A Dataset of Head and Eye Movements for 360° Videos

Erwan David[1], Jesús Gutiérrez[1], Antoine Coutrot[1], Matthieu Perreira Da Silva[1], Patrick Le Callet[1]

[1]LS2N UMR CNRS 6004, Université de Nantes, France

erwan.david@univ-nantes.fr

Introduction

- We created a new dataset of head and eye movements in Virtual Reality. Participants explored videos (non-stereoscopic) wearing a HTC-Vive with embedded eyetracking. We released saliency videos and scanpaths computed from raw data, along with a toolbox to compute similarity measures.

- Our dataset will help the community in the development of new theoretical and applied research such as foveated rendering, visual attention modeling, data compression, viewport adaptive streaming, etc.

Omnidirectional content (360°)

Sphere space Virtual sphere on which content is projected

Viewport Images displayed in the headset

Left eye Right eye

Data gathering

Material and method

- 57 observers (25 women; mean age: 25.7 years)
- 19 videos (duration: 20 sec.)
- HTC-Vive (90fps) + SMI Eyetracker (250Hz)
- Task: free-viewing, sitting on a rolling-chair
- experimental condition: start exploration either at longitude 0° or 180°

Statistics

- 5 video categories (Indoor, Outdoor, Urban, Rural, People)
- video framerate: 24 to 30 fps
- 380 sec. of saliency videos
- > 55,000 fixations

Dataset structure

ROOT
- **Stimuli**
  - video files
- **HE** (Head+Eye)
  - **SalMaps**
    - binary saliency videos
  - **Scanpaths**
    - **L**
      - fixation lists
    - **R**
      - fixation lists
- **H** (Head-only)
  - **SalMaps**
    - binary saliency videos
  - **Scanpaths**
    - head trajectories
- **Tools**
  - saliencyMeasures.py
  - scanpathMeasure.py
  - readBinarySalmap.py

Toolbox

Similarity measures: compare together...

- saliency videos → AUC, CC, KLD, NSS, SIM
- scanpaths → MultiMatch

Particularities: dynamic

- Scanpaths alignement → temporal de facto
- Pool saliency frames together by 200ms windows

Particularities: omnidirectional

- Correcting saliency frames for equirectangular map distortion: weighting maps with a sine function according to the latitude ($\sin y$ for $y \in [0, \pi]$)
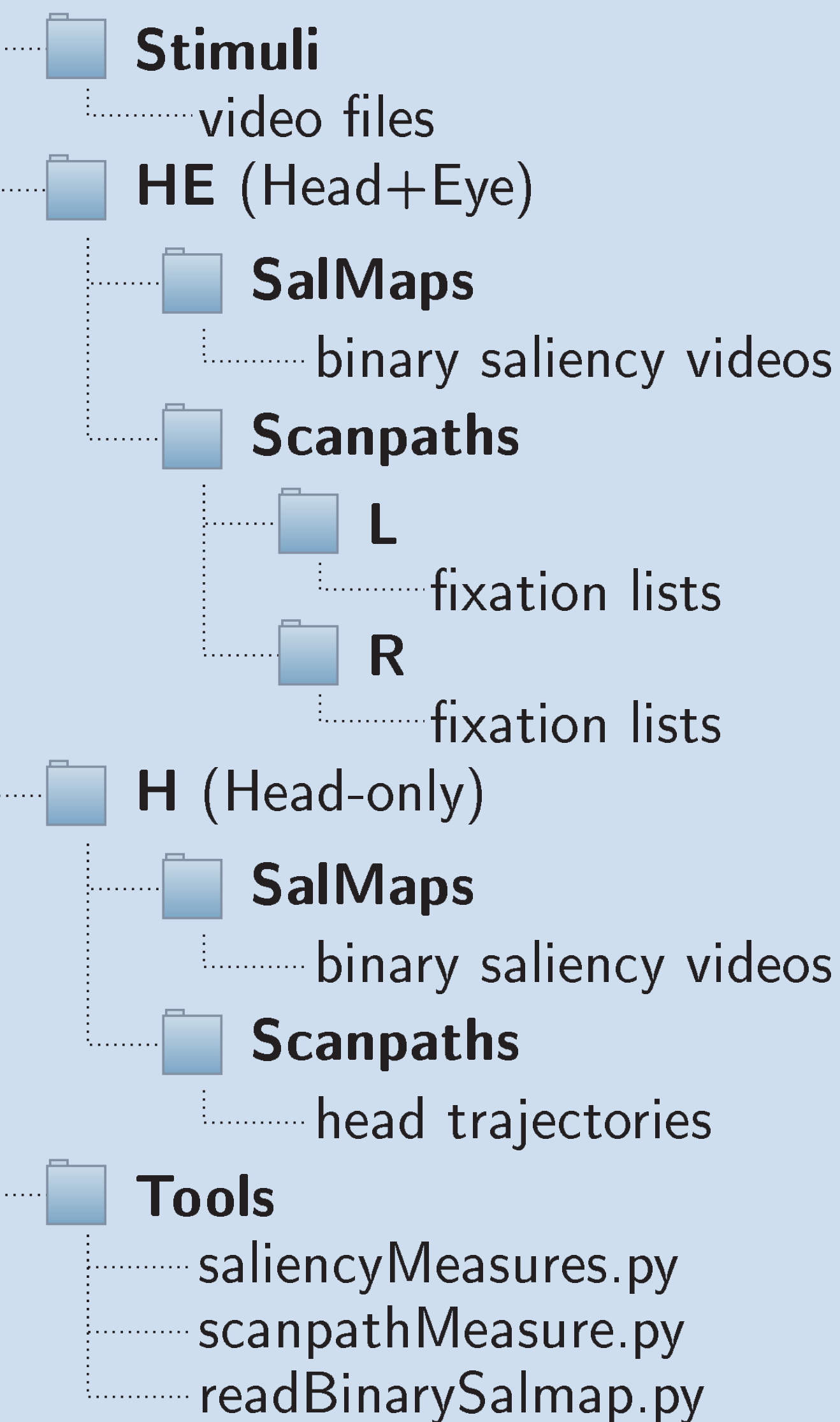
Gaze processing

