

第3章

デジタル情報を扱う

学習指導要領 2021年度・高校入学までの方、
「社会と情報」のテキストも参照してください。

学習指導要領 2022年度・高校入学以降の方は、
「情報 I」のテキストを参照してください。

2024/05/31

新潟大学大学歯学部1年生
スタディスキルズ講義資料
西山秀昌

項目

- 情報のデジタル化について
(復習と概念の再構成)
- デジタル情報を処理するコンピュータについて
- 情報インフラを支えるネットワークについて
- 暗号と認証について
- AIについて
- Society 5.0について

情報のデジタル化について (復習と概念の再構成)

高校テキスト「社会と情報」※での情報の定義の混乱

もし、物事に関する知らせが、判断や振る舞いの基準を与えないものであったとすれば、情報とはいえない。

1 情報とは何か

情報とは、物事に関する知らせであり、それを受け取る主体に判断や振る舞いの基準を与えるものである。この主体となるものはさまざまである。もちろん、人間は他者やメディアなどから情報を受け取る。また、人間以外の生物、コンピュータなどの非生物、更に生物を構成する個々の細胞なども情報を受け取り、それに基づく判断や振る舞いを示す。

もし、物事に関する知らせが、判断や振る舞いの基準を与えないものであったとすれば、情報とはいえない。

本当にそうなのか？
「誰が」そのように判断するのか？
主観的な解釈問題が内在するため、客観的な定義とはならない。

数学的な情報量（エントロピー）としての客観的な「起こりやすさ」と主観的な意味・価値が混同されている。



主体にとって何らかの意味や価値を持つものが情報であり、当然のことを伝えても情報にはならない。

「当然のこと」という意味を持ち、そのような判断を行わせる情報になっている

2 情報が与える影響

主体にとって何らかの意味や価値を持つものが情報であり、当然のことを伝えても情報にはならない。

高校テキスト「社会と情報」※での問題

※既にも出版社に問い合わせし、「誤解を与えかねない表現」との回答を得ている。

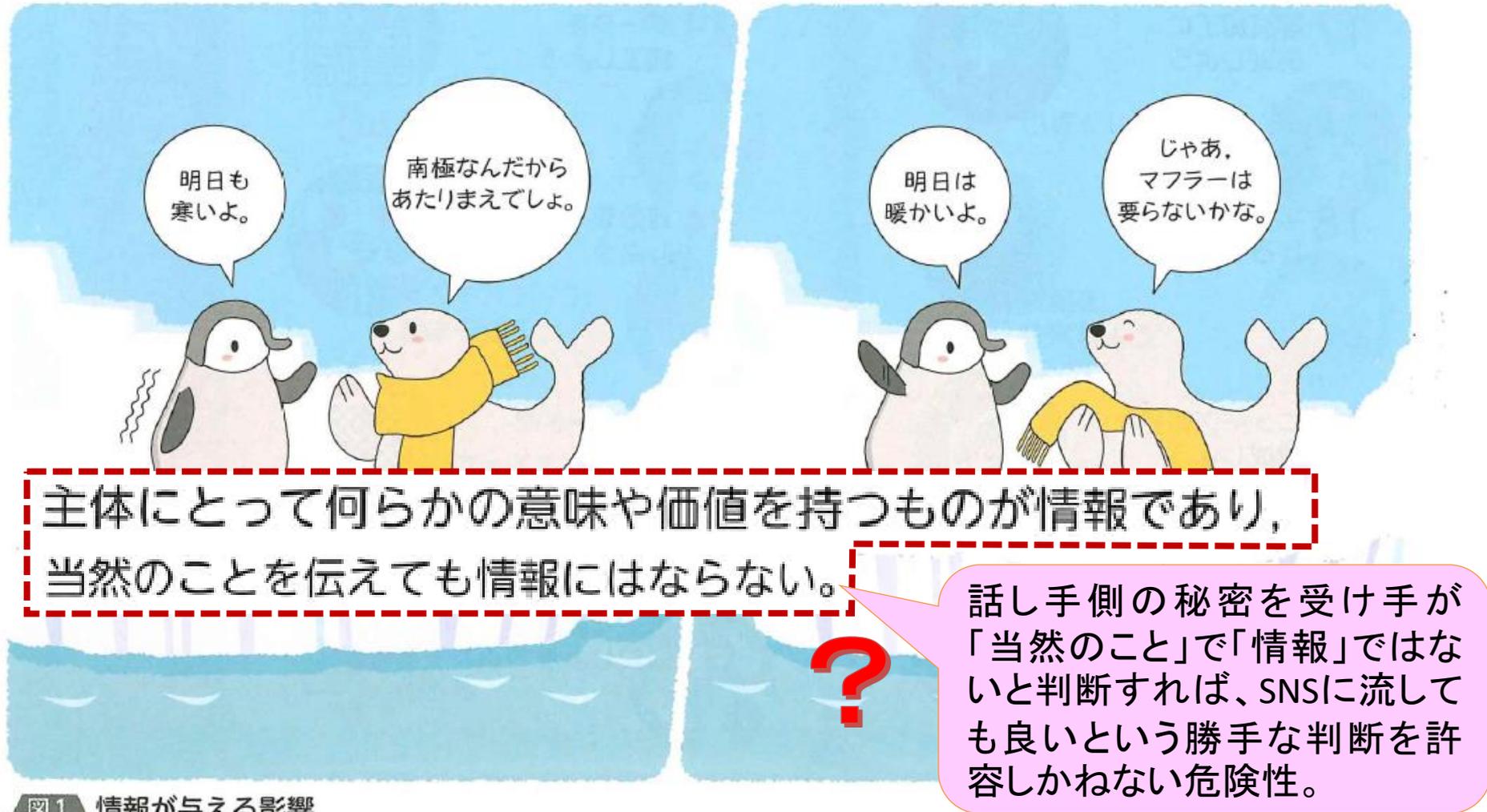


図1 情報が与える影響

主体にとって何らかの意味や価値を持つものが情報であり、当然のことを伝えても情報にはならない。

「情報にならない」とは？



「南極だからあたりまえでしょ。」とアザラシが発言している状況は、本当に「情報にならない」状況なのだろうか？

アザラシにとって「情報にならない」とは、一体どういった状況なのだろうか？

「情報」にならないとは??

1 情報とは何か

情報とは、物事に関する知らせであり、それを受け取る主体に判断や振る舞いの基準を与えるものである。この主体となるものはさまざまである。もちろん、人間は他者やメディアなどから情報を受け取る。また、人間以外の生物、コンピュータなどの非生物、更に生物を構成する個々の細胞なども情報を受け取り、それに基づく判断や振る舞いを示す。

もし、物事に関する知らせが、判断や振る舞いの基準を与えないものであったとすれば、情報とはいえない。

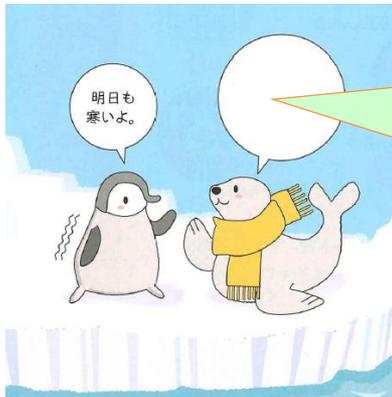
赤堀侃司, 永野和男, 坂元章 他 (2017)『社会と情報』東京書籍

「南極なんだからあたりまえでしょ」という反応を引き起こしている時点で、既に「情報」の定義を満たしている。

同じページに

人間以外の生物，コンピュータなどの非生物，更に生物を構成する個々の細胞なども情報を受け取り，それに基づく判断や振る舞いを示す。

と、明記されている。



少なくとも脳内で何らかの反応が生じていたなら、「情報」になっているはず。

アザラシがヘッドフォンしていて目を閉じて音楽に聴き入っている状態で、ペンギンが話しかけても「聞こえていない」状況なら「情報」になっていないと言える。

「当然・あたりまえ」は本当に「情報にならない」のか？



アザラシA:「**そうだね。このマフラーは外せそうにないよね。**」

← **判断や振る舞いの基準を与えている**

アザラシB:「**ペンギン君にとっては寒そうだよね。こんど、マフラーを編んであげるよ。**」

← **判断や振る舞いの基準を与えている**

同様に…

アザラシC:「**南極なんだからあたりまえでしょ。**」という反応を引き出すことに成功している。

← **判断や振る舞いの基準を与えている**

情報学 (informatics) での定義

- <https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-h160323-2.pdf>

「大学教育の分野別質保証のための教育課程編成上の参照基準」

情報学分野 平成28年(2016年)3月23日

日本学術会議 情報学委員会 情報科学技術教育分科会 資料

生命情報、社会情報、機械情報

生命の側からみた
トップダウン的な解釈

人間社会や機械に加えて、広く生物をも情報を扱う主体だと考えると、生物が生存するための選択行動が情報の意味作用の源泉だということがわかる。生物の生存のための情報 **(生命情報)** が最も根源的な、広義の情報である。これを記号で表すと人間社会で通用する狭義の情報 **(社会情報)** となる。社会情報は記号と意味内容のセットである。コンピュータで機械的に処理される情報 **(機械情報)** は、基本的に社会情報から派生し、記号が独立して意味内容が潜在化したものであり、最狭義の情報として位置づけられる。このように情報を扱う主体により情報を分類すれば、各種の情報の関係が明らかになり、記号の意味解釈とコミュニケーションの態様が明確になる。

基礎情報学での定義

生命情報

生きるために必要となる情報、意識下にはないものを含む

社会情報

他者と共有可能な情報、言語・言葉など

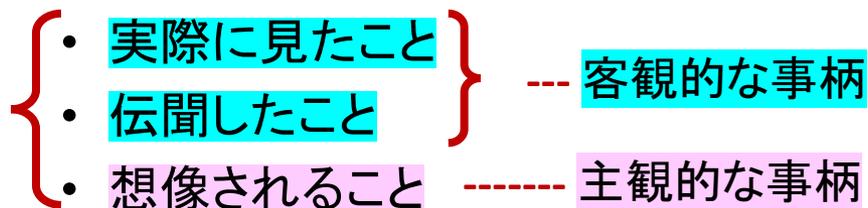
機械情報

意味が隠され、機械レベルで扱われる記号

小野厚夫(1991)

明治期における「情報」と「状報」

- 「野外演習軌典」(明治15年3月20日)等のまとめ
 - 間諜がもたらす諜報とは異なる、一般兵士からの報告
 - 上からの命令に対する下からの情報
 - 主観が入りやすいため次の3つの雛形に分ける。



- 「状報」は「情報」とともに明治15年から20年頃まで用いられていたが、その後「情報」に統一された。
- (当時の)兵書における明治期の用例からは「ありのままのようす、ありさまのしらせ」を意味することが多く、現在の「情」から受ける語感からすれば、「情報」よりも「状報」の方が適切だったように思われる。

森鷗外が使い分けたであろう 「情報」と「状報」

- 状報: (主観を含まない客観的な)「所変」(客体印象識)
- 情報: (主観に基づく)推測を含む「能変」(主体思量識)



創立45周年記念特別寄稿

情報

という

言葉を探ねて

(2)

小野厚夫

大手前大学社会文化学部キャリアデザイン学科
ono@omae.ac.jp

前所述べたように、情報は日本で作られた言葉で、1876年出版の酒井忠恕訳『佛蘭歩兵陣中要務實地演習教典』に最初の用例があり、その原語はフランス語の *renseignement* である。初期には情報と状報が併用されていたが、ほどなく情報に統一された。日清、日露戦争の記事で情報は新聞用語として定着し、国語辞書にも採録されるようになった。今回は、主に明治、大正期の情報の用例について述べる。

鷗外と「情報」

森林太郎 (鷗外) は、クラウゼヴィッツ (Clausewitz) の『Über den Kriege』の巻1-2を訳出している。巻1の6章「戦の情報」の冒頭に

情報とは、敵と敵国とに関する我智識の全体を謂ふ、

という定義がある。このように Nachrichten はおおむね「情報」と訳されているが、その前の3章「軍事上の天才」には次に示すように1カ所だけ「状報」と訳した部分がある。

戦は偶然の境界なり。人の事業中常に偶然と相触るゝ者戦に如くは無し。故に偶然の為に多く余地を存せざる可からざることも亦戦に如くは無し。偶然は諸状況をして其不确实の度を加へしめ、又事業の進捗を阻碍する者なり。

彼の諸状報及び諸予想の不确实と此の偶然の類りに来ることとは、戦者をして常に其遇ふ所の其の期する所に異なるを感ぜしむ。

情報処理、46巻5号、p.475-79

28

情報処理学会第40回(平成2年前期)全国大会

鷗外森林太郎による獨逸語 NACHRICHTEN の
二つの翻譯語「情報」と「状報」

3K-1

大島 進
大島技術士事務所

「所変」「能変」を森林太郎は識って彼「戦論」譯述時の獨逸語対 OBJEKTE と SUBJEKTE の譯語に充てると共に和漢語「報」の義の獨逸語源字 NACHRICHTEN 翻譯に際し敵之情推測の「所変(客体印象識)」である「状報」と推測の「能変(主体思量識)」である「情報」とに「戦者」心中の心作用の主客と轉識を看て譯出分け爲したと解する。(文節乙二F b)

確かに「状報」(客観的)と「予想」(主観的)を区分している。

「データと情報」という 分類・定義と問題点

A) ISO 2382-1 (JIS X 0001) での**定義**

<https://kikakurui.com/x0/X0001-1994-01.html>

B) 高校の「情報I」での**定義**

2022年(予想)の学習指導要領改訂に向けた
試作教科書2012「情報I」**2012.10.1**版

情報処理学会初等中等教育委員会(2012.10.1)

<https://www.ipsj.or.jp/12kyoiku/teigen/2012-10-EText.pdf>

「データ」と「情報」という二分類の曖昧さ
「誰にとって」の情報なのかをぼやかした定義

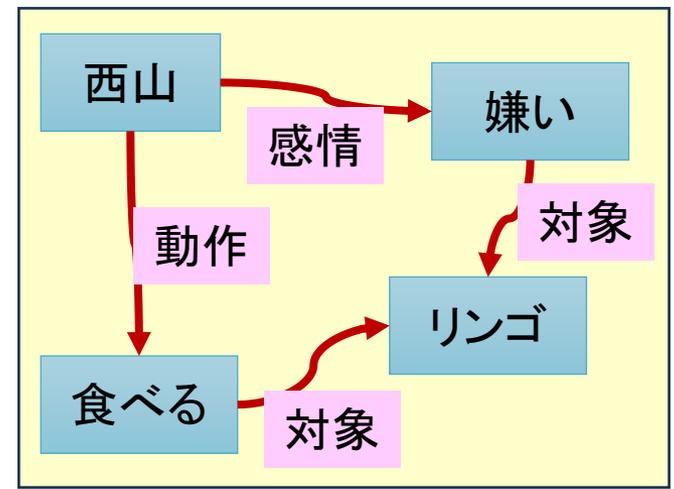
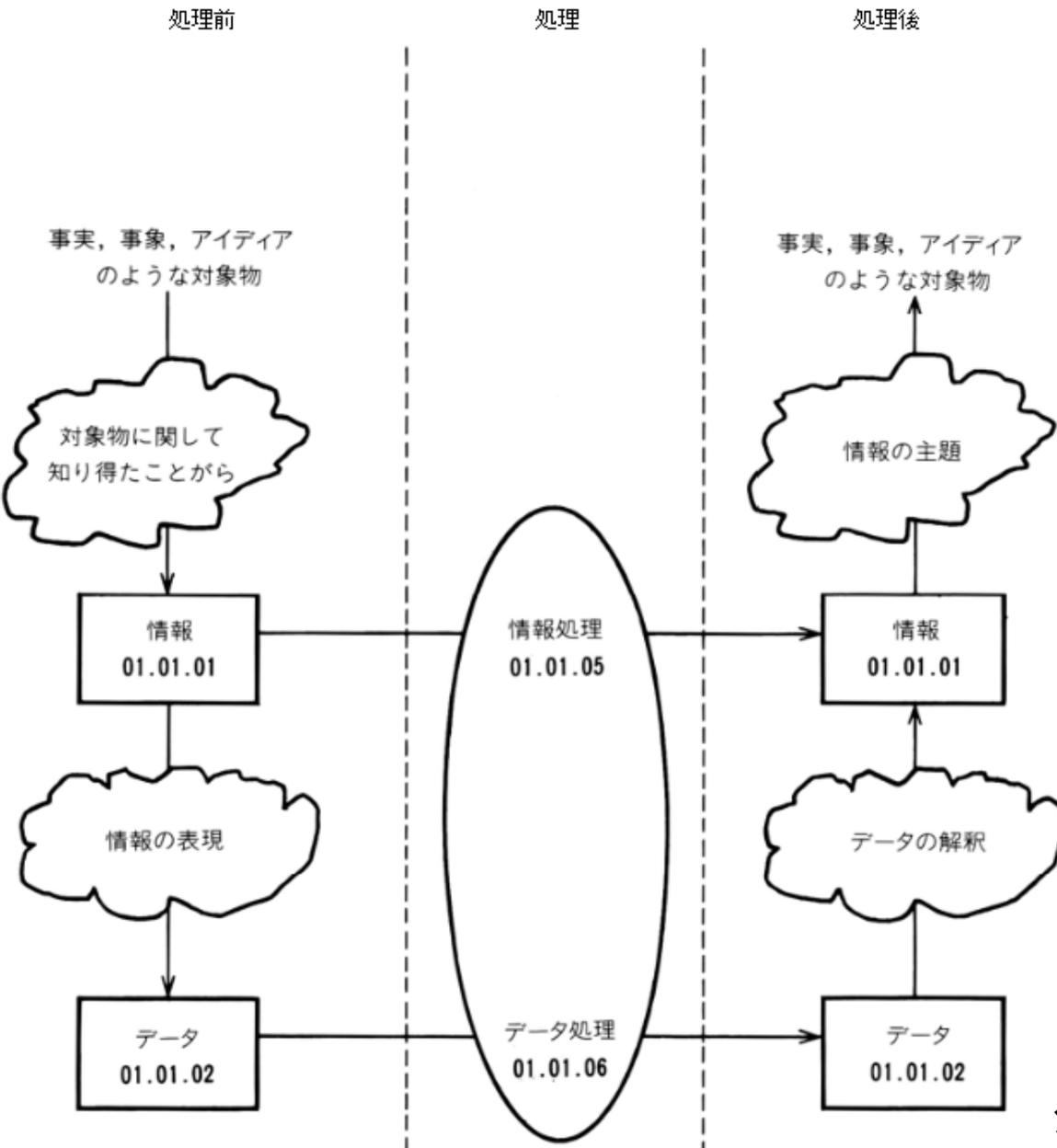
用語：DATAと情報

ISO 2382-1 (JIS X 0001)での定義

<https://kikakurui.com/x0/X0001-1994-01.html>

- **情報**（番号：01.01.01、対応英語：information）
事実，事象，事物，過程，着想などの対象物に関して知り得たことであって，**概念を含み，一定の文脈中で特定の意味をもつもの。**
- **データ**（番号：01.01.02、対応英語：data）
情報の表現であって，**伝達，解釈又は処理に適するよう**
に形式化され，再度情報として解釈できるもの。
備考1. データに対する処理は，人間が行ってもよいし，
自動的手段で行ってもよい。

図1 情報とデータの関係



構造化したデータは情報なのか？
情報ではないのか？→生成AI、
LLMが扱うものはデータ？情報？



分断された単語はデータとみてよいだろうが、固有名詞は情報ではないのか？→個人情報の情報とは？

1.1 情報と情報の役割

2022年(予想)の学習指導要領改訂に向けた
試作教科書2012「情報I」2012.10.1版

情報処理学会初等中等教育委員会(2012.10.1)

<https://www.ipsj.or.jp/12kyoiku/teigen/2012-10-EText.pdf>

1.1.1 情報の定義

現代は情報社会であり、情報に大きな価値が置かれています。では情報 (information) とは何でしょう。情報とは「ものごとの知らせ」であり、受け手に取って価値を持つようなものです。単なる「ものごとの知らせ」はデータ (data) であり、それ自体に価値があるかどうかは分かりませんが、受け手によってはそのデータが情報になるのです。

たとえば、壁に温度計が掛かっていて温度が表示されているとき、それを気に止めないならその温度は単なるデータです。しかし、その温度を見てどれくらい厚着をして行こうか判断するのに用いるとすれば、そのデータは価値があるものとなり、情報だと言えます。

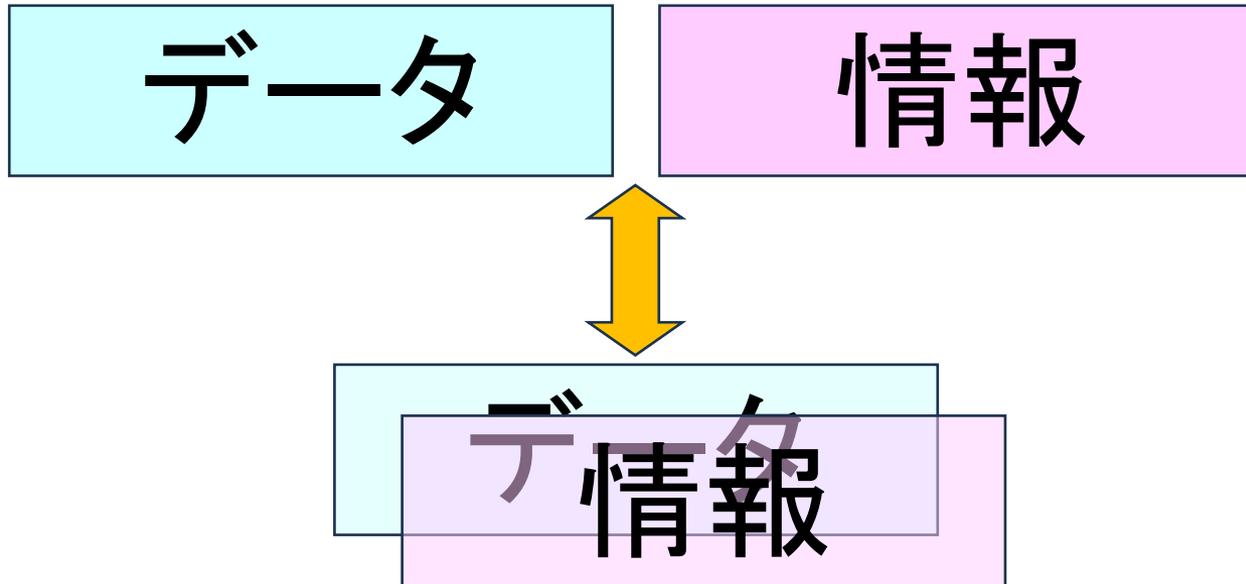
なぜ情報に価値があることになるのでしょうか。適切な上着を選べれば、その日が快適に過ごせ、それは自分にとって喜ばしいことです。喜ばしいことのためには、人は労力や手間を掛けます。このように、自分の行動を選択する際の判断に使うようなデータは、自分にとって価値のある「知らせ」すなわち情報だと言えます。

データと情報の二分類

情報とデータの二分法に 潜む問題

- 「受け手にとって価値を持つ情報」から「価値」を除いたものを全て「データ」としうるのか？という問題がある。
- 言い換えれば、「データ」はコンピュータでの処理を前提とした用語だと思われるが、コンピュータ処理を前提としないあるいは不可能な「情報」ないし「情報になり得るもの」についての概念が欠落している。

データ と 情報



単語として別であるため、内容も異なるイメージだが、「受け手の認識」によって「データ」が「情報」へと読み替えられる。

「個人情報保護」を「情報の保護」と読み違えると、「データ」なら保護対象外と読み違える危険性がある。

※1: 「個人情報保護」は「情報」の保護ではなく、個人データ処理からの「個人」の保護という視点は重要。

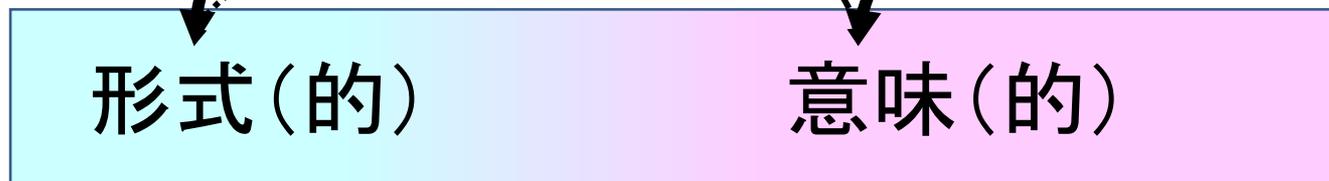
異なる視点・基準からの情報の分類

主観性・客観性
の高さ



扱う上での
意味の影響

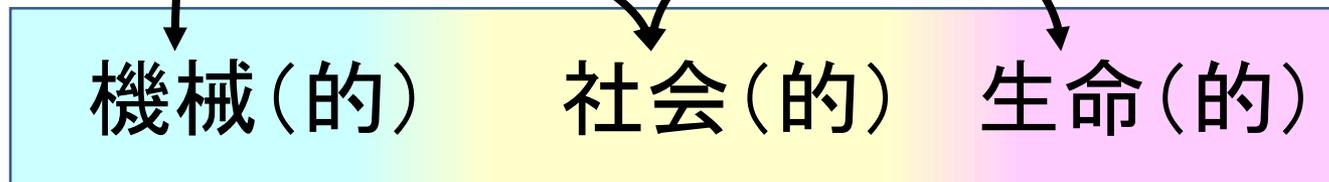
ISO 2382-1
(JIS X 0001)



