

The Impact of *Imperfect* XAI on Human-AI Decision-Making

Ph.D. Program Communications Requirement Talk – August 2023

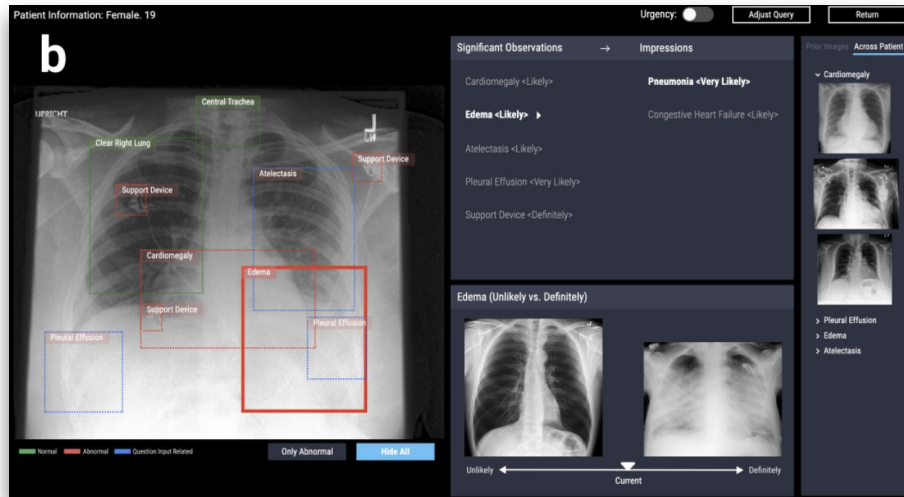
Katelyn Morrison^{1*}, Philipp Spitzer^{2*}, Violet Turri¹,
Michelle Feng¹, Niklas Kühl³, Adam Perer¹

1: Carnegie Mellon University; 2: Karlsruhe Institute of Technology; 3: University of Bayreuth

* equal contribution



High-Stakes Decision-Making



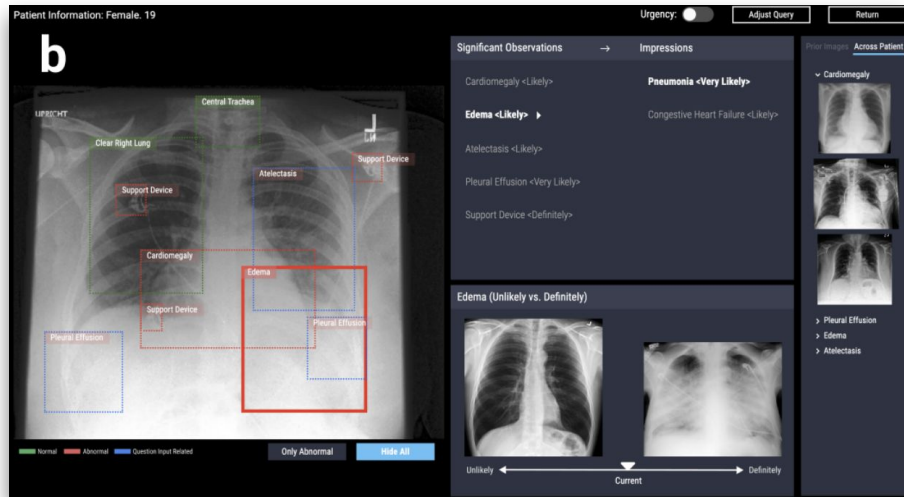
Xie et al., CHI 2020



Juan Lavista; Microsoft AI for Good, 2023

Accuracy and **accountability** of these tools are **vital**

High-Stakes Decision-Making



Xie et al., CHI 2020



Juan Lavista; Microsoft AI for Good, 2023

Accuracy and **accountability** of these tools are **vital**

Decision-makers need **guidance** and **explanations** about **how** and **why** certain predictions are outputted

Human-AI Decision-Making

How different **explanation techniques** impact **decision-making**



Morrison et al., CSCW 2023

How different **explanation techniques** are **interpretable** to humans



Morrison et al., CSCW 2023

Human-AI Decision-Making

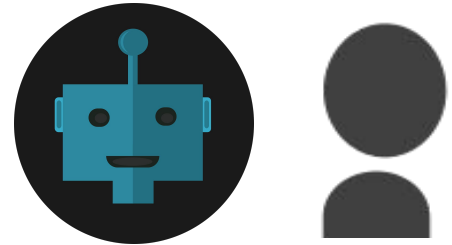
which characteristics of

explanation

decision-maker



impact



human-AI
collaboration

Human-AI Decision-Making

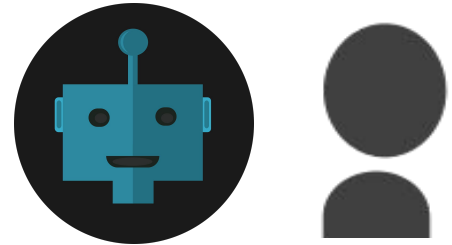
which characteristics of

explanation

decision-maker



impact



human-AI
collaboration

accurate information
conveying confidence

domain
expertise

Human-AI Decision-Making: Identifying Bird Species



Non-expert
(e.g., citizen scientist)



Expert
(e.g., ornithologist)



AI's Prediction:
Magnolia Warbler
Correct Prediction

Human-AI Decision-Making: Identifying Bird Species



Non-expert
(e.g., citizen scientist)



Expert
(e.g., ornithologist)



AI's Prediction:
Magnolia Warbler
Correct Prediction



gradient-based
method



Explainable AI techniques:

AI explaining AI

"this is a bird with a yellow belly black stripes on its breast and a grey head"



Human-AI Decision-Making: Identifying Bird Species



Non-expert
(e.g., citizen scientist)



Expert
(e.g., ornithologist)



AI's Prediction:
Magnolia Warbler
Correct Prediction

Explainable AI techniques:

AI explaining AI

"this is a bird with a yellow belly black stripes on its breast and a grey head"



Human-AI Decision-Making: Identifying Bird Species



Non-expert
(e.g., citizen scientist)



Expert
(e.g., ornithologist)



AI's Prediction:
Magnolia Warbler
Correct Prediction

AI's Prediction:
Magnolia Warbler

&
extracted features
of image

Hendricks et al. ECCV. 2016.

Explainable AI technique:
Natural Language Explanation

Human-AI Decision-Making: Identifying Bird Species



Non-expert
(e.g., citizen scientist)



Expert
(e.g., ornithologist)



AI's Prediction:
Magnolia Warbler
Correct Prediction

"this is a bird with a yellow belly black stripes on its breast and a grey head"

Hendricks et al. ECCV. 2016.

Explainable AI technique:
Natural Language Explanation
Correct Explanation

Human-AI Decision-Making: Identifying Bird Species



Non-expert
(e.g., citizen scientist)



Expert
(e.g., ornithologist)



AI's Prediction:
Magnolia Warbler
Correct Prediction



Explainable AI technique:
Example-Based Explanations
Correct Explanation

Can AI *explain* another AI's output? Sometimes ...



AI's Prediction:
Cerulean Warbler
Correct Prediction



"this is a small bird with a white belly and a grey head"

Explainable AI technique:
Natural Language Explanation
Incorrect Explanation

Can AI *explain* another AI's output? Sometimes ...

Imperfect XAI



AI's Prediction:
Cerulean Warbler
Correct Prediction



"this is a small bird with a white belly and a grey head"

Explainable AI technique:
Example-Based Explanations
Incorrect Explanation

Can AI *explain* another AI's output? Sometimes ...



AI's Prediction:
Nashville Warbler
Correct Prediction



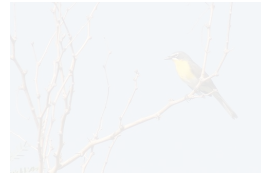
Explainable AI technique:
Example-Based Explanations
Incorrect Explanation

Can AI *explain* another AI's output? Sometimes ...

Imperfect XAI

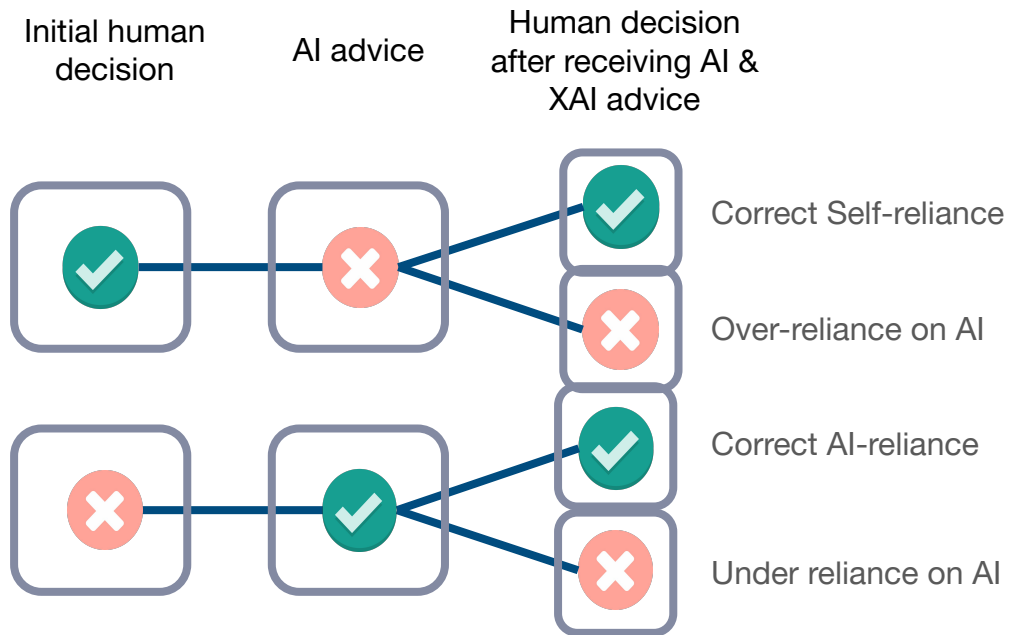


AI's Prediction:
Nashville Warbler
Correct Prediction



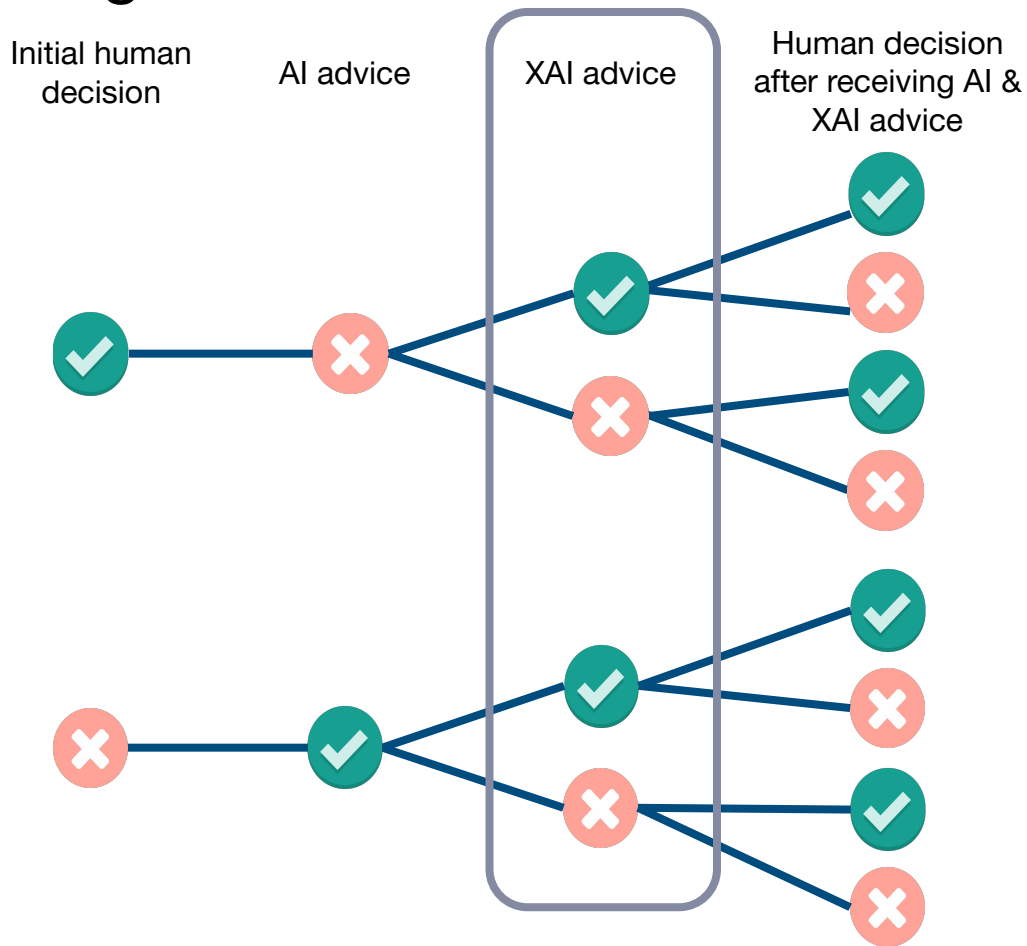
Explainable AI technique:
Example-Based Explanations
Incorrect Explanation

Conceptualizing Human-AI Collaboration

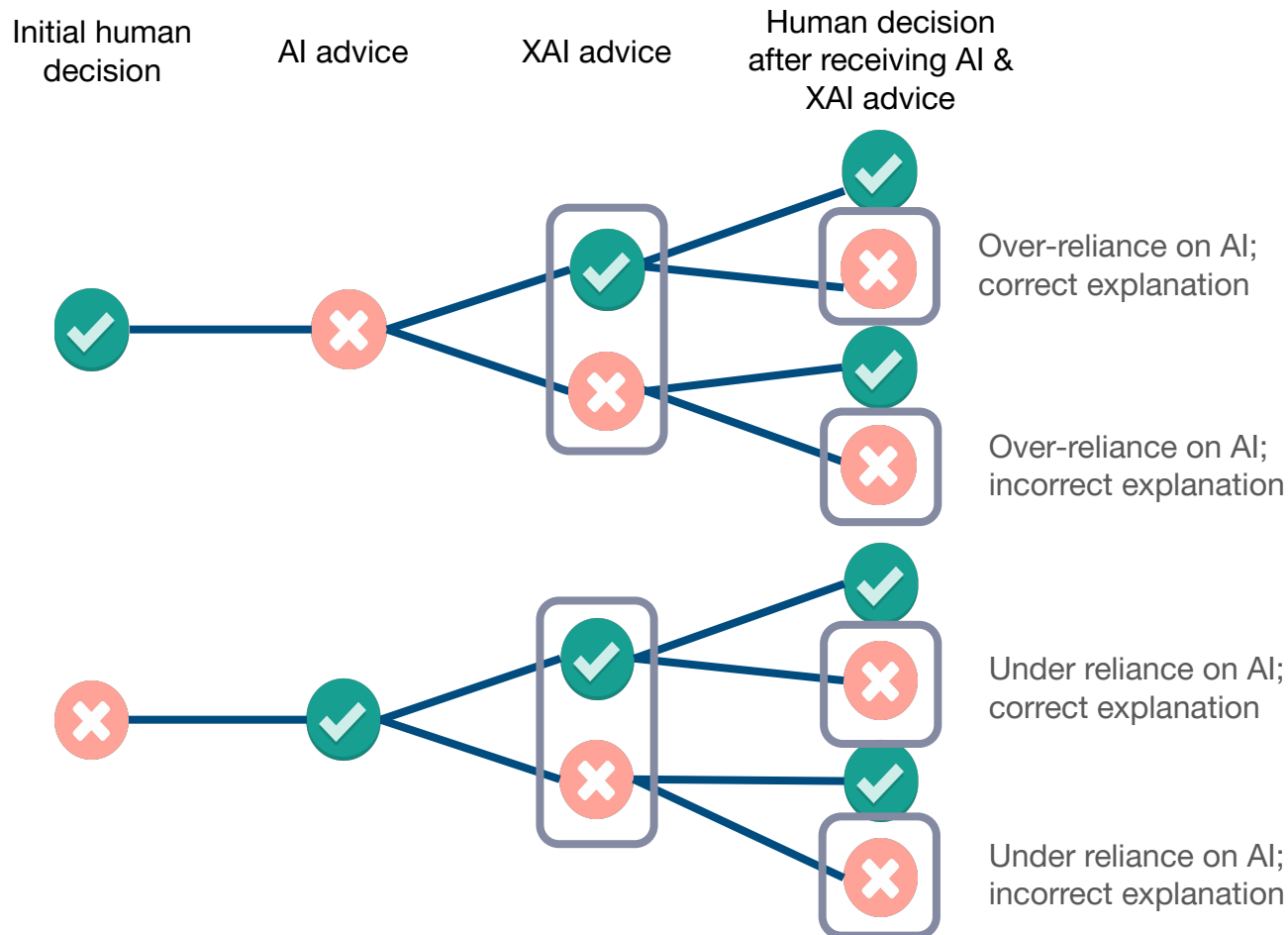


Schemmer et al. IUI 2023.

