

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

一般財団法人日本科学技術連盟
第36年度(2020年度) ソフトウェア品質管理研究会
成果発表会

研究コース5 人工知能とソフトウェア品質 XAIチーム

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)

2021年2月26日

アジェンダ

- 1. 現場の問題**
- 2. 研究の目的**
- 3. 検証方法**
- 4. 検証結果と結論**
- 5. 今後の展望**

1. 現場の問題

はじめに

機械学習技術の進展(Deep Learningの登場)で,
自動運転・医療診断・金融サービスなどでAI技術のシステム導入
が進んでいるが...

AIは従来に比べて非常に高精度である半面, 課題もある

- ・ 常に完璧 (精度100%) ということはない
- ・ どうして正しく判定できなかったか, 根拠をうまく説明できない

顧客のAIに対する期待は高いが, 提供側は説明に苦慮..
人はAIより人間による予測を信頼するという研究結果^[※]もある..

え? AIってなんでもできるんじゃないの?

なんで失敗するの?

顧客

できないケースもありまして..

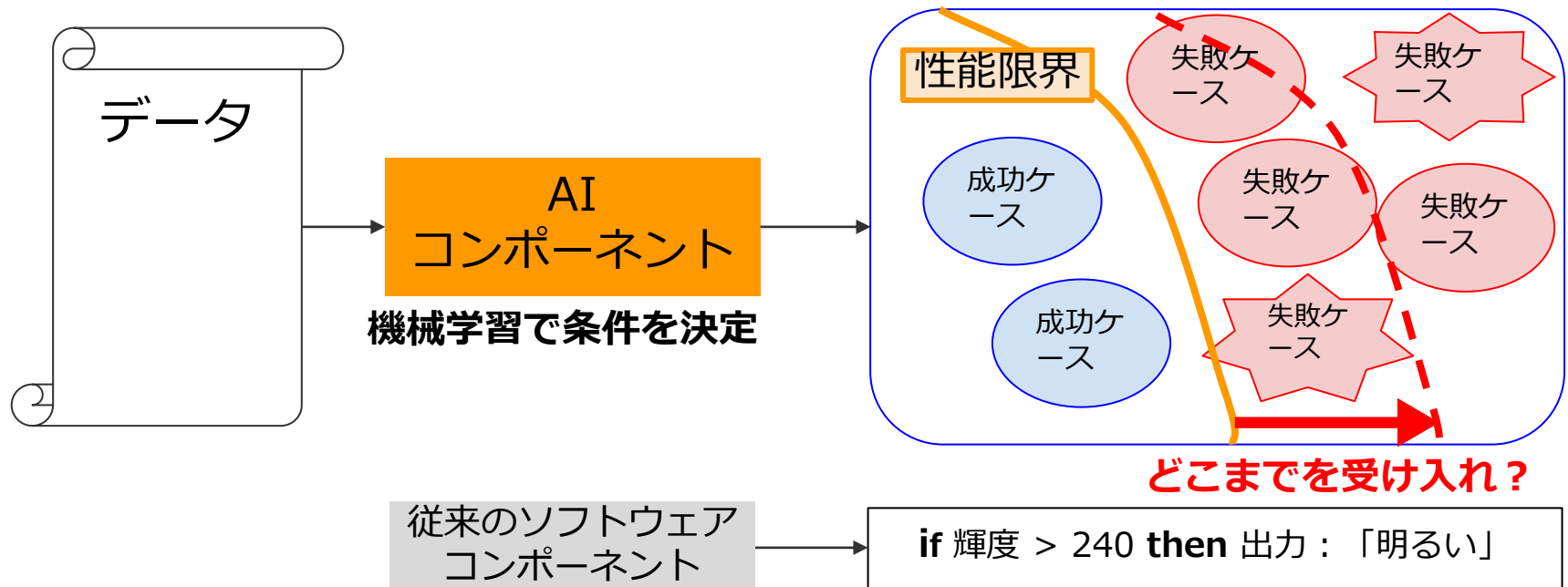
ええっと..不鮮明な画像が入力されると..

