Supplementary Materials

Factors Shaping Public Perception of the Impact of Artificial Intelligence on Employment: A Nationally Representative Study

S1. Sample Derivation and Missing Data

The 2024 General Social Survey cumulative data file contains 75,699 records. The artificial intelligence module was administered to one ballot subset; 1,536 respondents were administered and answered the outcome item (AIWORRY). Listwise deletion was applied across all model covariates, excluding 462 respondents with missing data on one or more covariates and yielding an analytic sample of N = 1,074. Respondents who reported never using the internet (INTRNETUSE = 7, n = 64 of the ballot subsample) were treated as missing because they cannot be placed on the 1–6 internet-use frequency scale used in the model.

Table S0 reports the number of respondents with missing data on each variable, among the 1,536 ballot respondents who answered the outcome item. Because a respondent may be missing on more than one variable, these counts do not sum to the total exclusion of 462.

Table S0. *Missingness by variable (of 1,536 ballot respondents).*

Variable	n missing
GENDTECH (men vs. women benefiting)	221
EDUCTECH (educated vs. less educated benefiting)	186
CLASSTECH (rich vs. poor benefiting)	141
INTSKILL (software-learning skill)	67
INTRNETUSE (never, or missing)	59
AGE	41
HARMGOOD1 (technology does more harm)	35
AIMED (comfort: medical robot)	34
RACECEN1	28
TECHESY (technology provides opportunities)	27
AIDRIVE (comfort: driverless car)	17
HISPANIC	7
EDUC	4
SEX	2
AIWORRY (outcome)	0

S2. Variable Recoding

The outcome (AIWORRY) was dichotomized: “very worried” or “somewhat worried” were coded as worried (1); “neither worried nor unworried,” “not very worried,” and “not at all worried” were coded as not worried (0). For the ordinal analysis (S6), the five-level response was retained as an ordered factor.

Race/ethnicity was collapsed into four categories (Hispanic; non-Hispanic White; non-Hispanic Black; non-Hispanic Other), with non-Hispanic Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, and other identifications combined into non-Hispanic Other because individual cell counts were too sparse for stable estimation. The “strongly disagree” and “disagree” categories of the technology-opportunities item were collapsed (combined n = 35), and the “very bad” and “bad” categories of software-learning skill were collapsed (combined n = 36). Internet-use frequency was recoded to a 1–6 scale (1 = less often, 6 = almost all the time), excluding “never.” Age was grouped into 18–29, 30–54, 55–64, and 65+. Years of schooling was retained as continuous. The complete recoding and analysis script is provided as a separate file.

S3. Weighted Descriptive Characteristics (Table 1)

Table S1. *Weighted descriptive characteristics of the analytic sample (N = 1,074). Categorical variables: unweighted n (weighted %). Continuous variables: median (Q1, Q3).*

Characteristic	n / median	Weighted % / IQR
Age group		
18–29	166	16%
30–54	486	45%
55–64	176	16%
65+	246	23%
Sex		
Female	599	56%
Male	475	44%
Race/ethnicity		
Hispanic	153	14%
NH White	681	64%
NH Black	175	16%
NH Other	65	9%
Years of schooling	14	(12, 16)
Rich vs. poor benefiting		
Rich more	654	61%
Equal	276	26%
Poor more	38	3%
Neither	106	10%
Educated vs. less educated benefiting		
Educated more	563	53%
Equal	327	30%
Less educated more	84	8%
Neither	100	9%
Men vs. women benefiting		
Men more	109	10%
Equal	621	58%
Women more	145	14%
Neither	199	18%
Technology does more harm than good		
Strongly disagree	47	4%
Disagree	213	20%
Neither	325	30%
Agree	378	35%
Strongly agree	111	10%

Characteristic	n / median	Weighted % / IQR
Skill at learning new software		
Bad or worse	36	3%
Neither	120	11%
Good	301	28%
Very good	617	57%
Internet use frequency (1–6)		
1 (less often)	12	1%
2	20	2%
3	29	3%
4	46	4%
5	397	37%
6 (almost all the time)	570	53%
Technology provides more opportunities		
Strongly agree	353	33%
Agree	555	52%
Neither	131	12%
Disagree or less	35	3%
Comfort: medical robot (0–10)	3	(0, 6)
Comfort: driverless car (0–10)	2	(0, 5)

Note. Weighted percentages use the GSS post-stratification weight (wtssps) and may not sum to 100% due to rounding.

S4. Weighted Prevalence of Worry (Table 2)

Table S2. *Weighted prevalence of worry that AI will take over many jobs, by characteristic. Confidence intervals are logit-transformed and bounded within 0–100%.*

Group	Level	Prevalence	95% CI
Overall	Overall	69.1%	65.2–72.8%
Age	18–29	59.5%	49.6–68.7%
Age	30–54	69.8%	64.3–74.9%
Age	55–64	76.0%	67.1–83.1%
Age	65+	73.3%	65.5–79.9%
Sex	Female	75.2%	70.1–79.6%
Sex	Male	62.5%	57.1–67.7%
Race/ethnicity	Hispanic	69.6%	59.2–78.3%
Race/ethnicity	NH White	70.1%	65.3–74.6%
Race/ethnicity	NH Black	65.9%	55.8–74.8%
Race/ethnicity	NH Other	65.6%	41.1–83.9%
Tech opportunities	Strongly agree	58.6%	52.0–64.9%
Tech opportunities	Agree	72.9%	67.4–77.8%
Tech opportunities	Neither	76.0%	63.6–85.2%
Tech opportunities	Disagree or less	86.5%	— †
Driverless car comfort	0–2	77.4%	73.0–81.3%
Driverless car comfort	3–5	60.9%	53.0–68.2%
Driverless car comfort	6–10	58.5%	47.0–69.1%

† *The "disagree or less" category of the technology-opportunities item has n = 35 with an observed prevalence near the boundary; the logit-transformed confidence interval did not converge and is therefore not reported. This estimate should be interpreted with caution.*

S5. Primary Model: Design-Based Binary Logistic Regression (Table 3)

Odds ratios from a design-based binary logistic regression (svyglm, quasibinomial family) incorporating stratification (vstrat), clustering (vpsu), and post-stratification weights (wtssps). Variance was estimated by Taylor-series linearization; single-PSU strata were centered at the grand mean. All tests are two-tailed at $\alpha = 0.05$.

