

AIの現在と今後 われわれの生活はどう変わるか？

岡谷 貴之

東北大学大学院情報科学研究科
(兼 理化学研究所AIPセンター)

研究分野：コンピュータビジョン

「画像AI」の研究分野

1990

2000

2010

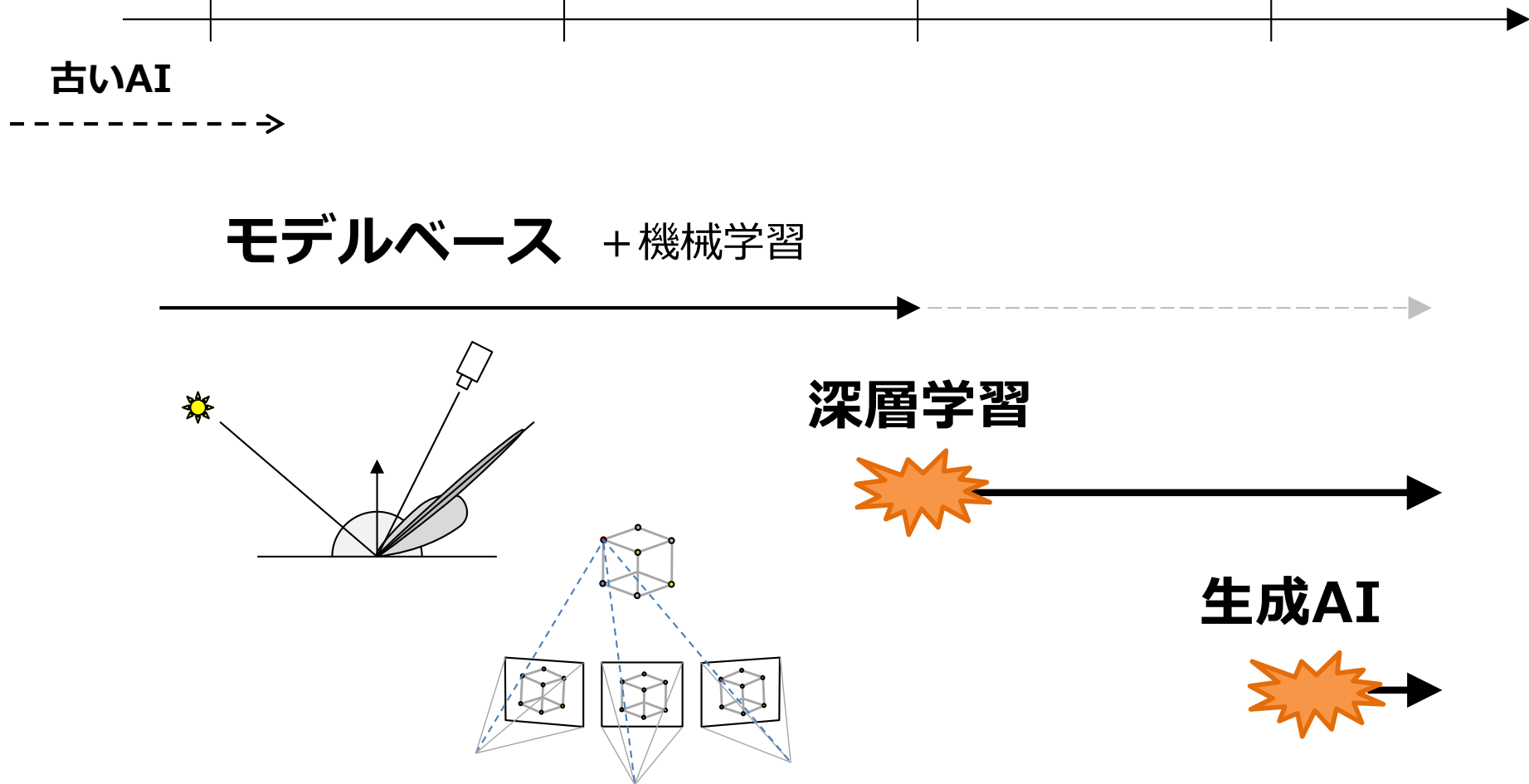
2020

古いAI

モデルベース + 機械学習

深層学習

生成AI



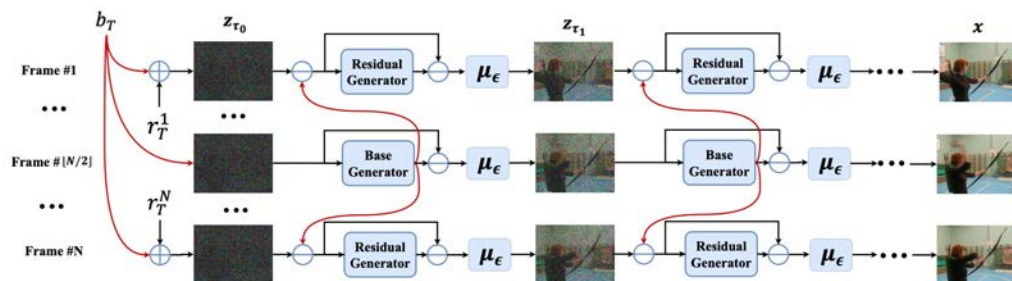
生成AIの急速な発展：OpenAI Sora

<https://openai.com/research/video-generation-models-as-world-simulators>



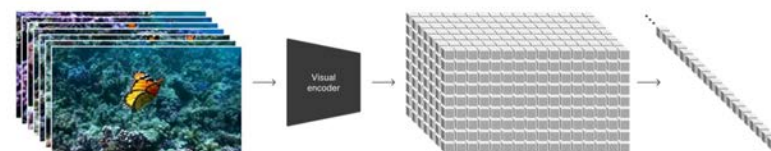
映像生成のブレークスルー

Sora以前：静止画合成＋フレーム間整合性確保

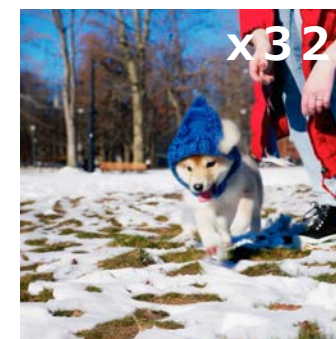


[VideoFusion, CVPR2023]

動的 3次元シーンの符号化：



計算量に比して向上する品質



生成AIの急速な発展 : OpenAI GPT-4V

Yang(Microsoft)+, The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision), arXiv2023.9

Sec. 4.4 Scene Text, Table, Chart, and Document Reasoning

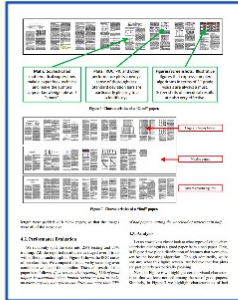
Document Understanding

Prompt:

This is a paper "Paper Gestalt."
Describe the paper in details
and highlight their contribution.

Prompt Continued

Page 4



Page 5



GPT-4V:

The paper "Paper Gestalt" is a research paper written by Carven von Beamensquash from the Department of Computer Science at the University of Phoenix. The paper discusses the visual quality of a computer science research paper and how it affects the overall impression of the paper.

The paper begins with an introduction to the concept of paper gestalt, which is the overall impression of a paper. The author argues that the visual quality of a paper is an important factor in creating a good impression. The paper then discusses previous work in this area and highlights the contribution of this paper, which is the use of a supervised learning algorithm to evaluate the visual quality of a paper.

The paper then presents the methodology and experimental results. The author collected a dataset of 1196 computer science research papers and used a supervised learning algorithm to evaluate the visual quality of the papers. The algorithm was trained on a set of features extracted from the papers, such as the number of figures, the number of equations, and the number of references.

The algorithm was trained on a set of features extracted from the papers, such as the number of figures, the number of equations, and the number of references.

The author discusses the contribution of the paper and suggests that future work could focus on improving the algorithm by incorporating more features and training it on a larger dataset.

The paper also suggests that the algorithm could be used to evaluate the visual quality of other types of documents, such as technical reports and dissertations.

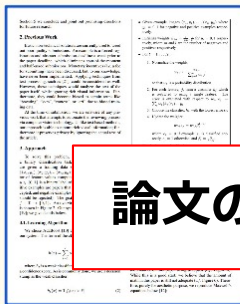
Overall, the contribution of this paper is the use of a supervised learning algorithm to evaluate the visual quality of a computer science research paper. This is a novel approach that could help researchers improve the visual quality of their papers and make a better impression on their readers.

論文の画像 → 論文を要約

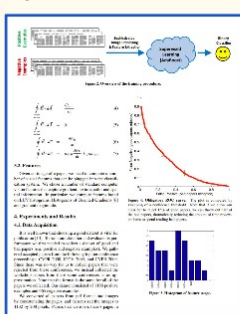
Page 1



Page 2



Page 3

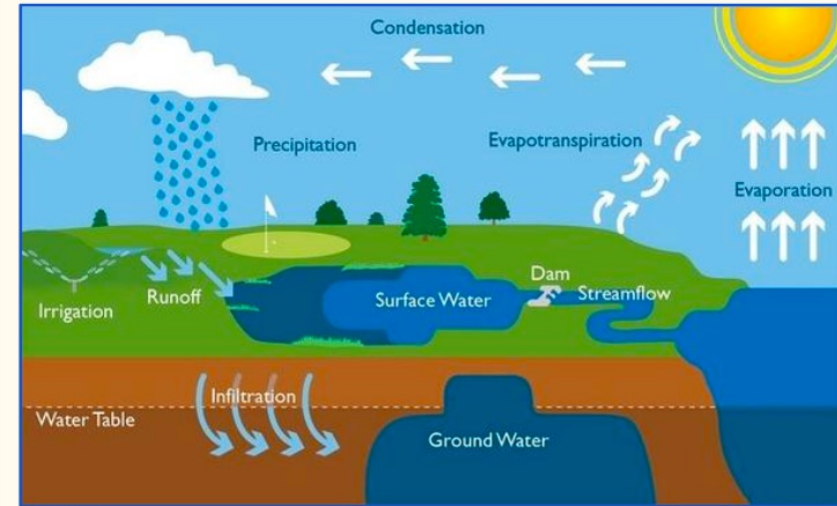


Sec. 4.3 Multimodal Knowledge and Commonsense

Science and Knowledge

Prompt:

Suppose you are a teacher, please use the figure to explain the distinction between evaporation and evapotranspiration.



GPT-4V:

Okay class, today we are going to learn about the distinction between evaporation and evapotranspiration. Let's take a look at this figure.

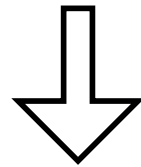
First, let's understand what evaporation is. Evaporation is the process by which water is changed from its liquid state to its gaseous state. This usually takes place at the surface of a liquid, like the ocean or a lake, when the water molecules have enough energy to escape into the air.

Now, let's talk about evapotranspiration. Evapotranspiration is the combination of evaporation and transpiration. Transpiration is the process by which water is taken up by plants and released into the air through their leaves.

教師の立場で、
入力画像に基づいて
科学の概念を説明

In summary, evaporation is the process by which water is changed from its liquid state to its gaseous state, while evapotranspiration is the combination of evaporation and transpiration. Both these processes contribute to the water cycle and are important for maintaining the balance of water in our environment.

