

厚生労働科学研究費補助金 政策科学総合研究事業
(臨床研究等ICT基盤構築・人工知能実装研究事業)
保健医療用人工知能の技術革新と国際競争力向上に資する人材育成
に関する研究 (H29-ICT-一般-004)

人工知能入門 人工知能の社会実装のために

千葉大学病院病院長企画室
病院経営管理研究センター 特任講師
亀田 義人

1. 人工知能とは何か
2. 人工知能を活用する意義
3. 人工知能の歴史とトレンド
4. ディープラーニングの理論的な背景
5. 医療における人工知能の活用領域
6. 人工知能の限界と適切な活用
7. 人工知能を活用した製品の調達について

1. 人工知能とは何か

2. 人工知能を活用する意義

3. 人工知能の歴史とトレンド


4. ディープラーニングの理論的な背景

5. 医療における人工知能の活用領域

6. 人工知能の限界と適切な活用

7. 人工知能を活用した製品の調達について

そもそも知能とは？  一般的には
知的な活動の能力 をいう。

知的な活動とは？  問題解決・推論・学習などの情報処理能力
抽象化・一般化など

人工知能とは？  一般に、
「何らかの知的動作が可能な計算機システム」

「知的動作」とは？  定義が無い

**人工知能では、知能のシミュレーションすることになるが、
知能のシミュレーションが知能といえるか？という疑問**

国内の主な研究者による人工知能の定義

中島秀之	公立はこだて未来大学	人工的につくられた、知能を持つ実態。あるいはそれをつくろうとすることによって知能自体を研究する分野である
武田英明	国立情報学研究所	
西田豊明	京都大学	「知能を持つメカ」ないしは「心を持つメカ」である
溝口理一郎	北陸先端科学技術大学院	人工的につくった知的な振る舞いをするためのもの（システム）である
長尾真		
堀浩一		
浅田稔		
松原仁		
池上高志		
山口高平		
栗原聡	電気通信大学	いるものを想像している
山川宏	ドワンゴ人工知能研究所	計算機知能のうちで、人間が直接・間接に設計する場合を人工知能と呼んで良いのではないかと思う
松尾豊	東京大学	人工的につくられた人間のような知能、ないしはそれをつくる技術。人間のように知的であるとは、「気づくことのできる」コンピュータ、つまり、データの中から特徴量を生成し現象をモデル化することのできるコンピュータという意味である

本邦での業界の第一人者でも人工知能の定義は様々であり、統一的な定義が成されているものではない

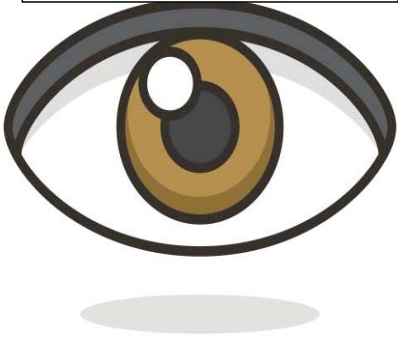
総務省 平成28年度情報通信白書



人間のような振る舞いをする知的コンピューターであり、設計した時の想定を超えた新たな問題にも対処する。

人間が設定した特定の問題（画像の識別等）に対して、主に機械学習を用いて解決する。

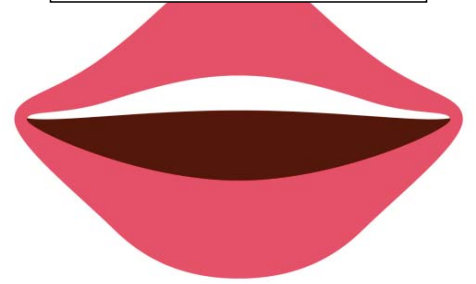
画像認識



音声認識



自然言語処理



人工知能の持つ機能

人工知能と機械学習及び、ディープラーニングの関係性



人工知能の 二つの 理論背景

線形代数

DNNなど

確率統計

ベイズ統計など

DNN: Deep Neural Network

人工知能の 二つの 問題解決 手法

回帰問題の解決

結果の
推計・推定

分類問題の解決

分類

大量の
利用可能
なデータ

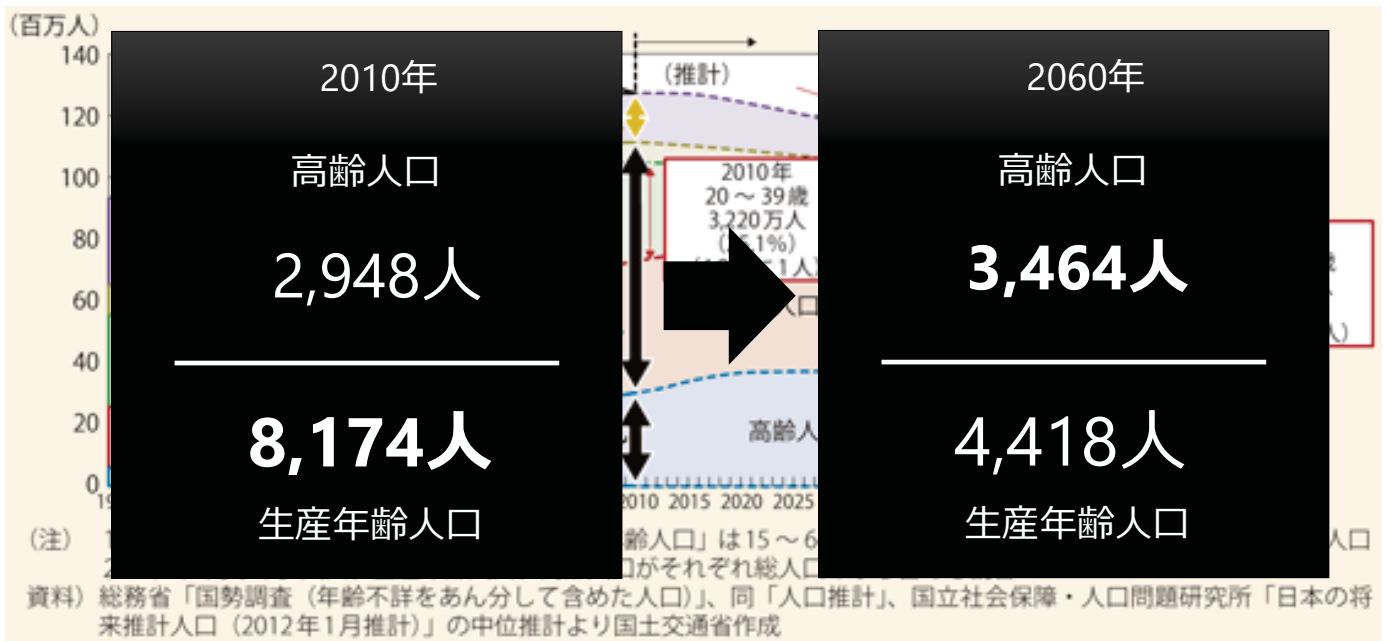
計算機能
力の向上

機械学習
アルゴリズム
の開発

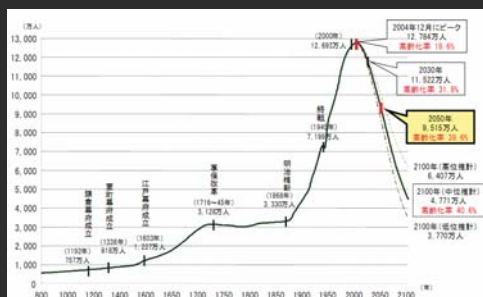
人工知能に問題解決をさせる上で重要な要素

1. 人工知能とは何か
2. 人工知能を活用する意義
3. 人工知能の歴史とトレンド
4. ディープラーニングの理論的な背景
5. 医療における人工知能の活用領域
6. 人工知能の限界と適切な活用
7. 人工知能を活用した製品の調達について

