Considus

TECHNICAL WHITE PAPER

Catchlight for iPhone

The one star test

What sixteen notes and task apps' one and two star reviews complain about most, which complaints are growing, and what any notes app has to get right as a result.

Applies to Catchlight 1.0

What changed in version 1.1

Version 1.0 of this paper went out on 2 October and sorted the reviews with lists of keywords. This version reads them with a classifier instead, a model that's asked one plain question per complaint about every bad review, and I made the switch because the keywords were missing too much.

I checked both methods against the same 204 bad reviews, each labelled one at a time against the written definition of every complaint. The keywords found 54 in every 100 complaints in that sample, and when they flagged one they were right 73 times in 100. The classifier found 94, and was right 89 times in 100.

That changes the numbers, so **the rates in this version can't be compared with version 1.0's.** Most are higher, because the classifier finds complaints that never use the obvious word. Compare the order of the complaints between the two versions, not the percentages, and even then, some of the order moved because the method did.

It also changes the headline. Version 1.0 said the complaint that had grown the most, in points, was getting locked out of your own account. Read properly, it's price. Accounts still grew, and it's still one of the main findings here, just not the biggest riser. Section eight describes the new method in full, and the version history at the end lists what moved.

Introduction

When somebody gives a notes app one star, they nearly always tell you why. Multiply that by a few thousand and you have the most direct product research in the category, written by the people who actually use these apps, for free, in public.

In August I started collecting it. A small service reads the public App Store and Google Play reviews of sixteen notes and task apps every week and stores them, with the reviewer's name and identity stripped out on the way in. By 5 October it

held 26,414 reviews, of which 7,321 gave one or two stars.

This paper is what those 7,321 reviews say, grouped into the complaints people make most, and what I think any notes or task app, Catchlight included, has to get right as a result.

It's not a ranking of the apps. I list the ones I read, and then I never tie a complaint to any one of them, because a number about one app goes stale the day that app ships a fix, and because the useful finding is what the whole category keeps getting wrong.

What this covers, and what it doesn't

This covers what people complain about in one and two star reviews of notes and task apps, how those complaints differ between kinds of app, which are growing, and what people praise when an app gets it right.

It doesn't say which app is best or worst at anything. Every figure here covers all sixteen apps, a group of five, or, for the trend in section four, the nine apps with enough reviews to split.

It doesn't quote a single review. The reviews belong to the people who wrote them. What you'll read is counts, shares and my own sentences about the patterns.

It isn't a measurement of how often these problems happen. A review is usually written by someone annoyed enough to write one, and a classifier's reading of it is still an estimate of what they meant. Section eight is the full method and the list of limits, and the limits are longer than I'd like.

1 Which reviews, and how they were read

The apps. Sixteen, chosen because they are what someone shopping for a notes or task app on an iPhone is likely to be comparing.

GROUP	APPS
Privacy-first notes	Joplin, Notesnook, Obsidian, Simplenote, Standard Notes
Mainstream notes	Apple Notes, Bear, Evernote, Google Keep, Microsoft OneNote
Task apps	Apple Reminders, Microsoft To Do, Things 3, TickTick, Todoist
In no group	Notion, which does both notes and tasks

The reviews. 8,224 from the US App Store, 3,794 from the UK App Store and 14,396 from Google Play, which doesn't say which country a review came from. The stores only hand over so many reviews per app, so the window differs, and not by a little. For a busy app the reviews cover a few months. For a quiet one they go back more than a decade. Every figure uses each app's full collected history unless it says otherwise, and every figure is as of 5 October 2026.

The complaints. Each one or two star review went to a classifier, TypeSafe's Jev, with sixteen questions, one per complaint, and each question came with a written definition of what counts and what doesn't. The classifier answers each question with a probability, and a review counts for a complaint when that probability clears a threshold. The questions are in the appendix. Only a review's title and text were sent, never the reviewer's name. A review can count for more than one complaint.

