Do-It-Yourself Artificial Intelligence Supporting the DoD and the IC

CSIAC Webinar Series

June 29, 2022

Dr. Anthony Hoogs
Vice President of AI
Kitware, Inc.
anthony.hoogs@kitware.com

Topics

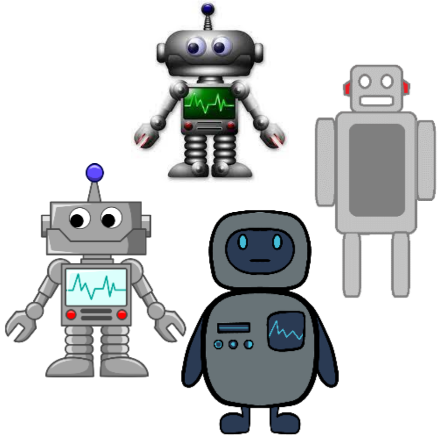
- Do-It-Yourself AI in practice
- Explainable AI for interactive search
- The XAI Toolkit

Many thanks to NOAA and the DARPA Explainable AI program for funding this work!

What is Do-It-Yourself AI?

What is in my data?

- How many objects of type X?
- How many events of type Y?
- How often do X and Y occur together?
- Are there trends I should know about?
- What is common and what is rare?



Can AI answer my questions?

- Which AI algorithm?
- How do I use it?
- Will it work?
- How do I know if it works?

Analytic results

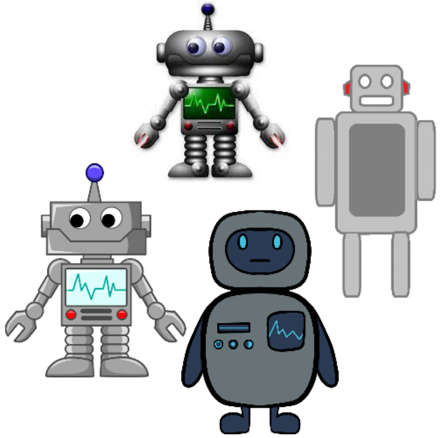
- Object counts and locations
- Event counts and durations
- Scaling to massive datasets
- Discovery of anomalies and novelties
- Customized, helpful answers to specific questions

Do-It-Yourself AI is…

An implemented methodology enabling end-users to build customized AI solutions to their specific problems with no programming or knowledge of how AI works

User inputs, customization and guidance

- Interactive feedback
- Selective labeling
- Algorithm selection



Library of AI algorithms

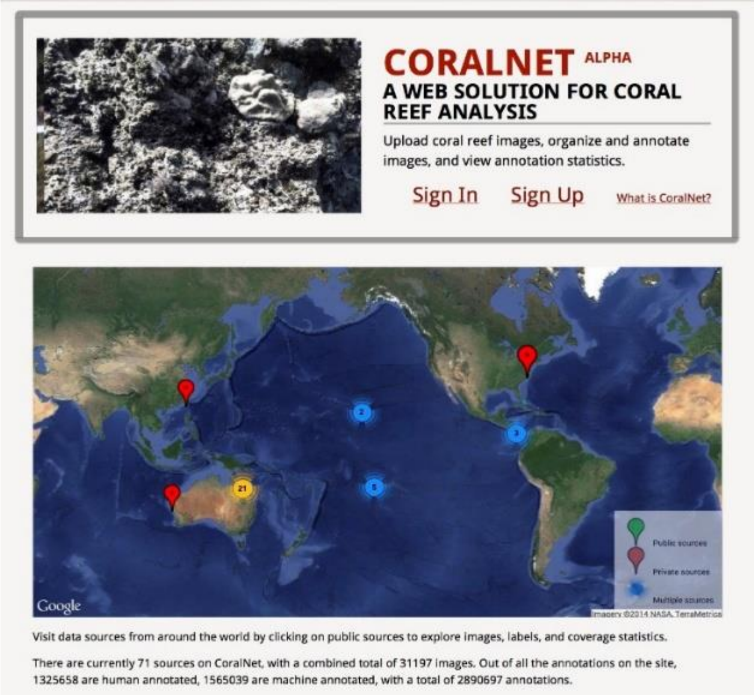
- Detection & classification
- Tracking
- Measurement
- Explainability

Analytic results

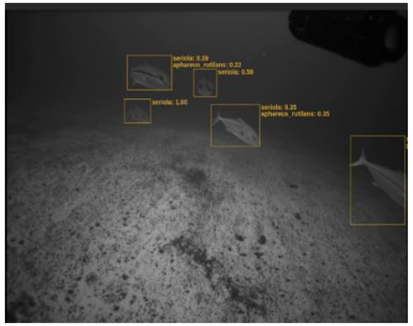
- Object counts and locations
- Event counts and durations
- Scaling to massive datasets
- Discovery of anomalies and novelties
- Customized, helpful answers to specific questions

NOAA National Marine Fisheries Service Strategic Initiative on Automated Image Analysis

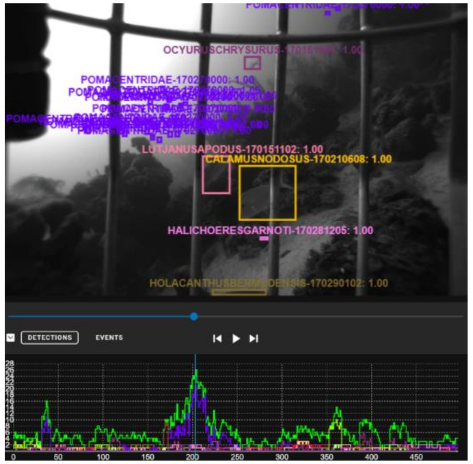
Mission: Develop guidelines, set priorities, and fund projects to develop broad-scale, standardized, and efficient automated analysis of still and video imagery for use in underwater stock assessment. 2014 – 2019.



http://coralnet.ucsd.edu



Funded VIAME and CoralNet from 2015 to present



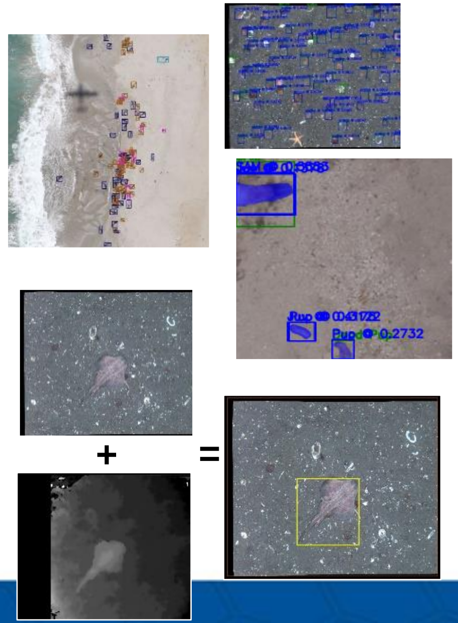
2019 Department of Commerce Gold Medal Awarded to NOAA Members of AIASI for VIAME and CoralNet

Video and Image Analytics for Marine Environments (VIAME)

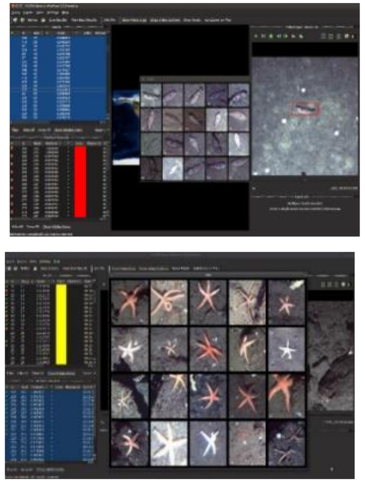
viametoolkit.org

- A do-it-yourself AI toolkit for virtually all types of imagery or video
- Designed for users with no programming or machine learning experience
- Sponsored by NOAA Fisheries
- Released as fully open-source with a permissive license
- Specializations to maritime processing such as stereo measurement and image enhancement
- In operational use at dozens of NOAA and marine science labs worldwide

Object Detection



Video Search and Rapid Model Generation



Object Tracking

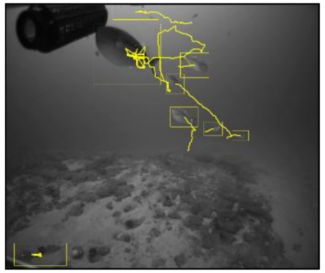
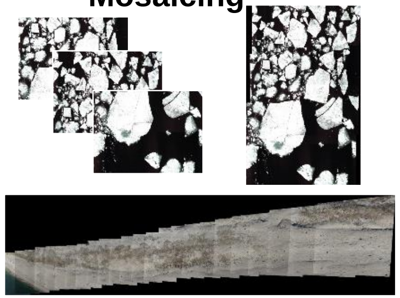
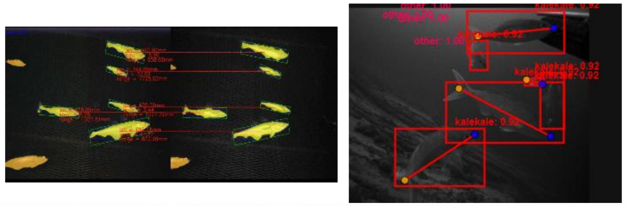


Image Enhancement

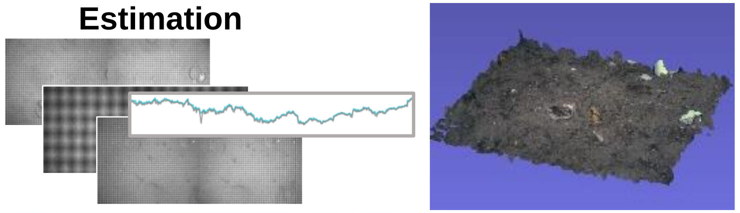
Image Registration and Mosaicing



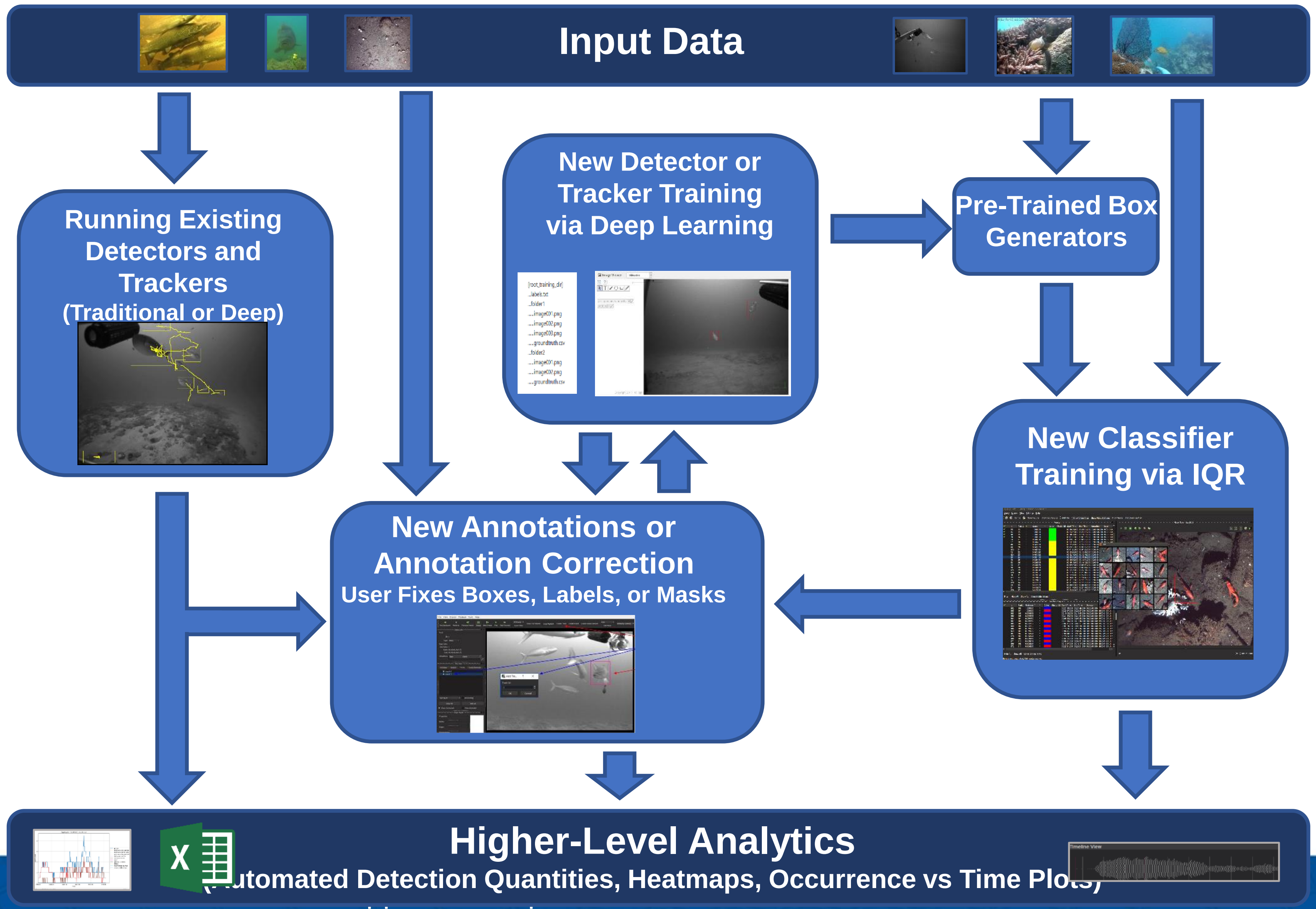
Stereo Measurement



Calibration, 3D and Altitude Estimation



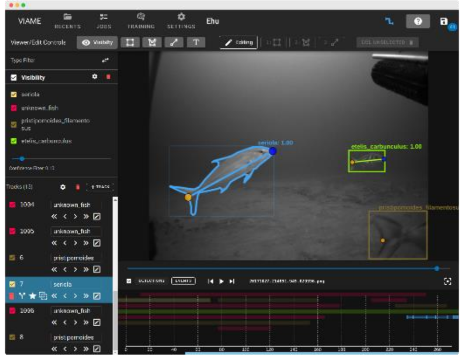
VIAME Methods to Create AI Analytics

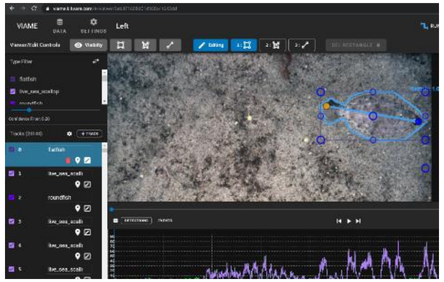


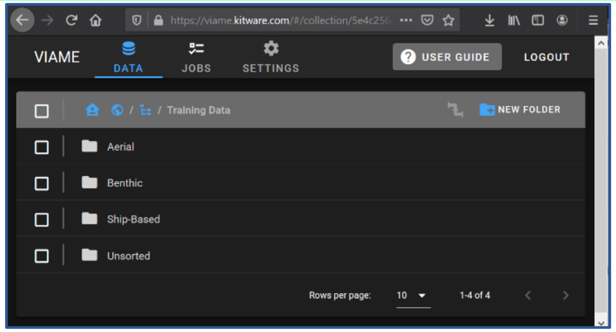
Kitware

Web Applications and Annotation Archive

User Complexity,
Algorithm Customization





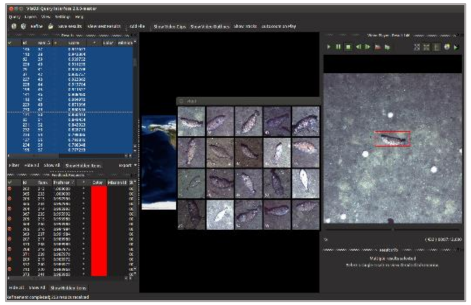


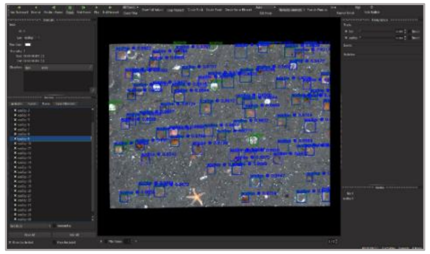
Online Example: **viame.kitware.com**

Server Manages Data and Annotations

15 Tb open data store (raided), 2 GPUs for training

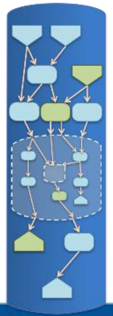
Desktop Applications

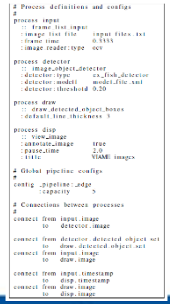


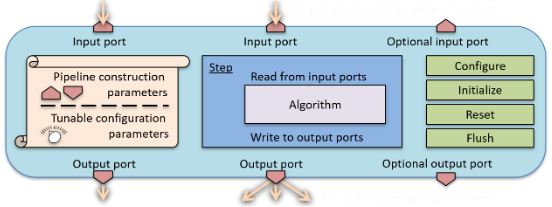


User Manages Data and Annotations

Command-Line Tools and APIs:







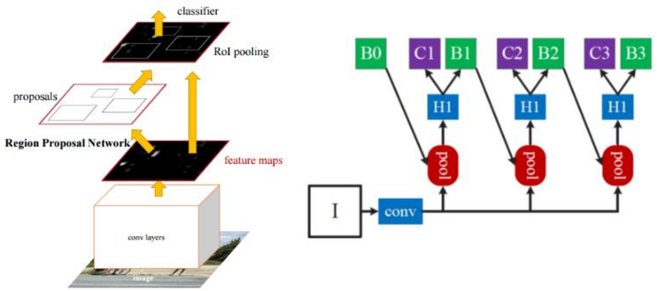
Full Feature Support

More Customization Ability

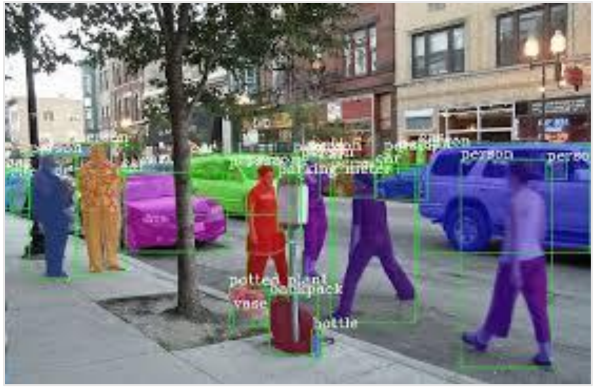
Useful for Embedded Platforms

Baseline Object Detectors

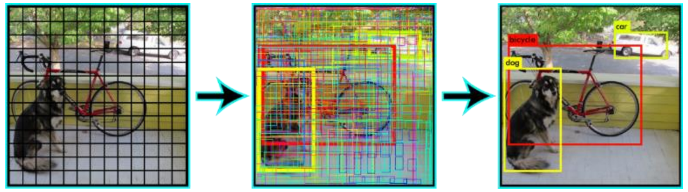
Cascade Faster R-CNN [1]



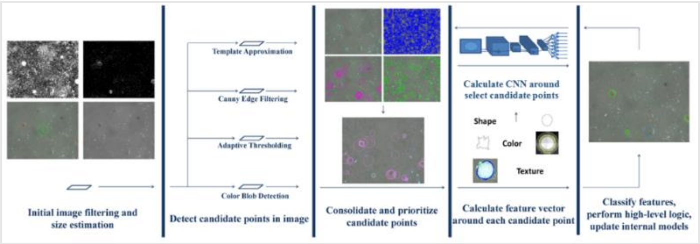
Cascade Mask Faster R-CNN [2]



