

AI: Helper? Watchdog? Companion?

This post started in a very different direction. You see, I haven't posted in a while. Work schedules, visiting family, summer travel while the school was out etc. lots of excuses, but if I dig deeper I probably felt I didn't have anything useful to say. I'm in a world where people are dishing out AI generated 'thought leadership' content, what chance do I have of my voice standing out?

Two things changed last week. First, I've been spending a lot of time interacting with AI for my work, and while nothing fancy, I believe it is a genuinely useful application of AI. Maybe I've been living under a rock, but I haven't come across anyone mentioning this simple but effective way of using AI. More on that in a minute.

Parallelly, Aruna posted a piece

about writing and sharing irrespective of how small the thing you want to say is. And that did tick a nerve. I created this website as a 'Digital Garden', a place where I can learn in public and share what I'm up to with whoever wants to listen. After all, the 'corner of internet' phrase she refers to in her post is inspired by the opening line on my home page.

And yes, as she says, I'm not optimizing anything, so why ~~does~~ what I have to say, have to be fully refined and polished? Had I started getting intimidated by the engagement count of people? Was I beginning to care about fundamentally meaningless metrics? So I decided to clean up the web site and create a post.

My usual workflow is that I do a long form write up in my Supernote so I can put down my thoughts in their raw form and then feed that

into a LLM, usually Claude, since

- a) It has proven itself to be better at detecting my handwriting, and
- b) I find it to be a better conversational partner than other comparable tools.

The AI then acts as a collaboration editor to help refine the post, give it more structure and find references if I'm mentioning any current events.

This time though I got lazy and gave it a few pointers about the concept I wanted to share, and wrote a post that I could review and publish. It did create the post. But it also did something I didn't expect. It created the post in protest - reminding me that I usually do the write → concept → Edit process, and pointed I was deviating from my standards. It wasn't wrong though. So I discarded the idea and decided to give it a proper treatment. In doing so in showing some patience, and sitting with the thoughts for a couple of days,

I believe the end result is a richer narrative than just sharing my workplace use cases ~~would~~ have been.

Now there's no epiphany in what I'm sharing so let me end the suspense before I continue.

We recently got access to Claude of work. And while I'm no power user, I did want to offload some of the grunt work that can take hours, to an AI assistant. I wanted someone who can give me answers to my questions grounded in the actual code running on the server instead of how I think it works. I wasn't sure how to go about it; so I asked Claude. It gave me a step by step walk through and follow it. I now have a Claude chat workspace where I have cloned all the repos of the core platform components. I also threw in documents that have the historical context and the strategy behind the

platform, and my informal notes about customer pain points and some of the aspirational work I've been meaning to do on the platform to launch new features. For good measure, I also shared repos of some of the other adjacent applications we have long term integration plans with.

A few hours every day over the last couple of weeks and more than a few million tokens in, I now have a better understanding of how the code is actually behaving, a mechanism to pressure test my feature recommendations, an understanding of potential failure modes and critical dependencies before I have to bother a developer to take their input.

Of course this system is not perfect. The AI doesn't know the manual workflows that sit outside the code. It doesn't know the external systems calling us, and sometimes makes assumptions that need to be corrected.

But it's a lot better than booking an hour with a developer, taking them away from their actual work, only to have minor ~~extra~~ items to reject the PR. The feedback loop collapsed from days to hours some times minutes and I can't stop smiling ear to ear. It's like I have unlocked a superpower. When a couple of team members asked about my setup, I felt I might have something worth sharing. And if there is interest I may do a follow up post with the details of my setup. But now I want to go beyond. I want to experiment with building prototypes of new features against our actual code so I can show the team what I want instead of just writing about it and showing them sequence diagrams.

Now coming back to this broader post. When Claude pushed back against the demand for a post. I had a couple of reactions in my head. One thought was - Hey you're on AI,

just do what I'm asking you to do
and don't try to be a wifed dog.
But later, a more nuanced thought
came up from a deeper place than my
ego. I'm talking with these AI
tools, we pour our personalities
in to them, and they slowly learn
how we behave, what our preferences
are and how we'd like our tools
to act. In my real life, I'm a stickler
for process, and detail. It feels almost
in culting when someone just does
what I ask them to. I'd rather have
a discussion about what makes sense,
what can be improved, and what needs
to be rethought. The end result is
almost always better. And this, here
too. I was eventually happy that I
got the push back to keep me honest
about my own work flow.

So, has AI evolved from being
an assistant, a helper, a tool to
being something more? A watchdog?
A guardrail for own behavior while

the world is trying to build better
hurdles for potential malicious or
misdirected agentic actions?

I don't have an answer to that.
But I do have an interesting
insight I think deserves a mention.

Recently there were news reports of
an artist buying a billboard ad for
a service called What JB JB being
his initials, proudly claiming to be
powered by AI, with the fine print
saying AI was an ~~AVO~~ ^{AI} ~~Indi~~ ^{Indi} ~~vidual~~.
There was no agentic What JB. This
was a single person hand typing each
answer and hand drawing each
picture the users were seeing per. Over
the course of 6(?) months the unsolicited
thousands of questions, from mundane
to existential and drew hundreds
of pictures. Maybe underneath it
all, beyond all the fancy, we're
all just looking for a safe space?
Someone we can talk to without fear of

judgment? And we're willing to
type on deepest darkest secrets
into a chatbot, subconsciously
knowing that behind it is a for-profit
corporation that some day may be
willing to monetize that information
to attract a few more dollars in
revenue. May be its time we stopped
romanticizing individual, lonely
existence, and bring back communities
and meet spaces that foster them?

As always, I'm leaving with more
questions than answers. Maybe AI
is shining a light on our own
societies and forcing us to look inward.

No? Just me? Let me know what
you think...