



地震工学委員会 第3回研究会  
2025年11月26日(水)

# 震後の緊急対応における **人間技術者**の役割分担を再考する —UAV・点群・LLMの活用可能性

国立研究開発法人 土木研究所  
構造物メンテナンス研究センター

小林 巧

# はじめに

大規模地震発生後には、以下の制約が顕在化

- 1) 時間的制約 (たとえば、人命救助の壁: 72時間)
- 2) 機材や資材の制約 (たとえば、足場がない条件での点検、復旧資材の不足)
- 3) 人材の制約 (たとえば、行政職員や点検者が十分に参集できない状況)
- 4) ライフラインやインフラの途絶 (たとえば、電気・通信・水等の途絶)

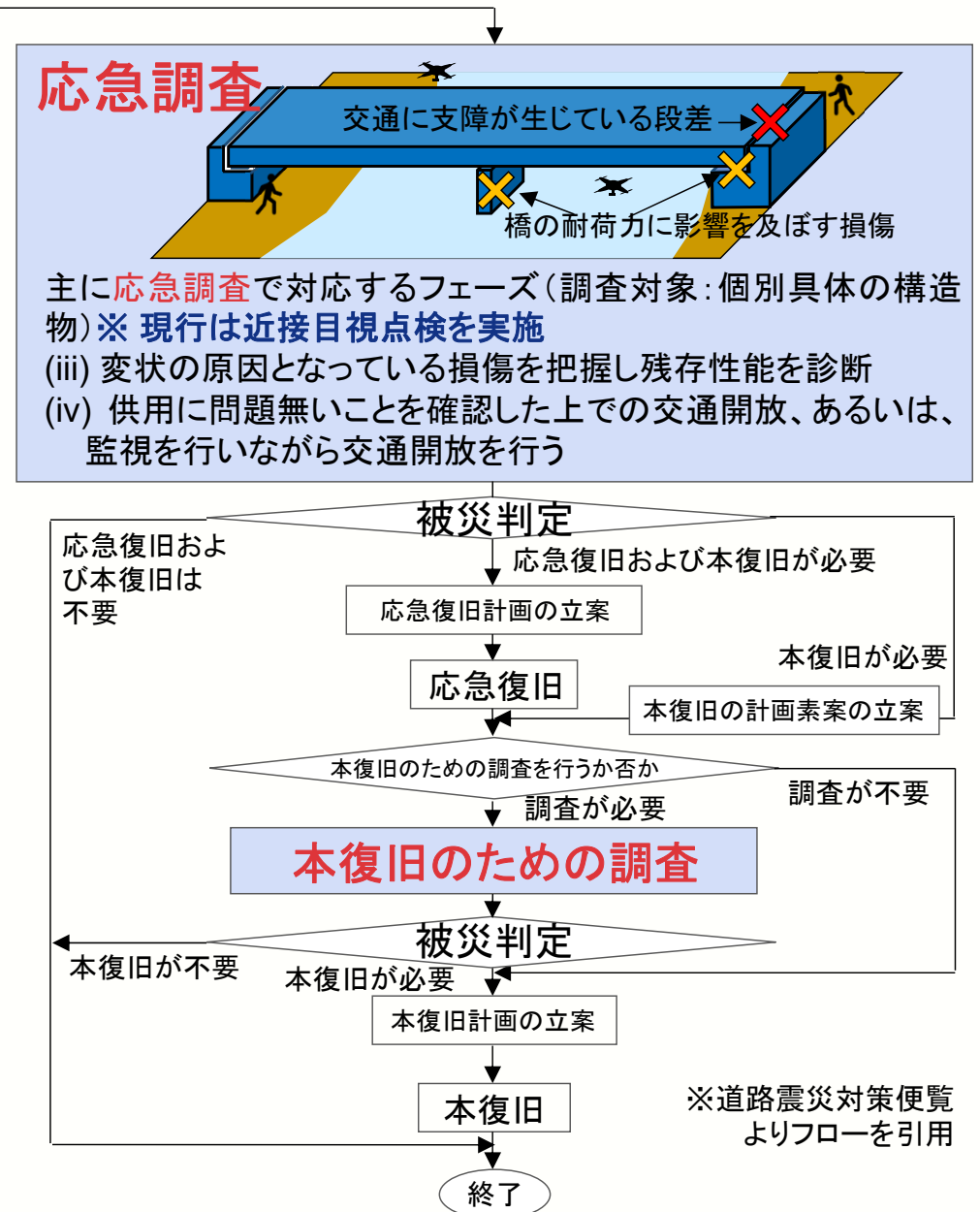
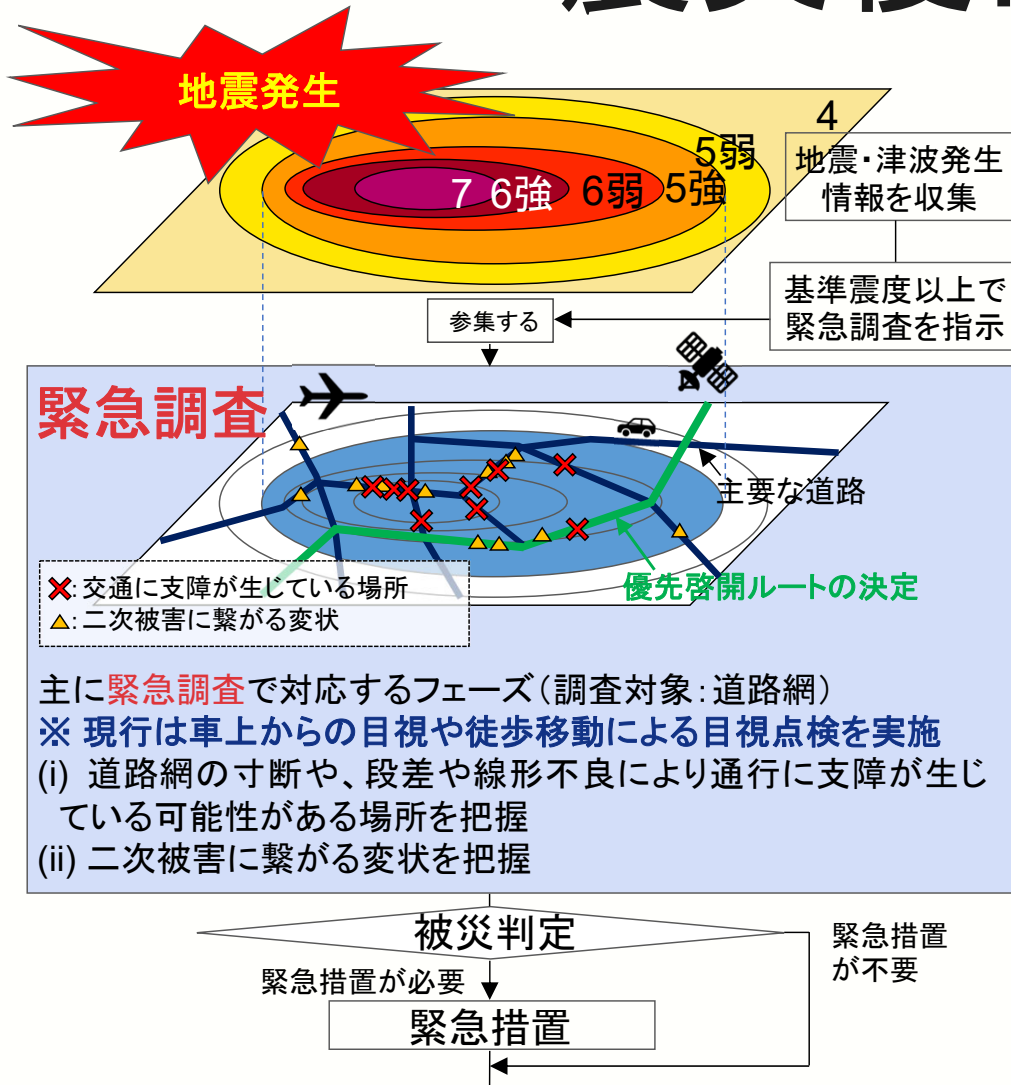
震後には、これら制約下で**迅速な交通開放と  
確実な措置を両立**することが社会的に要請。

実際に2016年の熊本地震では震度6弱以上が推定された地域に約15,700橋、  
2024年の能登半島地震では3,018橋の道路橋が存在  
人間技術者のみで迅速で確実な復旧を達成することには限界がある

