






CHALLENGE CALL & ERA-NET

REVIEW OF CHALLENGE INITIATIVES

Guillaume Avrin – Head of the « AI Evaluation » Department

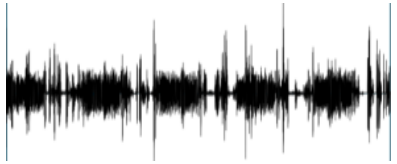
LABORATOIRE NATIONAL DE MÉTROLOGIE ET D'ESSAIS

11/23/20

LNE EXPERTISE IN EVALUATION OF AI SYSTEMS

SPEECH AND TEXT

Transcription, Keyword spotting, Speaker comparison, Named entities recognition, Translation, etc.

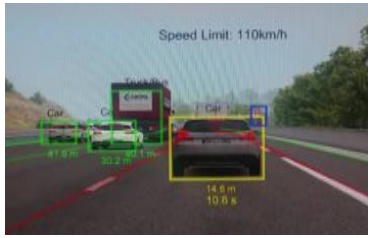


ترحيب ، يسعدنا أن نرحب بكم

Bienvenue, nous sommes ravis de vous accueillir

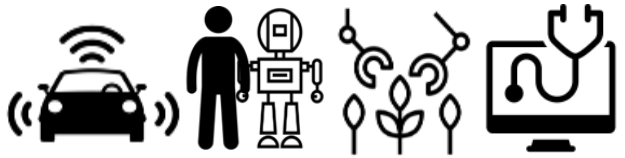
IMAGE

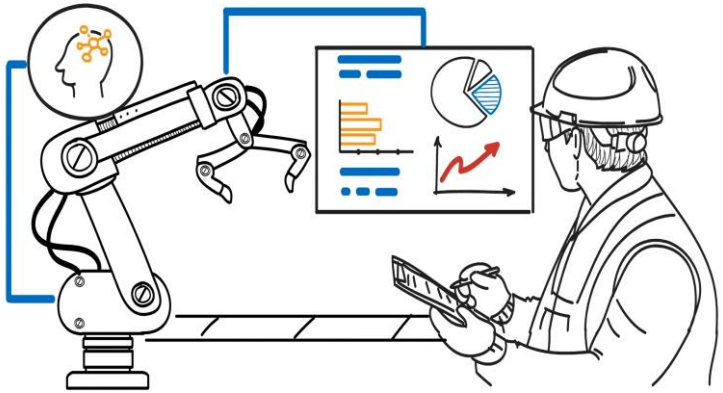
Object recognition, head tracking, OCR, etc.



MULTIMODAL

ROBOTICS





LNE: more than 10 years of experience in AI evaluation and more than 950 systems evaluated

Challenge	N° of participants
ETAPE	16
MAURDOR	6
TRAD	6
REPERE	16
QUAERO	32
IMM	14
ROSE	4
METRICS	12+

WHAT IS A CHALLENGE?

Evaluation campaigns to benchmark the performance of competing technologies, whose metrological rigor, collective emulation and knock-on effect generate progress in the field.

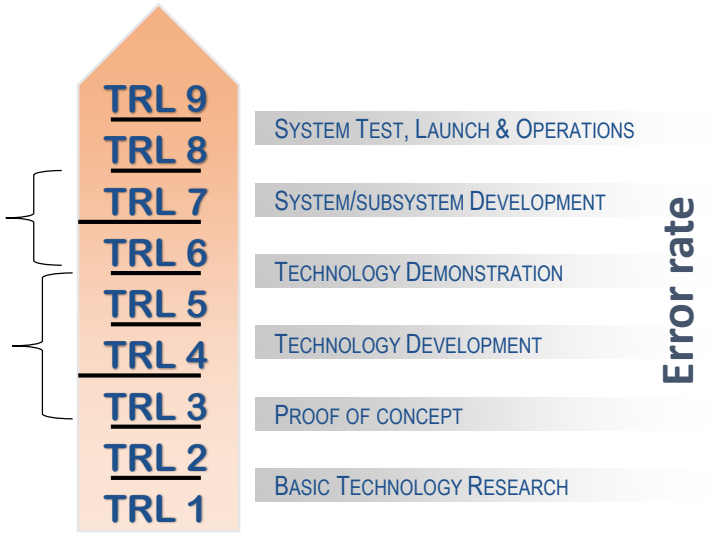
Repeated campaigns over time to assess progress:

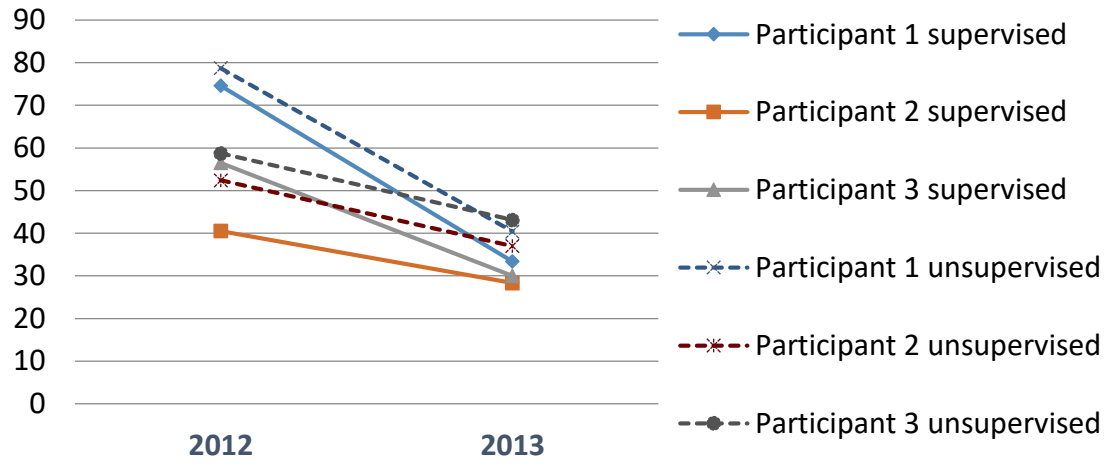
Preparation of the evaluation tools → Dry Run → 1st Campaign → 2nd Campaign

R&D funding tool to bridge the TRL death valley (with monitored R&D efforts and targeted participant profiles):

CLOSED CHALLENGES
The sponsor select the participants

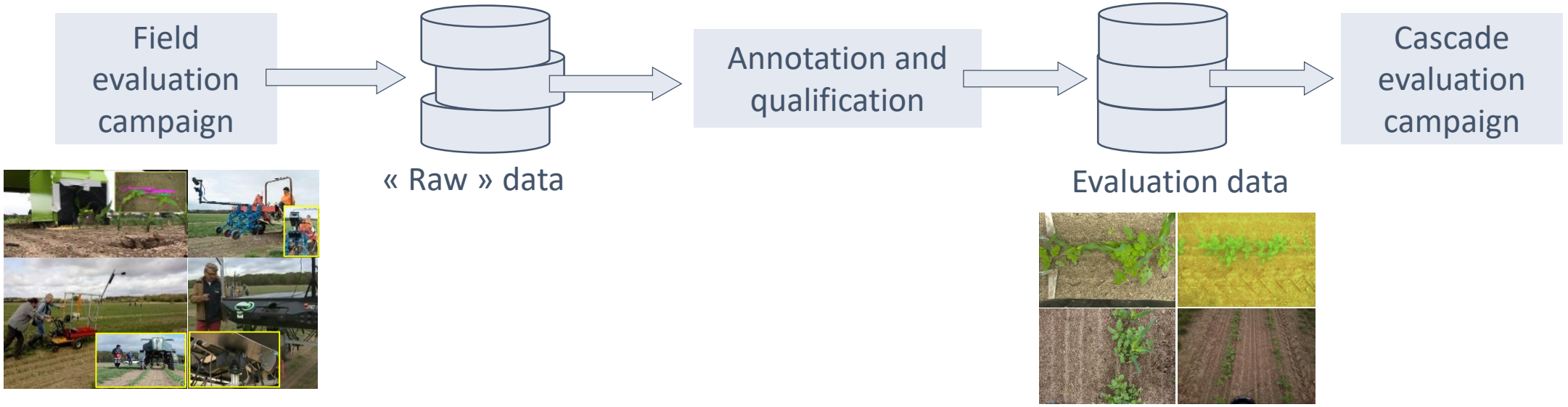
OPEN CHALLENGES
Everybody can participate





A CHALLENGE NEEDS HIGH QUALITY TESTING DATASETS

1) A challenge needs high quality datasets to attract participants
2) One option: the dataset used to evaluate the AI algorithms of the robots during the cascade evaluation can be the one which is collected during the filed evaluation.