YOLOv3 and v4 [3,4]



Scallop-TK



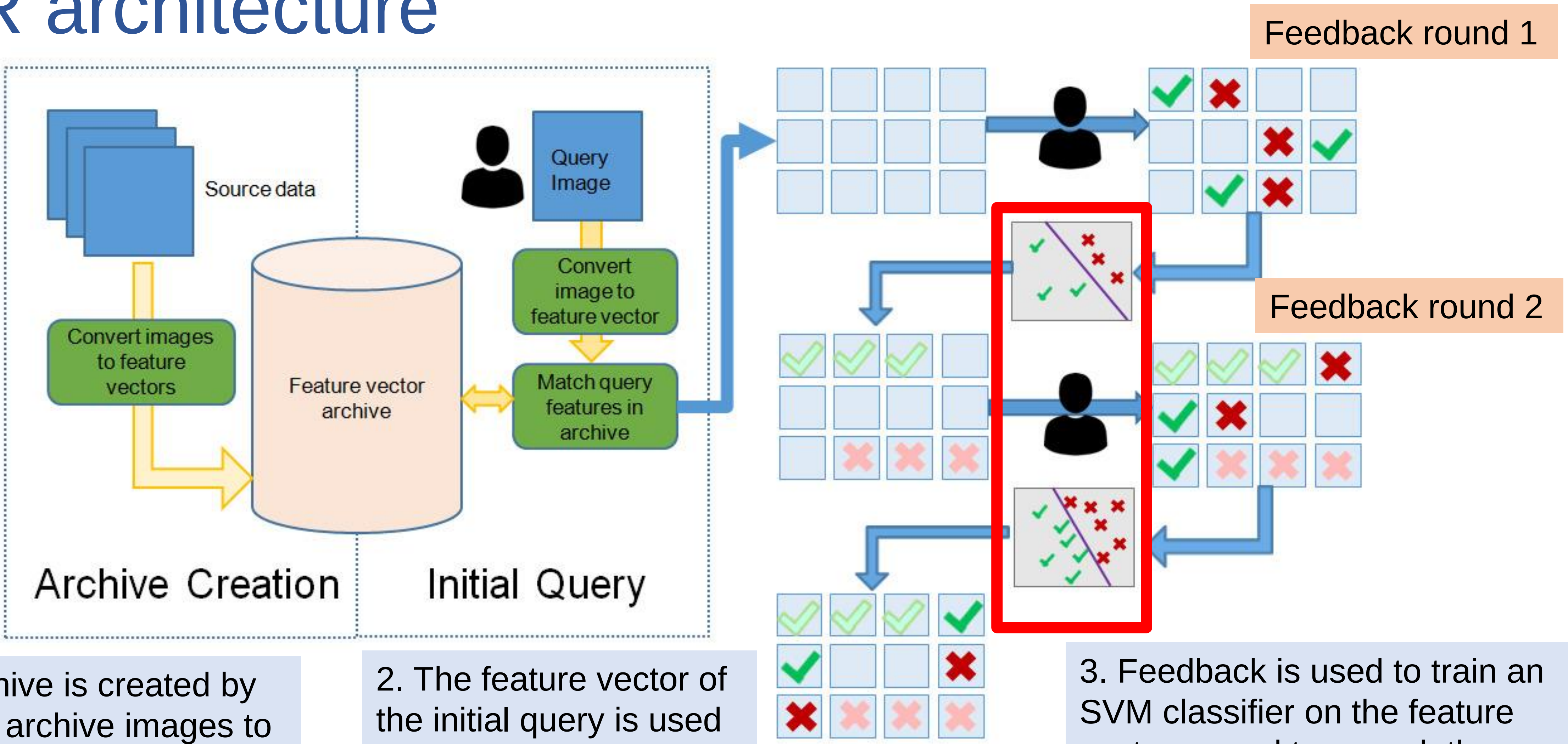
VIAME contains multiple baseline general purpose detectors from the larger computer vision community for wide applicability, but then specializations and other functionality added specific to domains of interest

[1] Cai, Zhaowei, et al. "Cascade R-CNN: Delving into High Quality Object Detection." CVPR 2018.
[2] Chen, Kai et al. “MMDetection: Open MMLab Detection Toolbox and Benchmark.” arXiv preprint 2020.
[3] Redmon, Joseph, and Ali Farhadi. "YOLOv3: An Incremental Improvement." arXiv preprint 2018.
[4] Bochkovskiy, Alexey et al. "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv preprint 2020.

Object Search and Rapid Detector/Classifier Training with Interactive Query Refinement (IQR)

Airplane detector & classifier

IQR architecture

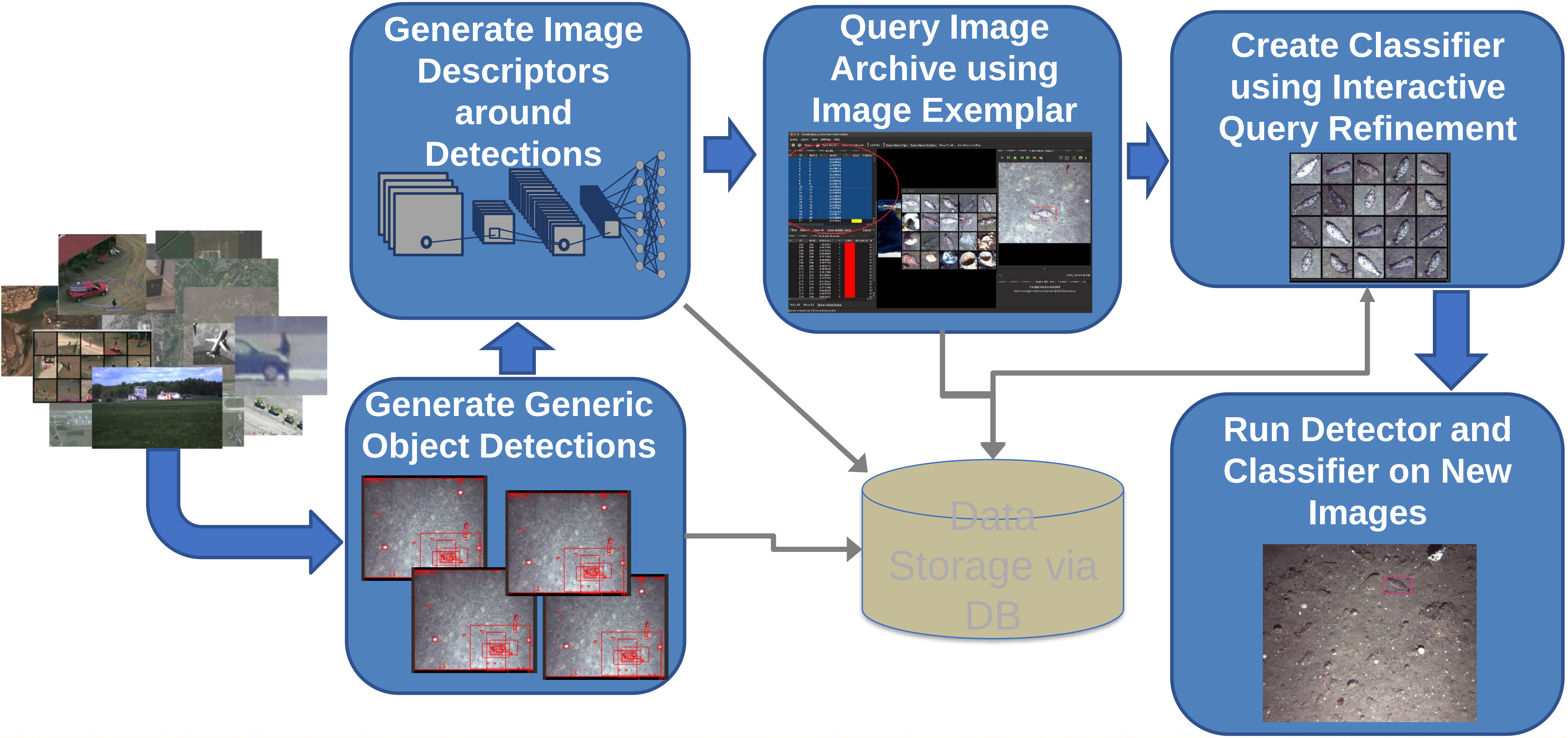


1. The archive is created by converting archive images to feature vectors

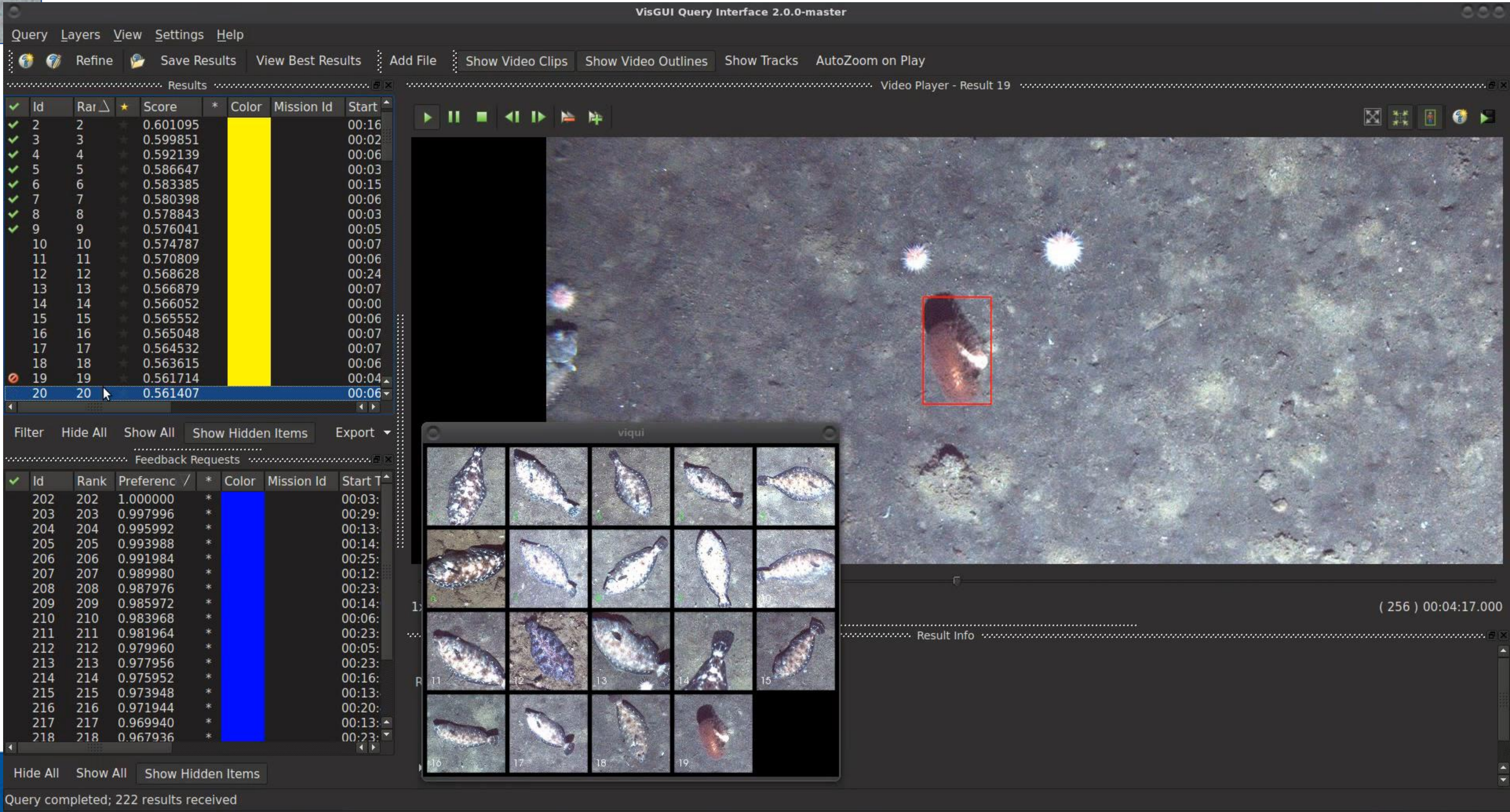
2. The feature vector of the initial query is used to return the initial result set

3. Feedback is used to train an SVM classifier on the feature vectors, used to re-rank the results so that higher results are more relevant

Interactive Search and Rapid Model Generation



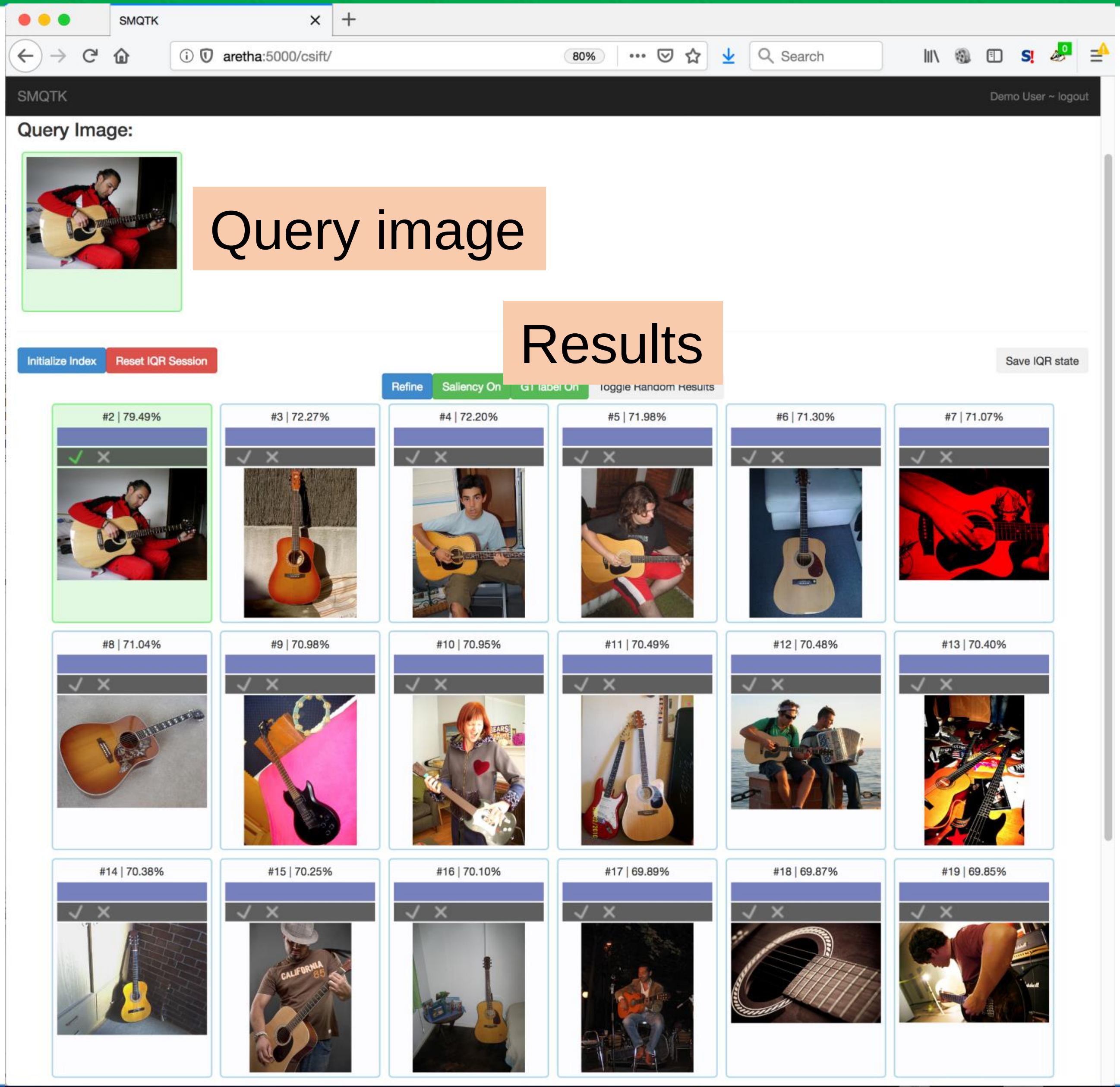
Interactive Search and Rapid Model Generation



SMQTK

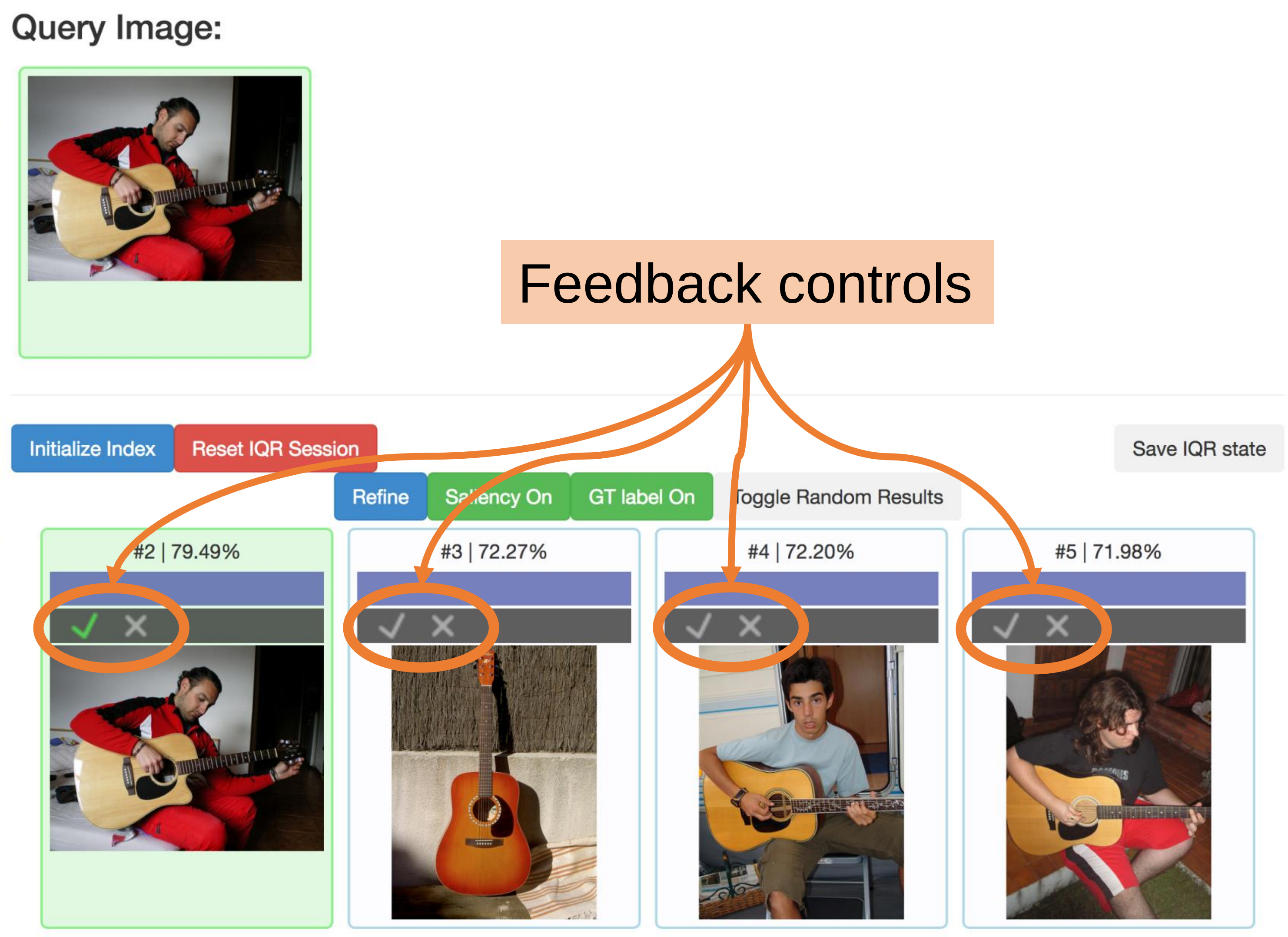
- Social Multimedia Query Toolkit (SMQTK) is an open-source, python-based app for interactive data search and IQR
- Aimed at images and videos with a web-based GUI
- The user starts the session by supplying a query image; the system selects the initial result set based on visual similarity
- This example uses an archive of ~50k images from ImageNet and feature vectors from ResNet

