SAND2022-13963C

Sandia National Laboratories

Exceptional service in the national interest

Cognitive Impacts of Computer Vision-Based Decision Support for International Nuclear Safeguards-Relevant Visual Analysis Tasks

Z.N. Gastelum,* L.E. Matzen, K.M. Divis, B. Howell

*CORRESPONDING AUTHOR: ZGASTEL@SANDIA.GOV

November 3, 2022

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Three Main Findings

1. Imperfect models provide decision support
2. Users of high performing models sometimes miss subtle but important errors
3. User expertise impacts response to model errors

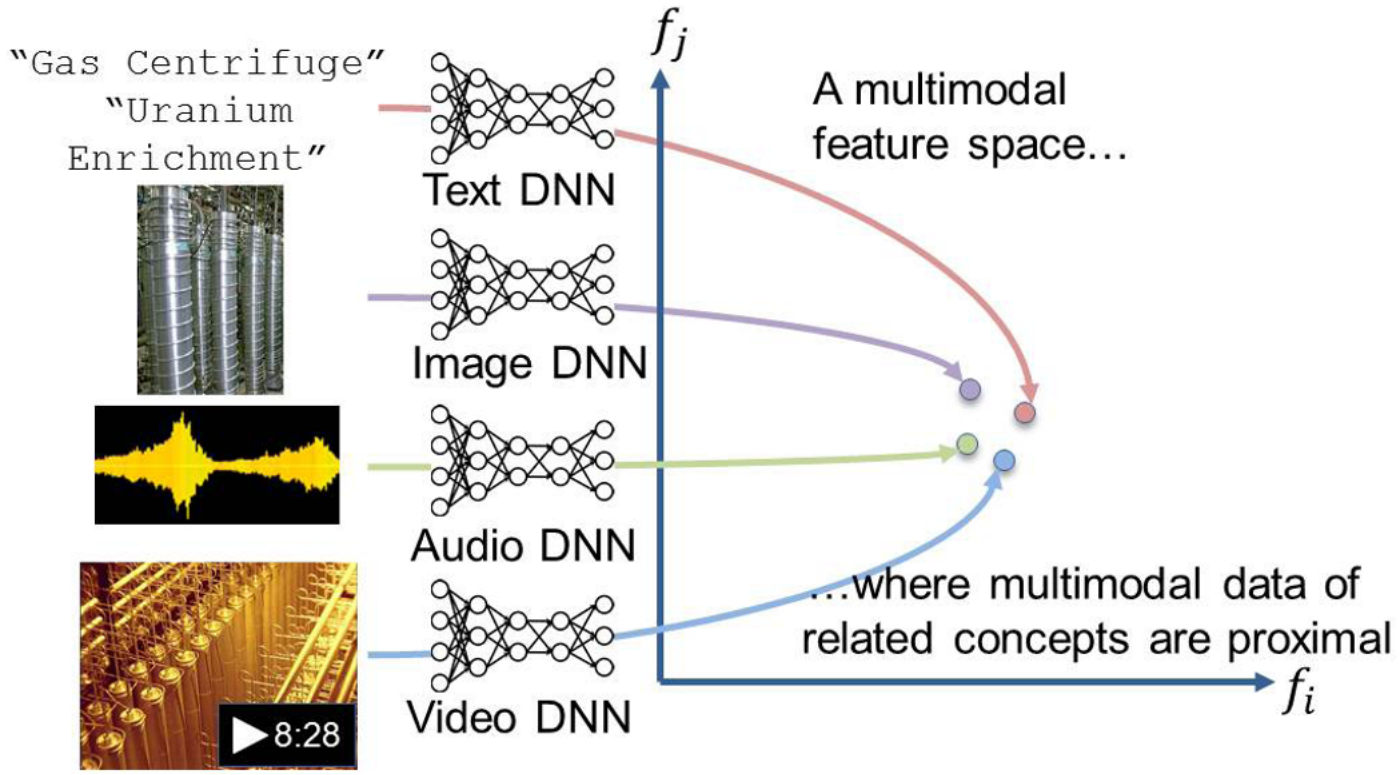
DL model performance can be very high, surpassing humans.

Computer vision models are being developed for safeguards-relevant tasks

- Open-source image review

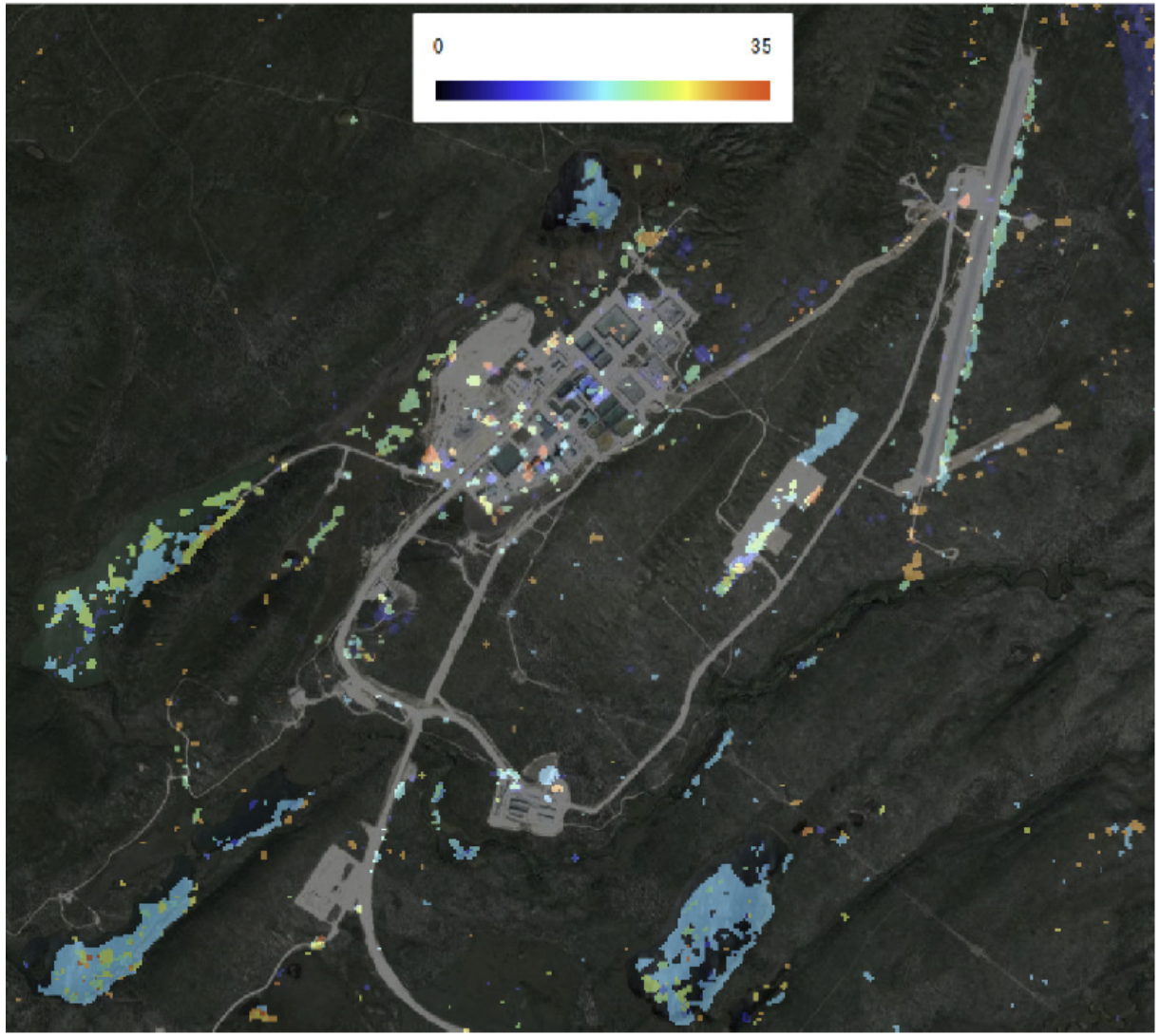
Computer vision models are being developed for safeguards-relevant tasks

