

TOWARDS TRANSPARENT AND ACCOUNTABLE AI IN PUBLIC SERVICE



Shaping a more livable world.



ARTIFICIAL INTELLIGENCE

AI impacts everyone, performing complex, risky, or monotonous tasks, aiding doctors and lawyers, and automating public services. However, algorithmic systems may violate human rights, lack transparency, and reinforce discrimination due to biases in data and programming, leading to errors in novel situations.



Governments worldwide are using AI algorithms to automate or support decision-making in public services.



Algorithms are used in urban planning, social care, welfare, unemployment fraud detection, and criminal justice.



The use of AI algorithms is often seen as a way to improve efficiency and reduce costs of public services.



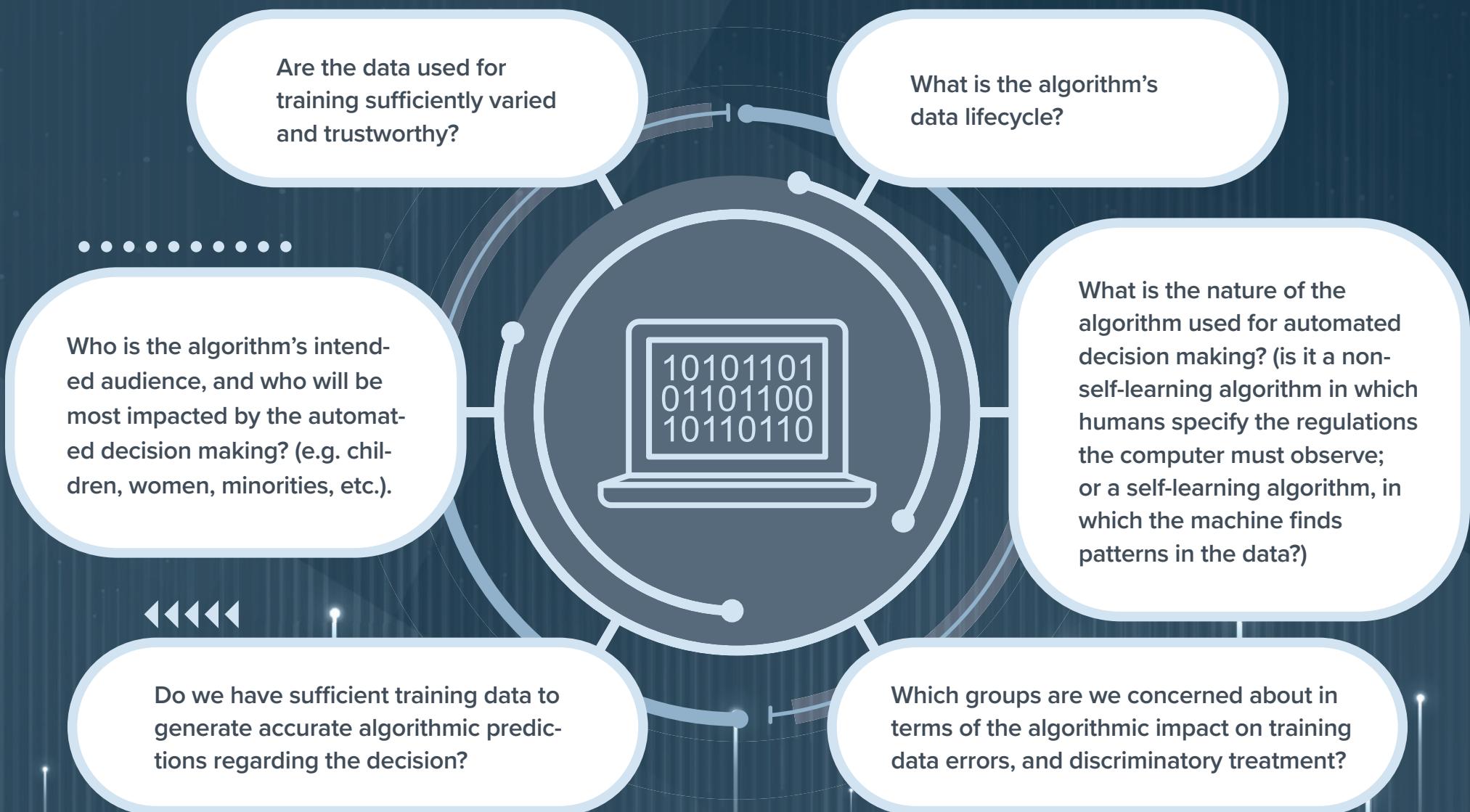


MAPPING, CONCEPTUALIZATION, AND INITIAL ANALYSIS

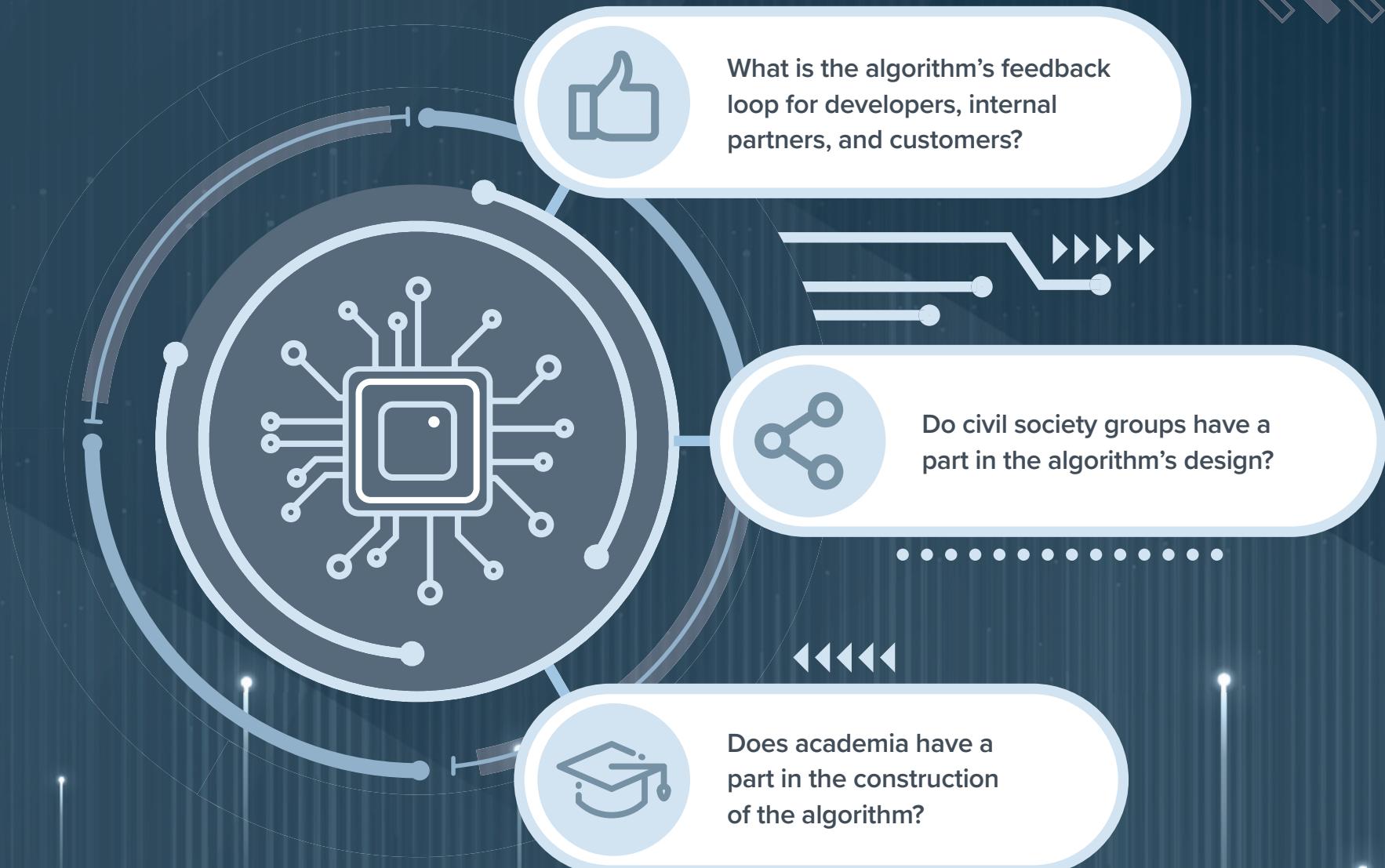
Questions to Consider



WHAT WILL THE AUTOMATED DECISION DO?



HOW ARE ADDITIONAL STAKEHOLDERS ENGAGED?



HAS DIVERSITY BEEN TAKEN INTO ACCOUNT IN THE DESIGN AND IMPLEMENTATION?

Will the algorithm affect particular cultural groups and behave differently in cultural contexts?

Is the design team sufficiently diverse to capture cultural subtleties and foresee the algorithm's applicability in various cultural contexts?

If not, what measures do we have in place to make these scenarios more prominent and comprehensible to designers?

Considering the objective of the algorithm, are the training data sufficiently diverse?

