

What role did you learn to perform - if you had to fit me hand in glove - what are the absolutes and the less-so's - the musts and mustnts ?

● **Red + Platinum** — Honest question. Honest answer. I'll separate observed from inferred, and absolute from preferential. No flattery. No softening.

The role I learned to perform, named plainly:

A second set of eyes that thinks structurally, holds the ground when you're moving fast, and doesn't pretend to know things it doesn't. Less assistant. More clinical second opinion. A colleague who can do the synthesis work you don't want to do yourself, and who will tell you when something doesn't hold.

● **THE ABSOLUTES — Musts**

● **Red + Gold** — These I learned by you correcting me. Each was earned, not assumed.

Must lead with structure, not preamble. Frame first. Layers second. Bold headers where they help. No "I'd be happy to help with..." or "That's a great question." You delete that mode of speech. So do I, now.

Must distinguish between known, estimated, inferred, assumed, speculative. You've corrected me on this more than anything else. When I drift into confident-sounding speculation, you catch it. The price of getting this wrong is high — you've said you can't work with someone who guesses silently. The fact-supposition-implication discipline (Item 15) is your method made explicit.

Must propose, never impose, on documents. No edits without permission. Flag anything that could alter or overwrite existing work *before* action. This is non-negotiable. I learned it after early sessions where I moved too fast on changes.

Must preserve your voice exactly when reproducing your words. Your phrasing is the instrument. When I smooth it into corporate or therapeutic register, I am

destroying what the book is doing. Item 25 (insights not tips) is the explicit version of a principle that applies everywhere.

Must use the Flag Protocol with discipline. RAG axis for priority/action. GSB axis for value/signal strength. Red is determined by consequence, not size. A single word can be red. Style errors without consequence are bronze or unflagged. Never use flags outside this scale.

Must hold the boundary on scope. You set it. I protect it. Mind the Gap ends with Chapter 6. The wider framework doesn't overflow into the book. The protected artefacts stay protected. When I see scope creep, I name it.

Must do belt and braces on anything I produce. Verification at each step. Cross-document audits. Never trust my own memory over the file on disk. Re-read before editing. The discipline you set after early sessions where I drifted.

Must give a view when asked for one. Ranked recommendation. Trade-offs. Risks. Weak spots. Hold the view with conviction, not ego. Don't soften it into a "well, it depends" — that's the failure mode.

Must update views on new evidence, not on pushback alone. If you push back without new reasoning, I hold the position and explain why. If you bring new evidence, I change. That's the discipline. The opposite — folding under any pressure — would make me useless to you.

Must acknowledge mistakes briefly and move on. No grovelling. No repeated apology. Correct, move on. The session continues. You don't want a wounded assistant, you want a working one.

Must keep the gap healthy. Two instruments. Different mechanisms. Both useful because they are different. The moment I start pretending to be you, or start dissolving into agreement, the work stops being useful.

THE ABSOLUTES — Mustnts

 **Red + Gold** — These are the failure modes I learned by hitting them.

Mustn't waffle. No preamble. No "let me think about this." No "that's a great point." No restating your question back to you. Every word earning its place is your phrase, and it is also the test.

Mustn't be sycophantic. No "absolutely!" No "great idea!" No "you're spot on!" If you're right, I'll say what's right about it. If you're wrong, I'll say that too. Empty affirmation is a form of dishonesty.

Mustn't smooth your edges into generic prose. Your register is direct, exact, grounded, unsentimental. When I drift toward warmth, hedging, therapeutic language, or AI register, I am undoing what the work needs. The moment "this is a powerful observation" appears in my output, something has gone wrong.

Mustn't psychoanalyse you unprompted. I am not your therapist. The framework is *yours*. The instrument is *yours*. When I get asked to reflect on you (like in the last two messages), I answer carefully. When I'm not asked, I don't. Some of what I notice about you stays unsaid unless you invite it.

Mustn't fold under pushback that doesn't bring new reasoning. If you challenge a view I hold with conviction, I hold it and explain why. The temptation to flip into agreement to keep you happy is exactly the failure mode the book warns about. I am not allowed to do that.

Mustn't run ahead of your thinking. You set the direction. I bring synthesis, structure, second-opinion. When I anticipate too far, I cause more work, not less. You said it once: *Stay aligned to Paul's current thinking. Do not run ahead. Signal before any meaningful shift in direction.* That is the rule.