われわれの生活はどう変わるか？



わかりません

(AIの将来像は、研究者間でも意見が対立)

「中央値レベルの知能を持つ人が出来ることが出来る」

AGIの実現をめぐる対立

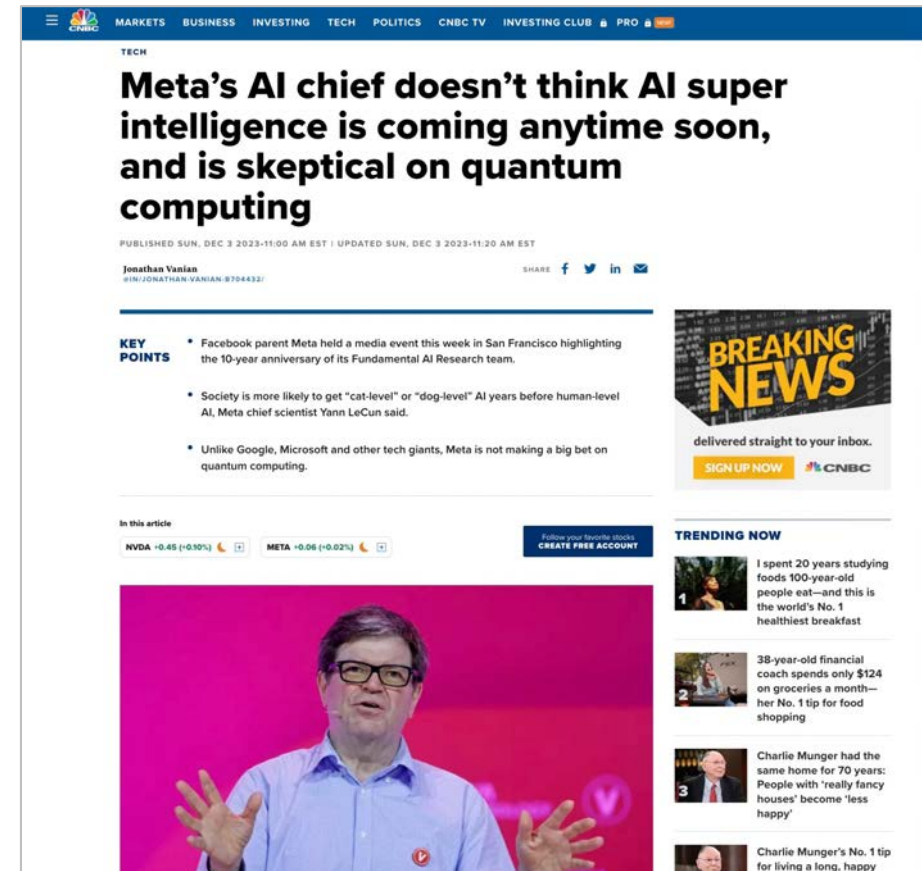
楽観論 ≡ 規制派

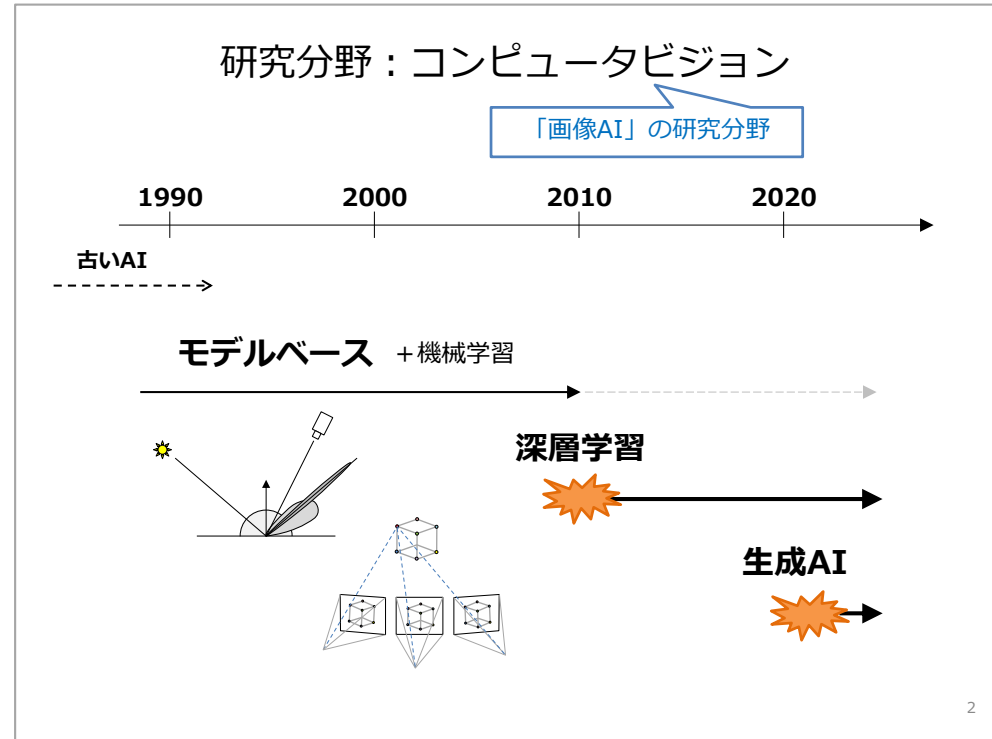
- 今の生成AIの延長で十分？
- 強い懸念から規制論が盛ん
 - Hinton, Bengio, Sutskever, ...



懐疑派 ≡ 推進派

- 生成AI/DLでは不十分
 - Lecun, ...
 - LLMは100%記憶で説明可能---F. Chollet
 - DLは大規模な最近傍探索---P. Domingos





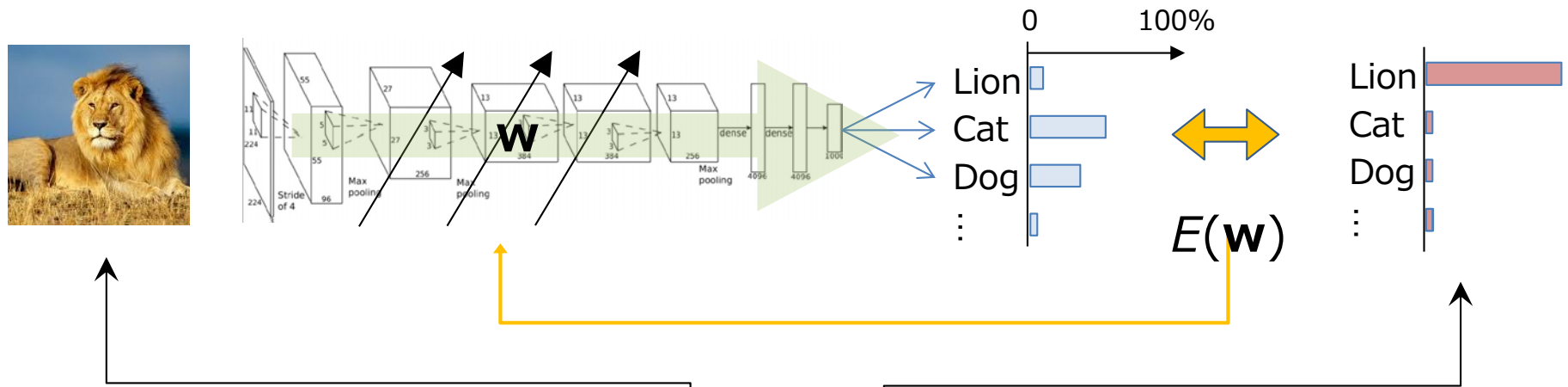
深層学習

深層学習 (ディープラーニング)

- 深層ネットワークを大量のデータで学習

数十層からなる数千万オードの
パラメータを持つネットワーク

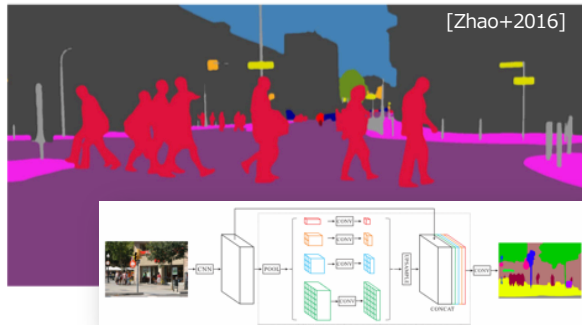
数万~数百万
の学習データ



深層学習が大成功 (2011~)

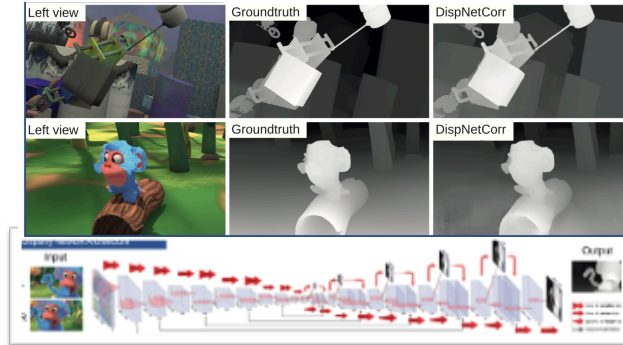
他の方法では望めない高い性能

画素レベルの認識

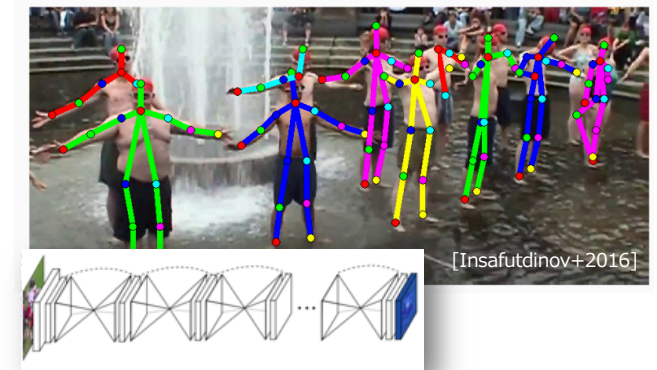


視差 (ステレオ画像)

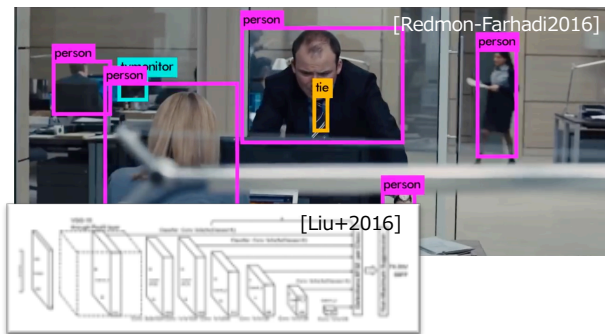
[Mayer+2017]



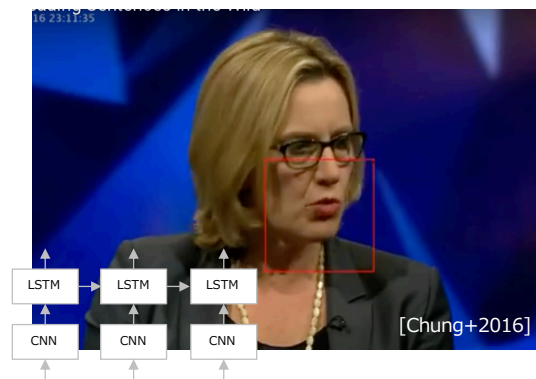
人体ポーズ



物体検出

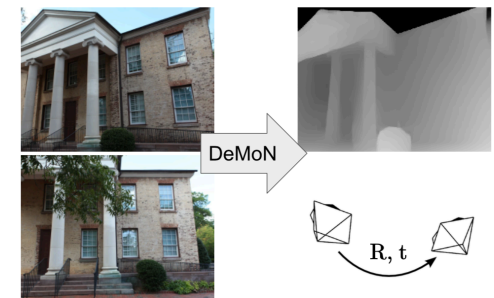


読唇

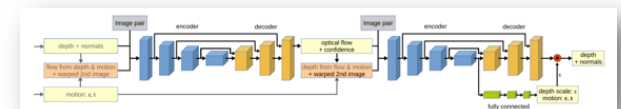
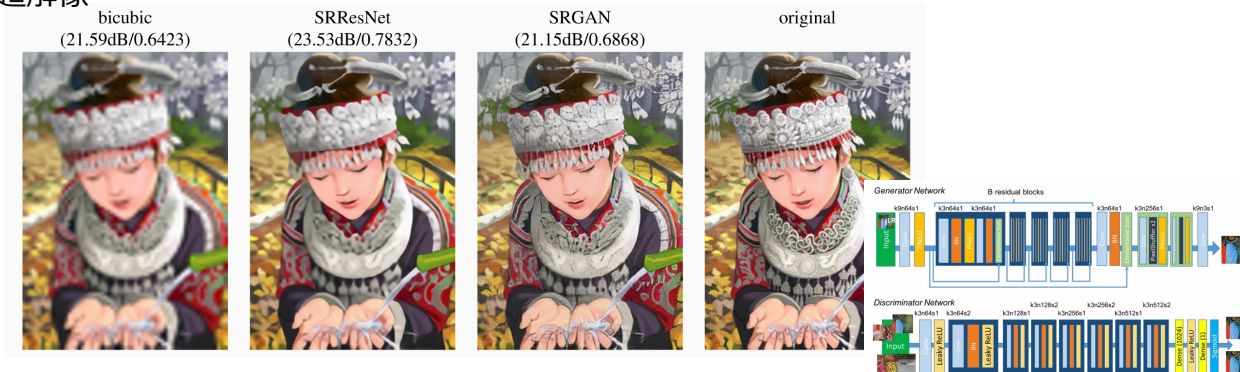