AIシステム提供側

AIシステム開発の難しさ

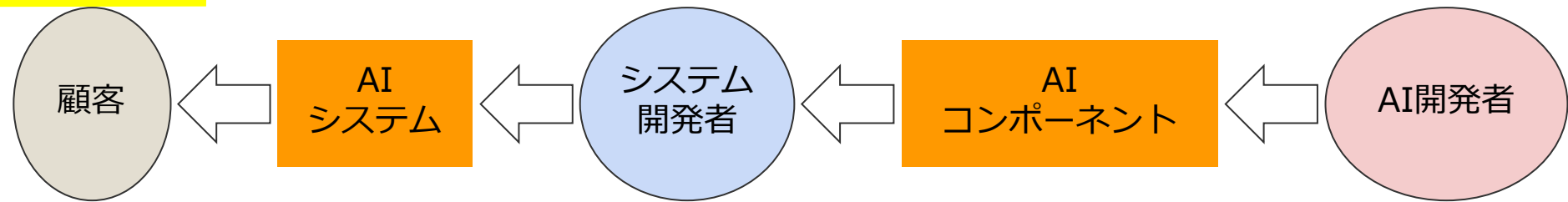
AIコンポーネントの特性

- ・ 何がどこまでできるかはデータで決まる（帰納的）
- ・ 判断根拠を説明できない（ブラックボックス）
- ・ 詳細仕様を決めるには個別のケースを見ていくほかない
→ **失敗ケースの受け入れの判断は難しい**



人間が条件を決めてプログラムを作成

AIシステム開発の現場で起こっている課題



最終成果物の責任者 (AIの非専門家)

- ❖ 顧客の問題解決
- ❖ AIコンポーネントの受入と改善要求の判断

なんでこれができないの？
人間だったら間違えないよ。
納得いかないんだけど...

この失敗ケースは、顧客の信頼をなくす...



どこまでを受け入れ？

人が見て
容易か

AIの
知見

AIコンポーネントの提供者 (AIの専門家)

- ❖ AI技術の選定
- ❖ AIコンポーネント構築

このケースは難しい...上手く
いったという論文も見ただ
けない..**納得**するしかない

学習データが足りてないな..
集めるあてもない、どうし
ょう..?

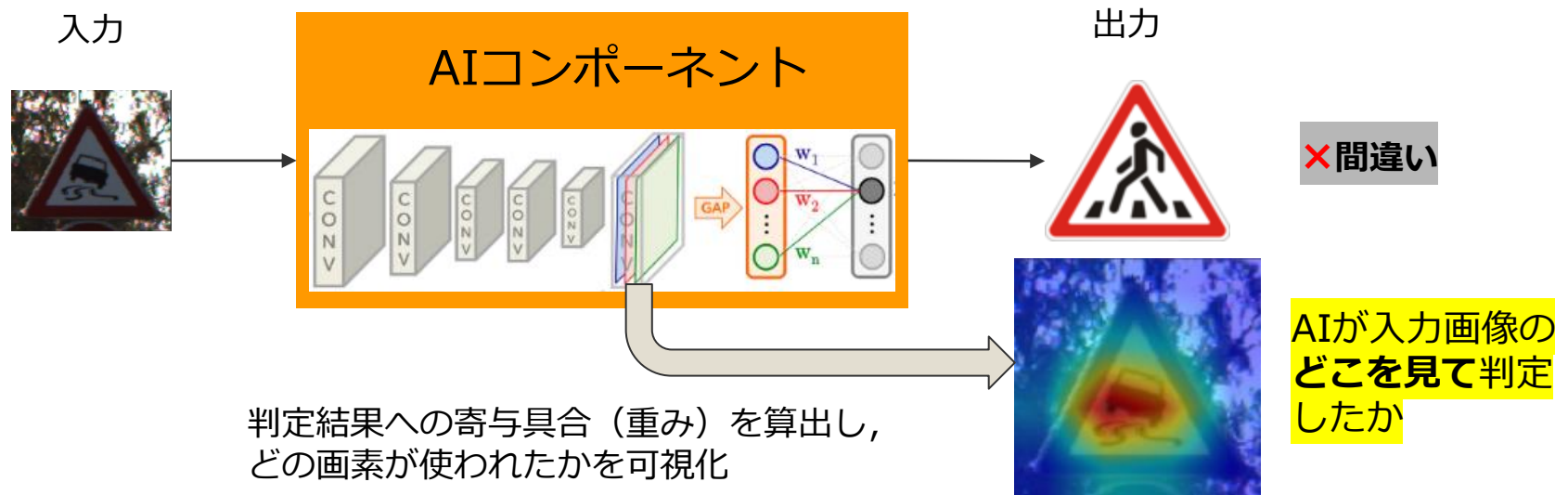
システム開発者とAI開発者で、失敗ケースの捉え方にギャップがあり、
AIコンポーネントの受け入れ判断で意見の一致に時間がかかっている