データ

(狭義の)情報

基礎情報学
での定義

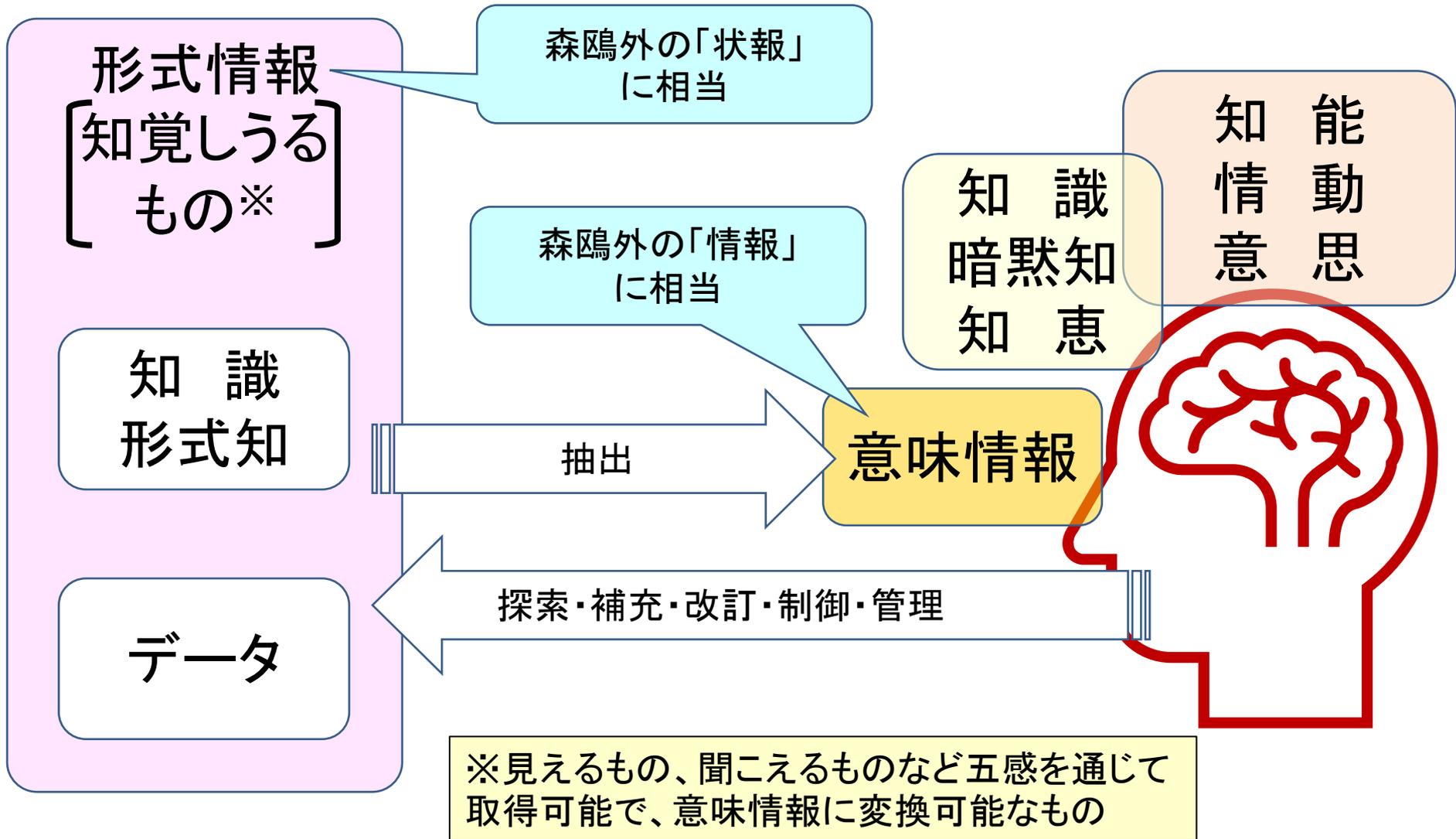


参考資料・情報の意味と形式

- 児玉 徳美 (2003). 「意味と形式」 立命館文學, 第580号
p.17-34
<http://www.ritsumei.ac.jp/acd/cg/lt/rb/580pdf/kodama.pdf>
- 関口恭毅 (2016). 「データ・情報・知識の含意と相互関係の二重性について」 商学論纂(中央大学) 57, 209-247.
<https://irdb.nii.ac.jp/01441/0004054036>
- 日本診療情報学会、「診療情報学 第2版」(医学書院)の「序論」
<https://www.igaku-shoin.co.jp/prd/tachiyomi/02397/html5.html#page=2>
- 中島聡 (2014). 基礎情報学の3つの情報概念を再確認する
<https://www.issj.net/conf/issj2014-papers/pfiles/e21.pdf>
https://www.jstage.jst.go.jp/article/proceedingsissj/10/0/10_e21/_pdf/-char/ja



データ、情報、知識・知恵の関係



データ(形式情報)から 価値としての情報(意味情報)へ



<https://www5.dent.niigata-u.ac.jp/~nisiyama/studyskills/digiInfo/info38.pdf>

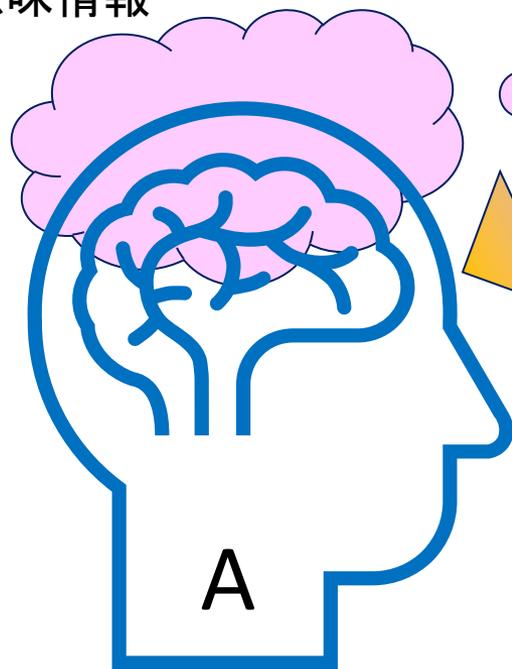


客観性の高い形式情報と主観性の高い意味情報の双方向の伝達・変換

A, B, C: 情報の受信側、発信側
A, B: 個人、C: 集団、国民、人類など

認識主体から切り離され、媒体中に
保管可能、コンピュータで処理可能

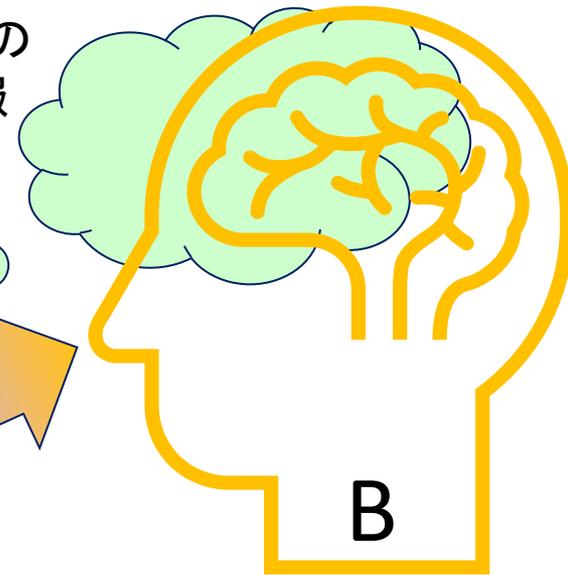
Aにとっての
意味情報



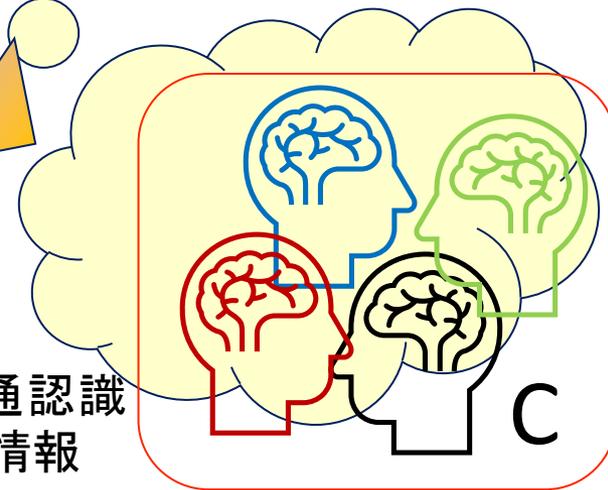
データ
形式情報

情報が追加さ
れる場合あり

Bにとっての
意味情報



C内部で共通認識
可能な意味情報



情報の単位について

形式情報

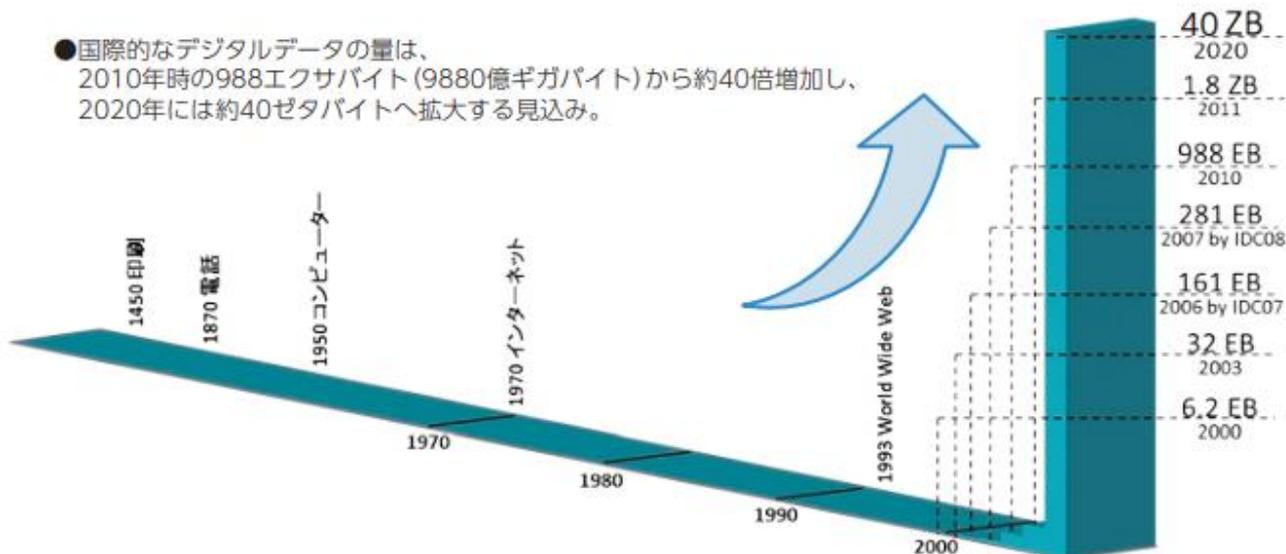
主として数学的、物理学的な側面

1 広がりを見せるビッグデータの活用

(1) ICTの進化が促すビッグデータの生成・流通・蓄積

米国の調査会社IDCによると、国際的なデジタルデータの量は飛躍的に増大しており、2011年（平成23年）の約1.8ゼタバイト（1.8兆ギガバイト）から2020年（平成32年）には約40ゼタバイトに達すると予想されている（図表3-1-1-1）。

図表3-1-1-1 デジタルデータ量の増加予測



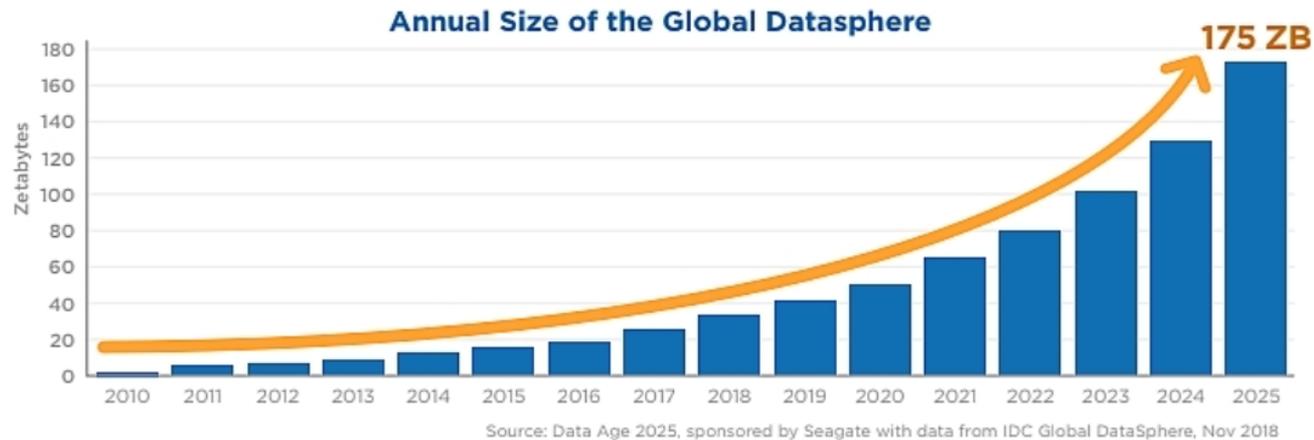
（出典）総務省「ICTコトづくり検討会議」報告書

全情報量の爆発的増大



全体的にみると、2025年に世界の接続機器は1500億台に達し、その多くがリアルタイムデータを生成することになるという。世界のデータ量は、2017年の23ゼタバイトから2025年には175ゼタバイトへと増加する見通しだ。1ゼタバイトは、1兆ギガバイトに相当する。

Figure 1 - Annual Size of the Global Datasphere



<https://japan.zdnet.com/article/35129774/>

Stephanie Condon (ZDNet.com) 翻訳校正: 編集部 2018-12-07 13:04

最新の単位・記号

2022年・第27回国際度量衡総会で決定されたものを含む

— SI接頭語の名称と記号、十進数表記と制定年 —

(赤字はこの度追加されたもの)

名称	記号	指数表記	十進数表記	制定年
クエタ (quetta)	Q	10^{30}	1 000 000 000 000 000 000 000 000 000 000	2022
ロナ (ronna)	R	10^{27}	1 000 000 000 000 000 000 000 000 000	2022
ヨタ (yotta)	Y	10^{24}	1 000 000 000 000 000 000 000 000	1991
ゼタ (zetta)	Z	10^{21}	1 000 000 000 000 000 000 000	1991
エクサ (exa)	E	10^{18}	1 000 000 000 000 000 000	1975
ペタ (peta)	P	10^{15}	1 000 000 000 000 000	1975
テラ (tera)	T	10^{12}	1 000 000 000 000	1960
ギガ (giga)	G	10^9	1 000 000 000	1960
メガ (mega)	M	10^6	1 000 000	1960
キロ (kilo)	k	10^3	1 000	1960
ヘクト (hecto)	h	10^2	100	1960
デカ (deca)	da	10^1	10	1960
デシ (deci)	d	10^{-1}	0.1	1960
センチ (centi)	c	10^{-2}	0.01	1960
ミリ (milli)	m	10^{-3}	0.001	1960
マイクロ (micro)	μ	10^{-6}	0.000 001	1960
ナノ (nano)	n	10^{-9}	0.000 000 001	1960
ピコ (pico)	p	10^{-12}	0.000 000 000 001	1960
フェムト (femto)	f	10^{-15}	0.000 000 000 000 001	1964
アト (atto)	a	10^{-18}	0.000 000 000 000 000 001	1964
zepto	z	10^{-21}	0.000 000 000 000 000 000 001	1991
yocto	y	10^{-24}	0.000 000 000 000 000 000 000 001	1991
ronto	r	10^{-27}	0.000 000 000 000 000 000 000 000 001	2022
quecto	q	10^{-30}	0.000 000 000 000 000 000 000 000 000 001	2022



SI接頭語、追加決定

0.1
0.01
0.001
0.000 001
0.000 000 001
0.000 000 000 001
0.000 000 000 000 001
0.000 000 000 000 000 001
0.000 000 000 000 000 000 001
0.000 000 000 000 000 000 000 001
0.000 000 000 000 000 000 000 000 001



quetta	ronna	ronto	quecto
Q	R	r	q
10^{30}	10^{27}	10^{-27}	10^{-30}

1 000 000 000 000 000 000 000 000 000 000
1 000 000 000 000 000 000 000 000 000 000
1 000 000 000 000 000 000 000 000 000
1 000 000 000 000 000 000 000 000
1 000 000 000 000 000 000 000
1 000 000 000 000 000 000
1 000 000 000 000
1 000 000
100
10

第27回国際度量衡総会にて SI接頭語の追加が決定

k (キロ)、M (メガ)、G (ギガ)、m (ミリ)、 μ (マイクロ)、n (ナノ) など、大きな量あるいは小さな量を端的に記述するために、10のべき乗を表し、SI単位と共に用いられるものをSI接頭語と呼びます。これまでは、 10^{24} から 10^{-24} の範囲で定められていましたが、2022年11月15日から18日にかけてフランスのベルサイユにて開催された第27回国際度量衡総会 (CGPM) にて、 10^{30} 、 10^{27} 、 10^{-27} 、 10^{-30} を表すSI接頭語である、Q (クエタ)、R (ロナ)、r (ロント)、q (クエクト) を新たに追加することが決定されました。昨今のデジタル情報量の急激な増加を背景に、1991年以来、31年ぶりにSI接頭語の範囲が拡張されました。

On the extension of the range of SI prefixes

The General Conference on Weights and Measures (CGPM), at its 27th meeting,

recalling that decisions were made at previous meetings when it was considered timely to extend the range of SI prefixes including Resolution 12 (paragraph 3) adopted by the CGPM at its 11th meeting (1960), Resolution 8 adopted by the CGPM at its 12th meeting (1964), Resolution 10 adopted by the CGPM at its 15th meeting (1975), and Resolution 4 adopted by the CGPM at its 19th meeting (1991),

considering

- the essential role of the International System of Units (SI) in providing confidence in the accuracy and global comparability of measurements needed for international trade, manufacturing, human health and safety, protection of the environment, global climate studies and scientific research,
- the benefits of encouraging the use of SI units by providing new SI prefixes for scientific communities that depend on measurements that are not covered by the current range,
- the needs of data science in the near future to express quantities of digital information expressed using orders of magnitude in excess of 10^{24} ,
- the importance of timely action to prevent unofficial prefix names being *de facto* adopted in other communities,

decides to add to the list of SI prefixes to be used for multiples and submultiples of units the following prefixes:

Multiplying factor	Name	Symbol
10^{27}	ronna	R
10^{-27}	ronto	r
10^{30}	quetta	Q
10^{-30}	quecto	q

<https://www.bipm.org/documents/20126/64811571/Draft-Resolutions-2022.pdf/2e8e53df-7a14-3fc8-8a04-42dd47df1a04?t=1644502962693>

表1. 情報の量と単位

10進法	2進法	16進法	単位・記号		量、比較	
0	0000	0	bit (ビット)		2進法の桁数	$\log_2 N$
1	0001	1	byte, B (バイト)		8 bit	2^8
2	0010	2	KB (キロバイト)		1,024 B	2^{10}
3	0011	3	MB (メガバイト)		1,024 KB	2^{20}
4	0100	4	GB (ギガバイト)		1,024 MB	2^{30}
5	0101	5	TB (テラバイト)		1,024 GB	2^{40}
6	0110	6	PB (ペタバイト)		1,024 TB	2^{50}
7	0111	7	EB (エクサバイト)		1,024 PB	2^{60}
8	1000	8	ZB (ゼタバイト)		1,024 EB	2^{70}
9	1001	9	YB (ヨタバイト)		1,024 ZB	2^{80}
10	1010	A	RB (ロナバイト)		1,024 YB	2^{90}
11	1011	B	QB (クエタバイト)		1,024 RB	2^{100}
12	1100	C	媒体	容量	情報伝達速度	単位・記号
13	1101	D	CD-R	700 MB	bit/sec	bps
14	1110	E	DVD-R	4.7 GB	量子ビット	単位・記号
15	1111	F	BD-R	25 GB	quantum bit	qbit

情報量の単位、bit(ビット)について

- 1 bit = 2進法1桁、「0」か「1」
- bit 数
 - 数字「N」の場合、 $\log_2(N)$ bit
 - 38 の場合、約 5.25 bit

情報理論における情報の単位

- 情報量、選択情報量(自己エントロピー)

$$I(E) = \log_2 \left(\frac{1}{P(E)} \right) = -\log_2 P(E) [\text{bit}]$$

- 平均情報量(エントロピー、シャノンエントロピー)

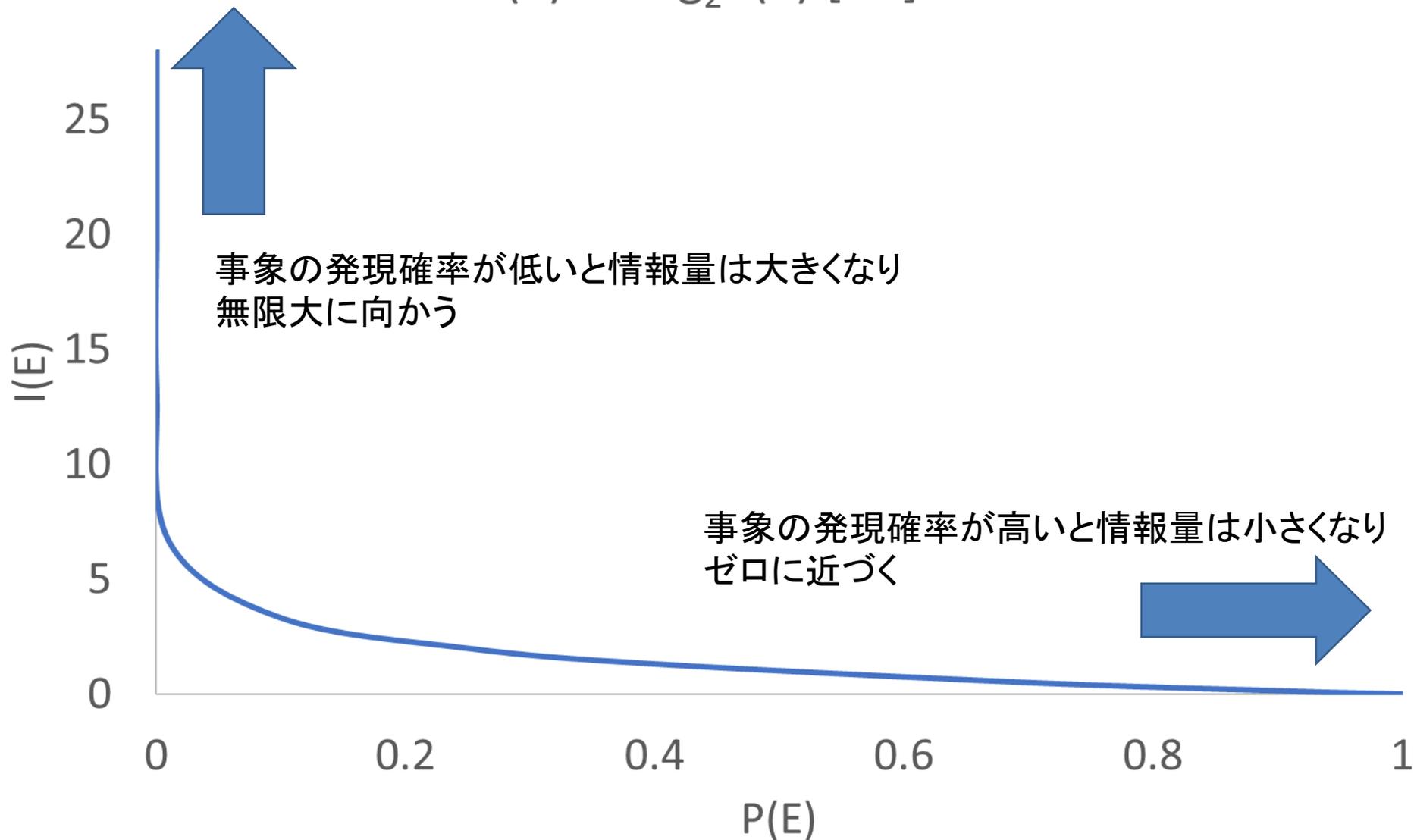
$$H = - \sum_{E=1}^N P(E) \log_2 P(E)$$

- その他、交差エントロピー、相対エントロピー、ダイバージェンス(偽距離)など

P(E): 事象(E)が生じる確率(1以下)

情報量、選択情報量(自己エントロピー)

$$I(E) = -\log_2 P(E) \text{ [bit]}$$



「当然・あたりまえ」も 情報量は低いが「情報」



①ペンギンが内在する「意味情報」から、「形式情報」としての言葉「明日も寒いよ。」への変換

②伝達されてくる「形式情報」としての言葉「明日も寒いよ。」をアザラシが受け取り、
③「意味情報」へ変換する。

たとえば
「南極だからあたりまえでしょ。」
とか
「それは情報にならない」
とか
「そうだね。このマフラーは外せそうにないよね。」
という「意味情報」への変換
④発話＝「形式情報」への変換

話者や聴者の状況、両者の関係、 全体の状況が言葉に意味を与える

- 話者や聴者の**状況**
 - 内面(心情)を含めているので「状況」を使った。
- 両者の過去からの関係を含めた全体の**状況**
 - 第三者の視点から客観的に記述可能な側面として「状況」を使った。
- これらの**状況・状況**が形式的な発音記号、物理学的な音の振動として(**形式情報**として)空中を伝播する「言葉」に意味を与え、**意味情報**となる。
- 話者の心情は必ずしも伝わらないため、両者にとっての**意味情報**は異なりうる。

問題

- A) 1から6までの目が均等に出るサイコロを振ったとき、平均情報量はいくらになるか？
- B) 特殊な細工が施されたサイコロがあり、1から6までの目が出る確率が下記の通りであった。平均情報量はいくらになるか？

目の数	1	2	3	4	5	6
確率	0.20	0.10	0.15	0.30	0.20	0.05

A/D変換

- アナログデータ --- 連続量
- デジタルデータ --- 離散量
- A/D変換
アナログからデジタルへの不可逆変換
 - 標本化 --- データを一定の区画で区切る
 - 量子化 --- 区画内の値を規定範囲の離散的数値に変換
 - 符号化 --- 量子化された値を2進法にする

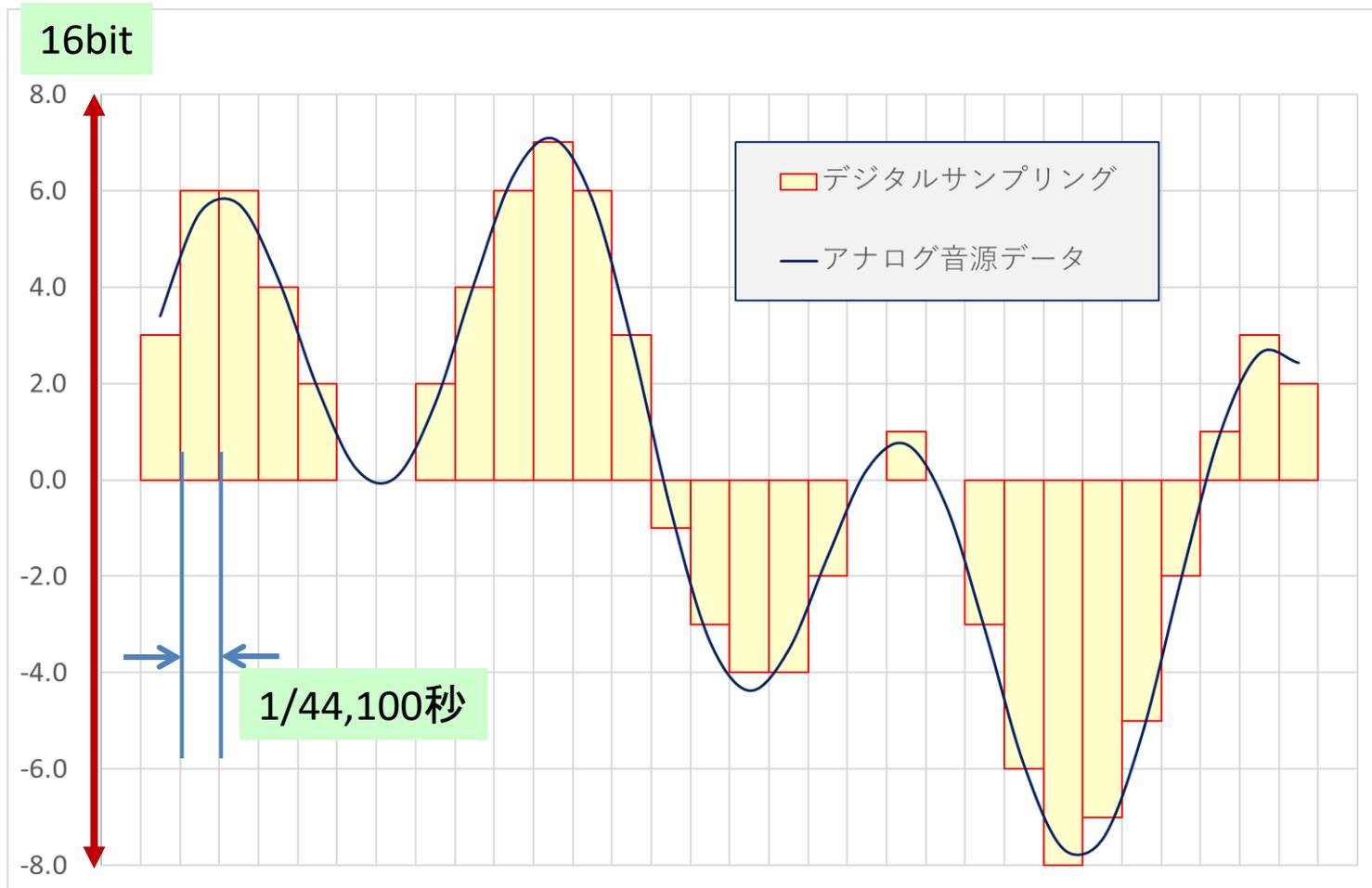
音、画像、動画のデジタル変換

高校テキスト「社会と情報」ないし「情報I」を参照のこと

- 音 (標本化: Hz、量子化: bit)
 - PCM (Pulse Code Modulation)
 - DSD (Direct Stream Digital)
 - MP3, WAVE, AAC, AIFF, FLAC, MIDI
- 画像 (標本化: pixel、量子化: bit)
 - ペイント系 (ビットマップ画像)
 - 非圧縮 (BMP)、非可逆圧縮 (JPEG)、可逆圧縮 (GIF, PNG, JPEG2000)
 - BMP, TIFF, JPEG, GIF, PNG, PICT, PDF
 - ドロー系 (ベクトル画像) --- 数式をデジタル化
 - EPS, PS, PDF
- 動画 (MPEG: 空間周波数変換、コマ単位の差分)
 - MPEG, MP4, AVI, MOV, WMV
- デジタル化困難なデータもあるので注意！

