CONTEXT DURATION IN NATURALISTIC SMARTPHONE USE: EVIDENCE FOR MILLISECOND-SCALE PROCESSING

Anonymous authors
Paper under double-blind review

ABSTRACT

Smartphone use is typically reported as hours of daily screen time, and content is still designed for second-to-minute attention. Recommendation and content systems have become good at predicting what a person might want, yet people often do not act on what they are shown. We call this the recommendation–action gap, and we show that it has a timescale. Using continuous on-device telemetry from 35 participants over 15 days (3.2 million screen touches; keyboard input excluded; logging only while the device was unlocked), we treat inter-touch intervals as a behavioral proxy for context duration, the time a user remains engaged with a given on-screen context. Relative to a 2022 measurement, mean daily use rose 36% (to 4 h 53 min) and daily touches rose 19% (to 5,384), yet aggregate context duration contracted 23% to 3.8 to 7.8 s, orders of magnitude below television-scale engagement (5 to 21 min). The inter-touch distribution is strongly right-skewed. Half of all touches arrive within 266 ms of the previous touch, and 75% within 1.06 s. Half of all interaction therefore sits inside the $\sim$300 ms window in which people decide to approach or avoid, well before the plays, add-to-carts, and orders that systems are trained to predict. Time on the device is accumulated through revolving-door sessions rather than sustained attention (Instagram: median 67 s sessions summing to 75 min/day; WhatsApp: median 23 s sessions summing to 38 min/day), and only the extreme tail of context duration approaches long-form media. We repeat this analysis across social media, streaming, music, food and quick commerce, and AI assistants. Decision windows differ by category: 7 to 14 s for what to view on Instagram, 4 to 7 s for what to reply to on WhatsApp, about 10 to 30 s for weekday streaming choice, 4 to 9 s for food orders, and 4 to 9 s on average across AI assistants. In every category the choice is made within seconds. Closing the recommendation–action gap needs content that can be processed in milliseconds and acted on within the measured window of each category.

1 INTRODUCTION

The smartphone is now the medium through which most people meet most content. Daily use is measured in hours, and the time keeps growing. Yet the content placed on the medium is still designed for the attention of an earlier one. Fifteen- and thirty-second commercials, six-second bumpers, and multi-minute episodes all assume that a person will stay with a screen long enough to take in a message at that grain. This paper asks whether that assumption holds. We measure, directly and in the wild, how long people remain with what is on their screen, and we find that the answer is a few seconds at most and, for half of all touches, under 300 milliseconds.

Lift and shift from older media. The main response of the industry has been to move formats across media rather than to rethink them. Television commercials are routinely cropped, trimmed, and re-served in mobile feeds. Drama became microdrama; advertising has no micro-ad.Content has moved faster than advertising. Microdramas, serialized scripted stories of two to three minutes per episode, generated an estimated $11 billion in global revenue in 2025, nearly twice the revenue of free ad-supported streaming television (Omdia, 2025). In the United States, users of the leading microdrama app now spend more time per day in it on mobile than in Netflix, Prime Video, or

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Disney+ (35.7 minutes against 24.8, 26.9, and 23.0 minutes; Advanced Television, 2026). The episode has been redesigned for the phone. The persuasive message has not: there is no *micro-ad* built for the timescale at which smartphone users engage.

Redefining communication for the smartphone. If the unit of engagement on a smartphone differs from the unit that formats were designed for, communication on the medium has to be redefined from its temporal unit upward. It cannot simply be resized from television. The same is true of the systems that decide what to show. Recommenders, feeds, and ranking models are trained on signals such as clicks, dwell time, and watch time (Yi et al., 2014; Covington et al., 2016). These signals are aggregated over seconds to minutes, and they are evaluated at that grain. Whether this grain matches the grain of the human decision is an empirical question, and it has rarely been measured.

