The Workplace GenAI Usage Divide: A Genetics-Informed Cognitive and Metacognitive Account

Supplemental Materials

- **S1** provides the protocol and accompanying scripts for genetic-data quality control, imputation, principal-component analysis, and PGS construction.
- **Table S1** compares participants and nonparticipants on observed demographic and job characteristics; no statistically significant differences were detected.
- **Table S2** reports the hypothesis-testing models without control variables; the focal relationships remain consistent with the primary analyses.
- **Table S3** reports robustness checks with alternative count models, using negative binomial (NB) and zero-inflated negative binomial (ZINB) regressions.
- **Table S4** reports sensitivity analyses excluding heavy users (top 5% or 10%), excluding nonusers, and restricting the sample to employee-months with at least one login. The focal associations remain robust.
- **Table S5** examines temporal trends in GenAI use and their interaction with the educational attainment PGS. GenAI use increases over time, but the trend does not vary significantly with the PGS.
- **Table S6** tests whether the associations between the educational attainment PGS and GenAI use vary by gender, age, job position, or department affiliation.
- **Table S7** compares the proposed and alternative directional pathways using AIC and BIC; the proposed ordering provides better relative fit.
- **Table S8** reports wild cluster bootstrap inference; the substantive conclusions remain consistent under this alternative procedure.
- **Figure S1** plots the distribution of GenAI use frequency over time, illustrating persistent heterogeneity across employees.

S1. Protocol for Genetic Data Processing

In this document, we provide a protocol, with accompanying scripts, for genetic data processing procedures including quality control (QC), imputation, and the calculation of the polygenic score (PGS).

We begin by outlining the QC procedures, which aim to minimize biases in subsequent analyses by identifying and excluding problematic samples and DNA markers. Next, we present the imputation procedures, where we show how to recover genotypes not measured in a specific cohort using a reference set of haplotypes. This step ensures that all cohorts have as many single-nucleotide polymorphisms (SNPs) as possible to best utilize the GWAS results. Lastly, we introduce the PGS calculation procedures, as PGSs serve as the primary independent variables in the analyses in our paper.

This protocol is based mainly on Anderson et al. (2010), Van Leeuwen et al. (2015), and Choi et al. (2020).

1. MATERIALS

Data

- Genome-wide SNP data. See the supplementary material in Anderson et al. (2010) for an example dataset: https://www.nature.com/articles/nprot.2010.116
- A reference panel for imputation. In our paper, we use 1000GP Phase 3 haplotypes, which can be downloaded from https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.html

Software

- PLINK: available at https://www.cog-genomics.org/plink/
- PRSice: available at https://choishingwan.github.io/PRS-Tutorial/prsice/
- SHAPEIT: available at https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
- IMPUTE: available at https://mathgen.stats.ox.ac.uk/impute/impute_v2.html
- GTOOL: documentation available at https://docs.rc.ufl.edu/software/apps/gtool/
- R: available at https://www.r-project.org/.

2. PROCEDURE

Pre-Imputation Quality Control

Let's call the original genetic datasets "raw-GWA-data.bed," "raw-GWA-data.bim," and "raw-GWA-data.fam."

This part refers to the following online tutorial: https://choishingwan.github.io/PRS-Tutorial/.

Step 1: Apply some of these QC metrics to the target data

```
./tools/plink \
    --bfile ./raw/raw-GWA-data \
--maf 0.01 \
--hwe 1e-6 \
--geno 0.01 \
--mind 0.1 \
--write-snplist \
```

```
--make-just-fam \
 --out ./raw/raw-GWA-data.QC
```

Step 2: Perform pruning to remove highly correlated SNPs

Extremely high or low heterozygosity in an individual may indicate DNA contamination or substantial inbreeding. Consequently, samples showing abnormal heterozygosity levels are usually excluded before conducting downstream analyses.

```
./tools/plink \
    --bfile ./raw/raw-GWA-data \
        --keep ./raw/raw-GWA-data.QC.fam \
        --extract ./raw/raw-GWA-data.QC.snplist \
        --indep-pairwise 200 50 0.25 \
        --out ./raw/raw-GWA-data.QC
```

