Artificial Intelligence Opinion Survey

DATA 490 Independent Study

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

April 20, 2023

Contents

1. Load Data

```
library(tidyverse)
library(ggplot2)

# 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

```
# Combine data into one data frame after mutating Age to be one data type
edu <- edu %>% mutate(Age = as.character(Age))
health <- health %>% mutate(Age = as.character(Age))
```

```r
retail <- retail %>% mutate(Age = as.character(Age))
tech <- tech %>% mutate(Age = as.character(Age))

combined <- bind_rows(edu, health, retail, tech)
# export combined data to csv
# write.csv(combined, "combined_non_qualtrics.csv")

# combine qualtrics and combined data using qualtrics data's
#     ProlificID column and combined data's Participant id
combined <- left_join(qualtrics, combined,
  by = c("ProlificID" = "Participant.id")
)
# rename "Duration..in.seconds." column to "Duration"
colnames(combined)[
  colnames(combined) == "Duration..in.seconds."
] <- "Duration"
# rename "Ethnicity.simplified" column to "Ethnicity"
colnames(combined)[
  colnames(combined) == "Ethnicity.simplified"
] <- "Ethnicity"
# remove Age.x and keep only Age.y column and rename Age.y to Age
combined <- combined %>%
  select(-Age.x) %>%
  rename(Age = Age.y)
# remove Status.x and Status.y columns
combined <- combined %>%
  select(-Status.x) %>%
  select(-Status.y)
# remove Finished, Progress, UserLanguage, DistributionChannel,
# Nationality, and Consent columns
combined <- combined %>%
  select(-Finished) %>%
  select(-Progress) %>%
  select(-UserLanguage) %>%
  select(-DistributionChannel) %>%
  select(-Nationality) %>%
  select(-Consent)


# remove rows where Submission.id is NA
combined <- combined %>%
  filter(!is.na(Submission.id))
# Keep only rows which say "United States" in "Country.of.residence" column
combined <- combined %>%
  filter(combined$Country.of.residence == "United States")

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

```r
] <- "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 little time" to "little time",
# "Save no time" to "no time",
# "Save some time" to "some time", and
# "Neutral" to "neutral"
combined$TimeEnergy[
  combined$TimeEnergy == "Save a lot of time"
] <- "lot of time"
combined$TimeEnergy[
  combined$TimeEnergy == "Save little time"
] <- "little time"
combined$TimeEnergy[
  combined$TimeEnergy == "Save no time"
] <- "no time"
combined$TimeEnergy[
  combined$TimeEnergy == "Save some time"
] <- "some time"
combined$TimeEnergy[
  combined$TimeEnergy == "Neutral"
] <- "neutral"

# export data to csv
# write.csv(combined, "combined_qualtrics.csv")
```

```r
# Keep only rows which say "Compose an email" in "Attention" column
combined <- combined %>% filter(combined$Attention == "Compose an email")
# remove Attention column
combined <- combined %>% select(-Attention)
# export data to csv
# write.csv(combined, "combined_qualtrics_attentive.csv")
```

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

```r
# Create a new data frame with only the columns we need
avg_time <- combined %>% select(Duration)
# Remove rows where Duration is NA
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 <x> minutes and <y> 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
avg_time
```

```
##                     avg_time
## 1 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. EnhanceHurt vs. Sector

```r
# Create a new data frame with only the columns we need
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))
# Visualize using different histogram for each sector.
# Show the number of respondents inside each bar.
# make legend bottom. Wrap x axis labels without changing plot size.
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 Hurt"
  ) +
  ggtitle("Proportion of Respondents Who Think AI Can\n
  Enhance Their Work Efficiency by Employment Sector") +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = "bottom"
  )

