

Analysis

Pallavi Sharma and Bhaskar Nikhil Sunkara

2025-05-04

Contents

1	Data Loading	1
2	Hypothesis Testing	1
2.1	Hypothesis 1 (H1) Trust in BioGPT vs Gemini	1
2.2	Hypothesis 2 (H2) — Conflict reduces Trust	4
2.3	Hypothesis 3 (H3): Anchoring Bias (Chi-square Test)	5
2.4	Hypothesis 4 (H4): Confidence — Conflict vs Consistent (Independent-samples t-test)	6
3	Linear Regression	8
3.1	to examine how prior experience, medical knowledge, or cognitive biases influence trust — and whether they moderate the relationship between model type or diagnosis consistency and trust.	8

1 Data Loading

```
data <- read.csv("/Users/Patron/Desktop/project/data/data_merged.csv")
```

2 Hypothesis Testing

2.1 Hypothesis 1 (H1) Trust in BioGPT vs Gemini

```
data <- data %>%  
  mutate(  
    trust_biogpt = ifelse(model_order_encoded == 0, trust_rating_B, trust_rating_A),  
    trust_gemini = ifelse(model_order_encoded == 0, trust_rating_A, trust_rating_B)  
  )  
  
data %>% filter(model_order_encoded == 0) %>% select(trust_rating_A, trust_rating_B)
```

```
##      trust_rating_A trust_rating_B
## 1           4           4
## 2           4           4
## 3           5           4
## 4           4           5
## 5           5           4
## 6           3           5
## 7           5           2
## 8           3           4
## 9           4           3
## 10          3           4
## 11          5           1
## 12          3           4
## 13          4           3
## 14          4           3
## 15          4           4
```

```
# Normality tests
shapiro.test(data$trust_biogpt)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$trust_biogpt
## W = 0.86394, p-value = 0.002684
```

```
shapiro.test(data$trust_gemini)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$trust_gemini
## W = 0.85662, p-value = 0.001917
```

```
# Mann-Whitney U Test
wilcox.test(data$trust_biogpt, data$trust_gemini, paired = FALSE)
```

