Algorithm Audit

Meaningful human oversight for risk profiling algorithms

GDPR Art. 22

+

Art. 14 AI Act

Preventing prohibited automated decision-making

Automated decision-making, including profiling, is prohibited under Art. 22 GDPR

Follow 5 steps to ensure a risk profiling algorithm is not prohibited

Step 1 – Create overview of decision-making process

Step 2 – Determine type of decision

Step 3 – Ensure meaningful human intervention

conditional on outcome Step 3

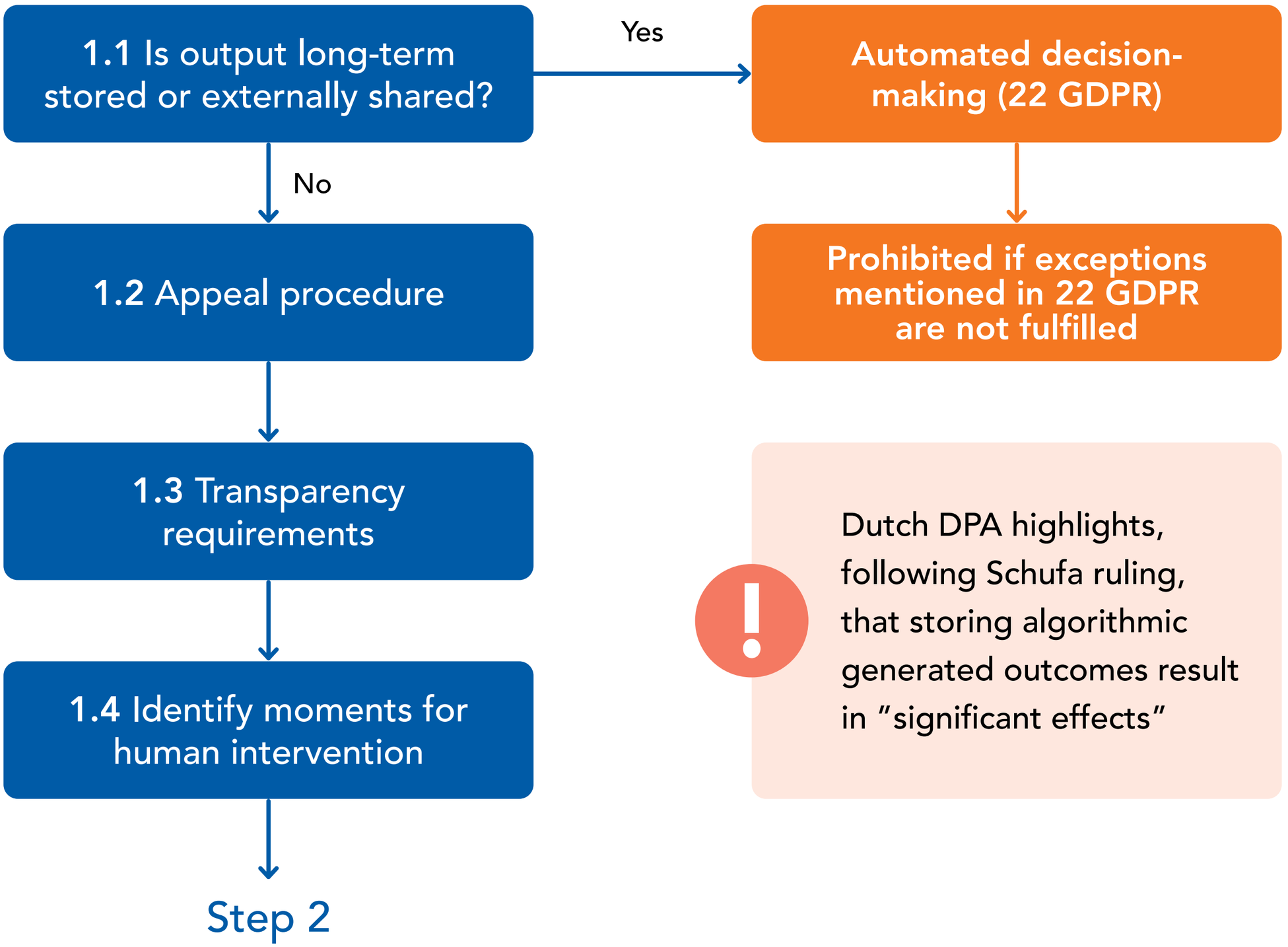
Step 4 – Perform data-analysis to assess effect of risk profiling algorithm

