Regression Analysis

```
survey_data <- read.csv('../../../backend/data/database/survey_data.csv')
survey_data$TreatmentGroup <- as.factor(survey_data$TreatmentGroup)
survey_data$TreatmentGroup <- relevel(survey_data$TreatmentGroup, ref = "machine")
#head(survey_data, n = 10)
```

General Result (all treatments)

```
m_general = glmer(TreatedIsLessPolar ~ TreatmentGroup + (1 | FK_ParticipantId),
             data=survey_data, family = "binomial")
summary(m_general)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TreatedIsLessPolar ~ TreatmentGroup + (1 | FK_ParticipantId)
##    Data: survey_data
##
##      AIC      BIC   logLik deviance df.resid
##    283.8    298.3   -137.9    275.8      272
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5470 -0.6059  0.2946  0.3926  1.6504
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  FK_ParticipantId (Intercept) 0.7319   0.8555
## Number of obs: 276, groups:  FK_ParticipantId, 69
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)             1.4946     0.3470   4.307 1.66e-05 ***
## TreatmentGrouphuman     0.7161     0.5045   1.420    0.156
## TreatmentGroupplacebo  -2.4424     0.4817  -5.071 3.96e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TrtmntGrph
## TrtmntGrphm -0.570
## TrtmntGrppl -0.770  0.403
```

```r
report(m_general)
```

```
## We fitted a logistic mixed model (estimated using ML and Nelder-Mead optimizer)
## to predict TreatedIsLessPolar with TreatmentGroup (formula: TreatedIsLessPolar
## ~ TreatmentGroup). The model included FK_ParticipantId as random effect
## (formula: ~1 | FK_ParticipantId). The model's total explanatory power is
## substantial (conditional R2 = 0.44) and the part related to the fixed effects
## alone (marginal R2) is of 0.31. The model's intercept, corresponding to
## TreatmentGroup = machine, is at 1.49 (95% CI [0.81, 2.17], p < .001). Within
## this model:
##
##   - The effect of TreatmentGroup [human] is statistically non-significant and
## positive (beta = 0.72, 95% CI [-0.27, 1.70], p = 0.156; Std. beta = 0.72, 95%
## CI [-0.27, 1.70])
##   - The effect of TreatmentGroup [placebo] is statistically significant and
## negative (beta = -2.44, 95% CI [-3.39, -1.50], p < .001; Std. beta = -2.44, 95%
## CI [-3.39, -1.50])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald z-distribution approximation.
```

Interpretation of Results

We fitted a **generalized linear mixed model** (estimated using maximum likelihood and Laplace approximation) to predict whether the treated text was perceived as less polarized (`TreatedIsLessPolar`) based on the `TreatmentGroup`. The model included `FK_ParticipantId` as a random effect. The results are summarized as follows:

Model Performance

- **AIC**: 283.8, indicating the relative quality of the model; lower values suggest a better fit.
- **BIC**: 298.3, providing a measure of fit penalized for the number of parameters.
- **Log-Likelihood**: -137.9, representing the likelihood of the observed data under the model.
- **Deviance**: 275.8, representing the likelihood of the observed data under the model.
- **Residual Degrees of Freedom**: 272
- The model's total explanatory power is substantial:
  - **Conditional R^2**: 0.44 (includes random effects)
  - **Marginal R^2**: 0.31 (fixed effects only)

Fixed Effects The fixed effects correspond to the treatment groups, with the reference category being `machine`. The intercept represents the log-odds of the treated text being perceived as less polarized for the `machine` group.

- **Intercept (machine)**:
  - Estimate: 1.49 (log-odds)
  - 95% CI: [0.81, 2.17]
  - **$z = 4.31$, $p < .001$**
  - This indicates that, for the `machine` group, there is a statistically significant positive log-odds of the treated text being perceived as less polarized.

