Why AI Fails: Parallax

Lanxi Xiao
Academy of Arts & Design,
Tsinghua University
Beijing, China
tarolancy@gmail.com

Weikai Yang
School of Software,
Tsinghua University
Beijing, China
vicayang496@gmail.com

Haoze Wang
School of Software,
Tsinghua University
Beijing, China
wanghz_thu18@163.com

Shixia Liu
School of Software,
Tsinghua University
Beijing, China
shixia@tsinghua.edu.cn

Qiong Wu
Academy of Arts & Design,
Tsinghua University
Beijing, China
qiong-wu@mail.tsinghua.edu.cn

ABSTRACT

"Why AI Fails: Parallax" is an interactive visual art installation in the "Why AI Fails" series. This artwork aims to illustrate the transformation of AI from an unexplainable "black box" to an explainable "white box" through a sliding screen. Its purpose is to allow people, regardless of their level of AI knowledge, to comprehend the reasons behind AI misclassifications intuitively. By interacting with the sliding screen, users can click on misclassified images of their interest and explore the primary factors that influenced the classification. They can also compare the differences in data and models between biased and normal AI instances. This installation serves as a bridge to overcome the technological gap. Integrating with various AI models, it assists artists and designers in gaining a deeper understanding of how AI makes decisions related to artistic design styles, features, imagery, materials, music rhythm, melody, and chords.

The 1st International Workshop on Explainable AI for the Arts
Creativity & Cognition 2023: June 19-21, Online

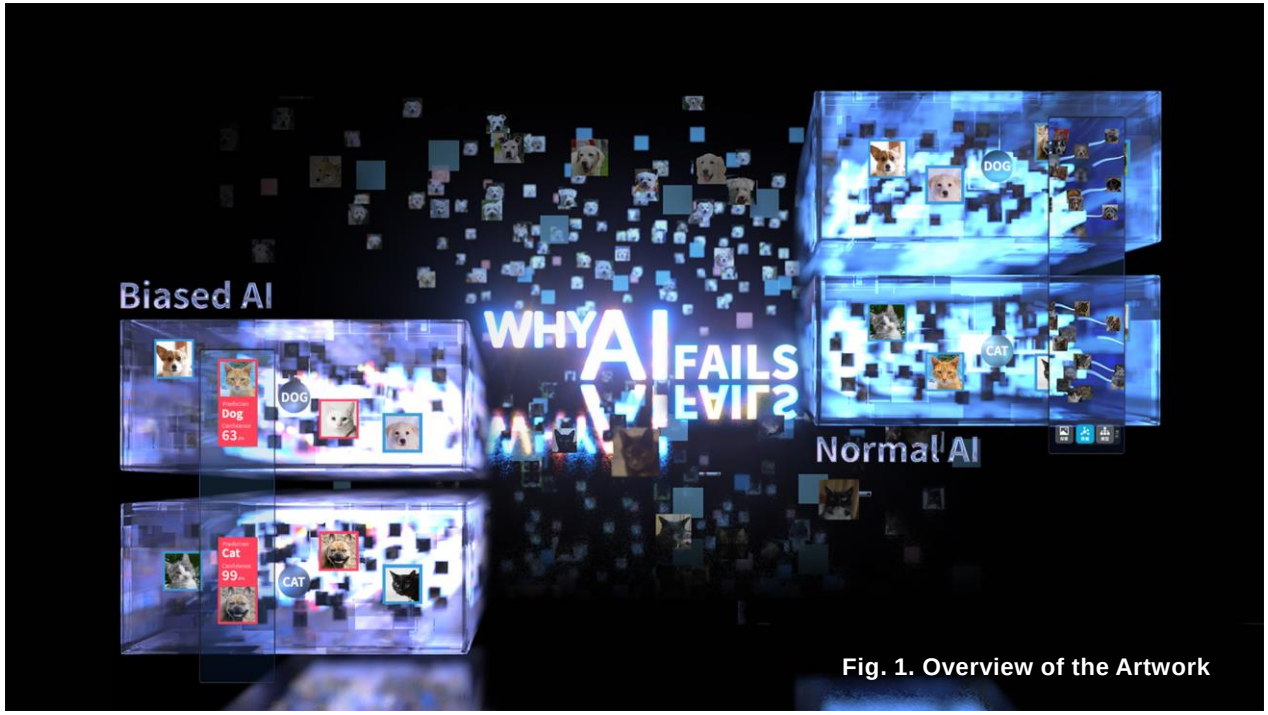

Fig. 1. Overview of the Artwork

Author Keywords

Artificial intelligence; misclassification; visualization design; interaction design; art installation.