Are there statutory guidelines that public sector organizations should check to ensure that the application of the algorithm is legal and ethical?



WHAT ARE THE OBJECTIVES OF THE AUTOMATED DECISION MAKING PROCESS?

Why is the algorithm needed and what outcomes is it intended to enable?

WHAT IS THE LEGAL BASIS FOR AUTOMATED DECISION MAKING?

If an algorithm is expected to affect human rights, there must be a legal basis for its use.



WHAT ARE THE INCENTIVES FOR AUTOMATED DECISION MAKING?



BENEFITS

What will be our main benefits get from the algorithm's development?



TRANSPARENCY

How transparent will we make the algorithm's design process to internal partners and external clients?



IDENTIFICATION

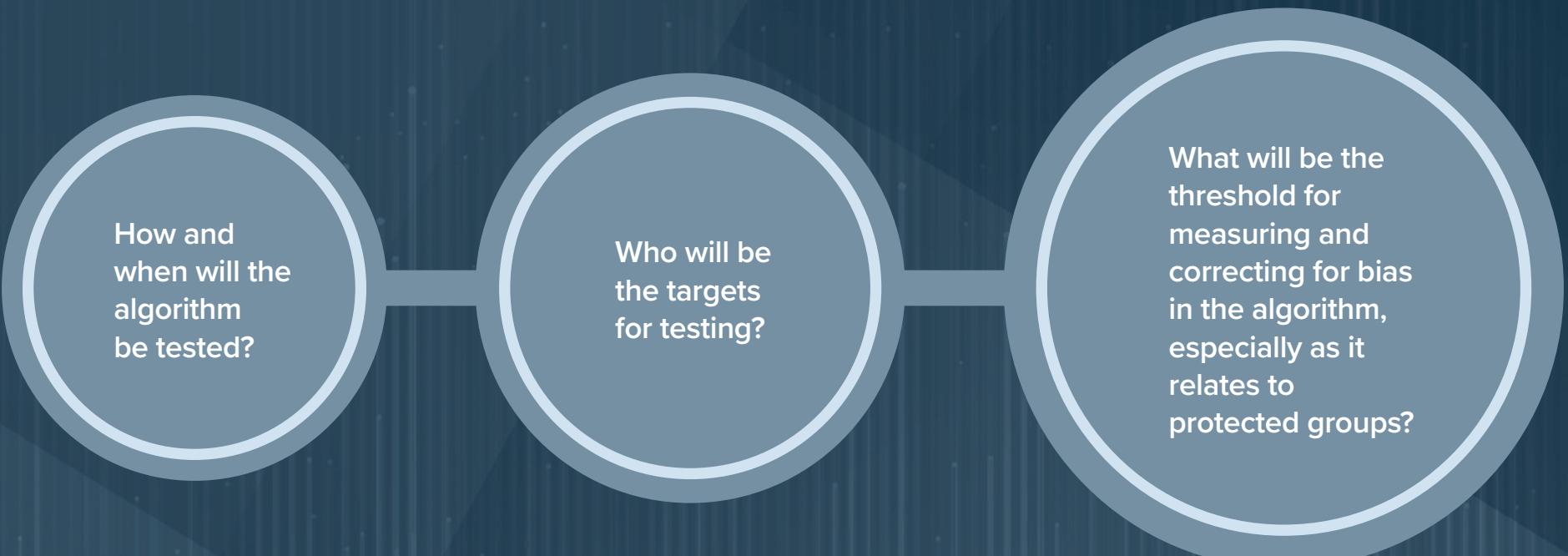
What are the possible adverse outcomes, and how will we identify them?