https://github.com/Kitware/SMQTK



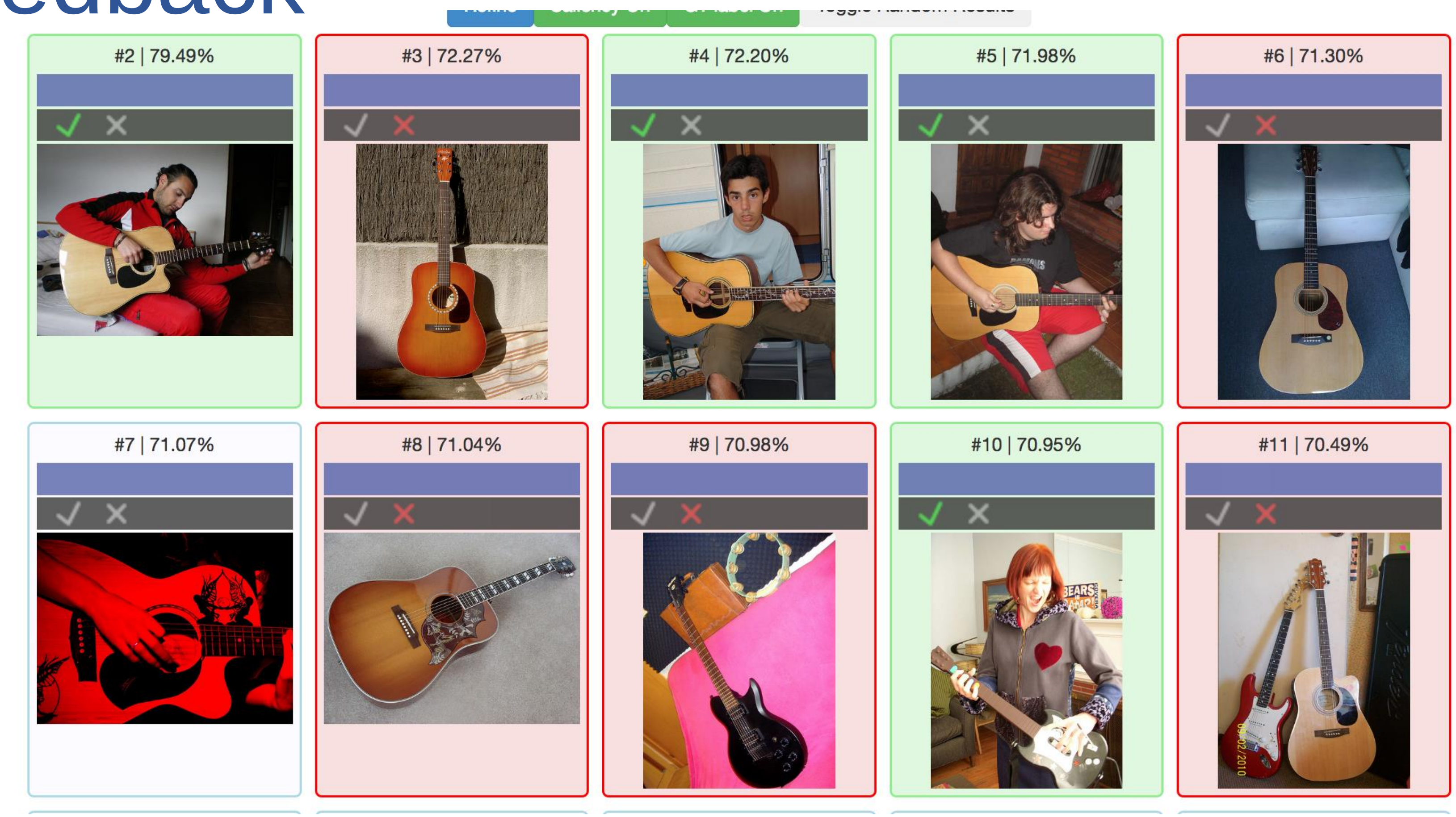
SMQTK GUI: IQR

- Each result has a positive / negative feedback indicator the user may set to indicate if that result is relevant to their query
  - The first result is the query, which is automatically set to “positive”
- For this example, **we will try to steer the results towards images with both people and guitars**



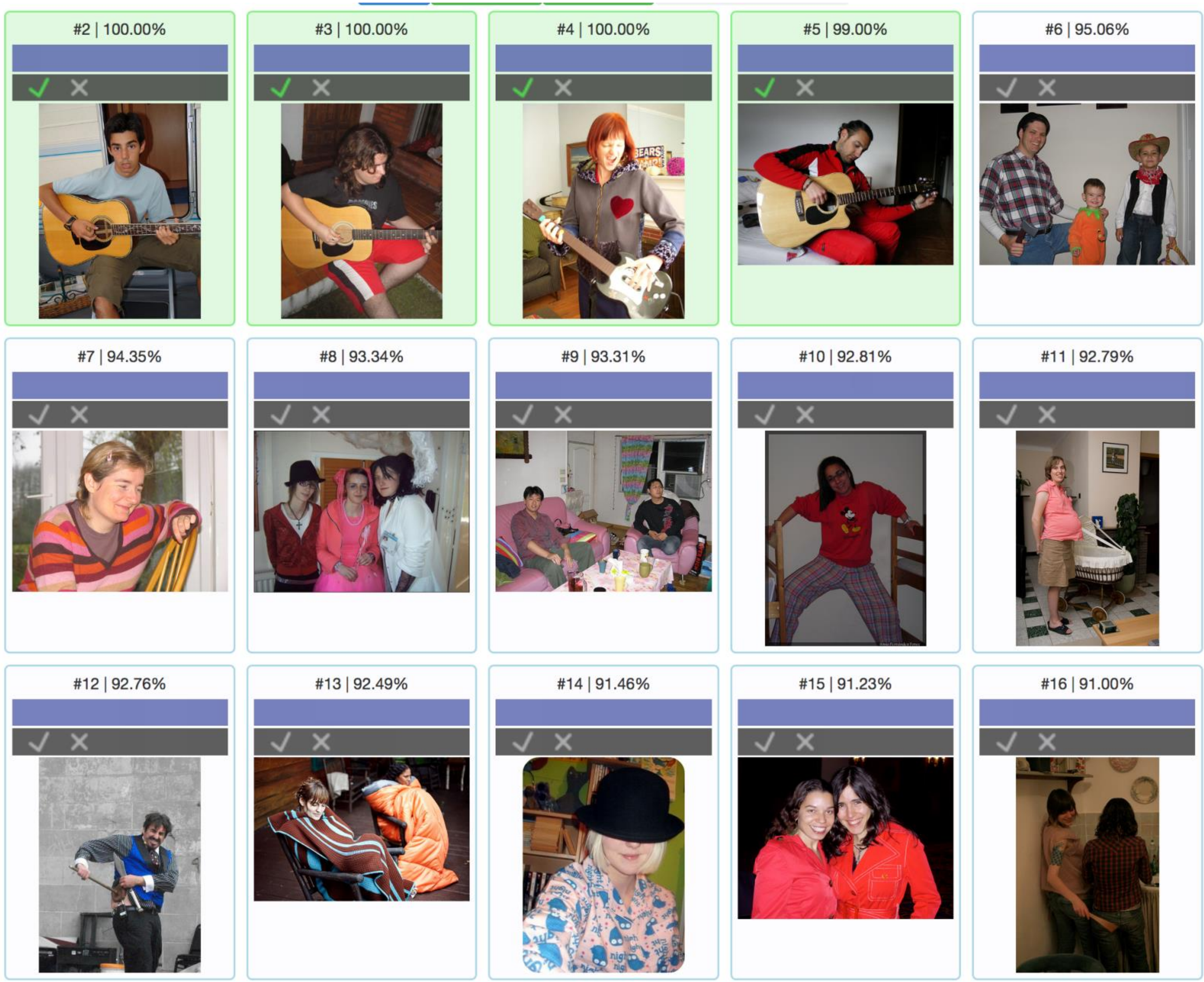
IQR Round 1: Feedback

- We give positive feedback to images with both people and guitars; negative feedback to guitars only
- No need to give feedback to every image
- Then click **REFINE** button to re-rank based on feedback



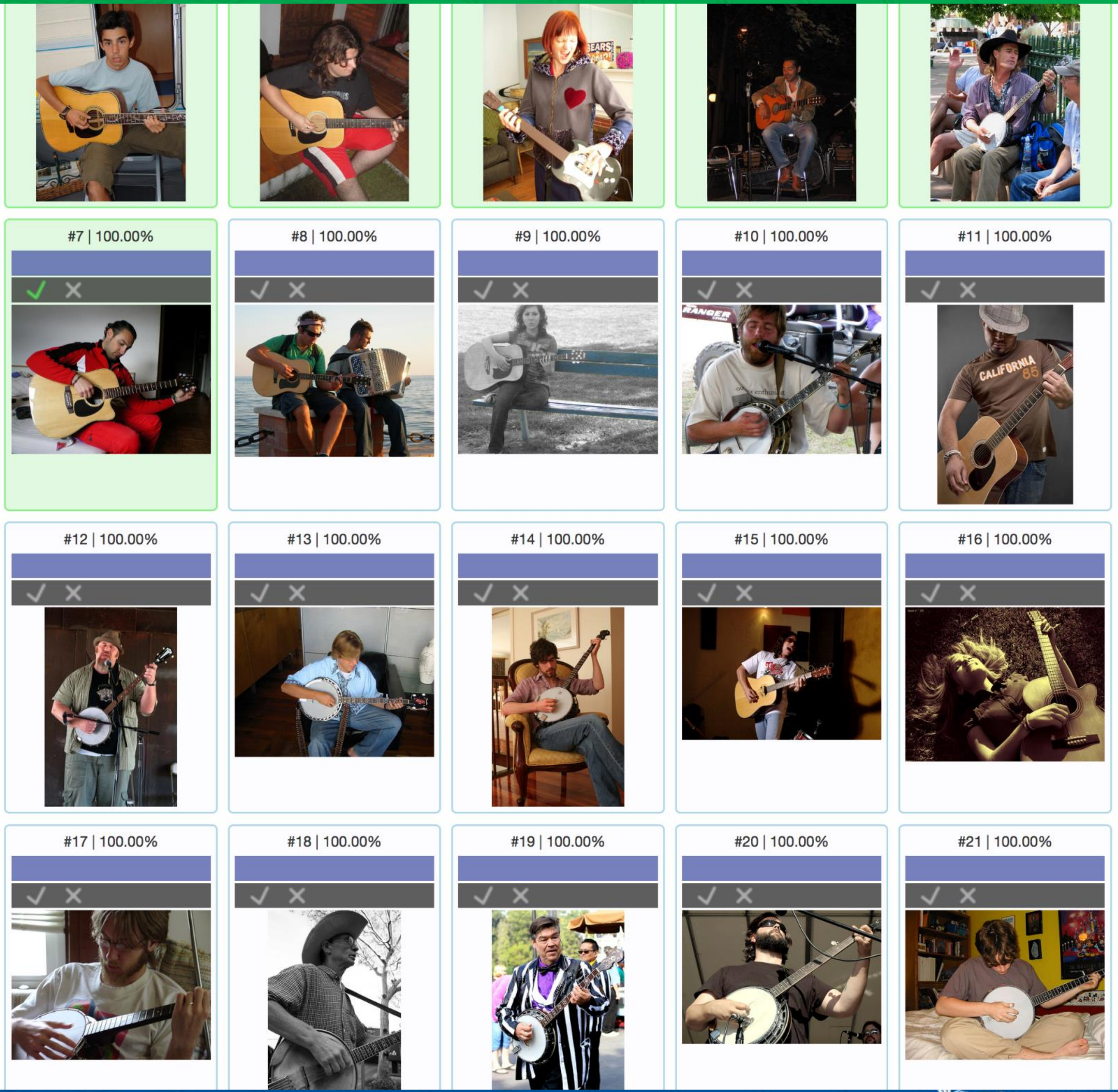
IQR Round 1: Results

- All novel images now have people but not guitars
  - Give negative feedback to new examples
  - Give positive feedback to two examples in next screen of guitar and banjo player
  - Click **REFINE** again



IQR Round 2: Results

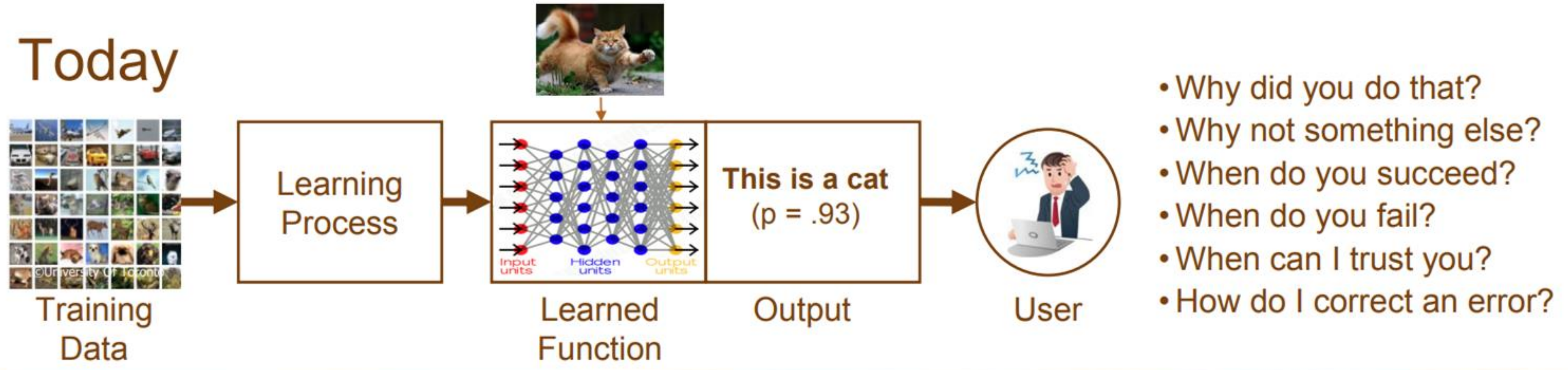
- Results now have a solid selection of people with guitars (and banjos)
- **Note that the classifier was trained using only one starting example and positive / negative feedback on results**



Topics

- Do-It-Yourself AI in practice
- Explainable AI for interactive search
- The XAI Toolkit

