

Before the Threshold: Search Behavior as an Early Signal of Emotional Distress

A Research Initiative Overview

Mark Hoashi, M.S.

The Uplift Index

Contact: mark@upliftindex.org | upliftindex.org

Abstract

Prevailing instruments for measuring population-level emotional distress are structurally dependent on the moment of disclosure — that is, on the point at which an individual self-identifies as struggling, seeks clinical care, or responds to a survey instrument. This paper argues that a meaningful and empirically tractable layer of emotional experience exists upstream of that threshold, and that anonymized internet search behavior constitutes a viable observational window into it. Drawing on a pilot study conducted in Los Angeles County and a structural analysis of search as a communicative medium, we propose that search query data captures pre-diagnostic emotional expression in ways that conventional survey instruments cannot. The paper presents preliminary findings, examines the methodological properties and limitations of this approach, identifies structural forces that currently distort the search environment for emotionally vulnerable users, and outlines a research agenda for developing search-derived signals into a responsible population-level early-warning system.

1. Introduction: The Gap Before Disclosure

Individuals in emotional distress seldom announce that distress publicly. They may lack the vocabulary to name what they are experiencing; they may sense that something is wrong without having determined whether it rises to the level of a clinically or socially recognized problem. Even when the feeling is clearly perceived, the social costs of disclosure — vulnerability, stigma, the obligation to account for oneself to others — frequently outweigh the anticipated benefits of speaking.

The result is a structural gap: a period of lived emotional experience that precedes the consolidation of meaning into identity, before an individual has categorized their state as

"depressed," "lonely," or "in crisis," and before any institutional system is configured to detect or respond to their condition.

This gap is not a marginal interval. For a substantial portion of the population it extends across months or years, and because existing measurement systems are designed to activate at the point of identifiable, disclosed distress, the emotional experience that occurs within this interval remains largely invisible to public health infrastructure.

The Uplift Index is a research initiative examining whether this gap can be observed — not through clinical intrusion or diagnostic inference, but through passive observation of signals that individuals generate privately and anonymously in the course of expressing concern to an internet search engine.

2. Why Search Behavior Is Different from Survey Data

The dominant instrument for measuring emotional distress at the population level is the survey. Survey-based methods offer considerable advantages: they support longitudinal trend analysis, enable demographic comparison, and generate data that is interpretable within established clinical and epidemiological frameworks.

Surveys are nonetheless subject to a structural constraint that limits their observational reach: they are administered on what sociologist Erving Goffman termed the "front stage" of social interaction (Goffman, 1959). When a respondent is asked to characterize their emotional state — even under conditions of anonymity — they engage in impression management, selecting not simply the most accurate account but the account that appears socially safe, emotionally stable, and consonant with the version of themselves they are willing to present to an implied audience.

Internet search queries, by contrast, occur in a structurally distinct communicative space. They are generated offstage — in the absence of an implied audience, social contract, or pressure toward narrative coherence or emotional composure. An individual who would characterize their wellbeing as satisfactory on a survey instrument may, within the same time period, submit a query such as "why do I feel so empty" to a search engine via a private device — a disclosure that would not surface in any conventional measurement context.

These are not equivalent data sources. They originate from distinct moments in the phenomenological arc of emotional experience and are sensitive to fundamentally different aspects of psychological state.

Survey instruments capture how individuals understand and represent themselves once an emotional state has consolidated into a stable self-conception. Search behavior, by contrast, provides observational access to the moment prior to that consolidation — the interval during which a feeling is affectively present but has not yet become socially expressible or clinically nameable.

Both data types are analytically valuable, but they index different temporal positions in the trajectory of emotional distress. For the purposes of early intervention — identifying and responding to distress before it reaches crisis threshold — the upstream, pre-disclosure signal represents a largely untapped observational resource.

3. Search as a Containment Technology

Search possesses an additional structural property that renders it particularly well-suited to capturing early emotional expression: the medium does not respond to the user.