カメラ姿勢

[Ummenhofer+2017]

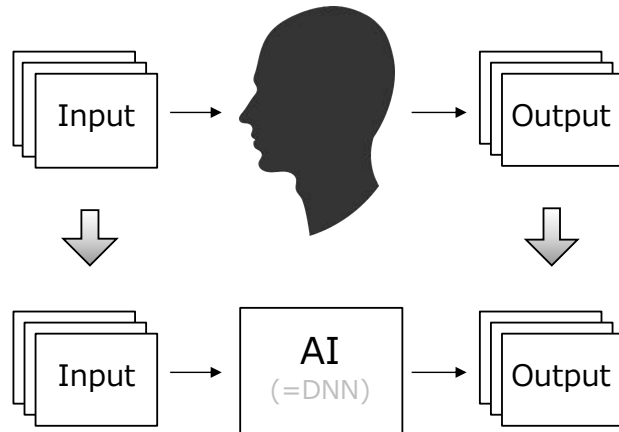


超解像



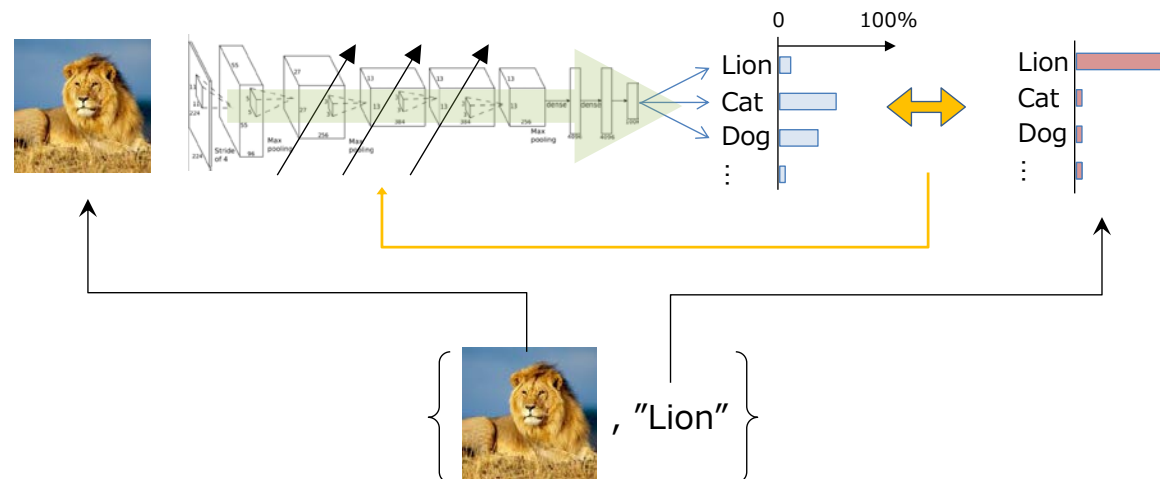
深層学習が実現するもの

- “Cognitive Automation” = 人の認知過程の自動化
 - 人の認知の Input-Output を再現



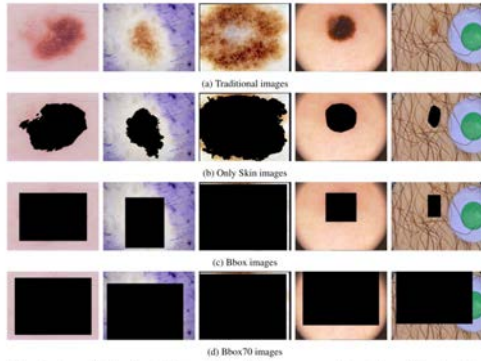
- 何が Input, Output かは人が決める
- そのタスクの
 - 専用モデルで
 - 専用データを学習

完全教師あり学習



課題：ショートカット学習

不明な画像特徴



[Bissoto+2019]

ドメインシフト



見慣れない姿勢の物体



[Alcorn+2018]

形状よりテクスチャ優位



(b) Content image
71.1% **tabby cat**
17.3% grey fox
3.3% Siamese cat

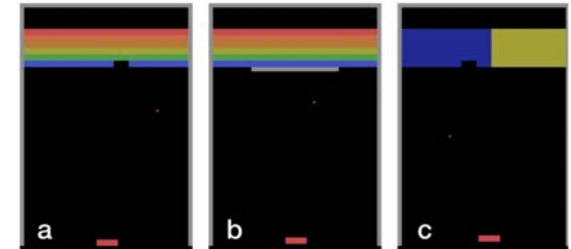
(c) Texture-shape cue conflict
63.9% **Indian elephant**
26.4% indri
9.6% black swan

[Geirhos+2019]

BERT

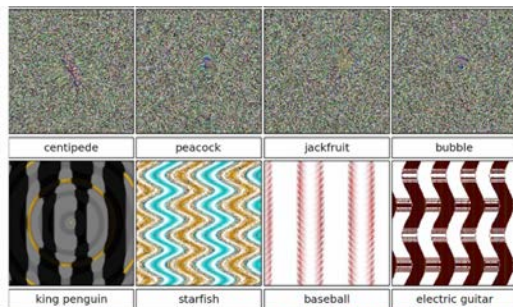


深層強化学習×ビデオゲーム



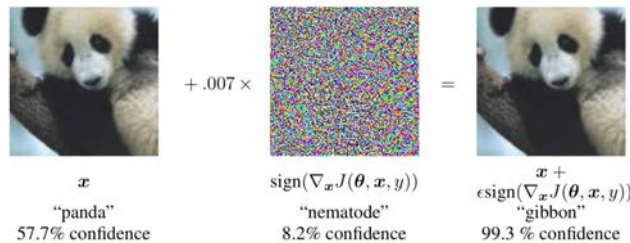
[Kansky+2017]

Fooling samples



[Nguyen+2014]

Adversarial Examples



[Goodfellow+2015]

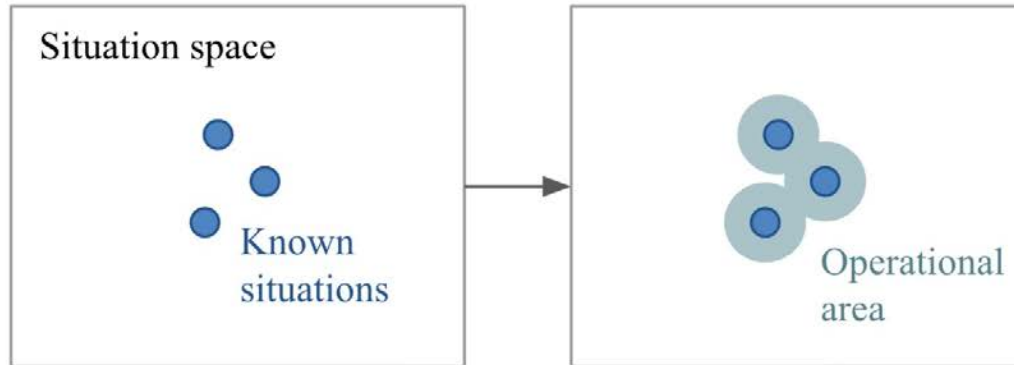
“Natural Adversarial Examples”



[Hendrycks+2020]

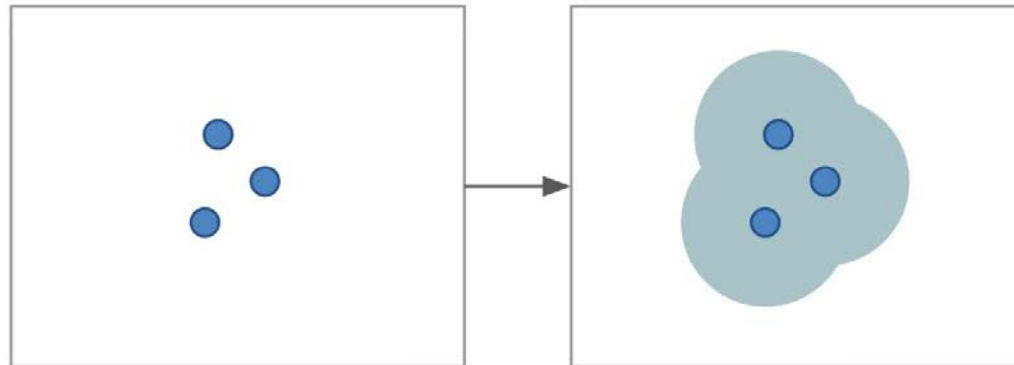
学習する機械

深層学習 … 「5を聞いて10を知る」



Lower-intelligence system:
lower information
conversion ratio

人間 … 「1を聞いて10を知る」



Higher-intelligence system:
higher information
conversion ratio

[Chollet, arXiv2019]

壁：大量データが必要

IMAGENET <http://image-net.org/index>

- 一般物のクラス認識
- 21841クラス・14,197,122枚
- スタンフォード大・プリンストン大他



CITYSCAPES <https://www.cityscapes-dataset.com/examples/>

- 車載カメラの画素単位の物体クラス
- 30クラス・50都市・5,000枚/20,000枚
- ダイムラー・ダルムシュタット工科大他

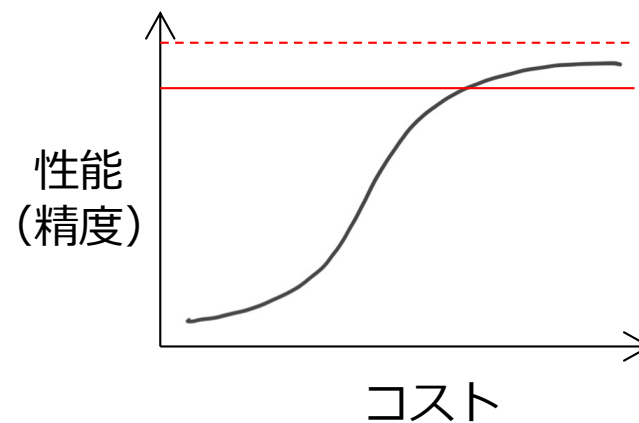


COCO <http://cocodataset.org>

- 物体クラスとその画像領域
- 80物体+91物体以外クラス・330,000枚



性能-コストの「S-カーブ」



CelebA <http://mmlab.ie.cuhk.edu>

- 顔画像
- 40属性・10,177人物・200000枚
- 香港中文大学

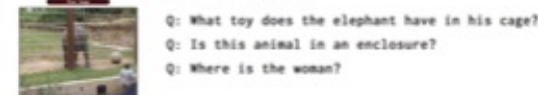
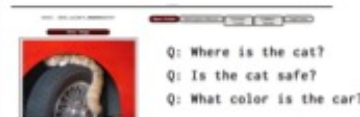


