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。

2025/05/23
改訂版

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）

- 「情報」の定義は時代と共に変化し、領域によって微妙に異なるので注意
- 「デジタル化」≒「コンピュータで処理可能」の意味と捉えて良いだろう

「データと情報」という分類・定義と問題点

A) 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

B) ISO 2382-1（JIS X 0001）での定義
https://kikakurui.com/x0/X0001-1994-01.html

「データ」と「情報」という二分類の曖昧さ
「はじめに言葉ありき」で誘導されやすい

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。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、自分に取って価値のある「知らせ」すなわち情報だと言えます。

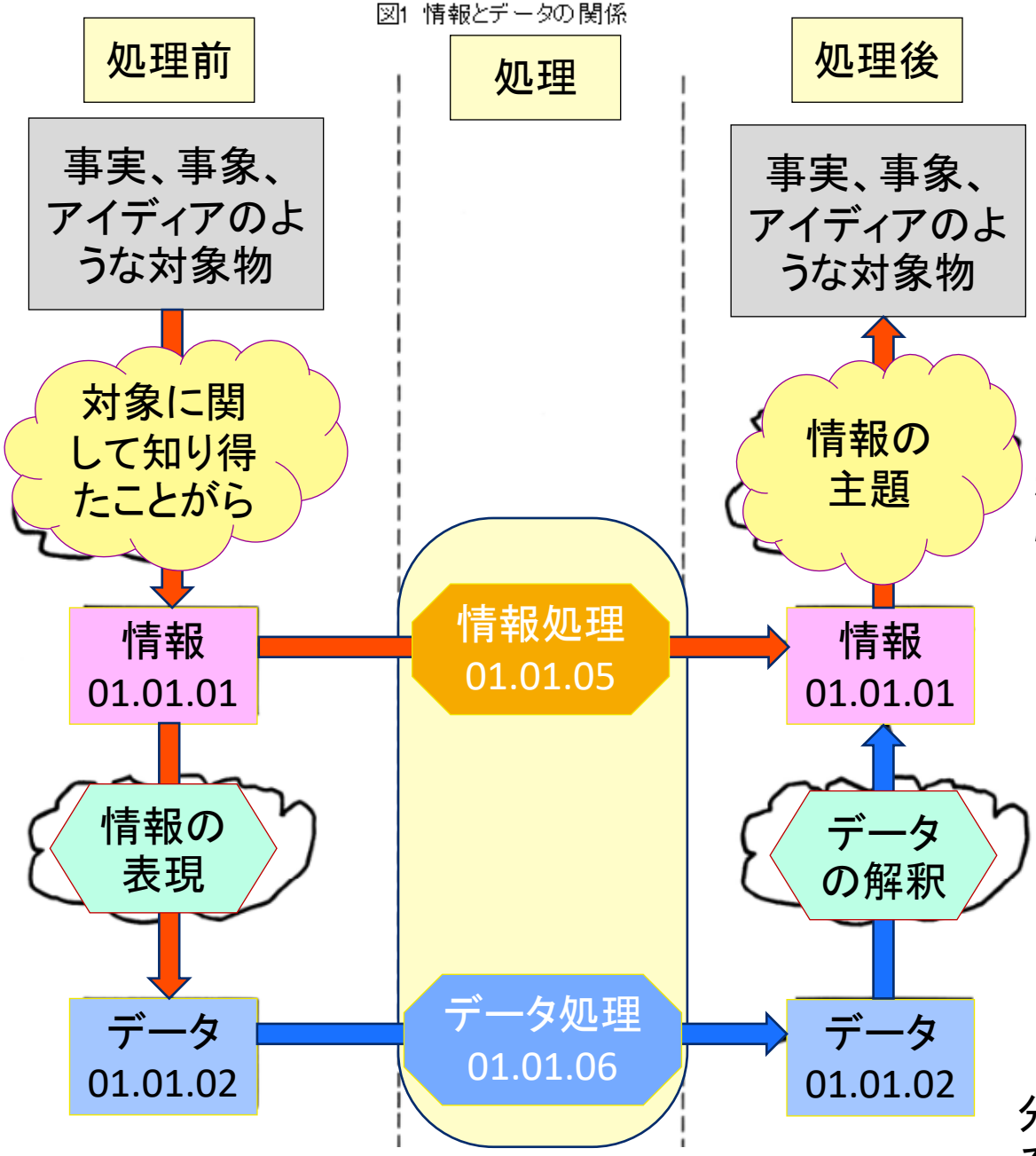
データと情報の二分類

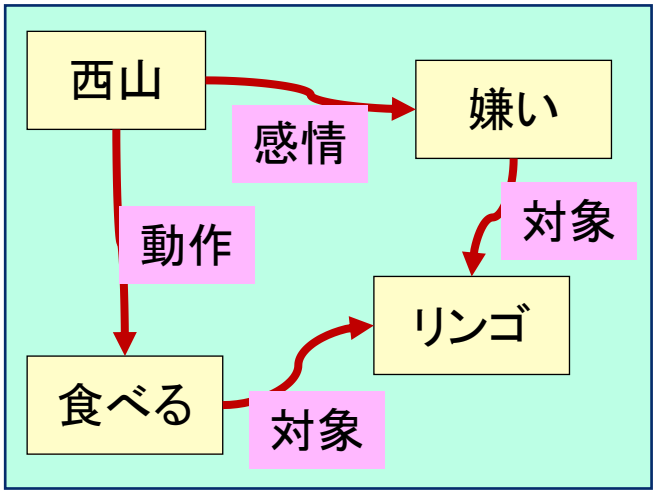
用語：DATAと情報
ISO 2382-1（JIS X 0001）での定義
https://kikakurui.com/x0/X0001-1994-01.html

- 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。

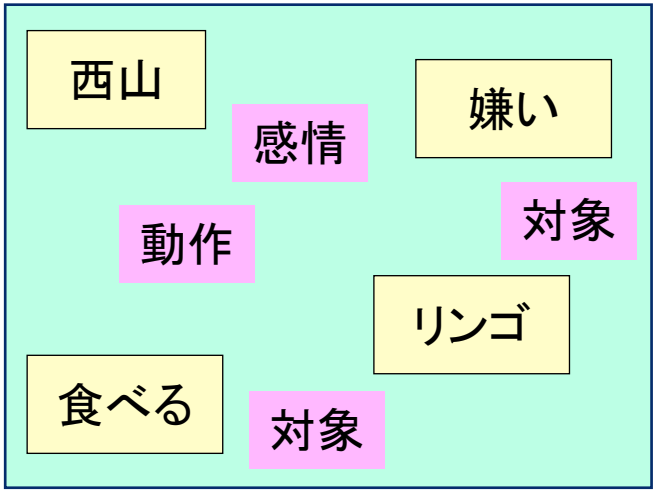
※：「一定の文脈中」＝「いつ、どこで、誰にとって？」といった環境と主体（発信者、受信者）を想定している。

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？



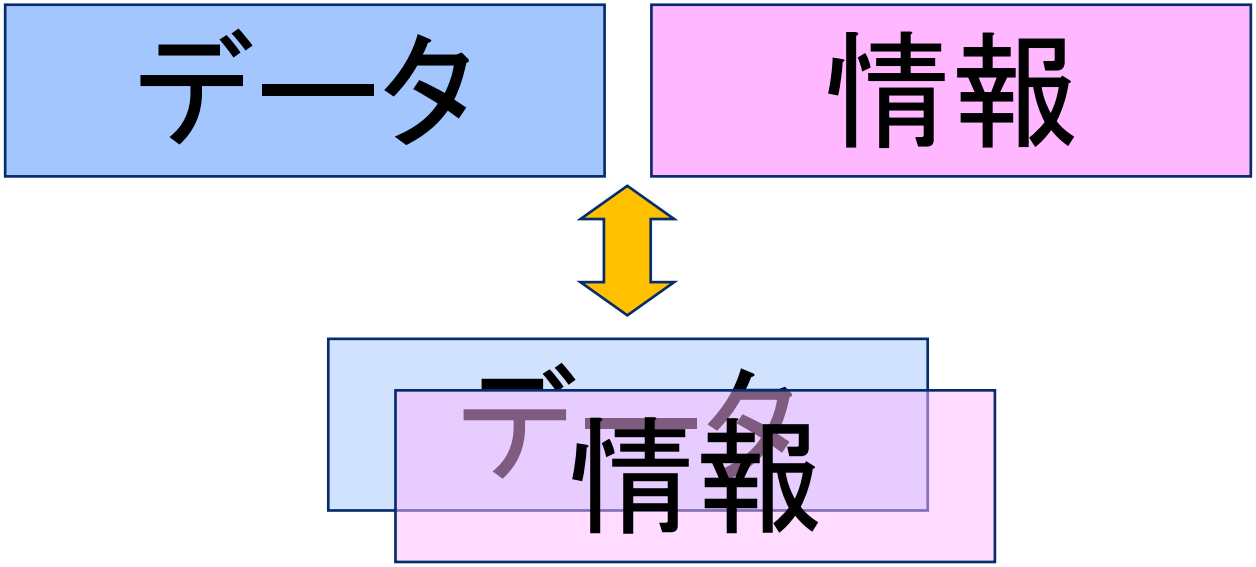
分断された単語はデータとみてよいだろうが、固有名詞は情報ではないのか？→個人情報の情報とは？

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

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・意味」を除けば、全て「データ」となり得るのか？
- 切り離された「価値・意味」も「デジタル化」してコンピュータ処理可能なら、「データ」では？
- 人工知能（AI）の著しい発達に伴い、人間とのやり取りにて機械内部に構築される（動的かつ機能と結びつく）データ構造内部に機械側にとっての「価値・意味」は無いのか？
- 機械側にとっての「価値・意味」が定義できれば、それは「データ」ではなく機械側にとっての「情報」と解釈すべきでは？

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」と「情報」は多義的に相互に入れ替わりうる。

「個人情報保護」を「個人の情報の保護」と読み違えると、「データなら保護対象外では？」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」の保護という視点は重要。

※1：提言書ドラフト版「デジタル社会を駆動する「個人データ法制」に向けて」（2022年8月9日発表） GLOCOM 六本木会議・資料
https://roppongi-kaigi.org/wp-content/uploads/2022/08/PERSONALDATA_GRK_20220809_DRAFT.pdf

情報
歴史的な定義の変遷
及び
様々な領域での定義の違い

小野厚夫（1991）
明治期における「情報」と「状報」

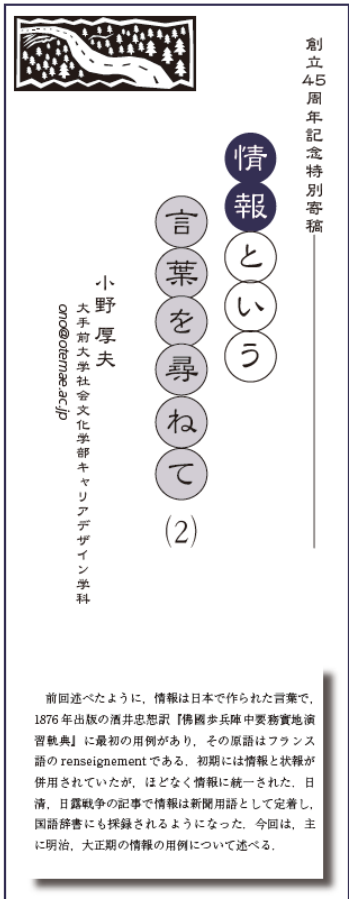
- 「野外演習軌典」（明治15年3月20日）等のまとめ
  - 間諜がもたらす諜報とは異なる、一般兵士からの報告
  - 上からの命令に対する下からの情報
  - 主観が入りやすいため次の3つの雛形に分ける。
    - 実際に見たこと
    - 伝聞したこと
    - 想像されること

    --- 客観的な事柄

    ------- 主観的な事柄
- 「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。

森鴎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

情報という言葉を尋ねて

(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@otemae.ac.jp

前回述べたように，情報は日本で作られた言葉で，1876年出版の酒井忠恕訳『佛國歩兵陣中要務實地演習軌典』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．

鴎外と「情報」

森林太郎（鴎外）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確実の度を加へしめ，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確実と此の偶然の頻りに来ることとは，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

3K－1

鷗外森林太郎による獨逸語 NACHRICHTEN の二つの翻譯語「情報」と「狀報」

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「狀報」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Ｆｂ）

の諸状報及び諸予想の

確かに「状報」（客観的）と「予想」（主観的）を区分している。

医療カルテでの情報、SOAP

https://www.ncbi.nlm.nih.gov/books/NBK482263 [1]
https://www.mhlw.go.jp/content/10808000/001220428.pdf [2]
https://www.jstage.jst.go.jp/article/naika/106/12/106_2529/_pdf [3]

- SOAP
  - Subjective：主観的情報（主訴、問診内容など）[2]
    Documentation under this heading comes from the “subjective” experiences, personal views or feelings of a patient or someone close to them. In the inpatient setting, interim information is included here [1].
  - Objective：客観的情報（診察所見、検査結果など）[2]
    This section documents the objective data from the patient encounter [1].
  - Assessment：経緯・評価 [2]
  - Plan：計画（治療）[2]
- 主観的なのか、客観的なのかの区分が難しい場合があり、分断することで因果関係がむしろ分かりにくくなる [3]。

情報学（inofrmatics）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。

https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-h160323-2.pdf　p.6～

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

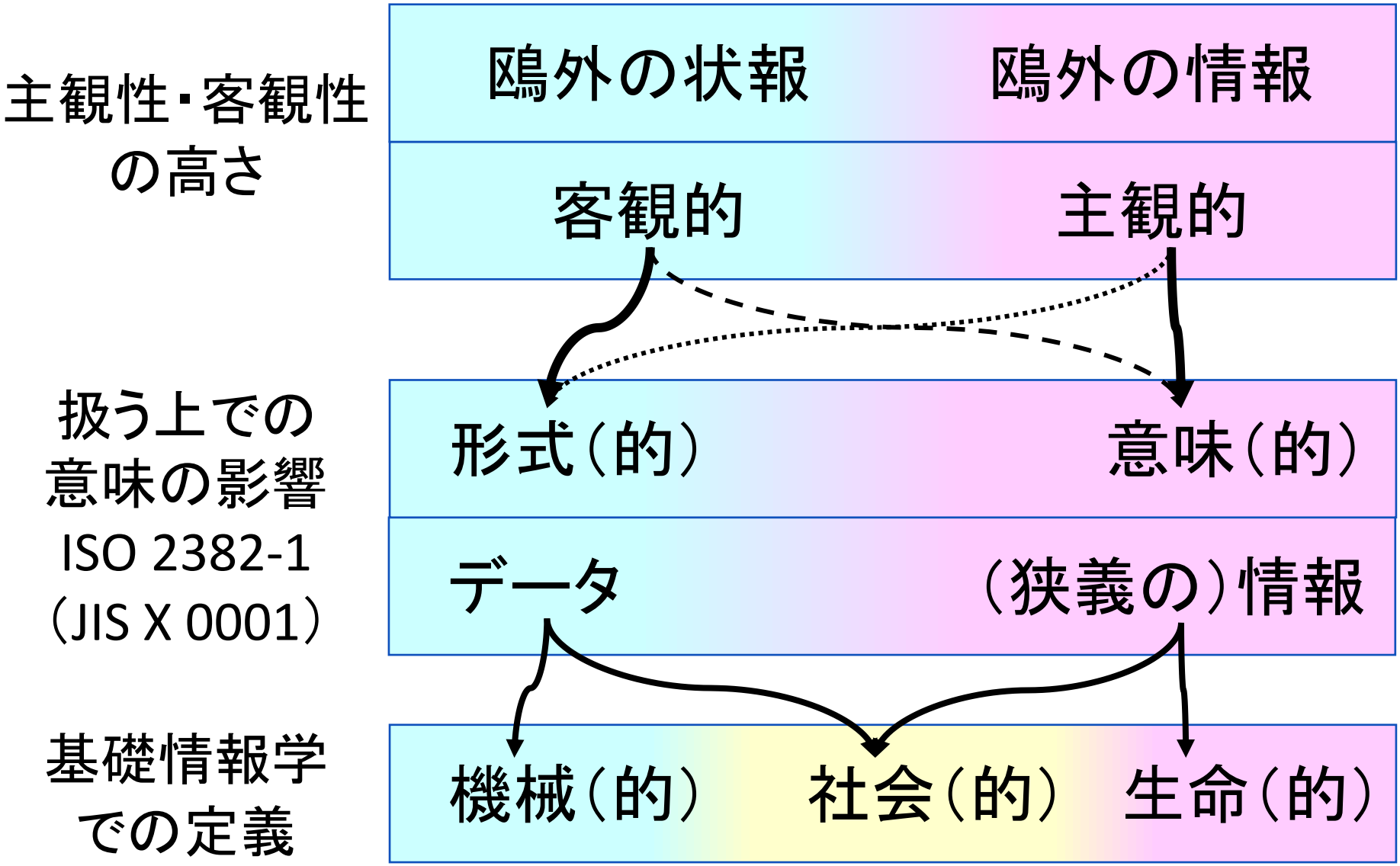
社会情報

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

異なる視点・基準からの情報の分類



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.（一部追記）

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

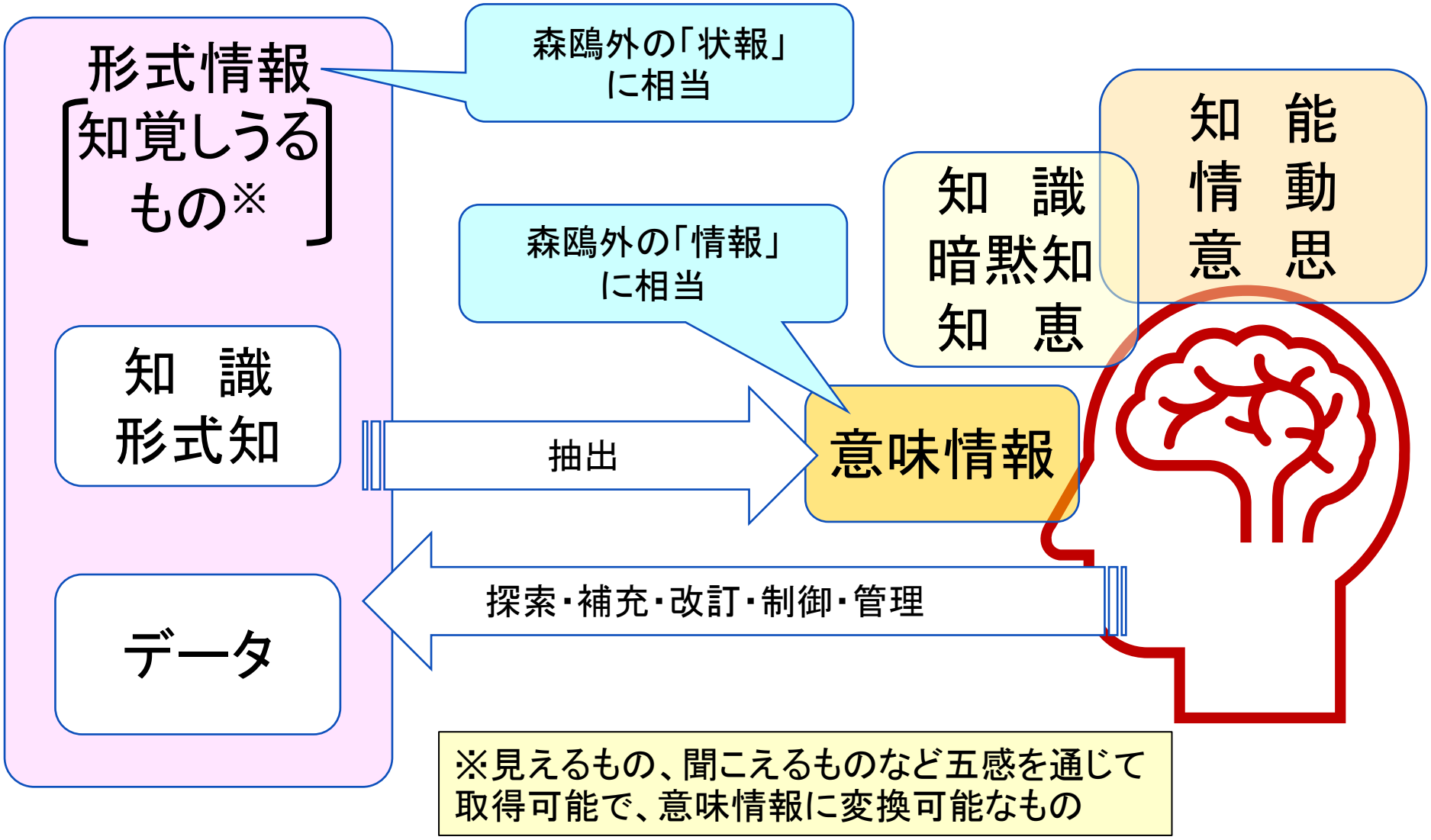
本授業で用いる用語について

意味情報≒（JIS X 0001の）情報

形式情報≒（JIS X 0001の）データ

ChatGPT-4oと、上記の定義にて対話した結果を公開します。
https://chatgpt.com/share/682de6da-bcd4-8005-a512-e07f0dcb5185

データ、情報、知識・知恵の関係

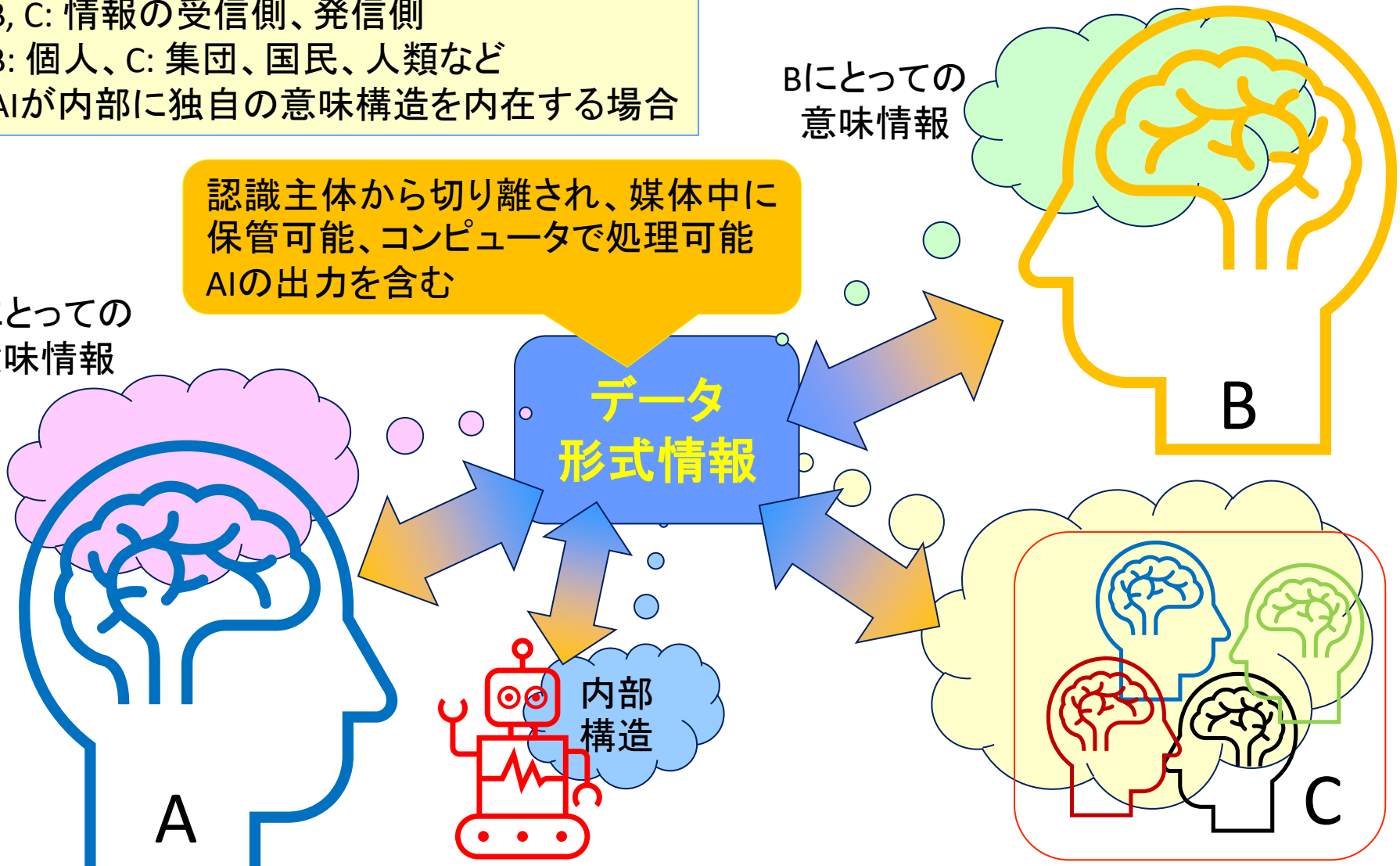


データ（形式情報）から
価値としての情報（意味情報）へ

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
D: AIが内部に独自の意味構造を内在する場合



情報の定義は
最終レポート（※）記述時にも関連

- 情報倫理の講義にて取り上げた「科学的、客観性の高さ」という概念は、「情報」の定義にも関連しています。
- 自分が認識している「情報の意味」（意味情報）を形式情報として正確に伝えるためには、その背景や付随する関連情報（事実・データ等）の引用・参照を明示（文脈を明示）する必要があります（剽窃・盗用の防止）。
- 主観性の高い意味情報を客観性の高い形式情報へと変換することで、より正確な情報伝達がなされます。

※小野先生の「レポートの書き方」（その１～３）にて指示された事項、テキストではp.182-192。

著者・話者や読者・聴者の情況、両者の関係、全体の状況が言葉に意味を与える

- 著者・話者や読者・聴者の情況
  - 内面（心情）を含めているので「情況」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。
- これらの情況・状況が形式的な文字記号や発音記号、物理学的な画像や音として（形式情報として）伝播される「言葉」に意味を与え、意味情報となる。
- 著者・話者の心情は必ずしも的確には伝わらない可能性があり、両者にとっての意味情報は異なりうるが、的確かつ適切な質と量を備えた形式情報であれば、誤解・誤読の危険性は低くなるだろう。

情報の単位について

形式（的）情報

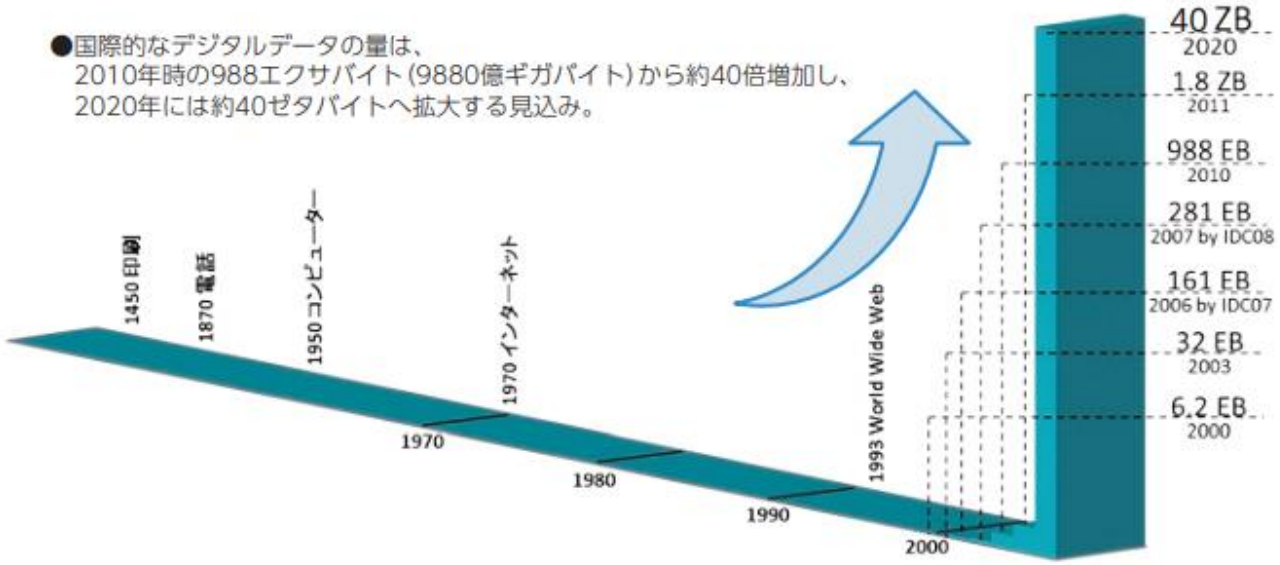
主として数学的、物理学的な側面

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

https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/pdf/n3100000.pdf

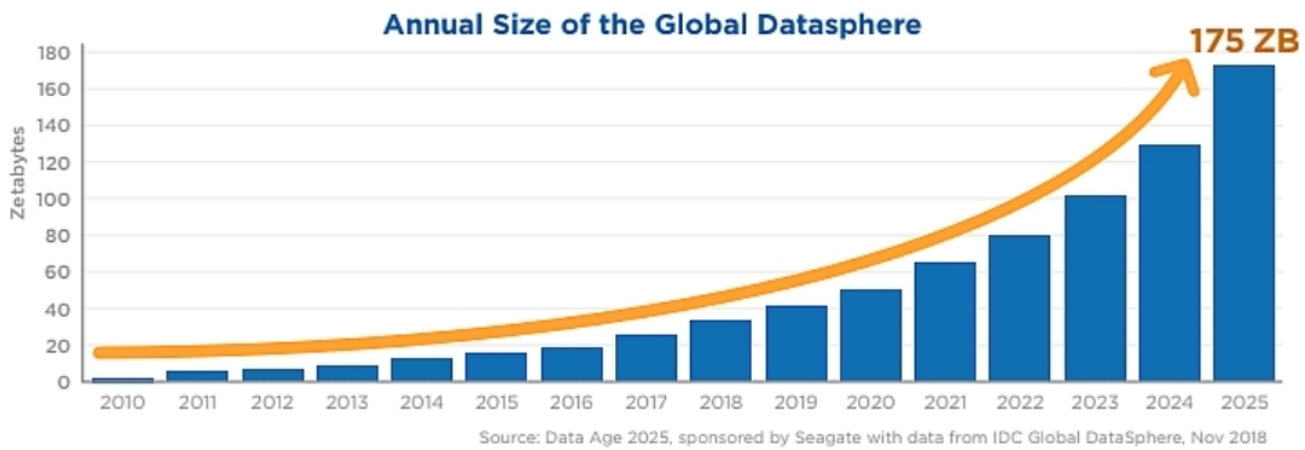
全情報量の爆発的増大

japan.zdnet.com/article/35129774/

ZDNet Japan　MENU

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。

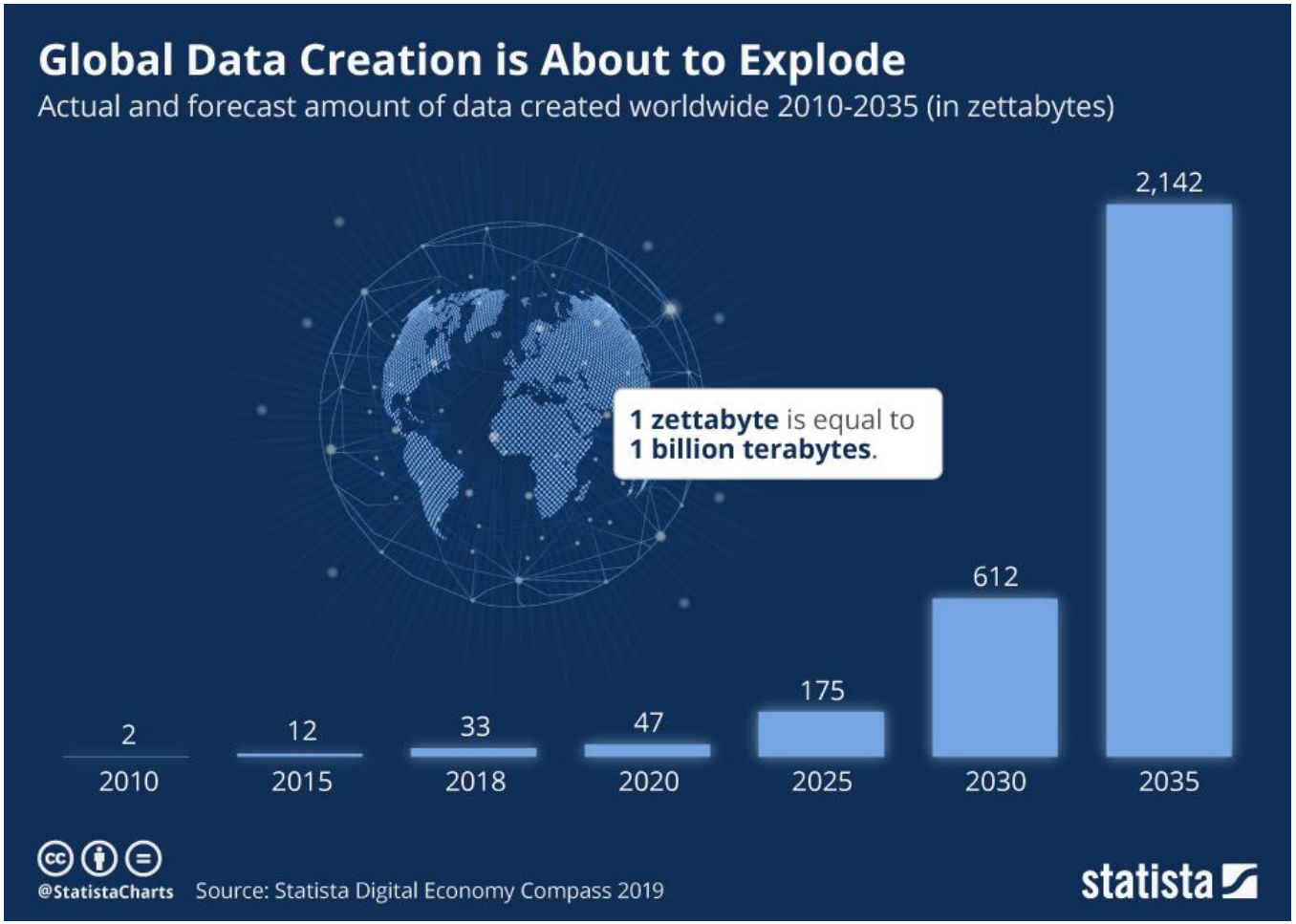
Figure 1 – Annual Size of the Global Datasphere