2. 研究の目的

研究の目的

学術分野でAI説明技術 (eXplainable AI)の研究が盛んである
個々のケースでAIの判断根拠に洞察を与える手法が提案されている

【GradCam系の着目領域の可視化技術】



XAIを失敗ケースに適用することでAIの非専門家であるシステム開発者もAI開発者に近い捉え方が可能になると考えた

→ 研究の目的：

XAI適用による，システム開発者とAI開発者のギャップ^o解消効果を検証

3. 検証方法

納得感への着目

- 失敗ケースの捉え方は、**納得の仕方の違い**という形で現れる

なんでこれができないの？
人間だったら間違えないよ。
納得いかないんだけど...

このケースは難しい...上手く
いったという論文も見たこと
ない..**納得**するしかない

そもそも納得とはどういうことか？

- 納得という概念は、様々な属性を持つ[※]が、特に**「理解の深化」**，「受け入れ」，「意思決定」は、AI システム開発の文脈と深い関係にある

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**」を指標とした

※参考論文：「納得の概念分析」、今井 et al.

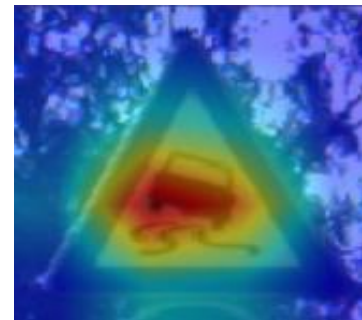
検証概要 (1/2)

【検証で想定したAIシステム】

- 題材：「運転支援システムにおける道路標識判定」
 - 自動車に付けられたカメラの映像から道路標識を判定してドライバーに通知するシステム
- 画像：ドイツの道路標識のデータセットGTSRB
(German Traffic Sign Recognition Benchmark)



- AI : ネットワーク : Resnet50. 学習データ : GTSRB中の学習用データ
- XAI : Grad-Cam++



XAI(Grad-Cam++)によるAIの着目領域
赤に近いほどAIが重視していることを示す

検証概要 (2/2)

システム開発者と AI 開発者にアンケート調査を実施

調査対象： システム開発者12名, AI開発者7名

アンケート内容：

失敗ケースの画像：15サンプル

設問 1： 回答者の属性(経験年数等)

設問 2： AIの判定結果と確信度を示し, 納得感とその理由を回答
(システム開発者とAI開発者で納得感に差があるか?)

設問 3： AIの判定結果と確信度にXAI画像を加え, 納得感とその理由を回答
(XAIの導入でシステム開発者とAI開発者の差は縮まるか?)

回答方式： 納得感…選択式 「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」

理由 …自由記述

使用したアンケート(一部)

設問 2 : XAI導入前

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%			

設問 3 : XAI導入後

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の注目領域（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%				

4. 検証結果と結論

XAI導入前の納得感の比較(1/2)