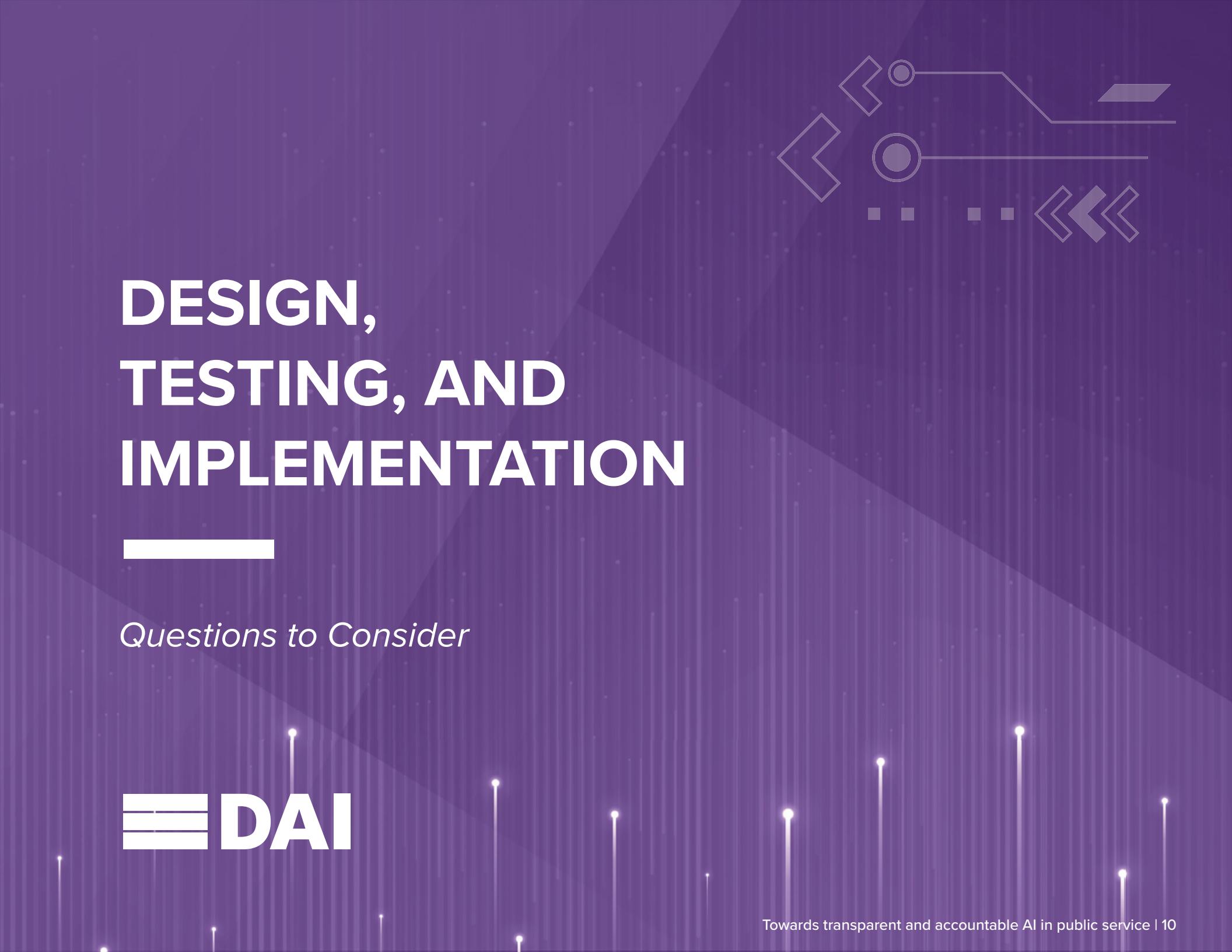


ACTION

What action will be done if it is predicted that the development or deployment of the algorithm may result in undesirable outcomes?

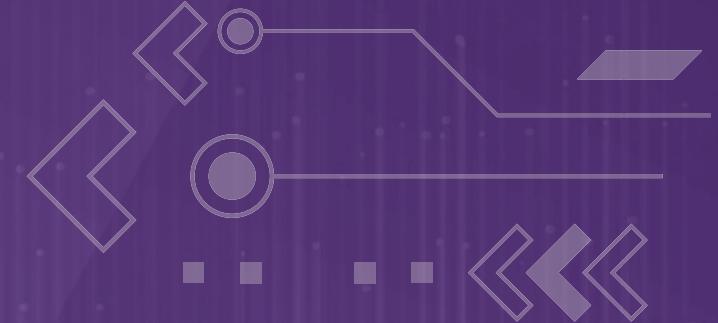
HOW WILL POTENTIAL BIAS BE DETECTED?



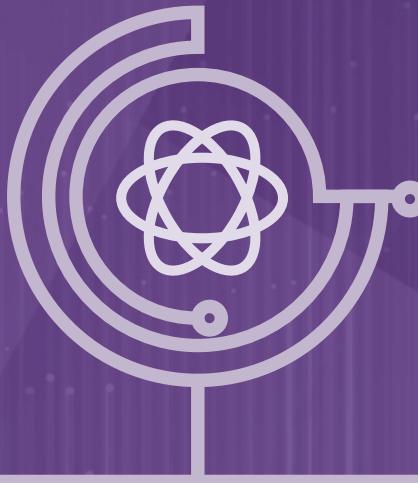


DESIGN, TESTING, AND IMPLEMENTATION

Questions to Consider



INFRINGED FUNDAMENTAL RIGHTS



Are the design testing and implementation going to impact fundamental rights such as, for example, privacy and data protection, freedom of expression, effective remedy and due process, rights to protection against

discrimination, the right to explanation, access to information, freedom of religion, freedom of association, and other fundamental rights as defined by the International Bill of Rights and national human rights law?

