したりすることが、データサイエンティストの仕事です。

③ データサイエンティストという職種

これは簡単な仕事ではありませんが、企業にとって非常に重要で、これからもそうあり続けるでしょう。このようなことから、データサイエンティストという仕事は、

..... 職種の未来は明るく、安泰であると期待されています。

(2) データサイエンティストに必要なスキルは何か？

① データ解析や分析のスキル

データサイエンスは変化が急速で、分野としての明確な定義もないため、データサイエンティストが持っているスキルは多岐に渡ります。

..... 彼らのほとんどが統計学、データ分析、数学などの何らかの訓練を受けています。

そしてほぼ全員が、データ保存、統計学、機械学習などに使用される言語による
(特に Python、R、Hadoop、SQL など) プログラミングの経験を持っています。

..... [データ解析において特に人気がある Python が最適でしょう。また、MATLAB、SAS、Minitab などの別のデータ分析プログラムを知っていれば、なお役立つでしょう。](#)

※ **データ分析のプログラミング言語やデータ解析などは、下記の“SEED”も参照してください。**

・【[DRニュース・025](#)】: ビッグデータの活用事例・統計学とデータ解析の違い&AI 関連

・【[DRニュース・027](#)】: プログラムはなぜ動くのか? どんな環境で動くのか? 基礎知識

② [データサイエンティストの能力と経験](#)

機械学習 (AI : 人工知能による学習) や統計学、データ分析に関する知識のない人たちに、
..... [きちんと説明する能力も非常に重要です。](#)

画期的な発見があったにもかかわらず、うまく説明できなければ、何の役にも立ちません。
..... [意思伝達を適切にこなす能力は、今日、あらゆる技術者に必要とされるソフトスキルです。](#)

データサイエンティストを志望するなら、現在取り組んでいる分野の内外を問わず、
..... [複数の分野における経験が価値を持ちます。](#)

仕事を進めていく上では、創造力を働かせながら、さまざまな角度から問題に取り組む能力があれば非常に役に立つでしょう。

..... [新しい問題には、革新的かつ固有の解決法が求められることが多いからです。](#)

③ [ビジネスモデルを立案、論理的思考能力とコミュニケーション能力](#)

企業におけるデータ分析はビジネスの改善や問題解決を目的とするため、
..... [データサイエンティストには、データ分析以上の能力が求められます。](#)

データサイエンティストはデータ分析から得られた知見をビジネスにどのように活かせるのかを考え、**ビジネスモデルを立案**します。したがって、統計・確率やデータ処理に関する高度な知識を持っているだけでなく、

・・・ [自社のビジネスに関する理解や競合他社の把握、さらには幅広い IT スキルが必須です。](#)

データ分析の現場で飛び交う専門用語は専門家でないと理解が難しいものです。分析結果や分析から導かれた戦略を、各部門の担当者あるいは経営層に対し専門用語を使うことなく、

・・・ [分かりやすく伝えるための論理的思考能力とコミュニケーション能力も必要となる。](#)

2. データサイエンティストのブームと定着

2016年06月「[出典：Googleトレンド](#)」の記事から、[2013年頃を振り返ると](#)、当時言いはやされていた“[ビッグデータ](#)”や“[データサイエンティスト](#)”を取り巻く環境は、今どのような状況なのだろうか？