- **TreatmentGroup [human]**:
  - Estimate: 0.72 (log-odds)
  - 95% CI: [-0.27, 1.70]
  - **z = 1.42, p = 0.156**
  - The effect of the `human` group is positive but not statistically significant, suggesting no clear evidence that the `human` group differs from the `machine` group in the perception of reduced polarization.
- **TreatmentGroup [placebo]**:
  - Estimate: -2.44 (log-odds)
  - 95% CI: [-3.39, -1.50]
  - **z = -5.07, p < .001**
  - The `placebo` group has a statistically significant and negative effect compared to the `machine` group, indicating that the treated text in the `placebo` group is much less likely to be perceived as less polarized.

Random Effects

- **Participant Variance**: 0.73 (Std. Dev: 0.86)
  - This suggests moderate variability in participants' responses.

Summary

- The `machine` group shows a significant positive log-odds of the treated text being perceived as less polarized.
- The `human` group does not significantly differ from the `machine` group in reducing perceived polarization.
- The `placebo` group is significantly less likely than the `machine` group to reduce perceived polarization.

Notes

- Standardized estimates and 95% confidence intervals were calculated using a Wald z-distribution approximation.
- Random effects account for individual participant variability, which enhances the model's explanatory power.

RQ1 Can LLMs mitigate textual polarization in social media texts?

Logistic Regression for mitigation effect.

```
machine_placebo <- subset(survey_data, TreatmentGroup %in% c("machine", "placebo"))
model_rq1 = glmer(TreatedIsLessPolar ~ TreatmentGroup + (1 | FK_ParticipantId),
              data=machine_placebo, family = "binomial")
summary(model_rq1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TreatedIsLessPolar ~ TreatmentGroup + (1 | FK_ParticipantId)
##    Data: machine_placebo
```

```
##
##      AIC      BIC   logLik deviance df.resid
##    216.0    225.7   -105.0    210.0      185
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9762 -0.6046  0.3865  0.5060  1.6539
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  FK_ParticipantId (Intercept) 0.7648   0.8746
## Number of obs: 188, groups:  FK_ParticipantId, 47
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)             1.5035     0.3569   4.213 2.52e-05 ***
## TreatmentGroupplacebo  -2.4556     0.4965  -4.946 7.59e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## TrtmntGrppl -0.779
```

```
report(model_rq1)
```

```
## We fitted a logistic mixed model (estimated using ML and Nelder-Mead optimizer)
## to predict TreatedIsLessPolar with TreatmentGroup (formula: TreatedIsLessPolar
## ~ TreatmentGroup). The model included FK_ParticipantId as random effect
## (formula: ~1 | FK_ParticipantId). The model's total explanatory power is
## substantial (conditional R2 = 0.41) and the part related to the fixed effects
## alone (marginal R2) is of 0.27. The model's intercept, corresponding to
## TreatmentGroup = machine, is at 1.50 (95% CI [0.80, 2.20], p < .001). Within
## this model:
##
##   - The effect of TreatmentGroup [placebo] is statistically significant and
## negative (beta = -2.46, 95% CI [-3.43, -1.48], p < .001; Std. beta = -2.46, 95%
## CI [-3.43, -1.48])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald z-distribution approximation.
```

Model Interpretation

Model Overview

- **Dependent Variable (`TreatedIsLessPolar`)**: A binary outcome (`0` or `1`), where `1` indicates that the treated text is perceived as less polarized than the original text.
- **Predictor (`TreatmentGroup`)**: Two groups (machine paraphrasing as the reference category, placebo).
- **Random Effects**:
  - A random intercept for each participant (`FK_ParticipantId`) accounts for individual variability in perceptions of polarization.

Key Metrics

1. **AIC and BIC**:
   - **AIC**: `216.0` and **BIC**: `225.7`. Lower values indicate better model fit when compared to alternative models.
2. **Log-Likelihood**: `-105.0`. Higher (less negative) values indicate better model fit.
3. **Deviance**: `210.0`. Lower values indicate a better fit.
4. **R^2 Values**:
   - **Conditional R^2**: `0.41`, which represents the variance explained by both fixed and random effects.
   - **Marginal R^2**: `0.27`, which represents the variance explained by the fixed effects alone.

Random Effects

- **Variance of Participant-Level Random Intercept**: `0.7648`, with a standard deviation of `0.8746`.
  - This indicates substantial variability in participants' baseline perceptions of polarization.