Step 3: Compute heterozygosity rate

```
./tools/plink \
    --bfile ./raw/raw-GWA-data\
      --extract ./raw/raw-GWA-data.QC.prune.in \
      --keep ./raw/raw-GWA-data.QC.fam \
      --het \
      --out ./raw/raw-GWA-data.QC
```

Use the following R command:

```
dat <- read.table("./raw/raw-GWA-data.QC.het", header=T)
m <- mean(dat$F)
s <- sd(dat$F)
valid <- subset(dat, F <= m+3*s & F >= m-3*s)
write.table(valid[,c(1,2)], "./raw/raw-GWA-data.valid.sample", quote=F, row.names=F)
q()
```

Step 4: Prune related individuals so that no retained pair has an estimated genomic relationship coefficient > 0.125 (the expected value for third-degree relatives)

```
./tools/plink \
    --bfile ./raw/raw-GWA-data \
        --extract ./raw/raw-GWA-data.QC.prune.in \
        --keep ./raw/raw-GWA-data.valid.sample \
        --rel-cutoff 0.125 \
        --out ./raw/raw-GWA-data.QC
```

Step 5: Generate final QC'ed target data file

```
./tools/plink \
    --bfile ./raw/raw-GWA-data \
      --make-bed \
      --keep ./raw/raw-GWA-data.QC.rel.id \
      --out ./source/clean-GWA-data \
      --extract ./raw/raw-GWA-data.QC.snplist
```

Imputation

Step 1: Split the data

Since the reference panel we used for our paper has been separated into chunks according to chromosome indexes, we also need to split the data file into corresponding chunk files. Researchers can skip this step if they use other reference panels.

```
for chr in {1..22}; do
  ./tools/plink \
    --bfile ./source/clean-GWA-data \
    --chr $chr \
    --recode \
    --out ./source/clean-GWA-data.chr${chr};
done;
```

Step 2: Imputation and transformation

We use SHAPEIT for phasing, IMPUTE for imputation, and GTOOL for converting .imputed2 files to .ped and .map files.

```
for chr in {1..22}; do
  namefile="clean-GWA-data.chr${chr}";
  ./tools/shapeit \
    --input-ped ./source/${namefile}.ped ./source/${namefile}.map \
    --input-map ./1000GP_Phase3/genetic_map_chr${chr}_combined_b37.txt \
    --output-max ./prephased/${namefile} \
    --thread 12 \
    --output-log ./prephased/${namefile};
done;
for chr in {1..22}; do
  namefile="clean-GWA-data.chr${chr}";
  maxPos=$(gawk '$1!="position" {print $1}' \
         ./1000GP_Phase3/genetic_map_chr${chr}_combined_b37.txt | \
       sort -n | \
       tail -n 1);
  nrChunk=$(expr ${maxPos} "/" 5000000);
  nrChunk2=$(expr ${nrChunk} "+" 1);
  start="0";
  countfile="0";
  for chunk in $(seq 1 $nrChunk2); do
    endchr=$(expr $start "+" 5000000);
    startchr=$(expr $start "+" 1);
    ./tools/impute2 \
      -known_haps_g ./prephased/${namefile}.haps \
      -m ./1000GP_Phase3/genetic_map_chr${chr}_combined_b37.txt \
      -h ./1000GP_Phase3/1000GP_Phase3_chr${chr}.hap.gz \
      -l ./1000GP_Phase3/1000GP_Phase3_chr${chr}.legend.gz \
      -int ${startchr} ${endchr} \
      -Ne 20000 \
      -o ./imputed/${namefile}.chunk${chunk}.impute2;
    start=${endchr};
    if test -f "./imputed/${namefile}.chunk${chunk}.impute2"; then
      let countfile=countfile+1;
      bedname="clean-GWA-data.imputed2.chr${chr}.chunk${chunk}";
      ./tools/gtool \
        -G \
        --g ./imputed/${namefile}.chunk${chunk}.impute2 \
        --s ./prephased/${namefile}.sample \
        --ped ./converted/${namefile}.chunk${chunk}.ped \
```

```
        --map ./converted/${namefile}.chunk${chunk}.map \
        --snp \
        --chr "${chr}";
      ./tools/plink \
        --file ./converted/${namefile}.chunk${chunk} \
        --make-bed \
        --out ./converted/$bedname;
      if (($countfile>1));
      then
        echo ./converted/$bedname >> ./converted/merge-list-chr${chr}.txt;
      else
        echo ./converted/$bedname >> ./converted/first-chunk-chr${chr}.txt;
      fi;
    fi;
  done;
done;
```

