



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Explainability in Deep Learning Through Communication

Prof. Dr. Zeynep Akata

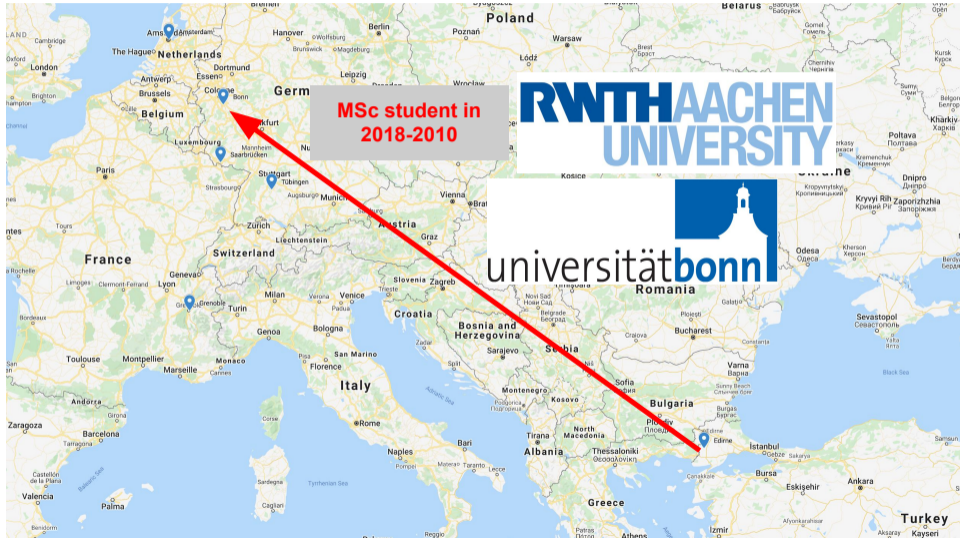
Alfried Krupp Award Lecture

19 October 2023

My Journey: from Edirne to Tübingen



My Journey: from Edirne to Tübingen

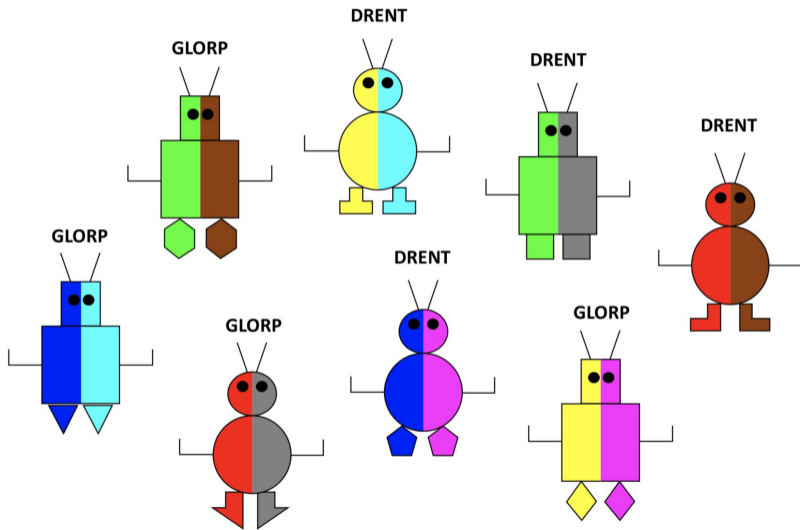


My Journey: from Edirne to Tübingen