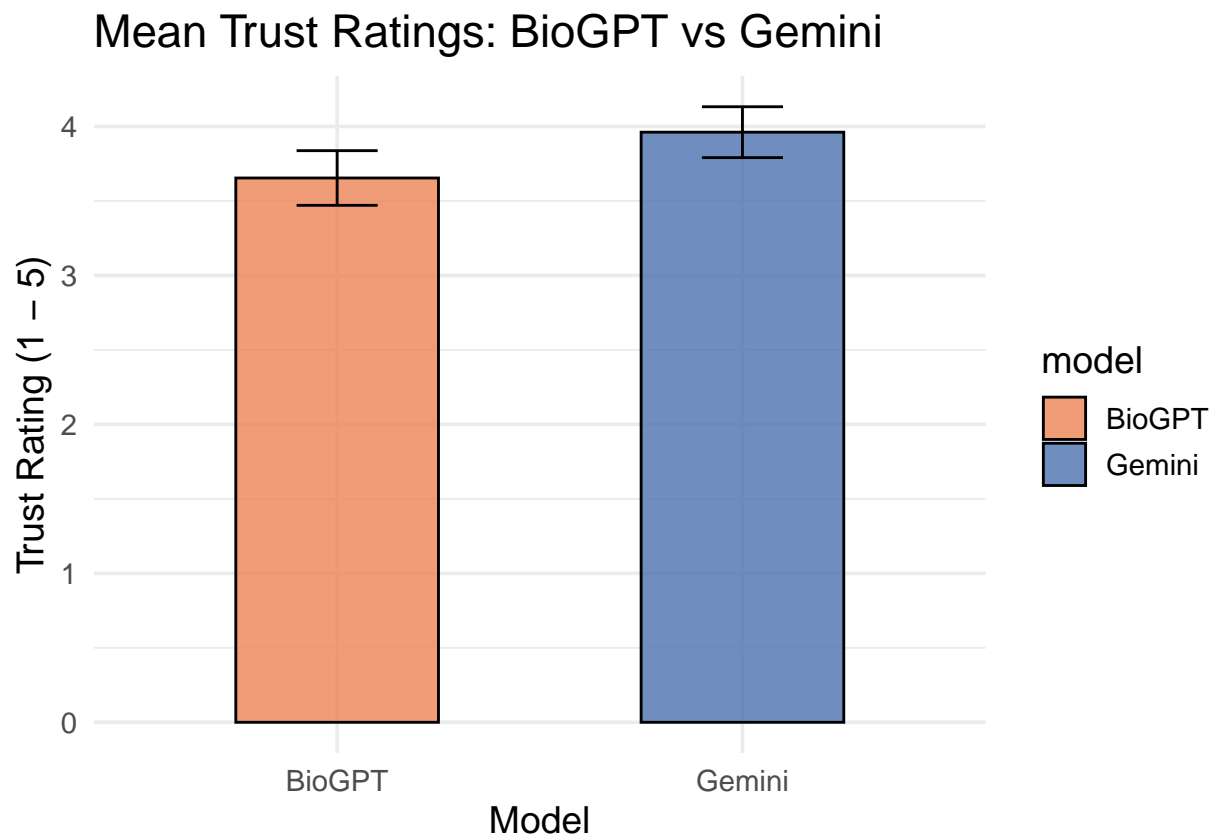
```
## Warning in wilcox.test.default(data$trust_biogpt, data$trust_gemini, paired =
## FALSE): cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  data$trust_biogpt and data$trust_gemini
## W = 280.5, p-value = 0.2684
## alternative hypothesis: true location shift is not equal to 0
```

```
# Effect size calculation
data %>%
  mutate(group = "BioGPT_vs_Gemini") %>%
  wilcox_effsize(trust_biogpt ~ trust_gemini, paired = FALSE)
```

```
## # A tibble: 6 x 7
##   .y.      group1 group2 effsize    n1    n2 magnitude
## * <chr>    <chr> <chr>   <dbl> <int> <int> <ord>
## 1 trust_biogpt 2      3      0.624     1     7 large
## 2 trust_biogpt 2      4      0.224     1    10 small
## 3 trust_biogpt 2      5      0.203     1     8 small
## 4 trust_biogpt 3      4      0.552     7    10 large
## 5 trust_biogpt 3      5      0.406     7     8 moderate
## 6 trust_biogpt 4      5      0.0557    10     8 small
```

```
# Bar plot (Mean ± SE) with colors
trust_data <- data.frame(
  model = factor(c(rep("BioGPT", nrow(data)), rep("Gemini", nrow(data)))),
  trust = c(data$trust_biogpt, data$trust_gemini)
)
# Updated bar plot for Trust (BioGPT vs Gemini)
ggplot(trust_data, aes(x = model, y = trust, fill = model)) +
  stat_summary(fun = mean, geom = "bar", width = 0.5, color = "black", alpha = 0.8) +
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.2) +
  scale_fill_manual(values = c("BioGPT" = "#EC8456", "Gemini" = "#4C72B0")) +
  labs(title = "Mean Trust Ratings: BioGPT vs Gemini",
       x = "Model",
       y = "Trust Rating (1 - 5)") +
  theme_minimal(base_size = 14)
```



2.2 Hypothesis 2 (H2) — Conflict reduces Trust

```
data$diagnosis_type_label <- factor(data$diagnosis_type_encoded,
                                   levels = c(0, 1),
                                   labels = c("Consistent", "Conflicting"))
```

```
# Normality test
```

```
shapiro.test(data$final_trust_rating)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: data$final_trust_rating
```

```
## W = 0.89951, p-value = 0.01527
```

```
# Mann-Whitney U Test
```

```
wilcox.test(final_trust_rating ~ diagnosis_type_label, data = data)
```

```
## Warning in wilcox.test.default(x = DATA[[1L]], y = DATA[[2L]], ...): cannot
```

```
## compute exact p-value with ties
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
```

```
##
```

```
## data: final_trust_rating by diagnosis_type_label
```

```
## W = 139, p-value = 0.001904
```

```
## alternative hypothesis: true location shift is not equal to 0
```

```
# Effect size
```

```
data %>%
```

```
  wilcox_effsize(final_trust_rating ~ diagnosis_type_label)
```

```
## # A tibble: 1 x 7
```

