CTSA
Clinical & Translational Science Awards Program

CLIC UN-MEETING

EINSTEIN
Albert Einstein College of Medicine

CLIC
Center for Leading Innovation & Collaboration

Trustworthy A.I.

Parsa Mirhaji MD, PhD

Albert Einstein College of Medicine, Montefiore Health System

The University of Rochester Center for Leading Innovation and Collaboration (CLIC) is the coordinating center for the Clinical and Translational Science Awards (CTSA) Program, funded by the National Center for Advancing Translational Sciences (NCATS) at the National Institutes of Health (NIH), Grant U24TR002260.

CTSA
Clinical & Translational Science Awards Program

CLIC
Center for Leading Innovation & Collaboration

Falsifiability

CTSA
Clinical & Translational Science Awards Program

CLIC
Center for Leading Innovation & Collaboration

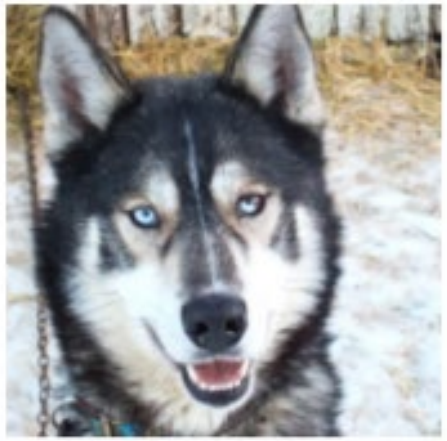
Hallucination

Model relying on non-robust features embedder in the training set

CTSA
Clinical & Translational Science Awards Program

CLIC
Center for Leading Innovation & Collaboration

Bias



(a) Husky classified as wolf

(b) Explanation

Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.

	Before	After
Trusted the bad model	10 out of 27	3 out of 27
Snow as a potential feature	12 out of 27	25 out of 27

Table 2: "Husky vs Wolf" experiment results.

CTSA
Clinical & Translational
Science Awards Program

CLIC
Center for Leading
Innovation & Collaboration

Overfit