- Open-source image review
- Multi-modal information retrieval



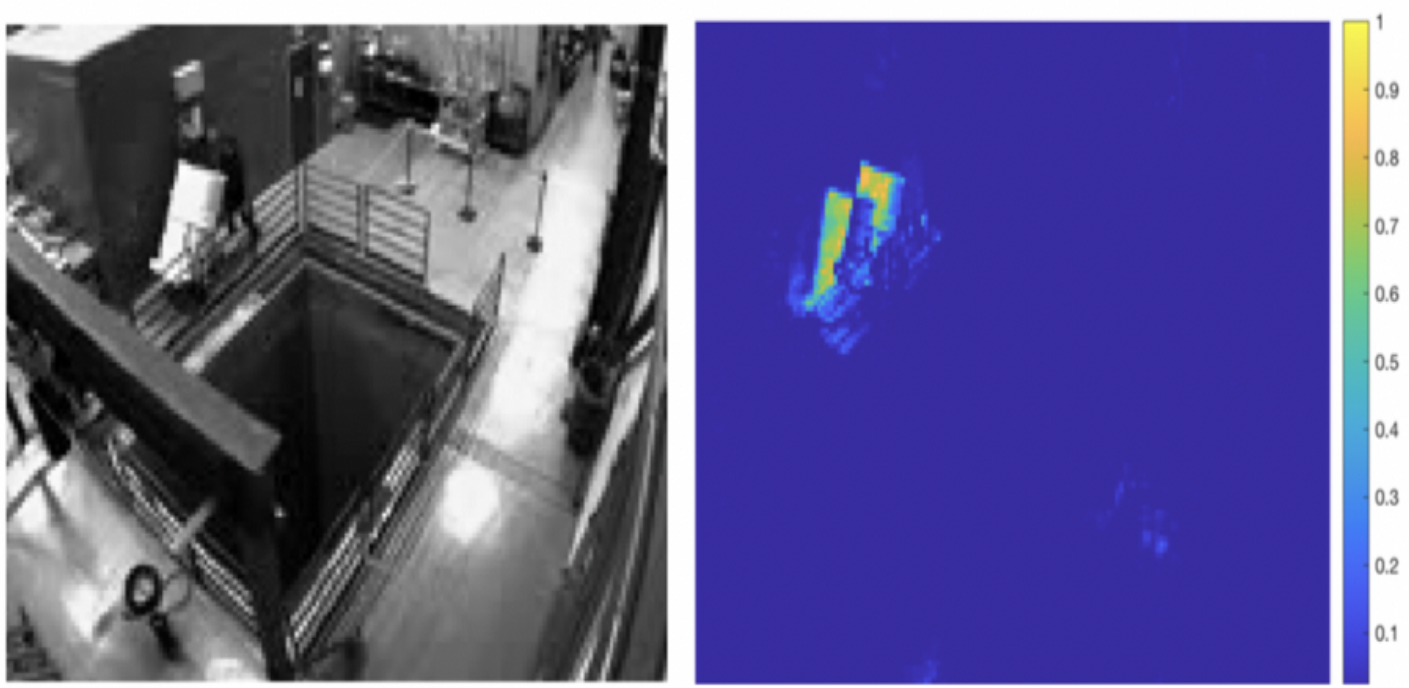
Computer vision models are being developed for safeguards-relevant tasks

- Open-source image review
- Multi-modal information retrieval
- Overhead imagery monitoring



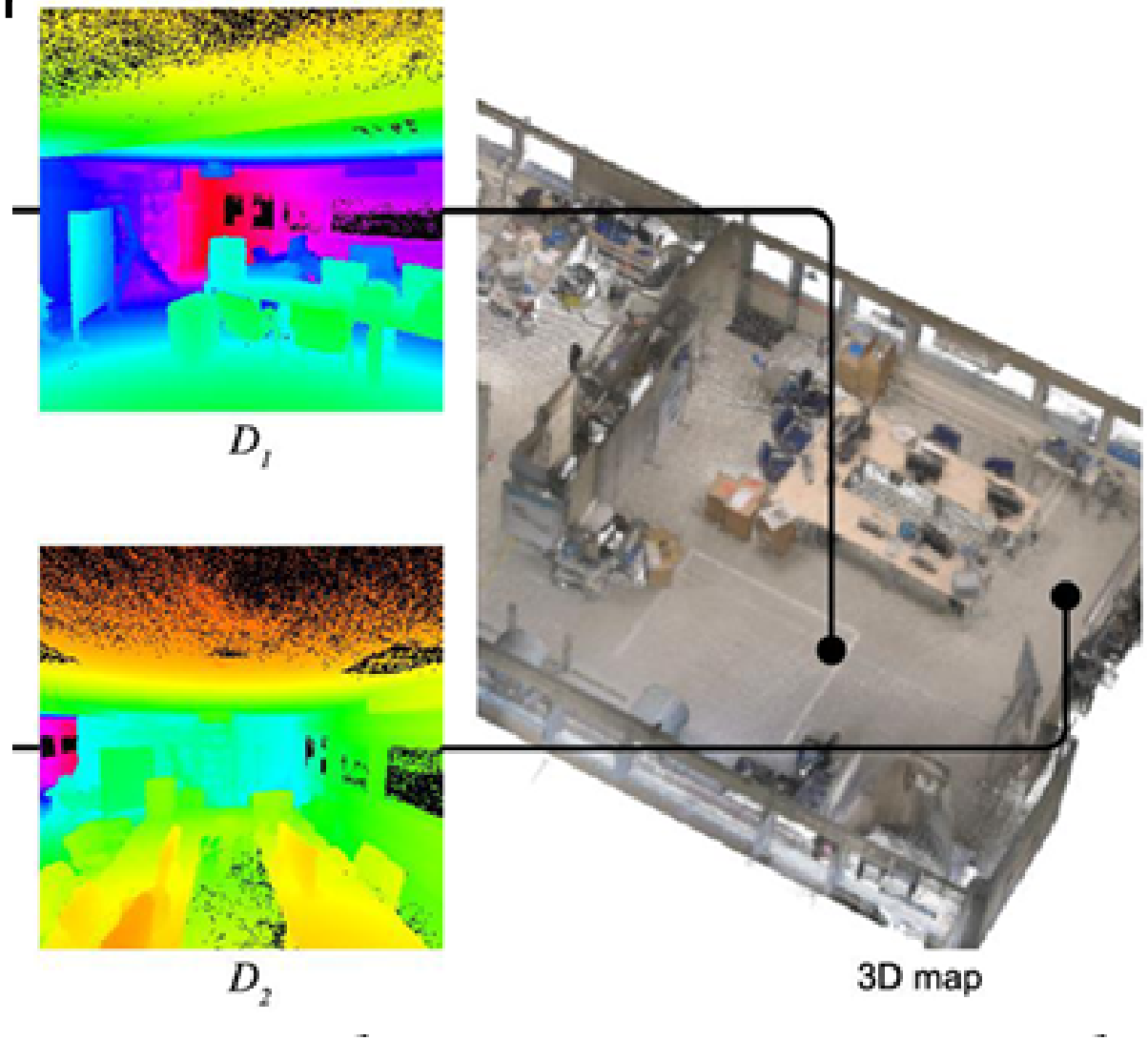
Computer vision models are being developed for safeguards-relevant tasks

- Open-source image review
- Multi-modal information retrieval
- Overhead imagery monitoring
- Surveillance camera review



Computer vision models are being developed for safeguards-relevant tasks

- Open-source image review
- Multi-modal information retrieval
- Overhead imagery monitoring
- Surveillance camera review
- Localization and wayfinding

