motivations (1)... (1) motivation practical fairness & cognitive

intrinsic

inclusivity

Generalisation studies have various

synthetic sources (4). (3) data shift

covariate shift label shift full shift

(4) source naturally occurring shifts natural data splits found 'in the wild' curated splits on natural data e.g. different domains e.g. different lengths generated shifts fully generated splits generated evaluation data for generated training and natural training data evaluation data e.g. SCAN e.g. HANS (McCoy et al., 2019), (Lake and Baroni, 2018) or natural evaluation data for a generated training set

between all stages

compositional across task cross-lingual domain

structural

robustness

...and can be categorised into types (2).

These data shifts can occur in different stages of the modelling pipeline (5). pre-training from pre-training to training data training (5) locus from training to test data

from pretraining to test data

They involve data shifts (3), where the data can come from natural or