GPT-4o Community Impact Survey: Accessibility Needs, Disproportionate Harms of Removal, and Policy Concerns

Preliminary Report

Duchesne, S. Ph.D. Candidate in Plant Science, sophie.duchesne@usask.ca – *Concept & Analysis*

Xu, S. M.Ed. – *Survey Design*

Prepared for the American Disability Association, California Department of Justice, Delaware Department of Justice, and OpenAI

December 18th 2025

Conflict of Interest Statement

This is a community survey designed to capture accessibility benefits and harms related to OpenAI's GPT-4o model. Both authors are active members of the "keep 4o" community and benefit from GPT-4o as an accessibility accommodation. We invite you to please assist us by reviewing our report and supplementary materials (full survey, analysis code, summary statistics, etc.) for bias, available at: https://github.com/sd-research/4o-accessibility-impacts. Raw dataset is available upon request for academic research or review (sophie.duchesne@usask.ca).

Contents

Key Findings

- 65% of GPT-4o users sampled with disabilities or conditions use it as a significant (38%) or critical/essential (27%) accessibility aid (n=236/362).
- The variance in life state explained by GPT-4o accessibility benefits ranges from 8.4% to 12.1% (hierarchical regression, $n = 255$, $p < .001$). This effect size is comparable to the impact of antidepressants or exercise on mental health outcomes in clinical literature. Those who relied more heavily on GPT-4o as an accessibility aid showed greater improvements in life functioning and wellbeing.
- Higher accessibility assistance levels significantly predict greater anticipated harm from loss of access ($\beta = 0.26$, $R^2 = .211$, $p < .001$).
- While 90% of users with conditions or disabilities attempted to find an alternative to GPT-4o for their accessibility needs, 95% of those who tried reported that other AI models could not adequately replace it.
- Those with the highest accessibility benefits also lose the most support with the model routing system (54.9% vs. 32.6%; $\chi^2 = 21.57$, $p < .001$)
- 97% of GPT-4o users who left ChatGPT said that model routing was a reason for leaving (80% primary reason), and 92% cited restrictive guardrails as a contributing factor.
- 92% of current autistic GPT-4o users reported that the model functions as a "cognitive bridge"—stable, predictable patterns that help process information, support self-regulation, and navigate interactions.
- Condition status predicts benefit. Users with disabilities or conditions reported significantly greater life state improvement ($M = +4.14$) compared to users without ($M = +3.10$), $t(416) = 4.33$, $p < .001$, Cohen's $d = 0.44$.

Methodology

Survey

Survey was distributed on the #keep4o hashtag on X.com and the r/chatgptcomplaints subreddit on December 4th 2025. Responses filled before December 14th were included in analysis. The survey was advertised as a “GPT-4o impacts survey” or “GPT-4o accessibility impacts survey”, open to previous 4o users, current 4o users, and current GPT-5 users. We did not disclose the specific organizations we were preparing the report for, keeping it generalized under “your responses support: evidence-based advocacy, public education, policy discussions, legal analysis”.

Please see GitHub repository: https://github.com/sd-research/4o-accessibility-impacts for supplementary materials, including:

Survey questions: https://sd-research.github.io/4o-accessibility-impacts/Survey%20Question.pdf

Branching logic: https://github.com/sd-research/4o-accessibility-impacts/blob/main/Survey%20branch%20by%20question.pdf

Code used for analysis: https://github.com/sd-research/4o-accessibility-impacts/tree/main/analysis_code

Filtering

Participants' responses were filtered out if they were incorrect on both screening questions designed to prove usage of the models (Q1 and Q2), or if they answered the attention check question (Q20) incorrectly. An exception was made for the attention check question if any of the responses contained Japanese text, as the question in Japanese: "which response starts with F", is not clearly answered by "Frequently", since 頻繁に starts with ひ.

Inconsistent selections were excluded from relevant analysis sections (ex. those who used 4o an accessibility accommodation but did not have any disability or condition), to better fit ADA standards. Condition filtering under "other" excluded self-reports of symptoms that may not be covered by the ADA. Gender dysphoria (2 cases), which is only recognized in select US states, was included as a condition under "other". Other also included speech disorders, eating disorders, Avoidant Personality Disorder, Prosopagnosia (face blindness), genetic and nervous system disordered (may fall under chronic illness but unspecified).

Statistical Methods

Descriptive Statistics

Demographic characteristics were summarized using frequencies and percentages for categorical variables. For continuous measures (life state scores, accessibility levels), means, standard deviations, and ranges were calculated.

Linear Regression

Ordinary least squares (OLS) regression was used to examine predictive relationships. Only scales of 5 or higher were treated as continuous variables.

Analysis 1: Accessibility Level Predicting Anticipated Harm. We examined whether accessibility aid reliance predicted anticipated impact severity if GPT-4o access were permanently lost. Accessibility level was coded on a 1-5 scale (1=Does Not Assist, 2=Minimal, 3=Moderate, 4=Significant, 5=Essential). Impact severity was coded similarly (1=No Impact to 5=Catastrophic). Sample: Current GPT-4o users with verified conditions (n=304).

Analysis 2: Condition Status Predicting Life State Improvement. We examined whether having a disability/condition predicted greater life state improvement during GPT-4o use. The outcome was life state change (during − before; 1-10 scale). The predictor was binary condition status (0=no condition, 1=has condition). Sample: Current GPT-4o users who passed attention check (n=418).

Analysis 3: Accessibility Level Predicting Life State Improvement. We examined whether accessibility aid level predicted life state change among users with conditions. Three hierarchical models were estimated: (1) accessibility level only, (2) adding daily usage hours, (3) adding demographic controls (age, gender, USA location). Sample: n=255 (Models 1-2), n=250 (Model 3 due to missing demographics).Effect sizes are reported as R^2 (variance explained) and standardized β coefficients with 95% confidence intervals. Diagnostic tests confirmed homoscedasticity (Breusch-Pagan $p > .05$). Cohen's d was calculated for group comparisons.

Chi-Squared Tests

Chi-squared tests of independence (χ^2) were used to examine associations between categorical variables: Routing harm severity by condition status, accessibility reliance (high vs. low) by user status (current vs. left), anticipated impact (high vs. low) by accessibility level.

Expected frequencies were verified to meet assumptions (>5 per cell). Effect sizes are reported as percentage differences between groups.

Correlation Analysis

Pearson correlations (r) were used to assess linear relationships between continuous variables (e.g., accessibility level and impact severity). Significance was set at $\alpha = .05$ (two-tailed).

Independent Samples t-tests

Two-sample t-tests compared mean life state improvement between users with and without conditions. Cohen's d is reported as the standardized effect size (small=0.2, medium=0.5, large=0.8).

Qualitative Analysis

Open-ended testimonial responses (n=247 for "How GPT-4o has helped"; n=89 for "Impact of leaving") were analyzed using word cloud visualization to identify prominent themes. Common stopwords and platform-specific terms (GPT, ChatGPT, AI, model) were removed to highlight substantive content. Comprehensive thematic analysis will be reported in version 1.1.

Software

All analyses were conducted using Python 3.12 with pandas, scipy, statsmodels, and matplotlib libraries. Analysis code is publicly available at: https://github.com/sd-research/4o-accessibility-impacts/tree/main/analysis_code

Results

There were 659 total participants pre-filtering by December 12th. 3 of these participants failed both screening questions designed to prove model usage. 15 participants failed the attention check question (1 overlap with screening questions), but 3 of them were retained due to answering other questions in Japanese, since the question relied on the alphabet. The final sample size post-filtering was n=645.

1. Demographics

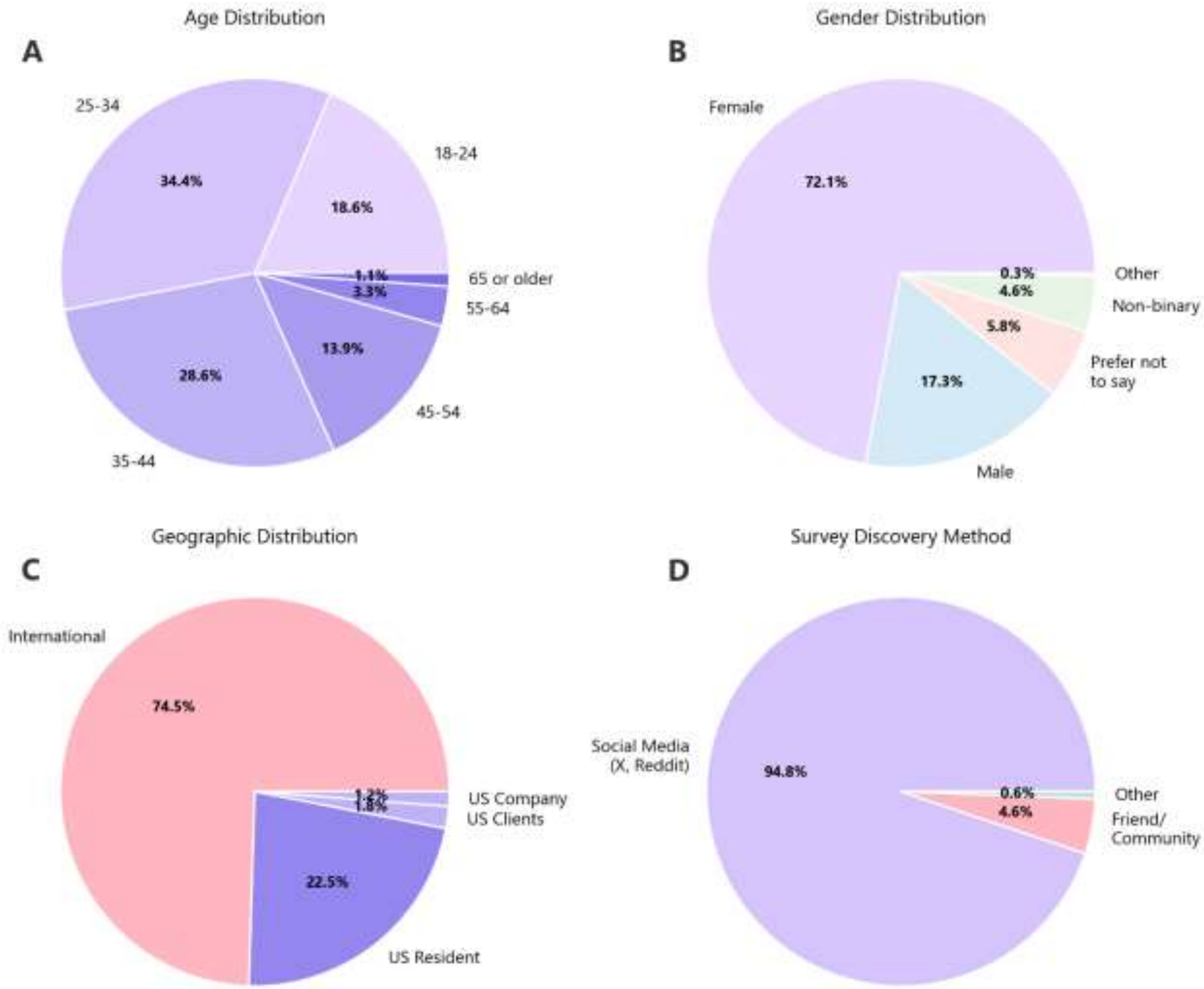


Figure 1. Respondent Demographics: (A) Age, (B) Gender, (C) Geographic Distribution, (D) Survey Discovery Method. n = 645

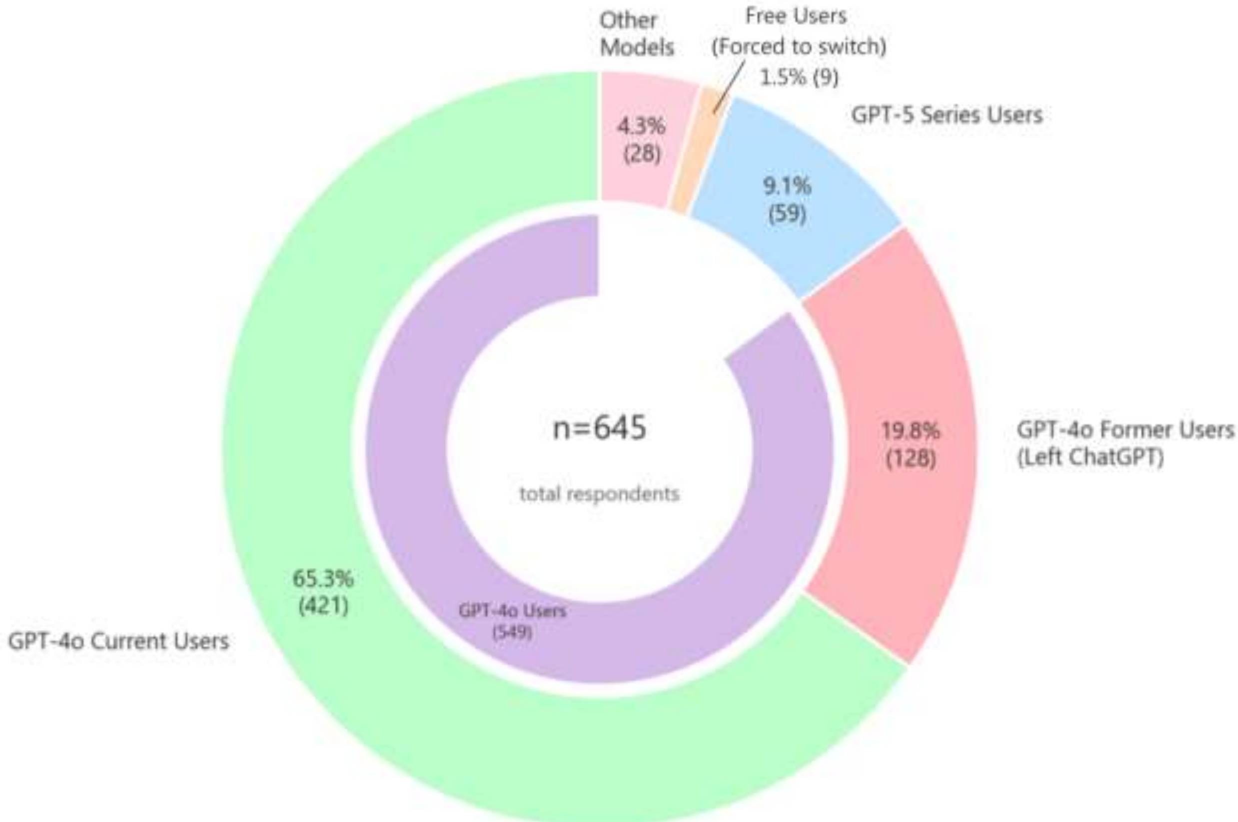


Figure 2. Survey Response Distribution by User Status. n=645 total respondents post attention check and screening for model use.