Step 3: Remove low-quality markers

To ensure data quality, we filter out imputed SNPs with info scores less than 0.8.

```
for chr in {1..22}; do
  namefile="./converted/first-chunk-chr${chr}.txt";
  mergefile="./converted/merge-list-chr${chr}.txt";
  firstfile=$(cat ${namefile});
  for fname in $(ls ./imputed | egrep chr${chr}\.chunk[[:digit:]]+\.impute2_info$); do
    gawk '$7>0.8 && NR>1 {print
$2}' ./imputed/${fname} >> ./imputed/info_filter${chr}.txt;
  done;
  ./tools/plink \
    --bfile ${firstfile} \
    --merge-list ${mergefile} \
    --make-bed \
    --extract ./imputed/info_filter${chr}.txt \
    --out ./CHR/chr${chr}.merge;
  rm ./imputed/info_filter${chr}.txt;
done;
```

Step 4: Merge all chunk files

```
for chr in {2..22}; do
  echo ./CHR/chr${chr}.merge >> ./CHR/merge-list.txt;
done;
./tools/plink \
  --bfile ./CHR/chr1.merge \
  --merge-list ./CHR/merge-list.txt \
  --make-bed \
  --out ./out/data.imputed;
```

Post-Imputation Quality Control

Step 1: Check the correctness of markers

We compare markers in our data with the dbSNP data in the UCSC Genome Browser to ensure they are correctly measured and consistent with the GWAS results.

Step 2: Remove low-quality markers before PGS computation

More details can be found at https://choishingwan.github.io/PRSice/quick_start/.

```
./tools/plink \
  --bfile ./out/data.imputed \
  --maf 0.01 \
  --mind 0.1 \
```

```
  --geno 0.01 \
  --hwe 1e-6 \
  --make-bed \
  --out ./out/data.imputed.qc
```

Step 3: Perform principal-component analysis

```
./tools/plink \
  --bfile ./out/data.imputed.qc \
  --pca 40 \
  --out ./out/data.imputed.qc
```

PGS Computation

Here we offer an example of PGS computation. Researchers can customize their scripts based on the instructions at https://choishingwan.github.io/PRSice/step_by_step/. For illustration, let's name the file that stores the GWAS results "gwas_stats.txt." Many GWAS results can be found in public databases or supplementary data of relevant papers.

```
./tools/PRSice_linux \
  --snp MarkerName \
  --a1 A1 \
  --a2 A2 \
 --chr CHR \
  --bar-levels 0.001,0.05,0.1,0.2,0.3,0.4,0.5,1 \
  --stat Beta \
  --pvalue Pval \
  --binary-target F \
  --base ./PGS_base/gwas_stats.gz \
  --target ./out/data.imputed.qc \
  --pheno ./pheno/data.pheno \
  --out ./PGS/PGS_result \
  --clump-kb 250kb \
  --clump-p 1.000000 \
  --clump-r2 0.100000 \
  --interval 5e-05 \
  --lower 5e-08 \
  --upper 0.5 \
  --num-auto 22 \
  --seed 2745672831 \
  --thread 2
```

References

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., & Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nature Protocols*, 5(9), 1564-1573.

Choi, S. W., Mak, T. S. H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, 15(9), 2759-2772.

Van Leeuwen, E. M., Kanterakis, A., Deelen, P., Kattenberg, M. V., Slagboom, P. E., de Bakker, P. I., ... & Hottenga, J. J. (2015). Population-specific genotype imputations using minimac or IMPUTE2. *Nature Protocols*, 10(9), 1285-1296.