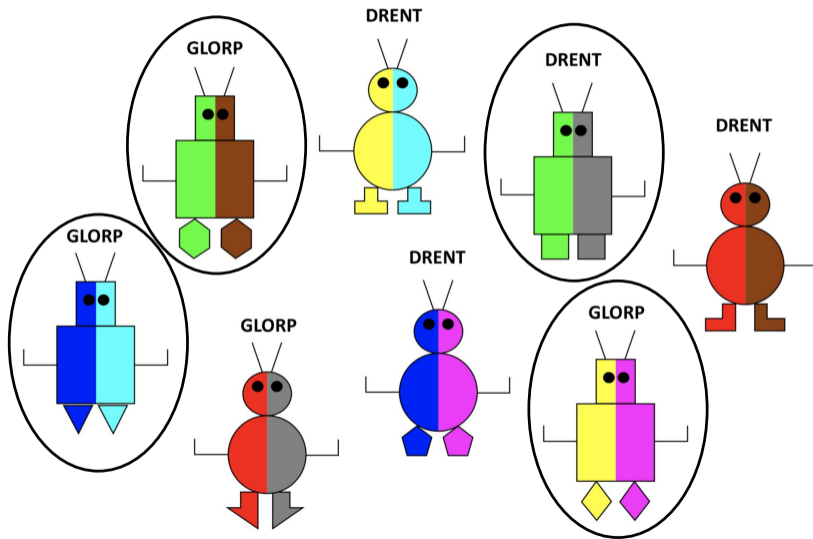
Learning and Explaining are Related

Lombrozo TICS'16



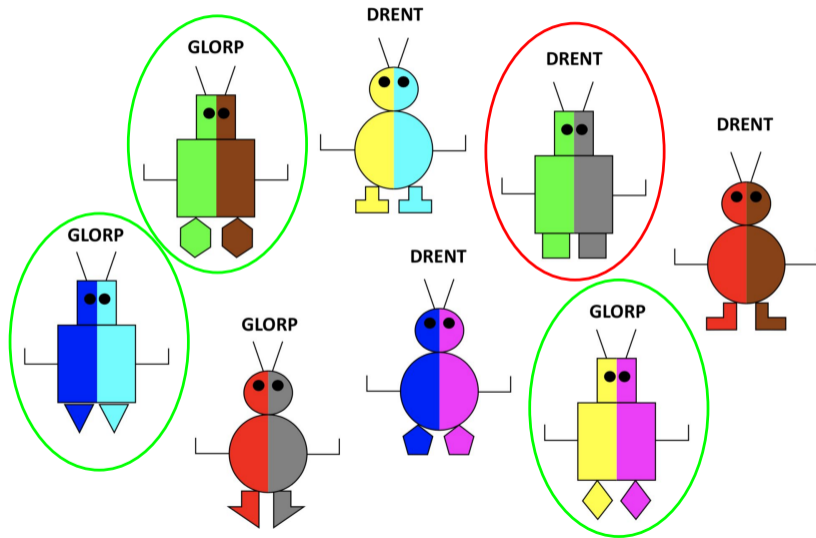
Learning and Explaining are Related

Lombrozo TICS'16



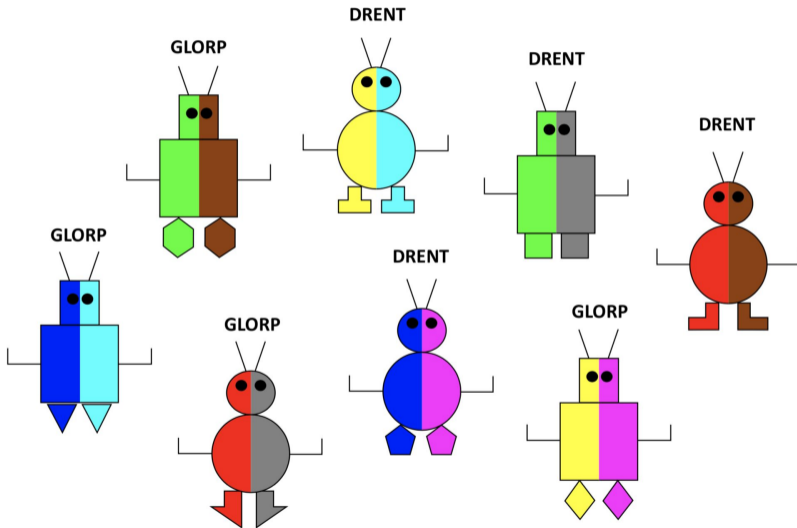
Learning and Explaining are Related

Lombrozo TICS'16



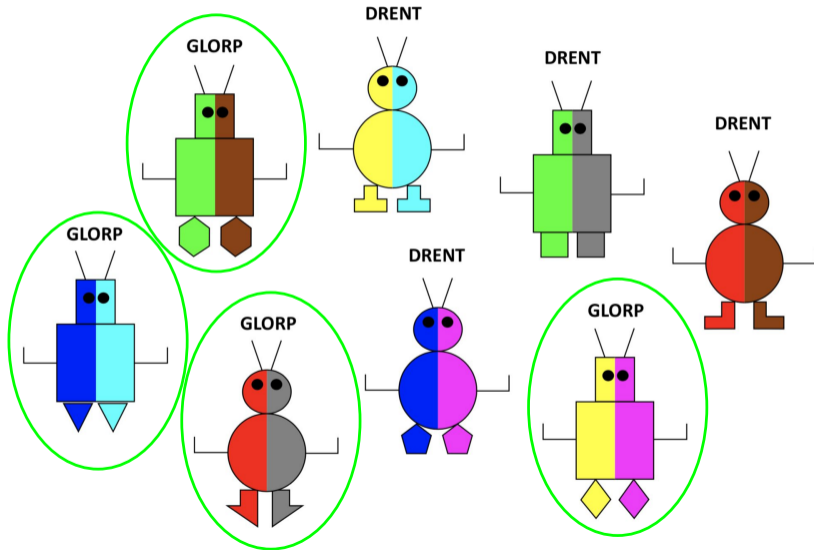
Learning and Explaining are Related

Lombrozo TICS'16