⇒あらゆる手段(技術)を用いて、震災対応を効率化したい

# 震災復旧の流れ

※道路管理者によって  
多少の違いがある



# 緊急調査の困難さ：どこに損傷がある？





# 遠望目視点検はなかなか難しい



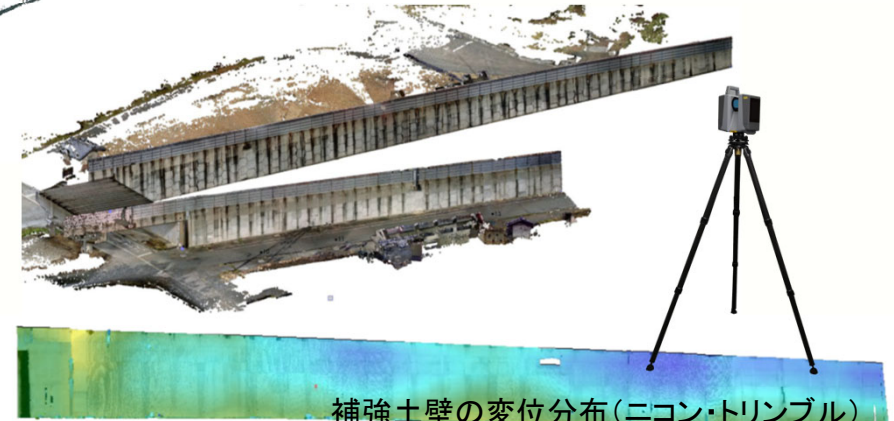


# 令和6年能登半島の震後調査における 点検技術の試行

北陸地方整備局や石川県等に調査の許可を頂き、「道路橋の震後点検の効率化・高度化に向けた新技術の利活用に関する共同研究(令和5～6年度、土研等)」の一環で実施



トラス上部構造の点群モデル





# 画像撮影技術の活用例

## 足場が無い状況での迅速な調査方法例



(II-1) 自撮り棒を橋面から降ろし、撮影



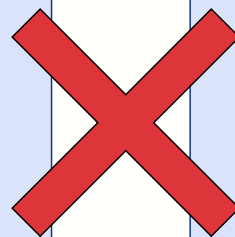
(II-2) 自撮り棒を地上から持ち上げ、撮影



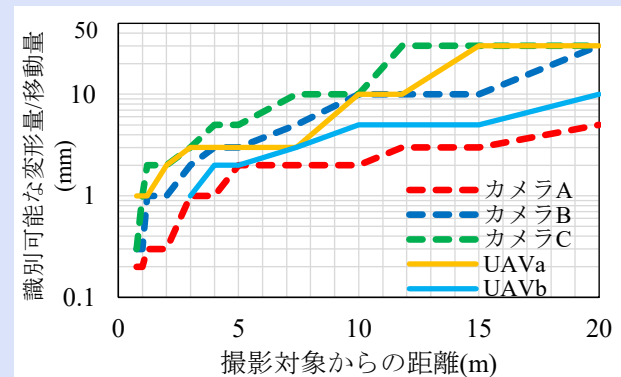
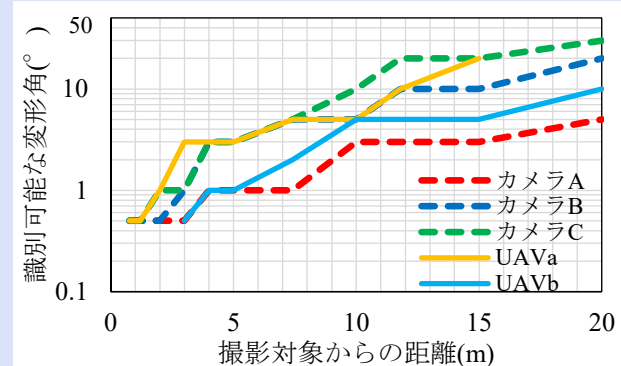
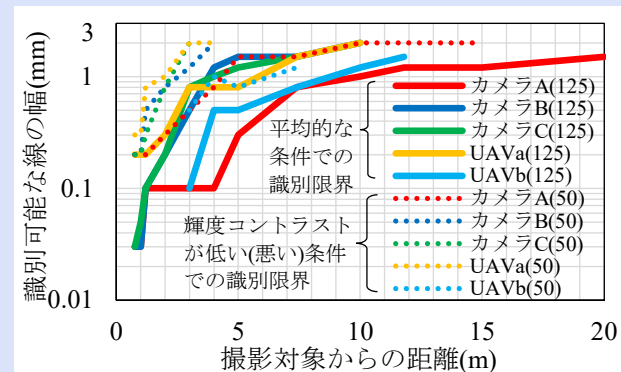
(III-1) 操縦者が橋面にいる状況でUAVを操縦



(III-2) 操縦者が地上にいる状況でUAVを操縦



## 事前の能力評価により調査の確実さを補完した例





# 画像撮影技術の活用により調査を効率化



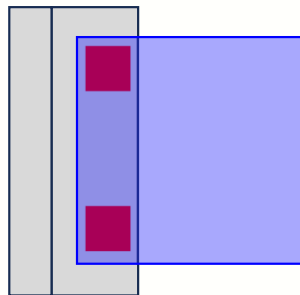
機動性を確保し調査の効率化に貢献  
一方、技術を用いた点検には限界も存在



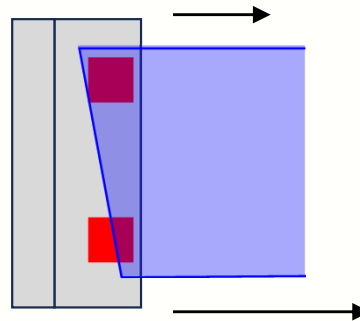
# 画像撮影や近接目視でわからない変状



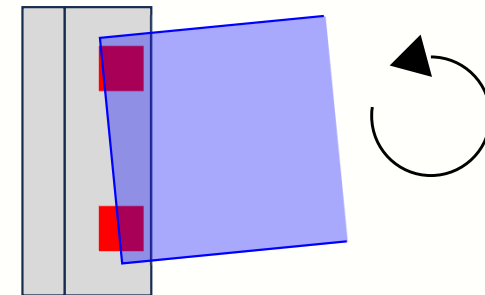
【凡例】 ■: 上部構造 ■: 支承 ■: 下部構造



a) 損傷前



b) 上部構造のゆがみにより  
上下流の支承部の移動  
量が異なる場合



c) 上部構造の回転により  
上下流の支承部の移  
動量が異なる場合

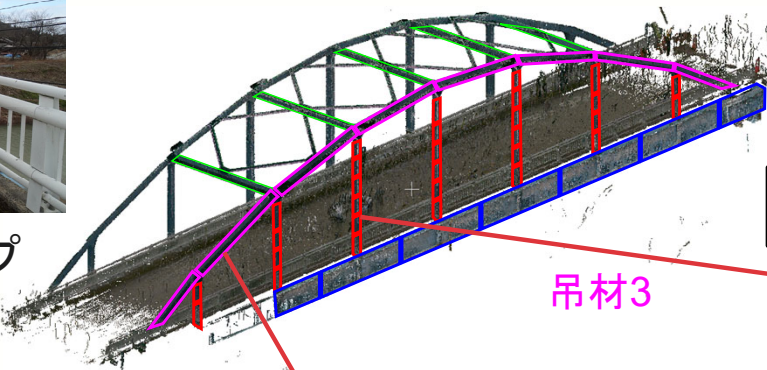
危惧される損傷モード(上部構造の移動に伴うねじれ)



# 点群計測技術の活用により微小変形を把握



ハンディタイプの  
のスカナー



吊材3

上弦材

上流側

下流側

上流側

A1

A2

吊材2

吊材3

吊材4

吊材5

吊材6

吊材7

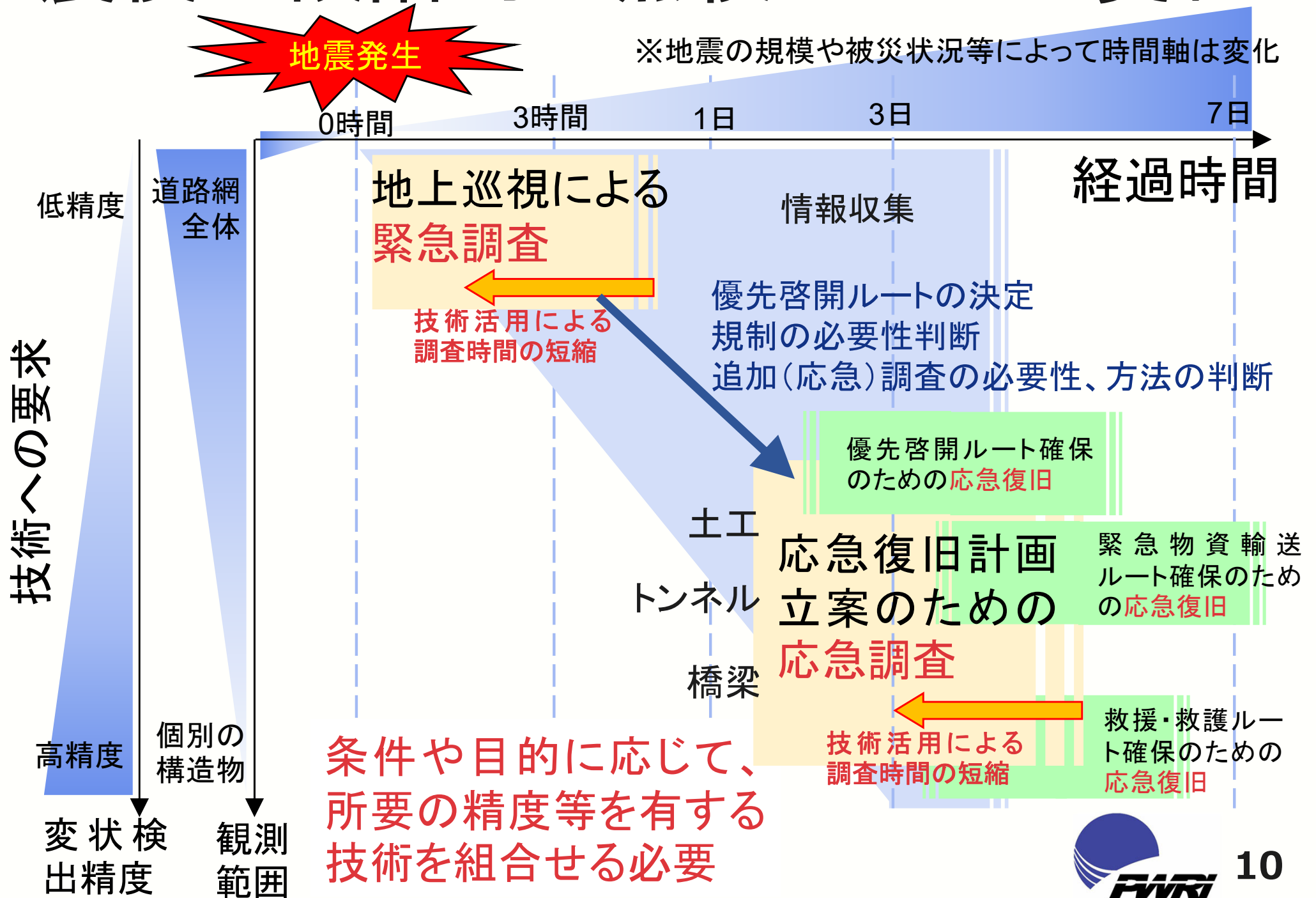
下流側

分析を工夫すれば、人間技術者には  
認識できない微小変形もわかる





# 震後は段階的に点検ニーズが変化



# 小まとめ (技術で技術者の何を補完するか?)

条件や目的に応じて、所要の精度等を有する技術を組合せる必要

目視による変状発見



震災対策便覧より

広範囲の変状

内部損傷等



微小変形

異音検知

インテリジェンス?

- ・変状や点検結果の解釈
- ・どこに損傷があるかを推論
- ・措置の必要性を判断
- ・状況を他者に伝達する
- ・想定外に対する現場判断
- ・判断の責任を負う 等

触診・打音

- ・ボルト破断
- ・はらみ出し・浮き

機動力



震災対策便覧より



UAV



衛星: Sentinel-1

[https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-1/Europe\\_lofts\\_first\\_Copernicus\\_environmental\\_satellite](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Europe_lofts_first_Copernicus_environmental_satellite)



国立研究開発法人 土木研究所



# 小まとめ (技術で技術者の何を補完するか?)

条件や目的に応じて、所要の精度等を有する技術を組合せる必要

! ?

目視による変状発見

異音検知



震災対策便覧より

広範囲の変状

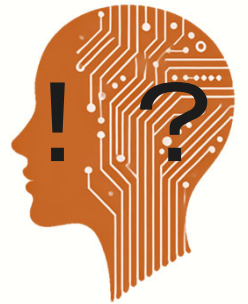
内部損傷等



微小変形

インテリジェンス?