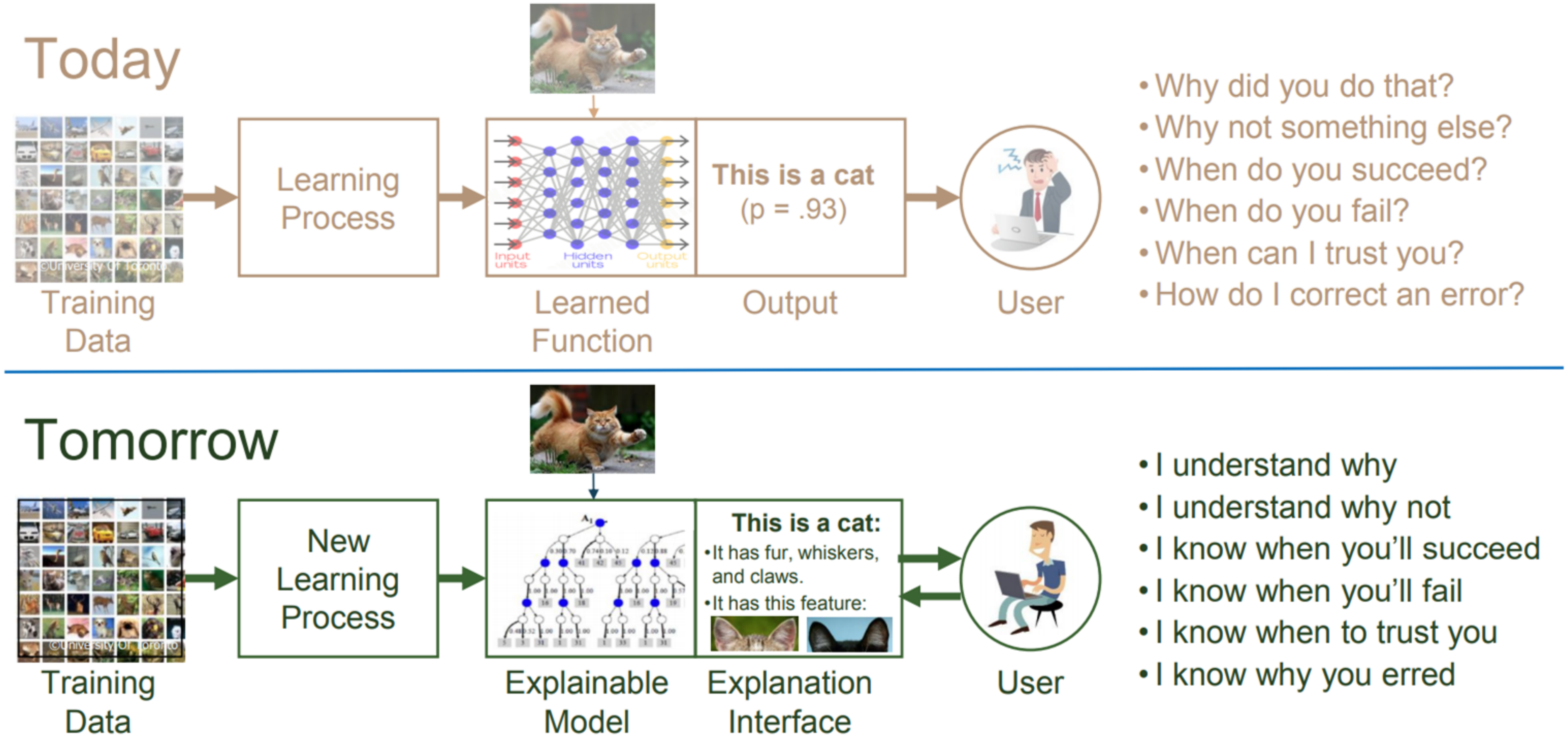
DARPA Explainable AI (XAI) Program



2016-2021

https://www.darpa.mil/program/explainable-artificial-intelligence

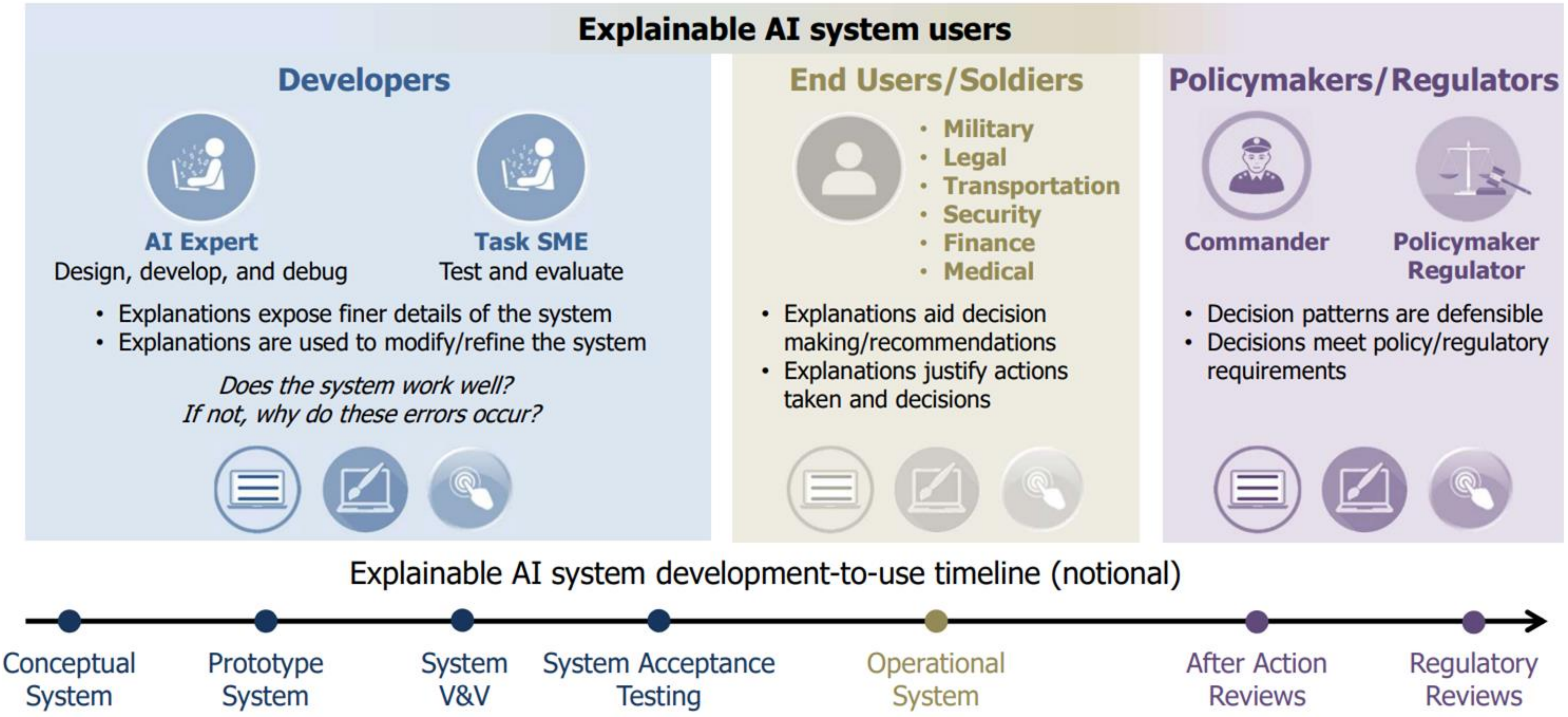
DARPA Explainable AI (XAI) Program



2016-2021

https://www.darpa.mil/program/explainable-artificial-intelligence

Diverse User Types for XAI



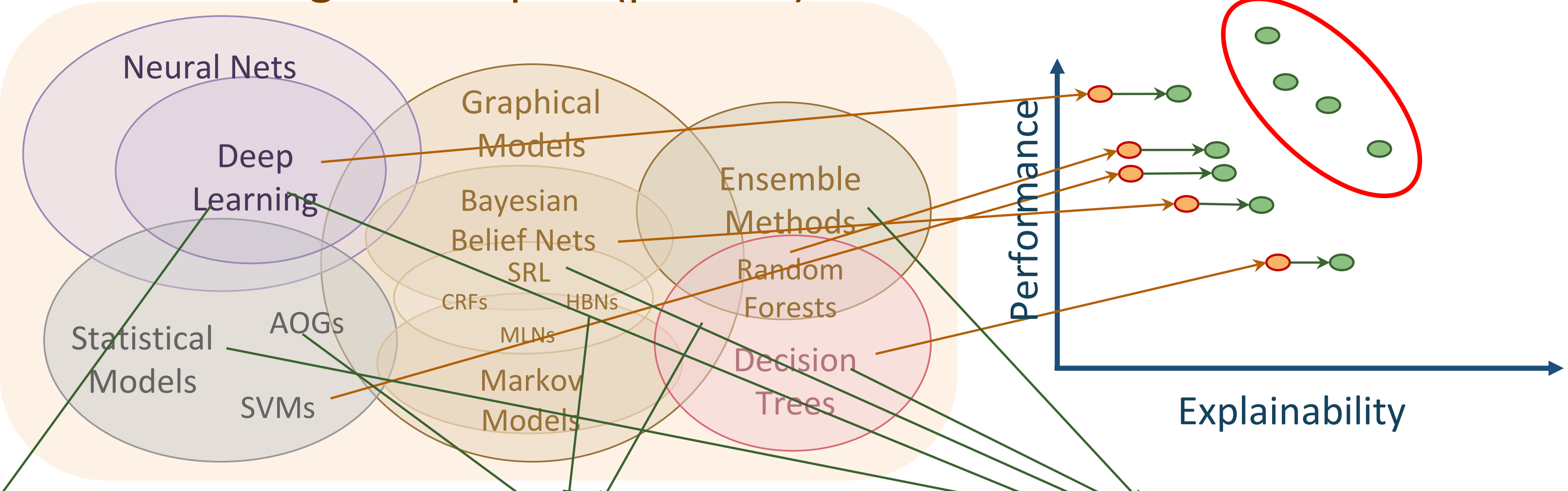
Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stump, S.; Yang, G.-Z. 2019. XAI—Explainable artificial intelligence. *Science Robotics* 18 Dec 2019: Vol. 4, Issue 37, eaay7120, DOI: 10.1126/scirobotics.aay7120.

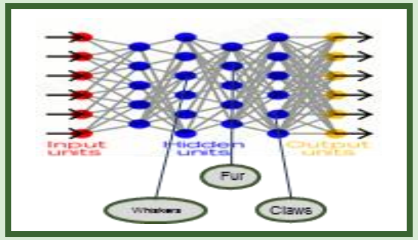
Performance vs. Explainability

New Approach

Create a suite of machine learning techniques that produce more explainable models, while maintaining a high level of learning performance

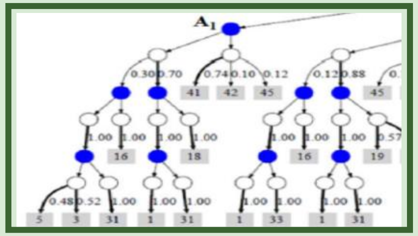
Can Explainability improve performance???





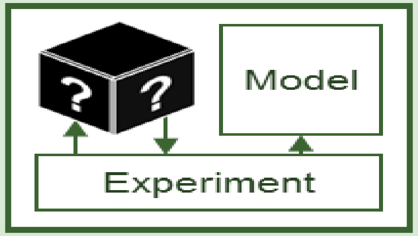
Deep Explanation

Modified deep learning techniques to learn explainable features



Interpretable Models

Techniques to learn more structured, interpretable, causal models



Model Induction

Techniques to infer an explainable model from any model as a black box

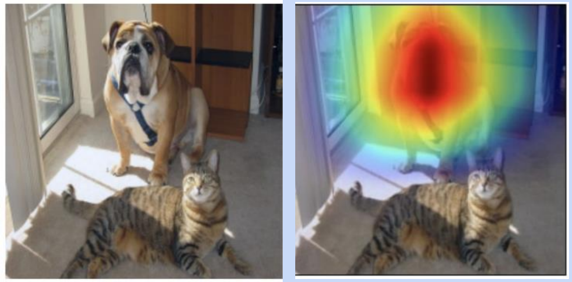
DARPA XAI slide

Saliency Maps as Explanations

Saliency maps are a form of ***visual XAI***, overlaying the input with (typically) a heatmap to **spatially indicate which areas of the input the AI found "important".** Saliency map algorithms, like other forms of XAI, can be classified as **black box** (model induction, or no knowledge of the associated AI) or **white box** (able to access the internal state of the AI).

White box examples

Saliency map for an algorithm on automatic text summarization (Tuckey et al.)



Grad-CAM result for "Dog" (Selvaraju et al.)

Black box examples

SBSM result showing where the image on the right is similar to the one on the left (Dong et al.)

RISE result for importance of "Sheep" (Petsiuk et al.)

Image Retrieval 1: Archive of Images (computed offline)

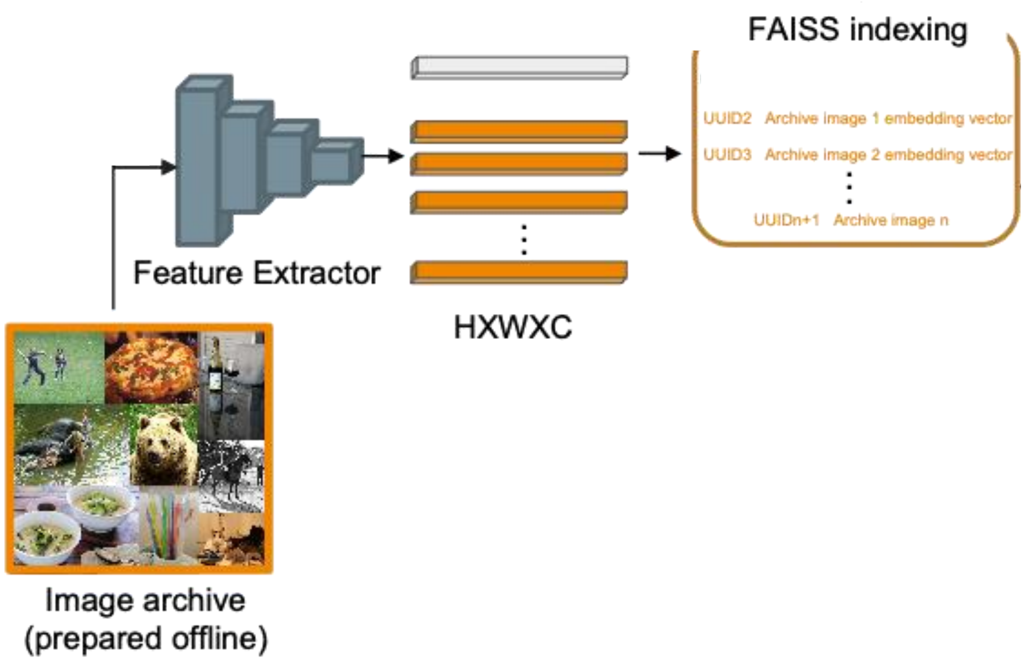


Image Retrieval 2: Query Image

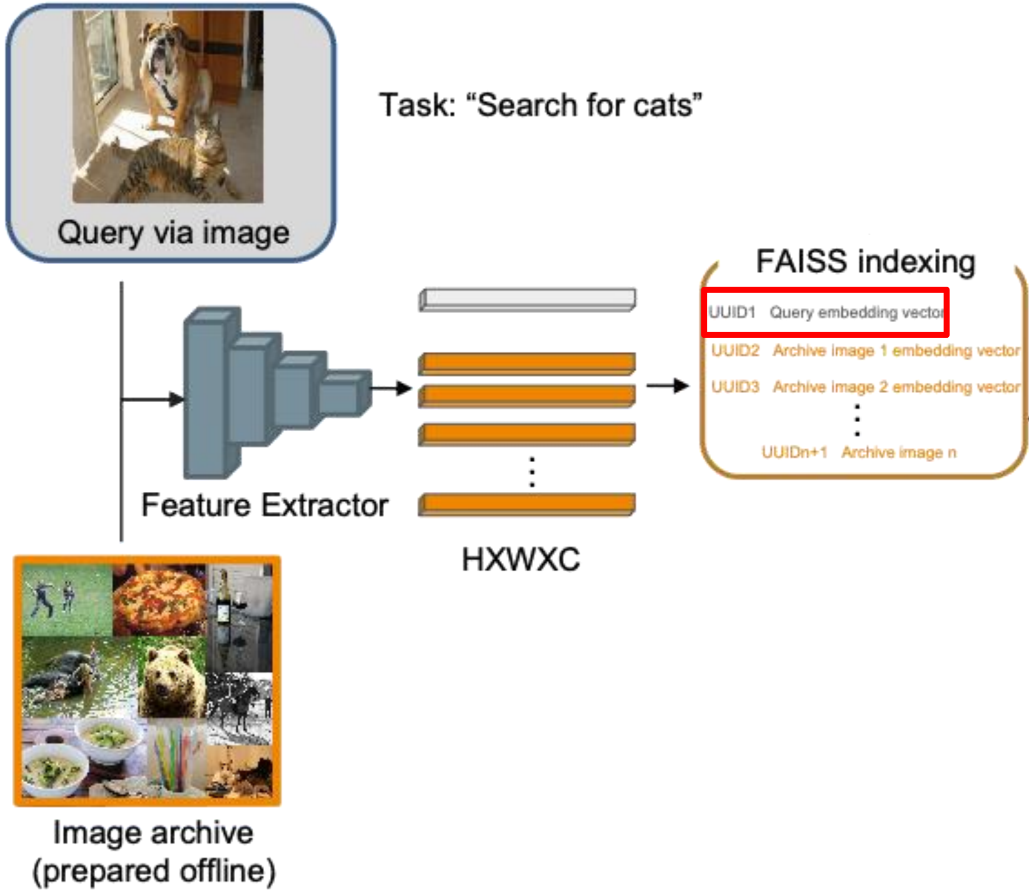


Image Retrieval 3: Feature-based Retrieval

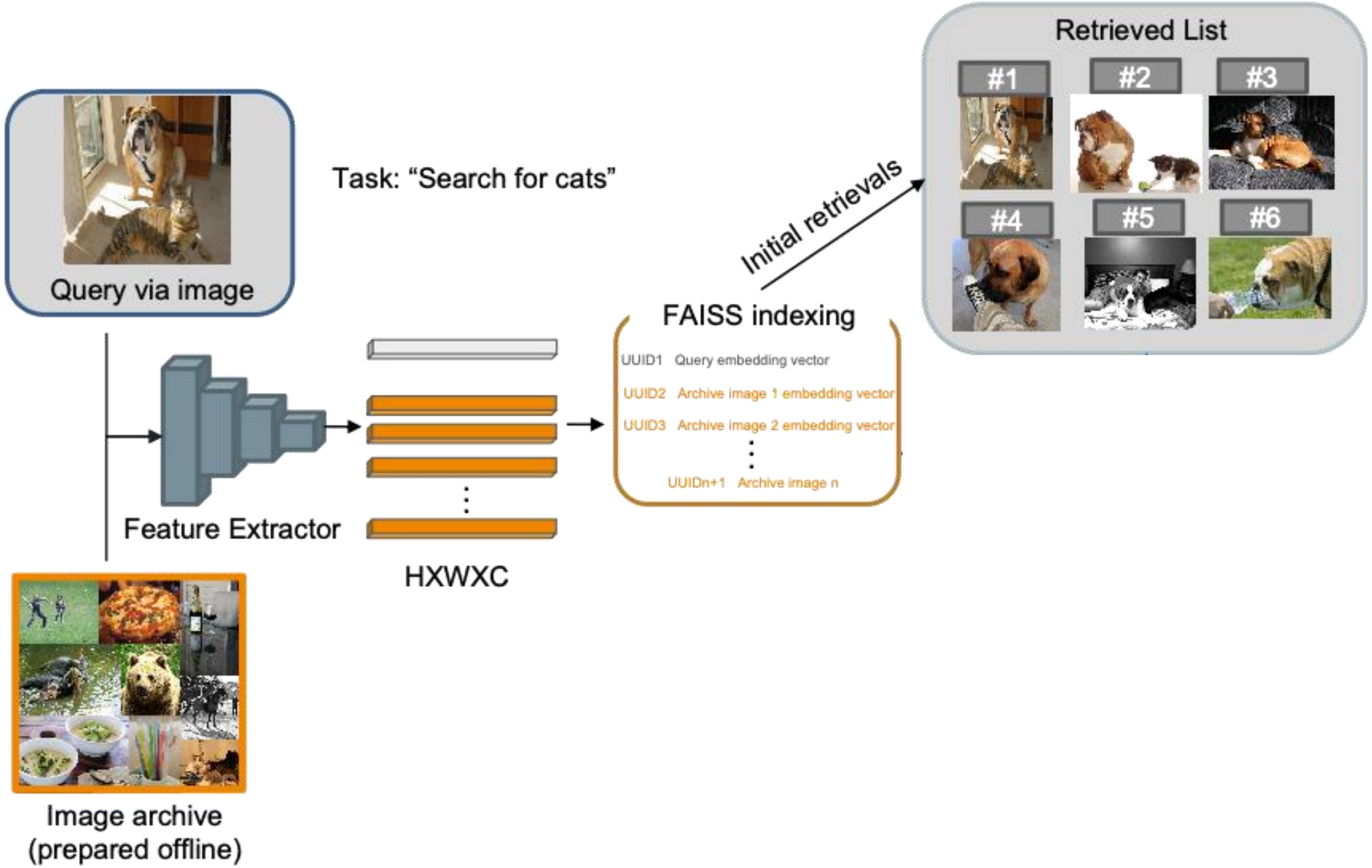


Image Retrieval 4: Feedback...

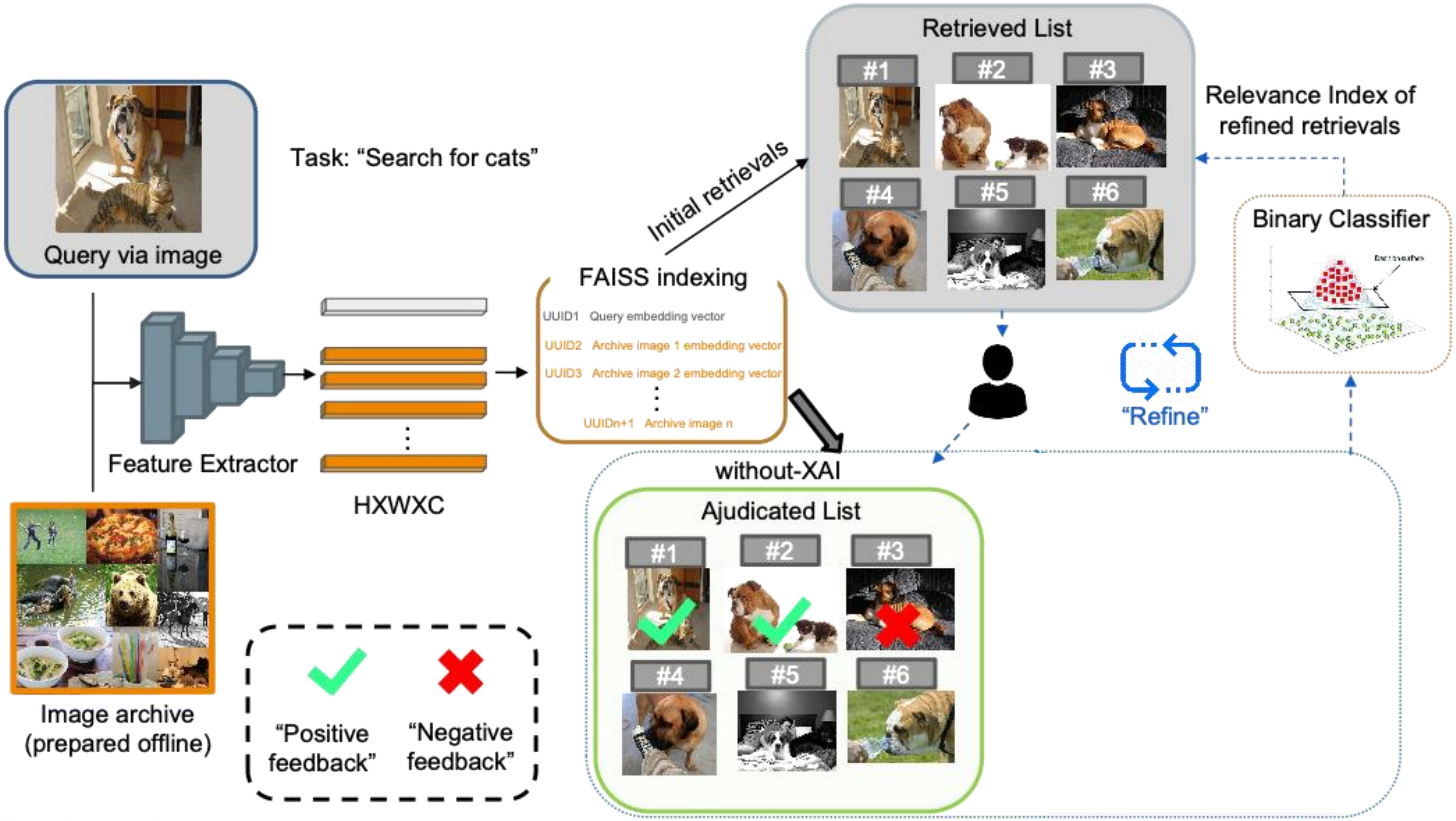


Image Retrieval 5: ...But why *that* image?