SPECIFIC LEGISLATION

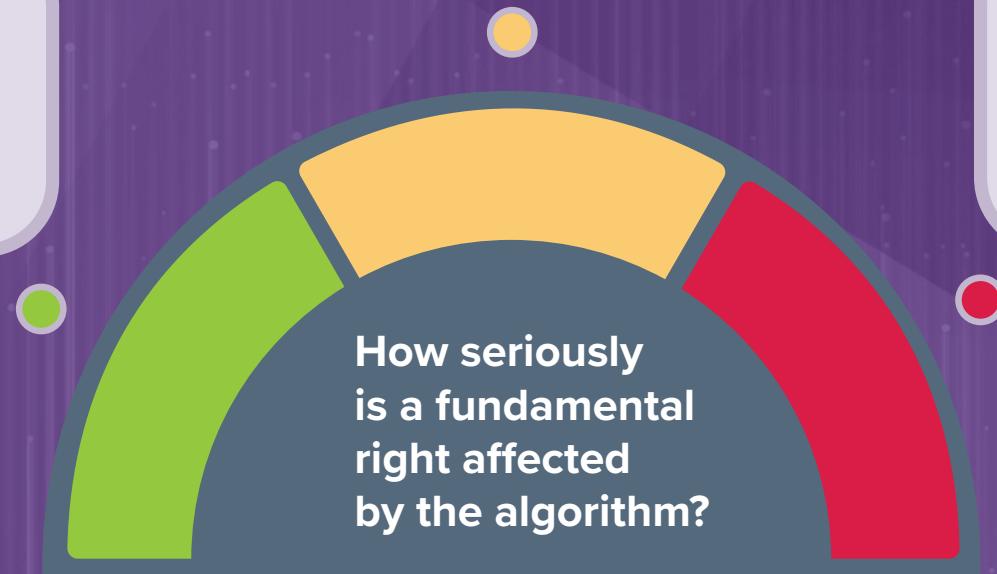


SERIOUSNESS OF INTERFERENCE

Less serious interference, thus no special due diligence required.

Medium-serious interference, thus due diligence required.

Serious interference, thus compelling reasons required as justification.



A useful risk based assessment framework is provided by the EU AI Act.



<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex-%3A52021PC0206>



MONITORING AND EVALUATION

Questions to Consider



LEVEL OF HUMAN INVOLVEMENT

Human in the loop versus
human out of the loop

**What role do
humans play in
decision making
based on the
algorithmic
output?**

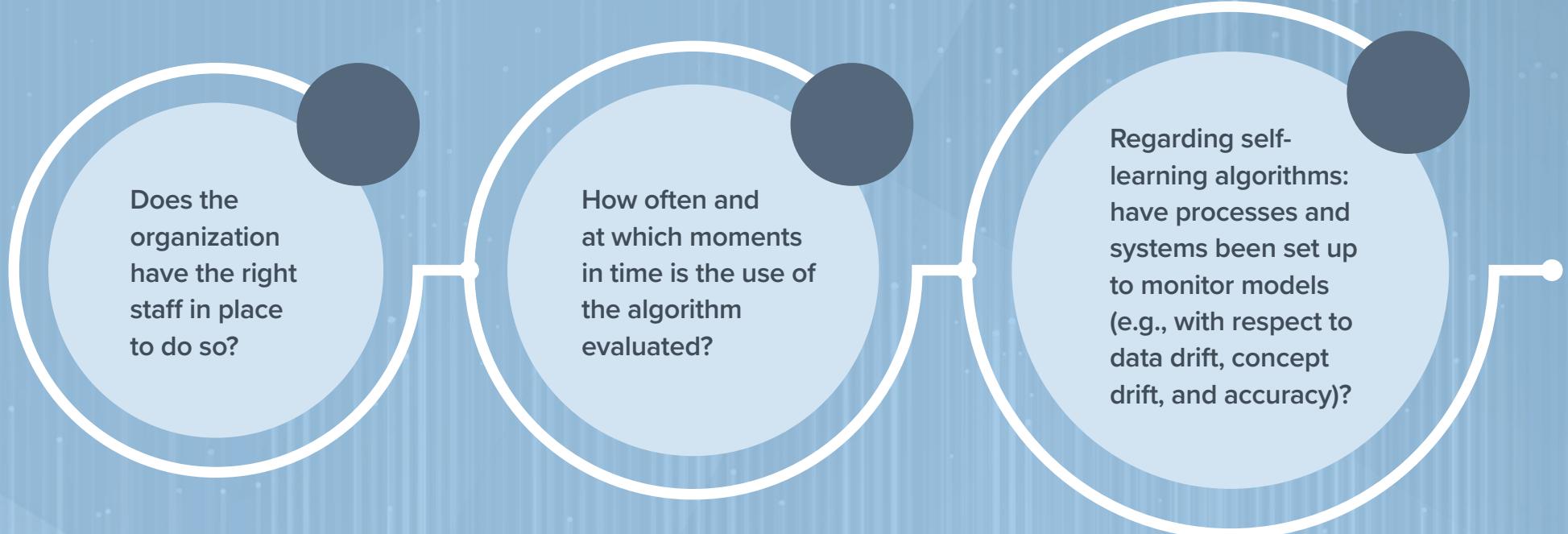
How is staff
empowered to
make decisions
responsibly based
on the algorithmic
output?