- ・変状や点検結果の解釈
- ・どこに損傷があるかを推論
- ・措置の必要性を判断
- ・状況を他者に伝達する
- ・想定外に対する現場判断
- ・判断の責任を負う 等



触診・打音

- ・ボルト破断
- ・はらみ出し・浮き

機動力



震災対策便覧より



UAV



衛星: Sentinel-1

[https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-1/Europe\\_lofts\\_first\\_Copernicus\\_environmental\\_satellite](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Europe_lofts_first_Copernicus_environmental_satellite)



国立研究開発法人 土木研究所

# AI等の飛躍的進化

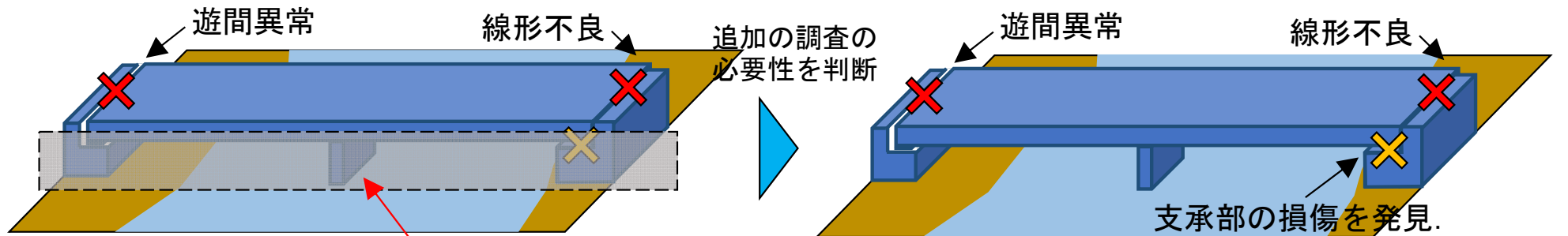
多くの業務がデジタル技術で**“技術的には”**代替・補完できる時代になりつつあるが、判断は依然として人間依存  
→倫理や説明責任等の観点では妥当な選択？

一方、大規模地震後の危機的状況下における、数千～数万の道路橋等に対する判断を少しでも技術で補助し、軽減したい。**人間技術者だけでは、正直厳しい。**

**何ができるばAIを安心・安全に使える？**  
**判断のブラックボックスを可視化することで、**  
**AIの信頼性も向上？学習も効率化？**



# そもそも緊急調査で熟練技術者は どのように判断を下すのか？



路面からの目視では確認できない範囲

（「遊間異常や線形不良がある場合は、支承部にも損傷が生じているはずだ」と帰納的に推論。観念で補間。）

支承部の損傷を根拠に、「支承部が損傷している場合は、上部構造を支持する能力に影響を及ぼす」ことを論拠に、「上部構造を支持する能力が不足している」ことを演繹的に論証。

(i) 路面からのスクリーニング的な目視点検（緊急調査）

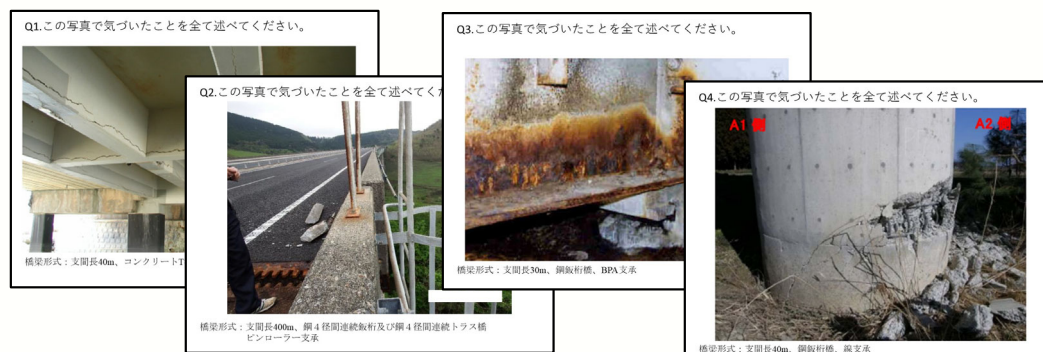
(ii) 追加の調査（応急調査）



研究の仮説：  
熟練技術者の推論過程は  
仮説演繹法に基づく論理モ  
デルで説明可能

# 仮説の検証手順

熟練技術者12名に半構造化インタビューを実施。



複数の画像を提示

## 半構造化インタビュー 想定質問：

Q1.あなたは今、点検をするために橋の現場に着きました。  
この写真で気づいたことを全て述べてください。

Q2. ○○を見つけ、あなたは思うか？

Q3.あなたの見つけた変状全ては何が原因で生じているか？

Q4.この写真から、この橋にはどのような性能(耐荷力・耐久性・その他の性能)の低下が疑われるか？

→Q5. ○○性能の低下を疑った理由は何か？

Q6.もし疑いを持った時、(緊急時には)どのような措置を取ることが望ましいか？

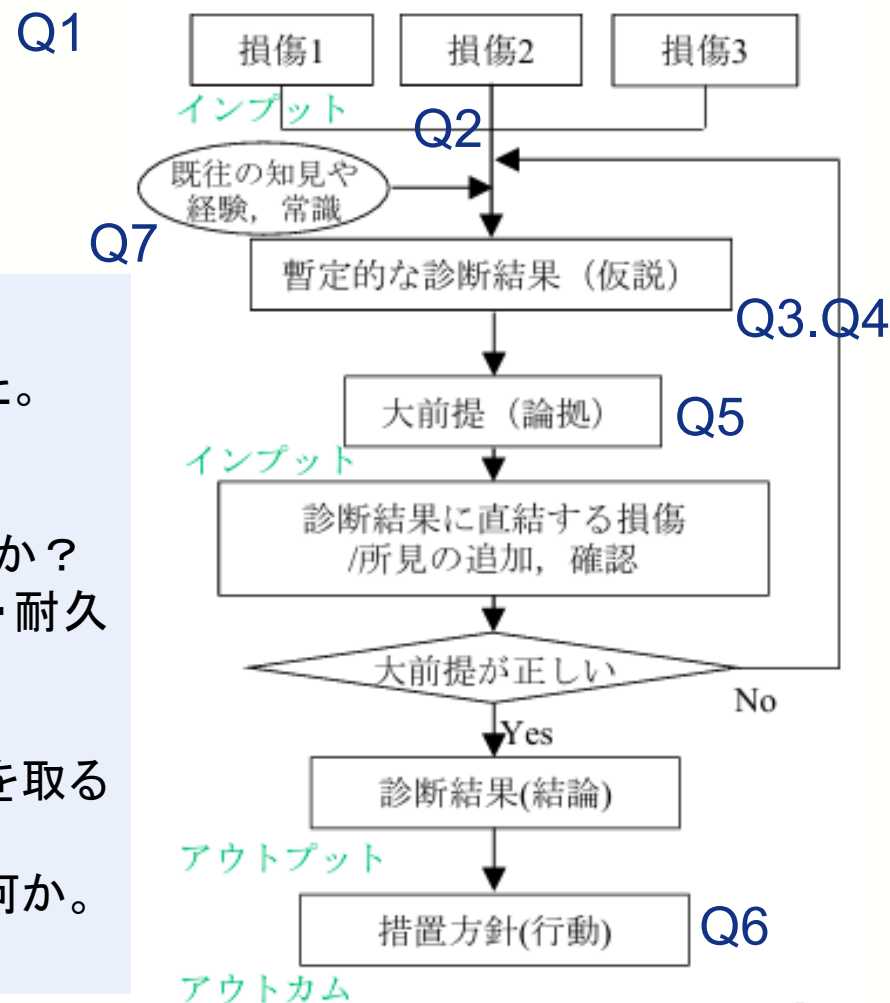
Q7.答えた見立てについて、その根拠となる考え方は何か。  
(知識、経験、常識、学習等)

認識の枠組みを押し付けることを防ぐため、非構造化→半構造化→構造化の順番で質問を行う。

非構造化質問(対象も反応も規定しない):この写真で何が特徴的ですか？

半構造化質問(具体的な対象が規定され、反応は自由):  
○○という変状を見つけ、あなたはどう思いましたか？

構造化質問(対象と反応の両方を規定):○○という変状から、何が原因でこの変状が生じており、橋のどのような性能の低下が予測されますか？



【仮説演繹法を震後点検時の論理に適用した例】



# 論理的視点での代表回答の選出基準例

- 1) 論理学は「演繹」を扱い、正しい思考の形式・法則を研究する学問であり、  
論理学で示された論理体系を用いることでその論理の正しさが保証
- 2) 論理学では論理体系の正しさを保証するが結論の真偽は保証しない  
結論が真であるためには前提が真であることを示す必要

**根拠**→写真からわかる事実に関する発言

**論拠**→写真のみではわからない推測や知識等を含む発言

**結論**→橋や部材等の性能に言及する発言

**措置方針**→措置方針に関する発言

これらが有機的に  
接続した回答  
が論理的と想定

- 1) 回答内容や結論に間違いが無い(真)ことを確認
- 2) 1) を前提にした上で、以下の式で得点が高い回答を選出

$$P = \sum_{i=1}^n (A_i \times B_i \times C_i \times D_i)$$

$P$ : 一連の推論の得点

$A$ : 根拠の個数

$B$ : 根拠と意味がつながる論拠の個数

$C$ : 論拠と意味がつながる結論の個数

$D$ : 結論と意味がつながる措置方針の個数



根拠、論拠、結論、措置方針の回答数が多いほど、  
根拠—論拠—結論—措置方針が多く結びつくほど、得点が上昇

# 代表回答の例1

Q2.この写真で気づいたことを全て述べてください。



橋梁形式：支間長400m、鋼4径間連続鈑桁及び鋼4径間連続トラス橋  
ピンローラー支承

凡例

根拠→写真からわかる事実(下線部)

論拠→写真のみではわからない推測や知識等

結論→橋や部材等の性能に言及する発言

措置方針→措置方針に関する発言

仮説法で思考していることを定性的に検証

Q. あなたは今、点検を行うために橋の現場に着きました。この写真で気づいたことを全て述べてください。

A. 壁高欄の軸線がずれている。検査路があるため、手前がアバットである可能性がある。壁高欄の軸線のずれと添架物用のパイプの切断が生じているため、上部構造と下部構造の相対変位が生じた可能性がある。縁石が飛んでいるため、径間側が動いた可能性がある。伸縮装置の遊間異常もあるため、地震による損傷が生じている可能性がある。伸縮装置は橋軸直角方向と上下方向がずれているため、下部工の移動やピン・ローラー支承の破壊が生じている可能性がある。