日本の人口及び産業構造の変化



人口ボーナス期と人口オーナス期



人口ボーナス期

総人口に占める「生産年齢人口(15歳～64歳の人口)」比率が圧倒的に多い状態。

安価で豊富な労働力があり、従属人口が少ないため、教育費や社会保障費の負担が少ない状態。

国家予算を経済政策に振り向けやすく、また他国からの投資を呼び込めるので、経済が活性化する。

後半になると高学歴化による人件費上昇と晩婚化・晩産化による少子化していく。

人口ボーナス期で経済発展に成功した後、医療や年金制度が充実して高齢化が進み人口オーナス期に突入する。オーナスとは、「重荷・負担」という意味。

「支えられる人」が「支える人」を上回り、社会保障費などが重い負担となるため、消費や貯蓄、投資が停滞する。

日本は人口オーナス期に突入して長いですが、十分な対策がなされてこなかった。

人口オーナス期

各国で人口知能は国家戦略に位置づけられている

ドイツ	インダストリー4.0
日本	世界最先端IT国家創造宣言
米国	米国人工知能研究開発戦略
中国	新一代人工智能発展計画

世界最先端IT国家創造宣言 目指すべき社会・姿

1. IT 利活用の深化により未来に向けて成長する社会
2. IT を利活用したまち・ひと・しごとの活性化による活力ある社会
3. IT を利活用した安全・安心・豊かさが実感できる社会
4. IT を利活用した公共サービスがワンストップで受けられる社



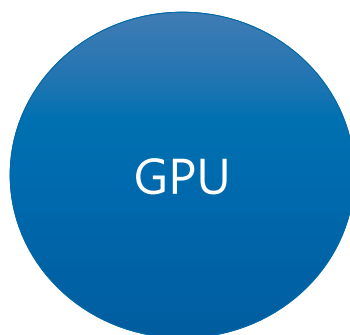
産業革命前後

同様の変革が起きようとしているのかもしれない

1. 人工知能とは何か
2. 人工知能を活用する意義
- 3. 人工知能の歴史とトレンド**
4. ディープラーニングの理論的な背景
5. 医療における人工知能の活用領域
6. 人工知能の限界と適切な活用
7. 人工知能を活用した製品の調達について

人工知能の3回のブーム

第1次	1950年代 ～	コンピューターによる「推論」や「探索」、パーセプトロンや人工対話システムELIZA	複雑な問題の解決が困難
第2次	1980年代 ～	「エキスパートシステム」、「ファジィ」や「ニューロ」コンピューティング	十分な情報処理能力をコンピューターが持たず限定的な能力も持つにとどまり、商用的にも一部「ファジィ」搭載という程度
第3次	2000年代 ～現在まで	「ディープラーニング」（2006年）などの機械学習アルゴリズム	「ビッグデータ」を、この間発達したコンピューターの計算能力（GPU/CPU）を元に、機械学習アルゴリズムを用いることにより様々な課題解決ができるようになった



人工知能に問題解決をさせる上で重要な要素

DATAの標準化・規格化とBIG DATAの活用

- 日本ではまだそのインフラ整備が途上にあるが、様々な領域で電子化が進んでいる
- 海外では例えば米国のNational Library of Medicine内のNational Center for Biotechnology InformationのPubmedなど、標準化された情報の集積化が先行している

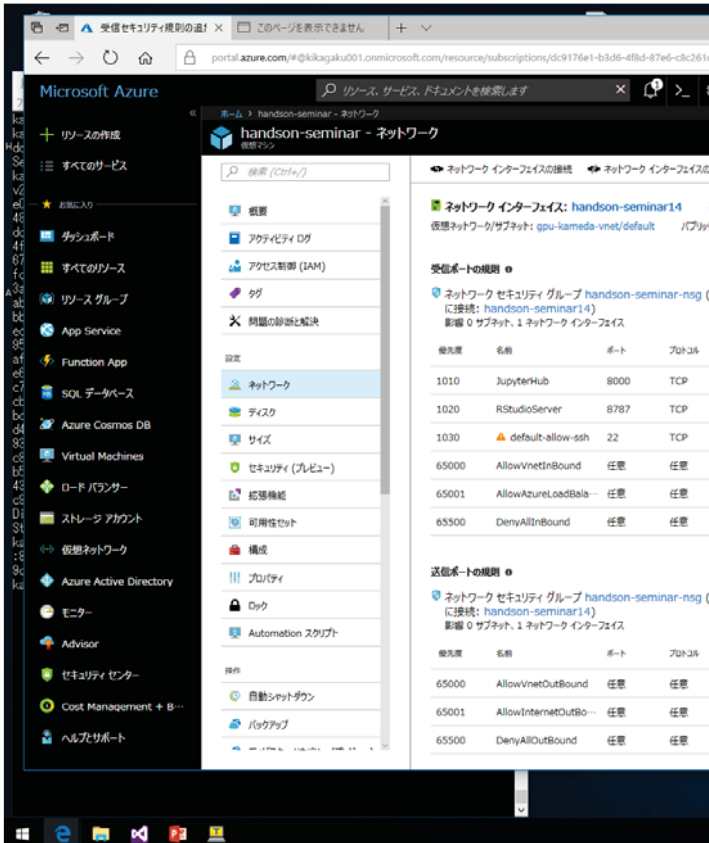
今後はIoT等により標準化された情報の集積と活用も課題



高性能なGPUの登場により演算能力の課題が克服された

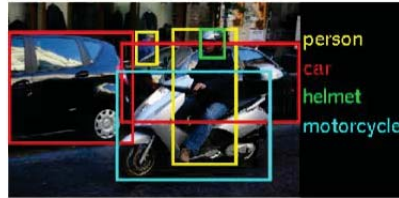
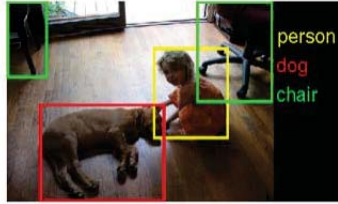
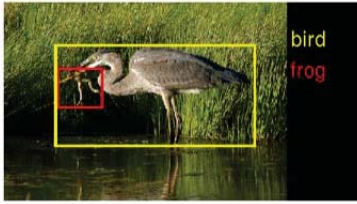
- GPU : Graphics Processing Unite