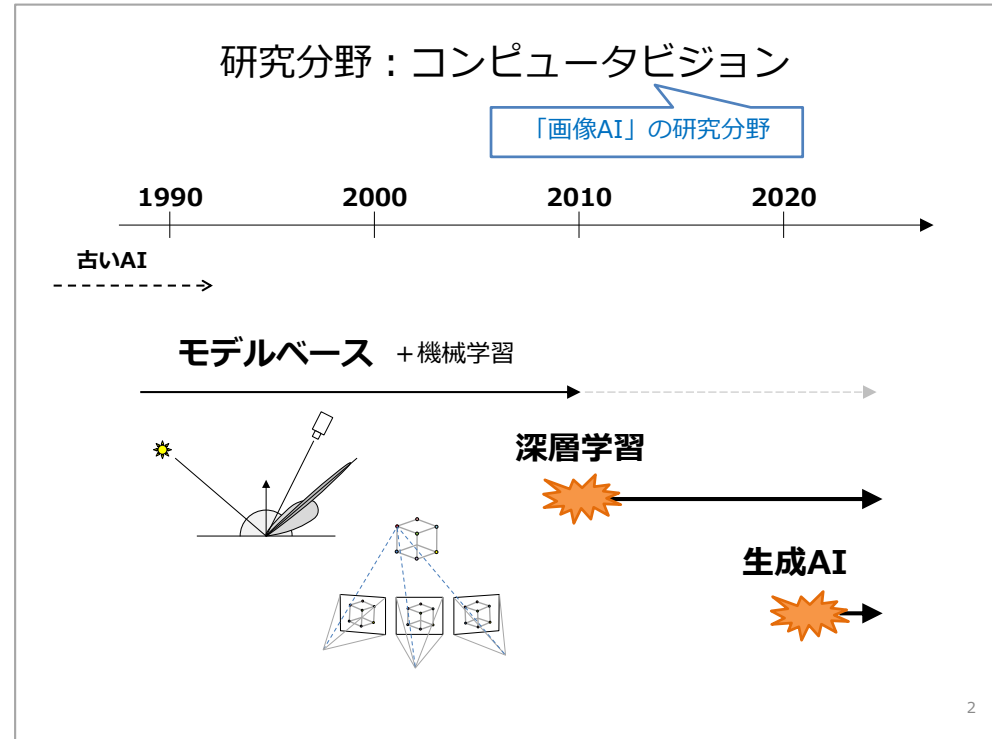
MPII Human Pose

- 人体ポーズ
- 前身関節位置
- Max Planck Institute for Computer Graphics and Vision



- 画像中のシーンに対する質問と答え
- 画像当たり平均5.4質問・10回答・265,016枚
- Virginia Tech., Georgia Tech.

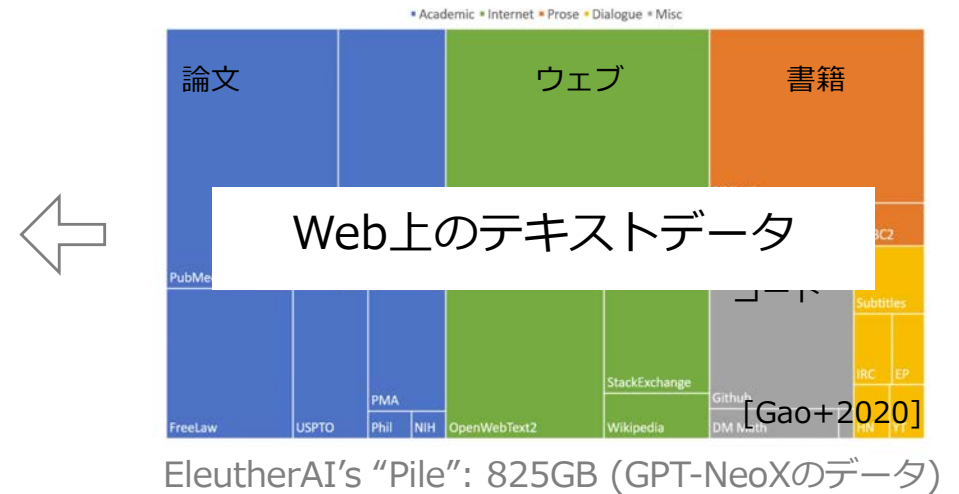
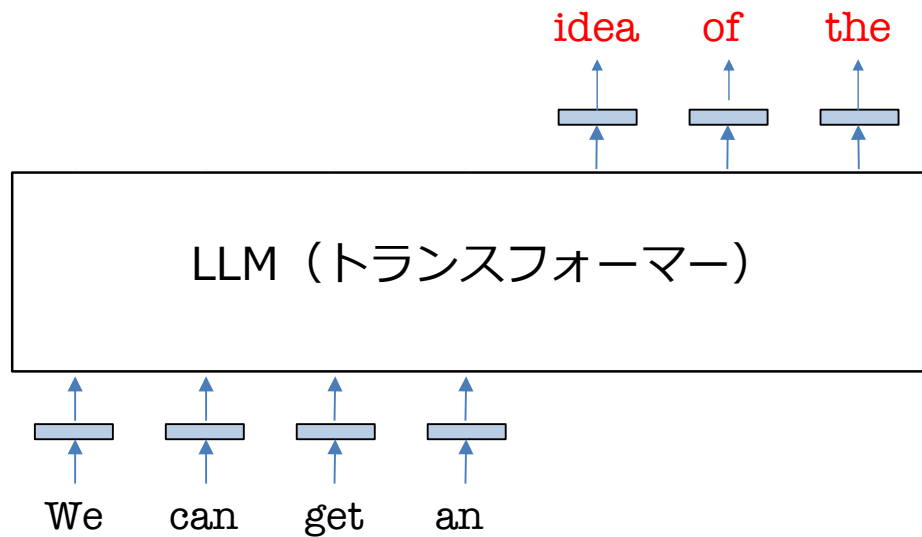




生成AI(LLM)の登場

大規模言語モデル(LLM)

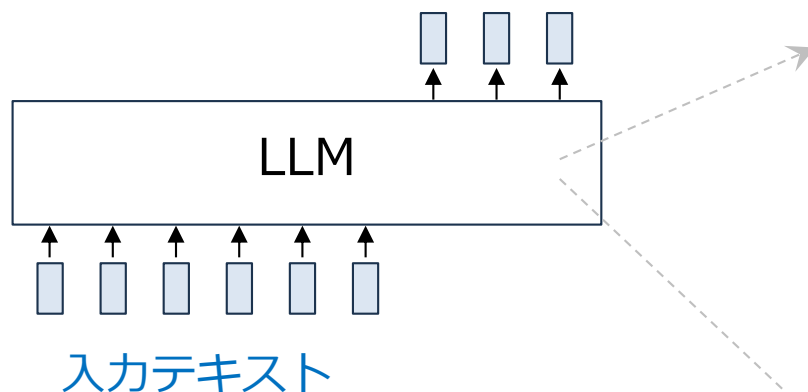
- 文の「次の単語」を予測するニューラルネット
 - 大量のテキストを用いた「自己教師学習」



- さらに、ユーザが求める振る舞いをさせるべく **ファインチューン**
 - アラインメント・Instruction Tuning

LLMのユニークさ

- 生成と推論が一体化
- 学習を容易に大規模化可能



タイトル → ニュース記事 [Brown+2020]

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.
The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the U.S.," but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10 percent of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

数学の問題を解く [Hendrycks+2021]

The following are multiple choice questions about high school mathematics.

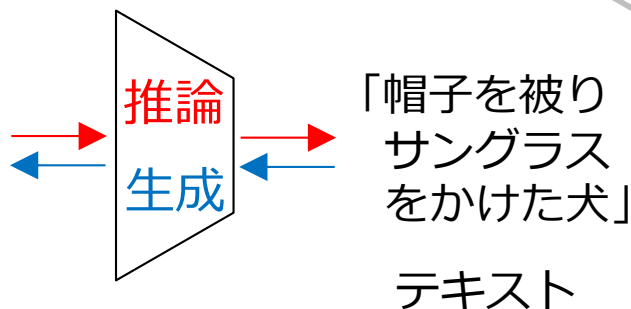
How many numbers are in the list 25, 26, ..., 100?
(A) 75 (B) 76 (C) 22 (D) 23
Answer: B

Compute $i + i^2 + i^3 + \dots + i^{258} + i^{259}$.
(A) -1 (B) 1 (C) i (D) $-i$
Answer: A

If 4 daps = 7 yaps, and 5 yaps = 3 baps, how many daps equal 42 baps?
(A) 28 (B) 21 (C) 40 (D) 30
Answer: C



画像



学習する機械

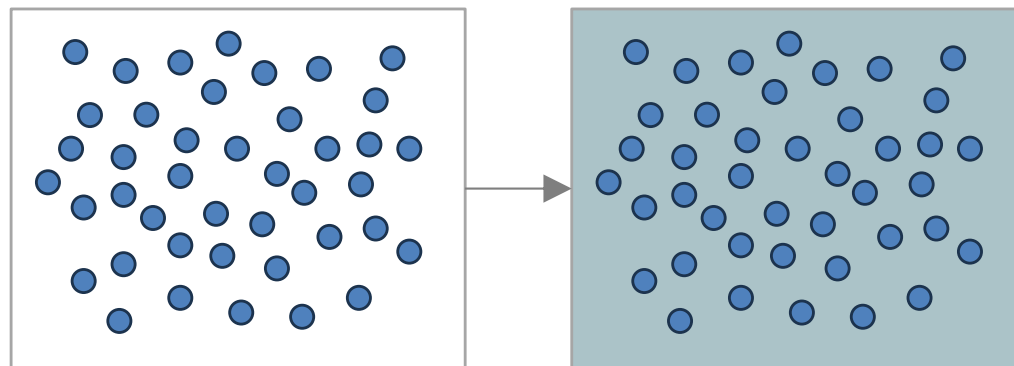
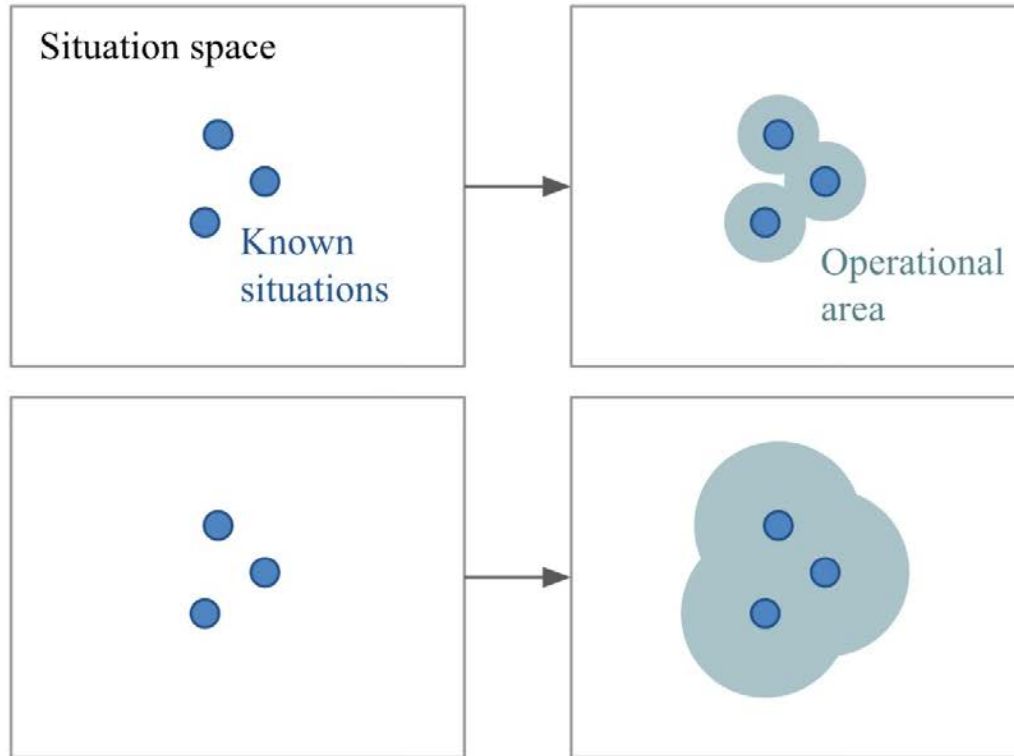
深層学習 … 「5を聞いて10を知る」