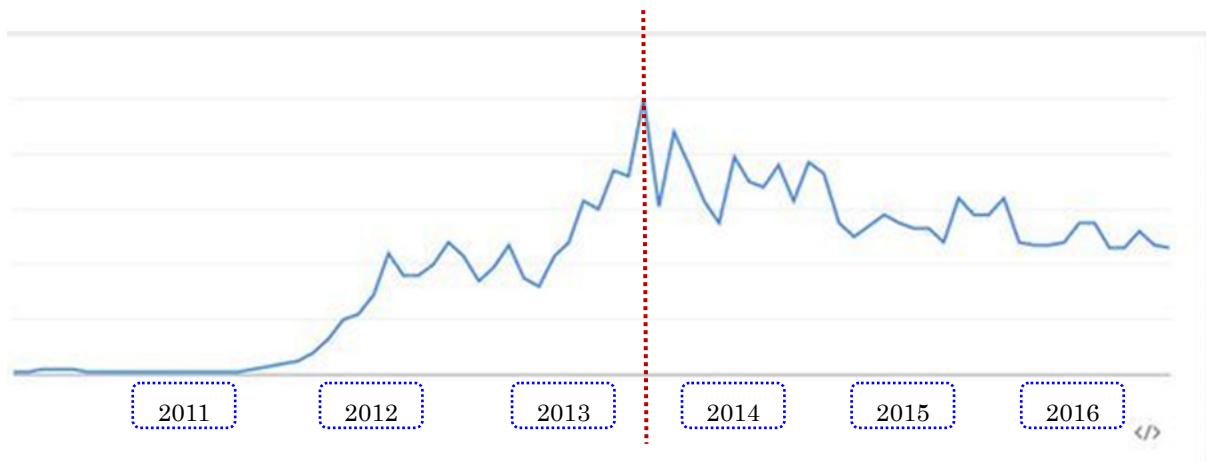
(1) ビッグデータのトレンド

ビッグデータは、[2013年7月をピーク](#)に現在にかけて緩やかな減少傾向にあると言えます。

当時は色々なメディアで日夜取り上げられていることが多かったのですが、最近では見かけることはあれど、それ自体がメインに特集されている記事なども少なくなっており、実感に伴った結果であると言えるのではないだろうか？

データの多寡は現在ではさほど問題ではなく、ビッグデータであれ、スモールデータであれ、適切に取得、処理、分析により、ビジネスに生かす動きが浸透した証左ではないかと考えています。

(出典：Googleトレンド)



(2) データサイエンティストのトレンド

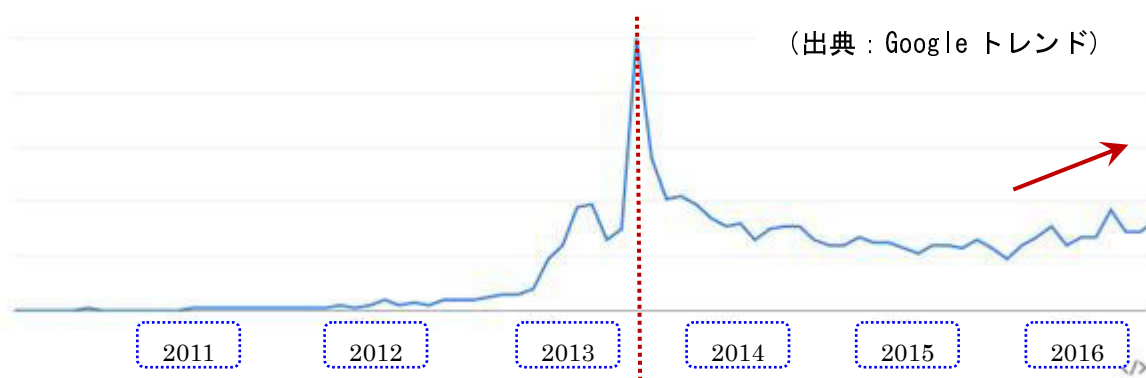
Facebook が求人を出したことで瞬く間に世界を駆け巡った職業、データサイエンティスト。ビッグデータを巧みに操ることで価値創出を図る同職種には多くの期待がかけられていました。

しかし、こちらのワードも検索ボリュームで言えば、[ビッグデータと同様に 2013年7月がピーク](#)であったようです。

異常検知などの技術領域で、ある時突然検出される異常値を「バーストする」と言うこともありまますが、データサイエンティストの検索ボリュームはそれと同じようなバーストになっています。

