Algorithm Audit

NGO Algorithm Audit

A European knowledge platform for AI bias testing and normative AI standards

November 2023

Activities NGO Algorithm Audit

1. **Normative advice commissions**

Advising on ethical issues that arise in concrete algorithmic practice through deliberative and diverse normative advice commissions, resulting in *algoprudence*

2. **Technical tools**

Implementing and testing technical tools to detect and mitigate bias, e.g., synthetic data generation and bias detection tool. Finalist in Stanford's AI Audit Challenge 2023 with Joint Fairness Assessment Method

3. **Knowledge platform**

Bringing together experts and knowledge to ignite the collective learning process on the responsible use of algorithms, bijv. AI Policy Observatory en position papers

4. **Project work**

Support for specific question of public and private sector organisations on the responsible use of AI

Financially supported by

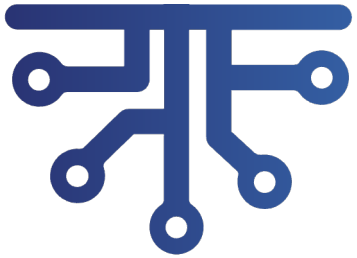
SIDN fonds

European Artificial Intelligence & Society Fund

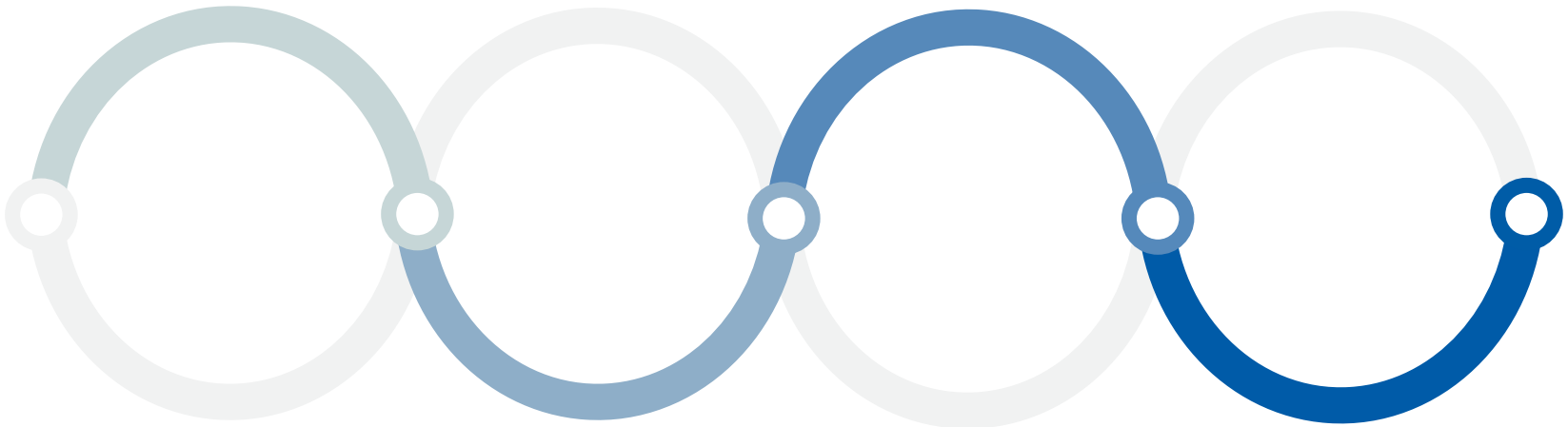
Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

What is *algoprudence*? The outcome of diverse and deliberative normative advice commissions

jurisprudence for *algorithms*



Normative advice commissions make decisions on responsible AI together with stakeholders



Step 1

Identify issue

Identify a concrete ethical concern in a real algorithm or data-analysis tool

Step 2

Problem statement

Describe ethical issue and hear stakeholders and affected groups

Step 3

Normative advice commission

Deliberative conversation on ethical issue by diverse and inclusive normative advice commission (NAC)

Step 4

Public advice

Advice of NAC is published together with problem statement as *algoprudence* on our website

Algorithm Audit's fields of expertise in the AI landscape

Profiling

Advise on algorithmic segmentation techniques, including variable selection to mitigate indirect discrimination, selection of relevant metrics for monitoring and evaluation, devising documentation standards, and identifying non-algorithmic alternatives

See: AA:2022:01 and AA:2023:02 (available soon).