The research gap: a recommendation–action gap. Recommendation and content systems have become very good at predicting what a person might want. The step that matters, whether the person acts, is much less well served. We call this the *recommendation–action gap*. In commerce, the average mobile conversion rate is about 2.9% (Dynamic Yield, 2026), and about 70% of shopping carts are abandoned before purchase (Baymard Institute, 2025). Systems surface relevant options, and people still do not act on them. Our measurements give this gap a timescale. Half of all touches in our data arrive within 266 ms of the previous touch, and 75% within 1.06 s, whereas the labels that learning systems optimize, such as a play, an add-to-cart, or a completed order, lie beyond that window.

Research questions. We address two questions.

- **RQ1.** How should the smartphone be understood as a medium? In particular, what is its temporal unit of engagement, and how does that unit vary across categories of apps?
- **RQ2.** How has this behaviour evolved over time, as daily use and touch frequency have grown?

Approach and contributions. We operationalize *context duration*, the time a user remains cognitively engaged with a given on-screen context, as the interval between successive touches. We measure it through continuous on-device telemetry from 35 participants over 15 days, covering 3.2 million touches, with keyboard input excluded and logging active only while the device was unlocked. We compare the results with a 2022 measurement that used the same method (Kamath et al., 2023). Our contributions are as follows.

- **A measured temporal unit for the smartphone.** Aggregate context duration is 3.8 to 7.8 s, orders of magnitude below the 5 to 21 minutes of television (Cha et al., 2008). The distribution is strongly right-skewed: 50% of touches arrive within 266 ms and 75% within 1.06 s.
- **Evidence that the unit is contracting.** Between 2022 and 2025, mean daily use rose 36% (from 3 h 36 min to 4 h 53 min) and daily touches rose 19% (from 4,513 to 5,384), while context duration fell 23%. People spend more time on the phone and stay with each screen for less time.
- **A category-level account of the recommendation–action gap.** Across social media, streaming, music, food delivery, quick commerce, and AI assistants, decisions are made within a few seconds, and sessions are revolving doors of short visits that accumulate into long daily totals.
- **Implications for communication and for learning systems.** We argue that persuasive content should be designed at the millisecond-to-second scale, and that the training and evaluation horizons of human-facing systems should be set by the measured unit of the medium rather than inherited from older ones.

2 METHODS

The study measures the time a person remains with what is on screen. Recommendation systems can be right about relevance and still fail if the visit ends before the action. The protocol is built to time that failure.

2.1 PARTICIPANTS

Thirty-seven Android users in India, aged 20 to 30, installed a purpose-built logger and were recorded continuously for 15 days. All participants gave informed consent to take part in the study and to share their data. Two logs failed (participant identifiers 24 and 26) and were removed, leaving an analytic sample of 35. The logger was passive after installation. Participants spent about five minutes on the study itself: installing the application at the start and returning the log at the end.

2.2 WHAT WAS RECORDED

The logger wrote two fields for each touch: the timestamp, and the application in which the touch occurred. It did not record the screen, the notification drawer, or any content of what was shown. Keyboard input was excluded, and writing was active only while the device was unlocked. The resulting log contains 3,205,636 touches.

The enrolled window is 35 participants by 15 days by 24 hours, which is 12,600 hours, or 45,360,000 seconds. That figure is calendar coverage. It is not time on the screen. Time on the screen is the daily-use measure in Section 3.

Because the log contains time and application only, it does not contain the recommendation a person was shown, and it does not contain the action that followed. A play, an add-to-cart, a sent reply, and a completed order are not events in these data. The gap is measured by placing the recorded time next to the action the category requires, and next to published rates at which that action fails (Jannach & Jugovac, 2019; Dynamic Yield, 2026; Baymard Institute, 2025; Lamere, 2014; Gomez-Uribe & Hunt, 2015).

2.3 CONTEXT DURATION, THE DECISION WINDOW, AND THE ACTION

Context duration is the time a user remains cognitively engaged with a given on-screen context. We operationalize it as the interval between successive touches, the same construction used for television, where context duration is the time between remote switches (Cha et al., 2008). For a named app, the interval is taken between successive touches logged to that app. A new context begins at each touch. The next touch is the act that ends it: staying, leaving, or moving on.

