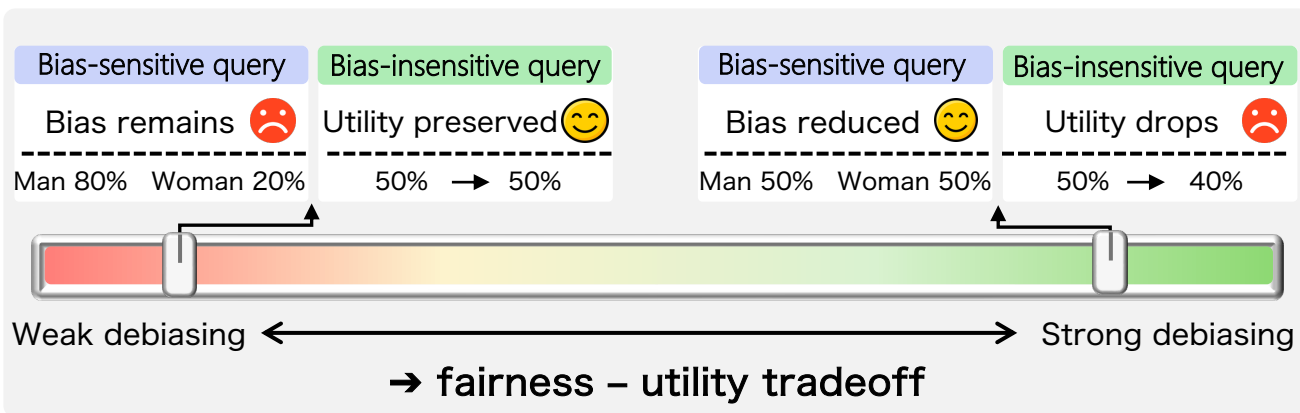
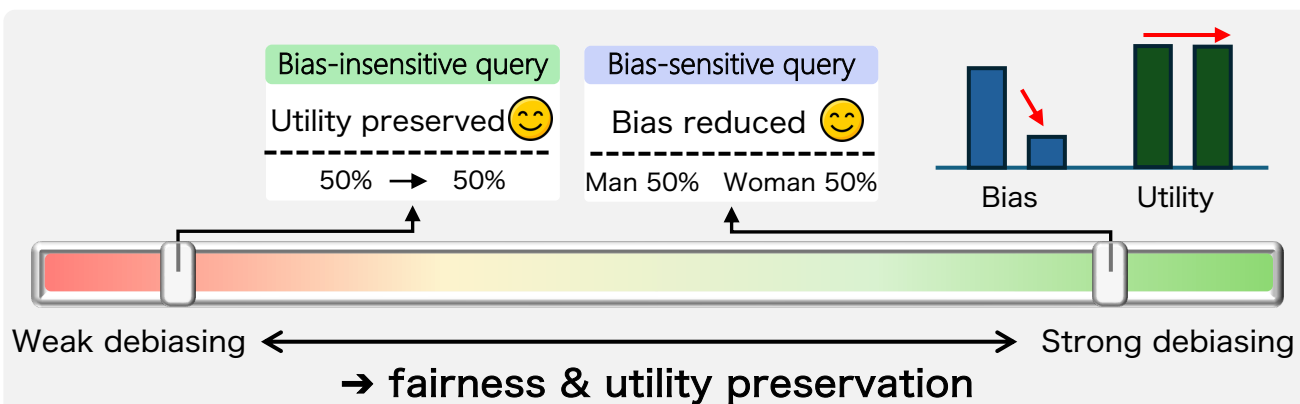




(a) Bias-sensitive and Bias-insensitive Inputs



(b) Existing methods: Uniform mitigation



(c) Ours: Selective mitigation via reward-gating for bias-sensitive inputs