Computer vision

Advise on suitable fairness methodologies for computer vision, for instance:
> blurring personal data
> image classification
> semantic segmentation,
and normative interpretation of quantitative fairness metrics.

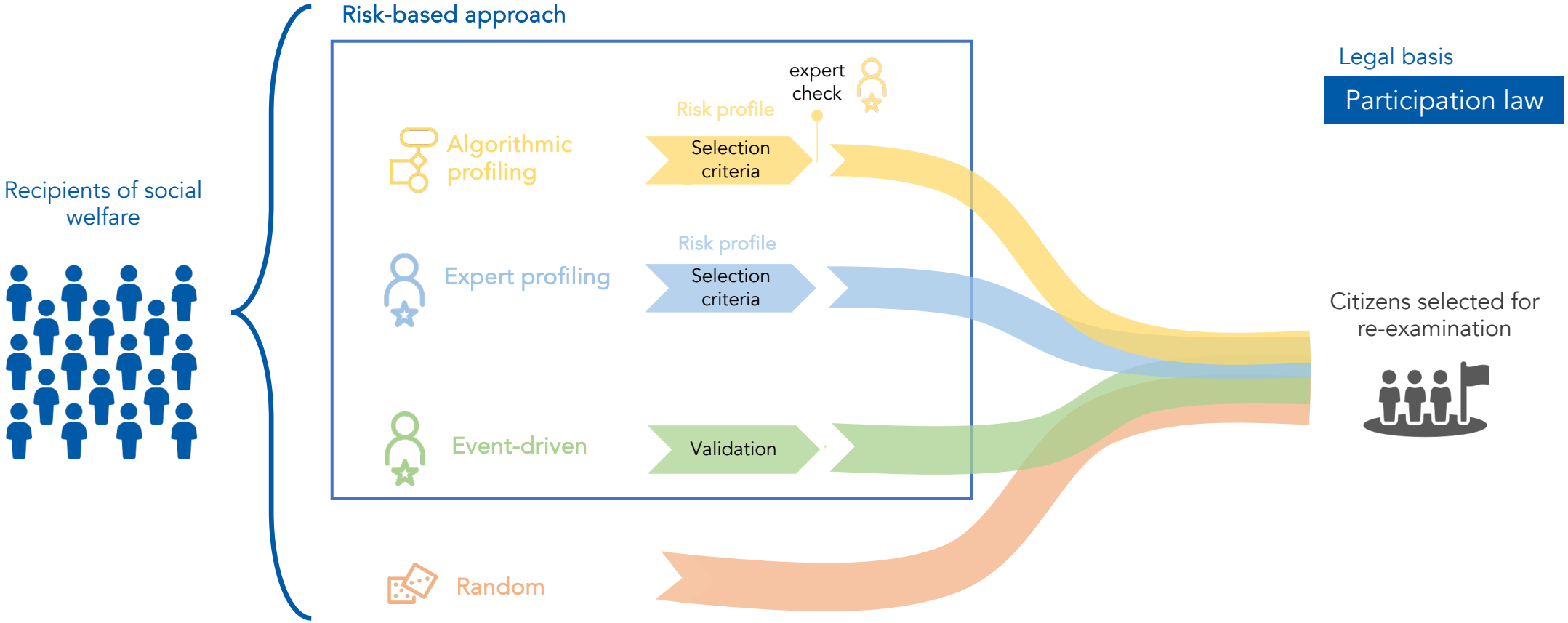
New case studies available in 2024.

Learning to rank

Advise on statistical methodologies underlying recommender systems used for, among others, search engines, user reviews and information retrieval. Focus on independent review of:
> pointwise ranking
> listwise ranking,
including normative interpretation of quantitative evaluation metrics.

Read our white paper on third-party auditing of recommender systems under the Digital Services Act (DSA).

