

## TRUSTING EXPERTS?

– Ezio Di Nucci –

**Abstract:** This paper argues against the claim that “expertise requires trust.” It does so by distinguishing between two versions of this claim, one according to which expertise requires *trustworthiness* and one according to which expertise requires *actual* trust. The paper then argues that the former version of the claim is obvious and therefore philosophically uninteresting, while the latter version of the claim is, actually, false. Five cases are deployed to defend these arguments, so that along the way we find out more about the nature of expertise, trust and also, incidentally, corruption and even ChatGPT.

**Keywords:** expertise, trust, trustworthiness

**Submitted:** 3 June 2024

**Accepted:** 19 December 2024

**Published online:** 27 February 2025

### Introduction

Does expertise require trust? This short paper argues for a negative answer, contra Giubilini et al (2025). Let us first understand the claim according to which “expertise requires trust” (otherwise said: trust is a necessary condition for expertise). The claim’s truth conditions are not obvious, at first glance: arguing that expertise requires trust appears to mean, at least, that you are not an expert if... and here we already have two plausible alternatives:

- i) you are not an expert if you are not trustworthy;
- ii) you are not an expert if you are not trusted.

The latter version of the claim needs itself further analysis in terms of specifying the subject of trust – you being the object here: are we talking of the general public, for example; or are we talking of a relevant group of peers; or maybe just policy-makers? Let us take this ambiguity of the latter version of the claim as a valid methodological reason to start from the former version instead. Do you need to be trustworthy in order to be an expert? We must get another ambiguity out of the way, namely what the “you” might refer to ontologically. Minimally, we ought not to commit ourselves to the narrow claim that the “you” here can only refer to human expertise, ChatGPT-4 might also be a candidate, say. This is, by the way, a different point from arguing that the eventual success of the conceptual relation between expertise and trust should depend on whether it is applicable beyond human experts (or, indeed, AI ones). The preliminary move is

---

Ezio Di Nucci  
University of Copenhagen, Copenhagen, Denmark  
(ezio@sund.ku.dk)

just to be inclusive about possible applications while evaluating the different versions of the original claim independently.

### **Expertise and trustworthiness**

Back to the former version of the “expertise requires trust” claim, according to which you cannot be an expert if you are not trustworthy. Here the paper argues that, on this version, the claim “expertise requires trust” is either obvious or meaningless (the latter because tautological or quasi-tautological). That’s just because expertise is what makes you trustworthy. One step at the time: can you imagine an expert who is not trustworthy?

Here are some candidates:

1. “true” expert, no formal qualifications;
2. “true” expert, in pharma’s pocket;
3. Omniscient creature;
4. Opaque ChatGPT-17;
5. Nerd.

We could continue but let us take these five candidates as representatives of some important disjunctions when it comes to expertise and trust: having formal qualifications, being independent, knowledge, transparency... oh, and being human. Here we need to distinguish between two different questions, which we will illustrate with the first candidate: the question of whether you need formal qualifications in order to be an expert, and the question of whether a “true” expert with no formal qualification is trustworthy. What we care about here is the latter question, because a true expert who is not trustworthy might constitute a counterexample to the former version of the claim that “expertise requires trust,” namely the subclaim that an expert must be trustworthy. Obviously, even though the questions are importantly different, they are related in the following way: one might argue that without formal qualifications, there is no expertise. But that’s just implausible, because it conflates epistemological considerations with ontological considerations – namely, formal qualifications are just (supposedly reliable) evidence of expertise, not expertise itself. Having ruled out the strong claim that formal qualifications are necessary for expertise, what about the idea that someone without formal qualifications might still be a true expert but not a trustworthy (true) expert, thereby providing the counterexample we were looking for? With this weaker claim, the conflation of expertise with evidence of expertise sounds less problematic, because the idea would be that the lack of formal qualifications affects your trustworthiness without affecting your actual expertise. Here we may consider a more obvious alternative, though – which is just to remind ourselves of the difference between the two versions of the “expertise requires trust” claim, so that instead of concluding that lack of formal qualifications affects your trustworthiness, we rather conclude that lack of formal qualifications affects whether or not you are actually trusted, not whether or not you are worthy of trust. This point has two methodological functions: not just to defuse the possible counterexample, but to start illustrating the difference between the two versions of the main claim, because readers probably won’t be surprised to already find out that while the former subclaim is either obvious or meaningless, the latter – being the only plausible interpretation of

the main claim left – is actually false. Before moving to the latter version of the “expertise requires trust” claim, we should check with our other candidates, because the analysis might be different. While a true expert without formal qualifications is still trustworthy, a so-called expert in pharma’s pocket is not trustworthy, but that does not constitute a counterexample because such a person has, literally, *sold* their expertise. Or at least they are renting it out – either way, it’s gone.<sup>1</sup>

