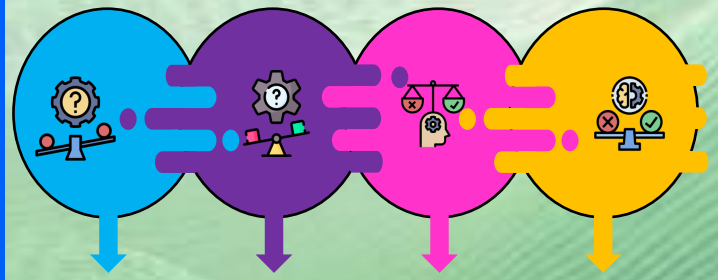
AI & Partners
Amsterdam - London - Singapore

Evaluating Bias in Algorithms (EDPB)

The AI: Complex Algorithms and effective Data Protection Supervision project clarifies bias evaluation tools for algorithms (e.g. social media).

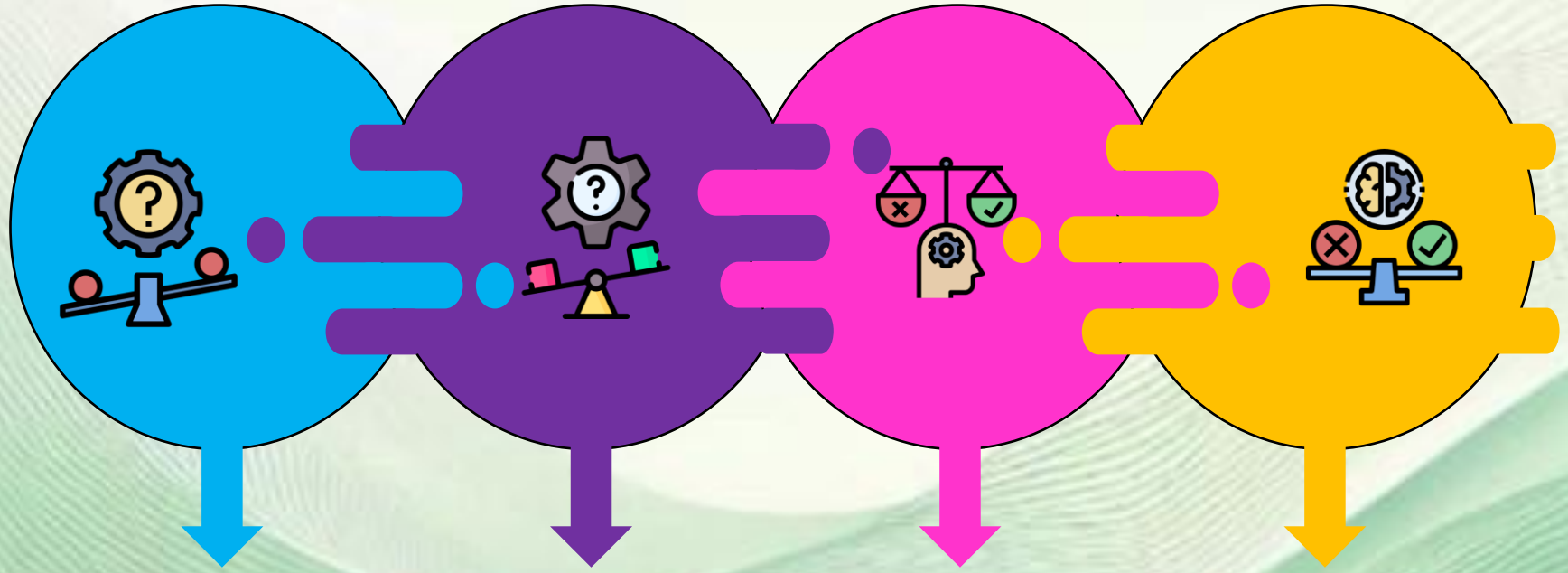
European Data Protection Board



What are the main sources of bias?
See **Slide 2** for details.

AI & Partners
Amsterdam - London - Singapore

Sources of bias:



Bias From Data (Historical)

When trained on historical data, AI systems reflect societal bias which are embedded in the dataset.

Algorithm Bias

Even when the datasets are not biased and are properly sampled, the algorithmic choices can contribute to biased decisions.

Evaluation Bias

Evaluation bias arises when assessments are made against a benchmark, or test dataset, because they could contribute to bias.

Bias in Facial Recognition Technology

Training and benchmark datasets are constructed from non-representative, publicly-available image datasets (e.g. web-scraping).

AI & Partners
Amsterdam - London - Singapore

Measures to address bias:

- **Pre-processing**: These methods adjust the training data before teaching the AI, making it harder to link sensitive factors to the results.
- **In-processing**: These methods add extra rules to the AI training process to reduce bias while the AI is learning.
- **Post-processing**: These methods try to fix bias in the AI model after it's already been trained.

What examples of bias reduction exist for algorithms?

Check the research report for more details.