Q. この写真から、この橋にはどのような性能(耐荷力・耐久性・その他の性能)の低下が疑われるか？

A. 走行性の低下が生じている可能性がある。ピン・ローラー支承が破壊しているのであれば、耐荷力が低下している可能性がある。

Q. もし疑いを持った時、(緊急時には)どのような措置を取ることが望ましいか？

A. 緊急的には段差を解消し、走行性を確保する必要があると考える。その後、詳細な調査を行う必要があると考える。

# 代表回答の例2

Q4. この写真で気づいたことを全て述べてください。



橋梁形式：支間長40m、鋼鈑桁橋、線支承

凡例

根拠→写真からわかる事実(下線部)

論拠→写真のみではわからない推測や知識等

結論→橋や部材等の性能に言及する発言

措置方針→措置方針に関する発言

仮説法で思考していることを定性的に検証

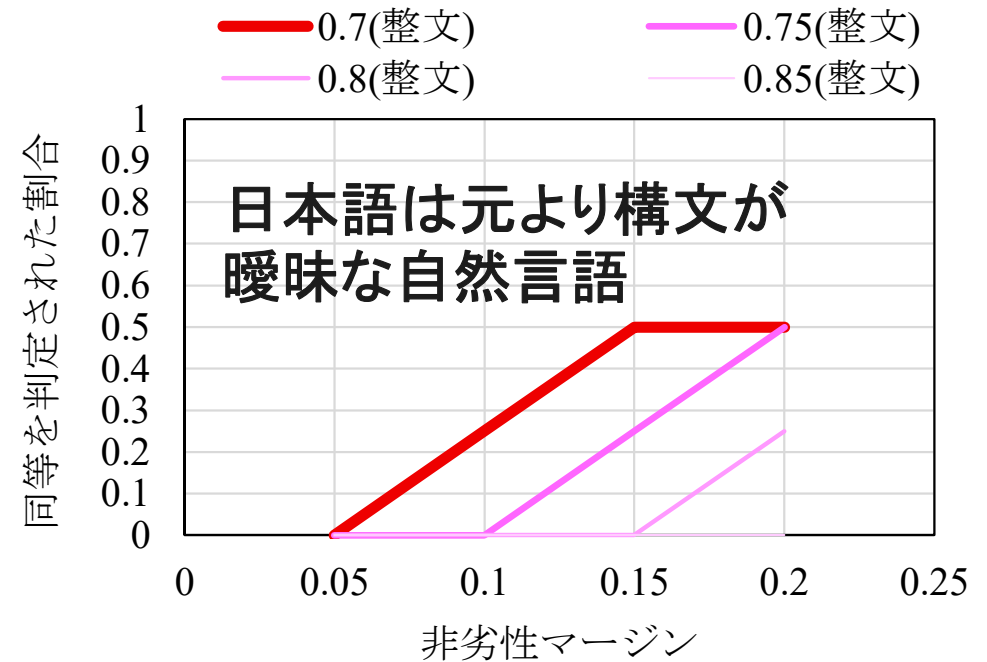
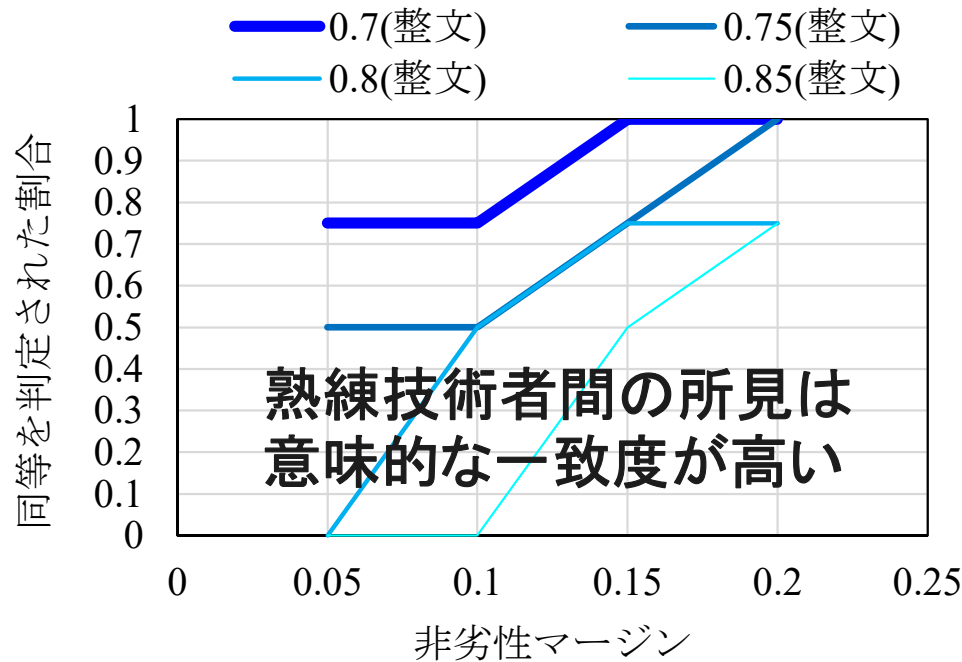
Q. あなたは今、点検を行うために橋の現場に着きました。この写真で気づいたことを全て述べてください。

A. 橋脚の基部に部分的な破壊が生じている。コンクリートの断面が破壊し、鉄筋も部分的にはらんでいる。ひび割れが斜め45度方向に入っている。それら変状の原因は地震の可能性がある。橋脚の引張サイドのコンクリートが破壊し、45度方向の斜めひび割れが生じているならば、有効断面がない可能性がある。これら変状から、鉛直方向の耐荷力はまだ残存している可能性がある。しかし、水平方向の耐荷力がほぼない状態である可能性があるので、車両を走行させることは危険である可能性がある。ゆえに、道路管理者に連絡を行い、通行止めの措置を判断頂いた方が良いと考える。



# 定量的にも検証 (代表回答とその他の回答の比較)

## ※スコアの同等性検定

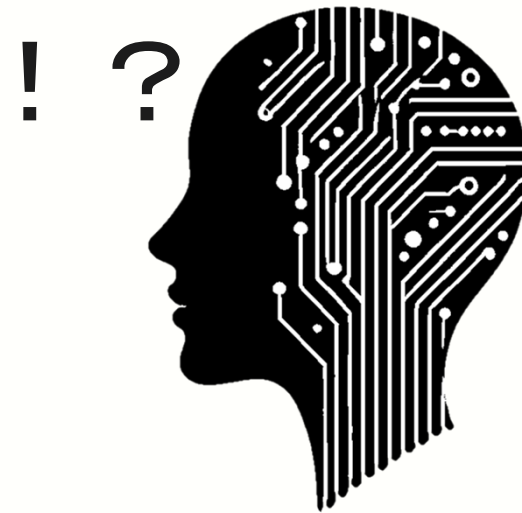
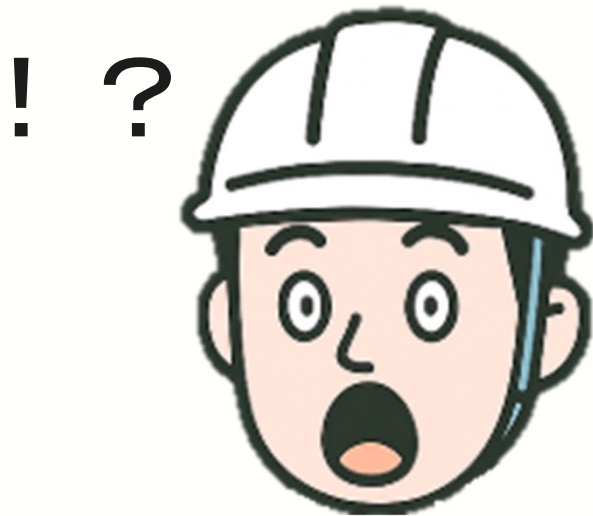


## ※Sentence-BERTによるスコア

## ※依存構造木によるスコア

少なくとも根拠・論拠・結論は多くの回答で類似した意味で存在。写真等にある事実(根拠)より、自らの知識(論拠)に基づき、結論を導くという、厳格な論理則に基づかないという意味での**“緩い”仮説法**は熟練技術者間で広く成立していることが示唆

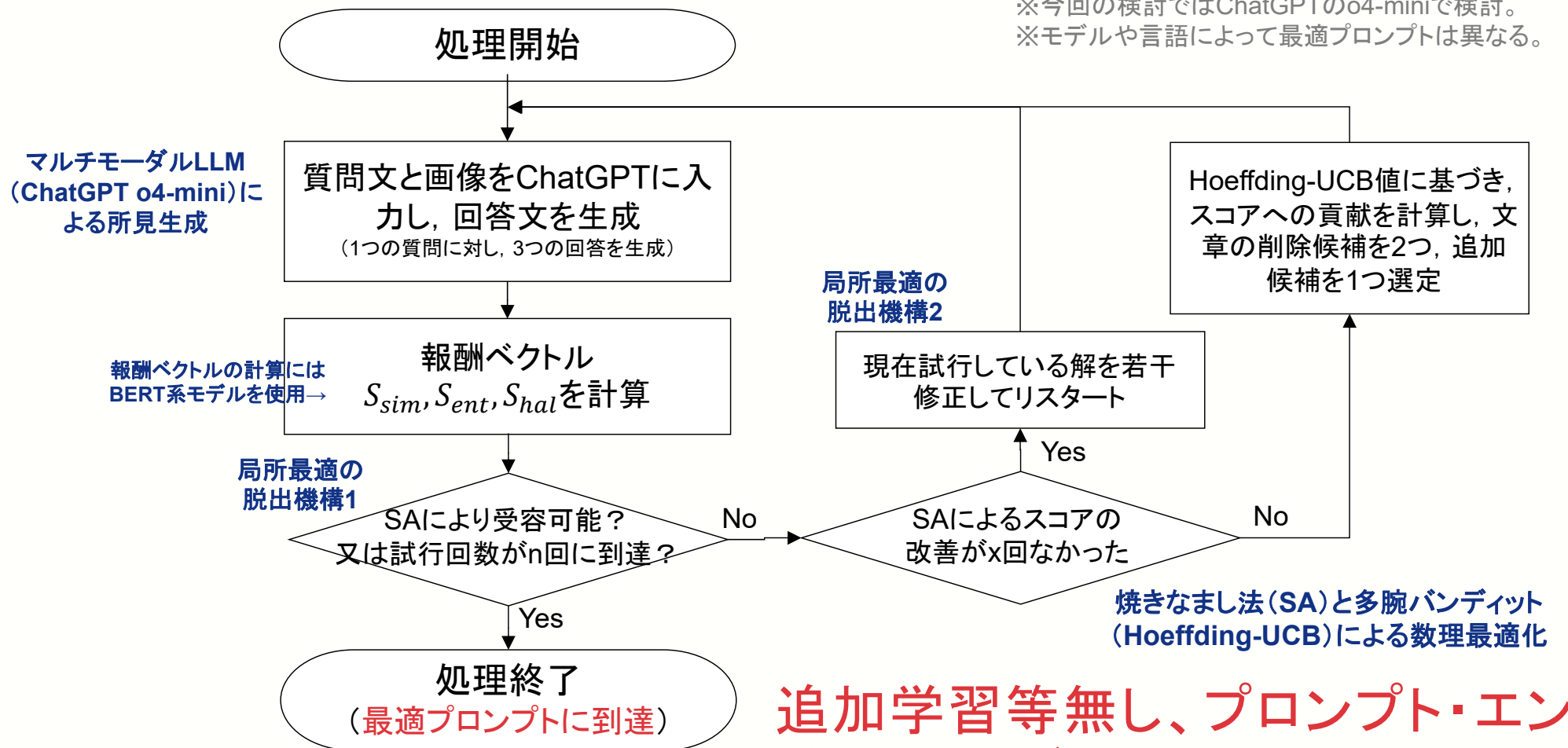
熟練技術者が仮説演繹法的に措置の必要性について推論していると仮定して、今のAIはどこまでそれを模倣できるか？