From a conventional design standpoint, which presupposes that more responsive systems produce better outcomes, this appears to be a deficiency. However, for individuals in early emotional distress who are not yet prepared to engage relationally — who are not yet ready to be seen, advised, reframed, or offered help — the non-relational character of search is not a limitation but a constitutive feature that enables a particular kind of expression.

Conversational AI systems and large language models are, by design, relational: they acknowledge the user, respond to queries, and implicitly constitute a dialogic exchange. Even when such systems are carefully calibrated for empathy and tact, the structural fact of being responded to alters the phenomenological character of the interaction. For individuals who are emotionally exhausted, ashamed, or simply not yet prepared for relational engagement, that interaction constitutes a form of labor.

Search, by contrast, functions as what might be termed a containment technology: it absorbs expression without reacting to it. A concern can be articulated without precipitating a conversation; a feeling can be externalized without being narrativized or subjected to interpretation. There is no follow-up prompt, no therapeutic reframing, and no implication that articulation constitutes commitment to a subsequent course of action.

This structural property offers a plausible account of why individuals in early or ambiguous distress disproportionately turn to search engines rather than to helplines, mental health resources, or conversational AI interfaces: not because the search engine provides superior information, but

because it does not respond, reframe, or require the user to position themselves within a relational context.

As conversational AI systems become increasingly empathetic, proactive, and relationally sophisticated, the comparative value of non-relational systems for this population may paradoxically increase. Where interactional costs are high, the communicative space that requires nothing of the user — that receives expression without producing response — may become correspondingly more valuable.

4. The Search Economy and Its Distorting Effects

Any serious analysis of search behavior as an emotional signal must contend with a foundational structural reality: the search environment is not a neutral observational medium. It is, in its prevailing commercial form, an auction.

Every query submitted to a major commercial search engine enters a marketplace in which the visibility of responses is determined not by therapeutic relevance or epistemic quality, but by the economic return they generate for platform operators. The advertiser capable of extracting maximum value from a given click secures the most prominent placement. Over time, this incentive structure does not merely shape what users encounter in response to their queries — it shapes which categories of human experience receive institutional acknowledgment at all.

For emotionally vulnerable users, this commercial architecture produces a specific and consequential distortion. Query strings associated with early emotional distress — expressions such as "why am I so lonely," "I feel empty," or "why does nobody reach out anymore" — overlap linguistically with clinical and pharmaceutical market categories that carry strong economic incentives to bid aggressively for those keyword positions.

The Uplift Index paid approximately \$8 per click on the query "why am I so lonely" during its Los Angeles pilot campaign — not because early-support resources are competitive, but because that query was priced against pharmaceutical advertisers whose business models could justify higher spend.

The cumulative effect is a form of conditioning. When an individual in early emotional distress repeatedly searches for understanding and is met first by clinical framing, pharmaceutical advertising, or pathologizing language, they are not receiving informationally neutral responses. They are receiving economically optimized responses that may prematurely foreclose the interpretive space available for self-reflection, social explanation, and meaning-making.

This matters because the population expressing these queries is heterogeneous. Not every individual who reports loneliness requires pharmacological intervention; not every experience of disconnection constitutes clinical illness. When the dominant response to emotional uncertainty is pharmaceutical, however, individuals learn — implicitly and repeatedly — that their affective experience is a treatable condition rather than a meaningful state warranting non-clinical forms of attention and response.

The search environment, as it is presently structured, may be systematically narrowing the interpretive options available to individuals at precisely the moment of greatest vulnerability — when the framing of an experience has the most significant downstream consequences for how it is understood, addressed, and resolved.

5. Pilot Study: Los Angeles County, December 2025

5.1 Overview

To assess the feasibility of search-based emotional distress signal detection, the Uplift Index conducted a preliminary observational study in Los Angeles County over a 72-hour period from December 14 through 17, 2025. The study employed keyword impression data collected through Google Ads infrastructure as a proxy for population-level search behavior associated with loneliness and early emotional distress.