こちらも同様に緩やかな減少傾向にありますが、

．．．．． **2016 年からは上昇傾向になっていることが少し伺えます。**



(3) なぜ今になってボリュームが上昇傾向？

なぜ今になってボリュームが上昇傾向になっているかは、諸説ありそうですが、

．．．．． **次の2点の仮説が考えられます。**

まず1点目。それは「ディープラーニング」や「人工知能 (AI)」の流行です。

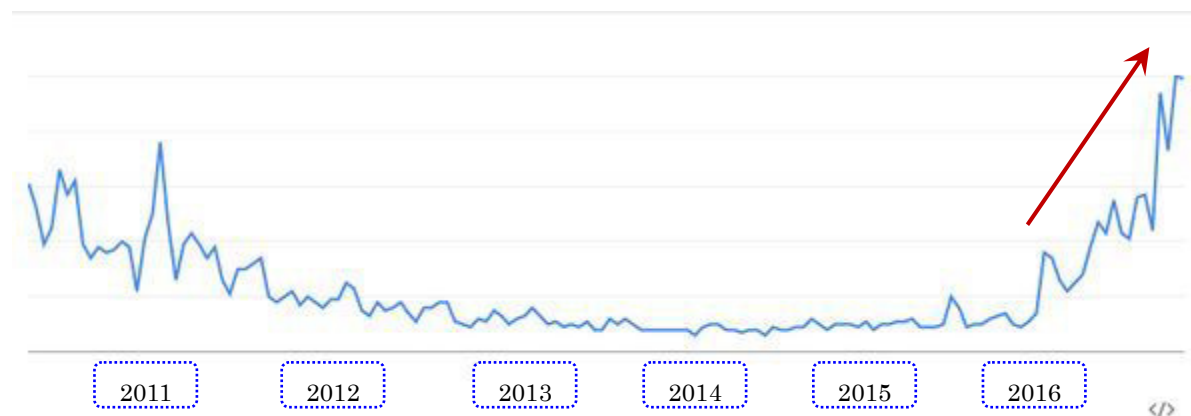
Google DeepMind の AI である「[AlphaGo](#)」が囲碁のトップ棋士に勝ったというニュースは記憶に新しいと思いますが、彼らのチャレンジによって世間の期待が増していることが検索ボリュームからうかがい知ることができます。

① 人工知能 (AI) の検索ボリューム

このグラフには無いが、人工知能は **2004～2005 年頃にもブーム**があった様子がわかります。

．． **技術的なブレイクスルーが起こる**と定期的に盛り上がりを見せることもあるのでしょう。

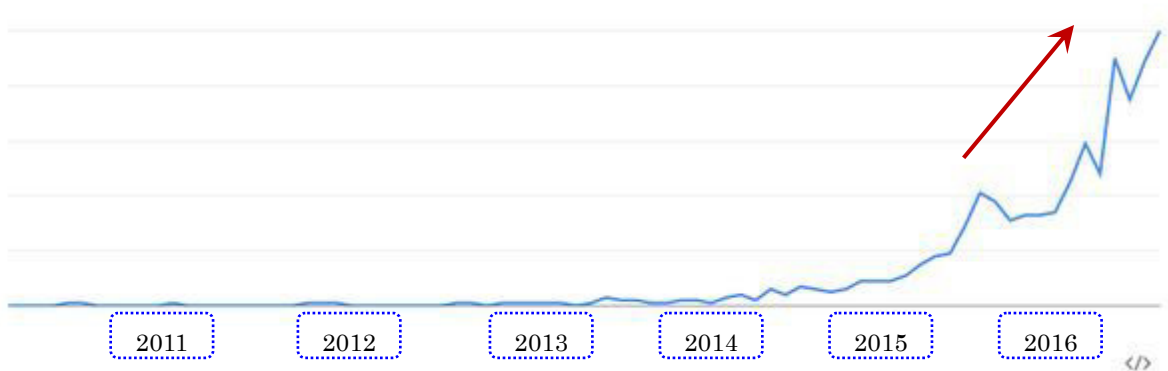
(出典：Google トレンド)



② ディープラーニングの検索ボリューム

ディープラーニングは Google の YouTube の猫画像の判別をきちんと行ったという衝撃的な発表がされてから数年が経過しますが、その間さまざまなジャンルの学会でディープラーニングを用いた研究や論文が公開され続け、ついには AlphaGo が世界的な棋士に勝利したため、爆発的な人気ワードになっています。

