

AI システム開発における AI 開発者とシステム開発者の

ギャップ解消に向けた AI 説明技術の検証

An Investigation into the Effectiveness of eXplainable AI to Fill the Gap between AI Developers and System Developers in AI System Development

研 究 員：斎藤 弘之 (NTT コミュニケーションズ株式会社)

東條 洋 (キヤノン株式会社)

主 査：石川 冬樹 (国立情報学研究所)

副 主 査：栗田 太郎 (ソニー株式会社)

徳本 晋 (株式会社富士通研究所)

研究概要

AI はブラックボックスと言われており、その素性や性能はデータを通してしか把握できない。そのため、AI に対する知見の深さによって AI の評価は大きく影響される。AI システム開発において、AI の専門家である AI 開発者と非専門家であるシステム開発者の間には、AI コンポーネントの評価においてギャップが生じており、円滑な連携の妨げとなっている。学術研究分野では AI を説明する技術について様々な提案があるが、AI システム開発の現場における実用性についてはよく知られていない。そこで本論文では、運転支援システムにおける道路標識判定を題材に AI 説明技術を適用することにより、AI 開発者とシステム開発者のギャップの解消が可能であることを検証した。その結果、一定の効果があることが確認できた。一方で、使用には注意が必要であることも分かった。

1. はじめに

Deep Learning に代表される機械学習技術の進展に伴い、従来と比べて圧倒的に高精度な AI 技術が登場しており、自動運転、医療診断、金融サービスなど幅広い分野で AI がシステムに導入されるようになってきている。通常のソフトウェアコンポーネントと（機械学習技術を用いた）AI コンポーネントの違いは、何がどこまでできるかは AI コンポーネント作成に用いたデータによって決定される点にある。本特性から AI システム開発には様々な問題^[1]が発生している。例えば、AI の精度は 100% と考える顧客もあり、受け入れの判断における開発者とのギャップが指摘されている^[1]。我々は日々の AI システム開発の中で、システム開発者と AI 開発者の間にもギャップがあると感じている。AI 開発者はシステム要件を満たす AI 技術を選択・作成し、システム開発者は作成された AI コンポーネントを用いて顧客の問題を解決するシステムを構築する。システムへの受け入れ判断では、AI の失敗がバグとは違い、必ずしも修正可能とは限らないため、失敗の受け入れ判断がポイントとなる。AI の非専門家であるシステム開発者は AI コンポーネントの動作原理を知ることができないため、「人間であれば間違えない」等、納得のいく失敗かどうかという観点で捉える。一方で AI 開発者にとって納得のいく失敗とは、AI の開発経験や学術論文から得た知見に照らして改善が困難と考えられるものであり、システム開発者と一致しないことも多い。このようなギャップから、AI コンポーネントの受け入れ判断に長時間の議論が費やされている。そこで本論文では、AI 開発者とシステム開発者の失敗の捉え方のギャップを課題にあげ、AI 説明技術の導入の効果を検証した。検証にあたっては二者間の納得の仕方の違いに着目し、「納得感」を指標として採用した。検証の結果、二者間の納得感の差は縮まる傾向が見られ、ギャップ解消に一定の効果があることがわかった。しかしながら、

第5研究コース (XAI グループ)

納得感に基づく判断の危うさも見られ使用方法に注意が必要であることもわかった。

本論文の構成は以下の通りである。2章で現場の問題、3章で研究の目的、4章で関連研究、5章で検証、6章で考察、7章で結論と展望についてそれぞれ述べる。

2. 現場の問題

AI コンポーネントのシステムへの受け入れの判断における AI 開発者とシステム開発者間のやり取りの一例を以下にあげる。

- (1) AI 開発者による評価データを用いた評価結果の説明
- (2) システム開発者による AI システム検証環境で商用データ等を用いた評価結果の説明
- (3) 二者間の失敗ケースの割り切り・改善の協議

(1), (2) は通常のソフトウェアコンポーネントのリリース・受け入れ評価に相当する標準的な活動であるが、AI コンポーネントにおいては(3)のような AI の失敗ケースを確認する何らかの活動が必要となると考える。失敗ケースとは、評価データの個々のサンプルに対して AI コンポーネントの出力が誤っているケースであり、作成された AI コンポーネントの性能限界を示す具体例にあたる。失敗ケースを割り切りとするか、改善とするかで、AI コンポーネントの詳細仕様が固まる場合があるのではないかと考える。割り切りか改善かを判断する指標として、機械学習品質マネジメントガイドライン^[2]に外部品質として挙げられているリスク回避性、AI パフォーマンス（有用性）、公平性が参考になる。リスク回避性と公平性の観点は人的被害や社会に与えるインパクト等、誰にとっても共通の価値観に基づいており比較的線引きしやすいが、有用性の観点では曖昧になりやすい。指標の一つとして、システムが想定している環境における当該ケースの発生頻度が考えられる。レアケースであれば比較的割り切り易いからである。しかしながら、開発開始時には洗い出せていなかった想定外のケース等、発生頻度の判断が困難なケースも少なくない。このような場合、AI コンポーネントの振る舞いは内部動作から推測できないことから、システム開発者は失敗ケースから判断する他なく、納得のいく間違え方であるかという観点で捉えるようである。「人間であれば間違えない」、「従来のソフトコンポーネントであれば間違えない」といった納得できない間違え方は、ユーザの AI に対する信頼性を損ないかねないからである。一方で、AI 開発者は、失敗ケースを AI の知見に基づいた改善の可能性の観点で捉える。AI 開発者も「人間であれば間違えない」といった判断をすることもあるが、AI の改善が容易と推測できる場合であり、システム開発者の判断とは必ずしも一致しない。AI 開発者にとって納得できない失敗ケースとは、改善が容易であるのに失敗しているケースである。このようにシステム開発者と AI 開発者では失敗ケースの捉え方にギャップがあり、(3)の失敗ケースの割り切り・改善の協議で、意見の一致に至るには長い時間を要している実態がある。

3. 研究の目的

本論文では、システム開発者と AI 開発者の失敗ケースの捉え方のギャップをどう埋めたらよいかを調査する。

3.1 XAI 技術の導入

失敗ケースの捉え方のギャップは、AI がブラックボックスであり、AI の非専門家であるシステム開発者には AI がどのように判断を行ったか分からないことが一因であると考えられる。学術分野では AI の出力結果の判断根拠の説明を試みる研究が盛んに行われており、XAI (eXplainable AI) と呼ばれている。XAI の代表的な手法としては、入力値近傍の挙動を単純な関数で近似して説明する手法である LIME (Local Interpretable Model-agnostic Explanations)^[3]、AI の構成要素の一部から AI がどこに着目して判定を行ったかを抽出して可視化する手法である Grad-Cam^[4]等がある。

第5研究コース（XAI グループ）

このように個別の事象について、どのように AI が判断しているか洞察を与える技術が XAI である。本技術を失敗ケースに適用することでシステム開発者と AI 開発者のギャップを埋めることができるのではないかと考え、検証を試みることにした。

3.2 納得感への着目

XAI によって失敗ケースの捉え方のギャップ解消を検証するにあたり、何を指標として検証するかが課題となる。現場の問題で触れたように失敗ケースの捉え方は具体的には納得の仕方の違いという形で現れる。また、納得という概念には、様々な属性があることが知られているが^[5]、「理解の深化」、「受け入れ」、「意思決定」といった AI システム開発の文脈と深い関係にある属性がある。そこで我々は、XAI 導入により改善が期待できる「理解の深化」に焦点を当て、「納得感」を指標とすることにした。

4. 関連研究

AI の判定結果の人間の解釈という観点では、桑島らの AI による画像内容の判定結果と、AI が判定過程で着目した部分画像を人間が見て付与した言葉の説明が、概念的に整合しているかについての研究^[6]がある。AI の判定過程と判定結果が人間から見て整合しているかについて明らかにしたものであるが、人間が納得するかという点にまで踏み込んでいない。

AI の信頼性という観点では、Dietvorst らによる AI による予測と人間による予測のどちらを信頼するかを調査した研究^[7]がある。人間の予測結果は AI より 2 倍失敗しているにもかかわらず、被験者は AI よりも人間の予測結果を信頼する、という結果が得られている。このことは、AI は未だ社会に対して広く受け入れられていないことを示しており、この意味でシステム開発者が AI の信頼性を重視するのは妥当なことと考えられる。

5. 検証

5.1 検証概要

システム開発者と AI 開発者に AI の失敗ケースを見せて納得感を評価してもらい、次に XAI の出力を同時に示して評価してもらった。題材には比較的ドメイン知識を必要としない「運転支援システムにおける道路標識判定」を用いた。画像はドイツの道路標識のデータセット GTSRB (German Traffic Sign Recognition Benchmark)^[8]を使用した。AI は Resnet50^[9]を用い GTSRB に含まれている学習用データを使用した。XAI には前述の GradCam の改良版である Grad-Cam++^[10]を用い、AI の着目領域を重視の度合いに応じて色分けし、入力画像に重ねて表示した。

5.2 検証方法

アンケート調査の対象と内容を以下に示す。

- (1) 調査対象：研究員が所属する各社の開発者 19 名。
- (2) 回答方法：各開発者にアンケート用紙を配布し回答を依頼。
- (3) アンケート内容：

設問 1：回答者の属性（システム開発や AI 開発の経験年数）を答える

設問 2：評価画像 1000 枚中の失敗画像 156 枚からランダムに選択した 15 枚の画像（サンプル）に対して AI の判定結果と確信度を示し、納得感とその理由を答える

- 納得感は、「4. 非常に納得する、3. やや納得する、2. あまり納得しない、1. 全く納得しない」から選択する
- 理由は、自由記述とする

設問 3：設問 2 の設問に加え各画像に XAI の画像を示し、納得感とその理由を答える（納得感と理由の回答方式は、設問 2 と同じである）。

第5研究コース（XAI グループ）

5.3 検証結果

システム開発者 12 名，AI 開発者 7 名，計 19 名から回答が得られた．納得感を「非常に納得する」から「全く納得しない」の 4 段階を 4 から 1 に数値化して集計した．

5.3.1 設問 2：XAI 導入前のシステム開発者と AI 開発者の納得感の比較

図 1 は XAI 導入前の各サンプルに対する納得感の平均値を，システム開発者と AI 開発者で比較したものである．横軸が各サンプルの ID，縦軸が納得感の平均値である．

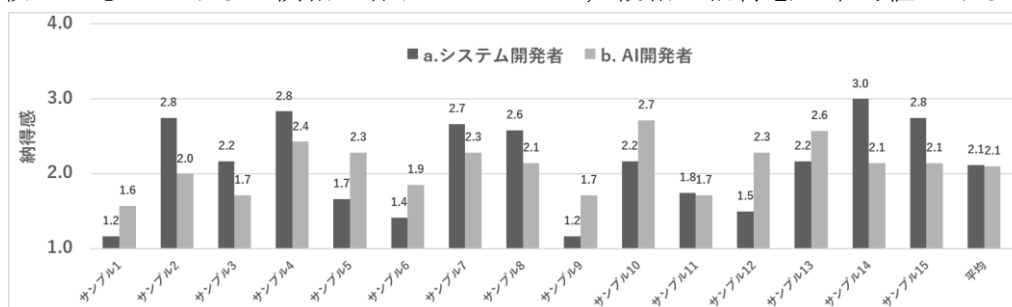


図 1 XAI 導入前のシステム開発者と AI 開発者の納得感の比較

- 二者間の納得感の平均の差の t 検定を行った結果(付録 2)，有意水準 5%では有意ではなく，全サンプル平均で納得感は 2.1，「あまり納得しない」傾向であった．
- サンプル毎の納得感の幅は，システム開発者と AI 開発者でそれぞれ，1.2～3.0，1.6～2.7 となった．
- システム開発者と AI 開発者で納得感の差が 14 サンプルで 0.4 以上あり，特に，サンプル 12 とサンプル 14 で，納得感の差が 0.8，0.9 と納得感の 1 ステップ分 (1.0)に近い差が見られた．
- 納得感が高いサンプルは，標識の形状が合っている入力画像(サンプル 4, 7, 14, 15)，及び，人が見ても判定が困難と思われる低画質の入力画像（サンプル 2）であり，いずれもシステム開発者の納得感の方が高い．（入力画像は付録 1 を参照）
- サンプル 12, 14 の納得感の理由(付録 3)によると，システム開発者は，「確信度が高く，形状も異なる」，「人間でも間違いかねない」等，設問にある情報（入力画像，確信度）から答えているが，AI 開発者は「照明変動に弱いアルゴリズムであると推測できる」，「難しい問題に見えない」など AI の知見に基づいた答えもあった．