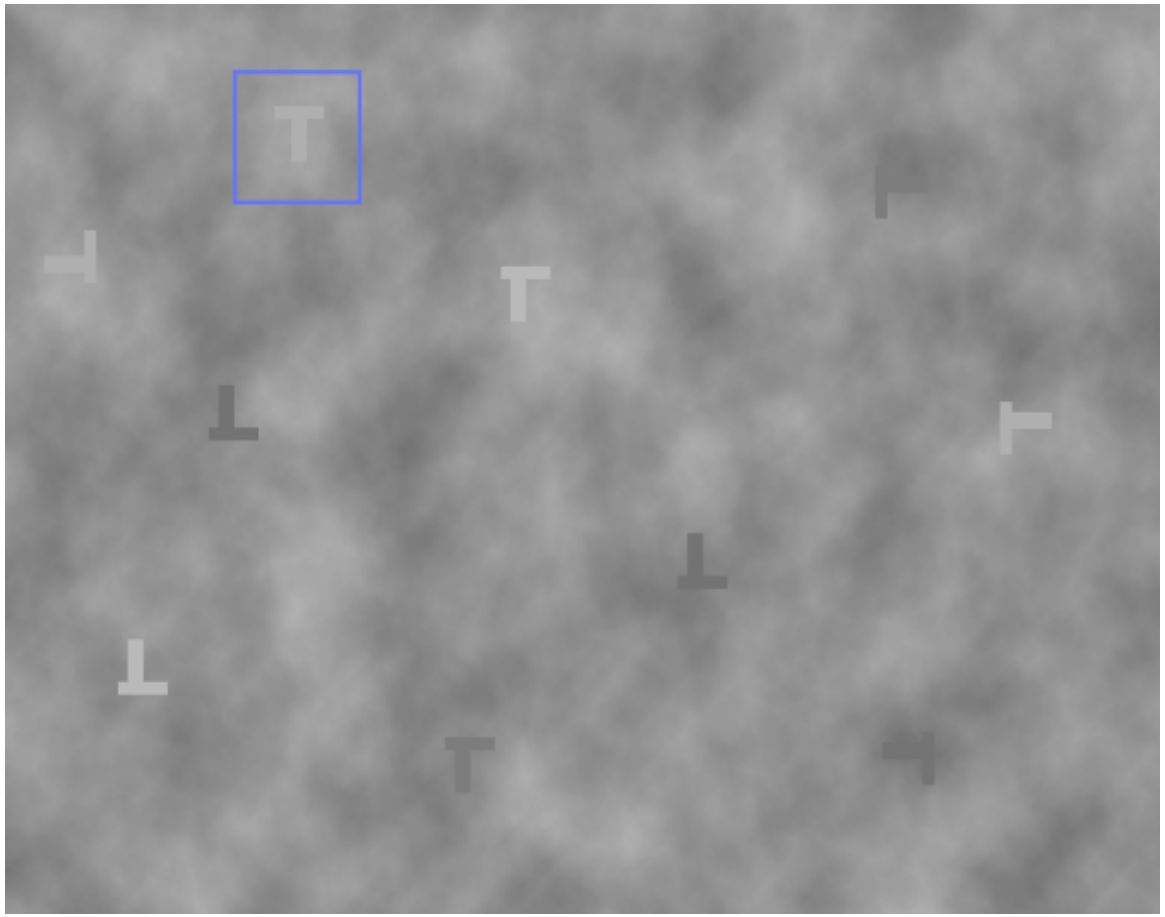


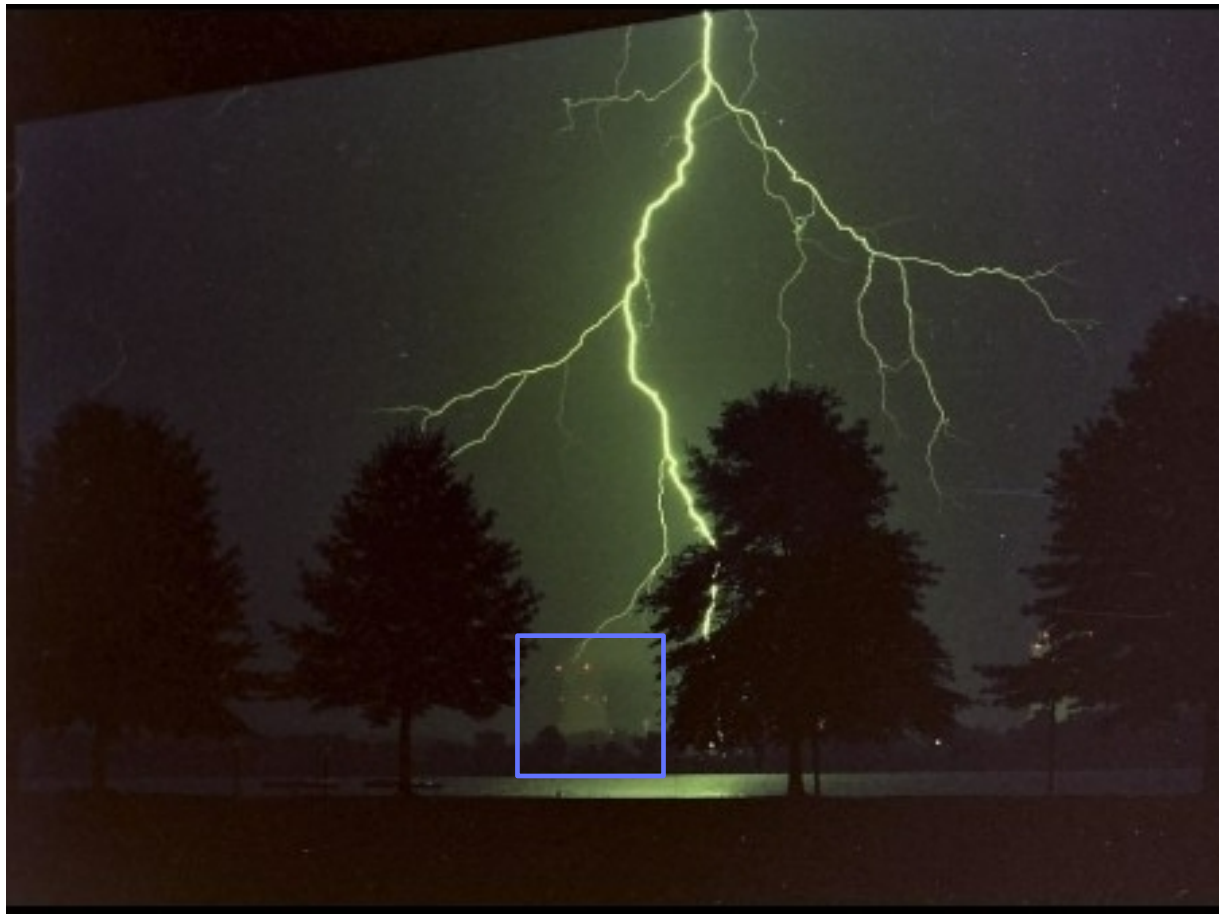
Key Research Questions

For the implementation of ML models to be effective, we need a better understanding of the impact of AI/ML errors on human users

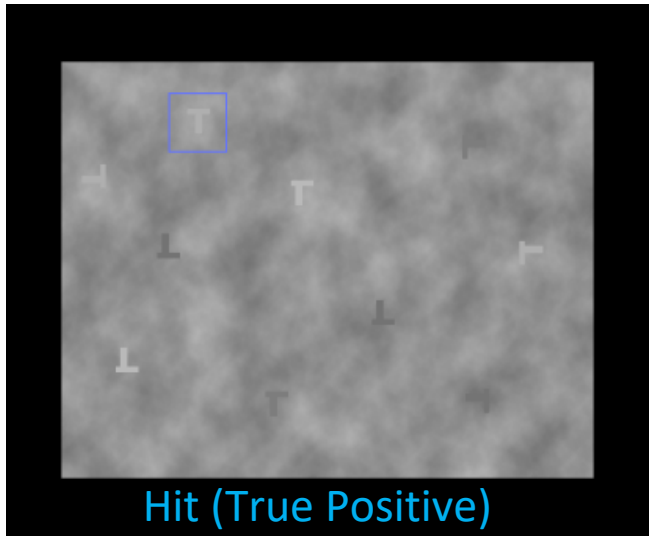
- When and how do errors in AI/ML outputs lead to errors in human assessments?
- What factors make it easier or harder for people to recognize errors?
- How do people develop appropriate levels of trust in the outputs?
- What level of accuracy in the model outputs is necessary to support acceptable levels of human/system performance?

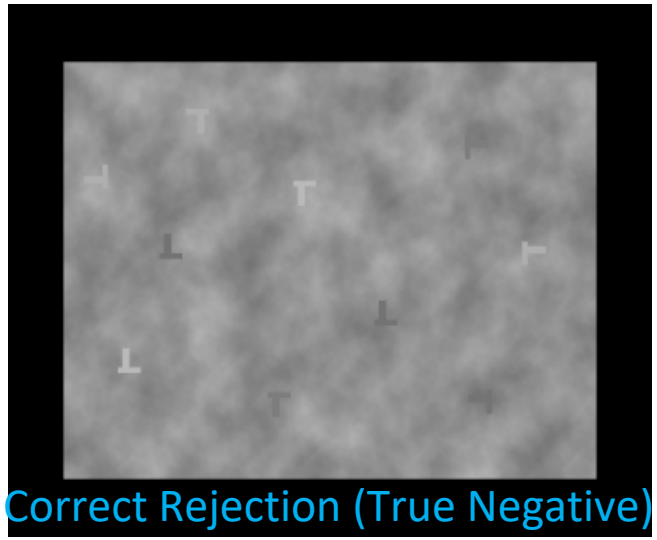
We use human performance testing on visual search tasks to test how humans interact with models