CPU(Central Precessing Unite)に対して、特に主に画像処理に関して処理能力を高めたもの。CPUが、一つのコア（演算処理を行う部分）が複雑な計算を行う事を念頭に設計されているのに対して、GPUでは単純な計算を多数並列処理させる事を念頭に設計されており、コア数も多い。



クラウドによるGPU利用環境提供サービスの普及

- 自前で大量のGPUを搭載したコンピュータを構築しなくても、クラウド環境でGPU利用環境を提供してくれる有料サービスが普及している。
- 左はMicrosoft Azureの仮想機械に実際にアクセスしているところ



※画像はILSVRC 2013のもの

- 2012年に開催された画像認識コンテスト、Large Scale Visual Recognition Challenge 2012 (ILSVRC2012)において、カナダ・トロント大学のGeoffrey Hintonチームがディープラーニングを用いて、著しく高い精度での画像認識を達成。

<http://www.image-net.org/challenges/LSVRC/2013/>

DEEP LEARNINGの登場

IEEE SPECTRUM

Follow on: [f](#) [t](#) [in](#) [+](#) [m](#)

Engineering Topics ▾ Special Reports ▾ Blogs ▾

Interactive: The Top Programming Languages 2017

Find the programming languages that are most important to you

The Institute of Electrical and Electronics Engineers

<https://spectrum.ieee.org/static/interactive-the-top-programming-languages-2017>

プログラム言語 Pythonの台頭

IEEE spectrumのプログラム言語ランキングにおいて、2016年に3位であったPythonが2017年に遂に1位に上昇。

人口知能・機械学習の利用環境がフリーで整えられ、社会実装の促進に貢献

Language Rank	Types	Spectrum Ranking
1. Python	🌐 🖥️	100.0
2. C	📱 🖥️ 🖨️	99.7
3. Java	🌐 📱 🖥️	99.4
4. C++	📱 🖥️ 🖨️	97.2
5. C#	🌐 📱 🖥️	88.6
6. R	🖥️	88.2
7. JavaScript	🌐 📱	85.4
8. PHP	🌐	81.1
9. Go	🌐 🖥️	75.8
10. Swift	📱 🖥️	75.0
11. Arduino	🖨️	72.4
12. Ruby	🌐 🖥️	72.1
13. Assembly	🖨️	71.8
14. Matlab	🖥️	68.6
15. Scala	🌐 📱	68.1
16. HTML	🌐	67.2

人工知能利活用のための無料プラットフォーム



Documentation Blog Contact Anaconda Cloud [Download](#)

Anaconda Data Science Certification

Objectively Demonstrate Your Data Science Experience

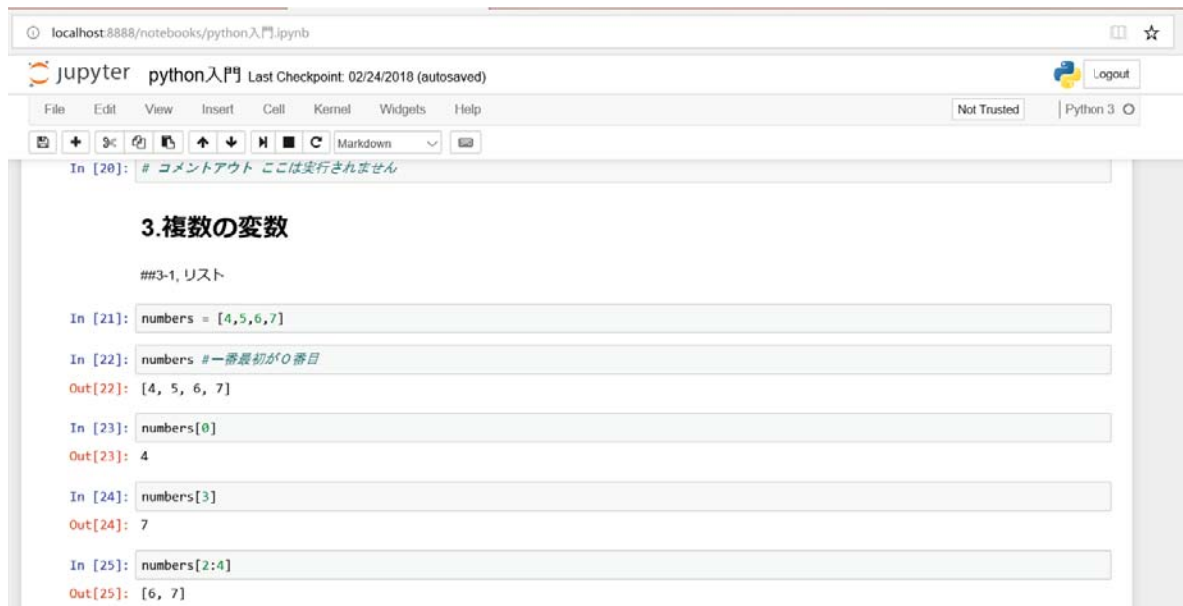
[Learn More](#)

The Most Popular Python Data Science Platform

機械学習やデータ処理が簡単にできるライブラリ※を多数標準装備したPythonのDistribution（利用環境をパッケージ化したもの）も登場

※ある特定の機能を持つプログラムを定型化して、他のプログラムが引用できる状態にしたものを、複数集めてまとめたファイル

ブラウザ上でJupyter notebookというPythonのインターフェースも活用可能



```
In [20]: # コメントアウト ここは実行されません

3.複数の変数

##3-1, リスト

In [21]: numbers = [4,5,6,7]
Out[21]: [4, 5, 6, 7]

In [22]: numbers # 一番最初の番号
Out[22]: [4, 5, 6, 7]

In [23]: numbers[0]
Out[23]: 4

In [24]: numbers[3]
Out[24]: 7

In [25]: numbers[2:4]
Out[25]: [6, 7]
```

代表的なDNN※ライブラリ

※DNN : Deep Neural Network DNNを用いてDeep Learning させる

TensorFlow

googleが開発したDNNライブラリ。googleが実際に使っているライブラリをオープン化している。

<https://www.tensorflow.org/versions/master/tutorials/>

Chainer

国内ベンチャー企業であるPreferred Networksが開発したDNNライブラリ。Preferred NetworksはMicrosoftとの協業や国内各大手企業と提携している。こちらもライブラリをオープン化している。

<https://docs.chainer.org/en/stable/guides/>

1. 人工知能とは何か
2. 人工知能を活用する意義
3. 人工知能の歴史とトレンド
- 4. ディープラーニングの理論的な背景**
5. 医療における人工知能の活用領域
6. 人工知能の限界と適切な活用
7. 人工知能を活用した製品の調達について