Conceptualizing Human-AI Collaboration w/ *Imperfect* XAI

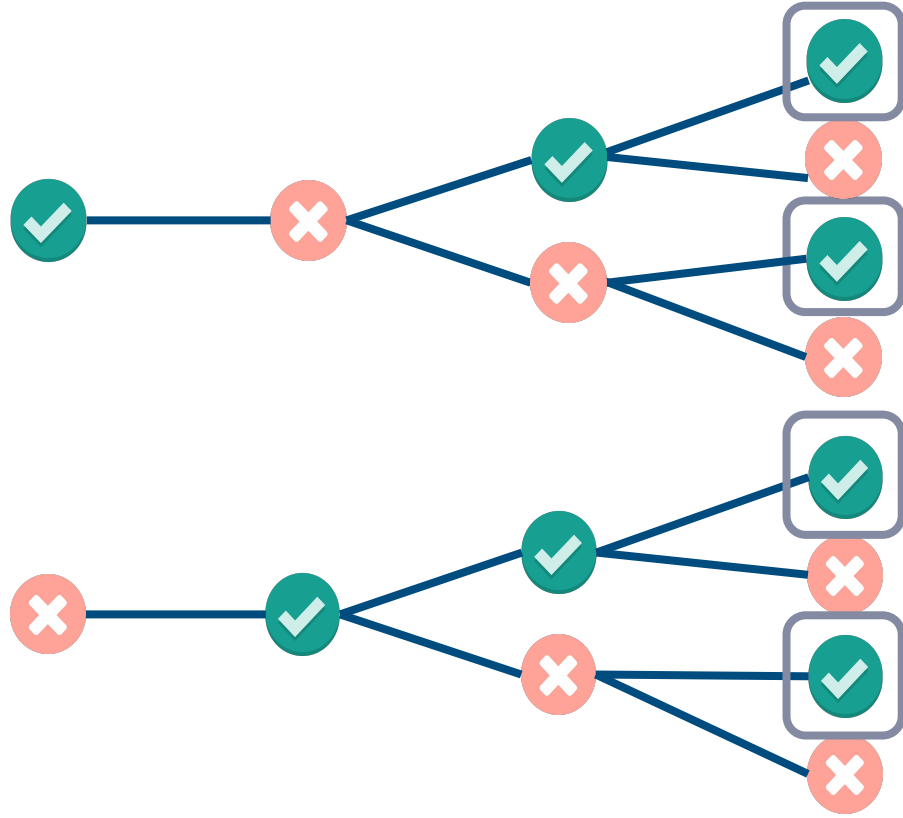


Conceptualizing Human-AI Collaboration w/ *Imperfect* XAI

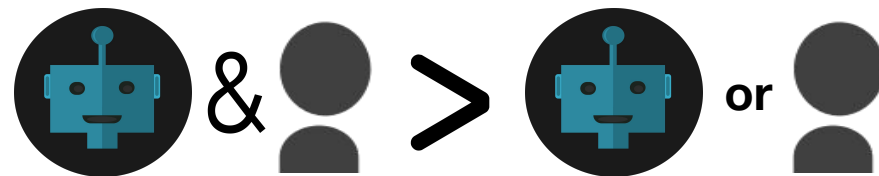


Conceptualizing Human-AI Collaboration w/ *Imperfect* XAI

Initial human decision AI advice XAI advice Human decision after receiving AI & XAI advice



Complementary Team Performance



Domain Expertise of User + *Imperfect* XAI = ... ?



Non-expert
(e.g., citizen scientist)



"this is a small bird with a **white**
belly and a **grey** head"



Expert
(e.g., ornithologist)



AI's Prediction:
Cerulean Warbler
Correct Prediction



Explainable AI technique:
Example-Based Explanations
Incorrect Explanation

Research Questions

Level of Expertise

How is **complementary team performance** impacted by the decision-maker's **level of domain expertise**?

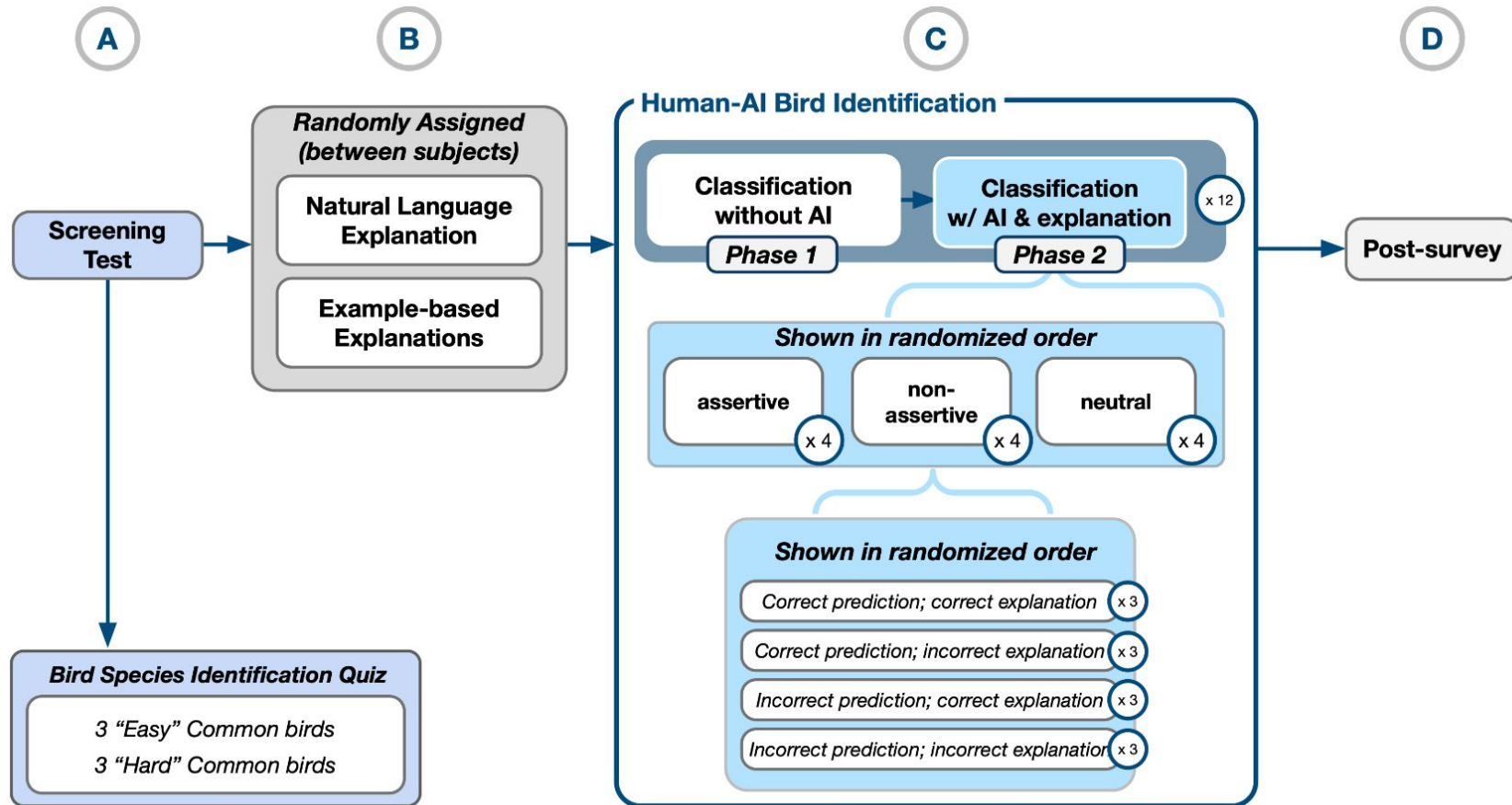
Correctness of Explanations

How is **complementary team performance** impacted by the **correctness of explanations**?

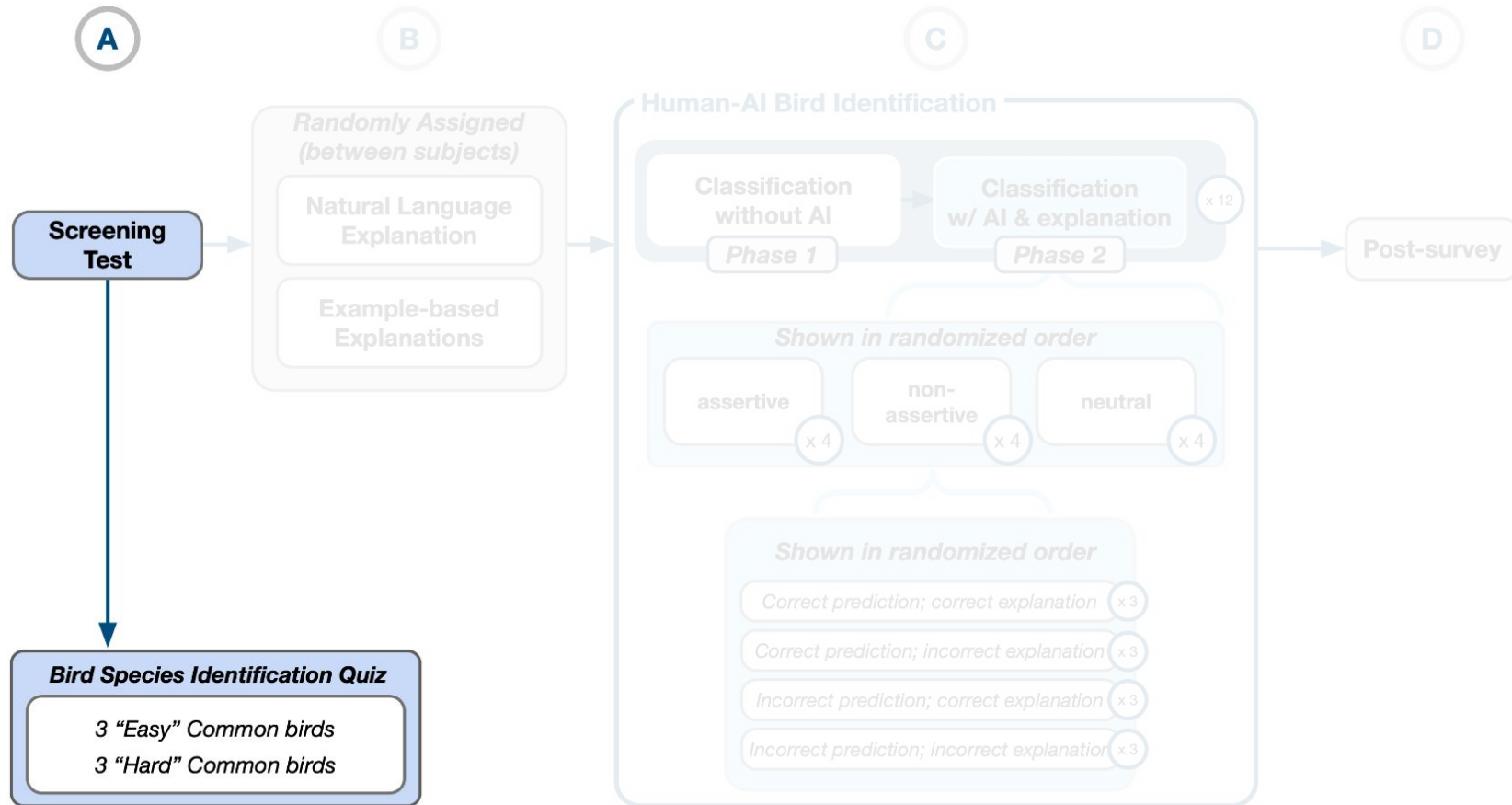
Deception

To what extent do **incorrect and correct explanations deceive** decision-makers with different levels of expertise?