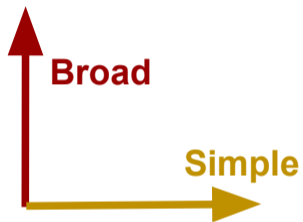
Learning and Explaining are Related

Lombrozo TICS'16

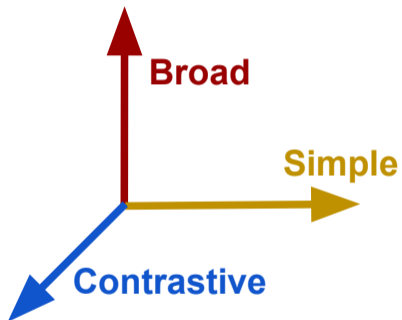




- Justify a range of observations



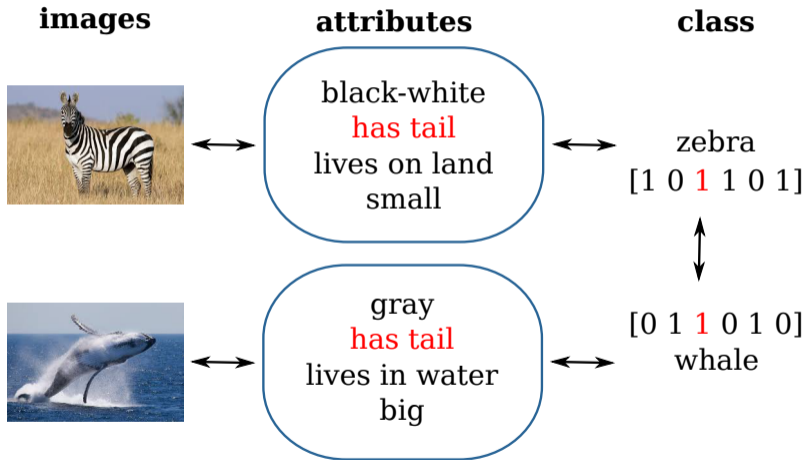
- Justify a range of observations
- Concise for the communication partner



- Justify a range of observations
- Concise for the communication partner
- Differentiate two alternative decisions