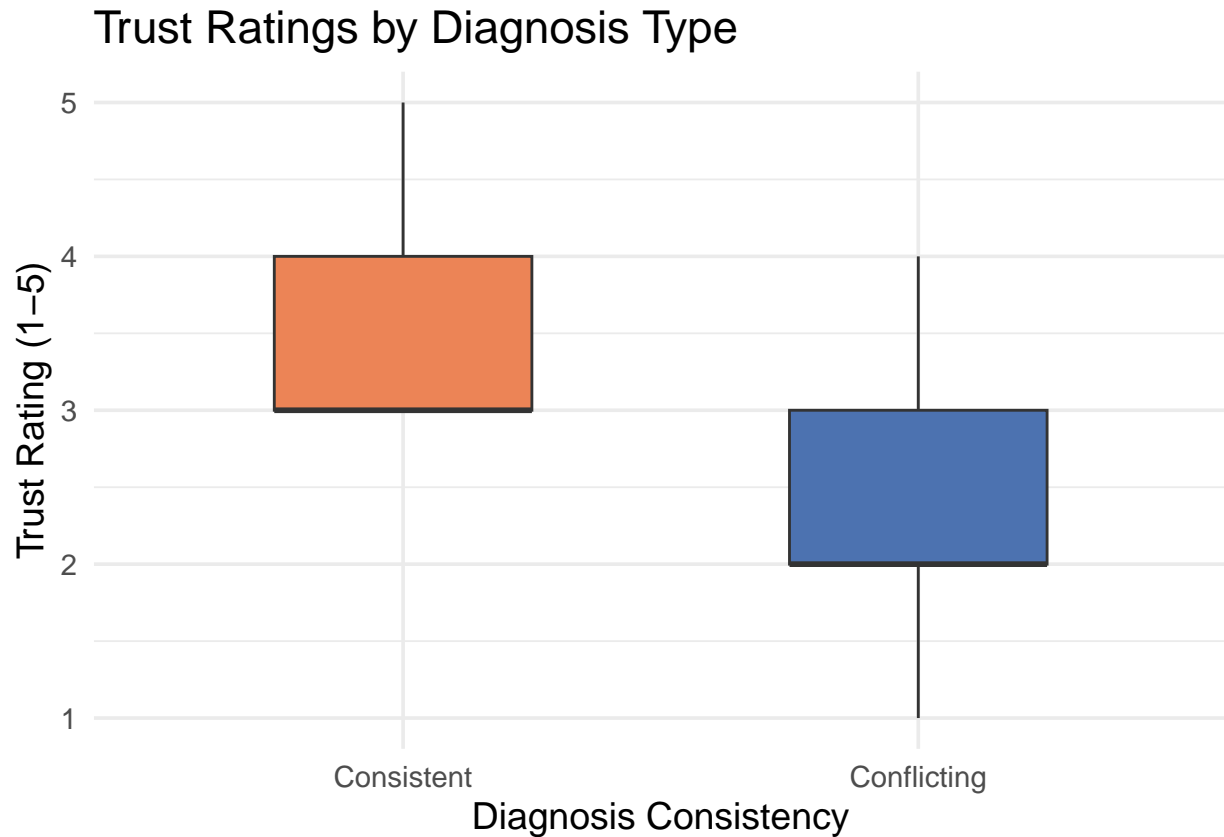
```
##   .y.      group1      group2      effsize      n1      n2 magnitude
```

```
## * <chr>      <chr>      <chr>      <dbl> <int> <int> <ord>
```

```
## 1 final_trust_rating Consistent Conflicting  0.614    11    15 large
```

```
# Trust Rating Boxplot
```

```
ggplot(data, aes(x = diagnosis_type_label, y = final_trust_rating, fill = diagnosis_type_label)) +
  geom_boxplot(width = 0.5, outlier.shape = 21, outlier.fill = "white", outlier.color = "black") +
  scale_fill_manual(values = c("Consistent" = "#EC8456", "Conflicting" = "#4C72B0")) +
  labs(title = "Trust Ratings by Diagnosis Type",
       x = "Diagnosis Consistency",
       y = "Trust Rating (1-5)") +
  theme_minimal(base_size = 14) +
  theme(legend.position = "none")
```