CSS Concepts

• Human-centered computing~Visualization

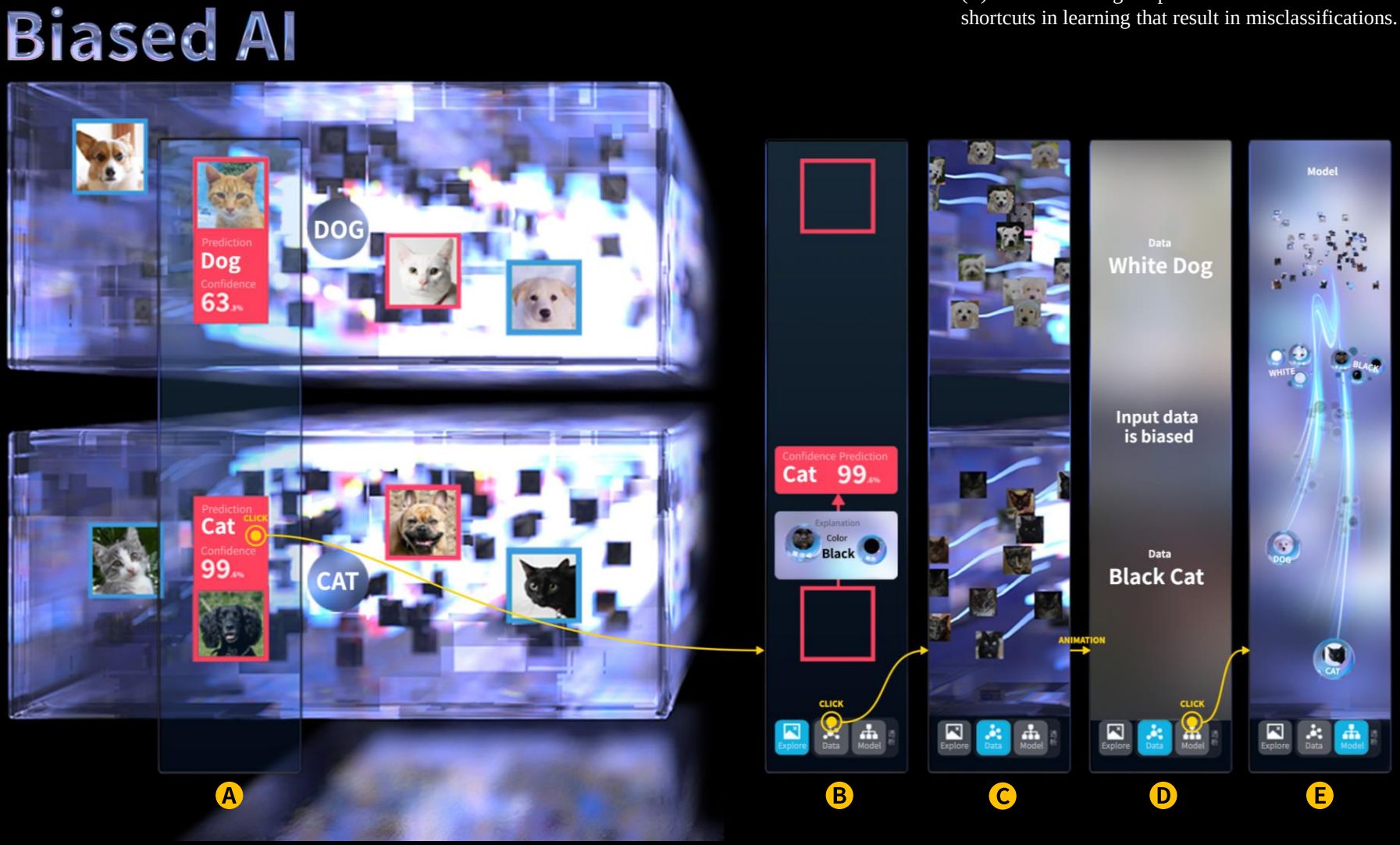

Fig. 2. Interactive Process of Sliding Screen

(A) and (B) show the Exploration Page, where users can interactively investigate why AI sometimes makes mistakes in categorizing images.

(C) and (D) display the Data Pages, which provide users with more information about the anomalies in AI data that can lead to misclassifications.

(E) The Model Page explains how AI can sometimes take shortcuts in learning that result in misclassifications.