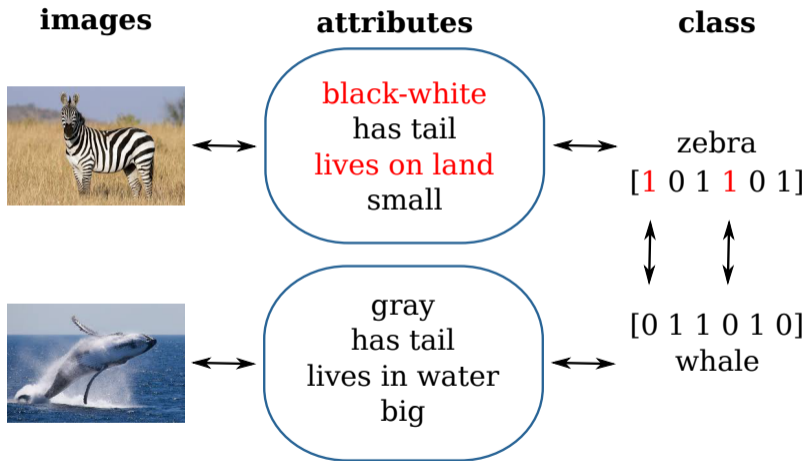
Attributes as Explanations

Lampert et al. CVPR'09



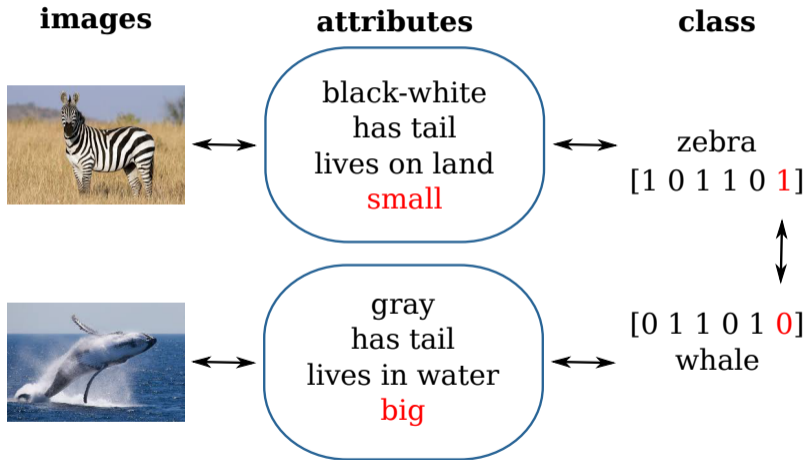
Attributes as Explanations

Lampert et al. CVPR'09



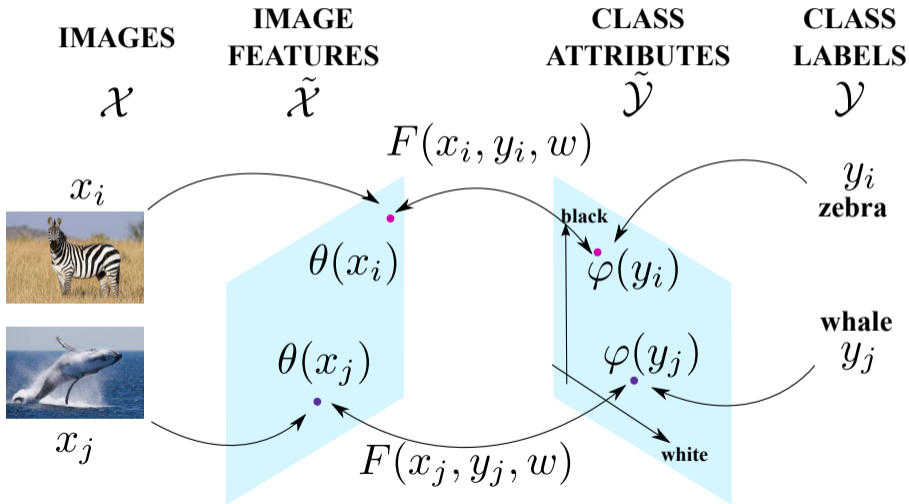
Attributes as Explanations

Lampert et al. CVPR'09



Multimodal Embeddings

Akata et al. CVPR'13 & TPAMI'16



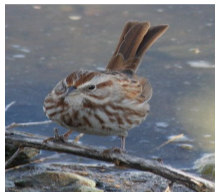
My Journey: from Edirne to Tübingen



Natural Language for Fine-Grained Classification



The bird has a white underbelly, black feathers in the wings, a large wingspan, and a white beak.



This bird has distinctive-looking brown and white stripes all over its body, and its brown tail sticks up.



This flower has a central white blossom surrounded by large pointed red petals which are veined and leaflike.