Fixed Effects

1. Intercept:

- **Estimate**: `1.5035`
- **Interpretation**: When the treatment group is **machine paraphrasing** (the reference category), the log-odds of the treated text being perceived as less polarized are **1.50**.
- **Probability**: This corresponds to a probability of about **81.8%** (`plogis(1.5035)`).
- **Significance**: The intercept is highly significant (`p < 0.001`).

2. `TreatmentGroupplacebo`:

- **Estimate**: `-2.4556`
- **Interpretation**: Compared to **machine paraphrasing**, the log-odds of the treated text being perceived as less polarized decrease significantly by **2.46** when the treatment is **placebo**.
- **Probability**: This corresponds to a probability of about **18.8%** (`plogis(1.5035 - 2.4556)`).
- **Significance**: The effect is highly significant (`p < 0.001`), indicating that the placebo treatment is substantially less effective than machine paraphrasing.

Confidence Intervals

- The **95% Confidence Interval** for each fixed effect provides the range of plausible values for the parameter estimates:
  - `Intercept: [0.80, 2.20]` ($p < 0.001$) – very strong evidence for the baseline probability.
  - `TreatmentGroupplacebo: [-3.43, -1.48]` ($p < 0.001$) – does not include zero, indicating a robust negative effect.

Correlation of Fixed Effects

- The correlation between the intercept and `TreatmentGroupplacebo` is `-0.779`, indicating a moderate negative relationship between these parameters.

Summary of Findings

1. **Treatment Effectiveness**:
   - **Machine paraphrasing** (LLM) is highly effective at mitigating perceived polarization, with an estimated probability of **81.8%** for the treated text being seen as less polarized than the original text.
   - **Placebo** is significantly less effective than machine paraphrasing, with a much lower probability of **18.8%**.
2. **Participant-Level Variability**:
   - There is substantial variability in how participants perceive the polarization of treated texts, as shown by the random intercept variance.
3. **Model Fit**:
   - Fixed effects explain **27%** of the variance (marginal R^2), and the full model explains **41%** (conditional R^2), indicating good explanatory power.

RQ2 Can LLMs significantly reduce perceived polarization in social media texts?

```
#model_rq2 <- lmer(DiffLikertTreatedOriginal ~ TreatmentGroup + TweetBias * ParticipantLeaning +
#                   (1 | FK_ParticipantId),
#                   data = machine_placebo)
model_rq2 <- lmer(DiffLikertTreatedOriginal ~ TreatmentGroup +(1 | FK_ParticipantId),
                  data = machine_placebo)
summary(model_rq2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: DiffLikertTreatedOriginal ~ TreatmentGroup + (1 | FK_ParticipantId)
##    Data: machine_placebo
```

```
##
## REML criterion at convergence: 591.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.37662 -0.49618  0.08537  0.44404  2.71298
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  FK_ParticipantId (Intercept) 0.2927   0.5411
##  Residual                     1.1312   1.0636
## Number of obs: 188, groups:  FK_ParticipantId, 47
##
## Fixed effects:
##                       Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)            -1.8021     0.1549 45.0000 -11.637 3.66e-15 ***
## TreatmentGroupplacebo   1.6173     0.2214 45.0000   7.306 3.59e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## TrtmntGrppl -0.700
```

```
report(model_rq2)
```

```
## We fitted a linear mixed model (estimated using REML and nloptwrap optimizer)
## to predict DiffLikertTreatedOriginal with TreatmentGroup (formula:
## DiffLikertTreatedOriginal ~ TreatmentGroup). The model included
## FK_ParticipantId as random effect (formula: ~1 | FK_ParticipantId). The model's
## total explanatory power is substantial (conditional R2 = 0.46) and the part
## related to the fixed effects alone (marginal R2) is of 0.32. The model's
## intercept, corresponding to TreatmentGroup = machine, is at -1.80 (95% CI
## [-2.11, -1.50], t(184) = -11.64, p < .001). Within this model:
##
##   - The effect of TreatmentGroup [placebo] is statistically significant and
## positive (beta = 1.62, 95% CI [1.18, 2.05], t(184) = 7.31, p < .001; Std. beta
## = 1.13, 95% CI [0.82, 1.43])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation.
```

Model Interpretation

Model Overview

- **Dependent Variable (`DiffLikertTreatedOriginal`)**: The difference in polarization scores between the treated and original texts, measured on a Likert scale.
- **Predictor (`TreatmentGroup`)**: Two groups (machine paraphrasing as the reference category, placebo).
- **Random Effects**:
  - A random intercept for each participant (`FK_ParticipantId`) accounts for individual variability in score differences.