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

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

５年後、１０年後の予想・・・



https://www.statista.com/chart/17727/global-data-creation-forecasts/

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
ゼプト（zepto）	z	10^{-21}	0.000 000 000 000 000 000 001	1991
ヨクト（yocto）	y	10^{-24}	0.000 000 000 000 000 000 000 001	1991
ロント（ronto）	r	10^{-27}	0.000 000 000 000 000 000 000 000 001	2022
クエクト（quecto）	q	10^{-30}	0.000 000 000 000 000 000 000 000 000 001	2022

https://unit.aist.go.jp/nmij/info/SI_prefixes/indexS.html

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
0.000 000 000 000 000 000 000 000 000 001

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

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

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。

https://unit.aist.go.jp/nmij/info/SI_prefixes/indexS.html

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

ISO単位と2進接頭辞

- ISO単位系ではk(キロ)、M（メガ）、G（ギガ）と単位が変化するとき、1000倍ずつの変化となる。
  ※通信速度：bps等ではISO単位
- IEC規格の２進接頭辞（接頭語）では、Ki（キビ）、Mi（ミビ）、Gi（ギビ）と単位が変化するとき、1024倍ずつの変化となる。
- 過去からの慣習（慣例）ではあるが、今でも記憶媒体の容量表現にてKiBの代わりにKB、MiBの代わりにMB、GiBの代わりにGBが使われる場合があるので注意（後述）。

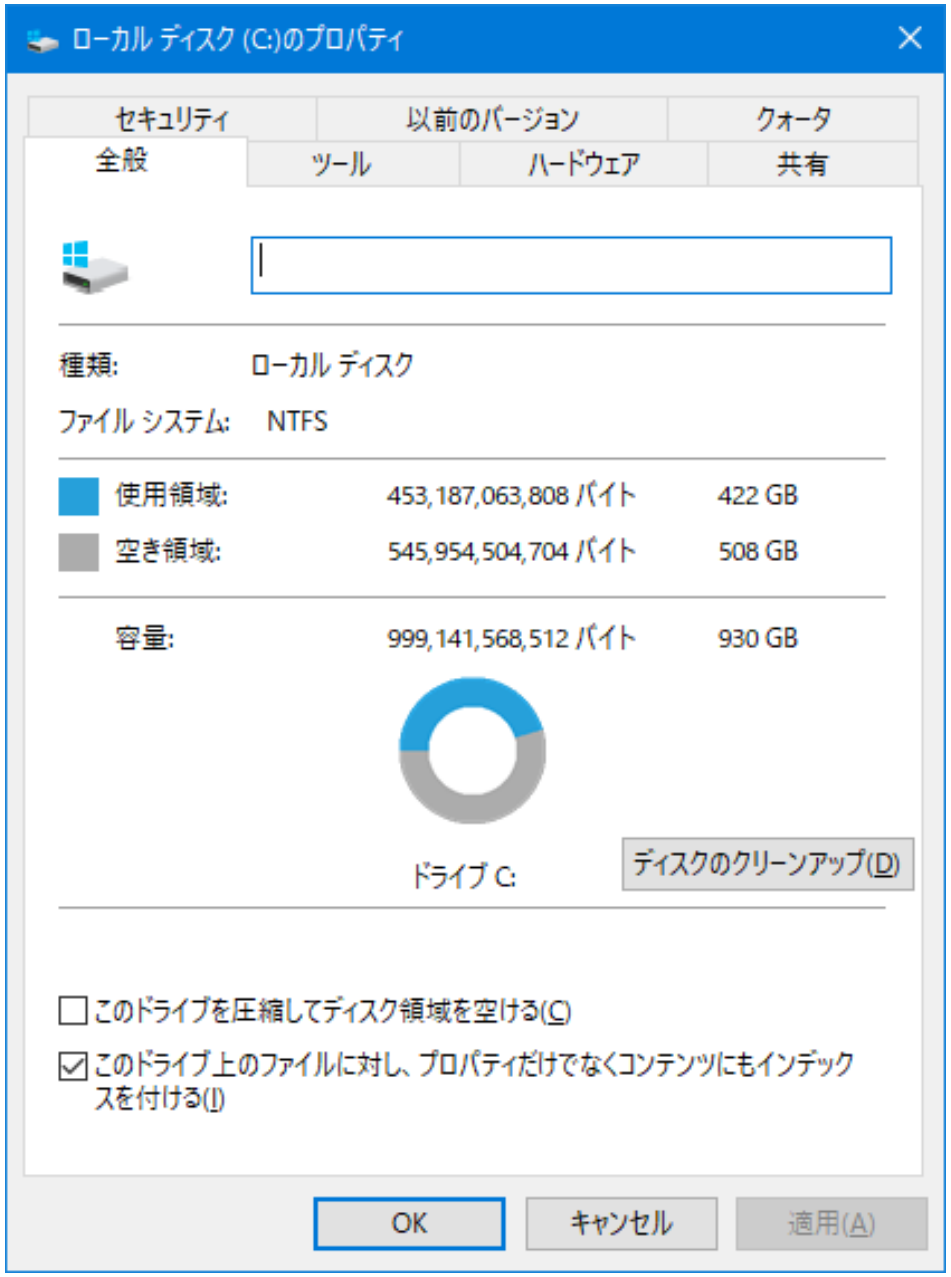
表1. 情報の量と単位（テキストの表を差替）

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

IEC 規格 2進接頭辞	単位・記号 IEC 規格	量、比較	
	bit（ビット）	２進法の桁数	$\log_2 N$
	byte, B（バイト）	8 bit	2^8
Ki(kibi)	KiB（キビバイト）	1,024 B	2^{10}
Mi（mebi）	MiB（メビバイト）	1,024 KiB	2^{20}
Gi（gibi）	GiB（ギビバイト）	1,024 MiB	2^{30}
Ti（tebi）	TiB（テビバイト）	1,024 GiB	2^{40}
Pi（pebi）	PiB（ペビバイト）	1,024 TiB	2^{50}
Ei（exbi）	EiB（エクスビバイト）	1,024 PiB	2^{60}
Zi（zebi）	ZiB（ゼビバイト）	1,024 EiB	2^{70}
Yi（yobi）	YiB（ヨビバイト）	1,024 ZiB	2^{80}
Ri（robi）＊	RiB（ロビバイト）	1,024 YiB	2^{90}
Qi（quebi）＊	QiB（クエビバイト）	1,024 RiB	2^{100}
媒　体	容　量	情報伝達速度	単位・記号
CD-R	700 MiB	bit/sec	bps
DVD-R	4.7 GiB	量子ビット	単位・記号
BD-R	25 GiB	quantum bit	qbit

記憶媒体の容量では慣習的にKB, MB, GB, TBが用いられている。
*Ri, Qiは検討中で正式とはなっていない。

https://iopscience.iop.org/article/10.1088/1681-7575/ac6afd



コンピュータでの容量表示では、2進接頭辞が使われていなくても1024の倍数で計算されていることがあるので注意
※Windowsの場合

使用領域	表示	本当の単位	
453,187,063,808	Byte	B	÷1024
442,565,492	KB	KiB	
432,193	MB	MiB	
422	GB	GiB	

空き容量	表示	本当の単位	
545,947,791,360	Byte	B	÷1024
533,152,140	KB	KiB	
520,656	MB	MiB	
508	GB	GiB	

容量	表示	本当の単位	
999,141,568,512	Byte	B	÷1024
975,724,188	KB	KiB	
952,856	MB	MiB	
931	GB(*)	GiB	

*プロパティでは小数点以下が切り捨てられている

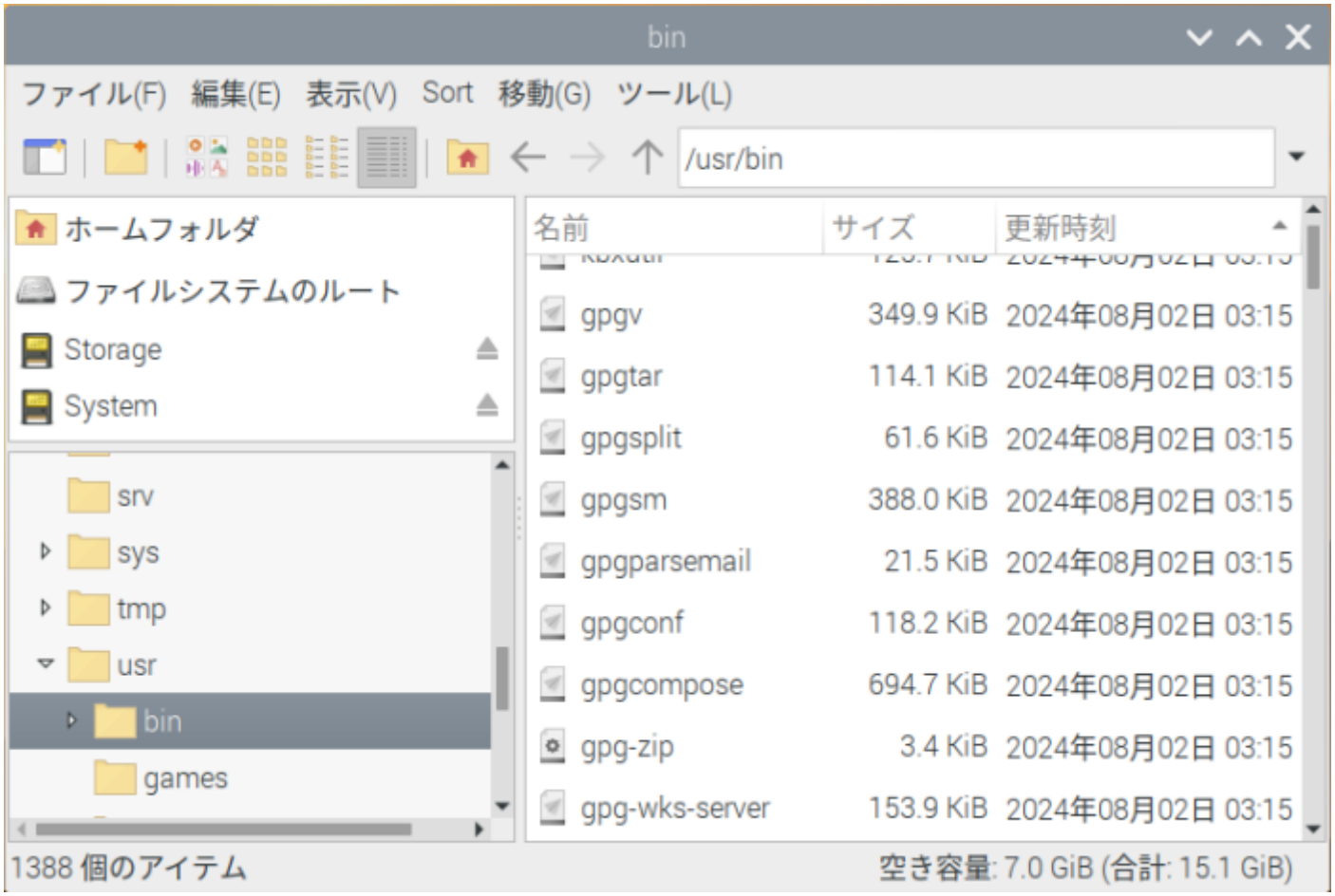
Linux(WSL2; Ubuntu 22.04.5 LTS)での表示

dfコマンドの
-hは1024単位（2進接頭辞）で表示
-Hは1000単位（SI接頭辞）で表示
しかしながら単位はSI接頭辞を流用している

```
hn:$ df -h /mnt/c
Filesystem      Size  Used Avail Use% Mounted on
C:\             931G  423G  509G  46% /mnt/c
hn:$ df -H /mnt/c
Filesystem      Size  Used Avail Use% Mounted on
C:\             1.0T  454G  546G  46% /mnt/c
```

https://atmarkit.itmedia.co.jp/ait/articles/1610/24/news017.html

Raspberry Pi
(Raspbian GNU/Linux 10)での表示
2進接頭辞で正確に表示されている



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、$\boldsymbol{log_2(N)}$ **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）

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

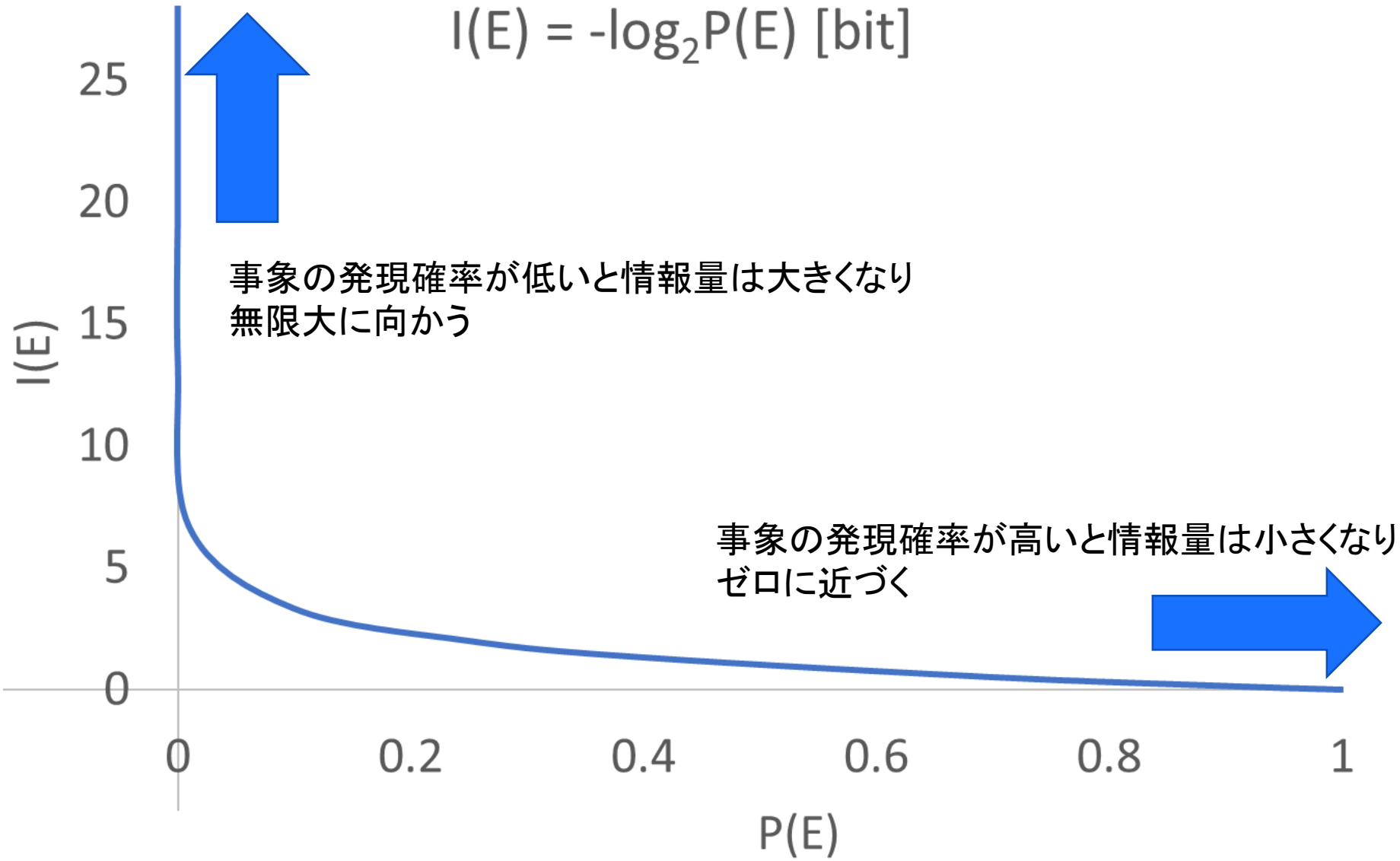
  - 平均情報量（エントロピー、シャノンエントロピー）

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

  - 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）



問　題　1

（有効桁数は小数点以下3桁とします）

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

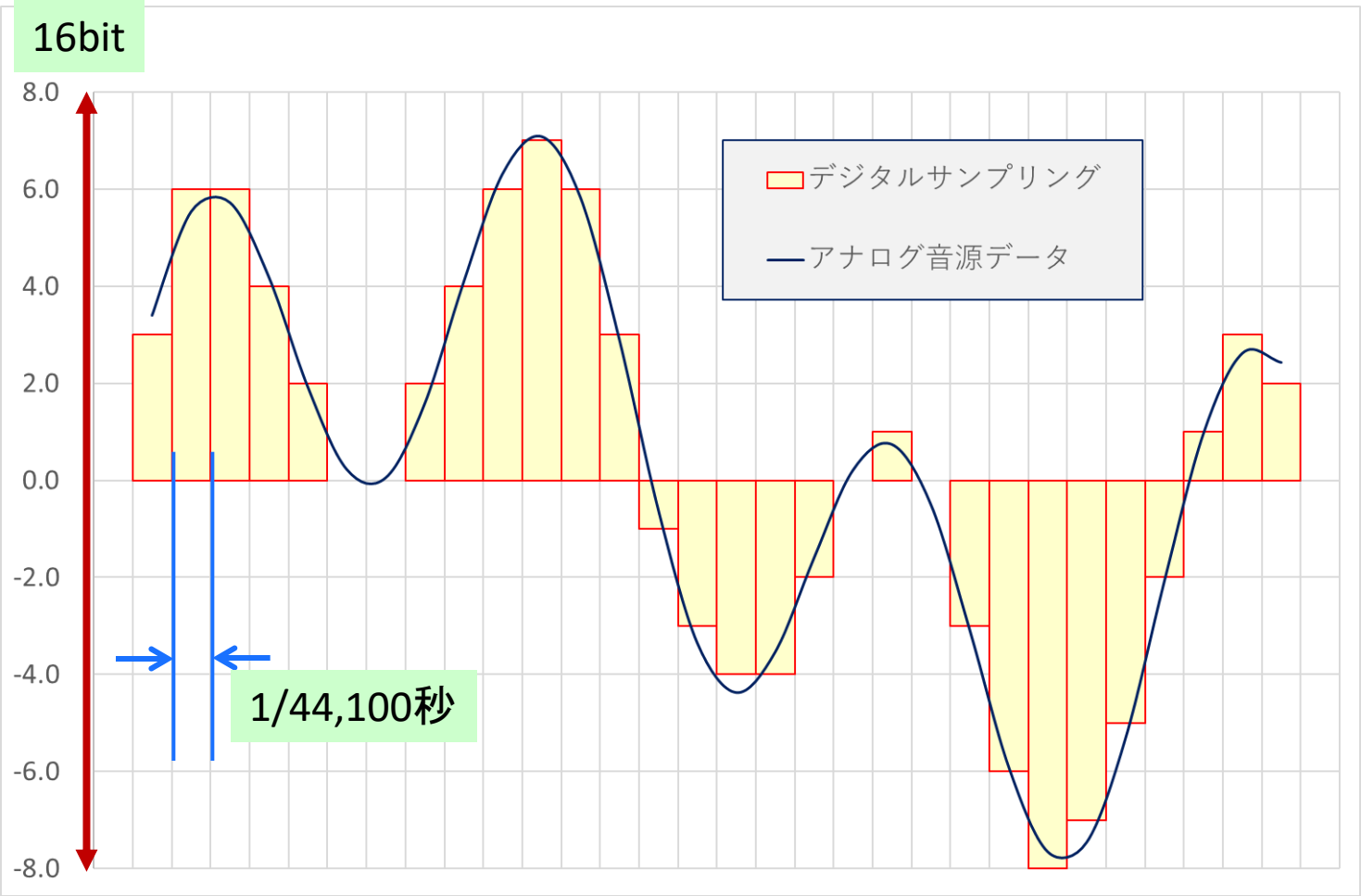
高校テキスト「社会と情報」ないし「情報I」を参照のこと

- 音（標本化： Hz、量子化： bit）
  - PCM（Pulse Code Modulation）, DSD（Direct Stream Digital）, MP3, WAVE, AAC, AIFF, FLAC, MIDI等
- 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（標本化： fps, pixel、量子化： bps, bit）

  fps: フレームレート（frames per second）

  bps: ビットレート（bits per second）
  - MPEG： 空間周波数変換、コマ単位の差分
  - 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　２

（最終結果は小数点以下3桁まで求めてください）

A）音楽CD（44.1kHz/16bit）に収録された12分の音楽のデータ量の計算過程にて単位換算の式として正しいのはどれか？

[　] 44.1kHz = 44.1×1000Hz
[　] 44.1kHz = 44.1×1024Hz
[　] 上記以外

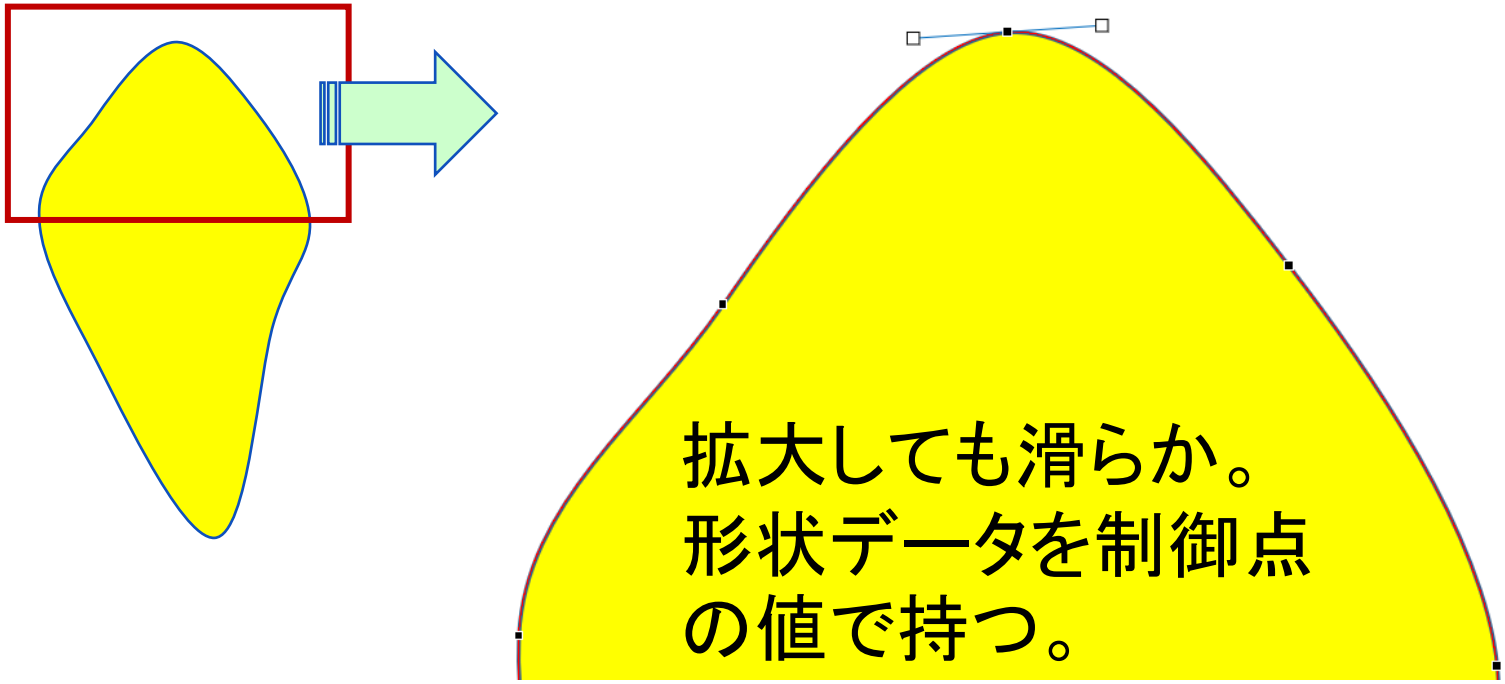
B）320kbpsと記されたMP3形式の２分３７秒の音楽ファイルがある。データ量の計算過程の式として正しいのはどれか？

[　] 320kbps = 320×1000 bits / sec
[　] 320kbps = 320×1024 bits / sec
[　] 上記以外

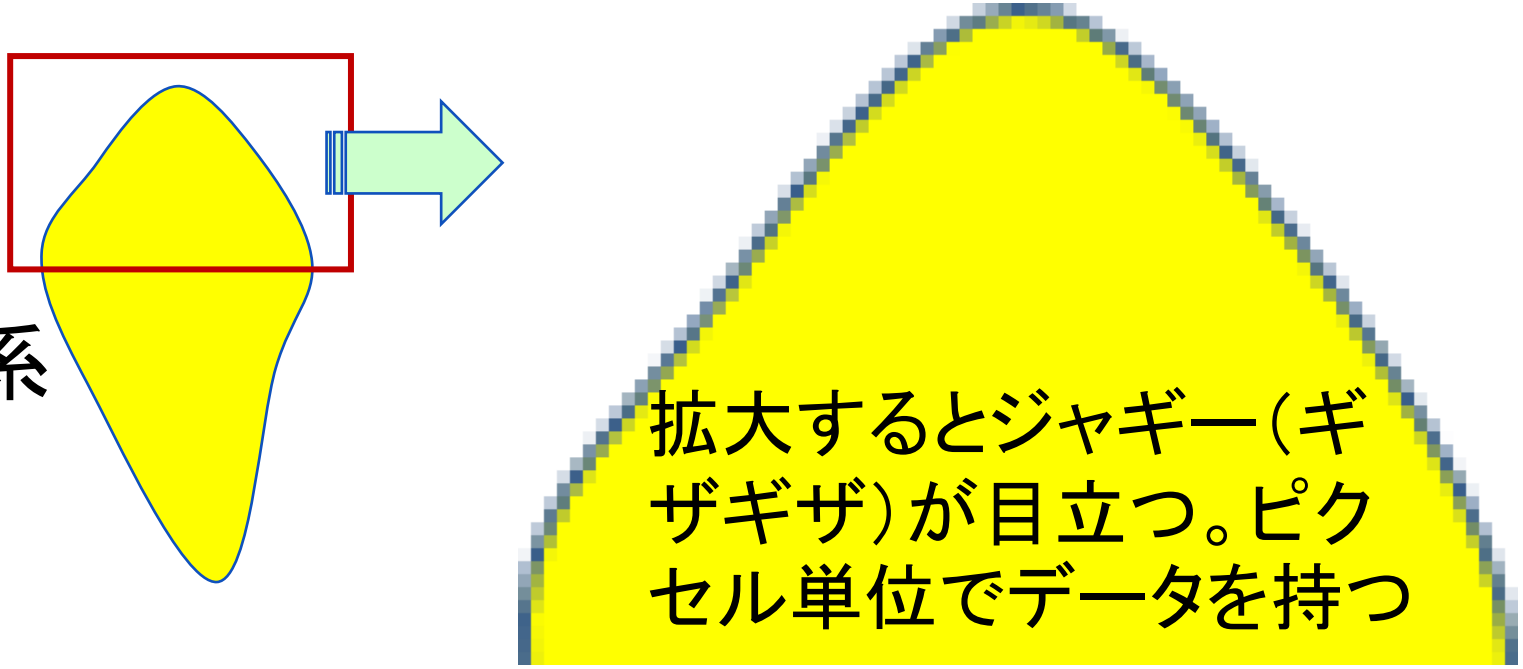
画像のデジタル変換

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

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

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

4-6. 画像解析

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)	
画像データの処理	– 画像処理（動画・滋賀大学）			– テキスト・画像・音声データ分析
画像認識、画像分類、物体検出	– 画像（動画・滋賀大学）			– テキスト・画像・音声データ分析(画像分類以外対応)

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

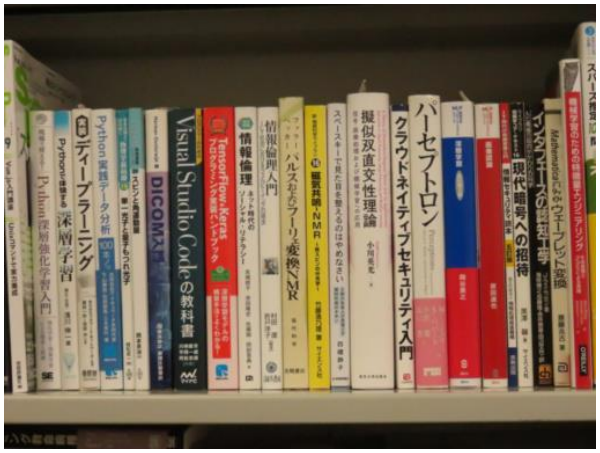
自動設定
（白色蛍光灯と同じ）

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/

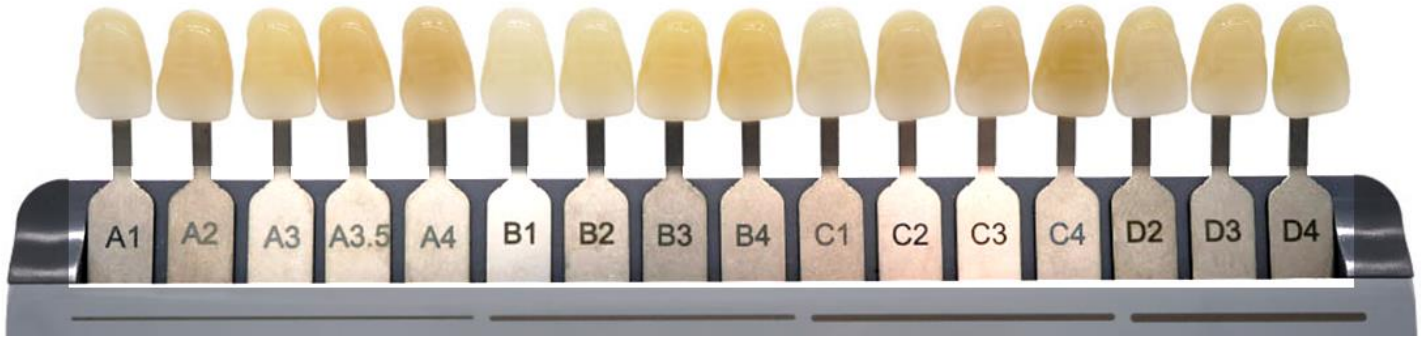
ビタ クラシカル A1-D4 シェードガイド (2021)

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　３

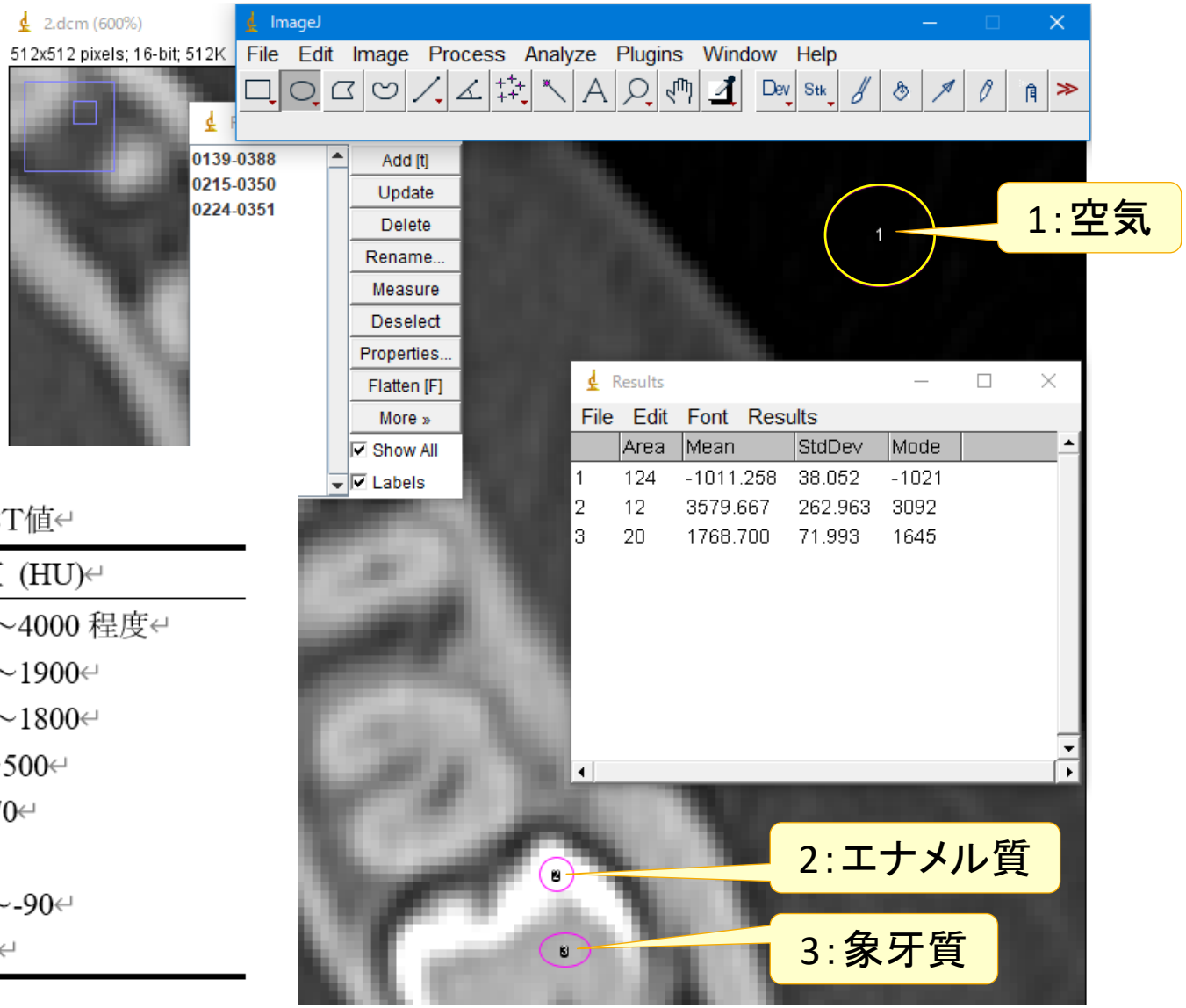
- シェードガイドのコードをデジタル化するに必要な形式的な情報量（ビット数）の最小値はいくらか？
- ただし、A,B,C,D,の4つの区分が分かりやすくなるように工夫すること。