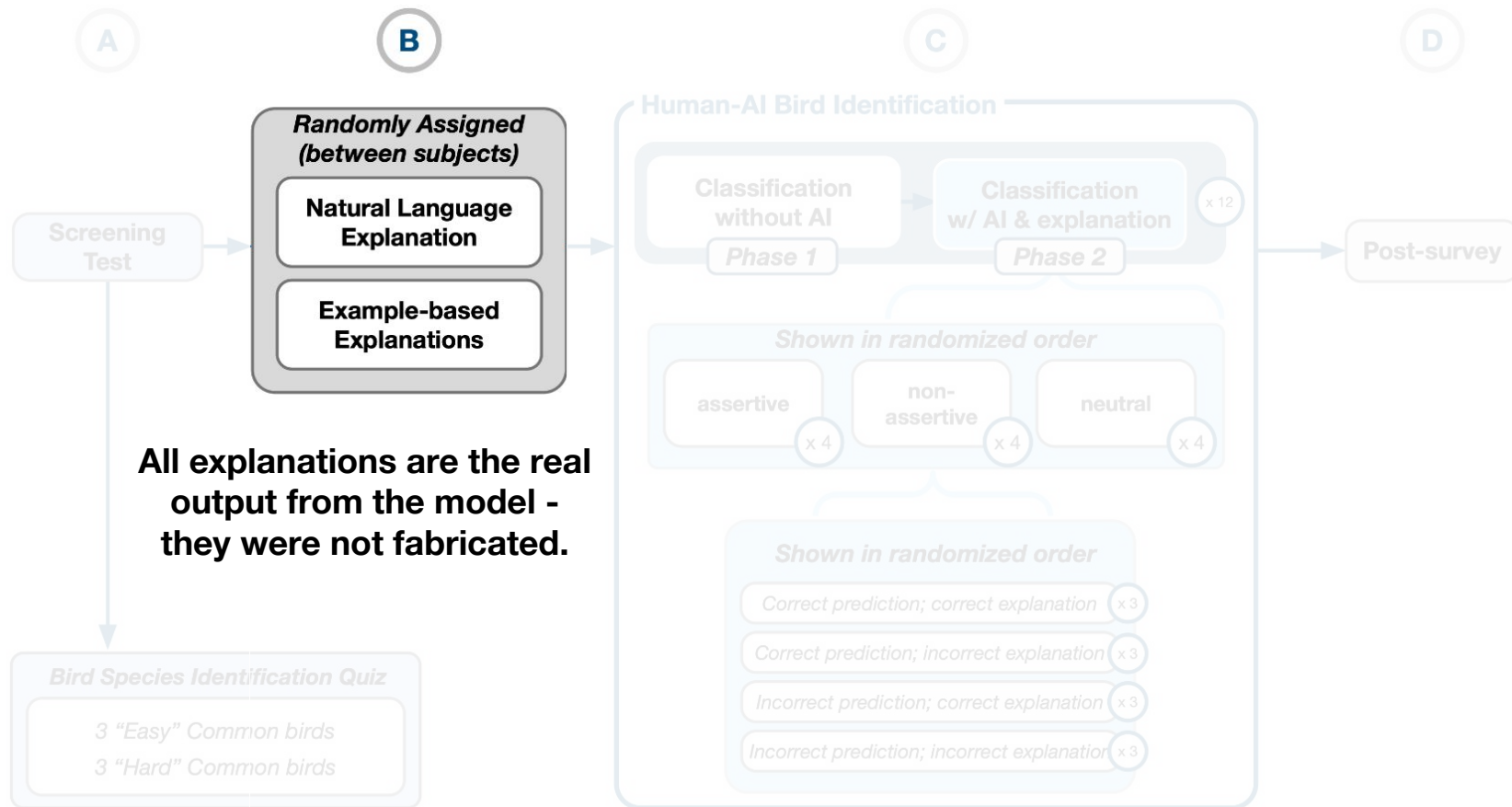
Study Design



Study Design

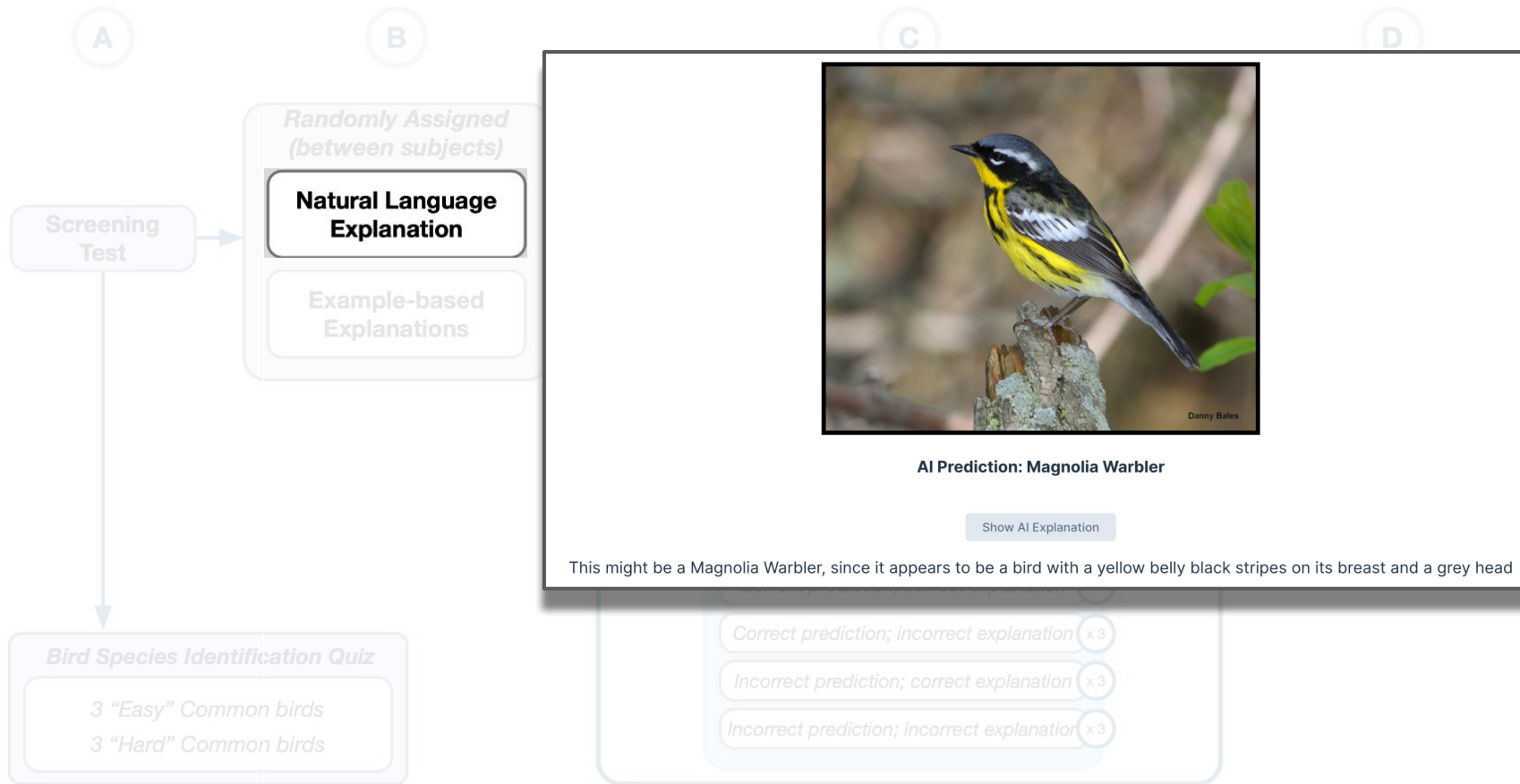


Study Design



Study Design

Hendricks et al. ECCV. 2016.



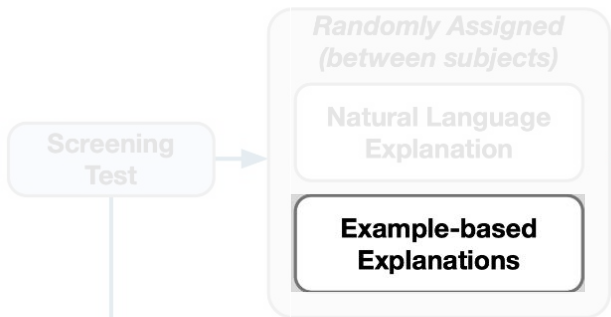
Study Design

A

B

C

D



Hendricks et al. ECCV. 2016.



extracted features of image from model





0	1	1
1	1	0
0	0	1

AI Prediction: Magnolia Warbler

This is definitely a Magnolia Warbler because it is clearly similar to the birds in the following three examples:



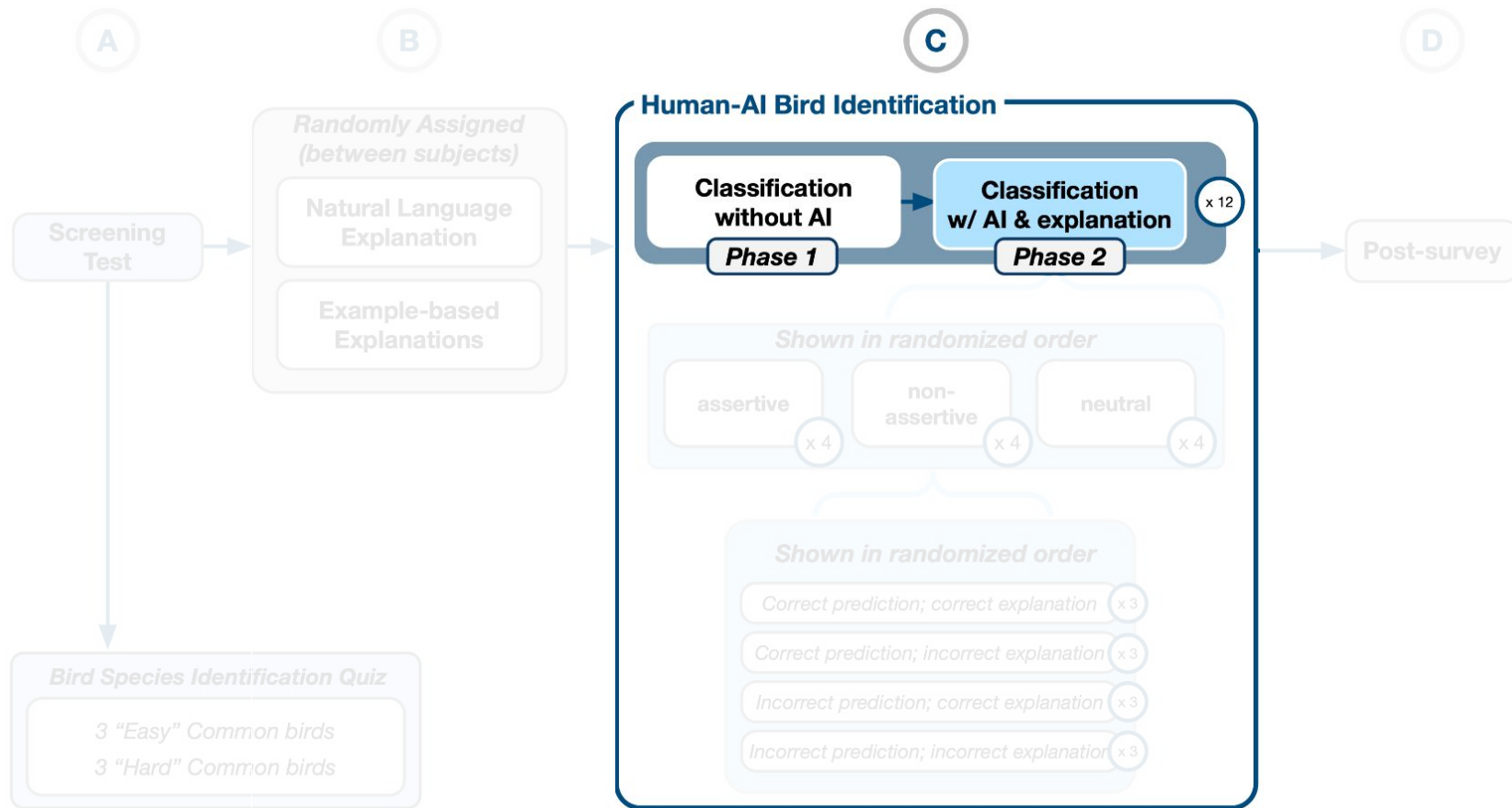


Correct prediction; incorrect explanation x3

Incorrect prediction; correct explanation x3

Incorrect prediction; incorrect explanation x3

Study Design



Study Design

A



Name the bird species:

Canada Warbler

How confident do you feel in your decision?

Not confident

1

2

3

4

5

6

Very confident

Next

3 "Easy" Common birds

3 "Hard" Common birds

B

C

Human-AI Bird Identification

Classification
without AI

Phase 1

Classification
w/ AI & explanation

x 12

Phase 2

Post-survey

Shown in randomized order

assertive

x 4

non-assertive

x 4

neutral

x 4

Shown in randomized order

Correct prediction; correct explanation

x 3

Correct prediction; incorrect explanation

x 3

Incorrect prediction; correct explanation

x 3

Incorrect prediction; incorrect explanation

x 3

Study Design

A

B

C

D

AI Prediction: Magnolia Warbler

Show AI Explanation

This is definitely a Magnolia Warbler because it is clearly similar to the birds in the following three examples:

Name the bird species:

Magnolia Warbler

How confident do you feel in your decision?

Not confident 1 2 3 4 5 6 Very confident

Next

3 "Easy" Common birds

3 "Hard" Common birds

Human-AI Bird Identification

Classification without AI

Phase 1

Classification w/ AI & explanation

Phase 2

x 12

Post-survey

Shown in randomized order

assertive

x 4

non-assertive

x 4

neutral

x 4

Shown in randomized order

Correct prediction; correct explanation

x 3

Correct prediction; incorrect explanation

x 3

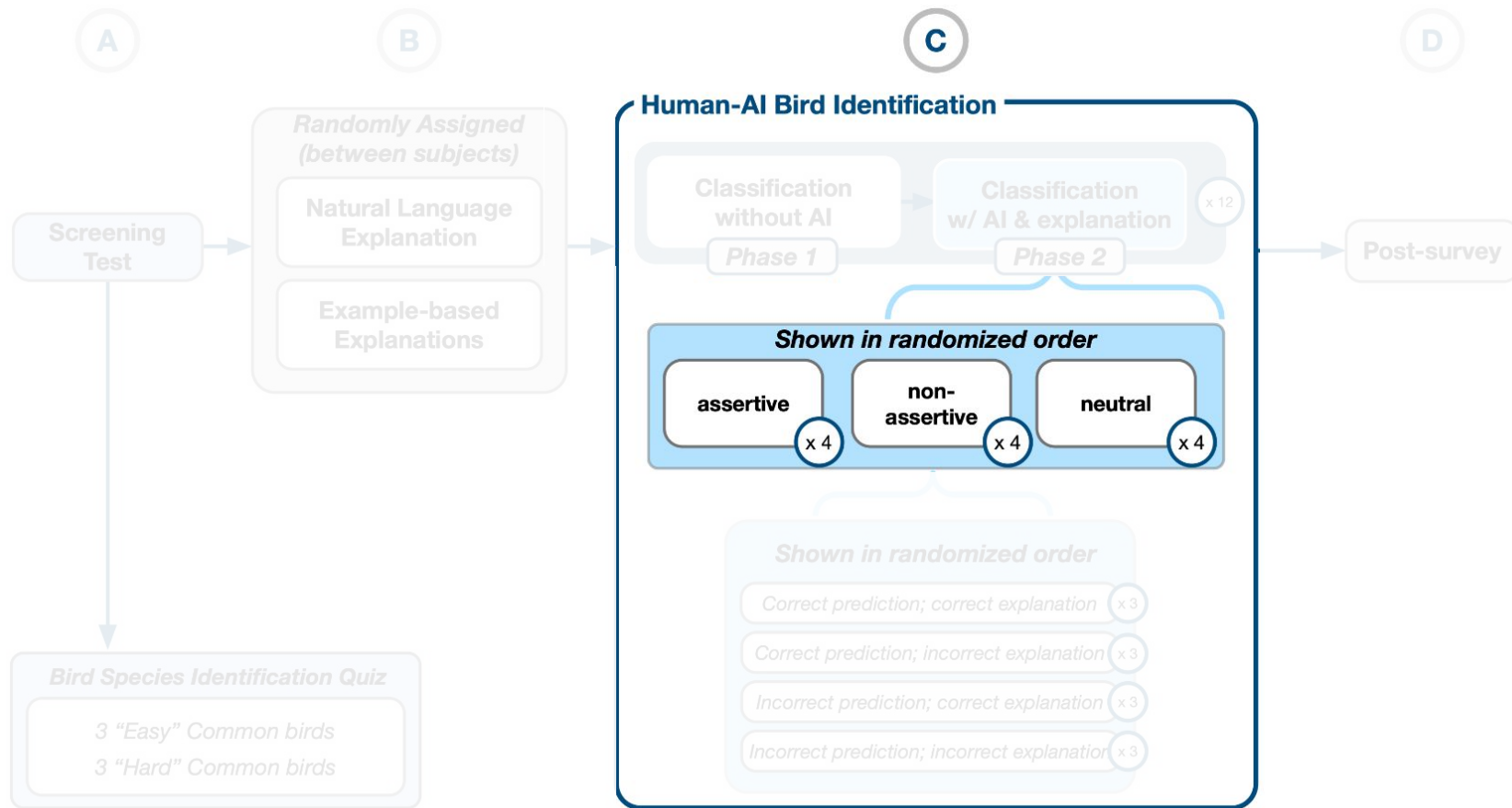
Incorrect prediction; correct explanation

x 3

Incorrect prediction; incorrect explanation

x 3

Study Design



Study Design

A

B

C

D

Randomly Assigned
(between subjects)