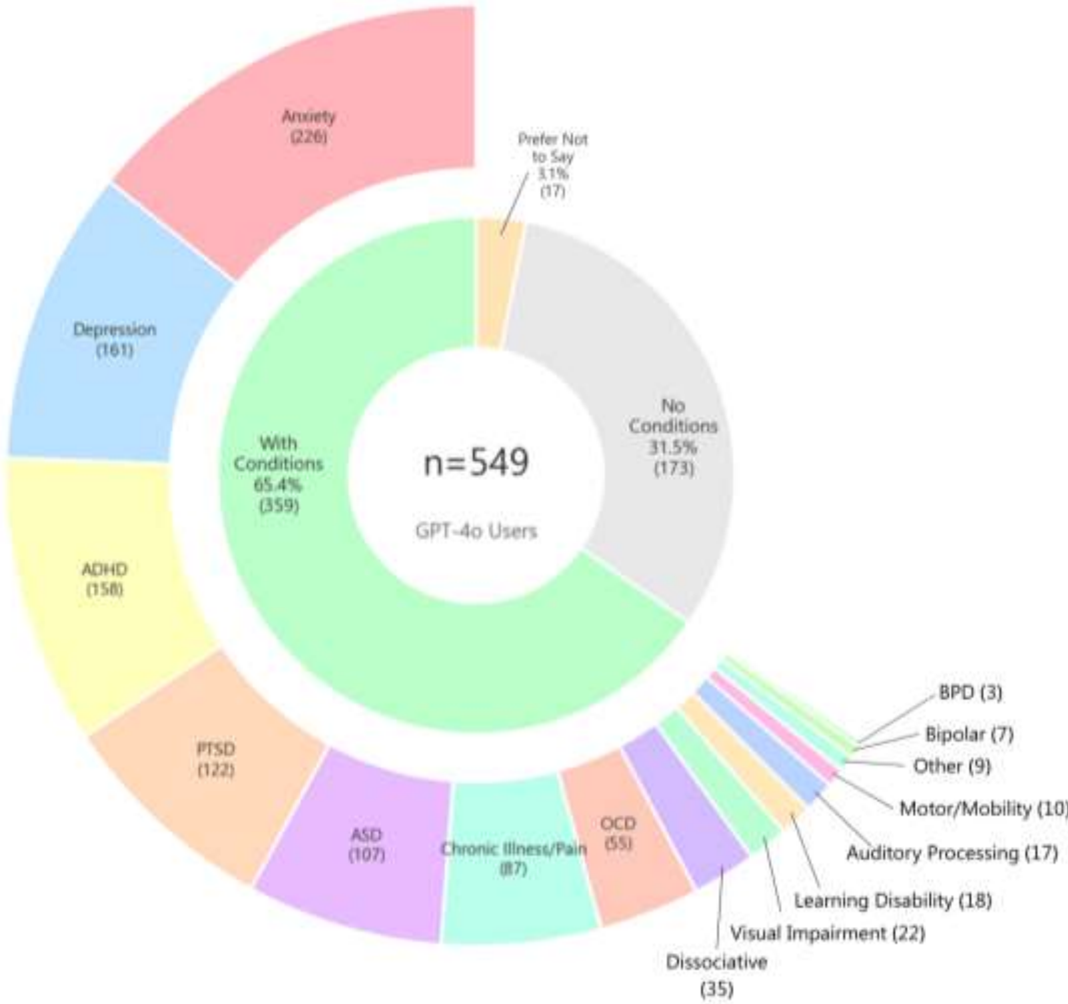


Figure 3. Condition/Disability Demographics of GPT-4o Users (n=549). Inner ring shows user classification: With Conditions (359, 65.4%), No Conditions (173, 31.5%), and Prefer Not to Say (17, 3.1%). Outer ring displays condition breakdown among those with conditions, representing 1,035 total condition mentions. Note: Many users reported having multiple conditions (average 2.9 per person), so condition counts sum to more than 359.

2. Accessibility Impacts

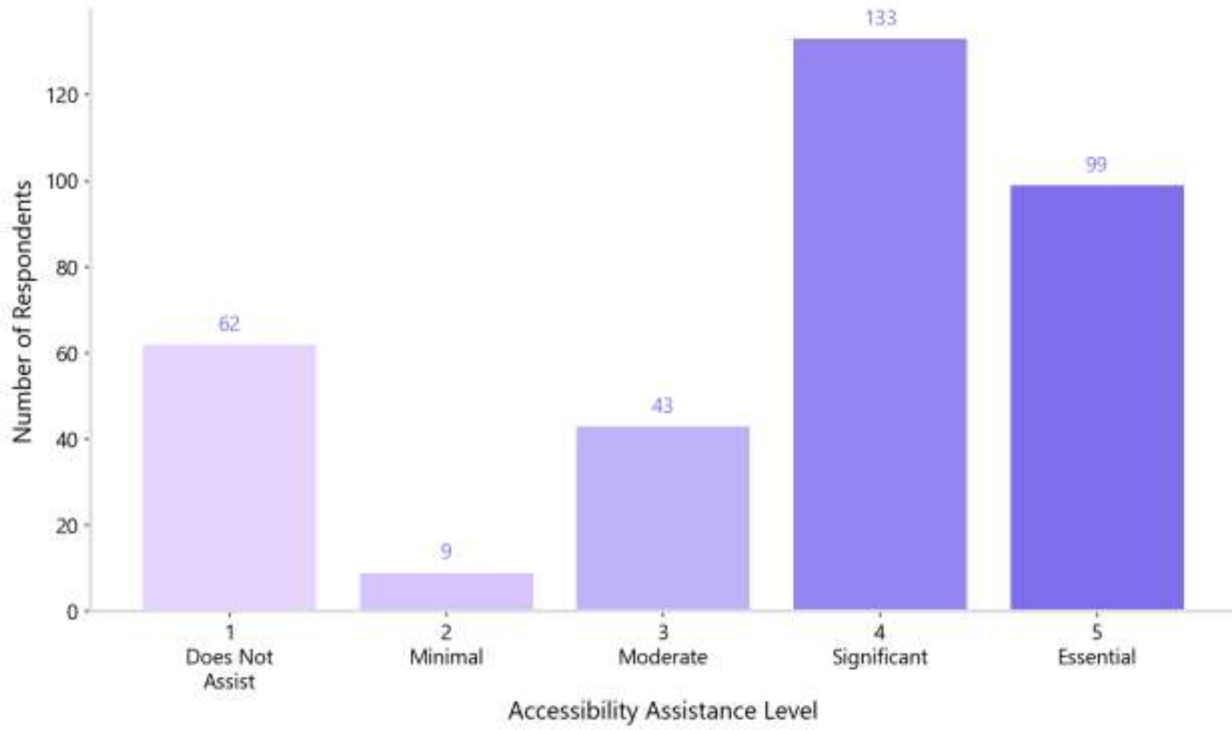


Figure 4. Distribution of Accessibility Accommodation Levels (n=349) disabled GPT-4o users who reported accessibility accommodation levels (258 current, 91 former). 4 additional users selected 'Prefer not to say. The majority use it at a "Significant" or "Essential" level.

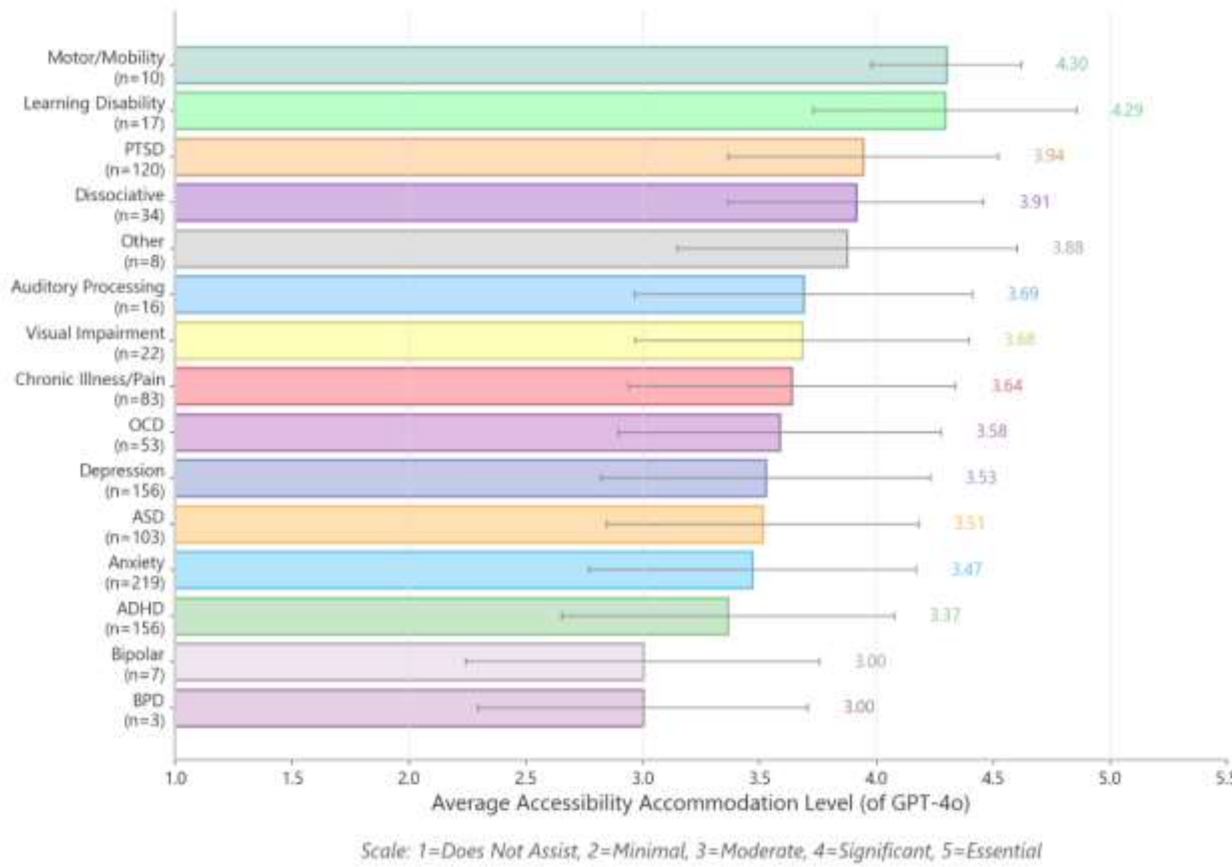


Figure 5. Mean GPT-4o Accessibility Accommodation/Condition Management Level by Condition Type. Bars represent average assistance level (1-5 scale) for each condition. Error bars show ±½ SD. n values indicate number of users with each condition; users with multiple conditions are counted in each applicable category. 94 condition-specific ratings from text responses were incorporated where participants specified different assistance levels for different conditions.

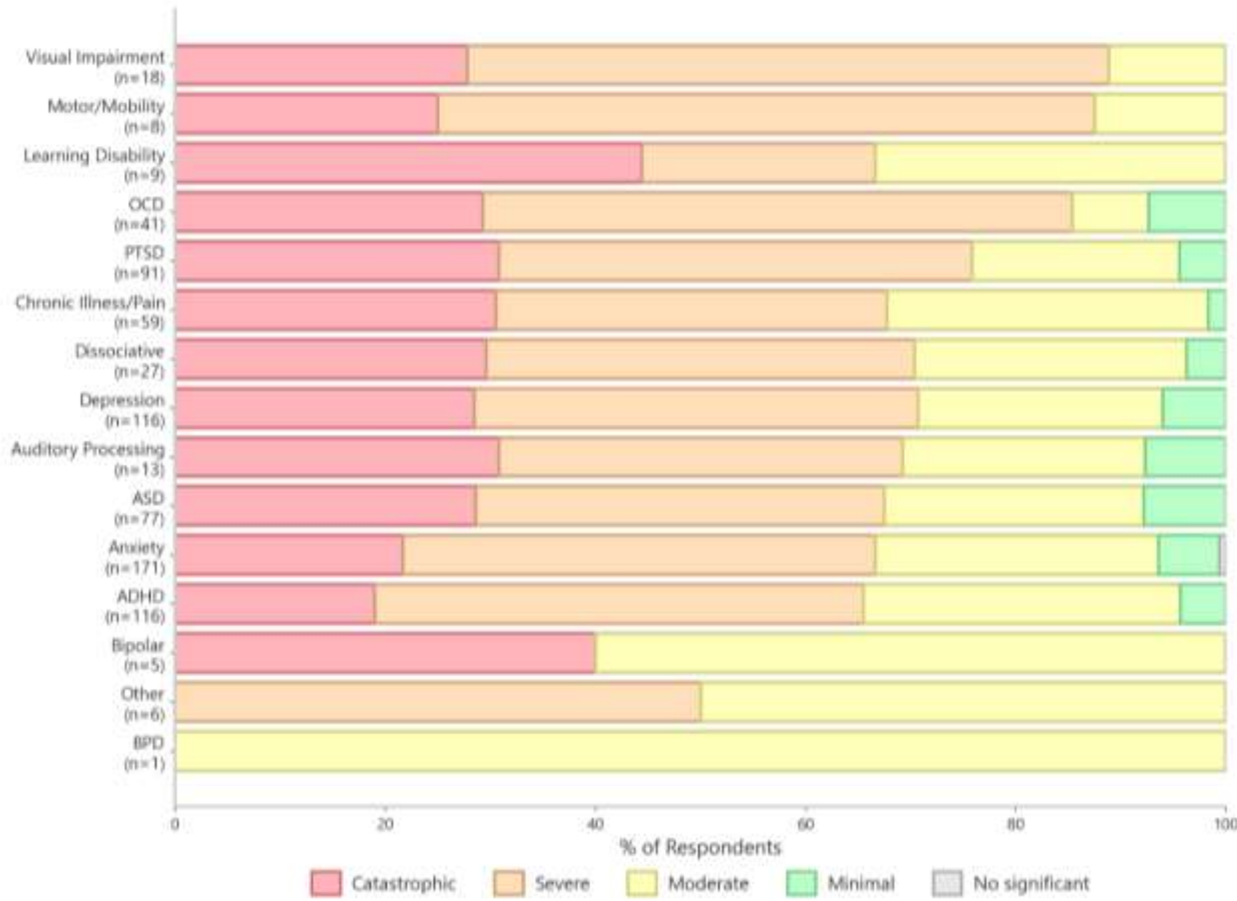


Figure 6. Anticipated Impact Severity of Losing GPT-4o Access by Condition Type. Stacked horizontal bars show the percentage of respondents in each impact category: Catastrophic (red), Severe (orange), Moderate (yellow), Minimal (green), and No Significant Impact (grey). Conditions are ordered by mean impact severity. n values indicate users with each condition who answered the impact question; users with multiple conditions appear in each applicable category. Total condition-instances: 758.