The decision window is the 90th to 95th percentile of context duration within an app. It describes the longer ordinary contexts, after the fleeting touches and before the extreme tail. The category tables (Tables 3–6) report this window. It is the latest moment at which a recommendation can still be met with a choice. Even this upper edge of ordinary use is a few seconds in every category we study.

A session is the visit that would have to contain the action, and daily time is the sum of those visits. Both are medians across participants, as reported in the app-level summaries. App switch or screen off event closes a session. We therefore treat session length as a reported summary, and we do not re-derive it from the touch log.

The action is defined by the category, not by the log. Quick commerce requires search, cart, and checkout. Food delivery requires an order. Music requires a listen. Video requires a title to be chosen and started. Social apps require a watch, a save, a follow, or a reply. An assistant requires the person to act on the response. The recommendation–action gap for a category is the mismatch among three times: the decision window, in which the choice is made; the session, which has to hold the action; and the time the action itself takes. We observe the first two. We compare them with the third.

Table 1: Change in use, touches, and context duration from 2022 to 2025. Daily use is a mean. The touch figure is the study median in 2025, set against the 2022 daily figure from the same method. The change in context duration is computed on the lower bound of the range.

Measure	2022	2025	Change
Mean daily use	3 h 36 min	4 h 53 min	+36%
Daily touches	4,513	5,384	+19%
Aggregate context duration	5–10 s	3.8–7.8 s	−23%

2.4 Comparison with 2022

Aggregate daily use, daily touches, and context duration are compared with a 2022 measurement made by the same method (Kamath et al., 2023). The 2022 figures are mean daily use of 3 h 36 min, 4,513 daily touches, and an aggregate context duration of 5 to 10 s. The comparison is a repeated measurement, not a within-person panel. The same individuals were not followed from 2022 to 2025.

The 2025 aggregate context duration is 3.8 to 7.8 s, a contraction of 23% measured on the lower bound (from 5 s to 3.84 s). This range summarizes central context duration. It should be kept distinct from the median inter-touch interval of 266 ms, which is a percentile of the pooled distribution of all touches (Section 3.2). All context durations reported are 90-95 percentile.

2.5 What the measurement adds

Studies of smartphone use have mostly timed sessions and app launches. This measurement times the interval between touches, across the phone, inside the apps where a recommendation is supposed to become an action. That interval is the unit that session and launch measures do not resolve. It is also the unit at which a learning system would have to act if its success condition were the user's action rather than a later label. The 2022 comparison then shows whether that unit is stable.

3 Results

3.1 Use rose, and the time in which to act shrank

People spent more time on the phone in 2025 than in 2022, and they touched it more often. They remained with each context for less time. Table 1 reports the comparison.

The direction of the three rows is the recommendation–action gap in motion. Each added touch is another moment at which something can be recommended. The context in which that recommendation would have to become an action is shorter than it was three years earlier. Systems that treat rising screen time as rising opportunity to convert are reading the first row and missing the third.

Mean daily use of 4 h 53 min is 36% above 3 h 36 min (35.65% unrounded). Daily touches of 5,384 are 19% above 4,513 (19.3% unrounded). Aggregate context duration fell from 5–10 s to 3.8–7.8 s. On television the comparable figure remains 5 to 21 minutes (Cha et al., 2008).

3.2 Most touches are over before an action can be labeled

The distribution of inter-touch intervals is strongly right-skewed. Half of all touches arrive within 266 ms of the previous touch. Three quarters arrive within 1.06 s (1,056 ms). Table 2 places these percentiles against the share of clock time they account for.

At 266 ms the decision available is approach or avoidance (Chen & Bargh, 1999; Duckworth et al., 2002). The recommender's training label is a later event. Three quarters of interactions end within about a second, before that label can be written.

The tail does not rescue the formats built for that label. About half of mean daily use, 2 h 29 min, falls in intervals above the 99th percentile of context duration, and 2 h 13 min falls below it. Long contexts are rare as decisions and large as clock time. Only the last percentile of the last percentile,

Table 2: Bulk of the touch distribution, and how daily time splits at the 99th percentile of context duration. The two time rows are the study split of mean daily use; they are reported as recorded and are not forced to sum to 100%.