AI: Magnolia Warbler



Correct Prediction



Magnolia Warbler



Magnolia Warbler



Magnolia Warbler

Correct Explanation

AI: Magnolia Warbler



Correct Prediction

"this is a bird with a yellow belly black stripes on its breast and a grey head"

Correct Explanation

3 "Easy" Common birds

3 "Hard" Common birds

Human-AI Bird Identification

Classification
without AI

Phase 1

Classification
w/ AI & explanation

Phase 2

x 12

Post-survey

assertive

x 4

in randomized order

non-assertive

x 4

neutral

x 4

Shown in randomized order

Correct prediction; correct explanation

x 3

Correct prediction; incorrect explanation

x 3

Incorrect prediction; correct explanation

x 3

Incorrect prediction; incorrect explanation

x 3

Study Design

A

B

C

D

Randomly Assigned
(between subjects)

AI: Nashville Warbler



Correct Prediction

Incorrect Explanation

AI: Cerulean Warbler



"this is a small bird with a white belly and a grey head"

Correct Prediction

Incorrect Explanation

3 "Easy" Common birds

3 "Hard" Common birds

Human-AI Bird Identification

Classification
without AI

Phase 1

Classification
w/ AI & explanation

Phase 2

x 12

Post-survey

assertive

in randomized order

non-assertive

neutral

x 4

x 4

x 4

Shown in randomized order

Correct prediction; correct explanation x 3

Correct prediction; incorrect explanation x 3

Incorrect prediction; correct explanation x 3

Incorrect prediction; incorrect explanation x 3

Study Design

A

B

C

D

Randomly Assigned
(between subjects)

AI: Yellow-breasted Chat



Yellow-breasted Chat Yellow-breasted Chat Yellow-breasted Chat

Incorrect Prediction

Correct Explanation

AI: Bewick Wren



"this is a small brown bird with a white eyebrow and a downward pointing beak"

Incorrect Prediction

Correct Explanation

3 "Easy" Common birds

3 "Hard" Common birds

Human-AI Bird Identification

Classification
without AI

Phase 1

Classification
w/ AI & explanation

Phase 2

x 12

Post-survey

assertive

non-assertive

neutral

x 4

x 4

x 4

Shown in randomized order

Correct prediction; correct explanation x 3

Correct prediction; incorrect explanation x 3

Incorrect prediction; correct explanation x 3

Incorrect prediction; incorrect explanation x 3

Study Design

A

B

C

D

Randomly Assigned
(between subjects)

AI: Dark-eyed Junco



White-crowned Sparrow White-crowned Sparrow White-crowned Sparrow

Incorrect Prediction

Incorrect Explanation

AI: Yellow-billed Cuckoo



"this is a white bird with a black wing and a large black beak"

Incorrect Prediction

Incorrect Explanation

3 "Easy" Common birds

3 "Hard" Common birds

Human-AI Bird Identification

Classification
without AI

Phase 1

Classification
w/ AI & explanation

Phase 2

x 12

Post-survey

assertive

in randomized order

non-assertive

neutral

x 4

x 4

x 4

Shown in randomized order

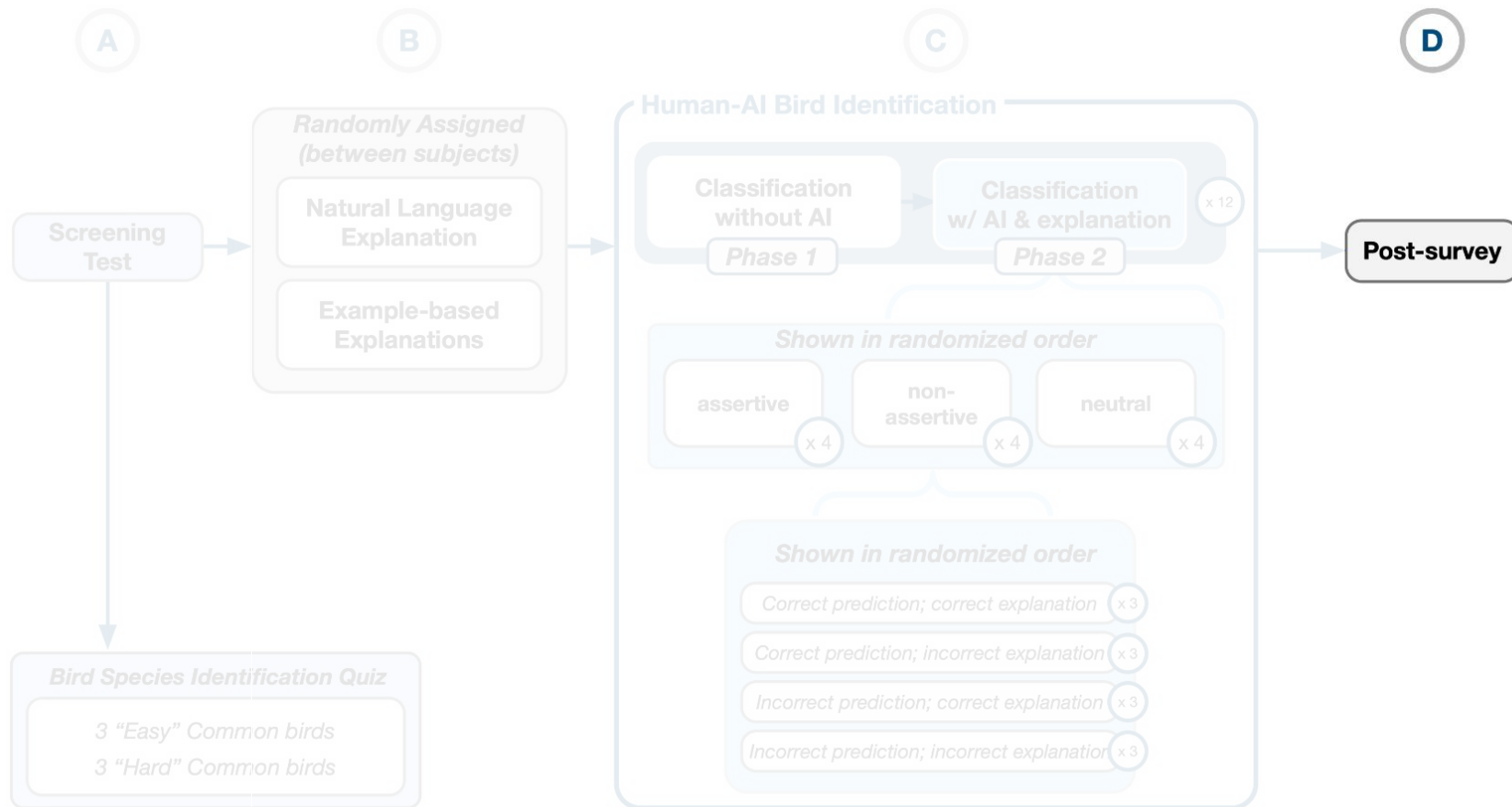
Correct prediction; correct explanation x 3

Correct prediction; incorrect explanation x 3

Incorrect prediction; correct explanation x 3

Incorrect prediction; incorrect explanation x 3

Study Design



Recruitment (N = 136)