Table S3. *Binary logistic regression odds ratios (N = 1,074).*

Variable	OR	95% CI	p
Age 30–54 (ref: 18–29)	1.34	0.83–2.17	0.219
Age 55–64	2.10	0.97–4.55	0.058
Age 65+	1.67	0.89–3.13	0.107
Sex: Male (ref: Female)	0.71	0.49–1.03	0.069
NH White (ref: Hispanic)	0.96	0.52–1.76	0.881
NH Black	0.68	0.37–1.26	0.215
NH Other	0.77	0.36–1.66	0.498
Years of schooling	0.96	0.90–1.02	0.174
Rich vs. poor: Equal (ref: Rich more)	0.91	0.53–1.54	0.713
Poor more	0.95	0.35–2.60	0.922
Neither	1.44	0.62–3.33	0.381
Educated vs. less: Equal (ref: Educated more)	0.64	0.41–1.01	0.056
Less educated more	1.06	0.47–2.38	0.880
Neither	0.43	0.20–0.94	0.036*
Men vs. women: Equal (ref: Men more)	1.07	0.58–1.99	0.829
Women more	0.89	0.41–1.89	0.748
Neither	0.92	0.44–1.95	0.826
Tech harm (linear contrast)	0.83	0.39–1.75	0.614
Tech harm (quadratic)	0.92	0.48–1.78	0.805
Tech harm (cubic)	1.62	0.95–2.78	0.076
Tech harm (quartic)	0.83	0.54–1.29	0.409
Software skill: Neither (ref: Bad or worse)	1.22	0.36–4.21	0.741
Good	1.38	0.44–4.33	0.568
Very good	1.31	0.41–4.13	0.639
Internet use frequency (continuous)	1.26	1.02–1.54	0.029*
Tech opportunities: Agree (ref: Strongly agree)	1.64	1.07–2.52	0.023*
Neither	1.69	0.82–3.46	0.148
Disagree or less	2.32	0.76–7.12	0.136
Comfort: medical robot (0–10)	0.95	0.89–1.01	0.122
Comfort: driverless car (0–10)	0.92	0.85–0.99	0.030*

* $p < 0.05$.

S5.1 Multicollinearity

Generalized variance inflation factors (GVIF) were computed on an unweighted logistic-regression analog of the primary model. All GVIF^(1/(2·Df)) values were below 1.25, indicating no meaningful collinearity.

Table S4. *Generalized variance inflation factors.*

Variable	GVIF	Df	GVIF^(1/(2·Df))
age_grp	1.43	3	1.06
sex	1.12	1	1.06
race4	1.31	3	1.05
educ_yrs	1.24	1	1.11
tech_rich	2.41	3	1.16
tech_educ	2.51	3	1.17
tech_sex	1.50	3	1.07
tech_harm	1.56	4	1.06
learn_sw	1.55	3	1.08
int_use_num	1.34	1	1.16
tech_opps	1.53	3	1.07
robot_med	1.46	1	1.21
robot_car	1.53	1	1.24

S6. Ordinal Model and the Proportional-Odds Assumption

An ordinal (proportional-odds) logistic regression was considered as an alternative specification, since the outcome is a five-level ordered measure of worry. We evaluated whether the proportional-odds assumption — that each predictor has a constant effect across all cumulative cutpoints of the outcome — was tenable. It was not, and the ordinal model was therefore not used for the primary analysis.

S6.1 Brant test of the proportional-odds assumption

A Brant test was conducted on an unweighted ordinal logistic regression (MASS::polr) analog of the model, since no design-based Brant test is available; this mirrors the unweighted approximation used for the collinearity diagnostic (S5.1). The omnibus test rejected the proportional-odds assumption for the model as a whole, $\chi^2(90) = 179.97$, $p = 5.8 \times 10^{-8}$. Examined term by term, the violation was concentrated in software-learning skill, the quadratic contrast of the technology-harm item, and non-Hispanic White race/ethnicity.

Table S5. *Brant test of the parallel-regression (proportional-odds) assumption. H_0: the assumption holds.*

Term	χ^2	df	p
Omnibus	179.97	90	<0.001
age_grp 30–54	0.78	3	0.854
age_grp 55–64	0.15	3	0.985
age_grp 65+	0.43	3	0.933
sex Male	3.72	3	0.293
race NH White	15.27	3	0.002*
race NH Black	0.33	3	0.955
race NH Other	4.03	3	0.258
educ_yrs	0.80	3	0.849
tech rich Equal	3.96	3	0.265
tech_rich Poor more	3.64	3	0.303
tech rich Neither	3.68	3	0.299
tech_educ Equal	0.42	3	0.935
tech_educ Less educated more	3.31	3	0.347
tech_educ Neither	8.91	3	0.031*
tech_sex Equal	2.46	3	0.483
tech_sex Women more	4.71	3	0.194
tech_sex Neither	0.75	3	0.862
tech_harm (linear)	7.81	3	0.050
tech_harm (quadratic)	17.99	3	<0.001*
tech_harm (cubic)	0.78	3	0.855
tech_harm (quartic)	6.10	3	0.107
learn_sw Neither	5.37	3	0.147
learn_sw Good	14.48	3	0.002*

Term	χ^2	df	p
learn_sw Very good	13.55	3	0.004*
int_use_num	1.83	3	0.609
tech_opps Agree	5.78	3	0.123
tech_opps Neither	2.80	3	0.423
tech_opps Disagree or less	0.87	3	0.833
robot_med	1.71	3	0.636
robot_car	1.31	3	0.727

** $p < 0.05$. Test conducted on an unweighted polr analog of the design-based model.*

S6.2 Cutpoint coefficient comparison

As an informal complement to the Brant test, the model was refit as four separate binary logistic regressions, one at each cumulative cutpoint of the ordered outcome. Under proportional odds, a predictor's coefficient should be approximately constant across the four columns. Several predictors instead show substantial movement — most clearly the software-skill terms, whose coefficients decay from strongly positive at the lowest cutpoint toward zero (and sign reversal) at the highest. The four cutpoints are: Cut1 = at least "not very worried"; Cut2 = at least "neither"; Cut3 = at least "somewhat worried"; Cut4 = "very worried."

Table S6. *Logistic regression coefficients across four binary cutpoints of the worry outcome (selected terms).*

Term	Cut1	Cut2	Cut3	Cut4
learn_sw Neither	2.21	1.49	0.20	-0.06
learn_sw Good	1.97	0.95	0.32	0.47
learn_sw Very good	2.01	0.97	0.27	0.77
tech_opps Agree	1.81	0.34	0.50	0.43
tech_opps Neither	2.20	0.93	0.52	0.35
tech_opps Disagree or less	2.58	1.05	0.84	0.71
tech_harm (linear)	0.27	-0.86	-0.19	-1.46
race NH White	0.33	0.12	-0.05	-1.06
int_use_num	0.06	0.29	0.23	0.08
robot_car	-0.13	-0.14	-0.08	-0.11

Note. Coefficients on the log-odds scale. Stable predictors (e.g., int_use_num, robot_car) vary little across cutpoints; unstable predictors (learn_sw, tech_opps) change substantially, consistent with the Brant test.

S6.3 Ordinal model estimates (not used)

For completeness, the ordinal proportional-odds model estimates are reported below. Because the proportional-odds assumption was rejected (S6.1–S6.2), these estimates are not interpreted in the main text and are provided only for transparency. Odds ratios represent the multiplicative change in the odds of being in a higher worry category; the ordered outcome runs from "not at all worried" to "very worried."