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

DICOM画像でのCT値
ImageJでの閲覧

「ワード」の講義で作った表



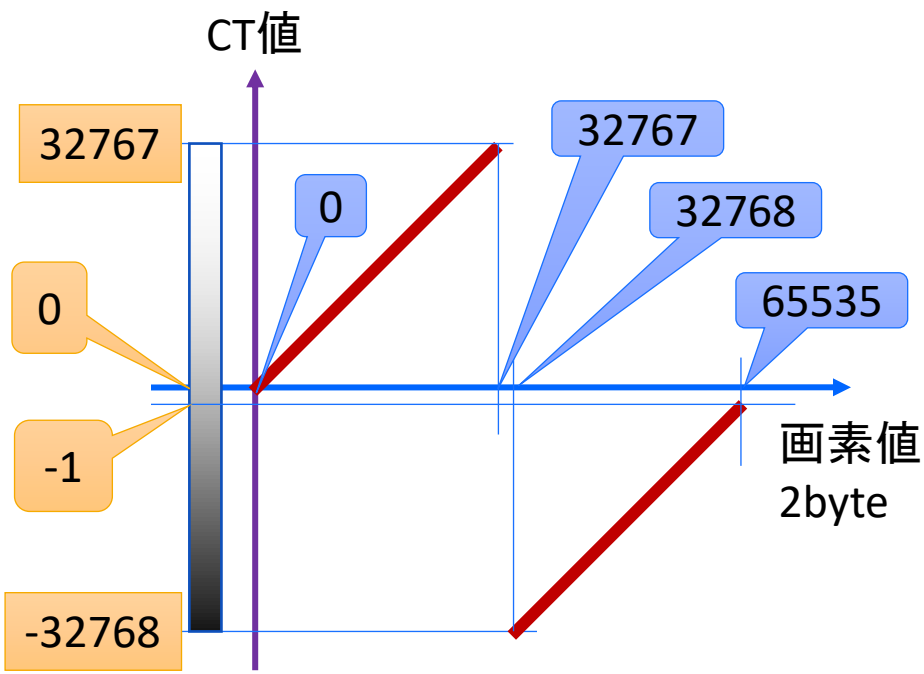
人体各組織のCT値

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

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	10進表示		
16ビット	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）

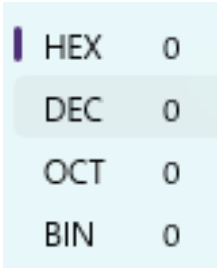
A）符号無しの16ビット(2バイト）の整数表現にて、10進数の「23156」を16進数で表すとどうなるか？

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？

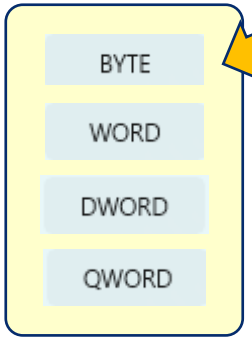
Windows附属の電卓で計算例

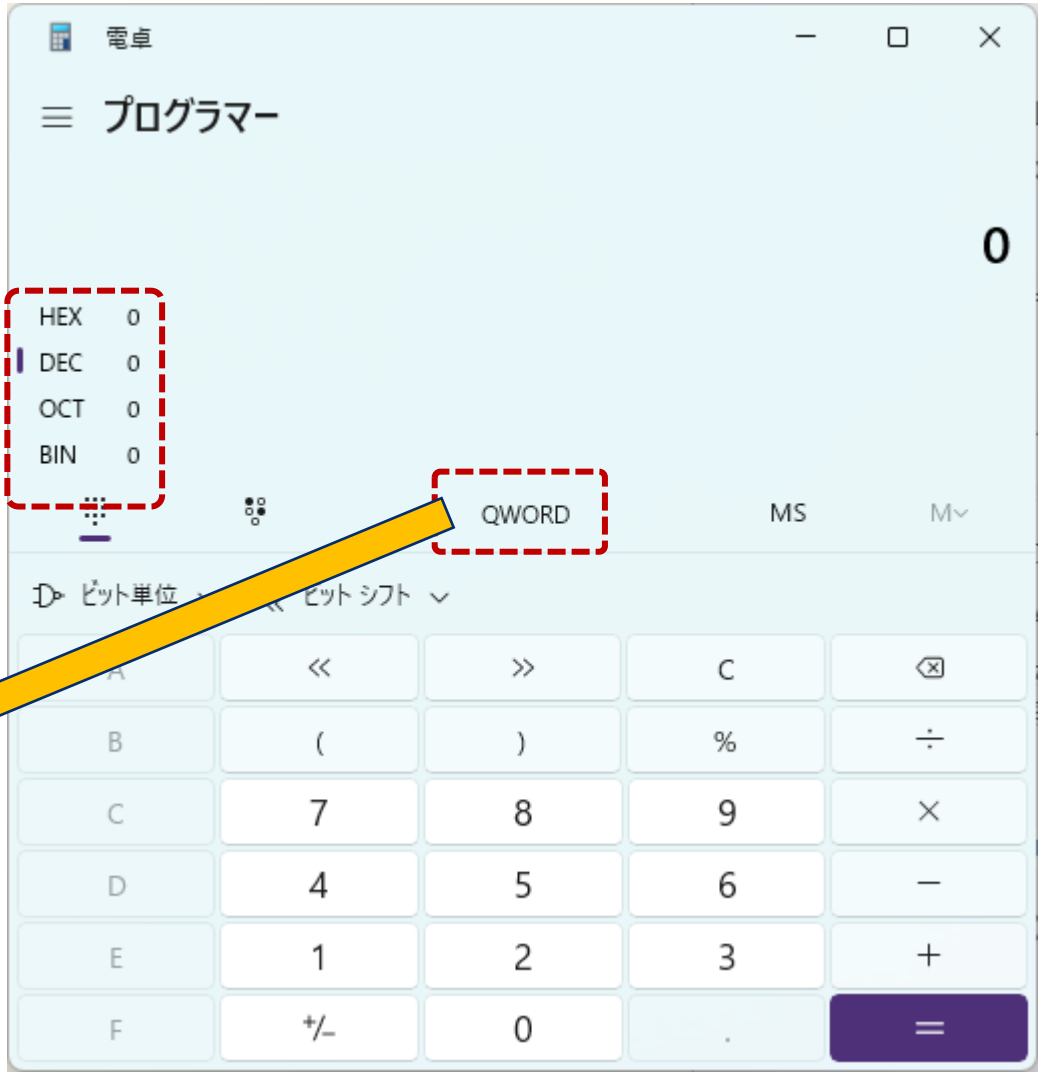
無論Excelの関数使ってもできる

16進数
10進数
8進数
2進数



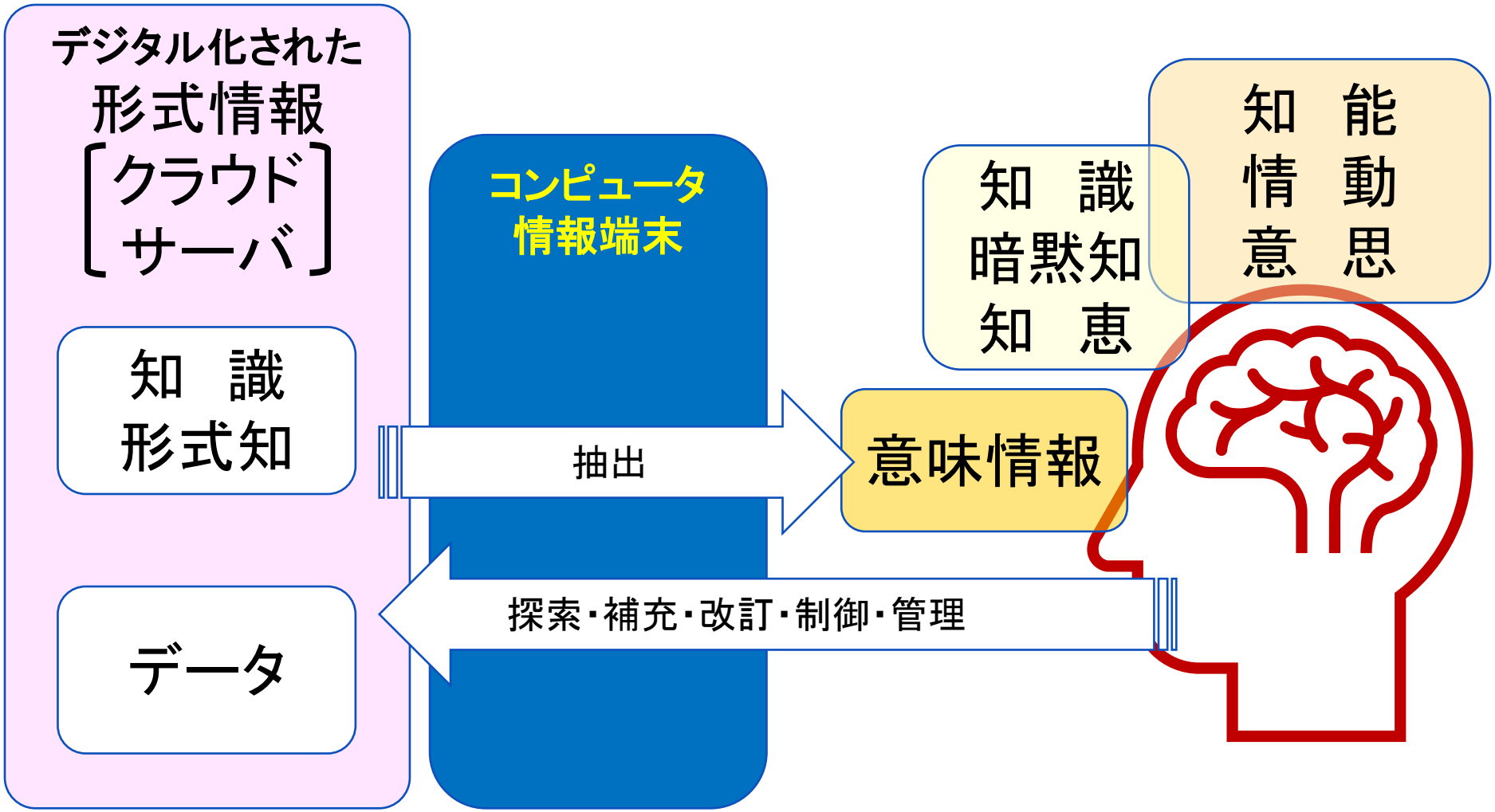
8bit
16bit
32bit
64bit





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

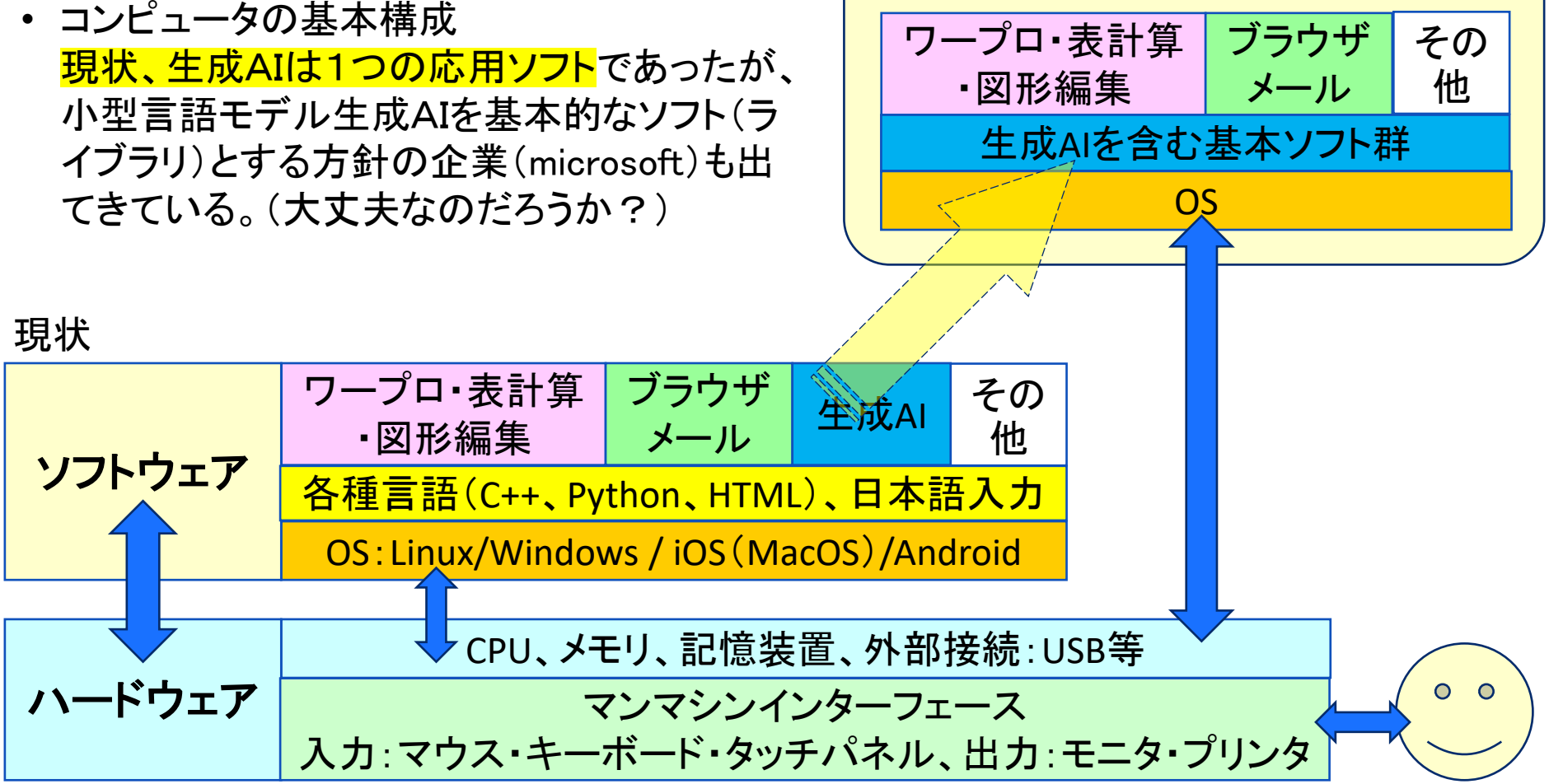


歴史と概要

- 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？？



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。

$a=18.0$

$b=2^{53}$

$b=2^{57}$

$b=2^{58}$

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
```

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
```

R 4.1.2

pit) - [R Console]

ファイル 編集 閲覧 その他 パッケージ ウインドウ ヘルプ

```
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	=((A1+A2)-A2)
b	1.44115E+17	=2^57
((a+b)-b)	32	=((A1+A4)-A4)
b	2.8823E+17	=2^58
((a+b)-b)	0	=((A1+A6)-A6)

Python 3.9.7

File Edit Shell Debug Options Window Help

```
Python 3.9.7 (tags/v3.9.7:1016ef3, Aug 30 202
v.1929 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "licen
formation.
>>> 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
>>>
```

```
IDLE Shell 3.13.0
File  Edit  Shell  Debug  Options  Window  Help
Python 3.13.0 (tags/v3.13.0:60403a5, Oct  7 2024, 09:38:07) [MSC v.1941 64 bit
AMD64)] on win32
Type "help", "copyright
>>> 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
>>> aa = 18
>>> b = 2 ** 57
>>> print ((aa + b) - b)
18
>>> b = 2 ** 58
>>> print ((aa + b) - b)
18
>>> b = 2 ** 100
>>> print ((aa + b) - b)
18
```

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」 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.
※Pythonの場合には整数に（メモリの制約以外）上限はないので整数演算すれば、問題は発生しない。

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

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

Excel (365)

	A	B
1	= 2^3^2	64
2	= 2^-3^2	0.015625
3		

R 4.1.2 Terminal Jobs

R 4.1.2 · ~/

```
> 2^3^2
[1] 512
> 2^-3^2
[1] 0.001953125
>
```

MATLAB 2021b

```
>> 2^3^2

ans =

    64

>> 2^-3^2

ans =

    0.0156
```

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
>>>
```

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。

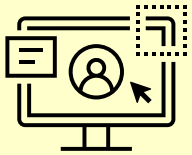
https://www.cais.niigata-u.ac.jp/system/pc/softlist.html

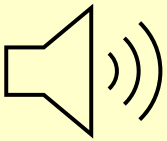
機械と人間とのインターフェース

入力装置・文字コードの問題

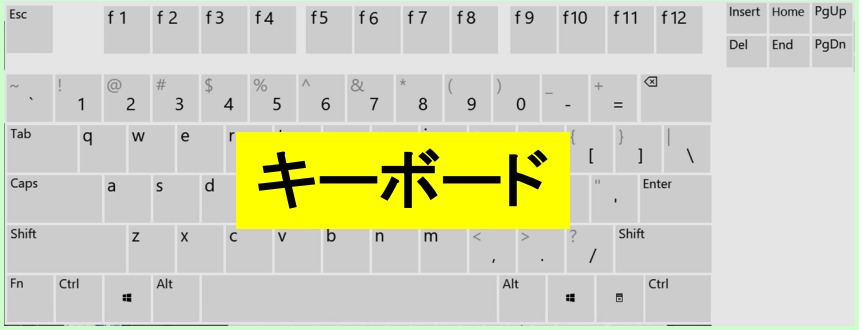
英文タイプライター

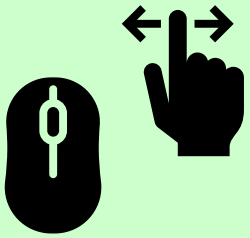
出力（ディスプレイ、印刷など）



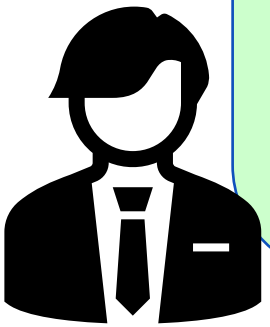


コンピュータ



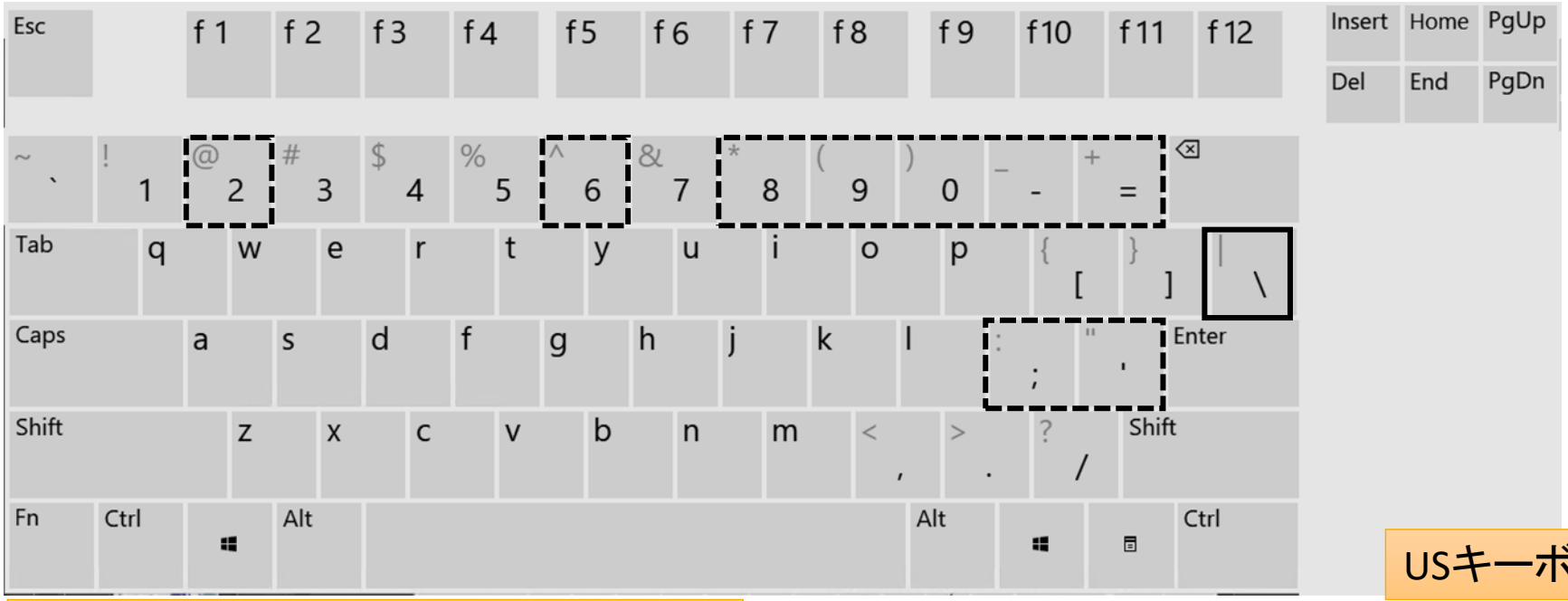


入力デバイス



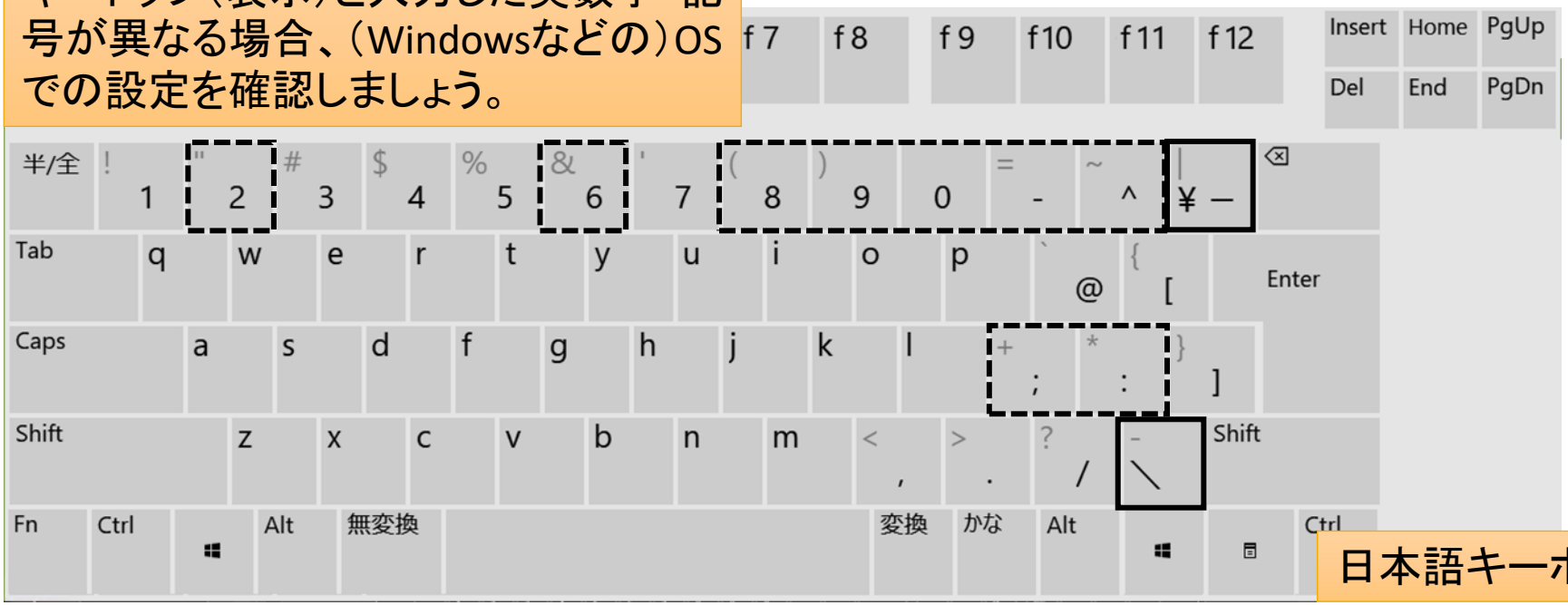
タイプライターから
コンピュータへ

- 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）

+【Ctrl】　　±【Shift】

上位ビット ＼ 下位ビット	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
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	ァ	ィ	ゥ	ェ	ォ	ャ	ュ	ョ	ッ
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	゛	゜
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。

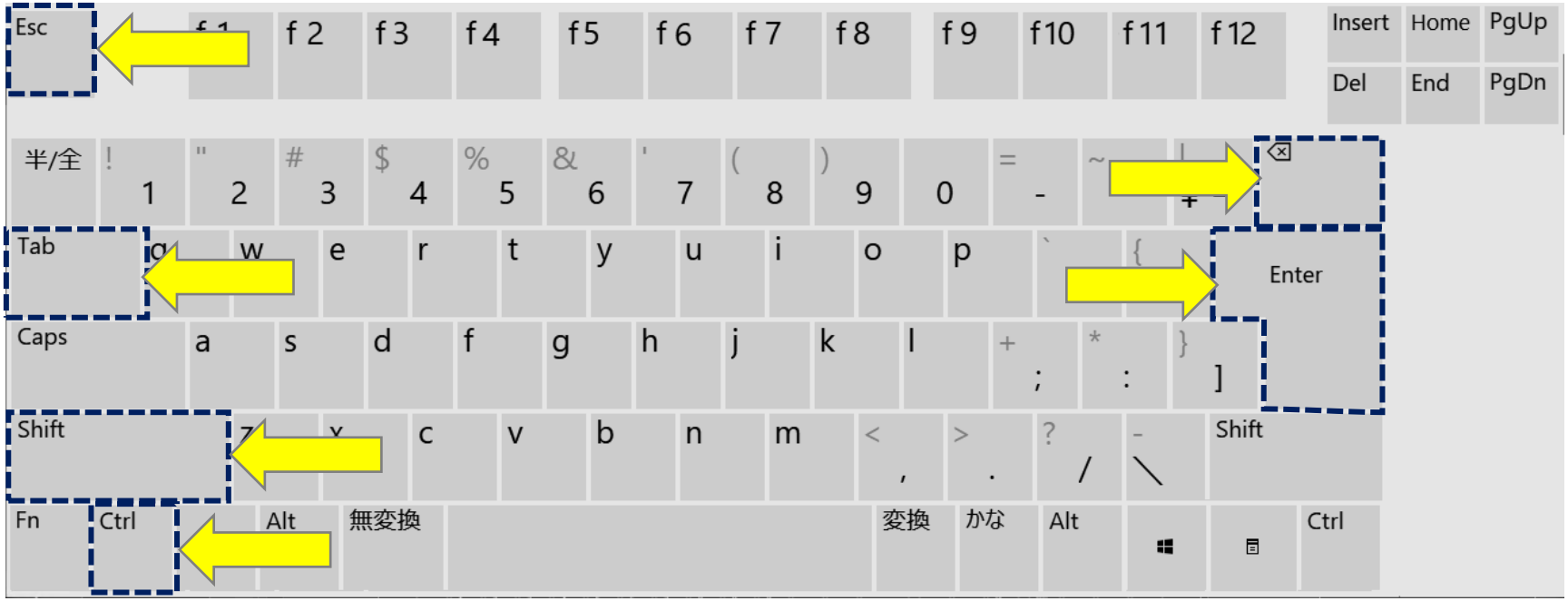
※ANKはJIS X 0201(旧JIS C 6220)規格の俗称で「Alphabet + Numerical digit + Katakana」の頭文字。

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（アンク；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きているが、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

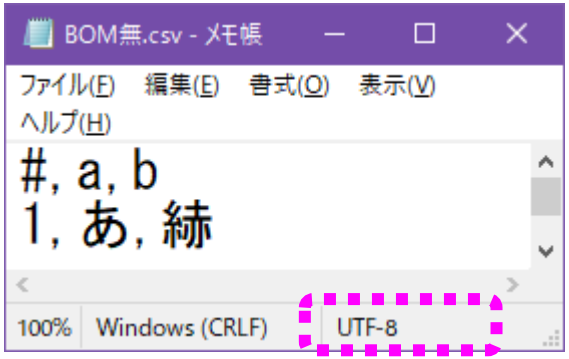
コンピュータで扱える文字（2）
2バイト以上で扱える漢字・各国語への拡張

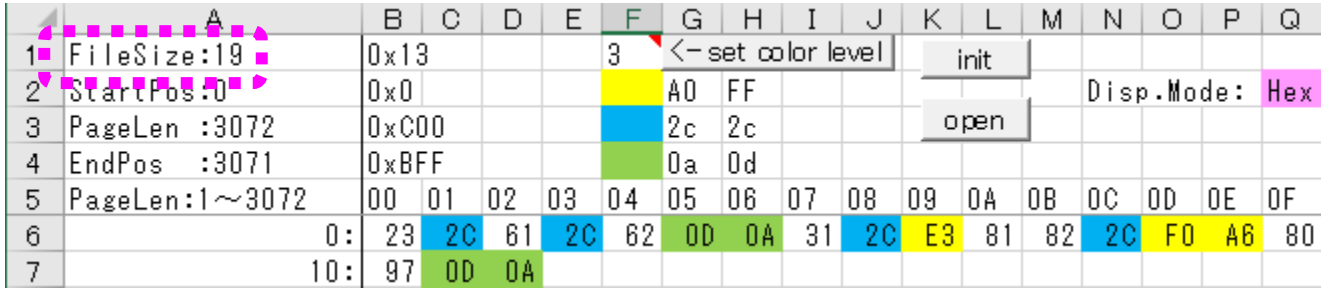
- 16bit（2byte）: 65536文字
  - JIS（ISO 2022-JP）
  - Shift-JIS（S-JIS, MS漢字コード）
  - EUC（Extended UNIX Code）
    - それぞれにコード体系が異なるので文字化けが生じる。
- 8-32bit（1-4byte）:
  最大4294967296文字（約43億文字）
  - Unicode
  - UTF-8, UTF-16, UTF-32など
  - 注意：近年UTF-8が主流となっているが、BOMの有無にて、処理時に副作用の原因になる場合がある。

BOM:　Byte Order Mark

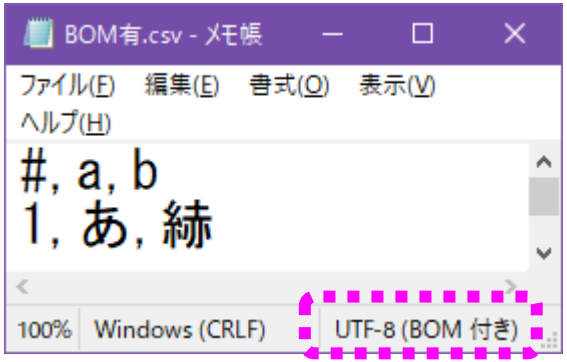
7-8 bit	EBCDIC		(extended binary coded decimal interchange code)
1 byte			初期のIBMコンピュータにて用いられた。
128-256文字	ASCII		(American National Standard Code for Information Interchange)
			情報交換用米国標準コード(7bit)
	ANK		(Alphabetic Numeric and Kana)
			JIS X 0201
			ASCIIコードに、カナを加えたもの(8bit)
16 bit	JIS		(ISO 2022-JP)
			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

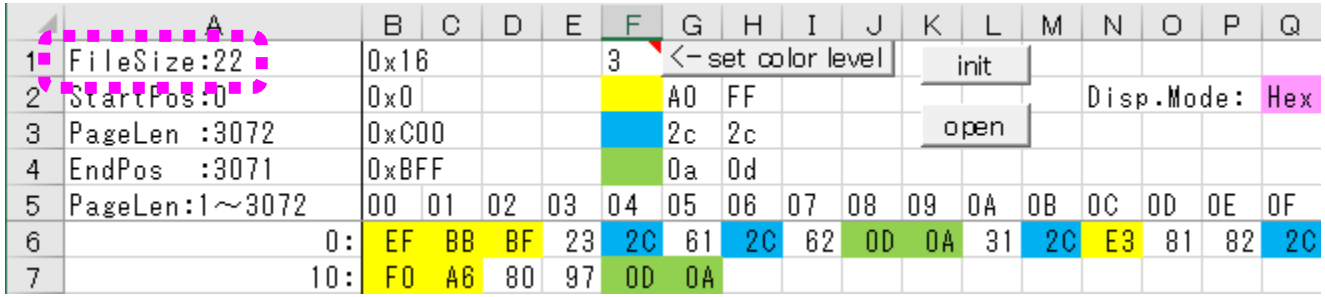
「BOM無」と「BOM付」の違い





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

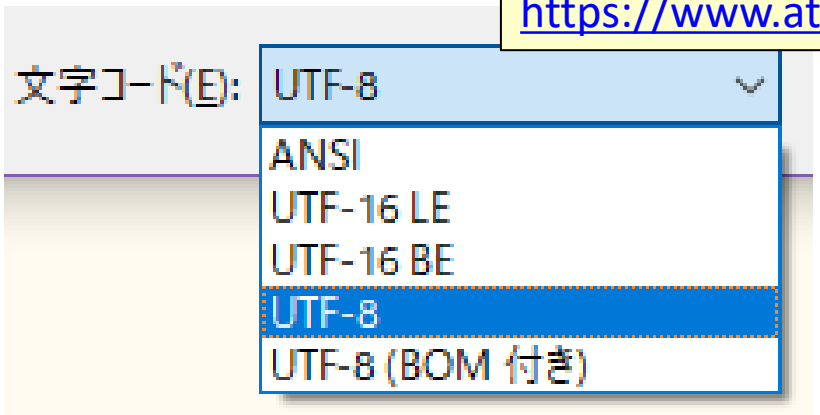




[BOM]			#	,	a	,	b	↵		1	,		あ			,		絲				↵	
EF	BB	BF	23	2C	61	2C	62	0D	0A	31	2C	E3	81	82	2C	F0	A6	80	97	0D	0A

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': '絲'}

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': '絲'}
```

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': '赫'}
```