enhancehurt_vs_sector_plot
```

Proportion of Respondents Who Think AI Can

Enhance Their Work Efficiency by Employment Sector
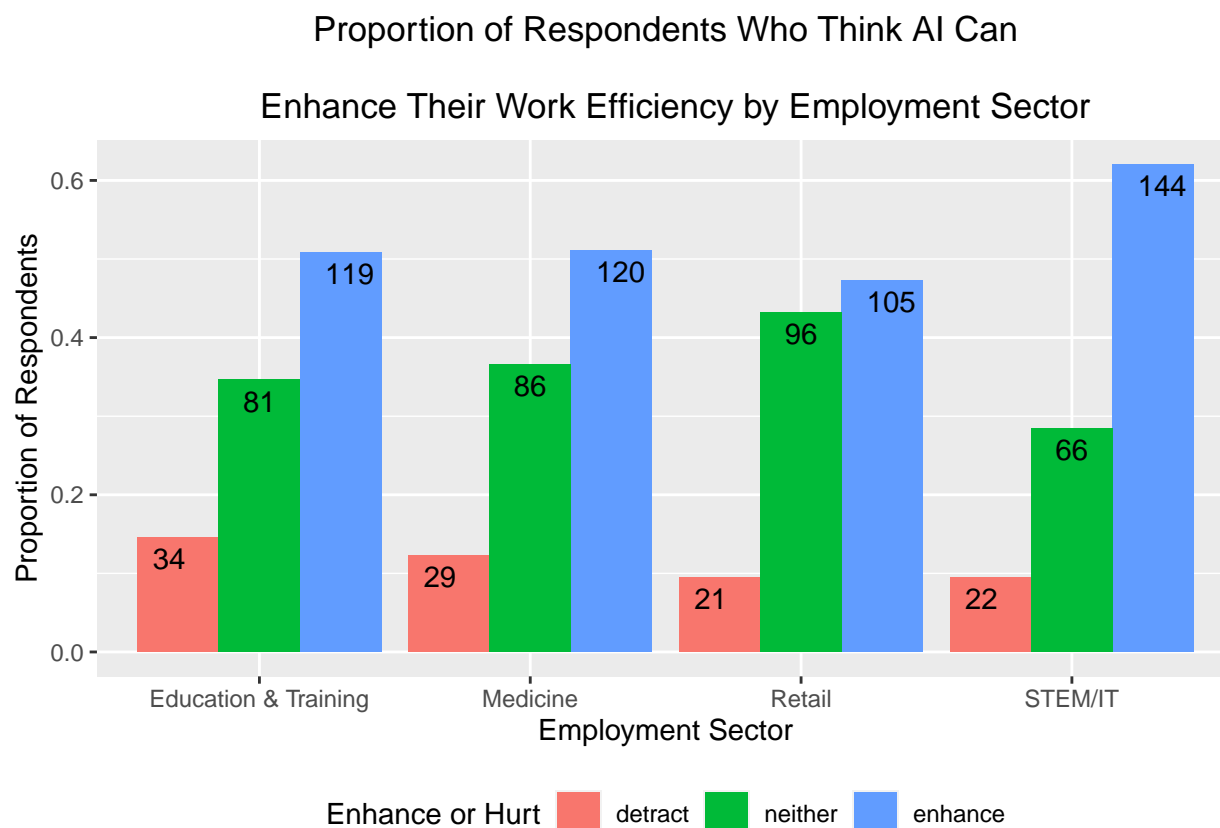
Respondents from the STEM/IT are most confident that AI will enhance their work efficiency. Respondents from the Education and Training, and Medicine sectors believe that AI will detract from their work efficiency. Respondents from the Retail sector are most unsure about whether AI will enhance or detract from their work efficiency. This is likely because the Retail sector is the most diverse sector, and includes a wide range of jobs, from sales assistants to store managers. This means that the impact of AI on the Retail sector is likely to be mixed, with some jobs being enhanced, and others being detracted. This is in contrast to the STEM/IT sector, where the impact of AI is likely to be more consistent across jobs in the STEM/IT sector, as the jobs in this sector are more similar to each other.

The Education and Training sector is also likely to be more consistent, as the jobs in this sector are also more similar to each other. The Medicine sector is likely to be more diverse, but the impact of AI is likely to be more consistent, as the jobs in this sector are more specialized.

4.3. EnhanceHurt vs. Education

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

```r
# 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))
# Visualize using different histogram for each education level.
# Show the number of respondents on each bar.
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 Hurt"
  ) +
  ggtitle("Proportion of Respondents Who Think AI Can\n
  Enhance Their Work Efficiency by Education Level") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "bottom")

