



Correctie van bias versterkende feedbackloop

Eric van Esch

Het probleem

de Correspondent

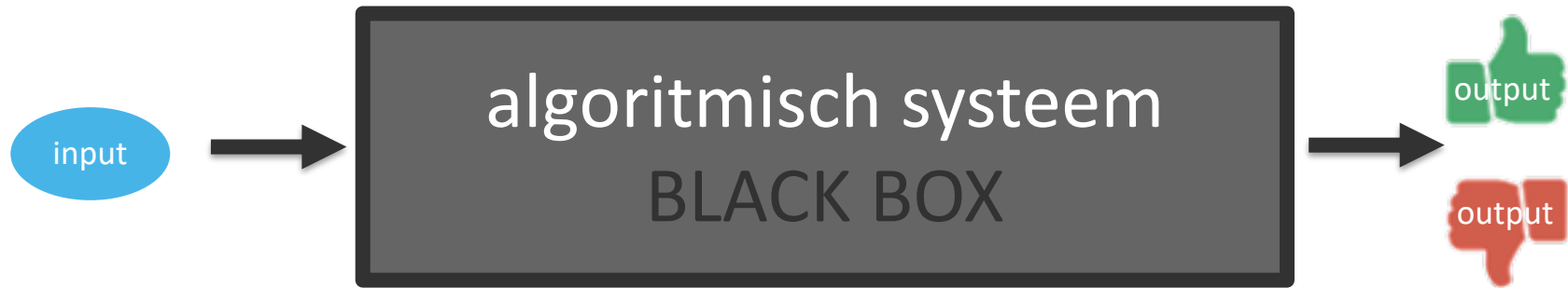
Kun je van algoritmes verwachten dat ze onbevooroordeeld zijn?

Een van de problemen met zo'n systeem: het is bevooroordeeld. SyRI werd letterlijk enkel ingezet in achterstandswijken. Maar ga je op bepaalde plekken zoeken, dan zul je ook alleen op die plekken iets vinden. De mensen die in rijke buurten de fout ingaan, zullen nooit in de database terechtkomen.