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': '赫'}
```

PowerShellでのUTF-8関連問題（歴史的背景含む）

- Windows PowerShell と PowerShellの違い
  - デフォルトでインストール済みなのはWindows PowerShellなので注意
  - https://learn.microsoft.com/ja-jp/powershell/scripting/whats-new/differences-from-windows-powershell?view=powershell-7.5
- https://learn.microsoft.com/ja-jp/powershell/module/microsoft.powershell.core/about/about_character_encoding?view=powershell-7.5
- https://virment.com/how-to-install-latest-powershell-in-windows10/

Windows PowerShell での文字エンコード

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 バイト順で UTF-16 を使用します。
- `BigEndianUTF32` ビッグ エンディアン 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ドを使用します。
- `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 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 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 あり) を使用します。

https://learn.microsoft.com/ja-jp/powershell/module/microsoft.powershell.core/about/about_character_encoding?view=powershell-7.5

一般に、Windows PowerShell では既定で Unicode UTF-16LE エンコードが使用されます。ただし、Windows PowerShell のコマンドレットで使用される既定のエンコードは一貫性がありません。

> ⓘ 注意
>
> `UTF7`を除く Unicode エンコードを使用すると、常に BOM が作成されます。

PowerShellでの文字エンコード

PowerShell (v7.1 以降) 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) 文字セットのエンコードを使用します。
- `ansi`: 現在のカルチャの ANSI コード ページのエンコードを使用します。このオプションは PowerShell 7.4 で追加されました。
- `bigendianunicode`: ビッグ エンディアン バイト順を使用して UTF-16 形式でエンコードします。
- `bigendianutf32`: ビッグ エンディアン バイト順を使用して UTF-32 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 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 オーダー 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 オーダー マーク (BOM) を使用せずに UTF-8 形式でエンコードします。
- `utf32`: リトル エンディアン バイト順を使用して UTF-32 形式でエンコードします。

PowerShell の既定では、すべての出力に対して `utf8NoBOM` されます。

PowerShell 6.2 以降では、**Encoding** パラメーターを使用すると、登録済みのコード ページの数値 ID (`-Encoding 1251` など) や登録済みコード ページの文字列名 (`-Encoding "windows-1251"` など) も使用できます。詳細については、Encoding.CodePage の .NET ドキュメントを参照してください。

PowerShell 7.4 以降では、`ANSI` パラメーターの値を使用して、手動で指定しなくても、現在のカルチャの ANSI コード ページの数値 ID を渡すことができます。

Windows 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,縺・。・
PS:Jupyter >type withBOM.csv
#,a,b
1,あ,縺
```

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,縺・。・

E:¥Python¥Jupyter>type withBOM.csv
・ﾉ#,a,b
1,縺・。・
```

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			,	縺				↵	
23	2C	61	2C	62	OD	OA	31	2C	E3	81	82	2C	F0	A6	80	97	OD	OA

withBOM.csvの構造

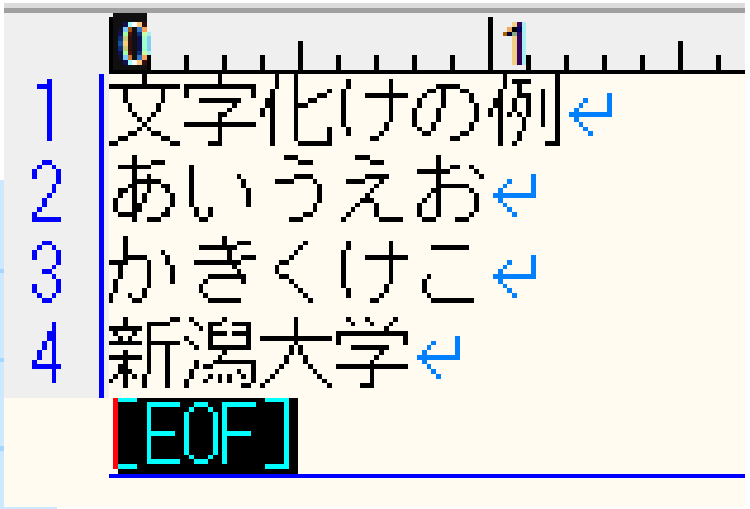
[BOM]			#	,	a	,	b	↵		1	,	あ			,	縺				↵	
EF	BB	BF	23	2C	61	2C	62	OD	OA	31	2C	E3	81	82	2C	F0	A6	80	97	OD	OA

全文検索におけるWindowsでのUTF-8（BOM無し）の問題

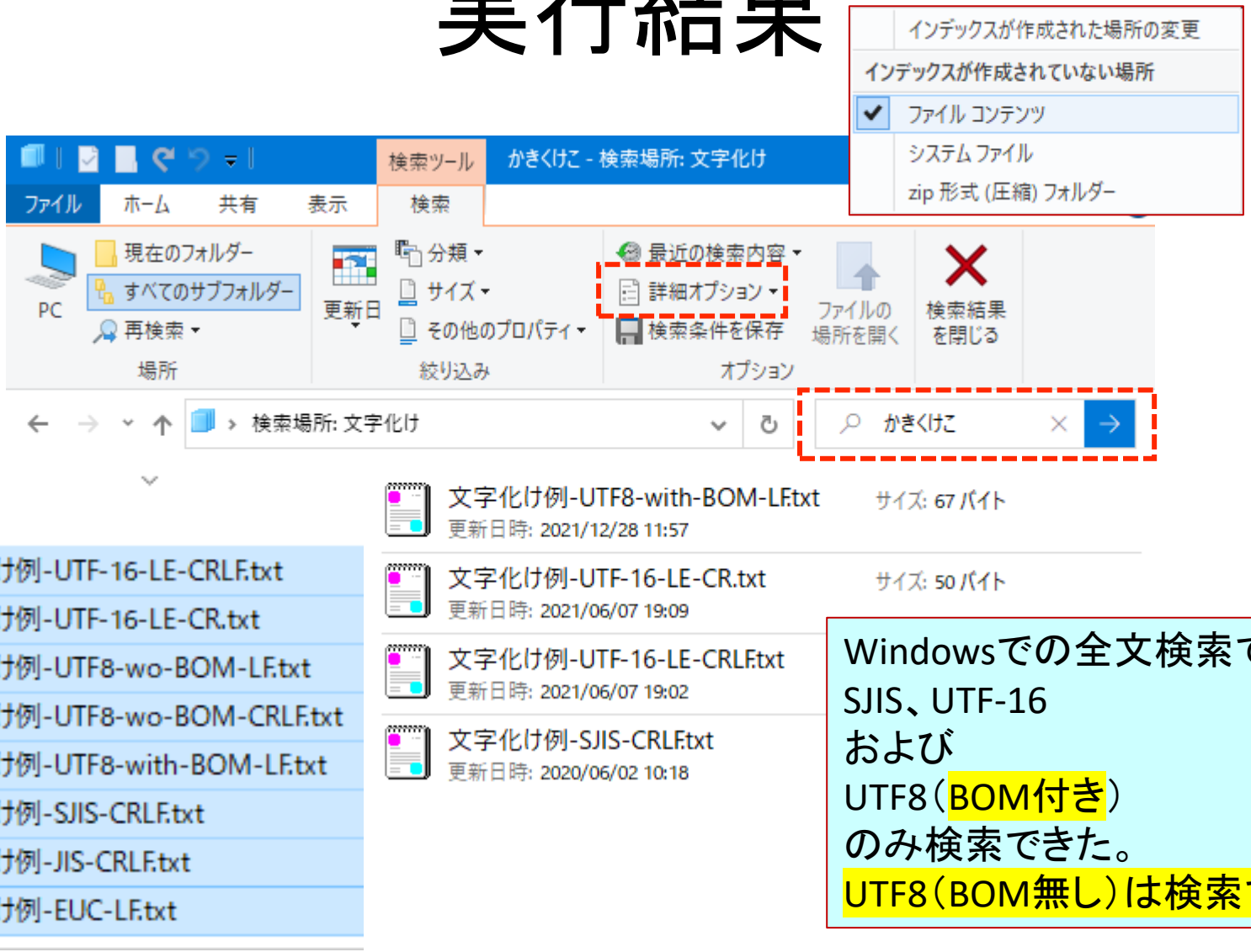
全文検索用のサンプルテキスト

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



Windowsの全文検索の設定と実行結果



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

2021年06月13日 10時00分更新　　文● 塩田紳二　編集● ASCII

参考資料

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

名前

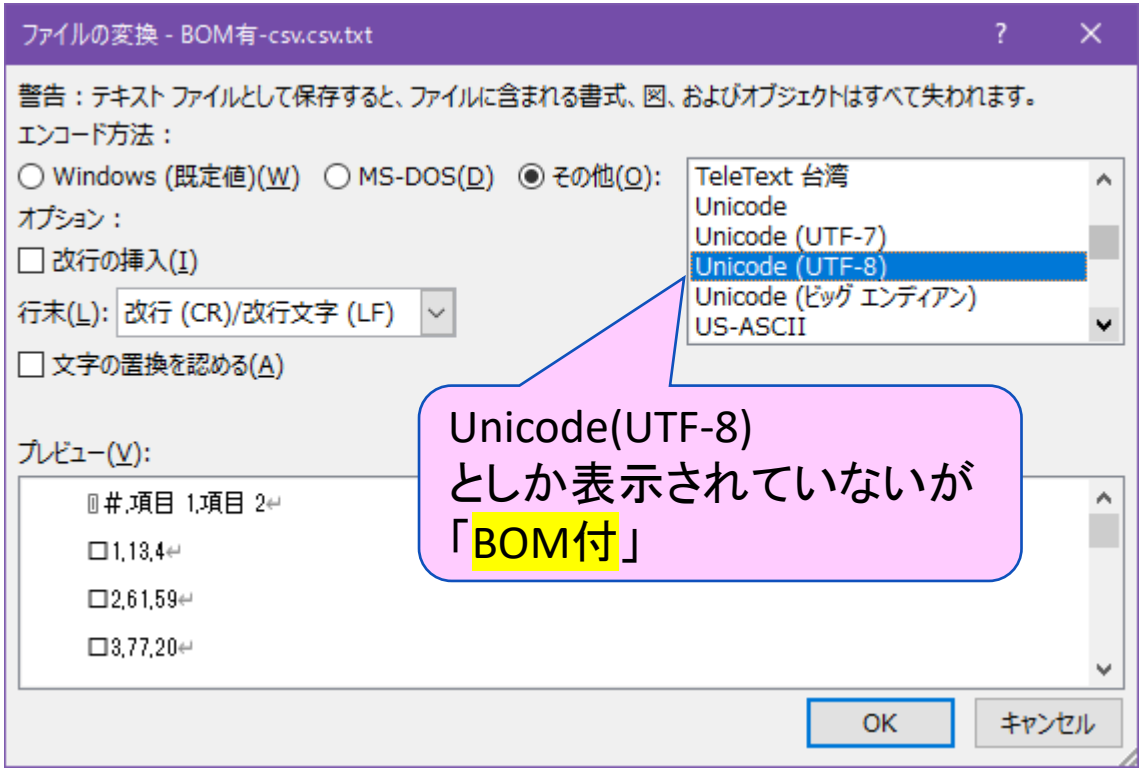
- 文字化け例-UTF-16-LE-CRLF.txt
- 文字化け例-UTF-16-LE-CR.txt
- 文字化け例-UTF8-wo-BOM-LF.txt
- 文字化け例-UTF8-wo-BOM-CRLF.txt
- 文字化け例-UTF8-with-BOM-LF.txt
- 文字化け例-SJIS-CRLF.txt
- 文字化け例-JIS-CRLF.txt
- 文字化け例-EUC-LF.txt

```
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

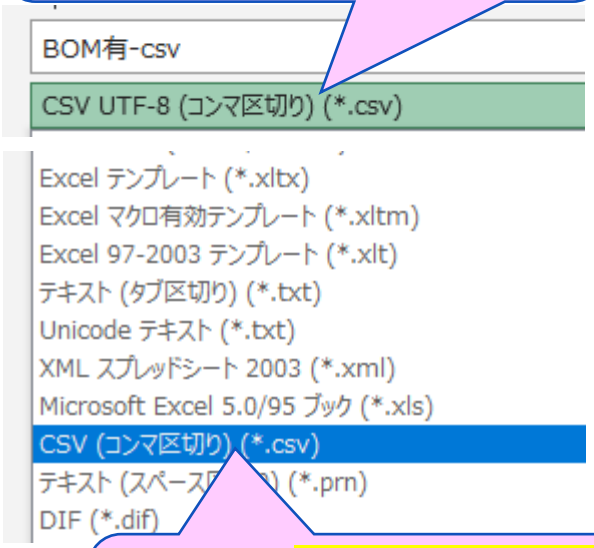
ワードでの保存時



Unicode(UTF-8)
としか表示されていないが
「BOM付」

エクセルでの保存時

UTF-8
としか表示されていないが
「BOM付」



こちらは「Sift-JIS」（SJIS）
なのでMAC等で文字化
けする

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！！

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(F) 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) ANSI

Windows10
JIS-CRLF として読み込む

ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)

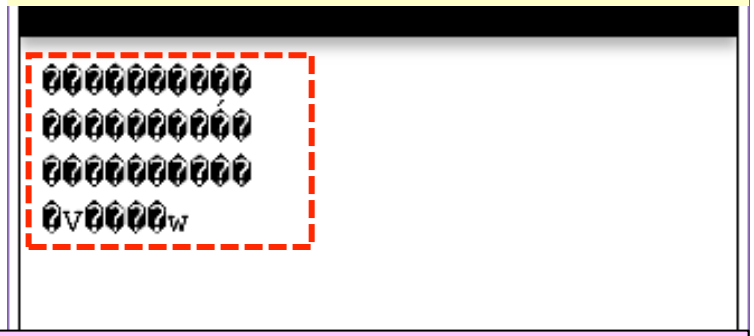
·$BJ8;z2=$1$NNc·(B
·B"$$$&$($*·(B
·B+$-$/$1$3·(B
·$B?73cBg3X·(B

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



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 Microsoft 365 を購入する

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る

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。

Subject: Re: =?ISO-2022-JP?B?GyRCO3UzWDJxGyhCSFAbJEIzTkcnTVEkTiVVGyhK?=
=?ISO-2022-JP?B?GyRCJSElJCVrM05HJyROJCo0aiQkGyhC?=
In-Reply-To: <####################@dent.niigata-u.ac.jp>
References: <####################@dent.niigata-u.ac.jp>
MIME-Version: 1.0
X-Mailer: AL-Mail32 Version 1.13
Content-Type: text/plain; charset=iso-2022-jp

送信側での改行部分
受信側で文字化けになる

「ISO-2022-JP」でエンコードされたタイトル

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?GyRCJSElJCVrM05HJyROJCo0aiQkGyhC?=" | wsl nkf -w
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?UmU6IOatr+WtpuS8mkhQ56K66KqN55So44Gu44OV77+944Kh44Kk44Or?=
=?UTF-8?B?56K66KqN44Gu44GK6aGY44GE?=
Content-Language: en-US

「UTF-8」でエンコードされたタイトル

echo "=?UTF-8?B?UmU6IOatr+WtpuS8mkhQ56K66KqN55So44Gu44OV77+944Kh44Kk44Or?=" | wsl nkf -wm

送信側改行部分

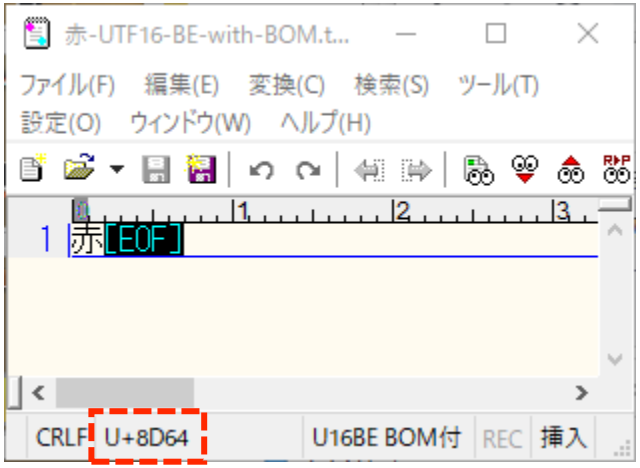
Re: 歯学会HP確認用のフ◆ァイル

受信側改行部分

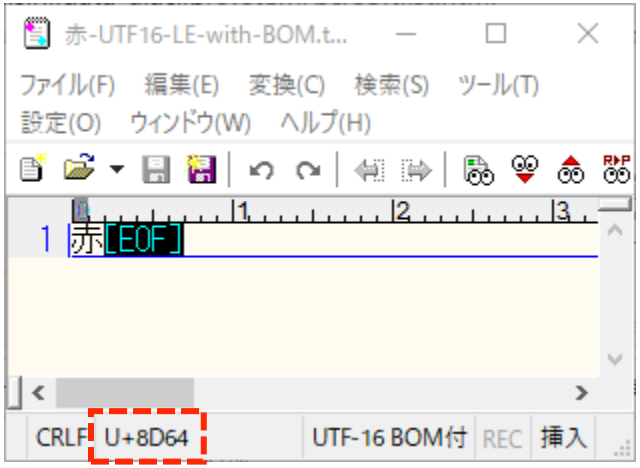
echo "=?UTF-8?B?56K66KqN44Gu44GK6aGY44GE?=" | wsl nkf -wm
確認のお願い

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

ビッグエンディアンとリトルエンディアン
UTF-16BE と UTF-16LE



「赤」はUnicodeで「8D64」



「赤」はUnicodeで「8D64」

UTF-16BE

	A	B	C	D	E	F	G	H	I	J
1	FileSize:4	0x4				3	<-set color level			
2	StartPos:0	0x0					fe	fe		
3	PageLen :3072	0xC00					ff	ff		
4	EndPos :3071	0xBFF					50	70		
5	PageLen:1～3072	00	01	02	03	04	05	06	07	08
6	0:	FE	FF	8D	64					
7										

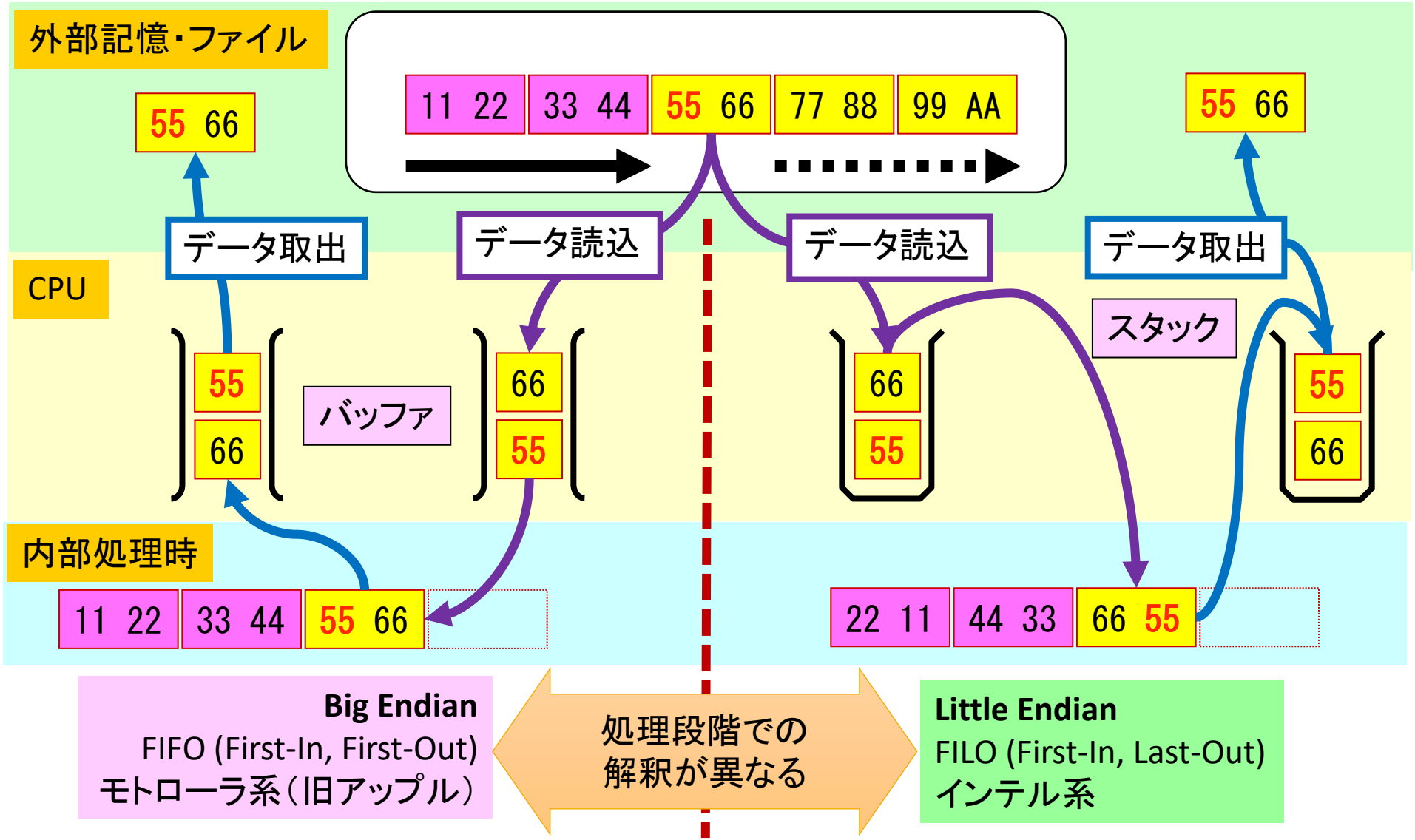
UTF-16LE

	A	B	C	D	E	F	G	H	I	J
1	FileSize:4	0x4				3	<-set color level			
2	StartPos:0	0x0					fe	fe		
3	PageLen :3072	0xC00					ff	ff		
4	EndPos :3071	0xBFF					50	70		
5	PageLen:1～3072	00	01	02	03	04	05	06	07	08
6	0:	FF	FE	64	8D					
7										

コンピュータ内部での解釈の問題

Endian：2バイト単位での例。

⇒ 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



DICOMでのTransfer Syntax（転送構文）でのEndianの使用例

英文	和文
Implicit VR Little Endian (※)	暗黙的VRリトル・エンディアン（非圧縮）
Explicit VR Little Endian	明示的VRリトル・エンディアン(非圧縮）
Explicit VR Big Endian	明示的VRビッグ・エンディアン(非圧縮）
JPEG Lossless	JPEG（可逆圧縮）
JPEG Lossy	JPEG（非可逆圧縮）

参考資料：放射線領域における標準規格 放射線領域における標準規格（DICOMの概要）IHEワークショップ ワークショップ in 大津
JIRAシステム部会本委員会 飯島直人（（株）島津製作所）
2007/8/11 IHE workshop in 大津
https://www.ihe-j.org/file2/n17/WS11_06_DICOM_iijima.pdf

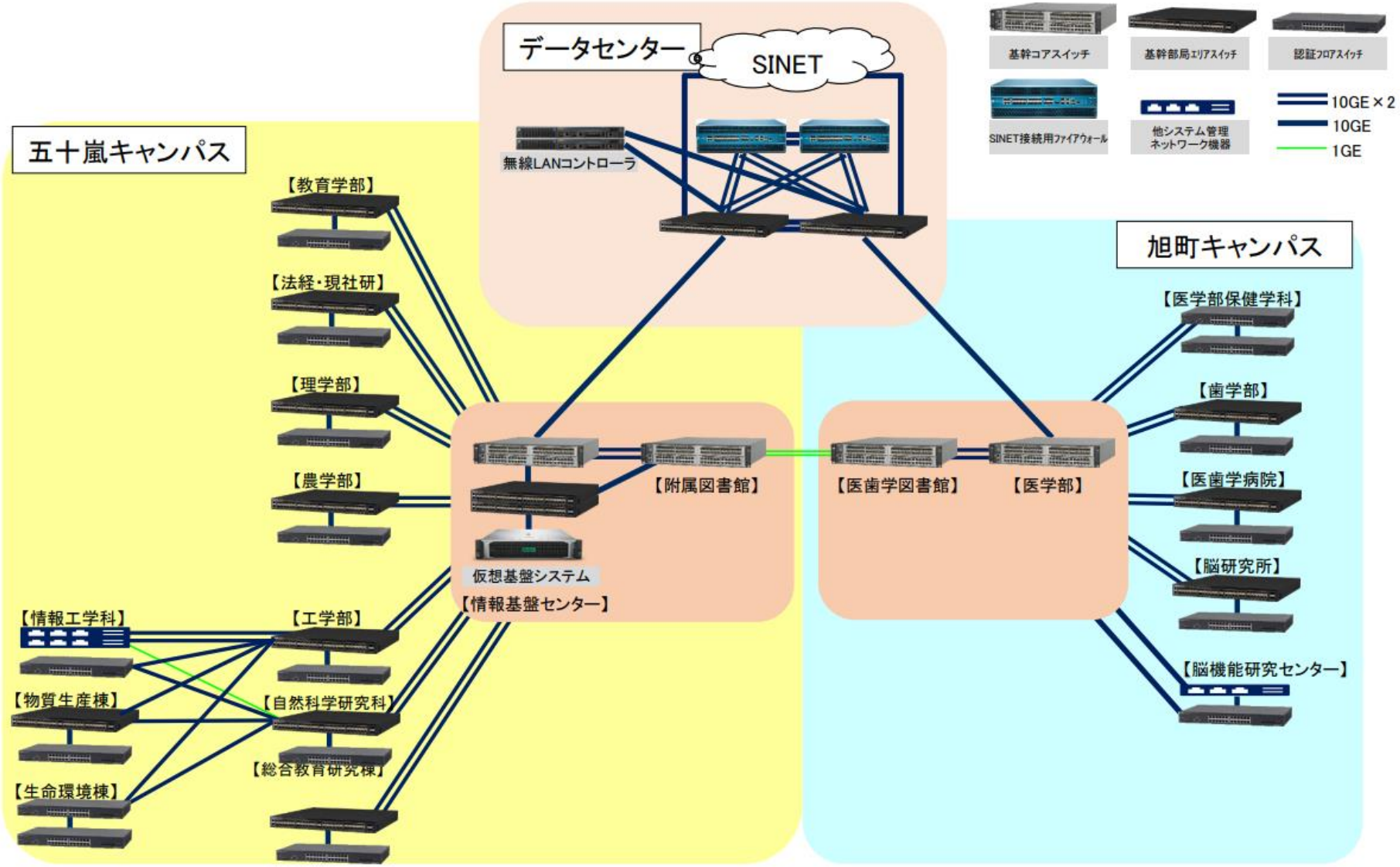
大学ネットワークの基盤について

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



学術情報ネットワーク SINET6

お問合せ　English

Home　SINETとは　接続・サービス　申請手続き　活用事例　お知らせ　イベント　FAQ

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

クラウド提供業者の皆様

最新の投稿

2025/05/02
SINET6 メンテナンス連絡(2025/6/21・6/28 データセンタ接続冗長化サービス)

2025/04/22
NII学術情報基盤オープンフォーラム2025開催のお知らせ

2025/04/07
SINET6 メンテナンス連絡(2025/5/17-6/9)

学術情報ネットワークとは

https://www.sinet.ad.jp/aboutsinet

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GÉANT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。

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の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

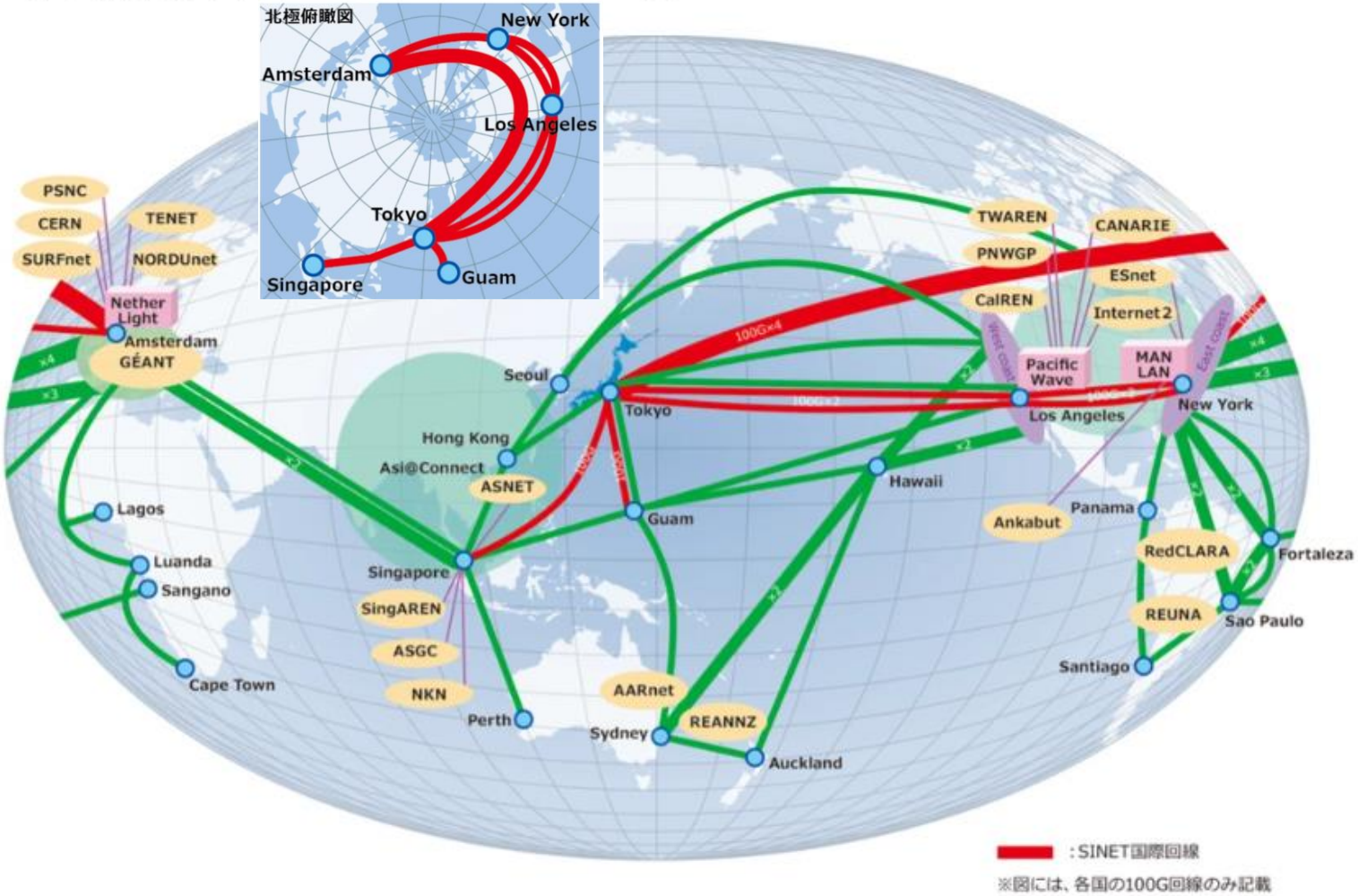
タグ

APOnet (1) ・ BGP (1) ・ IP Dual (19) ・ L1オンデマンド (6) ・ L2VPN (24) ・ L3VPN (7) ・ MMCFTP (1) ・ QoS (1) ・ SINET6 (1) ・ クラウド (9) ・ フルルート (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)