Table S1

Summary Statistics and Mean Comparisons of Participants and Nonparticipants

	Participants (N = 150) Mean (SD)	Nonparticipants (N = 353) Mean (SD)	Mean difference	t-statistic	p-value
Gender	0.53 (0.50)	0.54 (0.50)	-0.01	-0.16	0.874
Age	29.99 (5.71)	29.80 (4.90)	0.18	0.36	0.717
Job position	1.69 (0.95)	1.61 (0.86)	0.08	0.94	0.345
Education level	2.58 (0.66)	2.61 (0.61)	-0.03	-0.48	0.634

Notes. Values are reported as means with standard deviations in parentheses. Gender is coded as 0 = male and 1 = female. Job position is coded as 1 = entry-level employee, 2 = junior manager, 3 = mid-level manager, and 4 = senior manager. Education level is coded as 1 = vocational secondary school, 2 = junior college, 3 = bachelor's degree, and 4 = master's degree. Mean differences are calculated as participants minus nonparticipants. Reported p-values are based on two-sided independent samples t-tests. Participants are the 150 employees retained in the final analytical sample after genotyping quality control; nonparticipants are the 353 eligible employees who were invited but did not participate in genetic data collection. Supplementary tests yielded the same conclusions: Pearson chi-square tests for gender (p = 0.873), job position (p = 0.252), and education level (p = 0.660); Mann-Whitney tests for job position (p = 0.495) and education level (p = 0.727); and a Welch two-sample t-test for age (p = 0.734).

Table S2

Generalized Structural Equation Modeling Results Without Control Variables

	(1) Education level		(2) Strategic mindset		(3) GenAI credibility perceptions		(4) Metaknowledge of expertise location		(5) GenAI use frequency		(6) GenAI interaction intensity	
	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*
Intercept	2.58**	0.05	3.63**	0.06	4.95**	0.51	4.29**	0.42	-2.66	1.77	-0.66**	0.24
Independent Variable												
Educational attainment PGS	0.13*	0.05	0.13*	0.06	0.07	0.08	-0.01	0.06	0.23	0.15	0.03	0.02
Intermediate Variables												
Education level					-0.26*	0.11	-0.23*	0.09	-0.38	0.19	0.01	0.04
Strategic mindset					0.07	0.11	0.37**	0.09	0.62*	0.29	0.06*	0.03
GenAI credibility perceptions									0.34*	0.16	0.09**	0.03
Metaknowledge of expertise location									0.54*	0.23	0.11**	0.03
(Pseudo) R^2	0.04		0.03		0.04		0.14		0.33		0.35	

Notes. N = 3,150 employee-month observations. Coef. = coefficient; *SE* = standard error. All models include only the focal variables reported in the table and exclude control variables. GenAI use frequency in column (5) is measured as the monthly login count on the company's AI platform. GenAI interaction intensity in column (6) is measured as dialogue turns per login and is log-transformed. The educational attainment PGS is constructed using the GWAS results reported by Lee et al. (2018b) and is standardized to have a mean of 0 and a standard deviation of 1. Columns (1)–(4) present OLS estimates, column (5) presents Poisson regression estimates, and column (6) presents linear mixed-effects estimates with employee-level random intercepts. Cluster-robust standard errors at the employee level are reported. * $p < 0.05$; ** $p < 0.01$.

Table S3

Robustness Checks with Alternative Count Models

	(1) Education level		(2) Strategic mindset		(3) GenAI credibility perceptions		(4) Metaknowledge of expertise location		(5) GenAI use frequency (NB)		(6) GenAI use frequency (ZINB)	
	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*
Intercept	2.98**	0.39	3.56**	0.37	4.49**	0.69	3.74**	0.57	-4.92**	1.44	-4.72**	1.59
Independent Variable												
Educational attainment PGS	0.11*	0.05	0.13*	0.06	0.05	0.08	-0.03	0.06	0.23*	0.10	0.23*	0.10
Intermediate Variables												
Education level					-0.27*	0.12	-0.18	0.10	0.005	0.20	0.02	0.22
Strategic mindset					0.07	0.11	0.38**	0.09	-0.05	0.20	-0.06	0.22
GenAI credibility perceptions									0.28*	0.12	0.28*	0.12
Metaknowledge of expertise location									0.85**	0.18	0.84**	0.18
Control Variables												
Gender	0.09	0.10	-0.18	0.12	-0.003	0.15	0.23	0.14	-0.39	0.27	-0.40	0.27
Age	-0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.06*	0.03	0.06*	0.03
Job position FE	√		√		√		√		√		√	
Department FE	√		√		√		√		√		√	
Year and month FE	√		√		√		√		√		√	
Genetic principal components	√		√		√		√		√		√	
(Pseudo) R^2	0.22		0.19		0.14		0.23		0.07		0.07	