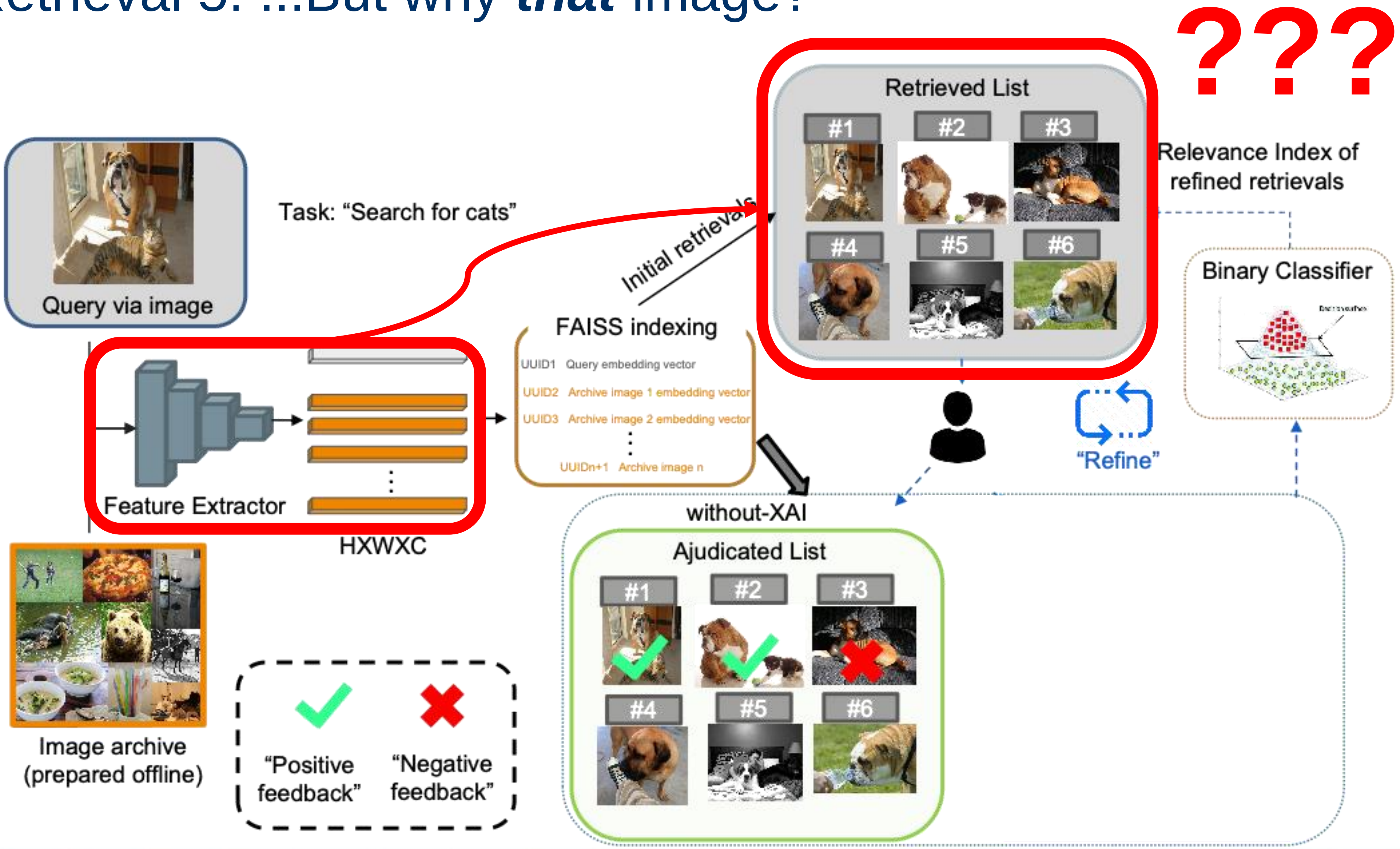
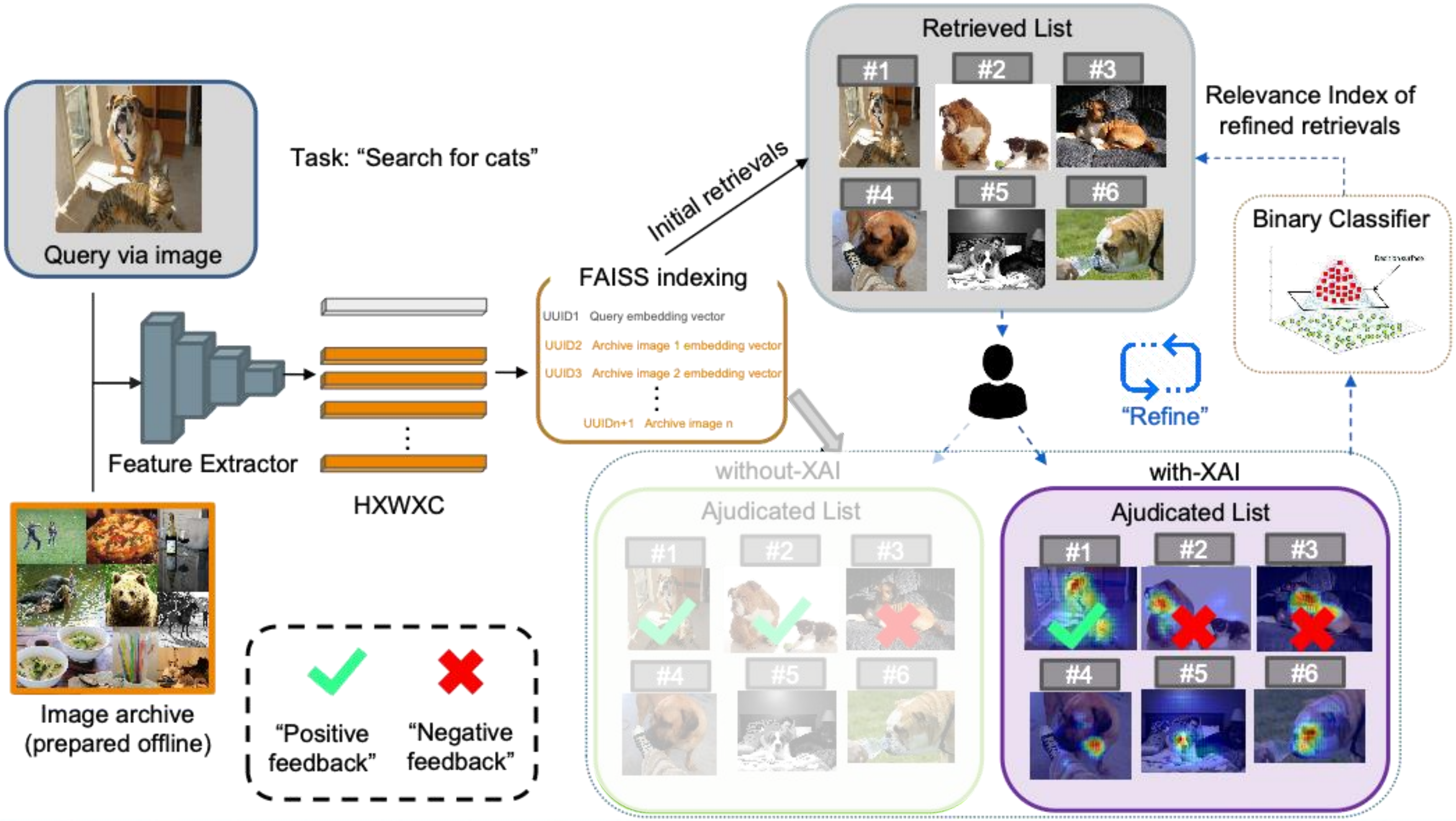
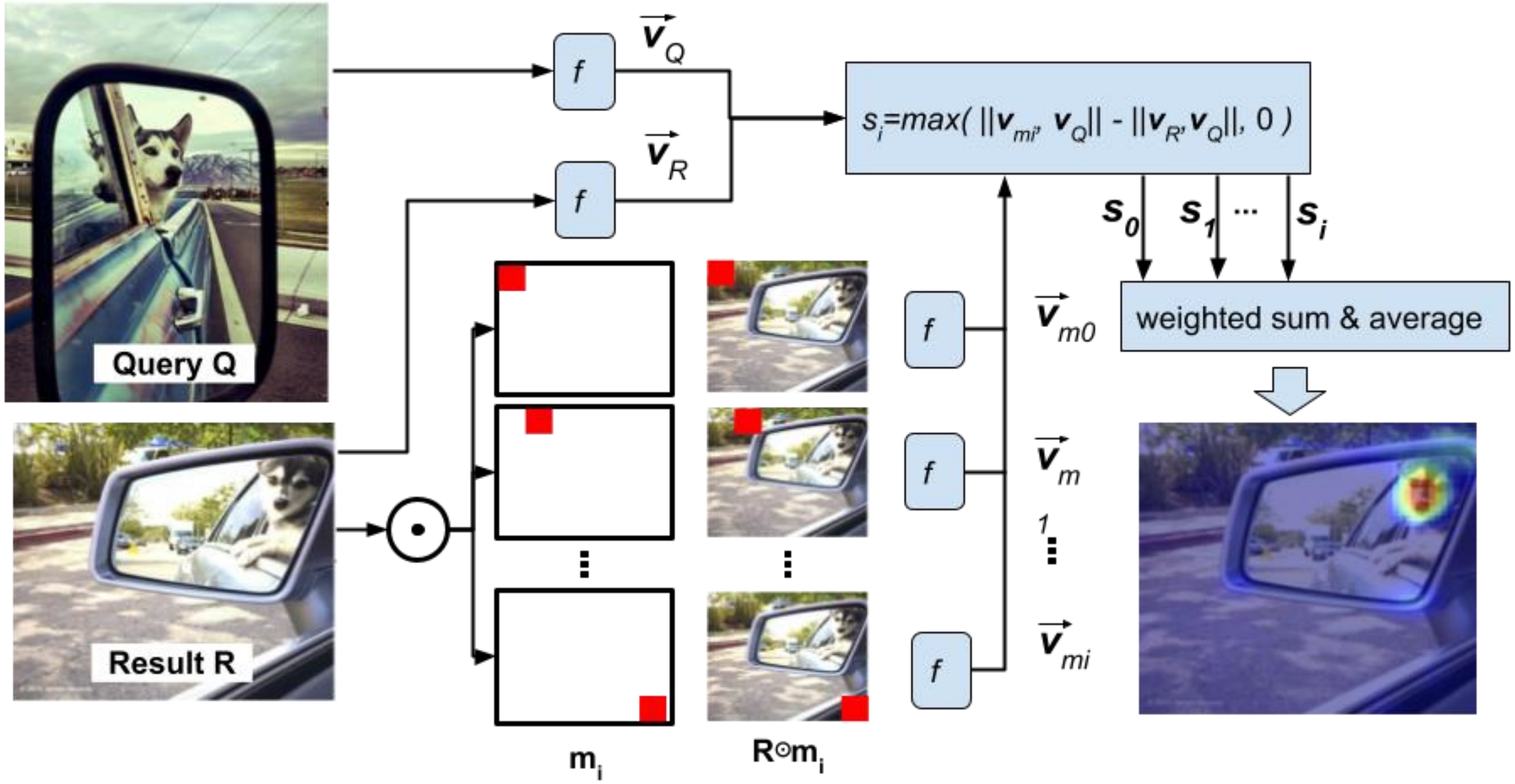


Image Retrieval 6: XAI to guide feedback



Similarity-Based Saliency Maps (SBSM)



Pros:

- \+ Black-box algorithm
- \+ Operates directly on image pairs
- \+ Reasonably fast

Cons:

- \- As implemented, only considers the feature vectors, rather than any higher-order model

[Dong et al., '19 (CVPR-W)]

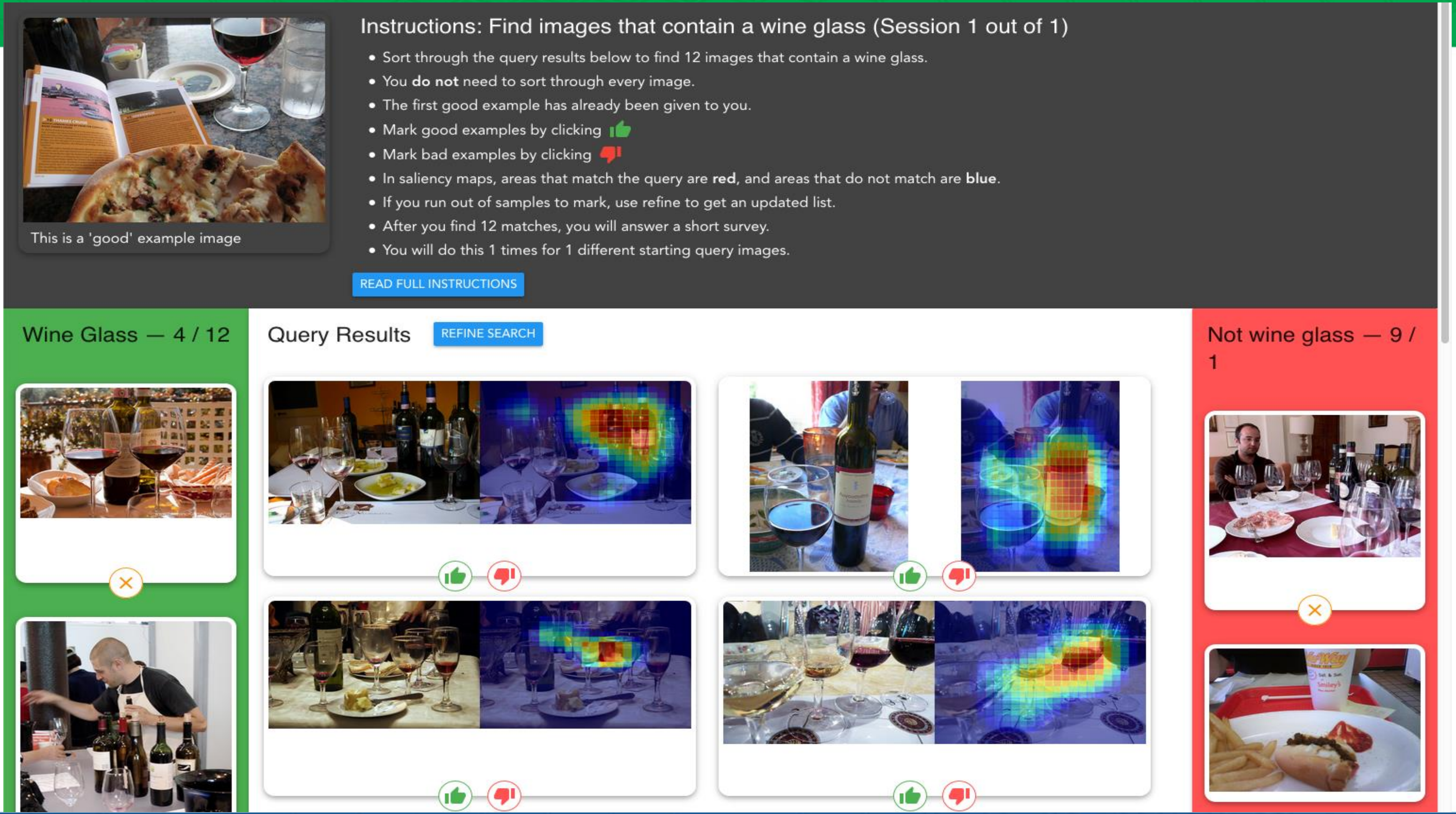
Pilot and Full Studies

Our utility-based evaluation focused on demonstrating that XAI could be integrated into an existing system and evaluating its performance relative to the baseline.