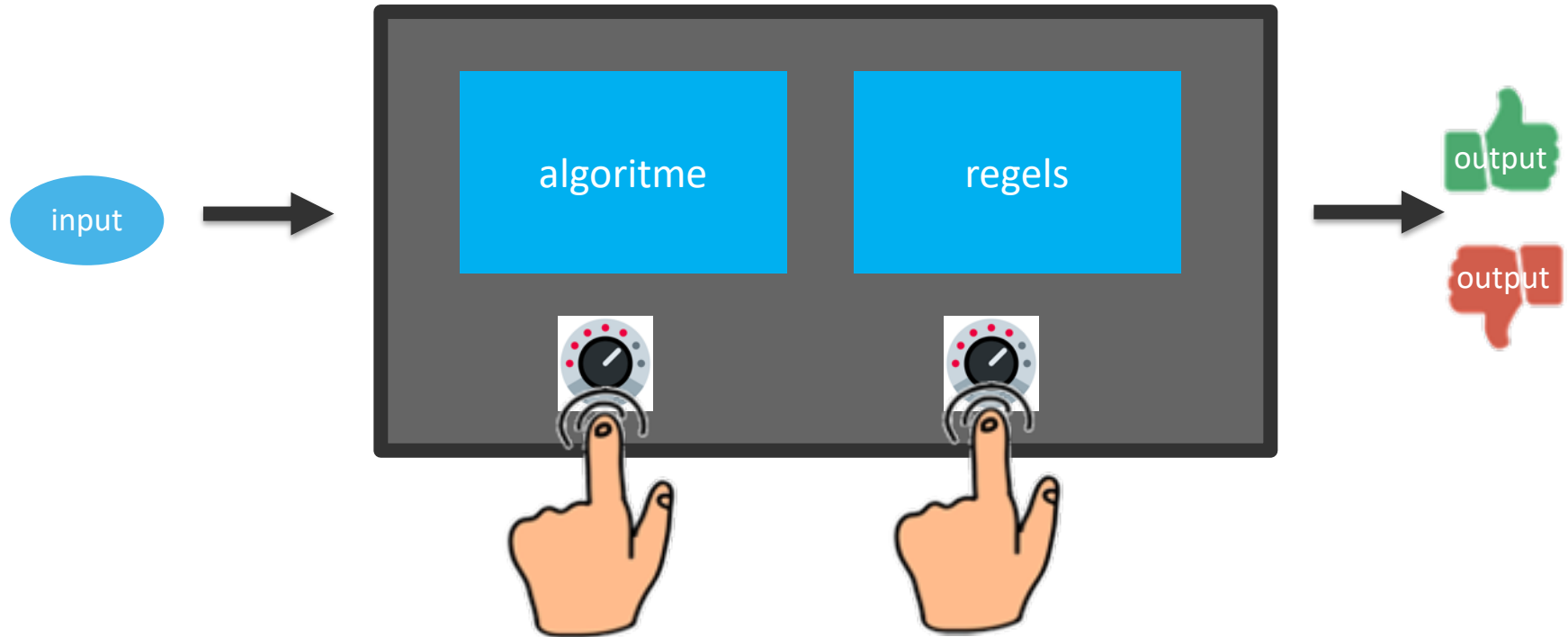


Bias in context

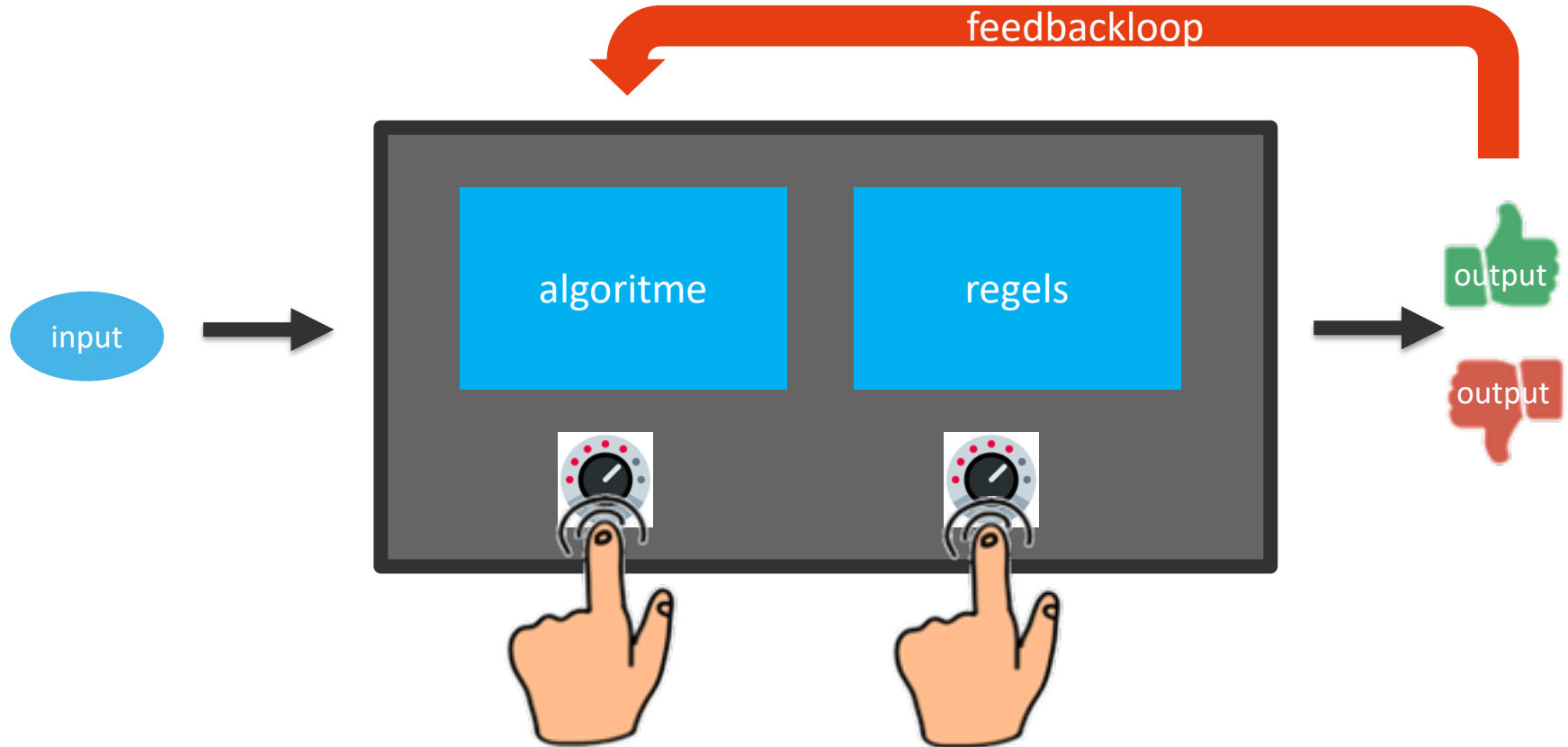