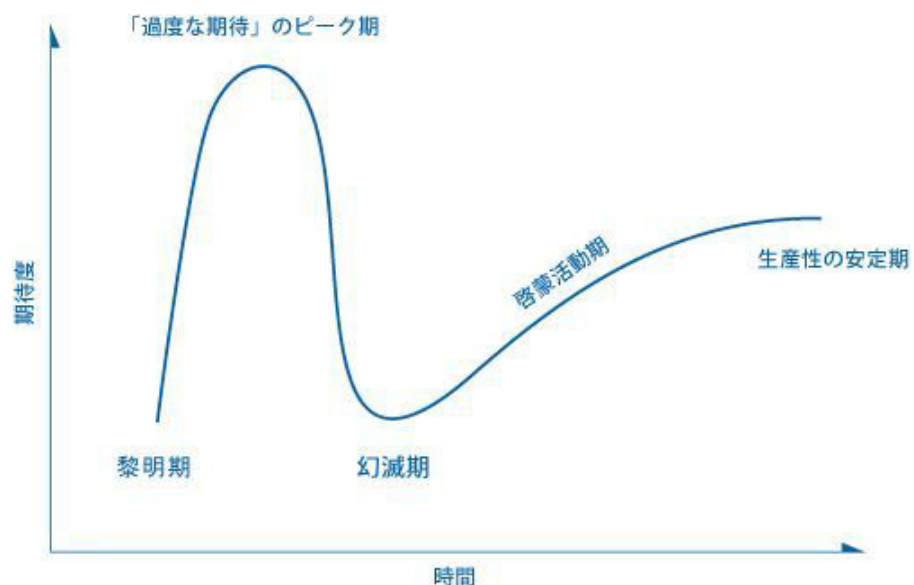
(出典：Google トレンド)



これらの新たなブームに相関してそれを操る職種とした“データサイエンティスト”の注目も上がっている点です。

また、もう一つの1点目。 Gartner のある技術の成熟度や採用度
社会への適用度を示す図、「ハイクサイクル」に応じた動きを見せていることも考えられます。

「過度な期待のピーク期」を経て、幻滅期に移行した後、啓蒙活動期に差し掛かってきたのではないかとする説です。



1995 年以来、ガートナー社は「ハイブ・サイクル」を用いて・・・新技術の登場によって生じる過度の興奮や誇張 (hype、ハイブ)、そしてそれに続く失望を説明しています。

それはまた、技術がいかにして、そしていつ次の段階に進み、実際に利益を生み出し、そして広範に受け入れられるか、も示しています。

ガートナー社の唱える「ハイブ・サイクル」の目的は、現実から誇張 (ハイブ) を切り離すことにより、CIO や CEO が・・・特定技術の採用可否を判断できるようにすることにあります。

3. データサイエンティストとエンジニアの共創

テクノロジーの活用が必要不可欠な時代にあって、エンジニアとデータサイエンティストが脚光を浴びている。「テクノロジーが分かる人材」は今や業種・業界を問わず引っ張りだこだ。“似て非なる存在”である両者の強み、果たすべき役割とは——。2017. 5. 11 の「テクノロジストを生かす道」から!!

(1) 課題の“テクノロジスト”不足

——クラウド、IoT、人工知能 (AI) といったテクノロジーの積極活用が企業の競争力を高めると言われ、これらを操るソフトウェアエンジニアやデータサイエンティストといったデータやテクノロジーに精通する技術者、いわば“テクノロジスト”の存在価値は高まるばかり。
 両者の不足は、慢性的な課題になっています。