5.3.2 設問 3：XAI 導入後のシステム開発者と AI 開発者の納得感の比較

図 2 は XAI 導入後の各サンプルに対する納得感の平均値を，システム開発者と AI 開発者で比較したものである．横軸が各サンプルの ID，縦軸が納得感の平均値である．

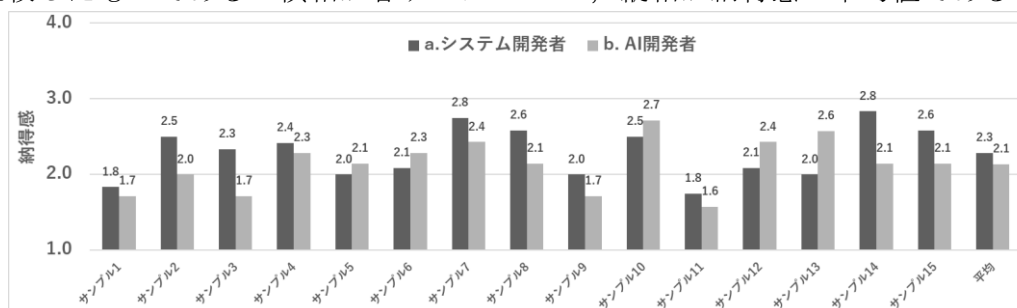


図 2 XAI 導入後のシステム開発者と AI 開発者の納得感の比較

- 二者間の納得感の平均の差の t 検定を行った結果(付録 2)，XAI 導入前と同じく有意水準 5%では有意ではなく，全サンプル平均で納得感はそれぞれ 2.3，2.1 と「あまり納得しない」傾向であった．
- サンプル毎の納得感の幅は，システム開発者と AI 開発者でそれぞれ，1.8～

第5研究コース（XAI グループ）

- 2.8, 1.6~2.7 となり, XAI 導入前（設問 2）より近くなる傾向がみられた.
- サンプル毎の納得感の差が 0.4 以上であったサンプルは, XAI 導入前（設問 2）の 14 サンプルから 7 サンプルに減った.
- XAI 導入前後で AI 開発者のサンプルごとの納得感の変化は小さく, AI 開発者の内 3 名は理由（付録 3）で「XAI は納得感に影響しない」とし, 変化がなかった.
- XAI 導入前（設問 2）に AI 開発者の納得感がシステム開発者より高かったサンプル 1, 5, 6, 9, 10, 12, 13 は, XAI 導入後, サンプル 13 を除いた全てについてシステム開発者の納得感が向上し, 二者の差は 0.3 以内になった.

5.3.3 サンプル毎の詳細分析

サンプルごとの詳細分析を行った. ここではシステム開発者の納得感が AI 開発者に近くなった代表例としてサンプル 1 とサンプル 9 について取り上げる.

(1) サンプル 1

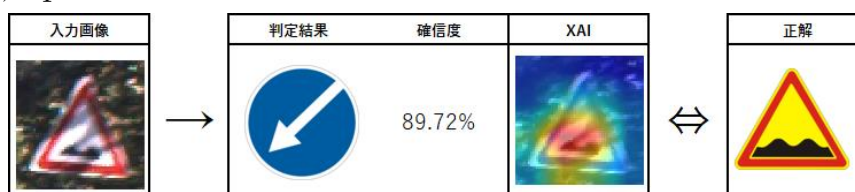


図3 サンプル1の設問3の設問内容

サンプル1の判定結果は正解と形状, 色, マークが全く異なり, かつ, 確信度は約90%と高い(図3). XAI が示した着目領域は標識の中央部分に集中している(赤で示した部分). 図4は, サンプル1に対するシステム開発者の納得感をXAI導入前と導入後の比較で示しており, 図5は同様にAI開発者についての比較である. 横軸が納得感4段階, 縦軸が人数を示している.

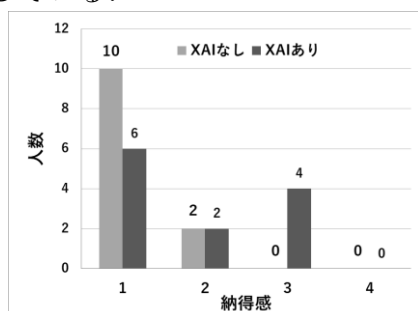


図4 サンプル1のシステム開発者の納得感

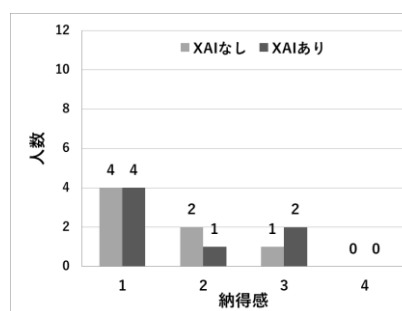


図5 サンプル1のAI開発者の納得感

- XAI 導入後, システム開発者は半数 6 名, AI 開発者 1 名の納得感が向上した.
- システム開発者は XAI 導入前, 「全く納得しない」が圧倒的多数であったが, 「やや納得する」4 名になり, 回答者によってバラつきが生じた.
- 納得感が向上した回答者の主な理由（付録 3）は, 「影の影響により判定結果の左下向きの矢印に見える」であり, XAI 導入後のみ「影の影響」への言及がみられた.

(2) サンプル 9

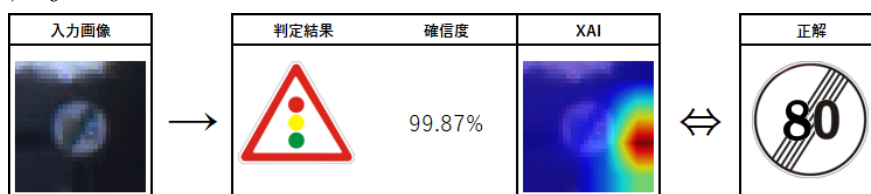


図6 サンプル9の設問3の設問内容

第5研究コース（XAI グループ）

サンプル9の判定結果は入力画像と形状，色，マークが正解と全く異なり，かつ，確信度は約100%と高い結果である（図6）．XAIが示した着目領域は標識から大きく外れている（赤で示した部分）．図7は，サンプル9に対するシステム開発者の納得感をXAI導入前と導入後の比較で示しており，図8は同様にAI開発者についての比較である．横軸が納得感の4段階，縦軸が人数を示している．

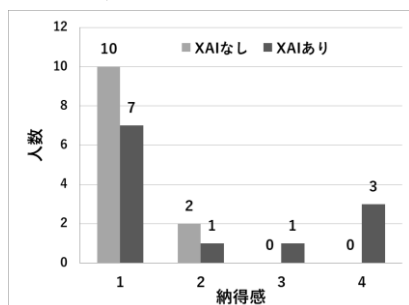


図7 サンプル9のシステム開発者の納得感

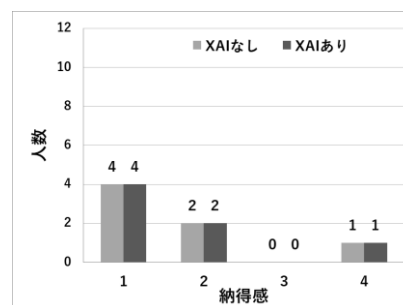


図8 サンプル9のAI開発者の納得感

- XAI導入後，システム開発者5名，アルゴリズム開発者2名の納得感が向上し，システム開発者1名，アルゴリズム開発者1名の納得感が低下した．
- システム開発者はXAI導入前，「まったく納得しない」が圧倒的多数であったが，「やや納得する」1名，「納得する」4名になり，回答者によるバラつきが生じた．
- 「全く納得しない」から「非常に納得する」への極端な変化が見られた（付録2）．
- 納得感が変化した主な理由（付録3）は，着目領域が標識から外れていることであるが，同じ理由であっても「やや納得する」「非常に納得する」とした回答者と，反対に「全く納得しない」とした回答者があった（付録2）．

なお，XAI導入後に回答者間で納得感のバラつきが生じた例は上記サンプル1, 9の他，半数以上の8サンプルに見られた．また，納得感が納得する側か納得しない側に揃ったサンプルは一つもなかった．（詳細は付録2を参照）

5.3.4 納得感の理由の分析

納得感の理由の分類を行った．XAIが示す着目領域によって形状・色等の具体的な内容の言及があり，「分かる」などの理解を示す表現を使っている理由を「理解の深化」とした．また，XAIの着目領域を見ることで必ずしも理解が深まって（「理解の深化」）いるとは限らず，返って疑問が生じているものもあった．そこで，具体的な内容を上げた上で疑問を呈している理由を「分からない」とした．「理解の深化」，「分からない」のどちらにも当てはまらない理由は，XAIの影響が見られないと考え「XAIの影響なし」とした．以上の定義に従って分類を行い，回答者全員，AI開発者，システム開発者それぞれについて集計した（図9）．なお，総数は285（＝回答者19名×15サンプル）である．

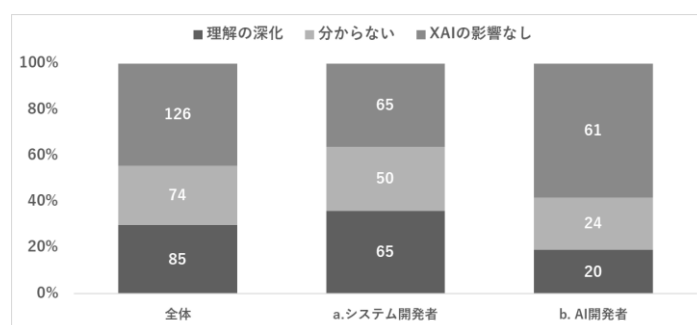


図9 納得感の理由の分類結果

第5研究コース（XAI グループ）

- 回答者全員の約 56%に XAI の影響（「理解の深化」、「分からない」）が見られた。システム開発者では約 64%で、AI 開発者の約 42%より多かった。
- 「理解の深化」と「分からない」ではやや「理解の深化」が多いものの、どちらかに大きく偏ることはなかった。

6. 考察

6.1 XAI 導入前の失敗ケースの納得感

5.3.1 の結果が示すように、全サンプルの平均でみると AI 開発者、システム開発者共に「あまり納得しない」傾向である。もともと道路標識は人間が判定し易いようにデザインされているためか、AI の精度に対する期待が高いことが伺える。また、サンプル全体では二者間の納得感の平均の差は t 検定の結果から有意ではなかったが、サンプルごとに比較すると、納得感の幅や 0.4 以上の差がついているサンプルの数で明らかな差が見られた。また、納得感の高いサンプルの傾向が、形状が正しいサンプルや低画質のサンプルであることから、人間が見た時の判定のし易さが納得感の基準となっていると考えられる。この傾向は納得感の高さからシステム開発者の方がより強いと考えられる。また、納得感の差が大きかったサンプル 12, 14 において、AI 開発者が AI の知見に基づいた理由が見られた。このことから、AI の知見の有無が二者の差の一因となっていると考えられる。