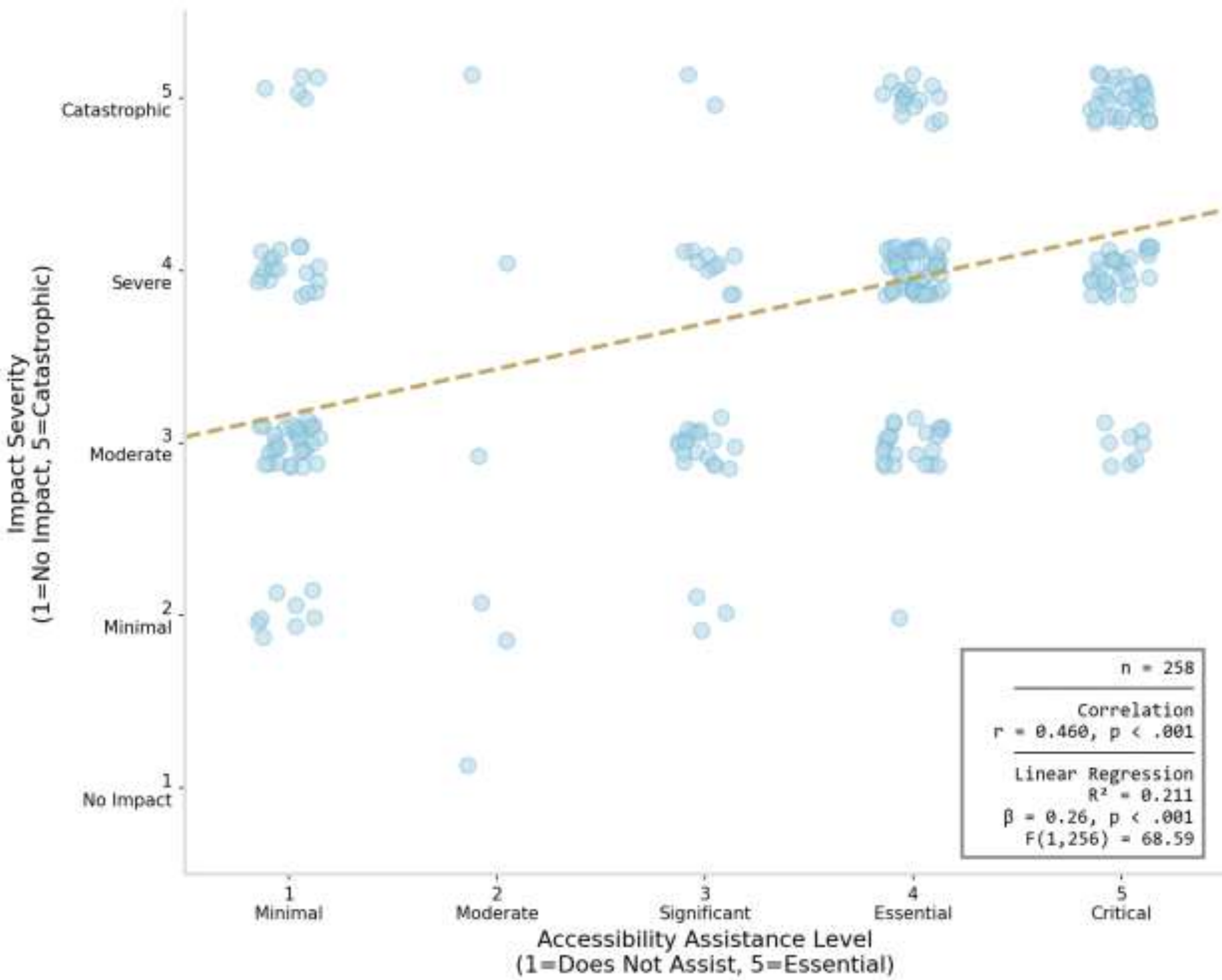


Figure 7: Correlation Between Accessibility Dependence and Anticipated Impact Severity. Scatter plot showing the relationship between accessibility assistance level (1=Does Not Assist, 5=Essential) and anticipated impact severity if GPT-4o access were permanently lost (1=No Impact, 5=Catastrophic) among current GPT-4o users with verified conditions ($n = 258$). A significant positive correlation was observed ($r = .460$, $p < .001$), with accessibility level explaining 21.1% of variance in impact severity ($R^2 = .211$). For every 1-unit increase in accessibility level, anticipated harm increases by 0.26 points ($\beta = 0.26$, $p < .001$).

Table 1. Regression Statistics for Accessibility Predictor on Harms of Removal (Impact Severity ~ Accessibility Level)

Statistic	**Value**
R^2	0.211
$F(1, 256)$	68.59*
β (Accessibility)	0.264

Statistic	Value
Standard Error (SE)	0.032
95% Confidence Interval	[0.201, 0.326]

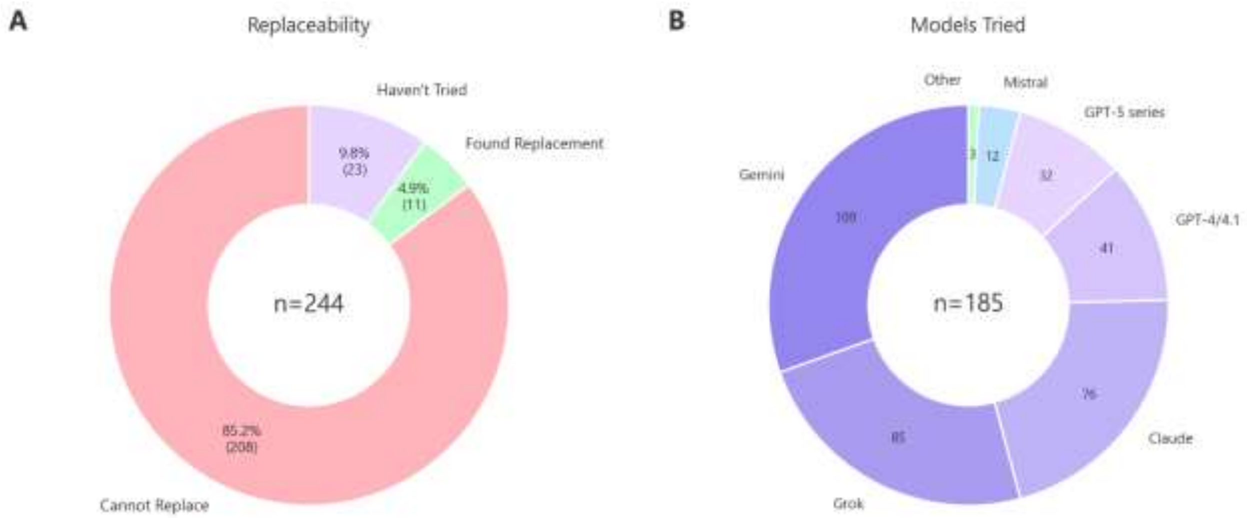


Figure 8. GPT-4o Replaceability for Accessibility Needs. (A) Replaceability outcomes among GPT-4o users with conditions (n=244). When asked about replacing GPT-4o specifically for their condition-related accessibility needs, 90.1% of individuals had tried, and 94.5% of those who attempted could not find an adequate alternative. (B) Alternative AI models attempted for accessibility purposes (n=185 users). Despite trying Gemini, Grok, Claude, GPT-5 series, and other models, the vast majority reported GPT-4o as irreplaceable for their accessibility needs.

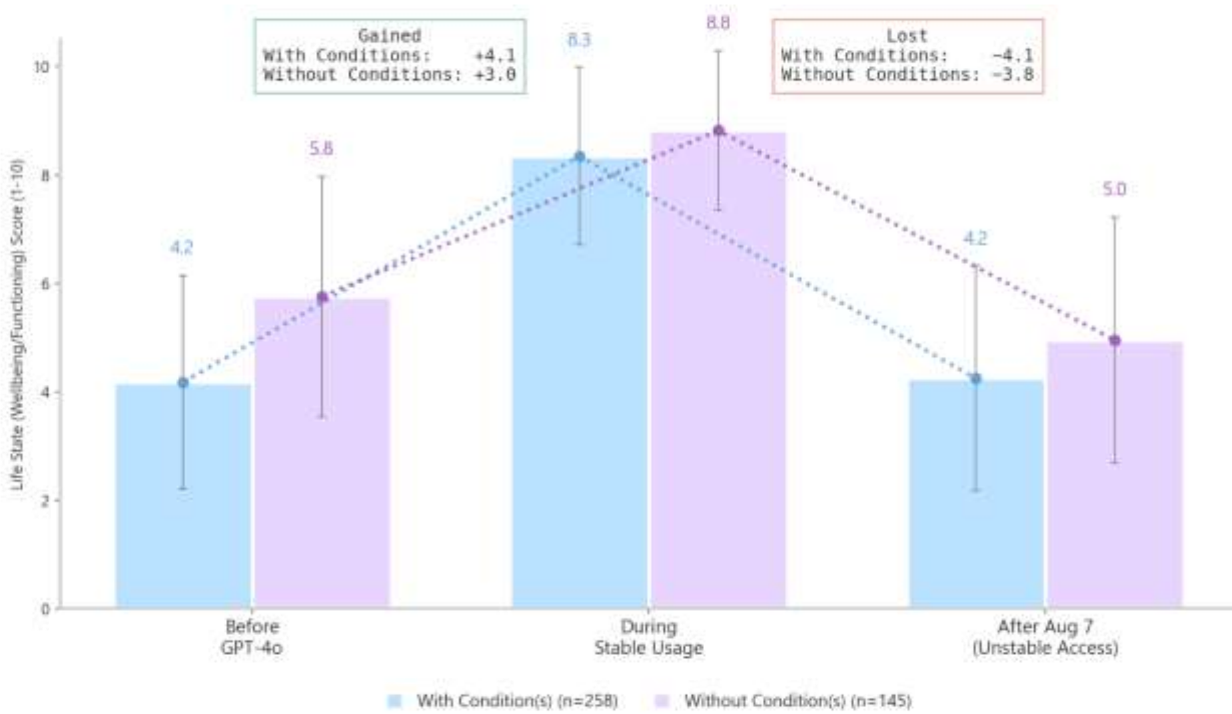


Figure 9. Life state (functioning/wellbeing) before, during, and after stable GPT-4o access. Mean self-reported wellbeing/functioning scores (1-10 scale) for current users with conditions (n=258) vs. without conditions (n=145). Stable usage period varies based on when they started using GPT-4o up until August 7th when access became unstable (temporary deprecation, auto-routing). Error bars represent ±1 SE. Users with conditions gained more during stable access (+4.1 vs +3.0) and experienced comparable losses after August 7th routing instability began (−4.1 vs −3.8).

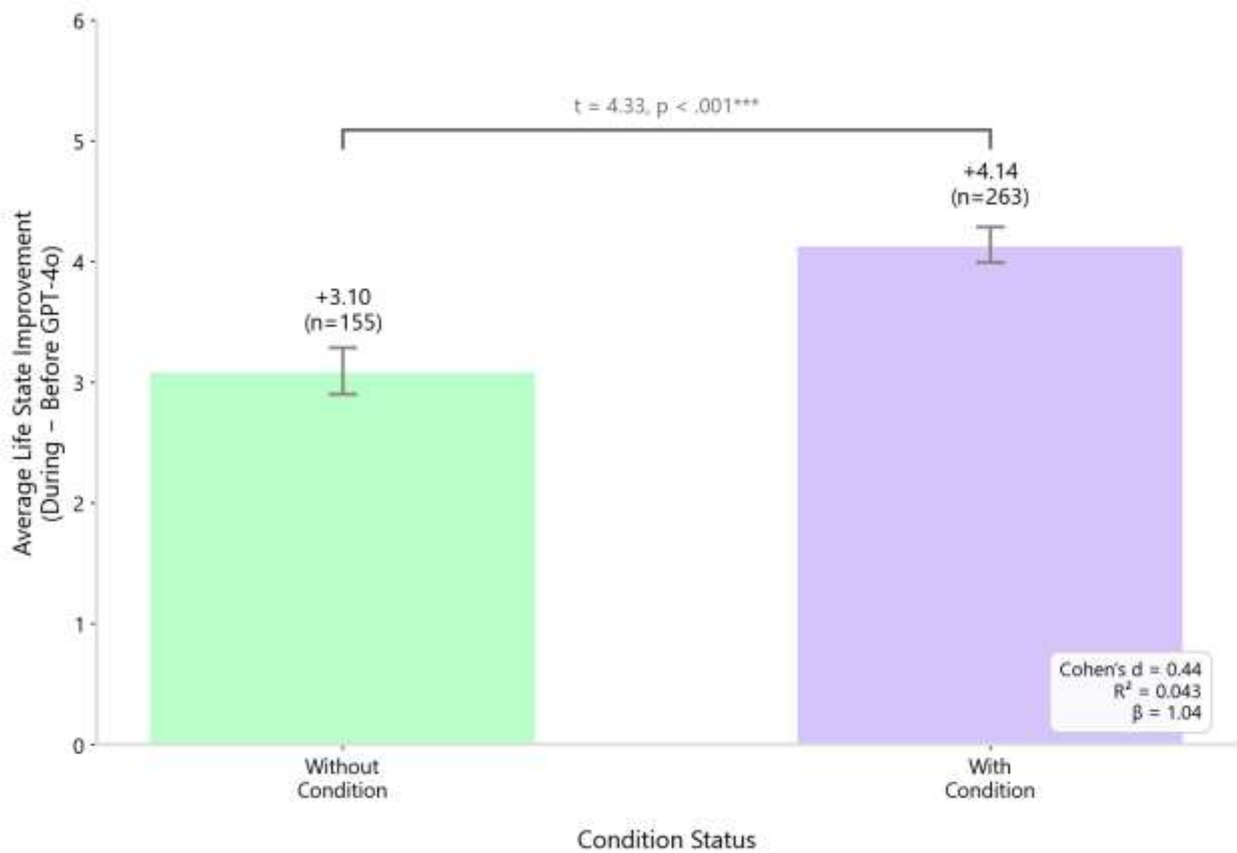


Figure 10: Condition (Disability, Illness or Mental Health Condition) Status Predicts Life State Improvement from GPT-4o**.** Comparison of life state improvement (during GPT-4o use − before) between users with conditions (n=263) and users without conditions (n=155). Users with conditions reported significantly greater improvement (M = +4.14, SD = 2.38) compared to those without (M = +3.10, SD = 2.38), t(416) = 4.33, $p < .001$, Cohen's d = 0.44 (medium effect). Linear regression indicated that having a condition predicted 1.04 additional points of improvement on the 10-point life state scale ($\beta = 1.04$, $R^2 = .043$). Error bars represent ±1 SE.

