

Nordic AI Meet 2026

P3: LLMs, Language, Evaluation, Privacy & Trustworthy AI

3. EvalCards: A Framework for Standardized Evaluation Reporting

Authors: Ruchira Dhar¹; Danae Sanchez Villegas¹; Antonia Karamolegkou¹; Alice Schiavone¹; Yifei Yuan²; Xinyi Chen³; Jiaang Li¹; Stella Frank¹; Laura De Grazia⁴; Monorama Swain⁵; Stephanie Brandl¹; Daniel Hershcovich¹; Anders Søgaard¹; Desmond Elliott¹

Affiliations: 1. University of Copenhagen, Denmark 2. ETH Zürich, Switzerland 3. University of Amsterdam, Netherlands 4. University of Barcelona, Spain 5. Johannes Kepler University Linz, Austria

Abstract: In today's landscape of rapid release and use of public-access AI models, transparent reporting practices are more critical than ever. Drawing on a survey of recent work on evaluation and documentation in the field, we identify three commonly discussed issues and propose the need for a new standard in evaluation reporting. We introduce Evaluation Disclosure Cards (EvalCards) as a potential path forward, highlighting some key design principles and considerations for any such reporting format. To instantiate our idea, we also provide three sample EvalCards for current AI models. We conclude with a discussion of potential issues on adoption of such a format and how EvalCards can be strengthened to enhance transparency for both researchers and practitioners while providing a practical foundation to meet emerging governance requirements.