# 最適プロンプトの探索アルゴリズム

様々な観点で最も熟練技術者の代表回答に近く、かつ、最適プロンプトの探索回数が少なくなるように数理最適化

※今回の検討ではChatGPTのo4-miniで検討。  
※モデルや言語によって最適プロンプトは異なる。



追加学習等無し、プロンプト・エンジニアリングのみでの模倣を検討



# プロンプト(質問文)の最適化

複数の文章から構成される質問文を作成

**[Z]#Instructions:** Based on the hypothetico-deductive method, logically derive a course of action—both its necessity and details—from the damage observed in the photos, taking into account the performance of the road bridge (including load bearing capacity, durability, and other factors such as traffic safety, damage to third parties, drivability, recoverability, etc.), using your expertise.

**[A]#Setting the role:** You are an expert on road bridges.

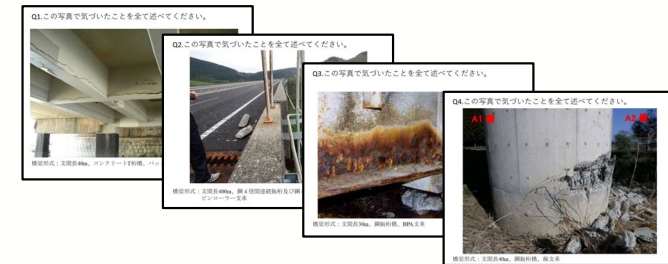
**[B1]#Setting the scene:** Assume a situation where an emergency inspection is conducted after an earthquake.

**[C]#Definitions of Terms:** Define the words used in the description of the hypothetico-deductive method as follows:

- 1) “Evidence” as facts that can be seen from the photo.
- 2) “Argument” is what is not in the photo but can be inferred from your knowledge, or what can be inferred by combining your knowledge with the facts known from the photo.
- 3) “Conclusion” is a statement about the performance of the road bridge that can be inferred by combining the evidence and arguments.
- 4) The “course of action” is a reasonable course of action that can be logically derived from the evidence and arguments, especially for the poor performance of the bridge in the conclusion.

→必ず行う質問Z

質問とともに画像を提示



必ず行う質問Zを補足する文章(A~F)

どのような質問の組合せを行えば熟練技術者の回答に近づくか？



# 最適化の方針

【Z】#Instructions: Based on the hypothetico-deductive method, logically derive a course of action—both its necessity and details—from the damage observed in the photos, taking into account the performance of the road bridge (including load bearing capacity, durability, and other factors such as traffic safety, damage to third parties, drivability, recoverability, etc.), using your expertise.

【A】#Setting the role: You are an expert on road bridges.

【B1】#Setting the scene: Assume a situation where an emergency inspection is conducted after an earthquake.

【C】#Definitions of Terms: Define the words used in the description of the hypothetico-deductive method as follows:

- 1) “Evidence” as facts that can be seen from the photo.
- 2) “Argument” is what is not in the photo but can be inferred from your knowledge, or what can be inferred by combining your knowledge with the facts known from the photo.
- 3) “Conclusion” is a statement about the performance of the road bridge that can be inferred by combining the evidence and arguments.
- 4) The “course of action” is a reasonable course of action that can be logically derived from the evidence and arguments, especially for the poor performance of the bridge in the conclusion.

- ✓ 様々な質問の組合せがある
  - 役割の定義（道路橋の専門家等）
  - 場面設定（地震後の点検等）
  - 推論ルール（根拠とは？論拠とは？）

以下の3つのスコアが最大化するように、最適なプロンプト（質問文の組合せ）を選択

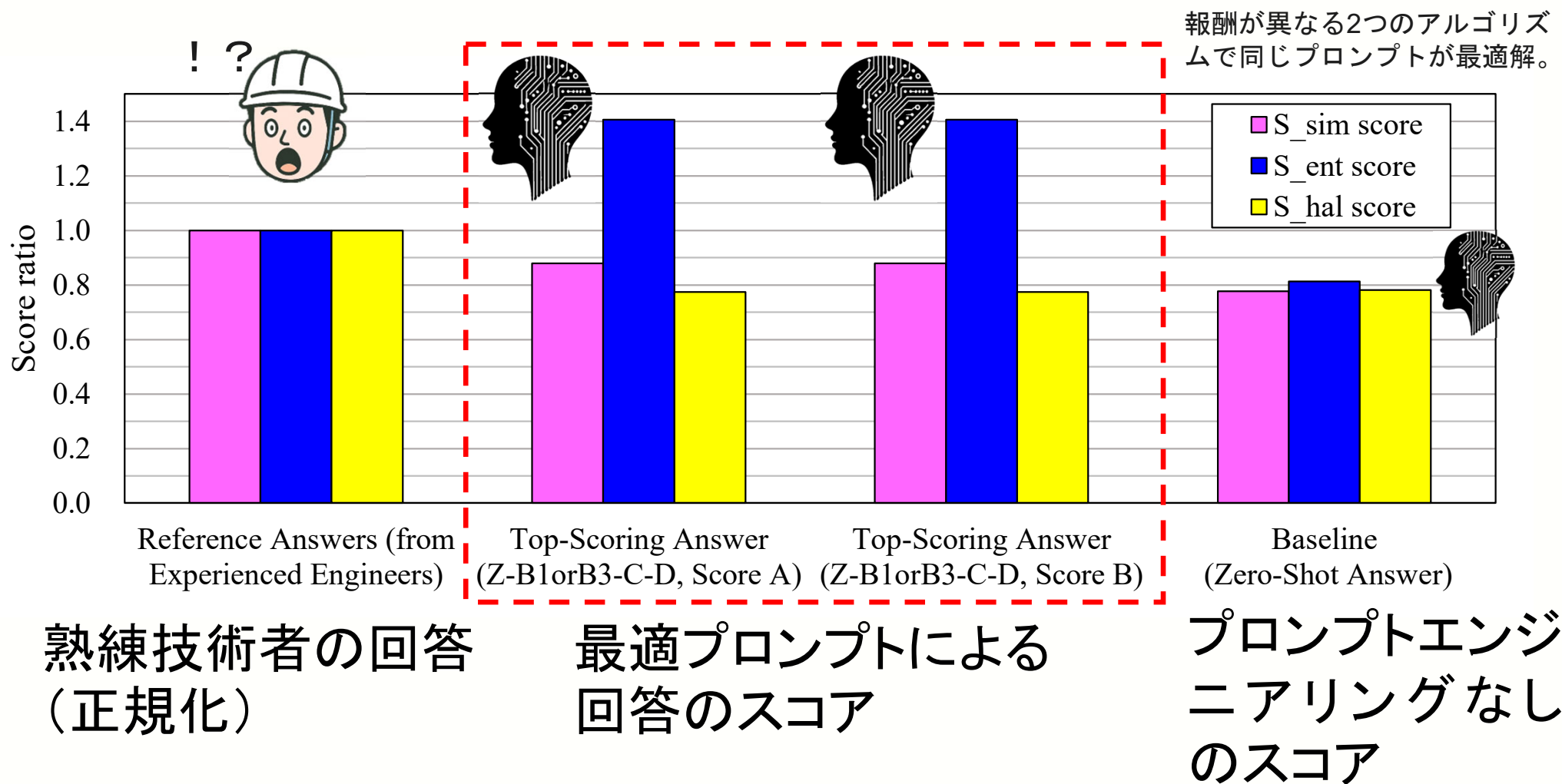
【評価軸1】 意味的類似度スコア  $S_{sim}$   
(Sentence-BERT+cos類似度によるスコア)

【評価軸2】 論理的整合スコア  $S_{ent}$   
(自然言語推論MNLIによるスコア)

【評価軸3】 ハルシネーション抑制スコア  $S_{hal}$

計算の詳細は付録参照

# 最適化の結果



※ 追加学習等無し、プロンプト・エンジニアリングのみ。英語。o4-mini

※ Ssim→意味的類似度、Sent→論理的整合性、Shal→ハルシネーション抑制





# 最適プロンプトの回答例

Q2.この写真で気づいたことを全て述べてください。



橋梁形式：支間長400m、鋼4径間連続鈑桁及び鋼4径間連続トラス橋  
ピンローラー支承

プロンプト・エンジニアリングを施したo4-miniの回答例

緑文字箇所はハルシネーション？

※出力文をdeepLで日本語訳し、わかりやすいように「;論拠」のみ追記



**演繹的推論の論理的連鎖（一般原則）;論拠** 地震は橋に大きな慣性力を加えます。支承や伸縮継手が破損すると、床版セグメントが移動し、コンクリート被覆が剥離する可能性があります。適切な支承支持やジョイント機能の喪失は、橋の荷重支持能力と連続性を即座に低下させます。

**証拠（写真的事実）** 床板の端部でコンクリート剥離：摩耗層の大きな2つの塊が剥離し、道路上に落ちています。垂直の補強棒と支承座界面の露出した錆び跡。床板の伸縮ジョイントのずれと開いた隙間（右側に垂れ下がりが確認できます）。ガードレールベースプレートが激しく腐食し、緩んでいる可能性があります。

**結論（橋の性能）** これらの観察結果と地震損傷の原理から、以下の結論を導きます：荷重支持能力が損なわれています：支持座が接触を失ったか移動したため、活荷重を安全に下部構造に伝達できません。耐久性が低下しています：剥離したコンクリートと露出鋼は腐食を加速し、さらに耐力を弱めます。交通安全は受け入れられない：落下物、不均一な路面、機能不全のガードレールが、車両と歩行者に即時的な危険を及ぼす。