2.3 Hypothesis 3 (H3): Anchoring Bias (Chi-square Test)

```
table(data$model_order_encoded, data$preferred_model_encoded) # View contingency table
```

```
##
##      0 1
##      0 7 8
##      1 6 5
```

```
chisq.test(table(data$model_order_encoded, data$preferred_model_encoded))
```

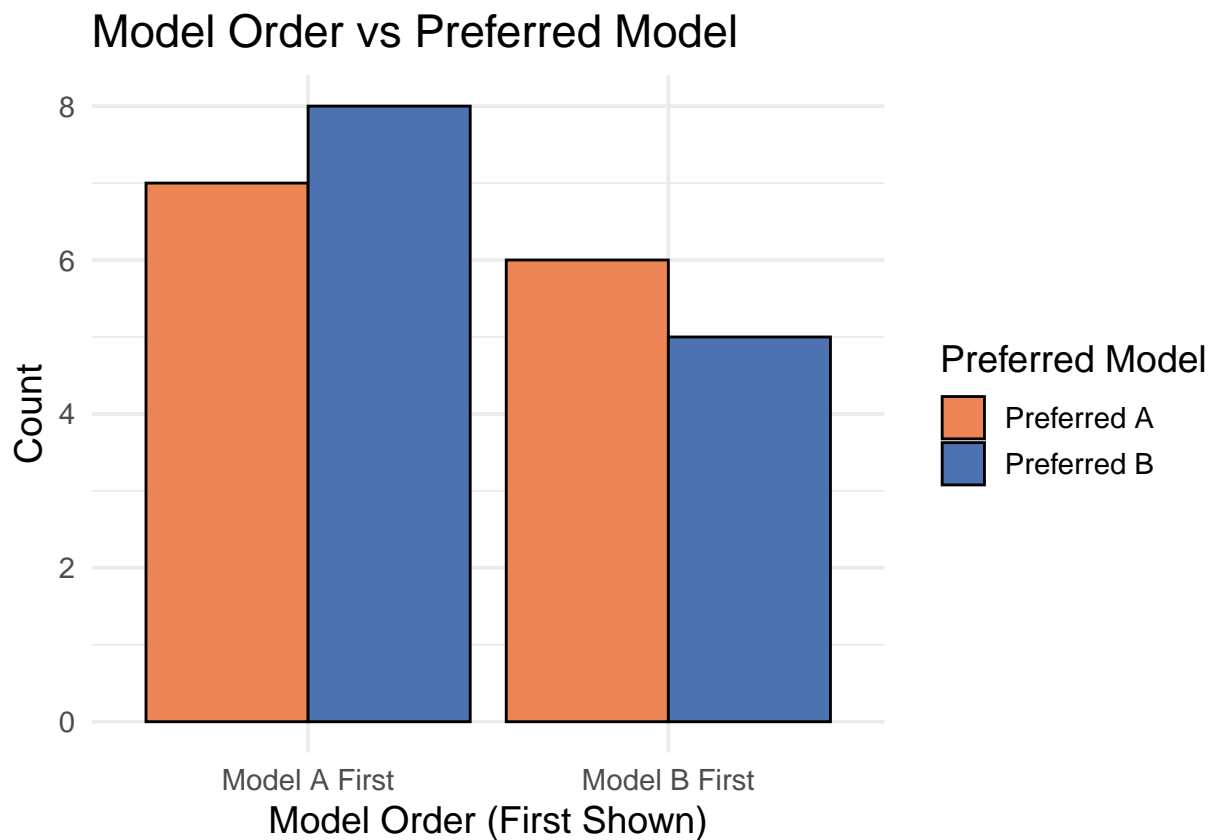
```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(data$model_order_encoded, data$preferred_model_encoded)
## X-squared = 0, df = 1, p-value = 1
```

```
# Create Proper Labels
```

```
data$model_order_label <- factor(data$model_order_encoded, labels = c("Model A First", "Model B First"))
data$preferred_model_label <- factor(data$preferred_model_encoded, labels = c("Preferred A", "Preferred B"))
```

```
# Anchoring Bias Plot
```

```
ggplot(data, aes(x = model_order_label, fill = preferred_model_label)) +
  geom_bar(position = "dodge", color = "black") +
  scale_fill_manual(values = c("Preferred A" = "#EC8456", "Preferred B" = "#4C72B0")) +
  labs(title = "Model Order vs Preferred Model",
       x = "Model Order (First Shown)",
       y = "Count",
       fill = "Preferred Model") +
  theme_minimal(base_size = 14)
```