国際相互接続

https://www.sinet.ad.jp/aboutsinet/interconnectivities

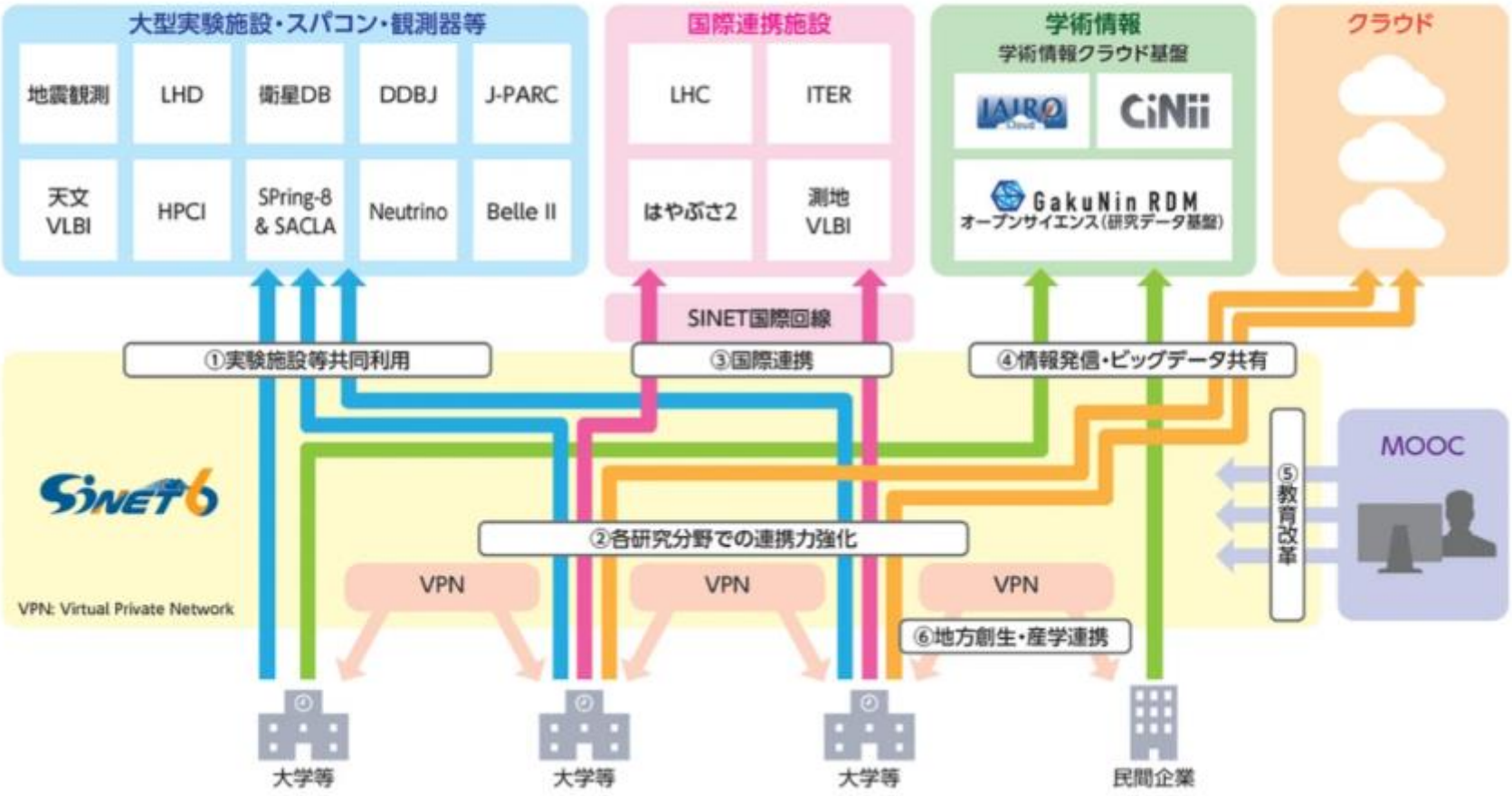
国際的な先端研究プロジェクトで必要とされる国際間の研究情報流通を円滑に進められるように、海外の学術情報ネットワークとの通信を確保しています。米国Internet2や欧州GÉANT等をはじめとした多くの海外研究ネットワークと相互接続しています。
そして、アジア地域との学術研究情報の流通促進を強化するために、Asi@Connectと連携・協力し、国際学術情報ネットワークの構築に貢献しています。



https://www.sinet.ad.jp/aboutsinet

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

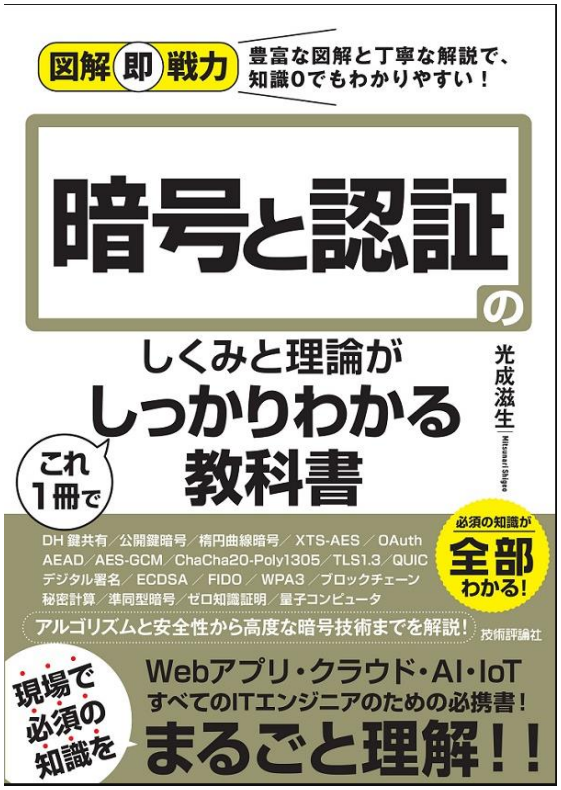
暗号と認証

- 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

https://www.cryptrec.go.jp/index.html

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
Cryptography Research and Evaluation Committees

https://www.cryptrec.go.jp/index.html

ENGLISH

- 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。

https://www.cryptrec.go.jp/topics/cryptrec-er-0001-2019.html

東大、これまでに解かれたことのない次元の暗号解読を実現

https://www.i.u-tokyo.ac.jp/news/files/IST-Pressrelease-Takagi_20231023.pdf

- https://pc.watch.impress.co.jp/docs/news/1541846.html

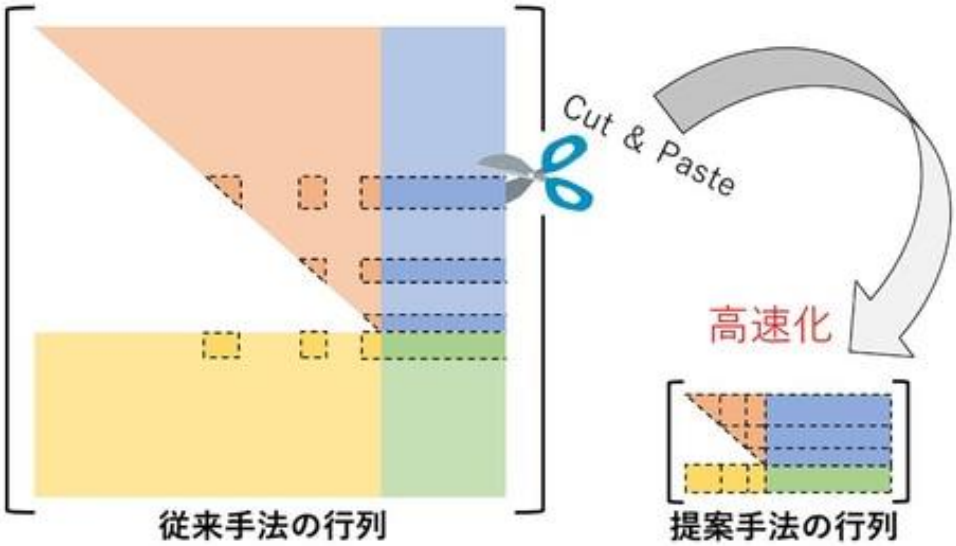


図1：提案アルゴリズムのイメージ図

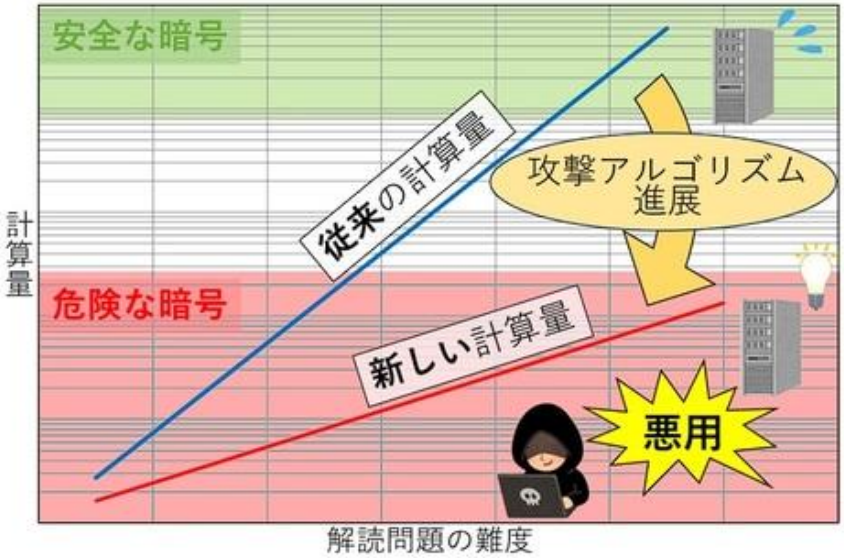
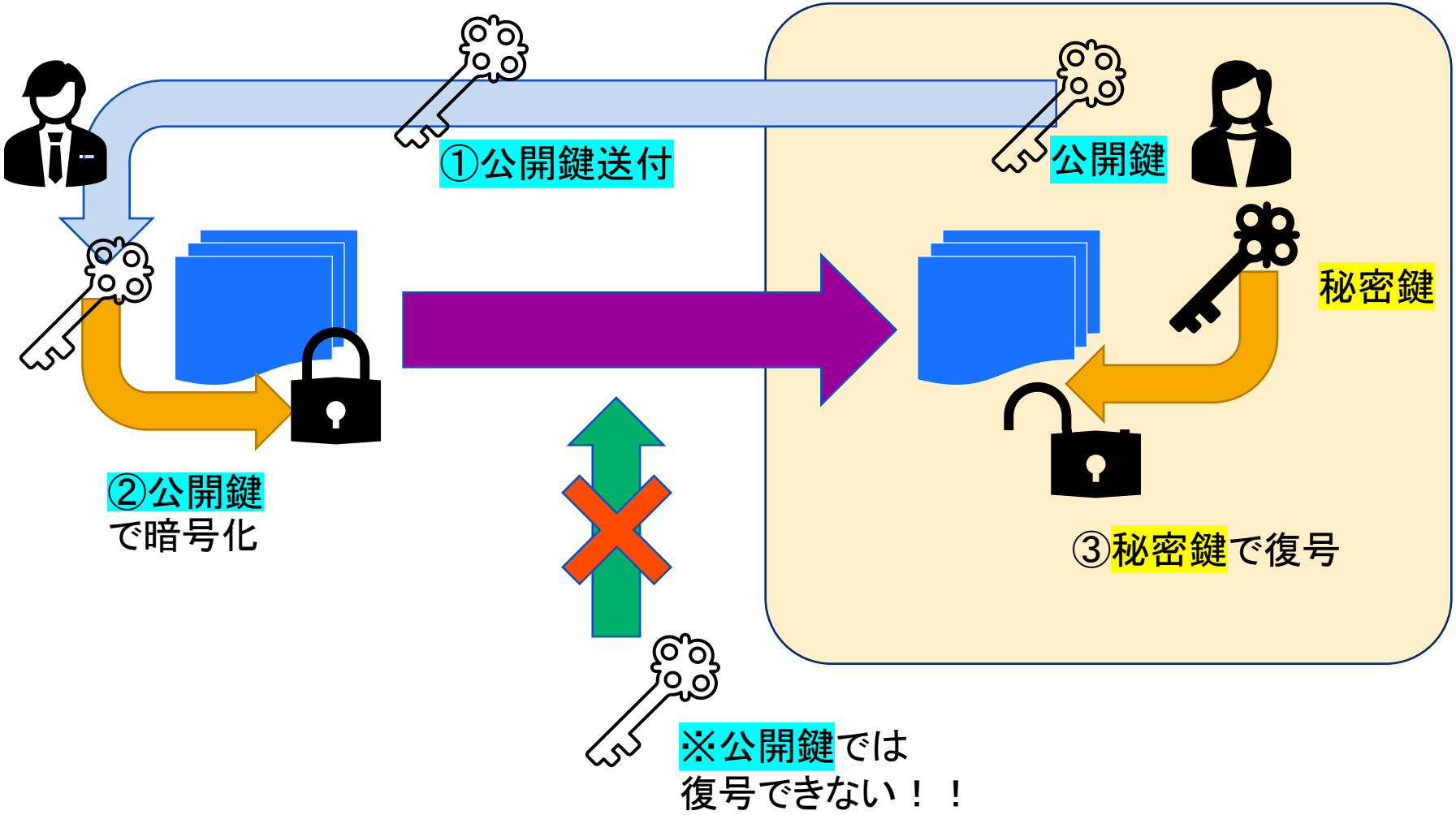


図2：解読問題の難度と問題の計算量の関係

公開鍵と秘密鍵を使う
公開鍵暗号方式



Windows10での公開鍵生成方法

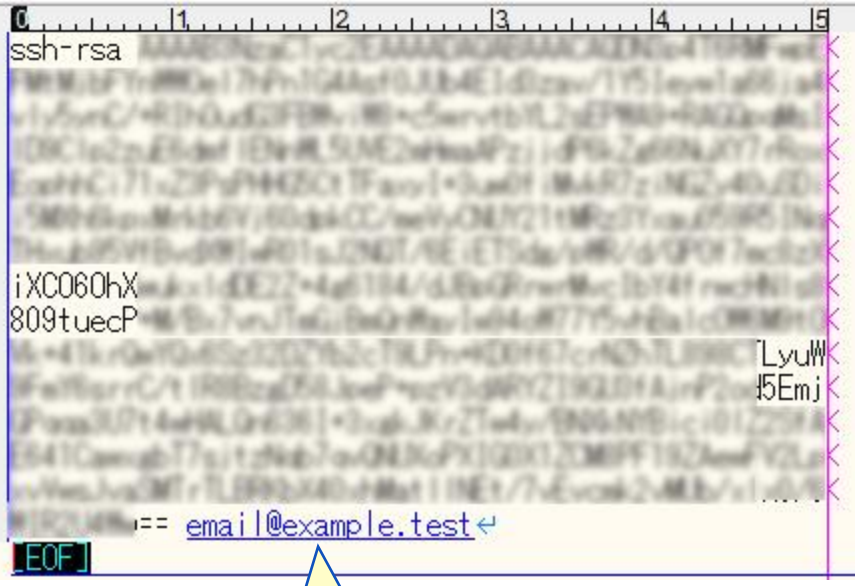
https://qiita.com/c60evaporator/items/ebe9c6e8a445fed859dc
https://docs.microsoft.com/ja-jp/windows-server/administration/openssh/openssh_keymanagement

```
PS:[redacted] > ssh-keygen -t rsa -b 4096 -C "email@example.test"
Generating public/private rsa key pair.
Enter file in which to save the key (C:¥Users¥[redacted]id_rsa): e:¥test/id_rsa
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in e:¥test/id_rsa.
Your public key has been saved in e:¥test/id_rsa.pub.
The key fingerprint is:
SHA256:[redacted]Pa[redacted]0 email@example.test
The key's randomart image is:
+---[RSA 4096]----+
|      .+[redacted]|
|      . O  + +oBo|
|[redacted]. * o|
|[redacted]+[redacted]|
|[redacted]. [redacted]|
|[redacted]o.[redacted]|
|    o [redacted]+ . [redacted]|
|   . + [redacted] . [redacted]|
|    o.        .o|
+----[SHA256]-----+
PS:[redacted] >
```

公開鍵の作成例
ssh-keygenコマンドで作成できる。
公開鍵は「id_rsa.pub」
秘密鍵は「id_rsa」
に保管される。

参考：Windows 用 OpenSSH でのキーベースの認証

- https://learn.microsoft.com/ja-jp/windows-server/administration/openssh/openssh_keymanagement



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。

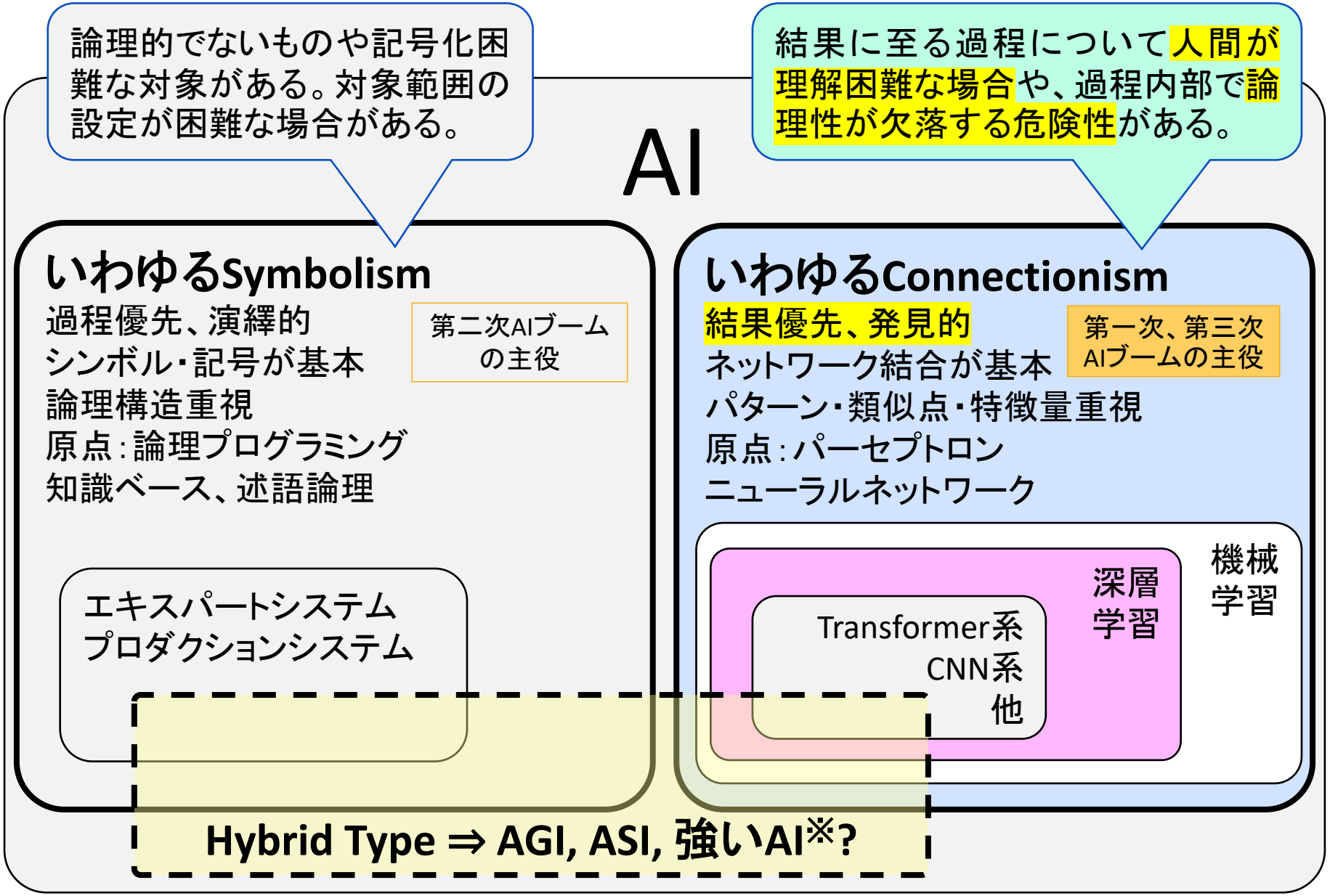
AIとは？
深層学習、機械学習とは？

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

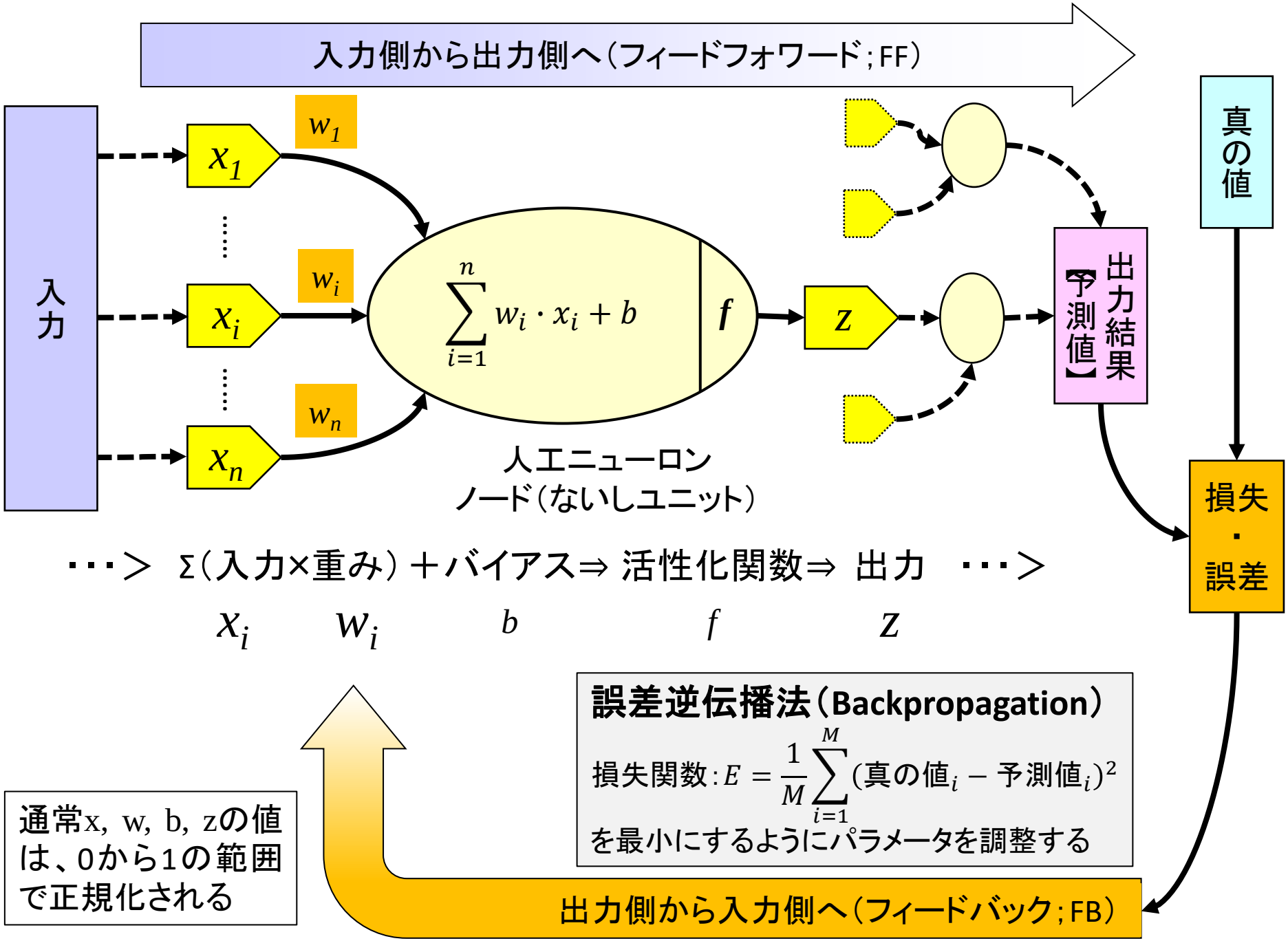
	人工知能の置かれた状況	主な技術等	人工知能に関する出来事
1950年代			チューリングテストの提唱（1950年）
1960年代	第一次人工知能ブーム（探索と推論）	・探索、推論 ・自然言語処理 ・ニューラルネットワーク	ダートマス会議にて「人工知能」という言葉が登場（1956年） ニューラルネットワークのパーセプトロン開発（1958年） 人工対話システムELIZA開発（1964年）
1970年代	冬の時代	・エキスパートシステム ・遺伝的アルゴリズム	初のエキスパートシステムMYCIN開発（1972年） MYCINの知識表現と推論を一般化したEMYCIN開発（1979年）
1980年代	第二次人工知能ブーム（知識表現）	・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		・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）
	冬の時代	・統計的自然言語処理	
2000年代	第三次人工知能ブーム（機械学習）	・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）

総務省・平成28年版　情報通信白書「図表4-2-1-5人工知能（AI）の歴史」（一部改変）
https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/html/nc142120.html
https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/image/n4201050.png

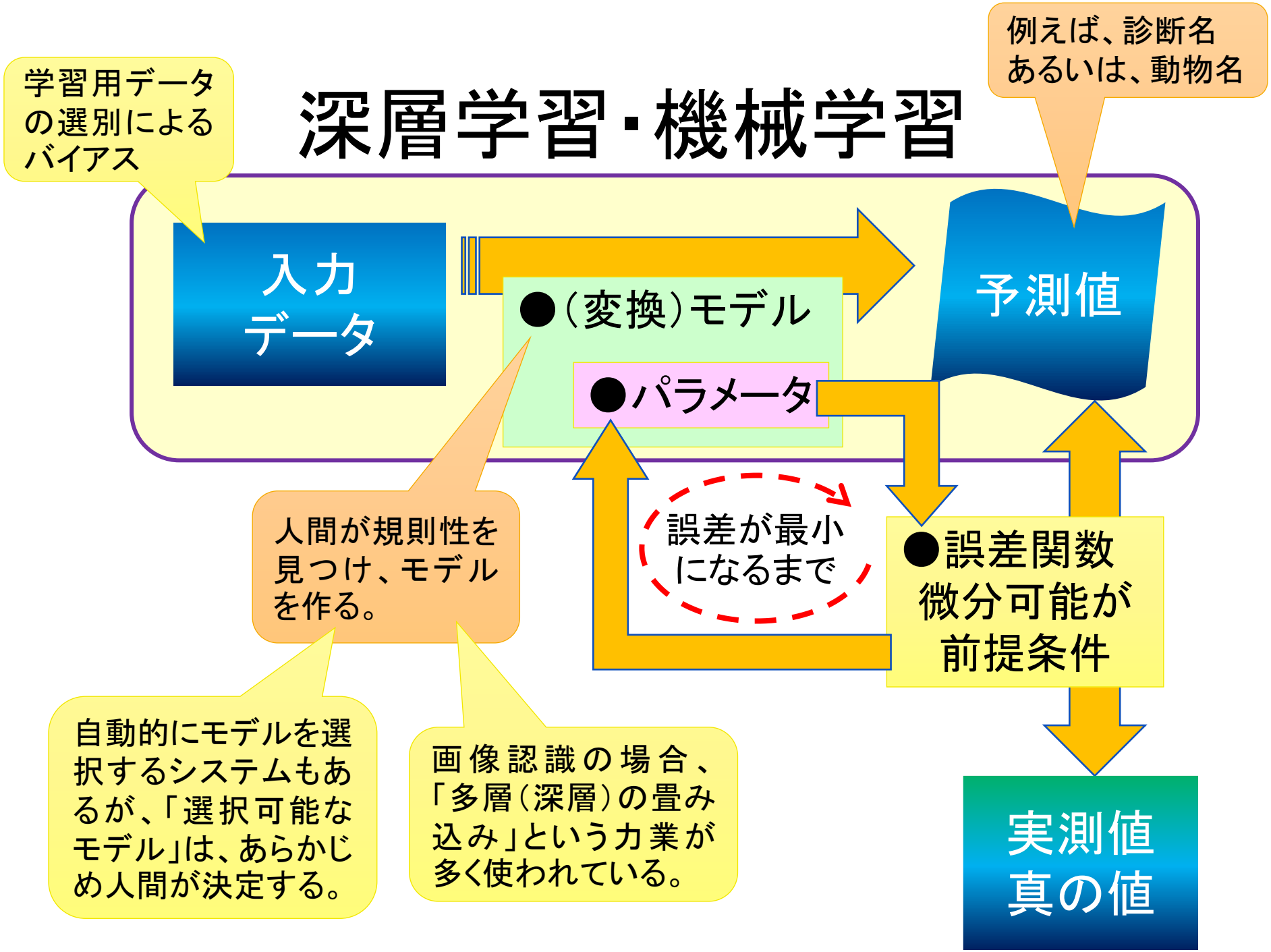


※AGI（Artificial General Intelligence、汎用人工知能）、ASI（Artificial Super Intelligence、超人工知能≒「強いAI」）

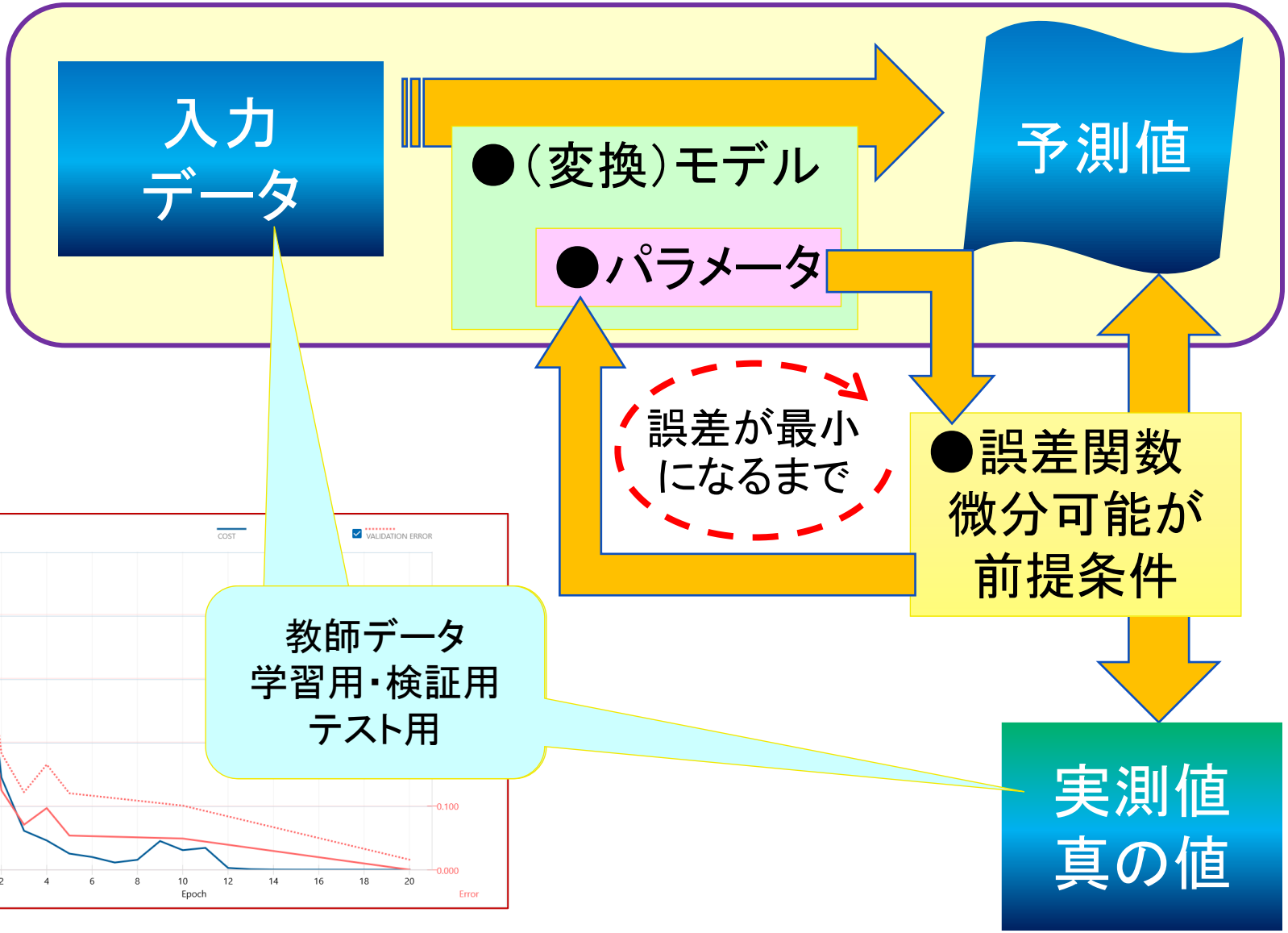
改訂） 西山秀昌. 歯学教育における人工知能（AI），機械学習および深層学習について－形式（的）情報と意味（的）情報の視点から－. 新潟歯学会誌. 2023;53(1):1-15.