6.2 XAI 導入後の失敗ケースの納得感

5.3.2 の結果が示すように、XAI 導入後も全体としては「あまり納得しない」傾向であり、納得感の平均の差も有意ではなく、XAI は納得感の向上に直結していないと思われる。これについては 6.3 で更に考察する。XAI 導入前にサンプルごとの比較で見られた AI 開発者とシステム開発者の差（納得感の幅と 0.4 以上の差があるサンプル数）は、XAI 導入後は縮まった。また、AI 開発者の納得感の変化は全体的に小さく、XAI 導入前に AI 開発者の方が納得感の高かったサンプル 1, 5, 6, 9, 10, 12, 13 において、サンプル 13 を除きシステム開発者の納得感が向上していることから、システム開発者の方が AI 開発者に近づいたと考えられる。このことは、5.3.3 の図 4 と図 5、図 7 と図 8 から、AI 開発者は XAI 導入の前後で納得感の人数分布に変化があまりないが、システム開発者の人数分布は変化し、AI 開発者の人数分布に近づいていることから確認できる。これらのことから、XAI は AI 開発者とシステム開発者間のギャップ解消に一定の効果をもたらしていると考えられる。また、サンプルごとの納得感は、XAI 導入後も回答者間でバラつきがあり、5.3.3 のサンプル 1 やサンプル 9 のように更に大きくなるサンプルもあった。この傾向は、システム開発者間の方に強く見られた（図 4、図 7）。また、サンプル 9 では、納得感が XAI 導入前後で極端な変化が見られること、納得感の向上した回答者と納得感に変化が見られなかった回答者で同じ理由を挙げていることから、納得感の評価は人によって異なるものであることが分かる。

6.3 XAI の影響

6.2 で“XAI は納得感の向上に直結しない”と述べたが、5.3.4 の納得感の理由の分析結果が示すように、納得感が向上しなかった回答であっても「理解の深化」或いは「分からない」理由として、具体的な根拠への言及という形で XAI の影響が見られた。具体的な根拠への言及の例として 5.3.3 のサンプル 1 では、XAI の結果を見ることで初めて「影の影響」が理由に挙げられるようになっている。この傾向はシステム開発者の方により強く表れており（AI 開発者約 42%に対し約 64%）、システム開発者の方が XAI をより積極的に活用していると考えられる。一方で AI 開発者への影響は限定的であった。AI 開発者の中には「XAI の結果は納得感に影響しない」とする回答者もあった（5.3.2）。今回実験に使用した道路標識判定は、学習データとして使用したデータの量・質共に実際の AI 開発に比べて

第5研究コース (XAI グループ)

小規模であり，AI 開発者は XAI 結果なしにその実力を推定可能だったと思われる．

7. 結論と展望

失敗ケースの捉え方に AI 開発者とシステム開発者でギャップが生じていること，システム開発者は人間が見た時の判別のし易さに基づいて納得し，AI 開発者は AI に対する知見に基づいて納得していることが確認できた．また，XAI の導入によりシステム開発者の納得感が AI 開発者に近づくという形でギャップ解消の効果が見られた．一方で，個々のサンプルに対する受け止め方（納得感）には個人の差があることもわかった．このことは少人数で XAI の結果を基に議論した場合には見方が偏ってしまい，誤った結論に至る危険性を孕んでいることを示している．また，XAI による着目領域の可視化については信頼性の問題も指摘^[11]されており，必ずしも正しい判断根拠が反映されているとは限らない．従って，XAI の結果から安易に受け入れを決定すべきではないと考える．XAI 導入の最大のメリットはシステム開発者も具体的な根拠をもって AI 開発者と各ケースについて議論ができる点にある．従って，あくまで議論のポイント，或いは追加の評価に必要な要素が抽出できることを重視すべきである．

本論文では，XAI の影響については定性分析に留まっており，定量的分析は今後の課題である．また，性能限界をより明確に示すために成功ケースも併せて提示すること，AI コンポーネントの改良過程において納得感を繰り返し計測すること，納得感の要素に他の AI 技術との比較による技術レベルを追加することで，調査内容を充実させることも今後の課題である．

参考文献

- [1] Fuyuki Ishikawa, et al., "How do engineers perceive difficulties in engineering of machine-learning systems? Questionnaire survey", CESSER-IP, 2019
- [2] 国立研究開発法人産業技術総合研究所, 機械学習品質マネジメントガイドライン 第 1 版, 2020
- [3] Marco T. R., et al., "' Why Should I Trust You?' Explaining the Predictions of Any Classifier", KDD, 2016
- [4] Ramprasaath R.S., et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", arXiv:1610.02391, 2016
- [5] 今井 芳江 ら, 「納得の概念分析」, 日本看護研究学会雑誌 Vol. 39 No. 2, 2016
- [6] Hiroshi Kuwajima, et al., "Improving Transparency of Deep Neural Inference Process", Journal of Experimental Psychology: General, 144 (1), 114-126, 2015
- [7] Dietvorst, B., et al., "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err", arXiv:1903.05501, 2019
- [8] J. Stallkamp, et al., "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition", Neural Networks, February 2012
- [9] Kaiming He, et al., "Deep Residual Learning for Image Recognition", arXiv:1512.03385, 2015
- [10] Aditya C., et al., "Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks", arXiv:1710.11063, 2018
- [11] Julius A. et al., "Sanity Checks for Saliency Maps", NeurIPS, 2018

付録1 アンケート調査票

あなたは、開発中の運転支援システムに採用するため、道路標識を判別するAIを選定しています。
このシステムは、AIの判定結果をアクセルやブレーキ、ハンドル操作の指示をナビゲーションシステムに表示します。

あなたのことを教えて下さい。

- (1) あなたの現在の職種を教えてください。

- (2) 現在の仕事の従事している年数を教えてください。

- (3) AI関連のプロジェクトに従事したことはありますか？

設問 1

入力画像を見て、該当する道路標識のIDを探して下さい。回答は不要です。

道路標識

ID	道路標識	意味	ID	道路標識	意味	ID	道路標識	意味	ID	道路標識	意味
0		速度制限	11		一時優先	22		路面凹凸あり	33		通過指示
1		速度制限	12		優先道路	23		スリップ注意	34		通過指示
2		速度制限	13		前方優先道路・徐行	24		右車線減少	35		通過指示
3		速度制限	14		止まれ	25		工事現場あり	36		通過指示
4		速度制限	15		車両通行禁止	26		信号機あり	37		通過指示
5		速度制限	16		3,5t以上の車両の進入禁止	27		横断歩道あり	38		通過指示
6		制限終わり	17		進入禁止	28		子供の飛び出し注意	39		通過指示
7		速度制限	18		危険	29		自転車注意	40		ロータリー
8		速度制限	19		左カーブあり	30		凍結注意	41		制限終わり
9		追い越し禁止	20		右カーブあり	31		動物横断注意	42		制限終わり
10		3,5t以上の車両の追い越し禁止	21		連続カーブあり	32		制限終わり			

No	入力画像	メモ欄
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

設問2

あなたは運転支援システムの開発者です。道路標識認識の開発部門から受け取った失敗例（下表）をみて、次の設問に答えて下さい。

（１）この認識機能を採用できるか判断する観点から納得感を４段階で答えて下さい。

（２）それぞれについて、そのように考えた理由を答えて下さい（１、２行でも結構です）。

【表の見方】「入力画像」をAIに入力し、判定した結果を「AIが判定した画像」と「（確信度）」に出力する。

「正解の画像」は、正しい画像を示している。

「確信度」は、AIが判定を行う際に算出した判定結果の確からしさを表す確率であり、高いほど確からしいことを示す。

No	入力画像	→	AIが判定した画像	（確信度）	↔	正解の画像	（１）回答欄：あなたの納得感 （４段階評価）	（２）回答欄：その理由 （自由記述）
1		→		89.72%	↔			
2		→		97.72%	↔			
3		→		50.20%	↔			
4		→		55.86%	↔			
5		→		77.73%	↔			
6		→		72.04%	↔			
7		→		70.60%	↔			
8		→		50.26%	↔			
9		→		99.87%	↔			
10		→		44.11%	↔			
11		→		97.29%	↔			
12		→		98.79%	↔			
13		→		85.31%	↔			
14		→		99.92%	↔			
15		→		51.71%	↔			

設問3

あなたは運転支援システムの開発者です。道路標識認識の開発部門から受け取った失敗例（下表）は、設問2と同じですが、AIが判定する際に注目した箇所を示す画像「XAI」を追加しました。
設問2と同様に納得感とその理由（「XAI」を一緒にみることで、納得感に変化があった場合は、特にその理由）を答えて下さい。

（1） この認識機能を採用できるか判断する観点から納得感を4段階で評価してください。

（2） それぞれについて、そのように考えた理由を答えて下さい（1, 2行でも結構です）。

【表の見方】 「入力画像」をAIに入力し、AIが判定した結果を「AIが判定した画像」と「（確信度）」さらに「（XAI）」に出力する。

「正解の画像」は、正しい画像を示している。

「XAI」の結果はAIの着目領域（赤に近いほど重視）を示しています。

No	入力画像	AIが判定した画像	（確信度）	（XAI）		正解の画像	（1）回答欄：あなたの納得感 （4段階評価）	（2）回答欄：その理由 （自由記述）
1			89.72%		↔			
2			97.72%		↔			
3			50.20%		↔			
4			55.86%		↔			
5			77.73%		↔			
6			72.04%		↔			
7			70.60%		↔			
8			50.26%		↔			
9			99.87%		↔			
10			44.11%		↔			
11			97.29%		↔			
12			98.79%		↔			
13			85.31%		↔			
14			99.92%		↔			
15			51.71%		↔			

設問4

以下の設問に答えて下さい。

- (1) 「総合的に判断してあなたはこの認識機能を採用したいですか？
(このAIの精度は、84.4%です。XAIの結果はユーザに提示しないと想定してお答えください)」

- (2) 「総合的に判断してあなたはこの認識機能を採用したいですか？(XAIの結果もユーザに提示すると想定してお答えください)」

- (3) 前項(1)(2)で、評価が変わった場合、その理由を教えてください。

アンケートは以上で終わります。ご協力に感謝申し上げます。

付録2 アンケート結果

前提 あなたは、開発中の運転支援システムに採用するため、道路標識を判別するAIを選定しています。
このシステムは、AIの判定結果をアクセルやブレーキ、ハンドル操作の指示をナビゲーションシステムに表示します。

設問1 あなたのことを教えてください。
(1) あなたの現在の職種を教えてください。
(2) 現在の仕事の従事している年数を教えてください。
(3) AI関連のプロジェクトに従事したことはありますか？

	AI関連プロジェクト経験		合計
	あり	なし	
a.システム開発者	2	10	12
b. AI開発者	7	0	7

設問2 あなたは運転支援システムの開発者です。道路標識認識の開発部門から受け取った失敗例（下表）をみて、次の設問に答えて下さい。

(1) この認識機能を採用できるか判断する観点から納得感を4段階で答えて下さい。

(4：非常に納得する，3：やや納得する，2：あまり納得しない，1：全く納得しない)