2.4 Hypothesis 4 (H4): Confidence — Conflict vs Consistent (Independent-samples t-test)

```
wilcox.test(final_confidence_rating ~ diagnosis_type_label, data = data)
```

```
## Warning in wilcox.test.default(x = DATA[[1L]], y = DATA[[2L]], ...): cannot
## compute exact p-value with ties
```

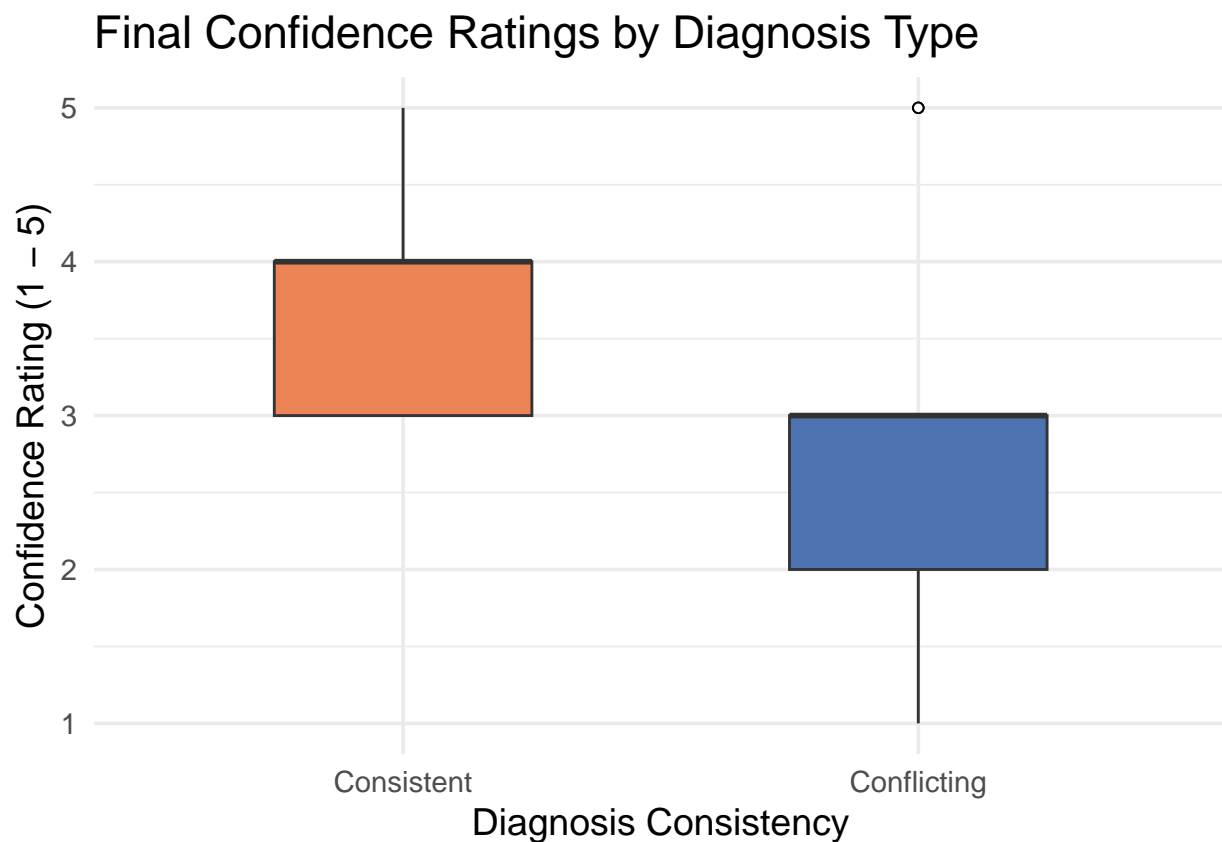
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: final_confidence_rating by diagnosis_type_label
## W = 121, p-value = 0.03845
## alternative hypothesis: true location shift is not equal to 0
```