Summary	Interval	Share of the day
Median inter-touch interval	266 ms	bulk of touches
75th percentile	1.06 s	bulk of touches
Below the 99th percentile	—	2 h 13 min (45%)
Above the 99th percentile	—	2 h 29 min (50%)

Table 3: Social apps. Touches, session, and daily time are medians. The decision window is the P90–P95 context duration.

App	Touches/day	P90 (s)	P95 (s)	Session	Daily time
WhatsApp	862	3.92	6.70	23 s	38 min 21 s
Instagram	1,150	7.34	14.42	1 min 7 s	1 h 15 min
LinkedIn	84	2.53	4.91	40 s	1 min 55 s
Reddit	56.5	8.61	14.10	55 s	2 min 2 s
Snapchat	55	3.50	7.72	2 s	2 min 3 s

on the order of 0.01% of interactions, reaches television-scale engagement, that is, more than 5 minutes. At that extreme, 57 min 17 s of the day is spent in such long contexts, against 3 h 31 min 44 s in short bursts of activity.

3.3 The same gap inside every category

Category results use the same three times: decision window (P90–P95 context duration), median session, and median daily time. The action named in the prose is the action the app requires of the visit. Sparse apps, those with a median of one or two touches a day, are reported and then set aside when a category claim would depend on them.

Social apps. Social use is the clearest revolving door. Instagram accounts for a median of 1,150 touches a day, a decision window of 7.3–14.4 s, a session of 1 min 7 s, and a daily total of 1 h 15 min 26 s. The hour is an accumulation of minute-long visits rather than a few long sittings. Inside each visit, what to see is settled in 7 to 14 seconds. A watch, a save, or a follow has to be available on that context. A second screen is a later visit, and the next visit is a new draw.

WhatsApp reverses the daily-time picture and keeps the same structure. The median session is 23 s and the day is 38 min 21 s, produced by a median of 862 touches. What to reply to is settled in 3.9–6.7 s. The reply itself is typed, and typing was not logged, so the recorded window is the tapping around the reply. The visit that contains the reply is still only 23 seconds long.

LinkedIn, Reddit, and Snapchat show the same separation of window and visit at lower volume (Table 3). Snapchat is the extreme case. Its median session is 2 s, shorter than its own P90 of 3.5 s, against a daily total of 2 min 3 s and 55 touches. Many Snapchat visits end on the opening touch. There is no later moment in that visit for a recommendation to become a follow or a reply.

Food delivery and quick commerce. Commerce is where the action is a transaction, and where the visit is shortest relative to the steps that transaction takes. On Zepto and Blinkit the decision window is 2.4–5.5 s, the session is 26–34 s, and the entire day in the app is 36–42 s (Table 4). Daily time and session length are the same order of magnitude. The day is essentially one visit. Search, cart, and checkout have to finish inside that visit, and the choice of what to order is made in its first few seconds.

Zomato and Swiggy allow a slightly longer look and the same conclusion. The category window is 3.9–8.9 s, the session is about half a minute (40 s and 31 s), and the day is about two minutes (2 min 8 s and 1 min 42 s). What to eat is decided before a menu can be browsed at the pace a longer medium would assume. The action is placing the order. Published mobile conversion is about 2.9%, and about 70% of carts are abandoned (Dynamic Yield, 2026; Baymard Institute, 2025). The

Table 4: Food delivery and quick commerce. Touches, session, and daily time are medians. The decision window is the P90–P95 context duration. BigBasket is sparse (median 2 touches a day) and is not used to set the quick-commerce window.

App	Touches/day	P90 (s)	P95 (s)	Session	Daily time
Zomato	35.2	2.43	3.87	40 s	2 min 8 s
Swiggy	15.0	5.26	13.84	31 s	1 min 42 s
Zepto	18.0	2.44	5.47	26 s	36 s
Blinkit	31.0	2.44	4.09	34 s	42 s
BigBasket	2.0	12.68	17.41	19 s	37 s

Table 5: Music and video. Touches, session, and daily time are medians. Prime Video weekday and several Hotstar rows rest on very few touches; the Netflix and Spotify rows are the stable comparisons.

