RETE Workshop Series on Coordination

Dai

September 29, 2022

Abstract

I will organize a series of workshops bridging major agendas in coordination research. Funding requirements include transport, accommodations, and groceries for participants, as well as a small incentive for operational assistance - see the budget section below.

1 Introduction

The motivation is a series of interlocking research agendas with partial overlap among multiple different orgs and researchers, but lacking the traction for explicit collaboration due to the "pre-paradigmatic" nature and multidisciplinary requirements of the content not meshing with publishing incentives.

The output of this synthesis is foundations for robust coordination that avoids the capabilities slant inherent in traditional game theoretic and machine learning approaches by focusing on communication bandwidth and stability analysis as key primitives.

A secondary goal is practically fostering robust intellectual coordination in the alignment-themed X-Risk ecosystem, to counteract the natural fragmenting effects of filter bubbling caused by lack of incentive for explicit third-party stewards of collaboration.

This agenda has been brewing informally for many years. I've spent the last year identifying key researchers as strategy co-founder at New Science. Prior to that I launched a number of interdisciplinary workshops while researching at the Broad Institute, Lincoln Lab, and independently over the past few years. Logistically, I have access to a couple talented operations managers eager to help, and individual lines of communication with each group, so the only bottleneck left is activation energy.

2 Sketches of the Technical Agenda

The technical agenda is roughly detailed here. I am extremely happy to express it live or via Q&A which I expect to be much higher bandwidth than guessing the reviewer's background.

3 Budget

Strategic use of these funds will make them go a long way, in contrast with major conferences or public workshops with more overhead. The aim is to have five main workshops addressing each overlap point between agendas, together with ongoing supplemental events for serendipitous progress with more diversity of participants.

Below is a rough "worst 2std" analysis, understanding that any undershot (fewer participants or days per event) will lead to proportionally more events.

- max 10 participants/staff
- round trip $1000/person = $10,000
- groceries (bulk) = $15/day/person × 7 days = $1050
- accommodations 300/night × 7 nights = $2100

= $13150 max per event

The high travel cost is due to international travel, but in practice venue will be chosen such that only about half the participants will need any travel so expect this to be the largest overshoot. Similarly, I expect to have access to a number of donated venues to offset accommodations cost.

Supplementary events will be limited to small-world shaped travel with up to 4 participants, no operational support and minimal accommodations.

- max 4 participants
- round trip $500 / person = $2000
- (supplemental) groceries, $10/day/person × 3-30 days = $1200
- emergency housing / transport, $100/day × 3-30 days = $300-$3000

So over the next year or so with 5 major workshops for communication groundwork and extended supplemental support to make deeper technical progress, the ideal budget sweet-spot amounts to $95k. It can scale back smoothly to about half that, understanding that there will be *lower breadth of expertise* as well as fewer brain-hours making. On the other hand, further funding will allow more major workshops for a longer research period, but probably will not speed up output short of 10x funding to access the next tier of talent (e.g. attracting machine-learning industry experts to do brute-force technical work rather than merely showing up to share ideas).

The budget as it currently exists is carefully chosen as a sweet-spot (scoped to one year) to access intellectual capacity bottle-necked by common-knowledge buy-in and daily opportunity cost, while avoiding the steeper costs of total career redirection - benefiting from compounding throughput of foundational common knowledge via normally inaccessible cross-industry expertise, rather than haphazardly optimizing time-to-publish.

4 Historical Motivation

Included in the appendix are unedited past fragments of highly exploratory communications to specific audience during the early formation of this agenda. They are included for historical interest but no longer fully represent the intended approach, being unmotivated by prerequisite context and containing some technical incompleteness.

A Logical Inductors as Proof Search

B Reflective Agents

C Co-knowledge

D Algebraic Ornaments (Lenses)