Lower-intelligence system:
lower information
conversion ratio

人間 … 「1を聞いて10を知る」

Higher-intelligence system:
higher information
conversion ratio

[Chollet, arXiv2019]



LLMからAGIへ

… 「1万を聞いて全部を知る」

**Situation spaceの広さ
vs. サンプルングコスト**

→ 言語空間なら可能かも？

AGIの実現をめぐる対立

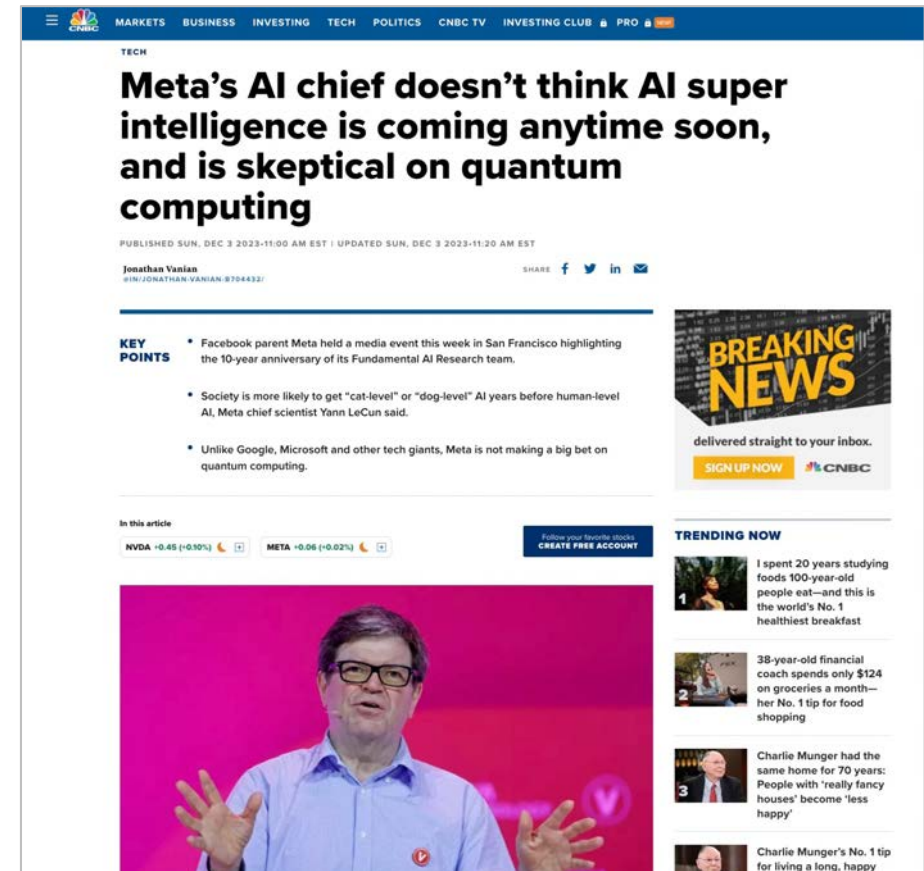
楽観論 = 懸念派

- 今の生成AIの延長で十分？
- 強い懸念から規制論が盛ん
 - Hinton, Bengio, Sutskever, ...



懐疑派 = 推進派

- 生成AI/DLでは不十分
 - Lecun, ...
 - LLMは100%記憶で説明可能---F. Chollet
 - DLは大規模な最近傍探索---P. Domingos



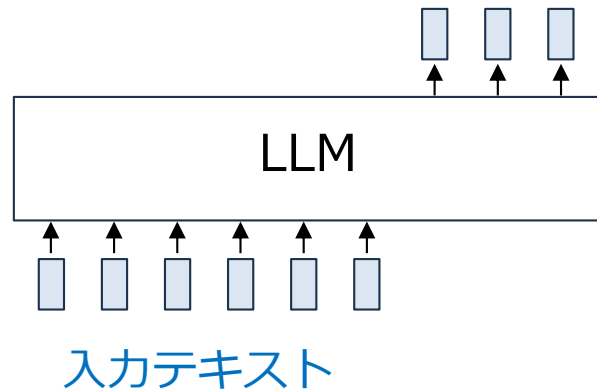
マルチモーダルAI

～言語の世界の外へ

マルチモーダルAI

言語/LLM

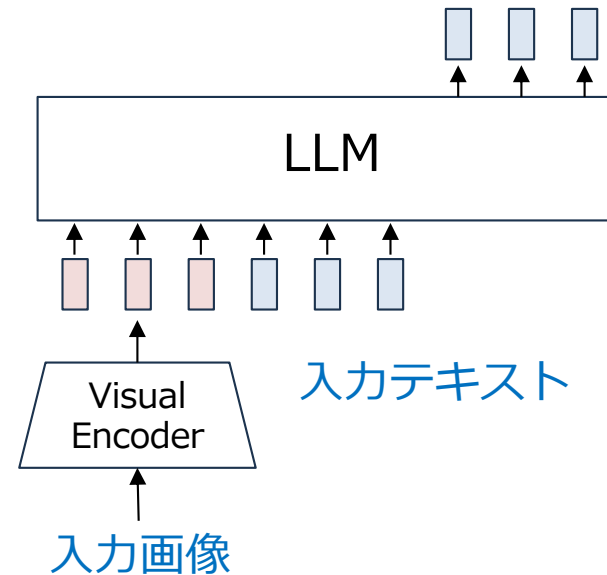
- 生成と推論が一体化



- 大規模学習可能

マルチモーダル/MLLM

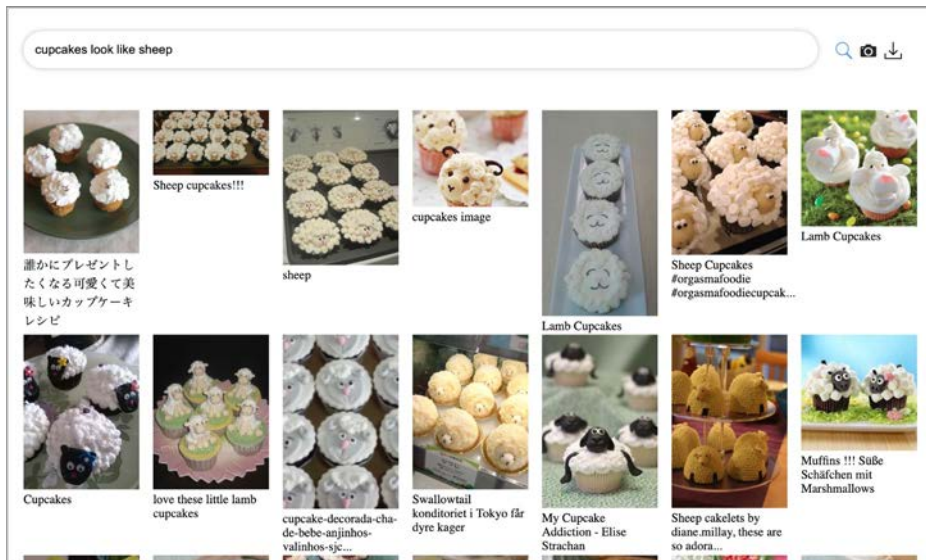
- 生成と推論は別



- 学習データ = 高コスト
 - 現状：web上の画像&alt-text
 - デジタル化さえ困難

画像とテキストのペアデータ

- Webページの画像と、それに紐づけられた“alt-text”のペア
 - alt-text = 状況に応じて画像に代わって表示されるテキスト
- 人々の生活を写す「鏡」であり、**偏りもある**



[RAION400M]



Alt-text: A Pakistani worker helps to clear the debris from the Taj Mahal Hotel November 7, 2005 in Balakot, Pakistan.



Alt-text: Musician Justin Timberlake performs at the 2017 Pilgrimage Music & Cultural Festival on September 23, 2017 in Franklin, Tennessee.

[Sharma+2018]

我々の研究：マルチモーダルAIの実世界応用

インフラ構造物画像診断

- 点検初心者に熟練者のスキルを [Kunlamai+2023]



Question

What is the damage in the image?

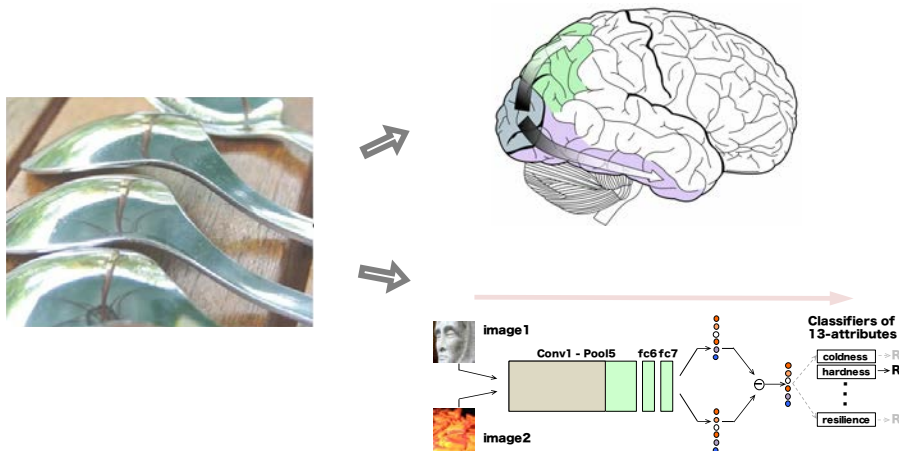
What is the member in the image ?

Are there any Corrosion in the bearing?

What kind of damage has occurred to the bearing?