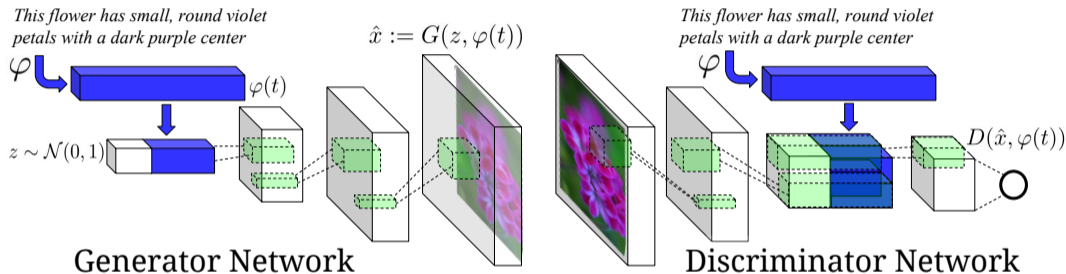


Light purple petals with orange and black middle green leaves

Learning Deep Representations of Fine-Grained Visual Descriptions;

Reed, Akata, Schiele, Lee at IEEE CVPR 2016

Generating Image Features with Natural Language

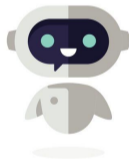


Generative Adversarial Text to Image Synthesis;
Reed, Akata, Yan, Logeswaran, Lee, Schiele at ICML 2016

My Journey: from Edirne to Tübingen



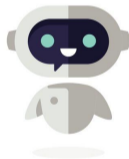
Generating Natural Language Explanations



Generating Natural Language Explanations



What type of bird is this?



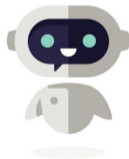
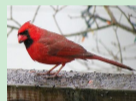
Generating Natural Language Explanations



What type of bird is this?



It is a **Cardinal**



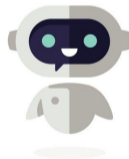
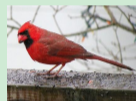
Generating Natural Language Explanations



What type of bird is this?



It is a **Cardinal** because it is a red bird with a red beak and a black face



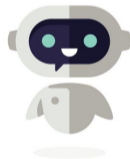
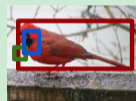
Generating Natural Language Explanations



What type of bird is this?



It is a **Cardinal** because it is a **red bird** with a **red beak** and a **black face**



Generating Natural Language Explanations

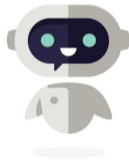


What type of bird is this?



Why not a **Vermilion Flycatcher?**

It is a **Cardinal** because it is a **red bird** with a **red beak** and a **black face**



Generating Natural Language Explanations



What type of bird is this?

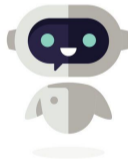


Why not a **Vermilion Flycatcher**?

It is a **Cardinal** because it is a **red bird** with a **red beak** and a **black face**



It is not a **Vermilion Flycatcher** because it does not have black wings.



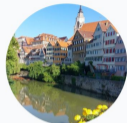
My Journey: from Edirne to Tübingen



Our Explainable Machine Learning Group



Zeynep Akata
Professor



Almut Sophia Koepke
Research Fellow



Stephan Alaniz
Postdoctoral Researcher



Maja Pantić
PhD Researcher



Otniel-Bogdan Mercea
PhD Researcher



Shyamgopal Karthik
PhD Researcher



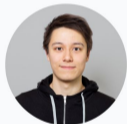
Thomas Hummel
PhD Researcher



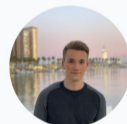
Anders Christensen
PhD Researcher



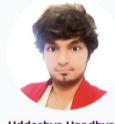
Jae Myung Kim
PhD Researcher



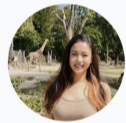
Karsten Roth
PhD Researcher



Lukas Thede
PhD Researcher



Uddeshya Upadhyay
PhD Researcher



Yiran Huang
PhD Researcher



Jessica Bader
HIWI MSc student



Leander Girrbach
PhD researcher



Leonard Salewski
PhD Researcher



Luca Eyring
PhD Researcher



Meghal Dani
PhD Researcher



Joschka Strüber
HIWI MSc student

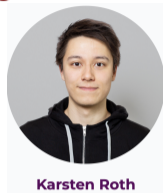
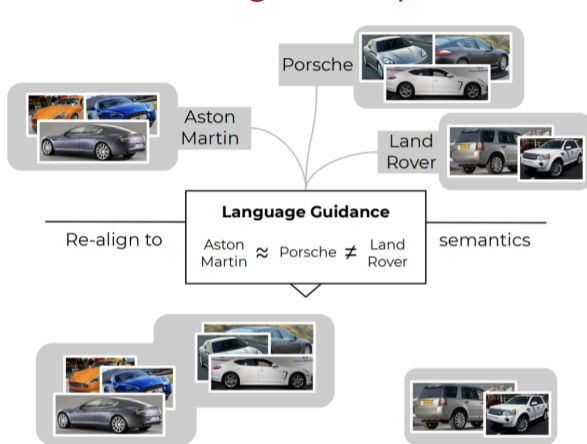


