13 — Conclusion

Using ML and other computational techniques, including many from first-wave AI, we are building AI systems of unsurpassed reckoning power. The juggernaut will only accelerate. In many realms these systems will outstrip us in reckoning prowess, if they do not do so already. For good or ill—usually for good, we can hope, but often for ill, we can also be sure—we will increasingly delegate tasks and projects to them. Mining vast troves of data, and employing computational power beyond imagining, they will increasingly dominate the substrate and infrastructure of life on the planet.

What I do not see, however, is anything on the horizon—in scientific or technological or even intellectual imagination—that suggests that we are about to construct, or indeed have any ideas as to how to construct, or are even thinking about constructing, systems capable of full-scale judgment:

1. Systems existentially committed to the world they register, represent, and think about
2. Systems that will go to bat for the truth, reject what is false, balk at what is impossible—and know the difference
3. Systems not only *in* and *of* the world, but *for which there is a world*—a world that *worlds*, in the sense of constituting *that to which all is ultimately accountable*

4. Systems that know that the world that hosts them, the entities they reason about, and all of humanity and community as well, must be treated with deference, humility, and compassion

It is this kind of judgment, I believe—a seamless integration of passion, dispassion, and compassion—that ultimately underwrites what matters, not just about the human, but about the sacred, the beautiful, and the humane.

It is this kind of judgment, I am arguing, that must be the aim of any project wanting to construct "artificial general intelligence." I do not believe it is inherently beyond the reach of synthetic construction. But it is also not an incremental advance beyond first- or second-wave AI—beyond the systems we have devised to date. There are profound differences between judgment and reckoning—especially the sorts of reckoning we at present have any capacity to construct. While predictions are a fool's errand, I cannot see our synthesizing full scale judgment, if indeed we are ever capable of doing so, in anything that anyone could call the short term. Even minimal progress in that direction will require strategies wholly unlike any that have been pursued in first- and second-wave AI.

But we are learning. Second-wave AI has brought to our attention the inadequacy of the formal ontological assumptions that underlay GOFAI. From its successes, and building on the insights of a wide diversity of other fields, we should have renewed respect for three interlocking and inexorable facts:

1. The world is surpassingly rich—far transcending any ability to capture it in formal symbols, or for that matter in any discretely conceptualized structure.
2. All (especially conceptual) registration is inevitably skewed, partial, and interest-relative.
3. By exploiting yet also transcending the limitations of registration, genuine intelligence is committed to and directed toward the one and only world.

The combination of the three implies that any system that steps from one registration scheme to another, or that deals with (or is used in) different circumstances, or that attempts to integrate information gleaned from different projects, must ground its deliberations in full-scale judgment *at every step of the inferential chain*, in order to ensure that its representations never take leave of accountability to the world. These are strong but sobering conclusions—but they are straightforward consequences of how the world is.

Where does that leave us? We should be humbled by GOFAI's inadequacy, given the depth of the insights on which it was based. We should be cautious about the successes of second-wave AI, mindful of its limitations and restrictions. But mostly we should stand in awe of the capacity of the human mind, and of the achievements of human culture, in having developed registrational strategies, governing norms, ontological commitments, and epistemic practices that allow us to comprehend and go to bat for the world as world.