The methodology relied on advertising campaign impression data — which records the frequency with which queries matching a defined keyword set were submitted within a specified geographic boundary — rather than direct search volume data. This approach carries known limitations, discussed in Section 5.3, but provides a usable population-level signal while fully preserving user anonymity.

5.2 Key Findings

Volume: Over the three-day observation window, keyword terms associated with loneliness and early emotional distress generated 21,128 total impressions within the Los Angeles metropolitan area.

Geographic distribution: Impressions were heavily concentrated in the city of Los Angeles (18,052 of 21,128 total), with meaningful signal in Long Beach (241), Vernon (230), Santa Monica (162), and West Hollywood (117), among 60+ additional municipalities.

Age distribution: The strongest signal came from adults aged 25–34, followed closely by 18–24, with the 65+ group showing notable representation as well. The pattern was distributed across age groups rather than concentrated at any single point in the lifespan.

Gender: Female users expressed these signals at higher rates than male users. More notably, women were nearly twice as likely to click through to a support resource when one was made available, a finding with potential implications for the design of intervention delivery.

Relationship status: Adults in relationships expressed emotional distress signals at significantly higher rates than single individuals, with married adults falling between these groups. This finding directly challenges the common intuition that loneliness is primarily a function of social isolation, suggesting instead that relational context does not straightforwardly buffer against the form of emotional experience indexed by these queries.

Parenting context: Parents of preschool-aged children and parents of teenagers (13–17) appeared at nearly identical rates, while parents of toddlers showed substantially lower representation. This suggests that specific parenting life stages carry elevated loneliness risk in ways that are not captured by coarse demographic categories.

Device: The substantial majority of searches were conducted on mobile devices, a finding consistent with the hypothesis that these query expressions represent private, low-friction affective disclosures rather than deliberate, desktop-initiated research behaviors.

Common search terms: The most frequently observed expressions included "lonely," "loneliness," "I am so lonely," "I'm so lonely," "how to stop feeling lonely," "I feel alone," and "how to overcome loneliness." The preponderance of affective, first-person language — as opposed to clinical terminology or explicit help-seeking formulations — is consistent with the theoretical framework advanced in this paper.

5.3 Limitations

Several significant methodological limitations must be acknowledged before the findings reported above can be appropriately interpreted.

The 72-hour observation window, conducted in December, does not support conclusions regarding trend direction, seasonal variation, or baseline prevalence. December represents a plausibly elevated period for loneliness-related search activity, and the findings cannot be extrapolated to other temporal contexts without additional longitudinal data.

Impression data does not constitute a direct measure of organic search volume. It reflects the frequency with which a query matching the defined keyword set was submitted during intervals in which an advertisement was eligible to serve, a quantity that is jointly determined by both ad targeting parameters and underlying search behavior. Users who employed different phrasing, or who searched in contexts in which no advertisement was eligible, would not be captured in these data.

The study is geographically bounded to a single metropolitan area and cannot be generalized to other regional or national populations. Los Angeles exhibits distinctive demographic and socioeconomic characteristics that may influence the observed pattern of results in ways that cannot be fully assessed without comparative data.

No causal claims are advanced. The data documents what was expressed at the population level, not the motivations, circumstances, or clinical status of the individuals who expressed it. The methodology provides no basis for inferring individual risk level or clinical need.

No personally identifiable information was collected, stored, or analyzed at any stage of this study.

6. What This Signal Is — and Is Not

It is important to be precise about the nature of the signal this research is attempting to identify and characterize.

Search-based emotional distress signals are not diagnostic at the individual level. A person who submits the query "I am so lonely" may be experiencing chronic and severe loneliness, may be having a momentarily difficult day, may be conducting research for an academic project, or may be doing any number of other things. At the level of the individual record, these expressions are not interpretable.

At the population level, however, aggregated patterns across thousands of search records begin to exhibit signal-like properties — not as indicators of individual clinical need, but as indices of the ambient emotional climate of a community. The analogy to epidemiological surveillance is instructive: in the same way that elevated emergency department admissions for respiratory illness can signal an incipient influenza wave before it reaches diagnostic peak, elevated search expression of emotional distress may index community-level psychological strain before it registers in clinical or social service systems.