Yavuz Durmazkeser
HIWI MSc student



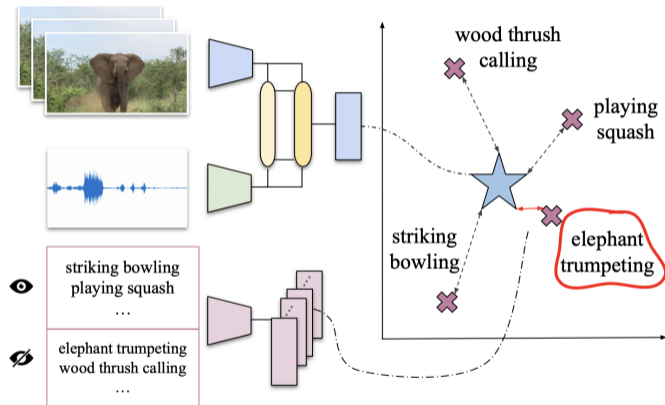
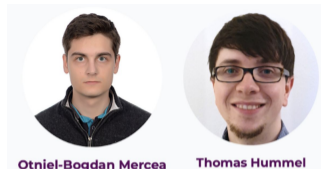
Michael Mergner
Foreign Language Secretary

Language as External Knowledge in Deep Metric Learning



Integrating Language Guidance into Vision-based Deep Metric Learning;
Roth, Vinyals, Akata at CVPR 2022

Language as External Knowledge in Audiovisual ZSL



Temporal and Cross-Modal Attention for Audio-visual Zero-shot Learning;
Mercea, Hummel, Koepke, Akata at ECCV 2022

Calibrating Uncertainty Values in Attribution Maps is Important



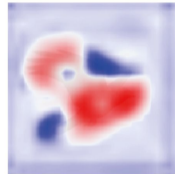
GT class: A



CAM for class A



CAM for class B



Why class A not B?



How is A \neq B ?



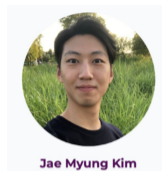
CALM for class A



CALM for class B



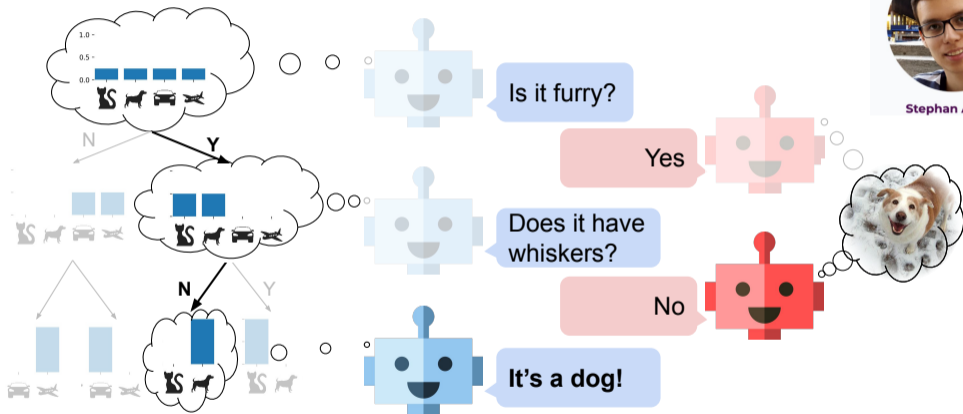
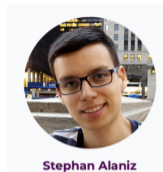
Why class A not B?