Notes. N = 3,150 employee-month observations. Coef. = coefficient; *SE* = standard error; FE = fixed effects; NB = negative binomial; ZINB = zero-inflated negative binomial. Columns (1)–(4) present OLS estimates. Columns (5) and (6) examine GenAI use frequency measured as the monthly login count on the company's AI platform, using negative binomial (NB) and zero-inflated negative binomial (ZINB) regressions, respectively. In column (6), the zero-inflation component includes an intercept only, and the reported coefficients are from the negative binomial count component. Cluster-robust standard errors at the employee level are reported. * $p < 0.05$; ** $p < 0.01$.

Table S4

Sensitivity Analyses Excluding Extremes of GenAI Use and Zero-Login Months

	(1) GenAI use frequency		(2) GenAI interaction intensity		(3) GenAI use frequency		(4) GenAI interaction intensity		(5) GenAI interaction intensity		(6) GenAI interaction intensity	
	Top 5% excluded		Top 5% excluded		Top 10% excluded		Top 10% excluded		Nonusers excluded		Observations with ≥ 1 login	
	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*
Intercept	-1.56	1.77	-0.28	0.28	-5.23**	1.77	-0.38	0.29	-0.40	0.33	0.29	0.23
Independent Variable												
Educational attainment PGS	0.26*	0.11	0.02	0.02	0.36**	0.10	0.02	0.02	0.03	0.02	-0.03	0.02
Intermediate Variables												
Education level	-0.35	0.22	-0.03	0.04	0.12	0.22	0.0006	0.04	-0.02	0.04	0.05	0.03
Strategic mindset	0.40	0.22	0.03	0.03	0.53**	0.20	0.03	0.03	0.04	0.03	-0.04	0.03
GenAI credibility perceptions	0.37**	0.11	0.07*	0.03	0.24*	0.11	0.07*	0.03	0.10**	0.02	0.09**	0.02
Metaknowledge of expertise location	0.40*	0.20	0.09**	0.03	0.63**	0.18	0.10**	0.03	0.07*	0.03	0.09*	0.04
Control Variables												
Gender	0.12	0.23	-0.01	0.05	-0.51*	0.20	-0.03	0.05	-0.05	0.06	0.004	0.04
Age	-0.03	0.03	-0.002	0.01	0.02	0.03	-0.0009	0.01	0.004	0.01	0.01	0.004
Job position FE	√		√		√		√		√		√	
Department FE	√		√		√		√		√		√	
Year and month FE	√		√		√		√		√		√	
Genetic principal components	√		√		√		√		√		√	
(Pseudo) R^2	0.43		0.35		0.36		0.31		0.32		0.48	
N	2,982		2,982		2,835		2,835		2,520		786	

Notes. The full sample consists of 3,150 employee-month observations. Coef. = coefficient; *SE* = standard error; FE = fixed effects. Columns (1) and (2) exclude employees in the top 5% of the corresponding employee-level GenAI use distribution, whereas columns (3) and (4) exclude employees in the top 10%. Column (5) excludes nonusers, defined as employees who recorded no valid logins during the observation period. Column (6) restricts the interaction-intensity analysis to employee-months with at least one valid login; zero-login employee-months are excluded rather than assigned an interaction-intensity value of zero. This restriction yields 786 employee-month observations. The resulting sample size for each specification is reported in the "*N*" row. GenAI use frequency in columns (1) and (3) is measured as the monthly login count on the company's AI platform. GenAI interaction intensity in columns (2), (4), (5), and (6) is measured as dialogue turns per login and is log-transformed. Columns (1) and (3) present Poisson regression estimates, whereas columns (2), (4), (5), and (6) present linear mixed-effects estimates with employee-level random intercepts. Cluster-robust standard errors at the employee level are reported. * $p < 0.05$; ** $p < 0.01$.