ROSE (AFB, ANR, 2017-2021)	METRICS (H2020, 2020-2022)
Weeding robots	Healthcare, agile production, inspection and maintenace, agrifood

A CHALLENGE NEEDS EFFICIENT COMMUNICATION TOOLS

Connexion with general public

- Through demonstration

Social media

- Youtube channel
- Twitter
- Medium
- Facebook
- Etc.

Robocup (1997 onwards)	Robotex international (2001 onwards)
Soccer, Industrial, Rescue, Home	Lego sumo, boat navigation, animal rescue, etc.

A CHALLENGE NEEDS TRUSTWORTHY ORGANIZERS

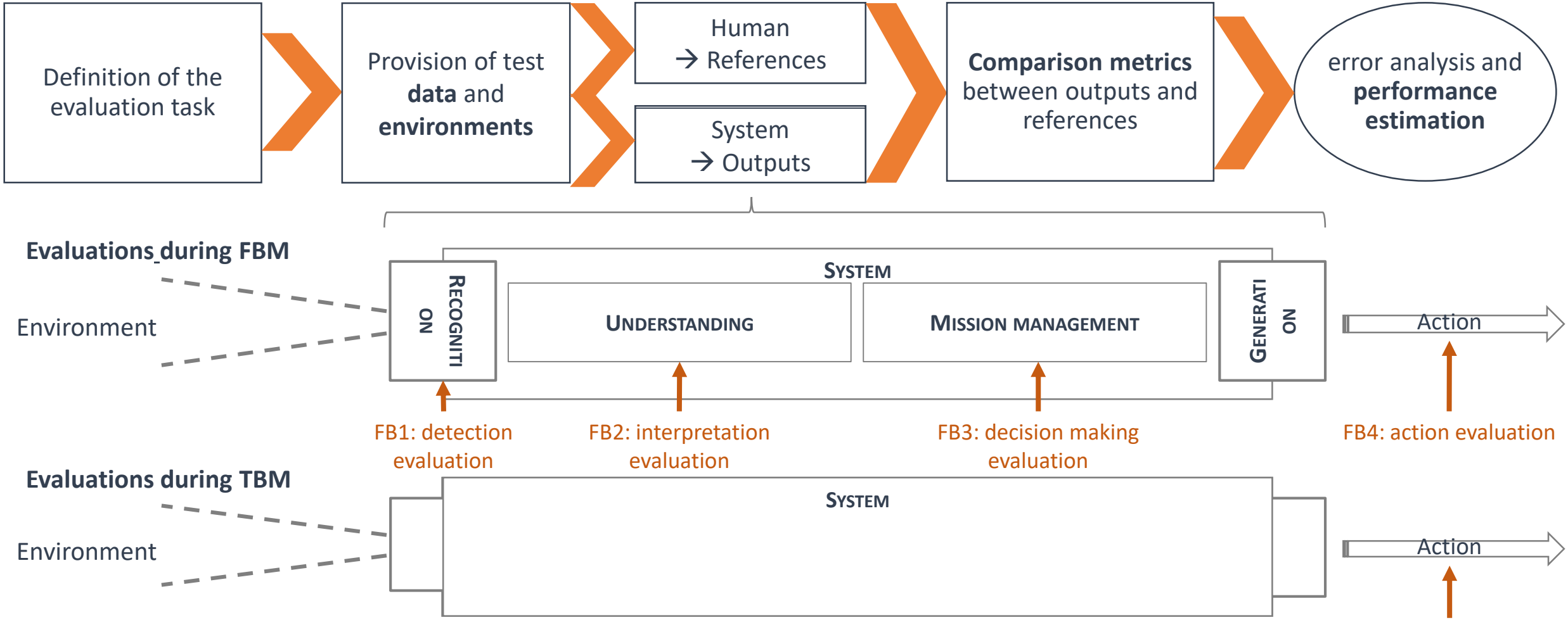
Organizers:

- Competent in organizing such challenges
- Expert in AI evaluation

Country coordinating	Challenge
Estonia	Robotex
France	Repere, Quaero, Rose, Metrics, Allies, Etiseo, Argos
Germany	Elrob
Italy	EuRoC, Promise (CLEF)
Netherlands	MediaEval
Portugal	RoCKIn
Spain	Albayzín evaluation
United-Kingdom	euRathlon

And many more, in Europe and abroad (cf. **NIST** presentation)

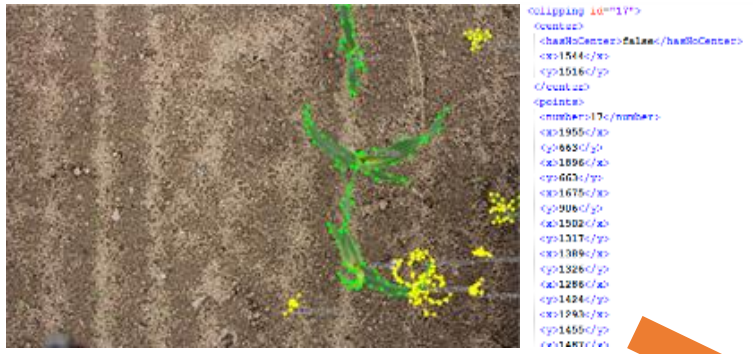
AI EVELUATION METHOD



PROMISE & CLEF (2010 onwards)	REPERE (2012-2014)	QUAERO (OSEO 2008-2014)	ROSE (AFB, ANR, 2017-2020)	Allies (CHIST-ERA, 2017-2020)	Kaggle (2010 onwards)
Multimedia&multilingual	People recognition	Search engines	Weeding robots	Lifelong learning	multidomain

AI EVELUATION METHOD

Acquisition of images by the 4 evaluated robots



References: human annotations

Hypotheses : outputs from detection systems

Metric:

$$EGER = \frac{\sum_{k=1}^{N} C_k + FP_k + FN_k}{\sum_{k=1}^{N} R_k}$$

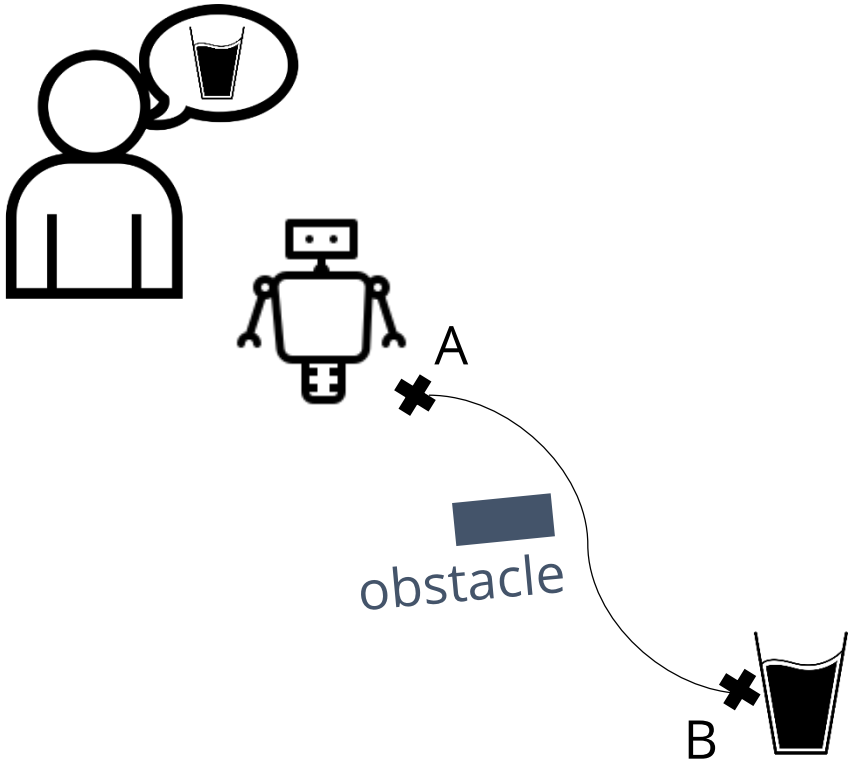
Comparaison

1. Mapping
2. Calculation of the error rate

FUNCTIONALITY AND TASK BENCHMARKS

Task benchmarks (TBM)