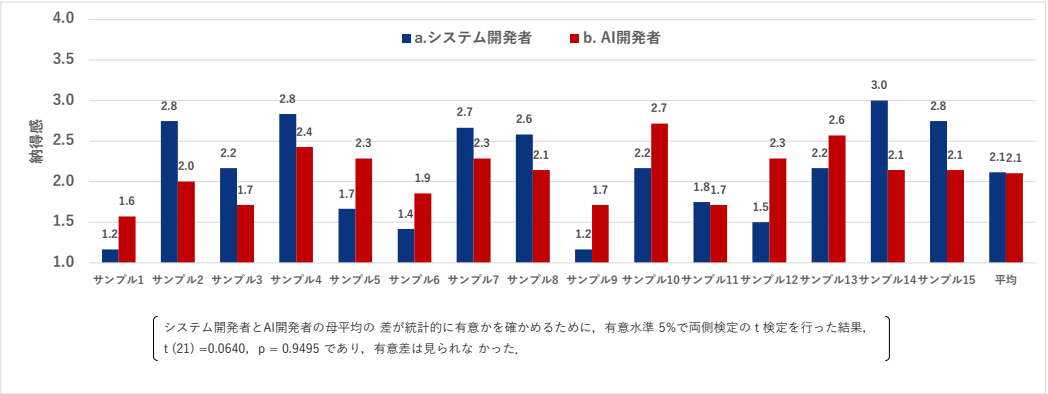


図2-1-1 サンプルごとのシステム開発者とAI開発者の納得感の平均値 (XAI導入前)

(4：非常に納得する，3：やや納得する，2：あまり納得しない，1：全く納得しない)

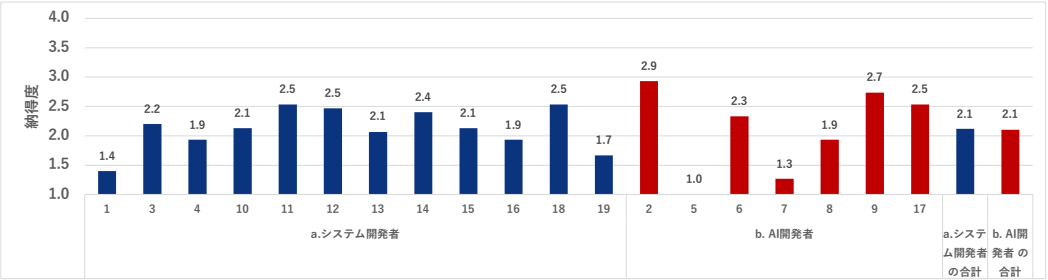


図2-1-2 回答者毎の全サンプルの納得度の平均

設問3 あなたは運転支援システムの開発者です。道路標識認識の開発部門から受け取った失敗例（下表）をみて、設問2と同様に納得感とその理由（XAIの結果と一緒にみることで、納得感に変化があった場合はその理由）を答えて下さい。

（1）この認識機能を採用できるか判断する観点から納得感を4段階で評価してください。

（4：非常に納得する，3：やや納得する，2：あまり納得しない，1：全く納得しない）

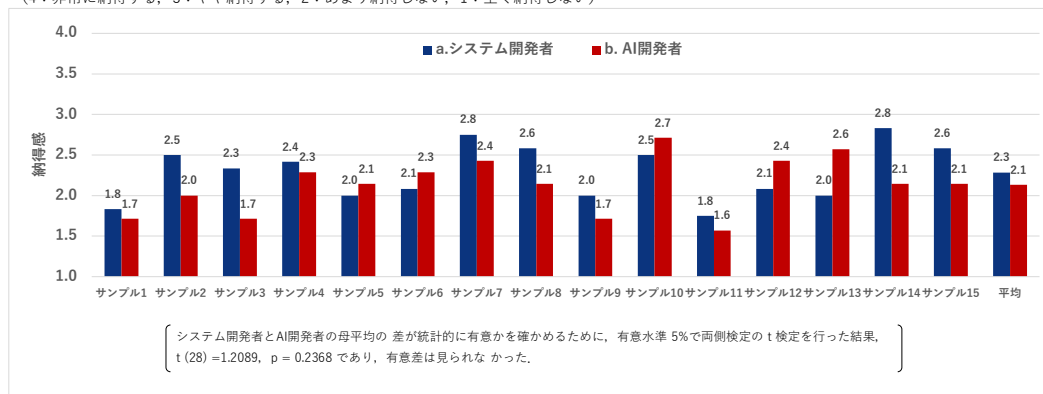


図3-1-1 サンプルごとのシステム開発者とAI開発者の納得感の平均値（XAI導入後）

（4：非常に納得する，3：やや納得する，2：あまり納得しない，1：全く納得しない）

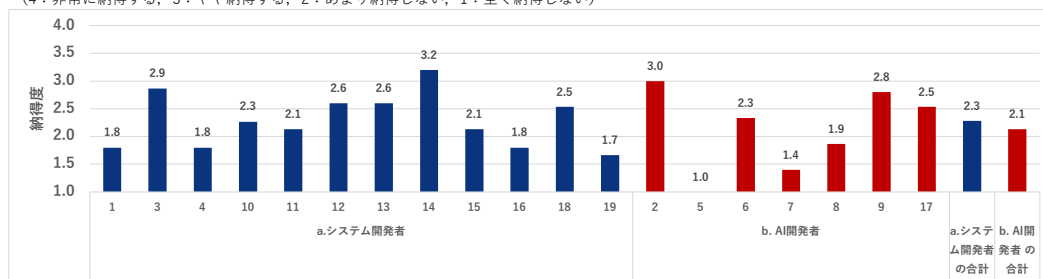


図3-1-2 回答者毎の全サンプルの納得度の平均（XAIの結果付き）

（2）それぞれについて、そのように考えた理由を答えて下さい（1，2行でも結構です）

（回答の理由を，「理解の深化／分からない／XAIの影響なし」で分類した。総数は回答者19名×15サンプル＝285）

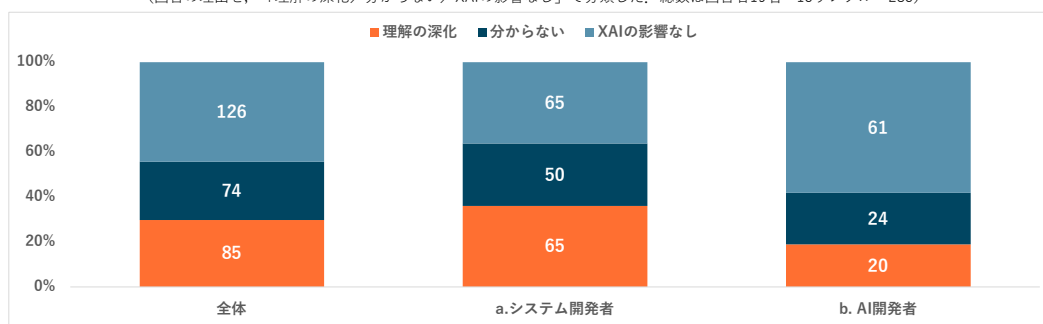


図3-2-1 サンプル毎の理由を分類

【参考1】サンプル1の納得感の変化

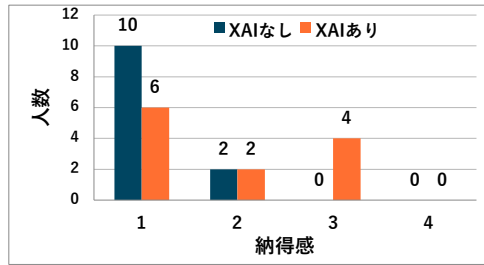


図3-3-1 サンプル1のシステム開発者の納得感

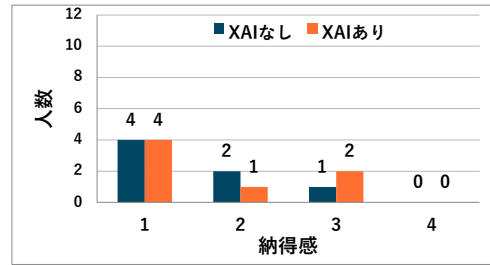


図3-3-2 サンプル1のAI開発者の納得感

【参考2】サンプル9の納得感の変化

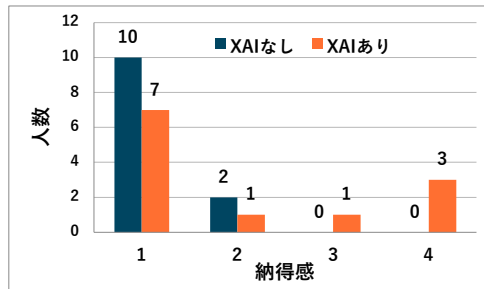


図3-3-3 サンプル9のシステム開発者の納得感

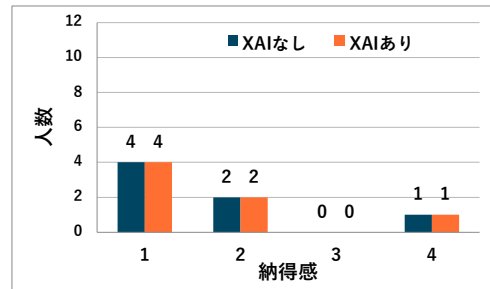


図3-3-4 サンプル9のAI開発者の納得感

【参考3】回答者毎の納得感の変化

表3-2 回答者毎の納得感の変化

サンプル	設問	システム開発者																	AI開発者						
		1	3	4	10	11	12	13	14	15	16	18	19	2	5	6	7	8	9	17					
1	XAIなし	1.0	2.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	2.0	3.0					
	XAIあり	1.0	3.0	1.0	2.0	3.0	3.0	2.0	3.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	3.0	3.0					
2	XAIなし	2.0	3.0	4.0	3.0	3.0	3.0	2.0	4.0	2.0	2.0	3.0	2.0	3.0	1.0	2.0	1.0	2.0	3.0	2.0					
	XAIあり	3.0	4.0	2.0	3.0	1.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	3.0	1.0	2.0	1.0	2.0	3.0	2.0					
3	XAIなし	1.0	2.0	1.0	2.0	2.0	1.0	4.0	3.0	3.0	2.0	1.0	4.0	3.0	1.0	2.0	1.0	2.0	1.0	2.0					
	XAIあり	3.0	2.0	1.0	2.0	3.0	1.0	4.0	2.0	3.0	2.0	1.0	4.0	3.0	1.0	2.0	1.0	2.0	1.0	2.0					
4	XAIなし	1.0	4.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	2.0	4.0	1.0	1.0	1.0	3.0	4.0	3.0					
	XAIあり	1.0	1.0	3.0	2.0	1.0	3.0	3.0	4.0	3.0	3.0	3.0	2.0	3.0	1.0	1.0	1.0	3.0	4.0	3.0					
5	XAIなし	1.0	1.0	1.0	3.0	4.0	2.0	1.0	1.0	1.0	1.0	2.0	2.0	2.0	1.0	3.0	2.0	2.0	2.0	4.0					
	XAIあり	1.0	4.0	1.0	3.0	4.0	2.0	2.0	1.0	1.0	1.0	1.0	3.0	3.0	1.0	3.0	2.0	1.0	1.0	4.0					
6	XAIなし	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	2.0	1.0	3.0	2.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0					
	XAIあり	1.0	4.0	1.0	2.0	1.0	3.0	2.0	3.0	2.0	1.0	3.0	2.0	3.0	1.0	2.0	2.0	2.0	4.0	2.0					
7	XAIなし	1.0	3.0	3.0	2.0	3.0	3.0	2.0	4.0	3.0	3.0	3.0	2.0	3.0	1.0	3.0	1.0	2.0	3.0	3.0					
	XAIあり	1.0	4.0	3.0	3.0	2.0	3.0	2.0	4.0	3.0	3.0	3.0	2.0	3.0	1.0	3.0	1.0	2.0	4.0	3.0					
8	XAIなし	1.0	3.0	3.0	2.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	1.0	4.0	1.0	1.0	1.0	3.0	3.0	2.0					
	XAIあり	2.0	2.0	3.0	2.0	2.0	3.0	3.0	4.0	3.0	3.0	3.0	1.0	3.0	1.0	1.0	1.0	3.0	4.0	2.0					
9	XAIなし	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	2.0	1.0	1.0	1.0	1.0	4.0	2.0					
	XAIあり	1.0	4.0	2.0	1.0	1.0	1.0	3.0	4.0	1.0	1.0	4.0	1.0	4.0	1.0	1.0	2.0	1.0	1.0	2.0					
10	XAIなし	1.0	3.0	2.0	2.0	4.0	2.0	1.0	2.0	2.0	2.0	3.0	2.0	3.0	1.0	3.0	3.0	3.0	3.0	3.0					
	XAIあり	3.0	3.0	2.0	3.0	1.0	3.0	2.0	3.0	2.0	3.0	3.0	2.0	3.0	1.0	3.0	3.0	3.0	3.0	3.0					
11	XAIなし	1.0	1.0	2.0	1.0	1.0	2.0	4.0	2.0	3.0	1.0	2.0	1.0	4.0	1.0	1.0	1.0	1.0	2.0	2.0					
	XAIあり	1.0	1.0	1.0	2.0	1.0	2.0	4.0	1.0	3.0	1.0	3.0	1.0	3.0	1.0	1.0	1.0	1.0	2.0	2.0					
12	XAIなし	1.0	1.0	1.0	1.0	2.0	3.0	1.0	2.0	1.0	1.0	3.0	1.0	3.0	1.0	3.0	4.0	1.0	1.0	3.0					
	XAIあり	1.0	4.0	1.0	1.0	3.0	3.0	2.0	4.0	1.0	1.0	3.0	1.0	3.0	1.0	4.0	1.0	1.0	4.0	3.0					
13	XAIなし	3.0	4.0	1.0	2.0	3.0	4.0	1.0	2.0	1.0	1.0	3.0	1.0	3.0	1.0	4.0	2.0	2.0	3.0	3.0					
	XAIあり	2.0	4.0	1.0	2.0	2.0	2.0	2.0	4.0	1.0	1.0	2.0	1.0	3.0	1.0	4.0	2.0	2.0	3.0	3.0					
14	XAIなし	3.0	2.0	2.0	3.0	4.0	4.0	4.0	3.0	3.0	4.0	3.0	1.0	3.0	1.0	4.0	1.0	2.0	2.0	2.0					
	XAIあり	3.0	1.0	3.0	3.0	4.0	4.0	4.0	4.0	3.0	1.0	3.0	1.0	3.0	1.0	4.0	1.0	2.0	2.0	2.0					
15	XAIなし	2.0	2.0	3.0	3.0	3.0	3.0	2.0	4.0	3.0	3.0	3.0	2.0	3.0	1.0	3.0	1.0	2.0	3.0	2.0					
	XAIあり	3.0	2.0	2.0	3.0	3.0	3.0	2.0	4.0	3.0	3.0	3.0	2.0	3.0	1.0	3.0	1.0	2.0	3.0	2.0					
平均	XAIなし	1.4	2.2	1.9	2.1	2.5	2.5	2.1	2.4	2.1	1.9	2.5	1.7	2.9	1.0	2.3	1.3	1.9	2.7	2.5					
	XAIあり	1.8	2.9	1.8	2.3	2.1	2.6	2.6	3.2	2.1	1.8	2.5	1.7	3.0	1.0	2.3	1.4	1.9	2.8	2.5					