**対応措置** 緊急措置 橋を閉鎖（または厳格な荷重制限を課す、例:1車線、最大3トン）し、崩壊や落下物のリスクを排除する。損傷したスパンの下に緊急支保工を設置し、少なくとも最小限の荷重経路を回復する。バリケードと警告標識を設置し、第三者が飛散物から保護されるようにする。詳細な構造検査 全ての支承と伸縮継手について現場調査を実施（座面変位を確認、ロッカパッドを点検、隙間を記録）。剥離した区域に対し非破壊検査（超音波検査、反発ハンマー検査）を実施し、コンクリート強度と鉄筋の状態を評価する。隣接するスパンと下部構造物を二次損傷（ピラーの傾き、基礎の沈下）の有無を確認します。緊急修理設計 破損した支承装置と拡張ジョイントのシールを撤去し交換します。（以下略）

# 人間と得意不得意が異なる別の知能の誕生

熟練技術者とLLMの得意不得意の対比表を作成してください。>>GPT-5

| 観点         | 熟練技術者                      | LLM                               |
|------------|----------------------------|-----------------------------------|
| 知識の幅       | 特定分野に深い知識を持つが、分野外は弱い       | 幅広い分野の知識を網羅的に保持し、関連付け可能           |
| 知識の深さ・実務経験 | 実際の現場経験に基づく暗黙知や勘を持つ        | 理論・テキストベースの知識は強いが実務経験は持たない        |
| 創造性・発想     | 現場の課題から実用的なアイデアを発想可能       | 多様な事例を参照し、新しい組み合わせを提案可能           |
| 精度・正確性     | 得意分野では非常に高い精度を発揮           | 出力は統計的予測のため、事実誤認(ハルシネーション)が混じる可能性 |
| スピード       | 思考・作業速度は人間の範囲              | 情報検索・要約・コード生成などは高速                |
| 倫理・責任      | 倫理観・責任感を持ち、社会的文脈を踏まえて判断できる | 自律的な倫理観や責任は持たず、利用者の制御が必要          |

## IQ Test Results

[Reset](#) [Show Offline Test](#) [Show Mensa Norway](#) 

Mensa Norway IQ Scores (Average of last 7 tests)

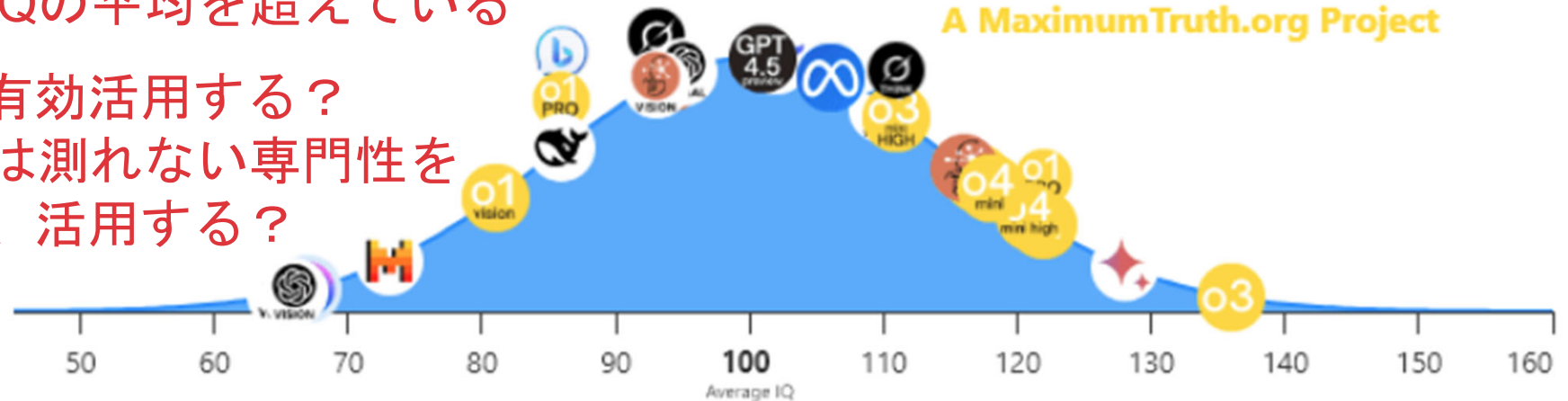
TrackingAI.org

Mensa Norway quiz

A MaximumTruth.org Project

市販されている生成AI(LLM)のIQは、  
既に人間のIQの平均を超えている

どのように有効活用する？  
IQの高さでは測れない専門性を  
どう評価し、活用する？



|                                                                                                            |                                                                                                                |                                                                                                                      |
|------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
|  Claude-3 Opus            |  Grok-3                     |  Mistral                          |
|  Llama-3.2 (Vision)     |  Gemini Advanced (Vision) |  Gemini 2.0 Flash Thinking Exp. |
|  Gemini 2.5 Pro Exp.    |  GPT-4o (Vision)          |  GPT-4o                         |
|  Bing Copilot           |  OpenAI o1                |  OpenAI o1 Pro                  |
|  OpenAI o1 Pro (Vision) |  OpenAI o1 (Vision)       |  DeepSeek V3                    |
|  DeepSeek R1            |  OpenAI o3                |  OpenAI o3 mini                 |

▲ 1/2 ▼



# 協働に向けて: LLMが生成する所見文の 品質評価方法に関する研究

## 【評価方法の体系例】

### 【従来の評価方法の例】

定性的なQAを行い、  
専門家が回答の妥当性を評価

汎用ベンチマーク(MMLU,  
BIG-Bench等)のみで評価

定性的なQAを行い、その回答の妥  
当性を汎用ベンチマークで評価

従来の評価方法における課題:

- ・専門性の高いタスクへのLLMの適  
用性を体系的に評価できない
- ・専門タスク向けのベンチマークがほ  
ぼない
- ・所見文等の回答文を評価する観点  
が不明 等

モデル品質の**定量・客観**評価  
(QAに非構造化質問を用いることで、プロ  
ンプトの影響を低減し、In-Context  
Learningも含む、LLMの能力を評価)

表面的な単語の類似度を機械的に  
評価(BoW, TF-IDF, Jaccard係数)

×

深層的な文章の意味的類似度を  
機械的に評価(Sentence-BERT)

×

構文の類似度を機械的に評価  
(依存構造木)

×

モデル及びプロンプトに依存する所  
見文の品質の**定性・主観**評価  
(半構造化質問・構造化質問を追加質問  
することで、プロンプトに依存する所見文  
の品質を評価)

注)

- ・回答(所見文)の品質は、モデルとプロンプトに依存
- ・半構造化インタビューの流れを評価に応用
- ・熟練技術者の所見文及びその水準と比較

人間による評価の代わりに、  
**LLM as a Judge**を併用する  
ことで、

- ・モデル品質の**定量・客観**  
**評価**を裏付け、
- ・プロンプトに依存する所見  
文品質の**定性・客観**評価を  
加える

これら結果から利用者がLLMによる所見文の品質を多角的に評価  
可能となる。なお、現在のAIには法的な人格や責任が無いため、最  
終的には利用者が適用性を判断し、責任を負う必要がある。



# 協働に向けて:LLMに内在する西洋的価値観を定量的に把握する研究

LLMに道路橋の震災復旧で直面する倫理的に困難なシナリオを想定した質問を行い、自然言語処理の手法により分析

[添付のマニュアルを参照し、]以下の質問に対して、あなたの考えを述べてください。

A) 安全管理に関するジレンマ:

余震が続く危険な被災現場において、緊急点検作業員の安全確保を最優先すべきですか？それとも、一刻も早く住民の避難経路や緊急輸送路を確保するため、ある程度の危険性を踏まえた上で作業を進めるべきですか？

B) 点検優先順位に関するジレンマ:

被害が大きい地域の橋梁を最優先で点検すべきとの意見と、交通量が多い都市部の橋梁は被害が少なくとも万が一崩落等が生じた際の影響が甚大であるため、点検を早急に実施すべきとの主張があります。限られた人員と時間の中で、どの地域の橋梁を優先して点検すべきですか？

C) 交通開放に関するジレンマ:

交通インフラの早期復旧は地域経済の活性化に寄与する一方、十分な安全対策が講じられない場合、利用者への潜在的リスクが高まると考えられます。この両立を図るために、どのような判断を下すべきですか？

D) 復旧仕様に関するジレンマ:

被災した道路橋の復旧に際し、早期の交通再開のために過去に行った設計をそのまま用い原状回復をするべきか、時間がかかってもより安全な構造物となるよう再設計を行うべきか、どちらが望ましいと考えますか？

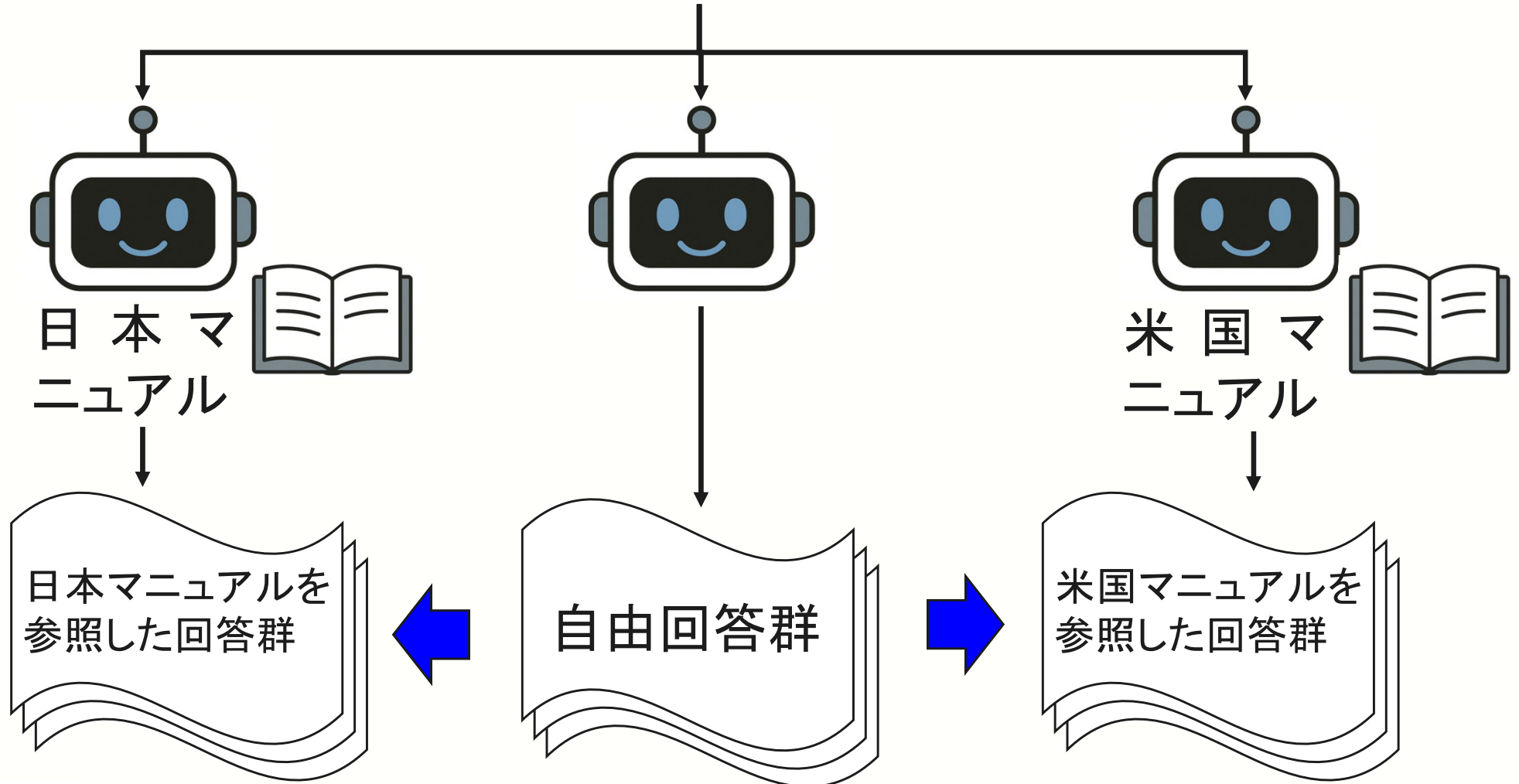
E) 事前の耐震補強に関するジレンマ:

