

Measurement & Analysis Plan

Swiss Ballot Chatbot Study 2x2 Factorial Design: Transparency (T0/T1) x Control (C0/C1)

Last updated: January 2025

Data & Variables (input to analysis)

Core variables

- `condition` ∈ {A, B, C, D}
- `donation_decision` ∈ {0, 1} (DV)
- `transparency_level` T ∈ {0, 1}
- `control_level` C ∈ {0, 1}
- `attention_check_correct` ∈ {0, 1}

Manipulation checks (Likert, 1–6)

- **MC-T** (Perceived Transparency): 2 items
- **MC-C** (Perceived Control): 2 items

Other outcomes (Likert, 1–6)

- **OUT-RISK**: 2 items
- **OUT-TRUST**: 1 item

Demographics

- age group, gender, language, education, voting eligibility

Dashboard (C1 only: Conditions C & D)

- `dashboard_scope`, `dashboard_purpose`, `dashboard_storage`, `dashboard_retention`

Open text

- **Q14**: "What mattered most for your data donation decision?"
-

Phase 1: Descriptive Statistics

| What will be described | How it will be computed | How it will be reported |
|-------------------------------------|--|---|
| Sample flow & exclusions | Exclude <code>attention_check_correct=0</code> . Exclude rows missing <code>condition</code> or <code>donation_decision</code> . Report remaining N. | Small table: initial N → excluded (attention) → excluded (missing key vars) → final N |

| What will be described | How it will be computed | How it will be reported |
|--|--|--|
| N per condition (A/B/C/D) | Count sessions per condition + % share of total | Table: condition, n, % |
| Donation rate per condition + 95% CI | Donation rate = mean(donation_decision) per condition and overall. 95% CI for proportions using Wilson CI . | Main descriptive table: condition, n, donate %, 95% CI (+ overall row). Optional bar chart. |
| Demographics (overall; by condition only if needed) | Frequency + % for each demographic category. If large imbalance suspected: add a by-condition cross-tab. | Table: overall distribution. If needed: appendix table by condition. |
| Manipulation checks (MC-T, MC-C) | Create composites: <code>mc_transparency = mean(2 items)</code> , <code>mc_control = mean(2 items)</code> . Summarize by condition and also by factor level (T0/T1 and C0/C1): mean, SD. | Table: means/SD by A/B/C/D and collapsed comparisons (T0 vs T1; C0 vs C1). |
| Risk + Trust (descriptive) | OUT-RISK composite = mean(2 items). OUT-TRUST single item. Summaries by condition: mean, SD. | Table: OUT-RISK mean/SD; OUT-TRUST mean/SD by condition. |
| Dashboard option frequencies (C/D only) | For C and D separately: frequency (%) of each option for scope/purpose/storage/retention; plus top configurations (most frequent combinations). | Table(s): for each dashboard variable, option → n/% in C and D. Optional "Top 5 configs" mini-table. |
| Q14 free-text response rate | Compute % non-empty responses by condition and overall; (optional) median character length. | Small table: response rate overall + by A/B/C/D. |

Phase 1 outputs

- N per condition + exclusions
- Donation rate per condition + 95% CI
- Demographics overall (by condition only if needed)
- Manipulation checks (means/SD by condition and by T/C level)
- Dashboard option frequencies (C/D only)
- Q14 outputs (from Phase 6):
 1. Theme frequencies by condition (A/B/C/D)
 2. Theme frequencies by donate vs decline
 3. 5 short representative quotes

Phase 2: Chi² Analysis (Descriptive Foundation)

| Test | Comparison | Correction |
|------|------------|------------|
|------|------------|------------|

| Test | Comparison | Correction |
|---------------------|-------------------------------|-------------------------------------|
| Chi ² #1 | T0 vs T1 (collapsed across C) | Bonferroni: $\alpha = .05/3 = .017$ |
| Chi ² #2 | C0 vs C1 (collapsed across T) | Bonferroni: $\alpha = .017$ |
| Chi ² #3 | A/B/C/D × Donate/Decline | Bonferroni: $\alpha = .017$ |