Table S7. *Ordinal logistic regression odds ratios (abandoned; proportional-odds assumption not met).*

Variable	OR	95% CI	p
Age 30–54	1.48	0.96–2.28	0.077
Age 55–64	2.09	1.19–3.69	0.011*
Age 65+	1.71	1.01–2.89	0.044*
Sex: Male	0.73	0.54–0.99	0.043*
NH White	0.56	0.31–1.03	0.061
NH Black	0.60	0.31–1.15	0.122
NH Other	0.43	0.23–0.81	0.008*
Years of schooling	0.95	0.90–1.00	0.066
Rich vs. poor: Equal	1.06	0.70–1.59	0.793
Poor more	0.93	0.42–2.09	0.865
Neither	0.97	0.53–1.77	0.917
Educated vs. less: Equal	0.63	0.46–0.87	0.004*
Less educated more	1.06	0.55–2.08	0.854
Neither	0.78	0.38–1.59	0.488
Men vs. women: Equal	1.00	0.53–1.87	0.994
Women more	0.87	0.40–1.89	0.728
Neither	0.89	0.44–1.77	0.732
Tech harm (linear)	0.42	0.19–0.96	0.039*
Tech harm (quadratic)	1.34	0.70–2.57	0.378
Tech harm (cubic)	1.21	0.74–2.00	0.445
Tech harm (quartic)	0.98	0.66–1.46	0.928
Software skill: Neither	1.36	0.39–4.81	0.631
Good	1.68	0.50–5.68	0.405
Very good	1.81	0.52–6.34	0.352
Internet use frequency	1.23	0.98–1.53	0.069
Tech opportunities: Agree	1.70	1.17–2.45	0.005*
Neither	1.72	0.96–3.10	0.070
Disagree or less	2.60	0.99–6.85	0.052
Comfort: medical robot	0.95	0.89–1.01	0.079
Comfort: driverless car	0.90	0.84–0.96	0.001*

* $p < 0.05$.

S7. Figures

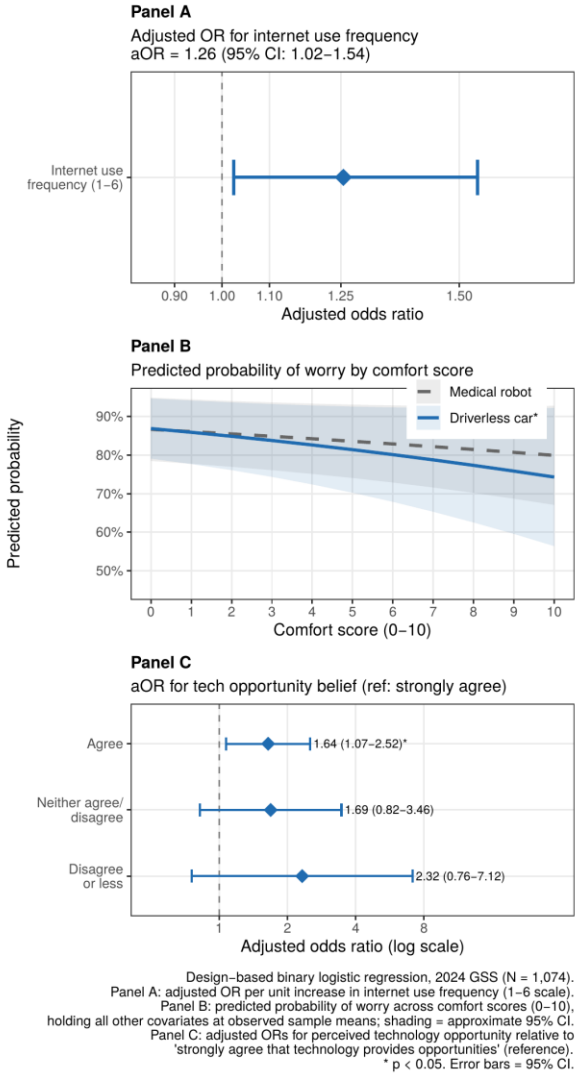

Figure S1. *Technology acceptance and trust as predictors of AI job-displacement worry. Panel A: adjusted odds ratio per unit increase in internet-use frequency (1–6 scale). Panel B: predicted probability of worry across comfort scores (0–10), holding other covariates at sample means; shading = approximate 95% CI. Panel C: adjusted ORs for perceived technology opportunity (ref: strongly agree). Design-based binary logistic regression, 2024 GSS (N = 1,074).*

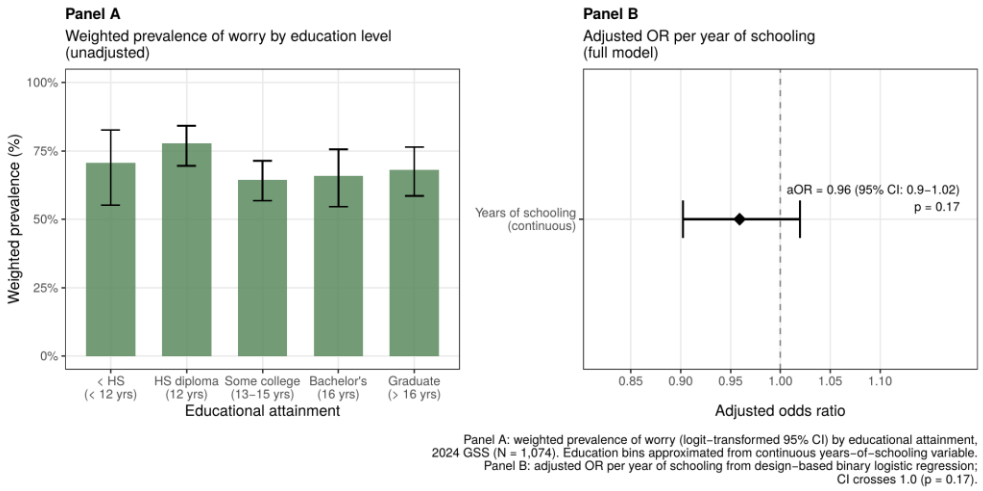

Figure S2. *Education and AI worry. Panel A: weighted prevalence by educational attainment (logit-transformed 95% CI). Panel B: adjusted OR per year of schooling; CI crosses 1.0 (p = 0.17).*