Table S5

Generalized Structural Equation Modeling Results with a Time Trend and Educational Attainment PGS Interaction

	(1) Education level		(2) Strategic mindset		(3) GenAI credibility perceptions		(4) Metaknowledge of expertise location		(5) GenAI use frequency		(6) GenAI interaction intensity	
	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*
Intercept	2.98**	0.39	3.56**	0.37	4.49**	0.69	3.74**	0.57	-5.02**	1.61	-0.50	0.27
Independent Variable												
Educational attainment PGS	0.11*	0.05	0.13*	0.06	0.05	0.08	-0.03	0.06	-0.01	0.24	0.01	0.03
Intermediate Variables												
Education level					-0.27*	0.12	-0.18	0.10	-0.18	0.16	-0.03	0.04
Strategic mindset					0.07	0.11	0.38**	0.09	0.15	0.18	0.03	0.03
GenAI credibility perceptions									0.29*	0.12	0.07**	0.03
Metaknowledge of expertise location									0.52**	0.20	0.10**	0.03
Temporal Effects												
Trend									0.05**	0.02	0.01**	0.003
Trend × Educational attainment PGS									0.01	0.02	0.0009	0.003
Control Variables	√		√		√		√		√		√	
(Pseudo) R^2	0.22		0.19		0.14		0.23		0.60		0.36	

Notes. N = 3,150 employee-month observations. Coef. = coefficient; *SE* = standard error. GenAI use frequency in column (5) is measured as the monthly login count on the company's AI platform. GenAI interaction intensity in column (6) is measured as dialogue turns per login and is log-transformed. The educational attainment PGS is constructed using the GWAS results reported by Lee et al. (2018b) and is standardized to have a mean of 0 and a standard deviation of 1. All models control for gender and age and include job position and department fixed effects, as well as the top 10 genetic principal components. The temporal fixed effects used in the primary analyses are replaced by the continuous Trend variable. Columns (1)–(4) present OLS estimates, column (5) presents Poisson regression estimates, and column (6) presents linear mixed-effects estimates with employee-level random intercepts. Cluster-robust standard errors at the employee level are reported. * $p < 0.05$; ** $p < 0.01$.

Table S6

Generalized Structural Equation Modeling Results Testing the Potential Moderating Effects

	(1) Education level		(2) Strategic mindset		(3) GenAI credibility perceptions		(4) Metaknowledge of expertise location		(5) GenAI use frequency		(6) GenAI interaction intensity	
	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*
Intercept	2.98**	0.39	3.56**	0.37	4.49**	0.69	3.74**	0.57	-6.04**	1.72	-0.47	0.26
Independent Variable												
Educational attainment PGS	0.11*	0.05	0.13*	0.06	0.05	0.08	-0.03	0.06	-0.19	0.83	-0.02	0.16
Intermediate Variables												
Education level					-0.27*	0.12	-0.18	0.10	-0.15	0.17	-0.01	0.04
Strategic mindset					0.07	0.11	0.38**	0.09	0.31	0.17	0.03	0.03
GenAI credibility perceptions									0.30**	0.11	0.08**	0.03
Metaknowledge of expertise location									0.58**	0.19	0.10**	0.03
Interaction Effects												
Educational attainment PGS × Gender									-0.29	0.22	-0.01	0.04
Educational attainment PGS × Age									0.01	0.03	0.002	0.01
Educational attainment PGS × Junior manager									-1.14**	0.30	-0.06	0.07
Educational attainment PGS × Mid-level manager									-0.32	0.31	-0.10	0.09
Educational attainment PGS × Senior manager									0.56*	0.23	-0.08	0.08
Educational attainment PGS × Operations									0.63	0.36	-0.02	0.06
Educational attainment PGS × Information technology									-0.33	0.53	0.13	0.12
Educational attainment PGS × Customer service									0.80	0.85	-0.19	0.10
Educational attainment PGS × Administration									0.96*	0.43	-0.002	0.06
Control Variables	√		√		√		√		√		√	
(Pseudo) R^2	0.22		0.19		0.14		0.23		0.64		0.37	

