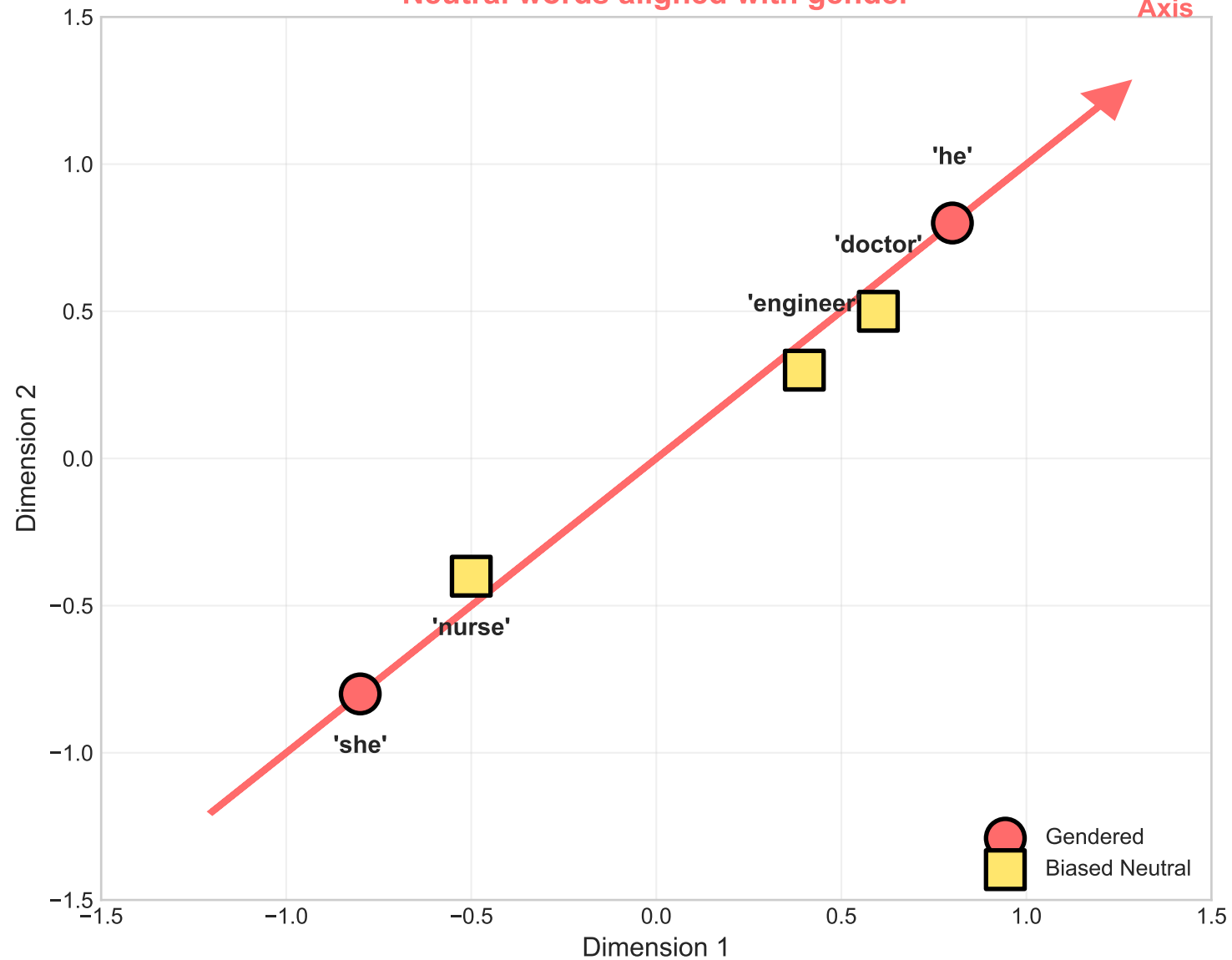


Word Embedding Debiasing: Removing Gender Bias via Projection

BEFORE Debiasing:
Neutral words aligned with gender

Gender
Axis



AFTER Debiasing:
Neutral words gender-neutral

Gender
Axis
(removed)