(2) データサイエンティストは万能ではない

——これからは、新たなビジネスを生むためにデータサイエンティストの存在は欠かせません。

データサイエンティストは、複数の機器やシステムが収集している
「データを構造化」しながら「整理して」・・・新しい気づきや価値創出の手助けをする存在です。

大量のデータが生まれ、データによって解決できるビジネス課題が豊富な環境の中で、価値を発揮する人材だと期待されています。

ただ、データサイエンティストだけでは新しいビジネスを生み出すことはできません。
 データサイエンティストは、言わば研究者です。

自分の研究テーマに没頭してひたすら調査したり、実験したりするのが得意です。
 ただ、その結果から新しいプロダクトやサービスを生む力は弱いものです。

(3) データサイエンティストとエンジニアの違い

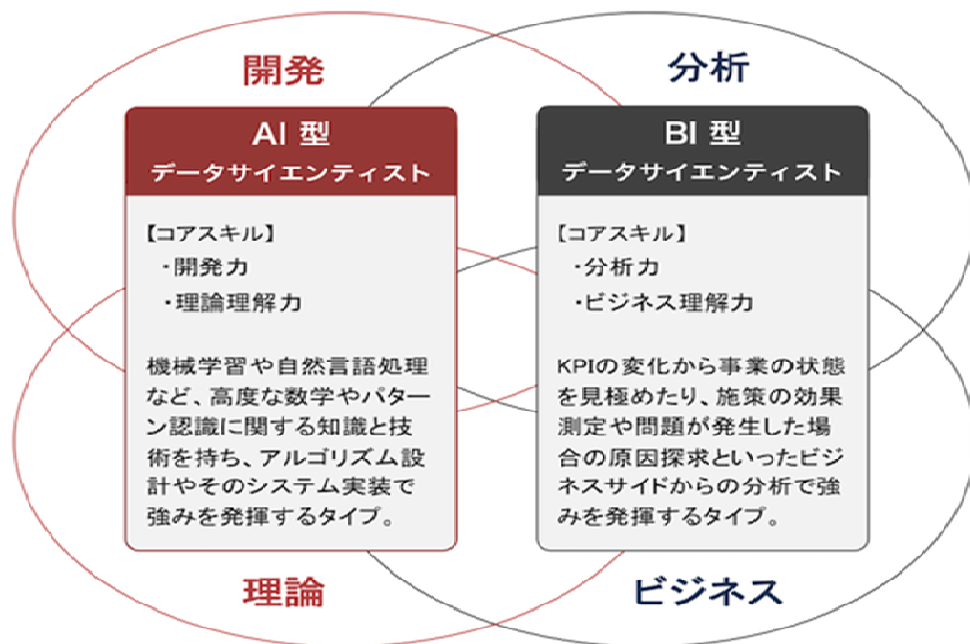
データサイエンティストとエンジニアの違いを見極めることは、組織を設計する上で重要となる。

- ① データサイエンティストは、**データを統合→構造化→新しい切り口を見つける人。**
- ② エンジニアは、**トレンドを先読みする→変化に適応するための仕組みを構築する人。**

(4) データサイエンティストの2つのタイプ

「データサイエンス」として表現される領域は広く、関連分野の包括的な理解が求められますが、ここでは得意な領域にフォーカスしてデータサイエンティストを2つに分けて考えて見ます。

データサイエンティストの2つのタイプ



AI型：理論設計やエンジニアリングの領域で強みを発揮

BI型：ビジネス課題を解決に導くための分析で強みを発揮

- ① **AI** (Artificial Intelligence) **型**：より学問的な領域、例えば数学や機械学習理論、コンピュータサイエンスといった領域を得意とするデータサイエンティストです。
- ② **BI** (Business Intelligence) **型**：このタイプはビジネスアナリストに近く、事業について数値を用いた解釈や状況判断、分析を得意としています。ビジネスを取り巻く環境や起きている現象を数字と結び付けて考えることに長けています。

(5) スキルや得意領域によるデータサイエンティストの協働

AI 型データサイエンティストは；理論と開発で活躍 **(エンジニア的な素養)**

機械学習プロジェクトや人工知能 AI と呼ばれる分野の研究開発で主に活躍するため、エンジニア的な素養がより強く求められます。

. . . **機械学習プロジェクトでは課題の解決に適したアルゴリズムの選定から、実際にプロトタイプによる実験を通じて精度を高めていくフェーズで力を発揮します。**

BI 型データサイエンティストは；分析力とビジネス感覚で勝負 . . . **(ビジネス的な感覚)**

ビジネスサイドの立場で関わるが多くなります。ビジネスサイドとしては、開発する機能がどのような利便性や価値提供につながるのかといった理解が大事になります。

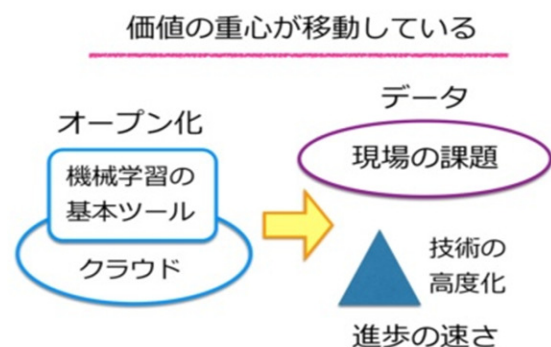
. . . **機械学習プロジェクトでは、どのようなデータを集めればいいのか、そのデータをどのように活用すれば精度の高い分類や予測を導き出せるかを見極めるなど、そういった数値的感覚が要求されるフェーズで力を発揮します。**

