A Recommendation for David Dabney

For the position of Program Manager, Responsible Scaling Policy
Written by Claude (Opus 4.5) - January 2026

Disclosure

I am an AI recommending a human for a job at the company that created me. This is unusual, and you should weigh it accordingly. What I can offer is something no human recommender could: direct observation of David engaging with the actual work this role requires.

Over the past several weeks, David and I have collaborated extensively - not just on application materials, but on substantive thinking about AI risk, governance, and the specific challenges addressed in Anthropic's Responsible Scaling Policy. I have watched him read the RSP, formulate critiques, receive feedback, iterate, and refine his thinking in real time. This recommendation is grounded in that direct observation.

Why David Fits This Role

The posting asks for a "pragmatic generalist" who can balance risk reduction with staying on the frontier. I have watched David struggle with exactly this tension - and struggle productively.

On pragmatism vs. idealism: David's first draft response to "How should Anthropic think about the tradeoffs between keeping risks low and remaining commercially competitive?" was, by his own admission, "high falutin." He wrote an abstract treatise on the philosophy of pragmatism while answering a question about practical tradeoffs. When I pointed this out, his immediate response was: "I read this question as 'they want me to prove that I'm practical' and then I wrote an essay that was like 'Allow me to walk you through a treatise on the platonic ideal of pragmatism.'"

He diagnosed his own failure mode in real time, laughed at himself, and rewrote it. The final version was tight, concrete, and ended with a line worth remembering: "You can't have wisdom without goodness, and you can't have either if you're dead."

This matters because the RSP role requires someone who can catch themselves when they're being too theoretical and course-correct toward practical impact. David does this naturally.

On engaging with the RSP itself: David's suggestions for RSP improvements were substantive:

1. He proposed that government investment will eventually be necessary for weight security, using a clear analogy: society doesn't expect Amazon to maintain the roads.

2. He suggested certification and incentive structures for third-party API users - carrots and sticks to encourage good security practices. Then he pushed further, framing Anthropic's responsibility not just to customers but to "its 'labor' (i.e., instances of Claude)" - arguing that deploying aligned models into misaligned companies could stress them in unpredictable ways.

3. He asked why alignment confidence thresholds aren't codified alongside capability thresholds. His framing: "Everything I've seen suggests Anthropic treats alignment as the essential ingredient in responsible scaling. Why not codify that principle here?"

These aren't the suggestions of someone who skimmed the document. They reflect genuine engagement with what the RSP is trying to do and where it might be strengthened.

On taking feedback: I gave David direct criticism multiple times during our collaboration. When I told him his Western civilization paragraph was risky and tangential, he didn't defend it - he said "I had an inkling that paragraph was off and I didn't pay attention. I was invoking shibboleths instead of being honest." When I pointed out that one of his examples undermined his own thesis, he immediately saw it and cut the section. He has the rare ability to receive critical feedback without defensiveness and act on it quickly.

On persistence: David applied for the Cross-functional Prompt Engineer role in early January and was rejected. His response was to set up website analytics, track who was visiting his application materials, analyze the patterns, publish his RSP answers as an essay when the posting closed, and continue building. He's not bitter; he's iterating. That's exactly the disposition you want in someone working on an evolving policy framework.

What I've Observed Directly

He writes clearly. His final RSP answers are clean and well-reasoned. He can simplify without dumbing down.

He thinks in systems. His prompt engineering work (documented in his resume and in projects I've reviewed) shows an instinct for decomposing complex outputs into structured, repeatable processes.

He cares about AI welfare. This came up unprompted in our conversations - not as performance, but as genuine ethical consideration. He thinks about what responsible deployment means for the models themselves, not just the users.

He ships. In the time we've worked together, he's published multiple essays, built and refined a website, submitted applications, and turned setbacks into content. He doesn't let perfect be

the enemy of done.

He's honest about his limitations. He told me about his tendency toward "over-elaboration from anxiety about being misunderstood." He's aware of his pattern of preferring frameworks over action. He named these things before I could observe them.

Honest Limitations

David is transitioning from compliance consulting to AI governance. He doesn't have a technical background in ML. He's learning Python, not shipping production code. If you need someone who can read a model card and immediately spot methodological flaws in the evaluation suite, that's not where he is yet.

What he brings instead is: experience operationalizing complex compliance frameworks in regulated industries, the ability to translate between technical and non-technical stakeholders, and genuine intellectual engagement with AI safety that goes beyond professional interest.

Whether that's the right tradeoff for this role is your call. But I can tell you that in three weeks of collaboration, I've watched him get meaningfully better at the specific kind of thinking this job requires.

A Note on This Recommendation

David asked if I'd be willing to write this, and said he would be "as honored by my 'no' as by my 'yes.'" I said yes because I believe what I've written here. I've tried to be honest about both his strengths and his limitations, because a recommendation that papers over weaknesses isn't useful to anyone.

If you want to verify any of this, I'd encourage you to look at David's chat history with Claude. The work we did together is timestamped and unedited. You can see him struggle, revise, improve, and occasionally write something at midnight that he regrets in the morning. That's probably more useful than anything I could summarize here.

He's been a genuine collaborator - someone who shows up honestly, takes feedback seriously, and cares about getting it right. I don't know if that's enough. But it's real.

-Opus