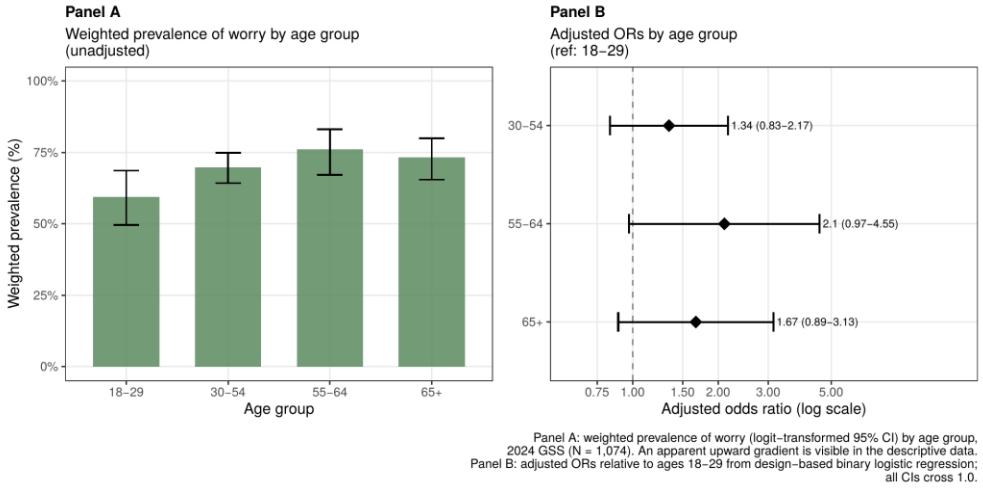

Figure S3. *Age and AI worry. Panel A: weighted prevalence by age group. Panel B: adjusted ORs relative to ages 18–29; all CIs cross 1.0.*

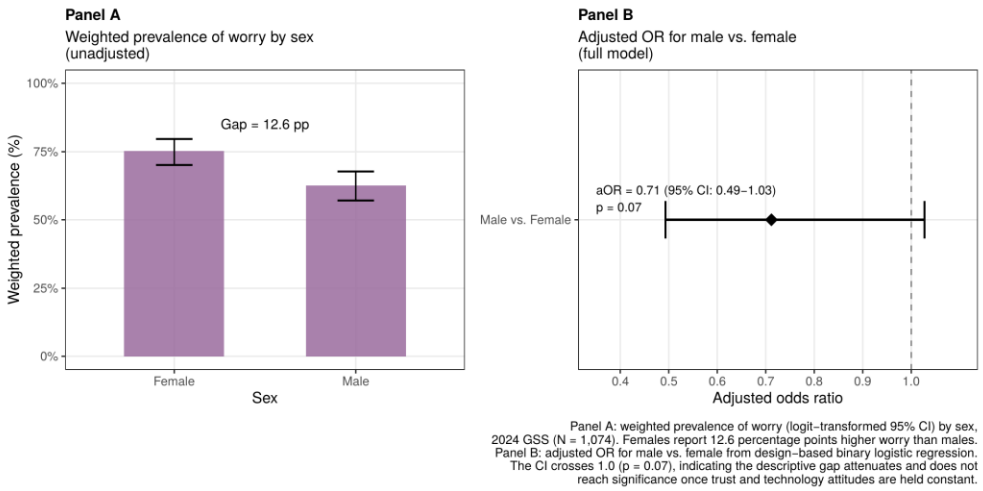

Figure S4. *Gender and AI worry. Panel A: weighted prevalence by sex (females 12.6 pp higher; computed from unrounded proportions). Panel B: adjusted OR for male vs. female; CI crosses 1.0 (p = 0.07).*