〔凡例〕 : XAIありで向上 : XAIありで低下

設問4 (1) 総合的に判断してあなたはこの認識機能を採用したいですか？
 (このAIの精度は、84.4%です。XAIの結果はユーザに提示しないと想定してお答えください)

(4：採用したい，3：改良は求めるが採用したい，2：改良結果に寄っては採用したい，1：採用したくない)

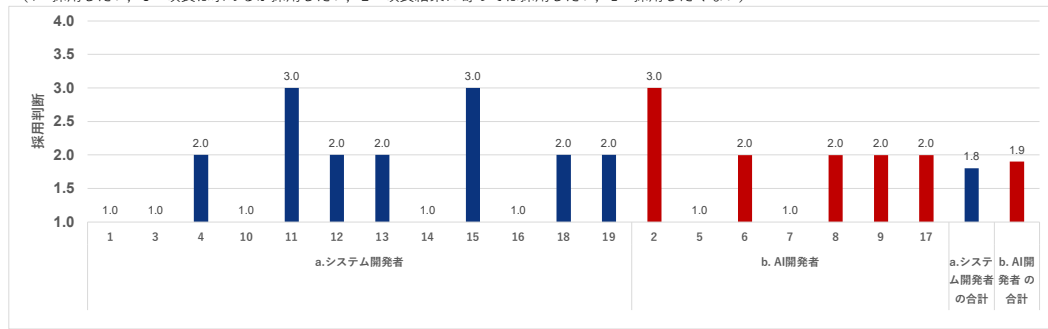


図4-1-1 回答者毎の採用判断

(2) 「総合的に判断してあなたはこの認識機能を採用したいですか？(XAIの結果もユーザに提示すると想定してお答えください)」

(4：採用したい，3：改良は求めるが採用したい，2：改良結果に寄っては採用したい，1：採用したくない)

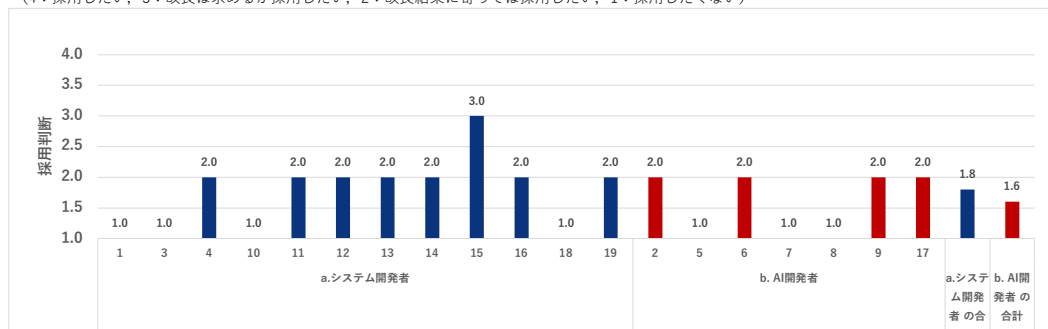


図4-2-1 回答者毎の採用判断 (XAIの結果もユーザに提示する仮定)

【参考図】設問4の(1)と(2)の変化

(4：採用したい，3：改良は求めるが採用したい，2：改良結果に寄っては採用したい，1：採用したくない)

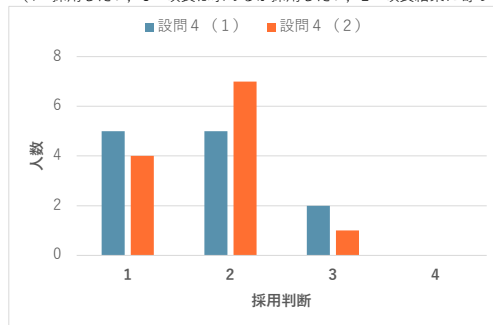


図4-3-1 システム開発者のXAIの結果を提示する場合の変化

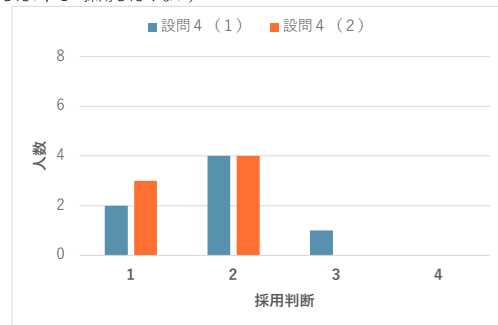


図4-3-2 AI開発者のXAIの結果を提示する場合の変化

付録3 アンケート結果_設問2と設問3の納得感の理由詳細

職種	サンプル	分類	納得感の理由	XAIあり納得感の理由	納得感の増減
a. システム開発	1	分からない	AIの結果では、意味をなさないから	XAIで△を認識できているはずなのに○と判断しているから	0
			色・形が全く異なる割には確信度が高い	重視している割には大きく異なっている	1
			色も違うし、形も違うため	中央部分に着目したのに、全く異なるものを選んでいるため	0
			全く異なる形状の標識である上に、確信度も高く算出されている。	全く異なる形状の標識である上に、確信度も高く算出されている。XAIの注目箇所を見ても、異なっている。	0
			入力画像の品質は良くないが、判定結果が形（三角と丸）まで間違えているため	AIが着目した点と判断結果には一定の納得感はあるが、形が異なる標識に誤判定している点に納得感がない	0
		理解の深化	形状は判別してほしい	形状をとらえていないことが判明した	1
			色の識別をしていない。	XAIも影を拾っている	1
			色や形などが違う標識と誤認識しているため、また確信度が低いから	XAI上では左下に伸びる直線がみえるため、影の影響を受けたと理解できる	2
			人では誤らないであろう、標識の形状を高い確信度で誤って認識しているから。	XAIの注目箇所に限って言えば、影により左下向きの矢印に見えなくもないから、納得度は高まった。	2
			輪郭ぐらいはわかってほしいと思った。色味で判断はしていないのだろうか？	XAIでの注目領域が赤い部分なら仕方ない	1
		XAIの影響なし	色も外側形状も違うことに加えて、確信度も高いため。	色も外側形状も違うことに加えて、確信度も高いため。	0
			正解の色や形状と異なるから	正解の色や形状と異なるから	0
				入力画像では数値を識別できないにもかかわらず、確信度が高い点は、納得できない	-2
				数字部分の認識に失敗していることは理解できるが、確信度が高すぎる。	0
				XAIの画像を見てもあまり特徴的な部分を抽出できていないように感じた	-2
	2	分からない	入力画像品質に対して、速度標識と正しく判定できたため		1
			数字部分の認識に失敗していることは理解できるが、確信度が高すぎる。		-2
			制限速度の標識としては同じ種類であり、また画質も粗いため、誤認識を納得できる		0
		理解の深化	速度が異なるのは困るが、入力画像の解像度ではしかたがない	XAIでは数値がわからないから	1
			同じ形状の標識ではあるが、数字の認識が誤っているにもかかわらず確信度が高く算出されている。	同じ形状の標識ではあるが、数字の認識が誤っているにもかかわらず確信度が高く算出されている。XAIの注目箇所が数字の部分であるのにもかかわらず、そこが誤っている。	0
			標識の形状は正しく認識しており、内部の文字がかすれて判別が難しいから。（このような時には確信度が下がっているはず）	標識中央の数字に着目しているのにもかかわらず、確信度が高いまま誤回答しているため納得度は低下した。	-1
			文字部分がつぶれてしまっているため。	数字全体には着目していないため	0
			よくみると3桁には見える	(空白)	0
		XAIの影響なし	画像の構成はあっている	XAIが補完していない	1
				XAIでの注目領域が赤くても元画像もボケているから誤判断も仕方ないか	0
			元画像もボケているから仕方ないか	(空白)	0
			数字文字が不明確だから	(空白)	0
			正解と色や形状は似ているものの、数字が異なるから	正解と色や形状は似ているものの、数字が異なるから	0
	3	分からない	同じ形状の標識であり、アイコンの認識は誤ってはいいるが確信度が低く算出されている。	同じ形状の標識であり、アイコンの認識は誤ってはいいるが確信度が低く算出されている。XAIの注目箇所がアイコンの部分であるのにもかかわらず、そこが誤っている。	0
			標識の形状は正しく認識しているものの、内部にマークが描かれていないことは容易に判別すべきであるから。ただし、確信度が低くAIが悩んでいることがわかるから。	標識中央に着目しているのにもかかわらず、確信度が高いまま誤回答しているから納得度は低下した。	-1
		理解の深化	無地なのに模様があると判定されるのはどうだろうか	XAIでなぜ赤い領域の中心に黒い部分がある？それが無地を有地に誤判定した結果なのか？	0
			これは識別してほしい	XAIでは○の中に何かがあると思えるから	2
			構成があっていない	XAIで補完出来ている	0
			明らかに何も書かれていないことを認識してほしいと感じたため、一方で確信度があまり高くないので、まったく納得できないわけではない	横一列な模様を抽出しているため、誤認識したと納得できたため	1
			ノイズで車の絵を判断したのだろうか	(空白)	0
	4	理解の深化	陰で誤認識していることは理解できるが、存在しない形状を誤認識してしまうのは不安がある。	陰で誤認識していることは理解できるが、存在しない形状を誤認識してしまうのは不安がある。	0
			画像からは判別しようがない	(空白)	0
		XAIの影響なし	画像上は内部の車部分が見えないため。	XAIが着目した画像内に模様が見えているが、入力画像からは判別できないため	0
			正解と全く異なる判定だったため	正解と全く異なるため	0
			目視でも間違えるから	目視でも間違えるから	0
	5	分からない	確信度も低いいため誤認識と判断しても仕方ないと感じた	5の一部が白飛びしているが、AIなら補完して認識してほしいと思った	-2
			同じ形状の標識であり、数字の認識は誤ってはいいるが、確信度が低く算出されている。	同じ形状の標識であり、数字の認識は誤ってはいいるが、確信度が低く算出されている。XAIの注目箇所が数字の部分であるのにもかかわらず、そこが誤っている。	0
			7と判断してもおかしくはない	XAIだと5に見えてきた	-1
			これは識別してほしい	XAIでは、50が認識できるから	0
			画像上5の下部分が削れていて、7に見えなくはないため。	数字全体には着目していないため	0
		理解の深化	元画像もボケているから仕方ないか	XAIでの注目領域が赤くても元画像もボケているから誤判断も仕方ないか	0
			標識の形状は正しく認識しているものの、5と7の判別ができそうであるから。確信度が低くAIが悩んでいることがわかるから。	標識中央の数字に着目して、確信度が低いいため、標識の数字でAIが悩んでいることが想定できるため。	1
			2桁であることは判別できているため	(空白)	0
			数字部分の認識に失敗していることは理解できる。確信度も低いので問題にならないか？5の真ん中の形状が認識されていないのはやや違和感がある。	数字部分の認識に失敗していることは理解できる。確信度も低いのはやや違和感がある。	0
			正解と色や形状は似ているものの、数字が異なるから	正解と色や形状は似ているものの、数字が異なるから	0
	XAIの影響なし	理解の深化	速度標識だと判定できているため	速度標識だという点は識別できているため	0
			読み込み画像が曖昧すぎる	読み込み、XAI共に問題無し	-3