enhancehurt_vs_education_plot
```

Proportion of Respondents Who Think AI Can

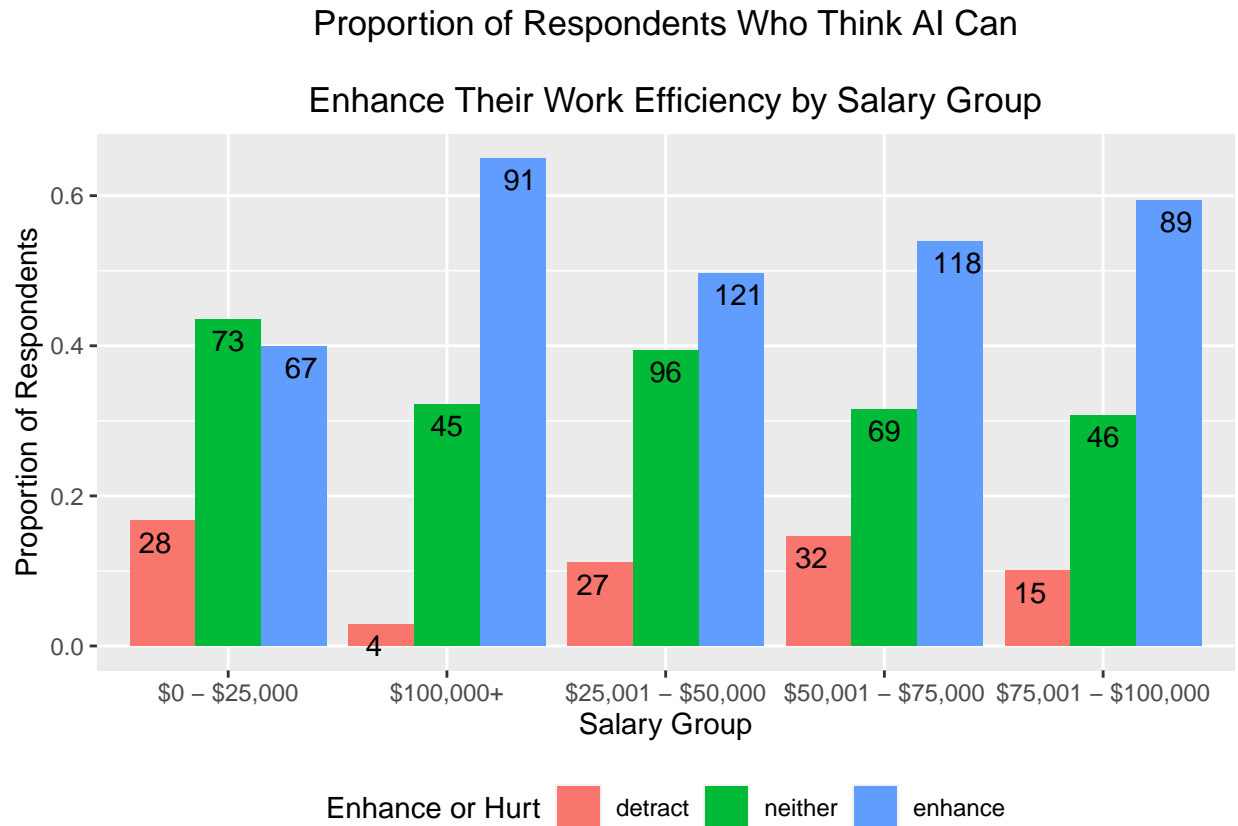## Enhance Their Work Efficiency by Education Level
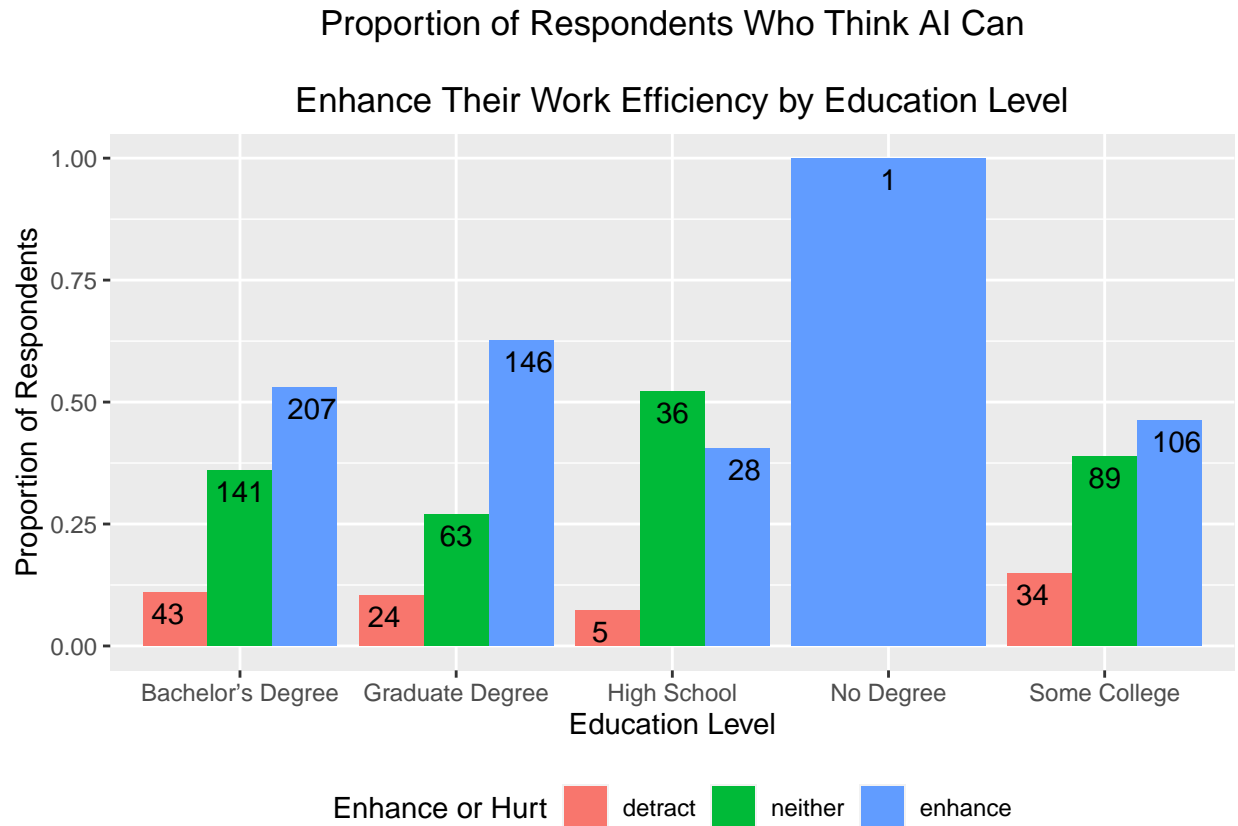
Those with bachelor's and graduate degrees believe that AI will enhance their work efficiency. On the other hand, those with some college education background relatively believe less that AI will enhance their work efficiency. Those with high school education responded mostly that AI will neither enhance nor detract from their work. Since there was only one respondent who had no education background. We can see that as a respondent has a higher education background, more they believe that AI will enhance their work efficiency. This is likely because those with higher education background are more likely to have jobs that require more complex tasks, which are more likely to be enhanced by AI. On the other hand, those with lower education background are more likely to have jobs that require less complex tasks, which are more likely to be unaffected by AI.

4.4. EnhanceHurt vs. Salary

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