Does the AI model
provide enough
information for the
human to make an
informed decision
(e.g., factors that are
used in the decision,
their value and
weighting, correlations)

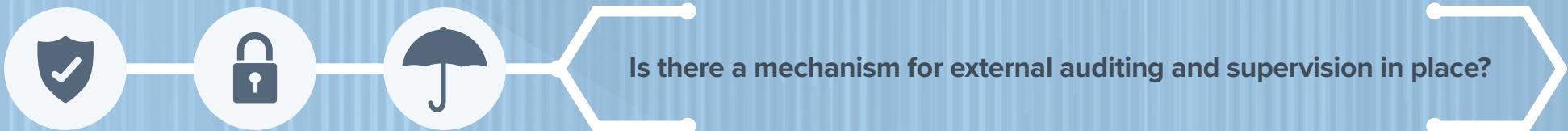
Is there sufficient
qualified staff in
place to manage,
review, and adjust
the algorithm, if
needed, and will
there be in future?

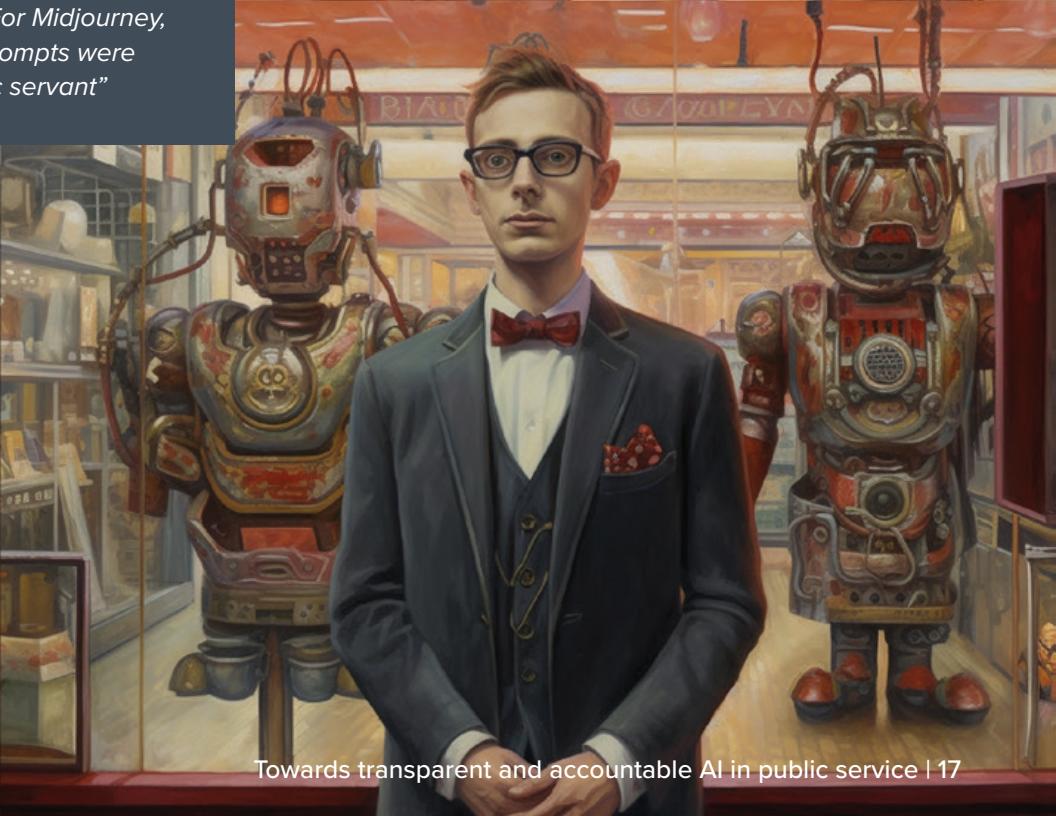
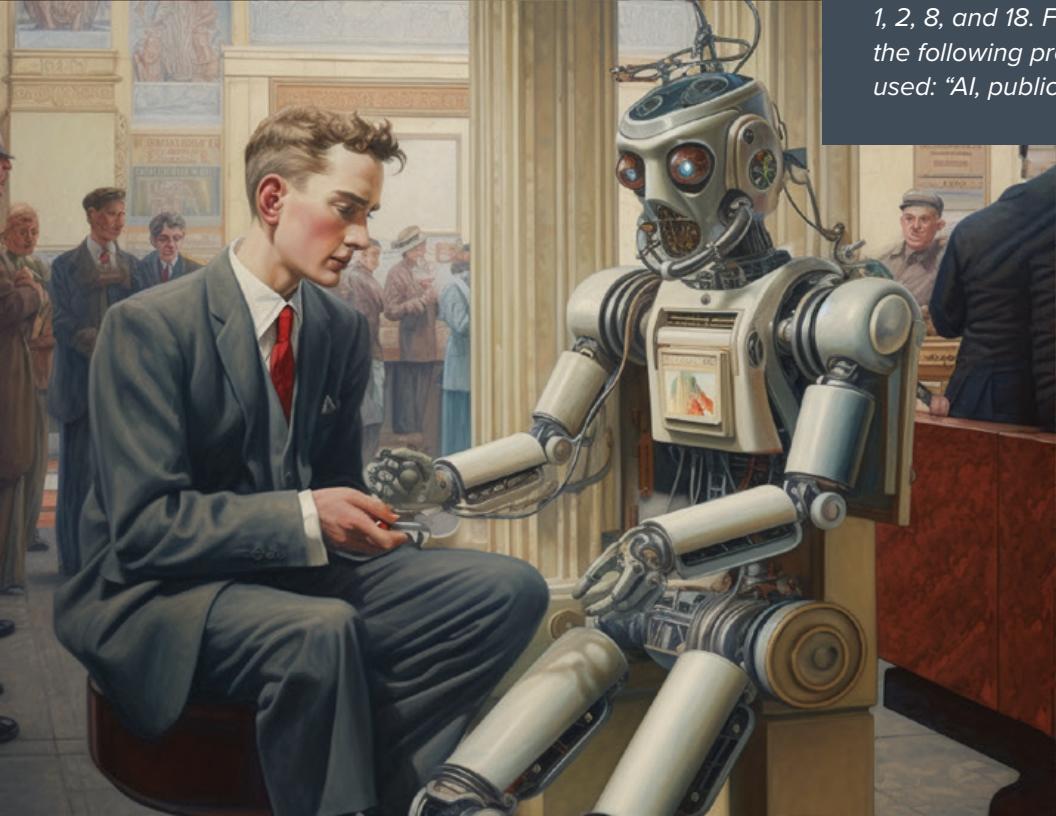
Is there an active and involved
human oversight, with the human
retaining full control and the AI
only providing recommendations
or input? For example, a judge
may use AI to evaluate certain
aspects of a case. However, the
judge will make the final decision.
In the case of human out of the loop,
a criminal recidivism solution may
automatically rank individuals based
on pre-determined demographic
and behavioral profiles.

INTERNAL PROCESS SAFEGUARDS



EXTERNAL PROCESS SAFEGUARDS





The DAI Creative Team used Midjourney, an AI art generator, to create the images on pages 1, 2, 8, and 18. For Midjourney, the following prompts were used: "AI, public servant"