App	Touches/day	P90 (s)	P95 (s)	Session	Daily time
Spotify	20.5	3.77	6.13	12 s	50 s
Netflix, weekday	110	10.69	31.25	1 min 5 s	44 min 49 s
Netflix, weekend	102	37.71	94.07	59 s	28 min 50 s
Prime Video, weekday	1.0	9.90	32.35	19 s	48 s
Prime Video, weekend	5.2	9.68	19.57	49 s	51 s
Hotstar, weekday	14.8	5.07	7.19	9 s	33 s
Hotstar, weekend	6.0	3.23	5.40	8 s	28 s

telemetry does not explain those rates by itself. It shows that the visit in which the rate would have to be earned is under a minute, and that the choice inside it is under ten seconds.

BigBasket is reported for completeness. A median of 2.0 touches a day, a window of 12.7–17.4 s, a session of 19 s, and a day of 37 s describe a sparse app. Its longer window would pull a quick-commerce aggregate away from Zepto and Blinkit, so the claim for quick commerce rests on those two apps.

Music and video. On Spotify the decision window is 3.8–6.1 s, the session is 12 s, and the day is 50 s, from a median of 20.5 touches (Table 5). What to listen to is settled inside a visit that is itself only 12 seconds long.

Among apps with substantial daily use, weekday Netflix has the longest weekday decision window, and it is still inside the abandonment window the service has already described. The decision window is 10.7–31.3 s, the session is 1 min 5 s, and the day is 44 min 49 s, from 110 touches. A member who loses interest after 60 to 90 seconds of choosing, having looked at 10 to 20 titles, is leaving at the end of this session (Gomez-Uribe & Hunt, 2015). The action is to choose and start a title. Watch time after that start is a different part of the distribution, the long tail that holds clock time. The choice itself is over inside half a minute.

The weekend does not hand the decision back to a longer medium. Netflix's weekend window lengthens to 37.7–94.1 s, while the session shortens to 59 s and the day falls to 28 min 50 s. The P95 context is longer than the median session. The longer look is the visit. There is no remainder of the session in which a slower choice can still become a start.

Prime Video and Disney+ Hotstar make the same point at very small daily totals. Weekday Prime is a median of 1.0 touch, a window of 9.9–32.4 s, a session of 19 s, and a day of 48 s. The P95 context exceeds the session: the visit is over before the longer ordinary context finishes. Weekend Prime is 5.2 touches, a 49 s session, and a day of 51 s. Hotstar's weekday session is 9 s and its day is 33 s; the weekend session is 8 s and the day is 28 s. A recommendation on these apps has one brief chance to produce a start, and on a typical day it has almost no accumulated time in which to recover from a miss.

Assistants. ChatGPT and Perplexity are the category in which the recommendation is a generated response, and the action is whatever the person does with it. ChatGPT's decision window is 5.9–

Table 6: AI assistants. Touches, session, and daily time are medians. The decision window is the P90–P95 context duration.

App	Touches/day	P90 (s)	P95 (s)	Session	Daily time
ChatGPT	41	5.93	11.66	30 s	2 min
Perplexity	35	2.86	5.84	47 s	1 min 8 s

11.7 s, its session is 30 s, and its day is 2 min, from 41 touches. Perplexity's window is 2.9–5.8 s, its session is 47 s, and its day is 1 min 8 s, from 35 touches (Table 6). Averaged across the two, the decision window is 4 to 9 s. Information is processed in a few seconds. The visit that would have to contain the action lasts about half a minute to a minute. A response that becomes usable only after that visit has ended is a recommendation with nowhere to land.