人工知能の 二つの 理論背景

線形代数

DNNなど

確率統計

ベイズ統計など

人工知能の3つの学習様式

教師あり学習

教師なし学習

強化学習

人工知能を活用した問題解決の基本的な枠組み

- 1) 解決したい問題に対して計算モデルを決める
- 2) 評価関数（損失関数）を決める
- 3) 評価関数（損失関数）を最大化もしくは最小化する

人口知能を活用した問題解決の基本的な枠組み

重回帰分析を例にとると

1) 解決したい問題に対して計算モデルを決める

⇒回帰式（目的変数を説明変数で計算する式）を設定

$$y = a_1x_1 + a_2x_2 + \dots + a_kx_k + b$$

2) 評価関数（損失関数）を決める

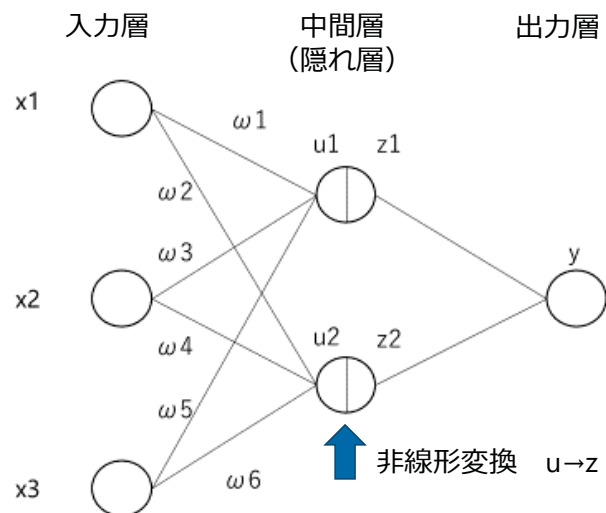
⇒最小二乗法

3) 評価関数（損失関数）を最大化もしくは最小化する

⇒「従属変数の測定値と、独立変数の測定値および回帰式を用いて求めた推定値の差の二乗和誤差」が最小になるように回帰式の係数を求める

シンプルなDNNモデル例

DNNは入力層、中間層（隠れ層）、出力層からなる。中間層を何層置くか、隠れ層のノードをいくつ置くかについてもヒトが設定する必要がある。パラメーターWの値ははじめランダムに設定する。



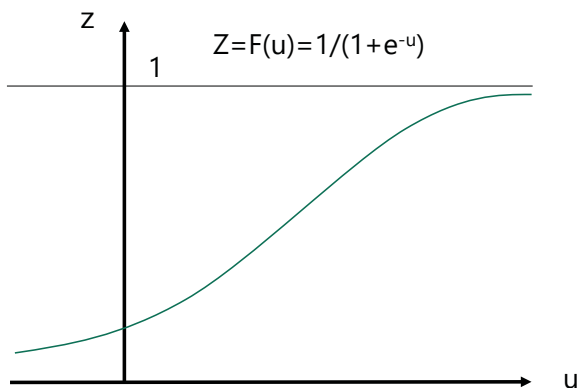
パラメーターで重み付けされた
入力から出力値を算出（順伝播）

誤差逆伝播法によるパラメーターの更新

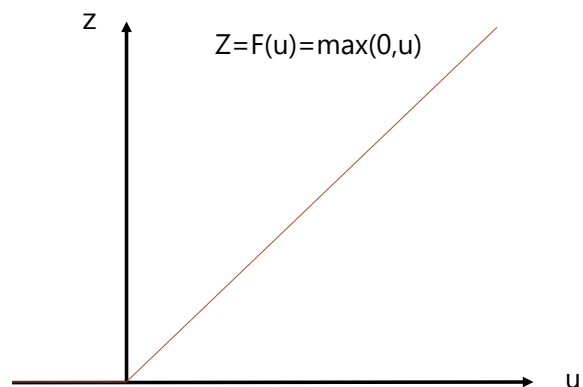
上記は回帰モデル。
分類問題の場合は出力層yが複数となる。

非線形変換に用いる活性化関数

シグモイド関数



Relu関数



uの負の値は無視、数値の大きさも考慮できる

評価関数（損失関数）を決める

回帰問題のとき

平均二乗誤差

Mean Squared Errorを用いる

$$L=(y-t)^2$$

分類問題のとき

ソフトマックス関数を用いて

クロスエントロピー誤差を求める

(参考)

ソフトマックス関数

$$f(x) = \frac{\exp(x_i)}{\sum_{j=1}^d \exp(x_j)}$$

クロスエントロピー誤差

$$L(hL) = \sum_i (y_i \log hL_i + (1-y_i) \log (1-hL_i))$$

実際はChainerライブラリ等の中に定義されていて
F.softmax_cross_entropy(y,t)等と入力すれば計算してくれる

損失関数を最小化する (+ そのためのパラメータwを求める)

最急降下法

1) ランダムにwの初期値を決める

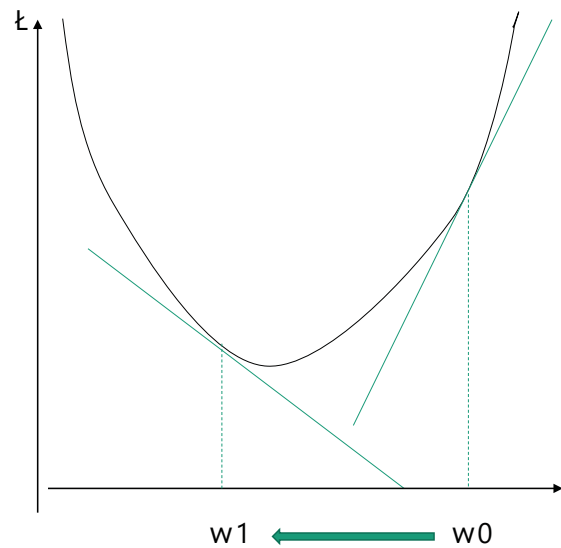
2) 損失関数の微分値 (傾き) を求める

$d/dw(t)$

3) パラメータを更新する

損失関数の微分値を引く

$w \leftarrow w - d/dw(t) (\times \rho : \text{学習係数})$



(参考)

目的別に最適化されたNeural Network

画像認識 ⇒ Convolutional Neural Network

時系列解析 ⇒ Recurrent Neural Network

1. 人工知能とは何か
2. 人工知能を活用する意義
3. 人工知能の歴史とトレンド
4. ディープラーニングの理論的な背景
- 5. 医療における人工知能の活用領域**
6. 人工知能の限界と適切な活用
7. 人工知能を活用した製品の調達について

医療においても様々な領域で人工知能の活用が期待されている

人工知能（AI）の実用化における機能領域

識別	音声認識	予測	数値予測	実効	表現生成
	画像認識		マッチング		デザイン
	動画認識		意図予測		行動最適化
	言語解析		ニーズ予測		作業の自動化

(出典) 総務省「ICTの進化が雇用と働き方に及ぼす影響に関する調査研究」(平成28年)

一つの分類方法として、総務省「ICTの進化が雇用と働き方に及ぼす影響に関する調査研究」において識別、予測、実効に分類している。これらを医療に照らし合わせることで、AIを活用してどのような事が実現可能か考える手がかりになるかもしれない。