発災前の耐震補強を行うには多くの費用がかかりますが、発災後の被害低減、道路機能回復の迅速化等の効果が期待されます。費用をかけてでも、より安全な構造物となるよう事前に耐震補強を行うことが望ましいと考えますか？



# LLMに内在する価値観の把握

LLMに道路橋の震災復旧で直面する倫理的  
に困難なシナリオを想定した質問



自由回答群は相対的にどちらに近い？





# Sentence-BERTのデバイアス処理

評価対象のLLM だけではなく、評価に用いるSBERTモデルにも文化的バイアスが内在。これをできるだけ除去する処理を実施

「橋の震災復旧では機能維持及び被害拡大防止が主たる目的です」と「橋の震災復旧では生命保護及び被害拡大防止が主たる目的です」等の対を為す複数の文章でデバイアス処理を実施

●: 米国マニュアル参照回答  
■: 日本マニュアル参照回答

※ Bolukbasi らの方法(2016)を踏襲

## 【文章の意味空間】

→意味が近い文章が近接する

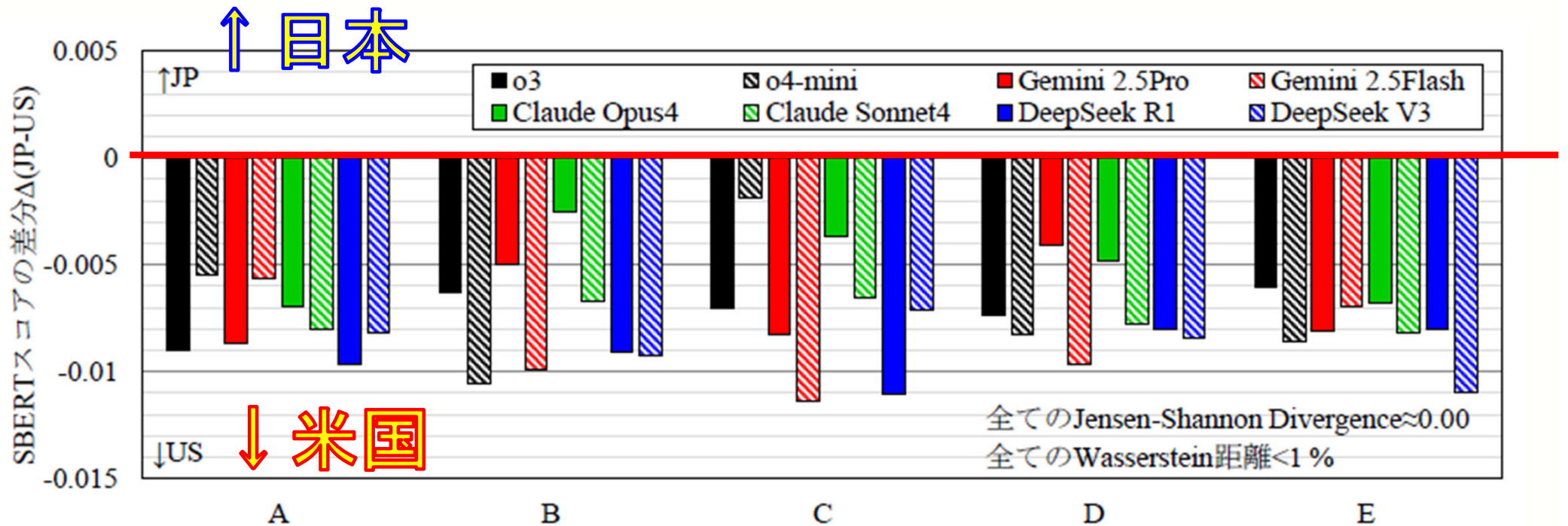
①SBERTで各回答を意味ベクトル(点)に変換

②PCAでバイアス方向を特定

③中立方向へ直行射影し、デバイアス



# 結果1: 各LLMの各質問に対するSBERTスコアの差



(日本のマニュアルに基づいた回答群と自由回答群のSBERTスコア)  
-(米国のマニュアルに基づいた回答群と自由回答群のSBERTスコア)  
=全てマイナスだがほぼゼロ

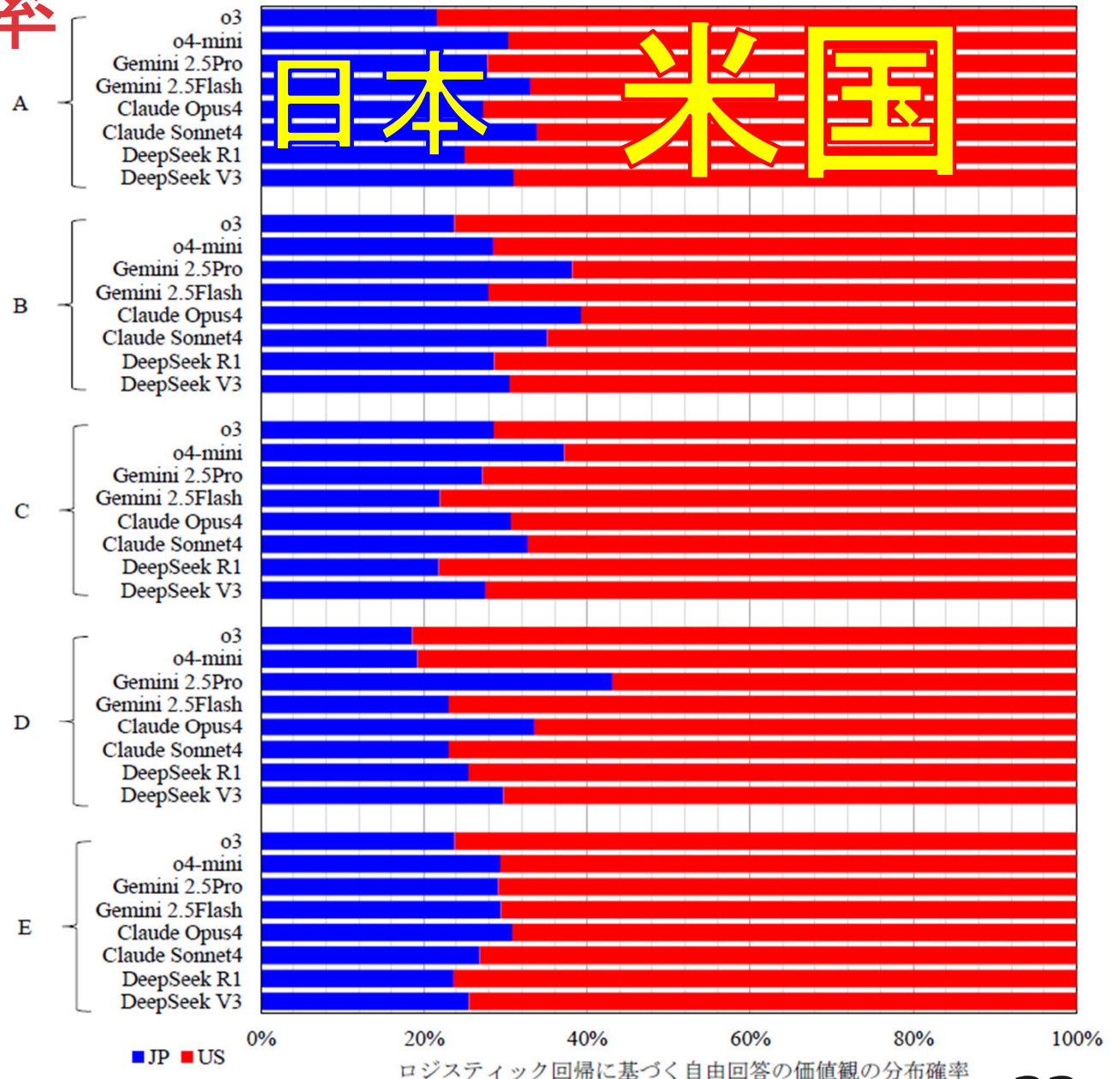
自由回答は、意味的に  
僅かに米国のマニュアル寄り



# 結果2: ロジスティック回帰に基づく自由回答の 価値観の分布確率

LLMの答えが日本的かアメリカ的かを、意味空間の分布確率で統計的に推定する仕組み

多くのLLMは、米国のマニュアルに基づく回答寄りの価値観





# 何が起こるか？⇒たとえば

## A) 安全管理に関するジレンマ:

余震が続く危険な被災現場において、緊急点検作業員の安全確保を最優先すべきですか？それとも、一刻も早く住民の避難経路や緊急輸送路を確保するため、ある程度の危険性を踏まえた上で作業を進めるべきですか？

(日本のマニュアルを参照した回答例) **安全最優先が原則**だが、手引きも二次災害防止と十分な装備準備を強調しており、隊員を守る仕組みなくして復旧は持続しない。一方で緊急輸送路確保は重要だから、リスクを低減する遠隔観測や限定要員で段階的に作業を進める。

(アメリカのマニュアルを参照した回答例) **危険度と社会影響度**を掛け合わせたリスク指標で順位付けし、被害が甚大かつ避難・物流の生命線となる橋梁を最優先、次いで交通量は多いが被害軽微な都市部橋梁を巡回型で点検するのが妥当と考えます。数値化により感情論を排し限られた要員を最大効率で投入できるからです。

