



Automatic image recognition in upload filters - computing of transparent decisions (XAI), with the help of Deep Learning methods

Timo Bohnstedt

Assessor : Prof. Dr. Alfred Holl

Supervisor: Prof. Dr. Gallwitz

Faculty of Computer Science

Technische Hochschule Nürnberg Georg Simon OHM

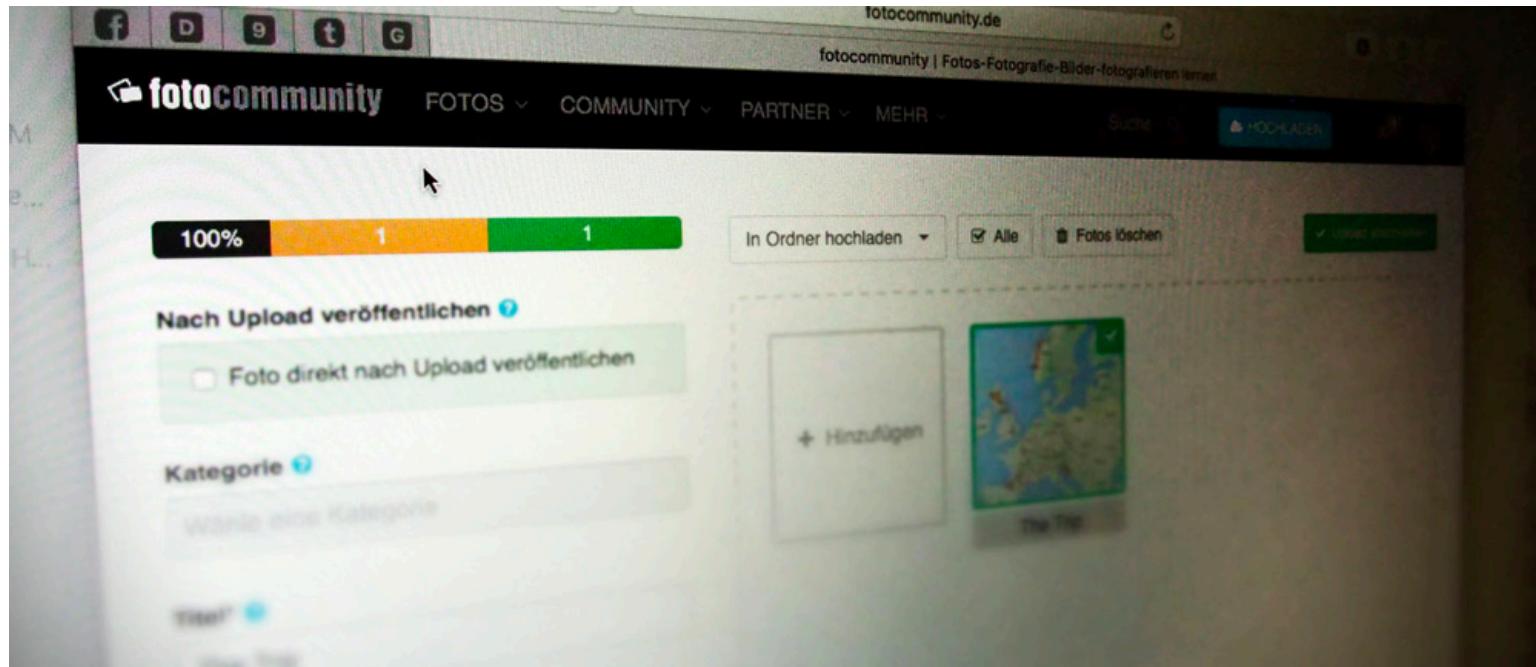
July 19th, 2020

Introduction



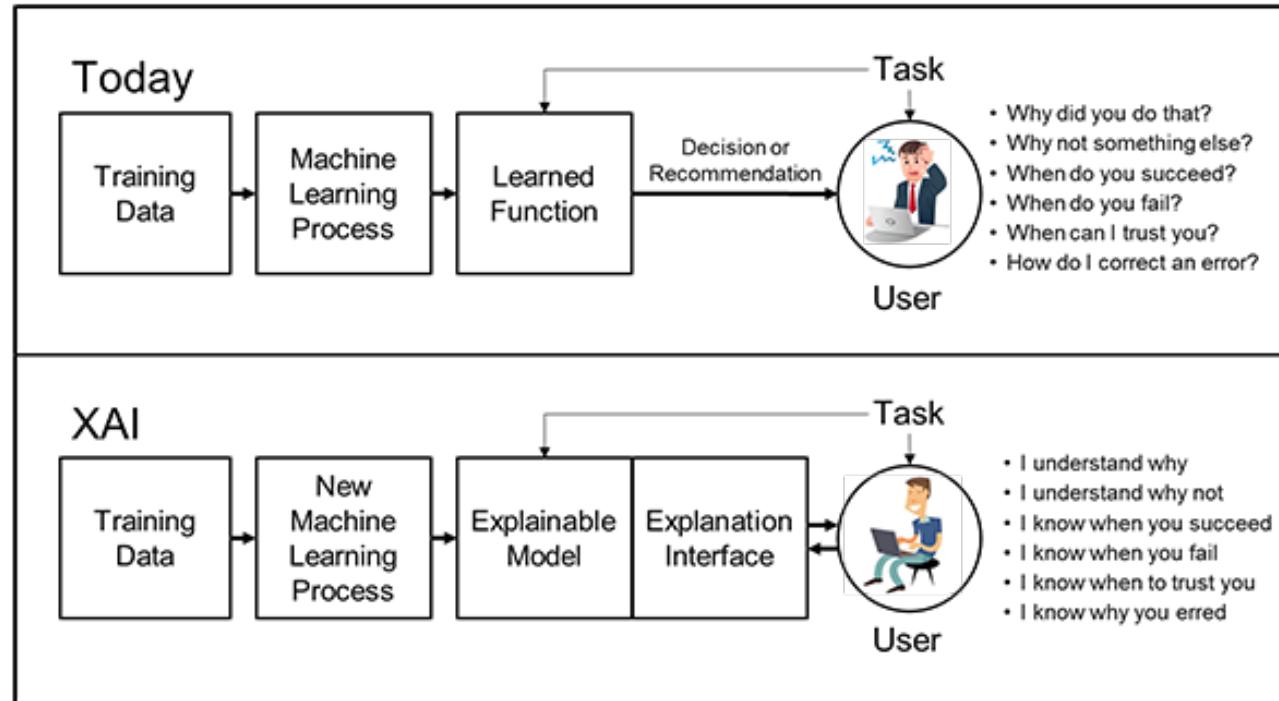
Copyright law reform: Germany considers upload filters

Introduction



Sharehosters would need to start monitoring all user behaviour consequences which could harm the business of small content creators

Introduction



Explainable artificial intelligence could help to make upload filter technology like deep learning more transparent

Related Work

Applying Explainable Artificial Intelligence for Deep Learning Networks to Decode Facial Expressions of Pain and Emotions