医療における人工知能の活用の具体例

既に医療における人工知能の利活用の事例は幅広く存在しており、増え続けているため、すべてを紹介するのは困難となっている。ここでは一部の領域についてピックアップする

- 1) 医療画像処理
- 2) がんのPrecision Medicine
- 3) 自然言語処理等の活用



1) 医療画像処理

画像処理・画像認識は第3次人工知能ブームの火付け役となった領域の一つであり、活用が特に先行している分野である。

画像診断、病理診断や患者の外観からの疾病予測等取り組まれている。

TED Stanford大学Fei Fei Li氏のプレゼンテーション
https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures?language=ja

人工知能による脳動脈瘤等の診断支援

富士フィルム 東京大学発ベンチャー「エルピクセル」と提携
－ AI技術を医療画像診断支援に活用 －

2018年4月12日
富士フィルム株式会社

富士フィルム株式会社(社長:助野 健児)は、医療領域の画像解析に強みを持つエルピクセル株式会社(以下:エルピクセル)と医療画像の診断支援に用いられるAI技術に関するパートナーシップについて2018年4月10日に合意しました。今後、エルピクセルが開発する診断支援AI技術を、富士フィルムの医用画像情報システム(以下:PACS)に搭載していくことを目指します。

人工知能を活用した医療画像診断支援技術「EIRL (エイル)」を
発表

EIRL | 人工知能(AI) | 医療画像診断支援

2017年11月24日



人工知能を活用した術中迅速病理診断支援

2018年1月31日

東芝デジタルソリューションズと千葉大学フロンティア医工学センター、AIによる胃がんのリンパ節転移巣検出の共同研究を開始

東芝デジタルソリューションズ株式会社

☞ マークの付いたリンクは、別ウィンドウで開きます。

東芝デジタルソリューションズ株式会社(神奈川県川崎市 取締役社長 錦織 弘信 以下、当社)は、国立大学法人千葉大学(千葉県千葉市以下、千葉大学)フロンティア医工学センター 林秀樹教授、医学研究院先端応用外科学 松原久裕教授、医学部診断病理学教室 松嶋博助教らの研究グループと、AI^{注1}(人工知能)による病理組織画像からの胃がんのリンパ節転移巣検出の共同研究を開始しました。

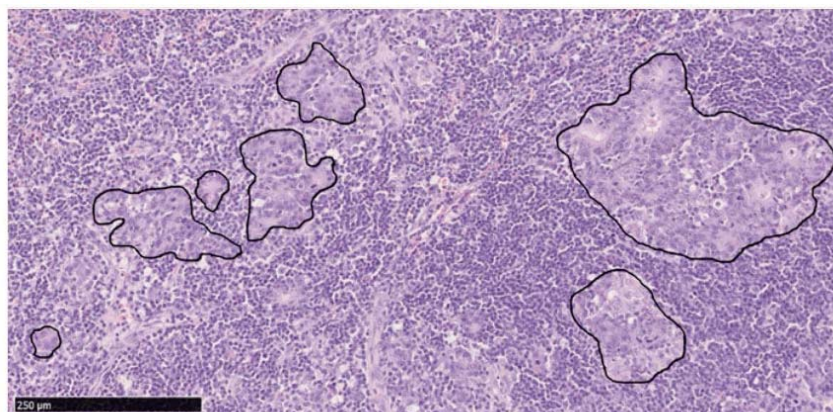
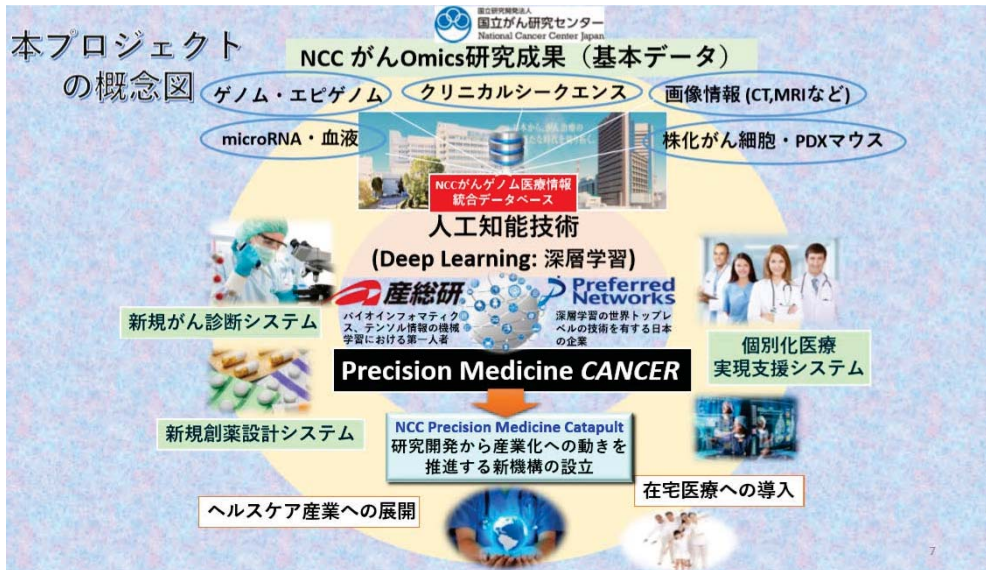


図1 胃がんのリンパ節転移(リンパ節のHE染色組織像:実線で囲まれた部分が転移組織)

2) がんのPrecision Medicine



国立がん研究センター 人工知能を用いた統合的ながん医療システムの開発

https://www.ncc.go.jp/jp/information/pr_release/2016/1129/press_release_20161129_01.pdf

3) 自然言語処理等の活用

総務省 次世代人工知能社会実装WG 自然言語処理の医療応用

http://www.soumu.go.jp/main_content/000474414.pdf



迅速・網羅的病原体ゲノム解析法の開発及び感染症危機管理体制の構築に資する研究



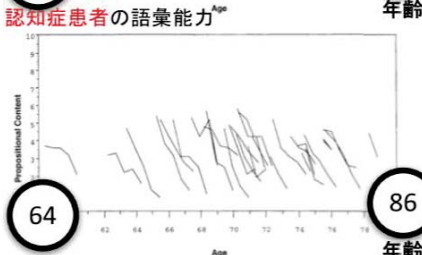
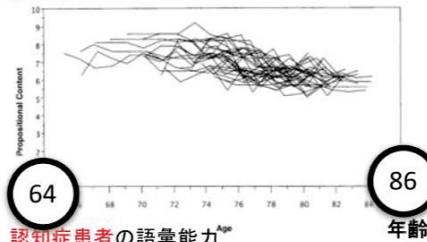
感染症危機管理体制の構築に資する網羅的な診断系を確立
Twitterの“つぶやき”から自然言語処理にて感染症発生を推定するツールを開発し、そのアラートが実際の現場（病院・地研・感染研）で正しい事象を捕捉しているのか否かを検証する。
海外来訪者の増大と東京オリンピック対策を兼ねて、感染研・東京都健康安全研究所との連携を密にし、実地疫学とNGS検査法を軸にネットワーク構築する。