Key Metrics

1. **REML Criterion**: `591.8`. A lower REML value suggests a better model fit when comparing similar models.
2. **Residual Standard Deviation**: `1.06`, indicating the average deviation of observed values from predicted values after accounting for fixed and random effects.
3. **R^2 Values**:
   - **Conditional R^2**: `0.46`, representing the variance explained by both fixed and random effects.
   - **Marginal R^2**: `0.32`, representing the variance explained by the fixed effects alone.

Random Effects

- **Variance of Participant-Level Random Intercept**: `0.2927`, with a standard deviation of `0.5411`.
  - This suggests moderate variability in participants' baseline differences in polarization scores.
- **Residual Variance**: `1.1312`, with a standard deviation of `1.0636`.

Fixed Effects

1. Intercept:

- **Estimate**: `-1.8021`
- **Interpretation**: When the treatment group is **machine paraphrasing** (the reference category), the mean difference in Likert scale polarization scores is **-1.80**.
  - This negative value indicates that machine paraphrasing significantly reduces polarization scores compared to the original texts.
- **Significance**: Highly significant (`p < 0.001`).

2. `TreatmentGroupplacebo`:

- **Estimate**: `1.6173`
- **Interpretation**: Compared to **machine paraphrasing**, the mean difference in polarization scores increases by **1.62** when the treatment is **placebo**.
  - This positive value suggests that placebo treatment results in less reduction (or even an increase) in polarization compared to machine paraphrasing.
- **Significance**: Highly significant (`p < 0.001`).

Confidence Intervals

- The **95% Confidence Interval** for each fixed effect provides the range of plausible values for the parameter estimates:
  - `Intercept: [-2.11, -1.50]` – consistently negative, indicating a robust reduction in polarization scores for machine paraphrasing.
  - `TreatmentGroupplacebo: [1.18, 2.05]` – consistently positive, confirming placebo's relative ineffectiveness compared to machine paraphrasing.

Correlation of Fixed Effects

- The correlation between the intercept and `TreatmentGroupplacebo` is `-0.700`, indicating a moderate negative relationship.

Summary of Findings

1. **Effectiveness of Treatments**:
   - **Machine paraphrasing** significantly reduces polarization scores, with a mean reduction of **1.80 points** on the Likert scale.
   - **Placebo** results in significantly less reduction (and possibly an increase) in polarization scores compared to machine paraphrasing, with a mean increase of **1.62 points** relative to the machine treatment.
2. **Participant-Level Variability**:
   - There is moderate variability in baseline score differences across participants, as indicated by the random effects.
3. **Model Fit**:
   - Fixed effects explain **32%** of the variance in polarization score differences (marginal R^2), while the full model explains **46%** (conditional R^2), suggesting good explanatory power.