設問 2： システム開発者とAI開発者で納得感に差があるか？
各サンプルごとに納得感をシステム開発者とAI 開発者
それぞれの平均値を算出して比較

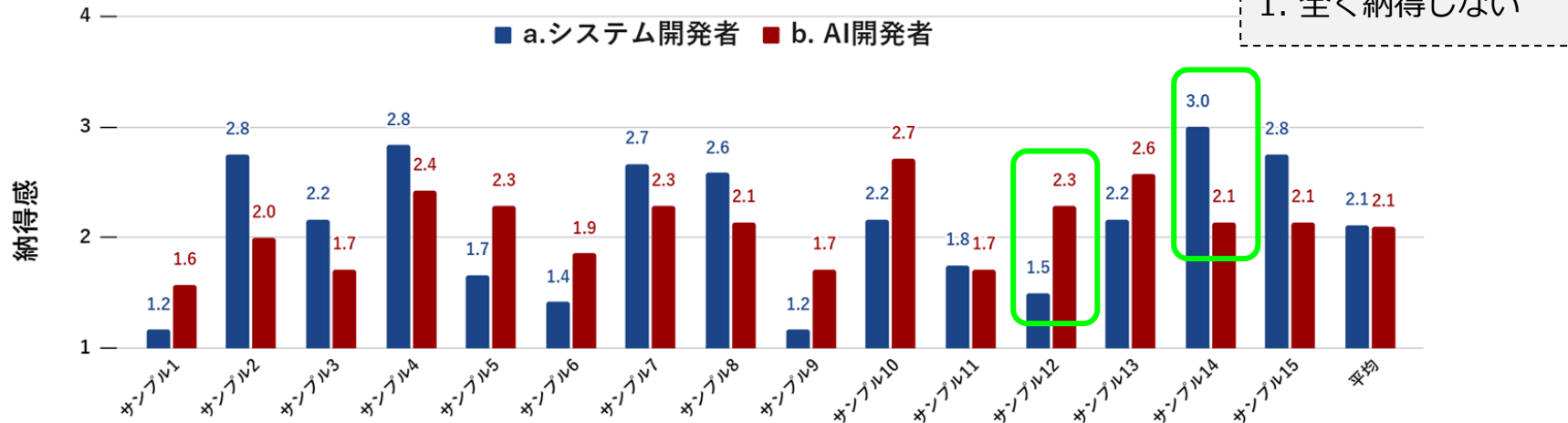


図1 XAI 導入前のシステム開発者と AI 開発者の納得感の比較

- ほとんどのサンプルで納得感の差が0.4以上
 - 15サンプル中14サンプル
- サンプル12, サンプル14で1ステップ分に近い差
 - 次ページで説明

XAI導入前の納得感の比較(2/2)