While the analysis is different, then, we still have the same outcome: no counterexample, this time because there can be no true but corrupt expert. If you are corrupt, you have relinquished your expertise, which might show that expertise is not a purely epistemological category. Not necessarily, though: because we could be able to analyze corruption in purely epistemological terms by pointing out that a corrupt so-called expert will not reliably give advice that correlates positively with evidence. Things can get quickly complicated here, because when does someone stop being a true expert, for example? When they take the money or when we find out? Or what about those more plausible so-called experts who are doing legitimate work, but their overall agenda is dictated by pharma’s funding? Sounds a lot like the difference between lying and not telling the truth, if you ask me. We don’t need to worry about these difficult questions, though, because the point of this paper is not a definition of expertise (or, for that matter, corruption or even trust), the more modest point of the paper is just the link between expertise and trust and the claim that “expertise requires trust,” and that particular claim is unaffected by the corrupt so-called expert, for the reasons we have just presented. What about omniscient creatures or ChatGPT? Let us do this quickly because we need to move on to the more interesting version of the claim. Omniscient creatures do not exist, but what about ChatGPT’s opaqueness, does that affect its trustworthiness without affecting its expertise? This is, I believe, the most promising of our candidates, as it is not as easy to dismiss the idea that a reliable version of AI can meet the requirements for genuine expertise but if it lacks transparency or explainability, then it ought not to be trusted and is, therefore, not trustworthy.<sup>2</sup> Do we finally have our counterexample?

Methodological caveat: we might not need a successful counterexample here. Having to rule out possible counterexamples might already be enough to show that the claim that “expertise requires trust” is not obviously true or meaningless, because then there would not have been any need for the work we have done so far. Still, what are we going to say about so-called AI systems, can they be untrustworthy experts? Here things are complicated both empirically and conceptually: empirically, because we at least need to have a rough distinction between LLMs (large language models) like ChatGPT and the way in which machine learning is used in healthcare, for example.<sup>3</sup> This is not necessarily for technical reasons, but because the kinds of systems that assist with, say, analyzing biopsies can plausibly be argued not to be experts because they

---

<sup>1</sup> A careful reader might here notice the similarity between the selling expertise and renting expertise distinction with the selling one’s body and renting one’s body distinction which is often used in debates around sex work. That careful reader might then take offence at the comparison between a corrupted so-called expert and a sex worker. That careful reader would be me, so nothing more to say here.

<sup>2</sup> Baum et al. (2022), Maclure (2021), Mittelstadt et al. (2019), Páez (2019) and Zednik (2021)..

<sup>3</sup> Di Nucci (2020).

provide decision-support rather than decision-making.<sup>4</sup> The idea here would be that their not being experts does not have anything to do with their knowledge or capacities but with their role and function.

The conceptual complexities are harder. When it comes to the question of whether AI systems can be considered to be untrustworthy experts, we oscillate between two ideas. On the one hand, if their advice is consistently reliable, then we ought to trust them even if we don't understand them. On the other, if we really don't know why AI systems are giving the advice they are giving, then it's not just that they are not trustworthy, but also they are useless (so-called) experts, because how are you going to use them productively in, say, the rarified political environment of policy-making, when their advice is so easy to contest because not accessible? It won't probably surprise you that the above dilemma isn't necessarily bad news for our purposes, because on the first horn, no counterexample, since these AI systems would indeed come out as trustworthy. And on the second horn, no counterexample either, since the same AI systems wouldn't be true experts – at least in a sense of expertise that does not include pure epistemic norms but also considerations of use, function and role in, say, policymaking (which starts to approach the sense of public trust that, as we see below, is the relevant one throughout). The risk for our argument is, admittedly, that once we apply our oscillating analysis to actual systems, some of them will turn out to be, in fact, in the middle, with enough epistemic power to make denying expertise less than fully plausible, while those same staggering computational capacities would make the system very hard to trust. But here we are already starting to cross the unstable bridge from trustworthiness to trust, which we will take as a symptom that it's time to move on to that version of the claim.

Let's conclude, then, by ruling out a final candidate, a brilliant nerd who is not a good communicator. This we can do quickly: this is a genuine expert who is trustworthy, and while might not be very presentable in policy circles, that should not affect judgments about their trustworthiness or expertise (still, think of the possible comparison between the nerdy bad communicator and ChatGPT: are they that different in the end?). In this section we have then analyzed and rejected five potential counterexamples to the claim that "expertise requires trustworthiness." Our discussion should have hopefully shown why that claim isn't just true, but obviously so in a way that makes it less than fully philosophically interesting.

