Artificial Intelligence Opinion Survey

DATA 490 Independent Study

Mentor: Dr. Nicholas Dietrich
Researchers: Aadarsha Gopala Reddy, Darren Lo,
Ethan Love, Jose Mancilla, Santo Sumo

May 07, 2023

Contents

1. Load Libraries & Data

```
library(tidyverse)
library(ggplot2)
library(knitr)

# Load data. Top row is column name.
edu <- read.csv("prolific_edu.csv")
health <- read.csv("prolific_health.csv")
retail <- read.csv("prolific_retail.csv")
tech <- read.csv("prolific_tech.csv")
qualtrics <- read.csv("qualtrics_data.csv")
```

2. Data Cleaning

```
# Mutate Age to be numeric data type
edu <- edu %>% mutate(Age = as.numeric(Age))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `Age = as.numeric(Age)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
health <- health %>% mutate(Age = as.numeric(Age))
retail <- retail %>% mutate(Age = as.numeric(Age))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `Age = as.numeric(Age)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
tech <- tech %>% mutate(Age = as.numeric(Age))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `Age = as.numeric(Age)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
# Combine all prolific data into one data frame
combined <- bind_rows(edu, health, retail, tech)
```

```
# Merge qualtrics data with combined data over qualtrics' ProlificID column
#   and combined's Participant.id column
combined <- left_join(qualtrics, combined,
  by = c("ProlificID" = "Participant.id")
)
# Rename the Duration..in.seconds. column to Duration
colnames(combined)[
  colnames(combined) == "Duration..in.seconds."
] <- "Duration"
# Rename the Ethnicity.simplified column to Ethnicity
colnames(combined)[
  colnames(combined) == "Ethnicity.simplified"
] <- "Ethnicity"

# Remove Finished, Progress, UserLanguage, DistributionChannel, Nationality,
#   Consent, Age.x, Status.x, Status.y columns AND rename Age.y to Age
combined <- combined %>%
  select(-Finished) %>%
  select(-Progress) %>%
  select(-UserLanguage) %>%
  select(-DistributionChannel) %>%
  select(-Nationality) %>%
  select(-Consent) %>%
  select(-Age.x) %>%
  select(-Status.x) %>%
  select(-Status.y) %>%
  rename(Age = Age.y)

# Remove rows which say "NA" in Submission.id column
combined <- combined %>% filter(!is.na(Submission.id))
# Remove rows which don't say "United States" in Country.of.residence column
combined <- combined %>% filter(combined$Country.of.residence == "United States")
```

```r
# Replace all cells that say
#   "Information Technology" and "Science, Technology, Engineering & Mathematics"
#   to "STEM/IT" in the Employment.sector column
combined$Employment.sector[
  combined$Employment.sector == "Information Technology"
] <- "STEM/IT"
combined$Employment.sector[
  combined$Employment.sector == "Science, Technology, Engineering & Mathematics"
] <- "STEM/IT"

# Replace all cells in column EnhanceHurt that say
#   "AI will neither enhance nor detract from my work" to "neither",
#   "AI will enhance my work" to "enhance", and
#   "AI will detract from my work" to "detract"
combined$EnhanceHurt[
  combined$EnhanceHurt == "AI will neither enhance nor detract from my work"
] <- "Neither"
combined$EnhanceHurt[
  combined$EnhanceHurt == "AI will enhance my work"
] <- "Enhance"
combined$EnhanceHurt[
  combined$EnhanceHurt == "AI will detract from my work"
] <- "Detract"

# Replace all cells in column TimeEnergy that say
#   "Save a lot of time" to "lot of time",
#   "Save some time" to "some time",
#   "Neutral" to "neutral"
#   "Save little time" to "little time", and
#   "Save no time" to "no time",
combined$TimeEnergy[
  combined$TimeEnergy == "Save a lot of time"
] <- "Lot of time"
combined$TimeEnergy[
  combined$TimeEnergy == "Save some time"
] <- "Some time"
combined$TimeEnergy[
  combined$TimeEnergy == "Neutral"
] <- "Neutral"
combined$TimeEnergy[
  combined$TimeEnergy == "Save little time"
] <- "Little time"
combined$TimeEnergy[
  combined$TimeEnergy == "Save no time"
] <- "No time"
```

```r
# Remove rows which don't say "Compose an email" in the Attention column
combined <- combined %>% filter(combined$Attention == "Compose an email")
# Remove the Attention column
combined <- combined %>% select(-Attention)
```

3. Data Exploration

The columns in the dataset are:

- *StartDate* - Date and time survey was started
- *EndDate* - Date and time survey was completed
- *IPAddress* - IP address of participant
- *Duration* - Duration of survey in seconds
- *RecordedDate* - Date and time survey was recorded
- *ResponseId* - Response ID
- *LocationLatitude* - Participant's location latitude
- *LocationLongitude* - Participant's location longitude
- *ProlificID* - Identification of the response on Prolific
- *Gender* - Gender of the participant
- *Education* - Education level of the participant
- *Salary* - Salary of the participant
- *AIKnowledge* - Knowledge of AI of the participant
- *UsedAI* - Whether the participant has used AI
- *TimeEnergy* - How much time and energy AI has saved the participant
- *SimilarTasks* - How much of the participant's tasks they think AI can do
- *EnhanceHurt* - Whether the participant thinks AI can enhance or hurt their work efficiency.
- *Comments* - Comments from the participant
- *Submission.id* - Submission ID
- *Started.at* - Date and time survey was started
- *Completed.at* - Date and time survey was completed
- *Reviewed.at* - Date and time survey was reviewed
- *Archived.at* - Date and time survey was archived
- *Time.taken* - Duration of survey in seconds
- *Completion.code* - Completion code
- *Total.approvals* - Total number of approvals
- *Employment.sector* - Employment sector
- *Age* - Age of the participant
- *Sex* - Sex of the participant
- *Ethnicity* - Ethnicity of the participant
- *Country.of.birth* - Country of birth of the participant
- *Country.of.residence* - Country of residence of the participant
- *Language* - Language of the participant
- *Student.status* - Whether the participant is a student

- *Employment.status* - Whether the participant is employed

4. Data Analysis

4.1. Average amount of time to complete survey

```
# Create a new data frame with only the columns we need
avg_time <- combined %>% select(Duration)

# Remove rows which say "NA" in the Duration column
avg_time <- avg_time %>% filter(!is.na(Duration))

# Calculate the average time taken to complete the survey
avg_time <- avg_time %>% summarise(avg_time = mean(Duration))

# Convert to minutes and seconds
avg_time$avg_time <- avg_time$avg_time / 60
avg_time$avg_time <- paste0(
  floor(avg_time$avg_time), " minutes and ",
  round((avg_time$avg_time - floor(avg_time$avg_time)) * 60), " seconds"
)

# Print the average time taken to complete the survey
kable(avg_time, row.names = FALSE)
```

avg_time
2 minutes and 13 seconds

The average time taken to complete the survey is 2 minutes and 13 seconds. This is a reasonable amount of time to complete the survey, and is not too long or too short. This means that the data collected is not rushed, and is of good quality.

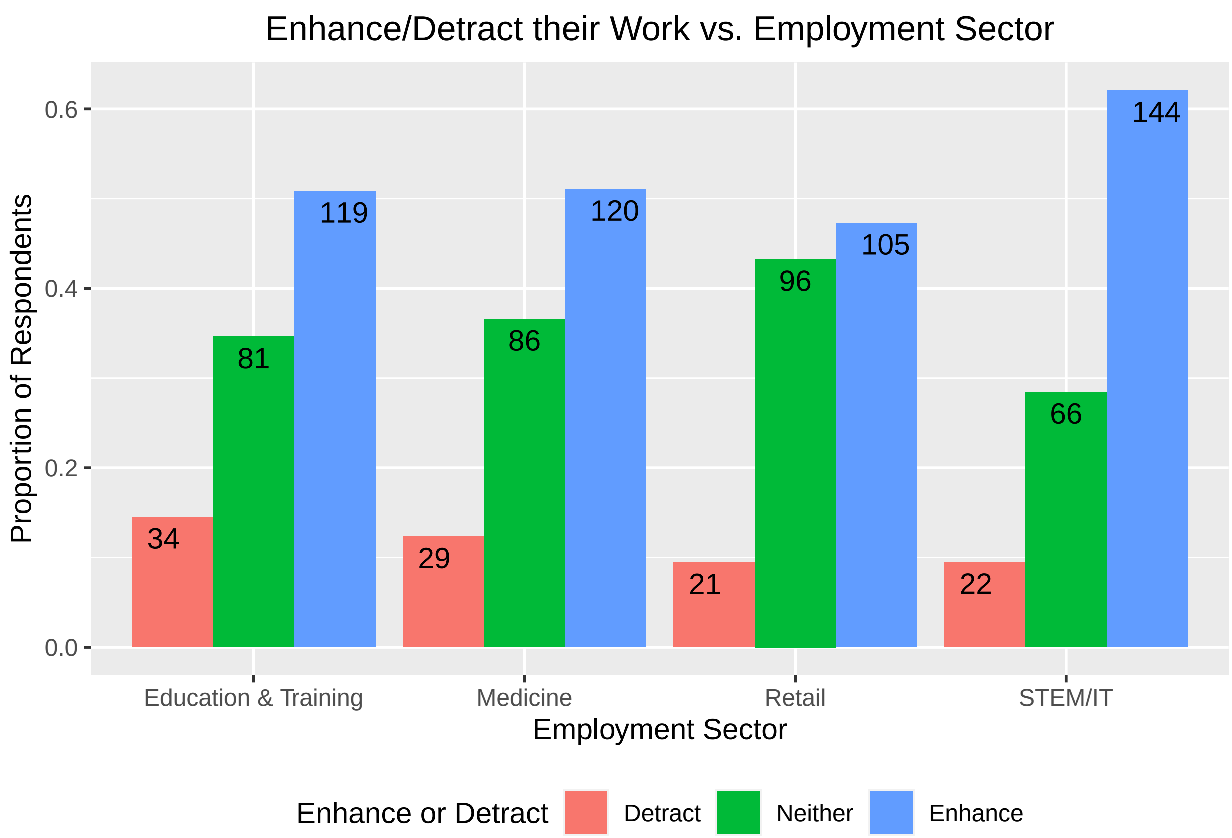
4.2. Enhance/Detract vs. Sector

```r
# Create a new data frame with only the necessary columns
enhancehurt_vs_sector <- combined %>% select(EnhanceHurt, Employment.sector)
```

```r
# For each sector, calculate the proportion of respondents who
#   think AI can enhance their work efficiency
enhancehurt_vs_sector <- enhancehurt_vs_sector %>%
  group_by(Employment.sector, EnhanceHurt) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))

# Create a bar chart to show the proportion of respondents who think AI
#   can enhance their work efficiency by employment sector
enhancehurt_vs_sector_plot <- ggplot(
  enhancehurt_vs_sector,
  aes(x = Employment.sector, y = prop, fill = factor(EnhanceHurt,
    levels = c("Detract", "Neither", "Enhance")
  )),
  ylim = c(0, 1)
) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = count), position = position_dodge(width = 1), vjust = 1.5) +
  labs(
    x = "Employment Sector", y = "Proportion of Respondents", fill = "Enhance or Detract"
  ) +
  ggtitle("Enhance/Detract their Work vs. Employment Sector") +
  theme(plot.title = element_text(hjust = 0.5), legend.position = "bottom")

# Print the bar chart
enhancehurt_vs_sector_plot
```



According to our survey results, respondents in the STEM/IT fields express the highest degree of confidence in AI's ability to improve work efficiency. In contrast, respondents from the Education and Training and Medicine sectors indicate that AI will likely detract from their work efficiency. The Retail sector is the least certain about the impact of AI on work efficiency, which can be attributed to its possibly diverse range of jobs, spanning from sales assistants to store managers. Consequently, the effects of AI on this sector are expected to vary, with some jobs being enhanced and others detracted. Unlike the Retail sector, the STEM/IT sector is anticipated to have a more consistent impact from AI as the jobs in this field are more alike. Similarly, the Education and Training sector is expected to be uniform in the effects of AI due to the jobs' similarities, while the Medicine sector's jobs are more specialized, but the impact of AI is expected to be consistent.

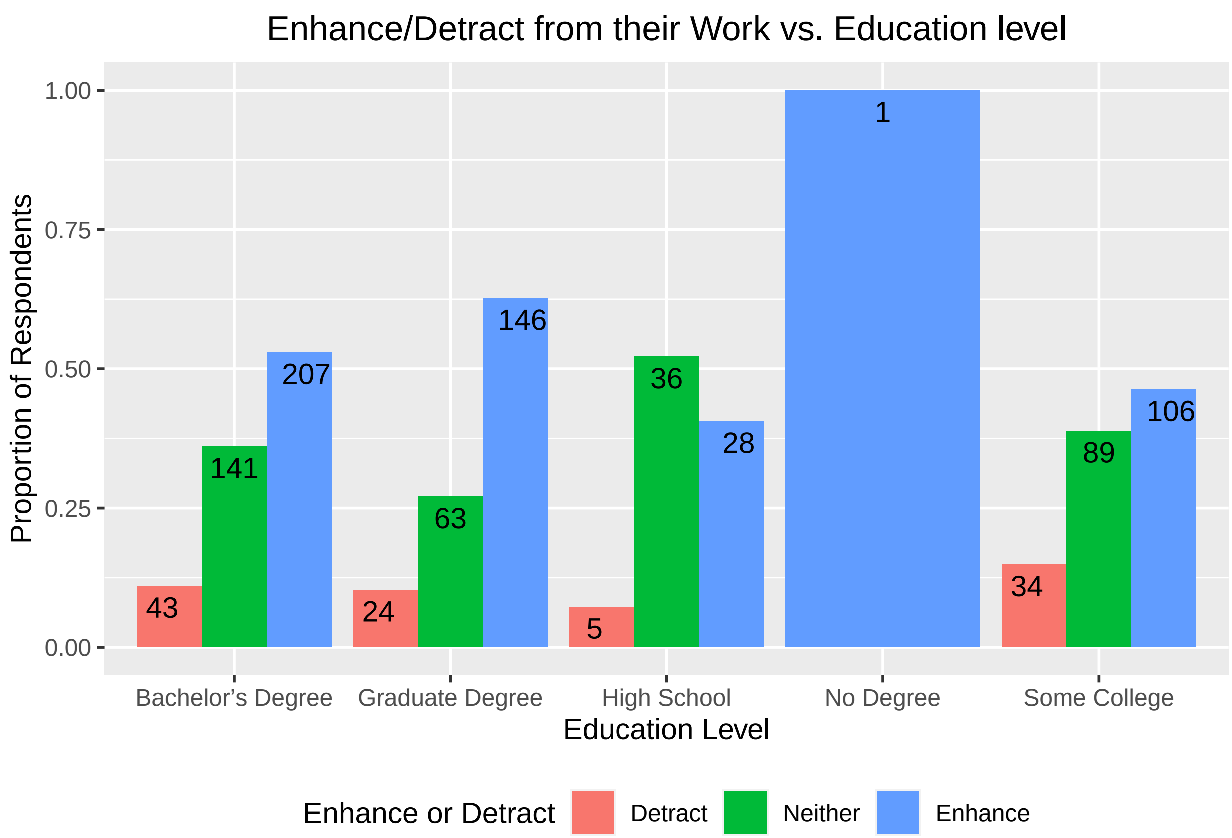
4.3. Enhance/Detract vs. Education

```
# Create a new data frame with only the necessary columns
enhancehurt_vs_education <- combined %>% select(EnhanceHurt, Education)
```

```
# For each education level, calculate the proportion of respondents who
#   think AI can enhance their work efficiency
enhancehurt_vs_education <- enhancehurt_vs_education %>%
  group_by(Education, EnhanceHurt) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))

# Create a bar chart to show the proportion of respondents who think AI
#   can enhance their work efficiency by education level
enhancehurt_vs_education_plot <- ggplot(
  enhancehurt_vs_education,
  aes(x = Education, y = prop, fill = factor(EnhanceHurt,
    levels = c("Detract", "Neither", "Enhance")
  )),
  ylim = c(0, 1)
) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = count), position = position_dodge(width = 1), vjust = 1.5) +
  labs(x = "Education Level", y = "Proportion of Respondents", fill = "Enhance or Detract") +
  ggtitle("Enhance/Detract from their Work vs. Education level") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "bottom")

# Print the bar chart
enhancehurt_vs_education_plot
```



The data shows that individuals with bachelor's and graduate degrees have a more positive outlook on the impact of AI on their work efficiency, while those with some college education are less convinced. Those with a high school education believe that AI will have a neutral effect on their work. It is noteworthy that only one respondent had no education background. It appears that as educational attainment increases, so does the belief that AI will enhance work efficiency. This may be because higher education is associated with jobs that require more complex tasks, which are more likely to be positively impacted by AI. Conversely, lower education levels are often associated with jobs that involve less complex tasks, which are less likely to be affected by AI.

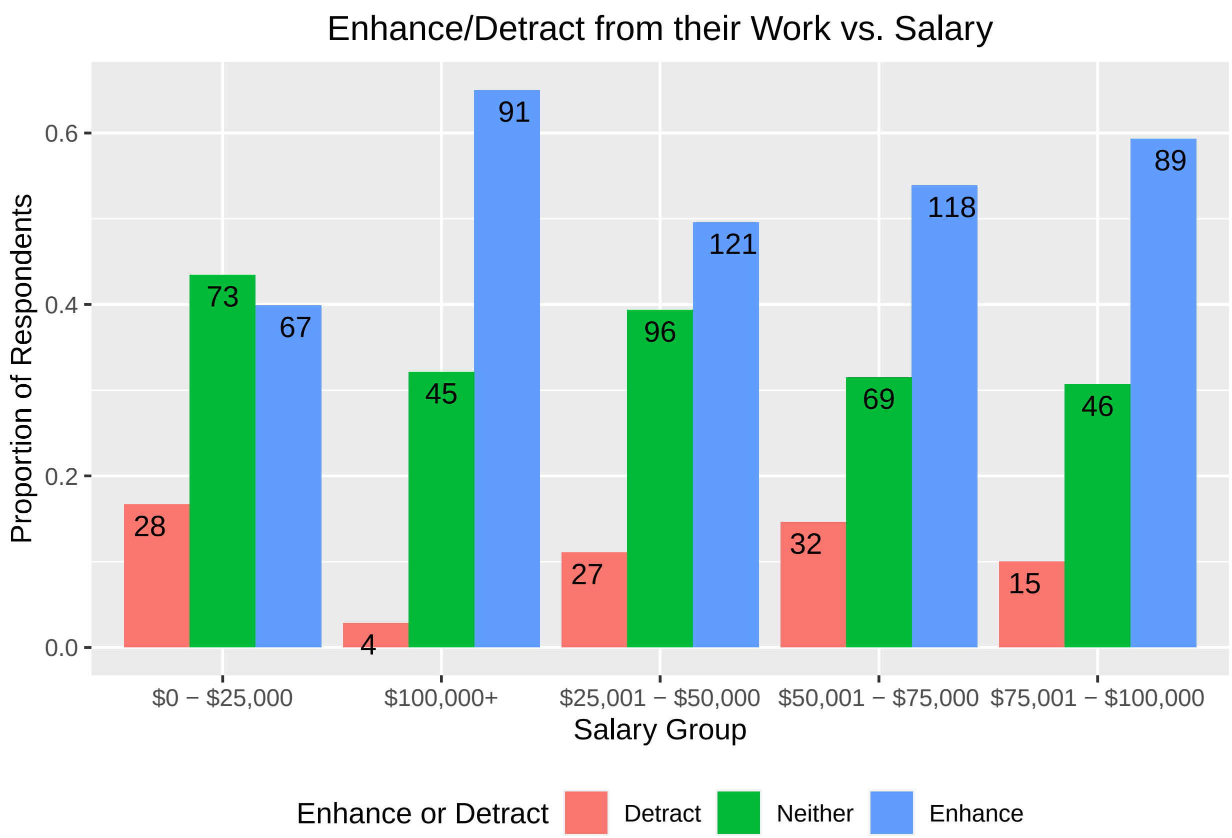
4.4. Enhance/Detract vs. Salary

```r
# Create a new data frame with only the necessary columns
enhancehurt_vs_salary <- combined %>% select(EnhanceHurt, Salary)
```

```r
# For each Salary range, calculate the proportion of respondents who
#   think AI can enhance their work efficiency
enhancehurt_vs_salary <- enhancehurt_vs_salary %>%
  group_by(Salary, EnhanceHurt) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))

# Create a bar chart to show the proportion of respondents who think AI
#   can enhance their work efficiency by salary range
enhancehurt_vs_salary_plot <- ggplot(
  enhancehurt_vs_salary,
  aes(x = Salary, y = prop, fill = factor(EnhanceHurt,
    levels = c("Detract", "Neither", "Enhance")
  )),
  ylim = c(0, 1)
) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = count), position = position_dodge(width = 1), vjust = 1.5) +
  labs(x = "Salary Group", y = "Proportion of Respondents", fill = "Enhance or Detract") +
  ggtitle("Enhance/Detract from their Work vs. Salary") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "bottom")

# Print the bar chart
enhancehurt_vs_salary_plot
```



The data presented highlights a correlation between salary and belief in the potential of AI to enhance work efficiency. The chart suggest that higher earners are more likely to hold this belief, potentially due to the nature of their jobs. Occupations that require more specific tasks, such as a doctor's role, are more likely to benefit from AI-enhanced tools, which could explain why individuals in higher-paying jobs are more optimistic about AI's potential impact. In contrast, roles that require less specific tasks, such as those of a cashier, may have fewer opportunities for AI to enhance work efficiency. This information underscores the importance of considering the nature of different occupations when evaluating the potential impact of AI on the workforce.

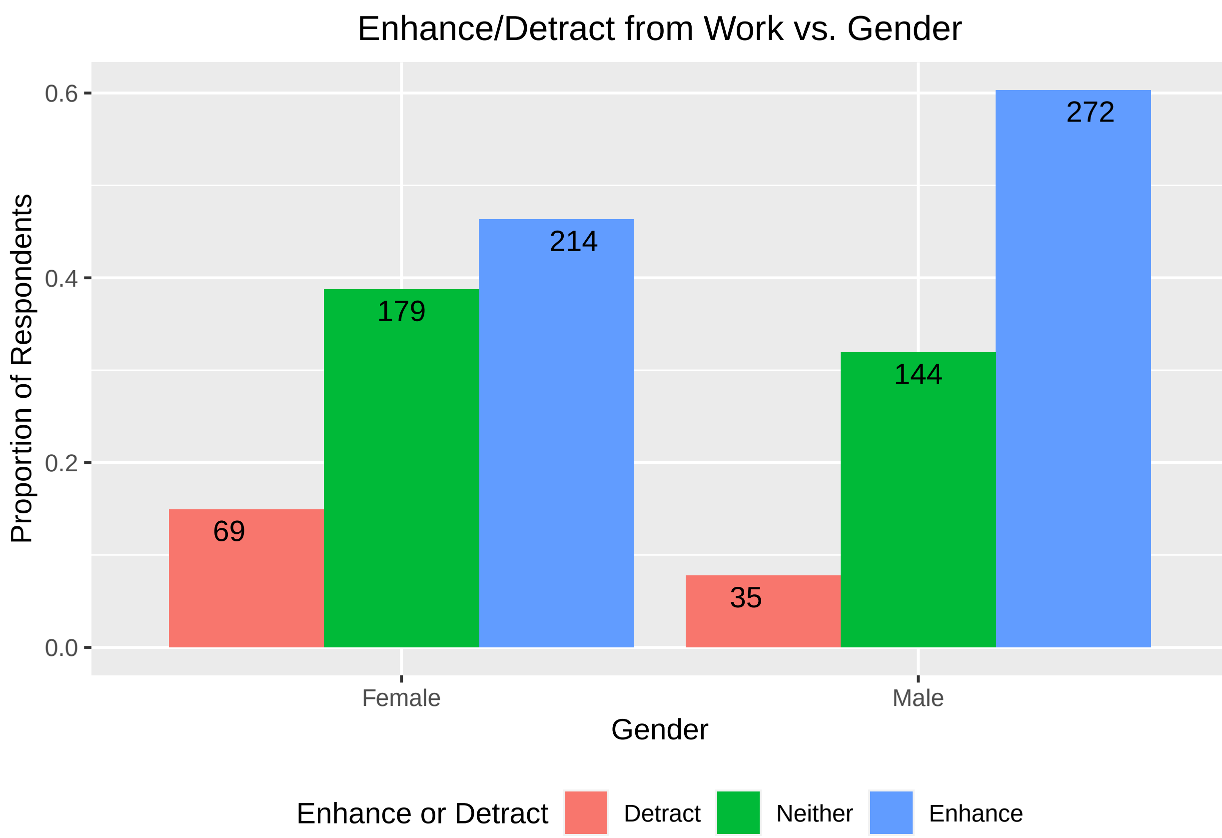
4.5. Enhance/Detract vs. Gender

```r
# Create a new data frame with only the necessary columns
enhancehurt_vs_gender <- combined %>% select(EnhanceHurt, Gender)
```

```r
# For each gender, calculate the proportion of respondents who
#   think AI can enhance their work efficiency
enhancehurt_vs_gender <- enhancehurt_vs_gender %>%
  group_by(Gender, EnhanceHurt) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))

# Create a bar chart to show the proportion of respondents who think AI
#   can enhance their work efficiency by gender
enhancehurt_vs_gender_plot <- ggplot(
  enhancehurt_vs_gender,
  aes(x = Gender, y = prop, fill = factor(EnhanceHurt,
    levels = c("Detract", "Neither", "Enhance")
  )),
  ylim = c(0, 1)
) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = count), position = position_dodge(width = 1), vjust = 1.5) +
  labs(x = "Gender", y = "Proportion of Respondents", fill = "Enhance or Detract") +
  ggtitle("Enhance/Detract from Work vs. Gender") +
  theme(plot.title = element_text(hjust = 0.5), legend.position = "bottom")

# Print the bar chart
enhancehurt_vs_gender_plot
```



The data indicates that a greater percentage of male participants believed that AI would improve their work performance, whereas a higher proportion of female respondents believed it would detract from their tasks. This gender disparity may be linked to the salary differences between men and women, as demonstrated in the chart displaying gender distribution in various salary brackets. The chart shows that a majority of male respondents hold higher-paying jobs than their female counterparts. Additionally, the EnhanceHurt vs. Salary chart in the previous section revealed that individuals with higher salaries were more likely to believe that AI would improve their work. This trend underscores the differences in preferences between men and women, which can be attributed to gender-based pay inequality.

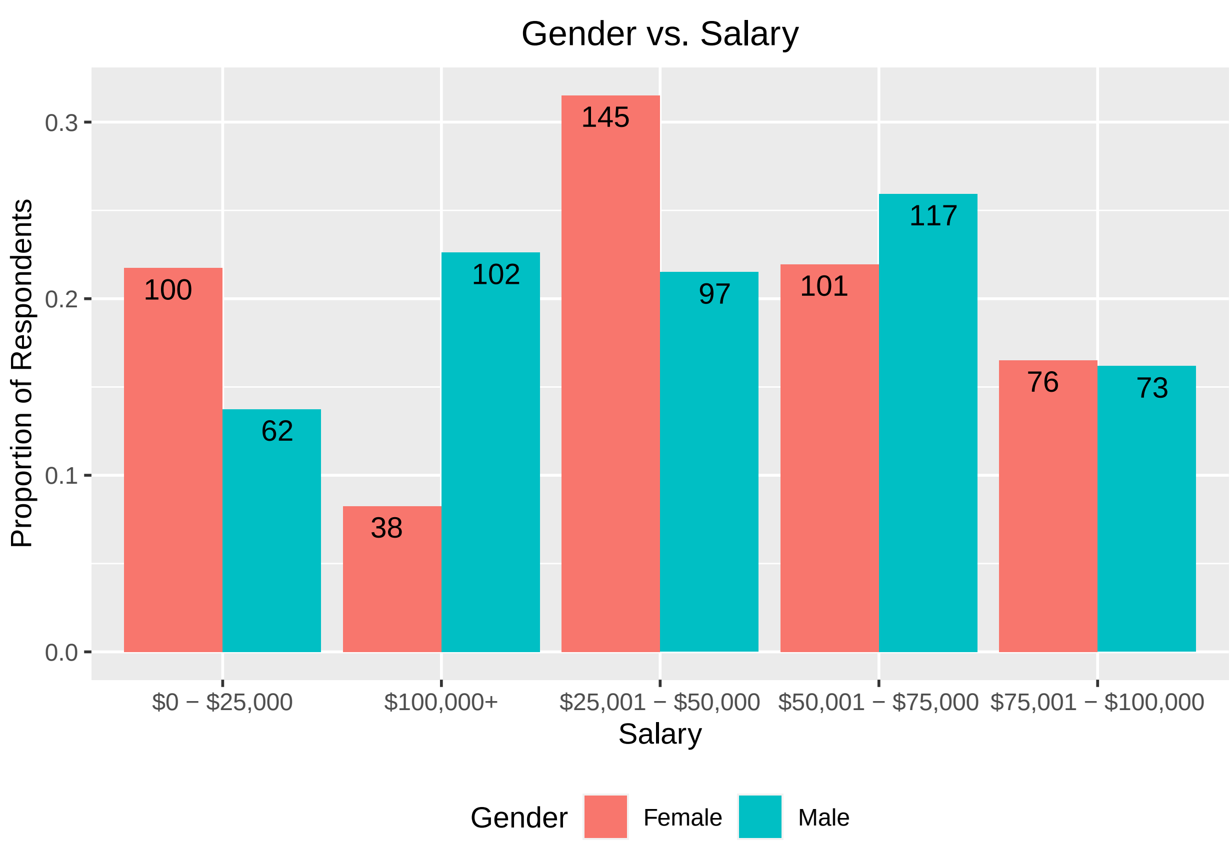
4.6. Salary vs. Gender

```
# Create a new data frame with only the necessary columns
salary_vs_gender <- combined %>% select(Gender, Salary)
```

```
# For each gender, calculate the proportion of respondents by salary
salary_vs_gender <- salary_vs_gender %>%
  group_by(Gender, Salary) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))

# Create a bar chart to show the proportion of respondents by salary
#   split by gender
salary_vs_gender_plot <- ggplot(
  salary_vs_gender,
  aes(x = Salary, y = prop, fill = factor(Gender)),
  ylim = c(0, 1)
) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = count), position = position_dodge(width = 1), vjust = 1.5) +
  labs(x = "Salary", y = "Proportion of Respondents", fill = "Gender") +
  ggtitle("Gender vs. Salary") +
  theme(plot.title = element_text(hjust = 0.5), legend.position = "bottom")

# Print the bar chart
salary_vs_gender_plot
```



This chart depicts the disparity in salaries between genders, highlighting the significant gap between male and female earnings. While the number of respondents is roughly equal in both categories, the data reveals a concerning trend, with males overwhelmingly occupying the higher salary brackets and females being disproportionately represented in the lower ones. The root causes of this discrepancy are complex and may include factors such as willingness to participate in surveys, discriminatory hiring practices and unconscious biases, gender pay inequality, and more.

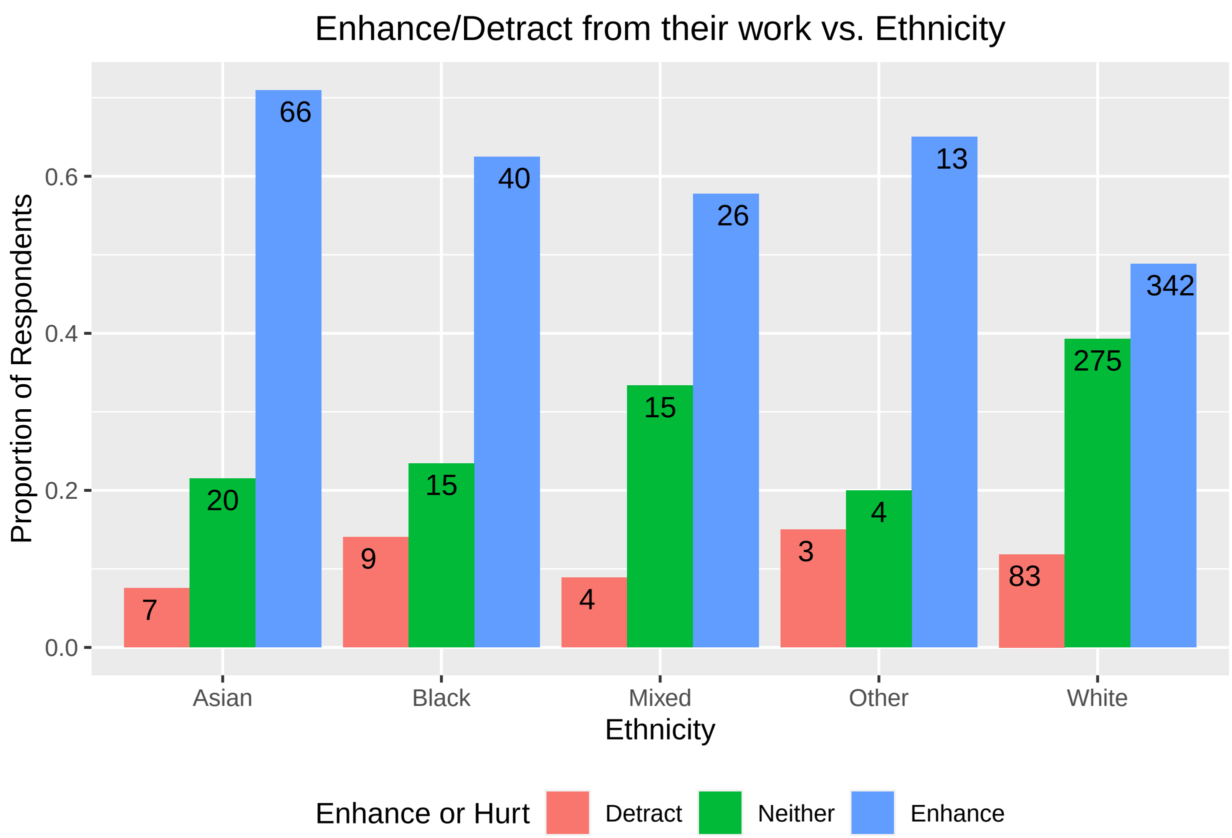
4.7. Enhance/Detract vs. Ethnicity

```
# Create a new data frame with only the necessary columns
enhancehurt_vs_ethnicity <- combined %>% select(EnhanceHurt, Ethnicity)
```

```
# For each Ethnicity, calculate the proportion of respondents who
#   think AI can enhance their work efficiency
enhancehurt_vs_ethnicity <- enhancehurt_vs_ethnicity %>%
  group_by(Ethnicity, EnhanceHurt) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))

# Create a bar chart to show the proportion of respondents who think AI
#   can enhance their work efficiency by ethnicity
enhancehurt_vs_ethnicity_plot <- ggplot(
  enhancehurt_vs_ethnicity,
  aes(x = Ethnicity, y = prop, fill = factor(EnhanceHurt,
    levels = c("Detract", "Neither", "Enhance")
  )),
  ylim = c(0, 1)
) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = count), position = position_dodge(width = 1), vjust = 1.5) +
  labs(x = "Ethnicity", y = "Proportion of Respondents", fill = "Enhance or Hurt") +
  ggtitle("Enhance/Detract from their work vs. Ethnicity") +
  theme(plot.title = element_text(hjust = 0.5), legend.position = "bottom")

# Print the bar chart
enhancehurt_vs_ethnicity_plot
```



Based on the data presented in the chart, it is clear that Asians have a high level of confidence in the ability of AI to enhance their work efficiency. On the other hand, the responses of White respondents are divided between those who believe that AI will enhance their work efficiency, and those who believe that it will neither enhance nor detract from it. The remaining racial groups exhibit a lower level of confidence in the potential of AI to enhance their work efficiency compared to Asians. One possible explanation for this discrepancy is that Asians are more likely to work in industries such as technology and healthcare, where they have more exposure to AI and its potential benefits. This idea is further supported by the ethnicity vs. employment sector chart presented in the next section.

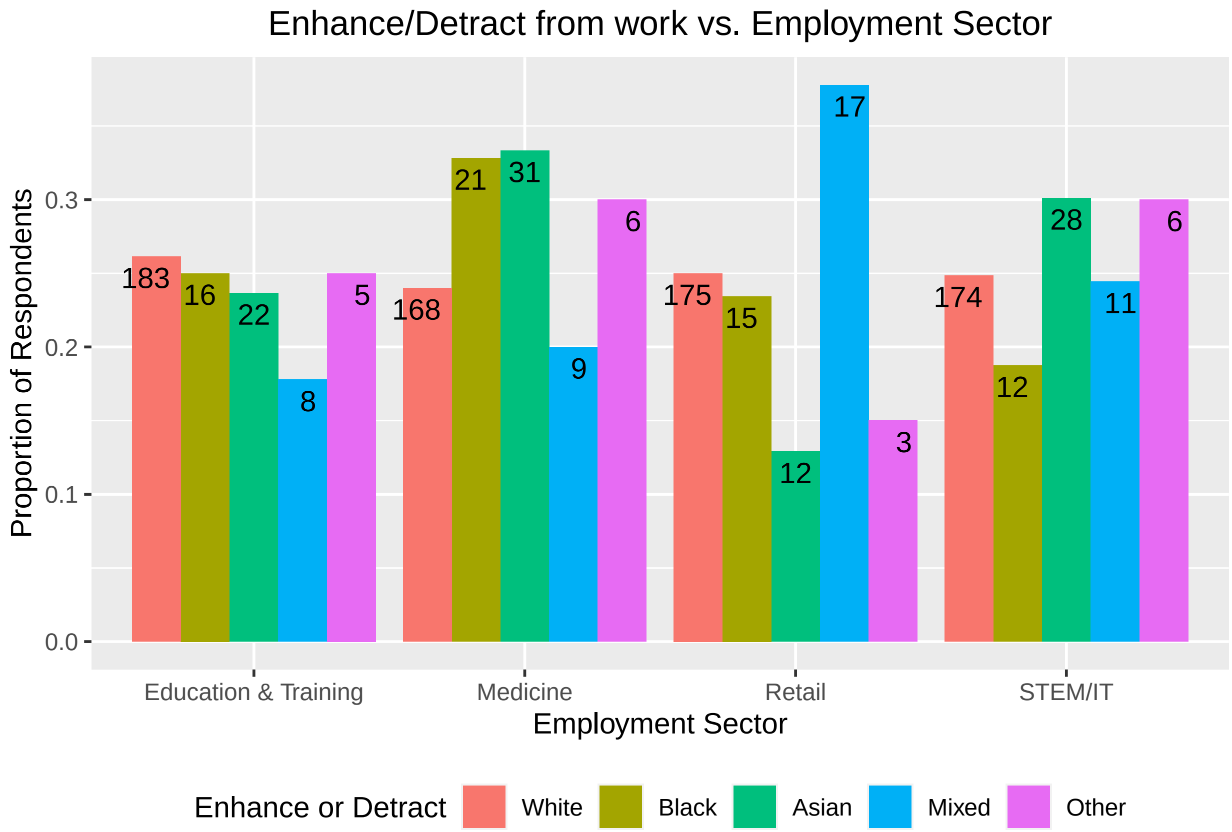
4.8. Employment Sector vs. Ethnicity

```r
# Create a new data frame with only the necessary columns
sector_vs_ethnicity <- combined %>% select(Employment.sector, Ethnicity)
```

```r
# For each Ethnicity, calculate the proportion of respondents by employment sector
sector_vs_ethnicity <- sector_vs_ethnicity %>%
  group_by(Ethnicity, Employment.sector) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))

# Create a bar chart to show the proportion of respondents by employment sector
#   split by ethnicity
sector_vs_ethnicity_plot <- ggplot(
  sector_vs_ethnicity,
  aes(x = Employment.sector, y = prop, fill = factor(Ethnicity,
    levels = c("White", "Black", "Asian", "Mixed", "Other")
  )),
  ylim = c(0, 1)
) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = count), position = position_dodge(width = 1), vjust = 1.5) +
  labs(
    x = "Employment Sector", y = "Proportion of Respondents", fill = "Enhance or Detract"
  ) +
  ggtitle("Enhance/Detract from work vs. Employment Sector") +
  theme(plot.title = element_text(hjust = 0.5), legend.position = "bottom")

# Print the bar chart
sector_vs_ethnicity_plot
```



The presented data reveals a significant overrepresentation of Asians in the healthcare and STEM/IT fields, while they are comparatively underrepresented in the retail sector. On the other hand, Black individuals are less prevalent in STEM/IT roles. The observed differences in ethnic proportions across the sectors may be attributed to various factors, including implicit biases in hiring practices, immigration trends, and education preferences. Additionally, the fact that white respondents are evenly distributed across all sectors may be a reflection of the majority of the survey participants being white. Overall, these findings highlight the ongoing need for greater diversity and equity in the workplace across all industries.

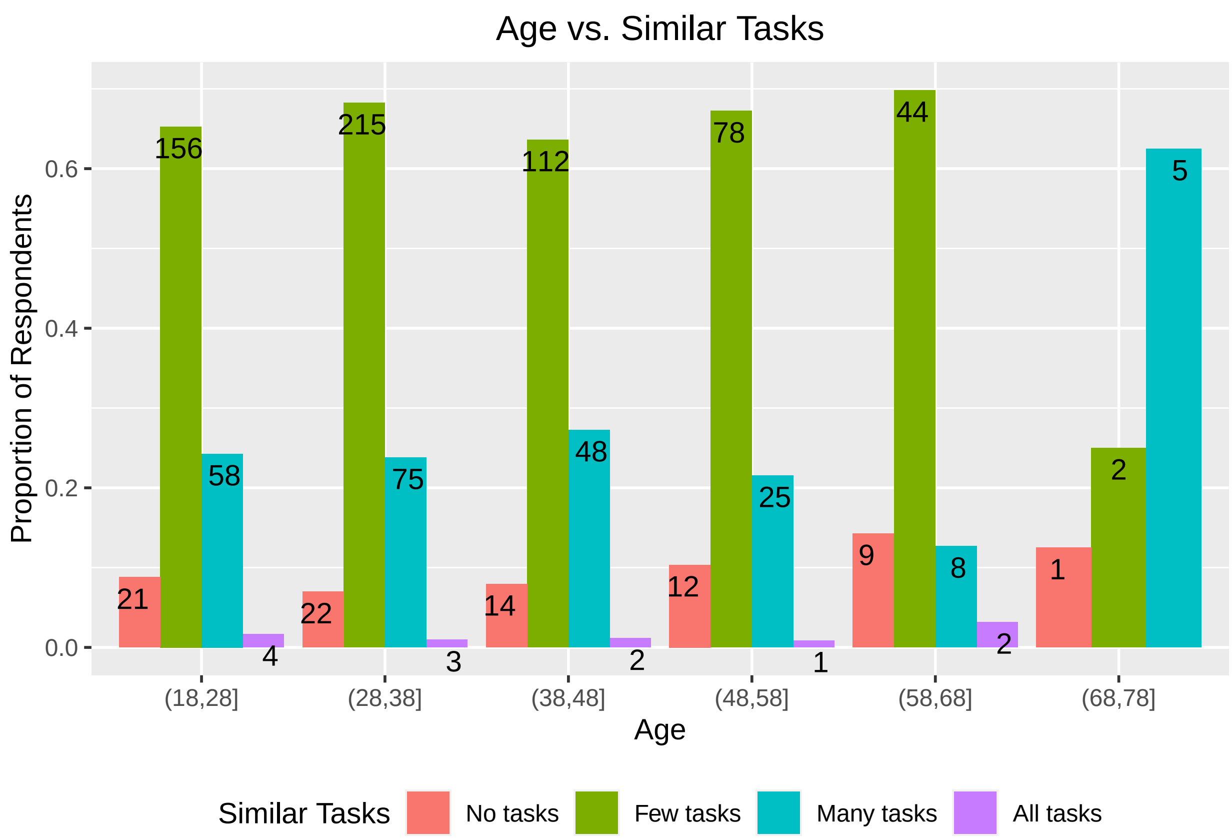
4.9. SimilarTasks vs. Age

```r
# Create a new data frame with only the necessary columns
similartasks_vs_age <- combined %>% select(SimilarTasks, Age)
```

```r
# For each Age group, calculate the proportion of respondents who
#   think AI can do similar human tasks
similartasks_vs_age <- similartasks_vs_age %>%
  group_by(Age, SimilarTasks) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))

# Create a bar chart to show the proportion of respondents who think AI
#   can do similar human tasks by age group
similartasks_vs_age_plot <- ggplot(
  similartasks_vs_age,
  aes(x = Age, y = prop, fill = factor(SimilarTasks,
    levels = c("No tasks", "Few tasks", "Many tasks", "All tasks")
  )),
  ylim = c(0, 1)
) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = count), position = position_dodge(width = 1), vjust = 1.5) +
  labs(x = "Age", y = "Proportion of Respondents", fill = "Similar Tasks") +
  ggtitle("Age vs. Similar Tasks") +
  theme(plot.title = element_text(hjust = 0.5), legend.position = "bottom")

# Print the bar chart
similartasks_vs_age_plot
```



This chart depicts a striking contrast between different age groups' perceptions of AI's capabilities. Specifically, the data shows that older participants are more skeptical of AI's potential and believe that it cannot perform any tasks that humans can. In contrast, middle-aged participants are more optimistic and believe that AI is capable of performing many tasks. Interestingly, all age groups are in agreement that AI can perform only a few tasks that humans can. This finding suggests that exposure to AI technology may be a key factor influencing one's perception of its capabilities. For instance, the older generation, which may have had less exposure to AI, is more skeptical, while the middle-aged generation, which may have more exposure, is more optimistic. On the other hand, the younger generation, who have grown up with AI, are also skeptical of its capabilities, likely because they have experienced its limitations firsthand.