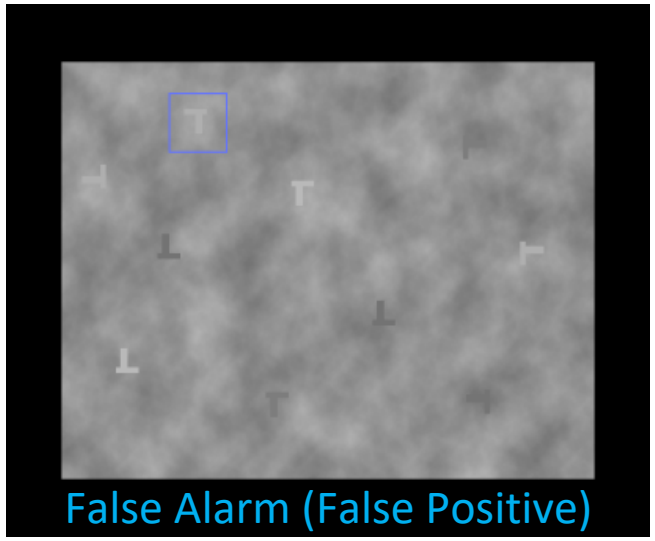


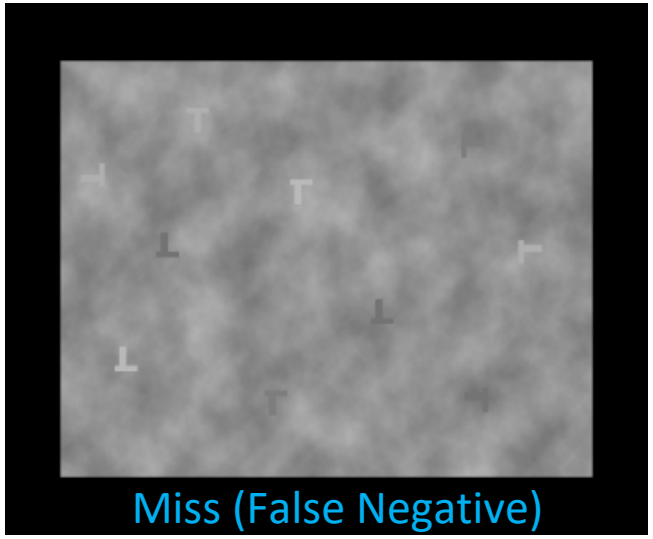


Manipulation of Model Output Accuracy

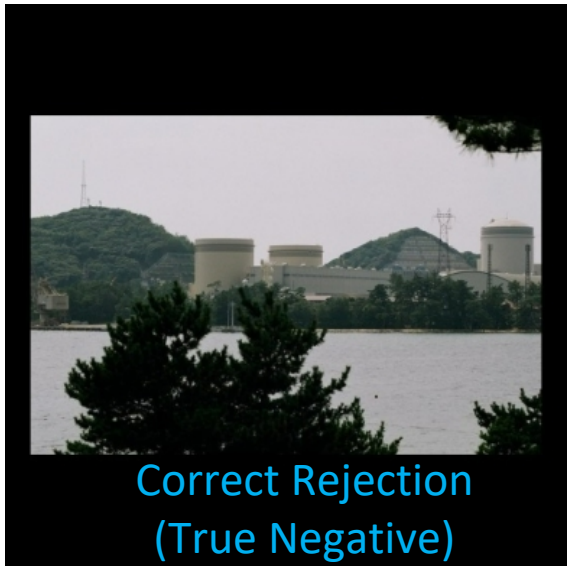

Hit (True Positive)


Correct Rejection (True Negative)


False Alarm (False Positive)


Miss (False Negative)


Hit (True Positive)

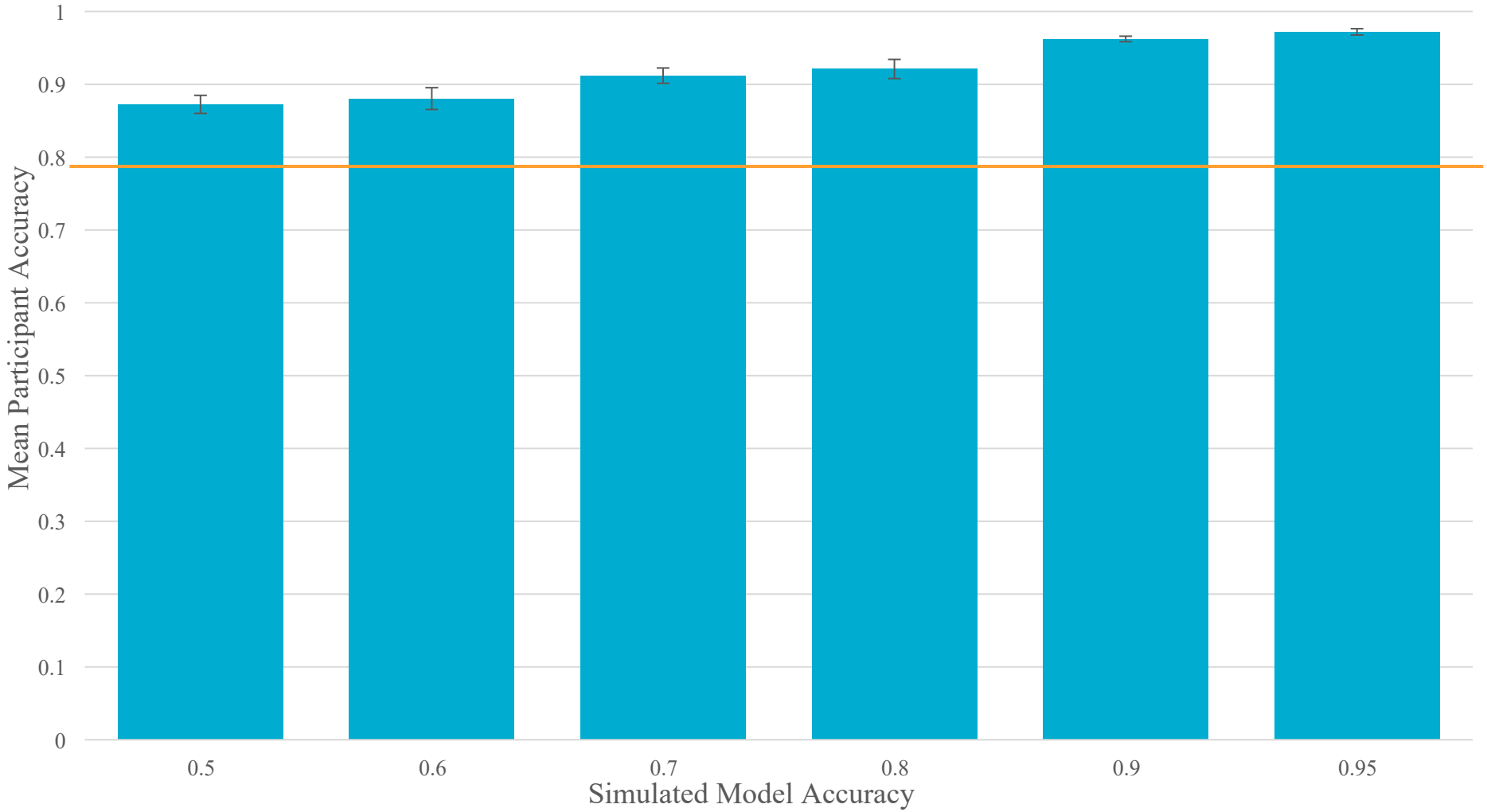

Correct Rejection (True Negative)


False Alarm (False Positive)


Miss (False Negative)

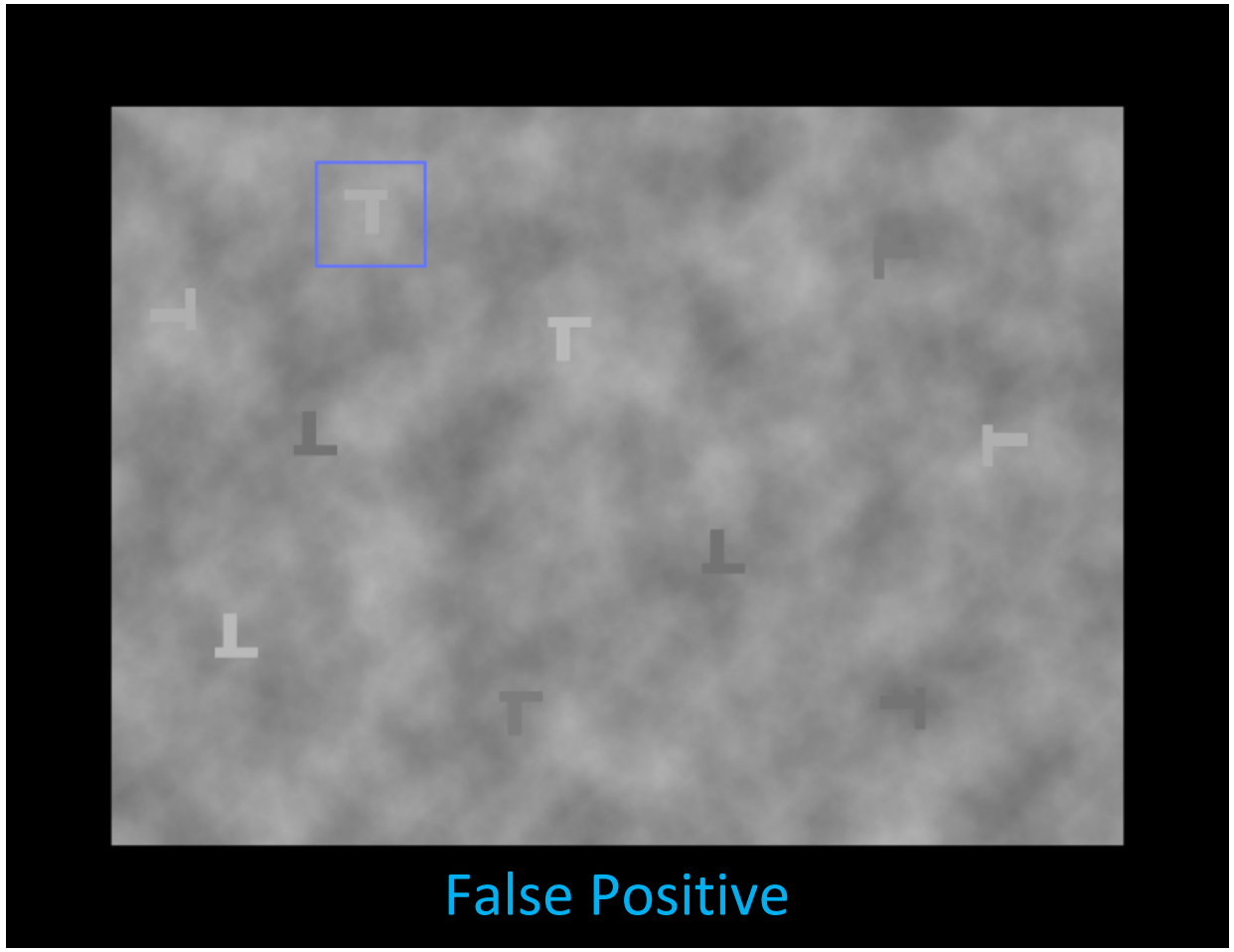
Finding 1: Imperfect models provide decision support