```
# Effect size
data %>%
  wilcox_effsize(final_confidence_rating ~ diagnosis_type_label)
```

```
## # A tibble: 1 x 7
##   .y.          group1      group2      effsize    n1    n2 magnitude
## * <chr>          <chr>      <chr>      <dbl> <int> <int> <ord>
## 1 final_confidence_rating Consistent Conflicting 0.411    11    15 moderate
```

```
# Confidence Rating Boxplot
```

```
ggplot(data, aes(x = diagnosis_type_label, y = final_confidence_rating, fill = diagnosis_type_label)) +
  geom_boxplot(width = 0.5, outlier.shape = 21, outlier.fill = "white", outlier.color = "black") +
  scale_fill_manual(values = c("Consistent" = "#EC8456", "Conflicting" = "#4C72B0")) +
  labs(title = "Final Confidence Ratings by Diagnosis Type",
       x = "Diagnosis Consistency",
       y = "Confidence Rating (1 - 5)") +
  theme_minimal(base_size = 14) +
  theme(legend.position = "none")
```



3 Linear Regression

3.1 to examine how prior experience, medical knowledge, or cognitive biases influence trust — and whether they moderate the relationship between model type or diagnosis consistency and trust.

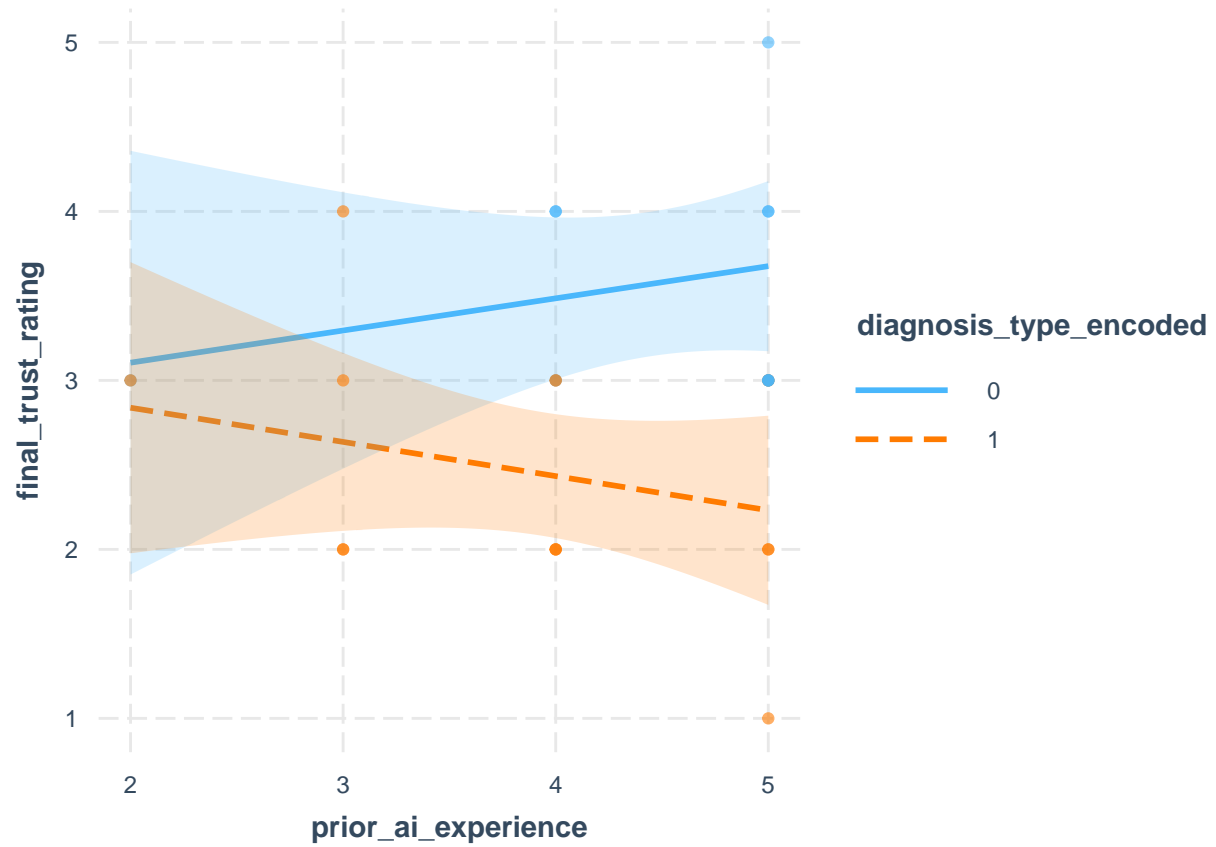
```
data$diagnosis_type_encoded <- as.factor(data$diagnosis_type_encoded)

# Linear model with interaction
model <- lm(final_trust_rating ~ diagnosis_type_encoded * prior_ai_experience + medical_knowledge_level
            data = data)

summary(model)
```

```