S8. Full Script

```
# Load required libraries for survey data analysis, modeling, and visualization
library(tidyverse)
library(survey)
library(srvyr)
library(gtsummary)
library(gt)
library(car)
library(ggplot2)
library(patchwork)
library(MASS)
library(brant)
library(haven)

# Set global option for handling lonely PSUs in survey estimation
options(survey.lonely.psu = "adjust")

# Import raw GSS SAS data and standardize column names to uppercase
dat_raw <-
read_sas("/Users/ankurbanerjee/Desktop/TAT/GSS_sas/gss7224_r3.sas7bdat")
dat_raw <- dat_raw %>% rename_all(toupper)

# Subset dataset to include only variables required for the analysis
```

```
vars <- c(
  "AIWORRY", "AGE", "SEX", "RACECEN1", "HISPANIC", "EDUC",
  "CLASSTECH", "EDUCTECH", "GENDTECH",
  "HARMGOOD1", "TECHESY", "INTSKILL",
  "INTRNETUSE", "AIMED", "AIDRIVE",
  "WTSSPS", "VPSU", "VSTRAT"
)
dat <- dat_raw %>% dplyr::select(any_of(vars))
n0 <- nrow(dat)
cat("Records in file:", n0, "\n")

# Filter for respondents who were administered and answered the primary
outcome
dat <- dat %>% filter(!is.na(AIWORRY))
n1 <- nrow(dat)
cat("Administered and answered AIWORRY:", n1, "\n")

# Evaluate and summarize missingness patterns across all model variables
miss_report <- dat %>%
  transmute(
    AIWORRY    = is.na(AIWORRY),
    AGE        = is.na(AGE),
    SEX        = is.na(SEX),
    RACECEN1   = is.na(RACECEN1),
    HISPANIC   = is.na(HISPANIC),
    EDUC       = is.na(EDUC),
    CLASSTECH  = is.na(CLASSTECH),
    EDUCTECH   = is.na(EDUCTECH),
    GENDTECH   = is.na(GENDTECH),
    HARMGOOD1  = is.na(HARMGOOD1),
    TECHESY    = is.na(TECHESY),
    INTSKILL   = is.na(INTSKILL),
    INTRNETUSE_never_or_na = is.na(INTRNETUSE) |
as.numeric(INTRNETUSE) == 7,
    AIMED      = is.na(AIMED),
    AIDRIVE    = is.na(AIDRIVE)
  ) %>%
  summarise(across(everything(), sum)) %>%
  pivot_longer(everything(), names_to = "variable", values_to =
"n_missing") %>%
  arrange(desc(n_missing))

cat("\nMissingness by variable (of", n1, "ballot respondents):\n")
print(miss_report)
```

```
# Recode variables, handle categorical groups, and drop missing cases
dat_coded <- dat %>%
  mutate(
    worry_bin = as.integer(as.numeric(AIWORRY) <= 2),

    worry_ord = factor(
      as.numeric(AIWORRY),
      levels = 5:1,
      labels = c("not at all worried", "not very worried",
                 "neither worried nor unworried",
                 "somewhat worried", "very worried"),
      ordered = TRUE
    ),

    age_grp = cut(
      as.numeric(AGE),
      breaks = c(17, 29, 54, 64, Inf),
      labels = c("18-29", "30-54", "55-64", "65+")
    ),

    sex = factor(SEX, levels = c(2, 1), labels = c("Female", "Male")),

    race4 = case_when(
      as.numeric(HISPANIC) > 1  ~ "Hispanic",
      as.numeric(RACECEN1) == 1 ~ "NH White",
      as.numeric(RACECEN1) == 2 ~ "NH Black",
      TRUE                      ~ "NH Other"
    ) %>% factor(levels = c("Hispanic", "NH White", "NH Black", "NH
Other")),

    educ_yrs = as.numeric(EDUC),

    tech_rich = factor(
      as.numeric(CLASSTECH),
      levels = c(1, 2, 3, 4),
      labels = c("Rich more", "Equal", "Poor more", "Neither")
    ),

    tech_educ = factor(
      as.numeric(EDUCTECH),
      levels = c(1, 2, 3, 4),
      labels = c("Educated more", "Equal", "Less educated more", "Neither")
    ),
```

```
    tech_sex = factor(
      as.numeric(GENDTECH),
      levels = c(1, 2, 3, 4),
      labels = c("Men more", "Equal", "Women more", "Neither")
    ),

    tech_harm = factor(
      as.numeric(HARMGOOD1),
      levels = 1:5,
      labels = c("Strongly disagree", "Disagree", "Neither",
                 "Agree", "Strongly agree"),
      ordered = TRUE
    ),

    tech_opps = case_when(
      as.numeric(TECHESY) == 1 ~ "Strongly agree",
      as.numeric(TECHESY) == 2 ~ "Agree",
      as.numeric(TECHESY) == 3 ~ "Neither",
      as.numeric(TECHESY) >= 4 ~ "Disagree or less"
    ) %>% factor(levels = c("Strongly agree", "Agree", "Neither", "Disagree
or less")),

    learn_sw = case_when(
      (6 - as.numeric(INTSKILL)) <= 2 ~ "Bad or worse",
      (6 - as.numeric(INTSKILL)) == 3 ~ "Neither",
      (6 - as.numeric(INTSKILL)) == 4 ~ "Good",
      (6 - as.numeric(INTSKILL)) == 5 ~ "Very good"
    ) %>% factor(levels = c("Bad or worse", "Neither", "Good", "Very
good")),

    int_use_num = case_when(
      as.numeric(INTRNETUSE) == 7 ~ NA_real_,
      TRUE                       ~ 7 - as.numeric(INTRNETUSE)
    ),

    robot_med = as.numeric(AIMED),
    robot_car = as.numeric(AIDRIVE)
  ) %>%
  drop_na(worry_bin, worry_ord, age_grp, sex, race4, educ_yrs,
          tech_rich, tech_educ, tech_sex, tech_harm, tech_opps,
          learn_sw, int_use_num, robot_med, robot_car)

n2 <- nrow(dat_coded)
```

```
n_lab <- format(n2, big.mark = ",")
cat("Complete-case analytic sample:", n2, "\n")
cat("Excluded for missing covariates:", n1 - n2, "\n")

# Define complex survey design using stratification, clusters, and weights
des <- svydesign(
  ids     = ~VPSU,
  strata  = ~VSTRAT,
  weights = ~WTSSPS,
  data    = dat_coded,
  nest    = TRUE
)
des_s <- as_survey_design(des)

# Generate Table 1: Weighted sample demographics and characteristics
tbl1 <- des_s %>%
  tbl_svysummary(
    include = c(age_grp, sex, race4, educ_yrs,
                tech_rich, tech_educ, tech_sex,
                tech_harm, learn_sw, int_use_num,
                tech_opps, robot_med, robot_car),
    statistic = list(
      all_categorical() ~ "{n_unweighted} ({p}%)",
      all_continuous()  ~ "{median} ({p25}, {p75})"
    ),
    label = list(
      age_grp     ~ "Age group",
      sex         ~ "Sex",
      race4       ~ "Race/ethnicity",
      educ_yrs    ~ "Years of schooling",
      tech_rich   ~ "Rich vs. poor benefiting from digital tech",
      tech_educ   ~ "Educated vs. less educated benefiting",
      tech_sex    ~ "Men vs. women benefiting",
      tech_harm   ~ "Technology does more harm than good",
      learn_sw    ~ "Skill at learning new software",
      int_use_num ~ "Internet use frequency (1=less often, 6=almost all the
time)",
      tech_opps   ~ "Technology provides more opportunities",
      robot_med   ~ "Comfort with medical robot (0-10)",
      robot_car   ~ "Comfort with driverless car (0-10)"
    )
  ) %>%
  modify_caption(paste0("Table 1. Weighted descriptive characteristics of the
analytic sample (N = ", n_lab, ")"))
```

```
tbl1

# Function to calculate logit-transformed survey prevalence estimates
prev_logit <- function(var) {