## **Trust and Expertise**

We can now move on to the latter interpretation of the claim that "expertise requires trust," which is more controversial because it requires genuine experts to be trusted. Here is a version of this claim by Giubilini and colleagues: "To the extent that it contributes to preserving public trust, transparency about expert disagreement and uncertainty is an essential aspect of what it means to be an expert. Indeed, it can be as important as being confident in what one, as an expert, believes to be true."<sup>5</sup> This is a plausible

---

<sup>4</sup> Here see Di Nucci (forthcoming). If you are curious over my other writings on the topic of AI, please see for example: Di Nucci (2019), Di Nucci (2024), Grote and Di Nucci (2020), Tupasela and Di Nucci (2020), and Di Nucci et al. (2020).

<sup>5</sup> Giubilini et al (2025): 23.

and useful target for our discussion here because it refers to “public trust” rather than trustworthiness, making the fascinating case that true expertise requires a version of, for lack of a better word, vulnerability on the part of so-called experts: being open about disagreement and uncertainty.

Let us start by emphasizing one of the benefits of our dual analysis in terms of trust and trustworthiness: namely that it speaks to the difficult but important hypothesis that it is a public duty of experts to make sure that they are believed (whether, again, that be believed by the public, policymakers or just peers). It is only by distinguishing between trustworthiness and trust that we can tackle this hard question, because a supposed duty to be believed can clearly not be cashed out purely in terms of trustworthiness – since it must also work in practice – but at the same time it cannot be reduced only to the contingent truth of whether or not, in the end, some bit of expert advice is believed or implemented on the ground. We are, again, in borderland here: but that’s the point of the paper, to move back and forth within the unstable equilibrium of trust and trustworthiness. The following, I believe, is not implausible: an expert who does not care whether or not they are believed, that’s not a true expert. Here you might be tempted to distinguish between a smart government adviser and an awkward guy who sleeps at the lab, in order to be inclusive about expertise and to say that caring about impact is not a requirement for expertise. But remember, we are also being inclusive on the receiving end, so that we can just distinguish between different groups that different experts want to be believed by, without necessarily rejecting the principle that they should care about being believed. In a certain way, that’s their bottom line, which is a helpful metaphor to go on and argue that a so-called expert whose bottom line was their actual bottom line, that’s not a true expert, because they care about profit more than they care about trust.

What is the relationship between our claim that true expertise requires a hands-on commitment to being believed rather than just an empty commitment to trustworthiness, and the overall claim that this paper is questioning, namely the idea that expertise requires trust? It would be tempting here to conclude that our considerations around an expert’s commitment to being believed are arguments for the claim that expertise requires trust – and they are certainly not empty. Take a brilliant researcher who will only share their data with a couple of trusted colleagues even though they know (or even: are made aware of the fact) that sharing more widely would impact public trust positively. They believe, correctly, that sharing more widely would have no benefits in terms of the quality of their study – since those trusted colleagues already checked the numbers – so that their tradeoff is merely between the extra work of sharing more widely and impact on public trust. What should we make of such a scientist, true expert or not? First of all, forgive me if you mind the simple dichotomies of classic analytic philosophy, but I guess if you made it so far you don’t mind as much as some people. To the point: this case is relevant and interesting, for us, because we have a scientist who is committed to being believed by their peers while at the same time not caring about public trust (or at least not prioritizing it, remember that it is a tradeoff, as in sharing more widely does have some costs in terms of resources, say time or electricity because of the servers, or whatever) in the sense that they will not go the extra mile in order to be believed by a wider public or, say, by policy-makers. This might remind some colleagues of the classic academic

dilemma of sexy titles, which on the one hand might increase impact (readership, citation, maybe even journal's impact factor, who knows these days) but on the other hand defeat the whole purpose of academic endeavor, namely that it is boring, time-consuming and all about the evidence, so sexy titles should be irrelevant and, indeed, it might even be argued that they might reduce public trust because they show that so-called experts care about attention, so they would not just care about evidence – by definition. We have, I believe, reached the climax of our argument: is a scientist who cares about public trust a scientist who cares about more than just evidence, and therefore not a pure scientist, who might for that very reason not deserve the label “expert”? Yes, you got it, dear reader, we have produced a paradox: in order to qualify as a true expert as opposed to so-called expert, you need to care about public trust (at least according to the argument we are addressing). But if you care about public trust, then you don't only care about evidence; and if caring only about evidence is what defines quality – and thereby trustworthiness – in a scientist, then we don't just have the superficial problem that someone who cares (too much?) about public trust might not be a true expert, but we have the deeper problem – maybe even the paradox – that a scientist who cares about public trust is not trustworthy. Indeed, a stronger conclusion is warranted: they are not trustworthy, according to our argument, exactly *because* they care about public trust.