Algoritmisch systeem & Bias



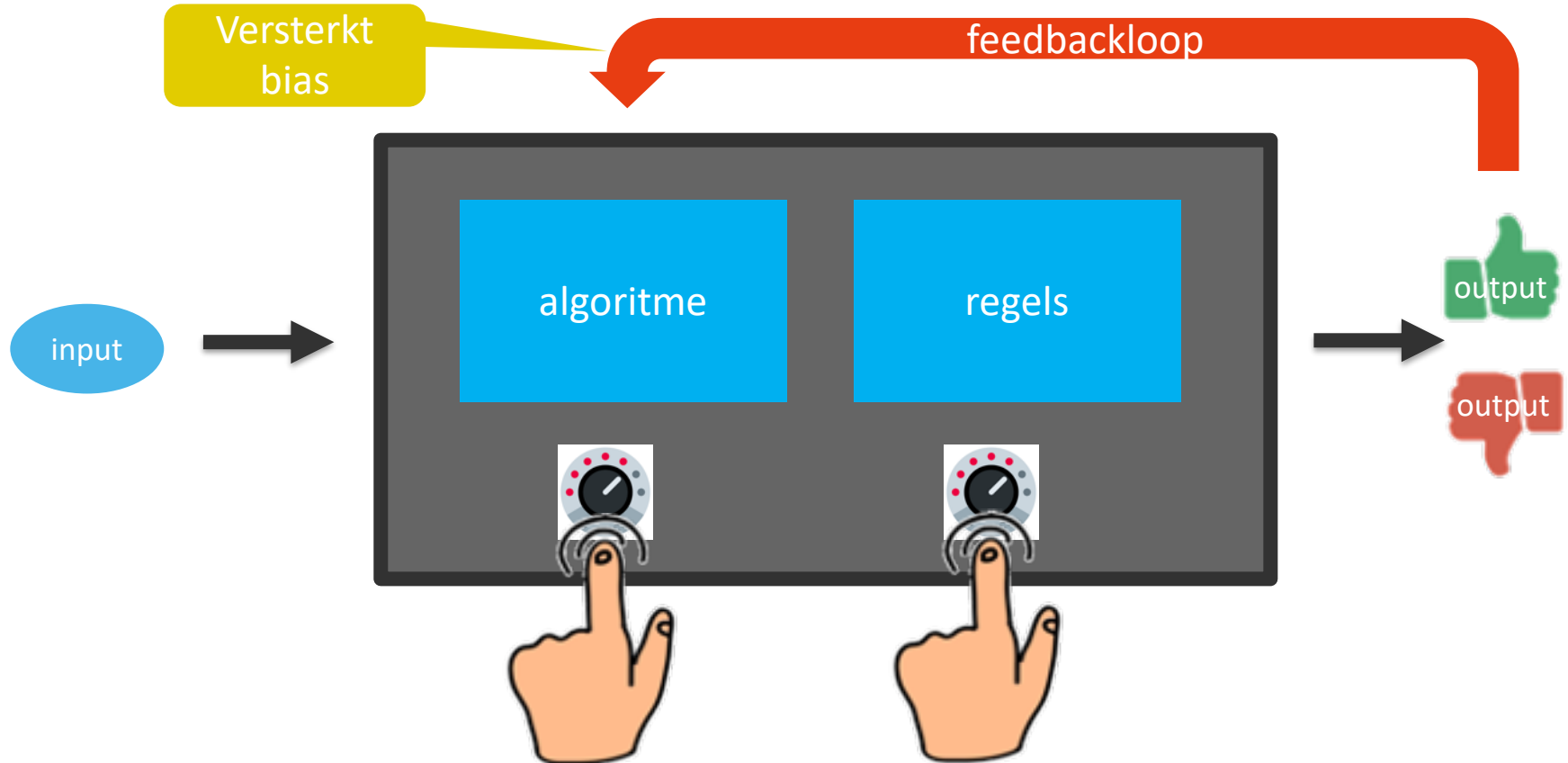
Algoritmisch systeem: algoritme + regels toepassing