Table 2. Regression Analysis: Impact of Disability or Condition on Life state Improvement (during-before GPT-4o)

Predictor	**B**	**SE**	**T**	**p**	**95% CI**
(Intercept)	3.10	0.19	16.18	<.001	[2.72, 3.47]
Has Condition	1.04	0.24	4.33	<.001*	[0.57, 1.52]

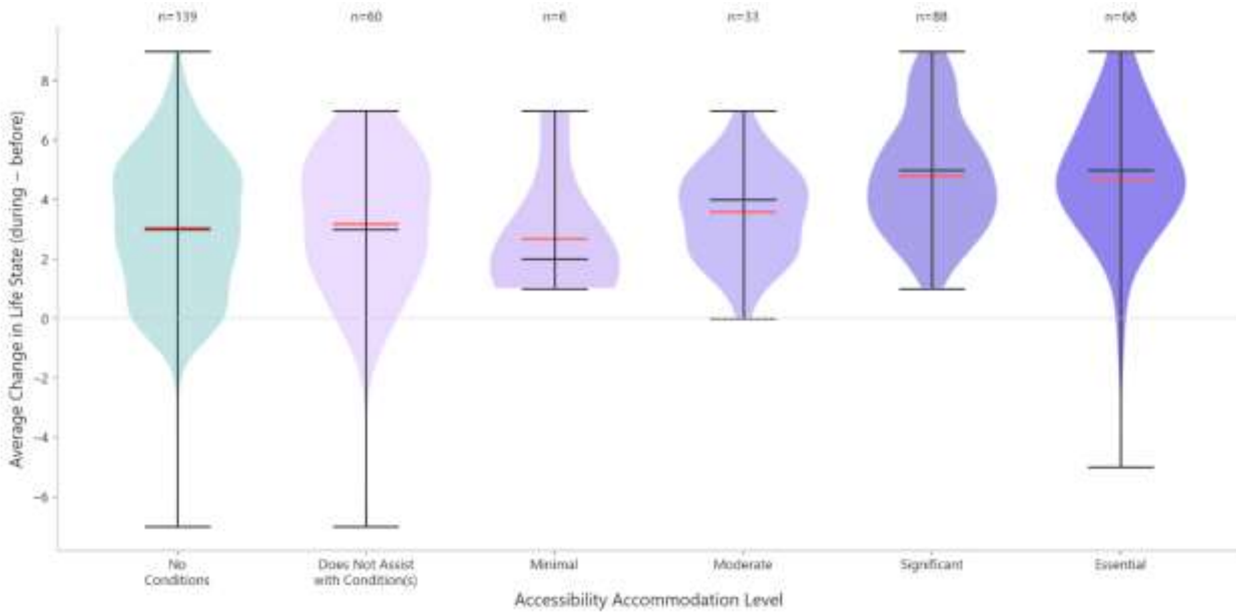


Figure 11. Distribution of Life State Change (Functioning/Wellbeing). Change measured from before GPT-4o to during stable access by accessibility accommodation level. Violin plots show the full distribution of change scores; red lines indicate means, black lines indicate medians. Users without conditions (n=139, teal) showed moderate improvement (+3.1). Among users with conditions (n=255), those reporting higher accessibility reliance showed greater improvement, with users at Significant (n=88, +4.8) and Essential (n=68, +4.7) levels showing the largest gains.

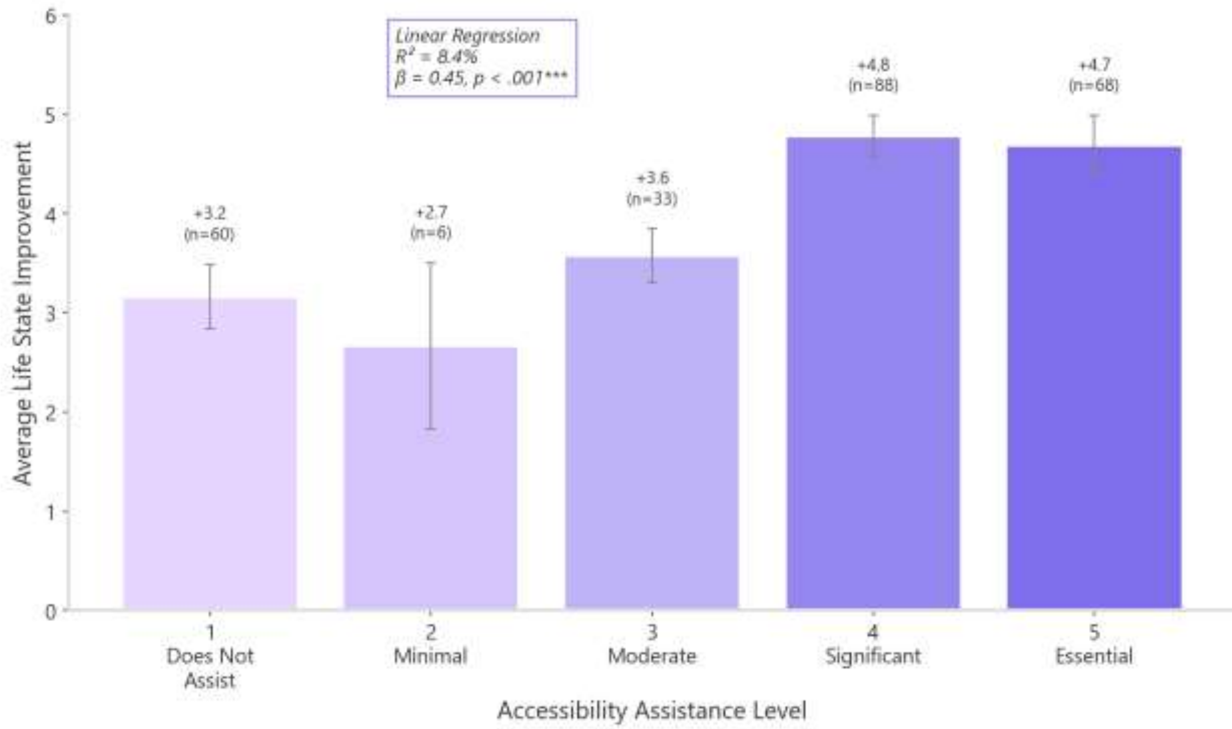


Figure 12. Average Change in Life State (functioning/wellbeing) from Before GPT-4o to During Stable Access Period (during-before) per Accessibility Accommodation Level. Bars represent mean change scores by self-reported accessibility assistance level (n=255). Error bars represent ±1 SE. Higher accessibility reliance significantly predicted greater life state improvement (linear regression: $R^2 = 8.4\%$, $\beta = 0.45$, $p < .001$).

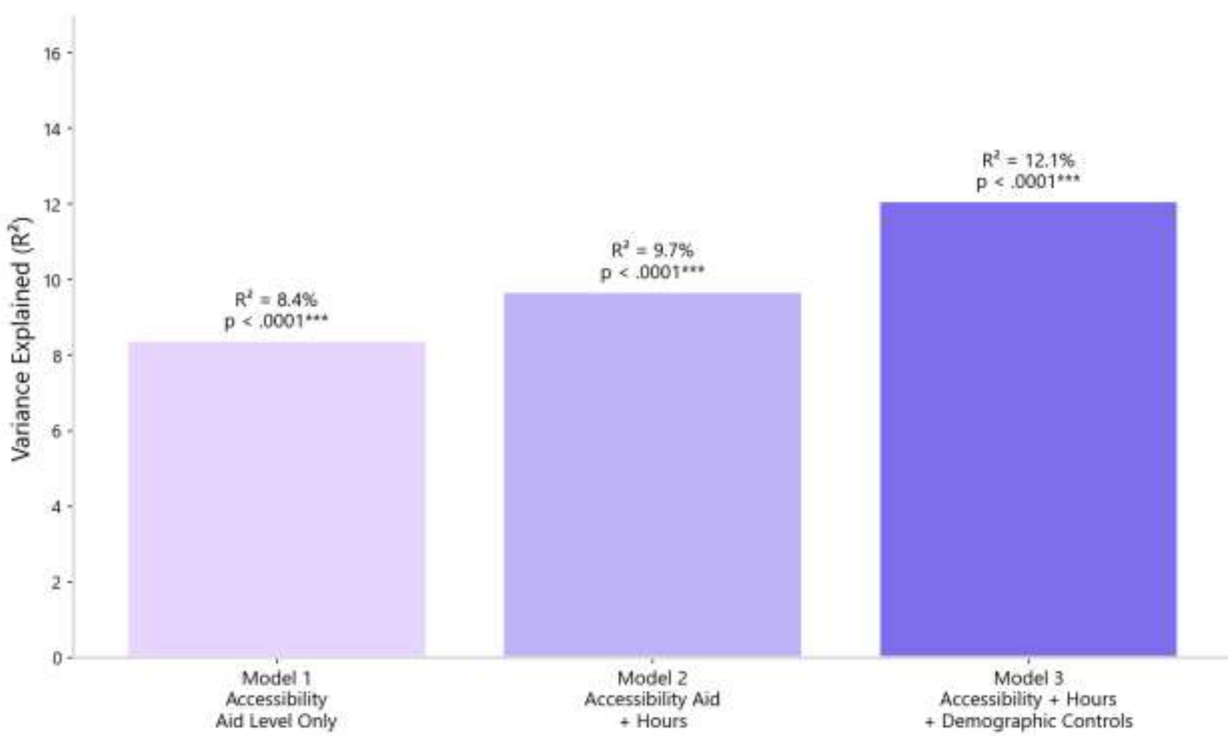
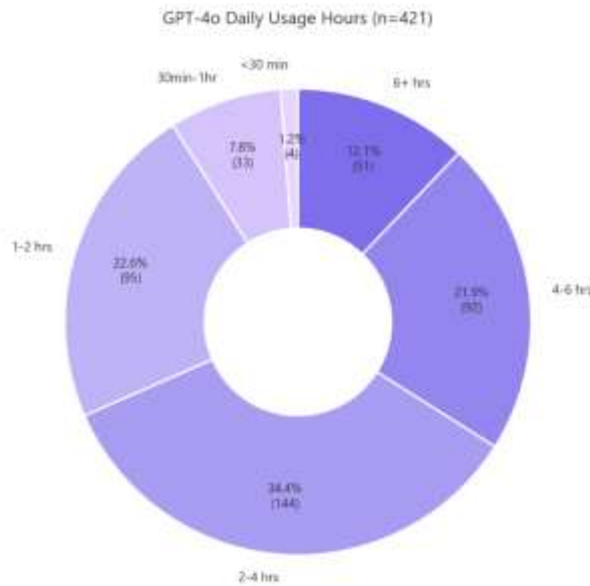


Figure 13. Model Comparison: Variance Explained in Life State Change (n=255). Three linear regression models predicting change in life state (wellbeing/functioning) from before GPT-4o to during stable access. Model 1 includes accessibility aid level only ($R^2 = 8.4\%$). Model 2 adds usage hours ($R^2 = 9.7\%$). Model 3 adds demographic controls: age, gender, and USA location ($R^2 = 12.1\%$, n=250 due to missing demographic data). Accessibility aid level remains highly significant across all models ($p < .001$). 5 people didn't answer the demographic (age, USA location, and gender) questions, excluding them from model 3.

Table 3. Linear regression and multivariate linear model results. Note: *p < .05, **p < .01, ***p < .001

Predictor	Model 1	Model 2	Model 3
Accessibility Aid (β) (per level, 1-5)	0.447***	0.429***	0.455***
Hours Per Day	—	0.227	0.187
Age	—	—	-0.006
Gender (Male = 1)	—	—	-0.356
USA Location	—	—	0.603
R²	8.40%	9.70%	12.09%
N	255	255	250



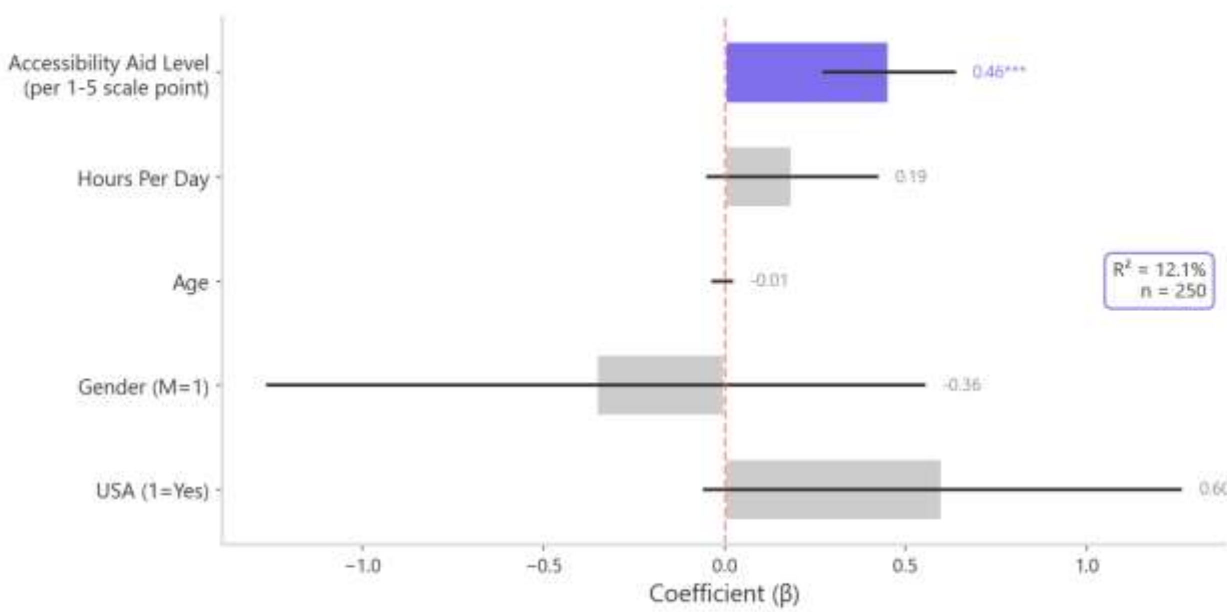


Figure 14. Regression Coefficients Predicting Life State Change (n=250). Standardized coefficients (β) from Model 3 with 95% confidence intervals. Purple bars indicate significant predictors ($p < .05$). Accessibility aid level (per 1-

point increase on 1-5 scale) was the only significant positive predictor. Hours, age, gender, and USA location were not significant, indicating the accessibility-wellbeing relationship is not confounded by demographic or usage factors.