  f <- as.formula(paste0("~", var))
  svyby(~worry_bin, f, des, svyciprop, vartype = "ci", method = "logit")
%>%
    as_tibble() %>%
    rename(prop = worry_bin, lower = ci_l, upper = ci_u)
}

# Compute descriptive survey prevalence values across groups
prev_overall   <- svyciprop(~worry_bin, des, method = "logit")
prev_age       <- prev_logit("age_grp")
prev_sex       <- prev_logit("sex")
prev_race      ~ prev_logit("race4")
prev_tech_opps <- prev_logit("tech_opps")
prev_car_grp   <- dat_coded %>%
  mutate(car_grp = cut(robot_car, breaks = c(-1, 2, 5, 10),
                       labels = c("0-2", "3-5", "6-10"))) %>%
  { svydesign(ids = ~VPSU, strata = ~VSTRAT, weights = ~WTSSPS,
              data = ., nest = TRUE) } %>%
  { svyby(~worry_bin, ~car_grp, ., svyciprop,
          vartype = "ci", method = "logit") } %>%
  as_tibble() %>%
  rename(prop = worry_bin, lower = ci_l, upper = ci_u)

# Specify and fit the primary design-based binary logistic model
form <- worry_bin ~ age_grp + sex + race4 + educ_yrs +
  tech_rich + tech_educ + tech_sex + tech_harm +
  learn_sw + int_use_num + tech_opps + robot_med + robot_car

m_bin <- svyglm(form, design = des, family = quasibinomial())

# Extract coefficients, calculate odds ratios, and fetch 95% confidence
intervals
bin_or <- exp(cbind(OR = coef(m_bin), confint(m_bin))) %>%
  as.data.frame() %>%
  rownames_to_column("term") %>%
  rename(lower = `2.5 %`, upper = `97.5 %`)

bin_pvals <- summary(m_bin)$coefficients[, "Pr(>|t|)"]

tbl3_data <- bin_or %>%
```

```
  mutate(p_value = bin_pvals[term]) %>%
  filter(term != "(Intercept)")

# Format Table 3 output using gt
tbl3 <- tbl3_data %>%
  gt() %>%
  fmt_number(columns = c(OR, lower, upper), decimals = 2) %>%
  fmt_number(columns = p_value, decimals = 3) %>%
  cols_label(
    term    = "Variable",
    OR      = "OR",
    lower   = "Lower 95% CI",
    upper   = "Upper 95% CI",
    p_value = "p-value"
  ) %>%
  tab_header(
    title = paste0("Table 3. Design-based binary logistic regression: odds
ratios for worry that AI will take over many jobs (N = ", n_lab, ")")
  ) %>%
  tab_footnote(
    footnote = "Survey-weighted logistic regression accounting for
stratification (vstrat), clustering (vpsu), and post-stratification weights
(wtssps). Variance estimated by Taylor-series linearization. Single-PSU
strata centered at grand mean. All tests two-tailed, alpha = 0.05."
  )

tbl3

# Multicollinearity validation using Generalized Variance Inflation Factors
(GVIF)
m_vif <- glm(
  update(form, . ~ .),
  data   = dat_coded,
  family = binomial(),
  maxit  = 100
)
gvif_vals <- vif(m_vif)
cat("\nGeneralized Variance Inflation Factors:\n")
print(gvif_vals)
cat("\nAll GVIF^(1/(2*Df)) values below 1.25:", all(gvif_vals[, 3] < 1.25),
"\n")

# Fit alternative ordinal regression model to check parallel lines assumption
form_ord <- worry_ord ~ age_grp + sex + race4 + educ_yrs +
```

```
  tech_rich + tech_educ + tech_sex + tech_harm +
  learn_sw + int_use_num + tech_opps + robot_med + robot_car

m_ord <- svyolr(form_ord, design = des)

cuts_bin <- list(
  c1 = as.formula("I(as.numeric(worry_ord) >= 2) ~ age_grp + sex + race4
+ educ_yrs + tech_rich + tech_educ + tech_sex + tech_harm + learn_sw +
int_use_num + tech_opps + robot_med + robot_car"),
  c2 = as.formula("I(as.numeric(worry_ord) >= 3) ~ age_grp + sex + race4
+ educ_yrs + tech_rich + tech_educ + tech_sex + tech_harm + learn_sw +
int_use_num + tech_opps + robot_med + robot_car"),
  c3 = as.formula("I(as.numeric(worry_ord) >= 4) ~ age_grp + sex + race4
+ educ_yrs + tech_rich + tech_educ + tech_sex + tech_harm + learn_sw +
int_use_num + tech_opps + robot_med + robot_car"),
  c4 = as.formula("I(as.numeric(worry_ord) >= 5) ~ age_grp + sex + race4
+ educ_yrs + tech_rich + tech_educ + tech_sex + tech_harm + learn_sw +
int_use_num + tech_opps + robot_med + robot_car")
)

po_coefs <- map(cuts_bin, ~ coef(svyglm(.x, design = des, family =
quasibinomial())))
po_check <- bind_cols(map(po_coefs, as_tibble)) %>%
  mutate(term = names(po_coefs[[1]])) %>%
  dplyr::select(term, everything()) %>%
  setNames(c("term", "Cut1", "Cut2", "Cut3", "Cut4"))

m_ord_unwt <- polr(form_ord, data = dat_coded, Hess = TRUE)
brant_result <- brant(m_ord_unwt)
print(brant_result)

cat("\nProportional-odds check: logistic coefficients across four binary
cutpoints\n")
cat("Instability in tech_harm and learn_sw indicates proportional-odds
assumption is not tenable.\n\n")
print(po_check)

ord_or <- exp(cbind(OR = coef(m_ord), confint(m_ord))) %>%
  as.data.frame() %>%
  rownames_to_column("term") %>%
  rename(lower = `2.5 %`, upper = `97.5 %`)

ord_pvals <- 2 * pnorm(abs(coef(m_ord) / sqrt(diag(vcov(m_ord)))),
lower.tail = FALSE)
```

```
supp_table <- ord_or %>%
  mutate(p_value = c(ord_pvals, rep(NA, nrow(ord_or) -
length(ord_pvals)))) %>%
  dplyr::filter(!grepl("\\|", term))

cat("\nSupplementary Table S1. Ordinal logistic regression results
(abandoned — proportional-odds assumption not met)\n")
print(supp_table)

# Set common theme configuration for visualizations
theme_fig <- theme_bw(base_size = 11) +
  theme(
    panel.grid.minor = element_blank(),
    strip.background = element_blank(),
    plot.title       = element_text(face = "bold", size = 11)
  )

# Figure 1
fig1_coefs <- tbl3_data %>%
  filter(term == "int_use_num") %>%
  mutate(label = "Internet use\nfrequency (1-6)")

panel_A <- ggplot(fig1_coefs, aes(x = OR, y = label)) +
  geom_vline(xintercept = 1, linetype = "dashed", color = "grey50") +
  geom_errorbarh(aes(xmin = lower, xmax = upper), height = 0.2,
                 linewidth = 1.0, color = "#1F6BB0") +
  geom_point(shape = 18, size = 5, color = "#1F6BB0") +
  scale_x_continuous(limits = c(0.85, 1.70),
                     breaks = c(0.90, 1.00, 1.10, 1.25, 1.50)) +
  labs(x = "Adjusted odds ratio", y = NULL,
       title = "Panel A",
       subtitle = paste0("Adjusted OR for internet use frequency\n",
                         "aOR = ", round(fig1_coefs$OR, 2),
                         " (95% CI: ", round(fig1_coefs$lower, 2),
                         "\u2013", round(fig1_coefs$upper, 2), ")")) +
  theme_fig

pred_comfort <- function(var_name, label) {
  newdata_base <- dat_coded %>%
    summarise(across(where(is.numeric), \(x) median(x, na.rm = TRUE)))
%>%
    slice(rep(1, 11))

  newdata_base[[var_name]] <- 0:10
```

```
  newdata_base$age_grp    <- factor("30-54",        levels =
levels(dat_coded$age_grp))