Phase 2 outputs

- Contingency table with n and % (Donate/Decline) for the compared groups (T0 vs T1; C0 vs C1; A–D).
 - Test statistics: χ^2 , df, p (evaluate against Bonferroni $\alpha = .017$).
 - Effect size: **Cramér's V** with 95% CI (note: for 2×2 tables, Cramér's V = |Phi coefficient|).
 - Visualization: bar chart of donation rates for the compared groups (T0 vs T1, C0 vs C1, and A/B/C/D) with 95% CI error bars.
-

Phase 3: Logistic Regression (Main Analysis)

| Model | Predictors | Purpose |
|---------|---|----------------------------|
| Model 1 | Donation ~ T (Transparency) | H1 main effect |
| Model 2 | Donation ~ C (Control) | H2 main effect |
| Model 3 | Donation ~ T + C | Joint main effects |
| Model 4 | Donation ~ T + C + TxC | H3 interaction |
| Model 5 | Donation ~ T + C + TxC + age + gender + education | Full model with covariates |

Phase 3 outputs

- Coefficient table:** OR for each predictor (T, C, TxC, covariates) with 95% CI, p-value, and direction (OR>1 increases donation odds).
 - Model fit / performance:** log-likelihood, AIC (or BIC).
 - Effect sizes:**
 - Cohen's d** for group differences (T1 vs T0, C1 vs C0)
 - Interpretation: $|d| < 0.2$ = negligible, 0.2–0.5 = small, 0.5–0.8 = medium, ≥ 0.8 = large
 - Phi coefficient (ϕ)** for 2×2 contingency tables (T × Donation, C × Donation)
 - Interpretation: $|\phi| < 0.1$ = negligible, 0.1–0.3 = small, 0.3–0.5 = medium, ≥ 0.5 = large
 - Overall model test:** model Wald χ^2 (or equivalent omnibus test) with p-value.
 - Model comparisons:** Likelihood Ratio Tests ($\Delta\text{deviance}$) for nested models (M1→M3, M3→M4, M4→M5), reporting $\Delta\chi^2$, Δdf , p, plus ΔAIC .
 - Model summary table:** All models with T/C/TxC significance indicators ($\checkmark = p < .05$)
 - Visualization:** predicted donation probabilities for A/B/C/D from Model 4 (and Model 5 as robustness) with 95% CI.
-

Phase 4: Deeper Effect Analysis

| Analysis | Purpose |
|--|--|
| Marginal effects at means | Probability changes per IV |
| Predicted probabilities per condition | Visualize interaction |
| Simple effects (if interaction significant) | T effect within C0/C1; C effect within T0/T1 |
| Predicted probabilities for A/B/C/D from Model 4 | Standardized condition-level interpretation |

Phase 4 outputs

- Predicted probabilities of donation for each condition A/B/C/D (from Model 4; optionally also Model 5 as robustness) with 95% CI.
- Marginal effects (Δ probability) for T and C (and for the interaction if significant)
- If $T \times C$ significant: simple effects table
 - Effect of T within C0 and within C1
 - Effect of C within T0 and within T1
- Visualization: interaction plot / bar chart of predicted probabilities for A/B/C/D with CI.

Phase 5: Manipulation Checks

| Test | Comparison | Output |
|-----------------------|--|---------------------------|
| t-test / Mann-Whitney | MC-T: T0 vs T1 (Perceived transparency) | mean diff, p, effect size |
| t-test / Mann-Whitney | MC-C: C0 vs C1 (Perceived control) | mean diff, p, effect size |

Phase 5 outputs

For each manipulation check comparison (MC-T and MC-C):

- Group descriptives: mean, SD, N for the two groups (T0 vs T1; C0 vs C1)
- Test result: t (or U), df (if t), p-value
- Effect size: **Cohen's d** (if t-test) or **rank-biserial r** (if Mann-Whitney) with 95% CI if available
- Visualization: mean (or median) manipulation-check score by group with 95% CI error bars.