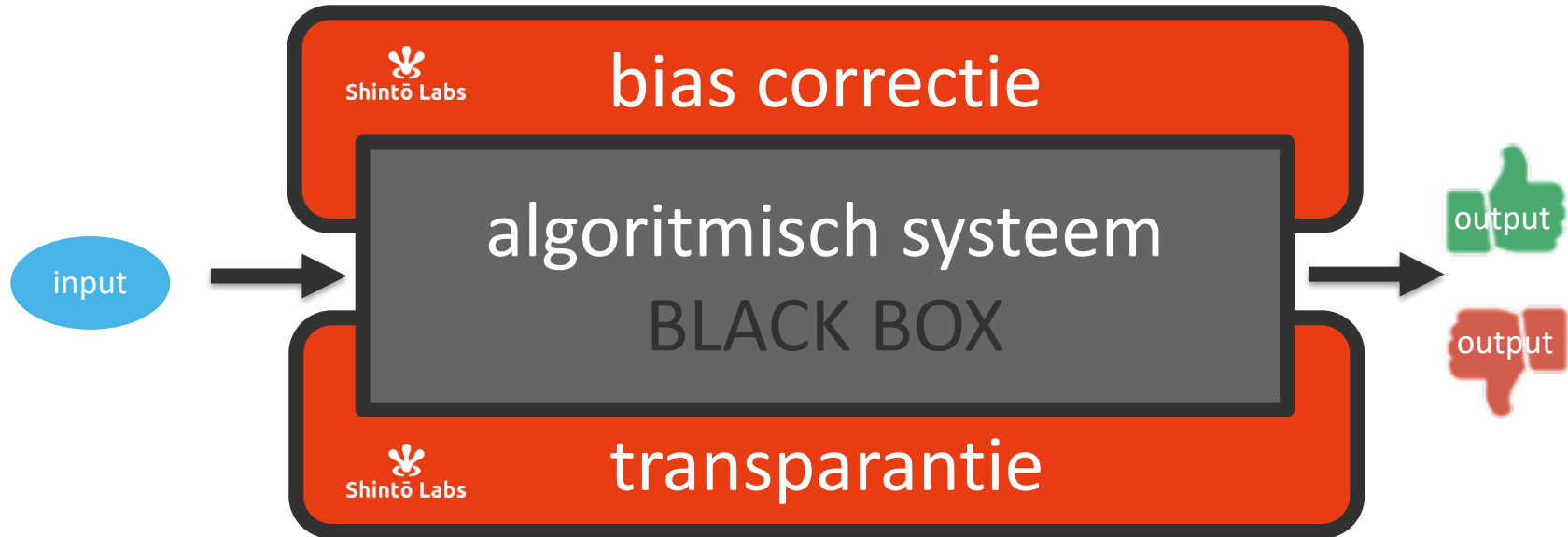
Algoritmisch systeem: algoritme + regels + feedbackloop