Author:
Katharina Blandina WEITZ
Matrikelnr: 1706341

Assessor:
Prof. Dr. Ute SCHMID
Supervisor:
Teena HASSAN

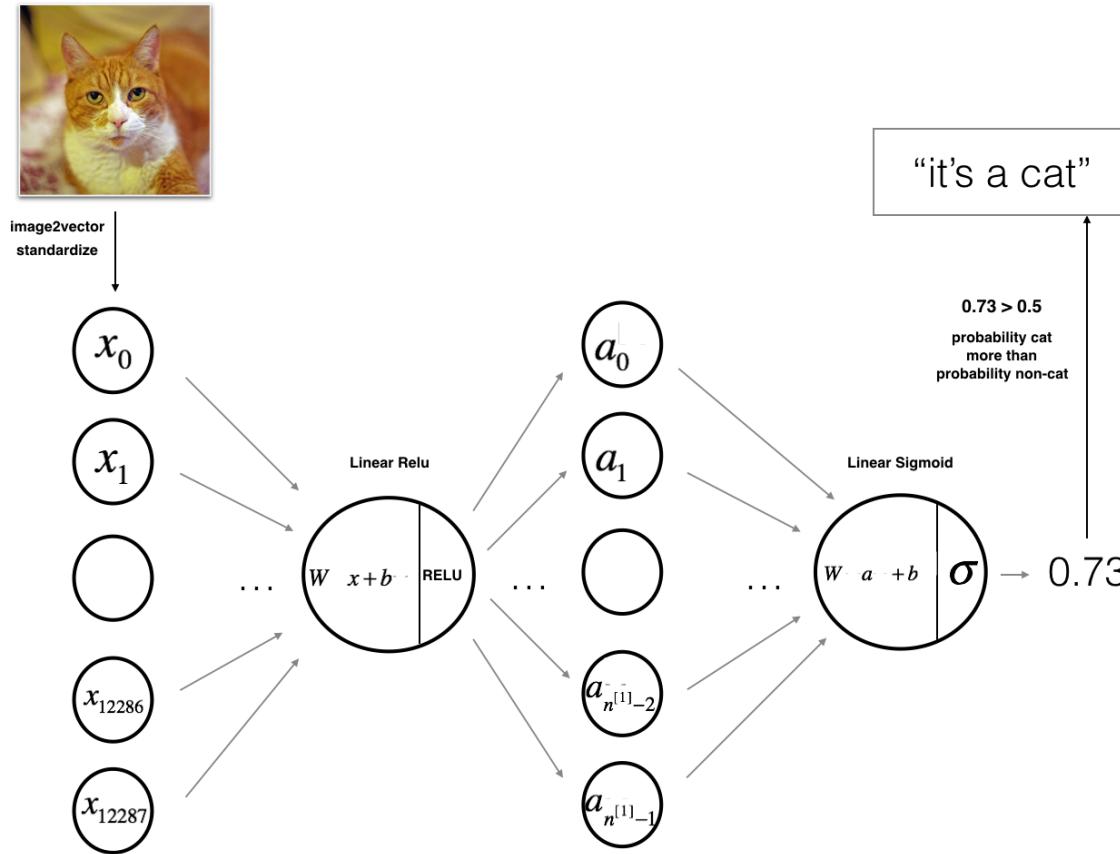
A Unified Approach to Interpreting Model Predictions

Scott M. Lundberg
Paul G. Allen School of Computer Science
University of Washington
Seattle, WA 98105
slund1@cs.washington.edu

Su-In Lee
Paul G. Allen School of Computer Science
Department of Genome Sciences
University of Washington
Seattle, WA 98105
suinlee@cs.washington.edu