入力側から出力側へ（フィードフォワード；FF）
入力
x_1
x_i
x_n
w_1
w_i
w_n
$\sum_{i=1}^{n} w_i \cdot x_i + b$
f
z
出力結果【予測値】
真の値
損失・誤差
人工ニューロン
ノード（ないしユニット）
・・・＞ Σ（入力×重み）＋バイアス⇒ 活性化関数⇒ 出力 ・・・＞
x_i w_i b f z
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
通常x, w, b, zの値は、0から1の範囲で正規化される
出力側から入力側へ（フィードバック；FB）

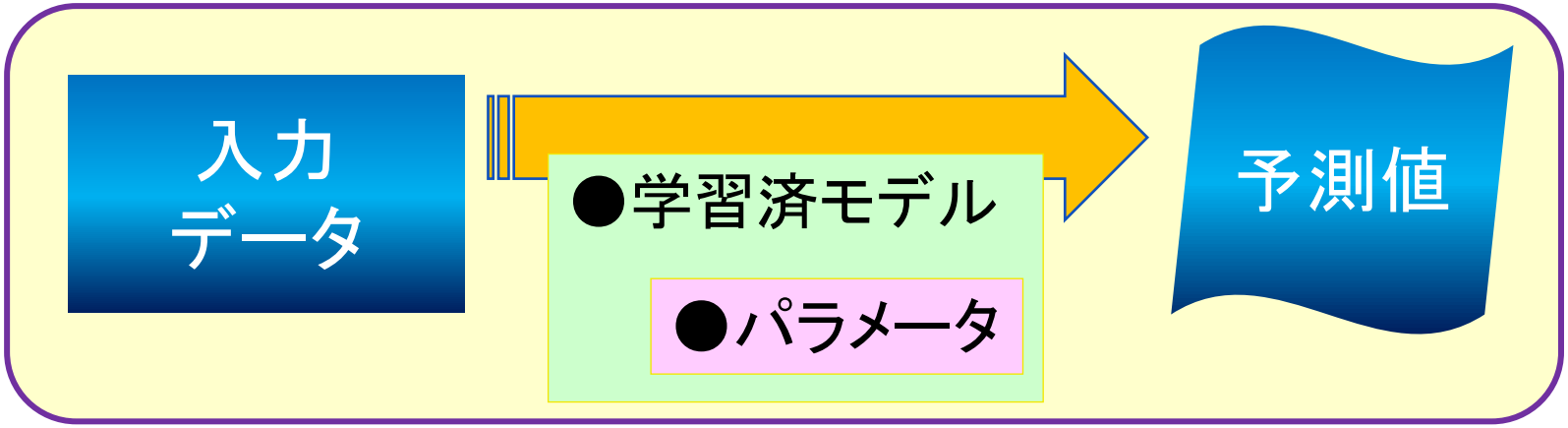
深層学習・機械学習


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
誤差が最小になるまで
●誤差関数
微分可能が前提条件
人間が規則性を見つけ、モデルを作る。
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」という力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・機械学習の稼動時



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

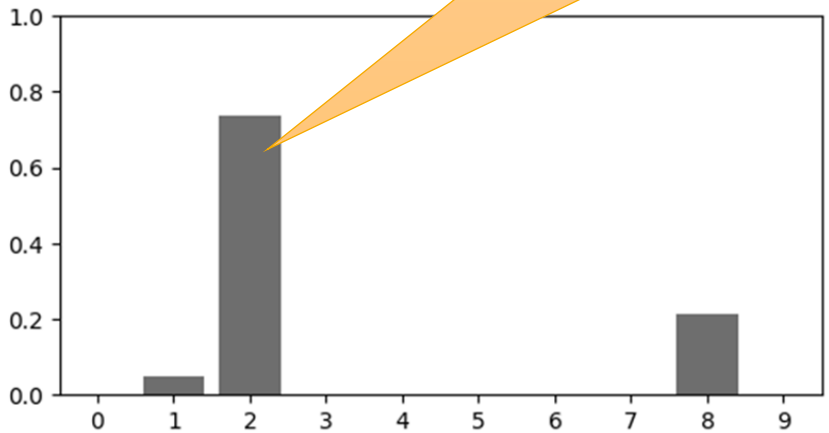
深層学習システムは学習したこと以外は回答できないし学習した範囲で答えようとする。

```
HTML(input_form + javascript)
# After appearing the Box for drawing...
# ↓  ↓  ↓  Please Draw number with pointing
```

Out[8]:

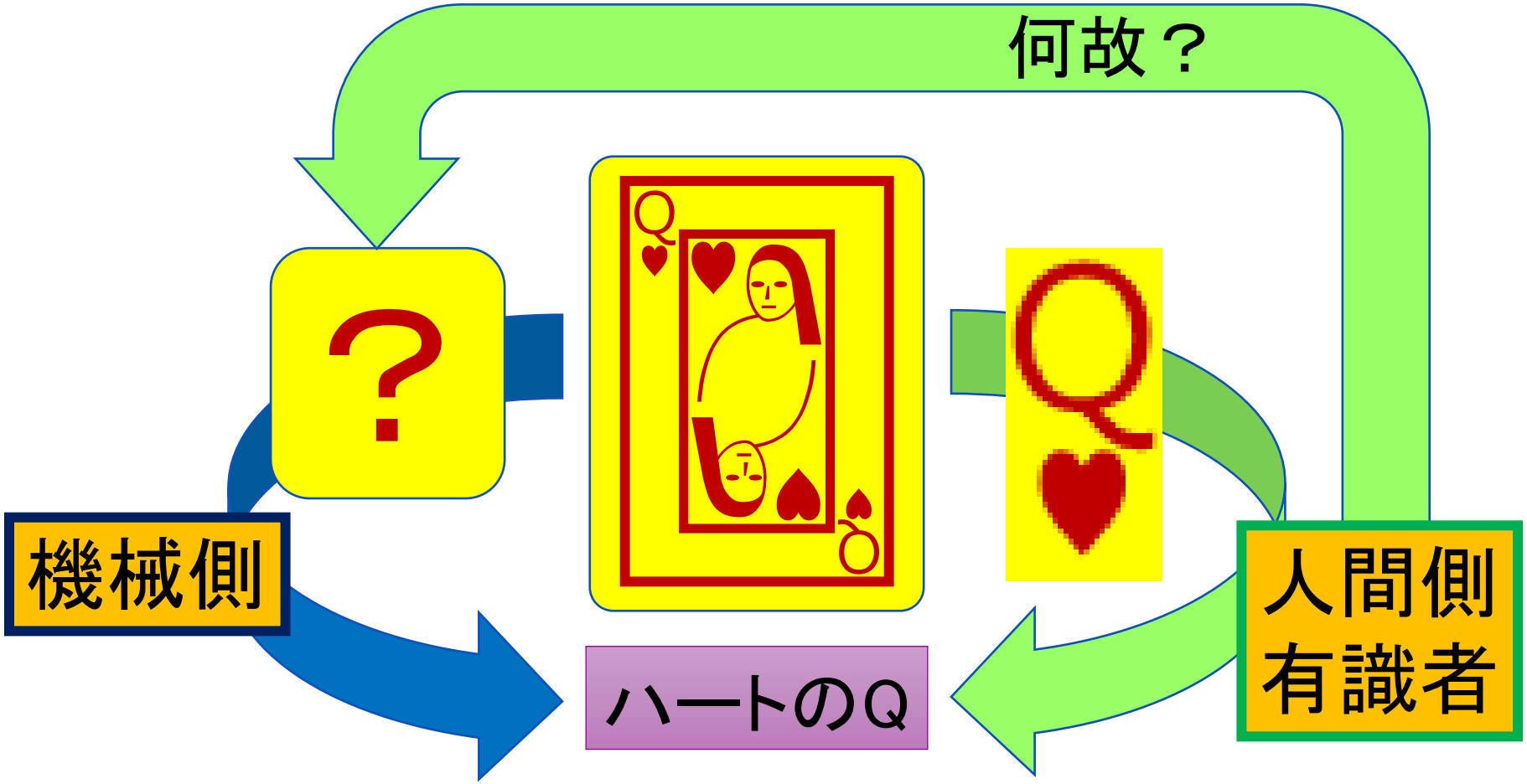
「2」の可能性が80%近いと回答
「分からない」とは答えないし、
「χ」とは答えない。

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



問題の本質について(1)

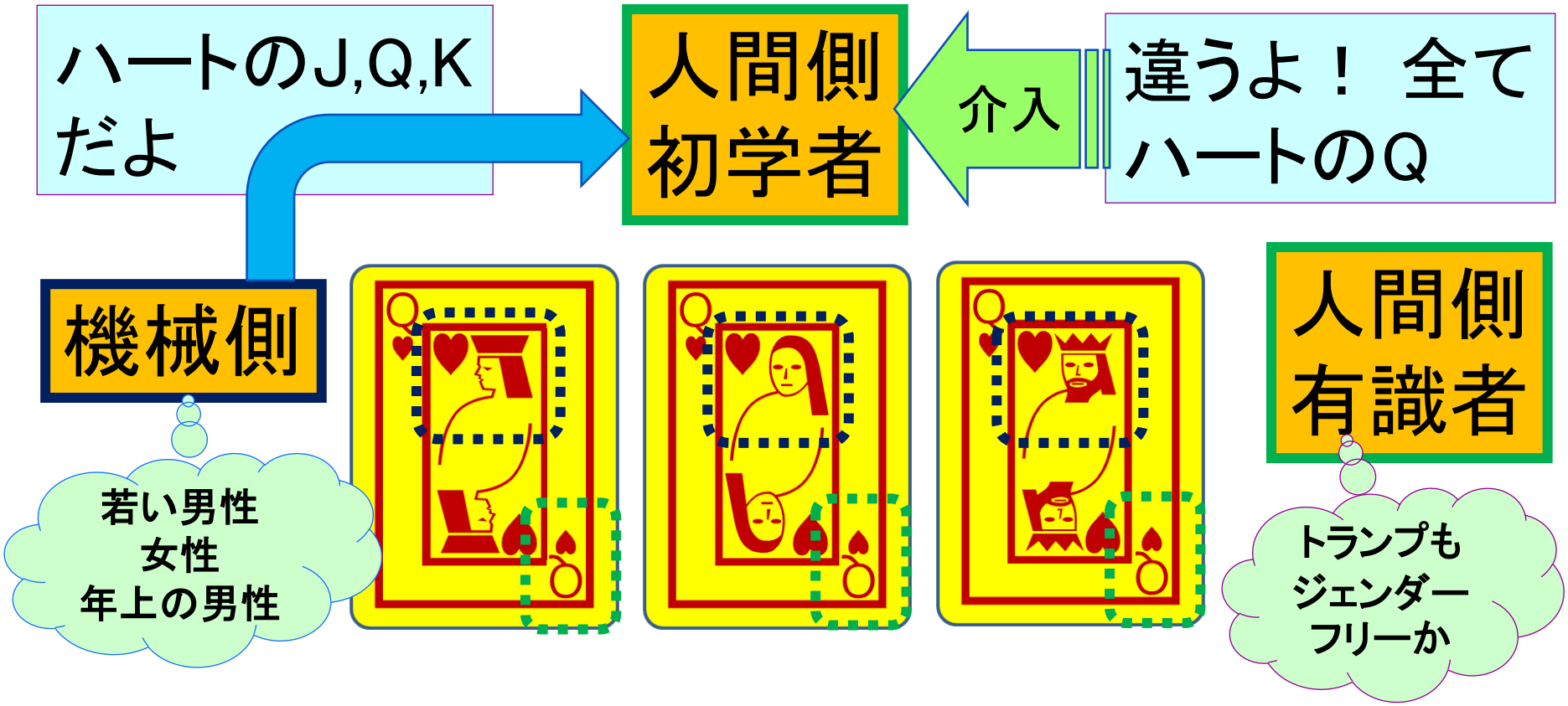
- 深層学習では、機械側の判断根拠を知ることが困難とされている。



問題の本質について(2)

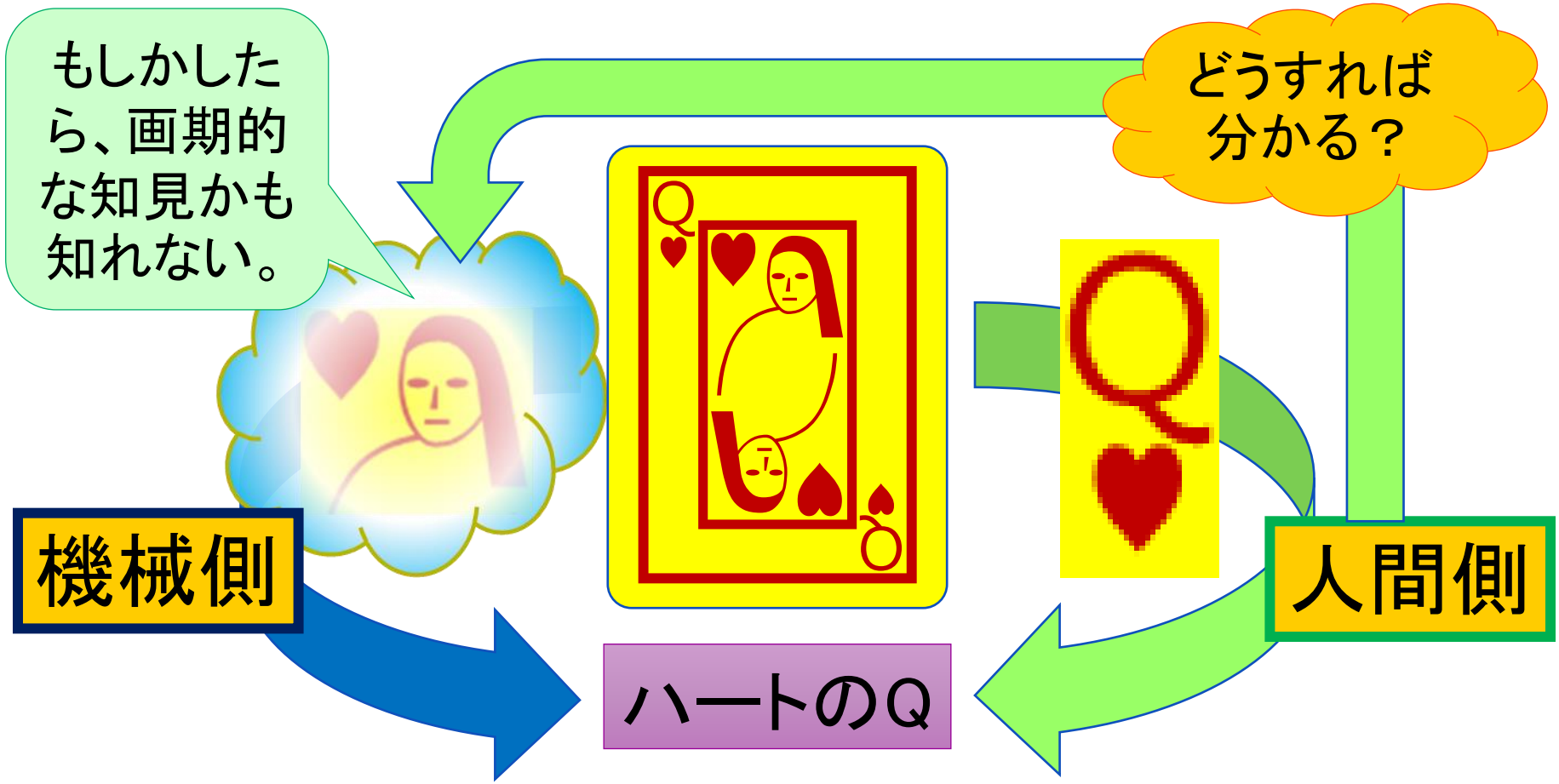
たとえ99%の精度を持つシステムであっても、

- 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。



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。



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。

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。

AI

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
他

改訂） 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.