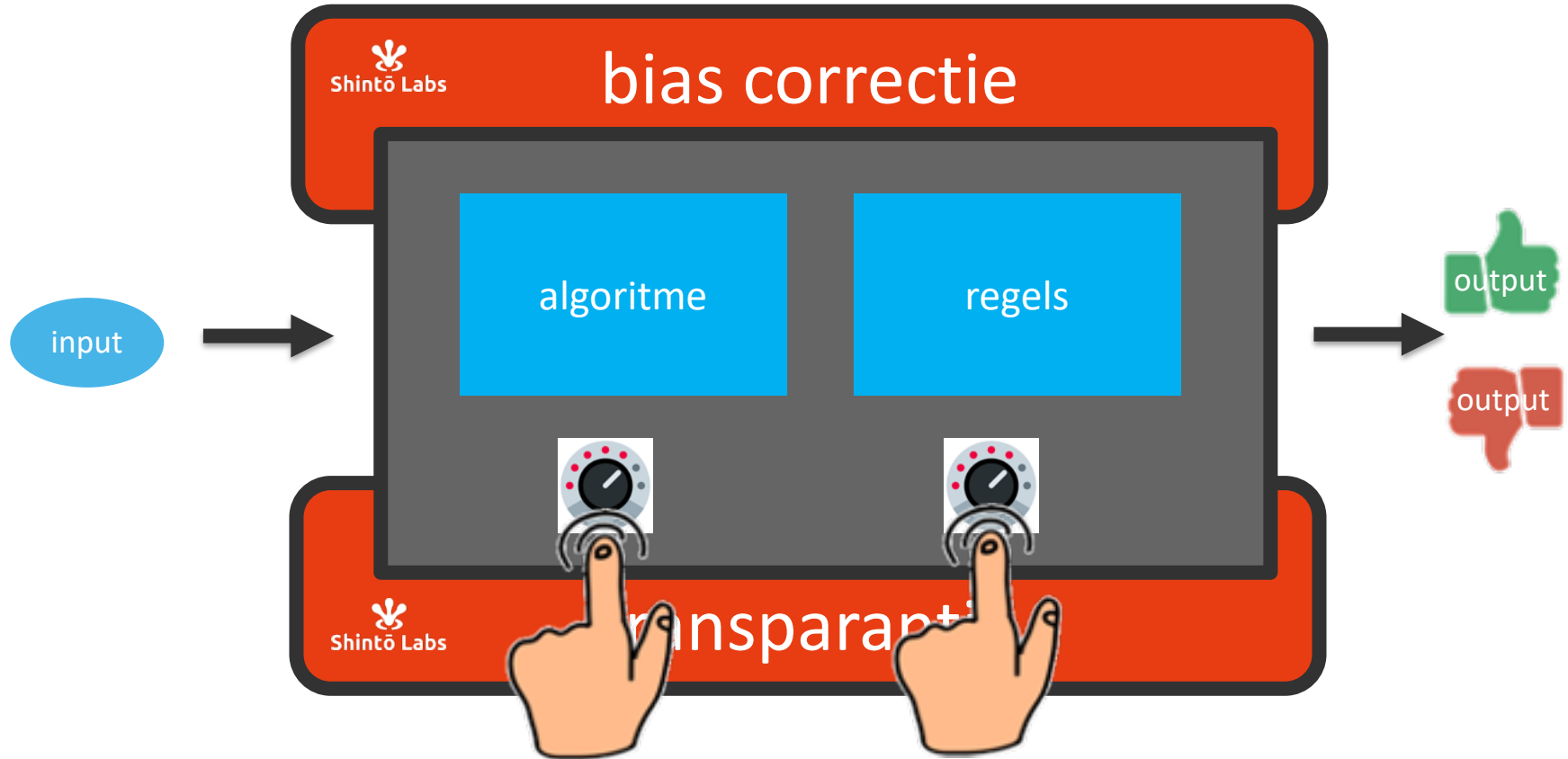
Algoritmisch systeem: algoritme + regels + feedbackloop



Algoritmisch systeem, bias correctie & transparantie



stelsel: algoritme + regels toepassing





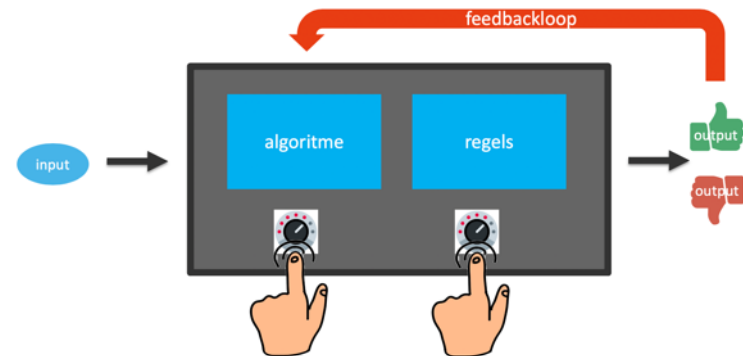
Onderscheid in Bias

Onderscheid in Bias

1. **Culturele bias** – op bv politiek of religie gebaseerde vooronderstelling of vooringenomenheid voor of tegen iets of iemand
 - Zit “verborgen” in de data

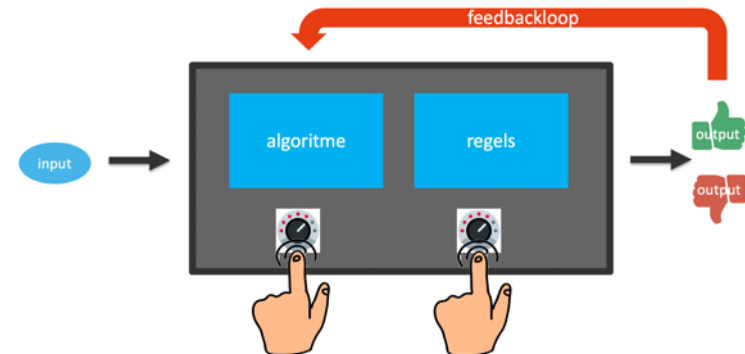
1. **Technische bias** – vertekening in het resultaat van een steekproef of algoritmisch systeem

- Algoritmische bias
- Bias in de gekozen regels
- Bias versterkende feedbackloop



Bias correctie

1. **Human In The Loop** – niet blind op het algoritmisch systeem vertrouwen. Gebruikers bepalen uiteindelijk.
2. **Transparantie algoritme & bias** – zodat gebruikers vertrouwen in het systeem hebben en de kennis hebben om aanpassingen te doen
3. **Bias correctie** – o.a. het “verborgen” versterkend effect van een feedback loop te corrigeren.