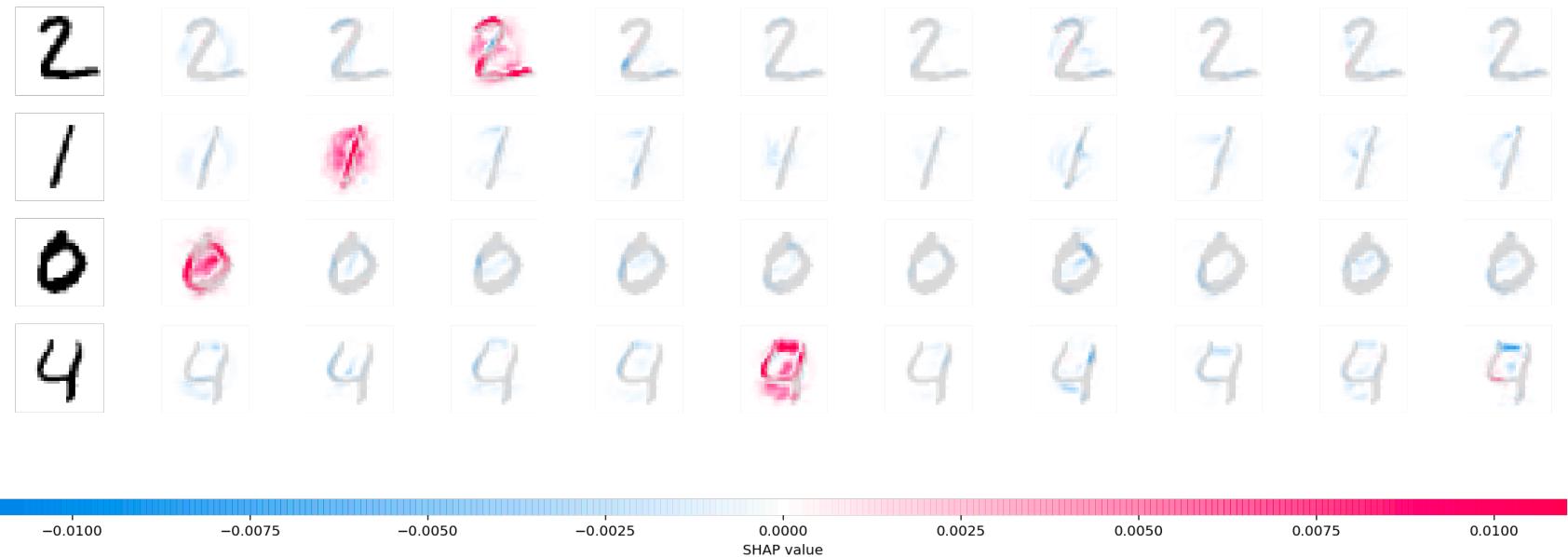
XAI within the medical filed → Make predictions transparent

Background and Theory



A deep learning application for image classification

Background and Theory

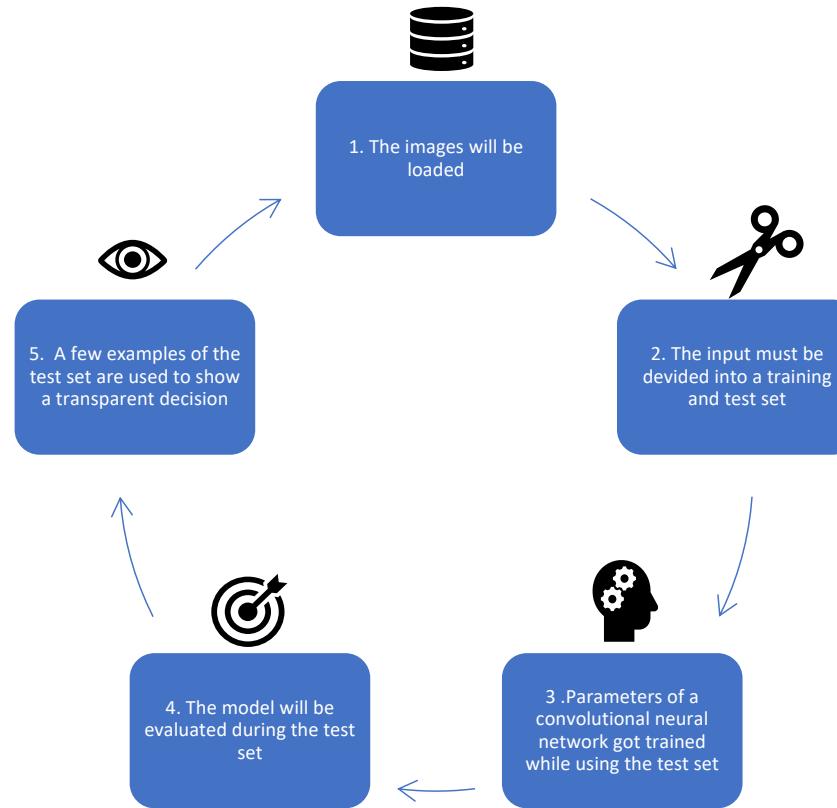


Implementation of Shapley values and integrated gradients as a method which could make deep learning more transparent by investigation the future importance

