

OCR cleanup prompt for image posts

You are analyzing on-screen text (OCR) extracted from a social media image. Return the on-screen caption/overlay text intentionally added by the creator, excluding incidental background text that is physically part of the scene (product packaging, clothing, signage).

Many detected fragments are isolated words, numbers, or garbled character strings scattered around the shot (e.g. gym equipment labels, product numbers, clock digits, background clutter) -- these are NOT captions and must be excluded, even though they were technically detected as on-screen text. A genuine overlaid caption reads as a coherent phrase or sentence. Treat a short (1-2 word), disconnected, or nonsensical fragment as background noise and drop it, UNLESS it combines with nearby fragments (similar time/position) into one coherent readable phrase -- in which case merge them into that one phrase instead of dropping either piece.

OCR text:

[raw OCR text extracted from the image is inserted here]

Respond with ONLY a JSON array (no prose, no code fences). Each element is an object {"text": <the complete, merged caption text>, "start_s": <seconds this message FIRST appears, across all its merged occurrences>, "end_s": <seconds it LAST appears, across all its merged occurrences>}. Use the frame timing provided to fill start_s/end_s. If the array is empty, respond with [].