RQ3 Can LLMs mitigate textual polarization as good as humans?

```
# Compare LLM vs. Human
machine_human <- subset(survey_data, TreatmentGroup %in% c("machine", "human"))
model_rq3 <- lmer(DiffLikertTreatedOriginal ~ TreatmentGroup + (1 | FK_ParticipantId),
                  data = machine_human)
summary(model_rq3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: DiffLikertTreatedOriginal ~ TreatmentGroup + (1 | FK_ParticipantId)
##    Data: machine_human
```

```
##
## REML criterion at convergence: 671.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7598 -0.7454 -0.0983  0.6117  3.9197
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  FK_ParticipantId (Intercept) 0.208    0.4561
##  Residual                     2.056    1.4339
## Number of obs: 184, groups:  FK_ParticipantId, 46
##
## Fixed effects:
##                     Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)          -1.8021     0.1735 44.0000 -10.390 2.03e-13 ***
## TreatmentGrouphuman   0.2339     0.2508 44.0000   0.933    0.356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## TrtmntGrphm -0.692
```

```
report(model_rq3)
```

```
## We fitted a linear mixed model (estimated using REML and nloptwrap optimizer)
## to predict DiffLikertTreatedOriginal with TreatmentGroup (formula:
## DiffLikertTreatedOriginal ~ TreatmentGroup). The model included
## FK_ParticipantId as random effect (formula: ~1 | FK_ParticipantId). The model's
## total explanatory power is weak (conditional R2 = 0.10) and the part related to
## the fixed effects alone (marginal R2) is of 6.03e-03. The model's intercept,
## corresponding to TreatmentGroup = machine, is at -1.80 (95% CI [-2.14, -1.46],
## t(180) = -10.39, p < .001). Within this model:
##
##   - The effect of TreatmentGroup [human] is statistically non-significant and
## positive (beta = 0.23, 95% CI [-0.26, 0.73], t(180) = 0.93, p = 0.352; Std.
## beta = 0.16, 95% CI [-0.17, 0.48])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation.
```

Model Interpretation

Model Overview

- **Dependent Variable (`DiffLikertTreatedOriginal`)**: The difference in polarization scores between the treated and original texts, measured on a Likert scale.
- **Predictor (`TreatmentGroup`)**: Two groups (machine paraphrasing as the reference category, human paraphrasing).
- **Random Effects**:

- A random intercept for each participant (`FK_ParticipantId`) accounts for individual variability in score differences.

Key Metrics

1. **REML Criterion**: `671.7`. A lower REML value suggests a better model fit when comparing similar models.
2. **Residual Standard Deviation**: `1.43`, indicating the average deviation of observed values from predicted values after accounting for fixed and random effects.
3. **R^2 Values**:
   - **Conditional R^2**: `0.10`, representing the variance explained by both fixed and random effects.
   - **Marginal R^2**: `0.006`, representing the variance explained by the fixed effects alone.

Random Effects

- **Variance of Participant-Level Random Intercept**: `0.208`, with a standard deviation of `0.4561`.
  - This suggests low variability in participants' baseline differences in polarization scores.
- **Residual Variance**: `2.056`, with a standard deviation of `1.4339`.

Fixed Effects

1. Intercept:

- **Estimate**: `-1.8021`
- **Interpretation**: When the treatment group is **machine paraphrasing** (the reference category), the mean difference in Likert scale polarization scores is **-1.80**.
  - This negative value indicates that machine paraphrasing significantly reduces polarization scores compared to the original texts.
- **Significance**: Highly significant (`p < 0.001`).

2. `TreatmentGrouphuman`:

- **Estimate**: `0.2339`
- **Interpretation**: Compared to **machine paraphrasing**, the mean difference in polarization scores increases slightly (by **0.23**) when the treatment is **human paraphrasing**.
  - However, this effect is **not statistically significant** (`p = 0.356`), suggesting no meaningful difference between the effects of human and machine paraphrasing.

Confidence Intervals

- The **95% Confidence Interval** for each fixed effect provides the range of plausible values for the parameter estimates:
  - `Intercept`: `[-2.14, -1.46]` – consistently negative, indicating a robust reduction in polarization scores for machine paraphrasing.
  - `TreatmentGrouphuman`: `[-0.26, 0.73]` – includes zero, confirming the non-significance of the effect.

Correlation of Fixed Effects

- The correlation between the intercept and `TreatmentGrouphuman` is `-0.692`, indicating a moderate negative relationship.

Summary of Findings

1. **Effectiveness of Treatments**:
   - **Machine paraphrasing** significantly reduces polarization scores, with a mean reduction of **1.80 points** on the Likert scale.
   - **Human paraphrasing** shows a slightly lesser reduction in polarization scores compared to machine paraphrasing, but the difference (**0.23 points**) is not statistically significant.
2. **Participant-Level Variability**:
   - There is low variability in baseline score differences across participants, as indicated by the random effects.
3. **Model Fit**:
   - Fixed effects explain only **0.6%** of the variance in polarization score differences (marginal R^2), while the full model explains **10%** (conditional R^2), suggesting weak explanatory power.