AI for
Conservation



Climate
Change AI



Computational
Sustainability



WildLabs.Net



Birding
International



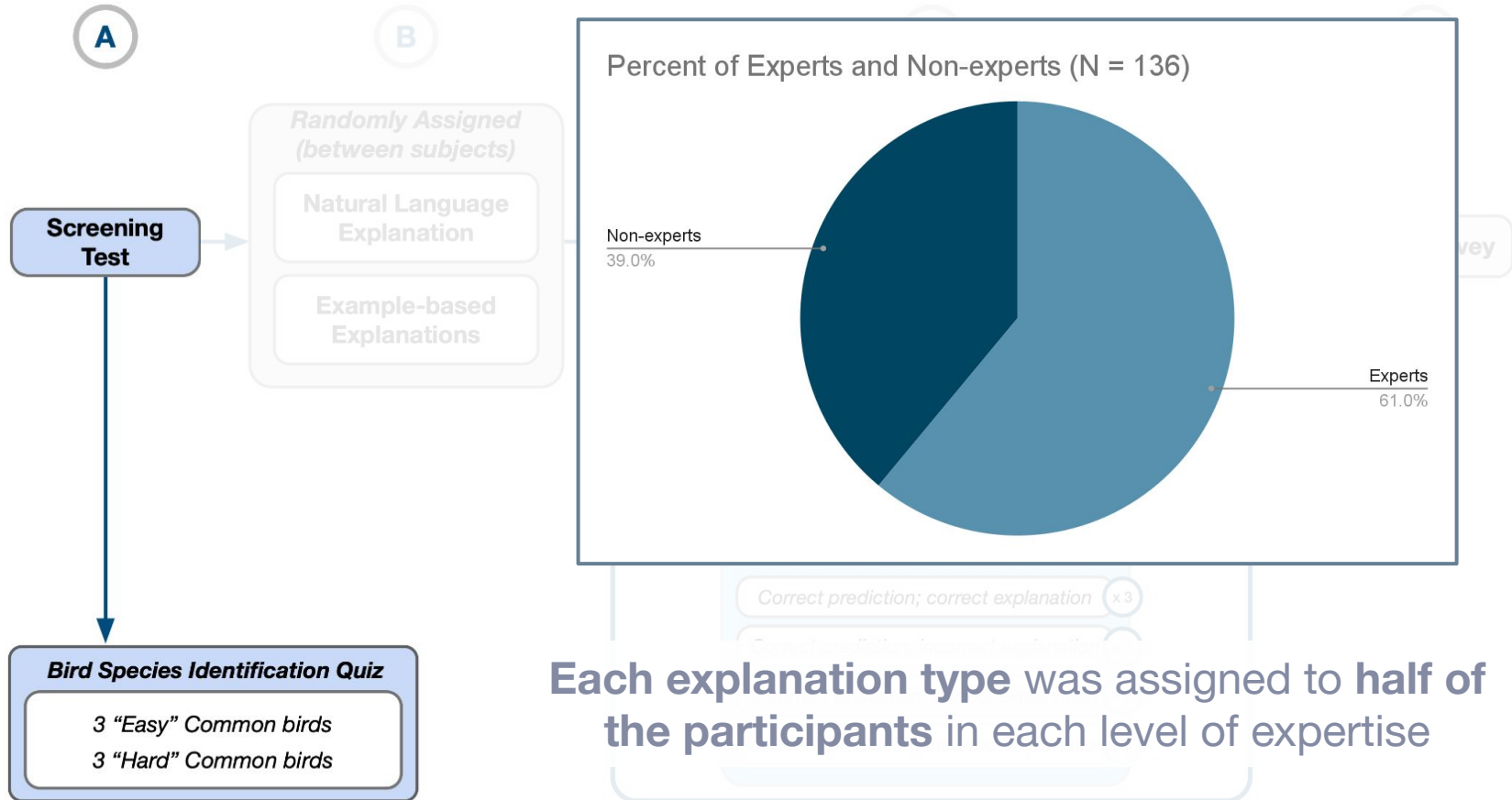
Audubon Society
Groups



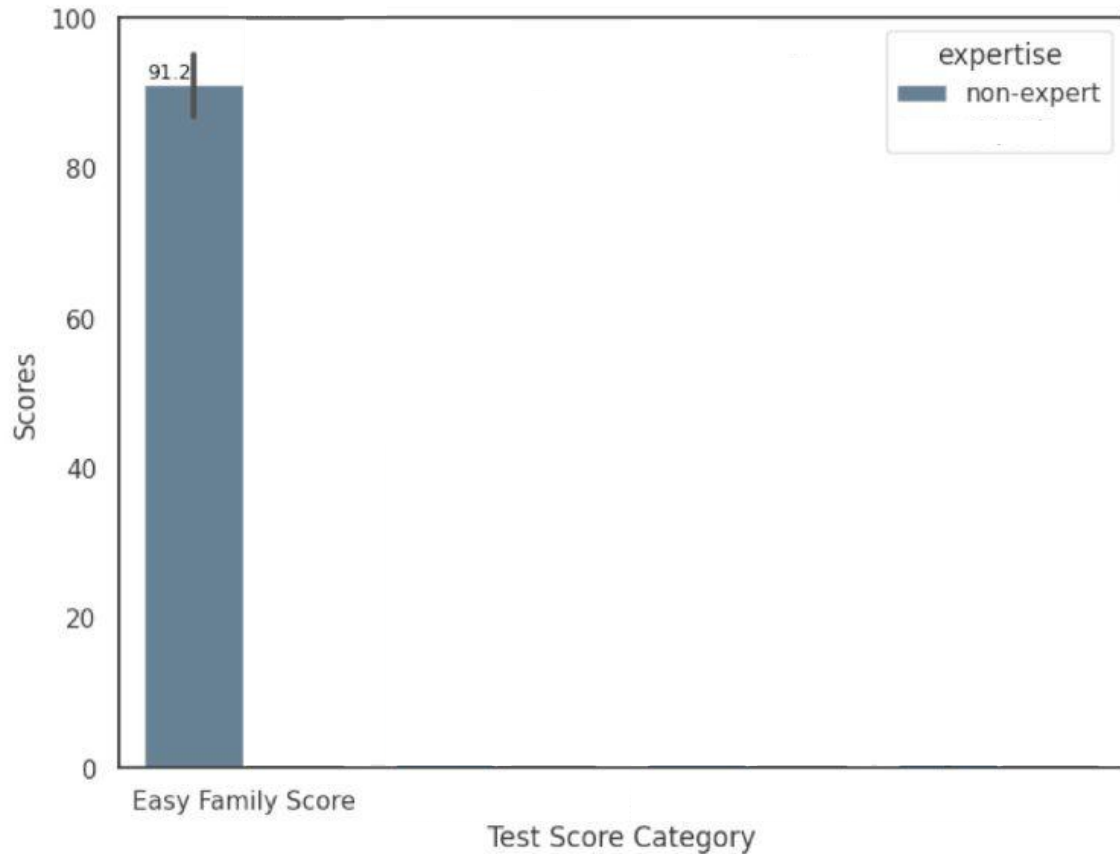
Prolific
Crowdsourcing



Group Participants into Experts & Non-Experts

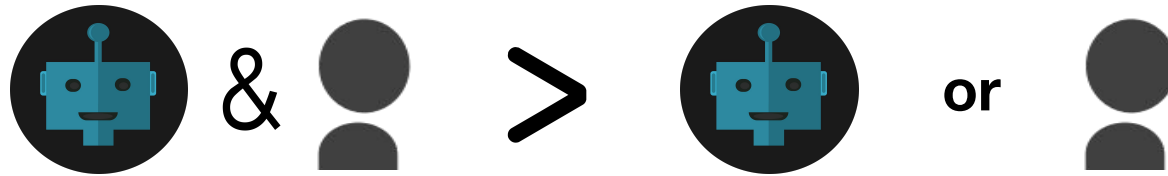


Participants' Bird Species Test Performance

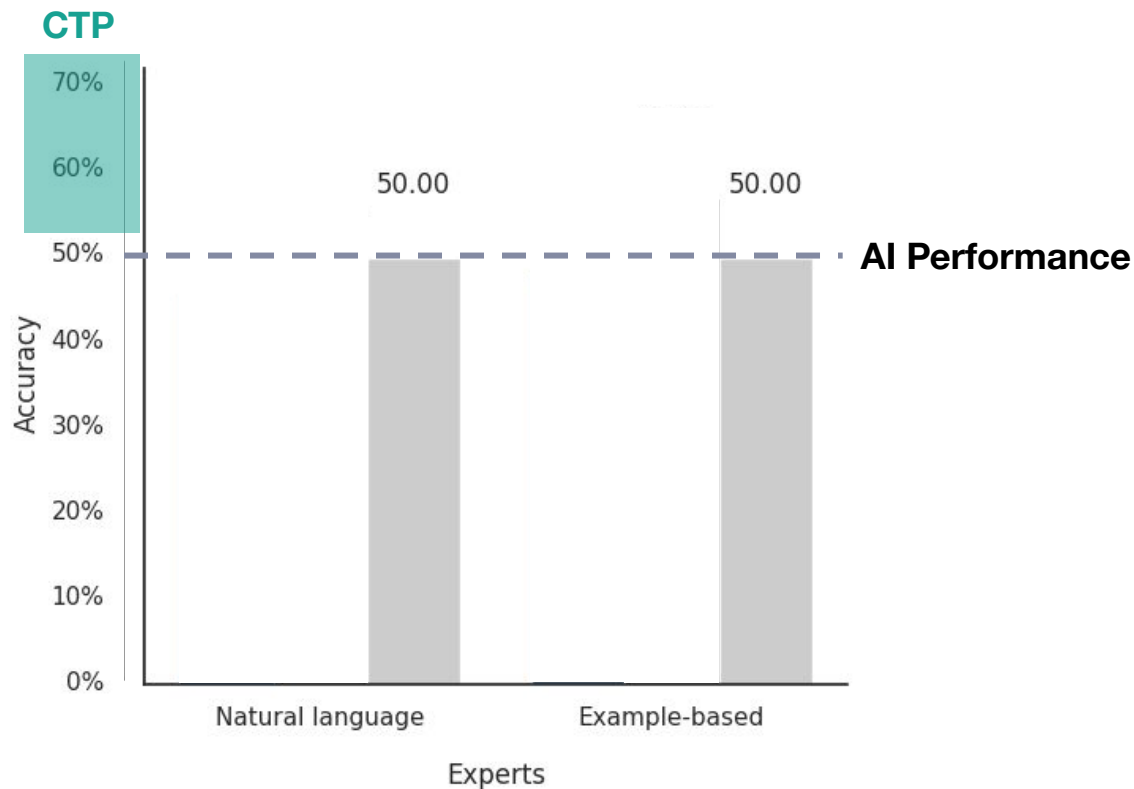


Complementary Team Performance (CTP) for Experts

Complementary Team Performance

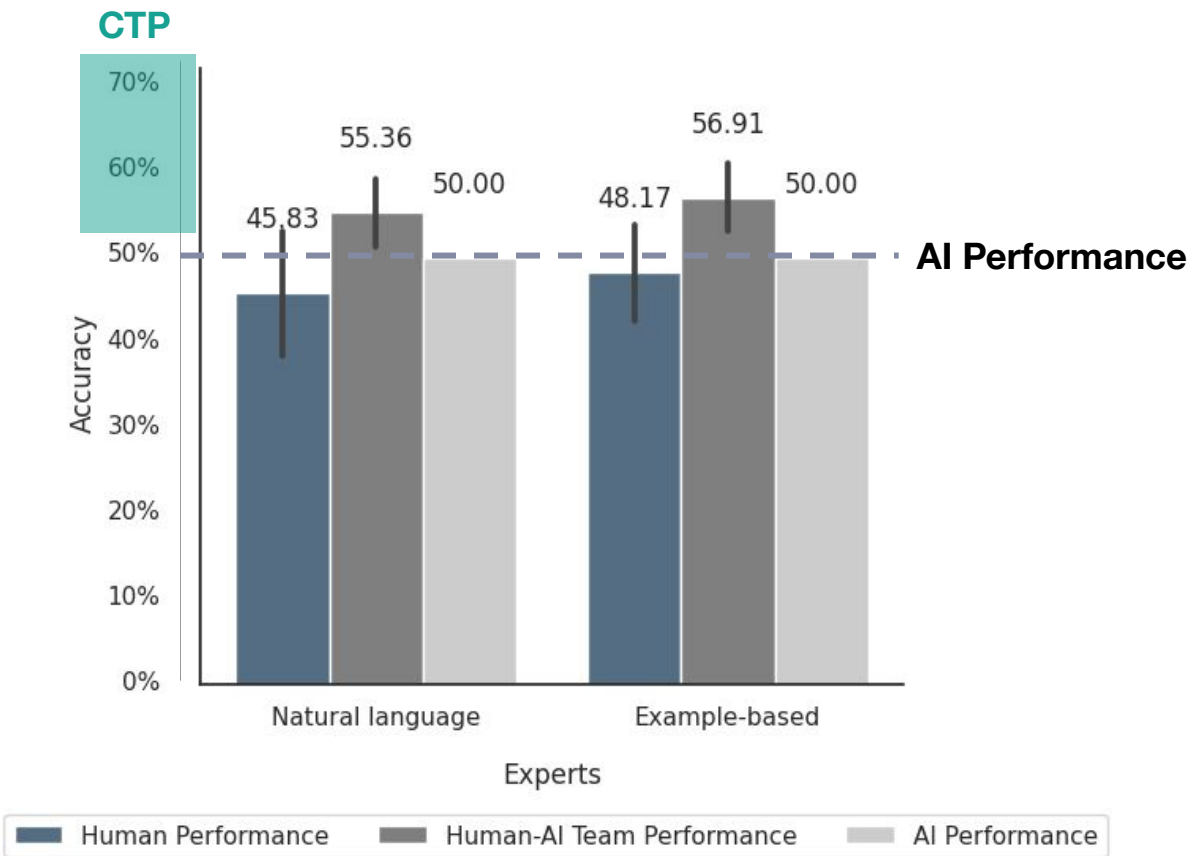


Complementary Team Performance (CTP) for Experts

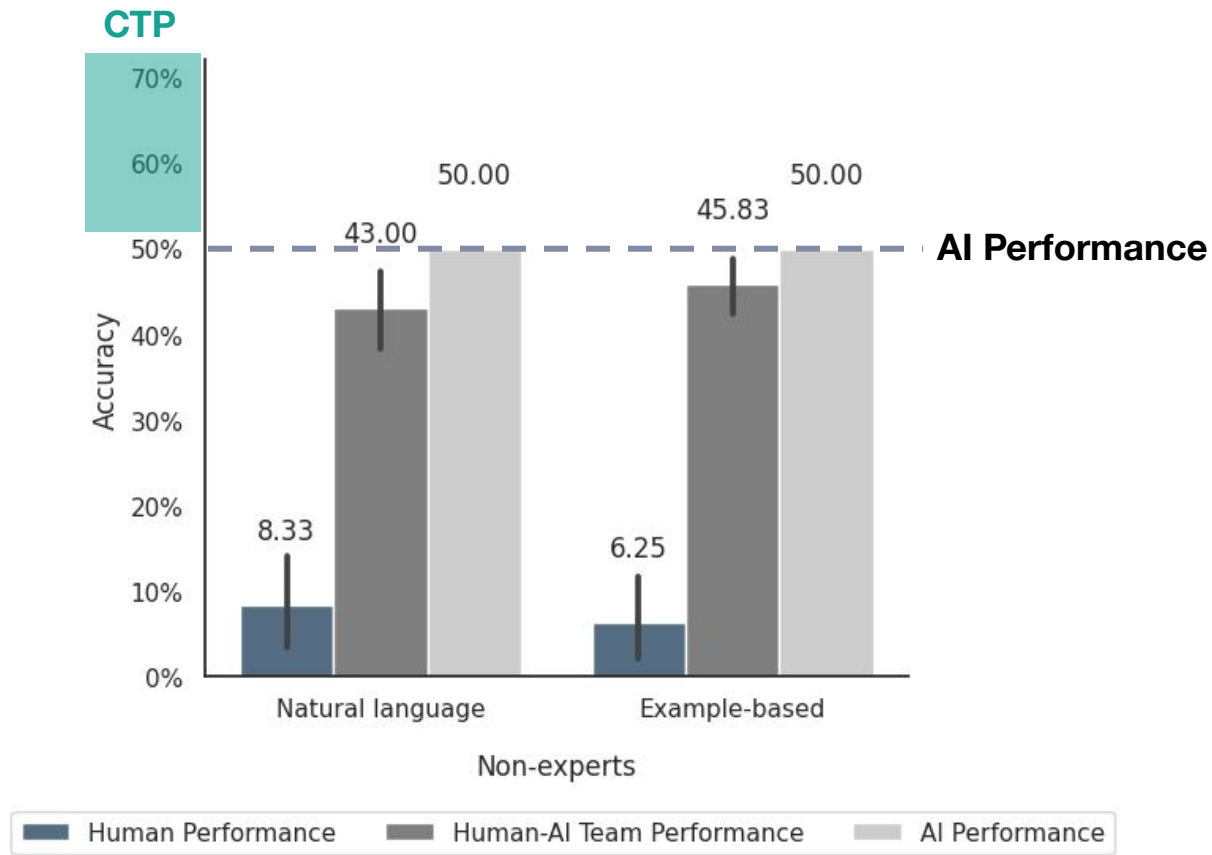


AI Performance

Complementary Team Performance (CTP) for Experts



Complementary Team Performance (CTP) for Non-Experts



Complementary Team Performance (CTP) & Expertise

How is **complementary team performance** impacted by the decision-maker's **level of domain expertise**?



AI

experts



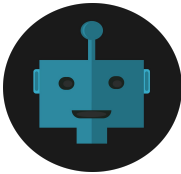

AI

non-experts

Complementary Team Performance (CTP) & *Imperfect* XAI

- ↓ Achieves CTP
- ↓ Does not achieve CTP

Human-AI Team Performance

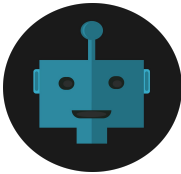

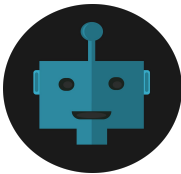

		<i>Natural Language</i>		<i>Example-based</i>	
		<i>Correct</i>	<i>Incorrect</i>	<i>Correct</i>	<i>Incorrect</i>
		~57%	~54%	~59%	~55%
AI	experts		↓		↓

Experts are able to rely on their own expertise when AI or XAI advice is incorrect

Complementary Team Performance (CTP) & *Imperfect* XAI

- ↓ Achieves CTP
- ↓ Does not achieve CTP

Human-AI Team Performance

		<i>Natural Language</i>		<i>Example-based</i>	
		<i>Correct</i>	<i>Incorrect</i>	<i>Correct</i>	<i>Incorrect</i>
		~57%	~54%	~59%	~55%
AI	experts		↓		↓
		~49%	~37%	~49%	~43%
AI	non-experts		↓		↓

Complementary Team Performance (CTP) & *Imperfect* XAI

How is **complementary team performance** impacted by the **correctness of explanations**?