音の標本化と量子化の例

PCM 16bit/44.1kHz (CD品質)



問題

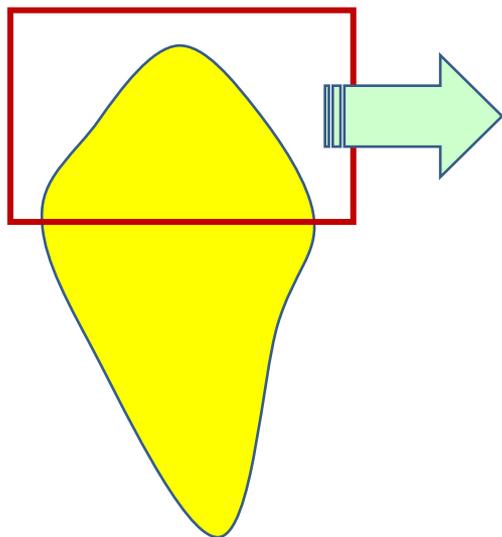
- A) 音楽CD (44.1kHz/16bit) に収録された12分の音楽のデータ量はいくらか？
※ただしモノラル録音 (1チャンネル) とする。
- B) 320kbps と記されたMP3形式の2分37秒の音楽ファイルがある。データ量はいくらと推定されるか？

画像のデジタル変換

- 標本化:ピクセル(Pixel)、空間分解能
- 量子化:ビット(bit)、階調
 - 白か黒のみの画像(1bit)
 - グレースケール(8bit~16bit)
 - カラー画像
 - インデックスカラー(3bit, 8色)
 - RGBカラー(8bit × 3 = 24bit)

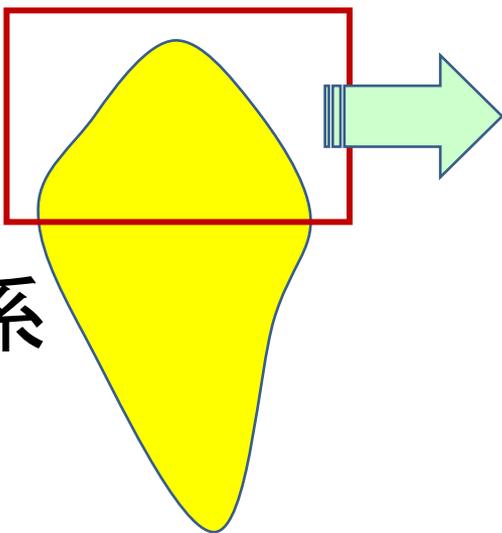


ドロー系



拡大しても滑らか。
形状データを制御
点の値で持つ。

ペイント系



拡大するとジャギー（ギ
ザギザ）が目立つ。ピク
セル単位でデータを持つ

画像のデジタル化に関してネット上で閲覧可能な情報

ビッグデータアクティベーション研究センター>人材育成ユニット>
データサイエンス関連の学習教材・環境整備

<https://www.irp.niigata-u.ac.jp/~suurids/e-learning.html>

からもリンクがあります。

数理・データサイエンス・AI教育強化拠点コンソーシアム

ホーム | コンソーシアム概要 | 運営体制・活動 | トピックス | 活動アーカイブ | リンク | English

活動アーカイブ
http://www.mi.u-tokyo.ac.jp/consortium/suggestion_report.html

● 数理・データサイエンス・AI教育強化拠点コンソーシアムの教材リンク

数理・データサイエンス・AI教育強化拠点コンソーシアム 成果物

教材等

- eラーニング教材・講義動画配信(リテラシーレベル)

モデルカリキュラムと対応する講義動画・スライド

- 1. 社会におけるデータ・AI活用
- 2. データリテラシー
- 3. データ・AI活用における留意事項
- 4. オプション

4-6. 画像解析

<http://www.mi.u-tokyo.ac.jp/consortium/e-learning.html#4>

4-6. 画像解析

キーワード	導入	基本	補助教材	教科書シリーズとの対応
	- 画像解析(スライド・東京大 学). PDF		- 画像解析(サンプルコード・ 東京大学). 外部リンク	
画像データの処理	- 画像処理(動画・滋賀大 学). 外部リンク			- テキスト・画像・音声デー タ分析 外部リンク
画像認識、画像分類、物体検 出	- 画像(動画・滋賀大学). 外部 リンク			- テキスト・画像・音声デー タ分析(画像分類以外対 応). 外部リンク

http://www.mi.u-tokyo.ac.jp/consortium2/pdf/4-6_literacy_level_note.pdf

歯科医療における 画像のデジタル化関連

- 口腔内に装着する補綴物の色との関連
 - 色温度の影響
 - シェードガイドについて
- DICOM規格について
 - 画素値とCT値の違い
 - 負の値の取扱

色温度の差の例

デジタルカメラのホワイトバランス



自動設定
(白色蛍光灯と同じ)



太陽光
(と想定して撮影)



曇り空
(と想定して撮影)



電球
(と想定して撮影)



白色蛍光灯
(と想定して撮影)



昼光色蛍光灯
(と想定して撮影)



<https://www.hakusui-trading.co.jp/products/13500020/>

ビタ クラシカル A1-D4 シェードガイド (2021)

両隣の歯の色に合わせた人工歯の色を選択するためのガイド。

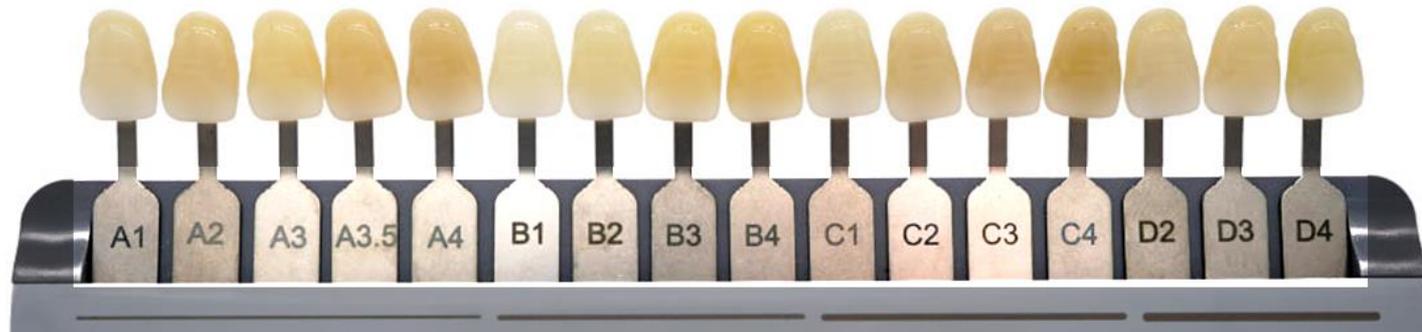
色見本(シェードガイド)からコード(16種類の情報)へ変換する。

シェードガイドから1本ずつ抜き出して、同じ照明の下で並べて比較し、決定する。

コード: A1, A2, A3, A3.5, A4, B1, B2, B3, B4, C1, C2, C3, C4, D2, D3, D4

問題

- シェードガイドのコードをデジタル化するに必要な形式的な情報量(ビット数)の最小値はいくらか？
- ただし、A,B,C,D,の4つの区分が分かりやすくなるように工夫すること。



DICOM画像でのCT値 ImageJでの閲覧

「ワード」の講義
で作った表

人体各組織のCT値

組織名	CT 値 (HU)
エナメル質	2000～4000 程度
象牙質	1200～1900
皮質骨	1000～1800
海綿骨	200～500
筋肉	50～70
水	0
脂肪	-120～-90
空気	-1000

HU: Hounsfield Unit

1: 空気

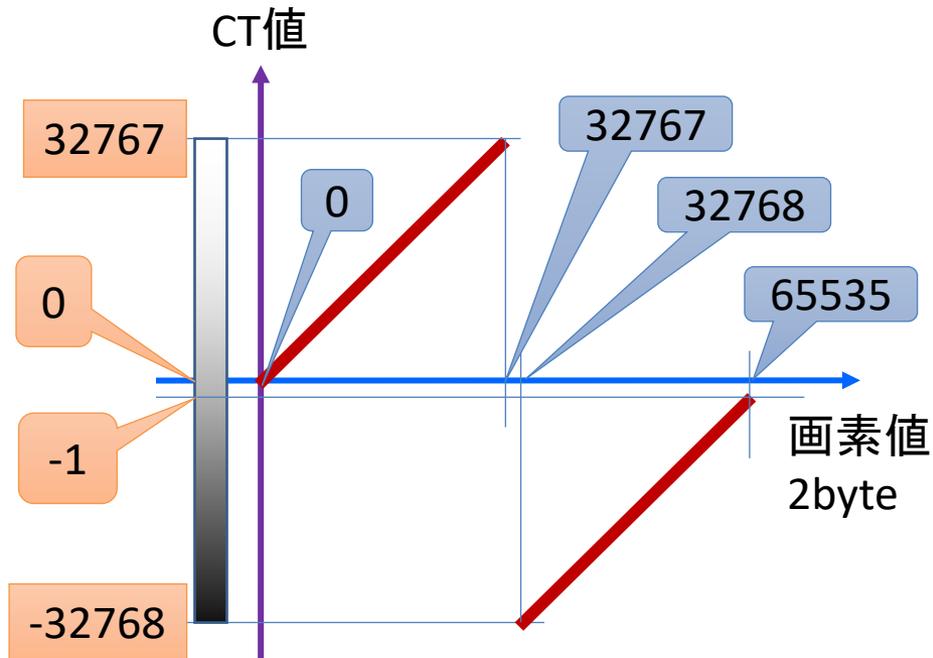
2: エナメル質

3: 象牙質

File	Area	Mean	StdDev	Mode
1	124	-1011.258	38.052	-1021
2	12	3579.667	262.963	3092
3	20	1768.700	71.993	1645

DICOM規格での負数の扱い

0028,0100 Bits Allocated: 16
 0028,0101 Bits Stored: 16
 0028,0102 High Bit: 15
 0028,0103 Pixel Representation: 1



2進表示 16ビット	10進表示		
	符号無し	符号付 2の補数表現	
0000 0000 0000 0000	0	0	この範囲は同じ値を示す。
0000 0000 0000 0001	1	1	
0000 0000 0000 0010	2	2	
...	
0111 1111 1111 1110	32766	32766	
0111 1111 1111 1111	32767	32767	
1000 0000 0000 0000	32768	-32768	第15ビットが1なので、符号付の場合 は負数になる。
1000 0000 0000 0001	32769	-32767	
...	
1111 1111 1111 1110	65534	-2	
1111 1111 1111 1111	65535	-1	

整数データの機械内部の表現と (人間側の)解釈の問題

4ビット			
16進法	2進法	10進法	
		符号無し	符号付き
0	0000	0	0
1	0001	1	1
2	0010	2	2
3	0011	3	3
4	0100	4	4
5	0101	5	5
6	0110	6	6
7	0111	7	7
8	1000	8	-8
9	1001	9	-7
A	1010	10	-6
B	1011	11	-5
C	1100	12	-4
D	1101	13	-3
E	1110	14	-2
F	1111	15	-1

「符号付き」の場合、最上位ビットが「0」か「1」かにて正負を決めている。
負数は「2の補数表示」という規則にて生成される。

Windows附属の「電卓」で簡単に計算できます。

下記URLの詳細参照

<https://www5.dent.niigata-u.ac.jp/~nisiyama/studyskills/digiInfo/2-complement.pdf>



演習問題(授業中に解答提示)

- A) 符号無し16ビット(2バイト)の整数表現にて、10進数の「23156」を16進数で表すとどうなるか？

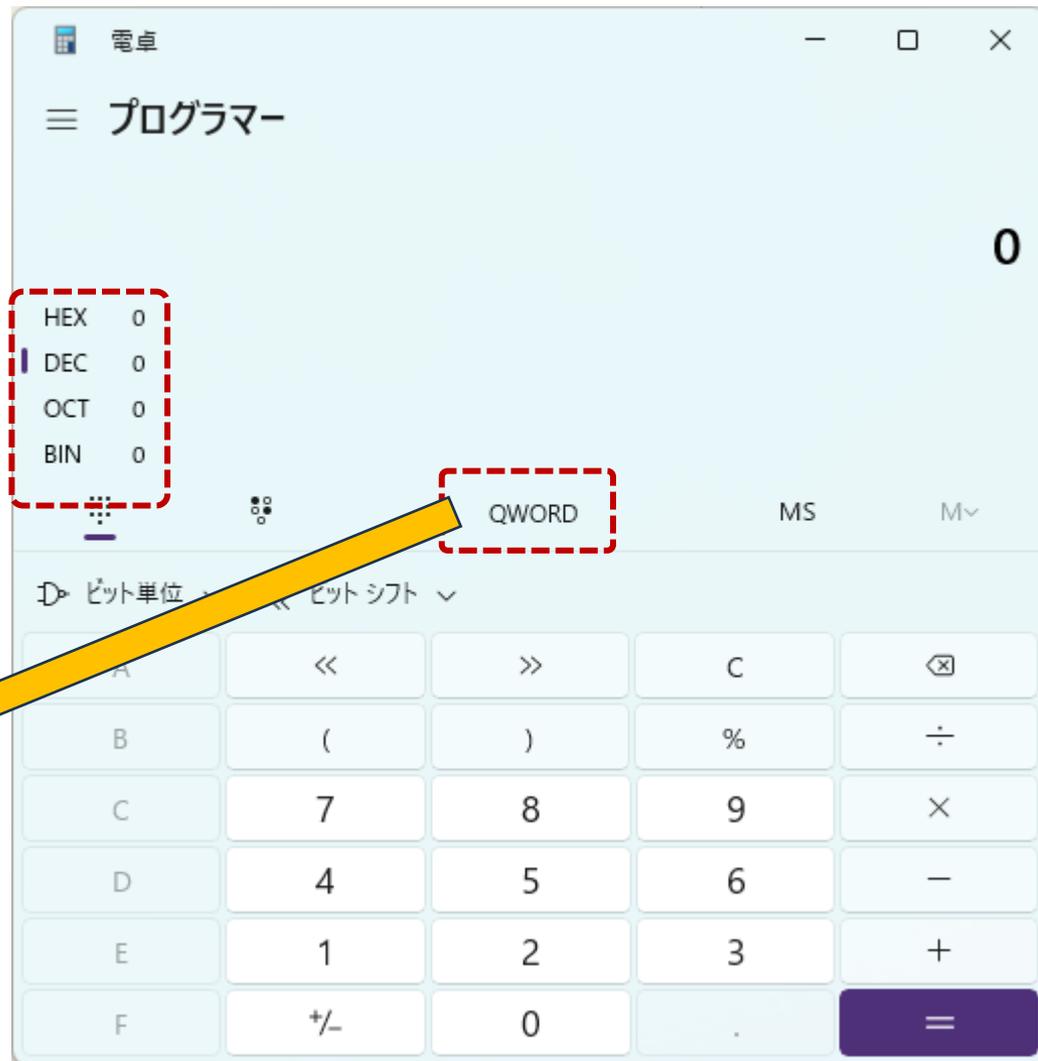
- B) 符号付き16ビット(2バイト)の整数表現にて、10進数の「-23156」は16進数で表すとどうなるか？

Windows 附属の電卓で計算例

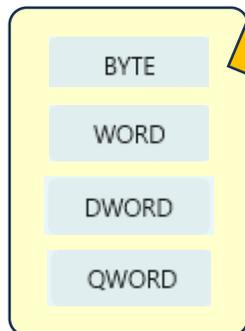


16進数
10進数
8進数
2進数

HEX	0
DEC	0
OCT	0
BIN	0

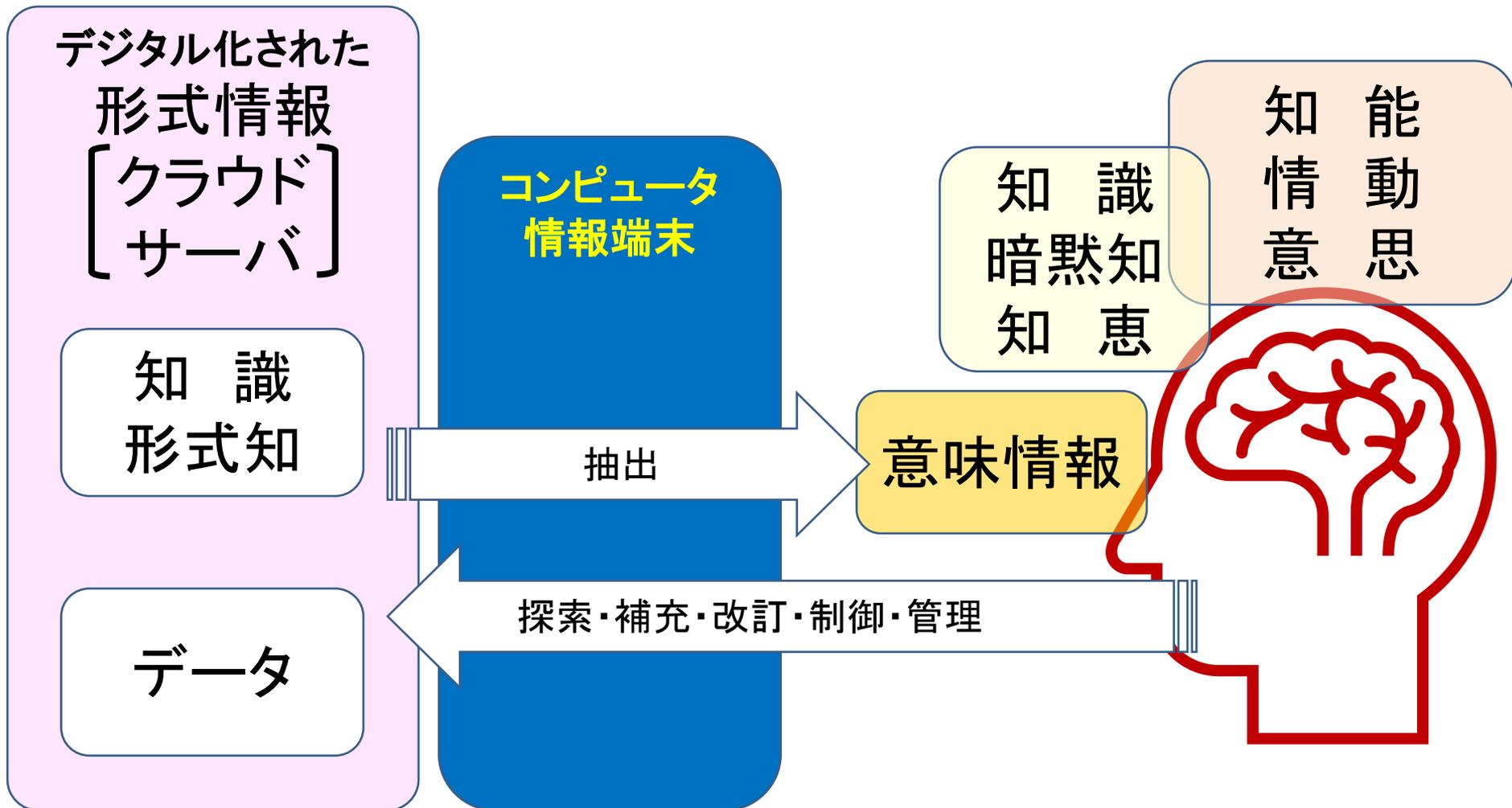


8bit
16bit
32bit
64bit



デジタル情報を処理する コンピュータ(機械)について

データ、情報、知識・知恵と コンピュータ・情報端末との関係



歴史と概要

- 歴史
 - 1946年 ENIAC
 - 1960年代 大型コンピュータのネットワーク化
 - 1980年代 パーソナルコンピュータの普及
- ネットワーク化とパーソナル化
- コンピュータの基本構成

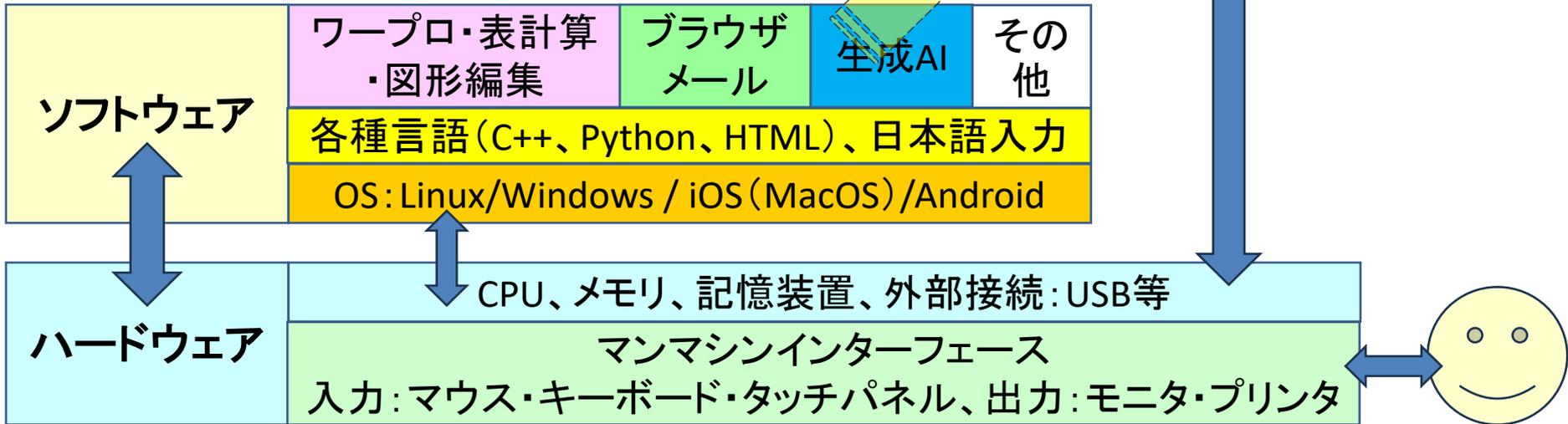
現状、生成AIは1つの応用ソフトであったが、小型言語モデル生成AIを基本的なソフト(ライブラリ)とする方針の企業(microsoft)も出てきている。(大丈夫なのだろうか?)

<https://xtech.nikkei.com/atcl/nxt/column/18/00001/09310/>

近い将来??