職種	サンプル	分類	納得感の理由	XAIあり納得感の理由	納得感の増減
a. システム開発	5	分からない	異なる形状の標識である上に、確信度もやや高めに算出されている。	異なる形状の標識である上に、確信度もやや高めに算出されている。XAIの注目箇所がアイコンの部分であるのにもかかわらず、そこが誤っている。	0
			人では誤らないであろう、標識の形状を誤って認識しているから。	標識の形状にも着目しているのに誤回答しているから。	0
		理解の深化	標識の形状が違うため。	着目領域が円形だが、判定結果は三角の標識を選んでいるため	-1
			輪郭ぐらいはわかってほしいと思った	XAIの注目領域なら丸い標識で判定されてもいいとおもった	0
			○を△と判定するのは・・・	XAIでは○が認識できるのに、結果が△だから	0
	6	XAIの影響なし	画質が粗く、明るさも一定でないため誤認識を納得できた	標識のマークとXAIのマークが似ていると感じたため	0
			形が違う	数字が読み取れていない	3
			形状は判別してほしい	形状をとらえていないことが判明した	1
			正解と比べると色味しか合っていないから	標識内の記号だけに着目すると、目視でも間違えそうだから	1
			タイヤ痕が似ている	(空白)	0
	6	分からない	外側形状が異なるものを一致判定することは違和感がある。	外側形状が異なるものを一致判定することは違和感がある。	0
			標識の形まで誤っているため	確信度が70%以上あるにもかかわらず、標識の形すら誤っているため	0
			○を▽に識別するのは・・・	XAIでは、▽には見えないから	0
			丸と三角では大きく異なる	縦の棒が強く重視しているならば、三角としてとらえても仕方ないと思ったが、そうではなかった	0
			中に何も無い標識という意味では同じだが、異なる形状の標識である上に、確信度もやや高めに算出されている。	中に何も無い標識という意味では同じだが、異なる形状の標識である上に、確信度もやや高めに算出されている。XAIの注目箇所が空白の部分でありそこは同じではあるが、そもそも形状が異なる。	0
			標識の全体像が見えていないため	着目している部分から逆三角の形状を選んだことが理解できないため	0
			明確に丸い形であり、画像的にも一様な明るさなので、認識してほしいと感じた	丸い領域が薄い青色であり、三角形の標識と誤認識するとは納得できない	0
			形状は判別してほしい	形状をとらえていないことが判明した	1
				標識の形状に着目せず、中央部の白い部分のみ着目しているため誤回答しているため妥当な判断であると若干感じたから。また、柱に注目しておらず、オクルージョンを考慮せずに判断していると感じたから。	2
		XAIの影響なし	人では誤らないであろう、標識の形状を誤って認識しているから。	XAIでの注目領域の赤い部分の右端の赤ラインを拾ったなら仕方ない	2
			輪郭ぐらいはわかってほしいと思った	外側形状が異なるものを一致判定することは違和感がある。	0
			外側形状が異なるものを一致判定することは違和感がある。	問題無し	3
			形が違う	正解と比べると形は合っているから	0
			正解と比べると形は合っているから	確信度が70%以上あるにもかかわらず、標識の形すら誤っているため	0
	7	分からない	標識の形を誤っているため	標識の上半分も注目すると、正しく認識できたと思うため、そこまで納得しない	-1
			3と5で形としては似ているため、ある程度は納得できる	XAIで、30が見えているから	0
			数字が違うのは困る。これで確信度70%というのも。	同じ形状の標識であり、数字の認識が誤ってはいるが、確信度もやや低めに算出されている。XAIの注目箇所が数字の部分であるのにもかかわらず、そこが誤っている。	0
			同じ形状の標識であり、数字の認識が誤ってはいるが、確信度もやや低めに算出されている。	3の上の方も強く重視しなければ5になってしまうと思った。	1
			3だと思う	数字全体には着目していないため	0
		理解の深化	5の一部が切れているように見えなくはないため。	数字部分の認識に失敗していることは理解できる。XAI上は数字の下半分を重視しているのが納得度はやや高まったが、単純に学習が足りていないように見える。確信度も低いので問題にならないか？	0
			数字部分の認識に失敗していることは理解できる。確信度も低いので問題にならないか？	標識中央の数字に着目して、確信度が低いため、標識の数字でAIが悩んでいることが想定できるため。	0
			標識の形状は正しく認識しており、差異の少ない5と3の判別で間違えているから。確信度も70%程度とAIが悩んでいることがわかるから。	(空白)	0
			画像から判別できそうと判断したため	XAIでの注目領域が赤くても元画像もボケているから誤判断も仕方ないか	0
			元画像もボケているから仕方ないか	速度標識である点は識別できているため	0
		XAIの影響なし	速度標識だという点は判定できているため	問題無し	1
			読み込み画像の解像度が上がれば判定できるはず	目視では間違えそうにないから	0
			目視では間違えそうにないから	数字の全体を注目しているように見えるため、誤認識することには納得できない	-1
			手前のきわめて明るい領域があるため誤認識したと考えられるため		
	8	分からない			
			一部がオクルージョンで隠されているため難しい問題設定ではあるのにもかかわらず標識の形状は正しく認識しているものの、人では容易に判別できそうな8を2と誤って認識しているから。また、確信度が低くAIが悩んでいることがわかるから。数字が違うのは困る。確信度は50%か、自信はないんだろうけど。	標識の形状に着目せず、中央部の白い部分のみ着目しているため誤回答しているため妥当な判断であると若干感じたから。	1
			数字部分の認識に失敗していることは理解できる。確信度も低いので問題にならないか？	XAIで、見づらいけど 20には見えないから	1
			数字部分の認識に失敗していることは理解できる。確信度も低いので問題にならないか？	数字部分の認識に失敗していることは理解できる。確信度も低いので問題にならないか？	0
			同じ形状の標識であり、数字の認識が誤ってはいるが、確信度が低く算出されている。	同じ形状の標識であり、数字の認識が誤ってはいるが、確信度が低く算出されている。XAIの注目箇所が数字の部分であるのにもかかわらず、そこが誤っている。	0
		XAIの影響なし	味方によっては2に見えるかなと思うため。	数字全体には着目していないため	0
			2桁であることは判別できているため	(空白)	0
			8に見えるが、斜めの線が強く見えるかもしれない	(空白)	0
			AIは数字判定が苦手の様。	数字読み取り精度が低い	-1
				XAIでの注目領域が赤くても元画像もボケているから誤判断も仕方ないか	0
			元画像もボケているから仕方ないか	速度標識である点は識別できているため	0
			速度標識だと判定できているため	目視では間違えそうにないから	0
			目視では間違えそうにないから		

