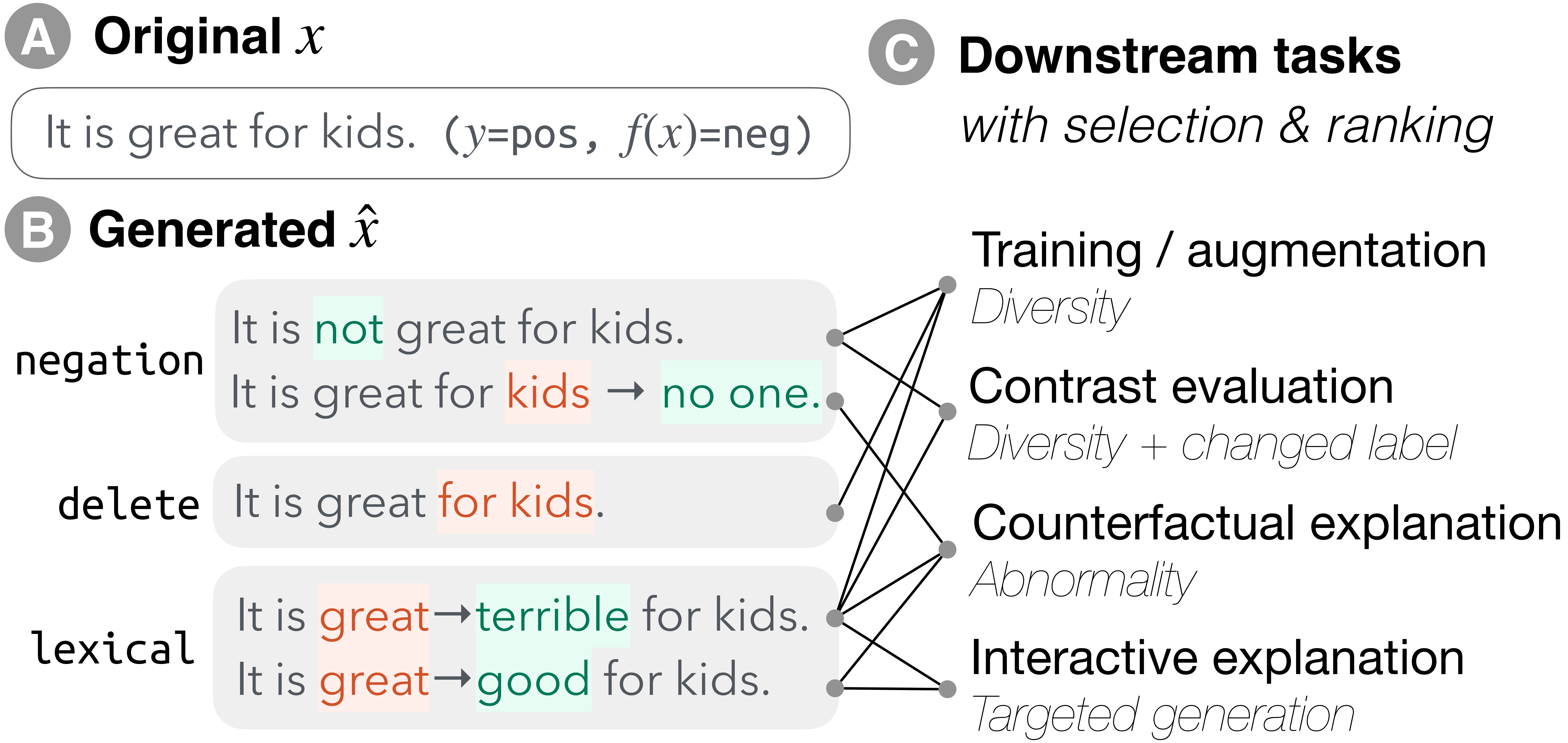

A **Original x**

It is great for kids. (y=pos, $f(x)$=neg)

B **Generated $\hat{x}$**

negation: It is not great for kids.

It is great for kids → no one.

delete: It is great for kids.

lexical: It is great→terrible for kids.

It is great→good for kids.

C **Downstream tasks**

with selection & ranking

Training / augmentation

Diversity

Contrast evaluation

Diversity + changed label

Counterfactual explanation

Abnormality

Interactive explanation

Targeted generation