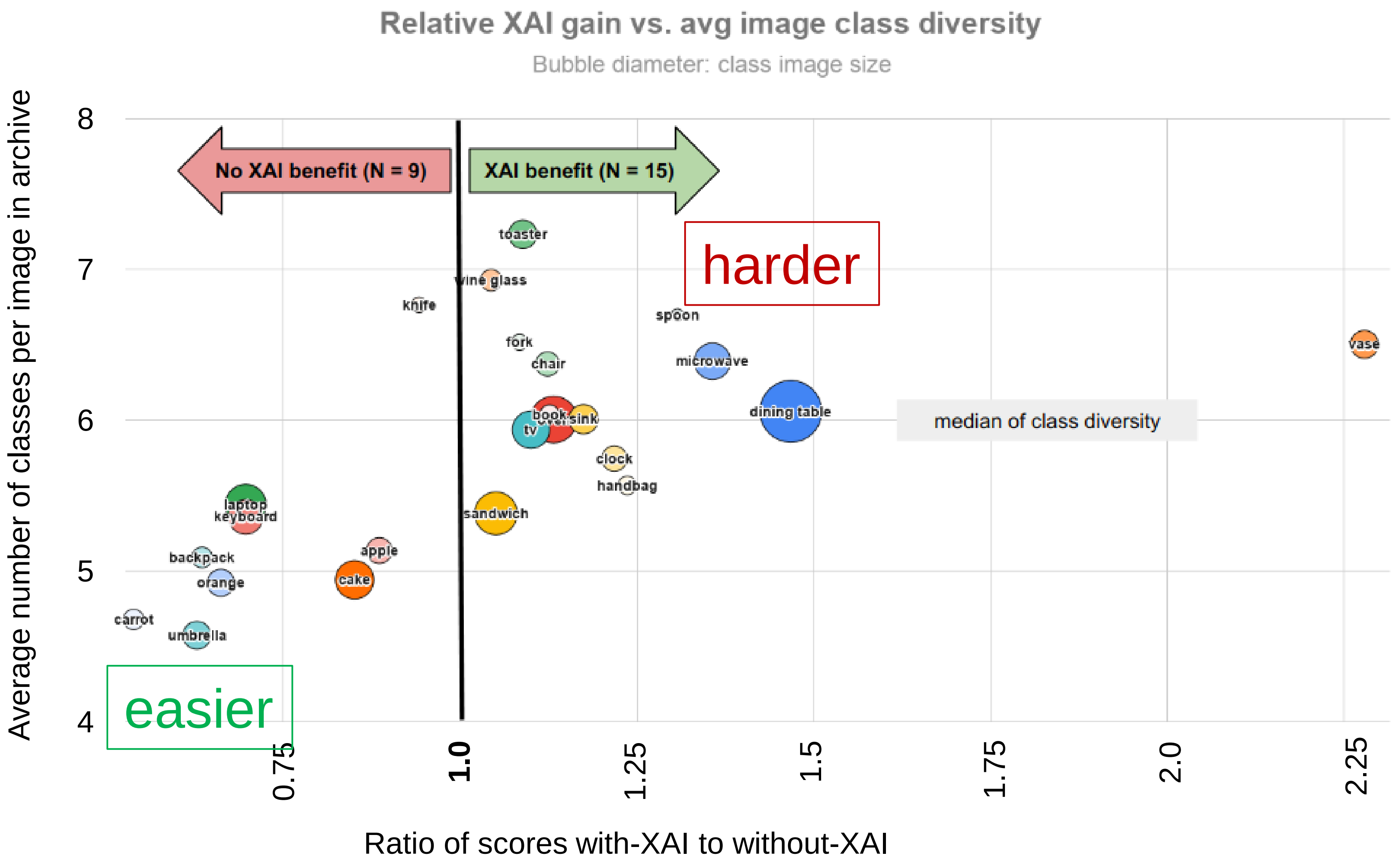
- **Goal:** Find 12 instances of an object class
- **Archive:** ~246,000 MS-COCO images
- **Users:** Amazon Mechanical Turk subjects

	Pilot	Full
Query pool classes	10	24
Query pool images	10	160
Subjects	104	476
Algorithm	F-Sal	SBSM

F-Sal: measures response of result against retrieval SVM
SBSM: measures response of result against query


Instructions: Find images that contain a wine glass (Session 1 out of 1)
Sort through the query results below to find 12 images that contain a wine glass.
You do not need to sort through every image.
The first good example has already been given to you.
Mark good examples by clicking
Mark bad examples by clicking
In saliency maps, areas that match the query are red, and areas that do not match are blue.
If you run out of samples to mark, use refine to get an updated list.
After you find 12 matches, you will answer a short survey.
You will do this 1 times for 1 different starting query images.
This is a 'good' example image
READ FULL INSTRUCTIONS
Wine Glass — 4 / 12
Query Results
REFINE SEARCH
Not wine glass — 9 / 1

Full Study Results: Quantitative



Each bubble is a query class; size is relative area of object in image.

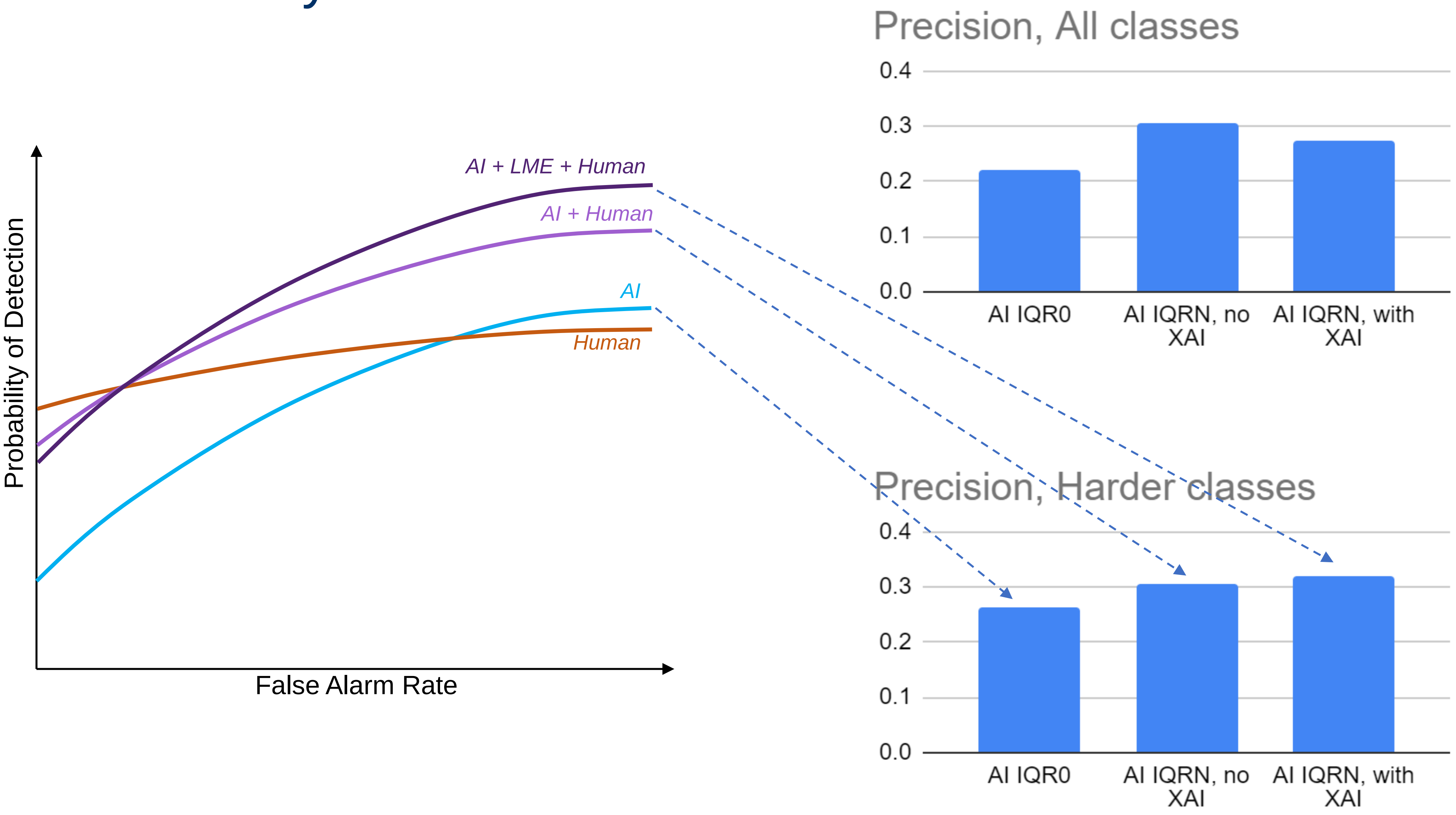
X-axis: ratio of count of retrieved images containing class with XAI vs. without (>1 → XAI benefit)

Y-axis: avg. number of classes per image in archive

Across 476 subjects, with-XAI found 2.7% more true positives (8356 vs. 8134).

In difficult cases, when the object is small and the scene is complex, 6.5% more.

XAI saliency measures



XAI degrades overall Precision

XAI improves Precision on more difficult problems

Full Study Results: Qualitative

Questions were on a 6-point Likert scale, 1=”strongly disagree”, 6=”strongly agree”; some observations:

• **XAI helps give feedback.** 60% agreed at some level that ”Overall, I feel saliency maps helped me give better feedback.”

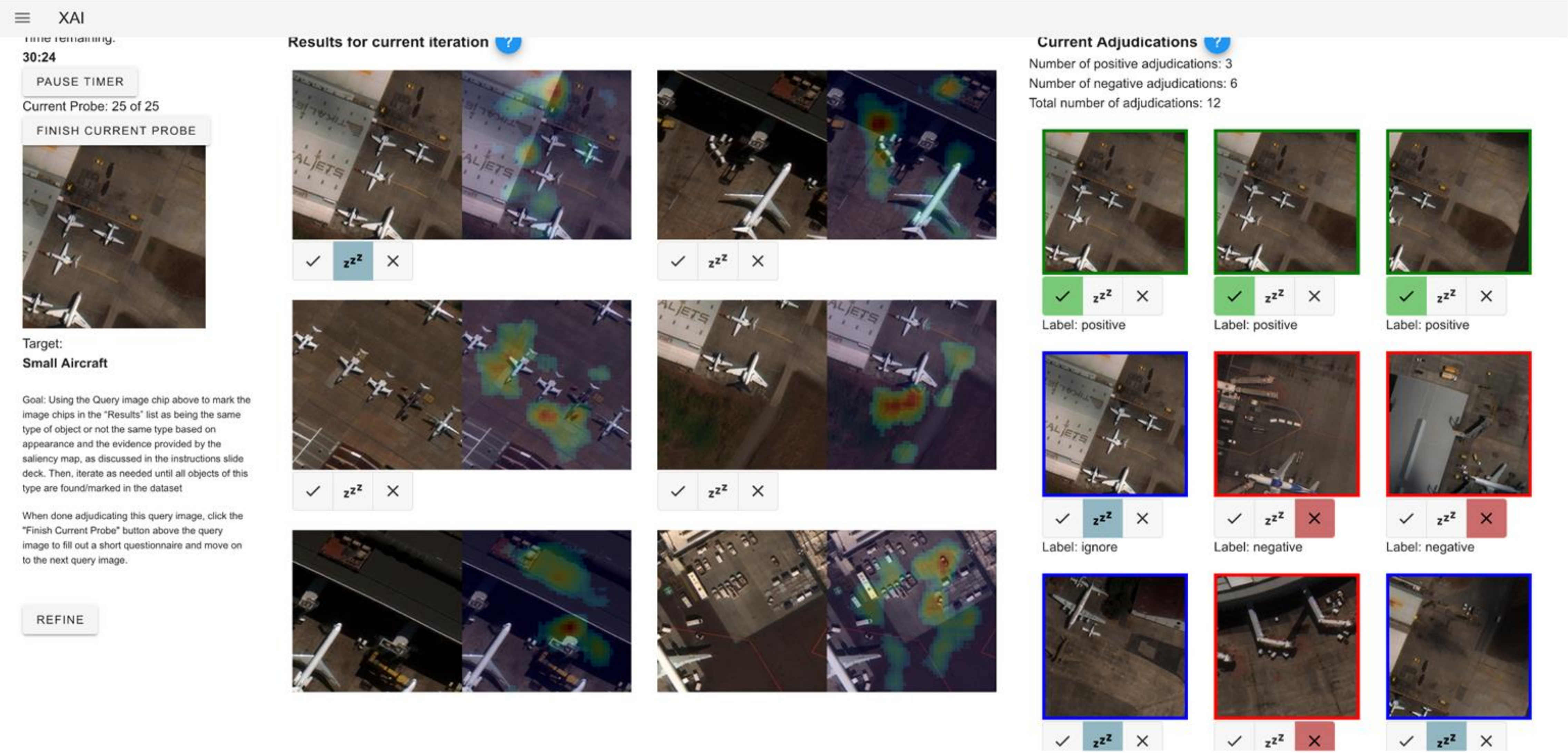
• **XAI helps understanding.** 83% agreed at some level, with 56% agreeing or strongly agreeing, that ”Saliency map helped me understand how the system ”thinks”.

• **XAI improves ease-of-use.** 62% agreed at some level, with 38% agreeing or strongly agreeing, that ”Saliency maps made the system easier to use.”

• **No clear signal on preference for saliency maps.** 58% agreed at some level that they would ”prefer to do [the task] with saliency maps rather than without”; however, 60% also agreed at some level with the opposite question that they would ”prefer to do [the task] without saliency maps rather than with.”

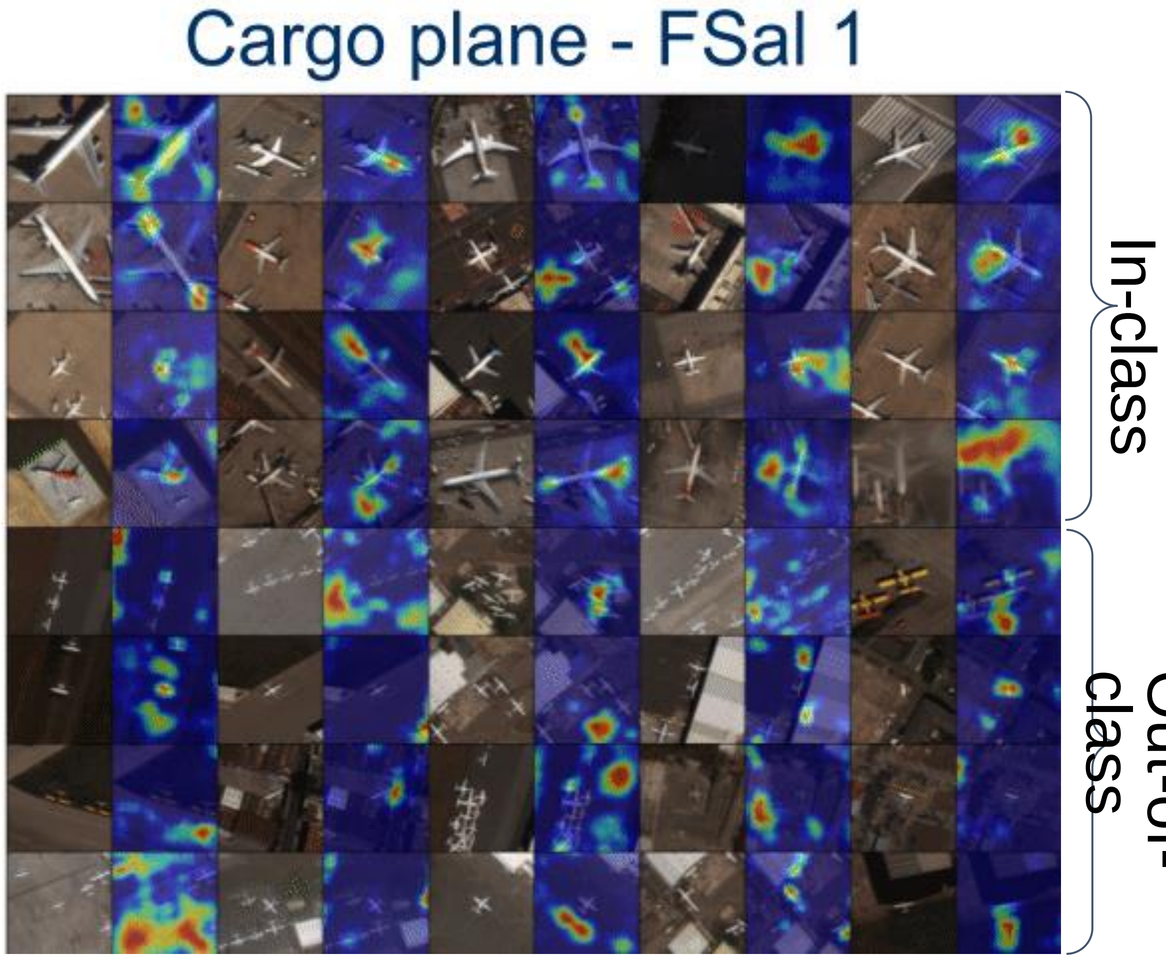
• **Responder confidence.** 95% agreed or strongly agreed that they understood the questions.

XAI IQR for User Study on xView



XAI Summary

- Saliency explanations can improve human-in-the-loop image retrieval when the problem is harder
  - Small objects, cluttered scenes
  - Highlights cases where the AI was right for the wrong reasons (e.g. matching the background)
- XAI helps users feel more confident in using the system
- Critical to identify the need for explanations within a given application
  - Easier cases may not need explanations
- Working with the JAIC and others to transition saliency for AI T&E and V&V
  - Aids in comparison of multiple black-box AI algorithms
- Saliency addresses Traceability and Reliability



Topics

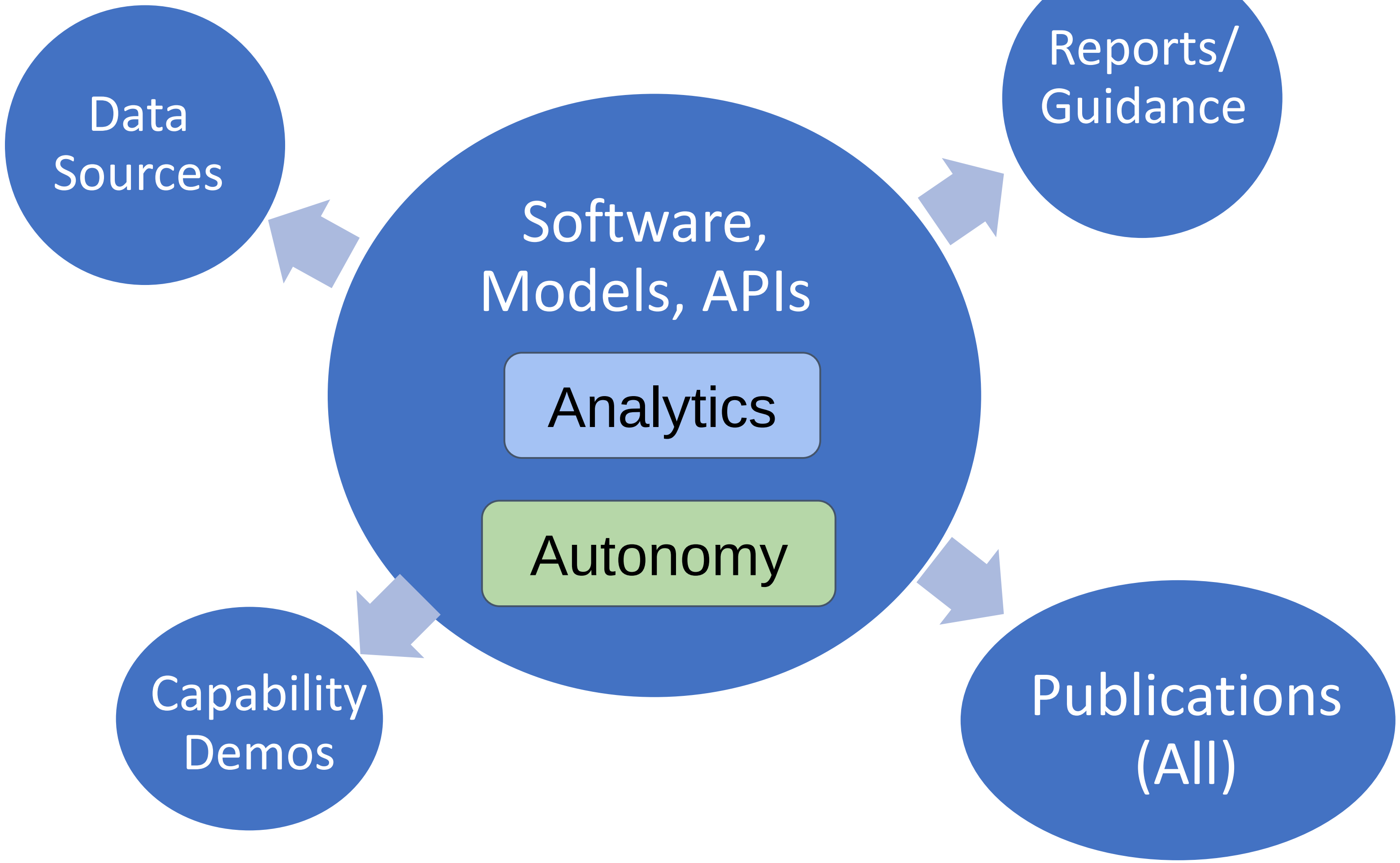
- Do-It-Yourself AI in practice
- Explainable AI for interactive search
- The XAI Toolkit

XAI-Toolkit Overview

https://xaitk.org

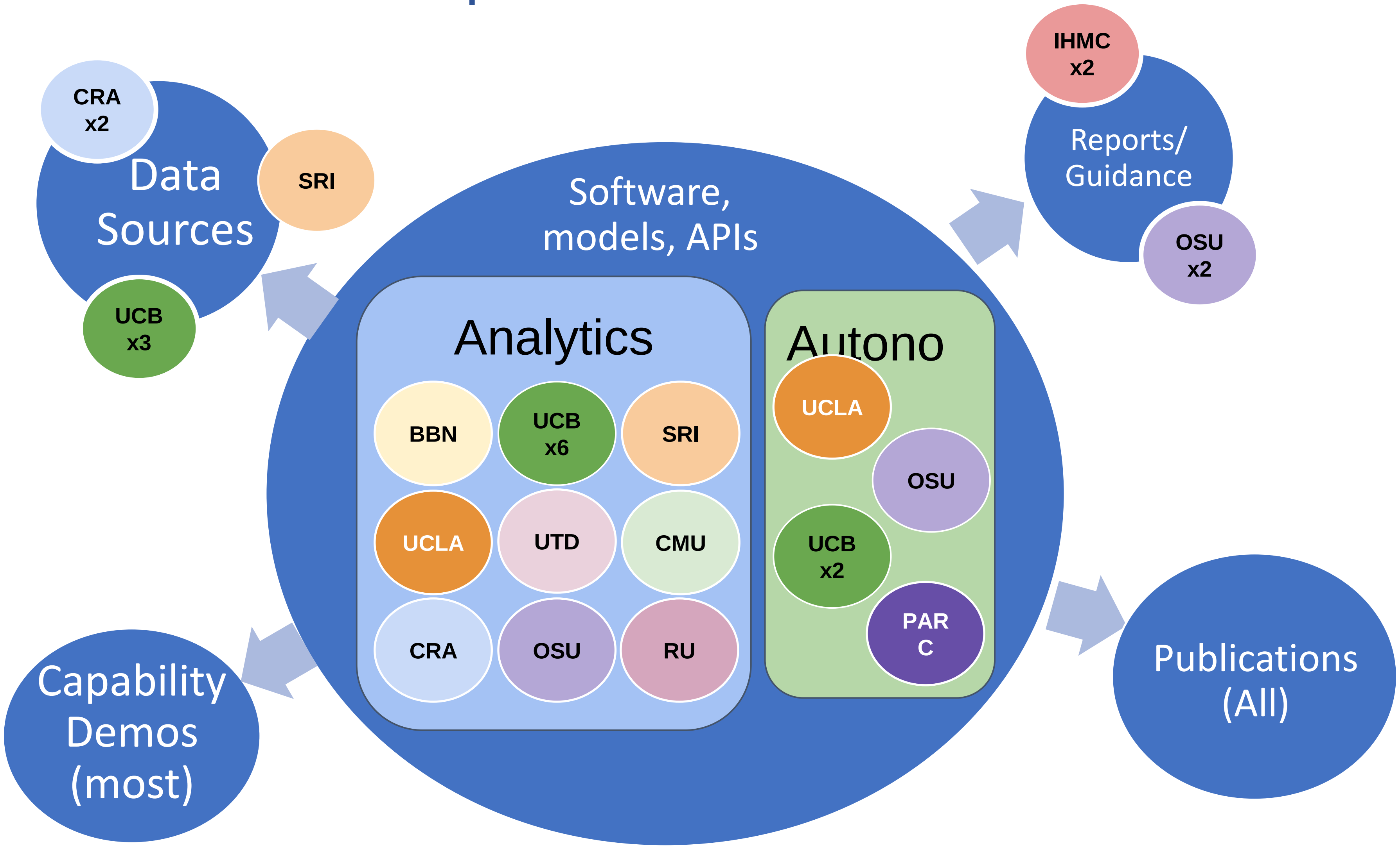
Objectives:

1. Create a single location for all XAI contributions
2. Create ontology from contributions
3. Create common software frameworks where possible
4. Easy to follow guide for what to use for a task
5. Make the XAI-TK publicly available and help transition



Supported by DARPA XAI program

XAI-Toolkit Components



Performer Team

- UCB — University of California Berkeley
- SRI — SRI International
- CRA — Charles River Analytics
- OSU — Oregon State University
- IHMC — Institute for Human & Machine Cognition
- UCLA — University of California Los Angeles
- UTD — University of Texas at Dallas
- PARC — PARC a Xerox Company
- CMU — Carnegie Mellon University
- RU — Rutgers University
- BBN — BBN/Raytheon

XAI Performer Saliency Explanations

Performer Team

UCB — University of California Berkeley

CRA — Charles River Analytics

OSU — Oregon State University

UCLA — University of California Los Angeles

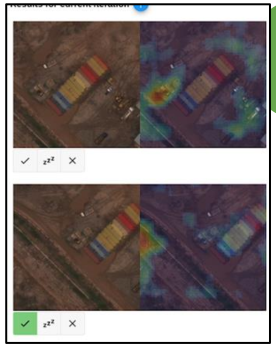
CMU — Carnegie Mellon University

Object Detection
(Bounding Box Feedback)

UCB

Image Classification
(Trinary Feedback)

UCB



UCLA

Image Classification
Fault-Line Detection

CRA

Image Classification
Distilling CNNs

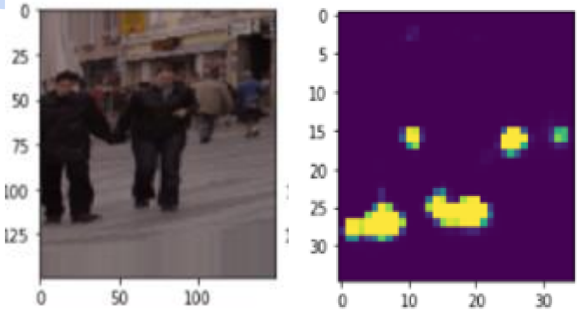


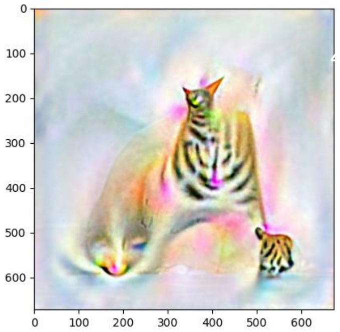
Image Classification
(NLP Fine-Grained)

UCB

This is a *Western Grebe* because it has a **long white neck**, **pointy yellow beak** and **red eye**.

CMU

Image Classification
Visualize Learned Classes



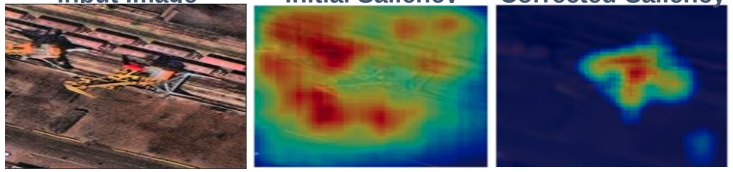
OSU

Image Classification
IGOS++ Saliency Maps

UCB

Image Classification
(Paintbrush Feedback)

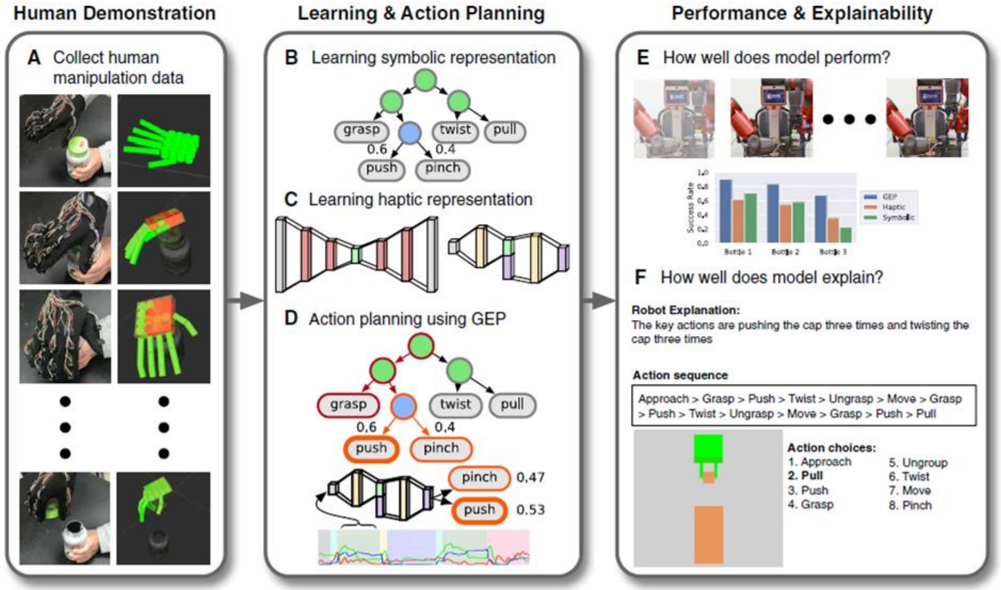
Input Image | Initial Saliency | Corrected Saliency



Autonomy: Performer Approaches

UCLA

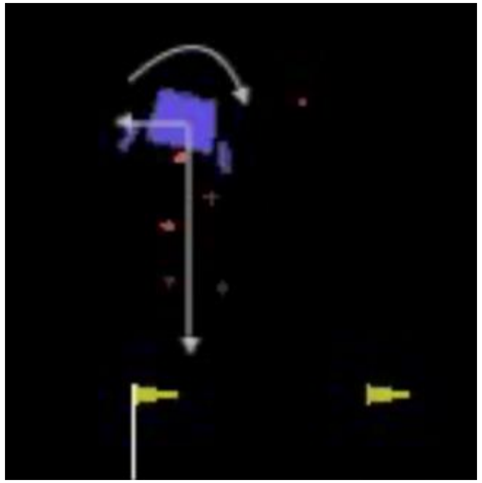
Explainable Robot Behavior

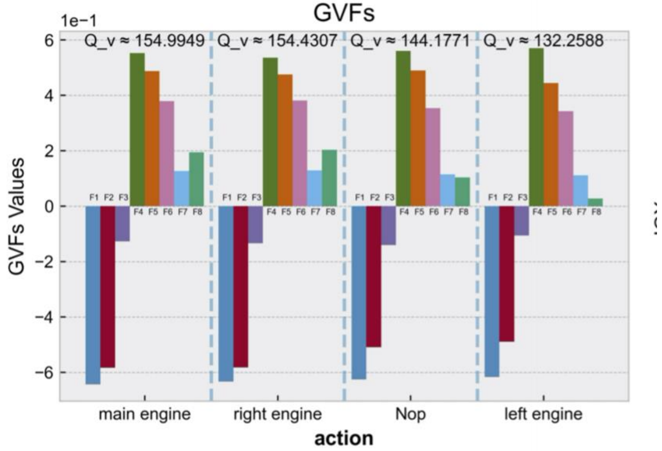


OSU

Model-Free Reinforcement Learning

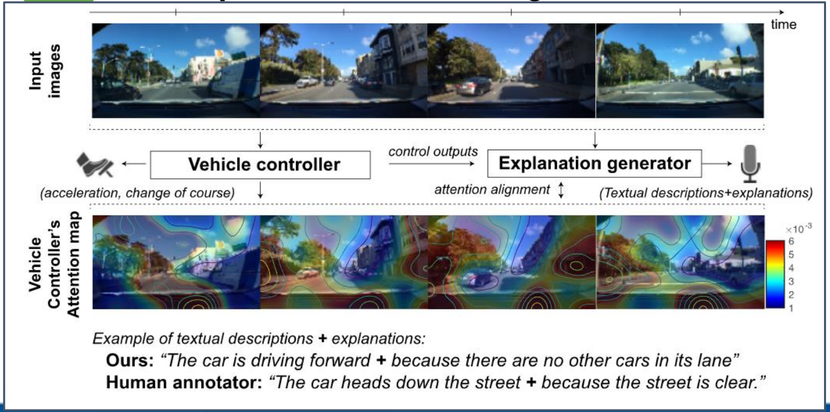
(Lunar Landing- Engine Burn: Discrete Action Domains)





UCB

Explainable Self-Driving Controller



UCB

Advisable Self-Driving Controller

Human: The car is steadily driving + now that the cars are moving

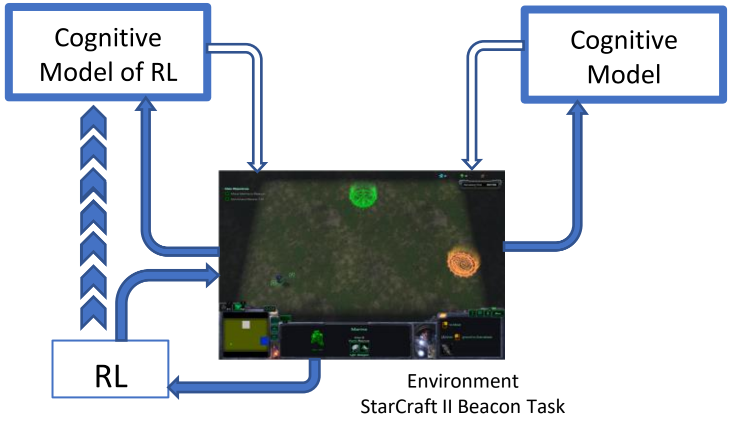
Description **Explanation**

Action ← **Condition**

PAR C

Image Classification

Cognitive Saliency

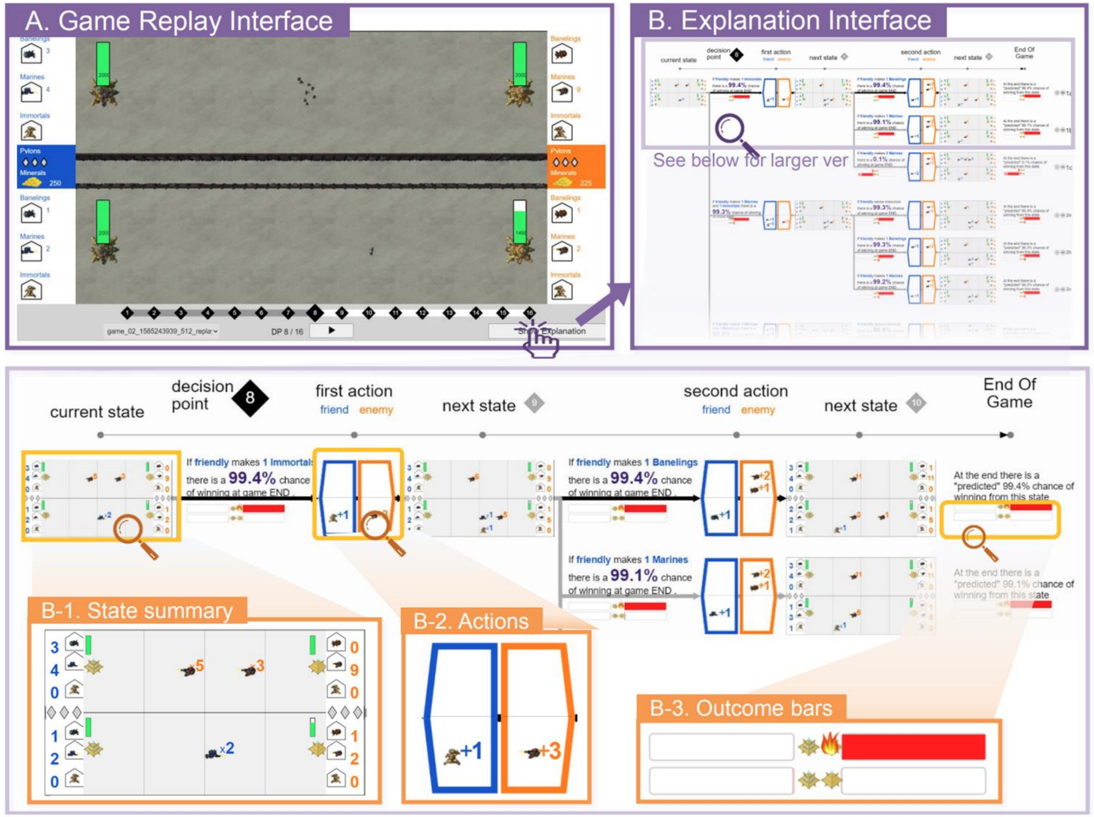


After Action Review for AI

- **AARfAI**: a structured series of steps to help users assess AI agents and formalize their understanding about these agents

- Focuses on **model-based explanations**, i.e. uses an explanation search tree at decision points to reveal patterns in AI's reasoning

- Designed to be a **model debugging** tool for AI agents at multiple levels, highlighted by two example use cases:
  - Finding bad decision points that cause loss of a game
  - Finding reasoning bugs using game-wide summaries of information



Demo Video

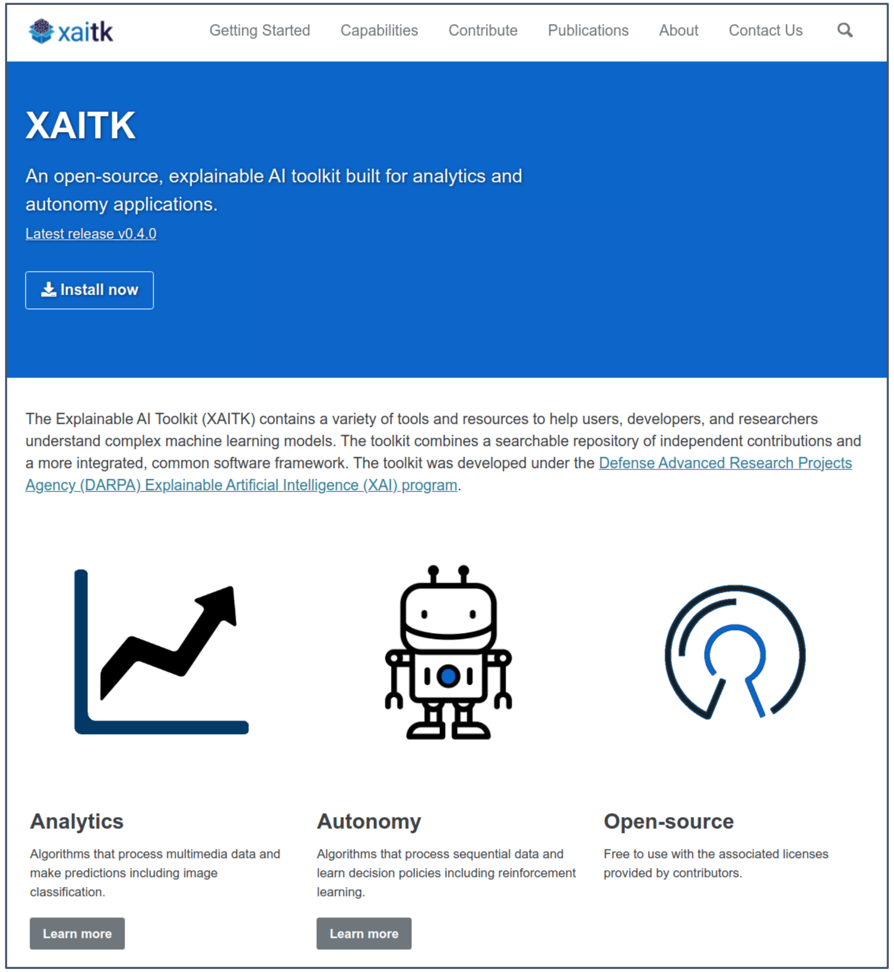
Current Status

- 1.5-year project, 3/31/2021 - 9/30/2022
- Website at https://xaitk.org
- *xaitk-saliency* package (https://github.com/xaitk/xaitk-saliency)
  - Set of black-box, occlusion-based saliency methods
- Manuscript on XAITK in Applied AI Letters
  - https://onlinelibrary.wiley.com/doi/10.1002/ail2.40

xaitk

XAITK Website

https://xaitk.org



xaitk

Getting Started | Capabilities | Contribute | Publications | About | Contact Us

XAITK

An open-source, explainable AI toolkit built for analytics and autonomy applications.