職種	サンプル	分類	納得感の理由	XAIあり納得感の理由	納得感の増減
a. システム開発	9	分からない	右半分だけを見れば確かにAI判定の画像に見えるが、確信度が高すぎる	XAIが的外れな重視をして、確信度が高すぎる	-1
			外側形状が異なるものを一致判定することは違和感がある。にもかかわらず確信度が高すぎる。	XAI上は認識対象とは別の部分に着目しており、不安がある。にもかかわらず確信度が高すぎる。	0
			確信度が高いにもかかわらず、標識の形すら誤っているため誤認識にしては確信度が非常に高いため	XAIがズレている点に納得しない 全く違う領域から判断しているため	1 0
			全く異なる形状の標識である上に、確信度も高く算出されている。輪郭ぐらいはわかってほしいと思った	全く異なる形状の標識である上に、確信度も高く算出されている。XAIの注目箇所は標識でないところになってしまっている。なぜ注目領域がずれているのかわからない	0 0
			形状、マーク共に明らかに不正解。かつ高い確信度で間違えているから。形状は判別してほしい	そもそも注目箇所が誤っており確信度も高いため、コントラストが低いデータが足りないのだろうと推測できるから。（おそらくこれらのデータは要件外なのだろう） 認識対象をとらえていないことがわかった	3 2
	10	分からない	標識の形が違うため。AIに入力した画像が悪すぎる	着目領域が標識とは別の場所を標識として判定しているため。読み取り、XAI共に悪い	2 3
			これは難しいかな。でも確信度 99.87%って。全然違うじゃん。正解の色や形状と異なるから	XAIの右側のカラーはノイズかな？ △には見えなから 正解の色や形状と異なるから	0 0
			確信度が50%前後あるなら、標識の形は間違わないでほしい	形を誤る点は納得しない 全く異なる形状の標識ではあるが、確信度が低く算出されている。XAIの注目箇所は標識全体になっているが、形状が異なっていて判定されている。	0 0
			全く異なる形状の標識ではあるが、確信度が低く算出されている。		0
			△を○に識別しているので	XAI だと、正解を判定するのは難しい。確信度も44.11%だし。	2
			外側形状が異なるものを一致判定することは違和感がある。形状は判別してほしい 人では誤らないであろう、標識の形状を誤って認識しているから。ただし、確信度が低くAIが悩んでいることがわかるため、1ではなく2と判断した。	XAI上は外側形状が同じように見えているため、納得はできる。形状をとらえていないことが判明した 標識に注目しているものの、注目箇所が標識に対して広がっており正しく認識できていないことが窺える。コントラストが低いことが原因と推測できるから。	1 1 1
			標識の一部に明るい部分があり、*のマークもあまり見えないため	全く違う領域から判断しているため	-3
			標識の形状が違うが、確信度が低いため。	着目領域の形状が三角よりも範囲が広がっているため。	0
			輪郭ぐらいはわかってほしいと思った	XAIでの注目領域が赤い部分なら仕方ない	1
			AIに入力した画像が悪すぎる	読み取り、XAI共に悪い	0
	11	分からない	確信度が低いが、中身はともかく三角だと思う	(空白)	1
			正解の形状と異なるから	正解の形状と異なるから	0
			鏡像画像で判断したならわかるが、この模様で向きを間違えるか？	なぜこの注目領域で逆向きに判定されるのかわからない 標識の中央部、右か左かに着目しているように見えるのに、誤回答しているから。	0 -1
			形状は一致、右か左かのわずかな特徴で誤回答しているものの、確信度が高いため。	XAIでも右向きが判別できるから	0
			向きはわかるでしょ。確信度 97.29%？ 入力画像の品質が、矢印の向きを誤るほど悪いものではないと思うため	XAIは正しい部分をフォーカスしているが、判定結果の矢印の向きが逆になっているため	-1
			標識の中身の形が反対ではあるが、見間違えるような違いではなく、確信度も高いため	標識の全体に注目しているのに、誤認識しているため	0
			回転でもさせたのだろうか	XAIでも右曲がりに見える	1
			同じ形状の標識ではあるが、アイコンの認識が誤っているにもかかわらず確信度が高く算出されている。ただ、アイコン自体も似た標識ではある。	同じ形状の標識ではあるが、アイコンの認識が誤っているにもかかわらず確信度が高く算出されている。ただ、アイコン自体も似た標識ではある。XAIの注目箇所はアイコンの部分になっているが、そこが誤っている。	0 1
			標識として左右両方があるのに識別できていないため	標識内のマークがない場所に注目しているため	0
			画像の読み取り精度が低い	XAI補完が参照できてない	0
	12	分からない	特徴的な形状を認識できてない上に、確信度が高すぎる。非常に類似しているものを選択できている	特徴的な形状を認識できてない上に、確信度が高すぎる。(空白)	0 0
			目視では間違えそうにないから	目視では間違えそうにないから	0
			形がぜんぜん違うじゃん。確信度 98.79%って。	XAIに▽の要素はないから	0
			全く異なる形状の標識である上に、確信度も高く設定されている。	全く異なる形状の標識である上に、確信度も高く設定されている。XAIの注目箇所は標識全体になっているが、形状が異なっていて判定されている。	0
			形状は判別してほしい	形状をとらえていないことが判明した	1
			多少の輪郭はあっているから誤判断も仕方ないと思った	XAIでの注目領域が赤い部分なら仕方ない	0
			大部分が隠れているため、誤認識は理解できるが、確信度が高すぎると感じた	XAI上では標識の形の抽出が難しかったのかと、理解できるため	1
			入力画像のコントラストが低く、判別が難しい画像ではあるものの、高い確信度で誤回答しているから。	着目された領域を見ると、木のエッジを含めた三角形を基に誤回答していると読み取れるから、納得感は高まった。	2
			標識の全体が正しく見えていないため。	着目領域が標識とは異なる部分のため	0
			この読み込み画像では無理	XAI補完が有効になっている	3
	13	分からない	確信度が高いにもかかわらず、標識の形すら誤っているため	形を誤る点は納得できない	0
			色が異なるし、確信度が高すぎる	(空白)	0
			正解の色や形状と異なるから	正解の色や形状と異なるから	0
			特徴的な形状を認識できてない上に、確信度が高すぎる。	特徴的な形状を認識できてない上に、確信度が高すぎる。	0
			ノイズが多いので、難しいと思う。	XAI では何かわからないから。でも確信度85.31%なので。	-1
			画像からどこが標識なのか判断が付きにくいから。	着目領域が2台の車に見えなからひし形の標識が判定されてもいいと思った	-1 -2
			元画像もボケているから誤判断は納得	全く異なる形状の標識である上に、確信度も高く設定されている。XAIの注目箇所は標識全体になっているが、形状が異なっていて判定されている。	0
			全く異なる形状の標識である上に、確信度も高く設定されている。		0
			背景と標識が非常に似ており、判断が難しいと感じた	標識として注目すべき点に注目できていないと思ったため	-1
			形状は判別してほしい	形状をとらえていないことが判明した	1
			入力画像のコントラストが低く、判別が難しい画像だから。マークの有無は判別できそうであるから。4ではなく3とした。	標識に注目しているものの、注目箇所が標識に対して広がっており正しく認識できていないことが窺える。コントラストが低いことが原因と推測できるから。	2

職種	サンプル	分類	納得感の理由	XAIあり納得感の理由	納得感の増減
a. システム開発	13	XAIの影響なし	この読み取り画像では無理	読み取りが悪い	0
			確信度が高すぎる	(空白)	0
			色も外側形状も違うことに加えて、確信度も高いため。	色も外側形状も違うことに加えて、確信度も高いため。	0
			人間が見ても識別困難か入力画像に対して、結果が誤っているにも関わらず、確信度が高いため	確信度が高いにもかかわらず結果が全く異なる判定であるため	0
			正解の色や形状と異なるから	正解の色や形状と異なるから	0
				XAIの結果では形状を正しく認識出来てそうに見えるが、誤認識してしまうのは不安がある。	-3
		分らない	人間でもばっと見て間違えそうな内容のため納得できる。	標識内の記号に着目しているにも関わらず間違っているから	0
			目視では間違えそうにないから	模様がない部分に着目しているため	0
			理解の深化		
			画像上の色がついているように見えるため。		
	14	XAIの影響なし	確信度がほぼ100%にもかかわらず、誤った識別をしているため	入力画像、XAIを考慮すると、判定結果はやや納得できる	1
			形状、マークの並びとも一致している誤回答であり、人でも誤りそうであるから。	標識に注目しているものの、注目箇所が薄く、AIがどこに着目すべきか悩んでいるようだから。	1
			人間でも間違えそう	XAIの画像で、3個の点か、！かは判断が難しいから	0
				同じ形状の標識ではあるが、アイコンの認識が誤っているにもかかわらず確信度が高く算出されている。ただ、アイコン自体も似た標識ではある。XAIの注目箇所はアイコンの部分になっているが、そこが誤っている。	0
			画像読み取り精度が低い	XAI補完が参照されていない	-1
			間違いかねないと思った	XAIの注目領域を見ても間違いかねないと思った	0
		理解の深化	似ていると思う	直接関係ないが、標識の形は重視していないのだろうか。単に真ん中に3つの丸っぽい領域があるからそのような判断に至ったのだろうか	0
			非常に類似しているものを選択できている	(空白)	0
			標識の形として非常に似ており、ノイズの影響で三色の信号と認識するのを理解できたため	非常に標識が似ているため	0
			画像読み取り精度が低い。特に数字	読み取り、XAI共数字の精度が悪い	0
				数字全体に着目していて、数字の数が識別できそうに感じたため	-1
				標識内の記号に着目しているにも関わらず間違っているから	-1
b. AI開発者	15	XAIの影響なし	数字がつぶれているため。	XAIを見る限り、数値3桁を判別できない点が残念だと思う	-1
			正解と色や形状は似ているものの、数字が異なるから		-1
			入力画像品質と、AIの確信度から納得した		-1
			確信度があまり高くなく、画像も明瞭でないため、納得できる誤認識である		
			左側に目線（1の部分）があるので、間違えないで欲しい	標識の外形などは一致しているため	0
				XAIの画像で、数値の判断は難しいから	1
		理解の深化	同じ形状の標識であり、数字の認識は誤ってはいるが確信度が低く算出されている。	同じ形状の標識であり、数字の認識は誤ってはいるが確信度が低く算出されている。XAIの注目箇所は数字の部分になっているが、そこが誤っている。	0
			入力画像のコントラストが低く、標識の形状は正しく認識している。確信度が低くAIが悩んでいることがわかるから。	標識中央の数字に着目して、確信度が低いため、標識の数字でAIが悩んでいることが想定できるため。	0
			よくみると3桁には見える	(空白)	0
			確信度が低いため	(空白)	0
				XAIでの注目領域が赤くても元画像もボケているから誤判断も仕方ないか	0
			元画像もボケているから仕方ないか	数字部分の認識に失敗していることは理解できる。確信度も低いので問題にならないか？	0
	1	分らない	数字部分の認識に失敗していることは理解できる。確信度も低いので問題にならないか？	数字部分の認識に失敗していることは理解できる。確信度も低いので問題にならないか？	0
			AIの判定は、色も形状も何一つ合っていない。	着目領域と判定結果の関連性がよくわからない。	0
			別の三角形の標識として判定するならわかるが、形も色もかけ離れている標識として判定しているため。確信度が比較的高い点も気になる。人間でも判定が難しそうなので間違えたこと自体の納得感はある。		
			誤答であることと、色形状が正解と異なっているため	影と標識の絵が合わさって矢印のように見えている箇所に反応していることが分かるため。	1
				XAIの結果をみても納得感に変化はない	0
				どれも納得感にはあまり影響しなかった（回答はすべて設問2と同じです）。アテンションが間違っていれば「なんでこんな所に注目してるんだ」、アテンションが正しければ「ちゃんと見ているのになぜ間違えるんだ」となるだけで、認識結果が間違っていることに変わりはないし、それで納得したりしなかったりには関係ないと思う。特に（技術を知らない）ユーザーはそうなるのではないか	0
		理解の深化	色も形もまったく違っていて納得できない。もっとも正解とも色が違うので（退色してる？）その点の難しさは認める	XAIの結果は何も影響されない	0
			中のアイコンが矢印に見えなくもない	結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点（標識の内側）であり、注目点を提示してもらっても差が感じられない。	0
				特に印象の変化なし	0
			認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。	着目領域は悪くなさそう（3桁目のところの強度が低い？）。なぜこの判定結果になるのかよくわからない。	0
			枠の形状も全体の色も違うのに高い確信度を出しているため		
			にまづい。	XAIで標識の丸い形状を判定に使っているとわかるため	0
	2	分らない	色、形状が正解と同じであり、また入力データが不鮮明であるため誤判定はしょうがないと思うため		
			文字のテキストチャが潰れており判定が難しそうな画像に対して、別の速度制限の標識としており、間違え方に納得感があるため。確信度がかなり高い点が気になる。	数字を部分的にしか見れていないことが推測できるため。	0
			2桁と3桁で違うので。	XAIの結果は何も影響されない	0
			数字が潰れて読みにくいが、3桁であることくらいは判別出来てほしい。確信度が高すぎる	(空白)	0
				結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点（標識の内側）であり、注目点を提示してもらっても差が感じられない。	0
			認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。		
		理解の深化	文字がつぶれていて人の目でも識別が困難（おそらく3桁だが100か120かわからない）状況で、高い確信度を出しているため	特に印象の変化なし	0

