

Responsible Data Science Workflows

Valentin Danchev @valdanchev



Building on FAIR principles & reproducible workflows to design Responsible Data Science Workflows embedding social & ethical considerations

WHAT A RESPONSIBLE
DATA SCIENCE WORKFLOW
WOULD LOOK LIKE?



REPRODUCIBLE REUSABLE



ENCODE COMMON

PATTERNS (OF ASSESSING

MODELS FOR SOCIAL &

ETHICAL RISKS) INTO

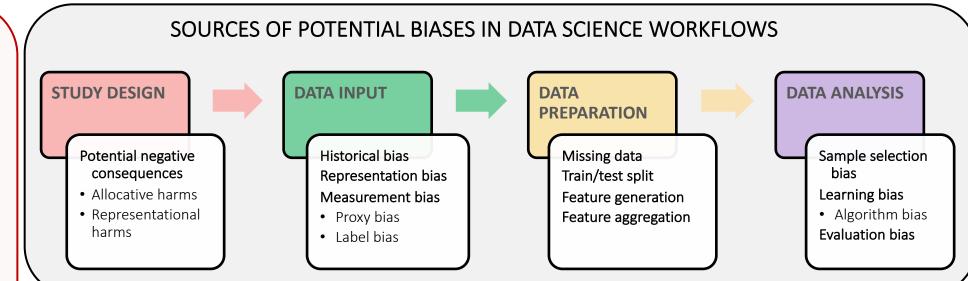
WORKFLOW MODULES



RESEARCHER-CENTRIC ≠
PIPELINE



MODULAR



BIAS MITIGATION & REPRODUCIBILITY

BIAS MITIGATION STRATEGIES
DEPEND ON THE SOURCE OF BIAS
Data augmentation
Fairness metrics
Domain knowledge
Causal inference

BIAS MITIGATION ADDS TO THE ARRAY OF CHOICES
IMPACTING REPRODUCIBILITY

Proliferation of fairness metrics: pros & cons

Proliferation of fairness metrics: pros & cor

Gaming fairness metrics (Goodhart's law)

(Assumed) fairness-accuracy trade-off

Researcher degrees of freedom

FAIR data & code Reproducible workflow

Pre-registration