RAISE status report
April-June 2018

Summary

Over the past three months, RAISE has continued to make course videos on corrigibility. Our study group started to enhance and publish a prerequisites track for AI safety. We discuss several strategies to increase our production speed further, most notably setting up a volunteer team at the EA Hotel around October. We received sufficient funding to maintain operations for about three months, and will continue to look for funding opportunities in the future.

Content

Milestones

After publishing our pilot lesson in March and processing the feedback we received, we set several goals for the way forward: we aimed to increase our funding, make more videos and assignments, find a content developer and receive more feedback. Our goals have been reached or partially reached, with a few changes in strategy that will be discussed here.

- **Received a prerequisites track ready for upgrading**
  RAISE was recently contacted by Erik Istre and Trent Fowler, who designed a prerequisite track that can take a student from high school level maths to a sufficient understanding of logic, computability theory and set theory in order to start studying AI safety. The track is meant as a supplement to a traditional education degree as well as a way to help people with a non-mathematical background switch tracks. Testing, upgrading and publishing the prerequisites track is a low-effort way to have direct impact.

- **Refurbished study group**
  As many people who are interested in our study group lack some prerequisite knowledge for AI safety research, we decided to let the study group focus on testing the prerequisites track and writing up complementary exercises. The effect of this approach are twofold: the study group applicants help each other to learn the prerequisite knowledge while designing exercises that will help future students. The study group is run by Philip Blagoveschensky, who recently joined the team.

- **Received funding**
  In March, we launched a fundraiser for RAISE. Within one month, we received €3,385 from our donors. Though we hoped to receive enough money to sustain ourselves for six months, we are very thankful to those who donated to us. Together with other donations received, their support will allow us to keep going for at least another three months. We are

currently applying for several grants, and will continue the search for funding opportunities.

- ▪ **Developed two course videos**
  We finished two videos for the corrigibility unit.

- ▪ **Started designing course structure**
  Johannes Heidecke has taken up the task of designing a course structure. We currently have a large overview of the entire course, as well as a more detailed overview of the corrigibility unit. We consulted several AI alignment experts for feedback on our course structure, and adjusted the structure accordingly.

Finances

Table 1 provides an overview of our costs and donations received between 2018-03-01 and 2018-06-25. MIRI has pledged to donate € 3983,32 (4650 USD) at the start of July, which will contribute to maintaining our operations for the next three months.

Table 1 Costs and Donations March-June 2018

Balance 2018-03-01	**€ 302.55**
Donations March-June	**€ 7,936.03**
Donations from individuals	€ 7,936.03
Costs March-June	**€ 5,505.85**
Lecturer salary	€ 2,160.00
Content dev. salary	€ 1,930.00
Prereqs development	€ 413.41
Operations salary	€ 400.00
Website and logo design	€ 300.00
Website hosting	€ 217.42
Miscellaneous	€ 85.02
Balance 2018-06-25	**€2,732.73**

Challenges and concerns

- **Production speed**
Over the last three months, we have struggled with our course content production speed. In order to resolve this, we hired a part-time content developer to help with the creation of course material. Even though this helped with course production, we found our major bottleneck to be our lecturer, as personalising and shooting videos takes up most of the time. Looking forward, we will employ two major strategies to increase course content creation:
    - A few weeks ago, our lecturer, Robert Miles decided to work full time for RAISE. This will allow him to focus completely making course content and videos, which is expected to double our production speed.
    - Toon will move to the EA Hotel in Blackpool, UK in September, and Johannes will probably join him at the end of October. Here they can focus on creating content for RAISE while being provided with free meals and a room to live in. We will search for more volunteers who want to join the team in Blackpool and work on RAISE. We think that living and working together in a team will greatly increase productivity, thereby further increasing the course production speed at a low cost.

Future

With the holiday season coming up, our work output will vary per month. In July, we will continue our current routine and plan to finish another lecture video. In August, operations will be scaled back to a minimum but we will keep working on the prerequisites track with the study group. In September, we will pick up our routine while preparing the start of our work at the EA Hotel.

- **Prepare for work at the EA hotel**
We aim to find 2 or more volunteers who plan to join RAISE at the EA hotel around October and potentially want to stay for a long period of time (3-6 months). In September, Toon will work at the EA Hotel himself and prepare an office for the new volunteers.

- **Increase funding levels**
In 2-4 months, we aim to increase our funding levels. Preferably, we would receive €30,000 euros or more. However, €10,000 in funds is sufficient to keep going for about half a year and produce more valuable content.

Our future track consists of several major longer-term goals:

- **Finish first course unit**
Finish the course unit on corrigibility, complete with lecture videos, accompanying text and course assignments.

- **Develop course structure**
We made a good start with course structure development over the last three months, and will continue this in the future.

- **Set up detailed timeline**
  If the EA Hotel turns out to improve the speed and quality of our content creation, we will start systematising our group approach there and make a better time estimates of deliverables.

The next status report will be published September 2018.