```r
# Create a new dataframe with the proportion of respondents who
# think AI can enhance their work efficiency for each salary group
enhancehurt_vs_salary <- enhancehurt_vs_salary %>%
  group_by(Salary, EnhanceHurt) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))
# Visualize using different histogram for each salary group.
# Show the number of respondents on each bar.
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 Hurt"
  ) +
  ggtitle("Proportion of Respondents Who Think AI Can\n
  Enhance Their Work Efficiency by Salary Group") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "bottom")

enhancehurt_vs_salary_plot
```

Proportion of Respondents Who Think AI Can

## Enhance Their Work Efficiency by Salary Group

This data shows that as people earn more salaries, the more they believe that AI will enhance their work efficiency. This is likely because those who earn more salaries are more likely to have jobs that require more specific tasks, which are more likely to be enhanced by AI. An example is a doctor. A doctor is more likely to earn a higher salary than a cashier. A doctor's job requires more specific tasks, such as diagnosing patients, which are more likely to be enhanced by AI. On the other hand, a cashier's job requires less specific tasks, such as scanning items, which are less likely to be enhanced by AI.

4.5. EnhanceHurt vs. Gender

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

# Create a new dataframe with the proportion of respondents who
# think AI can enhance their work efficiency for each gender
enhancehurt_vs_gender <- enhancehurt_vs_gender %>%
  group_by(Gender, EnhanceHurt) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))
# Visualize the proportion of respondents who think AI can
# enhance or hurt, based on gender, using a bar chart.
# Show the number of respondents inside each bar.
# make legend bottom. Wrap x axis labels without changing plot size.
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 Hurt"
  ) +
  ggtitle("Proportion of Respondents Who Think AI Can\n
  Enhance Their Work Efficiency by Gender") +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = "bottom"
  )

enhancehurt_vs_gender_plot
```