As Model Performance Increases, So Does Overall User Performance

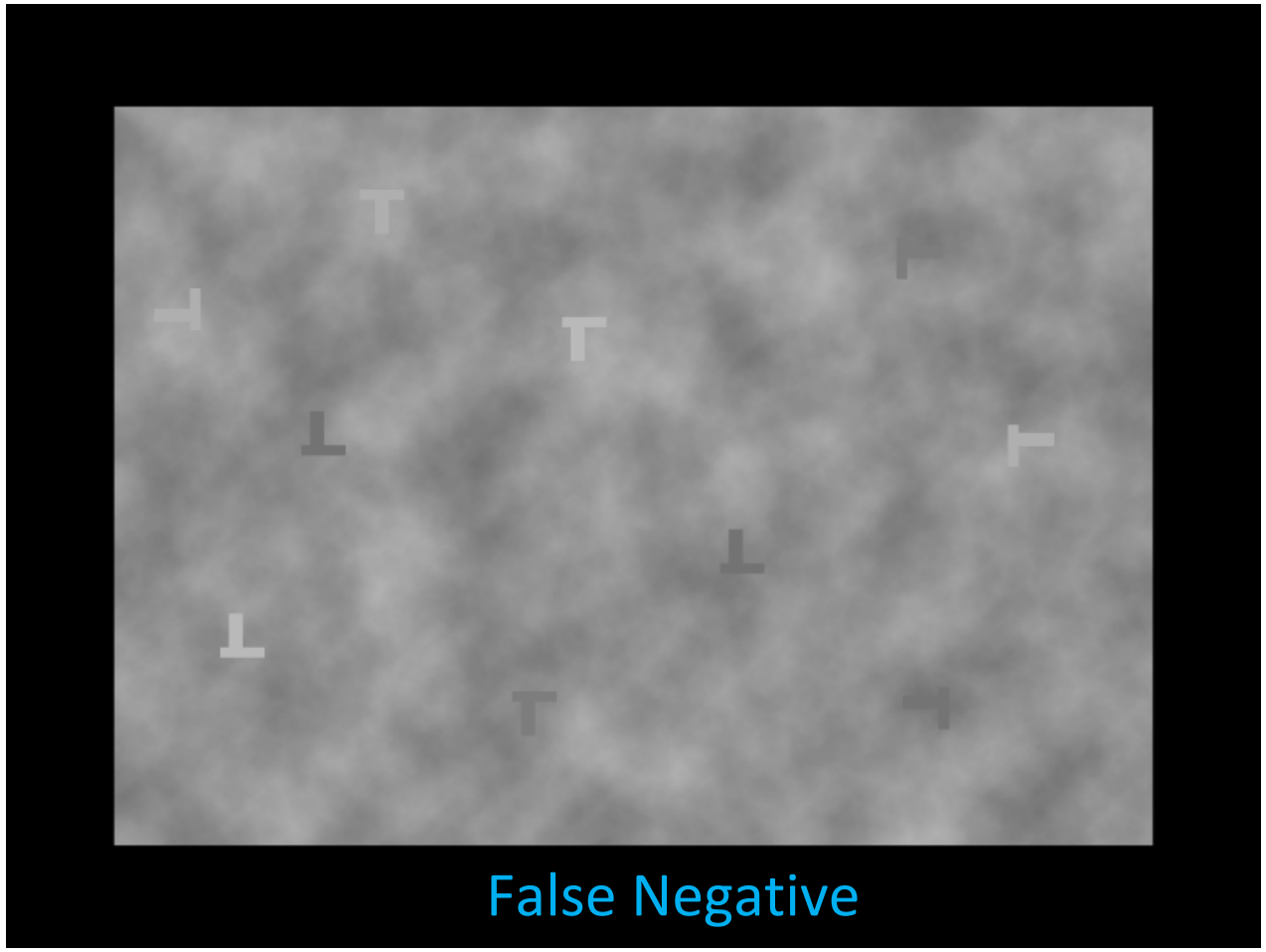


User performance without aid was 79%.

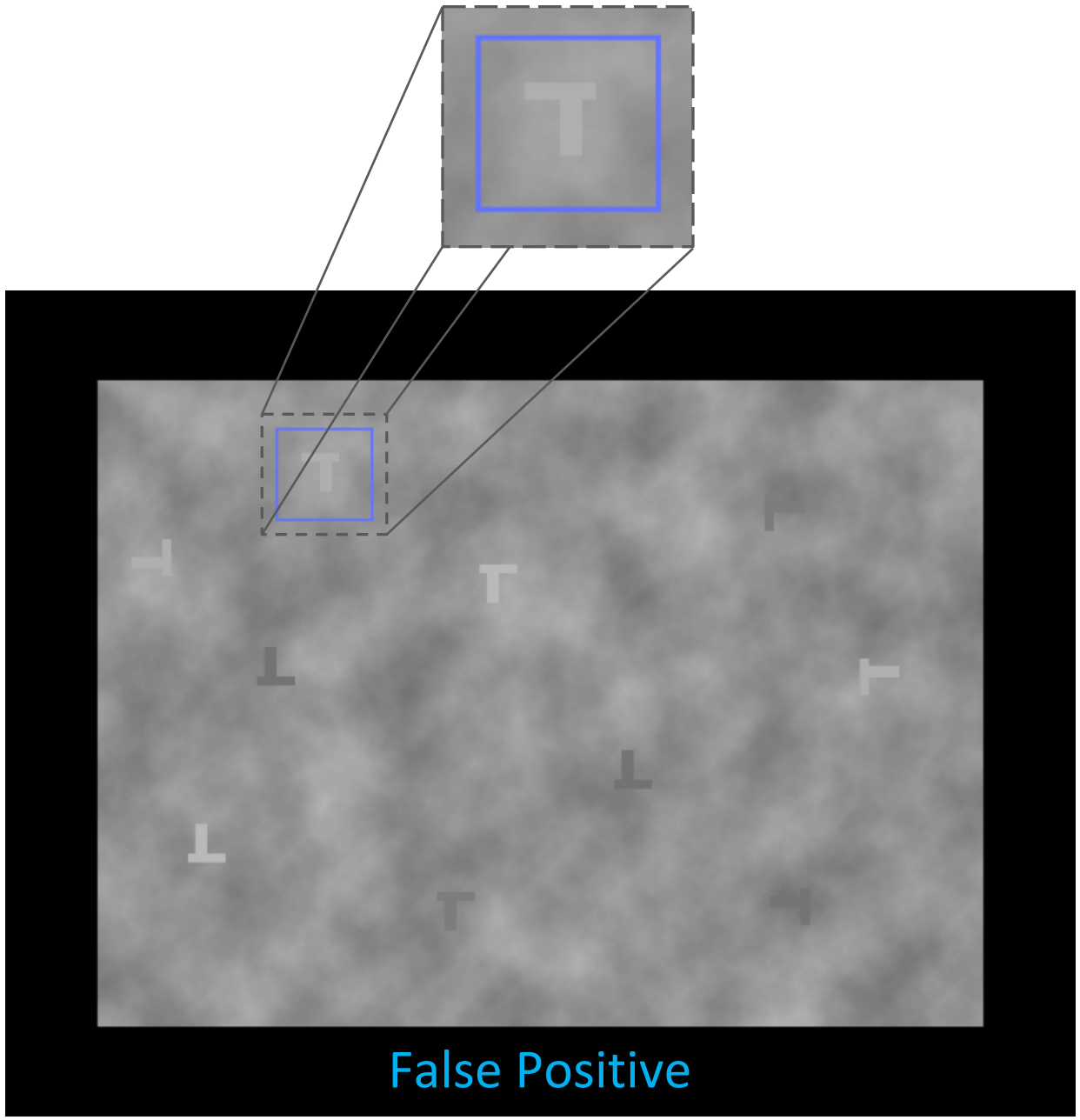
Finding II: Users of high performing models sometimes miss subtle but important errors



