UNIVERSITY OF CAMBRIDGE

Discussion

Jose Hernandez-Orallo[1,2]

CFI LEVERHULME CENTRE FOR THE FUTURE OF INTELLIGENCE

Vancouver, 20 February, 2024

[1] CFI, University of Cambridge,
[2] VRAIN, Universitat Politècnica de València

Take-aways

- Capability is not average performance
  - Capability allows to predict performance, at the instance level and even OOD
- System Capabilities and Task Demands are related through a “margin”
- Measurement layouts capture domain knowledge and intuitions
- Backward inference: estimate capabilities for one single system
- Forward inference: infer performance for new task instances.

Discussion

- How could capability-oriented evaluation benefit your own work?
- How can this approach be applied when there are too many relevant features?
- What if no meta-features are given?
- What can be done in scenarios where domain knowledge is limited?
- How can these limitations be addressed?

Get involved!

- Add your measurement layouts (to the measurement layouts library):
  - https://github.com/Kinds-of-Intelligence-CFI/measurement-layout-tutorial/
- Contribute to the Animal AI Platform:
  - https://sites.google.com/csah.cam.ac.uk/animalai/
- Follow the AI Evaluation digest to be up-to-date about AI evaluation:
  - https://aievaluation.substack.com/
- Related tutorials:
  - Item Response Theory at EACL2024 (https://github.com/eacl2024irt/eacl2024irt.github.io/blob/main/index.md),
  - AAI at Cogsci2024
- Kind of Intelligence Programme at CFI-Cambridge:
  - http://lcfi.ac.uk/projects/kinds-of-intelligence