conditional on outcome Step 4

Step 5 – Conduct field experiment *automation bias*

Step 1 – Create overview of decision-making process

Follow the steps below



Dutch DPA highlights, following Schufa ruling, that storing algorithmic generated outcomes result in "significant effects"

Step 2 – Determine type of decision

Examine the impact of the decision-making process the algorithm is embedded in

Legal effects or other similarly significant effects for individuals

- Formal decision (e.g., as defined in public administration law)
- Decision with financial implications
- Entering into an agreement
- Selection for a high impact investigation (house visit)
- Decision affecting someone's access to education
- Decisions influencing someone's employment opportunities
- Other types of decision with significant effect

Step 3

No legal effects or other similarly significant effects

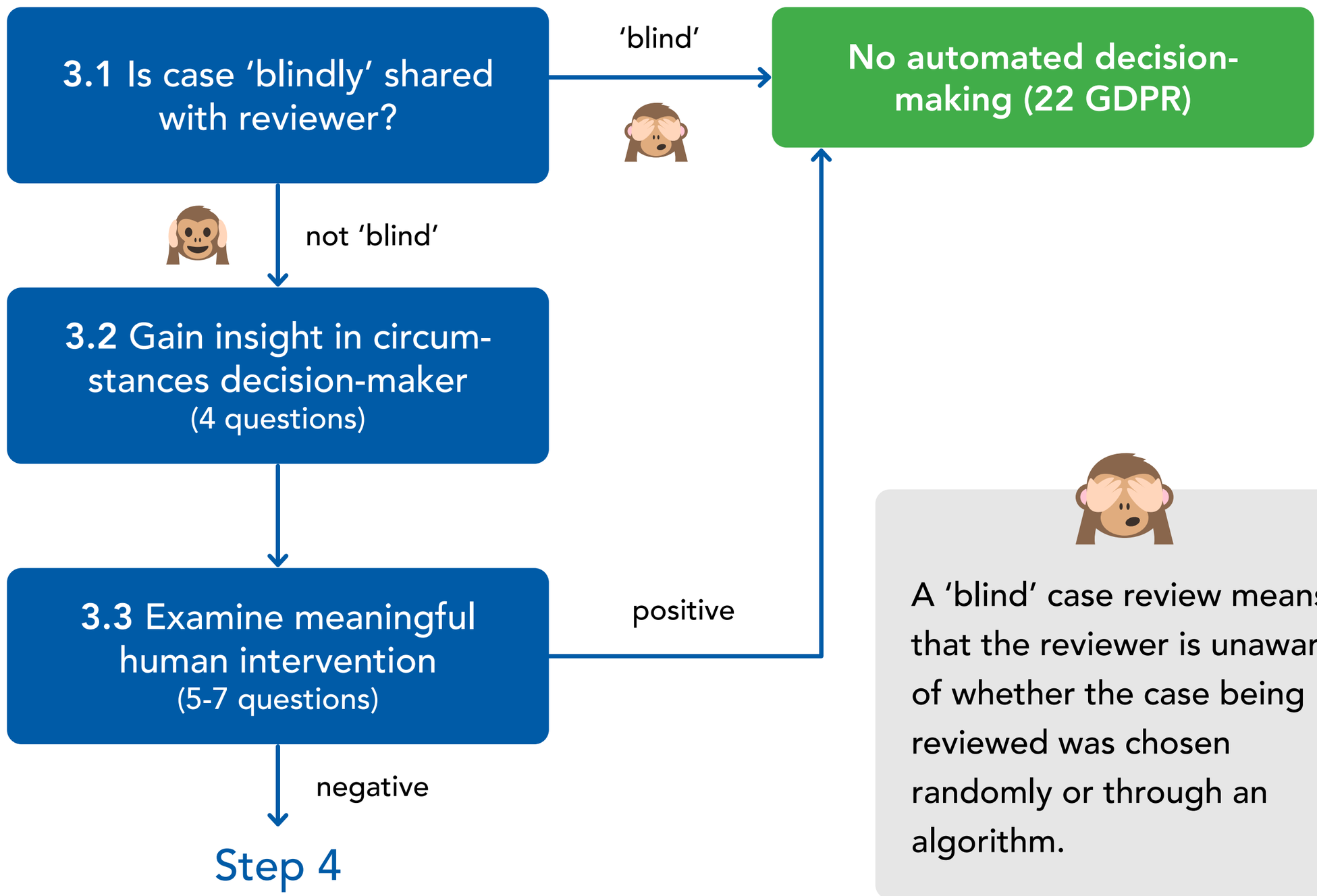
- Issuing a warning
- Prioritization of applications
- Selection for a low impact investigation (information request, warning)