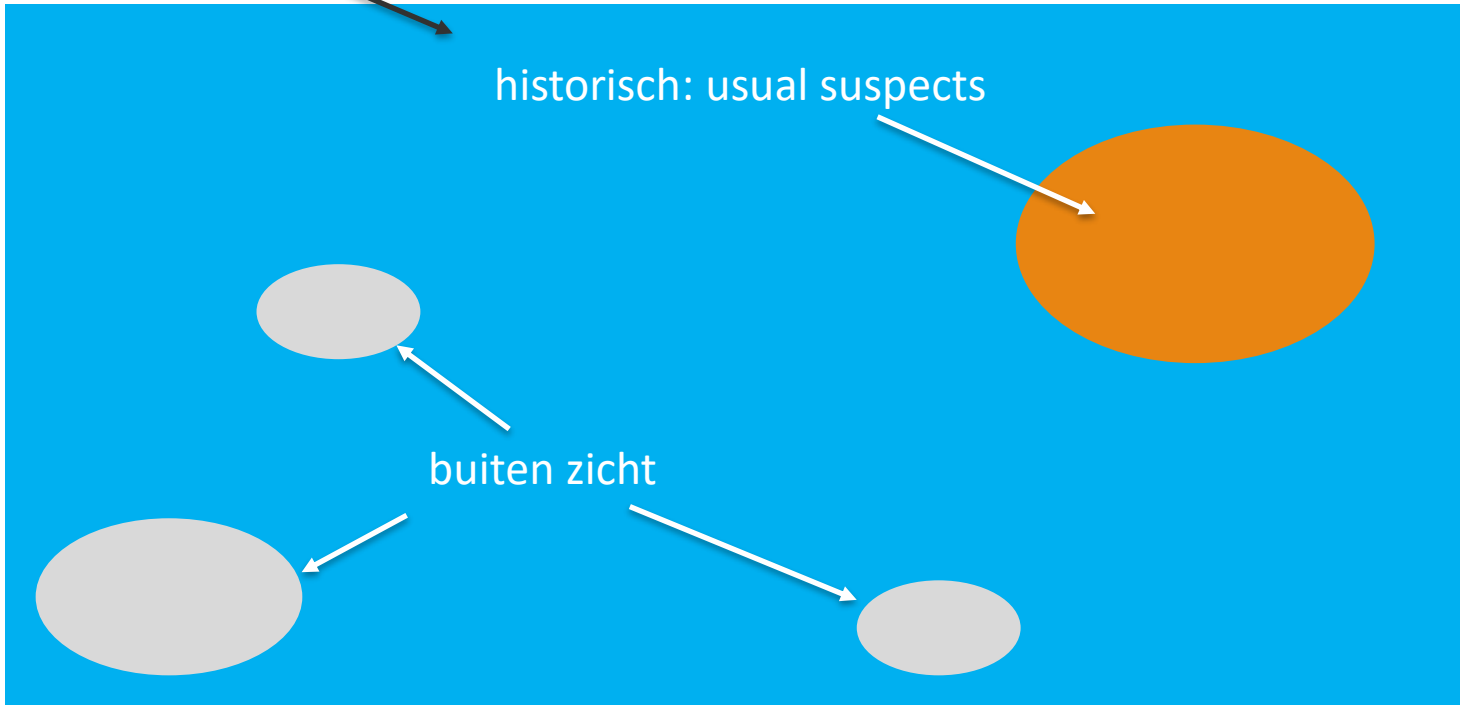




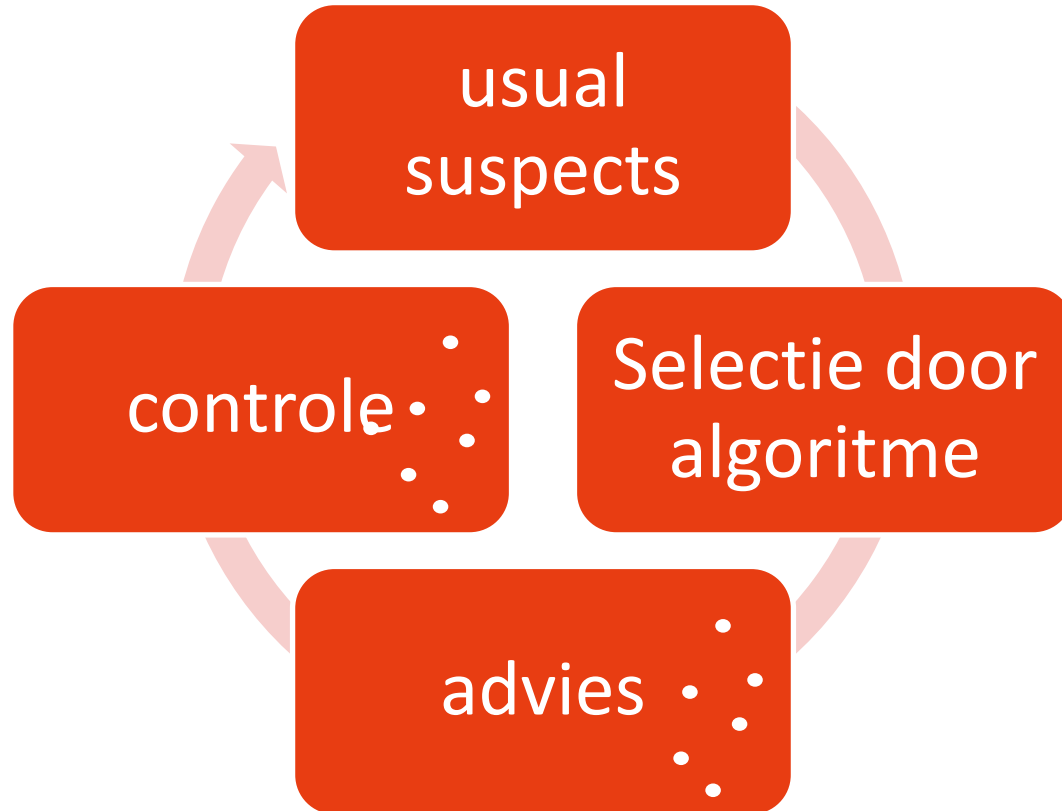
Feedbackloop & Bias

Historische bias

Hele populatie

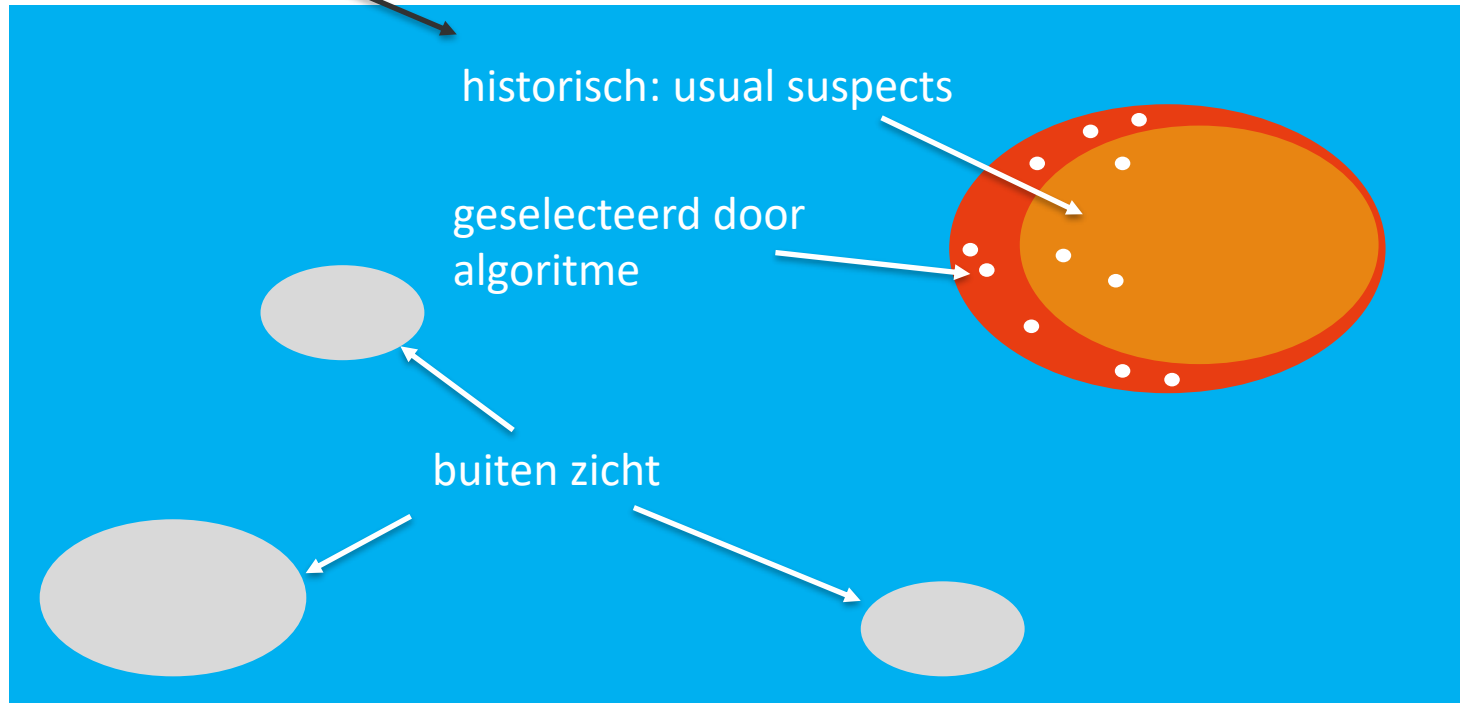


Bias versterkende feedbackloop



Bias bevestigende feedbackloop

Hele populatie



Corrigeren/bijsturen d.m.v. Intelligente sampling