- Viewport gaze position (2D) to Unit vector (gaze samples on unit sphere, 3D)
- Unit gaze samples to fixations and saccades (velocity algorithm, threshold: 80°/sec.)
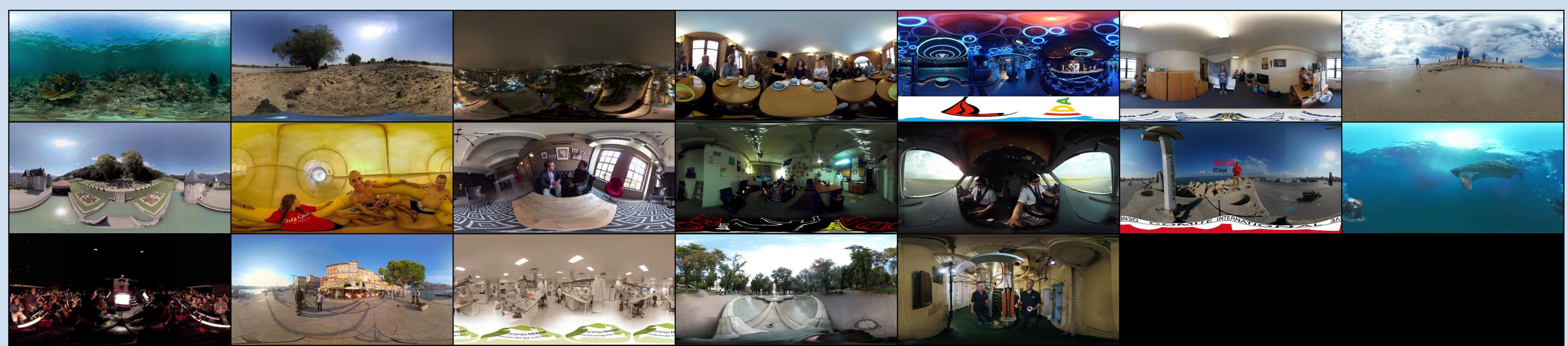
Head movement processing

- Head movements can be categorized as "fixations" (periods of stability, below ~25°/sec)
- With no gaze information we cannot make any assumptions about an observer's perception during head movements
- Solution: extract "head trajectories" sampled over 200ms windows (100 points per 20sec. stimulus)