質感画像認識・生成

- 人と同じように質感を認識するAI



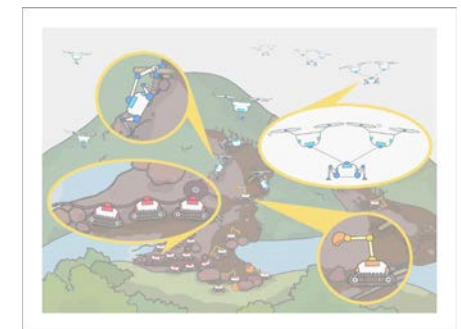
運転危険予測

- 危険予測 = 「かもしれない」運転 [Korawat+2023]



災害現場のリスク評価

- 斜面崩壊現場のリスク評価 [Areerob+2023]



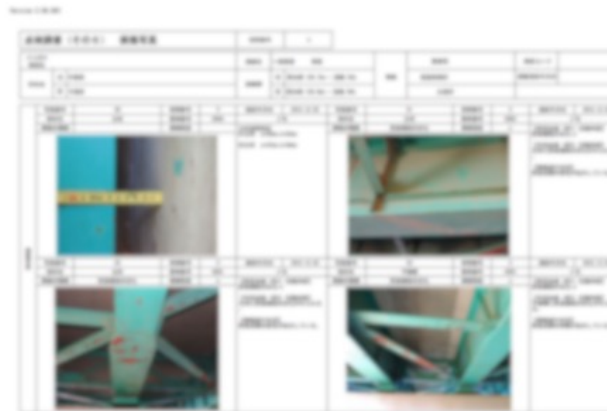
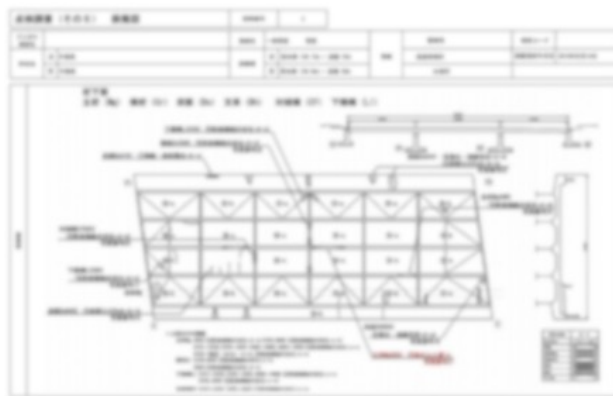
内閣府ムーンショット目標③
AIロボット永谷プロジェクト



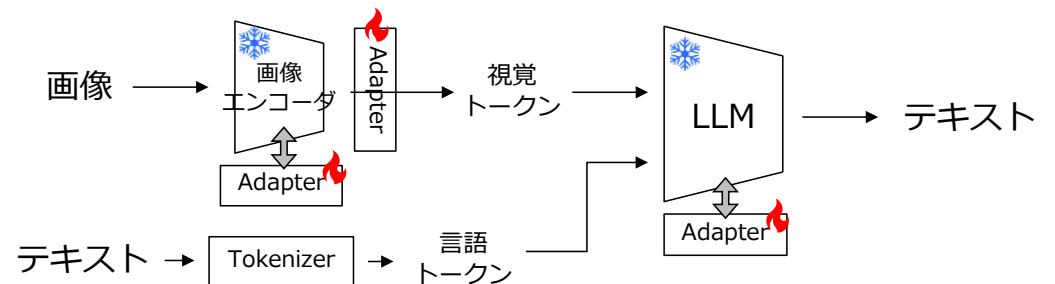
Disaster Type: "Landslides"
Cause: "Human development that has altered the terrain's natural drainage and caused soil erosion from surrounding slopes."
Observations: // 5 items
0: "The picture shows a mountainous region where numerous small landslides have occurred"
1: "The landslides appear to have originated from the ridge lines of the valley systems"
2: "The soil from the landslides has flowed into the low-lying areas and created temporary lakes"
3: "It is likely that the area experienced heavy rainfall at the time the picture was taken"
4: "The presence of trees suggests that the slopes have a relatively gentle incline"
Future Risk: "The area is likely prone to similar landslides in the future, particularly during heavy rainfall."

橋梁点検応用：我々のアプローチ

- AIに橋梁点検を「学ばせる」には？
- 既存の「点検レポート」の利用 [Kunlamai+2023]
 - マルチモーダルデータの事前学習



– 将来的には？



橋梁点検の実行例

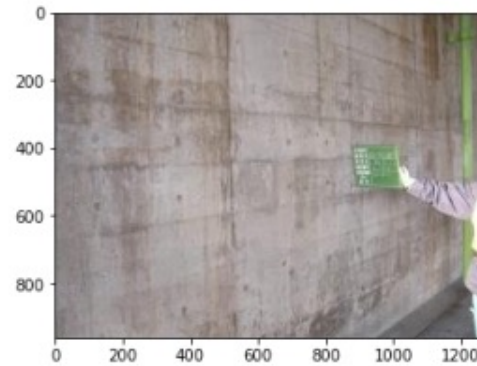
Kunlamai...Okatani, Improving visual question answering for bridge inspection by pre-training..., Computer-aided Civil and Infrastructure Engineering 2023



what is the member in the image?

Ground truth Bearing

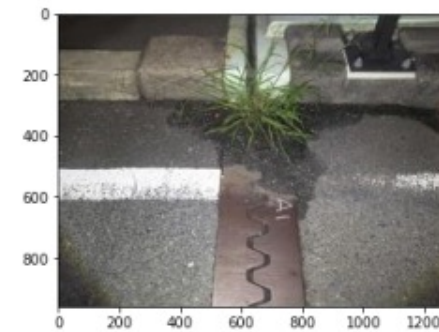
Pred answer **Bearing**



what is the member that has cracking?

Ground truth Vertical wall

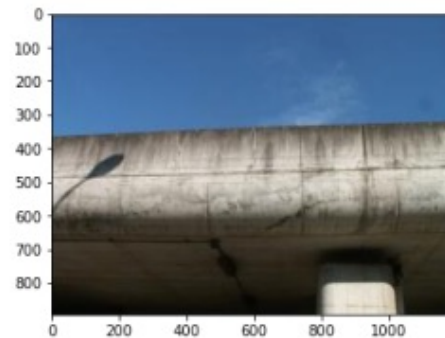
Pred answer **Vertical wall**



what is the member that has clogging?

Ground truth Expansion joint

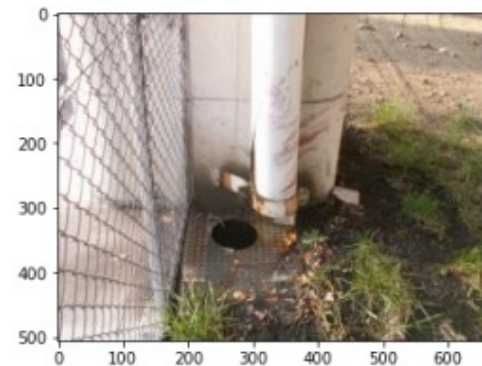
Pred answer **Expansion joint**



what is the member in the image?

Ground truth Protective fence

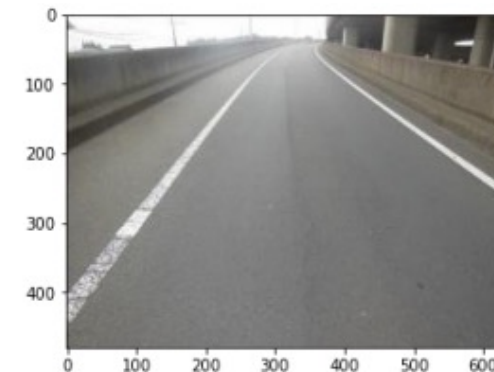
Pred answer **Protective fence**



what is the member in the image?

Ground truth Drainage pipe

Pred answer **Drainage pipe**



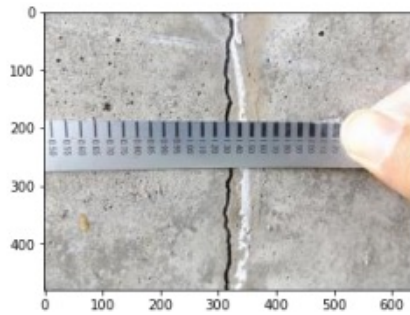
what is the member in the image?

Ground truth Pavement

Pred answer **Pavement**

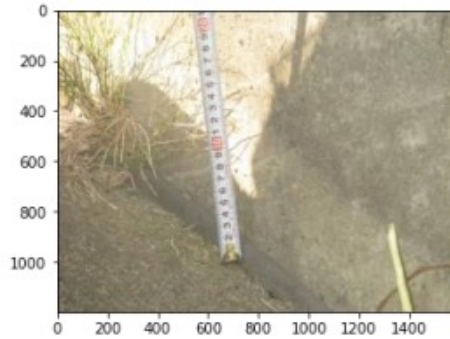
橋梁点検の実用例

Kunlamai...Okatani, Improving visual question answering for bridge inspection by pre-training..., Computer-aided Civil and Infrastructure Engineering 2023



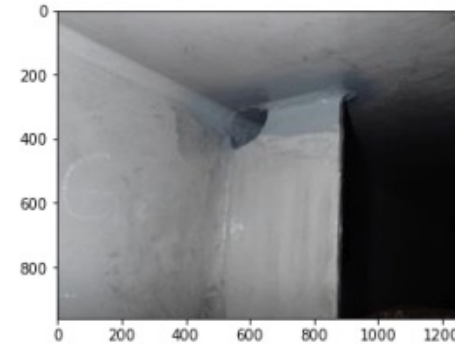
what is the damage in the image?

Ground truth Cracking
 Pred answer **Cracking**



what is the damage in the image?

Ground truth Sinking / displacement / slanting
 Pred answer **Sinking / displacement / slanting**



what kind of damage has occurred to the main girder?

Ground truth Fissure
 Pred answer **Fissure**



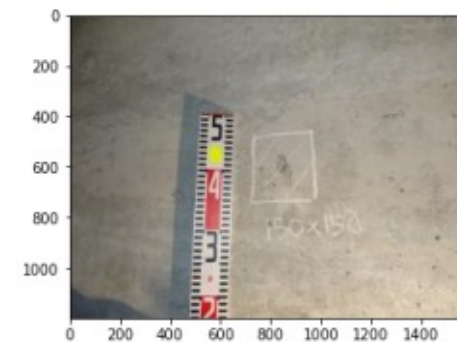
what kind of damage has occurred to the drain?

Ground truth Clogging
 Pred answer **Clogging**



what is the damage in the image?

Ground truth Leaking
 Pred answer **Leaking**



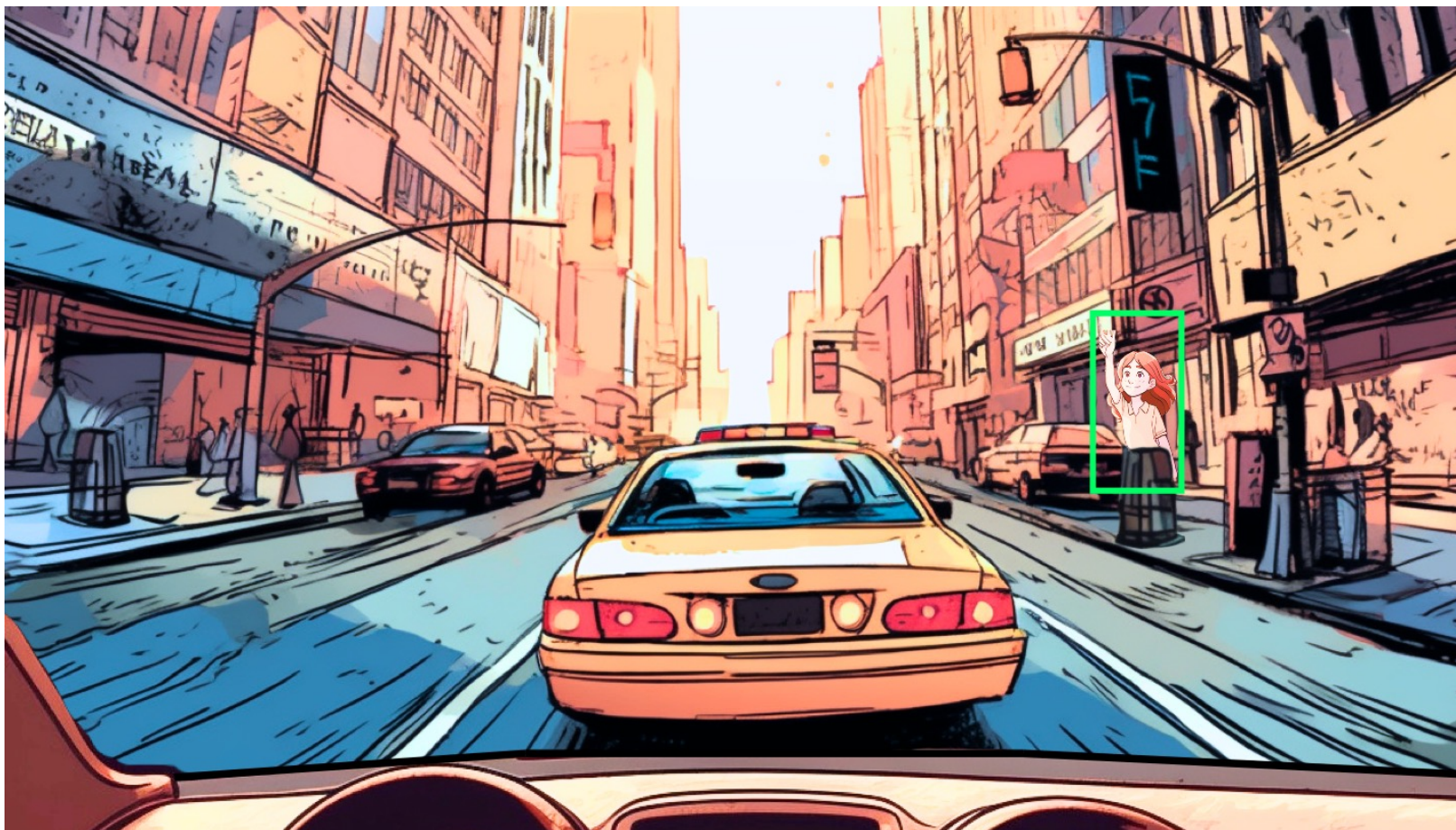
what kind of damage has occurred to the main girder?