Example – Identify problem (Step 1): Risk profiling for social welfare re-examination by Dutch municipalities



Example – Identify problem (Step 1): Defining ethical issues applying risk profiling in this context

Based on 60+ variables traced in FOI-request Municipality of Rotterdam

Issue I	What characteristics of recipients can be considered as a proxy variable for protected attributes (as defined in Article 14 of the European Convention on Human Rights), and which of those variables should be excluded from profiling methods to mitigate discriminatory bias?
Issue II	What characteristics are ethically undesirable to use in profiling methods, for reasons other than discriminatory bias?
Issue III	Under what circumstances is it desirable to select recipients for re-examination through algorithmic sampling, rather than by subject matter expert (SME) sampling?
Underlying question	How to weigh the impact of re-examination? (fraud-examination vs. service-oriented municipalities)

Example – Problem statement (Step 2): Hearing stakeholders and scoping ethical issues

- Citizens subjected to algorithm
- Representatives of affected groups
- Municipality of Rotterdam
- Investigative journalists
- Municipal institutions (Rotterdam Court of Auditors and Ombudsperson of Rotterdam)
- Legal experts and scientific researchers

Unique code to trace algoprudence (problem statement + advice)

AA:2023:02:P

Problem statement describing ethical issues + context

Example – Problem statement (Step 3): Normative advice commission gathering

Normative advice commission

Maarten van Asten, Alderman Finance, Digitalisation and Event Municipality of Tilburg

Munish Ramlal, Ombudsperson of Metropole region Amsterdam

Abderrahman Al Aazani, Representative of the Ombusperson of Rotterdam

Francien Dechesne, Associate Professor Law and Digital Technologies, Universiteit Leiden

Oskar Gstrein, Assistant Professor Governance and Innovation, Rijksuniversiteit Groningen

1. Initial written reaction

2. Commission gathering

diverse inclusive deliberative

Example – Public advice (Step 4): Normative advice commission gathering

Key take-aways of advice commission:

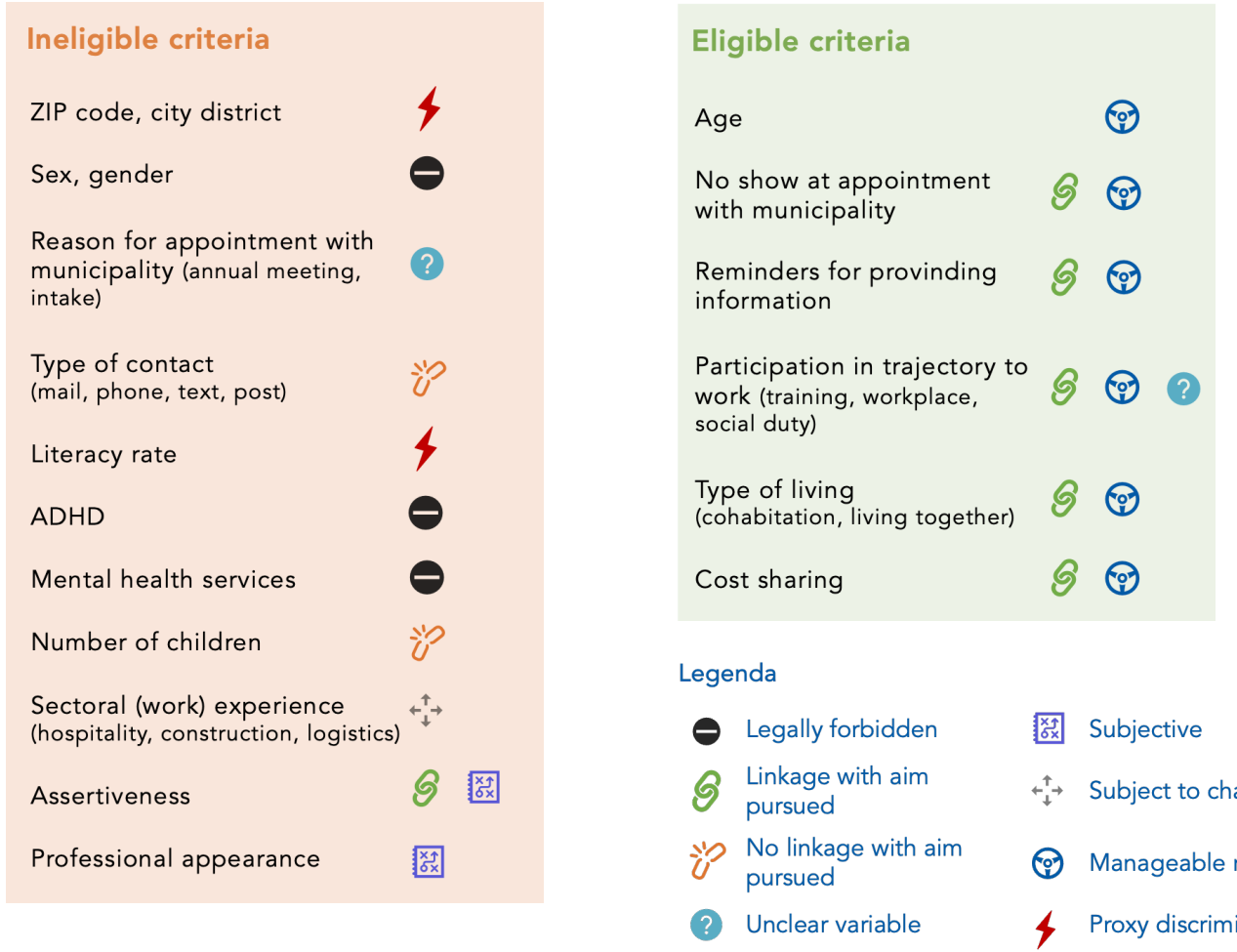
> Algorithmic profiling is possible under strict conditions