Stimuli – frame examples

Saliency maps – frames (videos are shown on tablets nearby)
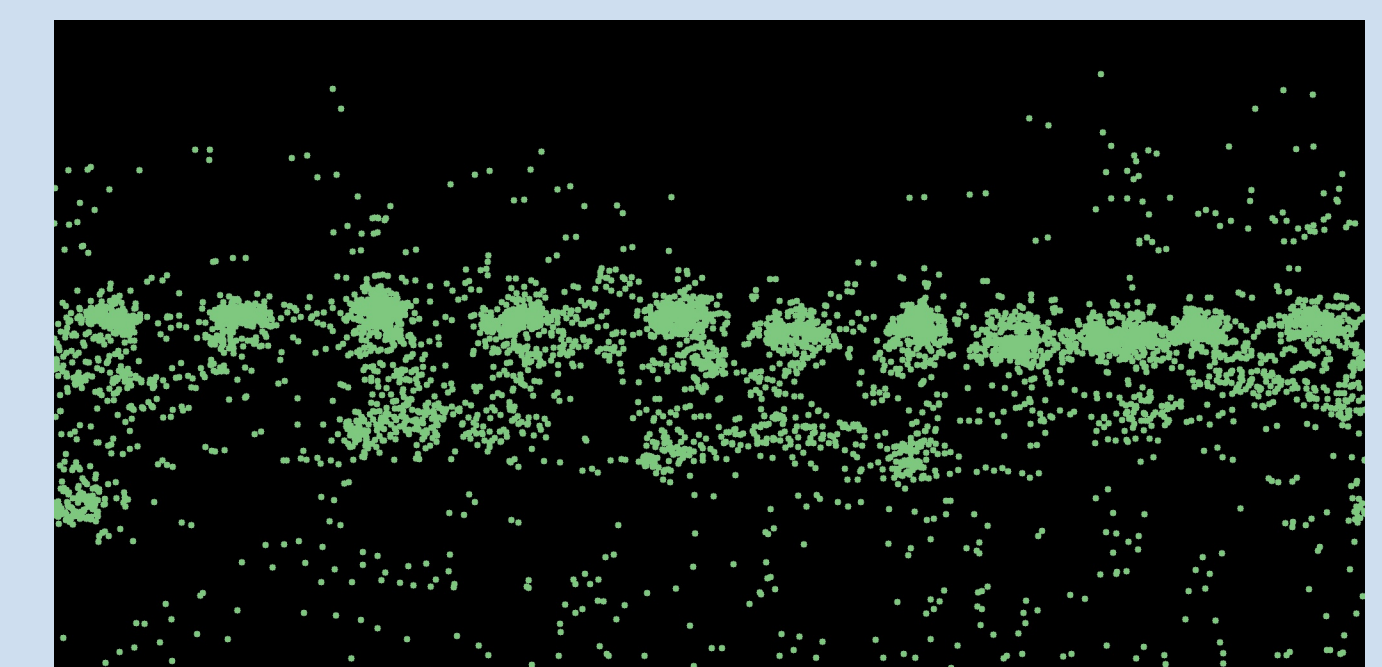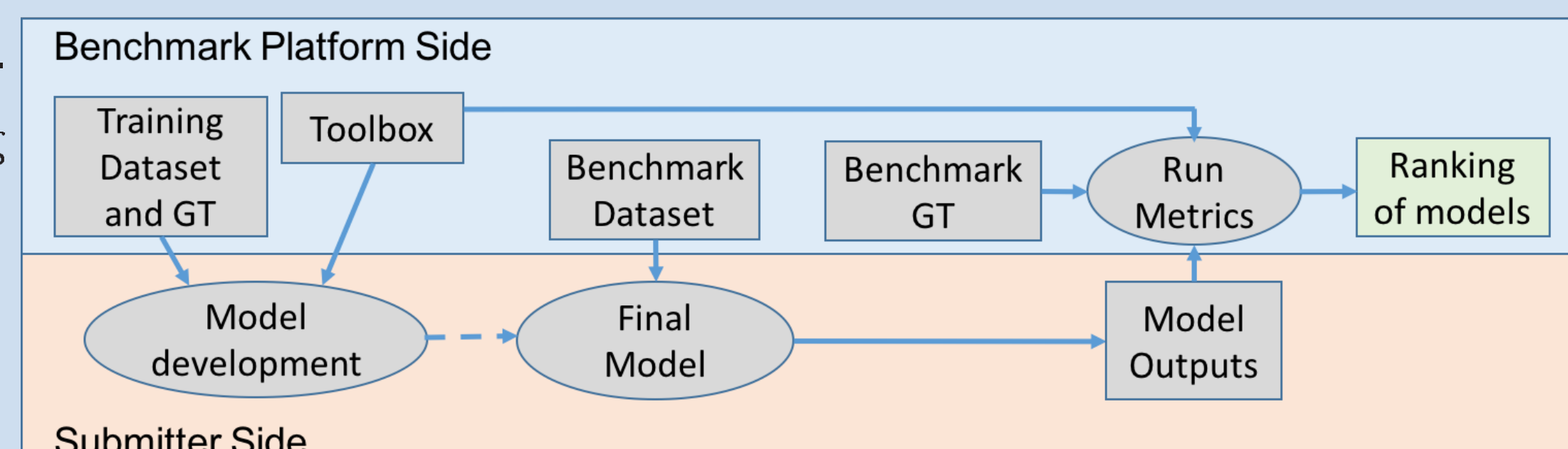
Scene   Head+Eye   Head-only

- A saliency map per frame → saliency videos (as many frames as the original stimuli)
- Saliency frames → unnormalized (free to normalize over individual frames or after pooling frames)

Scanpaths

Scanpaths are sequences of fixations, they retain temporal and individual information.

Fixation features:

- Fixation index
- Longitude
- Latitude
- Start timestamp
- Fixation duration
- Start frame
- End frame

Link

This dataset is part of the **UN Salient360! Benchmark**. A new benchmark for 360° attention modeling (image and video).

Come play with our dataset: **salient360.ls2n.fr/datasets/**

Benchmark Platform Side

Training Dataset and GT | Toolbox | Benchmark Dataset | Benchmark GT | Run Metrics | Ranking of models

Model development | Final Model | Model Outputs

Submitter Side

Acknowledgements