Notes. N = 3,150 employee-month observations. Coef. = coefficient; *SE* = standard error. GenAI use frequency in column (5) is measured as the monthly login count on the company's AI platform. GenAI interaction intensity in column (6) is measured as dialogue turns per login and is log-transformed. The educational attainment PGS is constructed using the GWAS results reported by Lee et al. (2018b) and is standardized to have a mean of 0 and a standard deviation of 1. All models control for gender and age and include job position, department, and year and month fixed effects, as well as the top 10 genetic principal components. Columns (1)–(4) present OLS estimates, column (5) presents Poisson regression estimates, and column (6) presents linear mixed-effects estimates with employee-level random intercepts. Cluster-robust standard errors at the employee level are reported. * $p < 0.05$; ** $p < 0.01$.

Table S7

Comparison of Proposed and Alternative Pathways

GenAI use outcome	Specification	Pathways	AIC	BIC	ΔAIC	ΔBIC
Frequency	Proposed	*Baseline cognitive*: Educational attainment PGS → education level → GenAI credibility perceptions → GenAI use *Metacognitive*: Educational attainment PGS → strategic mindset → metaknowledge of expertise location → GenAI use	118,948.40	119,777.90	—	—
	Alternative	*Baseline cognitive*: Educational attainment PGS → education level → GenAI use → GenAI credibility perceptions *Metacognitive*: Educational attainment PGS → strategic mindset → GenAI use → metaknowledge of expertise location	123,418.90	124,230.30	4,470.50	4,452.40
Intensity	Proposed	*Baseline cognitive*: Educational attainment PGS → education level → GenAI credibility perceptions → GenAI use *Metacognitive*: Educational attainment PGS → strategic mindset → metaknowledge of expertise location → GenAI use	31,539.66	32,369.22	—	—
	Alternative	*Baseline cognitive*: Educational attainment PGS → education level → GenAI use → GenAI credibility perceptions *Metacognitive*: Educational attainment PGS → strategic mindset → GenAI use → metaknowledge of expertise location	31,586.81	32,410.32	47.15	41.10

Notes. N = 3,150 employee-month observations. AIC = Akaike information criterion; BIC = Bayesian information criterion. Each specification includes the baseline and metacognitive pathways shown in the Pathways column. ΔAIC and ΔBIC are calculated as the alternative specification minus the corresponding proposed specification. Lower AIC and BIC values indicate better relative fit among the candidate models.

Table S8

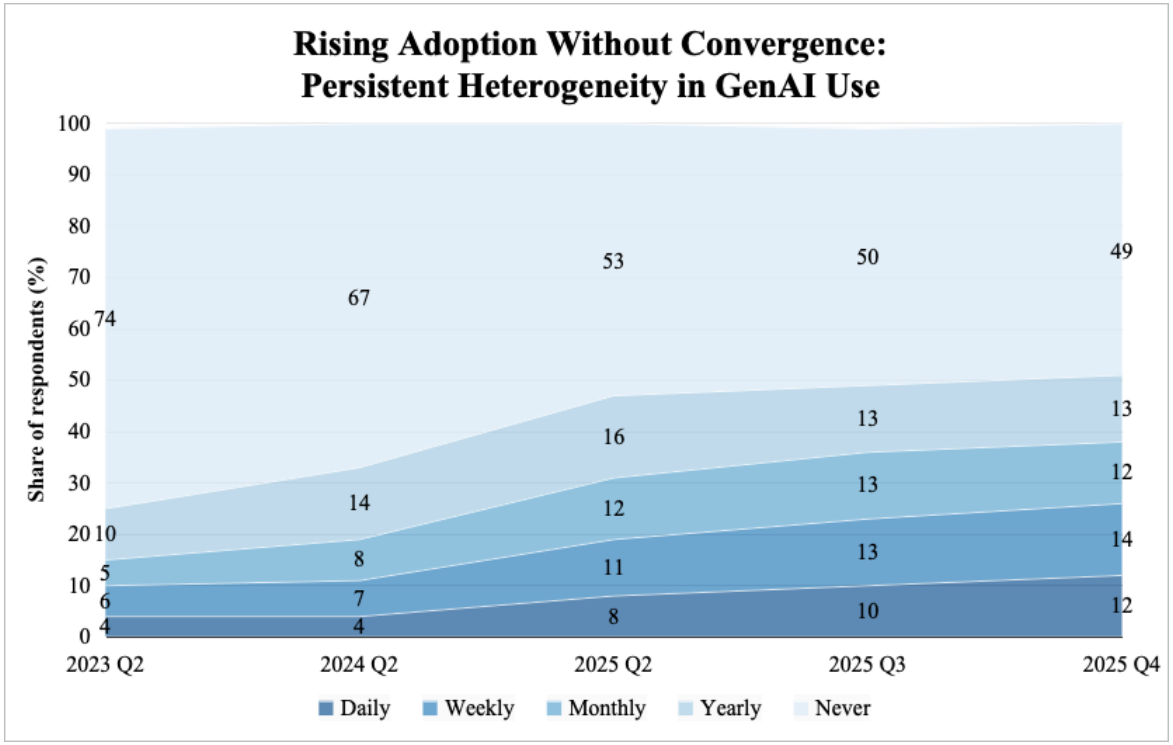
Robustness Checks Using Wild Cluster Bootstrap Inference