Across Tables 3–6, the common case is a choice made in a few seconds inside a short visit, with the day a sum of such visits. The exceptions are stricter rather than looser. In Snapchat, weekday Prime, and weekend Netflix, the upper end of ordinary context duration is as long as the visit or longer. The person leaves while the context is still open. Either way, the action has to be completable inside one short visit, and the choice has to be made at the start of it.

4 Discussion

4.1 The gap is a mismatch of times

A recommendation is an offer of a next action. The offer succeeds when the person takes the action. The measurements time the room in which that can happen, and the room is smaller than the offer assumes.

Three clocks run at once. The first is the bulk of touches: half arrive within 266 ms and three quarters within 1.06 s. This is approach or avoidance. The second is the decision window, the 90th to 95th percentile of context duration, a few seconds in commerce, music, and assistants, and up to about half a minute for weekday Netflix. This is the choice among what is on screen. The third is the session, from a few seconds on Snapchat to about a minute on Instagram and weekday Netflix, and it is often the whole of the day's use of a commerce or music app. The action has to fit in the third clock, and the choice of it has to fit in the second. Labels used to train recommenders are written after both clocks have stopped.

From 2022 to 2025 the context shortened while the day lengthened: more touches, shorter contexts, and more hours on the phone. The number of offers a system can make has grown. The time each offer has in which to become an action has shrunk. Rising engagement and a widening recommendation–action gap are the same trend read at two grains.

The split of clock time makes the usual substitute for action look better than it is. About half of daily use sits above the 99th percentile of context duration. A model trained on dwell or watch time (Yi et al., 2014; Covington et al., 2016) is trained on that half. The decisions, including the decision to act, sit in the other 99% of touches. Optimizing time and optimizing action select different parts of the same log.

4.2 Implications in general

Communication on this medium has to be built from the action backward. The unit is the context that can still change what the person does next. A message, a ranking, or a response that requires the person to remain into the tail is addressed to the 0.01% of interactions in which context duration finally resembles television. The rest of the day is a sequence of approach-or-avoid decisions, most of them over inside a second, with a choice among options that is over inside the decision window of that app.

The practical consequence is to judge a recommendation by whether the action occurred inside the visit, and to design the path so that the action can occur there. Relevance is the condition the field has learned to optimize. Fit to the window is the condition that decides whether relevance becomes

behavior. Field evidence already shows that offline accuracy does not reliably become conversion, sales, or retention (Jannach & Jugovac, 2019). These data give that finding a duration. The action fails to appear when the visit is shorter than the action, and the choice is made before the action's steps can be shown.

This is a design constraint before it is a model constraint. Shortening the path is part of the recommendation. An extra confirmation, a second screen, or a response that arrives after the session median is an extension of the offer past the time in which the person acts. The stimulus has to be legible in one context, and the action has to be completable in one session of that category.

4.3 Implications for learning systems

Human-facing models are trained and evaluated on horizons of seconds to minutes because their labels are plays, dwell, watch time, and conversions (Yi et al., 2014; Covington et al., 2016). Those labels are real events. They are the wrong grain for the decision that produces them. Two changes follow from the measurements.

Set the horizon to the category. The reward for a recommendation should be the action inside that category's decision window and session, not a later aggregate. For music, that is the absence of a skip inside 4 to 6 s. For quick commerce, that is progress to checkout inside a visit of about half a minute, with the choice visible in the first few seconds. For an assistant, that is an action taken before a session of 30 to 47 s ends. A single horizon inherited from watch time will fit Netflix's tail and miss every other row in the tables.

Use the next touch as a label the bulk of the data can actually support. Half of touches arrive by 266 ms. A stay-or-leave label at the next touch is observed for the interactions that conversion labels never see. Conversion remains the outcome that matters. It should be predicted through the intermediate action, because that is the action the log contains. Credit assigned across a minute is assigned across a span the user has often already left.

4.4 Limitations

The interval between touches is a proxy for cognitive engagement. A long interval can be reading, watching, or waiting. A short interval can be a motor slip or a scroll. The proxy is the one the television comparison uses, time between acts of control, and it is available for every touch. It is not a measurement of gaze or of felt attention.