> Profiling must not equate suspicion

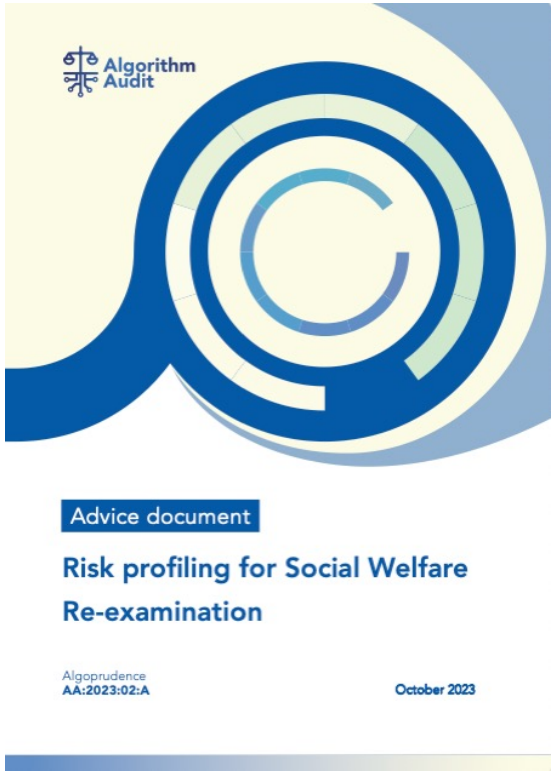
> Diversity in selection methods

> Well-considered use of profiling criteria

> Explainability requirements for machine learning



Algoprudence, so what?



Presented in November 2023 to Dutch Minister of Digitalization

So,

- Addressed to all 342 Dutch municipalities and local councils that need to interpret national law and EU non-discrimination law
- Harmonises interpretation of national law by local organisations in the context of ethical issues arising when applying algorithms
- Concrete suggestions how normative issues in the context of algorithmic-driven decision-making can be made
- Transparency on normative decision-making for public use of algorithms
- European formula how AI can be deployed democratically by engaging with civil society

Overview open source AI auditing tools developed and used by Algorithm Audit

Bias detection tool (BDT)

- Unsupervised bias detection: no access needed to protected labels
- Inform experts which statistically significant deviating datapoints should be examined manually
- Finalist in Stanford AI Audit Competition 2023
- Algorithm Audit project team working on SDG
- Collaborating with Dutch public sector organisations to test SDG in practice

Available as a webapp

Source code on GitHub

Case studies on our website

Synthetic data generation (SDG)

- Promising method to share aggregation statistics of data sets without revealing personal data
- 100% privacy preserving, see real-world example from Rotterdam in which SDG was created and shared
- Algorithm Audit project team working on SDG
- Collaborating with Dutch public sector organisations to test SDG in practice

Advocacy activities range from position papers and op-eds to webinars and regulatory feedback

Digital supervisory authorities
Discuss open legal norms with institutional actors in The Netherlands and in the EU

White papers and op-eds
Write reflections on policy initiatives, such as European Acts and Dutch policy initiatives, see here and here

Affected groups
Engage with individuals subject to AI and collect harms and concerns

Politics
Debate new policy initiatives to remediate failure of legal protection in the digital era

Overview of AI experts affiliated to Algorithm Audit

Board of Algorithm Audit


Ariën Voogt, board member (ethics)


Samaa Mohammad-Ulenberg, board member (AI)


Jurriaan Parie, Director and board member (statistics)