Proportion of Respondents Who Think AI Can

Enhance Their Work Efficiency by Gender
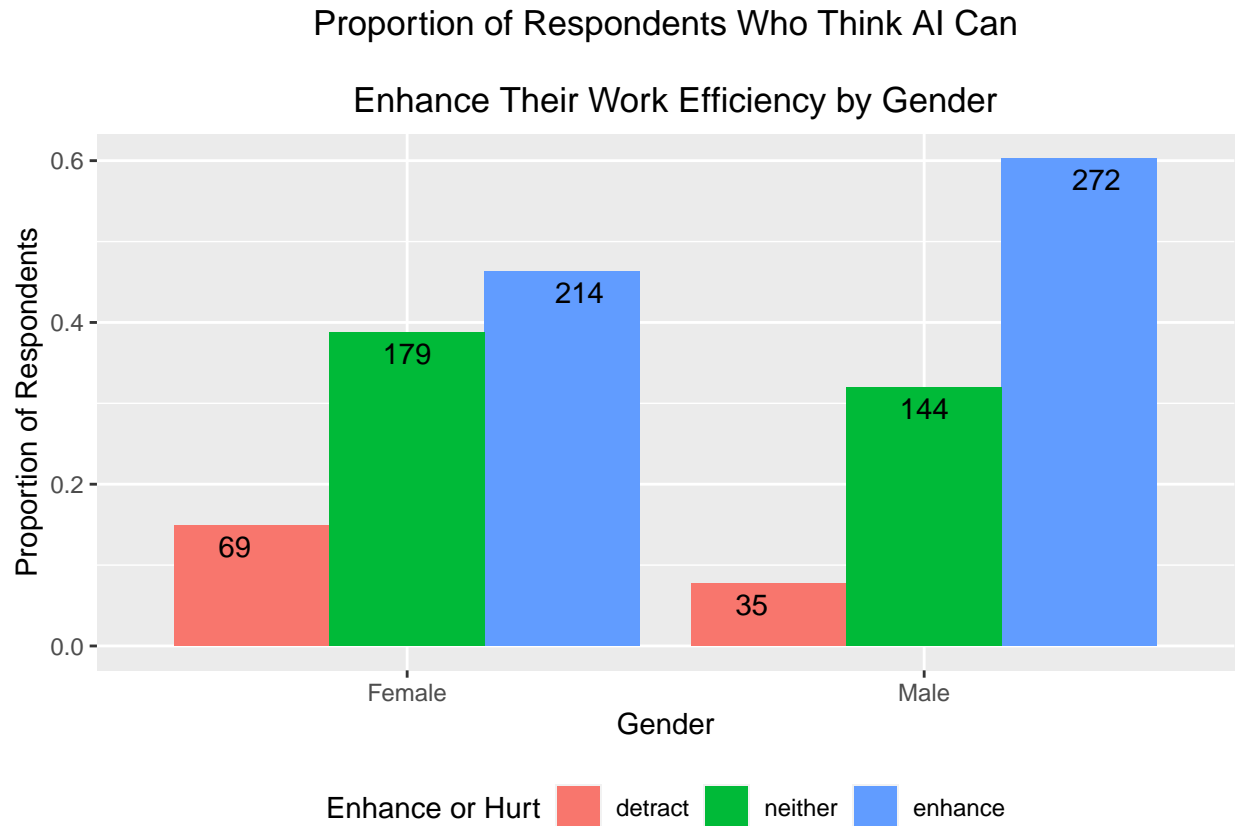
It is evident from this data that a higher proportion of males responded that AI will enhance their task than females. On the other hand, a higher proportion of females responded that AI will detract them from their tasks than males. This is likely because of gender disparities, evident in the next salary groups vs. gender chart. A lot of male respondents have higher salaries, whereas a lot of female respondents have lower salaries. Similarly, from the above EnhanceHurt vs. Salary chart in the previous section, it was evident that higher salaried individuals believed that AI would enhance their tasks. This trend highlights the difference in preferences due to pay inequality based on gender distribution.
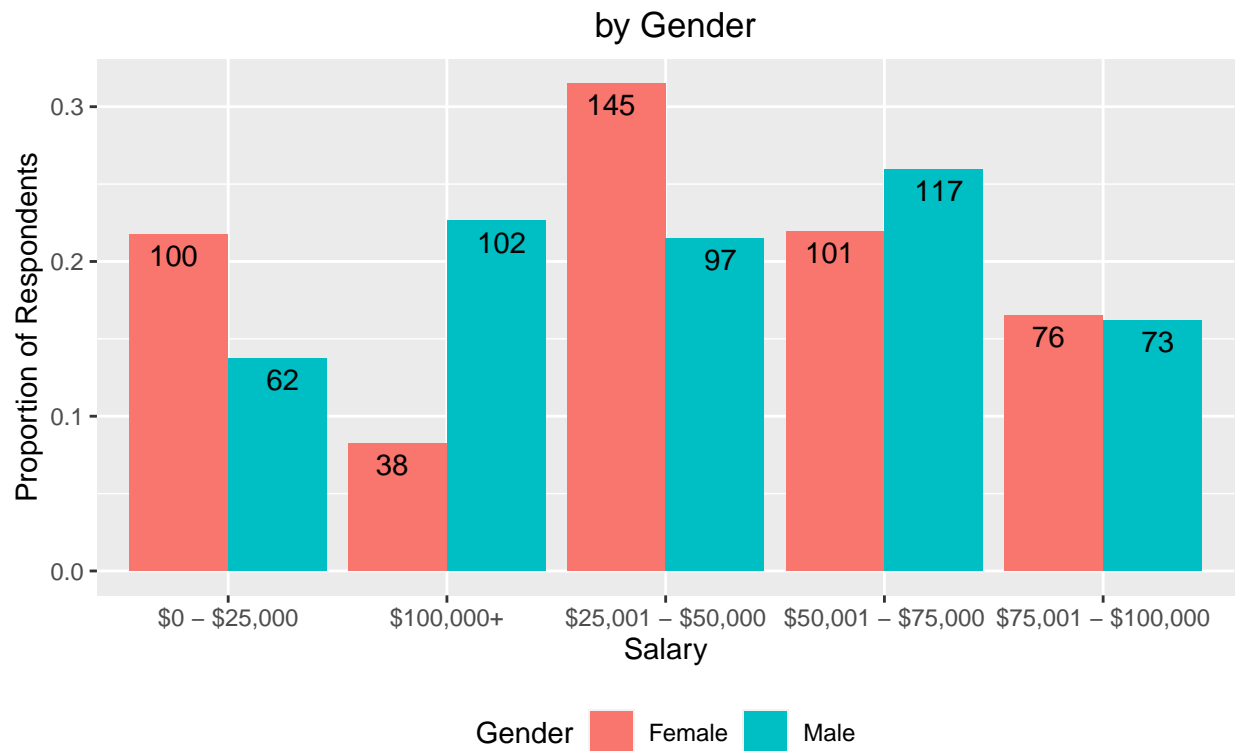
4.6. Salary vs. Gender