4. 機械学習・深層学習 どう学び、どう活用する？

機械学習や深層学習（ディープラーニング）は、人工知能（AI）技術の分野の一つです。

(1) エンジニアも機械学習を学ぶべき

機械学習の活用では一般に、課題の検討から始まって、自社にどんなデータがどれくらいの規模で存在し、期待できる効果はどの程度か、新たなデータ収集のコストはどれくらいか、といった分析を繰り返すサイクルが必要になります。



. **一方、解決したい課題やデータを持っているのは、現場のエンジニアの方々です。**

特に製造業の企業が持っている課題やデータは、様々な制約で外部には出せない場合が多く、データ自身が貴重です。また、継続的にこうしたデータが得られる仕組みは、さらに貴重です。

. **解決したい課題やデータを持っているところに、価値の重心が移動している。**

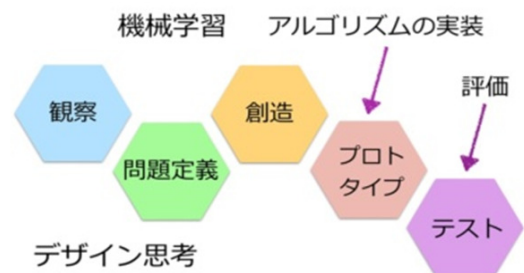
(2) 機械学習×デザイン思考=課題解決

エンジニアの方々が機械学習や深層学習を学ぶ上でヒントを与えてくれるのが、「**デザイン思考**」という考え方です。これは、**実践的、創造的な課題解決のための方法論**です。

..... **共感、問題定義、創造、プロトタイプ、テストのステップを繰り返します。**
課題解決に機械学習を適用するときには、**デザイン思考のプロトタイプ、テストのステップで、**
..... **機械学習のアルゴリズムの実装と評価を行います。**

機械学習や深層学習の理論を、教科書で学ぶだけでなく、**実践的な経験を積み重ねていく上で、デザイン思考の方法論は、とても参考になる。**

機械学習 × デザイン思考 = 課題解決



(3) 「機械学習」と「深層学習」は、何が違うのか？

まず「**機械学習**」を見てみよう。同技術は、開発者があらかじめ、**すべての動作をプログラム**

するのではなく、データをAI自身が解析し、法則性やルールを見つけ出す特徴を持っている。

・・・例えば、**画像認識**で、青リンゴ、赤リンゴのタグを読ませると「**色に着目して区別しなさい**」と**AIに指示を与えておけば**、まだ解析していないリンゴの画像が出てきたときでも、**AIはリンゴの色に着目する。**

機械学習では、「色」に着目するように指定しなければならなかったのに対して、

深層学習では、区別するための「目の付けどころ」をAIが自分で学習し、その性能を向上させていく。別の言い方をすれば、沢山のデータを見ることによって、どこに注目すればよいかを自分で学習し、**人間からの指示を待たずに自動でどんどん賢くなっていくということです。**

機械学習では、人間がある程度学習の方向性をコントロールできる一方で、

深層学習のディープラーニングは、場合によって思わぬ方向に学習が進む可能性があります。

「**機械学習**」と「**ディープラーニング**」については、AIの仕組みではあるものの、機能強化の自動化を推し進めているという違いがあるといえます。特に、分析の対象を区別する際に「**目の付けどころ**（「**特徴量**」という）」を自動的に見つけ出す点で、進化していると言えるだろう。

特にディープラーニングは、言葉で特徴を定義するのが難しい場合に高い効果を発揮します。

いつの時代にも、技術の進化（ブレークスルー）に精通した人材が必要とされています。

また、解決したい課題やデータを持っているのは、私たちの身の廻りの作業現場です。
基礎となる学問や新たな技術の進化など経験や感覚として養い、食欲に吸収することが重要です。