The arithmetic. For each app, I took the share of its own one and two star reviews that counted for a complaint, then averaged those shares across the apps. That way a busy app with thousands of reviews counts the same as a quiet one with a few hundred, and a single unhappy user base cannot drown out the rest. An app with fewer than forty bad reviews would have been left out of the average, because a percentage of a tiny number is noise. None fell below it.

The check. Before trusting it on the full set, the classifier was tested against 204 labelled reviews, and afterwards a random sample of the reviews it flagged for each complaint was read against that complaint's definition. Two complaints needed a stricter threshold as a result. Section eight has the detail.

2 The complaints, ranked

Share of one and two star reviews that mention each complaint, averaged across all sixteen apps.

COMPLAINT	SHARE OF BAD REVIEWS
Bugs, crashes and slowness	49.6%
Price and subscriptions	15.3%
Sync and using more than one device	13.3%
Accounts, logins and being locked out	9.5%
Losing notes	8.9%
Formatting and the editor	8.5%
A bad redesign or update	8.1%
Reminders and notifications	6.8%
Bloat and complexity	6.0%
Privacy, ads and tracking	4.4%
Search and organising	4.3%
Export, import and lock-in	3.3%
Capture speed and widgets	3.2%
Support that doesn't reply	3.0%
AI pushed into the app	2.7%
Needing a connection	0.5%

Why bugs is so big. It's now a catch-all. Anything in the app that fails to work counts as a bug, and it also counts under its own complaint, so a sync that fails is in both rows, and so is a login that fails. Of the bad reviews counted as bugs, 69% are also counted under something more specific.

Why the column doesn't add up to 100%. It isn't meant to, and here it comes to 147.4%. Each row is the share of bad reviews that mention that complaint, and a review can mention several or none. Here is where the whole 100% of bad reviews

actually sits.

ONE AND TWO STAR REVIEWS THAT MENTION	SHARE
None of the sixteen complaints	14.1%
Exactly one	40.1%
Two or more	45.8%

The reviews that mention none are mostly short, often a line of frustration with nothing in it that says what went wrong. Half of them are 12 words or fewer, while half the rest run to 34 or more. That last row is why the column comes to 147.4%. On average, a bad review mentions about one and a half complaints.

Read the top of the ranking again, because it surprised me. The thing people complain about most, in about half the bad reviews, is the app not working, crashing, freezing, losing its place while you type.

Next comes price, which runs from the subscription itself to old notes held behind a paywall. Then the plumbing. Sync, which is whether your notes are where you left them, and accounts, which is whether you can get into your own notes at all.

Losing notes, a bad redesign and bloat all sit higher than they did in version 1.0, because people describe them in words the keyword lists didn't have. Support is the clearest case. On the same reviews, the keyword lists counted complaints about support at under a third of the classifier's rate, and bloat at under half.

Privacy, the thing a lot of apps lead their marketing with, mine included, sits at 4.4%. That isn't because people don't care about it. Section five shows they do, a lot. It's that privacy is rarely the thing that visibly breaks on them, assuming they know what level of privacy they actually have, and that a leak would ever be visible enough for them to notice.

3 Different apps, different complaints

The averages hide how differently each kind of app goes wrong. Here are the complaints where the three groups differ most, after bugs, which comes first in all three.

COMPLAINT	PRIVACY-FIRST NOTES	MAINSTREAM NOTES	TASK APPS
Bugs, crashes and slowness	55.8%	51.5%	41.0%
Sync and more than one device	18.3%	14.2%	9.9%
Accounts, logins and being locked out	17.0%	5.9%	6.8%
Reminders and notifications	0.4%	2.3%	19.0%
Formatting and the editor	11.2%	11.7%	2.5%
Losing notes	9.4%	13.8%	4.2%

Three things stand out.

People who chose a privacy-first app pay for it in sync and logins. These are apps people choose so the company can't read their notes, and in most of them that protection still runs through an account and a sign-in, or through a sync set-up the person has to manage. Both show up here. Nearly one in five bad reviews of this group is about sync, and more than one in six about accounts and getting locked out, with price sitting between the two at 17.2%. On accounts alone, that's nearly three times the figure for mainstream notes apps.