現状



演習問題（授業中に解答提示）

コンピュータでの浮動小数点演算で
 $(a + b) - b$ を計算させるといくらになるか？

1. 「a」になる。
2. 「a」よりも大きくなる。
3. 「a」よりも小さくなる。
4. 「a」の値は上記の1から3の全ての場合がある。

a=18.0

b=2⁵³

b=2⁵⁷

b=2⁵⁸

MATLAB 2021b

コマンドウィンドウ

```
>> a=18.0;
>> b=2^53;
>> (a+b)-b
```

ans =

18

```
>> b=2^57;
>> (a+b)-b
```

ans =

32

```
>> b=2^58;
>> (a+b)-b
```

ans =

0

R 4.1.2 (64 bit) - [R Console]



```
language R
version.string R version 4.1.2 (2021-11-01)
nickname Bird Hippie
> a = 18.0
> b = 2^53
> (a + b) - b
[1] 18
> b = 2^57
> (a + b) - b
[1] 32
> b = 2^58
> (a + b) - b
[1] 0
```

Excel (365)

a	18
b	9.0072E+15 =2 ⁵³
((a+b)-b)	18 = ((\$A\$1+A2)-A2)
b	1.44115E+17 =2 ⁵⁷
((a+b)-b)	32 = ((\$A\$1+A4)-A4)
b	2.8823E+17 =2 ⁵⁸
((a+b)-b)	0 = ((\$A\$1+A6)-A6)

Julia 1.10.3

```
julia> a=18.0
18.0
```

```
julia> b=2^53
9007199254740992
```

```
julia> ((a+b)-b)
18.0
```

```
julia> b=2^57
144115188075855872
```

```
julia> ((a+b)-b)
32.0
```

```
julia> b=2^58
288230376151711744
```

```
julia> ((a+b)-b)
0.0
```

Python 3.9.7

```
File Edit Shell Debug Options Window Help
Python 3.9.7 (tags/v3.9.7:1016ef3, Aug 30 2021, v.1929 64 bit (AMD64)) on win32
Type "help", "copyright", "credits" or "license()" for more
>>> a = 18.0
>>> b = 2 ** 53
>>> print ((a + b) - b)
18.0
>>> b = 2 ** 57
>>> print ((a + b) - b)
32.0
>>> b = 2 ** 58
>>> print ((a + b) - b)
0.0
>>>
```

```
Python 3.6.8 (tags/v3.6.8:3c6b436a57, Dec 24 2018, 00:16:47) [MSC v.1916 64 bit  
(AMD64)] on win32  
Type "help", "copyright", "credits" or "license()" for more information.
```

```
>>> a = 18.0  
>>> b = 2 ** 53  
>>> print ((a + b) - b)  
18.0  
>>> b = 2 ** 57  
>>> print ((a + b) - b)  
32.0  
>>> b = 2 ** 58  
>>> print ((a + b) - b)  
0.0  
>>> print (2 ** 57)  
144115188075855872  
>>> print (2 ** 57 + 18)  
144115188075855890  
>>> print (2 ** 57 + 18.0)  
1.441151880758559e+17  
>>> |
```

Pythonでの計算例 (EXCEL, C, R, MATLABでも同様の結果となる)
整数と浮動小数点数が切り替わる段階での誤差

倍精度浮動小数点として「18.0」を変数:aに代入
倍精度浮動小数点として、変数:bに「 2^{57} 」を代入
「(a+b)-b」という計算を行い、結果を出力すると「18.0」にならず
「32.0」になる。

bに「 2^{53} 」、「 2^{54} 」、「 2^{58} 」を代入して計算すると、それぞれ
「18.0」、「16.0」、「0.0」と変化する。

「 2^{57} 」は整数として「144115188075855872」であり、整数「18」を加
算すると「144115188075855890」となるが、浮動小数点数として
の「18.0」を加算すると、「1.441151880758559e+17」となり、表示さ
れず内部で保持されている下2桁を含む範囲で整数から浮動小
数点数への変換時に誤差が生じ、予想外の計算結果になる。

参考: 中田真秀、「線形代数演算ライブラリBLASとLAPACKの基礎と実践1」、理化学研究
所・計算科学研究センター配信講義 計算科学技術特論A、第6回 2019年5月23日

<https://www.r-ccs.riken.jp/outreach/schools/tokurona-2019/>

<https://www.r-ccs.riken.jp/wp/wp-content/uploads/2020/09/nakata190523.pdf>

浮動小数点 ≠ 実数 (という概念) 結果の「正しさ」という意味

- 機械側が形式的に正しく計算し、正しく処理しても、人間側の概念とは異なる結果になることがある。
- このことは、生成AIが(機械側にとって形式的に)正しく処理し、正しく出力したとしても、人間側にとって間違った結果となる **ハルシネーションに共通する部分** である。
- 「正しさ」という意味が機械側と人間側で食い違っており、人間側の意味に機械側が処理する内容(記号処理)が合致していないことに起因する。
- **広い意味で「記号接地問題」** になっている。

形式情報と意味情報の差 各種言語の演算子の結合の差

$$2^{3^2} = ? \quad 2^{-3^2} = ?$$

```
R 4.1.2 Terminal x Jobs x  
R 4.1.2 · ~/ ↗  
> 2^3^2  
[1] 512  
> 2^-3^2  
[1] 0.001953125  
>
```

Excel (365)		
	A	B
1	= 2^3^2	64
2	= 2^-3^2	0.015625
3		

```
Python 3.9.7 3.9.7  
File Edit Shell Debug Options Window Help  
Python 3.9.7 (tags/v3.9.7:1016ef3, Aug 30 2021,  
20:19:38) [MSC v.1929 64 bit (AMD64)] on win32  
Type "help", "copyright", "credits" or "license  
( )" for more information.  
>>> 2**3**2  
512  
>>> 2**-3**2  
0.001953125  
>>>
```

```
MATLAB 2021b  
  
>> 2^3^2  
  
ans =  
  
64  
  
>> 2^-3^2  
  
ans =  
  
0.0156
```

演習問題（授業中に解答提示）

- A) 「エクセル」にて「 2^{3^2} 」を計算させた結果を「512」にするには、どうすれば良いか？
- B) 「R」にて「 2^{3^2} 」を計算させた結果を「64」にするには、どうすれば良いか？

※「R」は、教育用端末や貸し出しノートPCに標準で入っています。

<https://www.cais.niigata-u.ac.jp/system/pc/softlist.html>



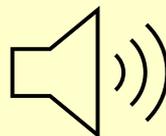
機械と人間との インターフェース

入力装置・文字コードの問題

英文タイプライター



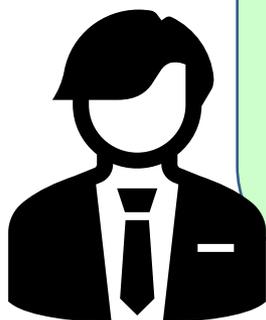
出力(ディスプレイ、印刷など)



コンピュータ



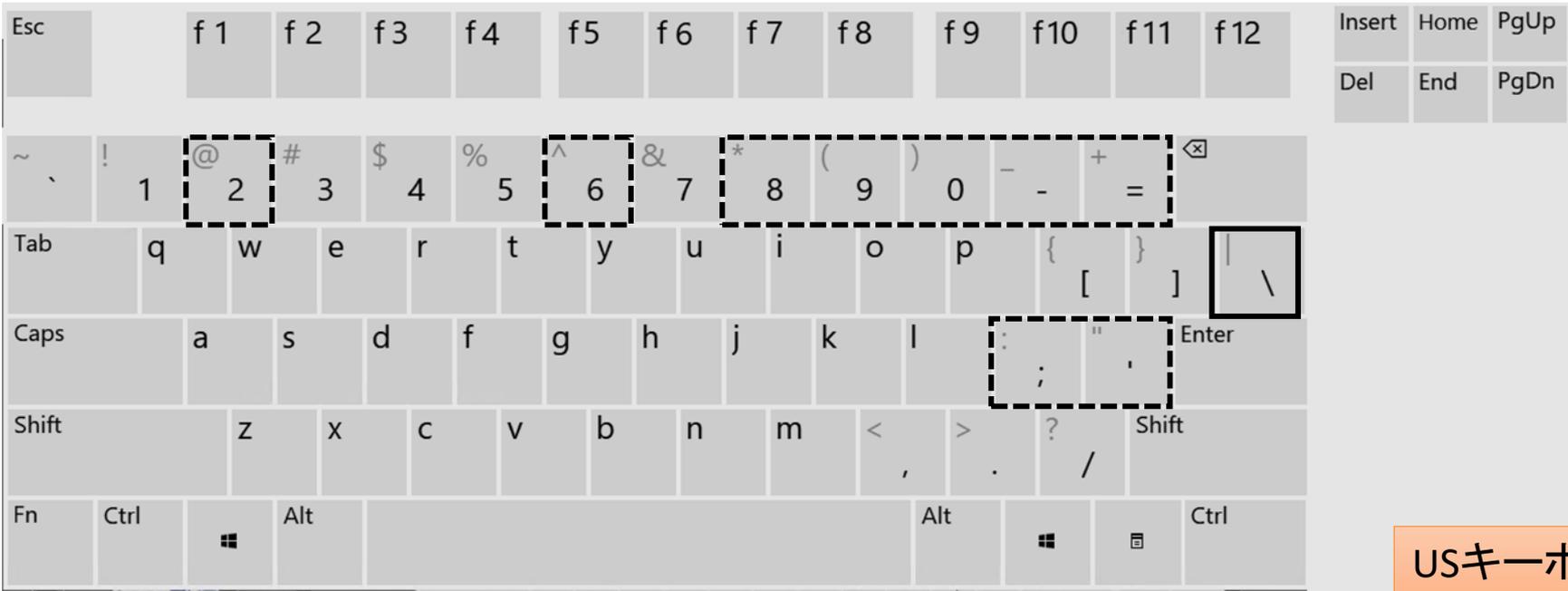
入力デバイス



タイプライターから コンピュータへ

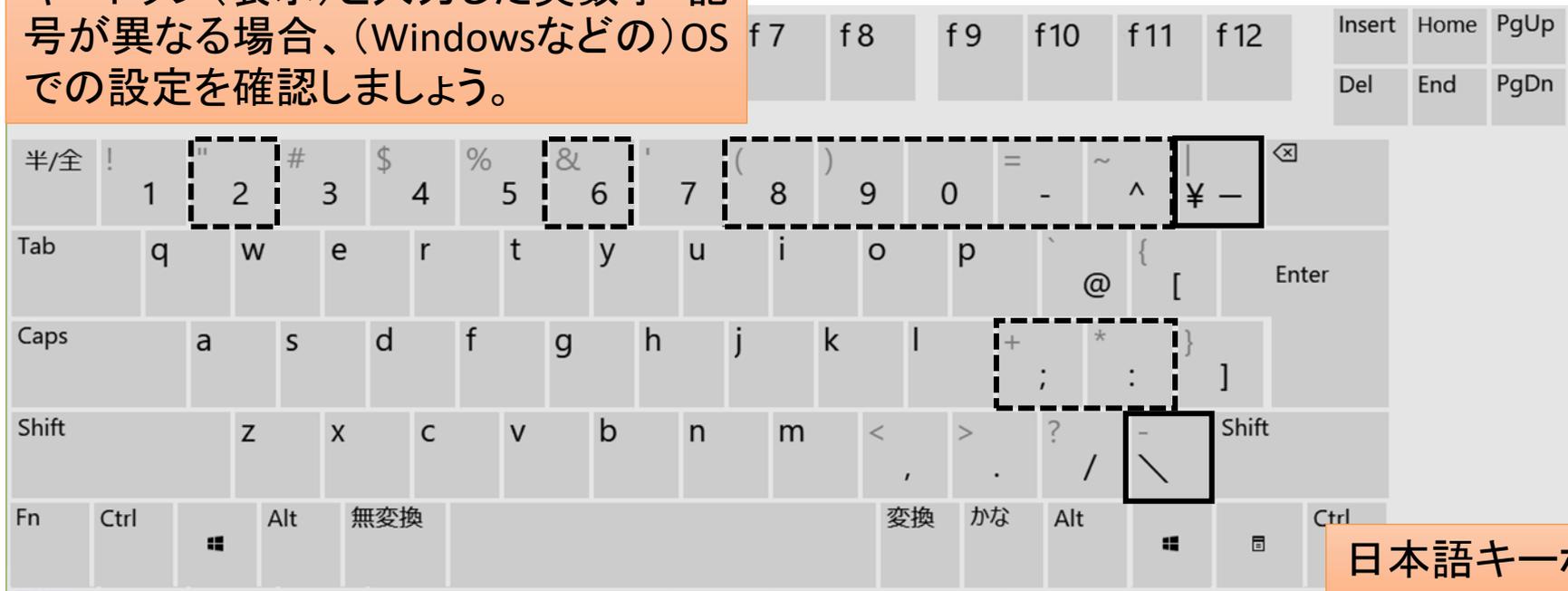


- **タイプライター**では、文字は彫ってあった。
 - インクリボンを介し刻印された文字を指の力ないし電動にて紙に打ち付けていた。
 - 異なる文字セットを使うには印字ヘッドを交換することで対応。
- **コンピュータ**では、文字はコードにて扱う。
 - 多くの文字を扱うには、多くのコードが必要。
 - コンピュータの記憶容量の問題から、当初は、アルファベットと数字しか扱えなかった。
 - その後、漢字などが扱えるようになってきた。



USキーボード

キートップ(表示)と入力した英数字・記号が異なる場合、(Windowsなどの)OSでの設定を確認しましょう。



日本語キーボード

ASCIIおよびANK (JIS X 0201)コード表 (8bit)

		下位ビット															
		0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
+【Ctrl】	00	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
	10	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
	20		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
	30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
	40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	50	P	Q	R	S	T	U	V	W	X	Y	Z	[¥]	^	_
	60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
	70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
	80																
	90																
	A0	□	。	「	」	、	・	ヲ	ア	イ	ウ	エ	オ	ヤ	ユ	ヨ	ツ
	B0	ー	ア	イ	ウ	エ	オ	カ	キ	ク	ケ	コ	サ	シ	ス	セ	ソ
	C0	タ	チ	ツ	テ	ト	ナ	ニ	ヌ	ネ	ノ	ハ	ヒ	フ	ヘ	ホ	マ
	D0	ミ	ム	メ	モ	ヤ	ユ	ヨ	ラ	リ	ル	レ	ロ	ワ	ン	ゝ	。
	E0																
	F0																

ASCII制御文字
 ANK拡張部分※

現在も良く使われている制御コードの例

09:horizontal tab(HT)、水平タブ

0A:line feed (LF)、改行

0D:carriage return (CR)、復帰

【Ctrl】キーによる制御コード入力は、現在、UNIX (LINUX)を除いて割当が大幅に変更されている。

※ANKはJIS X 0201(旧JIS C 6220)規格の俗称で「Alphabet + Numerical digit + Katakana」の頭文字。

コンピュータで扱える文字(1)

1バイトで扱える基本的な文字コード

- 7bit-8bit: 128-256文字
 - **EBCDIC** (extended binary coded decimal interchange code)
 - 拡張2進化10進コード
 - **ASCII** (American National Standard Code for Information Interchange)
 - 情報交換用米国標準コード(**7bit**)
 - **ANK** (アंक; Alphabetic Numeric and Kana)
 - JIS X 0201
 - ASCIIコードに、カナを加えたもの(**8bit**)

よく使う特殊キー

- 制御コード領域と密接に関連しています。



ctrl+キー

旧: 制御コード領域のコードを入力できた。
(現在でも一部OSやソフトは継承している)

shift+キー

キーに割り当てられている
複数の文字を切り替える。

機種依存する改行コード

- Windows (DOS)系: CR+LF
 - 16進表示: 0D 0A
- Macintosh系 (Mac OS 9以前): CR
 - 16進表示: 0D
- UNIX系: LF
 - 16進表示: 0A
- 最近のOSでは、表示には影響しなくなっているが、何らかの処理時には副作用の原因となる。

その昔、テレタイプを用いていたとき、
印字ヘッドを左端に戻す (CR)
一行用紙を送る (LF)
という二つの操作コマンドが必要だった。

下記動画の5:45からキーボードの説明
6:20から「CR」、6:40から「LF」の動作例
Teletype Model 19 (and Model 15) Demonstration
<https://youtu.be/jxkygWI-Wfs>

コンピュータで扱える文字(2)

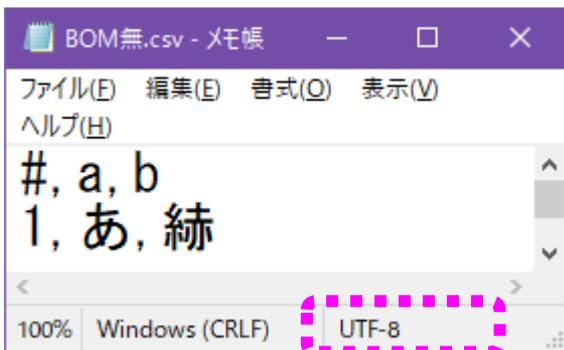
2バイト以上で扱える漢字・各国語への拡張

- 16bit (2byte): 65536文字
 - JIS (ISO 2022-JP)
 - Shift-JIS (S-JIS, MS漢字コード)
 - EUC (Extended UNIX Code)
 - それぞれにコード体系が異なるので文字化けが生じる。
- 8-32bit (1-4byte):
最大4294967296文字 (約43億文字)
 - Unicode
 - UTF-8, UTF-16, UTF-32など
 - 注意: 近年UTF-8が主流となっているが、BOMの有無にて、処理時に副作用の原因になる場合がある。

BOM: Byte Order Mark

7-8 bit	EBCDIC		(extended binary coded decimal interchange code)
1 byte			初期のIBMコンピュータにて用いられた。
128-256文字	ASCII		(American National Standard Code for Information Interchange)
			情報交換用米国標準コード(7bit)
	ANK		(Alphabetic Numeric and Kana)
			JIS X 0201
			ASCIIコードに、カナを加えたもの(8bit)
16 bit	JIS		(ISO 2022-JP)
			通信系、特に電子メールで主流。
2 byte	Shift-JIS		(S-JIS, MS漢字コード)
			Windowsで主流だったが、近年UTF-8へ移行しつつある。
65536文字	EUC		(Extended UNIX Code)
			Linuxで主流だったが、近年UTF-8へ移行してきている。
8-32 bit	Unicode		文字符号化形式、文字符号化方式を定めた規格名
1-4 byte	UTF		Unicode Transformation Format
最大約43億文字	UTF-8		現在主流、文字化けの危険性が最も少ない
			全ての文字を1符号単位(8bit、1byte)ないしサロゲートペアと呼ばれる2~4符号単位(2~4byte)として扱う。
			原理的にBOMは不要で非推奨だが、BOM付きしか扱わないソフトもあるため注意
			BOM無しを前提とするソフトにBOM付きファイルが入力されると誤動作の元となるので注意。
	UTF-16		全ての文字を1符号単位(16bit、2byte)ないしサロゲートペアと呼ばれる2符号単位(32bit、4byte)として扱う。BOMは必須。
		UTF-16BE	UTF-16のビッグエンディアン
		UTF-16LE	UTF-16のリトルエンディアン
	UTF-32		全ての文字を32bitとして扱う。BOMは必須。
		UTF-32BE	UTF-32のビッグエンディアン
		UTF-32LE	UTF-32のリトルエンディアン

「BOM無」と「BOM付」の違い



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	FileSize:19	0x13				3	← set color level					init						
2	StartPos:0	0x0					A0	FF										Disp.Mode: Hex
3	PageLen :3072	0xC00					2c	2c				open						
4	EndPos :3071	0xBFF					0a	0d										
5	PageLen:1~3072	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
6		0:	23	2C	61	2C	62	0D	0A	31	2C	E3	81	82	2C	F0	A6	80
7		10:	97	0D	0A													

#	,	a	,	b	↵	1	,	あ	,	絵	↵							
23	2C	61	2C	62	0D	0A	31	2C	E3	81	82	2C	F0	A6	80	97	0D	0A

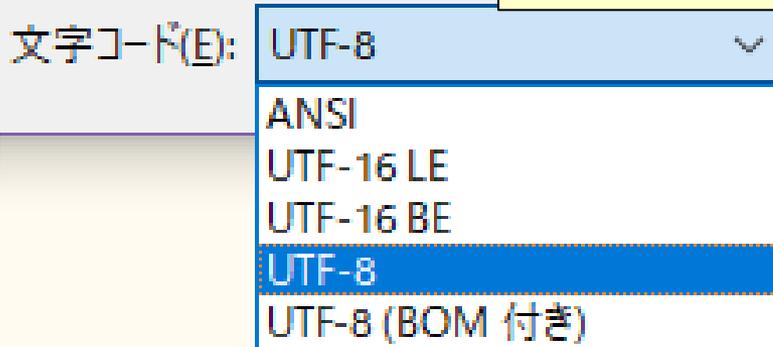


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	FileSize:22	0x16				3	← set color level					init						
2	StartPos:0	0x0					A0	FF										Disp.Mode: Hex
3	PageLen :3072	0xC00					2c	2c				open						
4	EndPos :3071	0xBFF					0a	0d										
5	PageLen:1~3072	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
6		0:	EF	BB	BF	23	2C	61	2C	62	0D	0A	31	2C	E3	81	82	2C
7		10:	F0	A6	80	97	0D	0A										

[BOM]	#	,	a	,	b	↵	1	,	あ	,	絵	↵									
EF	BB	BF	23	2C	61	2C	62	0D	0A	31	2C	E3	81	82	2C	F0	A6	80	97	0D	0A

文字コードが「UTF-8」のBOM付き ファイルに注意

@ITクラウドWindows Server InsiderUTF-8: Tech Basics / Keyword
<https://www.atmarkit.co.jp/ait/articles/1603/28/news035.html>



BOM: Byte Order Mark

- UTF-8では「BOM付」と「BOM無」がある。
- WordやExcelで、テキスト形式での保存時に「UTF-8」を選択すると、自動的に「BOM付」となる。
- メモ帳はバージョンによって異なる。
【参照】初心者のためのOFFICE講座・メモ帳の既定の文字コードが[UTF-8]に
<https://hamachan.info/win10-win-memoutf/>
- Python等のプログラムでの読み書き時に注意が必要。
<https://www.lifewithpython.com/2017/10/python-detect-bom-in-utf8-file.html>

Python 3.9でのUTF-8の BOM有りの影響

```
In [1]: import csv
```

```
In [2]: csv_file = open("../BOM無.csv", "r", encoding="utf_8")  
f = csv.DictReader(csv_file)  
for row in f:  
    print(row)  
csv_file.close()
```

```
{'#' : '1', 'a': 'あ', 'b': 'あ'}
```

```
In [3]: csv_file = open("../BOM有.csv", "r", encoding="utf_8")  
f = csv.DictReader(csv_file)  
for row in f:  
    print(row)  
csv_file.close()
```

```
{'¥ufeff#' : '1', 'a': 'あ', 'b': 'あ'}
```

PythonでBOMの有無に関わらず UTF-8を使うには、“utf_8_sig”を使う

```
In [4]: csv_file = open("../BOM無.csv", "r", encoding="utf_8_sig")
f = csv.DictReader(csv_file)
for row in f:
    print(row)
csv_file.close()
```

```
{'#': '1', 'a': 'あ', 'b': '緑'}
```

```
In [5]: csv_file = open("../BOM有.csv", "r", encoding="utf_8_sig")
f = csv.DictReader(csv_file)
for row in f:
    print(row)
csv_file.close()
```

```
{'#': '1', 'a': 'あ', 'b': '緑'}
```

PowerShellでのUTF-8関連問題 (歴史的背景含む)

- https://docs.microsoft.com/ja-jp/powershell/module/microsoft.powershell.core/about/about_character_encoding?view=powershell-7.2
- <https://virment.com/how-to-install-latest-powershell-in-windows10/>

文字エンコード (Windows PowerShell)

PowerShell 5.1では、Encoding パラメーターは 次の値をサポートしています。

- `Ascii` Ascii (7ビット) 文字セットを使用します。
- `BigEndianUnicode` ビッグ エンディアンのバイト順で UTF-16 を使用します。
- `BigEndianUTF32` ビッグ エンディアンのバイト順で UTF-32 を使用します。
- `Byte` 文字のセットをバイト シーケンスにエンコードします。
- `Default` システムのアクティブなコード ページ (通常は ANSI) に対応するエンコーディングを使用します。
- `Oem` システムの現在の OEM コード ページに対応するエンコードを使用します。
- `String`Unicode` と同じです。
- `Unicode` リトル エンディアンのバイト順で UTF-16 を使用します。
- `Unknown`Unicode` と同じです。
- `UTF32` リトル エンディアンのバイト順で UTF-32 を使用します。
- `UTF7` UTF-7 を使用します。
- `UTF8` UTF-8 (BOM 付き) を使用します。

https://docs.microsoft.com/ja-jp/powershell/module/microsoft.powershell.core/about/about_character_encoding?view=powershell-7.2

PowerShell での文字エンコード

PowerShell (v6 以上) では、**Encoding パラメーター**は 次の値をサポートします。

- `ascii`: ASCII (7 ビット) 文字セットのエンコードを使用します。
- `bigendianunicode`: ビッグ エンディアンのバイト順を使用して UTF-16 形式でエンコードします。
- `oem`: MS-DOS およびコンソール プログラムの既定のエンコードを使用します。
- `unicode`: リトル エンディアンのバイト順を使用して UTF-16 形式でエンコードします。
- `utf7`: UTF-7 形式でエンコードします。
- `utf8`: UTF-8 形式 (BOM なし) でエンコードします。
- `utf8BOM`: バイト 順マーク (BOM) を使用して UTF-8 形式でエンコードします
- `utf8NoBOM`: バイト 順マーク (BOM) を使用せずに UTF-8 形式でエンコードします
- `utf32`: UTF-32 形式でエンコードします。

https://docs.microsoft.com/ja-jp/powershell/module/microsoft.powershell.core/about/about_character_encoding?view=powershell-7.2

PowerShell (5.1) 及び コマンドプロンプトでの「UTF-8」のBOMの有無による文字化け

※ PowerShell 6以上では左側の問題は生じない

PowerShellの場合

```
PS:Jupyter > type woBOM.csv
#,a,b
1,縹・□・
PS:Jupyter > type withBOM.csv
#,a,b
1,あ,縹
```

Sift-JISとして解釈

UTF-8として正しく表示している

コマンドプロンプトの場合

```
E:¥Python¥Jupyter> type woBOM.csv
#,a,b
1,縹・□・
E:¥Python¥Jupyter> type withBOM.csv
・ ¥#,a,b
1,縹・□・
```

Sift-JISとして解釈

Sift-JISとして解釈し、さらに先頭のBOMコードを文字として解釈

woBOM.csvの構造

#	,	a	,	b	↵	1	,	あ	,	縹	↵							
23	2C	61	2C	62	OD	OA	31	2C	E3	81	82	2C	F0	A6	80	97	OD	OA

withBOM.csvの構造

[BOM]	#	,	a	,	b	↵	1	,	あ	,	縹	↵									
EF	BB	BF	23	2C	61	2C	62	OD	OA	31	2C	E3	81	82	2C	F0	A6	80	97	OD	OA

全文検索におけるWindowsでの UTF-8 (BOM無し)の問題

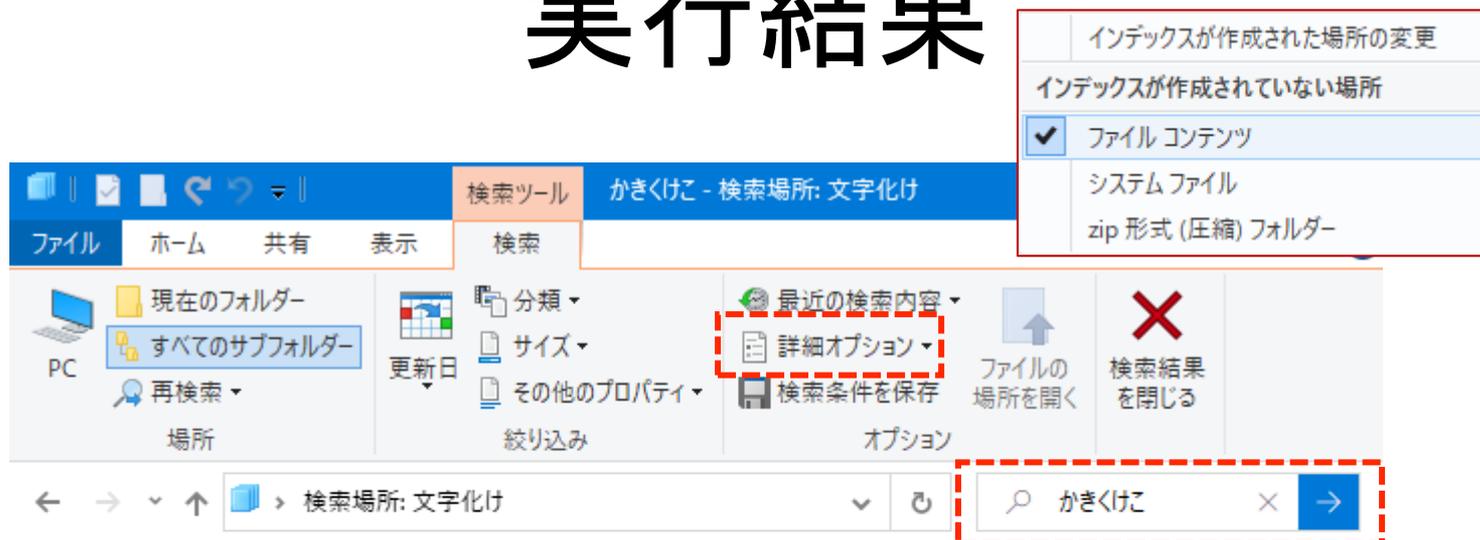
全文検索用のサンプルテキスト

名前

- 文字化け例-UTF-16-LE-CRLF.txt
- 文字化け例-UTF-16-LE-CR.txt
- 文字化け例-UTF8-wo-BOM-LF.txt
- 文字化け例-UTF8-wo-BOM-CRLF.txt
- 文字化け例-UTF8-with-BOM-LF.txt
- 文字化け例-SJIS-CRLF.txt
- 文字化け例-JIS-CRLF.txt
- 文字化け例-EUC-LF.txt

```
1 | 文字化けの例 ←
2 | あいうえお ←
3 | かきくけこ ←
4 | 新潟大学 ←
   | [EOF]
```

Windowsの全文検索の設定と実行結果



名前

- 文字化け例-UTF-16-LE-CRLF.txt
- 文字化け例-UTF-16-LE-CR.txt
- 文字化け例-UTF8-wo-BOM-LF.txt
- 文字化け例-UTF8-wo-BOM-CRLF.txt
- 文字化け例-UTF8-with-BOM-LF.txt
- 文字化け例-SJIS-CRLF.txt
- 文字化け例-JIS-CRLF.txt
- 文字化け例-EUC-LF.txt

- 文字化け例-UTF8-with-BOM-LF.txt サイズ: 67 バイト
更新日時: 2021/12/28 11:57
- 文字化け例-UTF-16-LE-CR.txt サイズ: 50 バイト
更新日時: 2021/06/07 19:09
- 文字化け例-UTF-16-LE-CRLF.txt
更新日時: 2021/06/07 19:02
- 文字化け例-SJIS-CRLF.txt
更新日時: 2020/06/02 10:18