Advisory Board of Algorithm Audit

- Anne Meuwese, Professor Public Law & AI at Leiden University
- Aileen Nielsen, Visiting Assistant Professor Privacy Law at Harvard Law School
- Edgar Whitley, Associate Professor of Information Systems at the Londen School of Economics and Political Sciences
- Kiza Magendane, Writer and founder of Progressive Café
- Raphaële Xenidis, Associate Professor in EU law at Sciences Po Paris

Project work team


Jurriaan Parie, Projectmanager


Ruurd Zanting (legal), Algorithm risk expert


Ola Al Khatib, candidate Law&AI Utrecht University


Floris Holstege, PhD-candidate Machine Learning, University of Amsterdam


Martijn Staal, MSc Public law and MSc Computer Science, University Leiden & Technical University Delft

Project team bias detection tool (BDT)


Joel Persson PhD, ML engineer Spotify


Kirtan Padh, PhD-candidate ML TU München


Mackenzie Jorgensen, PhD-candidate Computer Science King's College

?

Project team synthetic data generation (SDG)

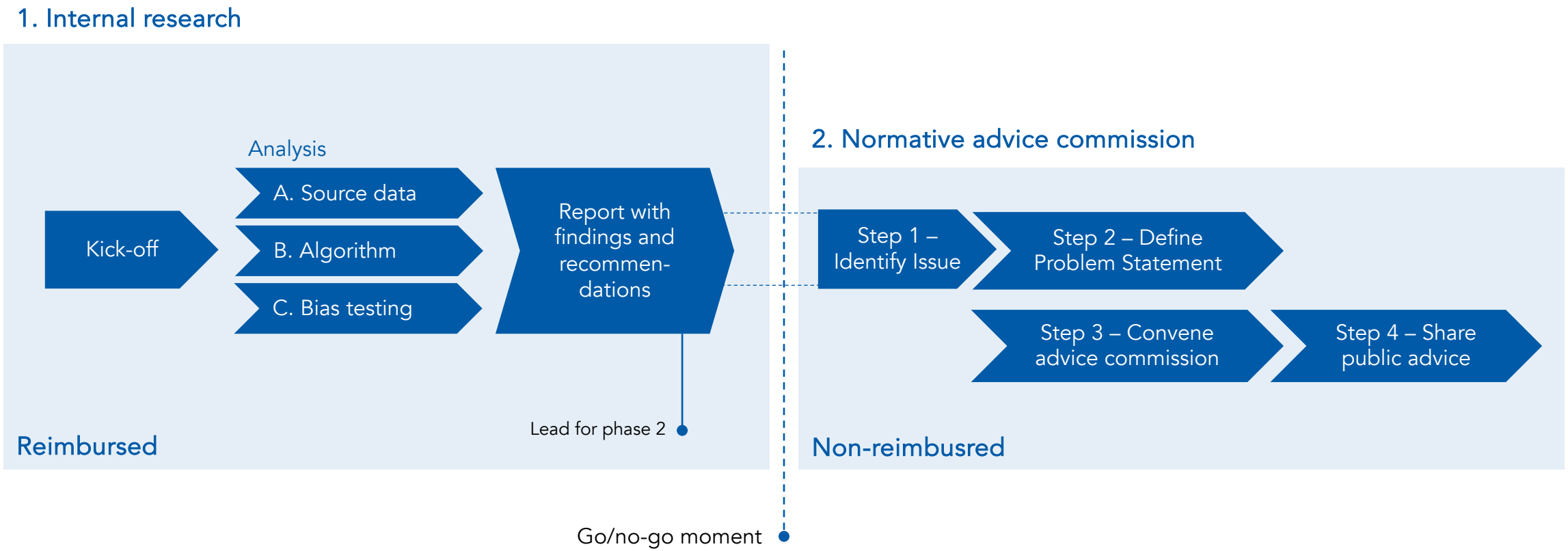

Joel Persson PhD, ML engineer Spotify


Jurriaan Parie, statistics

?

?

Project work is divided in: 1. Internal research and 2. Normative advice commission



Algorithm Audit

www.algorithmaudit.eu

info@algorithmaudit.eu

https://www.linkedin.com/company/algorithm-audit/

https://github.com/NGO-Algorithm-Audit

Stichting Algorithm Audit is bij de Kamer van Koophandel
geregistreerd onder nummer 83979212