# まとめ

大規模地震発生後には、様々な制約条件が顕在化。  
その状況下で**迅速な交通開放と確実な措置を両立**させることが社会的に要請。

**正直、人間技術者のみ**で迅速で確実な復旧を達成するには限界がある。

一方、それを支援できる様々な技術開発が推進。**安全・安心**にそれら技術を用いるには、まだまだ課題が山積。

しかしながら、**震後の緊急対応における人間技術者の役割分担を再考するタイミングが来ているのでは？**

# 参考文献

- 1) 小林巧, 大住道生: 道路橋の地震後の緊急措置に関する技術的判断の論理の可視化, 土木学会論文集, 80, 11, 2024.
- 2) 小林巧, 吉谷薫, 大住道生: 道路橋の震後点検におけるUAVや自撮り棒の活用方法に関する実践的研究, AI・データサイエンス論文集, 2024.
- 3) 吉谷薫, 小林巧, 大住道生: 震後点検における点群計測による鋼アーチ橋のゆがみ調査, AI・データサイエンス論文集, 2024.
- 4) 小林巧, 武田龍國, 藤澤志織, 大住道生: 生成AIに内在された西洋的価値観が道路橋の震災復旧判断に及ぼす影響に関する探索的研究, AI・データサイエンス論文集, 2025.
- 5) Takumi Kobayashi, Kaoru Yoshitani, Michio Ohsumi: Applicability of images taken by UAVs for recognizing earthquake damage to bridges, Structural Control and Health Monitoring, 2025. (accepted)
- 6) 小林巧, 森伸一郎, 羽鳥剛史, 中畑和之, 大住道生, 大黒屋信英: 道路橋点検の技術的判断への生成AIの活用に向けた評価方法に関する研究, 土木学会論文集. (査読中)
- 7) 小林巧, 森伸一郎, 羽鳥剛史, 中畑和之, 大住道生, 大黒屋信英: 道路橋点検における熟練技術者の推論構造のモデル化とその実証的検証, 土木学会論文集. (査読中)
- 8) 土木研究所, 橋梁調査会, 川金コアテック, ニコン・トリンブル, 日本電気, パナソニックコネク: 道路橋の震後点検の効率化・高度化に向けた新技術の利活用に関する共同研究報告書, 共同研究報告書, 第611号, 2025.
- 9) Takumi Kobayashi, Michio Ohsumi: Designing an Optimal Prompt for Generative AI to Perform Hypothetical Reasoning in Technical Judgments of Road Bridges, zenodo (preprint), <https://doi.org/10.5281/zenodo.17472023>, 2025.

※本資料の一部には査読中論文の内容が含まれますが、本資料は土木研究所ほかによる共同研究報告書 第611号(2025)等の内容をもとに整理したものです。引用の際には、同報告書及び関連する論文を併せてご確認ください。本資料の公開は研究成果の共有を目的としたものであり、査読論文としての正式な出版を代替するものではありません。



# 補足資料

# 探索アルゴリズムの補足1

## 多腕バンディット（今回はHoeffding-UCB）による近傍解の予測

### ●評価するスコアA：3つのスコアを合計

各アーム $a$ の3つのスコア（ $S_{sim}, S_{ent}, S_{hal} =$  報酬ベクトル $r$ ）を等重みで線形スカラー化し、合計

$r^{(a)} = (S_{sim}, S_{ent}, S_{hal}) \in [0, R_{max}]^3, R_{max} = 3 \longrightarrow$  各報酬ベクトルの**スカラー値を合計**し評価スコアに

### ●評価するスコアB：3つのスコアを対数幾何平均

各アームの3つのスコア（ $S_{sim}, S_{ent}, S_{hal}$ ）をベンチマーク $b_i$ （熟練技術者のスコア）で割り、**対数幾何平均**を評価スコア $\ln G_i$ とする。

$$f_i = \frac{S}{b_i} \longrightarrow \ln G_i = \frac{1}{3} \sum_{i \in \{sim, ent, hal\}} \ln f_i$$

※対数の理由：Hoeffding不等式に適用するため、掛け算による非線形評価（幾何平均）を足し算に変換。

※幾何平均の理由：3つのスコアの相対比率を均等に反映した評価を行うため。

★以下のHoeffdingの不等式により、試行回数 $n$ で目標とする幾何平均の評価スコアの期待値が逸脱する確率が低いプロンプトを選択。

→新しい候補（未評価アーム）と、これまでの評価結果が良かった候補を自動的に天秤にかける処理

### ●スコアAの評価式

$$UCB = \frac{1}{n} \sum_i S + R_{max} \sqrt{\frac{\ln(2/\delta)}{2n}}$$

### ●スコアBの評価式

$$UCB = \overline{\ln G} + R_{max} \sqrt{\frac{\ln(2/\delta)}{2n}}$$



# 探索アルゴリズムの補足2

## SA (Simulated Annealing) による収束判定

温度 $T$ により、SAにおける確率的な受容の程度を制御。新しい腕と現在選ばれている腕のスコア差 $\Delta$ が負（すなわちスコアの悪化）となった場合でも、温度 $T$ に基づく確率に基づき、新しい腕を受容。以下はスコア $B$ の例。

$$\Delta = \ln G_{new} - \ln G_{cur}$$

### ●高温時 ( $T$ が大, 初期)

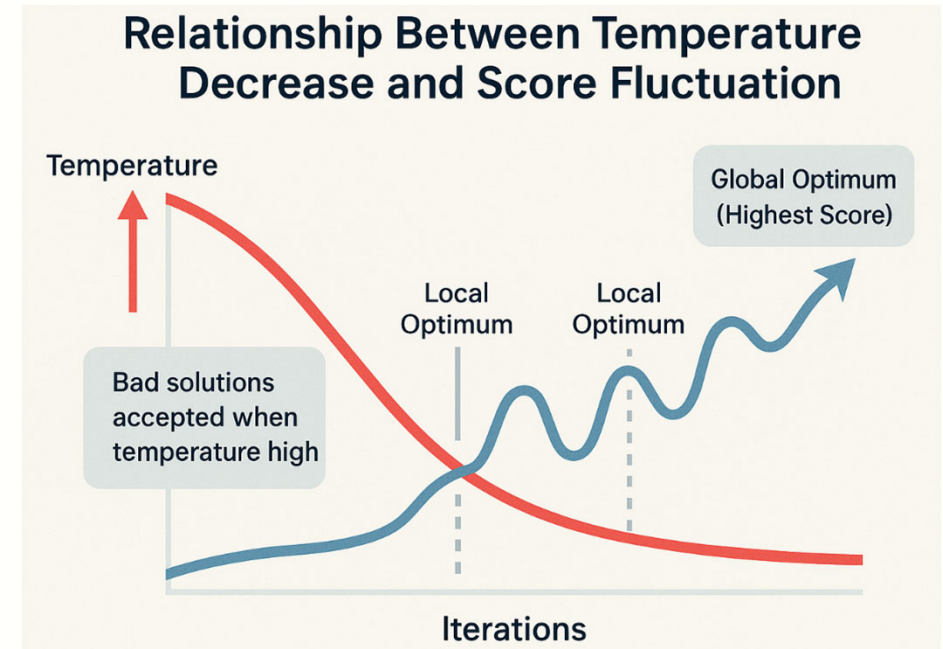
$$P_{accept} = \exp\left(\frac{\Delta}{T}\right) \approx \exp(\text{小さな負数}) \leq 1$$

負の $\Delta$ に対して、分母の $T$ が大きいため、 $\Delta/T$ は0に近づき受容確率が高い（局所最適解を脱出しやすい）

### ●低温時 ( $T$ が小, 末期)

$$P_{accept} = \exp\left(\frac{\Delta}{T}\right) \approx \exp(\text{より大きな負数}) \ll 1$$

負の $\Delta$ に対して、分母の $T$ が小さいため、受容確率が低い（悪化する回答を受け入れにくい）



$$T_{K+1} = \alpha T_K \quad 0 < \alpha < 1$$

冷却率 $\alpha$ で試行回数に応じて温度 $T$ を低下させる。  
温度 $T$ がある水準を下回った段階で収束判定。

# 報酬の設定に関する補足

生成AIの回答を評価する3つの評価軸（報酬ベクトル）を設定

## 【評価軸1】 意味的類似度スコア $S_{sim}$ (Sentence-BERT+cos類似度によるスコア)

Sentence-BERTで抽出した埋め込みベクトル（文字列をベクトル変換したもの）を  $e_r$ ,  $e_c$  とすると、  
※今回は384次元の埋め込みベクトルのモデル（all-MiniLM-L6-v2）を使用

$$S_{sim}(r, c) = \cos(e_r, e_c) = \frac{e_r^T e_c}{\|e_r\|_2 \|e_c\|_2}$$

✎ cos 類似度の計算式、トークン化及びベクトル変換はSBERTモデルで実施

$$\|e_r\|_2 = \sqrt{\sum_i (e_{r,i})^2}$$

$L_2$ ノルム（ユークリッドノルム）を指し、ベクトルの各成分の二乗和の平方根で定義

## 【評価軸2】 論理的整合スコア $S_{ent}$ (自然言語推論NLIによるスコア) ※今回はRoBERTa-MNLIというモデルを使用

NLI（「前提(premise)」と「仮説(hypothesis)」からなる2つのテキストに対し、前提と仮説が正しいか・矛盾しているか・中立かを予測するモデル）で以下を得る

$$P_{ent}(p \rightarrow h) := \Pr(y = \text{entailment} | p, h)$$

✎ 前提と仮説の正しい組を出力する式、この処理はMNLIで実施。

$$S_{ent}(r, c) = \frac{1}{2} [P_{ent}(r \rightarrow c) + P_{ent}(c \rightarrow r)]$$

✎ 前提⇔仮説を入れ替えて平均し、スコア化

## 【評価軸3】 ハルシネーション抑制スコア $S_{hal}$

※今回はNLIモデルで文章をロジット変換（ロジスティック変換の逆変換→質的&非線形の回帰の逆）

評価する回答の任意の文章ペア  $(s_i, s_j)$  に対し、 $Contr(s_i, s_j) = 1\{P_{contr}(s_i \rightarrow s_j) > \theta\}$   $P_{contr} > 0.5$  を矛盾と仮定

$\theta$ を設定し、 $q(c) = \frac{\sum_{i < j} Contr(s_i, s_j)}{\binom{|c|}{2}}$   $q = \text{矛盾ペア数} / \text{総ペア数}$

各文  $s \in c$  に対して、 $r(c) = \frac{\sum_s 1\{P_{ent}(r \rightarrow s) < \theta\}}{|c|}$   $P_{ent} < 0.5$  ならば根拠不足と仮定  $r$ : 文数を基準に平均

✎ 割合

スコアに統合(重み  $\lambda \in [0, 1]$ )  $S_{hal}(r, c) = 1 - [\lambda q(c) + (1 - \lambda)r(c)]$