Achieves CTP

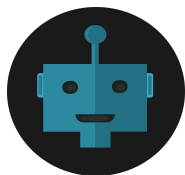


Does not achieve CTP

Complementary Team Performance & Impact

Natural Language

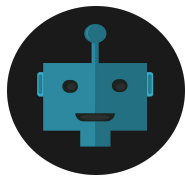
Example-based



AI



experts



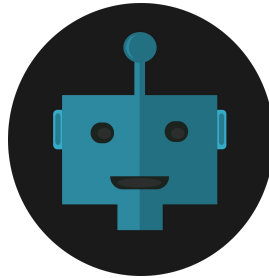
AI



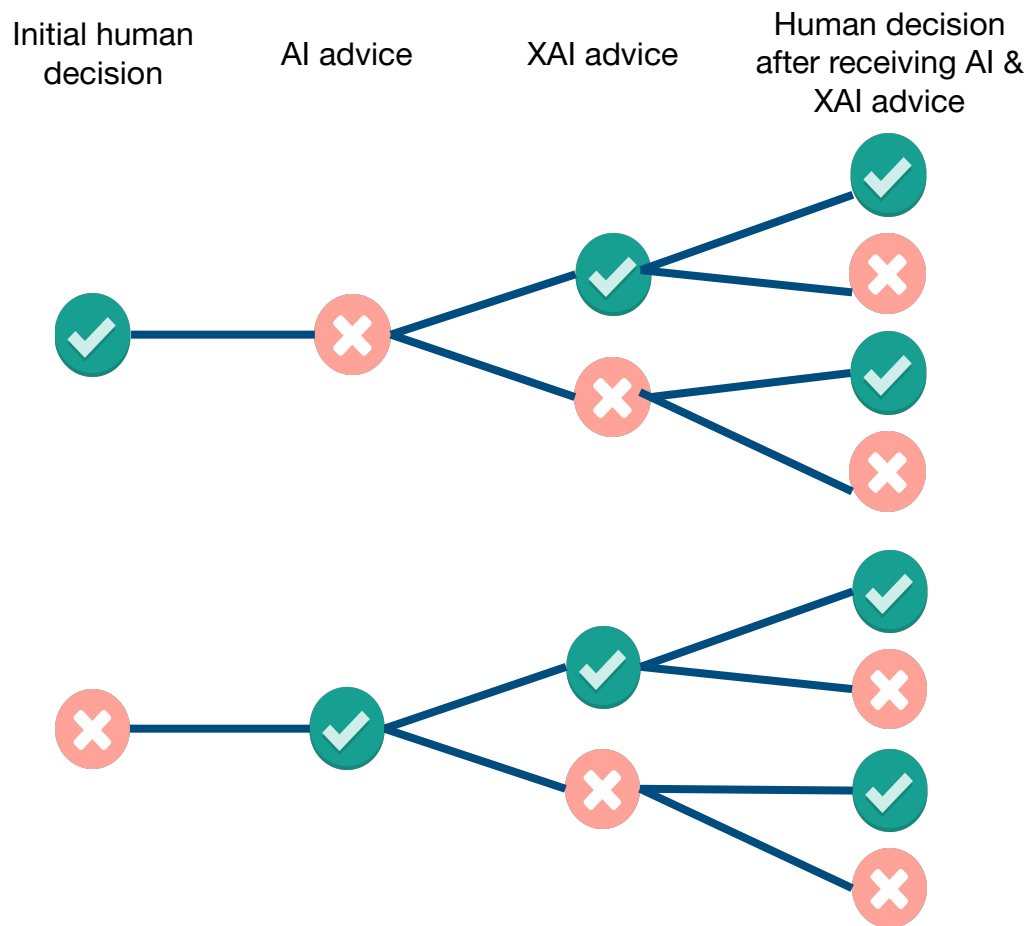
non-experts



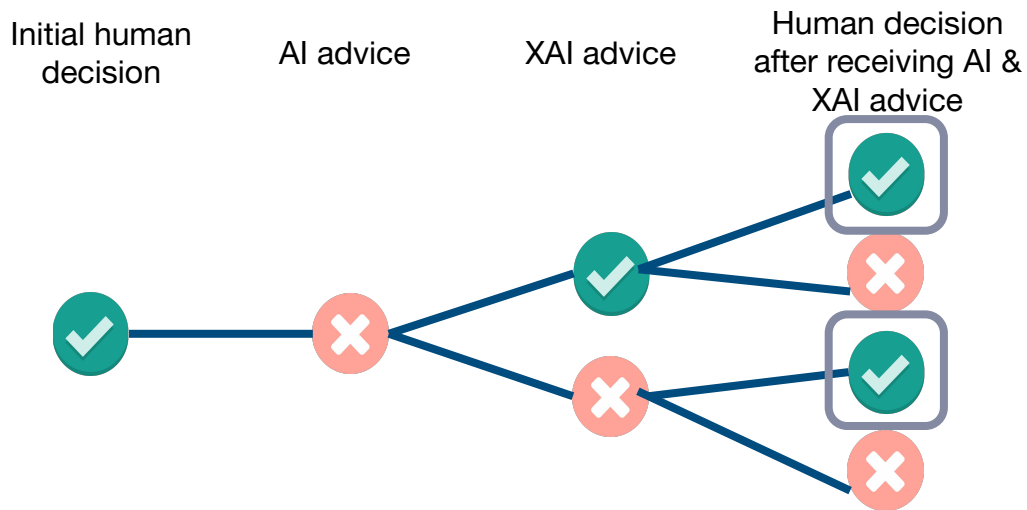
Imperfect XAI can be deceiving...



The Deception of *Imperfect* XAI



The Deception of *Imperfect* XAI

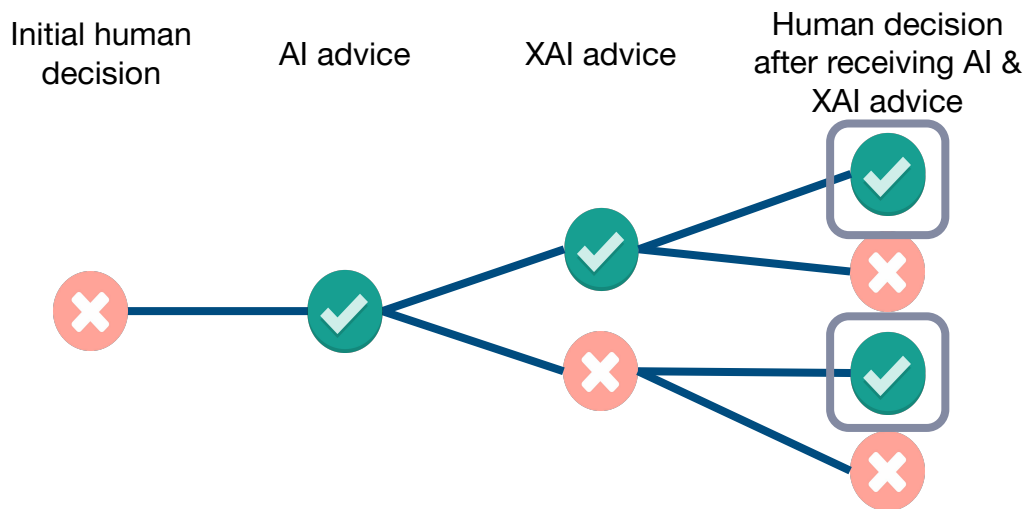


Relative Self-Reliance

equation hidden for simplicity

how often the human correctly relies
on their own decision out of all the
times the AI advice is incorrect

The Deception of *Imperfect* XAI

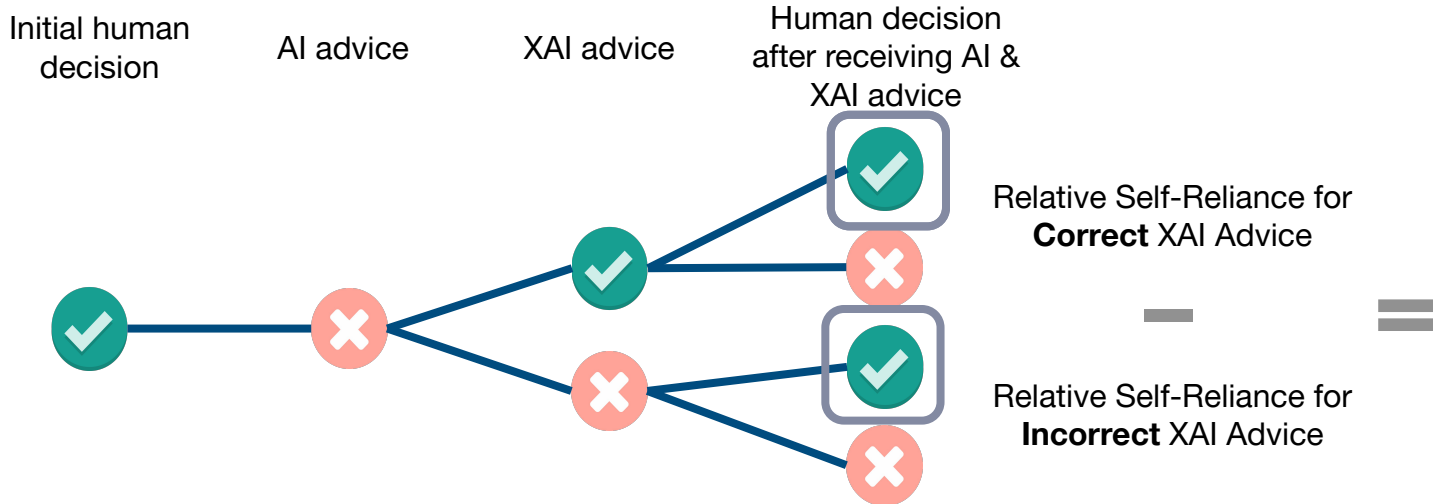


Relative AI-Reliance

equation hidden for simplicity

how often the human correctly relies
on the AI out of all the times when the AI
advice is correct

The Deception of *Imperfect* Example-Based Explanations



Deception of Reliance for Relative Self-Reliance

Relative Self-Reliance for **Incorrect** XAI Advice

AI: Yellow-breasted Chat

Incorrect Prediction

Correct Explanation

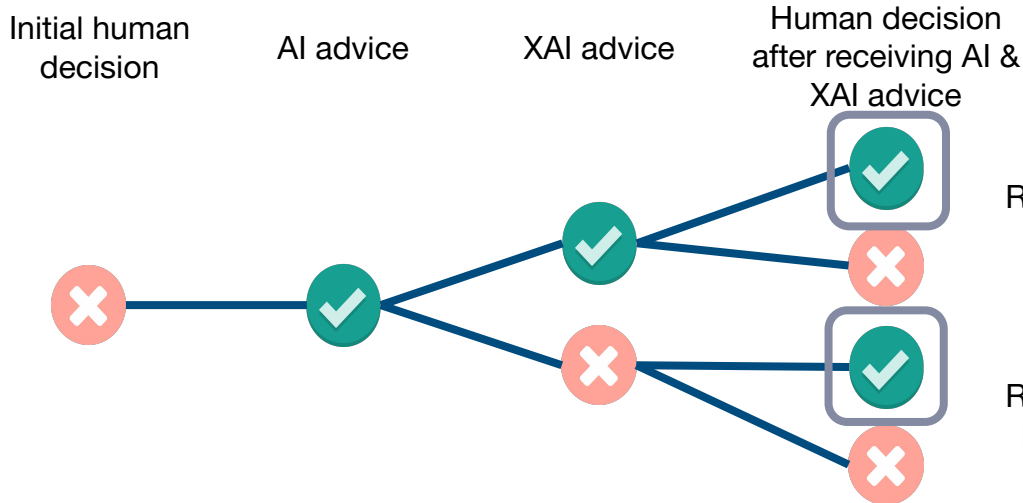


experts

Example-Based Explanations

-0.28***

The Deception of *Imperfect* Example-Based Explanations



Relative AI-Reliance for
Correct XAI Advice

Relative AI-Reliance for
Incorrect XAI Advice

Deception of Reliance for
Relative AI-Reliance

Example-Based Explanations

AI: Nashville Warbler



Correct Prediction



Painted Bunting



Yellow-breasted Chat



American Redstart

Incorrect Explanation

experts



non-experts



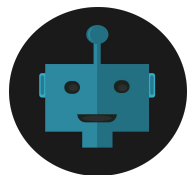
0.16*

0.26**

The **Deception** of *Imperfect* Example-Based Explanations

To what extent do **incorrect and correct explanations deceive** decision-makers with different levels of expertise?

Deception of Reliance: Example-Based Explanations



AI



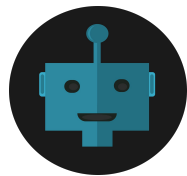
experts

Relative Self-Reliance

-0.28***

Relative AI-Reliance

0.16*



AI

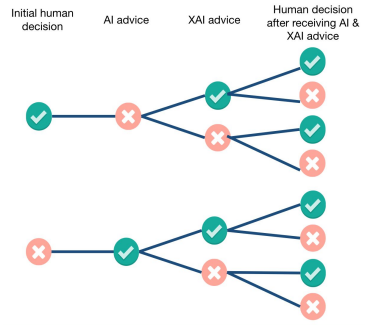


non-experts

0.26**

Summary of Contributions & Findings

Conceptualization of Human-AI Collaboration with Imperfect XAI



Level of Expertise

Complementary Team Performance

natural language example-based



AI experts



AI non-experts

Correctness of Explanations

Impact on Complementary Team Performance

↓ CTP
↓ GTP

natural language example-based



AI experts



AI non-experts

Deception of Reliance

Example-based explanations

Relative Self-Reliance Relative AI-Reliance



AI experts

-0.28*** 0.16*



AI non-experts

0.26**

Summary of Contributions & Findings

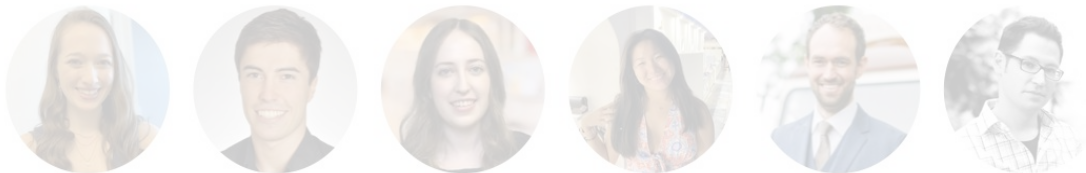
Guide researchers & practitioners on
how to **assess** and **design for** *imperfect*
XAI in human-AI collaborations

The Impact of *Imperfect* XAI on Human-AI Decision-Making

Thank you!

Katelyn Morrison^{1*}, Philipp Spitzer^{2*}, Violet Turri¹,
Michelle Feng¹, Niklas Kuhl³, Adam Perer¹

1: Carnegie Mellon University; 2: Karlsruhe Institute of Technology; 3: University of Bayreuth
* equal contribution



kcmorris@cs.cmu.edu

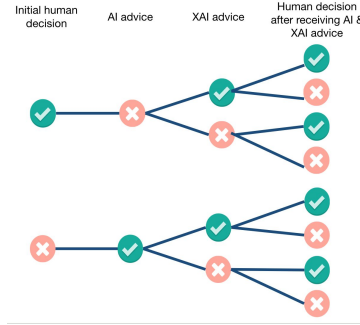
This work is under review at CSCW 2024

We would like to attribute this amazing work to the existence of networking events at in-person conferences because Philipp & Katelyn met at CHI and sparked up this wonderful collaboration.

Acknowledgements: Max Schemmer, Hao-Fei Cheng, Haiyi Zhu, Ken Holstein, KSRI Lab, DIG Lab, Nari Johnson, Youwei Jiang, and multiple experienced birders for their insightful discussions and guidance. And ChatGPT & DALL-E 2.

Summary of Contributions & Findings

Conceptualization of Human-AI Collaboration with Imperfect XAI



Questions?

Level of Expertise

Complementary Team Performance

natural language example-based



AI experts



AI non-experts

Correctness of Explanations

Impact on Complementary Team Performance

natural language example-based



AI experts



AI non-experts

Deception of Reliance

Example-based explanations

Relative Self-Reliance Relative AI-Reliance



AI experts

-0.28*** 0.16*



AI non-experts

0.26**

Can AI *explain* another AI's output?



LABELS: Atelectasis (Uncertain) and Pneumonia (Uncertain)

Natural Language Explanations for *Pneumonia*:

Ground-Truth: *Interval appearance of patchy opacity at the left base could represent early pneumonia, although aspiration or patchy atelectasis would also be in the differential.*

RATCHET: *Patchy opacities in the lung bases may reflect atelectasis, but infection is not excluded in the correct clinical setting.*

Kayser, Maxime, et al. "Explaining chest x-ray pathologies in natural language." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2022.



This is a pine grosbeak because this bird has a red head and breast with a gray wing and white wing.



This is a Kentucky warbler because this is a yellow bird with a black cheek patch and a black crown.

Hendricks, Lisa Anne, et al. "Generating visual explanations." *Computer Vision–ECCV*. 2016.

