



aiStudio: Benchmarking Model Guardian

Model Guardian – Designed to For In-Depth Analysis of Model Bias

An award-winning paper by Lee and Singh (2021) finds gaps in open source fairness toolkits resolved by Deloitte’s Model Guardian.

Open source fairness toolkits

Steep learning curve: all six toolkits require the user to “learn a lot of things before [he/she] could get going.” One interviewee said it would take “at least weeks of reading” to get up to speed on relevant literature

Awkward user interface: one-size-fits-all design that is either deceptively simple or too complex to navigate

Limited coverage of model pipeline: the tools only address biases in model build and evaluation phases, overlooking others, e.g. data collection

Incomplete mitigation strategy: the “de-biasing” pre-/in-/post-processing methods are controversial for only solving for a narrow version of fairness. These toolkits only offer technical mitigation, some of which may be inconsistent with anti-discrimination laws

Limited customisation: a lot of the tool is hard-coded to their data, and “a lot of extra work is needed” to adapt the toolkits to a particular use case

Complicated integration: practitioners found the toolkits difficult to “plug and play” with existing workflows, partially due to limited data security

Lack of consensus: each toolkit espouses a different vision of what “fairness” means with limited guidance on which approach is appropriate. Ex) see the framing difference between Fairness 360 and Fairlearn.

Model Guardian (quantitative) + Scorecard (qualitative)

No prior experience required: Model Guardian walks the user step-by-step through each stage of the analysis without assuming prior knowledge on algorithmic fairness

Intuitive design: Model Guardian highlights the key takeaways while offering the option to drill down into details

End-to-end model lifecycle coverage: the scorecard questionnaire addresses the biases introduced through people/process in model pipeline

Targeted mitigation: the best mitigation strategy may not be technical. If biases are introduced through data collection and labelling, they should be addressed through changes in marketing strategy and training of labellers. The scorecard questionnaire helps address the bias at its source

Bespoke, custom set-up: Model Guardian is set up as an accelerator to be easily adapted to contextual and regulatory considerations

Built for plug-and-play: Model Guardian contains building blocks in its code library to easily integrate into any organisational workflow

Bespoke expertise: with each Model Guardian implementation, we draw from Deloitte domain and regulatory experts and work with each organisation to provide a holistic understanding of the ethical issues



This communication contains general information only, and none of Deloitte GmbH Wirtschaftsprüfungsgesellschaft or Deloitte Touche Tohmatsu Limited (“DTTL”), its global network of member firms or their related entities (collectively, the “Deloitte organization”) is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication. DTTL and each of its member firms, and their related entities, are legally separate and independent entities.

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited (“DTTL”), its global network of member firms, and their related entities (collectively, the “Deloitte organization”). DTTL (also referred to as “Deloitte Global”) and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see www.deloitte.com/de/UeberUns to learn more.

Deloitte is a leading global provider of audit and assurance, consulting, financial advisory, risk advisory, tax and related services; legal advisory services in Germany are provided by Deloitte Legal. Our global network of member firms and related entities in more than 150 countries and territories (collectively, the “Deloitte organization”) serves four out of five Fortune Global 500® companies. Learn how Deloitte’s approximately 330,000 people make an impact that matters at www.deloitte.com/de.