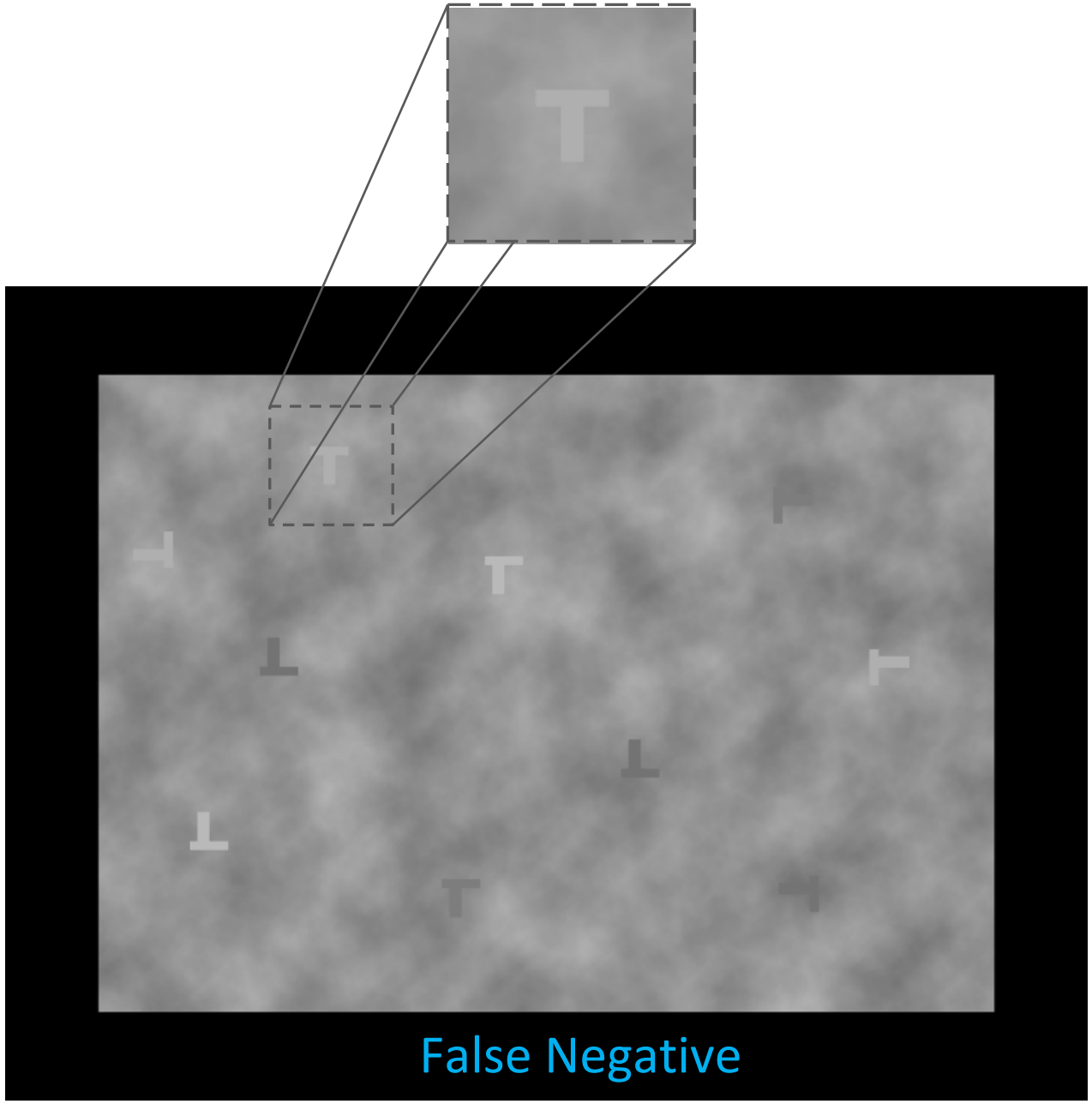
False Positive



False Negative

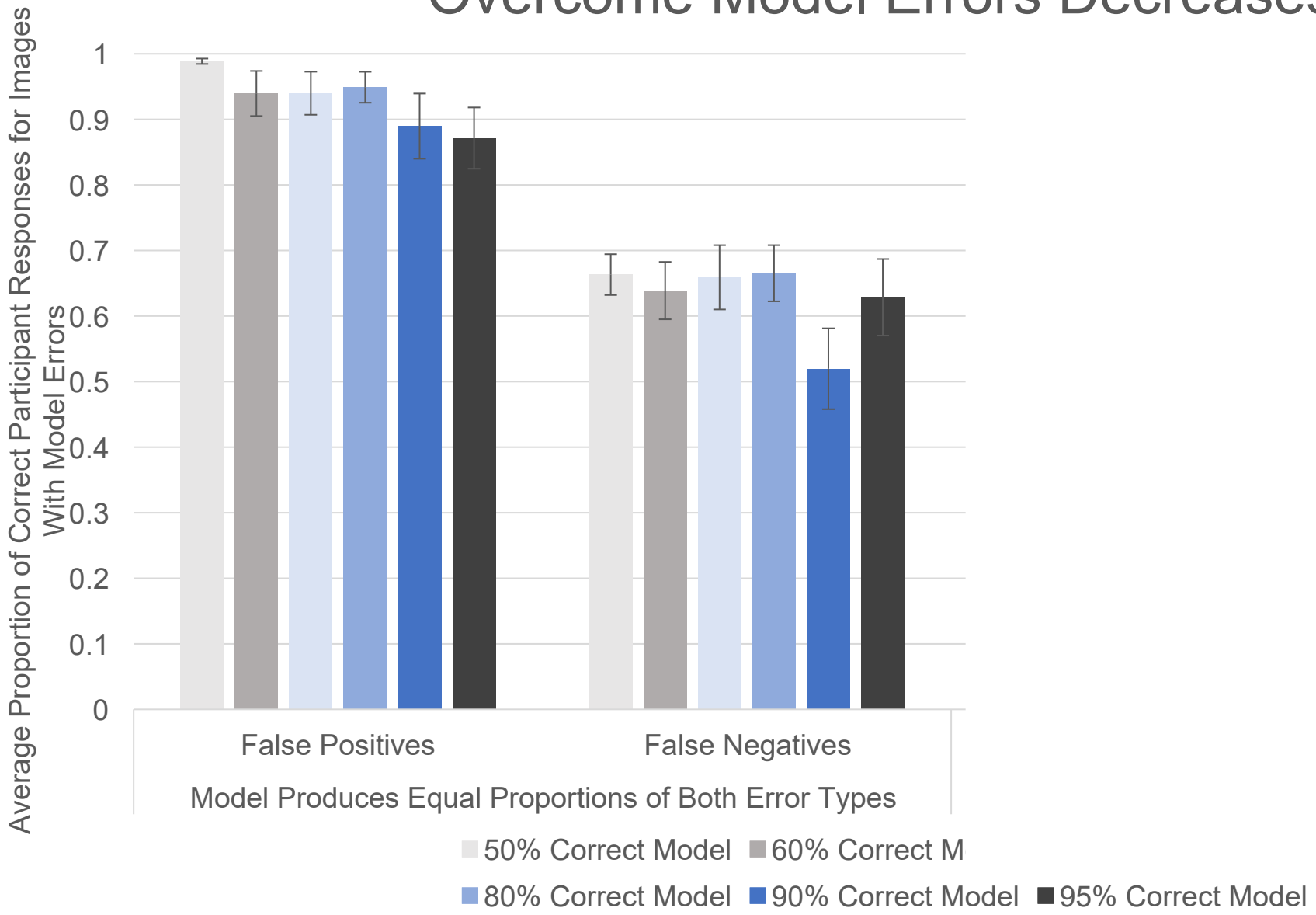


False Positive

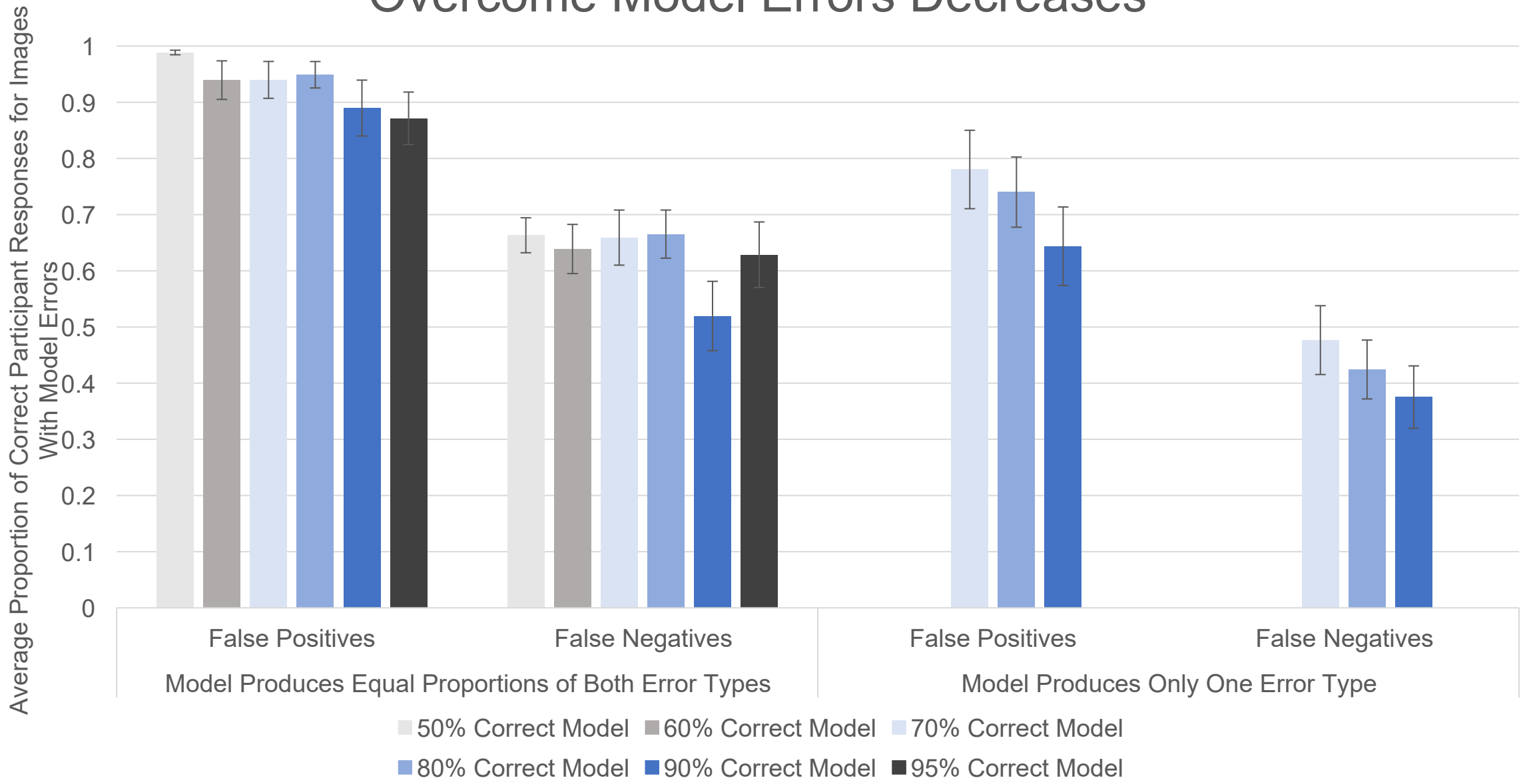


False Negative

As Model Performance Increases, User Ability to Overcome Model Errors Decreases



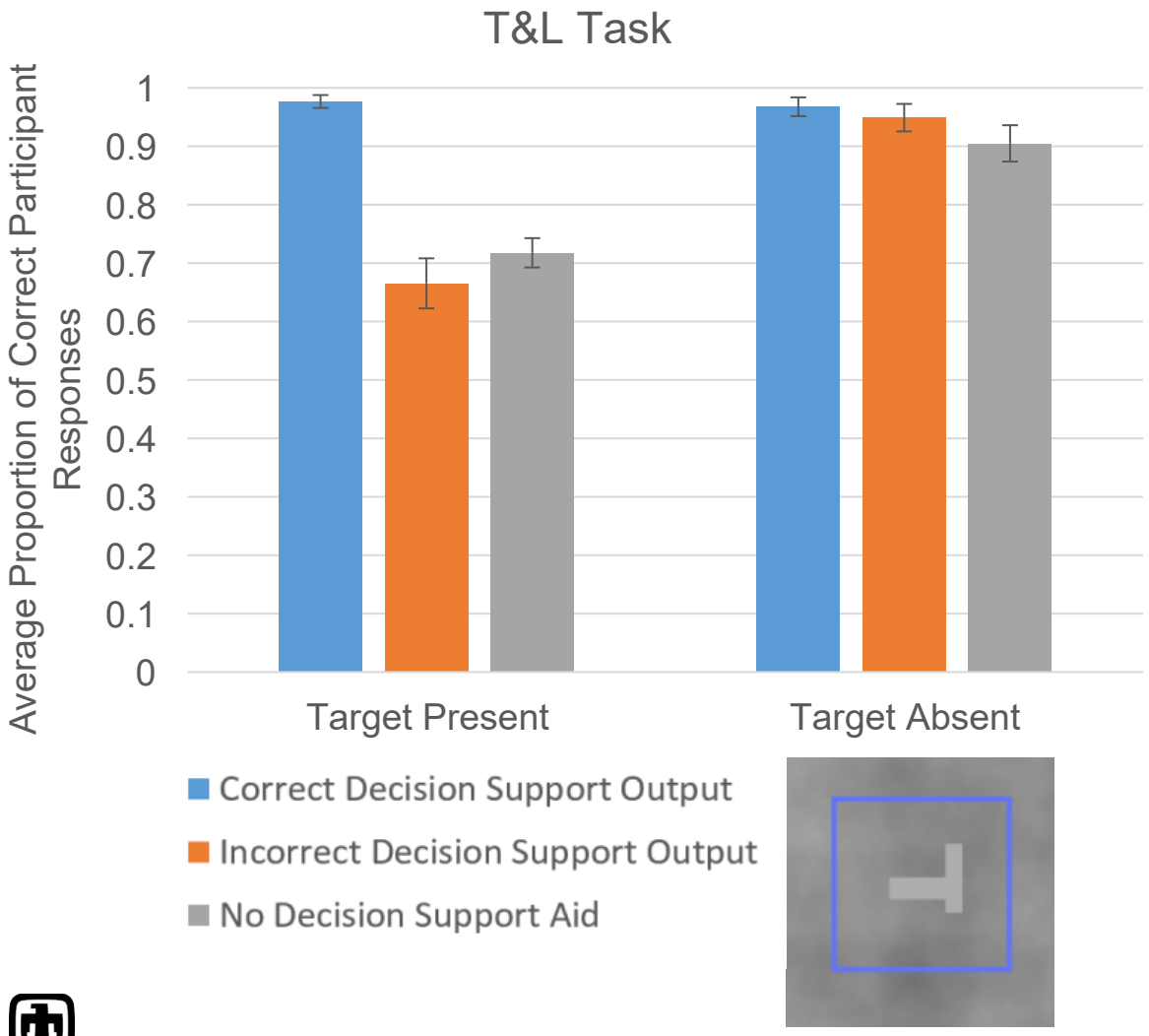
As Model Performance Increases, User Ability to Overcome Model Errors Decreases

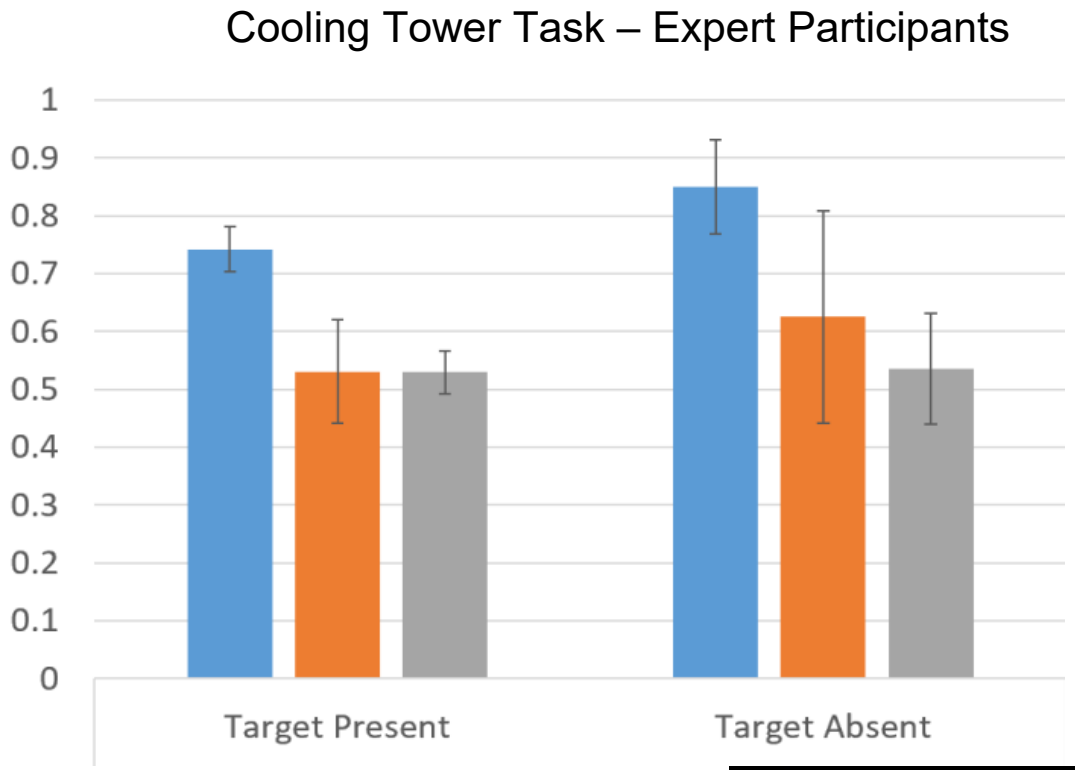


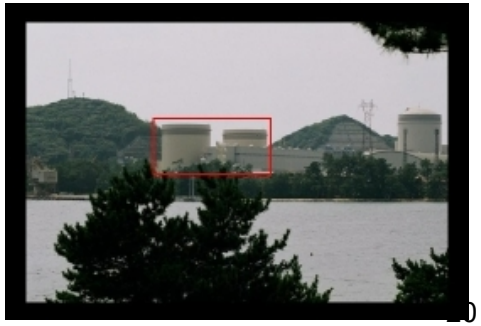
Finding 1I1: User expertise impacts response to model errors