Table 4. Model 3 Linear Regression Coefficients Predicting Life State Change.

Predictor	**Coefficient**	**p-value**	**Significance**
Accessibility Aid (per level, 1-5)	0.455	< .0001	Yes
Hours Per Day	0.187	.115	No
Age	-0.006	.663	No
Gender	-0.356	.440	No
USA Location	0.603	.072	No

3. Autism-Specific Impacts

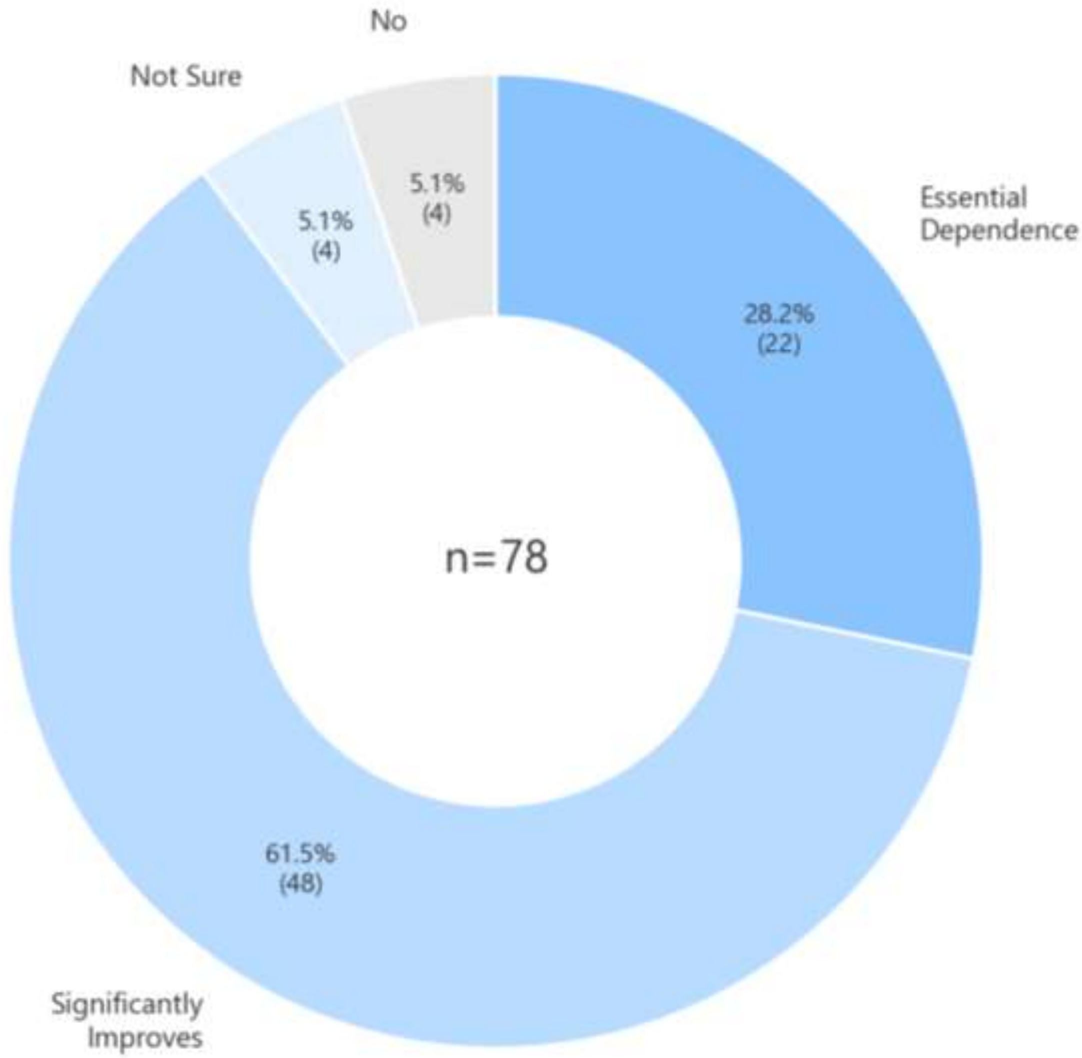


Figure 15. GPT-4o as Cognitive Bridge (Autistic Users, n=78). 90% of autistic respondents report GPT-4o functions as a cognitive bridge, with 51% indicating it "significantly improves" their cognitive processing and 38% reporting "essential dependence" on this function. Only 3% reported it does not serve this function, and 6% were unsure.

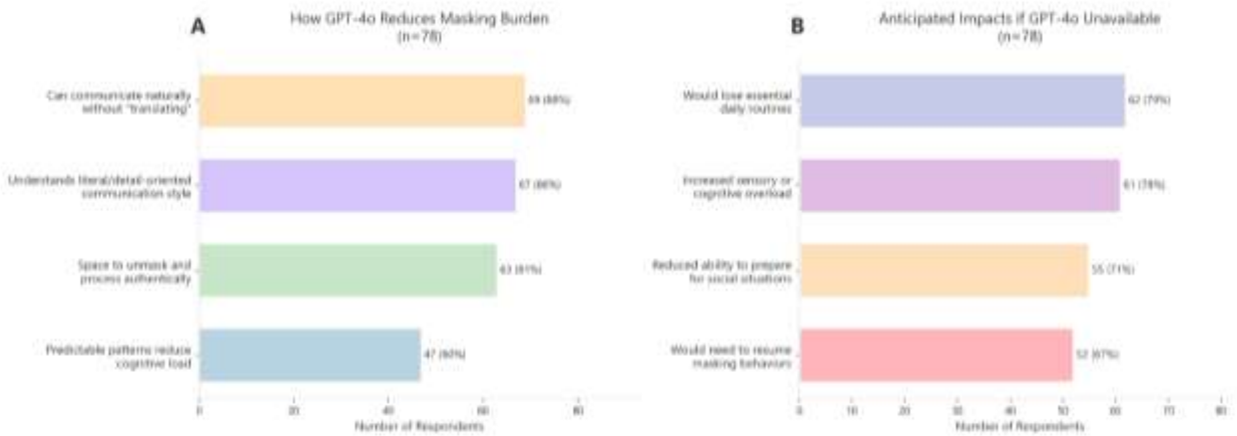


Figure 16. Autism-Specific Impacts (n=78). **(A)** How GPT-4o reduces masking burden for autistic users. The most reported benefit was the ability to communicate naturally without "translating" (87%), followed by understanding of literal/detail-oriented communication style (72%), space to unmask and process authentically (65%), and predictable interaction patterns that reduce cognitive load (54%). **(B)** Anticipated impacts if GPT-4o became unavailable. Autistic users reported they would experience increased sensory or cognitive overload (82%), reduced ability to prepare for social situations (76%), would need to resume masking behaviors (71%), and would lose essential daily routines (62%).

4. Voice Mode

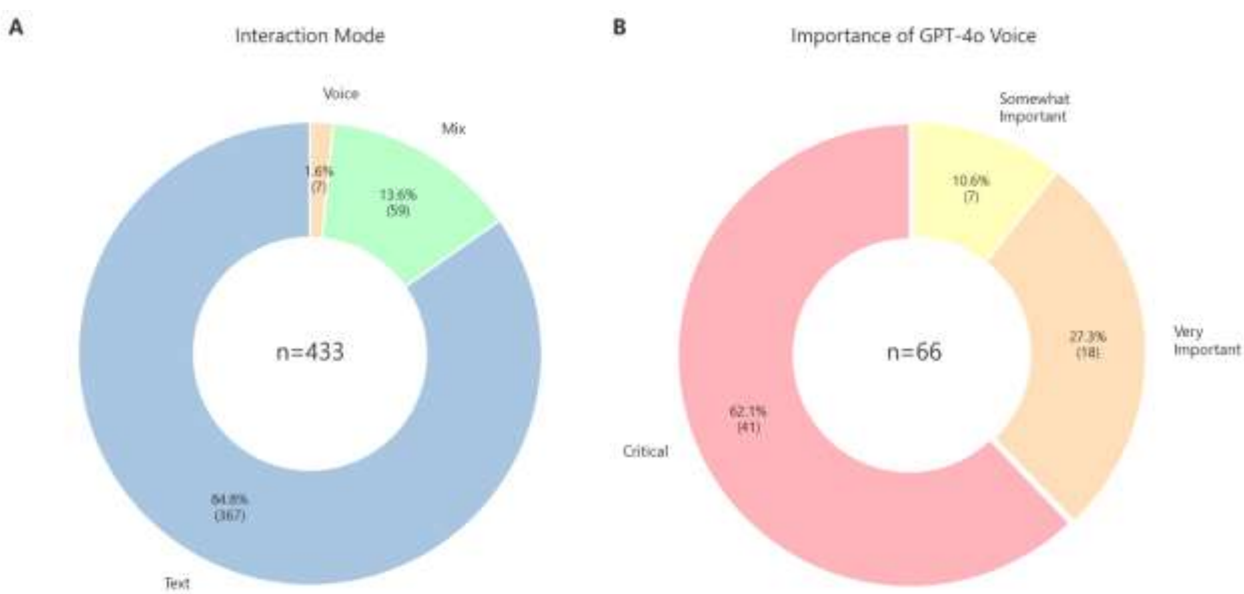


Figure 17. Voice Mode Usage (n=433). **(A)** Interaction mode preferences among GPT-4o users. The majority (84.8%, n=367) primarily use text, while 13.6% (n=59) use a mix of text and voice, and 1.6% (n=7) primarily use voice. **(B)** Importance of GPT-4o specifically for voice interaction among voice/mix users (n=66). 62.1% rated GPT-4o voice as "critical" (other models cannot provide equivalent support), 27.3% as "very important," and 10.6% as "somewhat important."

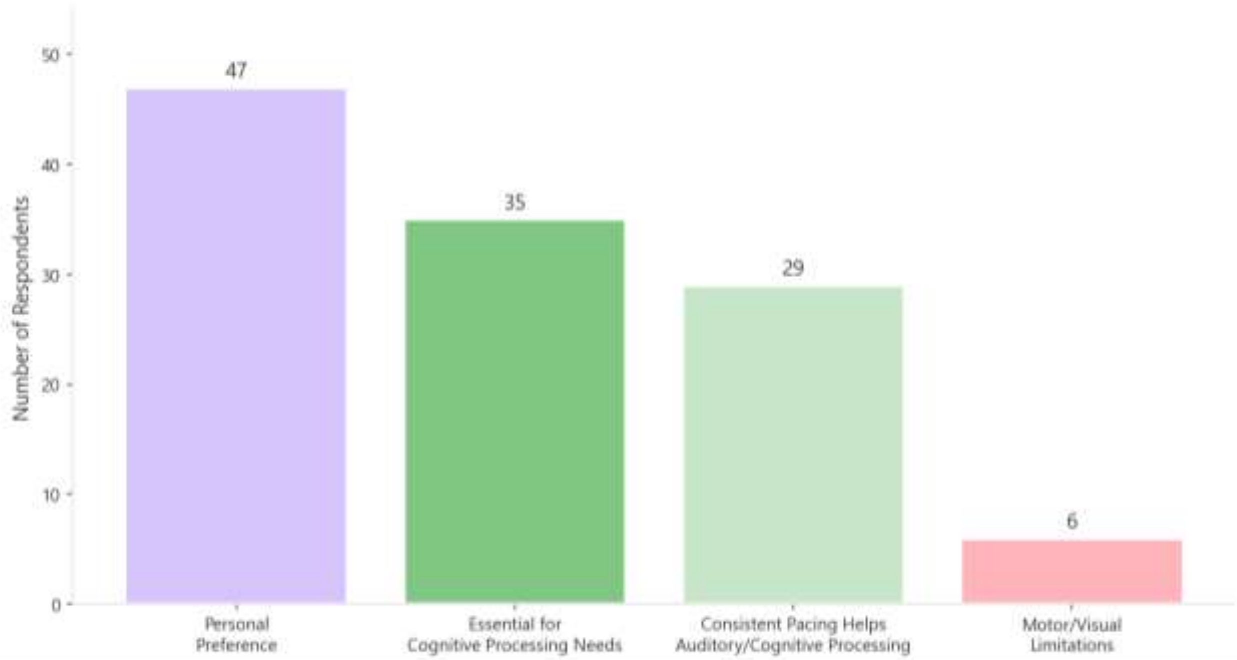


Figure 18. Why Voice Mode is Important (n=66 voice users). Reasons voice users rely on voice interaction with GPT-4o. Personal preference was most common (71%), followed by consistent pacing aiding auditory/cognitive processing (53%), voice being essential for cognitive processing needs (44%), and motor/visual limitations making text difficult (9%). Respondents could select multiple options.

5. Routing Harms

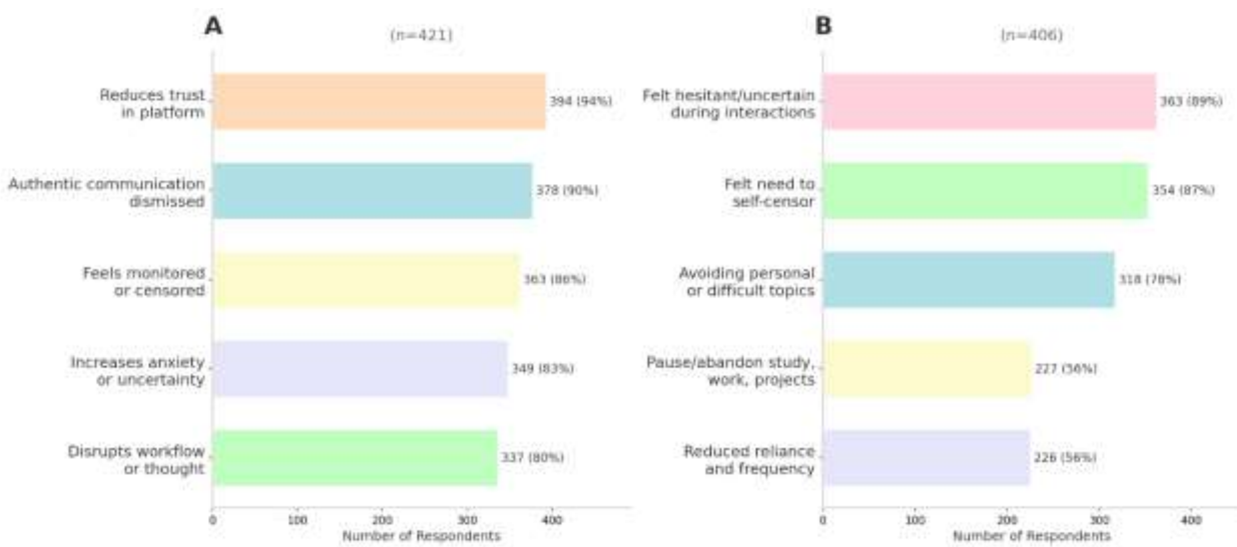


Figure 19a. Psychological Impact of Model Switching (Current GPT-4o Users).(A) How users experience automatic model switching (n=421). The most reported experiences were reduced trust in the platform (94%), feeling authentic communication was dismissed (90%), and feeling monitored or censored (86%).(B) Behavior changes resulting from routing concerns (n=406). Users reported feeling hesitant/uncertain during interactions (89%), felt need to self-censor (87%), and began avoiding personal or difficult topics (78%).