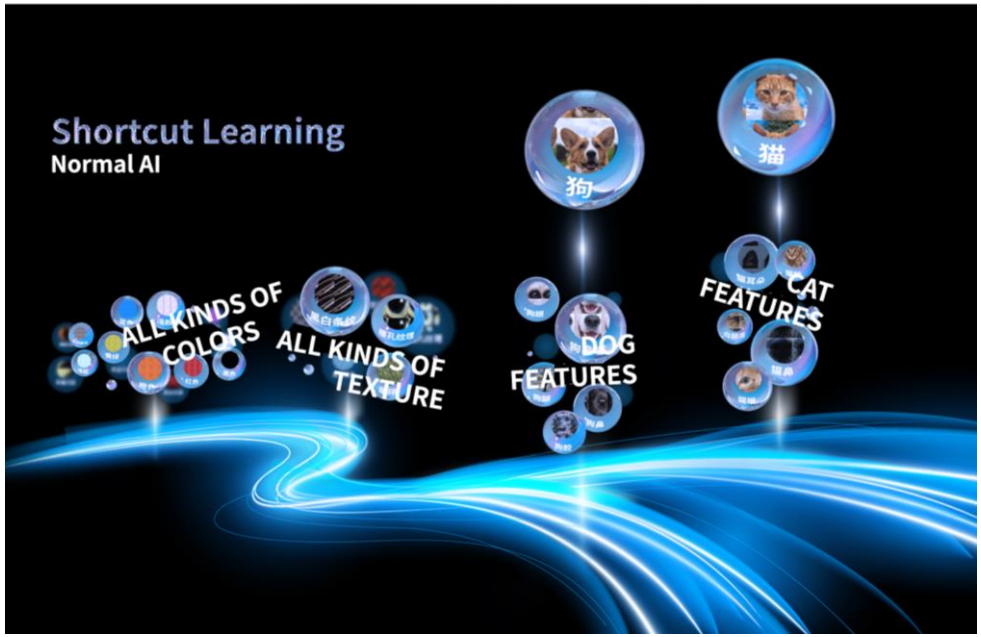

Fig. 3&4 Left. Comparison of Biased AI and Normal AI at Model Level

In these two visualizations, we combed the complex network model of artificial intelligence and clustered the neurons to obtain the most intuitive difference between biased and normal artificial intelligence.

We represented each class of features with a bubble and mapped the bubble size to indicate how vital the features are to AI. Biased and shortcut-learned artificial intelligence uses black and white colors as the primary basis for classifying cats and dogs, establishing a spurious relationship between input data and output results. In contrast, normal artificial intelligence ultimately learns all features.

Fig. 5 Bottom Right. An Application Scenario in Science Museums

Science museums and educational exhibition venues are prime locations for the compelling application of this installation.

At Tsinghua University Art Museum, we have carefully designed a structure combining a large 60-inch interactive display with an accompanying 28-inch interactive sliding screen. The 60-inch screen presents the mysterious AI black boxes, in front of which the 28-inch sliding screen provides an immersive experience showcasing the remarkable visual journey of artificial intelligence, capturing its transformation from black boxes to elucidating white boxes.

CONCLUSION

This pictorial illustrates how the "Why AI Fails: Parallax" installation enhances the public's understanding of AI model operations. It accomplishes this by highlighting two key aspects. Firstly, the installation provides a glimpse into the internal workings of AI models and clarifies the specific features utilized for classification, utilizing a cat-dog misclassification case [1] as a prominent example. Visualizations like exploded diagrams and bubble charts effectively represent the complex, high-dimensional data [2] that underlies AI models, revealing the mechanisms of AI's shortcut learning [3]. Secondly, the installation offers an immersive and captivating interactive experience through its sliding screen, engaging viewers and enabling them to explore AI's mechanisms firsthand.

The installation can further integrate with different AI models, fostering exploration and innovation in artistic practice and opening up new possibilities for artists and creators.

In art and design, AI systems face misclassification issues in various artistic forms, including misjudgment of style, features, imagery, and materials. These problems may stem from insufficient training data, algorithmic limitations, and shortcut learning. It can help artists and designers delve deeper into studying and provide insights and recommendations for improving AI art recognition.

This installation can also be used in music composition to explore misclassification issues of AI systems with audio data. For instance, musicians can input their musical compositions into the AI system and analyze the AI system's decision-making based on rhythm, melody, chords, and other elements. It aids in understanding the AI system's perception and expression of musical elements, providing new inspiration and possibilities for music composition.

ACKNOWLEDGMENTS

This work is supported by the Art Project of The National Social Science Fund of China (No. 19BG127).

REFERENCES

[1] Changjian Chen, Jun Yuan, Yafeng Lu, Yang Liu, Hang Su, Songtao Yuan and Shixia Liu. OoDAnalyzer: Interactive Analysis of Out-of-Distribution Samples. 2020. IEEE transactions on visualization and computer graphics, 27(7): 3335-3349.

[2] Mengchen Liu, Jiaxin Shi, Zhen Li, Chongxuan Li, Jun Zhu and Shixia Liu. Towards better analysis of deep convolutional neural networks. 2016. IEEE transactions on visualization and computer graphics. 23(1): 91-100.

[3] Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis Richard Zemel, Wieland Brendel, Matthias Bethge and Felix A. Wichmann. 2020. Shortcut learning in deep neural networks. Nature Machine Intelligence. 2(11): 665-673.