サンプル12

入力画像



AI判定結果



確信度：
98.79%

納得感の理由

「1.全く納得しない」理由

確信度が高く、形状も異なる

「2.やや納得する」理由

人間でも間違いかねない

システム開発者：
入出力の情報のみで判断

サンプル14

入力画像



AI判定結果



確信度：
99.92%

「3.やや納得する」理由

照明変動に弱いアルゴリズム
であると推測できる

「1.全く納得しない」理由

難しい問題に見えない

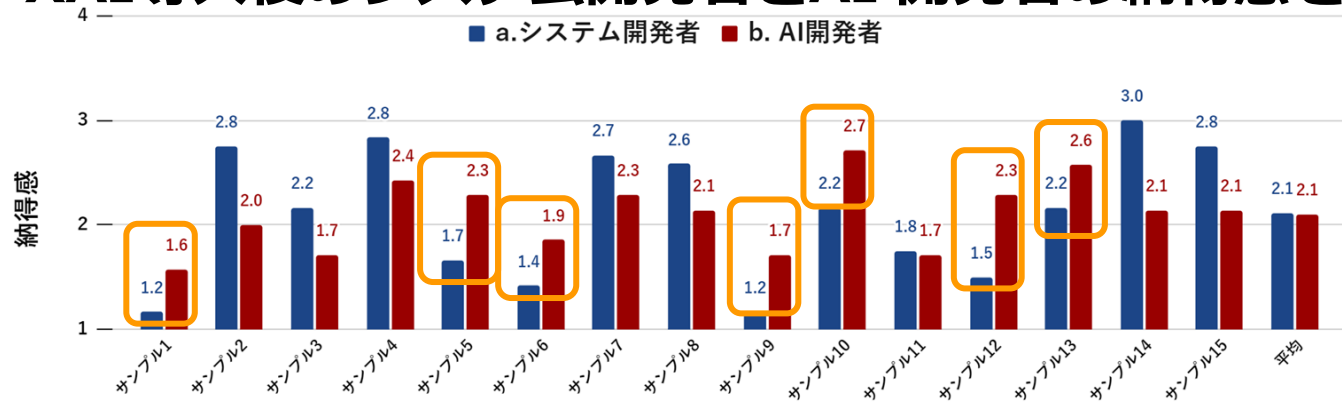
AI開発者：
AIの知見に基づく理由もあった

XAI導入後の納得感の比較

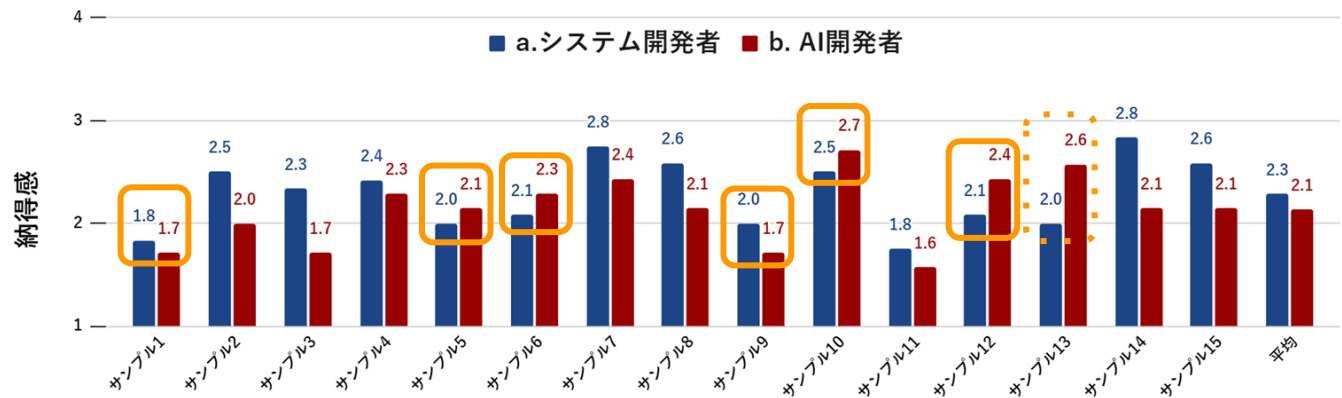
設問 3： XAIの導入でシステム開発者とAI開発者の差は縮まるか？

XAI導入後のシステム開発者とAI 開発者の納得感を比較

XAI導入前





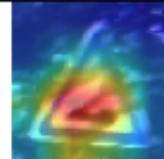

XAI導入後



- 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は、サンプル13を除き全てシステム開発者の納得感が向上し、AI開発者の納得感に近づいた