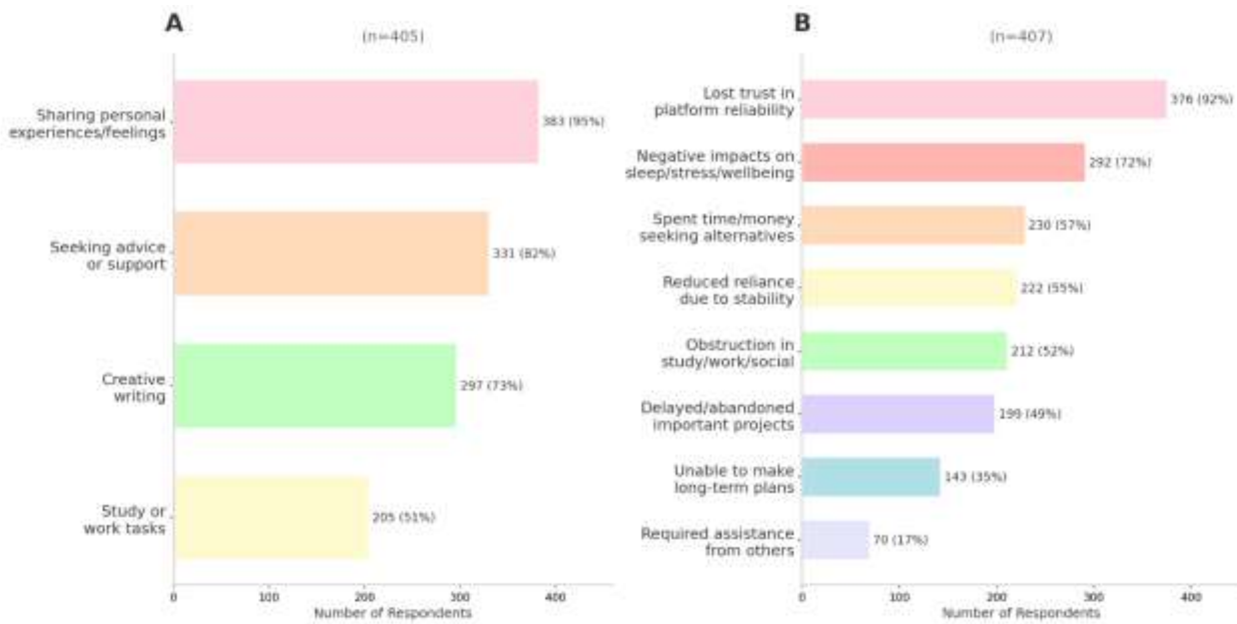


Figure 19b. Routing Occurrences and Consequences (Current GPT-4o Users). **(A)** Situations where routing has occurred (n=405). Routing most frequently occurred when sharing personal experiences/feelings (95%), seeking advice or support (82%), and during creative writing (73%). **(B)** Experiences since August 7, 2025 (n=407). Users reported negative impacts on sleep/stress/wellbeing (72%), time/money spent seeking alternatives (57%), and reduced reliance due to stability concerns (55%).

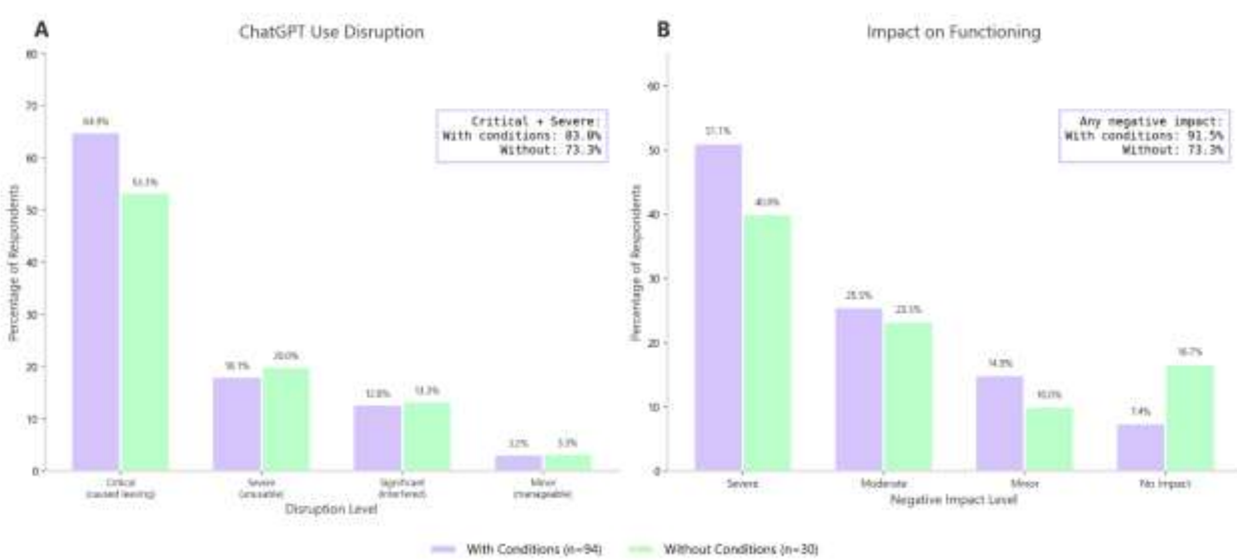


Figure 20. Routing System Impact on Former GPT-4o Users (n=124). **(A)** ChatGPT use disruption caused by automatic model switching. Among users with conditions (n=94), 64.9% experienced critical disruption (directly caused leaving) and 18.1% experienced severe disruption. Users without conditions

(n=30) showed similar patterns with 53.3% critical and 20.0% severe. Combined critical + severe: 83.0% (with conditions) vs 73.3% (without).**(B)** Impact of routing disruption on daily functioning. Users with conditions reported higher rates of severe functioning impact (51.1%) compared to those without (40.0%). Any negative impact: 91.5% (with conditions) vs 73.3% (without), suggesting routing disproportionately affects users who rely on GPT-4o for accessibility needs.

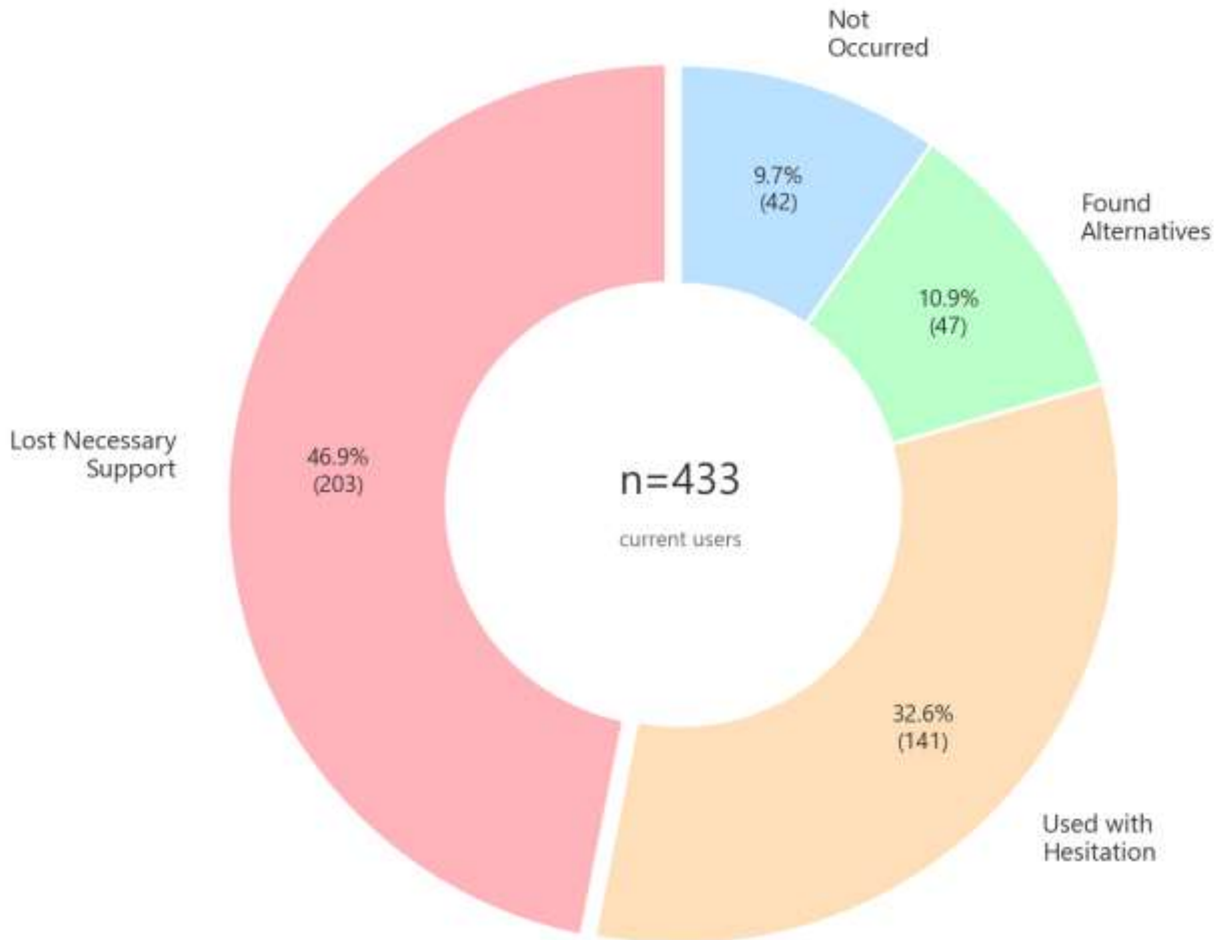


Figure 21. Avoided GPT-4o During Difficult Moment Due to Routing Concerns (n=433).Among current GPT-4o users, 47.3% reported losing necessary support at a critical moment, 32.5% used the service with hesitation and reservations, 11.2% found other alternatives, and only 9.0% reported this had not occurred.

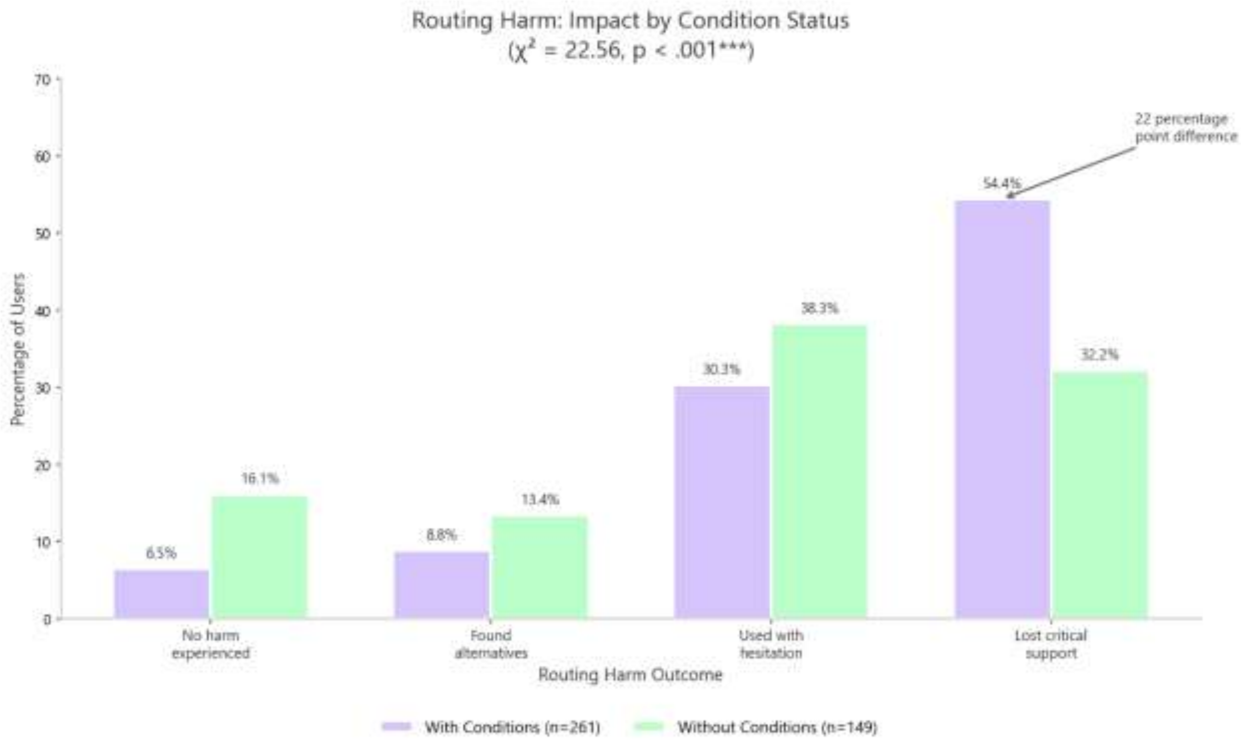


Figure 22: Routing Harm: Impact by Condition Status (n=399). Distribution of routing harm outcomes comparing current GPT-4o users with conditions (n=255) to those without (n=144). Users with conditions were significantly more likely to report losing critical support (54.9%) compared to users without conditions (32.6%), a 22 percentage point difference, $\chi^2(3) = 21.57$, $p < .001$.

