### Stats

* VRT environments most prevalent (31/73), with 52/73 (VRT+BLND) using at least some virtual component.
* 7:2 learning to training ratio
* STEM focus (55/73)
* 45:31 IND : MULT
* Most students are UNI (36) or K12 (30)
* Video present in most papers (61) and audio in roughy half (37). Logs and PPA also well-represented (33,30)
  + 59 out of the 73 total papers using some form of vision analysis
* 20 papers use sensors
* 45 (approximately 61.6\%) incorporate at least one of the human-centered modalities (QUAL, INTER, SURVEY, RPA, or PPA).
* Among these papers, an overwhelming majority—44 out of 45—incorporate more than one human-centered modality in their analyses
* 30 papers (40\%) from the corpus incorporated log data in their analysis
* Of the 73 papers, 54 perform early, mid, late, or hybrid fusion.
* mid fusion is the most prevalent, employed in 27 papers (36.99\%)
* Out of our paper set, 46 papers (63\%) use model-based methods, 16 papers (22\%) employ model-free methods, and 11 papers (15\%) use a combination of both.
* This distribution, with 78\% of papers favoring model-based analysis, indicates a strong preference in the MMLA community for developing models to explain the learning process.

### Results

* Human-centered:
  + researchers gained a more comprehensive understanding of the learning environment and the learning activities.
  + This human-centered approach offers insights into the participants' experiences, perceptions, and behaviors, often pinpointing subtle nuances that might be missed in a unimodal analysis.
  + QUAL provides rich contextual observations, PPA and RPA offer tangible artifacts, INTER captures in-depth discussions, and SURVEY provides multiple participant perspectives, collectively enriching the analysis.
  + The use of multiple modalities allows for triangulation and cross-verification, where findings from different sources are compared to enhance the validity of the results.

### SOTA

### Challenges

* Human-centered:
  + subjectivity, scalability, resource intensiveness, and potential limitations in generalizability
  + Manual collection and human analysis can be time-consuming and may not scale well, especially in large-scale educational settings
* Logs:
  + issues with time, limited data size, generalizability, and engineering expenses.
  + Temporal aspects
    - alignment,
    - Sampling rates,
  + Dataset scarcity and size
  + Generalizable findings
  + Time- and monetary cost

### Research Gaps

* Training envs, esp. physical ones like rehab therapy and athletic training, are not well-represented
  + Psychomotor gap (only 5 papers in corpus)
  + INST : TRAIN = 3:1 (45:31)
  + Sensor data present in less than ⅓ of papers (20)
* Professional development also lacking (only 5)
* Lack of text-based data (only 1 paper)
* least used modality: raw pixel data
  + Limitation of derived, observable features is their evaluation. Why do we trust them innately?
* Lack of mixed method (qual+quant) (20/73)
* Data visualization, especially interactive data viz
* XAI
* Temporality (logs): concentrate on overall learning outcomes within an assignment, but often overlooks the nuanced aspects of how student behaviors, emotions, and achievements evolve over time.
* Standardization:
  + standardized log format and consistent practices
  + absence of common coding conventions (human centered)
  + low adoption rate of established industry standards like xAPI, LTI, and Learning Management Systems (LMS) within educational technology.
  + This trend reflects a broader issue within the field to align with best practices and norms that have been established in the wider technology and education sectors.
* a crucial gap lies in the automation of human-coding processes (human centered)