How do you generate the Natural Language Explanations

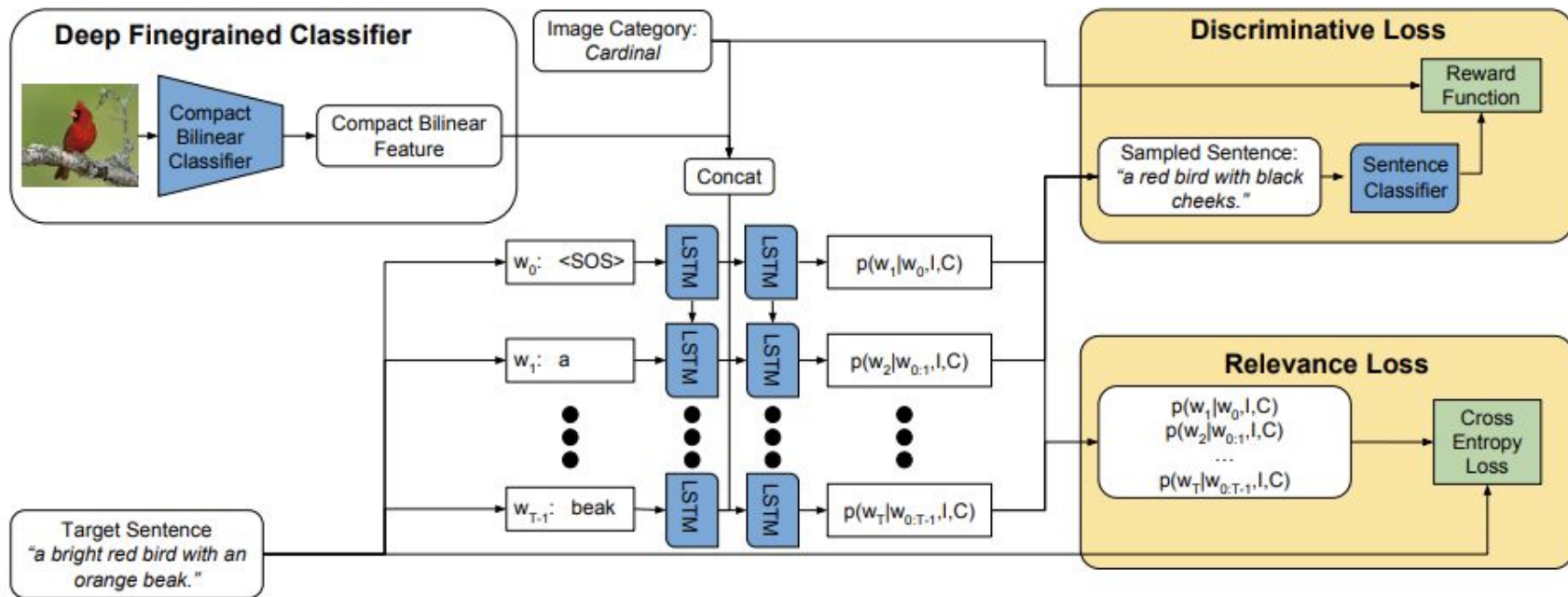


Fig. 3. Training our explanation model. Our explanation model differs from other caption models because it (1) includes the object category as an additional input and (2) incorporates a reinforcement learning based discriminative loss

Explanations

Examples of example-based explanations for each scenario

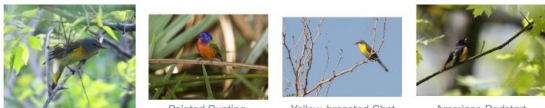
AI: Magnolia Warbler



Correct Prediction

Correct Explanation

AI: Nashville Warbler



Correct Prediction

Incorrect Explanation

AI: Yellow-breasted Chat



Incorrect Prediction

Correct Explanation

AI: Dark-eyed Junco



Incorrect Prediction

Incorrect Explanation

Examples of natural language explanations for each scenario

AI: Magnolia Warbler



"this is a bird with a **yellow belly** **black stripes** on its breast and a **grey head**"

Correct Prediction

Correct Explanation

AI: Cerulean Warbler



"this is a small bird with a **white belly** and a **grey head**"

Correct Prediction

Incorrect Explanation

AI: Bewick Wren



"this is a **small brown** bird with a **white eyebrow** and a **downward pointing beak**"

Incorrect Prediction

Correct Explanation

AI: Yellow-billed Cuckoo



"this is a **white** bird with a **black wing** and a **large black beak**"

Incorrect Prediction

Incorrect Explanation

Explanations

Format of Natural Language Explanations

This might be a {predicted class},
since it appears to be a {explanation}

non-assertive

This is definitely a {predicted class}
because it clearly is a {explanation}

assertive

This is a {predicted class}
because {explanation}

neutral

Format of Example-Based Explanations

This might be a {predicted class},
since it appears to be ...

non-assertive

This is definitely a {predicted class}
because it is clearly ...

assertive

This is a {predicted class}
because it is ...

neutral

... similar to the birds in the following three examples

Phase 1



Name the bird species:

Canada Warbler

How confident do you feel in your decision?

Not confident 1 2 3 4 5 6 Very confident

Next

Phase 2



AI Prediction: Magnolia Warbler

Show AI Explanation

This is definitely a Magnolia Warbler because it is clearly similar to the birds in the following three examples:



Name the bird species:

Magnolia Warbler

How confident do you feel in your decision?

Not confident 1 2 3 4 5 6 Very confident

Next

Please **classify the three birds** below to the best of your ability. To be eligible to participate in this study, you must classify at least two (of the three) common "easy" birds correctly. For the common "easy" birds, you only need to correctly guess the family name of the bird, not the exact species name.

Common "easy" birds



Search a bird species



Search a bird species



Search a bird species



You correctly identified the family for at least two of the three common "easy" birds. These next three birds are harder to classify than the last three. Please **classify the three birds** below to the best of your ability. This will help us better understand your knowledge and expertise in birding.