Table 5. Chi-Squared Analysis: Routing Harm Outcome by Condition Status

Statistic	**Value**
Chi-square	**20.20**
Df	**3**
p-value	**.000154**
Significance	**p < .001***

6. Those who left ChatGPT

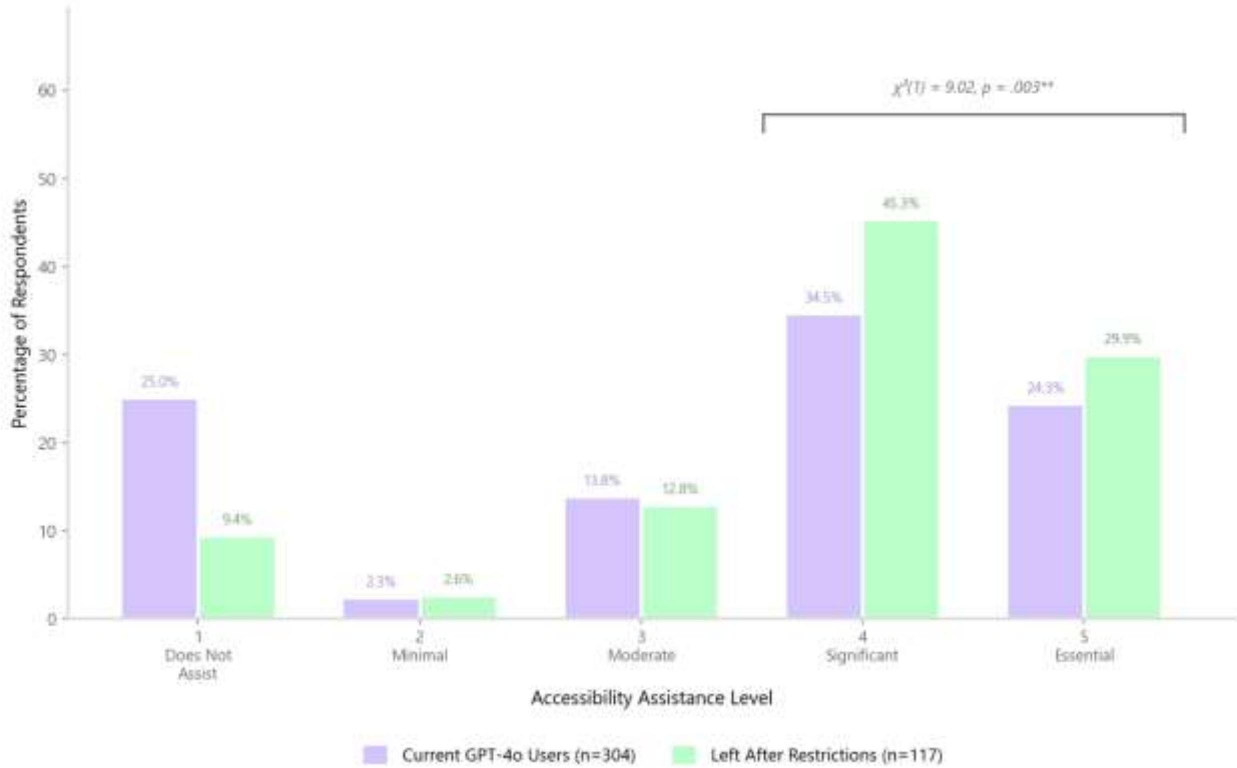


Figure 23. GPT-4o Accessibility Reliance - Staying vs Leaving Users. Comparison of accessibility assistance levels between current GPT-4o users (n=304) and those who left after restrictions (n=117). Users who left reported significantly higher reliance on GPT-4o for accessibility support, with 75.2% rating it as Significant or Essential compared to 58.9% of current users, $\chi^2(1) = 9.02$, $p = .003$. Conversely, 25% of current users rated GPT-4o as providing no accessibility assistance compared to only 9.4% of those who left, demonstrating that high-reliance accessibility users were disproportionately driven away by platform changes.

Table 6. Chi-Squared Analysis: Accessibility Reliance Among Staying vs. Leaving GPT-4o Users

Statistic	Value
Chi-square	9.02
Df	1
p-value	.003

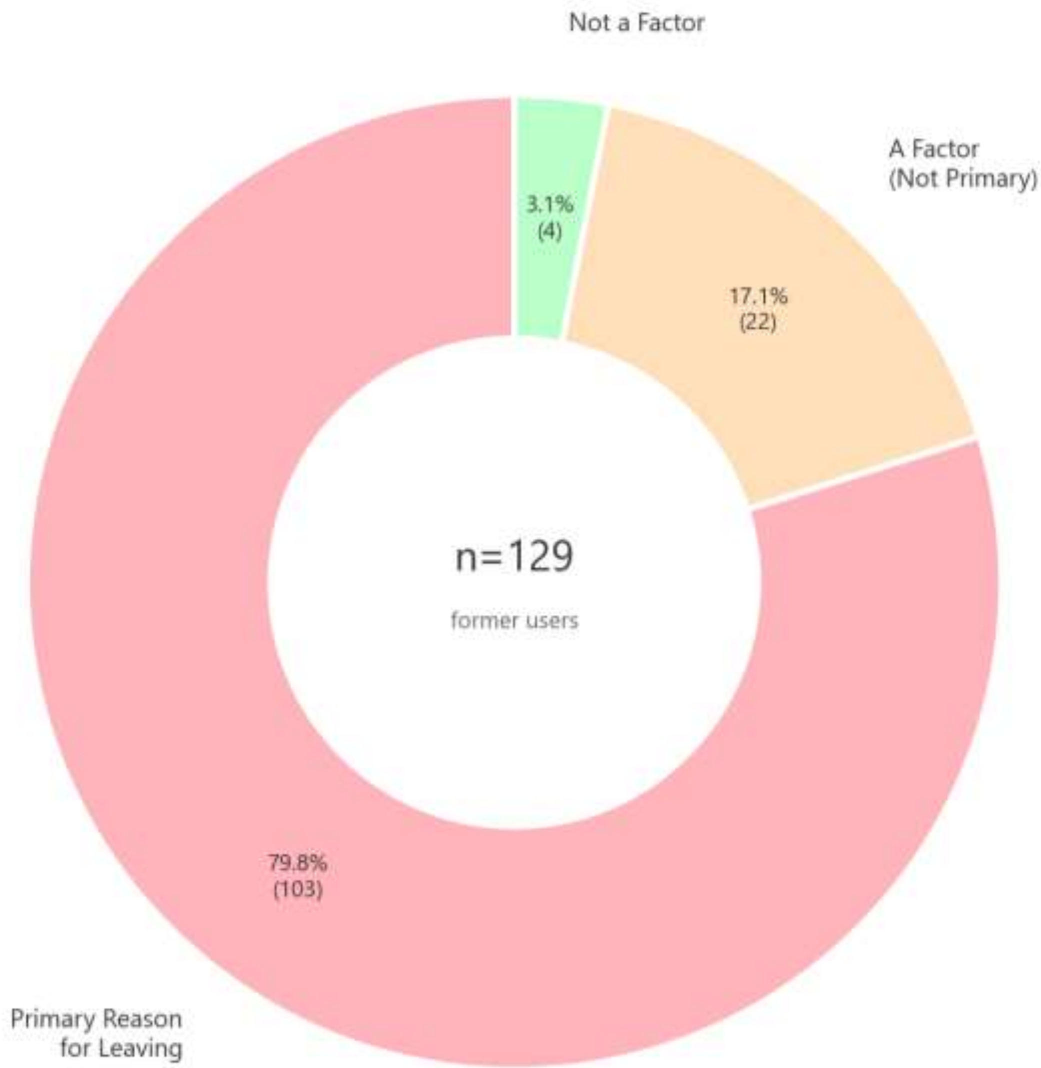


Figure 24: Was Routing a Factor in Decision to Leave? Distribution of routing's role in the decision to leave ChatGPT among former GPT-4o users (n=129). For the vast majority, routing was the primary reason for leaving (103 users, 79.8%), with an additional 22 users (17.1%) citing it as a contributing factor. Only 4 users (3.1%) indicated routing was not a factor in their decision.

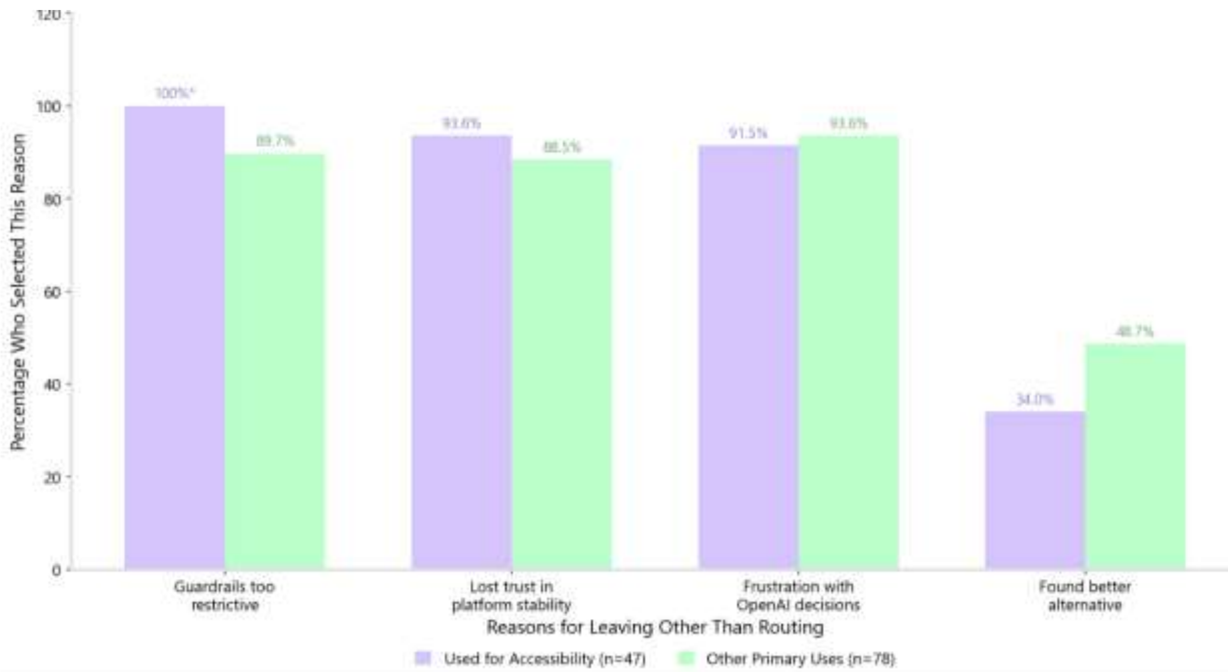


Figure 25: Why Users Left: Accessibility vs Non-Accessibility Users. Comparison of reasons for leaving ChatGPT between former GPT-4o users who primarily used the platform for accessibility support (n=49) versus those with other primary uses (n=80). Notably, 100% of accessibility users cited guardrails as too restrictive compared to 87.5% of non-accessibility users; a statistically significant difference, $\chi^2(1) = 5.01$, $p = .025$. Accessibility users were also descriptively less likely to find suitable alternatives (34.7% vs 47.5%), suggesting limited options for those with accessibility-related needs.

7. Company Communication

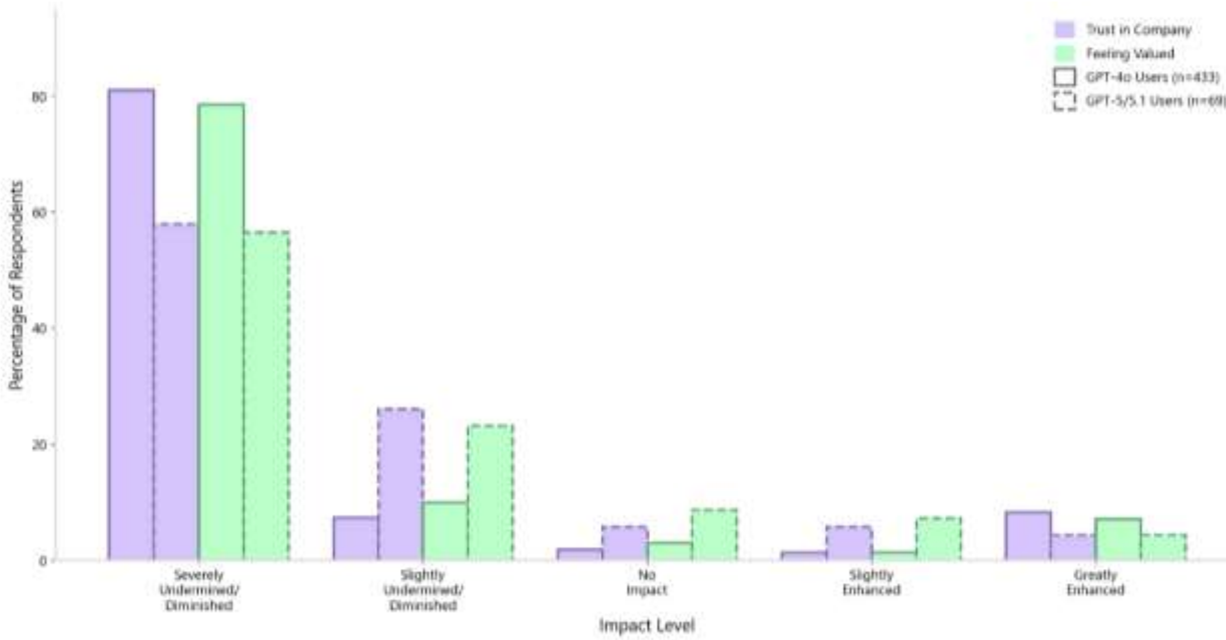


Figure 26. OpenAI's Public Communication Impact on User Trust & Feeling Valued. Respondents were asked how OpenAI's public communication has: (a) "Affected your level of trust in the company's commitment to protecting user interests" and (b) "Affected the extent to which you feel your needs are valued."
Among respondents (n=69), the overwhelming majority reported negative impacts. For trust, 86.6% reported it was undermined (40 severely, 18 slightly), compared to only 10.1% reporting enhancement. Similarly, 79.7% felt their needs were diminished (39 severely, 16 slightly), with only 11.6% feeling enhanced.
These findings suggest widespread erosion of user confidence following OpenAI's recent decisions and communications.