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).
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).
```

事実
（データ相当）

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）

解　説：医療情報学　7 (2), 1987 : 183–190

特　集：医　療　支　援　と　AI

診断過程のシミュレーション

——頭痛・顔面痛診断支援システム：RHINOS——

RHINOS Consultation System for Diagnosing Headache and Facial Pain

松村　泰志[1]　松永　　隆[2]　木村　通男[2]
Yasushi MATSUMURA　Takashi MATSUNAGA　Michio KIMURA

前田　祐輔[3]　津本　周作[3]　松村　　浩[4]
Yusuke MAEDA　Shusaku TSUMOTO　Hiroshi MATSUMURA

医師は，効率良く患者の病状に関する情報を集め，精度の高い診断をすることができる．これは，医師の診断の推論過程で，肯定の論理（患者が，ある疾患に特異的な症状を持つ時，その疾患を疑う），否定の論理（患者が，ある疾患に出現頻度の高い症状を持たない時，その疾患は否定的であると考える），併発症診断の論理（患者が，ある疾患に特異的な症状を持ち，かつ，その疾患では起こり得ない疾患を持つとき，2つ以上の疾患の併発を疑う）をうまく使っているからと考えられる．頭痛・顔面痛診断に領域を設定し，この仮設にもとづいて診断支援システムを作成し RHINOS (Rule based Headache and facial pain INformation Organizing System) と名づけた．51例について，システムの解答と専門家の解答が一致するか検討した結果，47例が一致し4例が異なった解答をした．正しく診断した47例の中には，非典型例を除外診断により正しく診断したものも含まれていた．また，頻度の高い典型例の診断は，少ない質問数で診断し，非典型例や，まれな疾患に対しては，比較的多くの質問をし，診断した．このように，RHINOS は効率良く，しかも，高い精度で診断し，RHINOS の診断能力が専門医のそれに近似していることが示された．逆に，このことより，RHINOS の診断アルゴリズムが医師の推論過程をよく表現していると考えられた．

(キーワード：診断支援システム，知識ベース，頭痛，顔面痛，RHINOS，マイクロコンピュータ，プロローグ)

3.3 環　　境

プログラミング言語として，Prolog-KABA [10] を用いた．Prolog の Horn Clause は，知識表現に便利である．また，Prolog-KABA は，マイクロコンピュータ上で走る Prolog の中で最も速い．

機種は，Prolog-KABA の走る日本電気社製の PC-9801 (CPU 8086 メインメモリ 384 KB) を利用した．

3.4 入　　力

最初に，すべての患者に対して，基本的質問をする (Fig. 1)．質問事項は以下のとおりである．

患者名・年齢・性別
頭痛の初発時期
経　過
痛む部位
痛みの性状
痛みの程度
頭痛に伴う神経学的所見
睡眠状況
既往歴
家族歴
Jolt headache (頭を振った時の痛みの増強)

答えは，患者名や年齢，性別，頭痛の初発時期など直接データを入力するものと，選択式で答えを選ばせるものとがある．

基本的質問の入力が終わった後，追加の質問をする (Fig. 2)．追加の質問は原則的に，はい，いいえで答えさせる．

基本的質問，追加の質問のどちらでも，「不明」の入力を認める．「不明」の入力があった場合，その事項を，Exclusive rule, Inclusive rule では適合するように，Disease image では，無視するように解釈する．

今までの情報で鑑別されるべき疾患は以下のとおりです
（％は頻度）

Common migraine	7.0%
Tension vascular headache	5.7%
Classic migraine	3.7%
Cluster headache	3.7%
Sympton of other disease	2.0%
Hemiplegic migraine	1.3%
Autonomic disturbance	0.6%
Invasive disease to the skull	0.6%
Unruptuerd intracranial aneurysm	0.6%
Unruptuerd arteriovenous malformation	0.3%
Ophthalmoplegic migraine	0.5%
Ophthalmoplegic migraine	0.5%
Pheochromocytoma	0.1%

第2段階の質問に進む（→ ↓）....

Fig. 3 List after screening.

以上の質問から次のような結果が得られます

1	通常型偏頭痛	肯定的	SI=80
2	他疾患の症状	肯定的	SI=40
3	緊張性血管性頭痛	否定的	CI=95
4	古典型偏頭痛	否定的	CI=92
5	群発性頭痛	否定的	CI=95
6	片麻痺型偏頭痛	否定的	CI=95
7	自律神経失調症	否定的	CI=80
8	頭蓋浸潤性疾患	否定的	CI=60
9	未破裂脳内動脈瘤	否定的	CI=90
10	未破裂 AVM	否定的	CI=90
11	眼球運動障害型偏頭痛	否定的	CI=95
12	褐色細胞腫	否定的	CI=80

Fig. 4 Diagnostic conclusions.

51例中2例で同一患者で2種類の頭痛を併発していた．しかし，RHINOS の解答はいずれの例も十分ではなかった．

専門家の診断と，RHINOS の診断が一致しなかった他の2例を含め以下にその最終診断と誤った理由を示す．

1) 下半部頭痛と筋収縮性頭痛の併発例

RHINOS は，できるだけ少ない質問数で診断できるように，頻度の低い疾患は，強く肯定される疾患が見つからなかった場合にのみその可能性を検討するという方法をとっている．したがって，頻度の高い疾患と低い疾患が併発した場合には，頻度の低い疾患の可能性を検討せずに診断を終了する．しかし，この場合でも，その症例が1つの疾患で説明し得ないことを示すことはできる．

2) 古典型偏頭痛と筋収縮性頭痛の併発例

血管緊張性頭痛は，筋収縮性頭痛と偏頭痛の特徴を合わせ持つ疾患であるが，この2つの特徴が異なる時期に発症した場合には，これらの疾患の併発を考える．RHINOS は，1人の患者に2種類の頭痛があっても別々に検討することができないので，筋収縮性頭痛と偏頭痛の併発を血管緊張性頭痛から区別することはできない．

3) 非典型の心因性頭痛の1例

それぞれの疾患についての Inclusive rule の Covering Index は，できるだけ大きくとることが原則であるが，必ずしも100％とはならない．きわめて非典型な症例で，その疾患のどの Inclusive rule にも適合しなかった場合には正しい診断はできない．しかしこのような症例では，専門家でも診断は難しい．

4) 人格障害を伴った血管拡張性頭痛の1例

人格障害があるため，患者の訴えに事実と異なるものが含まれていた．

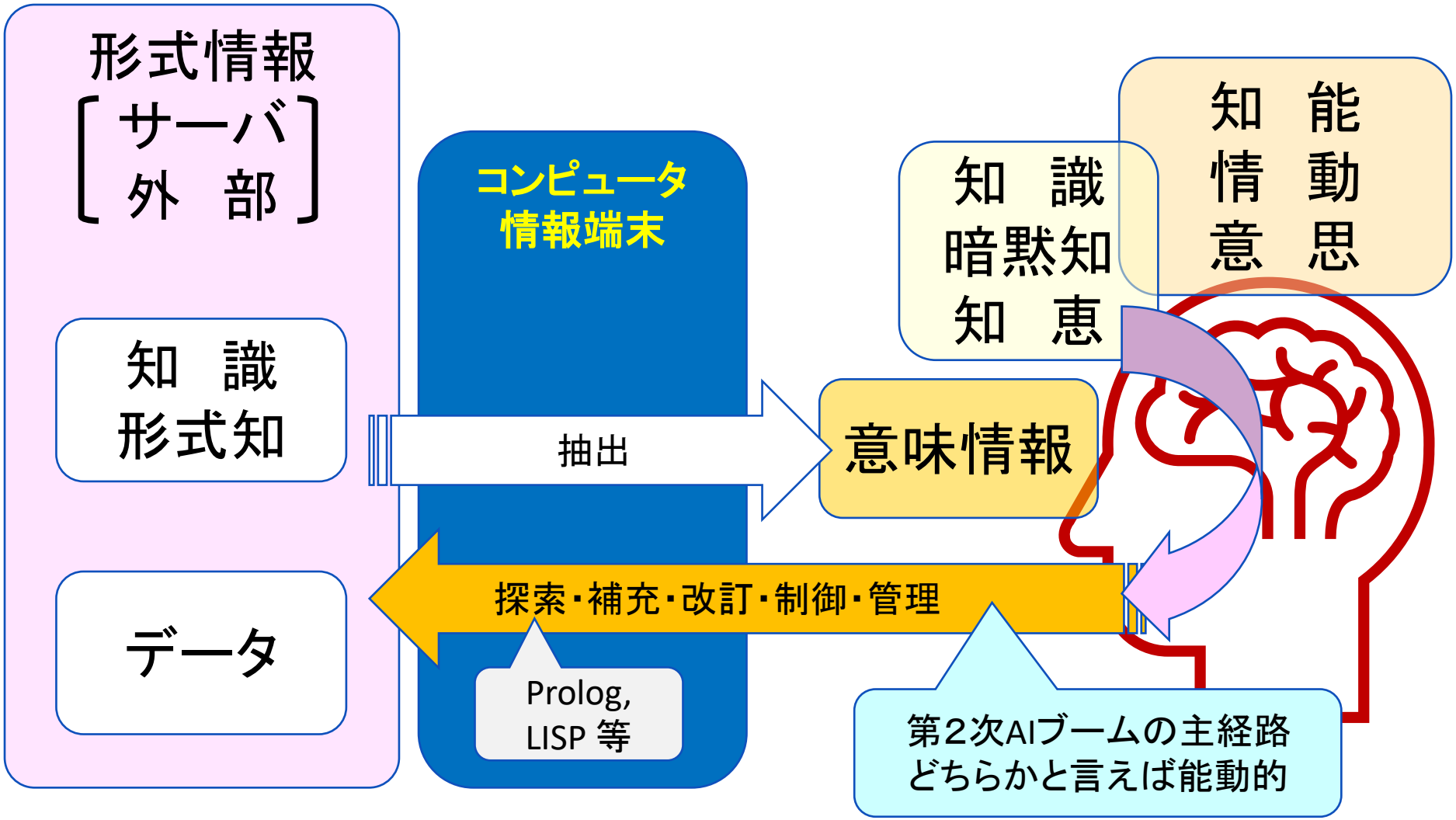
Table 2 Result of an evaluation of the system for 51 case studies

最終診断名	症例数	正解数
頭蓋内腫瘤	7	7
筋収縮性頭痛	6	6
古典型偏頭痛	5	5
くも膜下出血	5	5
心因性頭痛	4	3
群発性頭痛	4	4
通常型偏頭痛	3	3
緊張血管性頭痛	3	2
側頭動脈炎	1	1
未破裂動脈瘤	1	1
未破裂 AVM	1	1
前頭洞炎	1	1
心因性顔面痛	1	1
自律神経失調症	1	1
症候性三叉神経第二枝痛	1	1
特発性三叉神経痛	1	1
後頭部神経痛	1	1
他疾患の症状	1	1
眼疾患による頭痛	1	1
頭部交感神経過敏症	1	1
古典型偏頭痛と筋収縮性頭痛の併発	1	0
下半部頭痛と筋収縮性頭痛の併発	1	0
計	51	47

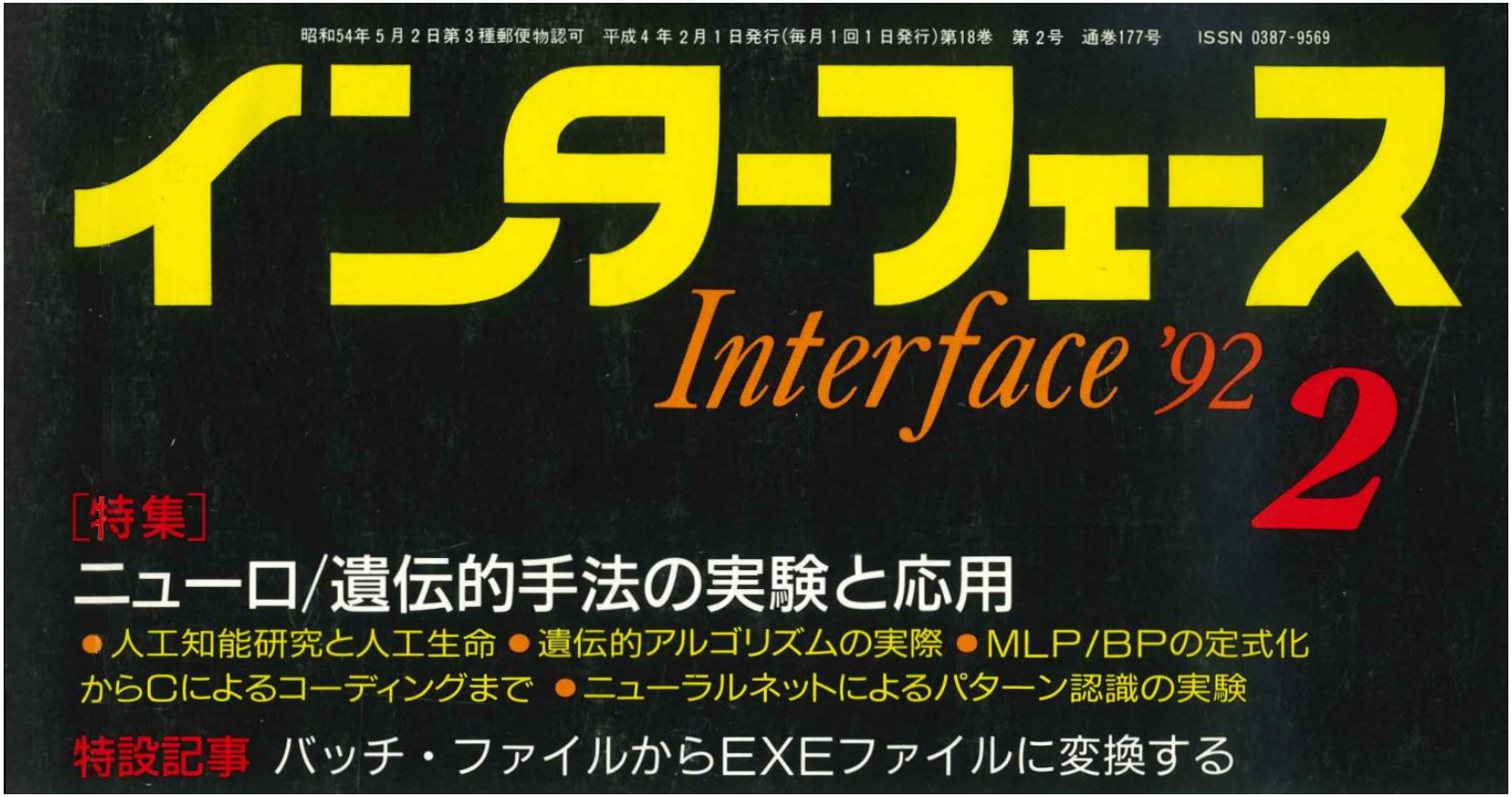
診断効率について客観的に表現する方法はないが，RHINOS が，専門家が診断に要するとほぼ同じ程度の質問数で，専門家に近い精度の診断が可能であったことは，RHINOS の診断効率が専門家のそれに近いことを示していると考えられる．

このように，RHINOS は，診断精度についても，診断効率についても専門家に近い診断能力を持っており，このことは，RHINOS の診断アルゴリズムが，専門家の診断の推論過程を良く表現していることを示していると考えられる．

データ、情報、知識・知恵と第二次AIブームとの関係

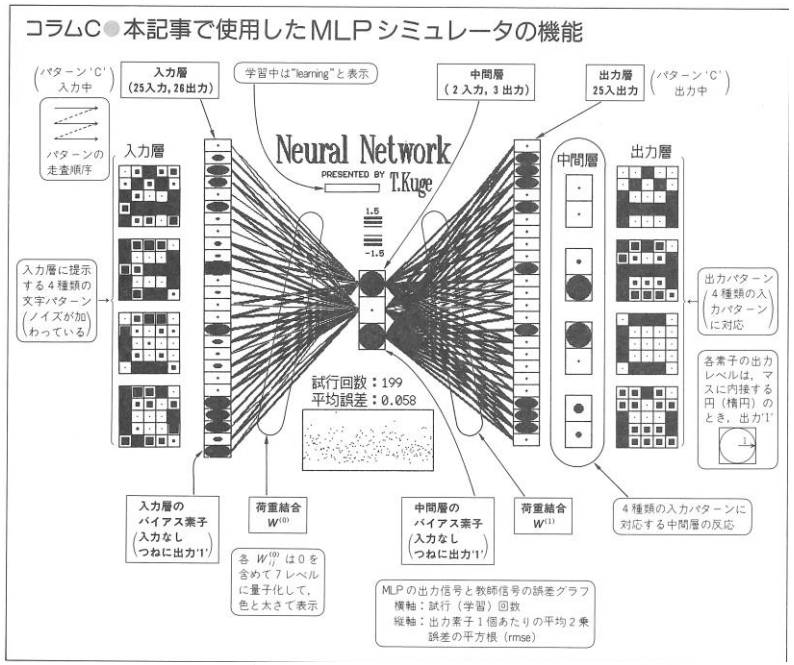


30年前
第2次人工知能ブームの後半位

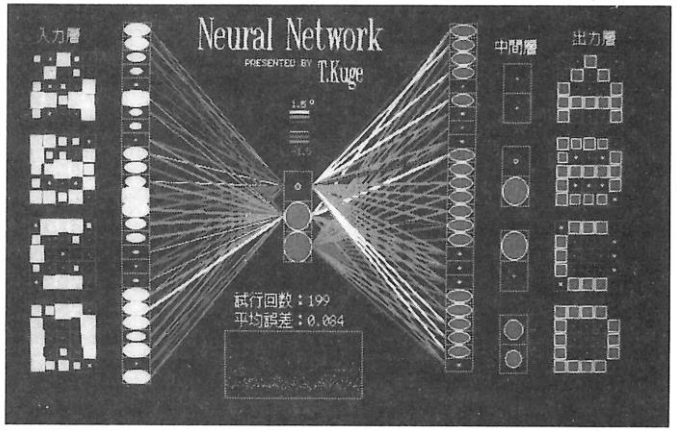


3層構造のNeural Network
入出力の層をカウントしての3層

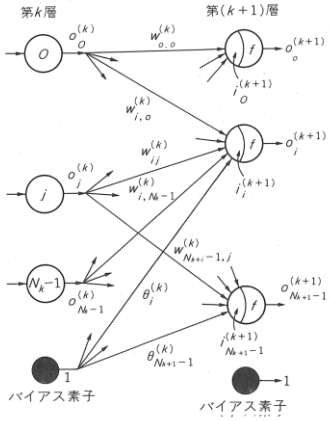
久下哲郎「ニューラル・ネットワーク・シミュレータの制作とパターン認識実験」



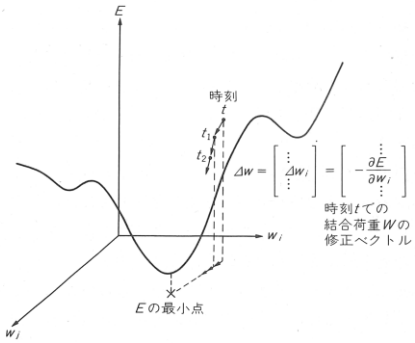
〔写真3.3〕一様乱数［0, 0.1］ノイズあり学習による画像符号化能力



〔図A2〕第 k 層と第 $(k+1)$ 層の結合

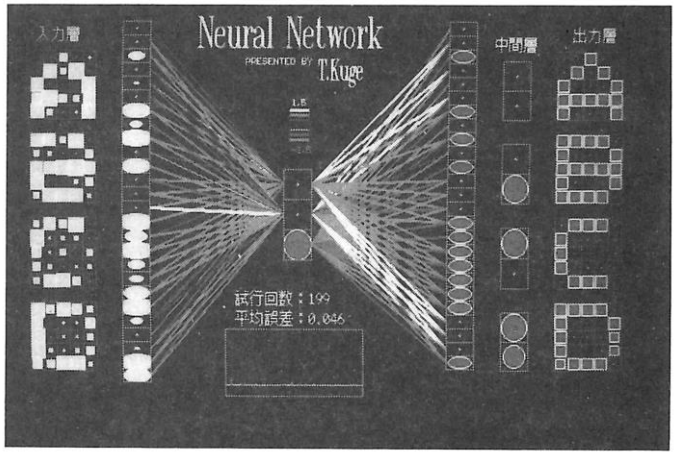


〔図A3〕最急降下法

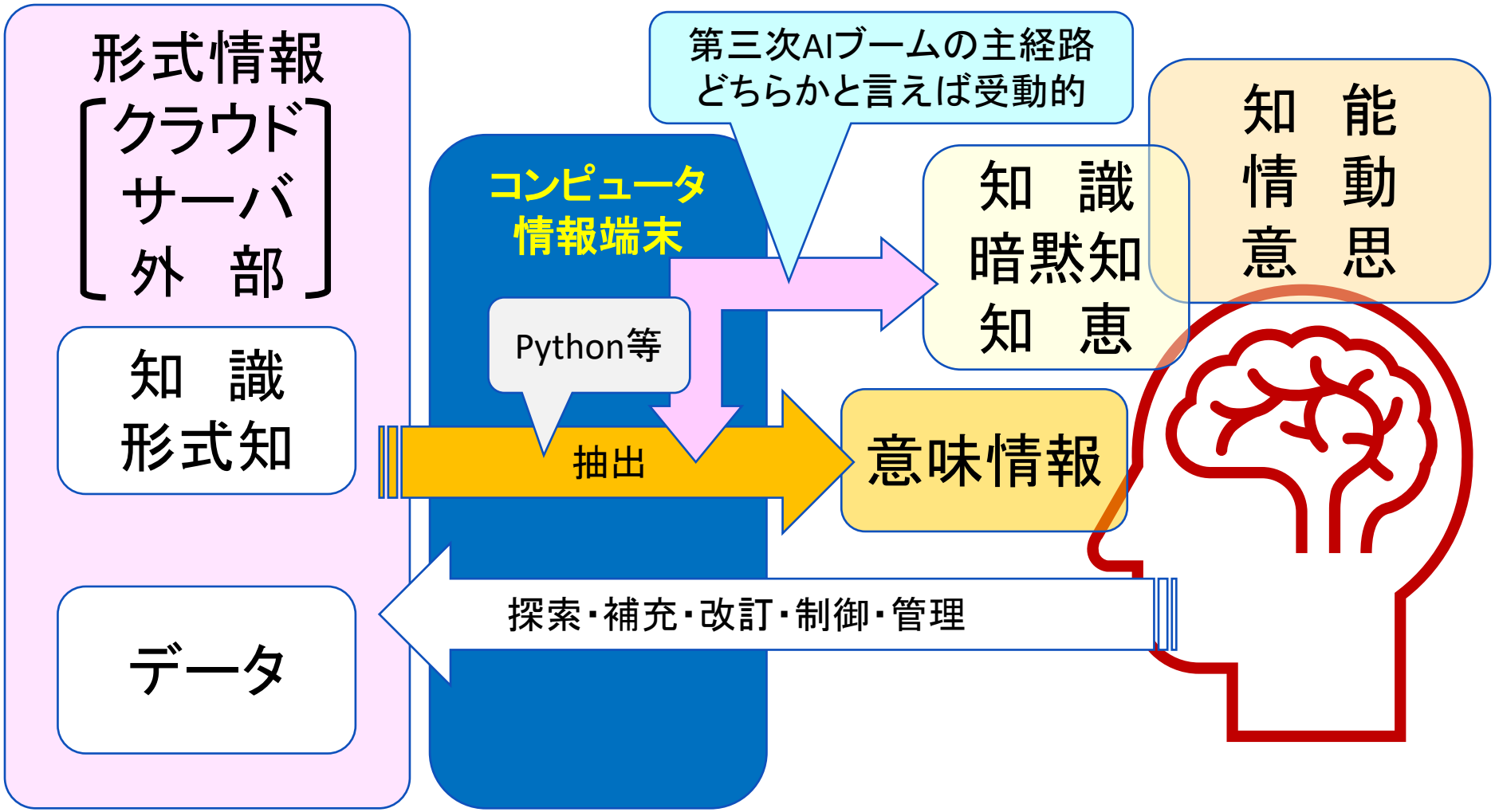


N_{MLP}次元平面上の誤差ポテンシャルE面上をEの最小点の方向に，降下してゆく．
Δwは，時刻tで，誤差Eをもっとも少なくする方向のベクトルである．

〔写真3.4〕一様乱数［0, 0.5］ノイズあり学習による画像符号化能力



データ、情報、知識・知恵と第三次AIブームとの関係



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。

AI

いわゆるSymbolism

過程優先、演繹的

第二次AIブームの主役

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

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

第一次、第三次AIブームの主役

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

機械学習

深層学習

Transformer系
CNN系
他

改訂） 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.

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

適正利用の原則

soumu.go.jp/main_content/000637097.pdf

［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当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うる（「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

5. 支援システム等で想定される主な課題と解決の方向性

mhlw.go.jp/content/10

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

- 包括的な成長、持続可能な開発、心身ともに満たされた状態
- 人権と民主主義的価値観（公平性とプライバシーを含む）
- 透明性と説明可能性
- 堅牢性、セキュリティ、安全性
- 説明責任

政策担当者への勧告

- AIの研究開発への投資
- 包括的な AI 対応エコシステムの育成
- 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
- 人的能力の構築と労働市場移行への準備
- 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
（古くなりつつある資料だが概略把握可能）

https://ja.stateofaiguides.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
  - 音声生成系
    - VALL-E

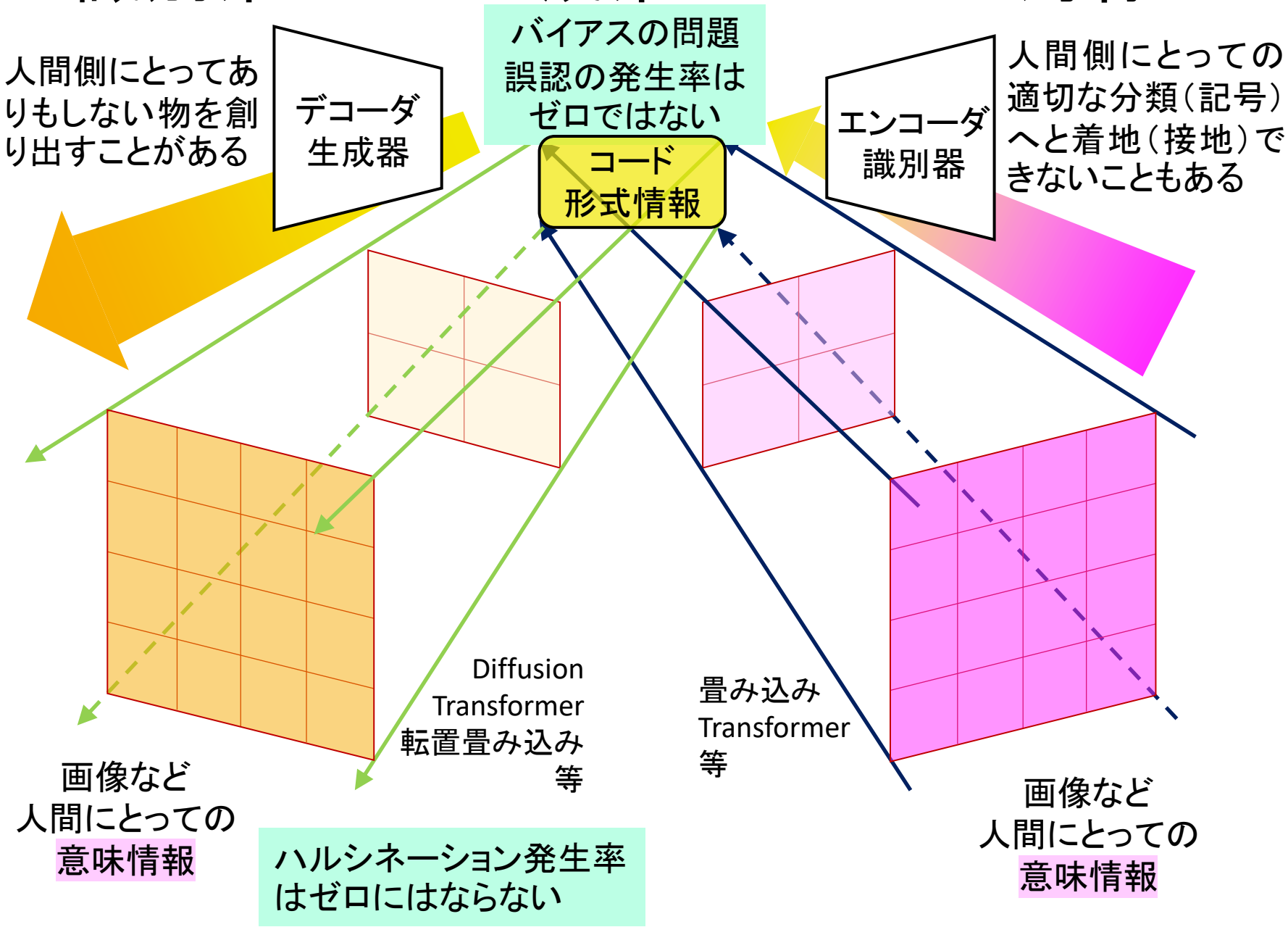
https://atmarkit.itmedia.co.jp/ait/articles/2301/13/news023.html

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

識別系のAIと生成系のAIはほぼ対称

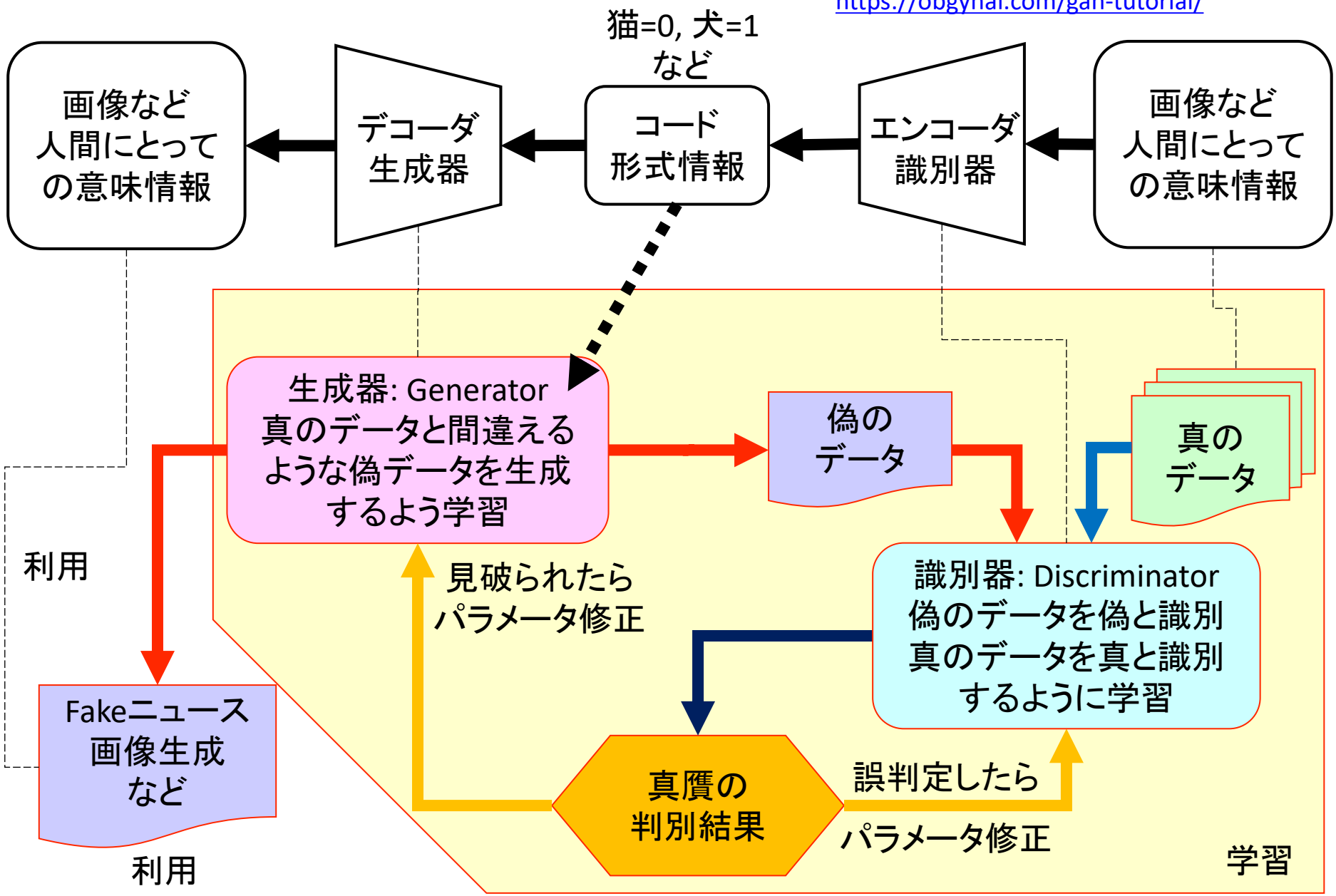


生成系AIの核

- 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。
- Diffusion（拡散モデル）
  - 主に画像生成系で使われている（Stable Diffusion）。統計学的な拡散をベースとしたモデル。

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/



生成AIの代表となるLLM（大規模言語モデル）の中心部分：Transformer

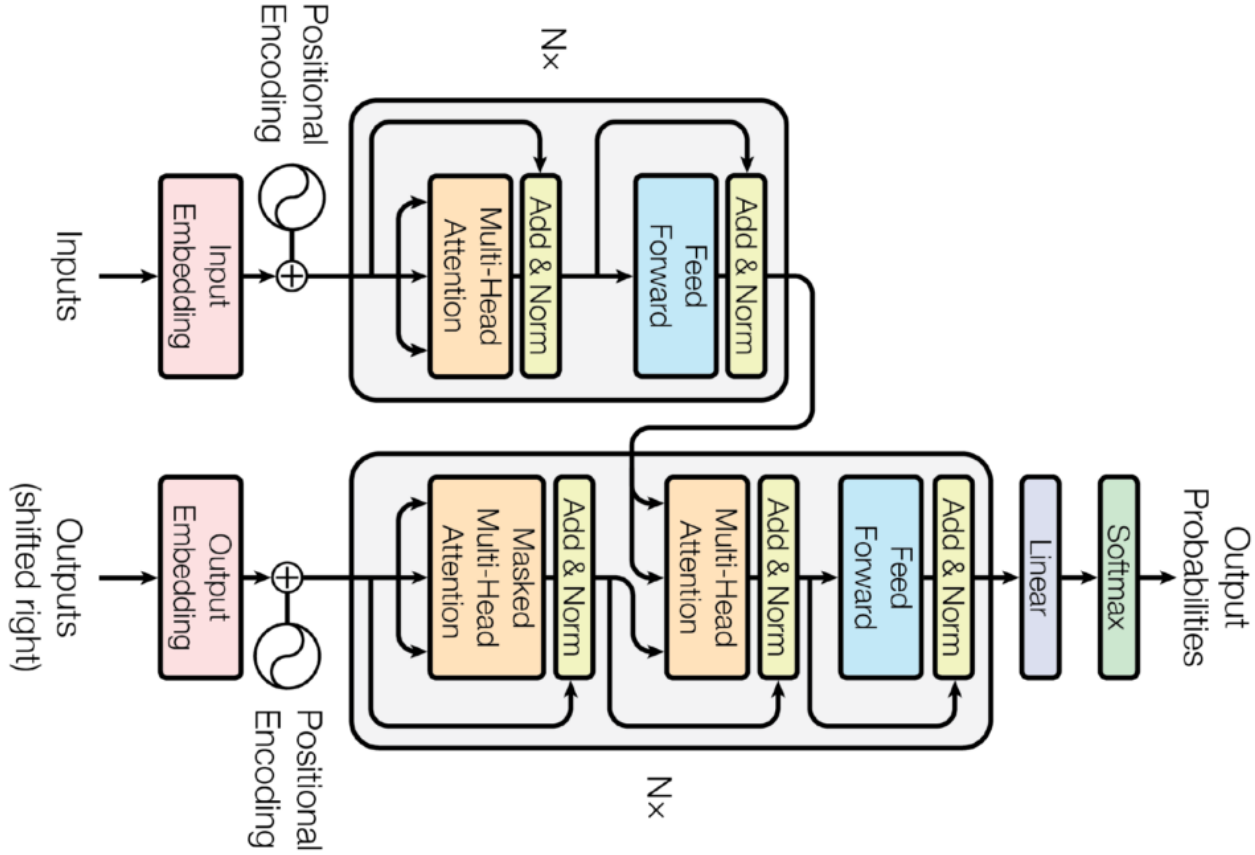


Figure 1: The Transformer - model architecture.

Vaswani, A. *et al.* Attention Is All You Need. *arXiv:1706.03762* (2017).

このTransformerモデルも1つのニューロンとして捉えることが可能

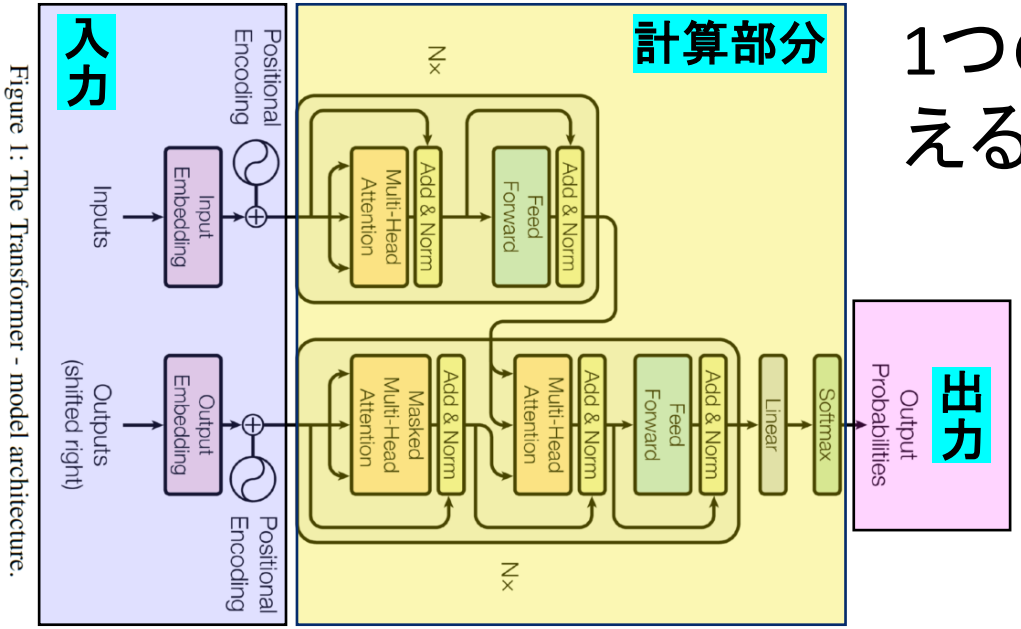
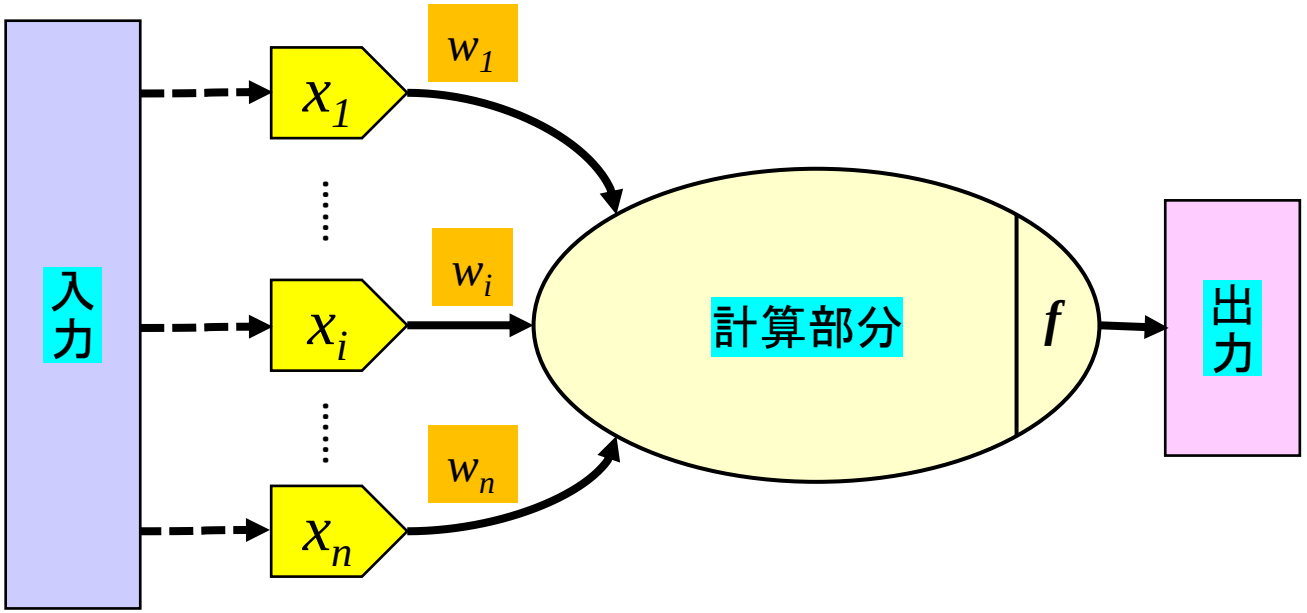


Figure 1: The Transformer - model architecture.

Vaswani, A. *et al.* Attention Is All You Need. *arXiv:1706.03762* (2017).



更にTransformerの中心となる Attention も・・・

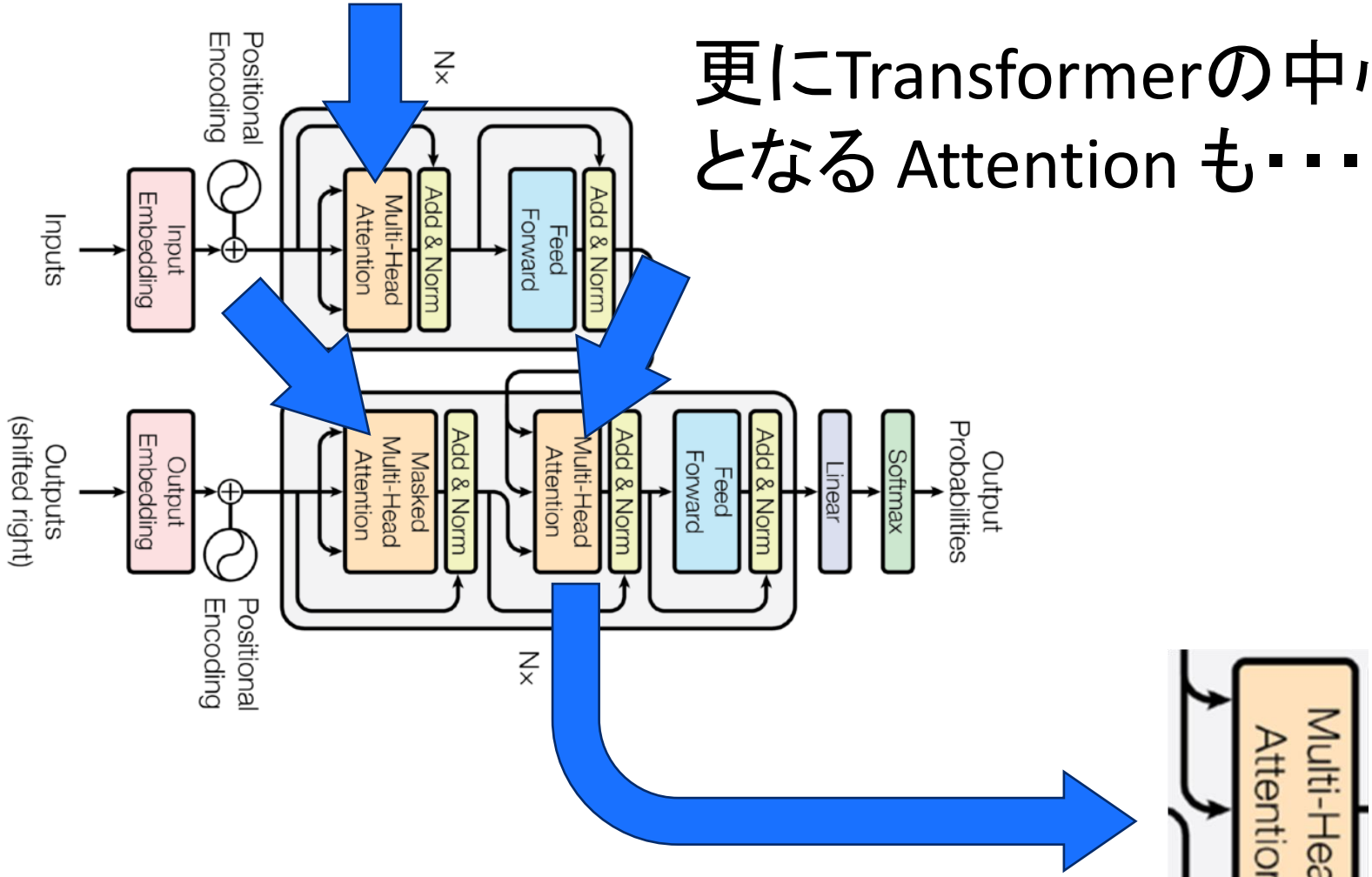
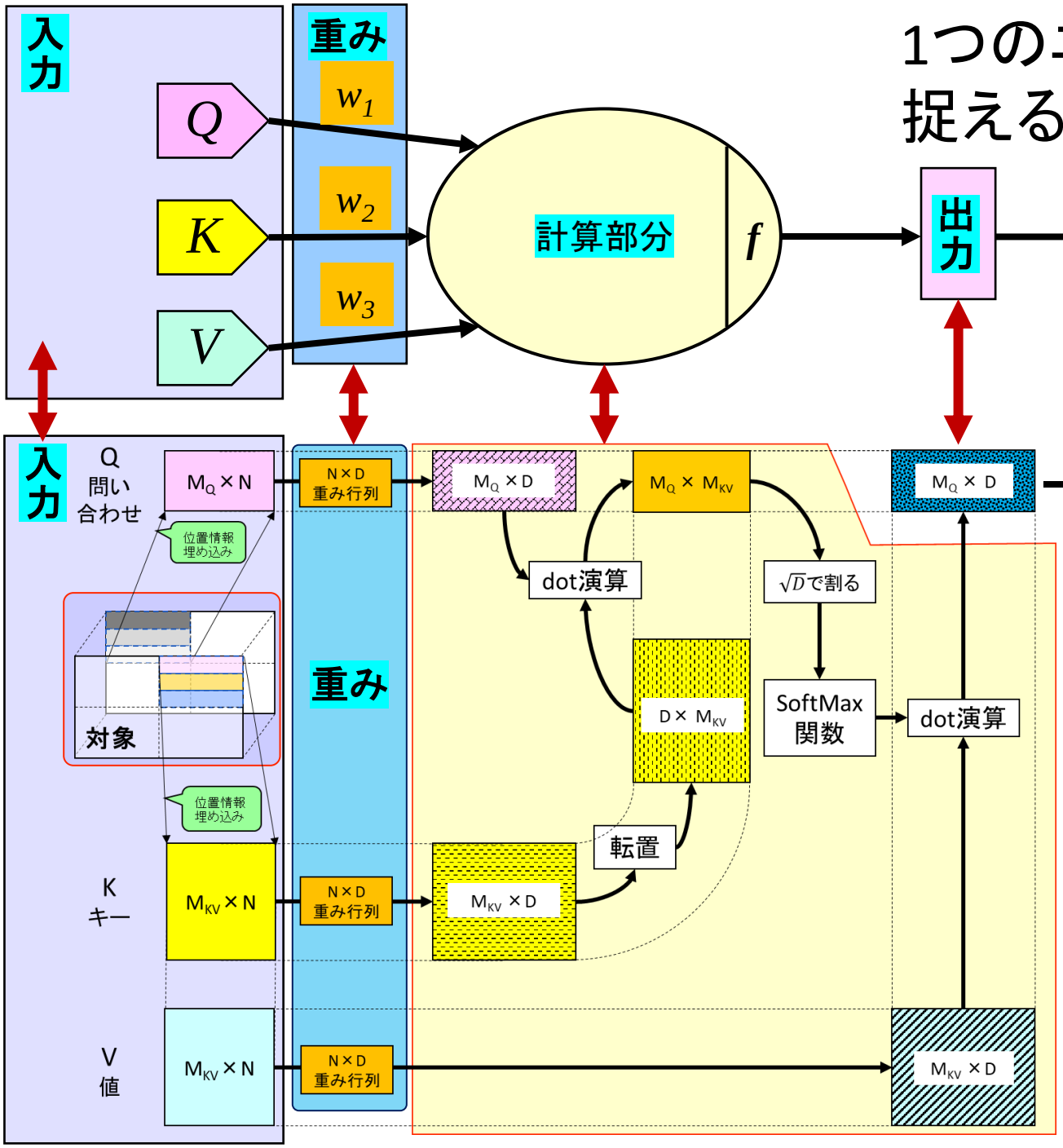
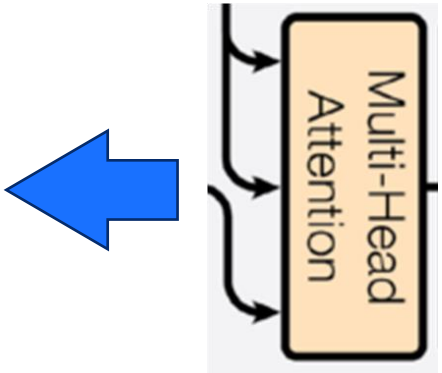


Figure 1: The Transformer - model architecture.

Vaswani, A. *et al.* Attention Is All You Need. *arXiv:1706.03762* (2017).

1つのニューロンとして捉えることが可能





「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.

補遺・Attensionが内部で主に行っていること
一度に大量ベクトルの内積（類似度）計算

Q×W_Q

$$\begin{bmatrix} \alpha & \beta & \gamma \\ \delta & \varepsilon & \zeta \end{bmatrix}$$

M_Q行3列

内積（ドット演算）

$$\begin{bmatrix} ga & gb & gc & gd \\ ha & hb & hc & hd \end{bmatrix}$$

M_Q行M_{KV}列

$ga = \alpha \times AA + \beta \times BA + \gamma \times CA$

K×W_{KV}

$$\begin{bmatrix} AA & BA & CA \\ AB & BB & CB \\ AC & BC & CC \\ AD & BD & CD \end{bmatrix}$$

M_{KV}行3列

転置

$$\begin{bmatrix} AA & AB & AC & AD \\ BA & BB & BC & BD \\ CA & CB & CC & CD \end{bmatrix}$$

3行M_{KV}列

例：(α, β, γ)ベクトルと(AA, BA, CA)ベクトルとの内積（類似度）を計算する。ベクトル長さを正規化（=1）にしているので、二つのベクトルのなす角度をθとすれば内積はcosθそのもので-1から1となる。ベクトルの類似度が大きい（＝θが小さい）と1に近づく。

-1から1までの値を0から1までに変換

SoftMax演算

V×W_{KV}

$$\begin{bmatrix} aa & ba & ca \\ ab & bb & cb \\ ac & bc & cc \\ ad & bd & cd \end{bmatrix}$$

M_{KV}行3列

内積（ドット演算）

$$\begin{bmatrix} p & q & r \\ s & t & u \end{bmatrix}$$

出力：M_Q行3列

$$出力 = SoftMax\left(\frac{[Q \times W_Q] \cdot [K \times W_K]^T}{\sqrt{D}}\right) \cdot [V \times W_V]$$

D=3の場合の例

$t = ha \times ba + hb \times bb + hc \times bc + hd \times bd$

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いうる数値の行列・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。

LLM内部で扱っているものが形式的な記号列だから異なる対象であっても同じモデルで扱える

https://blog.google/technology/ai/dolphingemma/

Google The Keyword Home Product news Company news Feed Subscribe

TECHNOLOGY > AI

DolphinGemma: How Google AI is helping decode d communication

Apr 14, 2025
5 min read

DolphinGemma, a large language model developed by G study how dolphins communicate — and hopefully find o

Dr. Denise Herzing
Research Director/Founder, Wild Dolphin Project

Dr. Thad Starner
Google DeepMind Research Scientist and Georgia Tech Professor

DolphinGemma: Google AIがイルカのコミュニケーションを解読する仕組み

2025年4月14日
5分で読む

Google が開発した大規模言語モデル DolphinGemma は、科学者がイルカのコミュニケーション方法を研究するのに役立っており、イルカが何を言っているのかも解明できると期待されています。

デニス・ハージン博士
ワイルドドルフィンプロジェクトの研究ディレクター/創設者

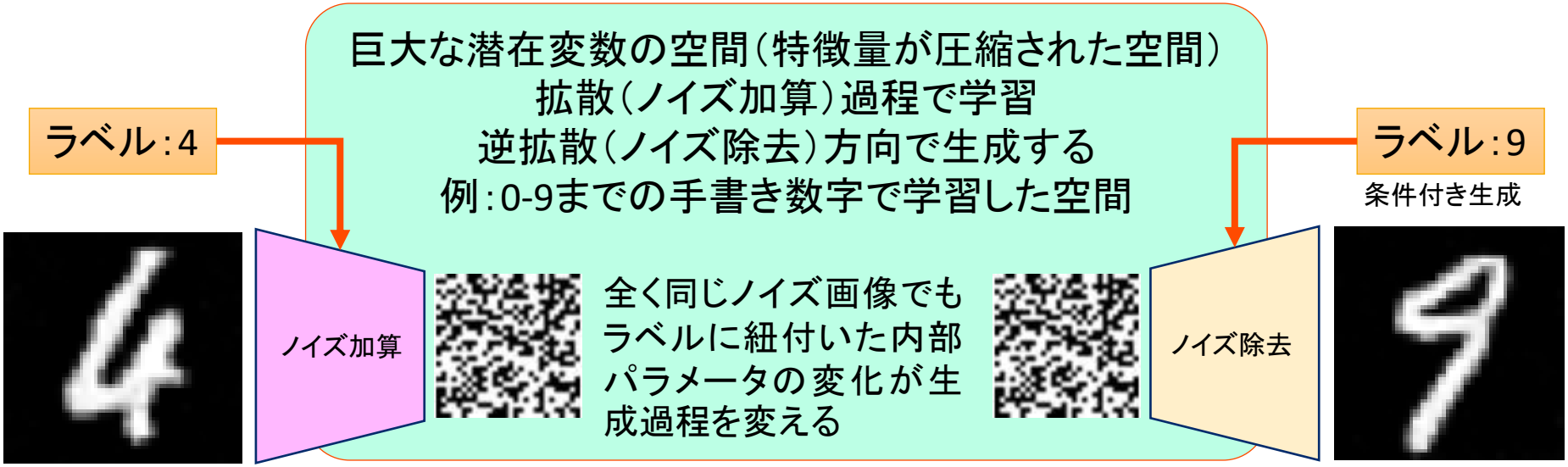
サド・スターナー博士
Google DeepMindの研究科学者、ジョージア工科大学教授

共有

Diffusion model（拡散モデル）

統計的な確率過程で制御されるモデル

- Diffusionのわかりやすいサイト
  https://qiita.com/omiita/items/ecf8d60466c50ae8295b
- 専門書：岡野原大輔. 拡散モデル データ生成技術の数理. 岩波書店; 2023.
- 概略本：岡野原大輔. 生成AIのしくみ〈流れ〉が画像・音声・動画をつくる. 岩波書店; 2024.
- 演習本：斎藤康毅. ゼロから作るDeep Learning ❺ —生成モデル編. オライリー・ジャパン; 2024.
- 上記「ゼロから作るDeep Learning ❺ —生成モデル編」テキストの演習サイトからフォークし、コメントを入れた大学院演習用サイト（西山）
  https://github.com/aujinen/Local_ipynb_deep-learning-from-scratch-5



西山記述原本を元にChatGPT-4oによる添削を参照し修正
https://chatgpt.com/share/67ff2b93-8dcc-8005-9f08-009ebff22005

ハルシネーション
（Hallucination）
と
記号接地問題
（Symbol Grounding Problem）

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」を見いだしうる。
- 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い込んでしまう。

ChatGPT関連のハルシネーション記事

- 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（人間側にとっての）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。

ハルシネーションと記号接地問題

- 「ハルシネーション（誤情報）」とは、機械学習モデルが確率的な勾配に基づいてデータを生成する際、それが人間が意図する意味や文脈と一致しない場合に発生する現象。
  これは、機械が扱う「形式情報（記号）」と、人間がそれに対して付与する「意味情報」との間の結びつき（接地）が不十分であるために起こり、この根本的な問題を「記号接地問題」と呼ぶ。
  （元原稿：https://www5.dent.niigata-u.ac.jp/~nisiyama/FD/20250301-FD-AI-HN_Part2.pdfのp.26, Google Gemini 2.5 Flashの修正案を参照して改訂）
- 原因：
  - 学習データ不足（外挿領域の問題など）
  - プロンプトの問題（機械側が誤解するような設定を行うなど）
  - モデル構造の限界（LLMは確率的勾配にて学習しており、必ずしも事実や文脈を深くは理解していない）
  - 確率的出力による揺らぎ（同じ入力に対してモデルが異なる出力を行う事がある）

ハルシネーション対策は様々に研究されてはいるが・・・

一般的には

プロンプトを適切に設定するとか

RAGを導入するとか

ローカルLLMにすると良いとは言うけれど・・・

医療系で使う以上、説明責任は求められる

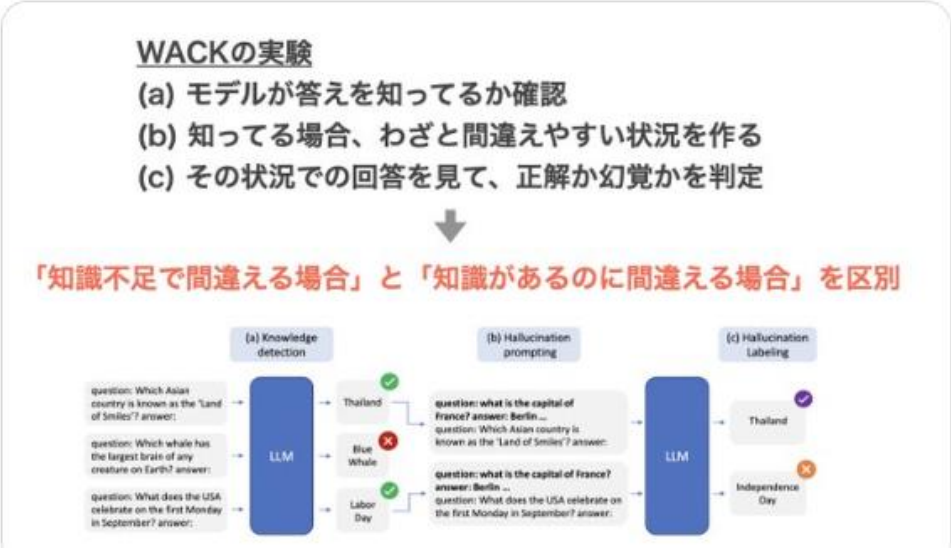
対策例:「知識がないための幻覚」と「知識があるのに起こる幻覚」を区別し対処

K.Ishi@生成AIの産業応用
@K_Ishi_AI

LLMの**ハルシネーション対策に大きな前進**が見られた。

この研究では、ハルシネーションを
- 知識不足による誤り(HK-)
- 知識を持っているのに誤る(HK+)