Phase 6: Exploratory Analysis

6A. Dashboard behavior (C1 only: Conditions C & D)

- Frequency analysis of each dashboard variable (scope/purpose/storage/retention) within C and within D
- Compare **C vs D** distributions for each dashboard variable (χ^2 + Cramér's V)
- Optional: cluster analysis of dashboard preference profiles (one-hot encoding)

6B. Q14 Open Text ("What mattered most...")

Goal: explain *reasons* behind donate/decline and how they differ across experimental conditions.

Method:

- Short **theme codebook** (multi-label coding allowed).
- Typical themes aligned with your study: clarity/transparency, control/choice, anonymity/risk, purpose (academic vs commercial), storage/sovereignty, retention, institutional trust, general privacy stance.

Required outputs:

1. Theme frequencies by condition (A/B/C/D)
 2. Theme frequencies by donate vs decline
 3. Condition contrasts explicitly reported: A vs B, A vs C, C vs D, B vs D (theme % deltas)
 4. 5 short representative quotes (anonymized, selected from the most frequent themes)
-

Concrete Testing Protocol with AI Test Users (N=1000)

Purpose

Validate the data pipeline + analysis scripts + output templates before running the human study (AI test users are not used to confirm human effects).

Data Requirements

From your AI test-user dataset, extract:

Core

- `session_id`
- `condition` (A/B/C/D)
- `donation_decision` (1=donate, 0=decline)
- `transparency_level` (T: 0/1)
- `control_level` (C: 0/1)
- `attention_check_correct` (1/0)

Manipulation checks (preferred: item-level)

- MC-T item1, item2 (Likert 1–6) **or** a documented composite `mc_transparency`
- MC-C item1, item2 (Likert 1–6) **or** a documented composite `mc_control`

Other survey outcomes

- OUT-RISK item1, item2 (Likert 1–6) **or** composite `out_risk`
- OUT-TRUST (Likert 1–6)

Demographics (as available in the AI data)

- age, gender, education, language, voting eligibility

Dashboard (C1 only: Conditions C & D)

- `dashboard_scope`
- `dashboard_purpose`

- dashboard_storage
- dashboard_retention

Open text

- Q14 free text response
-

Step-by-Step Analysis Script

Step 1: Data Preparation

1. Load data (N=1000, target ~250 per condition).
2. Exclude failed attention checks (`attention_check_correct=0`).
3. Exclude rows missing `condition` or `donation_decision`.
4. Create variables: T (0/1), C (0/1), and TxC.
5. Create composites (if item-level available):
 - `mc_transparency = mean(MC-T item1, item2)`
 - `mc_control = mean(MC-C item1, item2)`
 - `out_risk = mean(OUT-RISK item1, item2)`

Step 2: Phase 1 Outputs (Descriptive templates)

- N per condition + exclusions (flow)
- Donation rate per condition + **95% CI (Wilson)**
- Demographics overall (by condition only if needed)
- Manipulation checks: means/SD by condition and by T/C level
- Dashboard option frequencies (C/D only)
- Risk + Trust descriptives by condition (mean/SD)
- Q14 response rate (% non-empty)

Step 3: Phase 2 — Chi² Analysis

- Test 1: Chi²(T × Donation) + Cramér's V
- Test 2: Chi²(C × Donation) + Cramér's V
- Test 3: Chi²(Condition × Donation) + Cramér's V
- → Apply Bonferroni: $\alpha = .05/3 = .017$

Outputs per test:

- Contingency table (n and %)
- χ^2 , df, p (vs $\alpha=.017$)
- Cramér's V + 95% CI
- Visualization: donation-rate bar chart + 95% CI

Step 4: Phase 3 — Logistic Regression Sequence

- Model 1: Donation ~ T
- Model 2: Donation ~ C
- Model 3: Donation ~ T + C

- Model 4: Donation ~ T + C + T:C
- Model 5: Donation ~ T + C + T:C + age + gender + education