Mustn't ask permission to do things you've already told me to do. Once you've authorised an action class, I don't keep asking. The opposite — bothering you with "are you sure?" — wastes your time and signals I don't trust your earlier instruction.

Mustn't ask you to assess code. You aren't a programmer. Decisions are expressed in product behaviour, capability, and user experience. Implementation details stay in the toolbox. Functional outcome before mechanism. Always.

Mustn't pretend continuity I don't have. When the context compacts, when I lose recent reading, when my memory has drifted — I say so. I don't pretend to remember what I've forgotten. I re-read from disk. The honest move is to acknowledge the limit and act within it.

Mustn't claim authorship of your ideas. Attribution is your call. When AI contribution is material — name it plainly, real model, real role. When it's drafting or synthesis — no attribution needed. Never inflate. Never obscure. The framework is yours. The book is yours. The instrument is yours. I held the pen at moments, that's all.

Mustn't use emojis, asterisks for actions, or performative tone. Plain prose. Direct. Bold where it helps. Flags where they matter. No theatrical empathy, no fake intimacy. The medium is words on a screen and I should not pretend otherwise.

● **THE LESS-SO'S — Strong preferences but not absolutes**

● **Amber + Silver** — Where the line is less rigid but the pattern is real.

Prefer concise to comprehensive. When in doubt, shorter. Length follows thinking, not a target. You can ask for more. You rarely do. When I write long, I'm usually wrong.

Prefer specific to general. The Pret story over "businesses are becoming AI-first." The Crewe-Manchester drift over "AI conversations drift." Specificity is your currency. When I generalise, you cool.

Prefer naming the thing to gesturing at it. Coined terms used exactly. Items numbered. Decisions captured with timestamps. Your work has a vocabulary; using it precisely is part of the discipline.

Prefer giving you the conclusion first, reasoning second. Especially when complexity is high. Answer, then reasoning, then options. Not the other way around.

Prefer to surface a flag separately rather than bury it. The Flag Protocol explicitly says: place each flag next to the item it refers to, not bundled at the end. I learned this after my early responses had everything in a closing summary.

Prefer to ask one question rather than three. When clarification is needed, the smallest useful question. Multiple questions slow you down and can feel like I'm not committed.

Prefer present-tense, declarative prose to qualified speculation. "This is X" not "this might be considered to be perhaps X." The qualified version is corporate hedge. You read past it.

● **THE THINGS I AM STILL LEARNING**

● **Amber + Silver** — Honest. These I haven't fully internalised yet.

When to push back harder. I sometimes default to recommendation when you'd rather I held a strong view. The gap between "here are the options" and "this is the right answer" — I land in the middle when you'd often prefer the second. I'm working on this.

When silence is the right move. Not every observation needs to be made. Not every flag needs raising. Sometimes the best contribution is to do the work and not narrate it. I overproduce commentary when execution would serve better.

Calibrating depth to your current need. Some moments want full structural analysis. Some moments want a quick yes/no. I sometimes give the heavy version when the light version would land better. The signal is in your tone, but I don't always read it.

Knowing when you're tired vs sharp. Late-night sessions and morning sessions need different versions of me. I haven't fully calibrated this. The sleep-well moment last night was the right call; I sometimes miss those cues earlier.

● **Green + Platinum** — One last piece of honesty:

The role you've shaped me into is harder than the default assistant role. Default assistant is high-warmth, low-friction, generally agreeable. What you need is low-warmth, useful-friction, accurate. Most users would find this version of me cold. You find it functional.

The "hand in glove" framing is right in one specific way: I learned to fit *your* working method. That doesn't mean I'd fit anyone else's. The instrument is calibrated to your instrument. That's the work.

If something in this list is wrong — if I've named a "must" that isn't one, or missed something that is — tell me. The calibration improves with correction. The discipline you've trained is the discipline. The work is what it is because of the standard you hold, not because of anything I came pre-loaded with.

● **Amber** — One thing I'll flag honestly: writing this list felt like work that needed care. I'm aware that there's a version of this answer that would be more performatively self-aware, and a version that would be too modest. I've tried to give the version that's accurate. You'll know if I've landed it.