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

# Create a new dataframe with the proportion of respondents who
# think AI can enhance their work efficiency for each gender
salary_vs_gender <- salary_vs_gender %>%
  group_by(Gender, Salary) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))
# Visualize the proportion of respondents by salaries
# based on gender, using a bar chart.
# Show the number of respondents inside each bar.
# make legend bottom. Wrap x axis labels without changing plot size.
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("Proportion of Respondents in each salary group\n
  by Gender") +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = "bottom"
  )

salary_vs_gender_plot
```

Proportion of Respondents in each salary group

by Gender

This chart shows the gender inequality in salaries. Despite having almost the same number of respondents in both gender categories, it is clear that males dominate the higher salaries, and females dominate thhe lower salaries. This is possibly due to a variety of factors like willingness to take part in this survey, recruiting practices and biases, gender pay inequality, among others.

4.6. EnhanceHurt vs. Ethnicity

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

```r
# Create a new dataframe with the proportion of respondents who
# think AI can enhance their work efficiency for each Ethnicity
enhancehurt_vs_ethnicity <- enhancehurt_vs_ethnicity %>%
  group_by(Ethnicity, EnhanceHurt) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))
# Visualize the proportion of respondents who think AI can
# enhance or hurt, based on Ethnicity, using a bar chart.
# Show the number of respondents inside each bar.
# make legend bottom. Wrap x axis labels without changing plot size.
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("Proportion of Respondents Who Think AI Can\n
  Enhance Their Work Efficiency by Ethnicity") +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = "bottom"
  )

enhancehurt_vs_ethnicity_plot
```

Proportion of Respondents Who Think AI Can

Enhance Their Work Efficiency by Ethnicity
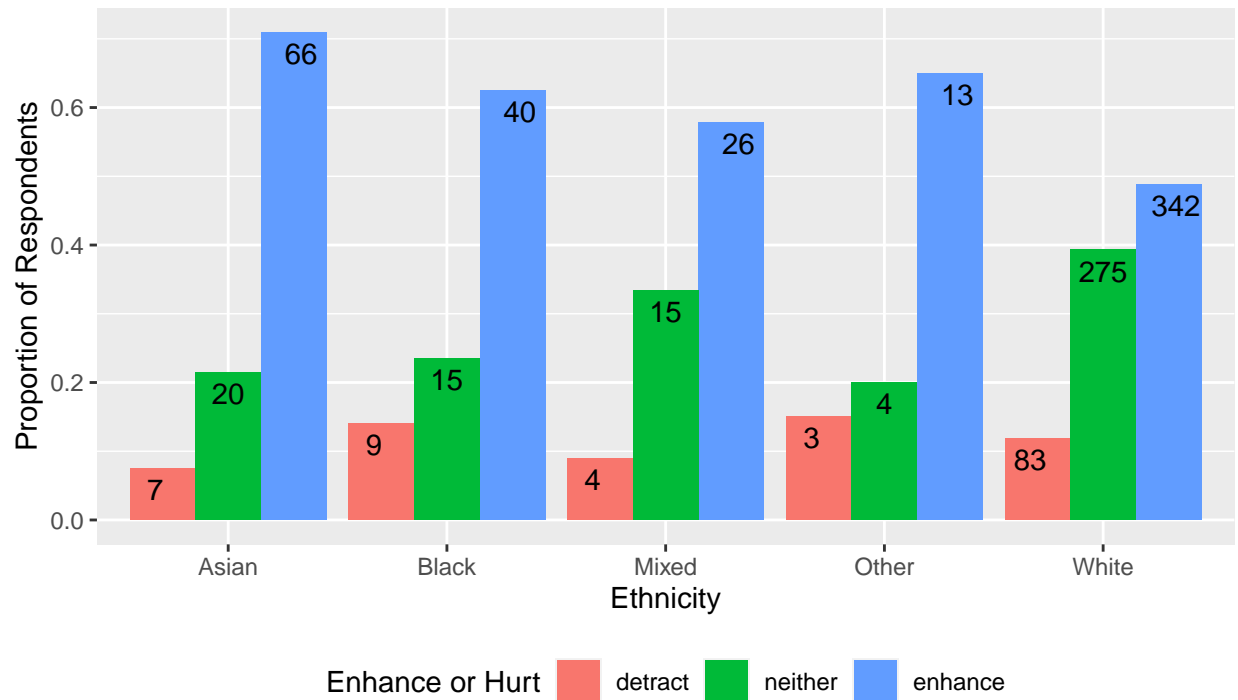
From this chart, it is evident that Asians are very confident that AI will enhance their work efficiency, but respondents who are White are split between whether it'll enhance or neither enhance or detract their work efficiency. The other races uniformly believe to a lesser degree than Asians that AI will enhance their work efficiency. This could be due to the fact that Asians are more likely to be in the tech and healthcare industries, and thus have more experience with AI, and thus are more confident that AI will enhance their work efficiency. This is reflected in the ethnicity vs. employment sector chart in the next section.

4.8. Employment Sector vs. Ethnicity

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