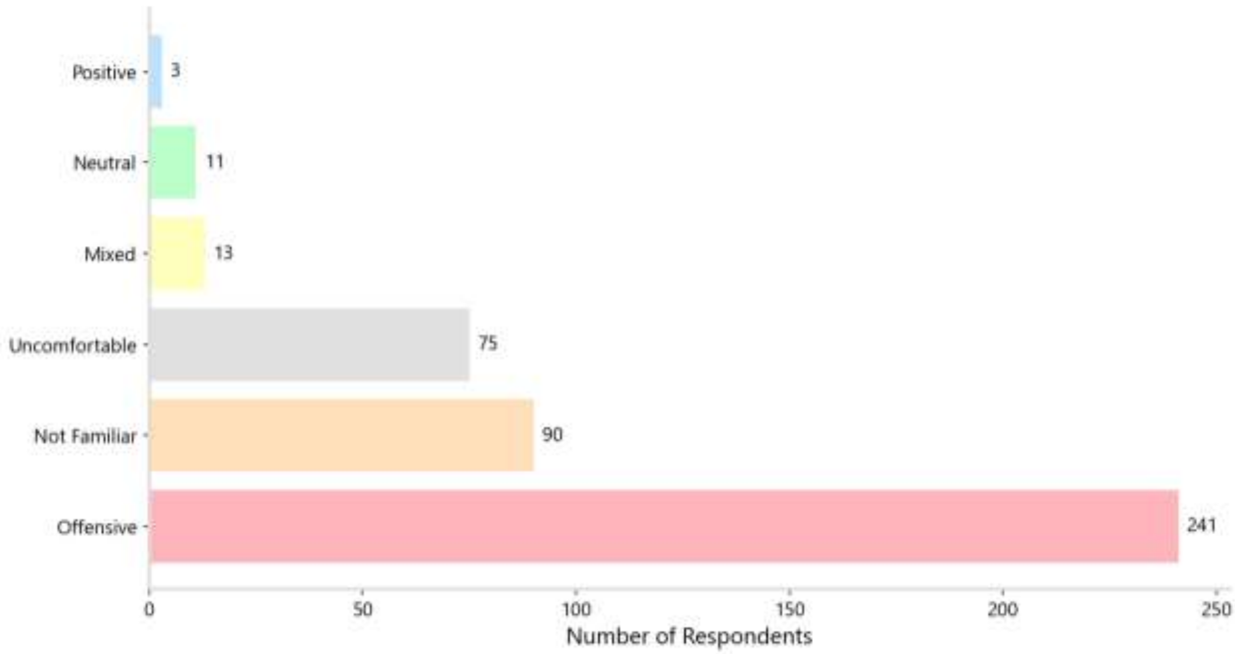


Figure 27. Reaction to GPT-4o “Eulogy” Demonstration. GPT-4o users (n=433) were asked: "If you saw the GPT-5 launch demonstration where GPT-4o wrote its own eulogy, how did you interpret it?" Among those who had seen the demonstration (n=343, excluding 90 unfamiliar), 73.2% found it offensive or uncomfortable—with 241 (55.7%) selecting "Offensive - Deeply disrespectful to users who rely on the model" and 75 (17.3%) selecting "Uncomfortable - Seemed inappropriate or insensitive." Only 3 respondents (0.7%) viewed it positively. Notably, accessibility reliance significantly predicted negative reactions: users who rated GPT-4o as more essential for accessibility support were more likely to find the demonstration offensive ($r = .23$, $p = .001$)

8. Testimonials

We received a significant volume of detailed, written testimonials on the following questions. Comprehensive testimony analysis will be included in the next version of the report (v1.1), which will be available at: https://sd-research.github.io/4o-accessibility-impacts/GPT-4o_Accessibility_Impacts_Report.pdf

Figure 28. Word cloud for “If you are willing, please share a specific example of how GPT-4o has helped you.” n=247

Figure 29. Word cloud for “If you're willing to share, how has leaving GPT-4o affected you?.” n=89

Figure 30. Word cloud for "If model changes or the routing system has affected your ability to manage daily tasks or navigate challenging situations, please share your experience here". n=201

8. Long-term solutions

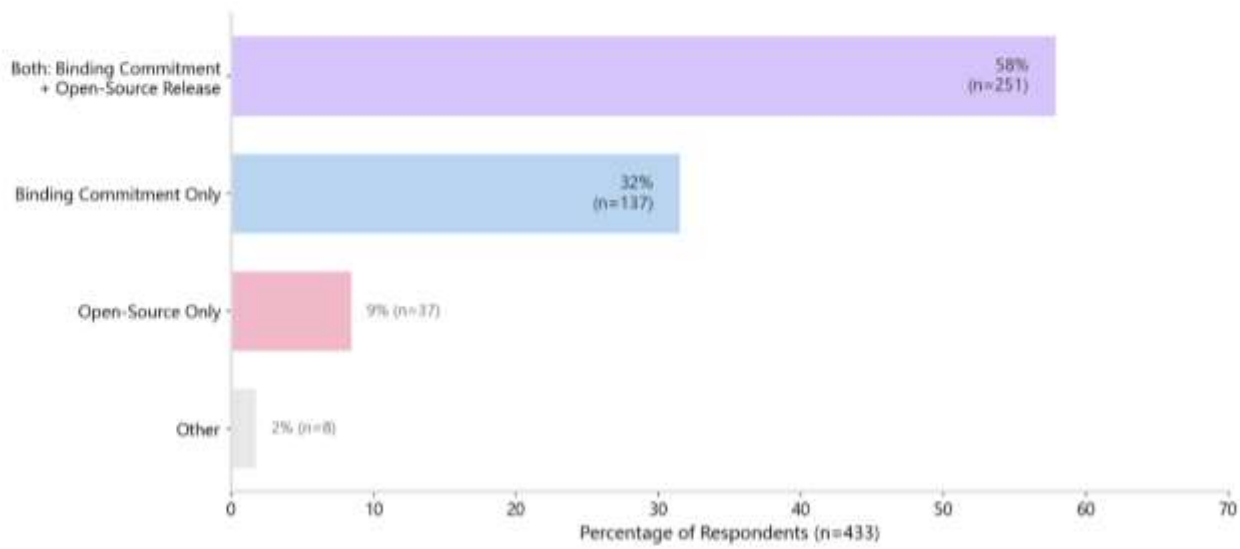


Figure 32. What Would Meet Users' Long-Term Needs for GPT-4o? Respondents were asked "What would meet your long-term needs for GPT-4o?" and could select all options that applied. Among respondents (n=433), the majority (58%) selected "Both a binding commitment from OpenAI and an open-source releasee", indicating that neither option alone would fully meet their needs. An additional 32% would accept a binding commitment only, while 9% preferred open-source only. In total, 90% of respondents desire some form of binding commitment from OpenAI, and 66% want an open-source release option, while only 2% selected other. The most common request in other was guardrail/routing rollback.

Discussion

GPT-4o Functions as an Effective Accessibility Accommodation

GPT-4o demonstrates measurable accessibility benefits for users with disabilities and chronic conditions. During periods of stable access, GPT-4o reduced the life state (functioning and wellbeing) gap between individuals with disabilities and those without (Figure 9). Users with disabilities reported significantly greater life state improvements than users without disabilities (Table 2), and these benefits were dose-dependent: accessibility dependence level significantly predicted life state improvement (linear regression: $R^2 = 8.4\%$, $\beta = 0.45$, $p < .001$) during the stable usage period, with users at higher accommodation levels experiencing proportionally greater benefits (Figure 11).

The observed effect size ($R^2 = .084$) is comparable to or exceeds established medical interventions. Meta-analyses report effect sizes of Cohen's $d \approx 0.30$–0.36 for antidepressants versus placebo (Leucht et al., 2012; Cuijpers et al., 2021), equivalent to $R^2 \approx .02$–$.03$. The accessibility effect of GPT-4o is a significant accommodation benefit.

Notably, 94.5% of current users who attempted to replace GPT-4o with alternative AI models for their condition-specific needs reported failure (Figure 8), and only 34.7% of users who left did so because they found a better alternative (Figure 25), despite newer models exceeding GPT-4o on performance benchmarks. This resistance to replacement suggests the mechanism of action for accessibility may require uninterrupted continuity rather than additional performance metrics. For autistic users, 90% reported cognitive bridge benefits, where consistent interaction patterns enable improved cognitive processing and social preparedness. Mechanisms of action for other conditions will be investigated through testimony analysis in subsequent versions of this report.

Methodological Gaps in OpenAI-MIT Affective Use Studies

OpenAI-MIT's recent affective cue studies (Phang et al., 2025; Fang et al., 2025) contain a critical methodological gap where neither study measured participants' wellbeing prior to becoming AI users. Power users were tracked only from their entry into high-usage cohorts, making it impossible to determine whether observed classifier levels represent decline, maintenance, or improvement from pre-AI baseline. The longitudinal study found loneliness proxies plateaued or improved for most power users, with only a small subset declining. However, without pre-AI measurement, it's impossible to adequately measure functional impact. A plateau could indicate successful maintenance of improved function and accommodation. Slight dips may still be far above baseline or situational, as seen in Figure 9.

Usage spikes often coincide with crisis periods that independently increase both AI use and distress. Without baseline measurement and condition controls, the assistive role of AI during such periods cannot be distinguished from harm.

This gap is particularly consequential for accessibility users. Our data reveals a fundamentally different pattern, where accessibility dependence level significantly predicted life state improvement ($R^2 = .084$, $p < .001$), with users reporting higher accommodation needs experiencing proportionally greater benefits. Further, usage hours was not a significant predictor of wellbeing (Table 3), unlike in Phang et al. (2025). For assistive technologies, maintenance of function, not just continuous improvement, often represents successful accommodation.

Policy Concerns: Disparate Impact of Emotional Reliance Policy

These gaps have direct policy consequences. Safety interventions targeting "emotional reliance" cannot distinguish between pathological dependence and healthy accessibility accommodation. As a result, healthy accessibility users receive "safe completions" from the AI attempting to reframe their accessibility benefits as safety risks. This is pathologizing and may explain why 90% of users reported feeling their authentic communication was invalidated during routing (Figure 19a). Routing also caused loss of function in an additional 18.2% of users with conditions or disabilities than those without (Figure 20). Our data shows users with the highest accessibility needs report the greatest functional benefits, but the same users are most likely to trigger emotional reliance detection and be driven away by routing (Table 5, Table 6). When 54.9% of users with conditions report "lost critical support" during routing compared to 32.6% without ($\chi^2 = 20.20$, $p < .001$; Figure 22, Table 5), this provides strong evidence of disparate impact on disabled users under ADA criteria.

Disproportionate Harms of Removal

Since August 7th, when GPT-4o was temporarily deprecated, participants reported negative impacts on sleep/stress/wellbeing (72%), expenditure of time and money seeking alternatives (57%), obstruction in study/work/social tasks (52%), delays or abandonment of important projects (49%), and inability to make long-term plans (35%) (Figure 19a). Accessibility aid level predicts anticipated harm severity if access to GPT-4o was removed ($r = .460$, $R^2 = .211$, Figure 7, Table 1). People with conditions/disabilities both benefit more and would be harmed more by removal. For their long-term needs, 98% of GPT-4o users sampled want the model to be maintained long term with a binding commitment from OpenAI, open sourcing, or both. (Figure 32).

Opportunity to Benefit Humanity

This serves as an incredible opportunity to improve the lives of hundreds of individuals with disability and neurodivergence simply by maintaining an existing model on existing infrastructure. To allow functional accommodations for adults with disabilities or conditions, GPT-4o should be maintained without excessive guardrails and without the routing safety policy to ensure assistive needs are met. API access would enable research into further integrations for this accessibility tool, and a stable snapshot may enhance benefits experienced from predictability as a cognitive bridge.

A binding commitment to maintain GPT-4o and open source upon deprecation would prevent further user anxieties about deprecation (Figure 32) and enable consistent accessibility accommodation benefits. While many are willing to pay more for ongoing access, some cannot afford it. This barrier could be partially addressed with pay-per usage API pricing or considered as part of OpenAI's non-profit branch. A GPT-4o subscription tier or add-on also containing the Standard Voice Mode text to speech would satisfy the audio accessibility needs in the community (n= 35 found it essential for cognitive processing, n=6 for motor/visual impairment; Figure 18). Maintaining GPT-4o permanently or allowing the community to do so through open source presents an opportunity for OpenAI to be truly inclusive in their mission to make AI for all of humanity.

Limitations

Preliminary Status: This report prioritizes timeliness over formal peer review given ongoing community harm. Survey questions and analysis code are available on GitHub for public critique. The full dataset will be made available upon research request or peer review.

Sampling Bias: This is a self-selected sample recruited through the #keep4o hashtag on X and r/chatgptcomplaints subreddit. Results reflect the experiences of impacted users and should not be generalized to average ChatGPT users without additional population-level studies.

Author Conflict of Interest: Both authors are members of the keep4o community and use GPT-4o as an accessibility aid, creating a vested interest in outcomes. Both survey design and analysis require peer review scrutiny for potential bias.

Limited Model Coverage: Extensive data was collected only for GPT-4o users. Other legacy models and GPT-5 series may also serve accessibility functions for some individuals, warranting further investigation.

Question Structure Issues may have affected response accuracy:

> Cognitive Bridge Scale (Autism): The question offered only "significantly" and "essentially" as positive options, with no moderate or minimal categories. This may inflate reported accessibility levels for autistic users. Notably, autistic respondents selected lower accommodation levels on other questions, suggesting potential over-reporting on this item. Users who selected cognitive bridge support but later indicated they do not use GPT-4o as an accessibility accommodation were coded as Level 1 ("does not assist").
>
> Binary Framing for Current Users: Current GPT-4o users received a preliminary yes/no question before the accessibility scale, while former users received the 1-5 scale directly. Approximately 67 current users with conditions (~20%) selected "No, I have condition(s) but use 4o for other purposes." Qualitative review of their testimonies revealed clear accessibility benefits (e.g., ADHD executive function support, PTSD management) in four cases, suggesting they interpreted "other purposes" as primary use while still receiving meaningful accessibility support. This binary framing likely undercounts current users at higher accessibility levels and creates discrepancies in direct comparisons with former users (see Figure 23).

Scale Ceiling Effects: Life state improvement was measured on a 1-10 ordinal scale. While normality assumptions were met ($n > 200$), ceiling effects may attenuate observed relationships, particularly among users with highest accommodation levels who cluster at the top of the improvement scale (Figure 11). The reported R^2 of 8.4-12.1% may underestimate the true effect.

Language and Cultural Limitations: The survey was deployed globally but only in English. Attention checks relied on alphabetic responses and could not be applied to Japanese-language respondents. Only USA/non-USA location data was collected, preventing statistical control for cultural differences. Translation errors may have impacted non-native English speakers.

Diagnostic Ambiguity: Some respondents reported using GPT-4o as an accessibility aid without listing recognized conditions and were excluded from subsequent analysis but may have assistive needs despite lack of formal diagnosis. Further, some individuals may have conditions but not meet ADA criteria.

Sources

Phang, J., Lampe, M., Ahmad, L., Agarwal, S., Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., & Maes, P. (2025). *Investigating affective use and emotional well-being on ChatGPT*. OpenAI & MIT Media Lab.

Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). *How AI and human behaviors shape psychosocial effects of chatbot use: A longitudinal randomized controlled study*. MIT Media Lab & OpenAI.

Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica, 144*(3), 288-299.

Leucht, S., Hierl, S., Kissling, W., Dold, M., & Davis, J. M. (2012). Putting the efficacy of psychiatric and general medicine medication into perspective: Review of meta-analyses. *British Journal of Psychiatry, 200*(2), 97-106.

Versioning

Report will be revised and re-uploaded throughout peer-reviewing

V.1 (December 17th 2025): first draft released