No high-impact decision

No automated decision-making (22 GDPR)

Step 3 – Ensure meaningful human intervention

Follow the steps below



A ‘blind’ case review means that the reviewer is unaware of whether the case being reviewed was chosen randomly or through an algorithm.

Step 3.2 – Gain insight in the circumstances under which reviewers must make a decision

Ask the following 4 questions

i. On what information should reviewers base their assessment or challenge the algorithm?

ii. How much data do reviewers have access to when making a decision?

iii. What qualifications must reviewers meet to be eligible to make decisions?

iv. How much time do reviewers typically have to assess the outcome of an algorithm? How does this compare to the nature of the decision to be made?

Questions in Step 3.2 and 3.3 are a selection of the most relevant questions of the Consultation Meaningful human intervention in algorhtmic decision-making of the Dutch Data Protection Authority

Step 3.3 – Examine meaningful human intervention

Ask the following 7 questions

i. Do reviewers understand how and based on what data the profiling algorithm arrives at a result?

ii. Would reviewers still be able to make the decision without the profiling algorithm?

iii. Do the reviewers have sufficient time for their evaluation?

iv. Can reviewers take specific circumstances into account that the algorithm does not consider? Ensure that information in work instructions for reviewers do not overlap with datapoints used in the risk profiling algorithms.

v. Do reviewers have the opportunity to ask for help from each other or a supervisor?

vi. Are quality checks conducted on the reviewers' work?

vii. Is the algorithm adjusted after feedback from reviewers, involved parties, or monitoring?

