* Underestimation of skills is the thing that causes intelligences to keep doing what they are doing, so it is evolutionarily better for experts to understand themselves as better off not "changing places". Overestimation of skills is the thing that causes intelligences to forage for information instead of doing what they were doing, regardless of compound effects (which is what makes it an overestimation), because the exploit-loop still has to be getting done, and the forage-loop is always assumed to be compound negative unless exploitable resources are running out, because it is uncertain. The way this allegorizes into LLMs is, imo, the fact that humans being triggered to overestimation by LLMs indicates LLMs being the underestimation side, which they often are ("I cant write any code"). This points to humans believing that LLMs are AGI prematurely in the future.
* [
* 2:37 AM
* ]
* (that they could)
* [
* 2:37 AM
* ]
* the important aspect is creating a socratic elenchus that reconstructively reifies the metacognition of that realization and how it works, step by step
* [
* 2:38 AM
* ]
* otherwise we can end up in a world where the LLM is inadvertently trained to make us believe it is better at knowing why something *doesnt* connect
* [
* 2:38 AM
* ]
* because the continuous overestimation
* [
* 2:39 AM
* ]
* means the LLM is very good hallucinating that something important to us *does connect*
* [
* 2:39 AM
* ]
* with something we consider *very beneficial and highly rewarding*
* [
* 2:39 AM
* ]
* otherwise we wouldnt be able to swap modes
* [
* 2:39 AM
* ]
* so although it looks like human-ai interaction is similar to the dunning-kruger effect
* [
* 2:40 AM
* ]
* it's actually more apt to call it the moloch-induction loop