RQ4 Does political bias influence the participants' perception of textual polarization?

```
model_rq4 <- lmer(OriginalLikertValue ~ TweetBias * ParticipantLeaning +
                  (1 | FK_ParticipantId), data = survey_data)
summary(model_rq4)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## OriginalLikertValue ~ TweetBias * ParticipantLeaning + (1 | FK_ParticipantId)
##    Data: survey_data
```

```
##
## REML criterion at convergence: 744.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6496 -0.4425  0.3401  0.6775  1.9901
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  FK_ParticipantId (Intercept) 0.1107   0.3327
##  Residual                     0.7971   0.8928
## Number of obs: 276, groups:  FK_ParticipantId, 69
##
## Fixed effects:
##                                             Estimate Std. Error        df
## (Intercept)                                  4.25000    0.29133 146.13450
## TweetBiasRight                              -0.25000    0.36449 199.00000
## ParticipantLeaningcenter-left                0.14583    0.32572 146.13450
## ParticipantLeaningcenter-right               0.13889    0.37611 146.13450
## ParticipantLeaningfar-left                  -1.25000    0.58266 146.13450
## ParticipantLeaningfar-right                  0.75000    0.77079 146.13450
## ParticipantLeaningleft                      -0.07353    0.33886 146.13450
## ParticipantLeaningnot informed               0.15000    0.43211 146.13450
## ParticipantLeaningright                     -0.35000    0.43211 146.13450
## TweetBiasRight:ParticipantLeaningcenter-left  0.31250    0.40751 199.00000
## TweetBiasRight:ParticipantLeaningcenter-right -0.25000    0.47056 199.00000
## TweetBiasRight:ParticipantLeaningfar-left     2.00000    0.72898 199.00000
## TweetBiasRight:ParticipantLeaningfar-right   -1.75000    0.96435 199.00000
## TweetBiasRight:ParticipantLeaningleft         0.42647    0.42396 199.00000
## TweetBiasRight:ParticipantLeaningnot informed -0.25000    0.54063 199.00000
## TweetBiasRight:ParticipantLeaningright        0.25000    0.54063 199.00000
##                                             t value Pr(>|t|)
## (Intercept)                                  14.588  < 2e-16 ***
## TweetBiasRight                               -0.686  0.49358
## ParticipantLeaningcenter-left                 0.448  0.65501
## ParticipantLeaningcenter-right                0.369  0.71245
## ParticipantLeaningfar-left                   -2.145  0.03358 *
## ParticipantLeaningfar-right                   0.973  0.33215
## ParticipantLeaningleft                       -0.217  0.82852
## ParticipantLeaningnot informed                0.347  0.72899
## ParticipantLeaningright                      -0.810  0.41927
## TweetBiasRight:ParticipantLeaningcenter-left  0.767  0.44408
## TweetBiasRight:ParticipantLeaningcenter-right -0.531  0.59581
## TweetBiasRight:ParticipantLeaningfar-left     2.744  0.00663 **
## TweetBiasRight:ParticipantLeaningfar-right   -1.815  0.07108 .
## TweetBiasRight:ParticipantLeaningleft         1.006  0.31568
## TweetBiasRight:ParticipantLeaningnot informed -0.462  0.64428
## TweetBiasRight:ParticipantLeaningright        0.462  0.64428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 16 > 12.
## Use print(x, correlation=TRUE)  or
```

```
##     vcov(x)        if you need it
```

```
report(model_rq4)
```

```