Jae Myung Kim

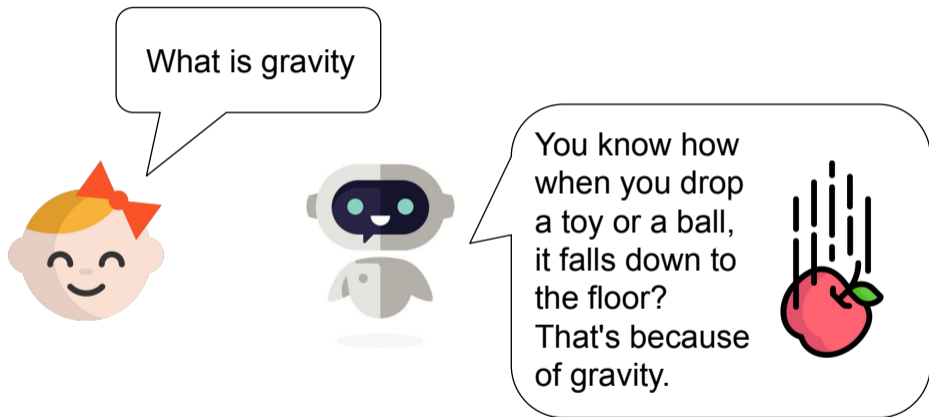
Keep CALM and Improve Visual Feature Attribution;
Kim, Choe, Akata, Oh at ICCV 2021

Transparency of Deep Learning Model via Disentanglement

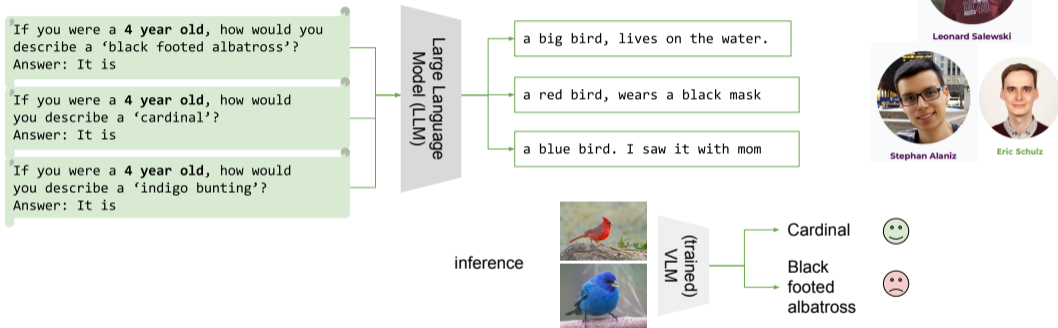


Learning Decision Trees Recurrently Through Communication;
Alaniz, Marcos, Schiele, Akata at CVPR 2021

Tailored Explanations for the Communication Partner

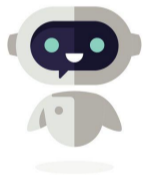
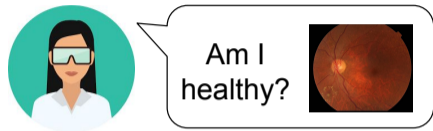


In-Context Impersonation of Large Language Models



In-Context Impersonation Reveals Large Language Models' Strengths and Biases;
Salewski, Alaniz, Rio-Torto, Schulz, Akata; NeurIPS 2023

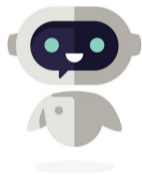
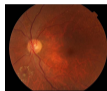
The Important Role of Explainability for Science



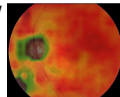
The Important Role of Explainability for Science



Am I healthy?



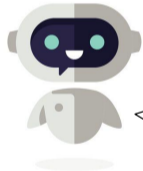
Green regions show enlarged veins. So there is a possibility that this indicates diabetes.



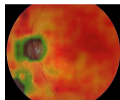
The Important Role of Explainability for Science



What is the diagnosis?



Green regions show **enlarged veins**. So there is a **possibility** that this indicates diabetes.

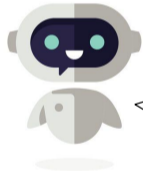


The patient has a superficial, **traumatic corneal abrasion** based on clinical signs and symptoms

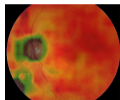
The Important Role of Explainability for Science



What causes it?



Green regions show **enlarged veins**. So there is a **possibility** that this indicates diabetes.



The patient has a superficial, **traumatic corneal abrasion** based on clinical signs and symptoms

Traumatic corneal abrasion due to **increased tearing** and **lack of epithelium** in the cornea

Thank you!