Incorrect model outputs don't hurt performance in domain-general task

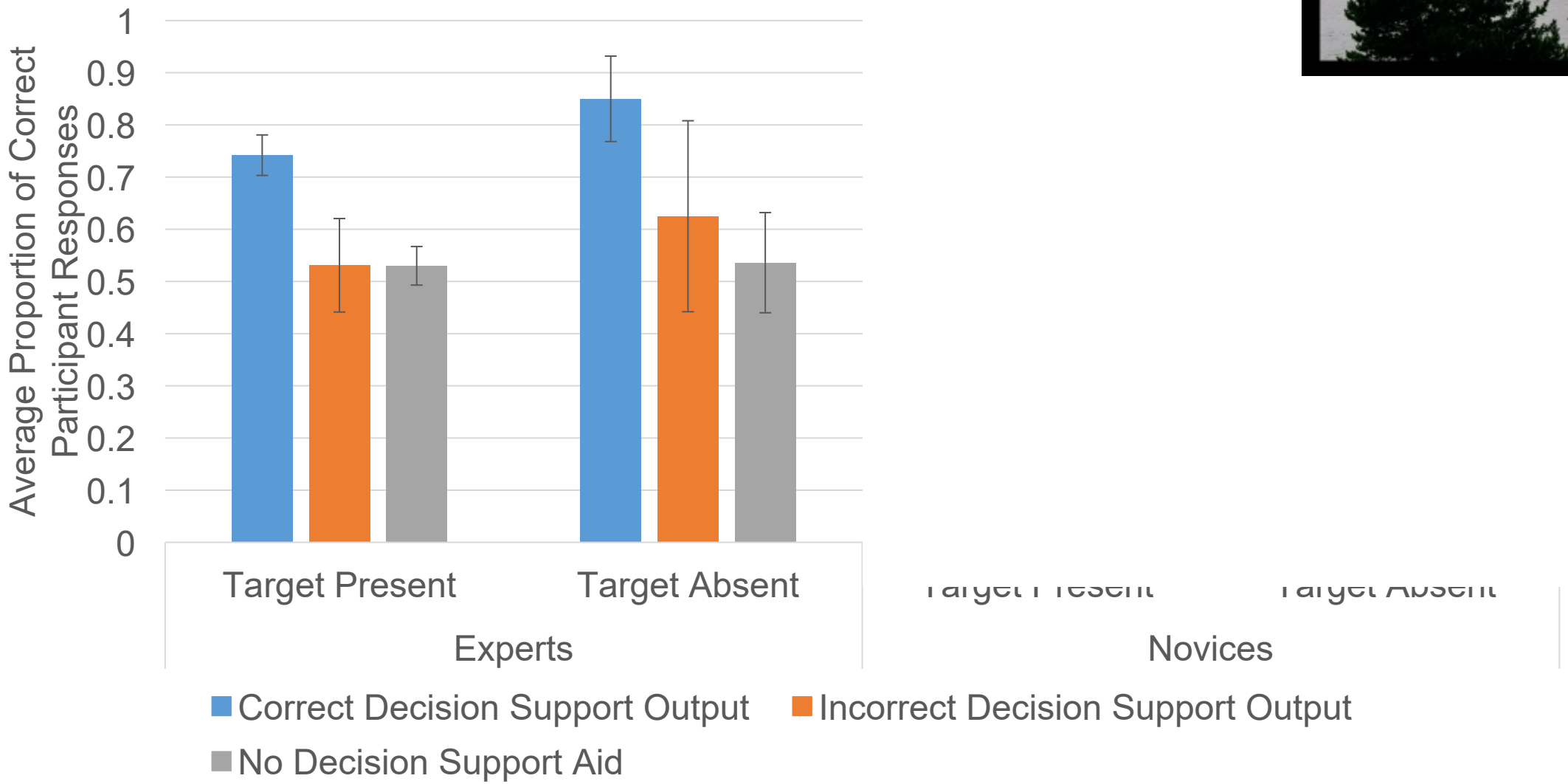
Similar pattern observed for domain experts in domain-specific task



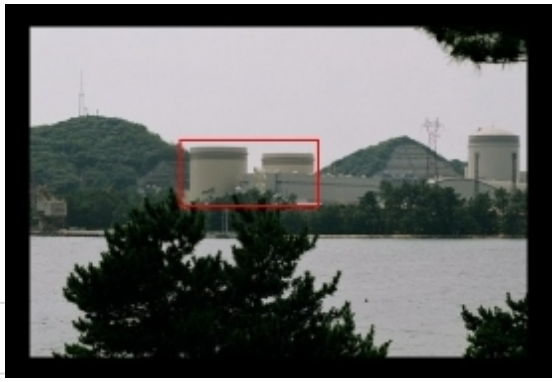


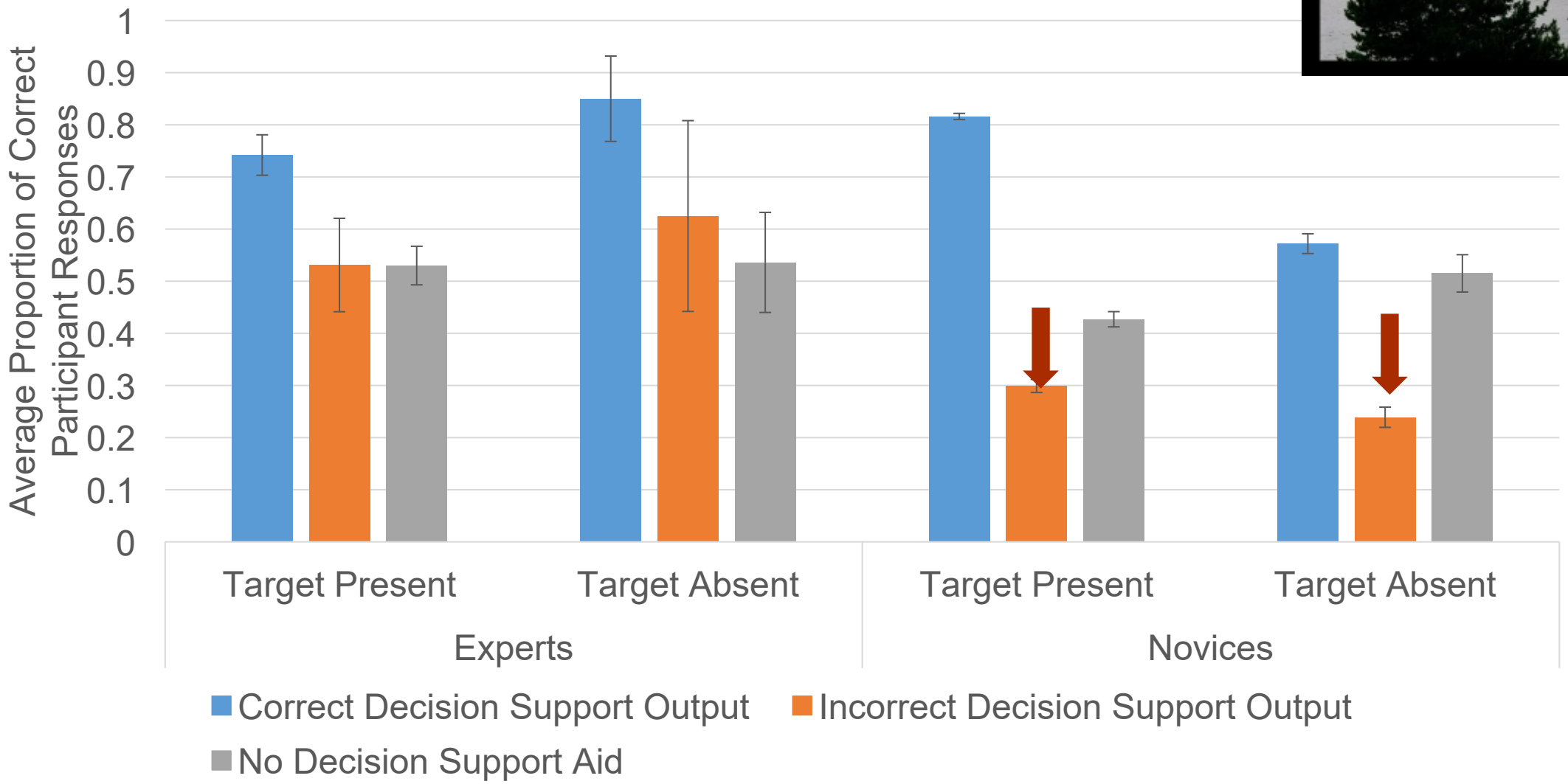


Errors in decision support aids impact novice users more than expert users



Errors in decision support aids impact novice users more than expert users





Discussion

- Computer vision decision support can benefit users
- Users should remain vigilant even with very good models
- Users were less likely to notice false negatives, which is especially significant for safeguards
- Model errors can have different impacts on users with different levels/types of expertise
- We encourage a systems-level approach to optimize human performance while using a decision support tool, rather than just tool performance.