**Comparing two approaches to calculating willingness to pay (WTP) Score**

The willingness-to-pay score was determined by averaging the ratings of four items: data protection, reduced use of data for marketing, assistance in controlling prolonged usage, and mitigation of fake news and radicalization on social media. Each item was rated on a 5-point Likert scale. In the original work by (Sindermann et al., 2020) participants were divided into three main categories based on their willingness-to-pay scores, as follows:

* Not Willing to Pay (1.00 <= WTP score <= 2.50); mathematical interval of [1, 2.5]
* Neutral (2.51 <= WTP score <= 3.50); mathematical interval of [2.51, 3.5]
* Willing to Pay (3.51 <= WTP score <= 5.00); mathematical interval of [3.51, 5.0]

This approach was based on the premise that cases with scores of 2.50 or 3.50, which could be rounded to belong to either the lower or higher group, were assigned to the lower group. Individuals with scores of 2.50 were classified as not willing to pay (rather than neutral), while those with scores of 3.50 were considered neutral (rather than willing to pay). However, the WTP score derived from four questions yields a multiple of 0.25 within the range of 1 to 5, therefore, the Willing to Pay group has one possible value less that counts for that group. Hence, including the value 3.5 in the Willing to Pay group helps balance the group sizes more effectively between those not willing to pay and those willing to pay. The mathematical description for this approach is outlined as follows:

* Not Willing to Pay (1.00 <= WTP <= 2.50); mathematical interval of [1, 2.5]
* Neutral (2.50 < WTP < 3.50); mathematical interval of ]2.5, 3.5[
* Willing to Pay (3.50 <= WTP <= 5.00); mathematical interval of [3.5, 5.0]

Figure S1 shows the alternative representation of the proportions of participants indicating willingness to pay for social media digital wellbeing services which more clearly illustrates the higher percentage in willingness to pay in the second dataset compared to the first dataset.

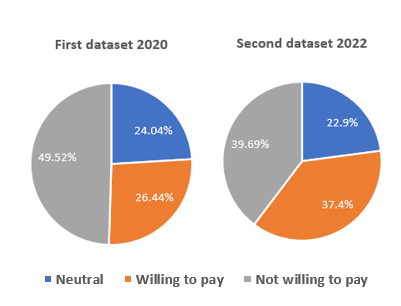


Figure S1. Proportions of participants indicating willingness to pay based on the new approach.