TBM1 : To fetch for a glass of water when asked

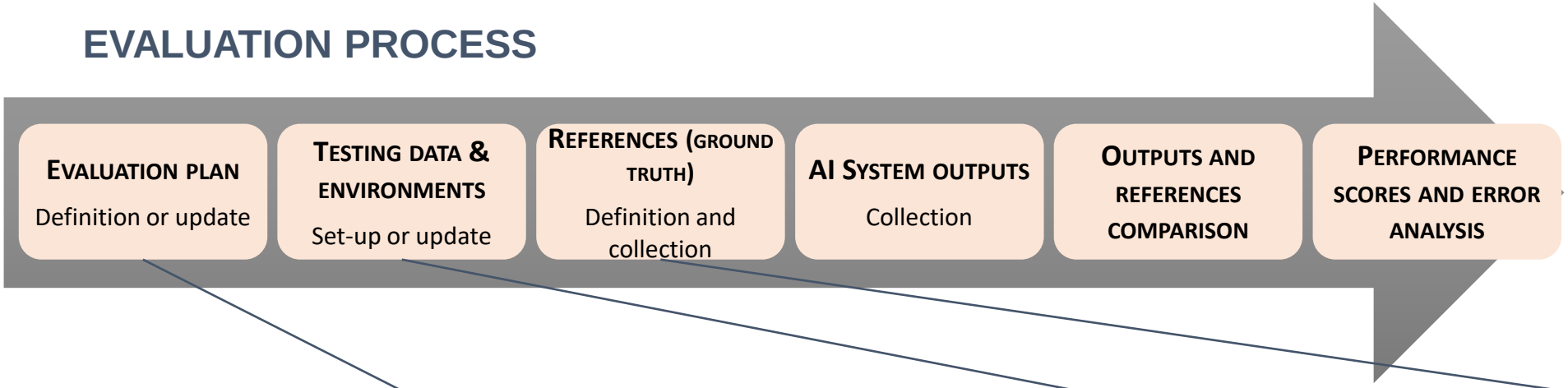


Functionality benchmarks (FBM)

FBM-1 : To understand fetching orders
FBM-2 : To detect obstacles
…
FBM-N : To grab a drink

Eurathlon (FP7, 2013-2015)	RoCKIn (FP7, 2013-2015)	EuRoC (FP7, 2014-2017)	SciRoc (H2020, 2018-2022)
Emergency response	Home, work	Manufacturing, logistics, inspection	Smart cities

EVALUATION PROCESS



(To be repeated at each evaluation campaign of the challenge)

Evaluation plan

1. Testing scenarios

- Identification of technoscientific barriers to be removed
- Definition of participation terms and conditions
- Definition of the evaluation tasks

2. Protocols, metrics

- Identification of influencing factors
- Definition of evaluation criteria and metrics
- Interpretation of results

Testing data and environments

3. Testing environments

- Development of adapted testing environments
- Control and measure of influencing factors
- Ensure reproducibility of experiments

4. Data

- Data selection: relevance, representativeness, quality
- Development of tools for data management and sharing (server)
- Development of tools for data collection