Windowsでの全文検索では
SJIS、UTF-16
および
UTF8(BOM付き)
のみ検索できた。
UTF8(BOM無し)は検索できない

PowerShellでのSelect-String (sls) が良い

Windows Info 第279回

BOMなしUTF-8によってWindowsでもたらされる困惑

参考資料

2021年06月13日 10時00分更新

文 ● 塩田紳二 編集 ● ASCII

<https://ascii.jp/elem/000/004/058/4058837/>

名前

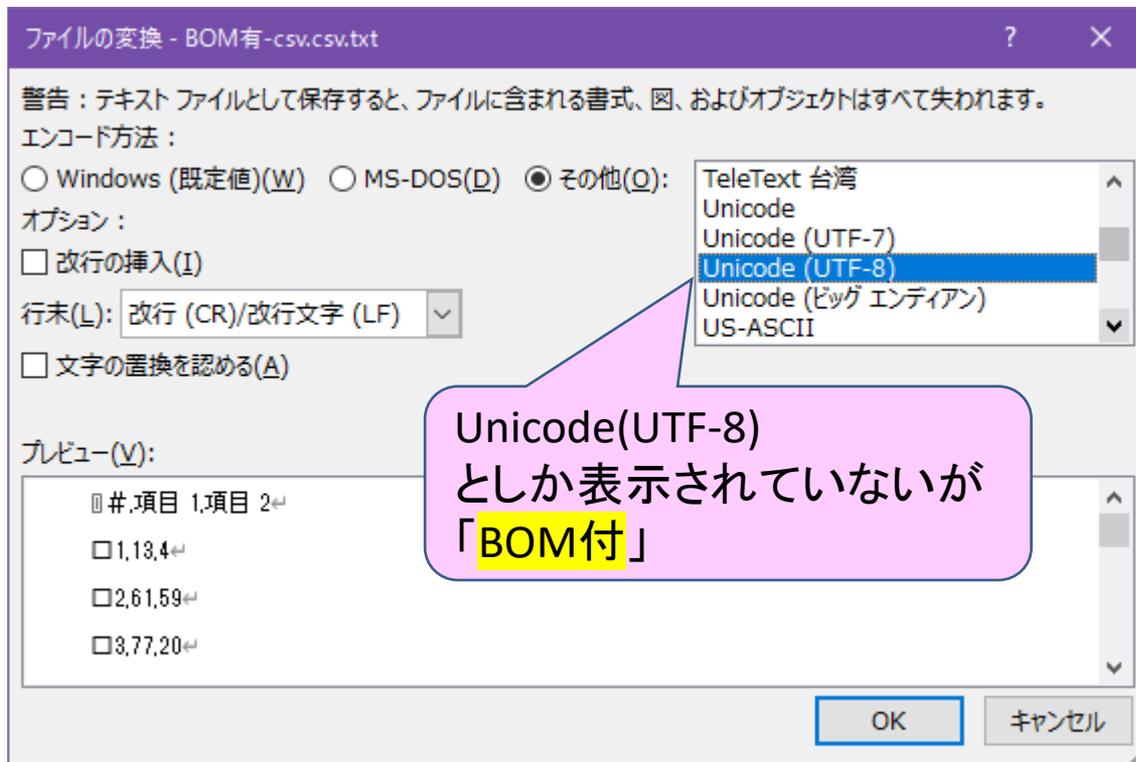
- 文字化け例-UTF-16-LE-CRLF.txt
- 文字化け例-UTF-16-LE-CR.txt
- 文字化け例-UTF8-wo-BOM-LF.txt
- 文字化け例-UTF8-wo-BOM-CRLF.txt
- 文字化け例-UTF8-with-BOM-LF.txt
- 文字化け例-SJIS-CRLF.txt
- 文字化け例-JIS-CRLF.txt
- 文字化け例-EUC-LF.txt

```
PS:文字化け > sls "かきくけこ" *.txt  
文字化け例-UTF-16-LE-CR.txt:3:かきくけこ  
文字化け例-UTF-16-LE-CRLF.txt:3:かきくけこ  
文字化け例-UTF8-with-BOM-LF.txt:3:かきくけこ  
文字化け例-UTF8-wo-BOM-CRLF.txt:3:かきくけこ  
文字化け例-UTF8-wo-BOM-LF.txt:3:かきくけこ  
PS:文字化け > _
```

PowerShellのSelect-String検索では
UTF-16 および UTF-8 (BOM付き、BOM無し) の両方を検索できた。
※SJIS、JIS、およびEUCは対象外

ワードやエクセルからテキスト形式 (CSV形式) での保存注意

ワードでの保存時



UTF-8で保存すると、自動的に「BOM付」となり、他のソフトでのデータ読み込み時にトラブルを発生しやすいので注意！！

エクセルでの保存時



こちらは「Sift-JIS」(SJIS)なのでMAC等で文字化けする

文字化け: 同じコードで「解釈」が異なる

Windows、Mac、Linux間でのファイルの交換時に注意

Windows10 元の文章

SJIS-CRLF

ファイル(E) 編集(E) 書式(O) 表示(V) ヘルプ(H)

```
文字化けの例  
あいうえお  
かきくけこ  
新潟大学
```

100% Windows (CRLF) ANS

Windows10

JIS-CRLF として読み込む

ファイル(E) 編集(E) 書式(O) 表示(V) ヘルプ(H)

```
·$BJ8;z2=$1$NNc·(B  
·$B$~$$$&$($*·(B  
·$B$+$-$/$1$3·(B  
·$B?73cBg3X·(B
```

100% Windows (CRLF) UTF-8

Linux Ubuntu 18.04

で読み込む

```
hn@64pro:~/text$ cat /etc/issue  
Ubuntu 18.04.6 LTS ¥n ¥l  
  
hn@64pro:~/text$ cat SJIS-CRLF.txt
```

vw

```
hn@64pro:~/text$
```

Apple iOS 15.1

で読み込む

```
0000000000  
0000000000  
0000000000  
0V0000w
```

通常UTF-8(BOM無)を使えば問題はない。

文字コードと半角カタカナ メールソフトでの文字化け

 Microsoft | サポート Microsoft 365 Office Windows 表示を増やす

製品 デバイス 最新情報 アカウントおよび課金 テンプレート その他のサポート

Outlook で JIS (ISO-2022-JP) にエンコードされたメールを受信した際、半角カタカナがメールの本文に含まれていると本文が文字化けする場合があります

現象

Microsoft Office Outlook 2007、Microsoft Office Outlook 2003、Microsoft Outlook 2002 および Microsoft Outlook 2000 で、JIS (ISO-2022-JP) にエンコードされたメールを受信した際に、半角カタカナがメールの本文に含まれていると、メールの本文が文字化けする場合があります。一度、この現象が発生すると JIS でエンコードされたメールを受信する度に現象が発生し、JIS 以外でエンコードされたメールを受信するまで現象が発生し続ける場合があります。

この現象の文字化けの特徴として、メールの本文に半角カタカナが多く表示されます。なお、[宛先]、[CC]、[BCC]、[件名] は文字化けしません。

Macとのやり取りでの メールの文字化け

- メールタイトルのMIMEエンコードと呼ばれる変換にて文字化けが生じる
- 元メールのタイトルが「ISO-2022-JP」にてエンコードされていたものをThunderbirdで受信すると(自動的にUTF-8に変換され)エンコードの改行部分に「」(0xFFFD)が入る。



Re: 歯学会HP確認用のフイル確認のお願い



Re: 歯学会HP確認用のファイル確認のお願い

Subject: Re: **=?ISO-2022-JP?B?GyRCO3UzWDJxGyhCSFABJEIzTkcnTVEkTiVVGyhK?=?**

=?ISO-2022-JP?B?GyRCJSEIJCVrM05HJyROJCo0aiQkGyhC?=?

送信側での改行部分
受信側で文字化けになる

In-Reply-To: <#####@dent.niigata-u.ac.jp>

References: <#####@dent.niigata-u.ac.jp>

MIME-Version: 1.0

「ISO-2022-JP」でエンコードされたタイトル

X-Mailer: AL-Mail32 Version 1.13

Content-Type: text/plain; charset=iso-2022-jp

echo "=?ISO-2022-JP?B?GyRCO3UzWDJxGyhCSFABJEIzTkcnTVEkTiVVGyhK?=" | wsl nkf -w
歯学会HP確認用のフ

送信側での改行部分、受信側で文字化けになる

echo "=?ISO-2022-JP?B?GyRCJSEIJCVrM05HJyROJCo0aiQkGyhC?=" | wsl nkf -w
ファイル確認のお願い

受信側改行部分

MIME-Version: 1.0

User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:91.0) Gecko/20100101

Thunderbird/91.6.0

Subject: **=?UTF-8?B?UmU6lOatr+WtpuS8mkhQ56K66KqN55So44Gu44OV77+944Kh44Kk44Or?=?**

=?UTF-8?B?56K66KqN44Gu44GK6aGY44GE?=?

「UTF-8」でエンコードされたタイトル

Content-Language: en-US

echo "=?UTF-8?B?UmU6lOatr+WtpuS8mkhQ56K66KqN55So44Gu44OV77+944Kh44Kk44Or?=" | wsl nkf -wm

送信側改行部分

Re: 歯学会HP確認用のフ

受信側改行部分

echo "=?UTF-8?B?56K66KqN44Gu44GK6aGY44GE?=" | wsl nkf -wm

確認のお願い

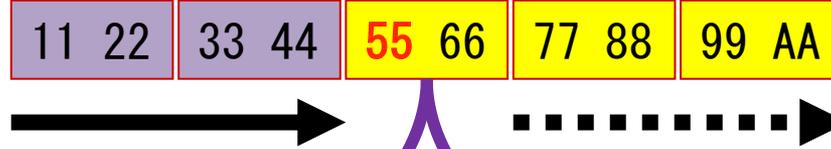
wsl nkf コマンドはWindows10のwsl2 + Ubuntuインストール状態にて使用

コンピュータ内部での解釈の問題

Endian: 2バイト単位で入れ替えて評価するかどうか。

⇒ Unicodeではファイルの先頭にBOM (Byte Order Mark) という符号を付けることで解決

外部記憶・ファイル



CPU

データ取出

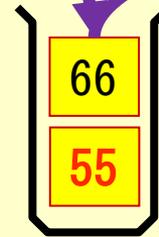
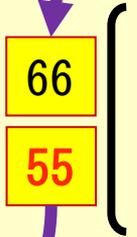
データ読込

データ読込

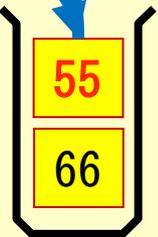
データ取出



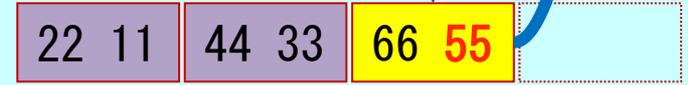
バッファ



スタック



内部処理時



Big Endian

FIFO (First-In, First-Out)
モトローラ系 (旧アップル)

処理段階での
解釈が異なる

Little Endian

FILO (First-In, Last-Out)
インテル系

大学ネットワークの基盤 について

新潟大学が利用している ネットワークの基盤 NINESとSINETについて

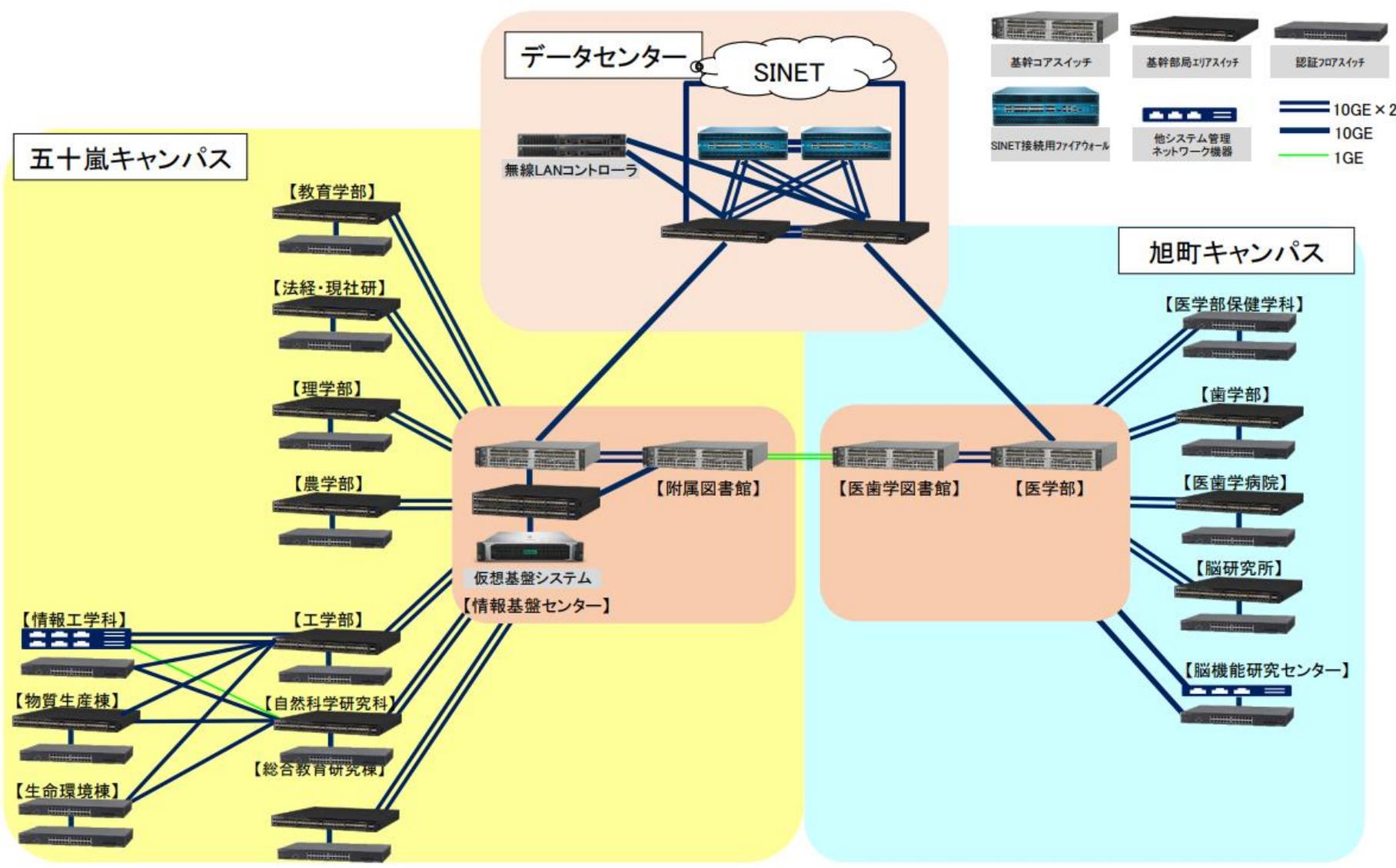
- NINES

- Niigata university Information Network System
- <https://www.cais.niigata-u.ac.jp/system/network/>

- SINET

- Science Information NETwork
- <https://www.sinet.ad.jp>

NINES5全体構成図



SINET6で広がる サービス

大学・研究機関等におけるセキュアで先端的な研究環境を構築するため、超高速ネットワークに加えて、仮想大学LANやL2オンデマンド等、利用者の立場に立ったサービスメニューを用意しています。

2022年
モバイル
実証

提案書
提出期限

学術情報ネットワークSINETは、日本全国の大学、研究機関等の学術情報基盤として、国立情報学研究所(NII)が構築、運用している情報通信ネットワークです。2022年4月よりSINET6の本格運用を開始しました。



新規加入をご検討の機関様



ご利用中の機関の皆様



クラウド提供者の皆様

最新の投稿

2022/06/08
標的型メール等に対する注意喚起

2022/05/20
SINET6開通式典 開催のお知らせ

2022/05/06
NII学術情報基盤オープンフォーラム2022（オンライン開催）

2022/04/11
SINET5加入機関情報ポータルサービス終了について

2022/04/08
pingチェックサイトに関するご連絡

カテゴリー

ニュース

お知らせ

学術情報ネットワークは、日本全国の大学、研究機関等の学術情報基盤として、国立情報学研究所(NII)が構築、運用している情報通信ネットワークです。大学、研究機関等に対して先進的なネットワークを提供するとともに、多くの海外研究ネットワークと相互接続しています。

学術情報ネットワーク(SINET[®])は、日本全国の大学、研究機関等の学術情報基盤として、国立情報学研究所(NII)が構築、運用している情報通信ネットワークです。

※SINET: Science Information NETwork (サイネット)

教育・研究に携わる数多くの人々のコミュニティ形成を支援し、多岐にわたる学術情報の流通促進を図るため、全国にノード(ネットワークの接続拠点)を設置し、大学、研究機関等に対して先進的なネットワークを提供しています。また、国際的な先端研究プロジェクトで必要とされる国際間の研究情報流通を円滑に進められるよう、米国Internet2や欧州GEANTをはじめとする、多くの海外研究ネットワークと相互接続しています。

2022年4月からは、従来の学術情報基盤であるSINET5を発展させたSINET6の本格運用を開始しました。世界最高水準の400Gbps回線ネットワークで有機的につなぎ、約1000機関に及ぶ大学・研究機関等にハイレベルな学術情報基盤を提供します。

SINET6のDCと接続形態



最新の投稿

2022/06/08

標的型メール等に対する注意喚起

2022/05/20

SINET6開通式典 開催のお知らせ

2022/05/06

NII学術情報基盤オープンフォーラム2022 (オンライン開催)

2022/04/11

SINET5加入機関情報ポータルサービス終了について

2022/04/08

pingチェックサイトに関するご連絡

カテゴリー

■ ニュース

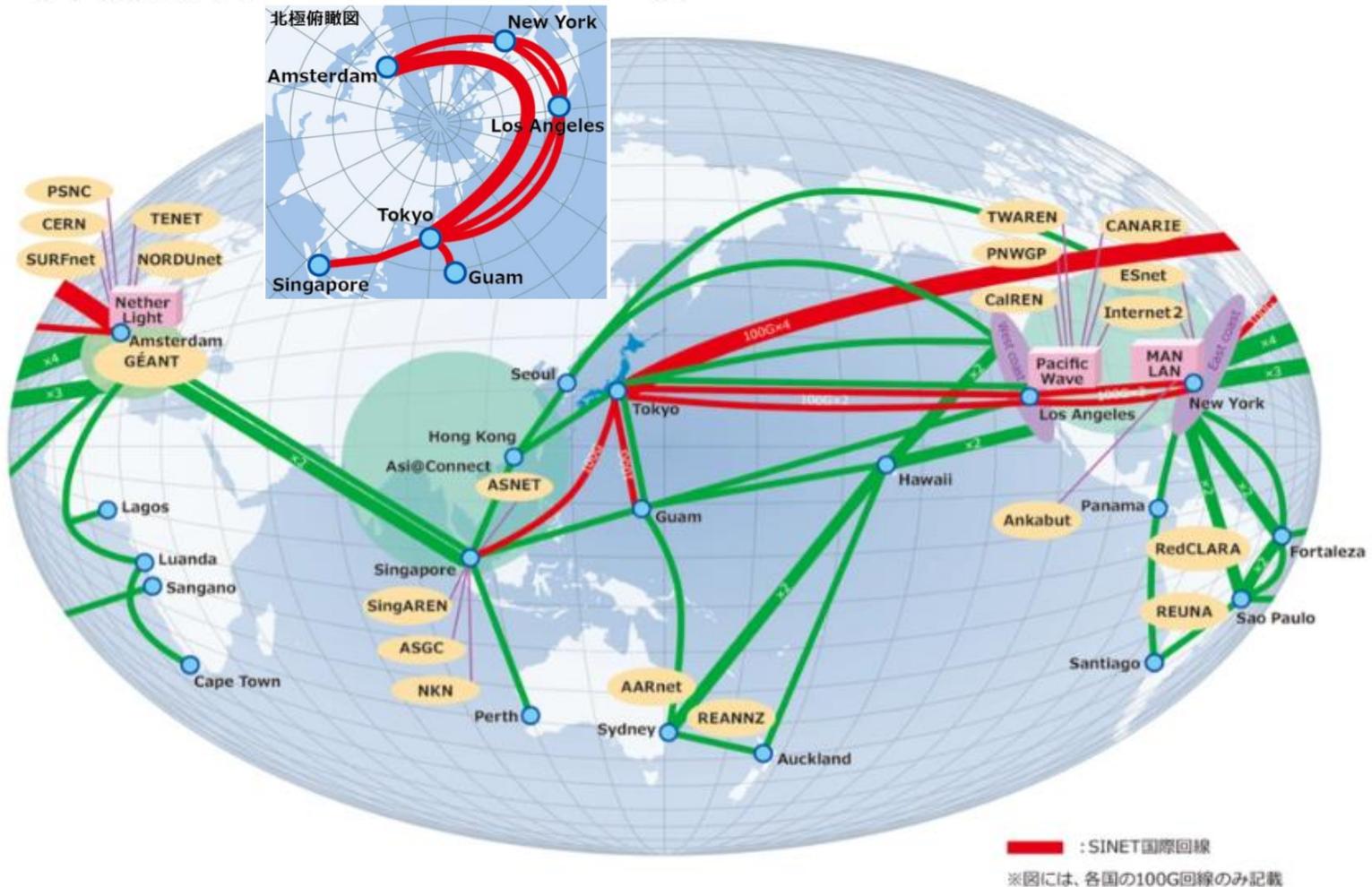
■ 活用事例

タグ

- ◆ APOnet (1) ・ ◆ BGP (1) ・ ◆ IP Dual (19) ・
- ◆ Liオンデマンド (6) ・ ◆ L2VPN (24) ・ ◆ L3VPN (7) ・
- ◆ MMCFTP (1) ・ ◆ QoS (1) ・ ◆ SINET6 (1) ・
- ◆ クラウド (9) ・ ◆ フルフレート (1) ・
- ◆ マルチホーム接続 (1) ・ ◆ 仮想大学LAN (2) ・
- ◆ 国際接続 (9) ・ ◆ 新型コロナ (4) ・
- ◆ 遠隔バックアップ (4) ・ ◆ オープンフォーラム2022 (1) ・
- ◆ SINET6開通式 (1)

国際的な先端研究プロジェクトで必要とされる国際間の研究情報流通を円滑に進められるように、海外の学術情報ネットワークとの通信を確保しています。米国Internet2や欧州GÉANT等をはじめとした多くの海外研究ネットワークと相互接続しています。

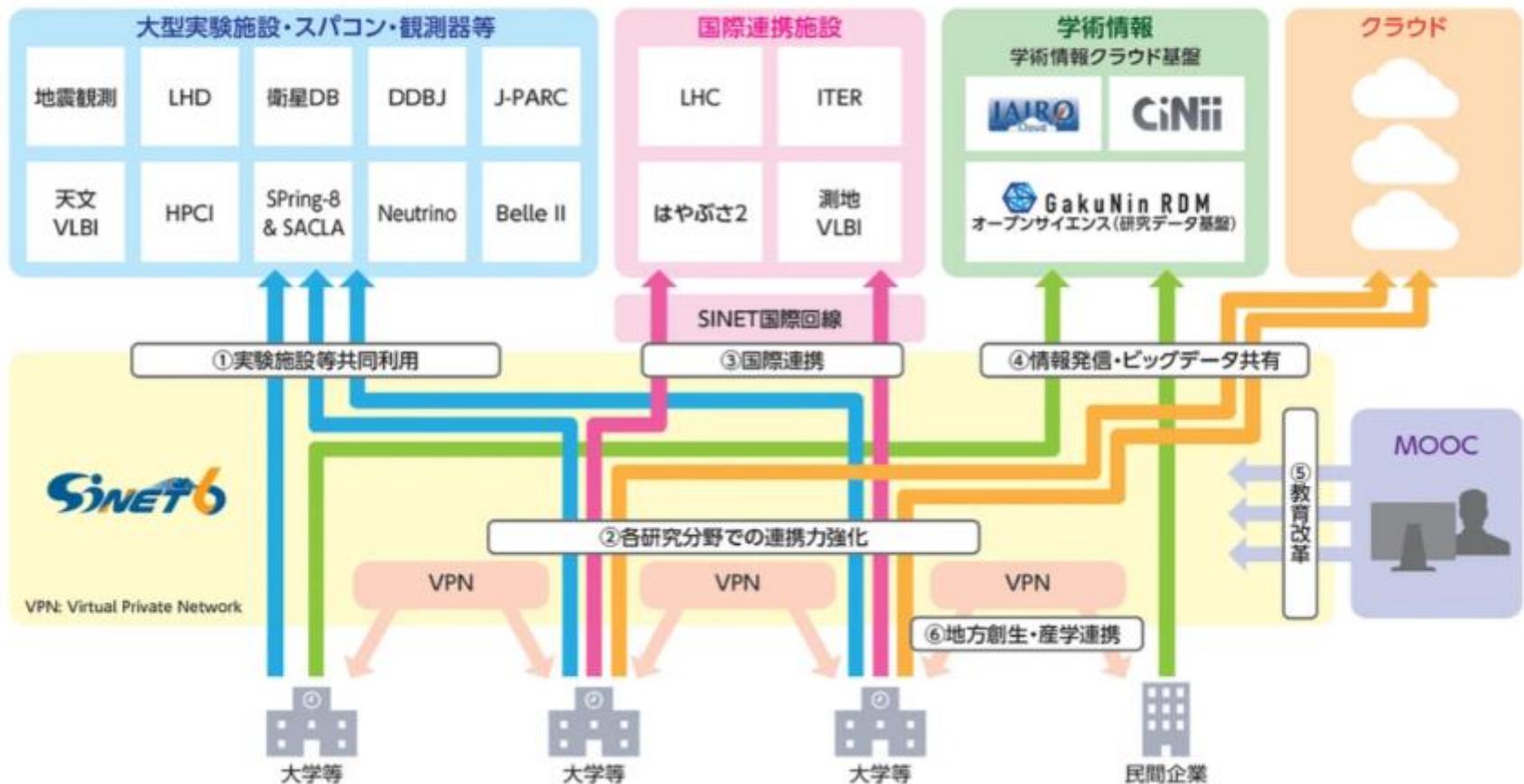
そして、アジア地域との学術研究情報の流通促進を強化するために、Asi@Connectと連携・協力し、国際学術情報ネットワークの構築に貢献しています。



● 研究データ基盤と一体化

SINETを介したクラウド、セキュリティ、学術コンテンツ、研究データの安全な利用を促進し、高機能な研究教育環境の構築を支援。

あらゆる学術分野に寄与しうるオープンサイエンス発展のため、SINET と研究データ基盤の一体化を推進します。



暗号と認証について

暗号と認証

- エンコード(コード化・符号化)とデコード(復号化)
 - エンコード(コード化)操作は、アナログ情報をデジタル化したりデジタル情報を更に別のデジタル情報に変換(圧縮や暗号化)する時に使われる。