  newdata_base$sex        <- factor("Female",       levels =
levels(dat_coded$sex))
  newdata_base$race4      <- factor("NH White",     levels =
levels(dat_coded$race4))
  newdata_base$tech_rich  <- factor("Rich more",    levels =
levels(dat_coded$tech_rich))
  newdata_base$tech_educ  <- factor("Educated more",levels =
levels(dat_coded$tech_educ))
  newdata_base$tech_sex   <- factor("Equal",        levels =
levels(dat_coded$tech_sex))
  newdata_base$tech_harm  <- factor("Neither",      levels =
levels(dat_coded$tech_harm), ordered = TRUE)
  newdata_base$tech_opps  <- factor("Agree",        levels =
levels(dat_coded$tech_opps))
  newdata_base$learn_sw   <- factor("Good",         levels =
levels(dat_coded$learn_sw))

  pred <- predict(m_bin, newdata = newdata_base, type = "response", se =
TRUE)
  data.frame(
    score = 0:10,
    prob  = as.numeric(pred),
    se    = as.numeric(attr(pred, "var")^0.5),
    label = label
  )
}

pred_med  ~ pred_comfort("robot_med", "Medical robot")
pred_car  ~ pred_comfort("robot_car", "Driverless car*")

pred_both <- bind_rows(pred_med, pred_car) %>%
  mutate(
    lower = prob - 1.96 * se,
    upper = prob + 1.96 * se,
    label = factor(label, levels = c("Medical robot", "Driverless car*"))
  )

panel_B <- ggplot(pred_both, aes(x = score, y = prob * 100,
                                 color = label, fill = label, linetype = label)) +
  geom_ribbon(aes(ymin = lower * 100, ymax = upper * 100), alpha =
0.12, color = NA) +
  geom_line(linewidth = 1) +
```

```
  scale_color_manual(values = c("Medical robot" = "#666666", "Driverless
car*" = "#1F6BB0")) +
  scale_fill_manual(values  = c("Medical robot" = "#666666", "Driverless
car*" = "#1F6BB0")) +
  scale_linetype_manual(values = c("Medical robot" = "dashed", "Driverless
car*" = "solid")) +
  scale_x_continuous(breaks = 0:10) +
  scale_y_continuous(limits = c(45, 95), labels = function(x) paste0(x, "%"))
+
  labs(x = "Comfort score (0\u201310)", y = "Predicted probability",
       title = "Panel B",
       subtitle = "Predicted probability of worry by comfort score",
       color = NULL, fill = NULL, linetype = NULL) +
  theme_fig +
  theme(legend.position = c(0.78, 0.92),
        legend.background = element_rect(fill = "white", color = NA))

opps_coefs <- tbl3_data %>%
  filter(grepl("tech_opps", term)) %>%
  mutate(
    level = case_when(
      grepl("Agree$", term)          ~ "Agree",
      grepl("Neither", term)         ~ "Neither agree/\ndisagree",
      grepl("Disagree or less", term) ~ "Disagree\nor less"
    ),
    level = factor(level, levels = c("Disagree\nor less", "Neither
agree/\ndisagree", "Agree")),
    sig   = p_value < 0.05,
    lab   = paste0(round(OR, 2), " (", round(lower, 2), "\u2013",
round(upper, 2), ")",
                   ifelse(sig, "*", ""))
  )

panel_C <- ggplot(opps_coefs, aes(x = OR, y = level)) +
  geom_vline(xintercept = 1, linetype = "dashed", color = "grey50") +
  geom_errorbarh(aes(xmin = lower, xmax = upper), height = 0.2,
                 linewidth = 0.7, color = "#1F6BB0") +
  geom_point(shape = 18, size = 4, color = "#1F6BB0") +
  geom_text(aes(x = upper, label = paste0(" ", lab)), hjust = 0, size = 2.8)
+
  scale_x_continuous(trans = "log", breaks = c(1.0, 2.0, 4.0, 8.0),
                     limits = c(0.5, 30)) +
  labs(x = "Adjusted odds ratio (log scale)", y = NULL,
       title = "Panel C",
```

```
      subtitle = "aOR for tech opportunity belief (ref: strongly agree)") +
  theme_fig

fig1 <- panel_A / panel_B / panel_C +
  plot_annotation(
    caption = paste0(
      "Design-based binary logistic regression, 2024 GSS (N = ", n_lab, ").\n",
      "Panel A: adjusted OR per unit increase in internet use frequency
(1\u20136 scale).\n",
      "Panel B: predicted probability of worry across comfort scores
(0\u201310),\n",
      "holding all other covariates at observed sample means; shading =
approximate 95% CI.\n",
      "Panel C: adjusted ORs for perceived technology opportunity relative
to\n",
      "'strongly agree that technology provides opportunities' (reference).\n",
      "* p < 0.05. Error bars = 95% CI."
    )
  )

ggsave("fig1_TAM.pdf", fig1, width = 5.5, height = 10, device = "pdf")

# Figure 2
edu_bins <- dat_coded %>%
  mutate(edu_grp = cut(
    educ_yrs,
    breaks = c(-Inf, 11, 12, 15, 16, Inf),
    labels = c("< HS\n(< 12 yrs)", "HS diploma\n(12 yrs)",
               "Some college\n(13-15 yrs)", "Bachelor's\n(16 yrs)",
               "Graduate\n(> 16 yrs)")
  ))

des_edu <- svydesign(ids = ~VPSU, strata = ~VSTRAT, weights =
~WTSSPS,
                     data = edu_bins, nest = TRUE)

prev_edu <- svyby(~worry_bin, ~edu_grp, des_edu, svyciprop,
                  vartype = "ci", method = "logit") %>%
  as_tibble() %>%
  rename(prop = worry_bin, lower = ci_l, upper = ci_u)

panel2A <- ggplot(prev_edu, aes(x = edu_grp, y = prop * 100)) +
  geom_col(fill = "#5B8A5F", alpha = 0.85, width = 0.65) +
```

```
  geom_errorbar(aes(ymin = lower * 100, ymax = upper * 100), width =
0.25, linewidth = 0.6) +
  scale_y_continuous(limits = c(0, 100), labels = function(x) paste0(x, "%"))
+
  labs(x = "Educational attainment", y = "Weighted prevalence (%)",
       title = "Panel A",
       subtitle = "Weighted prevalence of worry by education
level\n(unadjusted)") +
  theme_fig

educ_or <- tbl3_data %>%
  filter(term == "educ_yrs") %>%
  mutate(label = "Years of schooling\n(continuous)")

panel2B <- ggplot(educ_or, aes(x = OR, y = label)) +
  geom_vline(xintercept = 1, linetype = "dashed", color = "grey50") +
  geom_errorbarh(aes(xmin = lower, xmax = upper), height = 0.15,
linewidth = 0.7) +
  geom_point(shape = 18, size = 4) +
  scale_x_continuous(limits = c(0.82, 1.18),
                     breaks = c(0.85, 0.90, 0.95, 1.00, 1.05, 1.10)) +
  annotate("text", x = 1.18, y = 1,
           label = paste0("aOR = ", round(educ_or$OR, 2),
                          " (95% CI: ", round(educ_or$lower, 2),
                          "\u2013", round(educ_or$upper, 2), ")\np = ",
                          round(bin_pvals["educ_yrs"], 2)),
           hjust = 1, vjust = -0.3, size = 3) +
  labs(x = "Adjusted odds ratio", y = NULL,
       title = "Panel B",
       subtitle = "Adjusted OR per year of schooling\n(full model)") +
  theme_fig

fig2 <- panel2A + panel2B +
  plot_annotation(
    caption = paste0(
      "Panel A: weighted prevalence of worry (logit-transformed 95% CI) by
educational attainment,\n",