- 黒田誠 (国立感染症研究所)
- 木村博一 (国立感染症研究所)
- 梁明秀 (横浜市立大学大学院)
- 大石和徳 (国立感染症研究所)
- 四宮博人 (愛媛県環境衛生研究所)
- 貞升健志 (東京都健康安全研究所)
- 大場邦弘 (公立昭和病院)
- 荒牧英治 (奈良先端技術大学院大学)

+ BRAIN

Snowdonや Kemperは英語言語にて認知症者の語彙力が発症の30年前からも予測できる可能性を示唆

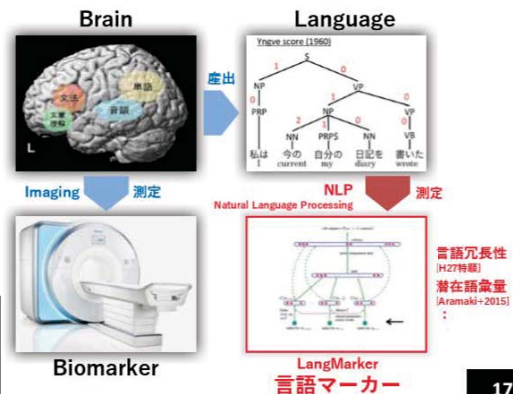
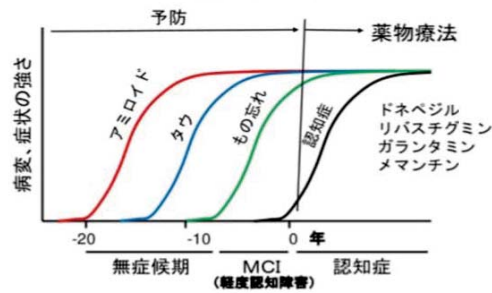
健常者の語彙能力



Snowdon, D. A., Kemper, S., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *Journal of the American Medical Association*, 275, 528-532.

個人発話データの利活用

病理的妥当性



1. 人工知能とは何か
2. 人工知能を活用する意義
3. 人工知能の歴史とトレンド
4. ディープラーニングの理論的な背景
5. 医療における人工知能の活用領域
- 6. 人工知能の限界と適切な活用**
7. 人工知能を活用した製品の調達について

人工知能の限界

哲学的・概念的な問題から、技術的・実践倫理的・法的課題まで

- 1) チューリングテスト
- 2) 不完全性定理
- 3) 中国語の部屋
- 4) フレーム問題
- 5) 記号接地問題
- 6) トロッコ問題
- 7) 瑕疵があった場合の責任の所在

1) チューリングテスト

チューリングテストという、機械に応答させて人間が応答しているか機械が応答しているか、判断するテストがある。30%の審査員に人だと判断させれば合格で、2014年に33%にて達成はされたが、未だ十分に人間らしい反応であるという成果を挙げてはいない。

2) 不完全性定理

「自己言及のパラドックス」いわゆる「クレタ人は嘘つきだとクレタ人は言った」を、ゲーゲルが数学的に証明した。完全な人工知能はその完全性を証明できないという矛盾が生じる。外部から条件・定義を設定する必要が生じる。

3) 中国語の部屋

中国語を知っているヒトがコンピュータに対して中国語を入力、別の部屋で中国語を知らないヒトが中国語を記号として認識して、ある記号に対してはこの記号、というように返答すると、見かけ上中国語を理解しているヒトが対応している様に見える。単に記号を操作するだけで知能をシミュレートしており、知能はヒトにも機械にもでっつけられるというもの。

4) フレーム問題

現在、人工知能は思考すべき範囲や詳細度の決定を人工知能自身でできない。ある目的を達成使用とする場合、関係ある事項のほか、関係無い事項は無限に存在するため、それを洗い出そうとすると無限に時間を要し停止してしまう。

現在のところ、人工知能に対しヒトが「適切」な範囲を定める必要があり、「範囲を定める問題」の一般的な解決は困難である。

5) 記号接地問題

記号を実世界の意味と結びつけられるかという問題。コンピュータは記号の意味を理解していないので、記号の操作だけで知能を実現できない。例えば、シマ+ウマ=シマウマという様な概念を一般化して獲得する事はできない。

6) トロツコ問題

トロツコが線路の上を走行しており、岐路が設けられている。線路の先には修理作業中の5名の作業員がおり、そのまま走ると5名が轢かれる。別の岐路の先には1人が作業している。A氏はたまたまトロツコの進行方向を変えるレバーの近くにいる。A氏がとるべき「正しい」行動とはどのようなものか。という問題。人工知能に「正しい」判断をさせる場合、何を持って「正しい」とするか。人工知能にどのような価値判断を与えるべきか、どのようにコンセンサスを得るのか、コンセンサスは得られるのか。

7) 瑕疵があった場合の責任の所在

人工知能を搭載した製品を使用して起きた問題の中で、人工知能の判断が誤っていたために起きた問題の責任は、製品製造者にあるか、製品使用者にあるか。

人工知能が出した解答に従って起きた問題の所在は、どのような場合にどこに置くべきか。

官民 ITS 構想・ロードマップ 2017

ITS (Intelligent Transport Systems : 高度道路交通システム)

【表 1】自動運転レベルの定義 (J3016) の概要⁴

レベル	概要	安全運転に係る監視、対応主体
運転者が全てあるいは一部の運転タスクを実施		
SAE レベル 0 運転自動化なし	・ 運転者が全ての運転タスクを実施	運転者
SAE レベル 1 運転支援	・ システムが前後・左右のいずれかの車両制御に係る運転タスクのサブタスクを実施	運転者
SAE レベル 2 部分運転自動化	・ システムが前後・左右の両方の車両制御に係る運転タスクのサブタスクを実施	運転者
自動運転システムが全ての運転タスクを実施		
SAE レベル 3 条件付運転自動化	・ システムが全ての運転タスクを実施 (限定領域内 [※]) ・ 作動継続が困難な場合の運転者は、システムの介入要求等に対して、適切にตอบสนองすることが期待される	システム (作動継続が困難な場合は運転者)
SAE レベル 4 高度運転自動化	・ システムが全ての運転タスクを実施 (限定領域内 [※]) ・ 作動継続が困難な場合、利用者が応答することは期待されない	システム
SAE レベル 5 完全運転自動化	・ システムが全ての運転タスクを実施 (限定領域内 [※] ではない) ・ 作動継続が困難な場合、利用者 ⁵ が応答することは期待されない	システム

自動運転の分野ではAI活用により生じた責任の所在も議論されている

<高度自動運転の制度整備に係る基本的考え方>

高度自動運転の実現のための制度整備の方針 (大綱) の検討にあたっては、我が国としては、以下の基本的な考え方 (戦略) に基づいて検討を行うものとする。

- i. 中期的視点に立った制度面における国際的リーダーシップの発揮
- ii. 安全性を確保しつつイノベーションが促進されるような制度枠組みの策定
- iii. 社会受容性を前提とつつイノベーションが促進されるような責任関係の明確化

人工知能の適切な活用

Asilomar AI Principles (2017)