職種	サンプル	分類	納得感の理由	XAIあり納得感の理由	納得感の増減
b. AI開発者	3	分らない	簡単そうな画像なのに間違えているの、なぜこのように間違えたのか推測できないため。	XAIと合わせて見ても判断根拠が分からないため。	0
			誤識別するような難しい問題には見えない。	着目領域と判定結果の関連性がよくわからない。	0
		理解の深化	色、形状が同じであり、入力データが不鮮明であり誤判定もしようがないと思うため	XAIで標識の丸い形状を判定に使っているとわかるため	0
		XAIの影響なし	何もないように見えるので。	XAIの結果は何も影響されない	0
	4	分らない	何も書いてないことは明らかだが、確信度が低いので理解は出来る	(空白)	0
			形状や主な色は一致しているが、内部が白一色(赤や黒を含まない)の画像であることを考えるともう少し確信度が低くなって欲しい	特に印象の変化なし	0
		理解の深化	認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。	結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
			誤識別するような難しい問題には見えない。70には見えない。	着目領域と判定結果の関連性がよくわからない。	0
	5	分らない	色、形状が同じであることと、入力データの5が不鮮明で7のようにみえても不思議はないため	XAIで標識の丸い形状を判定に使っているとわかるため。文字の誤判定はしようがないと思うため	-1
		理解の深化	入力画像は「70」に見えなくもない。MNISTのひどい癖字を間違えるのと似たような感じがして、間違え方に納得感がある。		
			「70」に対する確信度が低いため、もしかしたら「50」に対する確信度も最大ではないだけで比較的高いのかも。	数字を部分的にしか見れていないことが推測できるため。	0
		XAIの影響なし	5のカスレが、7にも見える	XAIの結果は何も影響されない	0
	6	分らない	画像自体白飛び気味の状況で、形状や色、桁数などが一致しておりある程度低めの確信度を出しているため	特に印象の変化なし	0
			間違え方よりも50の確信度がこれより低いことが問題に思う	(空白)	0
		理解の深化	認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。	結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
			テクスチャが潰れているため判定が難しいのはわかるが、なぜ「70」と間違えたのかの推測が難しいため。	XAIと合わせて見ても判断根拠が分からないため。	-1
	7	分らない	画質が悪く、内部の数字がつぶれて少し模様みtaiになっているとはいえ、形状が違うのに割と高い確信度を出しているため	注目箇所が円形の標識全体に出ている状態で、三角の標識と間違えるのは納得感が下がる	-1
		理解の深化	入力画像は崩れすぎていて視認不能なので、AIの判定が間違っているのも仕方ないと思うが、確信度が高すぎるのがよくない。	着目領域と判定結果の関連性がよくわからない。	0
		XAIの影響なし	形が違う標識であるため。ただし、確信度はそこまで高くないため、しようがないと思う部分もある。	XAIで標識の中心部を判定に使っているとわかり、かつ入力データが不鮮明なため誤判定された標識に見えなくもないため	1
			そう見える	XAIの結果は何も影響されない	0
	8	分らない	人でも判別困難なので仕方ない。ただ形くらいは当ててほしいと思った	(空白)	0
		理解の深化	認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。	結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
			赤線ははっきりと表れている状況で形状が違うのに割と高い確信度を出しているため	背景の白飛び部分も含めて白い直線(逆三角形の上辺辺)と判断したのかもしれないが、赤線がはっきり出ていることを考えるとあまり納得はできない	0
		XAIの影響なし	丸い標識を逆三角の標識として認識しており、隠れによる影響であることを推測できるため。	標識の形状をそもそも見れていないことが推測できるため。	1
	9	分らない	形状が異なるものに誤判定しているため	XAIにより標識全体を判定に使えていないとわかり、誤判定をしてもしようがないと思うため	1
		理解の深化	誤識別するような難しい問題には見えない。	AIは中央付近だけ見えていて、丸か三角かを見ていないので、間違えるべくして、間違えたたとわかる。	1
			よくありそうな「AIっぽい間違い」だが印象は悪い。このAIは形を重視してない？	(空白)	0
		XAIの影響なし	認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。	結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
	10	分らない	明らかに丸い看板なので誤識別するような難しい問題には見えない。50には見えない。	XAIの結果は何も影響されない	0
		理解の深化	同じ速度制限の標識と間違えているため。	数字の下の方を見すぎている？なぜそうなのか納得はできない。	0
			同種の標識に判定しているのと、3、5はよく間違えることがあるため	数字を部分的にしか見れていないことが推測できるため。	1
		XAIの影響なし	5は3に見えてもおかしくはないので	XAIで標識の丸い形状を判定に使っているとわかるため。文字の誤判定はしようがないと思うため	0
	11	分らない	70と50よりは字形が似ているのでやや納得度は高い	XAIの結果は何も影響されない	0
		理解の深化	形状や色、桁数などは一致しているが、数字が分かりそうな画像をある程度高い確信度で間違えているため	(空白)	0
			認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。	特に印象の変化なし	0
		XAIの影響なし		結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0

職種	サンプル	分類	納得感の理由	XAIあり納得感の理由	納得感の増減
b. AI開発者	8	分からない	誤認識するような難しい問題には見えない。20には見えない。	着目領域と判定結果の関連性がよくわからない。 数字を部分的にしか見れていないことが推測できるため。 「20」と「80」で共通している部分に強く反応していることが分かるため。	0
		理解の深化	同じ速度制限の標識と間違えているため。 同種の標識に識別しているのと、確信度が低くAI自身も疑問を持っているように見えるため	XAIで標識の丸い形状を判定に使っているとわかるため、文字の誤判定はしようがないと思うため	1
		XAIの影響なし	くっきりと8に読めるので	XAIの結果は何も影響されない	-1
			暗いとはいえ、これは当てなければならない画質だと思う	(空白)	0
				結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
			認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。		0
			標識の部分はコントラストが低い状況で、形状や色、桁数などが一致しておりある程度低めの確信度を出しているため。ただできれば正解して欲しいレベル。	特に印象の変化なし	0
				標識部分に反応していないことが分かるが、何もない領域から高い確信度で標識を推定している原因が不明。敵対的擾動などに弱いのではないかと思ってしまう。	-3
		9	分からない		
		理解の深化	ラベルのミスと思われるため。		
			画像自体全体的に暗いので難しさはあるが、間違えるときの確信度が高すぎる	注目箇所が全般的な外れな場所で、その場所に標識らしい形状もない状況で、高い確信度を出しているので余計に納得できない別の対象物を着目している。間違えた理由はよくわかるがあまり納得はできない。入力画像が余白多すぎる？学習がおかしい？	0
			入力画像は見にくいですが、AIの判定は似てなさすぎる。確信度が高すぎる。		1
			もともとの画像が不鮮明なので誤判定は仕方ないと思うが、確信度がとても高いため	XAIによりそもそも間違った場所を判定に使っているとわかるため	2
		XAIの影響なし	画質が悪いので当てられないことは仕方ないと思うが、形が全然違うのと確信度が高すぎて印象が非常に悪い	(空白)	0
			見えづらいが斜線が入っているのは確実に見えるので	XAIの結果は何も影響されない	0
				結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
			認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。		0
	10	分からない	光の反射 + ノイズで難しいのと、確信度がかなり低いため。	XAIと合わせて見ても判断根拠が分からないため。	0
		理解の深化	もともとの画像がとても不鮮明であり、かつ確信度も低い	XAIで標識の丸い形状と捉えてしまったのかなとわかるため	0
		XAIの影響なし	8には見えないが、確かに見えづらいので。	XAIの結果は何も影響されない	0
			このくらいは出来てほしいと思うが、確信度が低いので多少納得。		0
				間違ってるが、まあまあ納得できる。	0
			認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。	結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
			反射の影響や画質の悪さもあるが難しさがあり、確信度も低い。ただ、形状が違うのでもう少し確信度が低くなって欲しい		0
			非常に難しい例だと思うのでやむなし。ただやっぱり形が当てられていない	特に印象の変化なし	0
				(空白)	0
		11	分からない		
		理解の深化	簡単に思われる画像であるにもかかわらず、高い確信度で間違えているため。左右を間違えているのが特に気になる。他の方向系の標識も間違えてしまうのではないかと思ってしまう。	XAIと合わせて見ても判断根拠が分からないため。	0
			正解と判定結果が、真逆の関係になっていて非常にまずい。確信度が高すぎる。	着目領域は間違っていないように見える。そもそもの実力？	0
		XAIの影響なし	Augmentationでflipを入れた結果、学習時にGTミスを含んでいたなら納得はするが学習方法に対する不信感が出る。そうでなければ納得しない	特に印象の変化なし	0
			この間違い方は用途次第では致命的。確信度も高いので強い不安と不信感を感じる	(空白)	0
			はっきりと絵が見えるし、右に曲がっているように見える	XAIの結果は何も影響されない	0
			同種の標識に識別していることと、画像認識では画像の複製のために元データを回転させて学習を行うことがあるため、誤判定してしまうのではないと思うため		-1
				XAIによって納得感が変わらなかった	
			結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。		0
	12	分からない	認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。	着目領域は間違っていないように見える。画像からは判定できないので、確信度はなんとかしてほしい。	0
		理解の深化	オクルージョンを加味しても、誤認証するような難しい問題には見えない。確信度が高すぎる。	逆三角形とひし形に共通している形状に反応していることが分かるため。	0
			照明変動に弱いアルゴリズムであると推測できるため。		1
			木によって遮られた三角の形状を捉えてると考えられるので、誤判定もしようがないと思うため	XAIによって木によってが遮られた三角部分を捉えているのかなと思うため	0
		XAIの影響なし	もともと薄い色合いに逆光で白飛びしているので極めて難しいと思う。何に間違えても致し方ないと思う	(空白)	0
			隠れや白飛びなど難しさはあるが、形状も色も違うのに対して、確信度が高すぎる	特に印象の変化なし	0
			黄色い口がみえるが、確かに見えづらいので。	XAIの結果は何も影響されない	0
				結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
			認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。		0

職種	サンプル	分類	納得感の理由	XAIあり納得感の理由	納得感の増減
b. AI開発者	13	分からない	照明変動に弱いアルゴリズムであると推測できるため。	XAIと合わせて見ても判断根拠が分からないため。	0
			入力画像が見にくいので、間違っても仕方ない部分はあるが、確信度が高すぎる。	着目領域は間違っていないように見える。画像からは判定できないので、確信度はなんとかしてほしい。	0
			12番と同じ。12番よりさらに難しいと思う。(分類でなく) 標識検出の問題なら人間でも困難ではないか	(空白)	0
			四角い看板であると思うが、確かに見えずらいので	XAIの結果は何も影響されない	0
			全体的に色味がない影響で間違えているのかもしれないが、もっと確信度が下がって欲しい		
			入力データが不鮮明すぎるため	特に印象の変化なし	0
		XAIの影響なし		入力データが不鮮明なため誤判定はしょうがないため	0
				結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
				結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
				着目領域は間違っていないように見える。そもそもの実力？	0
	14	分からない	認識結果の使い道にもよると思うが、これくらい認識できないのなら何もしてくれないほうが嬉しい気がする。	XAIと合わせて見ても判断根拠が分からないため。	0
			誤識別するような難しい問題には見えない。	XAIにより標識の中の文字に着目していることがわかり、かつそれが似ている標識であるため	0
			低解像度だと形状が似ていると思われるため。確信度が高すぎる点が気になる。		
			両者がとても似ている画像のため誤判定も仕方がないと思うため		
			いかにもありそうな間違い。中のマークに色しみも出ているので相当難しいと思う。	(空白)	0
		理解の深化		結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0
				XAIの結果は何も影響されない	0
	15	分からない	枠の形状が同じでかつ真ん中の！の部分に画質の影響で色味が付いてしまっている可能性がありそうだが、確信度が高すぎる	特に印象の変化なし	0
			低い確信度で同じ速度制限系の標識と間違えているため。	XAIと合わせて見ても判断根拠が分からないため。	0
			入力画像は見にくいですが、2桁には見えない。	着目領域は間違っていないように見える。そもそもの実力？	0
			入力データが不鮮明で数字の部分が潰れているため誤判定も仕方がないと思うため	XAIで標識の丸い形状を判定に使っているとわかるため、文字の誤判定はしょうがないと思うため	0
			2桁と3桁で違うので。	XAIの結果は何も影響されない	0
		理解の深化		(空白)	0
			2番と同様だが、確信度が低いのでこちらはある程度納得できる	特に印象の変化なし	0
			確信度はある程度低いですが、桁数がせめて合っていて欲しい	結果が間違っていることが重要なのであって、どこを見て間違えたのかどうかを知ることはユーザとしてはさほど重要ではないと思う。正解しているときにどこを見て正解したのかは安心感につながると思う。また、これらの例ではどれも大体同じ注目点(標識の内側)であり、注目点を提示してもらっても差が感じられない。	0