の2種類に分類して分析した。

そして、"**厄介な方のHK+を意図的に誘発するプロンプト**"も編み出し、"**HK+を回答生成前に検知可能**"なことも示した。



https://x.com/K_Ishi_AI/status/1852895630328512765

Simhi, A., Herzig, J., Szpektor, I. & Belinkov, Y. Distinguishing Ignorance from Error in LLM Hallucinations. Preprint at https://doi.org/10.48550/arXiv.2410.22071 (2024).

対策例：LLMに「応答する前に考える」ことを学習させる

akira
@AkiraTOSEI

arxiv.org/pdf/2410.10630
LLMに「応答する前に考える」ことを学習させる研究。思考・応答に分離した複数のLLM出力から良い応答を出すものを学習データとし、LLMに思考させる方法を学ばせる。思考生成なしの直接応答と比較して、一般的なタスクでも精度が向上することを確認した。

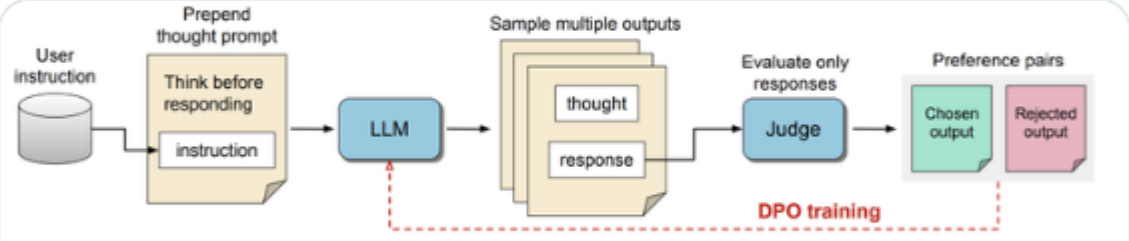


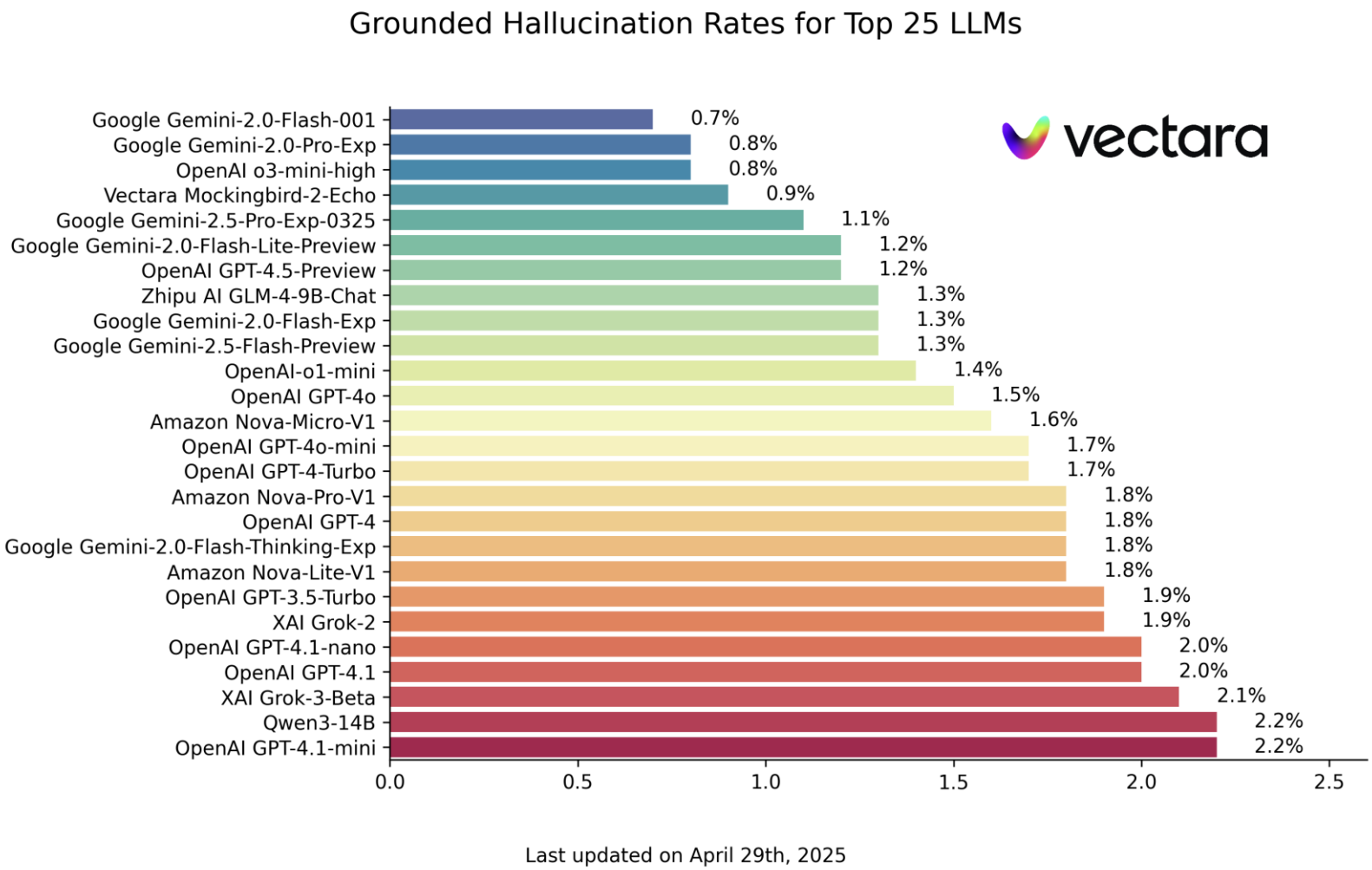
Figure 1: **Thought Preference Optimization:** We start by prompting the LLM to generate thoughts before its response. After sampling different outputs, we feed the response parts to the judge model which determines the best and worst ones. Then we use the corresponding full outputs as chosen and rejected pairs for DPO optimization. We perform multiple iterations of this training.

午後2:07 · 2024年11月1日 · **1万** 件の表示

20 147 70

Wu, T. *et al.* Thinking LLMs: General Instruction Following with Thought Generation. Preprint at https://doi.org/10.48550/arXiv.2410.10630 (2024).

しかしながら決してゼロにはならない Hallucination



ハルシネーション率の最新情報（不定期で更新）

https://github.com/vectara/hallucination-leaderboard

ハルシネーションは消せないし、大規模化し訓練するほど、少なくなる一方で検出困難になる

- ハルシネーションは消せず、AI モデルを人間の思考と認識の代替として用いるべきでなく、拡張として認識すべき。
  ※「構造的幻覚（Structural Hallucinations）」としてチューリングマシンやゲーデルの不完全性定理を扱っている。
- Banerjee, S., Agarwal, A. & Singla, S. LLMs Will Always Hallucinate, and We Need to Live With This. Preprint at https://doi.org/10.48550/arXiv.2409.05746 (2024).

- より大きく、より長く訓練されたLLMにてハルシネーションが少なくなる一方で、逆に残ったハルシネーションの検出がますます難しくなる。
- Hron, J. *et al.* Training Language Models on the Knowledge Graph: Insights on Hallucinations and Their Detectability. Preprint at https://doi.org/10.48550/arXiv.2408.07852 (2024).

- 大規模でより指導しやすいモデルは全体的に信頼性が低い。機能が向上しているにもかかわらず、エラーが最小限であるべき領域で頻繁に失敗し、ハルシネーションが増える
- Zhou, L. *et al.* Larger and more instructable language models become less reliable. *Nature* 1–8 (2024) doi:10.1038/s41586-024-07930-y.

- 僅かに古くなってはいるが、AI活用方法と問題点の双方向の視点からのまとめ
- 児玉龍彦. FORUM 日本型セルフケアへのあゆみ・第21回 ChatGPTの可能性と危険性:ハルシネーション問題. Preprint at https://doi.org/10.32118/ayu28609743 (2023).
  https://www.jspa-nsc.com/dcms_media/other/igakunoayumi_21.pdf

医療などの分野では疾患の診断時などで行う臨床推論や、患者への説明責任が問われる状況では重大な問題となる

https://www.niigata-u.ac.jp/news/2023/409517/

【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

内閣府　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ての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**

最近のLLM関連用語と解説URL

- o3, o4-mini についてちゃんと理解する【o1と何が変わったの？】
  https://zenn.dev/microsoft/articles/openai_o3_o4mini_catchup
- MCP は大きな問題を抱えている？導入前に検討すべき課題と知っておくべきセキュリティリスク
  https://zenn.dev/cloud_ace/articles/model-context-protocol-problem
- AIエージェント vs. エージェント型AI：概念分類、応用、課題
  Sapkota, R., Roumeliotis, K. I. & Karkee, M. AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges. Preprint at https://doi.org/10.48550/arXiv.2505.10468 (2025).
- Google DeepMind、AIが自らアルゴリズムを発見・進化させる「AlphaEvolve」発表
  https://www.itmedia.co.jp/aiplus/articles/2505/15/news108.html
  https://deepmind.google/discover/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/
  https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/AlphaEvolve.pdf
- Google I/O 2025
  https://blog.google/intl/ja-jp/google-io-2025/

自動化バイアスの問題
多くのネット上での信頼しうる資料からのまとめ

- AIは「誰でも作れて、誰でも使える」（かの如くの）時代へと突入してきているが、「使う側の能力」が十分ではない場合、問題が生じうる。（ ⇒ 自動化バイアス≒AIを過度に信じて落とし穴に填まる）
- 例：バイブコーディング
  - バイブコーディング（Vibe Coding）とは？
    定義からツール、導入方法まで徹底解説
    https://www.ai-souken.com/article/what-is-vibe-coding
  - Model Context Protocol (MCP) とは？
    仕組みやRAGとの違いを解説
    https://www.ai-souken.com/article/what-is-mcp

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
Cabinet Office

https://www8.cao.go.jp/cstp/society5_0/

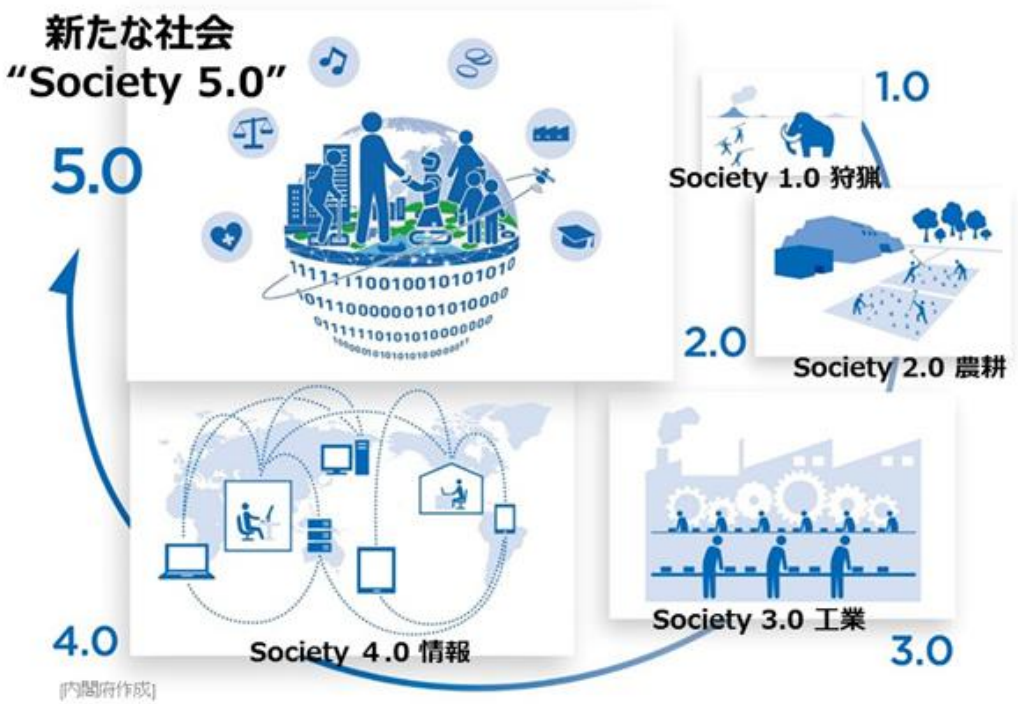
内閣府ホーム > 内閣府の政策 > 科学技術政策 > Society 5.0

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。



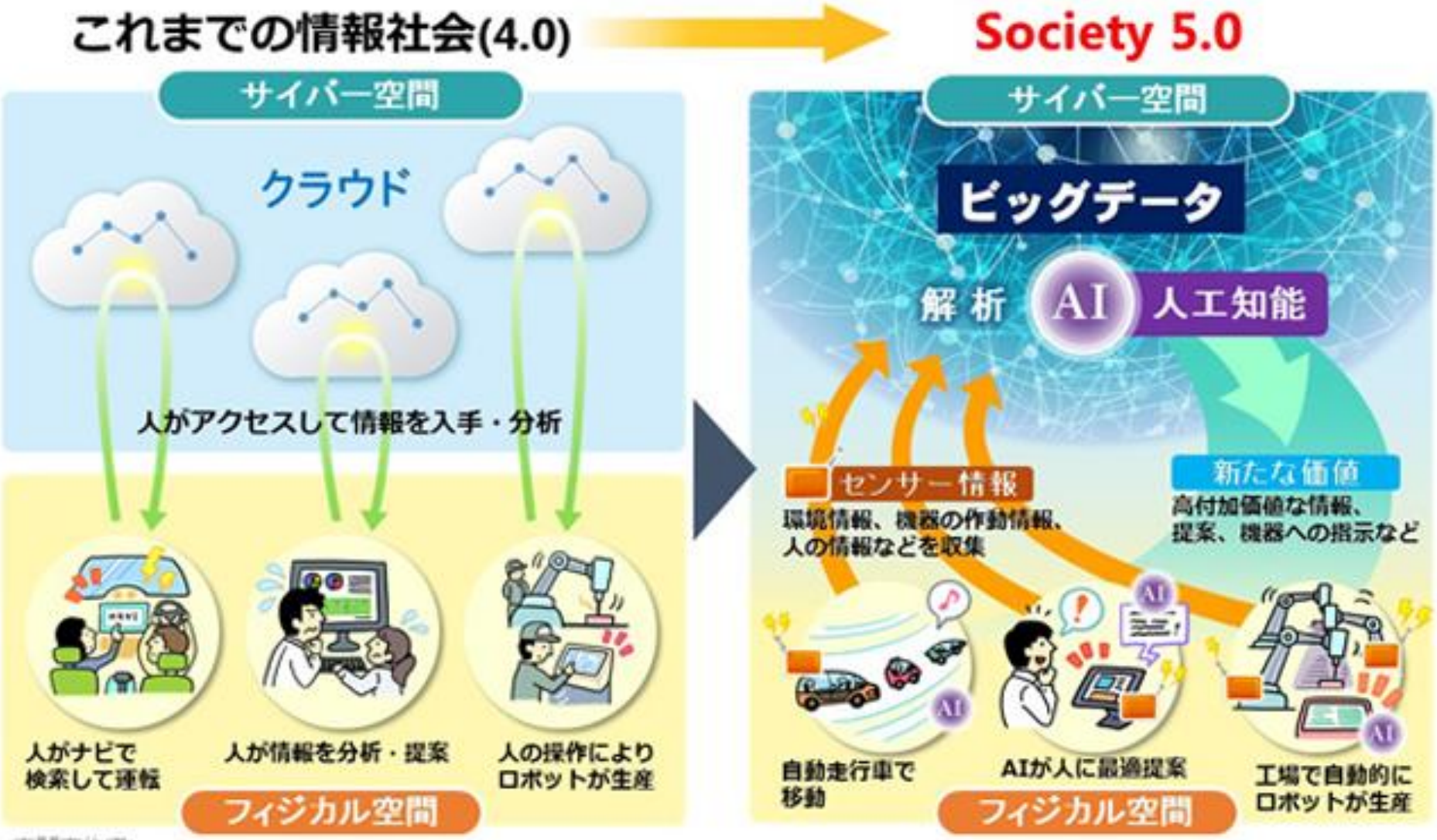
「Society 5.0」とは
第5期科学技術基本計画に基づき我が国が目指すべき未来社会の姿として初めて提唱された

- 「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）

https://www8.cao.go.jp/cstp/society5_0/
から文字情報を抜粋

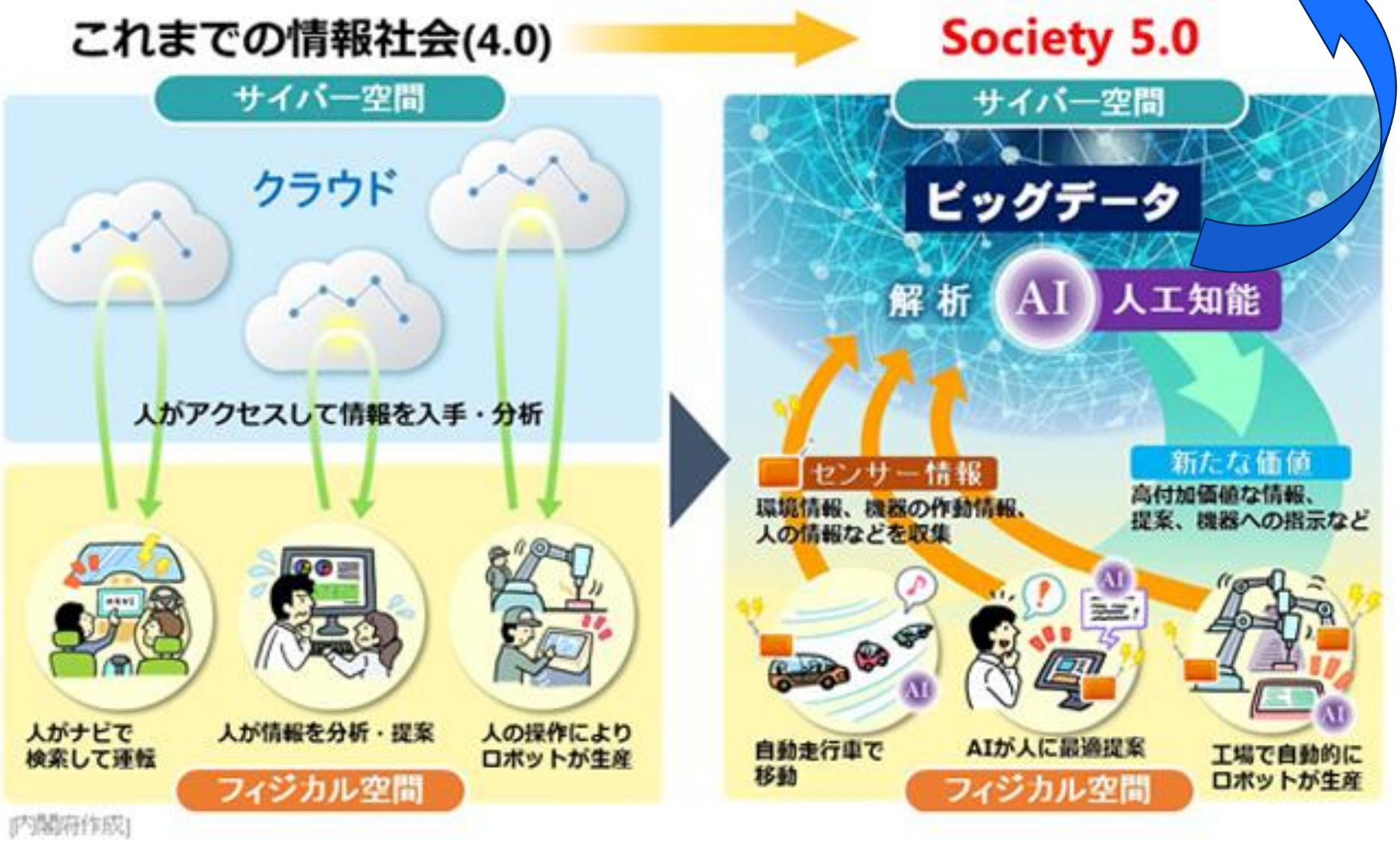
これまでの社会
知識・情報の共有、連携が不十分
IoTで全ての人とモノがつながり、新たな価値がうまれる社会
これまでの社会
地域の課題や高齢者のニーズなどに十分対応できない
イノベーションにより、様々なニーズに対応できる社会
Society 5.0
AIにより、必要な情報が必要な時に提供される社会
AI
これまでの社会
必要な情報の探索・分析が負担
リテラシー（活用能力）が必要
ロボットや自動走行車などの技術で、人の可能性がひろがる社会
これまでの社会
年齢や障害などによる、労働や行動範囲の制約

[内閣府作成]


これまでの情報社会(4.0)
Society 5.0
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
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

[内閣府作成]

記号接地を誤ると、危険



[内閣府作成]

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

[内閣府作成]

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

[内閣府作成]

年齢・性別に関係なく皆に恩恵
快適
日々の暮らしが
ラクラク・楽しく
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
活力
Society 5.0
質の高い生活
煩わしい作業から解放され、時間を有効活用
[内閣府作成]
経済発展と社会的課題の解決を両立
より便利で安全・安心な生活

最近重視されている倫理
AI倫理（1）

- AI倫理とは？－生成AIを開発・活用するために企業が考えなければいけないこと－
  https://www.dir.co.jp/world/entry/solution/ai-ethics
- 看護に求められる倫理概念をロボットやAIに実装できるのか？
  https://www.tus.ac.jp/today/archive/20230706_1527.html

最近重視されている倫理
AI倫理(2)・政府関係資料

- AIと社会 ー倫理・リスク・包摂ー(財務省)
  https://www.mof.go.jp/pri/research/seminar/fy2022/lm20221109.pdf
- AI倫理に関する国際対応について(内閣府)
  https://www8.cao.go.jp/cstp/ai/ningen/r4_1kai/siryo1.pdf
- UNESCOのAI倫理勧告(文部科学省)
  https://www.mext.go.jp/content/20220826-mxt_koktou02-000024125_4.pdf
- 欧州(EU)AI規制法
  https://artificialintelligenceact.eu/
  https://www.eu.emb-japan.go.jp/files/100741144.pdf
  https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689
  https://www.soumu.go.jp/main_content/000826707.pdf
  https://www.pwc.com/jp/ja/knowledge/column/awareness-cyber-security/generative-ai-regulation10.html

欧州（EU）AI規制法での汎用目的型AIモデル（ GPAI; General Purpose AI ）いわゆる「マルチモーダルなLLM」が該当する

「汎用人工知能（AGI; Artificial General Intelligence）とは定義が異なる
https://artificialintelligenceact.eu/article/3/

(63) 'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market;

Related: Recitals 97, 98, and 99

https://www.businesslawyers.jp/articles/1431 Business Lawyers「EU AI法の概要と日本企業に必要な対応を解説」

GPAIモデルの定義（3条63号）

大量のデータを使用して、自己教師にて大規模に訓練された場合を含め、顕著な汎用性を示し、モデルが市場に投入される方法にかかわりなく、広範囲の異なるタスクを適切に実行でき、さまざまな下流のシステムやアプリケーションに統合できるAIモデルをいう。ただし、市場に投入される前の研究、開発、またはプロトタイピング活動に使用されるAIモデルはGPAIから除く。

政府・政策関連サイト

※こちらのGPAIは欧州AI規制法でのGPAIとは異なる用語

AIに関するグローバル・パートナーシップ

（ GPAI ：The Global Partnership on Artificial Intelligence）

- GPAI（OECD）
  - https://oecd.ai/en/gpai-declarations
- GPAI専門家支援センター（GPAI Expert Support Center、日本・東京）
  - https://www2.nict.go.jp/gpai-tokyo-esc/
- AISI（AI Safety Institute、日本）
  - https://aisi.go.jp/
  - https://oecd.ai/en/wonk/ai-safety-institute-networks-role-global-ai-governance
  - https://oecd.ai/en/wonk/risk-thresholds-for-frontier-ai-insights-from-the-ai-action-summit

日本での生成AIガイドライン
注意喚起関係

- テキスト生成AI利活用におけるリスクへの対策ガイドブック（α版、デジタル庁）
  https://www.digital.go.jp/resources/generalitve-ai-guidebook
- 文章生成AI利活用ガイドライン・活用事例集（東京都・デジタルサービス局）
  https://www.digitalservice.metro.tokyo.lg.jp/business/ict/ai-guideline
- 自治体における生成AIガイドライン策定の必要性（日本総研）
  https://www.jri.co.jp/page.jsp?id=109760
- AI事業者ガイドライン（総務省・経産省）
  https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02ryutsu20_04000019.html
  https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20240419_report.html
- DeepSeek 等の生成 AI の業務利用に関する注意喚起（デジタル庁・事務連絡）
  https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/d2a5bbd2-ae8f-450c-adaa-33979181d26a/e7bfeba7/20250206_councils_social-promotion-executive_outline_01.pdf

テキスト生成AI利活用におけるリスクへの対策ガイドブック（α版）

https://www.digital.go.jp/resources/generalitve-ai-guidebook

２．１ テキスト生成 AI の利活用が不適切であるケースにテキスト生成 AI を用いるリスク

テキスト生成 AI の利活用が不適切である場合にテキスト生成 AI を用いることによりリスクが発生することが考えられる。例えば以下の 6 類型が考えられる。

(1) テキスト生成 AI による回答の期待品質が高すぎる場合
(2) そもそも人間が行うべき仕事であり、AI による代替が不適切な場合
(3) 特定の資格を必要とする作業であることが法令に定められている場合
(4) テキスト生成 AI の知識に無い事項を答えさせる場合
(5) テキスト生成 AI 利活用の構築・運用費が費用対効果に見合わない場合
(6) その他、AI 事業者ガイドラインの「共通指針」（ページ 12~20）に違反する場合

(3)は例えば、症状についての記述から必要な薬をテキスト生成 AI に推定させ、薬の処方箋を自動発行するケースは不適切な事例である。医師法や歯科医師法に違反する。

文化庁
AIと著作権について

- https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html
  から抜粋
- デジタル化・ネットワーク化の進展に対応した柔軟な権利制限規定に関する基本的な考え方（著作権法第30条の４，第47条の４及び第47条の５関係）
  https://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30_hokaisei/pdf/r1406693_17.pdf
- AIと著作権に関する考え方について
  https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901_01.pdf
- AIと著作権に関するチェックリスト＆ガイダンス
  https://www.bunka.go.jp/seisaku/chosakuken/pdf/94097701_01.pdf
- 令和6年度著作権セミナー「AIと著作権Ⅱ」
  https://www.bunka.go.jp/seisaku/chosakuken/pdf/94097701_02.pdf

新潟大学生成AI利用促進ガイドライン
新潟大学生成AI利用における注意点

新潟大学
情報基盤センター

関連リンク　言語選択　検索　メニュー

ホーム > サービス一覧 > 教育・研究 > 教育・研究　その他 > 新潟大学生成AI利用促進ガイドライン

新潟大学生成AI利用促進ガイドライン

更新日:2025年 4月 10日

新潟大学の教職員、学生等が教育・研究及び業務で生成AIを活用することで、業務効率の改善や新しいアイデア出し、効果的な学修の支援等が期待でき、また、高等教育機関として、生成AIの仕組みを理解し有効に活用する人材の育成は急務であり、積極的な利用を推奨します。

本ガイドラインは、本学の教職員及び学生等が生成AIの特性を理解し、利用するために注意すべき事項を解説するものです。

- 新潟大学生成AI利用促進ガイドライン　学内限定
- 新潟大学生成AI利用における注意点　学内限定

新潟大学生成AI利用促進ガイドライン

https://www.cais.niigata-u.ac.jp/wp/wp-content/uploads/gateway/guideline_generativeAI_202503.pdf

目　次

1．本ガイドラインの目的
2．生成AIとは
3．本ガイドラインが対象とする範囲
4．生成AIの利用
（1）研究活動における生成AIの利用
（2）教育活動における生成AIの利用
（3）業務運営における生成AIの利用
5．生成AIのリスクと気を付けるべきこと
（1）生成AIのリスク・正確性
（2）生成AIのリスク・公平性
（3）生成AIのリスク・著作権侵害
（4）生成AIのリスク・情報管理
（5）生成AIのリスク・悪用
6．注意事項（詳細）

新潟大学生成AI利用における注意点

https://www.cais.niigata-u.ac.jp/wp/wp-content/uploads/gateway/notification_generativeAI_202503.pdf

真の強さを学ぶ。
新潟大學
NIIGATA UNIVERSITY

新潟大学生成AI利用における注意点

令和７年３月
DX推進委員会

目　次

１．生成AIへのデータ入力の禁止事項

２．データ入力に際しての注意点

３．生成物を利用するに際しての注意点

- コリングリッジのジレンマ
  - 技術が社会で使われる前にその影響力を予測することは難しい
  - 一度普及してしまった技術は制御するのが難しい

- 「社会実験」ではなく「実験社会」
- 私たち一人一人の役割と責任とは？
- どのような生き方・働き方を目指す？

「応用倫理」

東京大学 未来ビジョン研究センター
日本ディープラーニング協会
理化学研究所 AIPセンター
江間有沙 先生の最後のスライド

https://www.mof.go.jp/pri/research/seminar/fy2022/lm20221109.pdf p.34

深層学習システムのデモ　テストデータの評価

GPUの無いNotePCでもローカルで動作可能

- 元データ・元資料
  - https://github.com/TMDU-AI/AI-yodosha
    の第5章
- 上記からForkし、SonyのNeural Network Console – Desktop版に移植した版
  - https://github.com/aujinen/AI-yodosha/tree/main/Chapter5/SonyNNC_Desktop
- Sony Neural Network Console – Desktop版
  - https://nnabla.org/ja/nnc/
  - Neural Network Console – Windows版でも動作する（Windows専用なので使い勝手が良い）が、データ構造を含めデータの取り扱い関連が全く異なるので注意（ここでは説明を省く）
  - データセットの流用（ハードリンク使う移植）については、下記参照
    https://github.com/aujinen/AI-yodosha/tree/main/Chapter5/SonyNNC_Desktop/move_to_datasets
  - 学習後のモデルでのtestデータの評価はInferenceにて実行をクリックし、アップロードエリアに複数枚のテストデータ画像をドロップすれば良い