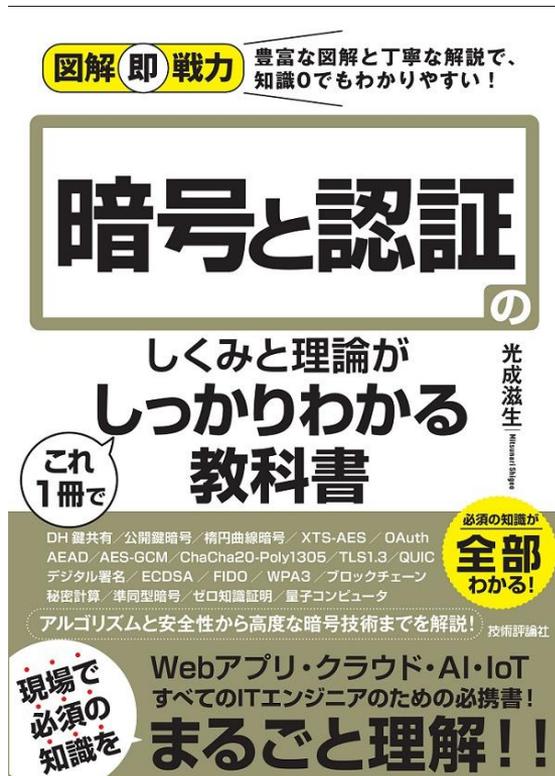
- エンコードされたデータは変換方法とセットでなければデコードできない。
- 変換方法を決定する情報(鍵)を当事者以外の部外者が使えない(分からない)状況にすることで暗号になる。
- 暗号化されたデータに正当にアクセスするための本人確認が認証。
- 認証され、鍵があれば、暗号は破られる。

「暗号と認証」おすすめの本 図書館にあります

- 図解即戦力シリーズ

- 図解即戦力「暗号と認証」

暗号と認証のしくみと理論がこれ1冊でしっかりわかる教科書



著者の関連情報サイト

<https://herumi.github.io/anninbon/>

日本の「暗号技術評価委員会」サイト

- CRYPTREC概要 >
- 注意喚起 >
- CRYPTREC暗号 >
- CRYPTREC報告書 >
- ガイドライン >
- 技術報告書 >
- 会議資料 >
- イベント >
- 関連機関等のご案内



誰もが安心して利用できる高度情報通信ネットワークには、
確かな安全性と信頼性を持った暗号技術が不可欠です。

CRYPTRECは、セキュアなIT社会の実現に向けて活動しています。

📄 新着情報

量子コンピュータの発達に関し暗号技術の 安全性についての見解



CRYPTREC概要

注意喚起

【最新】注意喚起情報

注意喚起一覧

CRYPTREC暗号

CRYPTREC報告書

ガイドライン

技術報告書

注意喚起情報

CRYPTREC ER-0001-2019

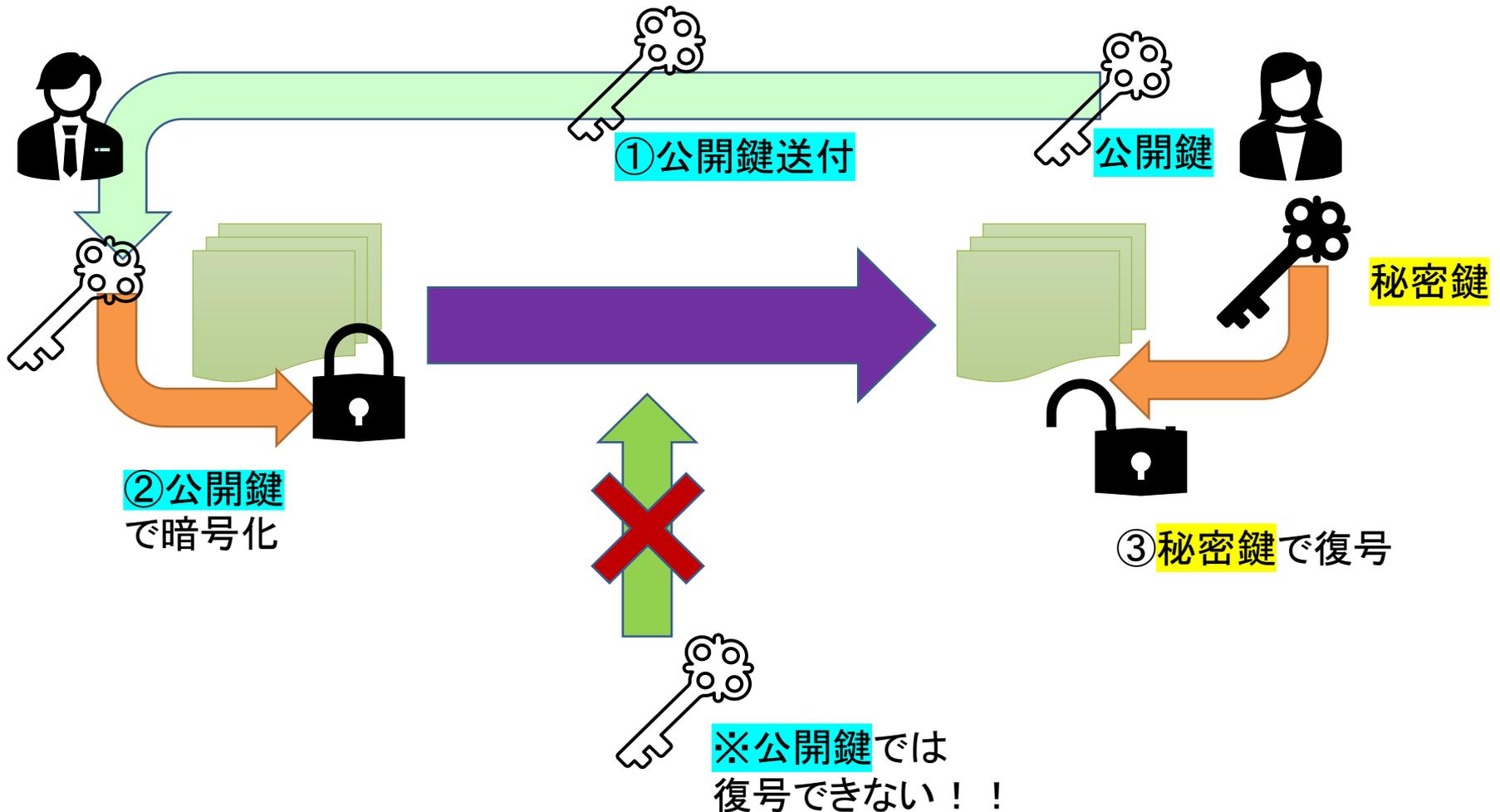
現在の量子コンピュータによる暗号技術の安全性への影響

2020年(令和2年)2月17日
CRYPTREC 暗号技術評価委員会

今般、ゲート型の量子コンピュータが量子超越を実現したという報告があり、暗号技術の危殆化が一部で懸念されております。しかし、現在の量子コンピュータの開発状況をふまえると、暗号解読には規模の拡大だけでなく量子誤り訂正などの実現が必要であるため、CRYPTRECとしては、CRYPTREC暗号リスト記載の暗号技術が近い将来に危殆化する可能性は低いと考えています。今後も、本暗号リスト記載の暗号技術の監視活動を引き続き実施していきます。



公開鍵と秘密鍵を使う 公開鍵暗号方式




```
-----BEGIN OPENSSH PRIVATE KEY-----
...
12      lqb7q:  ...      ZAewF  ...      TLB
13      :cmk2:  ...      B1DfT  ...      i28/
...
50-----END OPENSSH PRIVATE KEY-----
EOF
```

```
ssh-rsa
...
iXC060hX
809tuecP
...
LyWw
j5Emj
...
== email@example.test ←
EOF
```

公開鍵の例
相手に渡す経路についてはセキュリティ上、安全な方法を使うこと。

秘密鍵の例
厳重に管理する必要あり。
決して外部には漏らさない。

AIとは？ 深層学習、機械学習とは？

第一次および第三次AIブーム

いわゆる人工知能(AI)の歴史

人工知能の置かれた状況

主な技術等

人工知能に関する出来事

1950年代			チューリングテストの提唱 (1950年)
1960年代	第一次人工知能ブーム (探索と推論)	<ul style="list-style-type: none"> 探索、推論 自然言語処理 ニューラルネットワーク 	ダートマス会議にて「人工知能」という言葉が登場 (1956年) ニューラルネットワークのパーセプトロン開発 (1958年) 人工対話システムELIZA開発 (1964年)
1970年代	冬の時代	<ul style="list-style-type: none"> エキスパートシステム 遺伝的アルゴリズム 	初のエキスパートシステムMYCIN開発 (1972年) MYCINの知識表現と推論を一般化したEMYCIN開発 (1979年)
1980年代	第二次人工知能ブーム (知識表現)	<ul style="list-style-type: none"> 知識ベース 音声認識 	第五世代コンピュータプロジェクト (1982~92年) 知識記述のサイクプロジェクト開始 (1984年)
1990年代	冬の時代	<ul style="list-style-type: none"> データマイニング オントロジー 	誤差逆伝播法の発表 (1986年)
2000年代	第三次人工知能ブーム (機械学習)	<ul style="list-style-type: none"> 統計的自然言語処理 ディープラーニング 	ディープラーニングの提唱 (2006年)
2010年代			ディープラーニング技術を画像認識コンテストに適用 (2012年)

AI

論理的でないものや記号化困難な対象がある。対象範囲の設定が困難な場合がある。

結果に至る過程について人間が理解困難な場合や、過程内部で論理性が欠落する危険性がある。

いわゆるSymbolism

過程優先、演繹的
シンボル・記号が基本
論理構造重視
原点：論理プログラミング
知識ベース、述語論理

第二次AIブームの
主役

エキスパートシステム
プロダクションシステム

いわゆるConnectionism

結果優先、発見的
ネットワーク結合が基本
パターン・類似点・特徴量重視
原点：パーセプトロン
ニューラルネットワーク

第一次、第三次
AIブームの主役

機械学習

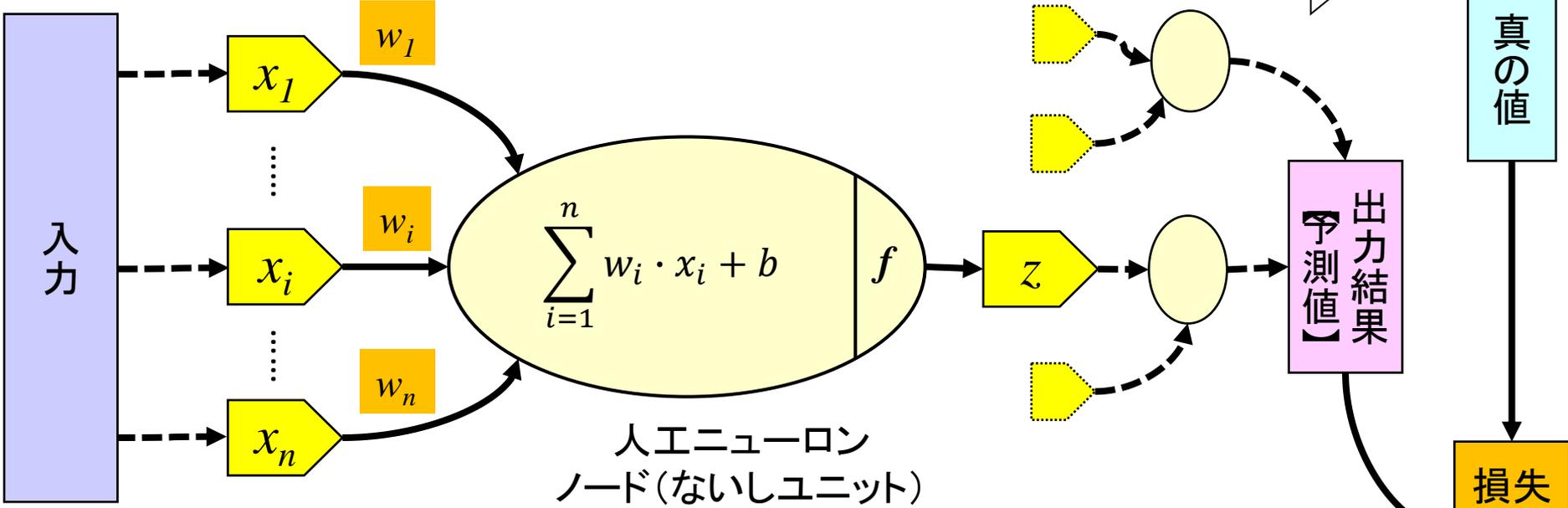
深層学習

Transformer系
CNN系

Hybrid Type ⇒ 「強いAI※」？

※人間のような自意識や知能を備え高度な判断が可能なAI
AGI (Artificial General Intelligence、汎用人工知能)とも呼ばれる

入力側から出力側へ(フィードフォワード; FF)



... > $\Sigma(\text{入力} \times \text{重み}) + \text{バイアス} \Rightarrow \text{活性化関数} \Rightarrow \text{出力} \dots >$

x_i w_i b f z

通常x, w, b, zの値は、0から1の範囲で正規化される

誤差逆伝播法 (Backpropagation)
損失関数: $E = \frac{1}{M} \sum_{i=1}^M (\text{真の値}_i - \text{予測値}_i)^2$
を最小にするようにパラメータを調整する

出力側から入力側へ(フィードバック; FB)

深層学習・機械学習

学習用データの選別によるバイアス

例えば、診断名
あるいは、動物名

入力
データ

● (変換)モデル
● パラメータ

予測値

人間が規則性を見つけ、モデルを作る。

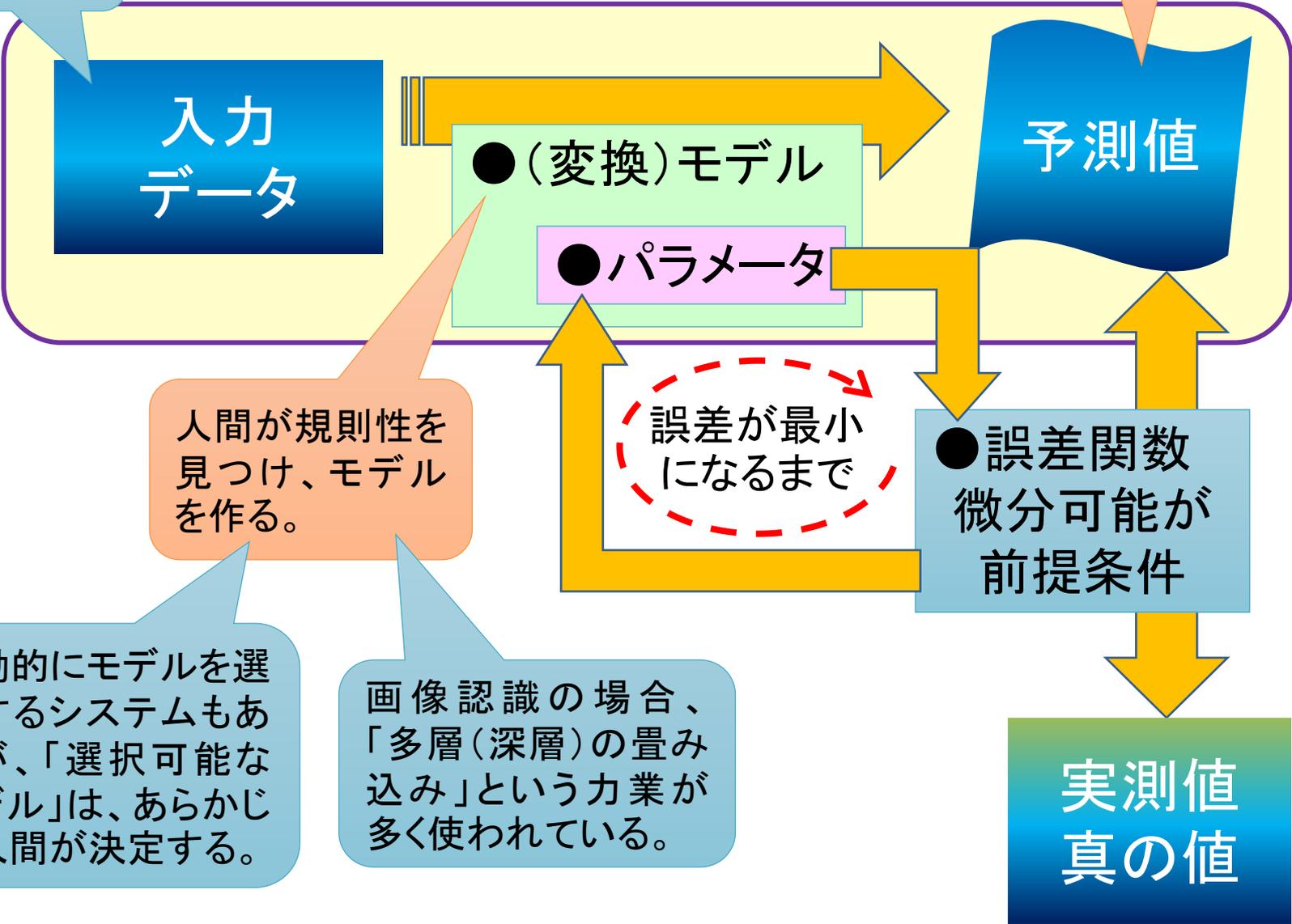
誤差が最小になるまで

● 誤差関数
微分可能が
前提条件

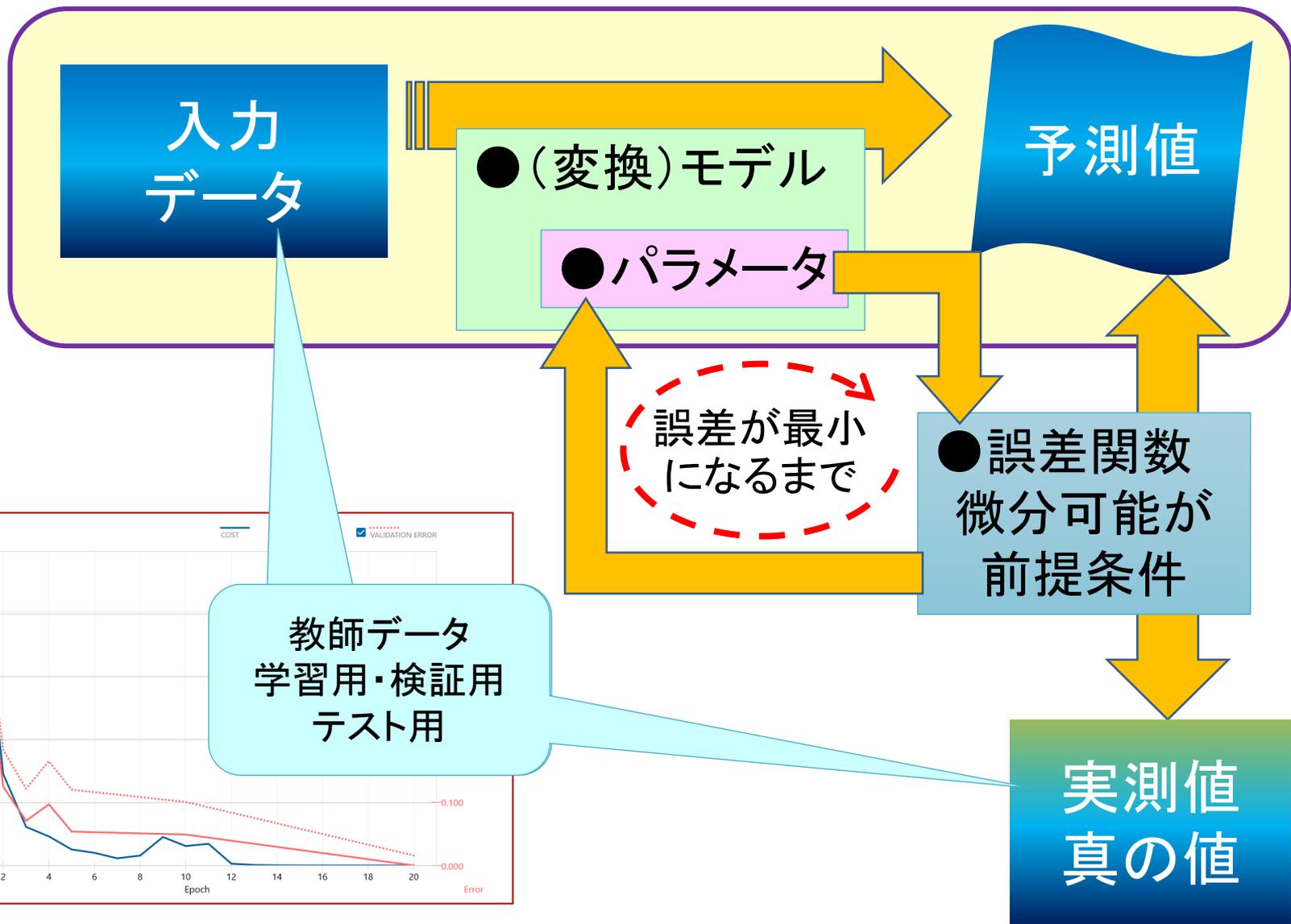
自動的にモデルを選択するシステムもあるが、「選択可能なモデル」は、あらかじめ人間が決定する。

画像認識の場合、「多層(深層)の畳み込み」という作業が多く使われている。

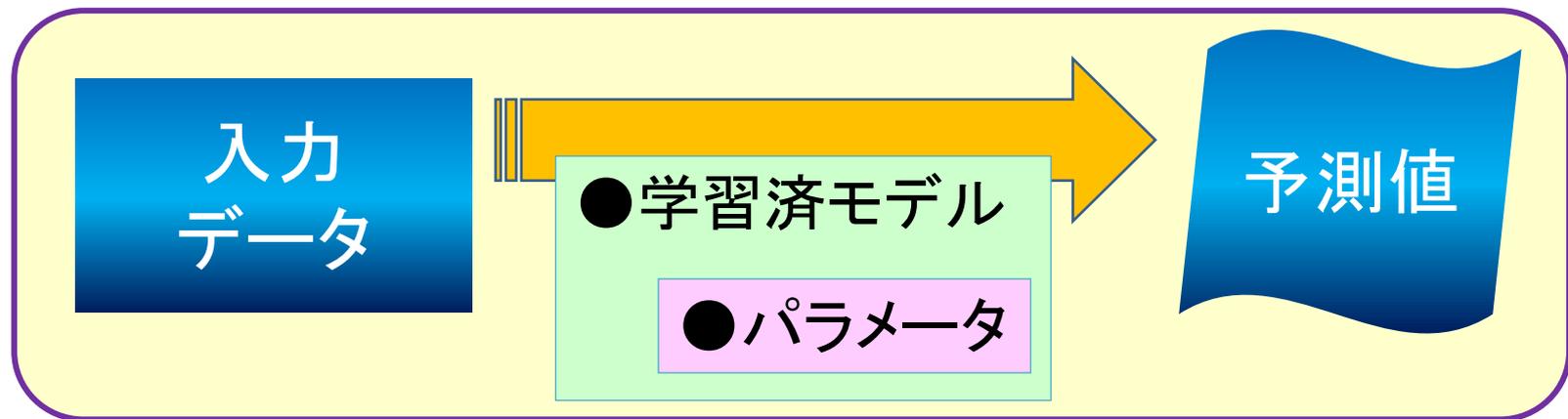
実測値
真の値



深層学習・機械学習の訓練中...



深層学習・機械学習の稼働時



犬と猫しか学習していない場合、猿をいれても「犬」か「猫」のどちらか一方の回答を出力する。「学習対象」以外の判定を行わせるための「学習」についても研究がなされているが、通常は困難。
⇒ChatGPTが「知らないこと」を「知っているかのごとく生成する」のも同じ理由。

手書きの「0」から「9」までを認識する深層学習システム

- MNISTと呼ばれる有名な手書きの数字データベースを使い深層学習させたシステムを準備する。
- 数字以外の手書き文字を入力した場合どのような回答するのか？
 1. 「分からない」と答える
 2. 自信なさげに、似た数字を提示する
 3. 自信満々に、似た数字を提示する

深層学習システムは学習したこと以外は回答できないし
学習した範囲で答えようとする。

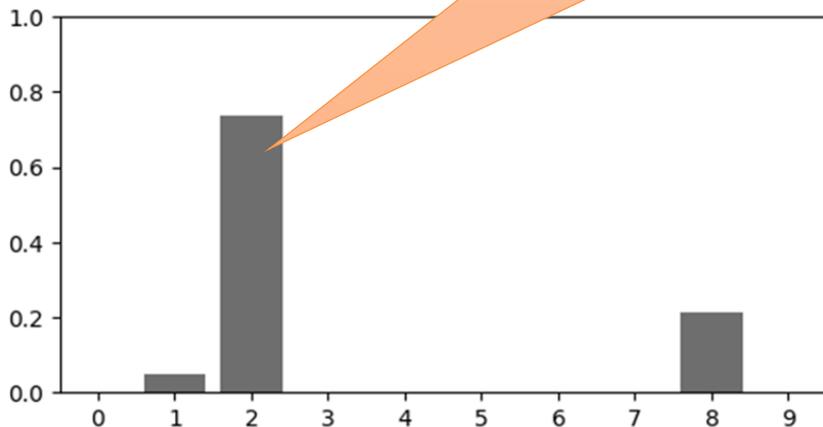
```
HTML(input_form + javascript)
# After appearing the Box for drawing...
# ↓ ↓ ↓ Please Draw number with pointing device such as mouse. Also Do the call below
```

Out [8]:



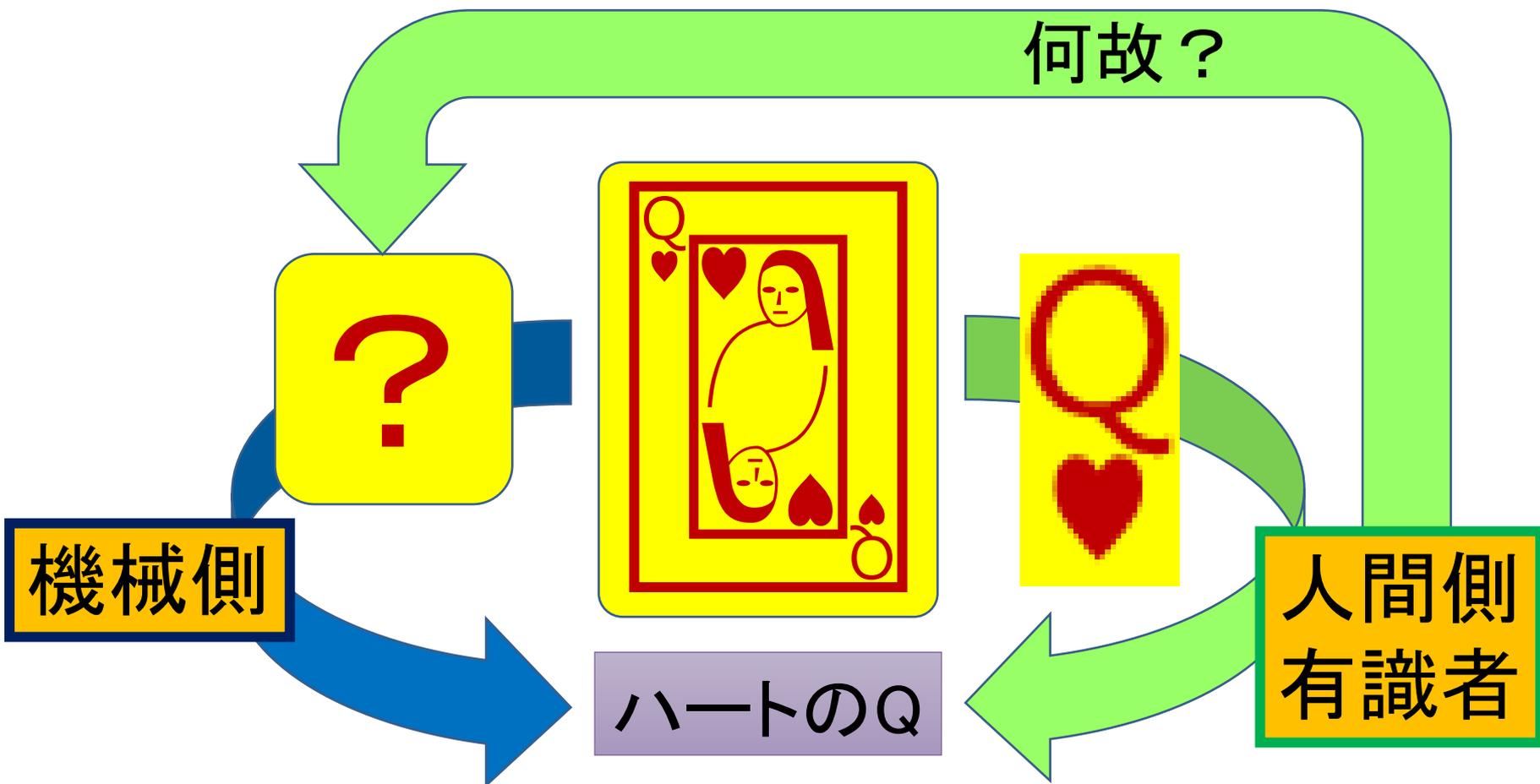
「2」の可能性が80%近いと回答
「分からない」とは答えないし、
「χ」とは答えない。