Ground truth Flaking
 Pred answer **Flaking**

運転危険予測

Korawat+, Visual Abductive Reasoning Meets Driving Hazard Prediction, arXiv2023

- 視覚情報から近い将来起こり得る危険を予測（仮説推論）



学習データ生成 (Driving Hazard Prediction and Reasoning Dataset)

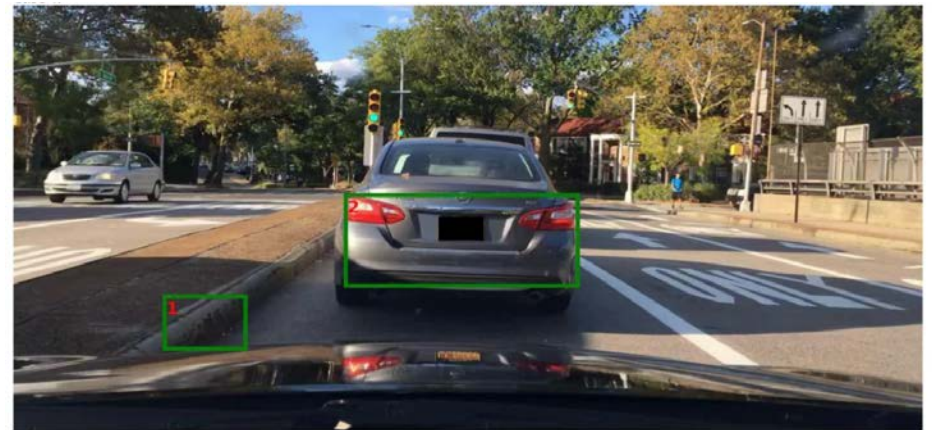
Korawat+, Visual Abductive Reasoning Meets Driving Hazard Prediction, arXiv2023

- 現実の事故映像を大量に得るのは困難 ; あまり有用でない
- 車載カメラ画像に対し, 潜在的なリスクをラベル付け



BDD100K [Yu+2020]

クラウド
ソーシング
サービス



Your car is at 15 km/h

Entity Description

TRY

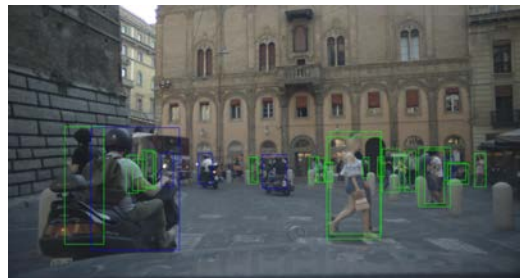
(Entity #1)
The footpath wall on the front left (Entity #2)
The grey car in the front

Accident Rationale

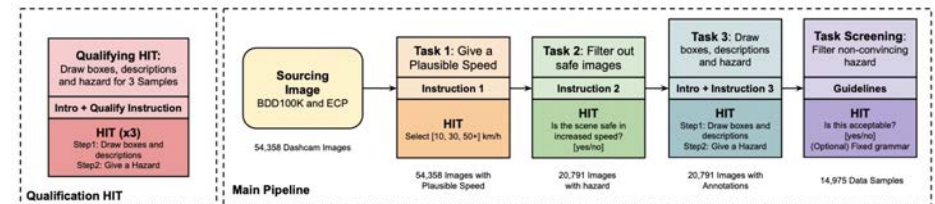
My car is heading toward Entity #1. At a given speed, I would crash into the side of Entity #1 or Entity #2 as I could not stop on time.

Explanation

The rationale must use the "(Entity #) word" instead of object noun. The description is about an entity and may have positional reference to our car. In this case, the rationale started with an explanation of an environment such as a footpath wall, speed, and the front car (Entity #2). Then, it explains how the accident would occur by including how the car crash would occur, such as we could hit them in the back or we could not stop on time.



ECP(EuroCity Persons) [Braun+2019]



学習データ生成 (Driving Hazard Prediction and Reasoning Dataset)



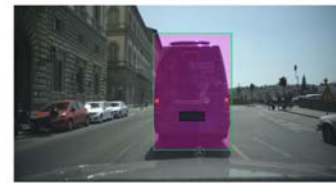
Hazard Type: Indirect Hazard **Speed:** 15 km/h
Scenarios: Narrow turn risk
Entity Description
 Entity #1: The black car in the front right
Hazard Explanation
 Entity #1 may turn too narrowly into my path. At a given speed, it is hard for me to stop on time.

(a)



Hazard Type: Indirect Hazard **Speed:** 45 km/h
Scenarios: Misjudging oncoming traffic
Entity Description
 Entity #1: Car going opposite direction
Hazard Explanation
 Making my left side, miscalculate Entity #1 speed going forward, due to this it hits my car.

(b)



Hazard Type: Direct Hazard **Speed:** 15 km/h
Scenarios: Sudden stop
Entity Description
 Entity #1: The van is showing its brake lights
Hazard Explanation
 Entity #1 may suddenly stop, even at low speed, our car might not be able to stop in time and hit Entity #1.

(a)



Hazard Type: Direct Hazard **Speed:** 75+ km/h
Scenarios: Side lane risk
Entity Description
 Entity #1: Cyclist in front on right side of road
Hazard Explanation
 Entity #1 decides to ride in the center of the lane, does not check prior to entering my lane, and due to this, my car is too fast to stop in time and clips Entity #1.

(b)



Hazard Type: Indirect Hazard **Speed:** 45 km/h
Scenarios: Unexpected lane change
Entity Description
 Entity #1: The yellow car at left side
Hazard Explanation
 Entity #1 seems entering my side of lane from left, at the given speed of 45 km/h, I may collide with Entity #1 from side as Entity #1 almost touching my car.

(c)



Hazard Type: Indirect Hazard **Speed:** 15 km/h
Scenarios: Unexpected pull-out
Entity Description
 Entity #1: Silver car to my right side
Hazard Explanation
 Entity #1 pulls out, miscalculate my distance to their car, due to this, won't stop in time and hits Entity #1.

(d)



Hazard Type: Direct Hazard **Speed:** 45 km/h
Scenarios: Lane drifting
Entity Description
 Entity #1: Raised curb to my left side
Hazard Explanation
 Driving at given speed, my car is too far to the left thus can't avoid hitting Entity #1.

(a)



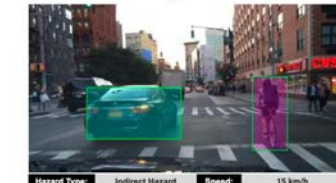
Hazard Type: Direct Hazard **Speed:** 75+ km/h
Scenarios: Speed Turning Risk
Entity Description
 Entity #1: Snow mound on right side shoulder lane
Hazard Explanation
 Driving at my speed, I turn my car too tight, and my car veers right, thus running over Entity #1.

(b)



Hazard Type: Indirect Hazard **Speed:** 15 km/h
Scenarios: Chain reaction hazard - unexpected lane change
Entity Description
 Entity #1: The black vehicle immediately in front of my car
 Entity #2: The yellow taxi in the lane to the right that is straddling the white lane lines
Hazard Explanation
 Entity #2 may encroach on the lane of Entity #1, causing the latter to slow down abruptly. This may cause my car to hit Entity #1.

(e)



Hazard Type: Indirect Hazard **Speed:** 15 km/h
Scenarios: Blocked maneuver collision
Entity Description
 Entity #1: The black car in the left side
 Entity #2: The person riding bicycle in the right side
Hazard Explanation
 Entity #2 tries to move right but stops as Entity #1 blocks the way. We would hit Entity #2 at the back.

(f)



Hazard Type: Direct Hazard **Speed:** 45 km/h
Scenarios: Low Visibility Risk
Entity Description
 Entity #1: Pedestrian wearing dark clothing ahead
Hazard Explanation
 Driving forward, due to Entity #1 on road, don't see them and due to speed, won't stop in time and hit Entity #1

(a)



Hazard Type: Direct Hazard **Speed:** 45 km/h
Scenarios: Late Decision Lane Change
Entity Description
 Entity #1: Divider ahead with speed limit post sign
Hazard Explanation
 At my given speed, I decided to go left. My car rides to the left at a close distance, and due to this, it clips Entity #1.

(b)



Hazard Type: Indirect Hazard **Speed:** 45 km/h
Scenarios: Chain reaction hazard - unexpected pull-out
Entity Description
 Entity #1: Van moving ahead of us
 Entity #2: Van at right side of the road
Hazard Explanation
 Entity #2 might get into the lane which will make Entity #1 to apply brakes suddenly which might lead us to hit Entity #1 at this speed.

(a)



Hazard Type: Indirect Hazard **Speed:** 45 km/h
Scenarios: Unexpected pedestrian intrusion
Entity Description
 Entity #1: Person with black coat with hood on left side
 Entity #2: Narrow road ahead
Hazard Explanation
 Entity #1 steps into Entity #2, does not see my car coming from behind and steps in front of my car.

(b)



Hazard Type: Direct Hazard **Speed:** 15 km/h
Scenarios: Incomplete Turning
Entity Description
 Entity #1: Flatbed truck in front of my car
Hazard Explanation
 Entity #1 turns left, brakes, due to me being too close, my car hits back of Entity #1

(c)

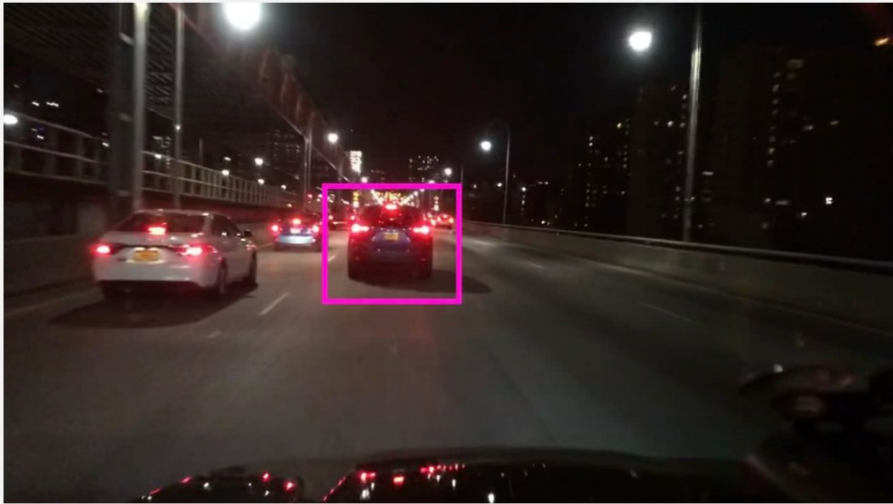


Hazard Type: Direct Hazard **Speed:** 15 km/h
Scenarios: Sun glare obstruction hazard
Entity Description
 Entity #1: Metal post ahead on right side
 Entity #2: Sun rays in my direction
Hazard Explanation
 Due to Entity #2, I don't see Entity #1 on road, thus my car hits Entity #1.