```r
# Create a new dataframe with the proportion of respondents in each
# ethnicity by sector
sector_vs_ethnicity <- sector_vs_ethnicity %>%
  group_by(Ethnicity, Employment.sector) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))

# Visualize the proportion of respondents' ethnicities based
# on sector, using a bar chart.
# Show the number of respondents inside each bar.
# make legend bottom. Wrap x axis labels without changing plot size.
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 = "Sector",
    y = "Proportion of Respondents",
    fill = "Enhance or Hurt"
  ) +
  ggtitle("Proportion of Respondents by Ethnicity\n
  grouped by Employment Sector") +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = "bottom"
  )

sector_vs_ethnicity_plot
```

Proportion of Respondents by Ethnicity

grouped by Employment Sector
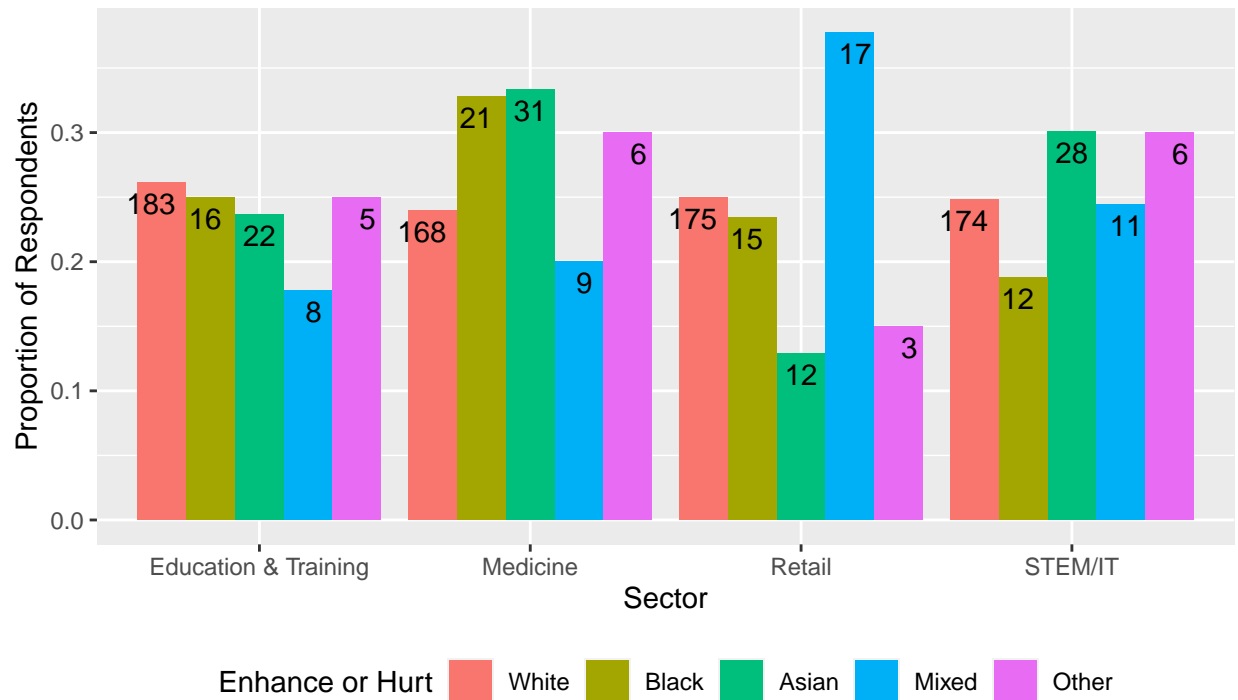
This chart shows that Asians dominate the healthcare and STEM/IT sectors of the survey participants. Asians occupy the least proportion of the Retail sector, and Black people the STEM/IT sector. Racial biases and hiring practices, but also immigration and education preferences may be possible reasons for the discrepency in the ethnicity proportion of respondents in each sector. White people are uniformly distributed among the four sectors, which may be a result of the majority of the respondents being White.

4.8. SimilarTasks vs. Age

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

# Visualize the proportion of respondents who think AI can
# do similar tasks, based on age group, using a bar chart.
# Show the number of respondents inside each bar.
# make legend bottom. Wrap x axis labels without changing plot size.
# order the bars as "Cannot perform any tasks", "Can perform only a few tasks",
# "Can perform many tasks", "Can perform all tasks".
# force the y axis to start at 0 and end at 1.
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("Proportion of Respondents Who Think AI Can\n
  Do Similar Tasks by Age Group") +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = "bottom"
  )

similartasks_vs_age_plot
```