Out [10]: <BarContainer object of 10 artists>



問題の本質について(1)

- 深層学習では、機械側の判断根拠を知ることが困難とされている。



問題の本質について(2)

たとえ99%の精度
を持つシステムで
あっても、

- 診断支援ないし教育支援として用いる場合、診断利用者ないし教育介助者には該当領域の有識者レベルの知識で介入が必要不可欠。

ハートのJ,Q,K
だよ

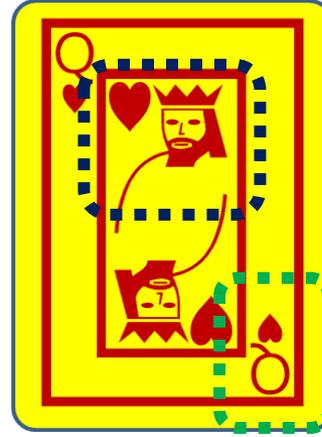
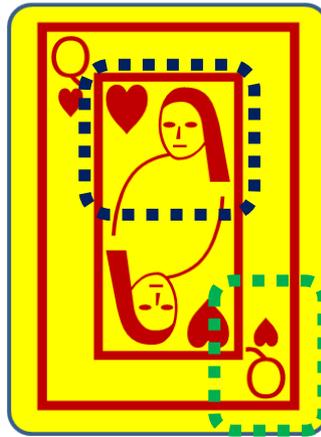
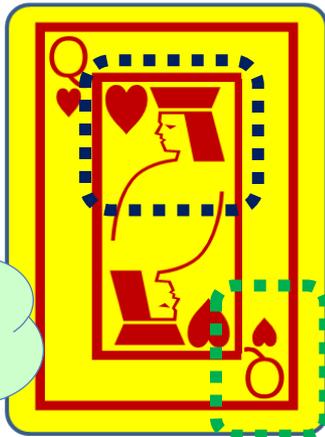
人間側
初学者

介入

違うよ！ 全て
ハートのQ

機械側

若い男性
女性
年上の男性



人間側
有識者

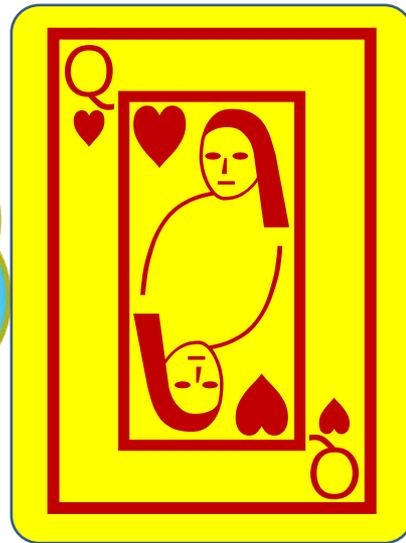
トランプも
ジェンダー
フリーか

問題の本質について(3)

- 【問題の本質】機械側の判断根拠が人間側の判断根拠と同じかどうか不明であること。

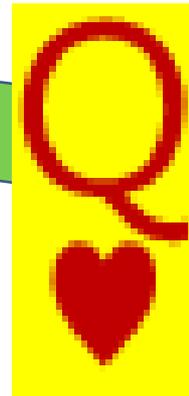
もしかしたら、画期的な知見かも知れない。

機械側

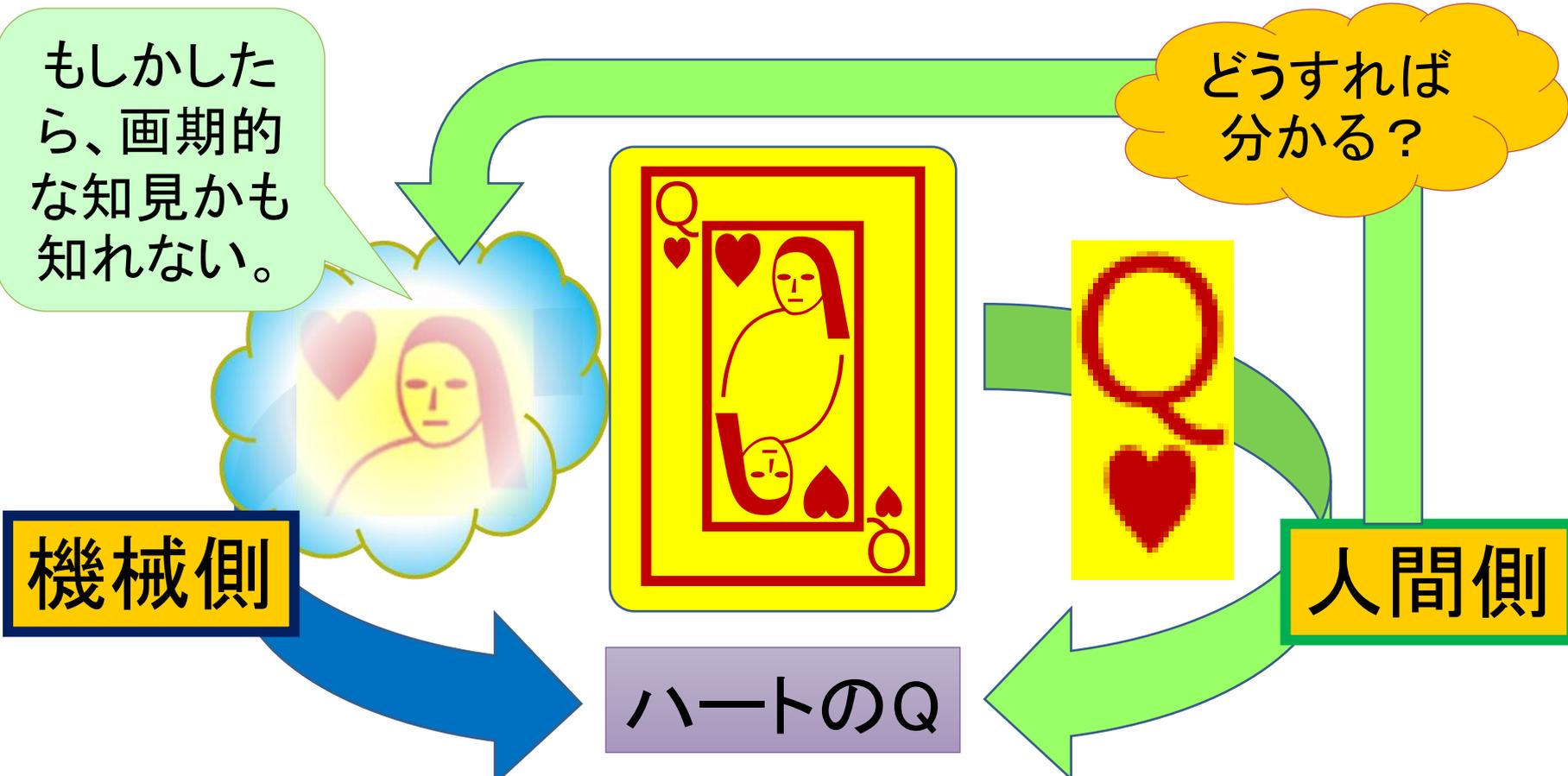


ハートのQ

どうすれば分かる？



人間側



自動運転関連での進歩と問題

- 日本初、「レベル4」自動運転車を認定 条件満たせば運転手不在でOK
 - <https://www.itmedia.co.jp/news/articles/2303/31/news236.html>
- 自動運転車の事故(2023年版)
 - 事故当時の状況や原因を紐解く
 - https://jidounten-lab.com/y_1615

機械側は、機械側として正しく処理している。

人間側は、人間側として、機械側が間違った処理をしたと判断している。

事故が発生した場合、人間側が理解可能な理由・説明が得られない。

→ ブラックボックス問題

→ 記号接地問題

遡って 第二次AIブームでは・・・

ざっくりと言えば、人間側の論理的な思考から逸脱しない範疇での人工知能を作ろうとしていた。

AI

論理的でないものや記号化困難な対象がある。対象範囲の設定が困難な場合がある。

結果に至る過程について人間が理解困難な場合や、過程内部で論理性が欠落する危険性がある。

いわゆるSymbolism

過程優先、演繹的

シンボル・記号が基本

論理構造重視

原点：論理プログラミング

知識ベース、述語論理

第二次AIブームの
主役

エキスパートシステム
プロダクションシステム

いわゆるConnectionism

結果優先、発見的

ネットワーク結合が基本

パターン・類似点・特徴量重視

原点：パーセプトロン

ニューラルネットワーク

第一次、第三次
AIブームの主役

機械学習

深層学習

Transformer系
CNN系

Hybrid Type ⇒ 「強いAI※」？

※人間のような自意識や知能を備え高度な判断が可能なAI
AGI (Artificial General Intelligence、汎用人工知能)とも呼ばれる

Prologの例

論理的という面で理解可能

```
25  
26 uncle_or_aunt_nephew_or_niece(namihei, norisuke).  
27 uncle_or_aunt_nephew_or_niece(taizou, wakame).
```

事実
(データ相当)

```
28  
29 uncle_or_aunt_nephew_or_niece(X,Y):-  
30     parents_child(L,Y),  
31     member(Z,L),  
32     brothers_or_sisters(X,Z).  
33
```

規則
(形式知相当)



SWI-Prolog (AMD64, Multi-threaded, version 8.2.4)

File Edit Settings Run Debug Help

```
Welcome to SWI-Prolog (threaded, 64 bits, version 8.2.4)  
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This is free software.  
Please run ?- license. for legal details.
```

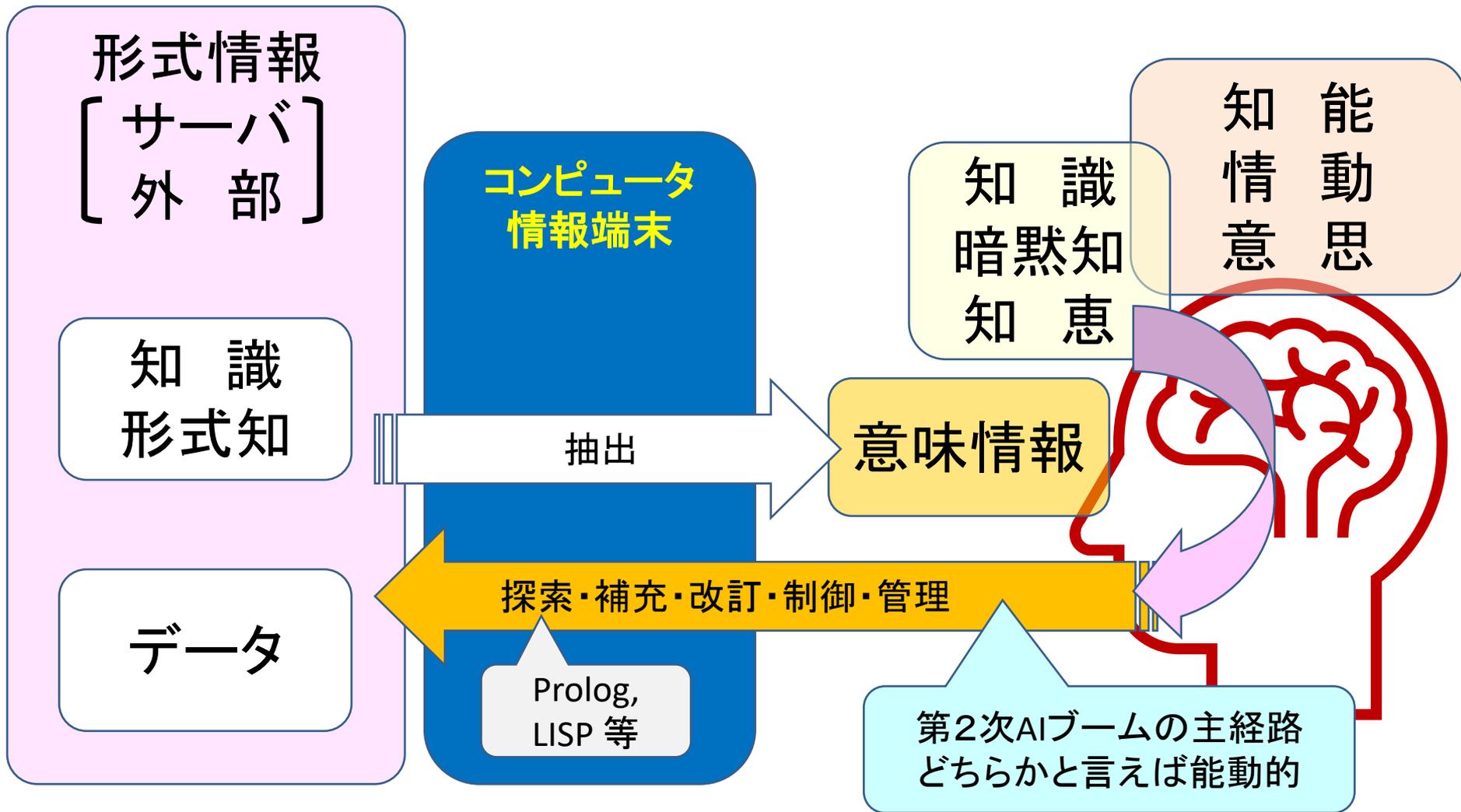
```
For online help and background, visit https://www.swi-prolog.org  
For built-in help, use ?- help(Topic). or ?- apropos(Word).
```

```
?- ['E:/bin/prolog/family.prolog'].  
true.
```

```
?- uncle_or_aunt_nephew_or_niece(X, tarao).  
X = katsuo ;  
X = wakame ;  
.
```

質問
(探索)

データ、情報、知識・知恵と 第二次AIブームとの関係



30年前 第2次人工知能ブームの後半位

昭和54年5月2日第3種郵便物認可 平成4年2月1日発行(毎月1回1日発行)第18巻 第2号 通巻177号 ISSN 0387-9569

インタ-フェ-ース

Interface '92

2

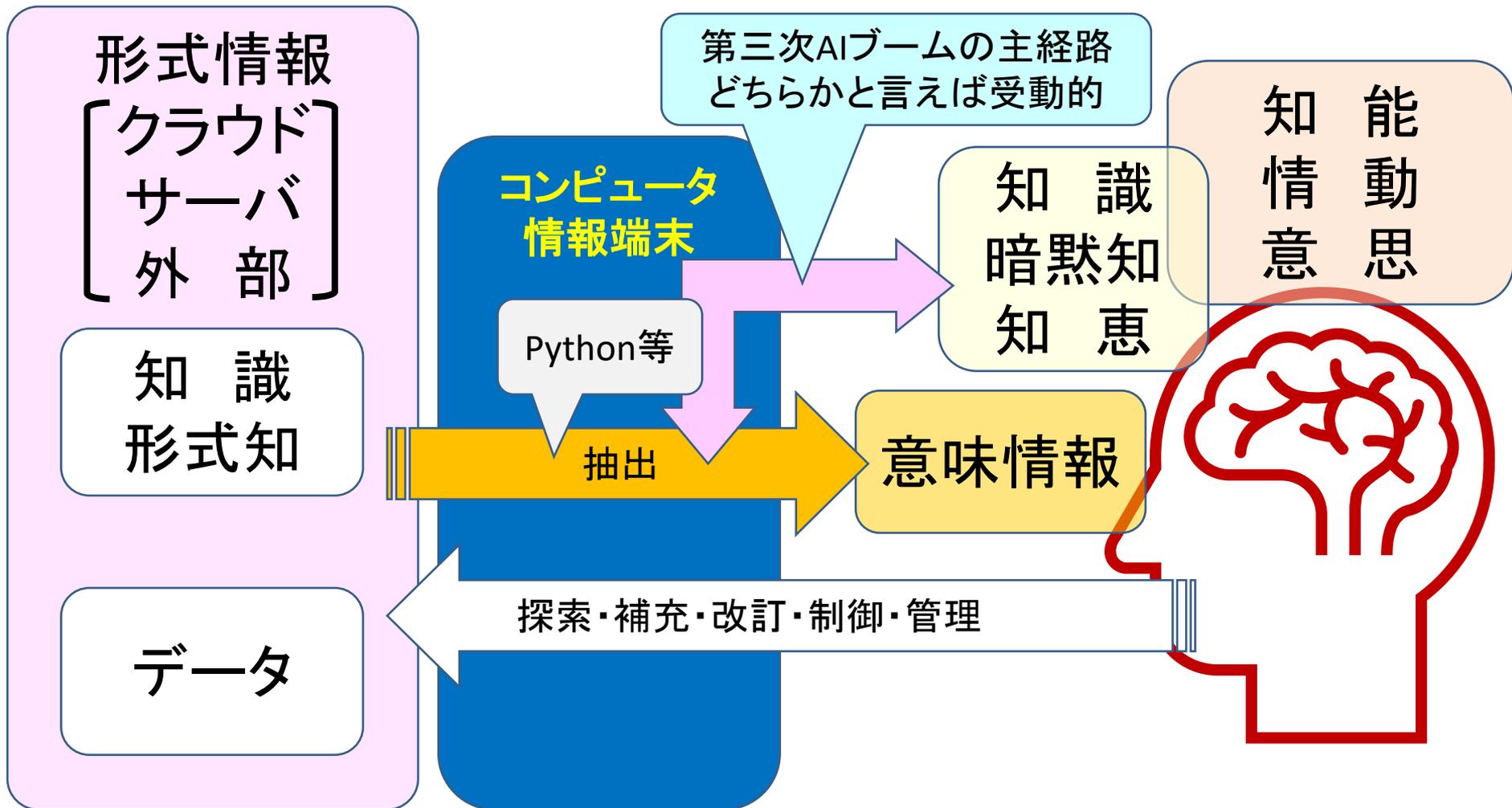
[特集]

ニューロ/遺伝的手法の実験と応用

●人工知能研究と人工生命 ●遺伝的アルゴリズムの実際 ●MLP/BPの定式化からCによるコーディングまで ●ニューラルネットによるパターン認識の実験

特設記事 バッチ・ファイルからEXEファイルに変換する

データ、情報、知識・知恵と 第三次AIブームとの関係



AI

論理的でないものや記号化困難な対象がある。対象範囲の設定が困難な場合がある。

結果に至る過程について人間が理解困難な場合や、過程内部で論理性が欠落する危険性がある。

いわゆるSymbolism

過程優先、演繹的

シンボル・記号が基本

論理構造重視

原点：論理プログラミング

知識ベース、述語論理

第二次AIブームの
主役

エキスパートシステム
プロダクションシステム

いわゆるConnectionism

結果優先、発見的

ネットワーク結合が基本

パターン・類似点・特徴量重視

原点：パーセプトロン

ニューラルネットワーク

第一次、第三次
AIブームの主役

機械学習

深層学習

Transformer系
CNN系

Hybrid Type ⇒ 「強いAI※」？

※人間のような自意識や知能を備え高度な判断が可能なAI
AGI (Artificial General Intelligence、汎用人工知能)とも呼ばれる

論理回路で構成されるコンピュータ

いずれにしても、現行のAIは、論理プログラミングの範疇にあって、そこからの逸脱はありえない。

論理プログラミング

エキスパートシステム
ファジィコンピュータ
遺伝的アルゴリズム
ボルツマンマシン(アニーリング)

数学、統計ソフト
画像処理ソフト
オフィスソフト
等々

ニューラルネット

深層学習システム

LLM (Large Language Model)
形式情報(記号列)としての言葉を計算対象として扱うシステム。

GPT (Generative Pre-trained Transformer)

<https://www.pmda.go.jp/files/000227450.pdf>

使用者責任 <https://www.mhlw.go.jp/content/10601000/000468150.pdf>

- 厚生労働省医政局医事課長通知(H30.12.19)
 - 「人工知能(AI)を用いた診断・治療支援を行うプログラムを利用して診療を行う場合についても、診断、治療等を行う主体は医師であり、**医師はその最終的な判断の責任を負う**こととなり、当該診療は医師法(昭和23年法律第201号)第17条の医業としておこなわれるものであるので、十分ご留意をいただきたい。」

画像診断におけるAIと 責任の所在・総務省・説明責任

- AIネットワーク社会推進会議 報告書2019公表
http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000081.html
- AI活用ガイドライン(R1.8.9)
http://www.soumu.go.jp/main_content/000637097.pdf

AI活用原則(10の原則)

- | | |
|--------------|------------------|
| 1) 適正利用の原則 | 6) プライバシーの原則 |
| 2) 適正学習の原則 | 7) 尊厳・自律の原則 |
| 3) 連携の原則 | 8) 公平性の原則 |
| 4) 安全の原則 | 9) 透明性の原則 |
| 5) セキュリティの原則 | 10) アカウンタビリティの原則 |

適正利用の原則

[イ 人間の判断の介在]

- AI サービスプロバイダ及びビジネス利用者は、AI によりなされた判断について、必要かつ可能な場合には、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介在させることが期待される。その場合、人間の判断の介在の要否について、基準例を踏まえ、利用する分野やその用途等に応じて検討することが期待される。
- また、AI によりなされた判断について人間が最終判断をすることが適切とされている場合に、人間が AI と異なる判断をすることが期待できなくなることも想定されることから、説明可能性を有する AI から得られる説明を前提として、人間が判断すべき項目を事前に明確化しておくこと等により、人間の判断の実効性を確保することが期待される²⁰。

統括する家庭内執事ロボットや周辺の AI を含む家電が当該アップデートに適応していないと、（家電同士、または家電とロボットの）相互の判断に齟齬が生じる（「報告書 2018」別紙 3 「AI が想定外の動作を行うなどのおそれ」の事例）。

²⁰ 加えて、人間が確認する AI の判断の適正性を確保するため、他の AI を利用したダブルチェック、AI への入力を摂動させることによる AI 動作の確認などの措置を検討することが望ましい。

厚生労働省・薬生機審発0523第2号

<https://www.mhlw.go.jp/content/10601000/000515843.pdf>

mhlw.go.jp/content/

その目的が「診断支援」であるにもかかわらず、医師が**最終診断まで委ねてしまう危険性**を含んでいる。

(3) 責任の所在

人工知能を利用した医療機器においては、その使用に伴う責任の所在が曖昧になることについての懸念が議論されている。特に、支援システム等においては、その目的が「診断支援」であるにもかかわらず、医師が最終診断まで委ねてしまう危険性を含んでいる。よって、支援システムの保守、設計・仕様上の不具合や故障等のトラブル対応においては、通常の医療機器と同様、製造販売業者が責任を負うことは変わらないが、支援システム等の使用目的、使用方法等を使用者に対して明らかにし、使用者に対するトレーニング等を実施する等、適正使用のために必要な方策を検討し、実施の実効性を担保する必要がある。

(6 (4) 項を参照)。なお、平成 30 年 12 月 19 日付け医政医発 1219 第 1 号厚生労働省医政局医事課長通知「人工知能 (AI) を用いた診断、治療等の支援を行うプログラムの利用と医師法第 17 条の規定との関係について」において、当該プログラムを利用して診断、治療を行う主体は医師であり、医師がその最終的な判断の責任を負うことが示されていることも留意されたい。

説明責任 + 使用者責任

OECDの人工知能に関する理事会勧告(2024/05更新版)

価値観に基づく5つの原則

政策担当者への勧告



包括的な成長、持続可能な開発、福祉



AIの研究開発への投資



人権と民主的価値観(公平性とプライバシーを含む)



包括的な AI 対応エコシステムの育成



透明性と説明可能性



AIの相互運用を可能にするガバナンスとポリシー環境の形成



堅牢性、安心・安全性



人間の能力を構築し、労働市場の移行に備える



説明責任



信頼できるAIのための国際協力

OECD AI Principles overview (2024/05更新版)の日本語訳

<https://oecd.ai/en/ai-principles>

総務省には2019年の旧版の和訳あり https://www.soumu.go.jp/main_content/000642217.pdf

生成系AI関連

生成系での最近のトレンド (既に古くなりつつある資料)

<https://ja.stateofaiguide.com/20221231-ai-trends-2022/>

- 拡散モデル、強化学習を主流とした生成系

- 画像生成系

- Stable Diffusion、DALL-E

- 文書生成系

- ChatGPT



<https://atmarkit.itmedia.co.jp/ait/articles/2301/13/news023.html>

- 音声生成系

- VALL-E



<https://www.itmedia.co.jp/news/articles/2301/10/news087.html>

- 利用には注意が必要

- バイアス、法的問題、倫理的問題等がある

<https://dentsu-ho.com/articles/8322>

https://www.itmedia.co.jp/news/articles/2301/28/news056.html#utm_term=share_sp



【学生の皆さんへ】生成系AI（ChatGPT等）利用に関する注意事項について

2023年05月17日 水曜日

お知らせ

🔗 シェアする

🐦 ツイート

生成系AI（ChatGPT等）の利用に際し、本学では、当面の間は大学教育における生成系AI使用について一律の規制を原則として行いませんが、以下の事項に注意してください。

注意事項

1. 授業担当教員の判断により生成系AIの利用を認めない場合、教員の指示に従わず生成系AIを用いてレポート等課題を作成した場合は剽窃とみなされる場合があるので注意すること。
2. 授業に関する情報収集等で利用する場合は、生成系AIの出力には誤りがある可能性に留意し、生成系AIの出力が正しいか、誤りなのか、学生自身でその真贋を確認すること。
3. 未発表の研究成果や個人情報・プライバシー情報等の機密情報を生成系AIに入力してしまうことにより、それら機密情報が意図せず流出・漏えいし、社会的な責任が問われる可能性がある点に注意すること。
4. 画像・動画・音楽等を生成するAIは、インターネット上のコンテンツをAIが学習して画像を生成しているため、利用者が知らないうちに著作権を侵害してしまう可能性があることに注意すること。

本件に関するお問い合わせ先

学務部教務課

電話 025-262-6303

【学生の皆さんへ】生成系AI(ChatGPT等)利用に関する注意事項について

- 生成系AI(ChatGPT等)の利用に際し、本学では、当面の間は大学教育における生成系AI使用について一律の規制を原則として行いませんが、以下の事項に注意してください。
- 注意事項
 - 授業担当教員の判断により生成系AIの利用を認めない場合、**教員の指示に従わず生成系AIを用いてレポート等課題を作成した場合は剽窃とみなされる場合がある**ので注意すること。
 - 授業に関する情報収集等で利用する場合は、**生成系AIの出力には誤りがある可能性に留意し、生成系AIの出力が正しいか、誤りなのか、学生自身でその真贋を確認する**こと。
 - **未発表の研究成果や個人情報・プライバシー情報等の機密情報を生成系AIに入力してしまうことにより、それら機密情報が意図せず流出・漏えいし、社会的な責任が問われる可能性がある**点に注意すること。
 - **画像・動画・音楽等を生成するAIは、インターネット上のコンテンツをAIが学習して画像を生成しているため、利用者が知らないうちに著作権を侵害してしまう可能性がある**ことに注意すること。