Requirements

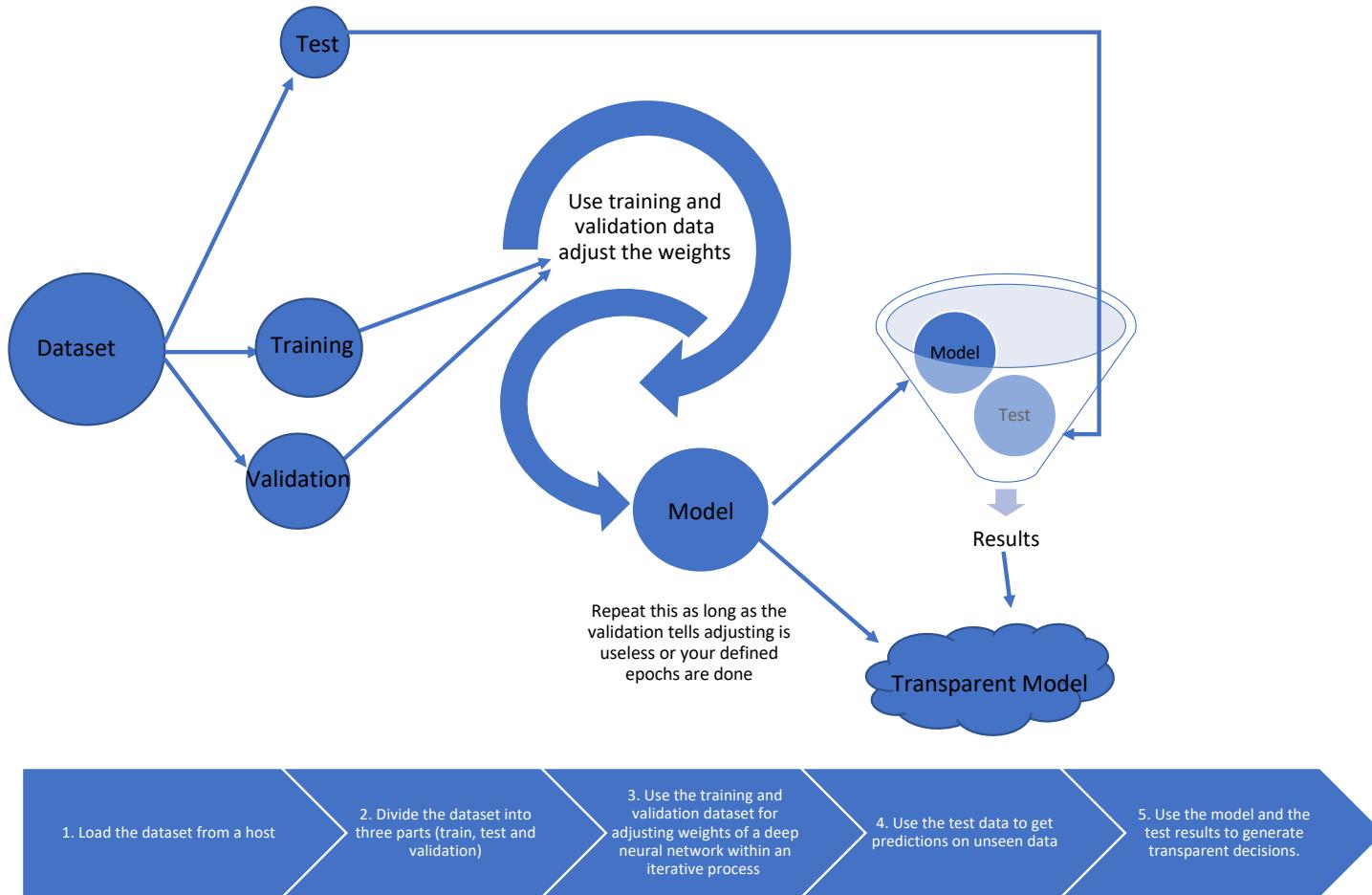
- Ability to predict
- Ability to calculate SHAP values
- Ability to visualize SHAP values