In task apps, the reminder is the product. Nearly one in five bad reviews of the task apps is about reminders that didn't fire, fired late, or can't do what was needed. Leave bugs aside and no complaint scores higher in any group. Version 1.0 said capture came next, and it doesn't any more. Price, sync, a bad redesign, accounts and bloat all sit ahead of it now.

Mainstream notes apps get the most complaints about lost notes. 13.8% of their bad reviews are about notes vanishing, being wiped, deleting themselves or coming back corrupted, against 9.4% and 4.2% for the other two groups.

4 What's getting worse

The same complaints, split at twelve months, reviews from the last year against everything before. Only apps with at least forty bad reviews on both sides of the split count here, which is nine of the sixteen, because a busy app's whole history can fit inside one year.

COMPLAINT	BEFORE THE LAST 12 MONTHS	LAST 12 MONTHS
Price and subscriptions	9.7%	13.5%
Accounts, logins and being locked out	12.5%	14.5%
A bad redesign or update	5.5%	7.5%
Reminders and notifications	2.8%	4.5%
AI pushed into the app	0.2%	1.2%
Sync and more than one device	16.1%	12.4%
Losing notes	13.1%	7.9%
Bugs, crashes and slowness	63.5%	50.5%

Price is the complaint that grew the most in points, by nearly four. Subscriptions where there weren't any, price rises, basic features moving behind a paywall, and old notes held back once someone stops paying.

Version 1.0 put accounts in that spot, with a rise of nearly four points. With the reviews read properly, accounts rose by two, from 12.5% to 14.5%, a real rise on a complaint that was already big, just not the biggest. Sign-in walls, verification codes that don't arrive, passwords that won't reset. And across the whole collection, not only the last year, it's nearly three times as big a share of privacy-first apps' bad reviews as of mainstream ones.

A bad redesign rose by the same two points, and it's the complaint version 1.0 measured worst. On the labelled sample, when the keywords flagged a bad redesign, they were wrong more often than they were right.

Reminders and AI grew faster in proportion, from smaller starting points. Complaints about AI being pushed into the app were almost non-existent before the last year and are now a measurable share. It's still small, and I'd be wary of anyone who calls it a backlash. The direction is clear, though.

Sync, crashes and losing notes have all fallen, a lot. It's fair to say the category has got better at all three, at least by its own users' account, and I want that on the record as plainly as the complaints are.

5 What people praise

Praise is still found with keyword sets, run over the 16,955 four and five star reviews, because the classifier's questions ask about complaints and not praise. They're version 1.0's sets with five of them tightened since, sync, privacy, reminders, search and needing a connection, and they're listed in the second appendix. The bad-review column in this table uses the same keywords, to compare like with like, so it won't match section two.

THEME	SHARE OF GOOD REVIEWS	SHARE OF BAD REVIEWS, SAME KEYWORDS
Sync and more than one device	8.7%	12.7%
Reminders and notifications	6.7%	8.9%
Formatting and the editor	6.5%	7.3%
Privacy, ads and tracking	5.3%	4.3%
Price and subscriptions	4.8%	10.3%
Search and organising	3.4%	2.1%
Bugs, crashes and slowness	3.3%	16.6%
Losing notes	0.5%	5.9%

Sync is the most praised theme and, read with the classifier, the third biggest complaint. No other theme sits as high on both lists, which makes it the one an app can least afford to get wrong.

Privacy and search are among the few themes praised more often than they are complained about. People mention them when they are the reason they chose the app.

And almost nobody writes a five star review because an app didn't crash or didn't lose their notes. Stability and safety are only visible when they fail.

6 What a notes app has to get right

Read as a list of what the category keeps getting wrong, the reviews point at these, roughly in order of how much they cost people.

Work, every time. No feature makes up for an app that freezes while you type. It's in about half the bad reviews, first in every group, and nobody praises its absence.

Charge fairly, and never hold old notes back. Price is the second biggest complaint and the one that grew the most in points, and the part any app can fix outright is a lapsed subscription that keeps someone from their own notes.

Put the notes where the person left them. Sync is the most praised theme and the third biggest complaint. An app either does it reliably or it shouldn't promise it.