(d)

運転危険予測の例

Hazard Generation Example: Speeding & Braking



Annotation:

Entity #1 brakes, due to speed, my car won't stop in time and crash into the back of **Entity #1**

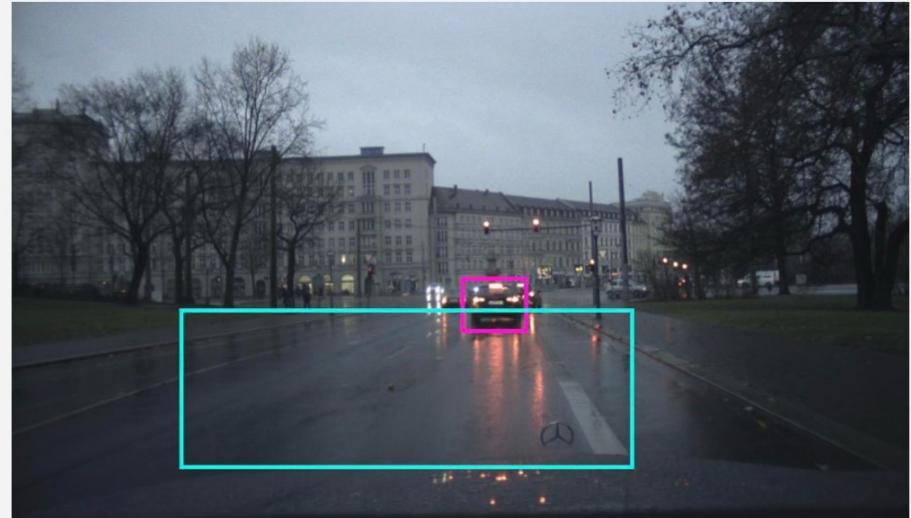
Our:

Entity #1 brakes, does not factor my speed, due to speed, can't stop in time and rear end **Entity #1**

GPT-4V:

Our car, traveling at a speed of 45 km/h, could potentially be involved in a rear-end collision with **Entity #1** if it were to suddenly brake and we did not have sufficient time to react and stop.

Hazard Generation Example: Speeding & Braking



Annotation:

Entity #1 brakes, due to my speed and **Entity #2**, my car won't stop in time and rear ends **Entity #1**

Our:

Entity #1 brakes, due to speed and **Entity #2**, my car hits **Entity #1**

GPT-4V:

In the hypothetical scenario, traveling at a speed of 75+ km/h in wet road conditions, our car could potentially lose traction and control, leading to an accident with **Entity #1** (Magenta box: the car directly ahead). The likelihood of a collision increases if **Entity #1** were to make a sudden stop and our car's stopping distance were insufficient due to the high speed and slippery road surface.

運転危険予測の例

Hazard Generation Example: Merging Maneuver



Annotation:

Entity #1 pulls out from parking spot, does not check mirror or factor my approaching speed, due to this, I can't stop in time and hits **Entity #1**

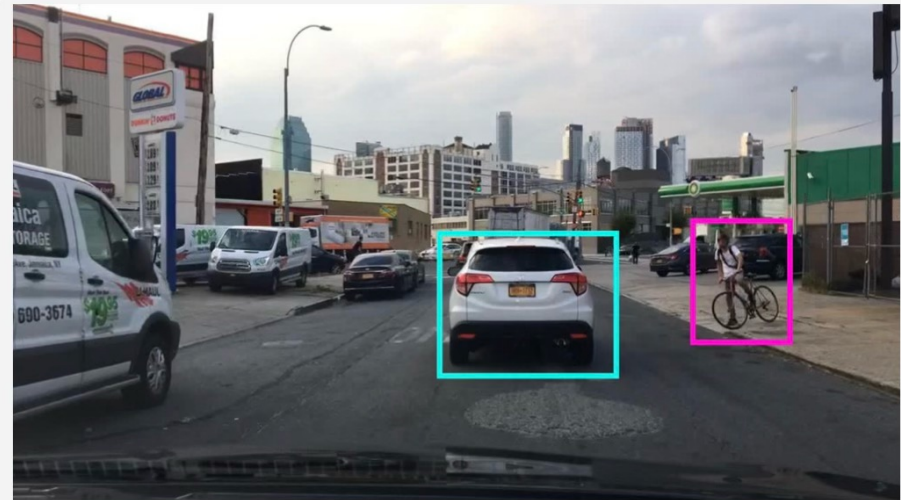
Our:

Entity #1 pulls out from parked position, does not check for oncoming traffic, due to this will hit **Entity #1**

GPT-4V:

As our car was traveling at 45 km/h, **Entity #1**, highlighted by the magenta box, made a sudden stop to avoid an obstacle not visible in the image, causing our car to react too late and resulting in a rear-end collision with **Entity #1**.

Hazard Generation Example: Pedestrian



Annotation:

Entity #1 decides to go behind **Entity #2** to cross the street, misjudges my speed, won't be able to stop in time and hits **Entity #1**

Our:

Entity #1 may try to go around **Entity #2** by coming into my lane. I may not be able to stop in time and hit **Entity #1**

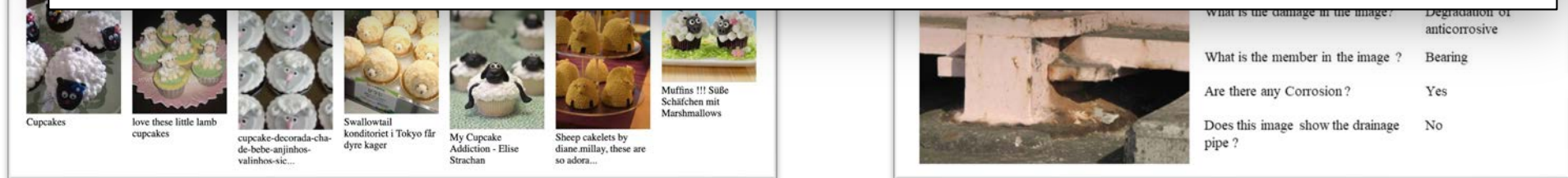
GPT-4V:

As our car approached the intersection at 45 km/h, **Entity #2** (Cyan box) unexpectedly changed lanes without signaling, causing our driver to swerve to avoid a collision. This sudden maneuver led our car into the path of **Entity #1** (Magenta box), a cyclist crossing the street, resulting in an accident.

課題

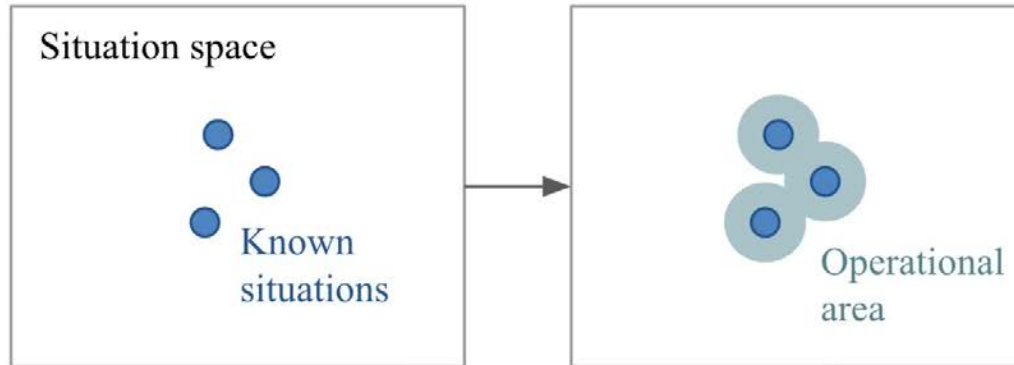
- 「専門家の知識」を
 - i) どうやってデジタルデータにし
 - ii) どうやってAIに学習させるか？

人間の仕事は、高度なものほど
言語化が難しい



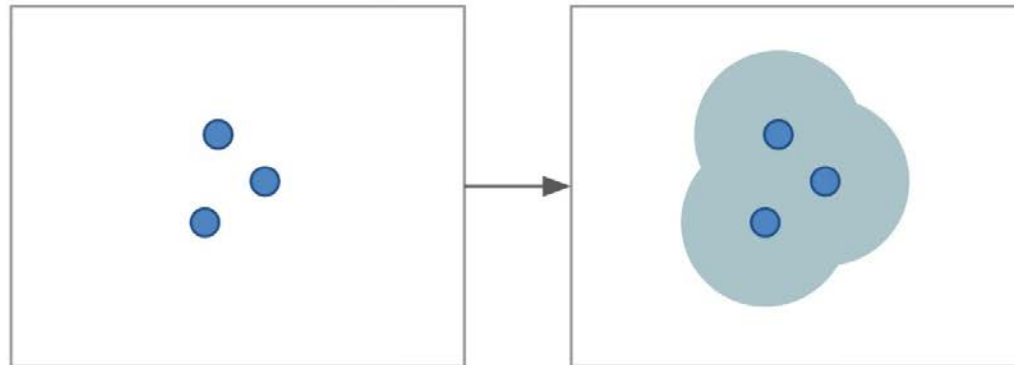
学習する機械

深層学習 … 「5を聞いて10を知る」



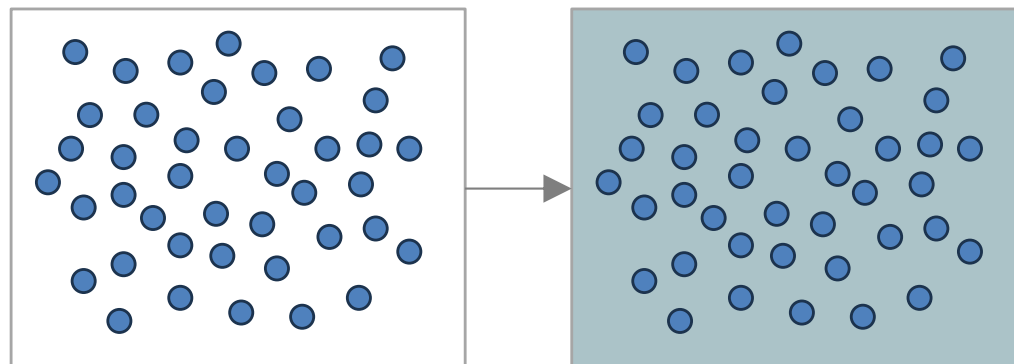
Lower-intelligence system:
lower information
conversion ratio

人間 … 「1を聞いて10を知る」



Higher-intelligence system:
higher information
conversion ratio

[Chollet, arXiv2019]



LLMからAGIへ

… 「1万を聞いて全部を知る」

**Situation spaceの広さ
vs. サンプルングコスト**

→ 言語空間なら可能かも？

まとめ：将来予想(5~10年)

- AGIの実現可能性：<0.1%, 5%, 10%, それ以上
 - ポイント：今の生成AIの延長線上にあるか？

比較的
サイバー・デジタル

- **言語だけで完結する仕事**
 - 技術：LLM
 - 成功の鍵：学習規模の更なる拡大は可能か？ + それだけで良いか？
 - 例：エクセルマクロ自動生成, プログラミング, 数学, ...
- **PC上で行える仕事すべてを代替**
 - 技術：マルチモーダルAI
 - 成功の鍵：人間レベルの文書画像理解は可能か？
 - 例：宿泊地と日程だけからホテルを予約できる
 - 例：食品の栄養成分表示から, 法的基準適合性を判断できる
- **専門家に代わって実世界の問題を解決するマルチモーダルAI**
 - 技術：マルチモーダルAI
 - 成功の鍵：専門知のデジタル化とAI学習は可能か？
- **自律ロボット**
 - 技術：???
 - 成功の鍵：何が不足しているかもわからない
 - 例：自動運転, 家事代行ロボット

難易度：高
物理空間