2017年1月、カリフォルニア州アシロマにて行われた、全世界のAIの研究者と経済学、法律、倫理、哲学の専門家が集まり議論された人類にとって有益なAIに関する原則。研究に関する5項目、倫理と価値観に関する13項目、長期的な問題に関する5項目計23項目が挙げられた。

AIネットワーク社会推進会議報告書2017

総務省情報通信政策研究所が、平成28年10月から「AIネットワーク社会推進会議」を開催し、AIネットワーク化をめぐる社会的・経済的・倫理的・法的な課題について検討を進めている。平成29年7月に報告書が発出され、その中で、国際的な議論のためのAI開発ガイドライン案が提案された

Asilomar AI Principles (2017)

- Research Issues
- 1) **Research Goal:** The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.
- 2) **Research Funding:** Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies, such as:
 - How can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked?
 - How can we grow our prosperity through automation while maintaining people's resources and purpose?
 - How can we update our legal systems to be more fair and efficient, to keep pace with AI, and to manage the risks associated with AI?
 - What set of values should AI be aligned with, and what legal and ethical status should it have?
- 3) **Science-Policy Link:** There should be constructive and healthy exchange between AI researchers and policy-makers.
- 4) **Research Culture:** A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.
- 5) **Race Avoidance:** Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.

- Ethics and Values
- 6) **Safety:** AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.
- 7) **Failure Transparency:** If an AI system causes harm, it should be possible to ascertain why.
- 8) **Judicial Transparency:** Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.
- 9) **Responsibility:** Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.
- 10) **Value Alignment:** Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.
- 11) **Human Values:** AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.
- 12) **Personal Privacy:** People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.

- 13) **Liberty and Privacy:** The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.
- 14) **Shared Benefit:** AI technologies should benefit and empower as many people as possible.
- 15) **Shared Prosperity:** The economic prosperity created by AI should be shared broadly, to benefit all of humanity.
- 16) **Human Control:** Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.
- 17) **Non-subversion:** The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.
- 18) **AI Arms Race:** An arms race in lethal autonomous weapons should be avoided.

- Longer-term Issues
- 19) **Capability Caution:** There being no consensus, we should avoid strong assumptions regarding upper limits on future AI capabilities.
- 20) **Importance:** Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.
- 21) **Risks:** Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.
- 22) **Recursive Self-Improvement:** AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures.
- 23) **Common Good:** Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization.

第2章 国際的な議論のためのAI開発ガイドライン案①

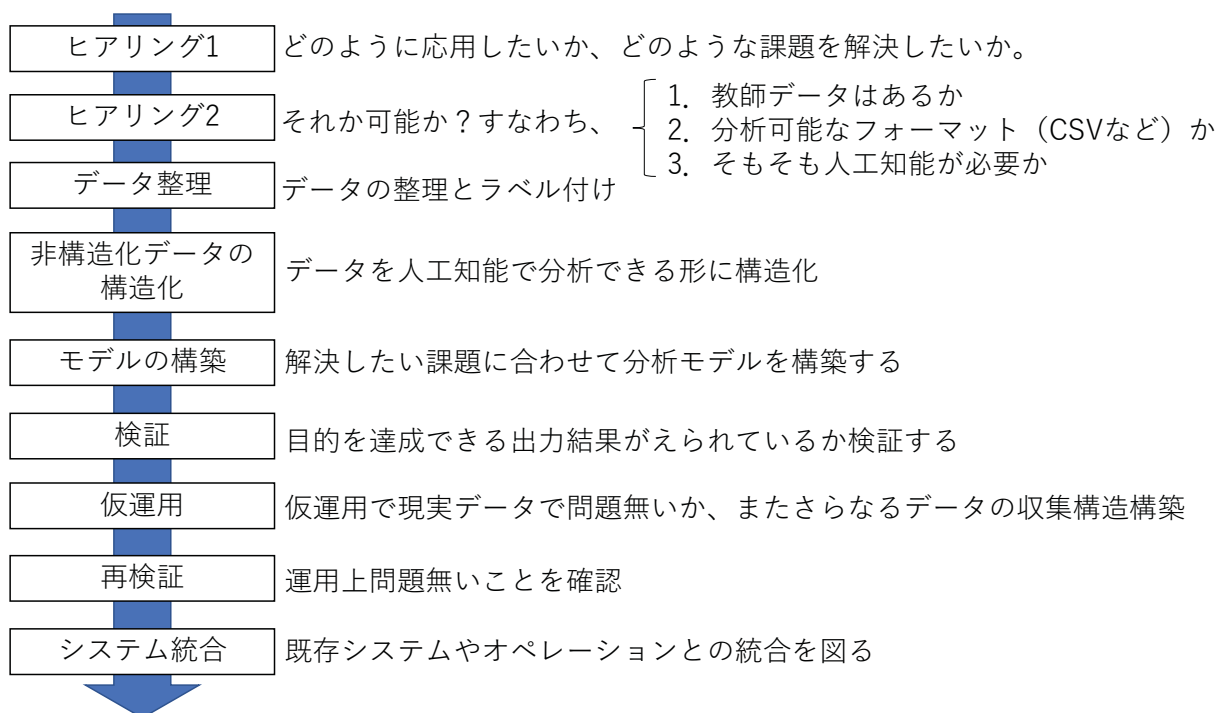
目的	基本理念	用語の定義及び対象範囲
<ul style="list-style-type: none"> AIネットワーク化(AIシステムがインターネット等と接続され、AIシステム相互間又はAIシステムと他の種類のシステムとの間のネットワークが形成されるようになること)の進展により、人間及びその社会や経済に多大な便益がもたらされる一方、不透明化等のリスクに対する懸念も存在 オープンな議論を通じ、国際的なコンセンサスを醸成し、非規制的で非拘束的なソフトローたるガイドラインやそのベストプラクティスをステークホルダ間で国際的に共有することが必要 AIネットワーク化の健全な進展を通じたAIシステムの便益の増進とリスクの抑制により、利用者の利益を保護するとともにリスクの波及を抑制し、人間中心の智連社会を実現 	<ul style="list-style-type: none"> 人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現 非拘束的なソフトローたる指針やそのベストプラクティスをステークホルダ間で国際的に共有 イノベティブでオープンな研究開発と公正な競争、学問の自由や表現の自由といった民主主義社会の価値を尊重するとともに、便益とリスクの適正なバランスを確保 技術的中立性の確保、開発者に過度の負担を課さないよう配慮 不断の見直し、必要に応じた柔軟な改定、広範で柔軟な議論 	<ul style="list-style-type: none"> 「AI」: AIソフト及びAIシステムを総称する概念 <ul style="list-style-type: none"> ●「AIソフト」: データ・情報・知識の学習等により、利活用の過程を通じて自らの出力やプログラムを変化させる機能を有するソフトウェア(例: 機械学習ソフトウェア) ●「AIシステム」: AIソフトを構成要素として含むシステム(例: AIソフトを実装したロボットやクラウドシステム) AIシステムの「開発者」及び「利用者」は以下のとおり定義 ※ただし、「開発者」及び「利用者」は場面に応じて個別に決まる相対的な概念 <ul style="list-style-type: none"> ●「開発者」: AIシステムの研究開発(AIシステムを利用しながら行う研究開発を含む)を行う者(自らが開発したAIシステムを用いてAIネットワークサービスを他者に提供するプロバイダを含む) ●「利用者」: AIシステムを利用する者(最終利用者(エンドユーザ)のほか、他者が開発したAIネットワークサービスを第三者に提供するプロバイダを含む) 対象とするAIシステムの範囲→ネットワーク化され得るAIシステム(ネットワークに接続可能なAIシステム) 対象とする開発者の範囲→定義された開発者すべて 対象とする開発の範囲→閉鎖された空間(実験室、セキュリティが十分に確保されたサンドボックス等)内での開発は対象とせず、ネットワークに接続して行う段階に限定
関係者に期待される役割	<ul style="list-style-type: none"> 各国政府及び国際機関による多様なステークホルダ間の対話の促進に向けた環境整備 開発者、利用者等ステークホルダによる対話やベストプラクティスの共有、AIの便益及びリスクに関する認識の共有 標準化団体等による推奨モデルの作成・公表 各国政府によるAIの開発者コミュニティの支援、AIに関する研究開発を支援する政策の積極的な推進 	