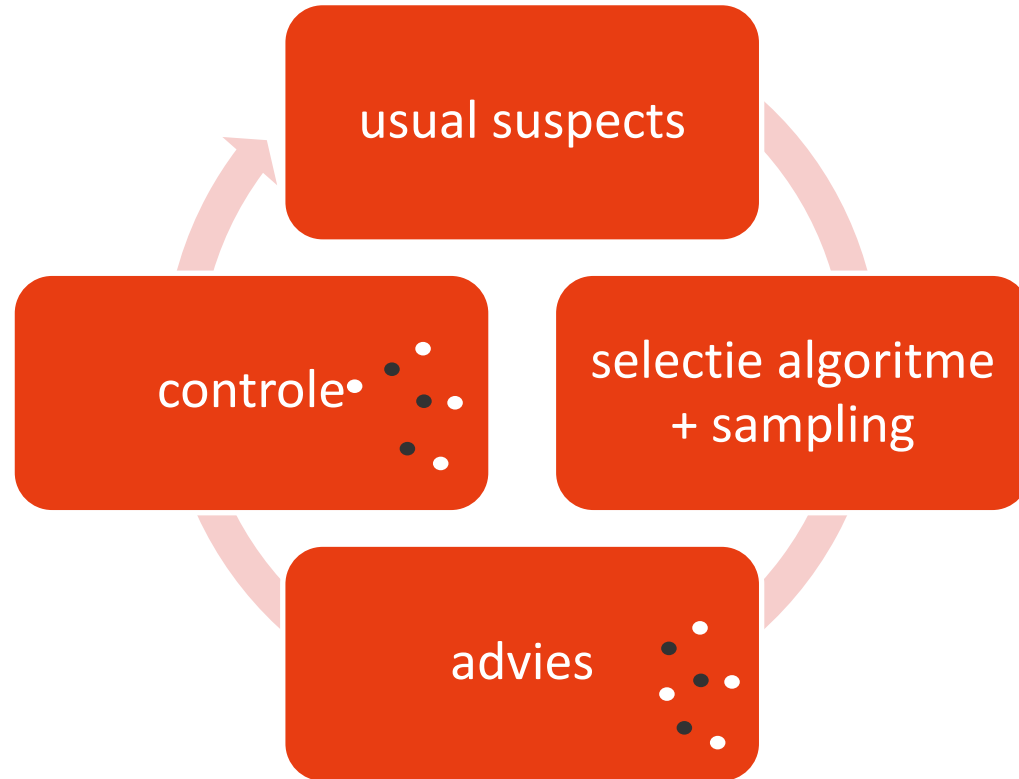
- In eerste instantie vooral random sampling verspreid over hele populatie
- Maar er zijn andere sampling mogelijkheden om eventueel sterker te sturen
- Na verloop van tijd zal de bias gecorrigeerd worden en een completer beeld van en een “eerlijkere” selectie uit de populatie ontstaan.

