



**RAISE status report  
July-October 2018**

# Summary

Due to holidays, transitions and lessons learned, the production of videos by RAISE has seen a slowdown over the course of July - October. The period was an opportunity to change our production strategy from one based on faith to one based on thoughtful project management. We have good reason to expect it to pay off, but our current publications do not yet reflect that.

# Content

<b>Summary</b>	<b>1</b>
<b>Milestones</b>	<b>1</b>
<b>Finances</b>	<b>2</b>
<b>Challenges and concerns</b>	<b>2</b>
<b>Future</b>	<b>3</b>

# Milestones

- **Established a team at the Blackpool hotel**  
Last report, we mentioned our plan establishing a production team in the EA Hotel. We have received 9 applications to join us there. Right now, 5 people live in the hotel that have been involved in content development at some point. One decided that it wasn't for them, and two are still trialing. Another will arrive in a week's time. Each member of the team commits at least 30 hours per week to development.
- **Significantly refactored our management and production strategy**  
Previously, content development work was largely done without a deliberate strategy. As one can read in the challenges section, we have learned the hard way that this can lead to problems being resolved much later than they come up. We have now started to implement systems of tracking and streamlining work, partly inspired by project management methodologies like scrum.  
As for production strategy, we have started using a new tool, Roam, that is designed to introduce clarity to intellectual work. We have designed an elaborate workflow that allows us to use this tool to map out the landscape of ideas in AI Safety in an order that best fits the learner's curiosity.
- **Received just enough funding to continue for the next period**  
Richard Ngo has stepped in at a good moment to provide RAISE with a donation of at least £5000. This happened when our runway was as short as one week. We are grateful that he has allowed us to keep up our production for at least another 10 weeks.
- **Welcomed Diana to the operations team**  
Diana Gherman has enthusiastically volunteered to take on part of operations. Her background is in recruiting, which has already proved helpful in getting more volunteers involved in the study group. Other activities she is involved in are helping to register RAISE as a CIO in the UK, hosting meetings, updating the website, and more.

- **Welcomed Davide to the content development team**

Davide Zagami has moved from Italy to the EA Hotel, now working more than full-time to summarize Inverse Reinforcement Learning papers and writing scripts that explain the concepts that underlie them. His background is in Computer Engineering.

An up to date account of the people working on RAISE can always be found at

<http://aisafety.camp/team>

## Finances

Table 1 provides an overview of our costs and donations received between 2018-06-25 and 2018-11-09.

**Table 1 Costs and Donations July-October 2018**

<b>Balance 2018-06-25</b>	<b>€ 2,732.73</b>
<b>Donations</b>	<b>€ 11,528.77</b>
MIRI	€ 5,800
Richard Ngo	€ 5,728.77
<b>Costs</b>	<b>€ 8,785</b>
Lecturer salary	€ 7,080
Prereqs development	€ 450
Website hosting	€ 154.33
Textbooks	€ 275.70
Video editing	€ 135
Miscellaneous	€ 240
<b>Balance 2018-11-09</b>	<b>€ 5,476.50</b>

## Challenges and concerns

- **Short runway**

Last report, we aimed to increase funding. Unfortunately, none of our three major grant proposals has been met. Those were with BERI, CEA and SIDN. A CEA grantmaker mentioned our proposal was “fairly strong”. BERI classified us as “promising”, ranking us somewhere between 18th and 38th out of 56 on their preference ordering. We have not received any specific feedback.

We hope that our continued work will continue to inspire support, and we will continue to talk to donors to see if our work might fit their philanthropic interest. A longer runway would allow us to aim for long term results, and we see plenty of opportunities to buy value.

- **Production speed**

We have not succeeded in increasing our production speed in the last 4 months. No videos have been published in this period. Needless to say, we are not satisfied with this. We have identified a few causes:

- On a managerial level, Toon found himself overcommitted to other projects including the EA Hotel, his thesis, holidays, and more. This left him with just enough

time to spend on RAISE operations, but not enough to do any oversight on or contribution to content development. He left Robert to his own devices.

- On an implementation level, Robert rightly worried that his knowledge in AI Safety was inadequate to finalize and publish some of the scripts. A lack of clear feedback from experts made it impossible to continue production without making obvious mistakes. He could not let that happen, as the field of AI Safety already runs the risk of not being taken seriously by academia.

However, there is good reason to believe that production will increase over the coming months:

- Toon has reduced his hours at the EA Hotel from 30 to 10, finished his thesis, cancelled some holidays, and some of his operations work is now taken up by Diana, so that he has at least 2 full days per week to focus on content development.
- We have implemented a system of daily and weekly reports that will ensure that any production bottleneck that arises will swiftly be noticed and dealt with. We have also implemented daily meetings where the team coordinates their work. This is a stark improvement over last period, where deliberate project management was not implemented.
- We now keep a list of AI Safety experts that have agreed to proofread our scripts, so that Robert will no longer find himself unable to proceed with his work.
- With Davide joining our team, there is considerably more scriptwriting being done in the system. He has joined us in the last week of October, meaning that this increase in production speed is not reflected in any publications yet.

We hope that the combination of these developments will allow us to get considerably closer to our goal of publishing one 10-minute video with accompanying exercises per 2 weeks.

## Future

### ▪ **Continue course production**

We now find ourselves in a position where most major roadblocks to production have been cleared. What's left is to do the work. We will spend the coming months building course material.

### ▪ **Further establish RAISE as a charity**

Now most of our work is done in the UK, we will register RAISE as a CIO. We hope to amass enough funding to keep our operations running at current speed, or even expand our work towards different strategies for building the Road to AI Safety Excellence. We also hope to further engage with the various AI Safety research institutions to get a better sense of their needs.

The next status report will be published March 2019.