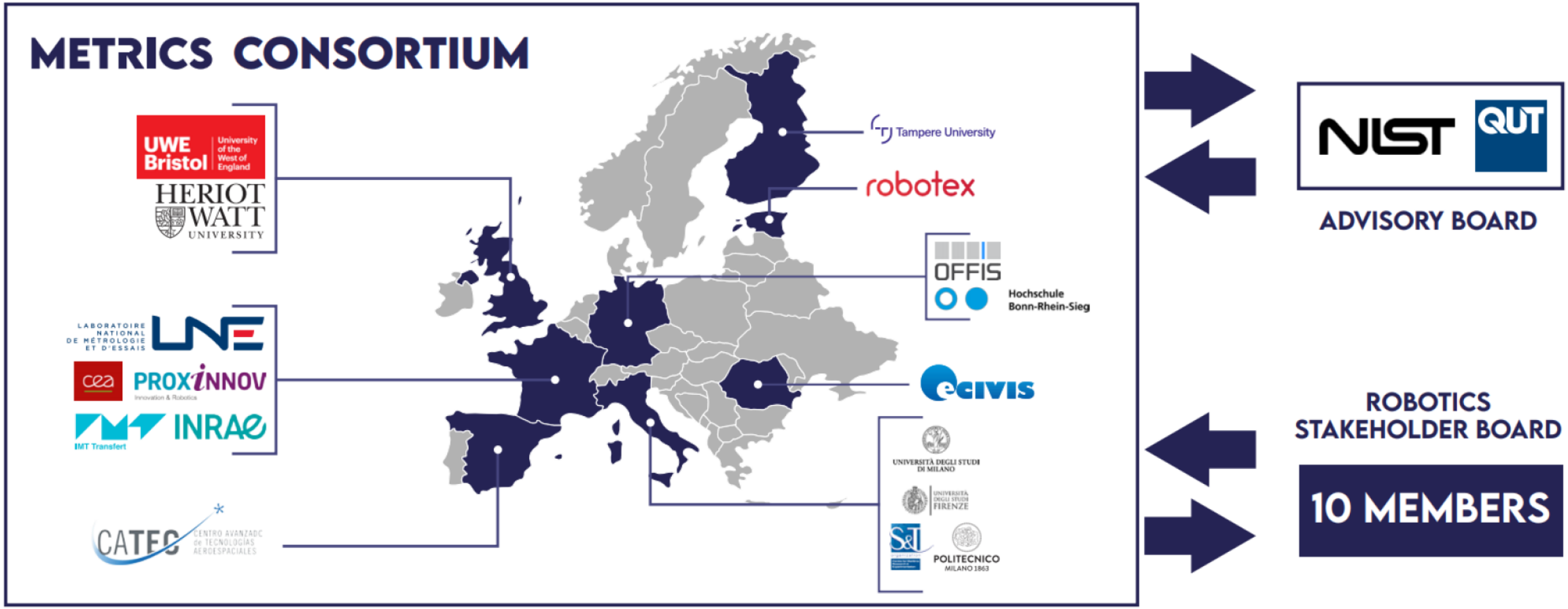
References

5. Ground truth

- Development of annotation systems
- Data annotation or supervision of data annotation
- Qualification of annotations and annotators

METRICS PROJECT

www.metricsproject.eu
guillaume.avrin@lne.fr



METRICS consortium relies on the collaboration of **17 partners from 8 EU countries** (Estonia, Finland, France, Germany, Italy, Romania, Spain, United Kingdom), which will contribute to strengthening the European AI and robotics communities, including in EU Widening countries

METRICS COMPETITIONS

HEART-MET
Healthcare

Assistive robots
1. Assess activity state
2. Item delivery
3. Area coverage
4. Prepare drink
5. Receive and transport drink

RAMI
Inspection and maintenance

Inspection autonomous robots
1. Underwater: pipeline area inspection and intervention
2. Aerial: punctual and repetitive inspection in difficult access areas

ACRE
Agri-food

Weeding robots
1. Intra-row weeding
2. Crop mapping

ADAPT
Agile production

Collaborative assembly robots
1. Collaborative programming for assembly
2. Collaborative assembly of complex parts

HOW TO GET INVOLVED?

As a participant:

- **What:** take part in one of the METRICS competitions by registering your technological solution to the corresponding evaluation campaigns (a robot for field evaluations and/or an AI algorithm for cascade evaluations).
- **Why:** take advantage, free of charge, of the evaluation tools made available by the consortium, test your system, position it in relation to those of the other participants and set up new collaborations.
- **How:** contact the coordinator of the corresponding competition (e.g. acre@metricsproject.eu for the agri-food competition).