サンプル1の詳細分析

サンプル1

入力画像		AIが判定した画像 (確信度)	(XAI)		正解の画像
	→	 89.72%		↔	

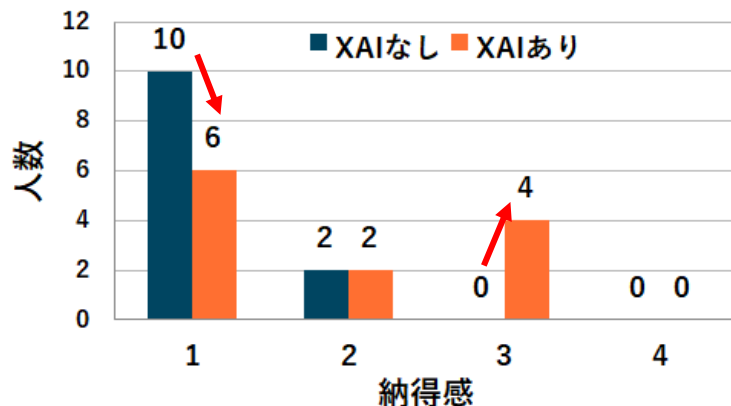
納得感の理由

設問3(XAI導入後)で初めて「影の影響」に言及が見られた

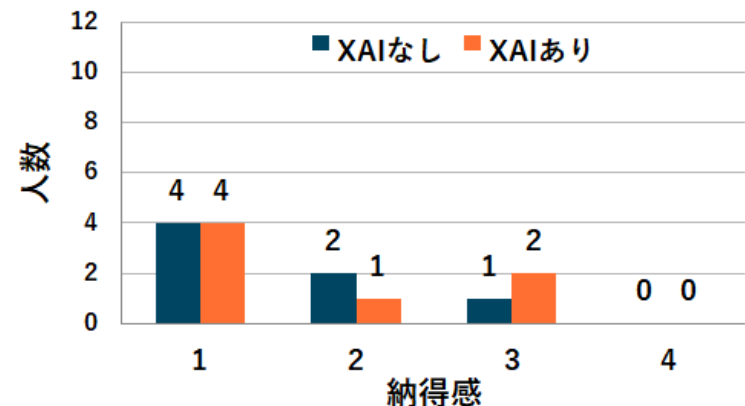
例：影の影響により左下向きの矢印に見える（システム開発者）

納得感の人数分布

システム開発者



A I 開発者



- システム開発者の大多数は、XAI 導入前「全く納得しない」であったが、XAI導入後はバラつき（個人差）が大きくなった

納得感の理由の分析

納得感の理由を分類し，XAIがどのように影響したかを見る

分類	定義
理解の深化	具体的な内容（形状・色等）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
分からない	具体的な内容（形状・色等）の言及があり，「分からない」などの疑問を示す表現を使っている
XAIの影響なし	上記のいずれにも当てはまらない

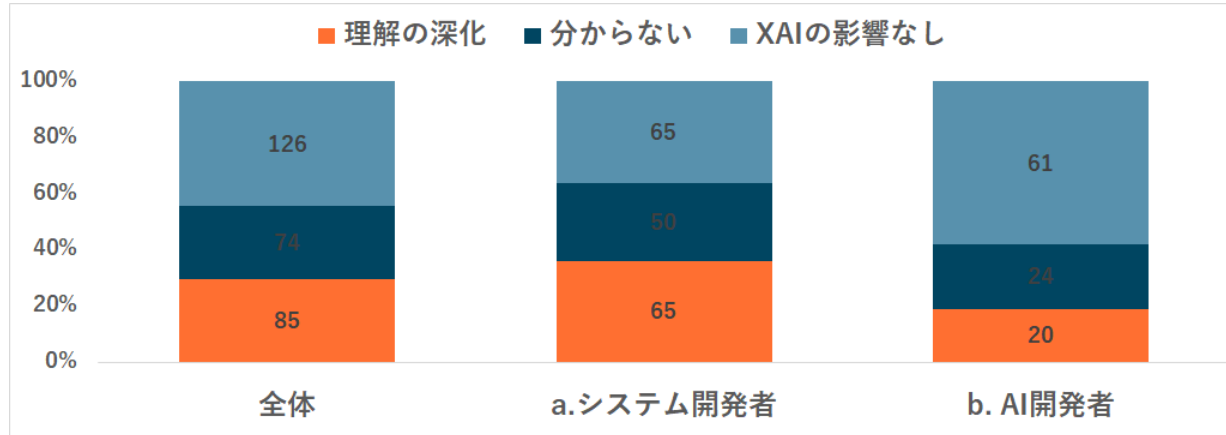


図 納得感の理由の分類結果

回答者全員の約56%にXAIの影響が見られた。

また，システム開発者は約64%であり，より影響が大きかった

結論

わかったこと

- 失敗ケースの捉え方はAI 開発者とシステム開発者で異なる
- XAIの導入はギャップ解消に一定程度効果があった：
 - － システム開発の納得感がAI開発者に近づいた
 - － 理解できる・できない理由の「具体的な根拠」が言えるようになる
- 受け止め方（納得感）には個人差がある

言いたいこと

- 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 開発者と議論ができること
 - 議論のポイントの明確化, 追加評価項目の抽出
- 少人数の議論は意見が偏るため危険！

5. 今後の展望

今後の展開

- **XAIの影響の定量分析**

- ① 被験者数の拡大
- ② 納得感の理由の定量分析（整理・分類）

- **調査内容の充実**

- ① 失敗ケースだけでなく成功ケースを併せて提示
- ② AIコンポーネントの改良過程に伴う納得感の変化を計測
- ③ 納得感の他の要素（他のAI技術との比較）の追加

**石川主査，栗本副主査，徳本副主査，
1年間ご指導ありがとうございました**

ご清聴ありがとうございました