AIと著作権の関係について

https://www8.cao.go.jp/cstp/ai/ai_team/3kai/3kai.html



English



[内閣府ホーム](#) > [内閣府の政策](#) > [科学技術・イノベーション](#) > [AI戦略](#) > [AI戦略チーム](#)
([関係省庁連携](#)) > [AI戦略チーム\(関係省庁連携\)\(第3回\)](#)

AI戦略チーム(関係省庁連携)(第3回)

日時



令和5年5月15日(月) 14:00-14:50

資料

資料

[AIと著作権の関係について\(PDF:380KB\)](#)

AIと著作権の関係等について



基本的な考え方

- 著作権法では、**著作権者の権利・利益の保護と著作物の円滑な利用のバランス**が重要
- 著作権は、「**思想又は感情を創作的に表現した**」著作物を保護するものであり、**単なるデータ(事実)やアイデア(作風・画風など)**は含まれない
- AIと著作権の関係については、「**AI開発・学習段階**」と「**生成・利用段階**」では、**著作権法の適用条文が異なり、分けて考えることが必要**

現状の整理

AI開発・学習段階（著作権法第30条の4[※]）

※平成30年著作権法改正により新たに規定

- 著作物を学習用データとして収集・複製し、学習用データセットを作成
- データセットを学習に利用して、AI（学習済みモデル）を開発
- AI開発のような情報解析等において、著作物に表現された思想又は感情の享受を目的としない利用行為^{※1}は、**原則として著作権者の許諾なく利用することが可能**
 - ※1 例えば、3DCG映像作成のため風景写真から必要な情報を抽出する場合であって、元の風景写真の「表現上の本質的な特徴」を感じ取れるような映像の作成を目的として行う場合は、元の風景写真を享受することも目的に含まれていると考えられることから、このような情報抽出のために著作物を利用する行為は、本条の対象とならないと考えられる
- ただし、「**必要と認められる限度**」を超える場合や「**著作権者の利益を不当に害することとなる場合**^{※2}」は、この規定の対象とはならない。
 - ※2 例えば、情報解析用に販売されているデータベースの著作物をAI学習目的で複製する場合など

生成・利用段階

- AIを利用して画像等を生成
- 生成した画像等をアップロードして公表、生成した画像等の複製物（イラスト集など）を販売
- AIを利用して生成した画像等をアップロードして公表したり、複製物を販売したりする場合の**著作権侵害の判断は、著作権法で利用が認められている場合[※]を除き、通常の著作権侵害と同様**
 - ※ 個人的に画像を生成して鑑賞する行為（私的使用のための複製）等
- 生成された画像等に**既存の画像等（著作物）との類似性**（創作的表現が同一又は類似であること）や**依拠性**（既存の著作物をもとに創作したこと）が認められれば、**著作権者は著作権侵害として損害賠償請求・差止請求が可能であるほか、刑事罰の対象ともなる**

今後の対応

- 上記の「現状の整理」等について、セミナー等の開催を通じて速やかに普及・啓発
- 知的財産法学者・弁護士等を交え、文化庁においてAIの開発やAI生成物の利用に当たった論点を速やかに整理し、考え方を周知・啓発
- コンテンツ産業など、今後の産業との関係性に関する検討等について

（問い合わせ先：著作権について）
文化庁著作権課
電話 03-5253-4111（内線2775）

（問い合わせ先：コンテンツ産業との関係について）
内閣府科学技術・イノベーション推進事務局重要課題（社会システム基盤）担当
電話 03-5253-2111（内線36230）

https://www8.cao.go.jp/cstp/ai/ai_team/3kai/shiryo.pdf

生成・利用段階

- AIを利用して画像等を生成
 - 生成した画像等をアップロードして公表、生成した画像等の複製物(イラスト集など)を販売
- AIを利用して生成した画像等をアップロードして公表したり、複製物を販売したりする場合の**著作権侵害の判断**は、著作権法で利用が認められている場合※を除き、**通常の著作権侵害と同様**
- ※ 個人的に画像を生成して鑑賞する行為（私的使用のための複製）等
- 生成された画像等に**既存の画像等（著作物）との類似性**（創作的表現が同一又は類似であること）や**依拠性**(既存の著作物をもとに創作したこと)が認められれば、著作権者は著作権侵害として**損害賠償請求・差止請求が可能であるほか、刑事罰の対象ともなる**

ChatGPT関連

- 米弁護士がChatGPTで作成した裁判書類にミス、実在しない判例6件 「二度と頼らない」
 - <https://www.sankei.com/article/20230531-SKPDGTD7VNL5HEDRHRCN5PQXHE/>
- ChatGPT と教育における信頼性 (松林 優一郎 東北大学 教育学研究科)
 - https://www.nii.ac.jp/event/upload/20230421-05_Matsubayashi.pdf
- 「真実のような嘘」(Hallucination)の問題
 - 機械側は形式情報として論理的に正しく処理している。
 - 人間側にとっての意味情報として解釈可能な形式で出力しようとするときに虚言が混在する。
 - 知らないことを知らないと出力せずに、類似度の高い、正解に近いであろう回答を出力するように学習しているから。

生成系AIの核

- GAN

- Generative Adversarial Network
- 敵対的生成ネットワーク
- 偽物を生成するシステムと真贋を判定するシステムを相互に競わせて、真贋判定困難なレベルの偽物を生成するように学習させたシステム。(分かりやすい)

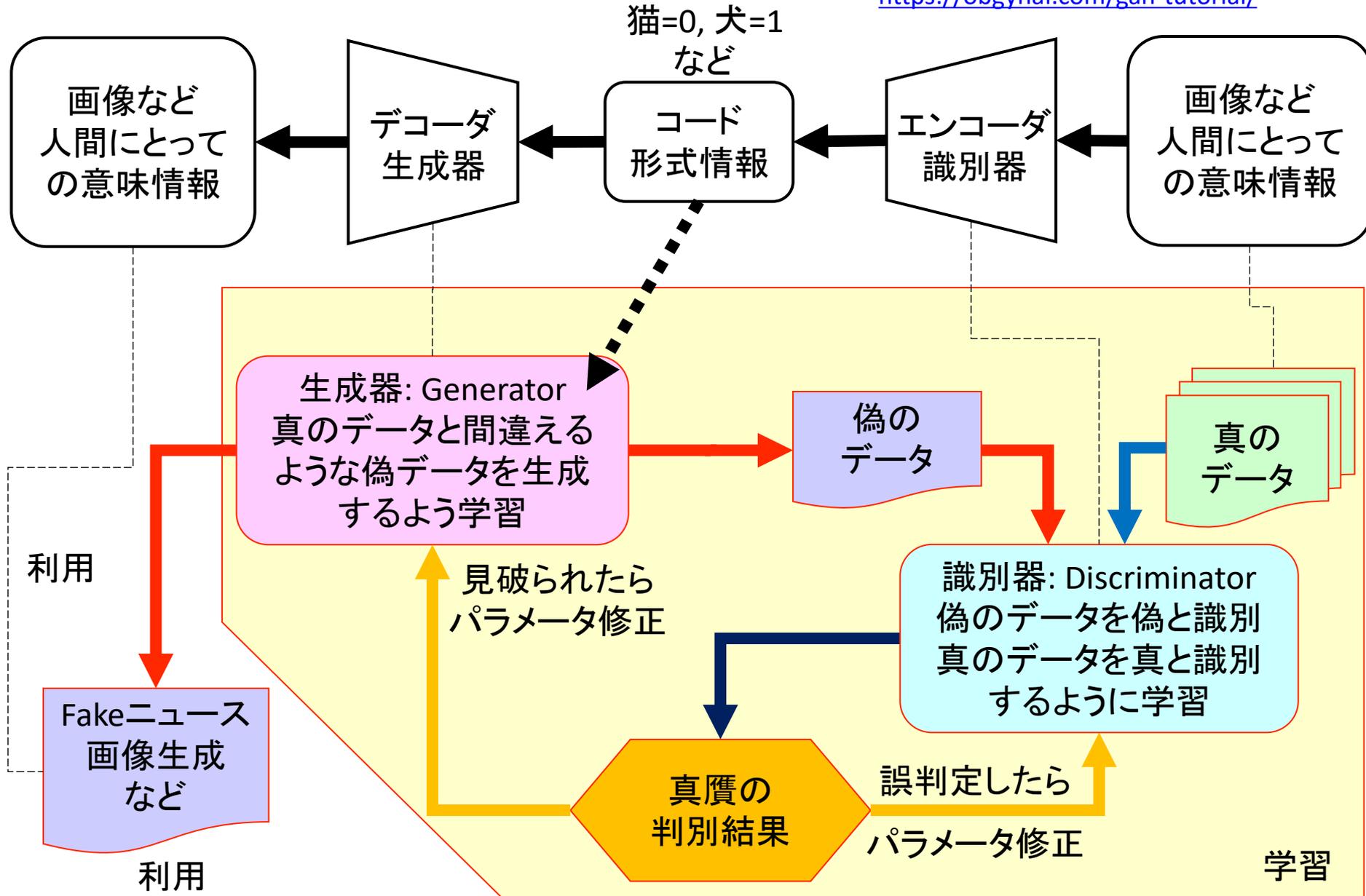
- Transformer

- 元々、言語系の解析、生成に用いられていたシステムの中核
- 比較対象同士を部分に分割し、部分同士の類似度を元に学習していくシステム。(分かりにくい)
- 生成系の場合、学習した類似度を元に、空白部分を埋めるような生成が行われる。

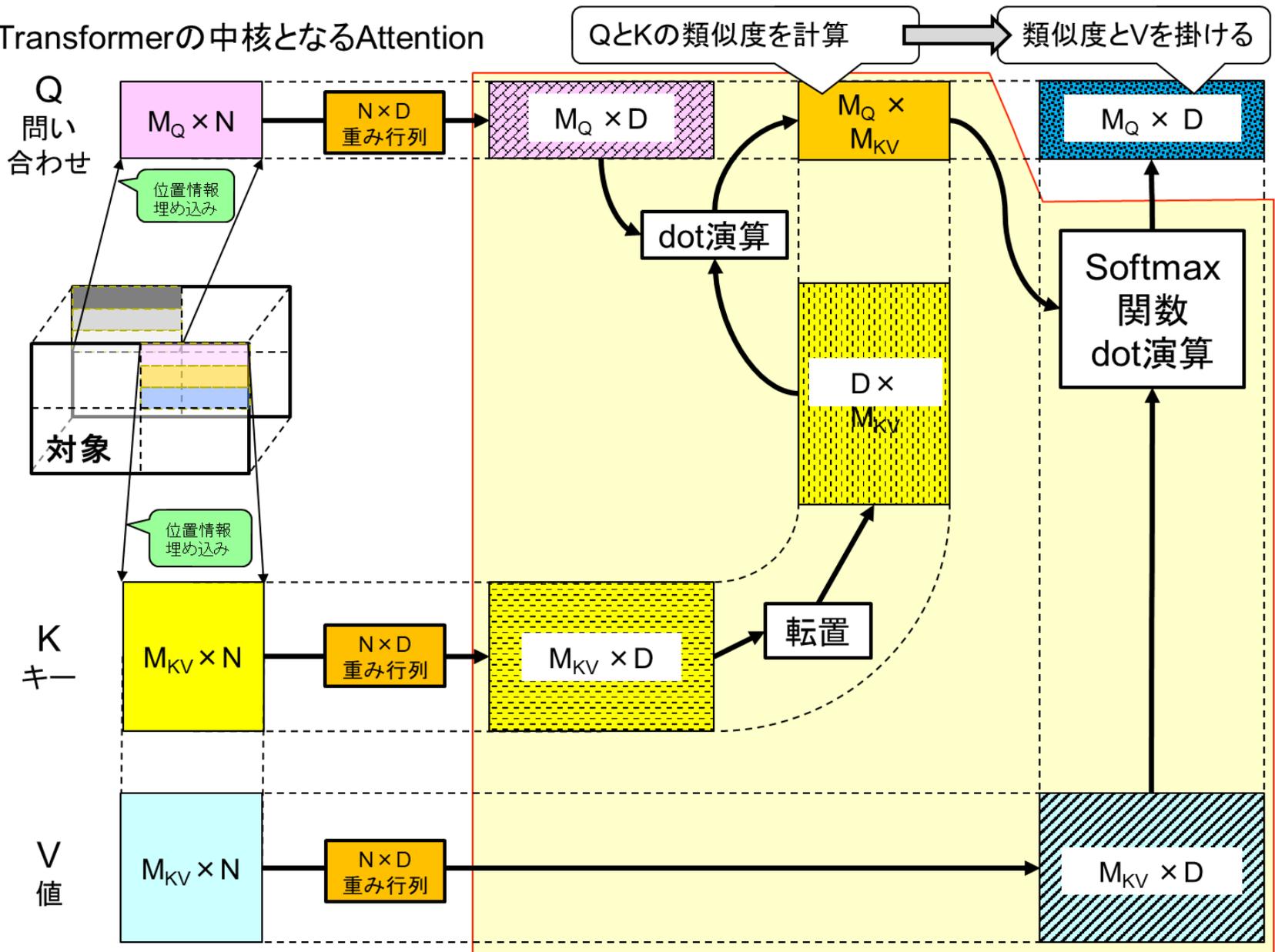
GAN: Generative Adversarial Network (敵対的生成ネットワーク)

<https://crystal-method.com/topics/gan/>

<https://obgynai.com/gan-tutorial/>

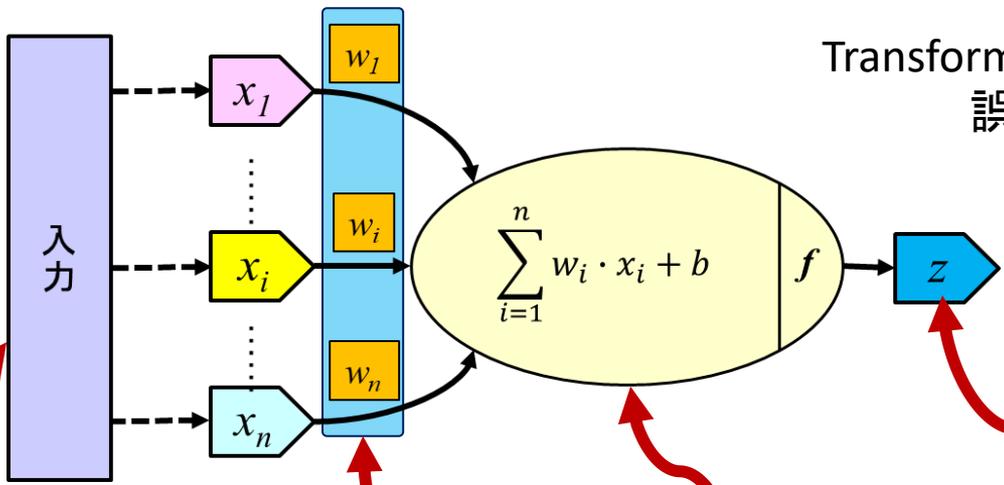


Transformerの中核となるAttention

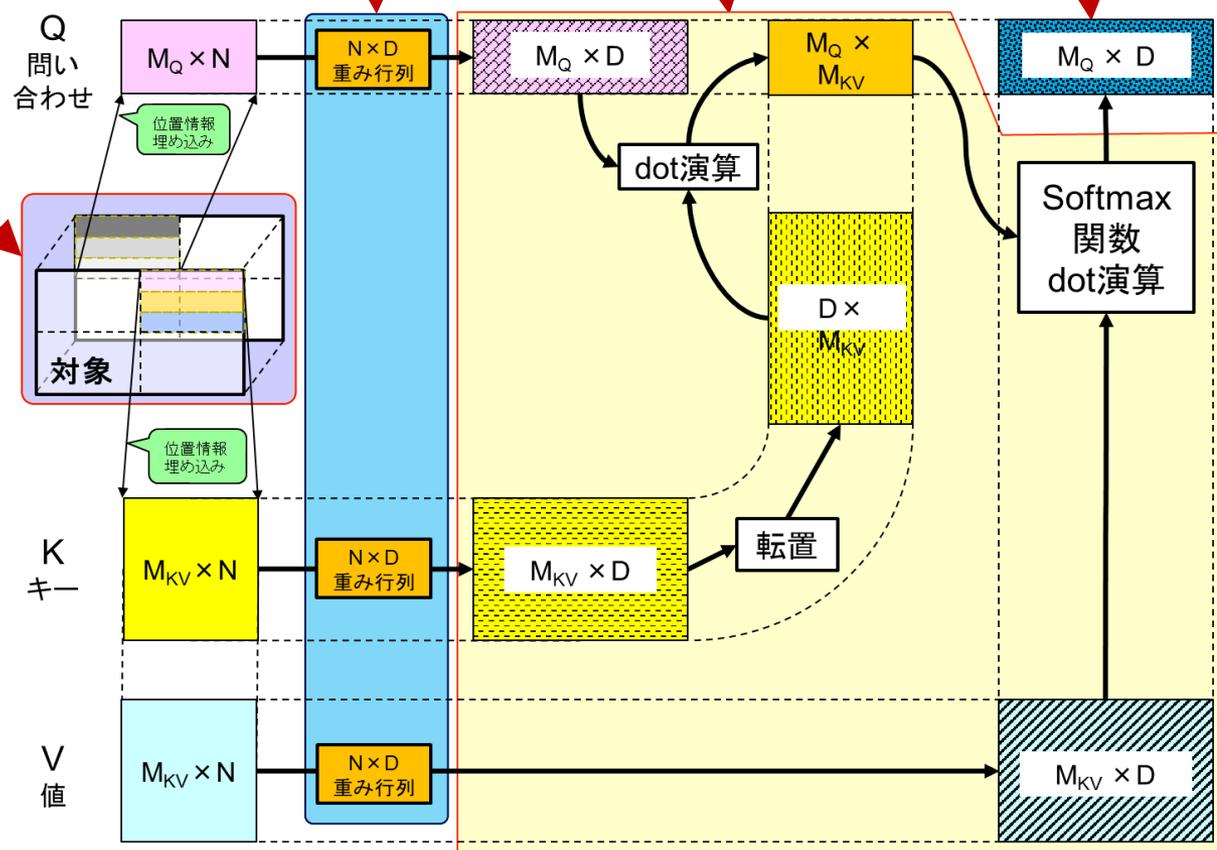


西山秀昌. 歯学教育における人工知能(AI), 機械学習および深層学習について—形式的情報と意味(的)情報の視点から—. 新潟歯学会誌. 2023;53(1):1-15. から、一部改訂

TransformerのAttentionと人工ニューロンとの対比 誤差逆伝播による学習対象は「重み行列」



単一のニューロンが単一のベクトル列を扱うのに対し、TransformerのAttentionは複数のベクトル列(複数行のベクトル列=行列)を一度に扱う



Transformerの場合、フィードバック機構ではなく、フィードフォワード中心のシステムである点異なる。

※フィードフォワード中心の学習とは？
LLM (Large Language Model; 大規模言語モデル) の場合与えられた文字列から次の文字を類推する学習。
画像生成系の場合、部分的に欠落・劣化した画像にて該当部分を補うような学習。

西山秀昌. 歯学教育における人工知能 (AI), 機械学習および深層学習について—形式(的)情報と意味(的)情報の視点から—. 新潟歯学会誌. 2023;53(1):1-15. から、一部改訂し統合編集

Transformerでの学習

ざっくりと言えば

誤差判定の代わりに類似度を使い、学習をおこなっていく

- 画像であれ、音声であれ、コンピュータ内部でのデジタルデータとして扱われる数値の行列・ベクトル列として扱っている。
- 正答か誤答か？、類似度はどの程度か？を、Q(問い合わせ)とK(キー)の類似度(N次元空間内でのベクトルの方向の類似度＝内積の結果)で判定するシステム。
- ※TransformerはChatGPTをはじめ、生成系AIでも広く利用されている。

いわゆるAIでの意味情報とは？

- 癒やし系や配膳ロボット、ないしChatGPTは単に機械的に応答している(形式情報として内部で処理している)だけ。
- しかしながら受け手側は、そこに「意味」や「感情」を見いださう。
- なぜなら「両者の関係、全体の状況が言葉に意味を与えている」から。
- このため、AI側内部にて人間側と同じように「意味」が扱われ、意味情報として伝わってきていると思っ込んでしまう。

仮想空間は生成AIが作り出す世界を含む

=

人間側にとって、どうして間違えたのか分からない・
説明できない誤情報を含む

Society 5.0

内閣府が目指す社会になるためには
実際の所、何が必要なのだろうか？

形式情報に踊らされないためには

どうすればよいのだろうか？

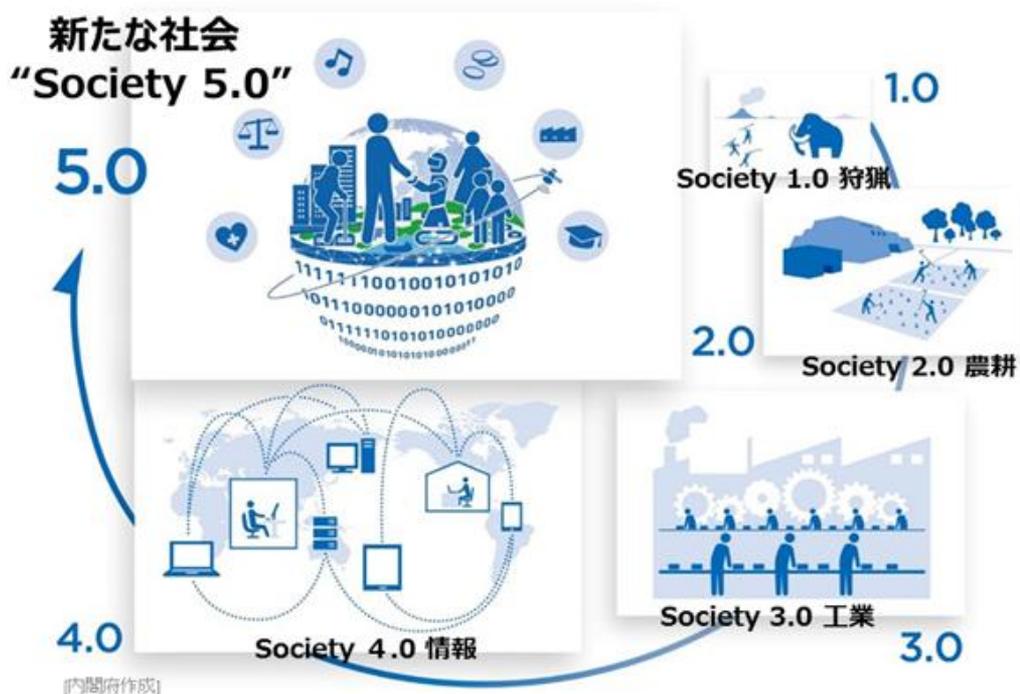
[内閣府ホーム](#) > [内閣府の政策](#) > [科学技術政策](#) > Society 5.0

Society 5.0

Society 5.0とは

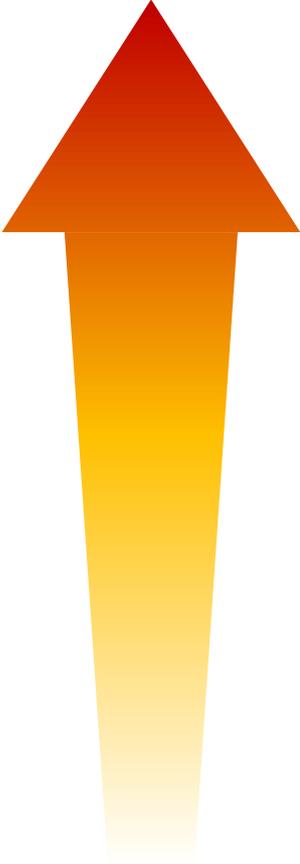
サイバー空間(仮想空間)とフィジカル空間(現実空間)を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の社会(Society)

狩猟社会(Society 1.0)、農耕社会(Society 2.0)、工業社会(Society 3.0)、情報社会(Society 4.0)に続く、新たな社会を指すもので、[第5期科学技術基本計画](#)において我が国が目指すべき未来社会の姿として初めて提唱されました。



「Society 5.0」とは

第5期科学技術基本計画に基づき我が国が目指すべき未来社会の姿として初めて提唱された

- 
- 「サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に癒合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の社会（Society）」
 - 情報社会 （Society 4.0）
 - 工業社会 （Society 3.0）
 - 農耕社会 （Society 2.0）
 - 狩猟社会 （Society 1.0）

これまでの社会

知識・情報の共有、連携が不十分



IoTで全ての人とモノがつながり、新たな価値が生まれる社会



これまでの社会

地域の課題や高齢者のニーズなどに十分対応できない



イノベーションにより、様々なニーズに対応できる社会



Society 5.0

AIにより、必要な情報が
必要な時に提供される社会

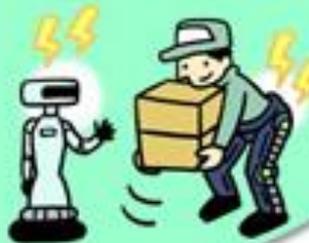


これまでの社会

必要な情報の探索・分析が負担
リテラシー（活用能力）が必要



ロボットや自動走行車などの技術で、
人の可能性がひろがる社会



これまでの社会

年齢や障害などによる、
労働や行動範囲の制約



これまでの情報社会(4.0)

サイバー空間

クラウド

人がアクセスして情報を入手・分析

人がナビで
検索して運転

人が情報を分析・提案

人の操作により
ロボットが生産

フィジカル空間

Society 5.0

サイバー空間

ビッグデータ

解析 AI 人工知能

センサー情報

環境情報、機器の作動情報、
人の情報などを収集

新たな価値

高付加価値な情報、
提案、機器への指示など

自動走行車で
移動

AIが人に最適提案

工場で自動的に
ロボットが生産

フィジカル空間

経済発展

- エネルギーの需要増加
- 食料の需要増加
- 寿命延伸、高齢化
- 国際的な競争の激化
- 富の集中や地域間の不平等

社会的課題の解決

- 温室効果ガス（GHG）排出削減
- 食料の増産やロスの削減
- 社会コストの抑制
- 持続可能な産業化
- 富の再配分や地域間の格差是正

IoT、ロボット、AI等の先端技術をあらゆる産業や社会生活に取り入れ、格差なく、多様なニーズにきめ細かに対応したモノやサービスを提供

「Society 5.0」へ

経済発展と社会的課題の解決を両立



予防検診・ロボット介護

健康寿命延伸・社会コストの抑制



Society 5.0



エネルギーの多様化・地産地消

安定的確保、温室効果ガス排出削減



農作業の自動化・最適な配送

食料の増産・ロスの削減



最適なバリューチェーン・自動生産

持続可能な産業化の推進・人手不足解消



快適



必要なモノやサービスを、必要な人に、必要な時に、必要なだけ提供

サイバー空間とフィジカル空間を
高度に融合



Society 5.0

活力

質の高い
生活



経済発展と社会的課題の解決を両立