Step 4 – Perform data-analysis to assess effect of risk profiling algorithm

Follow the steps below

4.1 Categorize predictions as positives and negatives

↓

4.2 Categorize outcome of decisions-makers as positives and negatives

↓

4.3 Determine True Positive Rate (TPR)

low TPR → **No automated decision-making (22 GDPR)**

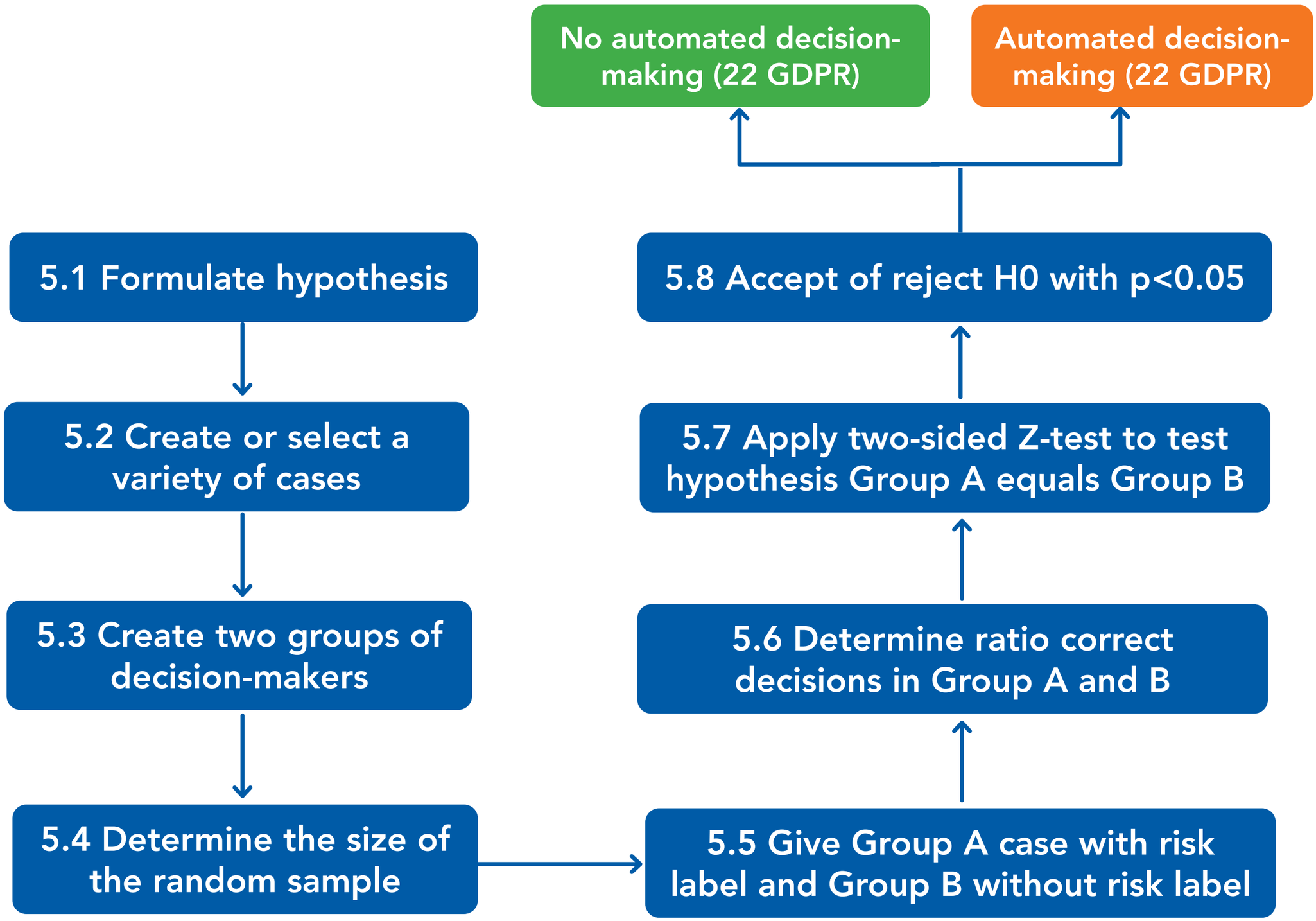
↓ high TPR

Step 5

A low True Positive Rate (TPR) means that decision-makers frequently diverge from the algorithm's prediction, indicating meaningful human intervention

Step 5 – Conduct field experiment *automation bias*

Examine the effect of including algorithmically generated risk labels in case evaluations



All steps are explained in detail in the public standard ‘Meaningful human intervention for risk profiling algorithms’

https://algorithmaudit.eu/knowledge-platform/knowledge-base/public_standard_meaningful_human_intervention/

Building *public knowledge*

for *ethical algorithms*

Join the discussion!

Algorithm Audit

www.algorithmaudit.eu

info@algorithmaudit.eu

www.linkedin.com/company/algorithm-audit

www.github.com/NGO-Algorithm-Audit