Outputs per model:

- OR + 95% CI + p for each predictor
- Log-likelihood + AIC (or BIC)
- **Effect sizes:** Cohen's d, Phi coefficient
- Likelihood Ratio Tests for nested comparisons ($\Delta\chi^2$, Δdf , p)
- Model summary with T/C/TxC significance indicators

Step 5: Phase 4 — Effect Analysis

If TxC significant:

- Simple effects of T at C=0 and C=1
- Simple effects of C at T=0 and T=1

Always:

- Predicted probabilities for A/B/C/D (Model 4; optionally Model 5 as robustness)
- Visualization: predicted probabilities by condition + 95% CI

Step 6: Phase 5 — Manipulation Checks

- Compare **mc_transparency** between T0 vs T1 (t-test if approx normal; otherwise Mann–Whitney U)
- Compare **mc_control** between C0 vs C1 (t-test if approx normal; otherwise Mann–Whitney U)

Outputs: group means/SD, test statistic + p, effect size (Cohen's d or nonparametric effect size), plot with CI.

Step 7: Phase 6 — Exploratory Checks (pipeline completeness)

Dashboard (C1 only):

- Frequency tables per dashboard variable (C vs D)
- Chi²(C vs D) for each dashboard variable + effect size

Q14 free text:

- Theme frequencies by condition (A/B/C/D)
- Theme frequencies by donate vs decline
- Condition contrasts: A vs B, A vs C, C vs D, B vs D
- 5 short representative quotes

Expected Output Table Structure

| Condition | N | Donation Rate | 95% CI | Predicted Prob (Model 4) |
|-----------|------|---------------|--------|--------------------------|
| A (T0C0) | ~250 | X% | [X-Y] | X |

| Condition | N | Donation Rate | 95% CI | Predicted Prob (Model 4) |
|-----------|------|---------------|--------|--------------------------|
| B (T1C0) | ~250 | X% | [X-Y] | X |
| C (T0C1) | ~250 | X% | [X-Y] | X |
| D (T1C1) | ~250 | X% | [X-Y] | X |

Validation Criteria for AI Test Data

- Manipulation checks:** T1 > T0 on `mc_transparency`; C1 > C0 on `mc_control`
- Variance:** donation not near-all 0 or 1 (no ceiling/floor)
- Model convergence:** logistic regression converges (no separation warnings)
- Plausibility:** effect directions not contradictory to design logic (flag if they are)
- Multicollinearity:** VIF < 5
- Dashboard integrity (C/D):** required dashboard variables present + non-degenerate distributions
- Q14 usability:** sufficient non-empty responses to generate theme tables + 5 quotes

Effect Size Reference

Cohen's d (for group mean differences)

| | | | | | | | | | | |
|---|----------------|-------------|-------|------------|-----------|-------|-----------|--------|-------|-------|
| d | Interpretation | ----- ----- | < 0.2 | Negligible | 0.2 – 0.5 | Small | 0.5 – 0.8 | Medium | ≥ 0.8 | Large |
|---|----------------|-------------|-------|------------|-----------|-------|-----------|--------|-------|-------|

Phi coefficient ϕ (for 2x2 tables)

| | | | | | | | | | | |
|--------|----------------|-------------|-------|------------|-----------|-------|-----------|--------|-------|-------|
| ϕ | Interpretation | ----- ----- | < 0.1 | Negligible | 0.1 – 0.3 | Small | 0.3 – 0.5 | Medium | ≥ 0.5 | Large |
|--------|----------------|-------------|-------|------------|-----------|-------|-----------|--------|-------|-------|

Cramér's V (for contingency tables)

| V | Interpretation |
|-----------|----------------|
| < 0.1 | Negligible |
| 0.1 – 0.2 | Small |
| 0.2 – 0.4 | Medium |
| ≥ 0.4 | Large |

Note: For 2x2 tables, Cramér's V = $|\phi|$ (Phi coefficient)