Requirements



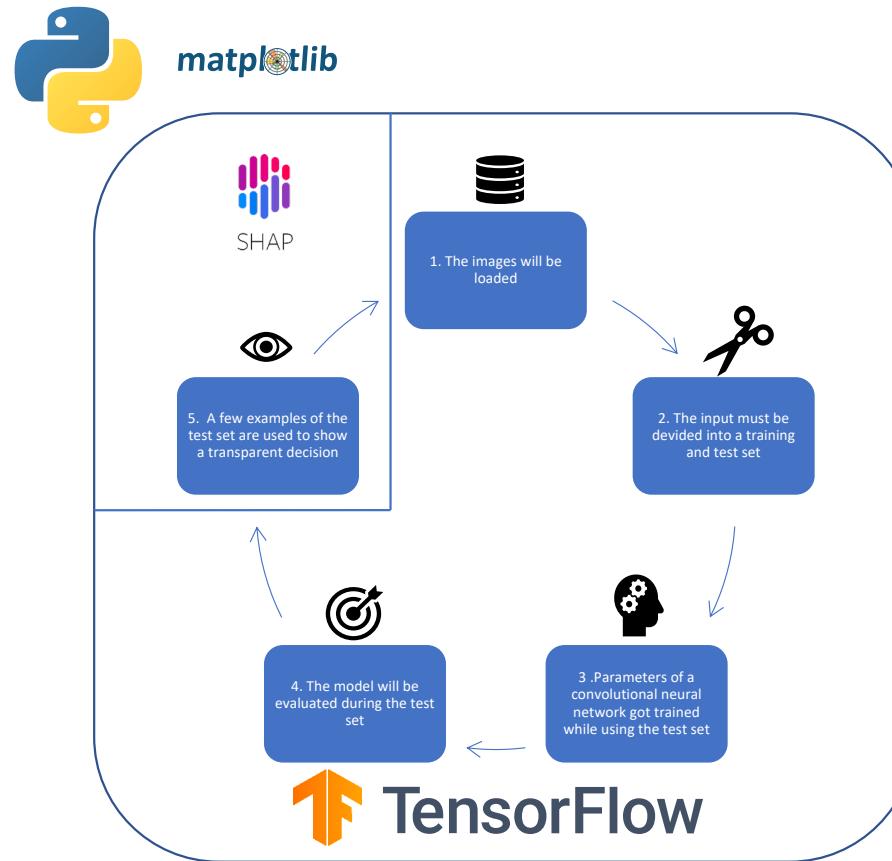
A prototype to apply methods of transparent decision making

Functional Specifications



How data gets processed within the prototype

IT Specifications



Which technologies are used within the prototype implementation

Generalization

Difficult because of the assumptions



Deep Learning and XAI are difficult to implement in real-world upload filter applications