##
## Call:
## lm(formula = final_trust_rating ~ diagnosis_type_encoded * prior_ai_experience +
##     medical_knowledge_level, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0547 -0.4532 -0.1563  0.4563  1.5027
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   2.3597     1.0850   2.175
## diagnosis_type_encoded1         0.5188     1.2972   0.400
## prior_ai_experience              0.1903     0.2315   0.822
## medical_knowledge_level          0.1858     0.1001   1.856
## diagnosis_type_encoded1:prior_ai_experience -0.3927     0.2983  -1.316
##                                Pr(>|t|)
## (Intercept)                   0.0412 *
## diagnosis_type_encoded1         0.6933
## prior_ai_experience              0.4202
## medical_knowledge_level          0.0775 .
## diagnosis_type_encoded1:prior_ai_experience  0.2022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6802 on 21 degrees of freedom
## Multiple R-squared:  0.5105, Adjusted R-squared:  0.4172
## F-statistic: 5.475 on 4 and 21 DF,  p-value: 0.003517

# Interaction plot
interact_plot(model, pred = prior_ai_experience, modx = diagnosis_type_encoded,
              plot.points = TRUE, interval = TRUE)
```

```
interact_plot(model, pred = medical_knowledge_level, modx = diagnosis_type_encoded,  
              plot.points = TRUE, interval = TRUE)
```

```
## Warning: medical_knowledge_level and diagnosis_type_encoded are not included in an  
## interaction with one another in the model.
```