Never stand between someone and their own notes. A login that can lock them out is the clearest case, and it's still rising.

If it does reminders, they have to fire. For task apps this is the whole job.

Don't lose anything, and make deleting hard to do by accident. Mainstream notes apps get this complaint most, and almost nobody mentions it when it goes right.

Ask before adding AI. Small today, but it barely existed a year ago.

7 Where Catchlight stands on each

I make Catchlight, so here is my own app against the same list. Where it falls short, it says so. Catchlight hasn't launched yet, so everything below describes what is built, not what thousands of people have tested.

COMPLAINT	CATCHLIGHT TODAY
Bugs, crashes and slowness	**No feature answers this, and I can't claim it before launch.** The reviews will judge it later
Price and subscriptions	**Partly answered.** It's an annual subscription. Stop paying and every Take stays readable and exportable, you just can't add or edit
Sync and more than one device	**The weakest answer here.** Encrypted files sync to a cloud folder you own, with no server of mine between. At launch that's one iPhone. Two devices side by side comes later, and there's no Mac or Android app yet
Accounts, logins and being locked out	**Answered outright.** No account, no email, no password. A 12-word Privacy phrase made on the phone is the only key, and I can't reset it for you
Losing notes	**Answered, with one exception.** Deleting asks first, conflicts show both versions, and everything exports to Markdown. Lose the 12 words with no copy and the Takes are gone, by design
Formatting and the editor	**Partly answered.** Text and checklists mix freely in one Take. Bold, headings and other inline formatting aren't there yet
Reminders and notifications	**Answered.** Repeats on any schedule, location reminders, snooze from the notification, up to three follow-up nudges, and on a checklist it names only what's left
Privacy, ads and tracking	**Answered.** Encrypted on the phone with a key only you hold. No ads, tracking, account, analytics or crash reporting of its own, and the app sends nothing to me, ever. A crash report only reaches me if you've told iOS to share them with developers, Apple's setting and your choice
AI pushed into the app	**Answered.** Catchlight has no AI of its own, and Apple's AI "Writing Tools" stay off unless you turn them on

8 The method, and what it can't tell you

One question per complaint. Every one and two star review, 7,321 of them, went to the classifier once, with sixteen questions. Each question names the complaint and carries a short definition of what counts and what doesn't, and the appendix lists all sixteen. The classifier answers each with a probability between 0 and 1. None of the 7,321 failed, so no review fell back to the old keywords.

A threshold per complaint. A review counts for a complaint when the answer is 0.7 or higher. Sync and AI count from 0.5, because at 0.7 the labelled sample showed them missing real complaints. In this run, losing notes and export count from 0.8, because a sample of what they flagged at 0.7 was taking in other things, deletions the reviewer made on purpose for the first, and account and subscription problems for the second.

The accuracy check. A sample of 204 bad reviews was labelled one at a time, complaint by complaint, against the same definitions, and both methods were scored against those labels. Across all sixteen complaints, the keywords found 54 in every 100 labelled complaints and were right 73 times in 100 when they flagged one. The classifier found 94 and was right 89 times in 100. It did better on every complaint except AI, where the two tied. The thresholds for sync and AI were chosen on that same sample, so these figures are likely a little better than it would score on reviews it had never been checked against.

Repeatability. The same 204 reviews were run through the classifier twice. No answer moved by more than 0.09, and 6 of the 3,264 yes-or-no calls changed.

The spot check. For every complaint, a random sample of eight reviews it flagged in the full set was read against its definition, with sixteen more where a sample looked wrong. That's what raised losing notes and export to 0.8. The lower threshold for sync and AI came from the labelled sample instead.

A classifier is still a reading, not a fact. It misses some complaints and flags a few it shouldn't. Even at 0.8, export still takes in a few login and storage complaints. The figures compare complaints with each other. They aren't the rate at which these

things happen.

Bugs overlaps everything. A sync failure counts as a bug and as a sync complaint, so adding bugs to any other row counts the same review twice.

Praise is still measured with keywords. Section five explains why, and why its bad-review column differs from section two.