Feedbackloop correctie door (random) sampling

Hele populatie

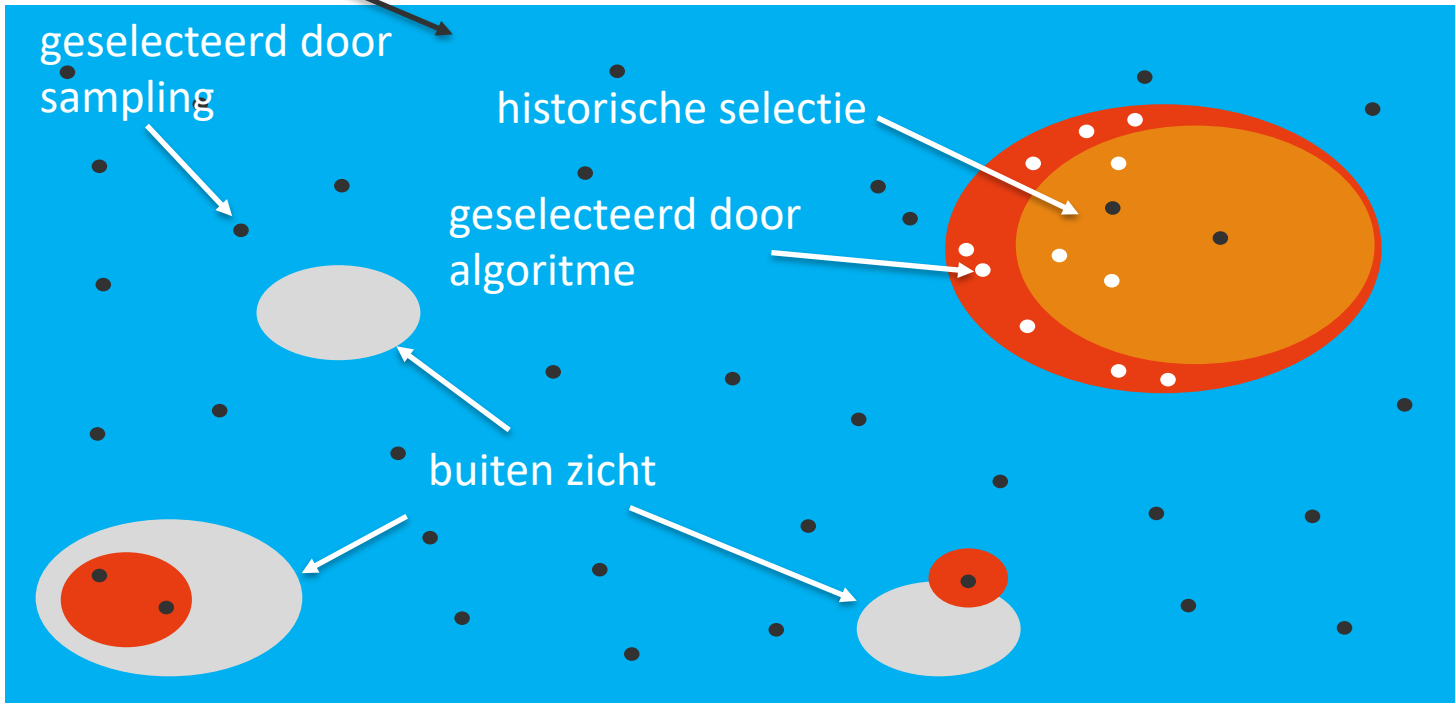


Feedbackloop + sampling

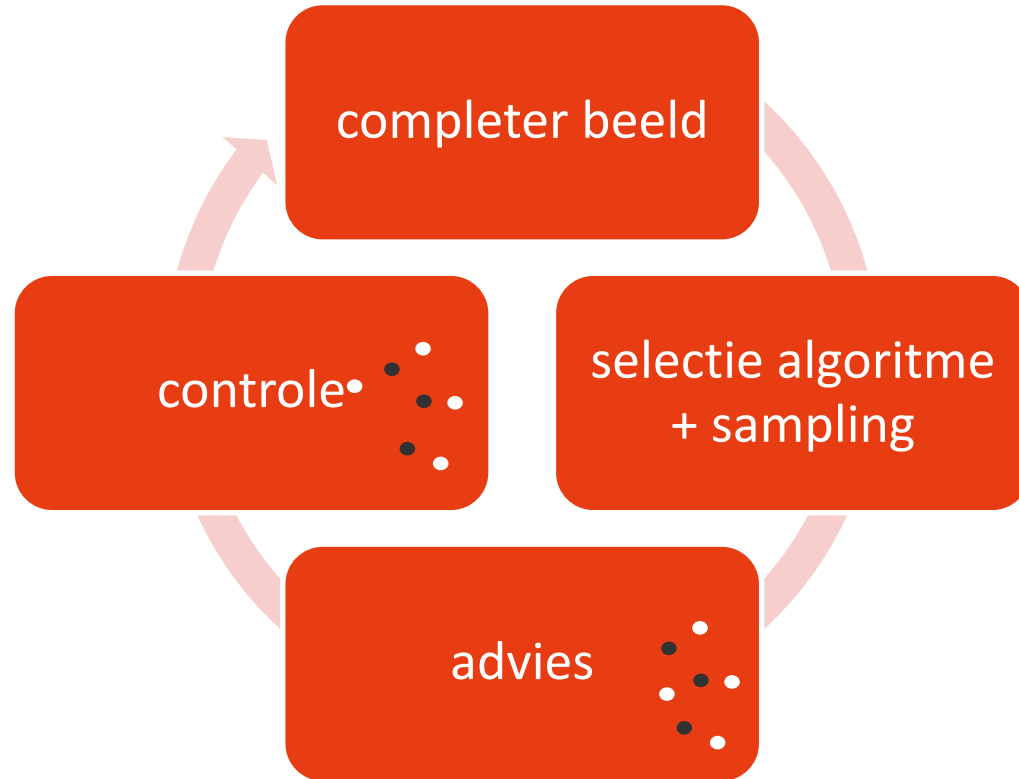


Feedbackloop correctie door (random) sampling

Hele populatie

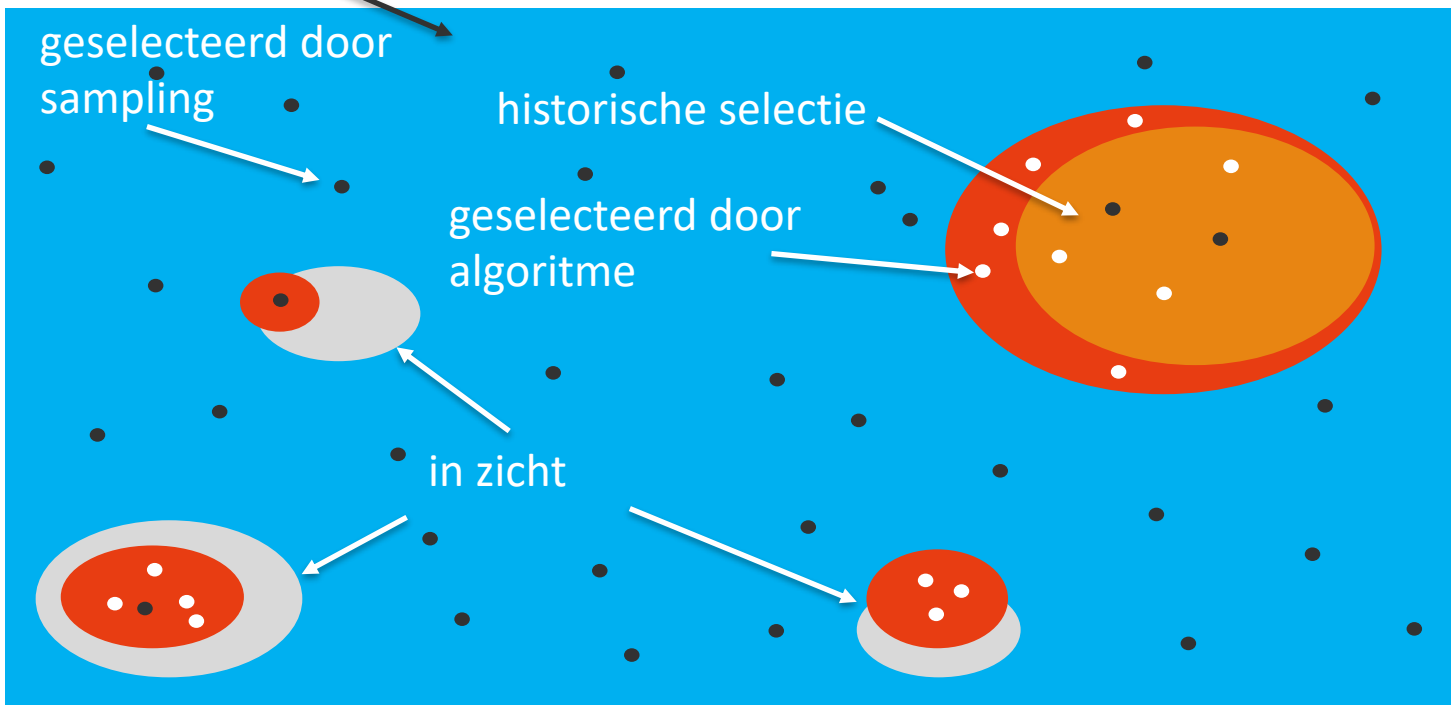


Feedbackloop + sampling => completer beeld



Feedbackloop + sampling => completer beeld

Hele populatie



Toepassing op het Stratum Analytics platform

- Milieunavigators – voorselectie van relevante branches en toekennen risicoindicator om onbekende bedrijven in beeld te krijgen.
- Risicoradar ondermijning (in ontwikkeling) – het vinden van nieuwe regio's en branches waar mogelijk ondermijning zou kunnen spelen



Meer weten?

info@shintolabs.nl

www.shintolabs.nl

<https://www.facebook.com/shintolabs>

<https://twitter.com/shintolabs>

06 - 21 89 55 78