Latest release v0.4.0

Install now

The Explainable AI Toolkit (XAITK) contains a variety of tools and resources to help users, developers, and researchers understand complex machine learning models. The toolkit combines a searchable repository of independent contributions and a more integrated, common software framework. The toolkit was developed under the Defense Advanced Research Projects Agency (DARPA) Explainable Artificial Intelligence (XAI) program.

Analytics

Algorithms that process multimedia data and make predictions including image classification.

Learn more

Autonomy

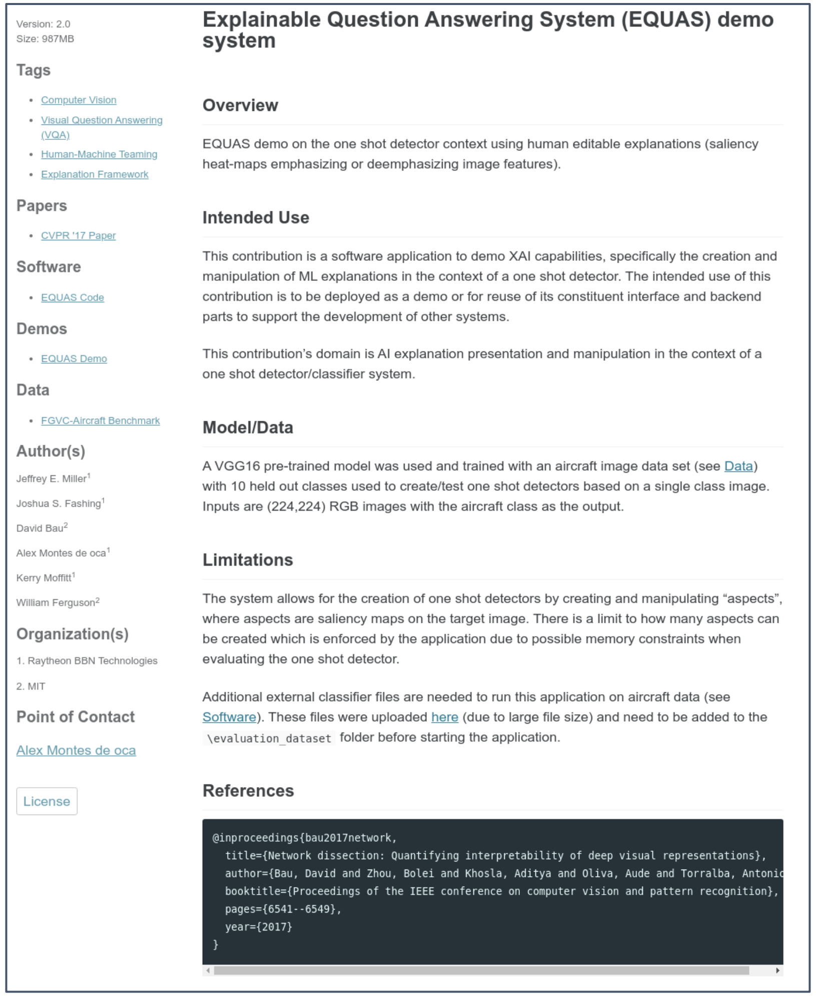
Algorithms that process sequential data and learn decision policies including reinforcement learning.

Learn more

Open-source

Free to use with the associated licenses provided by contributors.

Landing Page



Version: 2.0
Size: 987MB

Tags

- Computer Vision
- Visual Question Answering (VQA)
- Human-Machine Teaming
- Explanation Framework

Papers

- CVPR '17 Paper

Software

- EQUAS Code

Demos

- EQUAS Demo

Data

- FGVC-Aircraft Benchmark

Author(s)

Jeffrey E. Miller[1]

Joshua S. Fashing[1]

David Bau[2]

Alex Montes de oca[1]

Kerry Moffitt[1]

William Ferguson[2]

Organization(s)

1. Raytheon BBN Technologies
2. MIT

Point of Contact

Alex Montes de oca

License

Explainable Question Answering System (EQUAS) demo system

Overview

EQUAS demo on the one shot detector context using human editable explanations (saliency heat-maps emphasizing or deemphasizing image features).

Intended Use

This contribution is a software application to demo XAI capabilities, specifically the creation and manipulation of ML explanations in the context of a one shot detector. The intended use of this contribution is to be deployed as a demo or for reuse of its constituent interface and backend parts to support the development of other systems.

This contribution's domain is AI explanation presentation and manipulation in the context of a one shot detector/classifier system.

Model/Data

A VGG16 pre-trained model was used and trained with an aircraft image data set (see Data) with 10 held out classes used to create/test one shot detectors based on a single class image. Inputs are (224,224) RGB images with the aircraft class as the output.

Limitations

The system allows for the creation of one shot detectors by creating and manipulating "aspects", where aspects are saliency maps on the target image. There is a limit to how many aspects can be created which is enforced by the application due to possible memory constraints when evaluating the one shot detector.

Additional external classifier files are needed to run this application on aircraft data (see Software). These files were uploaded here (due to large file size) and need to be added to the `\evaluation_dataset` folder before starting the application.

References

```
@inproceedings{bau2017network,
  title={Network dissection: Quantifying interpretability of deep visual representations},
  author={Bau, David and Zhou, Bolei and Khosla, Aditya and Oliva, Aude and Torralba, Antonio
  booktitle={Proceedings of the IEEE conference on computer vision and pattern recognition},
  pages={6541--6549},
  year={2017}
}
```

Example Contribution

xaitk-saliency package

- Support for image classification, object detection, and image retrieval
- Modular design and easily extendable to support new algorithms
- Compatible with Pytorch and other deep learning/machine learning frameworks

https://github.com/XAITK/xaitk-saliency

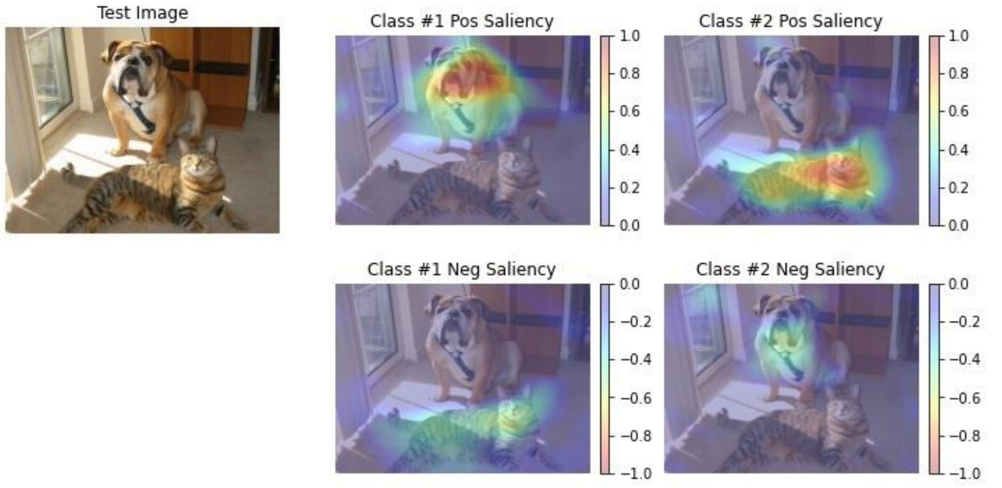
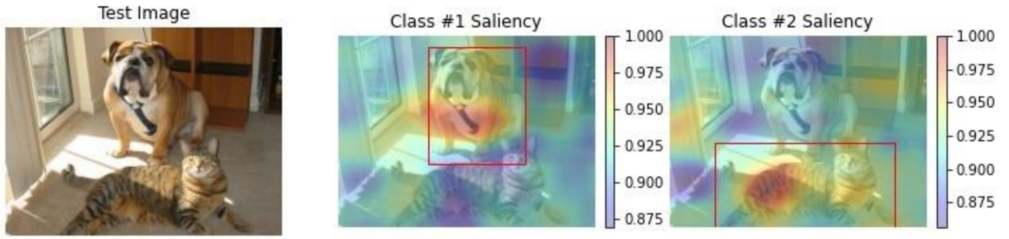


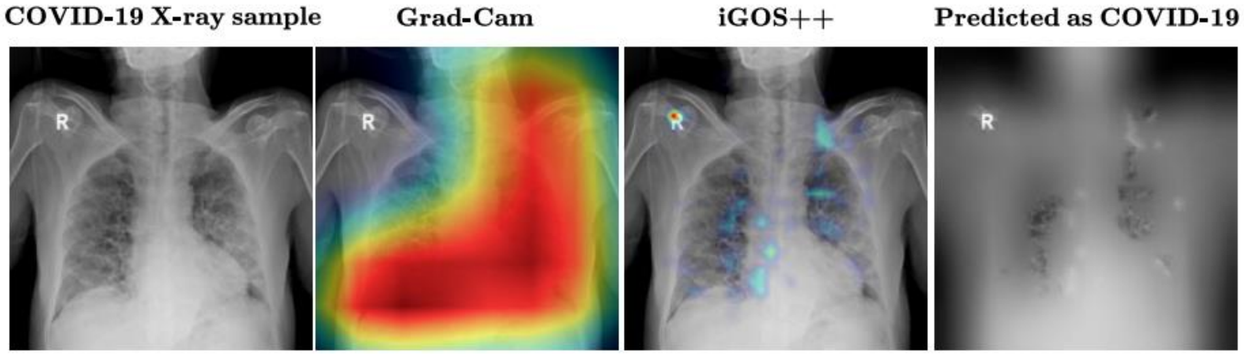
Image Classification



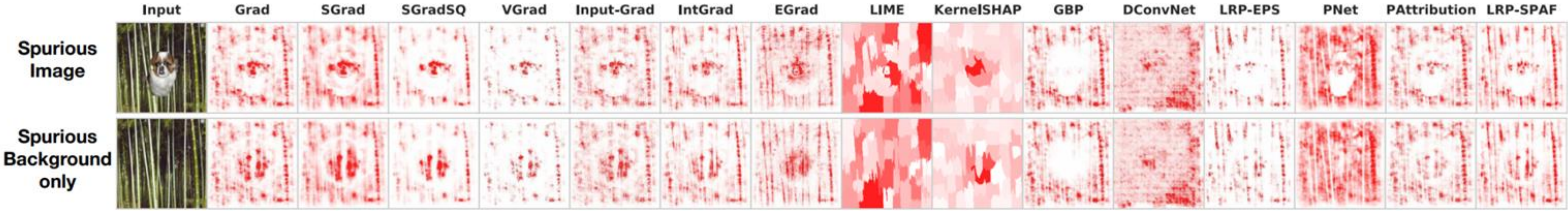
Object Detection

Saliency Maps for AI Model Debugging

Detector Data Poisoning
https://arxiv.org/pdf/2006.03204.pdf

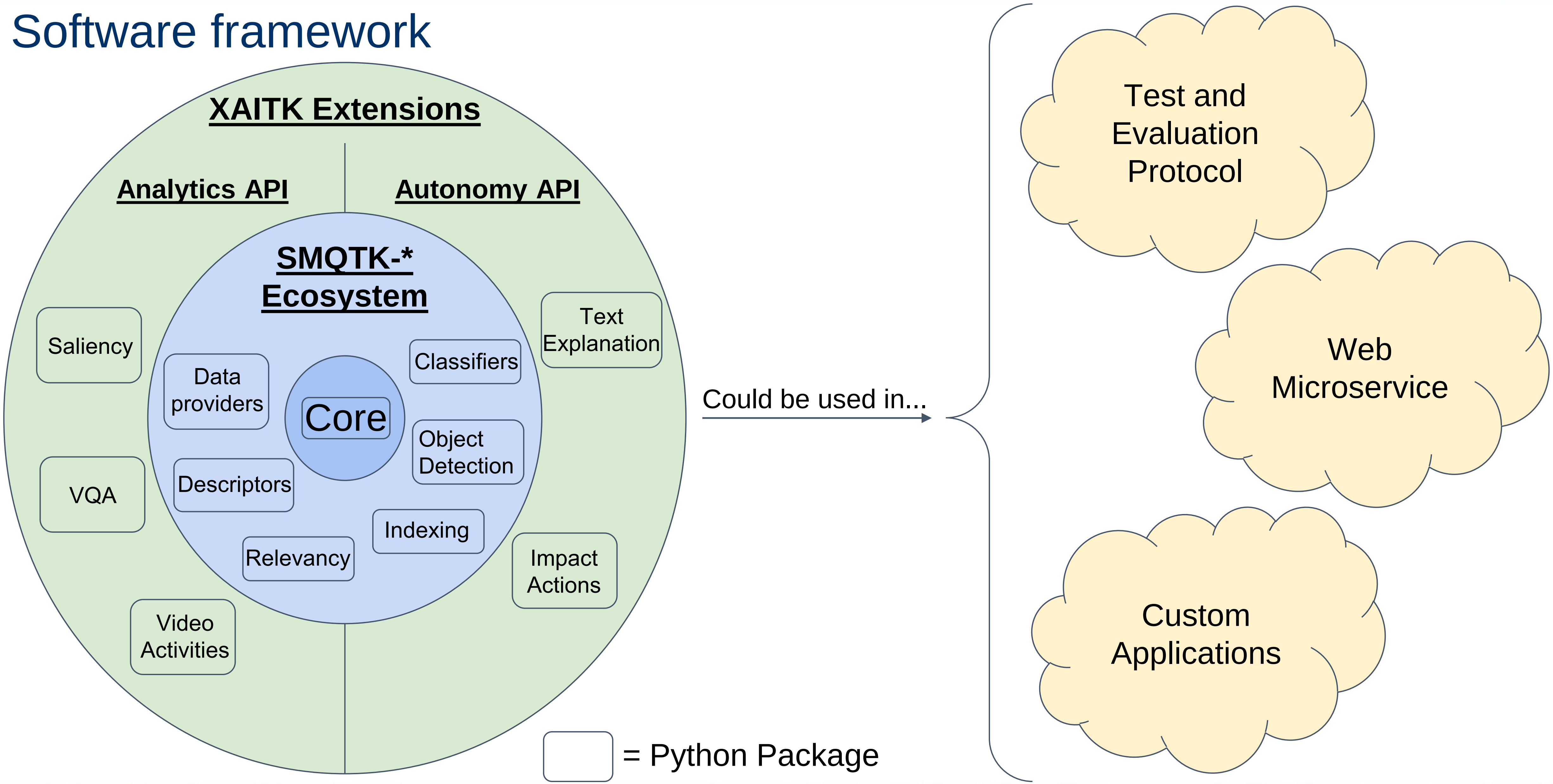


COVID-19 classification
https://arxiv.org/pdf/2012.15783.pdf

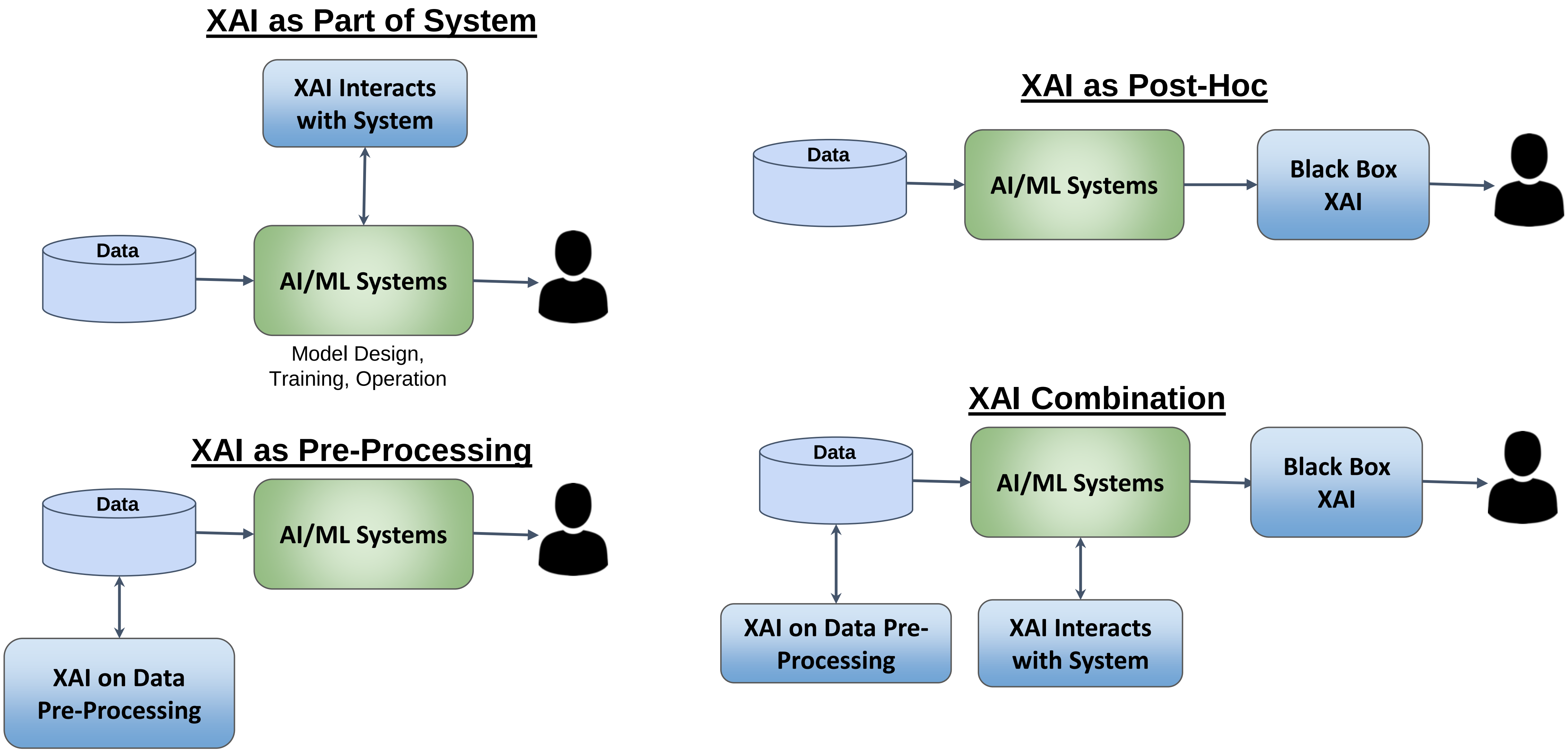


Spurious background correlation, https://arxiv.org/pdf/2011.05429.pdf

Software framework



Levels of XAI-Toolkit Integration



Conclusions

- Open-source DIYAI toolkit, VIAME, enables end-users to create cutting-edge AI analytics with no programming or ML experience
    - Initially for maritime but now applied to many domains
- Explainable AI can improve user performance for difficult problems in interactive domains
- XAITK is an open-source toolkit for explainable AI and visual saliency maps