	(1) Education level		(2) Strategic mindset		(3) GenAI credibility perceptions		(4) Metaknowledge of expertise location		(5) GenAI use frequency		(6) GenAI interaction intensity	
	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*	Coef.	*SE*
Intercept	2.98**	0.39	3.56**	0.37	4.49**	0.69	3.74**	0.57	-5.20**	1.63	-0.39	0.28
	[1.84, 3.96]		[2.70, 4.42]		[3.01, 6.05]		[2.53, 4.95]		[N.A.]		[-0.99, 0.21]	
Independent Variable												
Educational attainment PGS	0.11*	0.05	0.13*	0.06	0.05	0.08	-0.03	0.06	0.17	0.15	0.02	0.02
	[0.004, 0.22]		[0.01, 0.27]		[-0.11, 0.23]		[-0.15, 0.10]		[N.A.]		[-0.02, 0.06]	
Intermediate Variables												
Education level					-0.27*	0.12	-0.18	0.10	-0.18	0.16	-0.03	0.04
					[-0.55, -0.002]		[-0.41, 0.03]		[N.A.]		[-0.10, 0.05]	
Strategic mindset					0.07	0.11	0.38**	0.09	0.15	0.18	0.03	0.03
					[-0.17, 0.30]		[0.19, 0.58]		[N.A.]		[-0.03, 0.09]	
GenAI credibility perceptions									0.29*	0.12	0.07**	0.03
									[N.A.]		[0.01, 0.13]	
Metaknowledge of expertise location									0.52**	0.20	0.10**	0.03
									[N.A.]		[0.03, 0.16]	
Control Variables	√		√		√		√		√		√	

Notes. N = 3,150 employee-month observations. Coef. = coefficient; *SE* = standard error; N.A. = not applicable. GenAI use frequency in column (5) is measured as the monthly login count on the company's AI platform. GenAI interaction intensity in column (6) is measured as dialogue turns per login and is log-transformed. The educational attainment PGS is constructed using the GWAS results reported by Lee et al. (2018b) and is standardized to have a mean of 0 and a standard deviation of 1. All models control for gender and age and include job position, department, and year and month fixed effects, as well as the top 10 genetic principal components. Columns (1)–(4) present OLS estimates, column (5) presents Poisson regression estimates, and column (6) presents linear mixed-effects estimates with employee-level random intercepts. Cluster-robust standard errors at the employee level are reported. Values in square brackets in columns (1)–(4) and (6) represent 95% confidence intervals obtained using wild cluster bootstrap inference with Rademacher weights and 2,000 replications. N.A. in column (5) indicates that wild cluster bootstrap confidence intervals are not reported for the Poisson model, as the standard implementation of this procedure is typically applied to linear models; wild cluster bootstrap p-values remain available and are used to determine statistical significance. Significance levels (* $p < 0.05$; ** $p < 0.01$) are based on wild cluster bootstrap p-values.

Figure S1

Heterogeneous GenAI Use Persists Over Time



Note. The figure shows the distribution of AI usage frequency among employees over time, illustrating persistent heterogeneity in GenAI use. Values are reported as percentages. Data source: Gallup (Kemp, 2026).