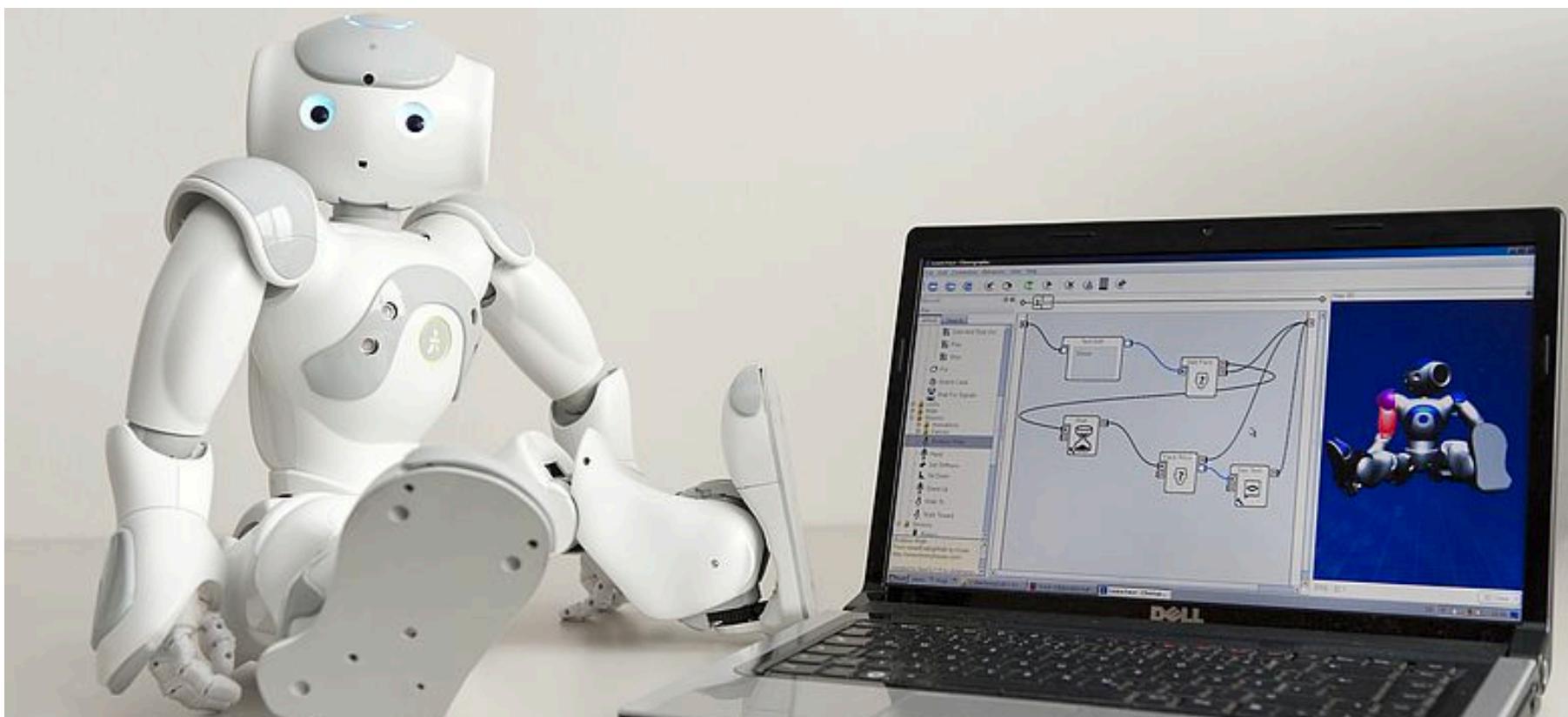
More research has to be done in order to make XAI applicable in upload filters

Conclusion

- Ability to predict
- Ability to calculate SHAP values
- Ability to visualize SHAP values



Thanks for your attention!





Sources

Figure 1: Photo by Tim Lüdemann, 2019, retrieved from <https://advox.globalvoices.org/2019/03/07/we-are-not-bots-in-berlin-thousands-protest-proposed-eu-regulation-on-internet-upload-filters/>

Figure 2: Photo by Julia Reda, 2019, retrieved from <https://juliareda.eu/eu-copyright-reform/censorship-machines/>

Figure 3: Photo by Nisarg Dave, 2019, retrieved from <https://www.linkedin.com/pulse/explainable-ai-understanding-decisions-data-driven-deep-nisarg-dave/>

Figure 4: Katharina Weitz in their 2019 paper titled "[Applying Explainable Artificial Intelligence for Deep Learning Networks to Decode Facial Expressions of Pain and Emotions](#)"

Figure 5: Scott Lundberg, et al. in their 2017 paper titled "[A Unified Approach to Interpreting Model Predictions](#)"

Figure 6: Created by Timo Bohnstedt Jan 2020

Figure 7: Scott Lundberg, et al. in their 2017 paper titled "[A Unified Approach to Interpreting Model Predictions](#)" retrieved from <https://github.com/slundberg/shap>

Figure 8: Created by Timo Bohnstedt June 2020

Figure 9: Created by Timo Bohnstedt June 2020

Figure 10: Created by Timo Bohnstedt June 2020