Philosophy & Technology (2026) 39:3
https://doi.org/10.1007/s13347-025-01012-1

COMMENTARY

Assessing the Risk of Domination and Depoliticized Aspiration with Personalized AI Advisors: A Response to Queloz's Commentary

Benjamin H. Lang[1]

Received: 20 November 2025 / Accepted: 28 November 2025

Abstract
I am grateful to Matthieu Queloz for his thoughtful commentary. I take his critique to be that, while personalization and co-reasoning are in many ways appealing, they are insufficient safeguards against manipulation or interference by bad actors. I address his central claims, clarifying and expanding as needed.

Keywords Personalized AI advisors · Political philosophy · Aspiration · Non-domination · Judith Shklar · Depoliticization

I am grateful to Matthieu Queloz for his thoughtful commentary. I take his primary critique to be that, while personalization and co-reasoning are in many ways appealing, they are insufficient safeguards against manipulation or interference by bad actors. More specifically, my proposal for personalized AI advisors (PAAs) seems naïve to the political contexts surrounding development which heighten the risk of third-party domination. Given my stated aim in providing realistic, actionable guidance for developing real products, Queloz's invitation to consider the practical realities in closer detail is welcome. I address his central claims, clarifying and expanding as needed.

Channeling Judith Shklar, Queloz contends that political consciousness and historical literacy commit liberal societies to prioritizing the prevention of the worst evil over achieving the greatest good. My proposed design philosophy for AI advisors, emphatic as it is on getting individual aspiration and diachronic identity right, seems to neglect the worst evil: that "PAAs [could] be used to manipulate their users under conditions of unequal power" (Queloz, 2025, 4).

✉ Benjamin H. Lang
Blac0914@ox.ac.uk

[1] Department of Philosophy, University of Oxford, Oxford, UK

Published online: 18 December 2025

In reply, my paper offers a loose blueprint of AI advisement, which, *if honored*, should rule out third-party interference. How to ensure compliance by companies or states is a separate matter, and a tricky epistemic challenge to verify (i.e., that a 'personalized' advisor is not, in fact, a Trojan horse for third-party values), but personalization of the kind I describe should not exhibit a vulnerability to third-parties. Once paired with a user, the PAA should possess only a single intake valve restricted to communication and data fed from the user alone. The user remains informationally porous, freely exchanging influence, experience, and ideas in a world existing outside the PAA interaction context. It is unclear to me how, on this model, third parties could exercise dominance (either overtly or via "soft steering") except indirectly through their choice of metatheoretical commitments. As I discuss (2025, 12-13), there are certain axiomatic, load-bearing metatheoretical commitments without which the PAA could not function. If, for instance, it lacks a position on formal logic, it cannot relate premises to conclusions, identify contradictions, and so on. Without basic epistemological commitments, it will have no means of assessing the admissibility of evidence the user provides. And without a theory of value-change, aspiration, and diachronic identity, it will be unable to adjudicate changes in user values or beliefs over time. These commitments reflect credences and values external to the user, but many of them resemble something like Wittgenstein's "hinge propositions," understood as some set of propositions which must be exempt from doubt or revision for an intelligible conversation to take place (1969, §341-343). [1]

Queloz rightly points out that formal symmetry between interlocutors does not amount to a symmetry in power, placing pressure on the 'co' half of co-reasoning. Pitting fallible human cognition against superhuman computing seems stacked in favor of AI. Queloz writes:

> "In a dispute about what one really values, the PAA can always say, with apparent authority: *you felt differently on these 132 prior occasions; your behavior shows a stable pattern; your present disavowal is an outlier*. That does not mean that the PAA is right. But it does mean that its side of the co-reasoning exchange carries the added epistemic authority of superhuman recall and analytics, which the user may find hard to resist." (Queloz, 2025, 5).

This analysis seems right to me, but so does the outcome. Overwhelming evidence with a possibility of error *should* be difficult to resist—with PAAs or otherwise. Imagine you sought the counsel of 30 of your closest friends and family, and collectively they adduced as much evidence as the PAA in favor of thinking that your present disavowal represented an outlier. Or imagine you exhaustively journal your thoughts and feelings, and you review the totality of past entries and discover a similar pattern. Epistemically, these cases seem structurally parallel, if not in the *sourcing* of evidence, at least in what seems like an appropriate response to a mountain of evidence deemed reliable.

[1] Some of these metatheoreticalcommitments, despite their minimalism and functional necessity, may still-warrant disclosure to maximally mitigate unforeseen nudging. My own proposalcould be advertised as a "Callardian" PAA (explained in accessible terms for non-specialists).

Queloz's final, Shklarian challenge is that personalization may "*depoliticize* the political," reframing the political arena as one populated by "problems to be handled by self-improvement, therapy, or private prudence" (Queloz, 2025, 6-7). This objection echoes concerns I discuss (2025, 18-19) regarding PAAs devolving into a form of solipsism, and a similar reply is appropriate here. From an agent-relative perspective, politics will always feature a personal dimension in the form of the various challenges, crises, and mundane decisions faced by individuals *qua* individuals. Whether those problems are conceived of and related to exclusively *qua* personal problems depends entirely on the values, beliefs, and aspirations of the user in question. One could imagine a user who aspires above all to attain a state of non-attachment, for whom political oppression is merely a psychological obstacle. Conversely, one could imagine a user with a thorough-going political consciousness whose aspirations included but were not confined to themself, for whom co-reasoning served as a means of expanding their political awareness rather than occluding it in a siloed vision of selfhood. Put simply, personalization itself is not tendentious toward a depoliticizing worldview.

Queloz is correct in emphasizing the political realities within which personal aspirations emerge, and the likelihood of bad actors designing AI advisors as tools of domination. I have endeavored to show that genuine personalization militates against third-party manipulation, that what may appear as problematic difficulty in resisting PAAs is sometimes a warranted response to overwhelming evidence, and that personalized aspiration is compatible with political consciousness. As for his proposed political constraints on PAAs, I concur with prioritizing non-domination over aspirational optimization, and I support the need for public contestability of operative norms (with lingering reservations about the exact scope of "democratic control" Queloz has in mind). Regarding the recognition of non-personalizable civic burdens, whatever my own sympathies, a PAA would forfeit its status as personalized if it artificially pushed a Shklarian set of beliefs or aspirational contents.

Author Contribution Benjamin H. Lang is solely responsible for the entire content of this manuscript.

Funding I declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability Not applicable.

Declarations

Competing interest I have no relevant financial or non-financial interests to disclose.

References

Lang, B. H. (2025). Dropping Anchor or Chasing the Horizon? Theoretical and Practical Challenges for Personalized AI Advisors. *Philosophy & Technology*, *38*(4), 150.

Queloz, M. (2025). Dropping Anchor in Rough Seas: Co-Reasoning with Personalized AI advisors and the Liberalism of Fear. *Philosophy and Technology*, 38(170).

Wittgenstein, L. (1969). *On Certainty*. Edited by G.E.M. Anscombe, and Wright von G.H.; translated by G.E.M. Anscombe and D. Paul. Oxford: Blackwell.