Can this outrageous conclusion really be valid? Answering this question goes beyond the scope of this short paper, so let me just remind the reader that the conclusion only follows, if it follows at all, under the premise that a true expert must only care about evidence and nothing else – and that premise is not something that we have independently argued for here. In a slogan, though, the argument wouldn't just be: evidence first. It would rather be: evidence, evidence, evidence.<sup>6</sup>

## Conclusion

Summing up, briefly: we have argued against the claim that true expertise requires trust. We have done so by distinguishing between the obvious claim that expertise requires trustworthiness and the dubious claim that expertise requires actual trust. We have concluded by speculating about a paradoxical conclusion according to which it is their very investment in public trust that makes a scientist less trustworthy.

**Acknowledgments:** Thanks to my colleague Lea de Chiffre Skovgaard and two reviewers for their comments and suggestions.

---

<sup>6</sup> Disclaimer for my local STS colleagues (thanks Lea!): you might object to this evidence-only account on two grounds. Firstly, you might resist the idea that evidence is ever “pure” or “real”; or, secondly, you might resist the idea that there is such thing as a “pure” or “real” researcher who is only guided by evidence and therefore deserves the “expert” label, the latter because flesh and blood researchers cannot, in real life, ever be guided by only one thing. The former because data and evidence are much more complex and dirtier than philosophers imagine them to be. Here I can clearly not engage with the whole STS tradition, but I just wanted to nod to that tradition by clarifying that my account can accommodate the idea that evidence is always dirty, as long as that's the only thing guiding a researcher while on the job.

**Funding:** None to declare.

**Conflict of Interests:** The author declares that he has no conflict of interest.

**License:** This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## References

- Baum K., Mantel S., Schmidt E., Speith T. (2022), "From responsibility to reason-giving explainable artificial intelligence," *Philosophy & Technology* 35 (1): 12.
- Di Nucci E. (2019), "Should we be afraid of medical AI?," *Journal of Medical Ethics* 45 (8): 556–558.
- Di Nucci E. (2020), *The Control Paradox: From AI to Populism*, Rowman & Littlefield, Lanham.
- Di Nucci E. (2024), "Too much control? Health, sex and war," [in:] F. Santoni de Sio, G. Meccacci (eds.), *Research Handbook on Meaningful Human Control of Artificial Intelligence Systems*, Edward Elgar Publishing, Cheltenham: 254–263.
- Di Nucci E. (submitted), "AI-supported Clinical Decision-Making," SSRN, URL = [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4851973](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4851973) [Accessed 04.02.2025].
- Di Nucci E., Jensen R. T., Tupasela A. (2020), "Kunstig intelligens og medicinsk etik: Tilfældet Watson for Oncology," [in:] R. T. Jensen, S. Andersen (eds.), *8 Cases i Medicinsk Etik*, Munksgaard, Copenhagen: 169–192.
- Giubilini, A., Gur-Arie, R., Jamrozik, E. (2025), "Expertise, Disagreement, and Trust in Vaccine Science and Policy: The Importance of Transparency in a World of Experts," *Diametros* 22 (82): 7–27.
- Grote T., Di Nucci E. (2020), "Algorithmic decision-making and the problem of control," [in:] M. Christen, B. Gordijn, M. Loi (eds), *Technology, Anthropology, and Dimensions of Responsibility*, Springer, Cham: 97–113.
- Maclure J. (2021), "AI, explainability and public reason: The argument from the limitations of the human mind," *Minds and Machines* 31 (3): 421–438.
- Mittelstadt B., Russell C., Wachter S. (2019), "Explaining explanations in AI," [in:] *Proceedings of the Conference on Fairness, Accountability, and Transparency*, ACM, New York: 279–288.
- Páez A. (2019), "The pragmatic turn in explainable artificial intelligence (XAI)," *Minds and Machines* 29 (3): 441–459.
- Tupasela A., Di Nucci E. (2020), "Concordance as evidence in the Watson for Oncology decision-support system," *AI & Society* 35: 811–818.
- Zednik C. (2021), "Solving the black box problem: A normative framework for explainable artificial intelligence," *Philosophy & Technology* 34 (2): 265–288.