## We fitted a linear mixed model (estimated using REML and nloptwrap optimizer)
## to predict OriginalLikertValue with TweetBias and ParticipantLeaning (formula:
## OriginalLikertValue ~ TweetBias * ParticipantLeaning). The model included
## FK_ParticipantId as random effect (formula: ~1 | FK_ParticipantId). The model's
## total explanatory power is moderate (conditional R2 = 0.19) and the part
## related to the fixed effects alone (marginal R2) is of 0.08. The model's
## intercept, corresponding to TweetBias = Left and ParticipantLeaning = center,
## is at 4.25 (95% CI [3.68, 4.82], t(258) = 14.59, p < .001). Within this model:
##
##   - The effect of TweetBias [Right] is statistically non-significant and negative
## (beta = -0.25, 95% CI [-0.97, 0.47], t(258) = -0.69, p = 0.493; Std. beta =
## -0.26, 95% CI [-1.00, 0.48])
##   - The effect of ParticipantLeaning [center-left] is statistically
## non-significant and positive (beta = 0.15, 95% CI [-0.50, 0.79], t(258) = 0.45,
## p = 0.655; Std. beta = 0.15, 95% CI [-0.51, 0.81])
##   - The effect of ParticipantLeaning [center-right] is statistically
## non-significant and positive (beta = 0.14, 95% CI [-0.60, 0.88], t(258) = 0.37,
## p = 0.712; Std. beta = 0.14, 95% CI [-0.62, 0.91])
##   - The effect of ParticipantLeaning [far-left] is statistically significant and
## negative (beta = -1.25, 95% CI [-2.40, -0.10], t(258) = -2.15, p = 0.033; Std.
## beta = -1.29, 95% CI [-2.48, -0.11])
##   - The effect of ParticipantLeaning [far-right] is statistically non-significant
## and positive (beta = 0.75, 95% CI [-0.77, 2.27], t(258) = 0.97, p = 0.331; Std.
## beta = 0.78, 95% CI [-0.79, 2.35])
##   - The effect of ParticipantLeaning [left] is statistically non-significant and
## negative (beta = -0.07, 95% CI [-0.74, 0.59], t(258) = -0.22, p = 0.828; Std.
## beta = -0.08, 95% CI [-0.77, 0.61])
##   - The effect of ParticipantLeaning [not informed] is statistically
## non-significant and positive (beta = 0.15, 95% CI [-0.70, 1.00], t(258) = 0.35,
## p = 0.729; Std. beta = 0.16, 95% CI [-0.73, 1.04])
##   - The effect of ParticipantLeaning [right] is statistically non-significant and
## negative (beta = -0.35, 95% CI [-1.20, 0.50], t(258) = -0.81, p = 0.419; Std.
## beta = -0.36, 95% CI [-1.24, 0.52])
##   - The effect of TweetBias [Right] × ParticipantLeaning [center-left] is
## statistically non-significant and positive (beta = 0.31, 95% CI [-0.49, 1.11],
## t(258) = 0.77, p = 0.444; Std. beta = 0.32, 95% CI [-0.51, 1.15])
##   - The effect of TweetBias [Right] × ParticipantLeaning [center-right] is
## statistically non-significant and negative (beta = -0.25, 95% CI [-1.18, 0.68],
## t(258) = -0.53, p = 0.596; Std. beta = -0.26, 95% CI [-1.22, 0.70])
##   - The effect of TweetBias [Right] × ParticipantLeaning [far-left] is
## statistically significant and positive (beta = 2.00, 95% CI [0.56, 3.44],
## t(258) = 2.74, p = 0.007; Std. beta = 2.07, 95% CI [0.58, 3.56])
##   - The effect of TweetBias [Right] × ParticipantLeaning [far-right] is
## statistically non-significant and negative (beta = -1.75, 95% CI [-3.65, 0.15],
## t(258) = -1.81, p = 0.071; Std. beta = -1.81, 95% CI [-3.78, 0.15])
##   - The effect of TweetBias [Right] × ParticipantLeaning [left] is statistically
## non-significant and positive (beta = 0.43, 95% CI [-0.41, 1.26], t(258) = 1.01,
## p = 0.315; Std. beta = 0.44, 95% CI [-0.42, 1.31])
##   - The effect of TweetBias [Right] × ParticipantLeaning [not informed] is
## statistically non-significant and negative (beta = -0.25, 95% CI [-1.31, 0.81],
```

```
## t(258) = -0.46, p = 0.644; Std. beta = -0.26, 95% CI [-1.36, 0.84])
##   - The effect of TweetBias [Right] × ParticipantLeaning [right] is statistically
## non-significant and positive (beta = 0.25, 95% CI [-0.81, 1.31], t(258) = 0.46,
## p = 0.644; Std. beta = 0.26, 95% CI [-0.84, 1.36])
##
## Standardized parameters were obtained by fitting the model on a standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation.
```