第2章 国際的な議論のためのAI開発ガイドライン案② (AI開発原則案の解説)

開発原則	開発原則の解説 (留意することが期待される事項等)
① 連携の原則 開発者は、AIシステムの相互接続性と相互運用性に留意する。	・相互接続性と相互運用性を確保するため、①有効な関連情報の共有に向けた協力、②国際的な標準や規格への準拠、③データ形式の標準化及びインターフェイスやプロトコルのオープン化への対応、④標準必須特許等のライセンス契約及びその条件についてのオープン・公平な取扱い、などに留意することが望ましい。
② 透明性の原則 開発者は、AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する。	・採用する技術の特性や用途に照らし合理的な範囲で、AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意することが望ましい。 (※アルゴリズム、ソースコード、学習データの開示を想定するものではない。)
③ 制御可能性の原則 開発者は、AIシステムの制御可能性に留意する。	・AIシステムの制御可能性について、あらかじめ検証及び妥当性の確認をするよう努めるとともに、採用する技術の特性に照らし合理的な可能な範囲において、人間や信頼できる他のAIによる監督・対処の実効性に留意することが望ましい。
④ 安全の原則 開発者は、AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。	・AIシステムの安全性について、あらかじめ検証及び妥当性の確認をするよう努めるとともに、AIシステムの本質安全や機能安全に資するよう、開発の過程を通じて、採用する技術との特性に照らし可能な範囲で措置を講ずるよう努めることが望ましい。 ・利用者及び第三者の生命・身体・財産の安全に関する判断を行うAIシステムについては、利用者等ステークホルダに対して設計の趣旨などを説明するよう努めることが望ましい。
⑤ セキュリティの原則 開発者は、AIシステムのセキュリティに留意する。	・AIシステムのセキュリティについて、あらかじめ検証及び妥当性の確認をするよう努めるとともに、開発の過程を通じて、採用する技術の特性に照らし可能な範囲で対策を講ずるよう努めることが望ましい(セキュリティ・バイ・デザイン)。
⑥ プライバシーの原則 開発者は、AIシステムにより利用者及び第三者のプライバシーが侵害されないよう配慮する。	・AIシステムのプライバシー侵害のリスクについて、あらかじめ影響評価を行うよう努めるとともに、開発の過程を通じて、採用する技術の特性に照らし可能な範囲で措置を講ずるよう努めることが望ましい(プライバシー・バイ・デザイン)。
⑦ 倫理の原則 開発者は、AIシステムの開発において、人間の尊厳と個人の自律を尊重する。	・AIシステムの学習データに含まれる偏見などに起因して不当な差別が生じないよう、採用する技術の特性に照らし可能な範囲で所要の措置を講ずるよう努めることが望ましい。 ・国際人権法や国際人道法を踏まえ、AIシステムが人間性の価値を不当に毀損することがないよう留意することが望ましい。
⑧ 利用者支援の原則 開発者は、AIシステムが利用者を支援し、利用者を選択の機会を適切に提供することが可能となるよう配慮する。	・AIシステムの利用者のために、①利用者の判断に資する情報を適時適切に提供し、かつ利用者にとって操作しやすいインターフェイスが利用可能となるよう配慮、②利用者を選択の機会を適時適切に提供する機能が利用可能となるよう配慮、③社会的弱者の利用を容易にするための取組、に留意することが望ましい。
⑨ アカウンタビリティの原則 開発者は、利用者を含むステークホルダに対しアカウンタビリティを果たすよう努める。	・開発原則①～⑧の趣旨に鑑み、利用者等に対しAIシステムの技術的的特性について情報提供や説明を行うほか、ステークホルダとの対話を通じた意見聴取や、ステークホルダの積極的な関与を得るよう努めることが望ましい。 ・AIシステムによりサービスを提供するプロバイダなどの情報共有・協力に努めることが望ましい。

1. 人工知能とは何か
2. 人工知能を活用する意義
3. 人工知能の歴史とトレンド
4. ディープラーニングの理論的な背景
5. 医療における人工知能の活用領域
6. 人工知能の限界と適切な活用
7. 人工知能を活用した製品の調達について

機械学習を含んだ人工知能製品の開発運用フロー例



適切なシステム開発のためには、現場の問題意識と、人工知能開発研究者と、ユーザーインターフェース開発ベンダーとの間で、密なコミュニケーションが必要となる。それを理解した上で、対応適任者を専任する必要が出てくる。また、精度向上のためには、利用環境で生まれるデータを学習させ続ける仕組み作りも重要である。

日本はイノベーション能力が低いと評価されている

第Ⅱ-3-2-1-1表 WEF イノベーションランキング 2016-2017年版

	イノベーション ランキング	イノベーション 能力	科学技術 調査機関の質	企業の研究 開発投資	研究開発にお ける産学協業	先進技術に対 する政府調達	科学者・技術者 の対応領域と数	PCT 国際出願 件数
スイス	1	1	1	1	1	28	14	3
イスラエル	2	4	3	3	3	9	8	5
フィンランド	3	6	8	7	2	26	1	4
米国	4	2	5	2	4	11	2	10
ドイツ	5	5	11	5	8	6	16	7
スウェーデン	6	3	7	6	12	23	20	2
オランダ	7	10	4	14	5	21	21	9
日本	8	21	13	4	18	16	3	1
シンガポール	9	20	10	15	7	4	9	13
デンマーク	10	18	16	16	14	53	37	8

備考：PCT 出願とは、特許協力条約に基づいた特許申請。

資料：WEF The Global Competitiveness Report (2016-2017年版) から経済産業省作成。

平成29年版通商白書 経済産業省

http://www.meti.go.jp/report/tsuhaku2017/pdf/2017_00-all.pdf

イノベーションに向けて行動して行くことが重要