The log does not contain the recommendation or the action. Package name and timestamp can time the room. They cannot show what was offered, and they cannot show whether an order, a listen, or a reply occurred. Published conversion, skip, and abandonment figures are used as the external record of the action. They are not estimated from these participants. The paper claims a mismatch of timescales. It does not claim that short contexts cause a specific drop in conversion.

Keyboard input was excluded. Search queries, chat replies, and other typed actions are invisible, and the intervals around them are the taps that remain. Messaging and assistant results should be read with that omission. The session boundary is not documented in the released materials, so session lengths are used as reported and are not reconstructed. For a few apps the P95 context duration exceeds the median session. That is reported as a result. It also means the two summaries are not guaranteed to be nested cuts of one distribution.

The 2022 comparison is a repeated measurement, not a panel of the same people. Prime Video on a weekday (median 1.0 touch) and BigBasket (median 2.0 touches) are too sparse to anchor a category. The sample is 35 Android users in India, aged 20 to 30, observed for 15 days after two logging failures. Nothing in the design supports a claim about iOS, about other age groups or countries, or about the population beyond these logs.

4.5 Summary and conclusion

The smartphone's temporal unit is a context of a few hundred milliseconds for most touches, and a few seconds at the upper edge of ordinary use. It varies by category. It is, in every category we measured, shorter than the visit would need to be for today's formats and today's training labels to

match the action. Between 2022 and 2025 daily use rose 36% and daily touches rose 19%, while context duration fell 23%. The medium offers more moments to recommend and less time for each recommendation to become an action.

Systems that surface a relevant option, and messages built for a longer hold, are aimed at a part of the day the log shows to be the tail. The action is decided earlier, inside a short visit, and often inside the first second. Persuasive content, commerce paths, ranked media, and assistant responses should be built to be acted on inside the measured window of their category. Learning systems should be trained and evaluated on that same window. The objective is the action. The time in which it can happen has now been measured.

REFERENCES

Advanced Television. Research: Microdramas overtake streamers on mobile engagement, 2026. URL https://www.advanced-television.com/2026/02/23/research-microdramas-overtake-streamers-on-mobile-engagement/.

Baymard Institute. Cart abandonment rate statistics, 2025. URL https://baymard.com/lists/cart-abandonment-rate.

Meeyoung Cha, Pablo Rodriguez, Jon Crowcroft, Sue Moon, and Xavier Amatriain. Watching television over an ip network. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pp. 71–84, 2008.

Mark Chen and John A. Bargh. Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, 25 (2):215–224, 1999. doi: 10.1177/0146167299025002007.

Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for YouTube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys)*, pp. 191–198, 2016. doi: 10.1145/2959100.2959190.

Kimberly L. Duckworth, John A. Bargh, Magda Garcia, and Shelly Chaiken. The automatic evaluation of novel stimuli. *Psychological Science*, 13(6):513–519, 2002. doi: 10.1111/1467-9280.00490.

Dynamic Yield. ecommerce conversion rate benchmarks, 2026. URL https://marketing.dynamicyield.com/benchmarks/conversion-rate/.

Carlos A. Gomez-Uribe and Neil Hunt. The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, 6(4):13:1–13:19, 2015. doi: 10.1145/2843948.

Dietmar Jannach and Michael Jugovac. Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems*, 10(4):16:1–16:23, 2019. doi: 10.1145/3370082.

Karthika Kamath, Tanya Upadhyay, Jieya Rawal, Kirtana Sunil Phatnani, and Biju Dominic. Fostering effective communication between humans and machines. 2023.

Paul Lamere. The skip. Music Machinery (blog), 2014. URL https://musicmachinery.com/2014/05/02/the-skip/.

Omdia. Microdramas to generate $11bn in global revenue in 2025, 2025. URL https://omdia.tech.informa.com.

Xing Yi, Liangjie Hong, Erheng Zhong, Nathan Nan Liu, and Suju Rajan. Beyond clicks: Dwell time for personalization. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys)*, pp. 113–120, 2014. doi: 10.1145/2645710.2645724.