# Symposium on Scaling AI Assessments

**SAIA 2024, September 30–October 1, 2024, Cologne, Germany**

Edited by

Rebekka Görge
Elena Haedecke
Maximilian Poretschkin
Anna Schmitz

*Editors*

**Rebekka Görge**
Fraunhofer IAIS, Sankt Augustin, Germany
rebekka.goerge@iais.fraunhofer.de

**Elena Haedecke**
Fraunhofer IAIS, Sankt Augustin, Germany
University of Bonn, Bonn, Germany
elena.haedecke@iais.fraunhofer.de

**Maximilian Poretschkin**
Fraunhofer IAIS, Sankt Augustin, Germany
University of Bonn, Bonn, Germany
Maximilian.Poretschkin@iais.fraunhofer.de

**Anna Schmitz**
Fraunhofer IAIS, Sankt Augustin, Germany
Anna.Schmitz@iais.fraunhofer.de

*ACM Classification 2012*
Computing methodologies → Artificial intelligence; Computing methodologies → Machine learning; Applied computing; Social and professional topics → Computing / technology policy; General and reference → General conference proceedings; General and reference → Cross-computing tools and techniques; Software and its engineering → Software creation and management

**ISBN 978-3-95977-357-7**

*Published online and open access by*
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at https://www.dagstuhl.de/dagpub/978-3-95977-357-7.

*Publication date*
January, 2025

*Bibliographic information published by the Deutsche Nationalbibliothek*
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at https://portal.dnb.de.



Digital Object Identifier: 10.4230/OASIcs.SAIA.2024.0
**ISBN 978-3-95977-357-7** **ISSN 1868-8969** **https://www.dagstuhl.de/oasics**

## OASIcs – OpenAccess Series in Informatics

OASIcs is a series of high-quality conference proceedings across all fields in informatics. OASIcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.



**ISSN 1868-8969**

**https://www.dagstuhl.de/oasics**

# Contents



Symposium on Scaling AI Assessments (SAIA 2024).
Editors: Rebekka Görge, Elena Haedecke, Maximilian Poretschkin, and Anna Schmitz
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

## Transparency and XAI



## Certification

# Preface

This volume presents scientific and practical contributions from the Symposium on Scaling AI Assessments (SAIA 2024). SAIA 2024 was held on September 30 and October 1, 2024 in Cologne, Germany. It gathered practitioners from the TIC sector (testing, inspection, certification), representatives from tech start-ups and AI deployers, as well as researchers in the field of trustworthy AI. Together, they discussed and promoted solution approaches towards scalable AI assessments.

Especially against the background of European AI regulation, AI conformity assessment procedures are of particular importance, both for specific use cases and for general-purpose models. But also in non-regulated domains, the quality of AI systems is a decisive factor as unintended behavior can lead to serious financial and reputation damage. As a result, there is a great need for AI audits and assessments and in fact, it can also be observed that a corresponding market is forming. At the same time, there are still (technical) challenges in conducting the required assessments and a lack of extensive practical experience in evaluating different AI systems. Overall, the emergence of the first marketable, commercial AI assessment offerings is just in the process and a definitive, distinct procedure for AI quality assurance has not yet been established. These outstanding challenges can be addressed from two perspectives which must be intertwined to enable scalable solutions:

- **Operationalization perspective:** AI assessments require further operationalization both at level of governance and related processes and at the product level. Empirical research is pending that applies and evaluates governance frameworks, assessment criteria, AI quality KPIs and methodologies in practice for different AI use cases.
- **Testing tools and implementation perspective:** Conducting AI assessments in practice requires a testing ecosystem and tool support, as many quality KPIs cannot be calculated without tool support. At the same time automation of such assessments is a prerequisite to make the corresponding business model scale.

Taking a pragmatic and market-oriented approach in bringing together the two perspectives, SAIA 2024 includes practitioner contributions in addition to academic papers. Specifically, the practitioner track was open for short abstracts of practice reports and case studies, some of which were extended to full papers after the conference. Regarding the academic track, SAIA 2024 places particular emphasis on the commitment of young researchers along more experienced participants. The detailed list of the topics of interest is provided below. Beyond the presentations from the academic and practitioner tracks, the conference program included keynotes by Prof. Dr. Bertrand Braunschweig, scientific coordinator of Confiance.ai, and Prof. Dr. Roberto V. Zicari, head of the Z-Inspection initiative, who shared their experience on implementing trustworthy and ethical AI in practice. In addition, a legal panel with Dr. Andreas Engel, Prof. Dr. Dimitrios Linardatos and Prof. Dr. Mark Cole dealt with questions such as what requirements the AI Act places on generative AI and how it interacts with other complementary legal frameworks such as the GDPR.

We thank the program committee very much for their contribution to the planning and organization of the Symposium on Scaling AI Assessments and for their effort in reviewing the papers with care and quality. We are especially grateful for the international cooperation in the program committee with with representatives of Confiance.ai, Confiance IA and CSIRO Australia. With your support, SAIA 2024 provided a framework for practitioners and researchers the field of AI assessment to become more connected as an interdiscip-

Symposium on Scaling AI Assessments (SAIA 2024).
Editors: Rebekka Görge, Elena Haedecke, Maximilian Poretschkin, and Anna Schmitz
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

linary community. In this sense, SAIA 2024 should also be seen as a contribution to the development of an international and interdisciplinary community on this topic, building on previous conferences and workshops, namely *AITA: AI Trustworthiness Assessment* and *RAIE: International Workshop on Responsible AI Engineering*[1,2]. We hope that our joint efforts have encouraged further cooperation also beyond the conference, since this is an important prerequisite for driving scalable AI assessment forward and expanding the scientific state of art at the same time. Last but not least, we thank all the participants for presenting their work and contributing to lively discussions.