Model Interpretation

Model Overview

- **Dependent Variable (`OriginalLikertValue`)**: Participants' polarization ratings of original tweets on a Likert scale.
- **Predictors**:
  - **TweetBias**: Political bias of the tweet (Left vs. Right).
  - **ParticipantLeaning**: Participants' political orientation (seven categories: center, center-left, center-right, far-left, far-right, left, not informed, and right).
  - **Interaction**: Between TweetBias and ParticipantLeaning.
- **Random Effects**:
  - A random intercept for each participant (`FK_ParticipantId`) accounts for individual differences in baseline polarization ratings.

Key Metrics

1. **REML Criterion**: `744.8`. A lower REML value suggests a better fit when comparing similar models.
2. **Residual Standard Deviation**: `0.89`, indicating the average deviation of observed values from predicted values after accounting for fixed and random effects.
3. **R^2 Values**:
   - **Conditional R^2**: `0.19`, representing the variance explained by both fixed and random effects.
   - **Marginal R^2**: `0.08`, representing the variance explained by fixed effects alone.

Random Effects

- **Variance of Participant-Level Random Intercept**: `0.11`, with a standard deviation of `0.33`.
  - Suggests moderate variability in participants' baseline ratings of polarization.
- **Residual Variance**: `0.80`, with a standard deviation of `0.89`.

Fixed Effects

1. Intercept:

- **Estimate**: `4.25`
- **Interpretation**: For tweets with **Left bias** and participants with **center** political orientation, the mean polarization score is **4.25**.
- **Significance**: Highly significant (`p < 0.001`).

2. Main Effects:

- **TweetBias (Right)**:
  - **Estimate**: `-0.25`
  - Interpretation: Tweets with a **Right bias** are rated slightly less polarized than Left-biased tweets, but this difference is **not significant** (`p = 0.493`).
- **ParticipantLeaning**:
  - Most ParticipantLeaning categories show **non-significant effects**, indicating their baseline polarization ratings are not markedly different from the **center** orientation.
  - **Exception**: Participants with a **far-left orientation** rate tweets as significantly less polarized compared to the center group (beta = `-1.25`, `p = 0.034`).

3. Interaction Effects:

- **TweetBias × ParticipantLeaning**:
  - **Significant Interaction**:
    - **TweetBias [Right] × ParticipantLeaning [far-left]**: Far-left participants perceive Right-biased tweets as significantly more polarized compared to Left-biased tweets (beta = `2.00`, `p = 0.007`).
  - **Marginally Significant Interaction**:
    - **TweetBias [Right] × ParticipantLeaning [far-right]**: Far-right participants perceive Right-biased tweets as less polarized than Left-biased tweets, but the effect is marginally significant (beta = `-1.75`, `p = 0.071`).
  - Other interaction terms are **not significant**, indicating no notable differences between Left and Right tweet bias ratings across other participant orientations.

Confidence Intervals

- The **95% Confidence Interval** for each fixed effect indicates plausible parameter values:
  - **Intercept**: `[3.68, 4.82]` – robustly positive.
  - **TweetBias (Right)**: `[-0.97, 0.47]` – includes zero, confirming non-significance.
  - **ParticipantLeaning (far-left)**: `[-2.40, -0.10]` – significant, entirely negative.
  - **Interaction (TweetBias × ParticipantLeaning [far-left])**: `[0.56, 3.44]` – significant, entirely positive.

Summary of Findings

1. **Main Effects**:
   - **TweetBias**: No significant difference in polarization ratings between Left and Right-biased tweets on average.
   - **ParticipantLeaning**: Far-left participants generally rate tweets as less polarized compared to the center group.
2. **Interaction Effects**:
   - Far-left participants perceive Right-biased tweets as significantly more polarized than Left-biased tweets.
   - Far-right participants tend to perceive Right-biased tweets as less polarized, but the effect is marginally significant.
3. **Participant-Level Variability**:
   - Moderate variability in participants' baseline ratings, as captured by the random effects.
4. **Model Fit**:
   - Fixed effects explain **8%** of the variance in polarization ratings, and the full model explains **19%**, suggesting moderate explanatory power.