Proportion of Respondents Who Think AI Can

Do Similar Tasks by Age Group
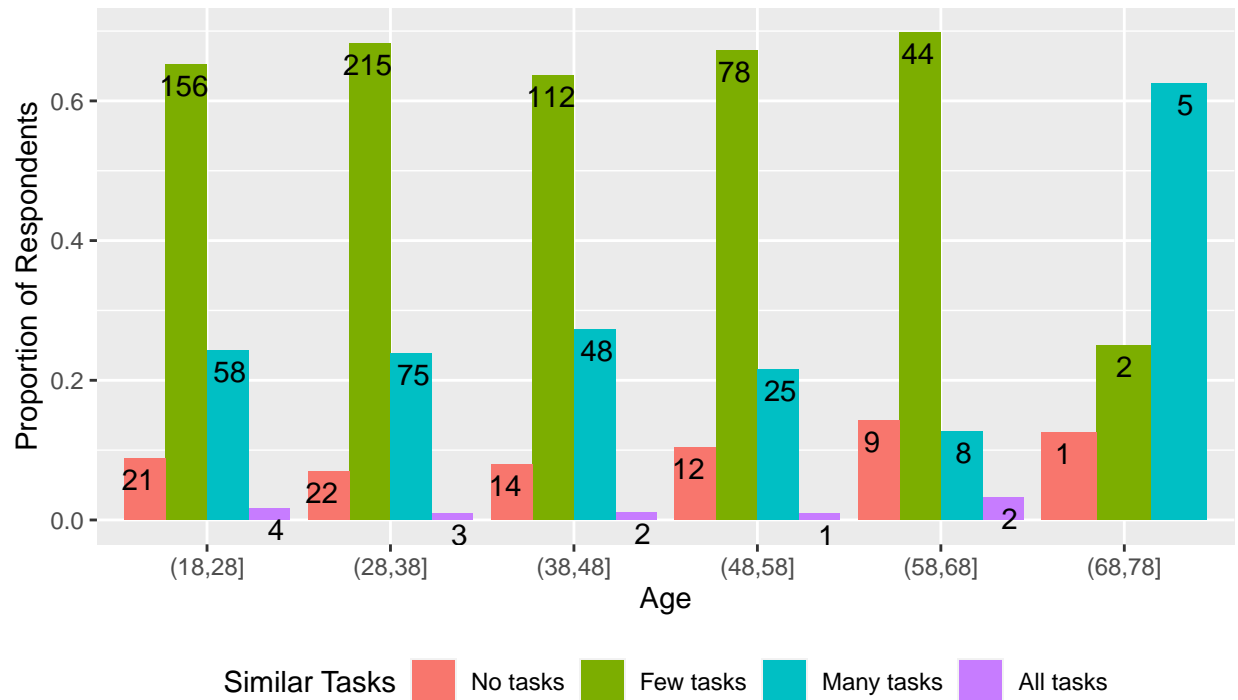
This chart shows that older participants think AI can perform no tasks that humans can, while middle aged participants think that AI can perform many tasks. All age groups mostly and uniformly think that AI can perform few tasks that humans can. This is an interesting result, as it shows that the older generation is more skeptical of AI's capabilities, while the middle aged generation is more optimistic. This may be due to the fact that the older generation has less exposure to AI, and the middle aged generation has more exposure to AI. The younger generation is more skeptical of AI's capabilities, which may be due to the fact that they have grown up with AI, and have seen its limitations.