The goal of this approach is not to identify individuals in distress. It is to observe, at the community level, where emotional strain is present, growing, or unevenly distributed — so that communities and public health planners can assess whether early-stage support resources are adequate, accessible, and appropriately targeted.

This constitutes a qualitatively different type of public health signal from those conventionally generated by institutional systems. It does not produce case counts, does not yield diagnoses, and does not identify individuals eligible for clinical intervention. What it does provide is evidence regarding whether a layer of pre-clinical distress is present within a given population — and preliminary characterization of the demographic and geographic distribution of that distress.

7. Research Agenda

The study described in Section 5 constitutes an initial feasibility assessment. A more rigorous and sustained research program would pursue the following priorities.

Longitudinal tracking: A 90-day or full-year observation window would enable researchers to identify seasonal patterns, establish trend direction, and assess whether the signal responds to identifiable community events. Loneliness-related search behavior that demonstrates systematic elevation following natural disasters, economic disruptions, or major cultural events would substantially strengthen the validity case for this signal as a meaningful population-level indicator.

Multi-city comparison. Comparing search-based distress signals across cities of similar size but different social and economic characteristics would allow for hypothesis testing about the structural factors that elevate community-level loneliness expression. Are signals higher in cities with lower social infrastructure? In cities with higher housing costs and longer commutes? These questions become answerable with comparative data.

Cross-validation against existing measures. The most important next step for methodological credibility is comparing search-derived signals against existing loneliness and social isolation prevalence data from validated surveys (e.g., the UCLA Loneliness Scale, CDC BRFSS social connectedness module). If the search signal correlates with known prevalence patterns at the geographic level, its validity as an early indicator is substantially strengthened.

Keyword expansion: The Los Angeles pilot study focused primarily on loneliness-related query terms. A more comprehensive research program should map the full semantic space of pre-clinical emotional distress expression — including query clusters associated with anxiety,

hopelessness, social withdrawal, and related affective states — and assess the signal properties of each.

Response environment audit: A complementary and necessary research direction involves systematic auditing of the response environment that users actually encounter when submitting these queries. Documenting the degree to which search results are pharmaceutical, clinical, commercial, or supportive — and examining whether that distribution varies by geography, demographic targeting parameters, and temporal context — would provide important empirical grounding for the concerns about commercial distortion raised in Section 4.

8. Conclusion

The Uplift Index is grounded in a straightforward empirical observation: the majority of emotional experience occurring within populations is invisible to the systems designed to support those populations, because those systems are architecturally configured to detect distress only once it has become speakable, diagnosable, or eligible for formal intervention.

Internet search behavior provides observational access to a quieter and earlier layer of emotional experience — one that exists before meaning has fully consolidated, before identity has organized itself around a diagnosis, and before the individual has determined that they are someone who requires help. The approach proposed here is not a replacement for clinical measurement; it is a complement to it, one that operates upstream of the threshold at which clinical systems become relevant.

The objective is not to intervene in that private space of pre-disclosure expression. It is to observe it — at the population level, using aggregated and anonymized data, under conditions of methodological transparency and ethical accountability — so that communities and public health systems have an opportunity to respond before the window for early, low-intensity intervention has closed.

This is a research initiative at its earliest stage of development. The pilot findings reported here are preliminary, the methodology carries significant limitations, and the translational path from observational signal to actionable public health application is necessarily long. The signal, however, appears to be real; the question it poses is of genuine public health significance; and the gap it addresses — the pre-disclosure interval of emotional distress — has remained largely invisible to the institutional infrastructure nominally responsible for population-level wellbeing.

The Uplift Index welcomes collaboration with researchers, public health practitioners, and institutions interested in exploring how this class of early-distress signal can be responsibly developed, validated, and applied.

All data referenced in this paper reflects aggregated, anonymized signals collected at the population level. No personally identifiable information was collected, stored, or analyzed at any stage of this research.

Contact: mark@upliftindex.org | upliftindex.org