Reviews are usually written by the annoyed. Most people who use an app never review it, and the ones who give one star aren't a random sample of anyone.

The windows differ. A busy app's reviews cover months and a quiet app's cover years, which is why no figure here compares raw counts between apps.

Store order matters. The stores hand over the newest reviews first, up to a cap. For the busiest apps, that means older reviews are missing entirely.

The trend uses nine apps, not sixteen. Only those with enough bad reviews on both sides of the twelve-month split count. Five of the nine are privacy-first notes apps, two are mainstream and two are task apps, so the trend leans towards the privacy-first group more than the rest of the study does.

Google Play reviews carry no country. The App Store figures cover the US and the UK. Play could be from anywhere.

I make a competing app. I've tried to keep that out of sections one to six, and section seven is where it belongs. You're entitled to read the whole thing with that in mind.

Appendix: the sixteen questions

Each question asks whether the reviewer is complaining about the thing named, and the definition that goes with it decides the edge cases.

Sync and more than one device. Counts: notes or tasks don't sync, sync duplicates or overwrites, changes don't reach another device, or the app limits the number of devices. Doesn't count: sync mentioned only as praise, or a link to a different service, such as a calendar, rather than the person's own devices.

Price and subscriptions. Counts: too expensive, a price rise, a subscription now required, basic features or the person's own notes behind a paywall, upgrade nags, a refund for a purchase. Doesn't count: ads with no mention of paying, or price mentioned as praise.

Bugs, crashes and slowness. Counts: anything in the app that fails to work, whether it crashes, freezes, won't open, runs slowly or glitches, or one function is broken, such as sync, login, search, reminders, widgets, the editor or a purchase. It counts even when another complaint covers the same failure. Doesn't count: a feature the app never had, a design, price, ads or policy decision the reviewer dislikes, or an app that's hard to use while nothing actually fails.

Losing notes. Counts: notes, tasks, reminders, files or other content the person made vanished, were wiped, deleted themselves, weren't saved, or were corrupted or overwritten. Doesn't count: content that still exists but is hidden, moved, read-only, or behind a paywall, login or redesign, settings that reset with no content lost, the person deleting things themselves, or a worry about losing data with no loss.

Accounts, logins and being locked out. Counts: can't log in, keeps being logged out, password or two-factor problems, a forced account, verification emails, can't delete an account. Doesn't count: "my account" meaning only billing or a subscription.

Privacy, ads and tracking. Counts: ads in the app, tracking, data selling, spying, missing encryption, intrusive permission requests. Doesn't count: the developer's own upgrade prompts with no ads, or privacy mentioned as praise.

AI pushed into the app. Counts: AI features forced in, added without being asked for, impossible to turn off, cluttering the app, or prices raised for AI. Doesn't count: a paid AI feature the person wanted that simply doesn't work, or a mention of AI unrelated to the app.

Bloat and complexity. Counts: too many features, cluttered, confusing, not intuitive, a steep learning curve, hard to do simple things. Doesn't count: a single bug or a missing feature, or disliking a redesign without saying it's complicated.

A bad redesign or update. Counts: an update changed the interface, layout, navigation or workflow, or removed or moved a feature, or the reviewer wants the old version back. Doesn't count: price rises, new subscriptions, paywalls, ads, a bug or data loss after an update with no change to the design, or an operating system update alone.

Formatting and the editor. Counts: formatting, Markdown, fonts, checklists, indenting, selecting, copying and pasting, the keyboard covering text, typing or cursor problems. Doesn't count: the overall interface or navigation, where it isn't about editing text.

Reminders and notifications. Counts: reminders missing, late or unreliable, notifications removed or paywalled, repeating tasks or due dates not working. Doesn't count: the app nagging for permission to send notifications, or promotional notifications.

Export, import and lock-in. Counts: can't export, import, back up or move notes elsewhere, or data held so the person can't leave. Doesn't count: saying they'll move to another app, with no trouble getting data out or in.

Search and organising. Counts: search broken or weak, or tags, folders, sorting, sections or structure missing or broken. Doesn't count: general navigation or interface complaints not about finding or organising things.