      "2024 GSS (N = ", n_lab, "). Education bins approximated from
continuous years-of-schooling variable.\n",
      "Panel B: adjusted OR per year of schooling from design-based binary
logistic regression;\n",
      "CI crosses 1.0 (p = ", round(bin_pvals["educ_yrs"], 2), ")."
    )
  )
```

```
ggsave("fig2_education.pdf", fig2, width = 10, height = 5, device = "pdf")

# Figure 3
prev_age_df <- svyby(~worry_bin, ~age_grp, des, svyciprop,
                     vartype = "ci", method = "logit") %>%
  as_tibble() %>%
  rename(prop = worry_bin, lower = ci_l, upper = ci_u)

panel3A <- ggplot(prev_age_df, aes(x = age_grp, y = prop * 100)) +
  geom_col(fill = "#5B8A5F", alpha = 0.85, width = 0.65) +
  geom_errorbar(aes(ymin = lower * 100, ymax = upper * 100), width =
0.25, linewidth = 0.6) +
  scale_y_continuous(limits = c(0, 100), labels = function(x) paste0(x, "%"))
+
  labs(x = "Age group", y = "Weighted prevalence (%)",
       title = "Panel A",
       subtitle = "Weighted prevalence of worry by age group\n(unadjusted)")
+
  theme_fig

age_ors <- tbl3_data %>%
  filter(grepl("age_grp", term)) %>%
  mutate(
    level = sub("age_grp", "", term),
    level = factor(level, levels = c("65+", "55-64", "30-54")),
    lab   = paste0(round(OR, 2), " (", round(lower, 2), "\u2013",
round(upper, 2), ")")
  )

panel3B <- ggplot(age_ors, aes(x = OR, y = level)) +
  geom_vline(xintercept = 1, linetype = "dashed", color = "grey50") +
  geom_errorbarh(aes(xmin = lower, xmax = upper), height = 0.2, linewidth
= 0.7) +
  geom_point(shape = 18, size = 4) +
  geom_text(aes(x = upper, label = paste0(" ", lab)), hjust = 0, size = 2.8)
+
  scale_x_continuous(trans = "log",
                     breaks = c(0.75, 1.0, 1.5, 2.0, 3.0, 5.0),
                     limits = c(0.6, 14)) +
  labs(x = "Adjusted odds ratio (log scale)", y = NULL,
       title = "Panel B",
       subtitle = "Adjusted ORs by age group\n(ref: 18-29)") +
  theme_fig
```

```
fig3 <- panel3A + panel3B +
  plot_annotation(
    caption = paste0(
      "Panel A: weighted prevalence of worry (logit-transformed 95% CI) by
age group,\n",
      "2024 GSS (N = ", n_lab, "). An apparent upward gradient is visible in
the descriptive data.\n",
      "Panel B: adjusted ORs relative to ages 18-29 from design-based binary
logistic regression;\n",
      "all CIs cross 1.0."
    )
  )

ggsave("fig3_age.pdf", fig3, width = 10, height = 5, device = "pdf")

# Figure 4
prev_sex_df <- svyby(~worry_bin, ~sex, des, svyciprop,
                     vartype = "ci", method = "logit") %>%
  as_tibble() %>%
  rename(prop = worry_bin, lower = ci_l, upper = ci_u)

delta_pp <- round((prev_sex_df$prop[prev_sex_df$sex == "Female"] -
                     prev_sex_df$prop[prev_sex_df$sex == "Male"]) * 100, 1)

panel4A <- ggplot(prev_sex_df, aes(x = sex, y = prop * 100)) +
  geom_col(fill = "#9B6B9E", alpha = 0.85, width = 0.55) +
  geom_errorbar(aes(ymin = lower * 100, ymax = upper * 100), width =
0.2, linewidth = 0.6) +
  annotate("text", x = 1.5, y = 85,
           label = paste0("Gap = ", delta_pp, " pp"), size = 3.5) +
  scale_y_continuous(limits = c(0, 100), labels = function(x) paste0(x, "%"))
+
  labs(x = "Sex", y = "Weighted prevalence (%)",
       title = "Panel A",
       subtitle = "Weighted prevalence of worry by sex\n(unadjusted)") +
  theme_fig

sex_or <- tbl3_data %>%
  filter(term == "sexMale") %>%
  mutate(label = "Male vs. Female")

panel4B <- ggplot(sex_or, aes(x = OR, y = label)) +
  geom_vline(xintercept = 1, linetype = "dashed", color = "grey50") +
```

```
  geom_errorbarh(aes(xmin = lower, xmax = upper), height = 0.15,
linewidth = 0.7) +
  geom_point(shape = 18, size = 4) +
  scale_x_continuous(limits = c(0.35, 1.10), breaks = seq(0.4, 1.0, 0.1)) +
  annotate("text", x = 0.35, y = 1,
           label = paste0("aOR = ", round(sex_or$OR, 2),
                          " (95% CI: ", round(sex_or$lower, 2),
                          "\u2013", round(sex_or$upper, 2), ")\np = ",
                          round(bin_pvals["sexMale"], 2)),
           hjust = 0, vjust = -0.3, size = 3) +
  labs(x = "Adjusted odds ratio", y = NULL,
       title = "Panel B",
       subtitle = "Adjusted OR for male vs. female\n(full model)") +
  theme_fig

fig4 <- panel4A + panel4B +
  plot_annotation(
    caption = paste0(
      "Panel A: weighted prevalence of worry (logit-transformed 95% CI) by
sex,\n",
      "2024 GSS (N = ", n_lab, "). Females report ", delta_pp, " percentage
points higher worry than males.\n",
      "Panel B: adjusted OR for male vs. female from design-based binary
logistic regression.\n",
      "The CI crosses 1.0 (p = ", round(bin_pvals["sexMale"], 2),
      "), indicating the descriptive gap attenuates and does not\n",
      "reach significance once trust and technology attitudes are held
constant."
    )
  )

ggsave("fig4_gender.pdf", fig4, width = 10, height = 5, device = "pdf")

# Print tables for verification
tbl1_df <- as_tibble(tbl1$table_body) %>%
  dplyr::select(variable, label, stat_0)
cat("\nTable 1: Weighted descriptive characteristics\n")
print(tbl1_df, n = 100)

tbl2_df <- bind_rows(
  tibble(group = "Overall", level = "Overall",
         prop  = as.numeric(prev_overall),
         lower = attr(prev_overall, "ci")[1],
         upper = attr(prev_overall, "ci")[2]),
```

```
  prev_age       %>% transmute(group = "Age",         level = age_grp,
prop, lower, upper),
  prev_sex       %>% transmute(group = "Sex",         level = sex,       prop,
lower, upper),
  prev_race      %>% transmute(group = "Race",        level = race4,
prop, lower, upper),
  prev_tech_opps %>% transmute(group = "Tech opps",   level =
tech_opps, prop, lower, upper),
  prev_car_grp   %>% transmute(group = "Car comfort", level = car_grp,
prop, lower, upper)
) %>%
  mutate(across(c(prop, lower, upper), \(x) round(x * 100, 1)))
cat("\nTable 2: Weighted prevalence of worry, %\n")
print(tbl2_df, n = 100)

cat("\nTable 3: Binary logistic ORs\n")
tbl3_print <- tbl3_data
tbl3_print$OR      <- round(tbl3_print$OR, 2)
tbl3_print$lower   <- round(tbl3_print$lower, 2)
tbl3_print$upper   <- round(tbl3_print$upper, 2)
tbl3_print$p_value <- unname(round(tbl3_print$p_value, 3))
print(as.data.frame(tbl3_print))

cat("\nOverall Prevalence Summary\n")
cat("Overall worried:", round(as.numeric(prev_overall) * 100, 1), "%\n")
cat("Overall CI lower:", round(attr(prev_overall, "ci")[1] * 100, 1), "%\n")
cat("Overall CI upper:", round(attr(prev_overall, "ci")[2] * 100, 1), "%\n")

sessionInfo()
```