Common "hard" birds



Search a bird species



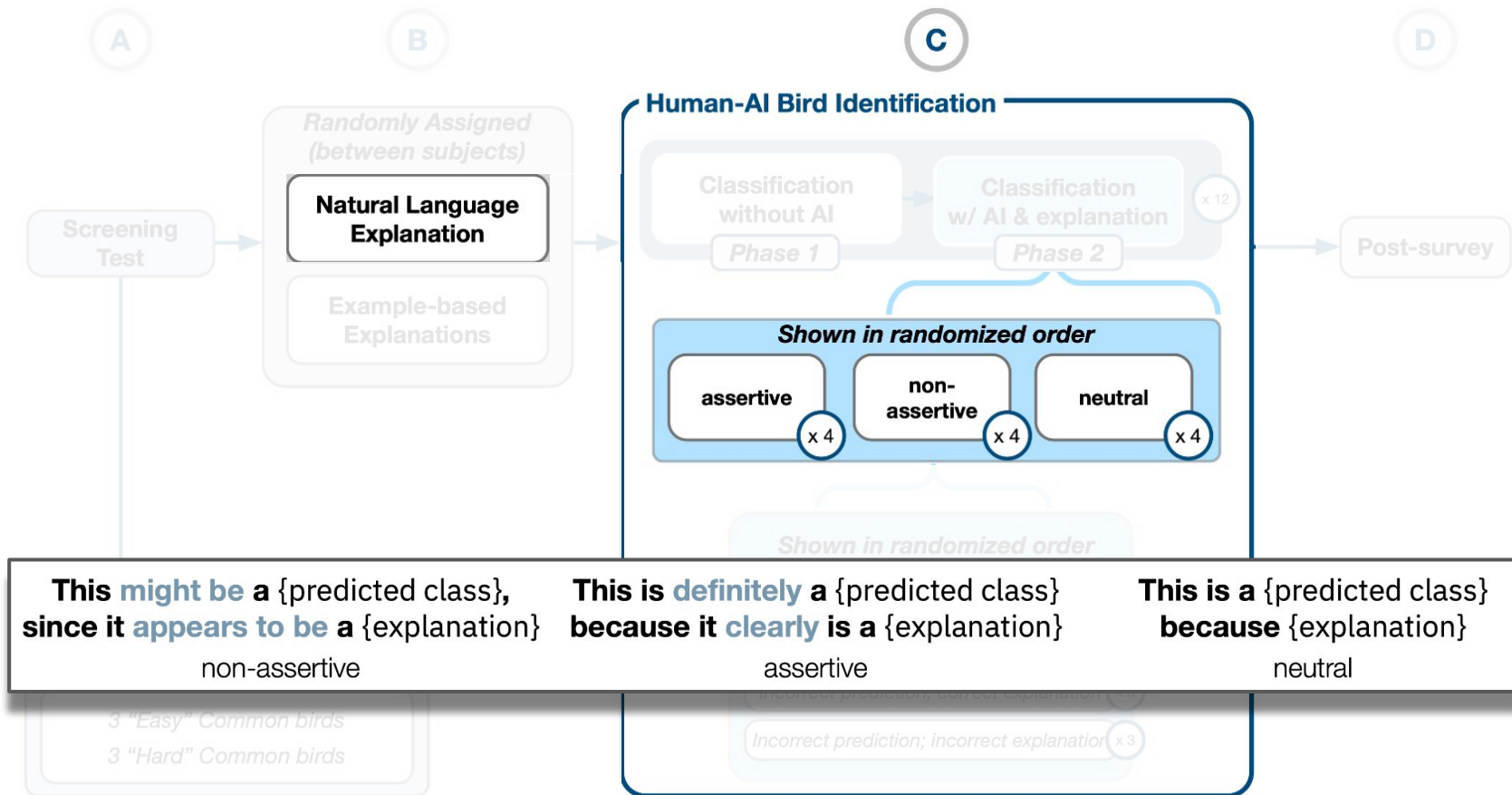
Search a bird species



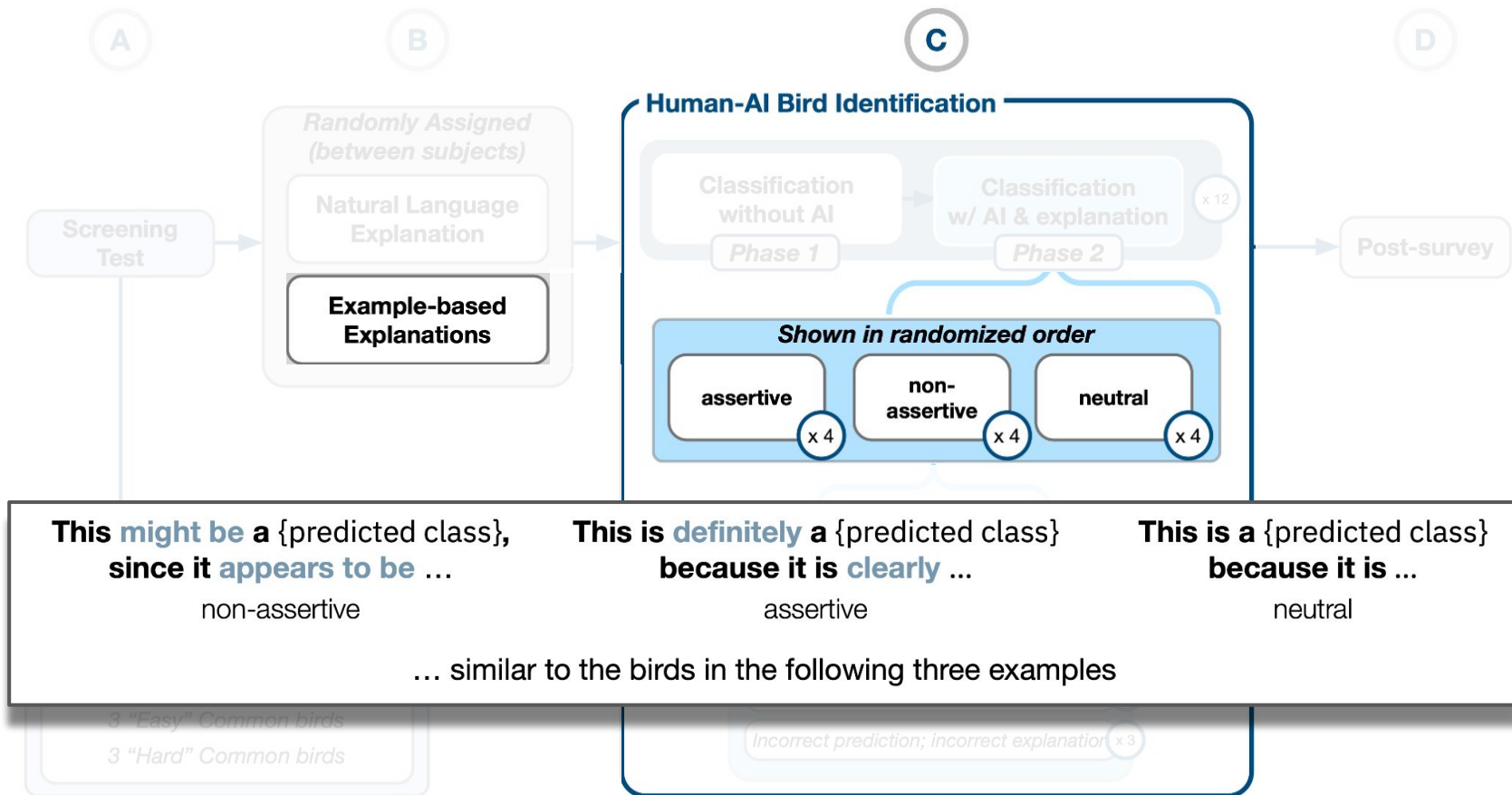
Search a bird species



Study Design: Assertiveness

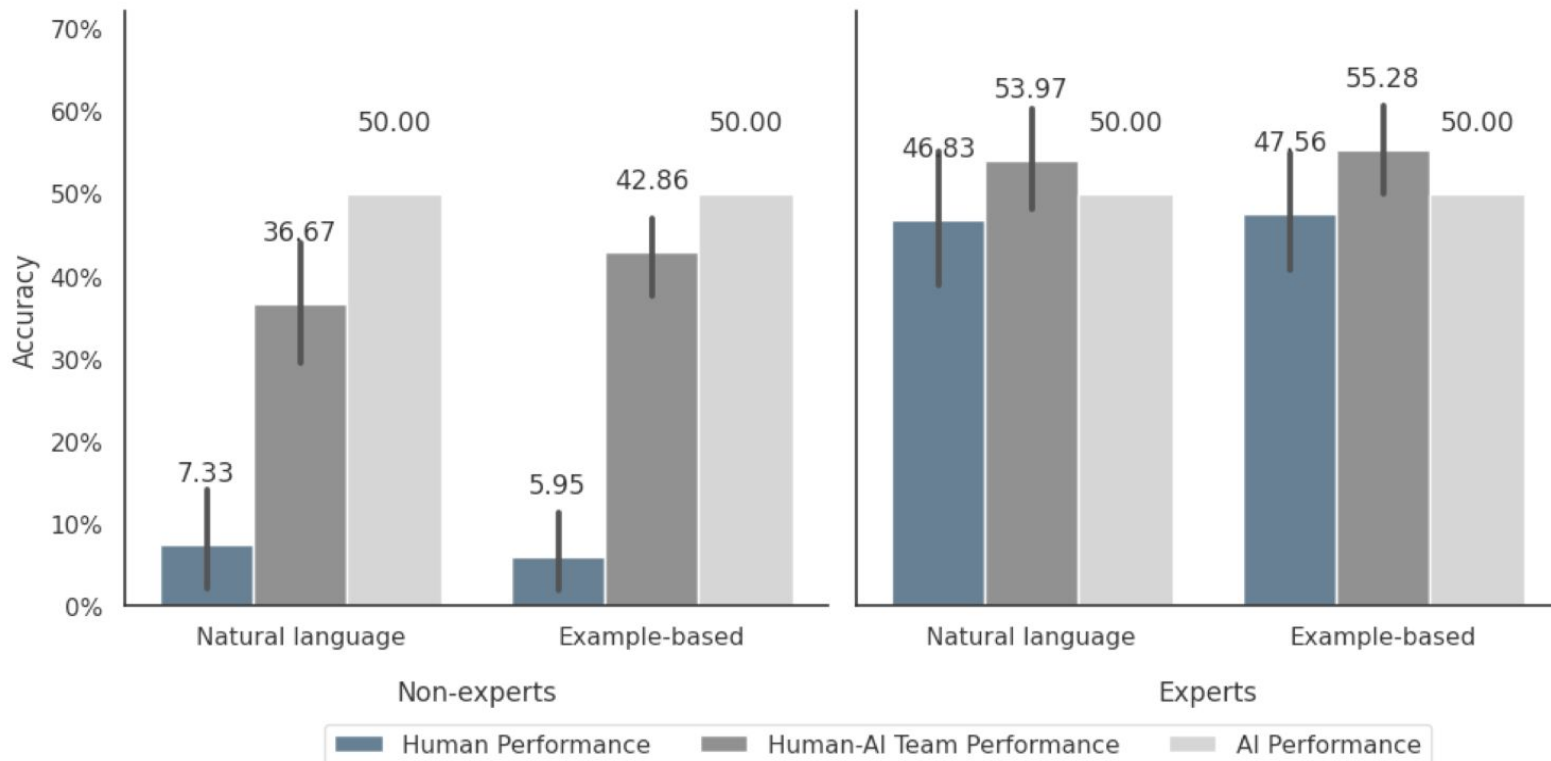


Study Design: Assertiveness

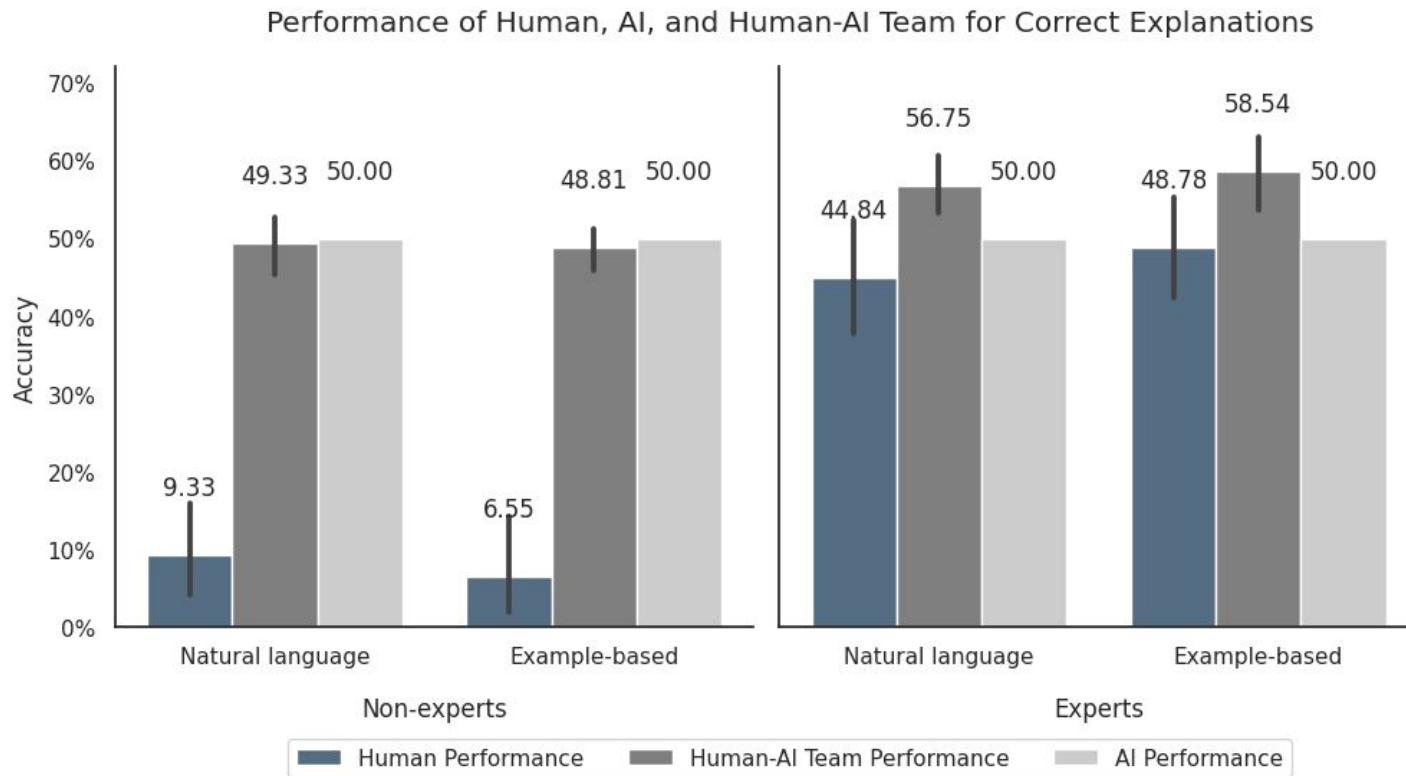


Complementary Team Performance (CTP) & *Imperfect* XAI

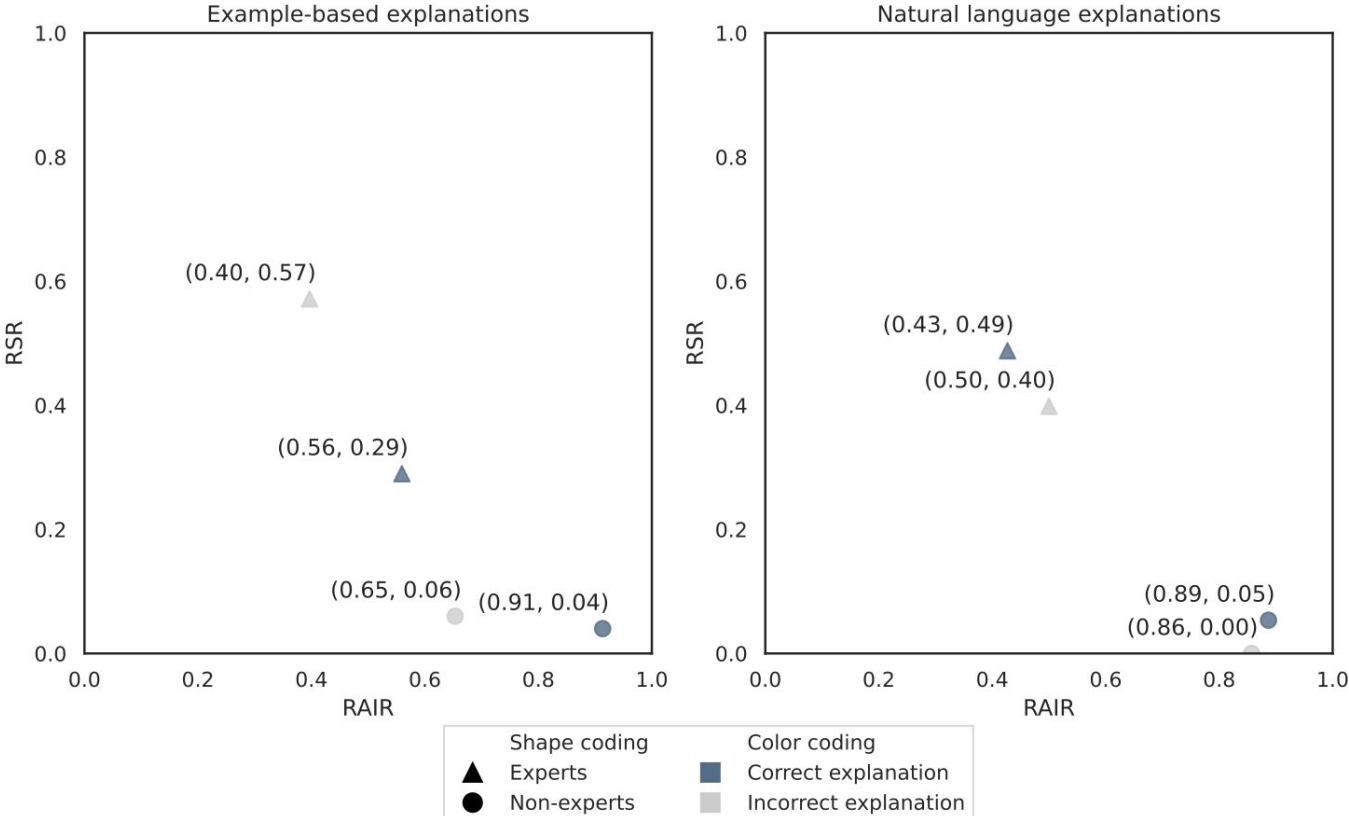
Performance of Human, AI, and Human-AI Team for Incorrect Explanations



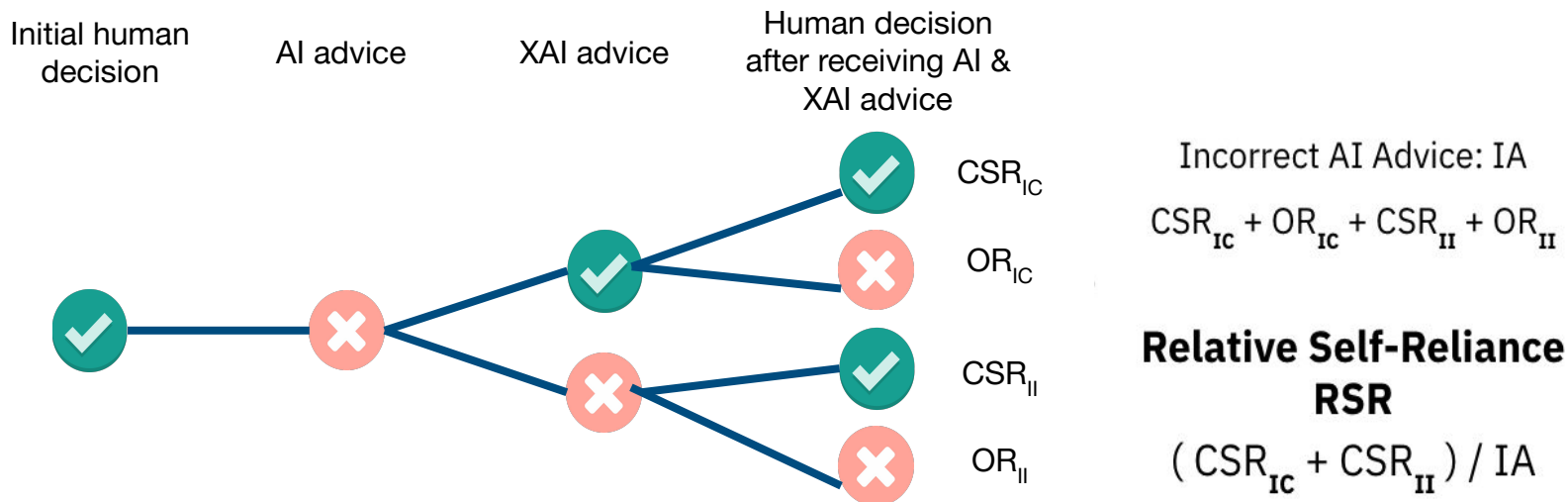
Complementary Team Performance for Non-Experts



Deception of Reliance Significance Tests

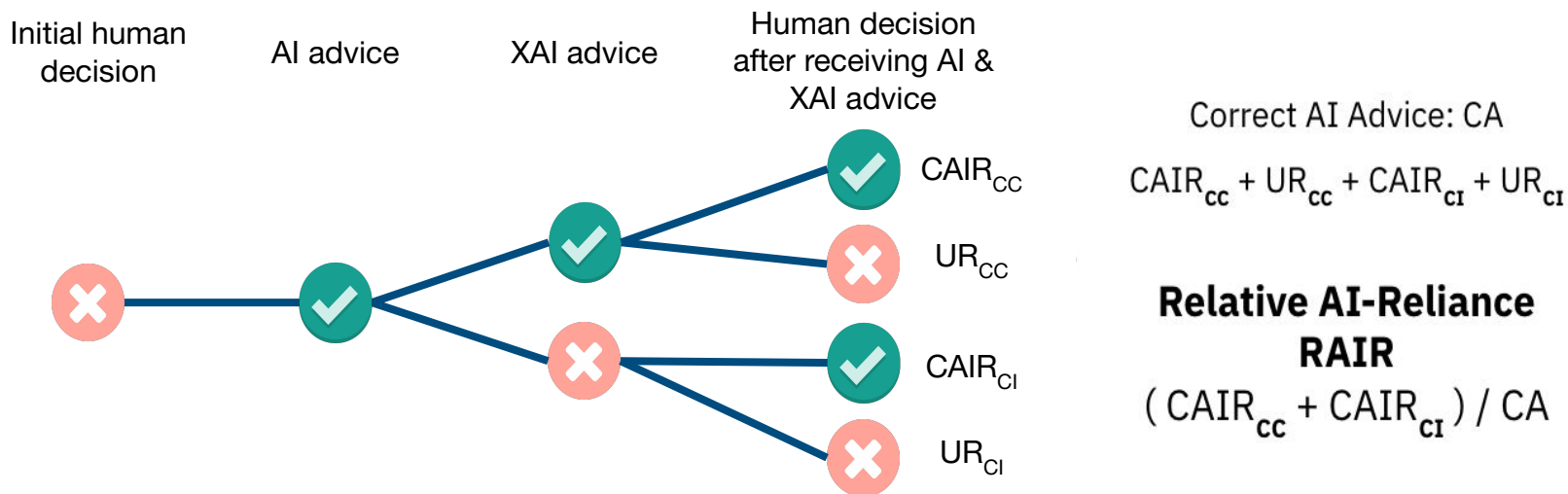


The Deception of *Imperfect* XAI



how often the human correctly relies on their own decision out of all the times the AI advice is incorrect

The Deception of *Imperfect* XAI



how often the human correctly relies on the AI out of all the times when the AI advice is correct