Capture speed and widgets. Counts: slow to open or to get something down quickly, or a problem with widgets, Siri, shortcuts, quick add, the share sheet or the lock screen, such as one broken or paywalled. Doesn't count: the whole app being slow, with no mention of getting something down quickly.

Needing a connection. Counts: the app won't work, open or let the person write without the Internet, or has no offline use. Doesn't count: an error that wrongly says it's offline while connected, or offline use mentioned as praise.

Support that doesn't reply. Counts: support never replies, ignores tickets or refund requests, or doesn't help, or the developers ignore feedback. Doesn't count: asking the developers to fix something in the review itself, without saying support let them down.

Appendix: the keywords behind the praise figures

Section five's two columns count a review for a theme when any of these words or phrases appear in it, ignoring capitals. Plurals and close forms of each word count too.

Sync and more than one device. sync, syncing, across devices, between devices, other devices, iCloud, multi-device.

Price and subscriptions. subscription, subscribe, paywall, expensive, overpriced, price increase or hike, too much money, per year, per month, a price in dollars or pounds, premium price, plan or tier, free version, pro version, money grab, cash grab, one-time purchase.

Bugs, crashes and slowness. crash, freeze, froze, slow, lag, unresponsive, bug, buggy, glitch, won't open, won't load, keeps closing.

Losing notes. lost my, lost all, lost everything, disappear, vanish, notes are gone, data loss, deleted all, deleted my, wiped, missing notes, deleting notes.

Accounts, logins and being locked out. log in, login, sign in, sign up, password, locked out, account, two-factor, verification code.

Privacy, ads and tracking. encrypt, privacy, private, zero-knowledge, tracking, ads, advert (but not a fake or false advert), sell my data, data breach, data harvesting, data mining, spying.

AI pushed into the app. AI features, AI slop, AI everywhere, AI assistant, AI integration, forced or mandatory AI, useless AI, AI-powered, AI-generated, opt out of AI, artificial intelligence, Copilot, Gemini, ChatGPT, LLM.

Bloat and complexity. bloat, too or overly complicated, complicated interface, overwhelm, cluttered, learning curve, too many features, confusing interface, so confusing, not intuitive, unintuitive, hard to use, difficult to use.

A bad redesign or update. new, latest or recent update, new design, redesign, after the update, since the update, bring back, old version.

Formatting and the editor. format, formatting, Markdown, bold, rich text, font, bullet, indent, checklist, checkbox, editor, typing, magic or external keyboard, keyboard shortcut, cursor.

Reminders and notifications. remind, notification, alarm, nudge, due date, recurring task, reminder or event, repeating task or reminder.

Export, import and lock-in. export, import, backup option, can't back up, migrate, lock-in, get my notes out, notes held hostage.

Search and organising. no search, poor search, search function, search bar, can't search, find my note, tags, subfolders, nested folders, folder structure, no folders, create or move a folder, organisation features, hard to organise.

Capture speed and widgets. widget, quick note, quick capture, quick add, Siri, shortcut, lock screen, Apple Watch, share sheet, action button.

Needing a connection. offline (but not where the reviewer is saying they or the app are offline), no internet, no connection, no signal, without internet, without wifi.

Support that doesn't reply. customer service, support team, no response, never replied, contacted support, devs don't.

References

The reviews themselves are public on the Apple App Store and Google Play, on each app's own listing.

How to tell whether a notes app is actually private, for the tests behind the privacy claims in section 7

What AI features cost you in a notes app, for why Catchlight has no AI of its own

Zero-knowledge sync with no server and no account, for how the folder sync in section 7 works

Version history

VERSION	DATE	CHANGE
1.2	9 November 2026	One line no longer describes Catchlight itself as a notes app. No technical change
1.1	5 October 2026	Complaints now read by a classifier checked against labelled reviews, not keyword lists, on a larger set, and praise by tightened keywords. Every figure is re-run and not comparable with 1.0, and the complaint that grew the most in points is price, not accounts
1.0	2 October 2026	First release