As a sponsor:

- **What:** help us drive the competition through sponsorship (cash or in-kind contribution) as well as active involvement in the definition of the scenarios, evaluation criteria and judging of the competitions.
- **Why:** a unique opportunity to shape the competition challenges, rules and evaluation criteria to make them meaningful to your business current and future needs in robotics.
- **How:** contact the METRICS coordinator at info@metricsproject.eu.

TAKE-HOME MESSAGES

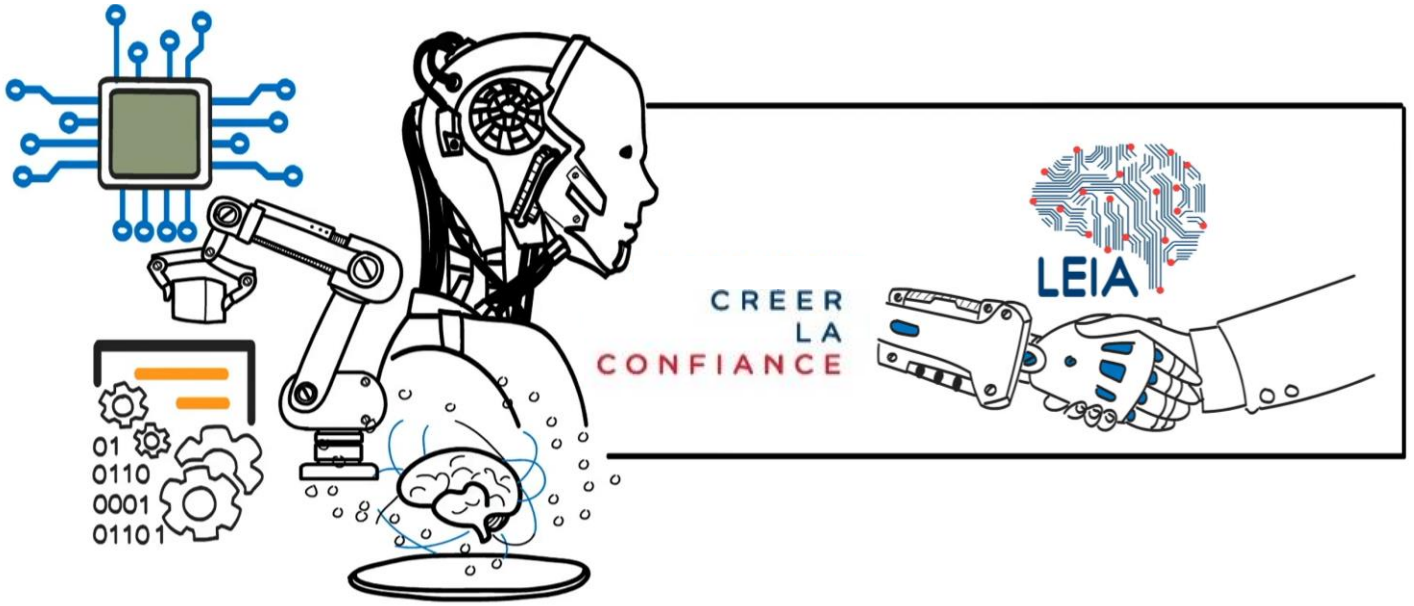
Why organize challenge ?

- To promote progress in AI with an unprecedented efficiency (explicited problems, identified obstacles, validated innovations, measured maturity, easier technology transfer, stimulated cooperations, encouraged competitiveness)
- To evaluate/monitor impact of R&D fundings

How to do it?

- Repeated campaigns over time
- Fair and reproducible evaluation
- Attract participants with high quality datasets
- Rely on modular assessments when possible
- Organizers competent in challenge organization and AI evaluation (protocols, metrics, etc.)

Thank you for your attention



Guillaume AVRIN, PhD
Head of the « AI Evaluation » department
Testing direction
Tel: +33 (0)1 30 69 13 62 - Mob: +33 (0)7 60 49 01 24
guillaume.avrin@lne.fr
Laboratoire national de métrologie et d'essais
29 avenue Roger Hennequin 78197 Trappes Cedex - lne.fr