SAIA 2024 and these proceedings were organized as part of the flagship project ZERTIFIZIERTE KI which is funded by the Ministry of Economic Affairs, Industry, Climate Action and Energy of the State of North Rhine-Westphalia, Germany. The editors would like to thank the consortium for the successful cooperation.

## Topics of Interest

- **Standardization of concepts and frameworks for AI assessment**
  - *Operationalization perspective:* How can basic concepts of AI assessments such as the target-of-evaluation, the operational environment and the operational design domain (ODD) be specified in a standardized way? How can compatibility with existing assessment or certification frameworks for other domains (e.g. safety, data protection) be guaranteed? How to deal with third party components, in particular general-purpose AI models, that are difficult to access during an assessment?
- **Risk assessment and safeguarding**
  - *Operationalization perspective:* What methodologies can be employed to effectively characterize and evaluate potential risks and vulnerabilities, considering both the technical aspects and broader implications? How must AI governance frameworks look like to mitigate those risks efficiently?
  - *Testing tools and implementation perspective:* What strategies or methods can developers employ to select suitable testing or risk mitigation measures tailored to the specific characteristics of their AI systems? What are novel techniques, tools or approaches for quality assurance? How can systematic tests be performed and what guarantees can these tests give? In particular, how can diverse test examples be generated, including corner cases and synthetic data, to enhance the robustness and quality of AI products? How can generative AI be used as part of assessment tools e.g., for generating test cases?
- **Conformity with Regulations**
  - *Operationalization perspective:* How can compliance with the AI Act and upcoming regulations be implemented into AI software and AI systems, particularly in specific use cases, and what steps are required for achieving and maintaining compliance? In other words, how does a trustworthy AIOps framework look like?

---

[1] Bertrand Braunschweig, Stefan Buijsman, Faïcel Chamroukhi, Fredrik Heintz, Foutse Khomh, Juliette Mattioli, Maximilian Poretschkin. AITA: AI Trustworthiness Assessment. In *AI and Ethics 4*, pages 1–3. 2024

[2] Qinghua Lu, Foutse Khomh, Apostol T. Vassilev, Maximilian Poretschkin. 2nd International Workshop on Responsible AI Engineering (RAIE'24). Foreword to RAIE 2024. In *IEEE/ACM International Workshop on Responsible AI Engineering (RAIE)*, pages 7-7. 2024

- **Business models and practical application of AI assessments**
  - *Operationalization perspective:* What are business models based on AI assessments and what are key success factors for them? How must assessment criteria be formulated and which KPIs are suitable to make AI quality and trustworthiness measurable in specific AI systems? How need AI quality seals be designed and how do they influence consumers' decisions?
- **Infrastructure and automation:**
  - *Testing tools and implementation perspective:* What infrastructure and ecosystem setup is necessary for effective AI assessment and certification, including considerations for data and model access, protection of sensitive information, and interoperability of assessment tools? Which approaches are there to automate the assessment (process) as much as possible?

# Organizers of the Workshop

## Organizing Committee

- Rebekka Görge, Fraunhofer IAIS, Germany
- Elena Haedecke, Fraunhofer IAIS, University of Bonn, Germany
- Fabian Malms, Fraunhofer IAIS, Germany
- Maxmilian Poretschkin, Fraunhofer IAIS, University of Bonn, Germany
- Anna Schmitz, Fraunhofer IAIS, Germany

## Program Committee

- Bertrand Braunschweig, Confiance.ai, France
- Lucie Flek, University of Bonn, Lamarr Institute for AI and ML, Germany
- Antoine Gautier, QuantPi, Germany
- Marc Hauer, TÜV AI.Lab, Germany
- Manoj Kahdan, RWTH Aachen, Germany
- Foutse Khomh, Polytechnique Montreal, Canada
- Julia Krämer, Erasmus School of Law in Rotterdam, Netherlands
- Qinghua Lu, CSIRO, Australia
- Jakob Rehof, TU Dortmund, Lamarr Institute for AI and ML, Germany
- Franziska Weindauer, TÜV AI.Lab, Germany
- Stefan Wrobel, University of Bonn, Fraunhofer IAIS, Germany
- Jan Zawadzki, Certif.AI, Germany

## Additional Reviewers

- Sujan Gannamaneni, Fraunhofer IAIS, Germany
- Anna Hake, QuantPi, Germany
- Yue Liu, CSIRO, Australia
- Max Losch, QuantPi, Germany
- Michael Mock, Fraunhofer IAIS, Germany
- Mahesh Chandra Mukkamala, QuantPi, Germany
- Maximilian Pintz, Fraunhofer IAIS, University of Bonn, Germany
- Boming Xia, CSIRO, Australia

Symposium on Scaling AI Assessments (SAIA 2024).
Editors: Rebekka Görge, Elena Haedecke, Maximilian Poretschkin, and Anna Schmitz
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany