
Google+
NASA

1
00:00:12,320 --> 00:00:10,580
hello and welcome with a collective

2
00:00:14,720 --> 00:00:12,330
minds and machines exploration challenge

3
00:00:16,129 --> 00:00:14,730
hang out I'm Jason cruzan I'm the

4
00:00:18,680 --> 00:00:16,139
director of an civics advanced

5
00:00:20,359 --> 00:00:18,690
exploration systems for NASA within our

6
00:00:23,390 --> 00:00:20,369
human exploration and operations mission

7
00:00:24,710 --> 00:00:23,400
directorate so in my group we are

8
00:00:26,900 --> 00:00:24,720
developing a lot of the next-generation

9
00:00:29,290 --> 00:00:26,910
technology and capabilities that we need

10
00:00:32,389 --> 00:00:29,300
to send humans beyond low-earth orbit

11
00:00:34,580 --> 00:00:32,399
and today we're excited to be talking

12
00:00:36,650 --> 00:00:34,590
about our latest in a series of

13
00:00:38,780 --> 00:00:36,660

challenges discussing our collective

14
00:00:40,580 --> 00:00:38,790
minds and machine exploration challenge

15
00:00:41,840 --> 00:00:40,590
it's a new open innovation and

16
00:00:43,819 --> 00:00:41,850
crowdsourcing challenge where we're

17
00:00:45,830 --> 00:00:43,829
asking you to help create a machine

18
00:00:48,590 --> 00:00:45,840
learning algorithm that will help us

19
00:00:51,830 --> 00:00:48,600
accelerate discoveries for NASA to use

20
00:00:55,340 --> 00:00:51,840
but also for the science communities

21
00:00:57,470 --> 00:00:55,350
here back on earth we're going to give

22
00:00:58,910 --> 00:00:57,480
you some brief background about what

23
00:01:01,220 --> 00:00:58,920
we're what we're doing and then we'll

24
00:01:02,420 --> 00:01:01,230
open up the floor for questions I do

25
00:01:04,490 --> 00:01:02,430
want to encourage you if you want to

26
00:01:10,070 --> 00:01:04,500
send a question in send it in through

27
00:01:12,590 --> 00:01:10,080
twitter using the pound ntl hashtag and

28
00:01:13,760 --> 00:01:12,600
again it's the pound and teal hashtag

29
00:01:16,090 --> 00:01:13,770
send your questions in that way and

30
00:01:20,630 --> 00:01:16,100
we'll be able to answer your questions

31
00:01:22,039 --> 00:01:20,640
live here on the Hangout starting off a

32
00:01:25,520 --> 00:01:22,049
little bit of the background why nasa

33
00:01:28,609 --> 00:01:25,530
uses challenges obviously we we have a

34
00:01:30,620 --> 00:01:28,619
lot of different challenges in order to

35
00:01:33,200 --> 00:01:30,630
send humans into a deep space and a lot

36
00:01:35,569 --> 00:01:33,210
of our science objectives at NASA one of

37
00:01:37,550 --> 00:01:35,579
the areas that piqued our interest in

38
00:01:39,280 --> 00:01:37,560

open innovation is how do we involve the

39
00:01:42,319 --> 00:01:39,290
public and helping solve those problems

40
00:01:44,840 --> 00:01:42,329
from a involvement category but also

41
00:01:48,020 --> 00:01:44,850
from wanting to get great excellent

42
00:01:50,510 --> 00:01:48,030
ideas and actual real solutions from the

43
00:01:52,999 --> 00:01:50,520
public as well we've seen this a lot of

44
00:01:54,530 --> 00:01:53,009
over the years of NASA and in fact every

45
00:01:56,389 --> 00:01:54,540
time that we have in a challenge or

46
00:01:58,130 --> 00:01:56,399
something out of a mission we actually

47
00:02:00,230 --> 00:01:58,140
get a lot of unsolicited feedback and

48
00:02:01,730 --> 00:02:00,240
such so we were looking at that as a

49
00:02:03,679 --> 00:02:01,740
mechanism of how do we actually do that

50
00:02:06,380 --> 00:02:03,689
up front that actually helped develop

51
00:02:10,630 --> 00:02:06,390
the systems with if we need to send and

52
00:02:13,280 --> 00:02:10,640
execute our very different missions so

53
00:02:13,730 --> 00:02:13,290
in order to facilitate this we stood up

54
00:02:16,040 --> 00:02:13,740
a think of

55
00:02:18,980 --> 00:02:16,050
nasa tournament lab because NASA doesn't

56
00:02:21,290 --> 00:02:18,990
do this just by ourselves and you'll see

57
00:02:23,600 --> 00:02:21,300
today we have our partners with Harvard

58
00:02:25,430 --> 00:02:23,610
and topcoder and the University of San

59
00:02:29,450 --> 00:02:25,440
Diego with us for this specific

60
00:02:32,090 --> 00:02:29,460
challenge and and they'll let them each

61
00:02:36,260 --> 00:02:32,100
explain a little bit about what they do

62
00:02:39,770 --> 00:02:36,270
um we'll start with dr. Albert Lin all

63
00:02:41,660 --> 00:02:39,780

right well thanks for having me I'm I'm

64
00:02:44,540 --> 00:02:41,670
excited this is a huge day for us you

65
00:02:47,330 --> 00:02:44,550
know it's the next evolution of how I

66
00:02:49,580 --> 00:02:47,340
think of the future of exploration I'm a

67
00:02:52,580 --> 00:02:49,590
research scientist here at UC San Diego

68
00:02:53,660 --> 00:02:52,590
and also i'm proud to say an emerging

69
00:02:56,510 --> 00:02:53,670
explore with the National Geographic

70
00:03:01,220 --> 00:02:56,520
Society which means that every day I try

71
00:03:03,110 --> 00:03:01,230
to take whatever I can from technology

72
00:03:04,670 --> 00:03:03,120
that exists to technology that we're

73
00:03:06,470 --> 00:03:04,680
inventing two ways in which we think

74
00:03:09,110 --> 00:03:06,480
about science and apply it to

75
00:03:12,290 --> 00:03:09,120
exploration to search and discovery to

76
00:03:16,070 --> 00:03:12,300
try to which limits what we know about

77
00:03:17,330 --> 00:03:16,080
life on this planet and you know over

78
00:03:24,080 --> 00:03:17,340
the years the things that I have focused

79
00:03:26,960 --> 00:03:24,090
on have been discovery uh I spent last

80
00:03:31,490 --> 00:03:26,970
four or five years now trying to

81
00:03:34,550 --> 00:03:31,500
investigate a place that has been

82
00:03:36,110 --> 00:03:34,560
forbidden to go to for 800 years the

83
00:03:39,080 --> 00:03:36,120
homeland and actually the ancestral

84
00:03:42,050 --> 00:03:39,090
lands of the Mongols Genghis Khan right

85
00:03:44,630 --> 00:03:42,060
and in doing so we've been trying to do

86
00:03:46,550 --> 00:03:44,640
this in a way which respects the local

87
00:03:47,780 --> 00:03:46,560
traditions of Mongolian people which

88
00:03:49,850 --> 00:03:47,790

mean that could be completely

89
00:03:52,610 --> 00:03:49,860
non-invasive in our archaeological

90
00:03:54,170 --> 00:03:52,620
service right but we're looking for a

91
00:03:56,900 --> 00:03:54,180
needle in a haystack we're looking for

92
00:04:00,170 --> 00:03:56,910
something where you know not only is it

93
00:04:02,750 --> 00:04:00,180
buried within a deep amount of you know

94
00:04:05,660 --> 00:04:02,760
data information where we don't really

95
00:04:07,910 --> 00:04:05,670
have a clue of where it might be but we

96
00:04:11,870 --> 00:04:07,920
also have to look in a way where we

97
00:04:14,060 --> 00:04:11,880
don't presume to know what what were

98
00:04:15,949 --> 00:04:14,070
what we will find is actually what we

99
00:04:17,539 --> 00:04:15,959
think it is right i mean we don't have a

100
00:04:19,970 --> 00:04:17,549
definition that the characteristic of

101
00:04:23,230 --> 00:04:19,980
what we're looking for so when we use

102
00:04:27,320 --> 00:04:23,240
satellite imagery that's clarity's of

103
00:04:29,629 --> 00:04:27,330
ground cover we have to try to

104
00:04:31,159 --> 00:04:29,639
happen to the minds of many of our

105
00:04:33,830 --> 00:04:31,169
friends and colleagues for their human

106
00:04:36,140 --> 00:04:33,840
perception look for what's different and

107
00:04:38,839 --> 00:04:36,150
in collaborating with ntl and with the

108
00:04:42,140 --> 00:04:38,849
group you know through the Harvard

109
00:04:43,790 --> 00:04:42,150
Business School and top coder NASA you

110
00:04:45,409 --> 00:04:43,800
know what I see is an exciting

111
00:04:48,170 --> 00:04:45,419
opportunity for us to take what we've

112
00:04:49,909 --> 00:04:48,180
learned from crowd sourcing from tapping

113
00:04:53,240 --> 00:04:49,919

into these massive pools of human

114
00:04:55,730 --> 00:04:53,250
perception and feed that into a new

115
00:04:58,100 --> 00:04:55,740
paradigm where machines and humans are

116
00:04:59,869 --> 00:04:58,110
working together where some of the

117
00:05:02,540 --> 00:04:59,879
challenges that we face don't have

118
00:05:04,040 --> 00:05:02,550
enough eyes to solve them we're trying

119
00:05:06,800 --> 00:05:04,050
to take on the challenges of our big

120
00:05:09,050 --> 00:05:06,810
data avalanches that we face today but

121
00:05:11,180 --> 00:05:09,060
in ways that are collaborative both

122
00:05:13,999 --> 00:05:11,190
across large networks of people and also

123
00:05:16,309 --> 00:05:14,009
in which we as those networks can

124
00:05:18,649 --> 00:05:16,319
leverage the most advanced technologies

125
00:05:20,390 --> 00:05:18,659
possible and the reason for connecting

126
00:05:21,680 --> 00:05:20,400
with MTL is that we have no idea what

127
00:05:24,260 --> 00:05:21,690
those advanced technologies would be

128
00:05:26,390 --> 00:05:24,270
anyway as well right so we have to tap

129
00:05:28,969 --> 00:05:26,400
into the open innovation forms that

130
00:05:31,580 --> 00:05:28,979
exist out there and think about how you

131
00:05:33,409 --> 00:05:31,590
know how you as a community can actually

132
00:05:34,730 --> 00:05:33,419
help us come up with the best answers

133
00:05:36,950 --> 00:05:34,740
and the best solutions to this challenge

134
00:05:38,839 --> 00:05:36,960
so it's a great pleasure to be here and

135
00:05:44,659 --> 00:05:38,849
look forward to spending the next after

136
00:05:47,719 --> 00:05:44,669
talking with you guys okay next I'm

137
00:05:49,519 --> 00:05:47,729
gonna have cream who is our director of

138
00:05:51,740 --> 00:05:49,529

the Harvard NASA tournament lab talk a

139
00:05:56,329 --> 00:05:51,750
little bit about his background and also

140
00:05:57,559 --> 00:05:56,339
how the NASA turma lab team sure so

141
00:06:01,279 --> 00:05:57,569
thanks Jason and thanks to all of you

142
00:06:05,269 --> 00:06:01,289
for joining us today so the National Lab

143
00:06:07,519 --> 00:06:05,279
came about as a way for us to take the

144
00:06:09,529 --> 00:06:07,529
lessons of around crowdsourcing around

145
00:06:12,050 --> 00:06:09,539
the use of contests and communities to

146
00:06:14,809 --> 00:06:12,060
solve problems and apply it to real

147
00:06:17,329 --> 00:06:14,819
problems from real organizations like

148
00:06:19,430 --> 00:06:17,339
NASA and the arrangement was basically

149
00:06:22,670 --> 00:06:19,440
that you know we've done some pilot

150
00:06:24,909 --> 00:06:22,680
prototype tests with with top coder and

151
00:06:28,309 --> 00:06:24,919
NASA and we found really good results

152
00:06:30,559 --> 00:06:28,319
from a technical point of view and what

153
00:06:32,570 --> 00:06:30,569
we decided to do was say how how about

154
00:06:35,209 --> 00:06:32,580
we scale this up in two ways one is

155
00:06:38,749 --> 00:06:35,219
skill this up so we can take in more

156
00:06:40,030 --> 00:06:38,759
problems for NASA and a diverse set of

157
00:06:43,030 --> 00:06:40,040
problems for NASA

158
00:06:45,580 --> 00:06:43,040
but also for us to do real social

159
00:06:47,680 --> 00:06:45,590
science around them so we're trying to

160
00:06:49,750 --> 00:06:47,690
build out both our understanding of how

161
00:06:51,850 --> 00:06:49,760
contests work what are the motivations

162
00:06:54,610 --> 00:06:51,860
for people to participate incentives

163
00:06:56,680 --> 00:06:54,620

when do contests fail and we can do that

164
00:07:00,370 --> 00:06:56,690
in the setup of a real life laboratory

165
00:07:02,770 --> 00:07:00,380
that is top coder from our relationship

166
00:07:04,390 --> 00:07:02,780
with NASA we've now actually worked with

167
00:07:05,830 --> 00:07:04,400
some other federal agencies as well and

168
00:07:09,040 --> 00:07:05,840
you've seen some of those challenges on

169
00:07:11,740 --> 00:07:09,050
the top coat or platform and what we're

170
00:07:16,450 --> 00:07:11,750
really trying to do is try to take

171
00:07:20,470 --> 00:07:16,460
mainstream crowd sourcing as a viable of

172
00:07:22,660 --> 00:07:20,480
a toolkit for organizations to use on a

173
00:07:24,040 --> 00:07:22,670
regular basis that we shouldn't be

174
00:07:26,200 --> 00:07:24,050
thinking about this as a substitute for

175
00:07:29,140 --> 00:07:26,210
what happens inside of a company or an

176
00:07:31,780 --> 00:07:29,150
organization but a real compliment and

177
00:07:33,600 --> 00:07:31,790
our work here is to really show this

178
00:07:36,610 --> 00:07:33,610
happening in a wide range of settings

179
00:07:39,670 --> 00:07:36,620
and with a relationship with NASA and

180
00:07:41,500 --> 00:07:39,680
the top coder you really sort of shown

181
00:07:44,020 --> 00:07:41,510
to the world to the Technic communities

182
00:07:47,580 --> 00:07:44,030
that we that we interact with that in

183
00:07:53,590 --> 00:07:47,590
fact you know we can create robust

184
00:07:56,860 --> 00:07:53,600
amazing solutions for for for some very

185
00:07:59,020 --> 00:07:56,870
significant technical challenges it's a

186
00:08:02,020 --> 00:07:59,030
ton of fun a real privilege for me to be

187
00:08:03,430 --> 00:08:02,030
to be working with NASA you know I

188
00:08:06,520 --> 00:08:03,440

always want to be an astronaut growing

189
00:08:08,200 --> 00:08:06,530
up never will become that I don't think

190
00:08:09,640 --> 00:08:08,210
unless know the space program expands

191
00:08:13,150 --> 00:08:09,650
greatly or ER and a few million more

192
00:08:16,000 --> 00:08:13,160
dollars to you know to go into space but

193
00:08:20,470 --> 00:08:16,010
certainly to work with NASA to work with

194
00:08:22,990 --> 00:08:20,480
Jason and to really take on concrete

195
00:08:25,510 --> 00:08:23,000
problems faced by NASA and then to have

196
00:08:27,990 --> 00:08:25,520
them participate in a system that helps

197
00:08:31,570 --> 00:08:28,000
solve them is quite the privilege um

198
00:08:33,130 --> 00:08:31,580
similarly a matar coder it's been an

199
00:08:35,320 --> 00:08:33,140
amazing relationship for us from a

200
00:08:38,440 --> 00:08:35,330
research point of view they have really

201
00:08:41,260 --> 00:08:38,450
helped us understand the dynamics of

202
00:08:44,200 --> 00:08:41,270
contest and tournaments from a real

203
00:08:46,180 --> 00:08:44,210
sense of there's a lot of Huey a lot of

204
00:08:48,280 --> 00:08:46,190
models about how contest really worked

205
00:08:49,720 --> 00:08:48,290
contest work but very little empirical

206
00:08:50,930 --> 00:08:49,730
evidence and we've been very fortunate

207
00:08:53,570 --> 00:08:50,940
about having more

208
00:08:55,970 --> 00:08:53,580
with the top coder community and the top

209
00:08:58,550 --> 00:08:55,980
clear company on this and then our

210
00:09:01,370 --> 00:08:58,560
relationship with Albert came about as

211
00:09:05,000 --> 00:09:01,380
our school was exploring new models for

212
00:09:07,790 --> 00:09:05,010
collaboration and we had a visit Oh to

213
00:09:09,500 --> 00:09:07,800

Albert's lad on San Diego and that's

214
00:09:11,810 --> 00:09:09,510
where we were exposed to his ideas and

215
00:09:14,000 --> 00:09:11,820
it was a natural fit was sort of like

216
00:09:16,430 --> 00:09:14,010
well we've got all this crowd label data

217
00:09:18,890 --> 00:09:16,440
and let's take the best of the crowd

218
00:09:22,010 --> 00:09:18,900
labeling at efforts that that albert is

219
00:09:23,750 --> 00:09:22,020
in so well and tie back to what the top

220
00:09:26,030 --> 00:09:23,760
coder to me does well which is machine

221
00:09:27,950 --> 00:09:26,040
learning and let's do top coat from but

222
00:09:29,960 --> 00:09:27,960
you know crowd sourcing for both ends

223
00:09:32,270 --> 00:09:29,970
both from the generation of data as well

224
00:09:35,810 --> 00:09:32,280
as your machine learning so I am super

225
00:09:37,550 --> 00:09:35,820
excited about this this this project and

226
00:09:42,160 --> 00:09:37,560
they looking forward to your questions

227
00:09:45,020 --> 00:09:42,170
at your participation more pork great

228
00:09:46,820 --> 00:09:45,030
last but not least wouldn't be able to

229
00:09:49,580 --> 00:09:46,830
do any of this without actually having a

230
00:09:51,770 --> 00:09:49,590
community of really excellent solvers to

231
00:09:54,260 --> 00:09:51,780
help us I can contribute to this and

232
00:09:57,460 --> 00:09:54,270
that is enabled through our partnership

233
00:10:00,560 --> 00:09:57,470
with top coder and Andy over to you

234
00:10:02,600 --> 00:10:00,570
thanks Jason hi I'm Manny lamora I am

235
00:10:05,540 --> 00:10:02,610
senior vice president of the government

236
00:10:08,030 --> 00:10:05,550
platforms of top coder and as creating

237
00:10:10,130 --> 00:10:08,040
Jason mentioned what we do is power the

238
00:10:12,200 --> 00:10:10,140

the ntl and that's a tournament lab and

239
00:10:15,110 --> 00:10:12,210
it's really exciting for us to be part

240
00:10:17,690 --> 00:10:15,120
of this because how many people get to

241
00:10:19,970 --> 00:10:17,700
work on a NASA problem so by bringing

242
00:10:22,670 --> 00:10:19,980
these challenges to our community we

243
00:10:24,320 --> 00:10:22,680
allow people all over to have a crack at

244
00:10:27,110 --> 00:10:24,330
a problem that they probably never

245
00:10:29,240 --> 00:10:27,120
thought they'd be able to and the types

246
00:10:30,950 --> 00:10:29,250
of solutions it brings are amazing and

247
00:10:34,570 --> 00:10:30,960
it's exciting and we just love being

248
00:10:37,910 --> 00:10:34,580
part of it so kind of in keeping with my

249
00:10:41,530 --> 00:10:37,920
company's function of bringing a crowd

250
00:10:44,360 --> 00:10:41,540
to ntl and NASA today I'll be bringing

251
00:10:47,450 --> 00:10:44,370
questions from a crowd to the panelists

252
00:10:52,360 --> 00:10:47,460
for this challenge so without further

253
00:10:56,240 --> 00:10:52,370
ado I think we can go to the crowd right

254
00:10:58,040 --> 00:10:56,250
so please follow along on google me know

255
00:11:01,630 --> 00:10:58,050
if you have any questions for us send

256
00:11:04,580 --> 00:11:01,640
them in on twitter with the hashtag ntl

257
00:11:12,340 --> 00:11:04,590
and as they come in will be read

258
00:11:19,730 --> 00:11:14,570
first things first I have a question

259
00:11:21,980 --> 00:11:19,740
here uh for for Albert so Albert has

260
00:11:25,760 --> 00:11:21,990
your research by using crowdsourcing

261
00:11:28,880 --> 00:11:25,770
produced any new discovery yeah

262
00:11:31,670 --> 00:11:28,890
absolutely you know I mean it's not just

263
00:11:34,160 --> 00:11:31,680

a I think of crowdsourcing I don't think

264
00:11:38,360 --> 00:11:34,170
of it as a as a as a solution to solving

265
00:11:42,350 --> 00:11:38,370
problems of scale alone I think about it

266
00:11:45,170 --> 00:11:42,360
as a new method for tapping into the

267
00:11:47,600 --> 00:11:45,180
unknown right because unlike machines

268
00:11:49,880 --> 00:11:47,610
humans are very good at trying to figure

269
00:11:51,650 --> 00:11:49,890
out what you know what something is when

270
00:11:56,500 --> 00:11:51,660
they first see if it's based on instinct

271
00:12:00,980 --> 00:11:56,510
and what we used in our last platform to

272
00:12:02,540 --> 00:12:00,990
tap into that was a process where we

273
00:12:05,060 --> 00:12:02,550
asked many many people to stare at a

274
00:12:06,500 --> 00:12:05,070
single piece of imagery and tell us in

275
00:12:08,690 --> 00:12:06,510
parallel what they saw on that imagery

276
00:12:14,180 --> 00:12:08,700
but thought it was something strange

277
00:12:18,020 --> 00:12:14,190
modern man-made natural and they would

278
00:12:20,390 --> 00:12:18,030
tag these things right then when they

279
00:12:21,980 --> 00:12:20,400
finish tagging this this little piece of

280
00:12:23,780 --> 00:12:21,990
information they would see what the

281
00:12:25,040 --> 00:12:23,790
people around them had said about the

282
00:12:26,960 --> 00:12:25,050
same piece of data but they couldn't

283
00:12:28,940 --> 00:12:26,970
change their answers right so now I have

284
00:12:32,360 --> 00:12:28,950
a method in which I can use a

285
00:12:34,790 --> 00:12:32,370
mathematical statistical method to to

286
00:12:37,910 --> 00:12:34,800
see where clusters of agreement emerge

287
00:12:39,200 --> 00:12:37,920
in in the highest density right and I

288
00:12:41,420 --> 00:12:39,210

can actually write out on horseback

289
00:12:42,680 --> 00:12:41,430
every single day into the field and

290
00:12:44,390 --> 00:12:42,690
check out exactly where that

291
00:12:48,320 --> 00:12:44,400
correspondents you on the ground and

292
00:12:51,140 --> 00:12:48,330
over the process of you know doing this

293
00:12:53,330 --> 00:12:51,150
in the in the amazing landscape of long

294
00:12:55,640 --> 00:12:53,340
ago Leah we discovered over 55

295
00:12:58,460 --> 00:12:55,650
archaeological sites that we identified

296
00:13:00,170 --> 00:12:58,470
these things across massive you know

297
00:13:01,880 --> 00:13:00,180
region some of which we are pretty

298
00:13:05,660 --> 00:13:01,890
excited about you know a lot of that

299
00:13:07,130 --> 00:13:05,670
work is now being being reviewed within

300
00:13:08,750 --> 00:13:07,140
the ministry of culture Mongolia and

301
00:13:12,320 --> 00:13:08,760
we'll figure out what we can say about

302
00:13:14,060 --> 00:13:12,330
that later but well we're you know what

303
00:13:15,800 --> 00:13:14,070
what I got even more excited about after

304
00:13:18,110 --> 00:13:15,810
that was the fact that this method

305
00:13:20,390 --> 00:13:18,120
didn't just apply to archaeology right

306
00:13:21,950 --> 00:13:20,400
I mean it could be used in archeology

307
00:13:23,960 --> 00:13:21,960
for other areas and that's what you know

308
00:13:26,980 --> 00:13:23,970
we continue to do places that range from

309
00:13:30,260 --> 00:13:26,990
you know the Mayan civilizations of

310
00:13:32,810 --> 00:13:30,270
Central America to Egypt but we could

311
00:13:34,850 --> 00:13:32,820
also use us beyond that right i mean the

312
00:13:38,090 --> 00:13:34,860
method of tapping into large quantities

313
00:13:40,640 --> 00:13:38,100

of human perception for discovery and

314
00:13:44,329 --> 00:13:40,650
the unknown can be applied to mapping

315
00:13:49,190 --> 00:13:44,339
neural pathways in the brain or two you

316
00:13:51,710 --> 00:13:49,200
know mapping the the the planets that we

317
00:13:54,350 --> 00:13:51,720
start to observe at higher sensitivities

318
00:13:56,060 --> 00:13:54,360
with better sensors and and better

319
00:13:57,560 --> 00:13:56,070
telescopes right i mean this is all an

320
00:13:59,510 --> 00:13:57,570
arrow but we're producing more and more

321
00:14:01,850 --> 00:13:59,520
information and the processes and the

322
00:14:04,220 --> 00:14:01,860
methods is which we use to analyze that

323
00:14:05,600 --> 00:14:04,230
data is going to is going to define the

324
00:14:07,820 --> 00:14:05,610
kinds of discoveries that we make in the

325
00:14:10,610 --> 00:14:07,830
level of discovery that would make you

326
00:14:12,920 --> 00:14:10,620
know most recently my friends over a

327
00:14:16,820 --> 00:14:12,930
digital globe have been using a very

328
00:14:18,680 --> 00:14:16,830
similar method to search for the

329
00:14:20,390 --> 00:14:18,690
schooner Nina this lost this law

330
00:14:23,300 --> 00:14:20,400
schooner that went missing in Tasmanian

331
00:14:25,160 --> 00:14:23,310
see and you know I I really hope that

332
00:14:27,140 --> 00:14:25,170
they end up finding this but it just

333
00:14:29,300 --> 00:14:27,150
shows you that you know we as a

334
00:14:31,790 --> 00:14:29,310
community can solve big problems real

335
00:14:33,680 --> 00:14:31,800
problems if we just think of new ways in

336
00:14:35,630 --> 00:14:33,690
which we network with each other to take

337
00:14:37,160 --> 00:14:35,640
on those challenges and the discoveries

338
00:14:39,140 --> 00:14:37,170

that we've made thus far with this

339
00:14:41,180 --> 00:14:39,150
methodology have just continued to

340
00:14:42,949 --> 00:14:41,190
astound me and I'm just excited to see

341
00:14:44,840 --> 00:14:42,959
what happens next when we can when we

342
00:14:47,690 --> 00:14:44,850
can tap into a collaboration with the

343
00:14:52,190 --> 00:14:47,700
machines to amplify even further what

344
00:14:54,500 --> 00:14:52,200
you know what this power is thanks

345
00:14:57,829 --> 00:14:54,510
Albert and the second question kind of

346
00:15:01,210 --> 00:14:57,839
conveniently relates to build on on your

347
00:15:05,180 --> 00:15:01,220
experiences there so with your work on

348
00:15:07,519 --> 00:15:05,190
the on the reviewing the are looking for

349
00:15:10,760 --> 00:15:07,529
ancient structures and given what we saw

350
00:15:14,390 --> 00:15:10,770
recently with the the Boston bombing

351
00:15:16,010 --> 00:15:14,400
where people were using images from from

352
00:15:18,560 --> 00:15:16,020
twitter from whatever sources were

353
00:15:21,530 --> 00:15:18,570
available I considering the search for

354
00:15:23,780 --> 00:15:21,540
the schooner um what are some of the

355
00:15:25,760 --> 00:15:23,790
greatest barriers to using this approach

356
00:15:28,730 --> 00:15:25,770
to using crowdsourcing works from the

357
00:15:31,119 --> 00:15:28,740
gut you know that's a very appropriate

358
00:15:33,530 --> 00:15:31,129
question and I think you

359
00:15:35,780 --> 00:15:33,540
it all comes down to the way in which we

360
00:15:37,340 --> 00:15:35,790
structure an architect these big

361
00:15:39,230 --> 00:15:37,350
networks and how they collaborate with

362
00:15:42,110 --> 00:15:39,240
each other careful structuring allows us

363
00:15:44,990 --> 00:15:42,120

to avoid turning a crowd of helpers into

364
00:15:49,100 --> 00:15:45,000
a mob which is you know something that

365
00:15:50,780 --> 00:15:49,110
was somewhat you know highlighted as a

366
00:15:53,240 --> 00:15:50,790
potential outcome of what happened in

367
00:15:54,319 --> 00:15:53,250
Boston's right so in Boston we had a

368
00:15:57,949 --> 00:15:54,329
case where people were looking at

369
00:16:00,499 --> 00:15:57,959
imagery online but in some cases then

370
00:16:02,569 --> 00:16:00,509
analysis was so emotionally driven that

371
00:16:05,150 --> 00:16:02,579
you know we didn't have the mechanisms

372
00:16:09,889 --> 00:16:05,160
of which to try to create these

373
00:16:11,290 --> 00:16:09,899
statistical valid verifications of the

374
00:16:14,240 --> 00:16:11,300
information that was coming out of that

375
00:16:16,819 --> 00:16:14,250
what we're doing here is we're using

376
00:16:18,559 --> 00:16:16,829
architecture in which you know you can

377
00:16:20,420 --> 00:16:18,569
see certain things and you can work and

378
00:16:24,019 --> 00:16:20,430
network in crush certain parallels but

379
00:16:26,360 --> 00:16:24,029
but the way in which your data evolves

380
00:16:28,819 --> 00:16:26,370
allows it to be something that's safe

381
00:16:31,579 --> 00:16:28,829
and secure and also is truly

382
00:16:34,220 --> 00:16:31,589
self-guiding in terms of you know this

383
00:16:37,280 --> 00:16:34,230
this crowd is no single individual in

384
00:16:39,860 --> 00:16:37,290
the crowd is deciding who might be or

385
00:16:41,960 --> 00:16:39,870
where this thing might be the anomaly

386
00:16:45,110 --> 00:16:41,970
that we're looking for rather it's the

387
00:16:47,509 --> 00:16:45,120
the group as a whole agreeing upon a

388
00:16:49,009 --> 00:16:47,519

location independently and and that

389
00:16:52,100 --> 00:16:49,019
emerging an answer which we have found

390
00:16:53,509 --> 00:16:52,110
to be far far greater percentage of

391
00:16:55,400 --> 00:16:53,519
accuracy from that kind of methodology

392
00:16:57,710 --> 00:16:55,410
so it's all about structure and that's

393
00:17:00,290 --> 00:16:57,720
the lesson that I've learned you know

394
00:17:02,030 --> 00:17:00,300
and you know I think we can continue to

395
00:17:03,199 --> 00:17:02,040
evolve those structures to to really

396
00:17:05,120 --> 00:17:03,209
push limits of what our

397
00:17:07,399 --> 00:17:05,130
interconnectivity can do we just have to

398
00:17:09,110 --> 00:17:07,409
be careful about how we do it yeah I

399
00:17:11,390 --> 00:17:09,120
finally add to this I think I think it's

400
00:17:13,549 --> 00:17:11,400
so I mean what we know is one of the

401
00:17:17,210 --> 00:17:13,559
failures for crowds is lack of

402
00:17:19,370 --> 00:17:17,220
governance so if you think about both

403
00:17:21,699 --> 00:17:19,380
sort of communities open source software

404
00:17:24,590 --> 00:17:21,709
communities there's a very strong

405
00:17:26,449 --> 00:17:24,600
implicit governance model that exists in

406
00:17:27,909 --> 00:17:26,459
open source software communities or even

407
00:17:30,610 --> 00:17:27,919
Wikipedia which is a very ornate

408
00:17:34,909 --> 00:17:30,620
governance model put into it I'm

409
00:17:37,490 --> 00:17:34,919
similarly a you know a in contests as

410
00:17:40,490 --> 00:17:37,500
well there is a governess required there

411
00:17:42,950 --> 00:17:40,500
it has to be some way to organize the

412
00:17:44,210 --> 00:17:42,960
inputs coming in and some way to detect

413
00:17:46,430 --> 00:17:44,220

false positives

414
00:17:48,350 --> 00:17:46,440
often times when we work together you

415
00:17:51,020 --> 00:17:48,360
know in all the projects we do a top

416
00:17:52,789 --> 00:17:51,030
coder and a big effort also with Albert

417
00:17:54,320 --> 00:17:52,799
boys to sort of say what is going to be

418
00:17:56,570 --> 00:17:54,330
the scoring algorithm how will we know

419
00:17:59,210 --> 00:17:56,580
and we get the right answer instead of a

420
00:18:01,580 --> 00:17:59,220
false positive or false negative so

421
00:18:05,330 --> 00:18:01,590
that's that's where you know governance

422
00:18:07,010 --> 00:18:05,340
matters a ton and that's what almost you

423
00:18:12,200 --> 00:18:07,020
know the size of a project will be

424
00:18:13,850 --> 00:18:12,210
successful or not so that brings up a

425
00:18:16,730 --> 00:18:13,860
question we talked about crowdsourcing

426
00:18:18,529 --> 00:18:16,740
we know popular does contests and

427
00:18:21,710 --> 00:18:18,539
tournament lab has the word tournament

428
00:18:27,440 --> 00:18:21,720
in it so one question is what's a

429
00:18:30,500 --> 00:18:27,450
challenge it really does like a question

430
00:18:32,149 --> 00:18:30,510
for you cream what's the challenge what

431
00:18:33,710 --> 00:18:32,159
is in the context of what we're doing

432
00:18:36,110 --> 00:18:33,720
here we'd use the word like contest and

433
00:18:37,730 --> 00:18:36,120
challenges so from the view of somebody

434
00:18:40,180 --> 00:18:37,740
who wants to participate what what is a

435
00:18:43,100 --> 00:18:40,190
challenge and how do we use it here Oh

436
00:18:45,200 --> 00:18:43,110
sociologist is it specifies the problem

437
00:18:49,039 --> 00:18:45,210
your we use these words in our most

438
00:18:51,260 --> 00:18:49,049

intermixed stably meaning that you know

439
00:18:53,510 --> 00:18:51,270
will refer to a contest or refer to a

440
00:18:55,850 --> 00:18:53,520
tournament will refer to a problem or

441
00:18:59,240 --> 00:18:55,860
refer to challenge and all those

442
00:19:02,750 --> 00:18:59,250
elements are basically saying that we

443
00:19:05,419 --> 00:19:02,760
are setting up a structure where there's

444
00:19:08,240 --> 00:19:05,429
a defined problem a defined set of

445
00:19:11,960 --> 00:19:08,250
evaluation criteria that are objective

446
00:19:13,880 --> 00:19:11,970
that we have developed and the challenge

447
00:19:17,000 --> 00:19:13,890
is to solve the problem as stated on the

448
00:19:19,220 --> 00:19:17,010
site I that creates the best score based

449
00:19:20,690 --> 00:19:19,230
on our evaluation criteria that's as

450
00:19:23,060 --> 00:19:20,700
simple as that people will use

451
00:19:25,580 --> 00:19:23,070
challenges people will use prizes people

452
00:19:28,850 --> 00:19:25,590
will say tournaments people in all those

453
00:19:30,860 --> 00:19:28,860
terms basically that the fundamental

454
00:19:33,200 --> 00:19:30,870
component is it's a race there is

455
00:19:34,789 --> 00:19:33,210
there's a time amount required in the

456
00:19:37,100 --> 00:19:34,799
sense of it's going to be over a certain

457
00:19:40,880 --> 00:19:37,110
period of time and then the best

458
00:19:42,620 --> 00:19:40,890
performing solutions will one yeah so in

459
00:19:44,840 --> 00:19:42,630
this case a little bit more on that in

460
00:19:48,289 --> 00:19:44,850
this case we have a singular kind of

461
00:19:51,130 --> 00:19:48,299
what's called a marathon match is

462
00:19:54,440 --> 00:19:51,140
looking at creation of this machine

463
00:19:58,130 --> 00:19:54,450

learning algorithm based on the data

464
00:20:00,920 --> 00:19:58,140
sets that Albert's team has

465
00:20:02,420 --> 00:20:00,930
provided to us I was actually recreating

466
00:20:03,620 --> 00:20:02,430
bishop and machine learning algorithm

467
00:20:05,810 --> 00:20:03,630
they can actually learn from these type

468
00:20:07,670 --> 00:20:05,820
of data sources out there and it's a

469
00:20:09,920 --> 00:20:07,680
singular kind of challenge or tournament

470
00:20:12,110 --> 00:20:09,930
in other in other cases we decompose

471
00:20:15,080 --> 00:20:12,120
problems into a series of smaller

472
00:20:17,300 --> 00:20:15,090
challenges in order to in order to kind

473
00:20:19,130 --> 00:20:17,310
of split apart the problem and sometimes

474
00:20:21,470 --> 00:20:19,140
we call those tournaments or individual

475
00:20:23,720 --> 00:20:21,480
contests that then go to an overall

476
00:20:25,220 --> 00:20:23,730
challenge type thing or getting it all

477
00:20:26,870 --> 00:20:25,230
depends on the scope within size of the

478
00:20:30,100 --> 00:20:26,880
problem we do end up using the

479
00:20:32,390 --> 00:20:30,110
terminology pretty intermittent airy but

480
00:20:34,280 --> 00:20:32,400
which you can you can actually this a

481
00:20:35,450 --> 00:20:34,290
bit of a hierarchy to this to in some

482
00:20:37,670 --> 00:20:35,460
cases when you want to split it apart

483
00:20:40,520 --> 00:20:37,680
but in this case it's a singular kind of

484
00:20:44,510 --> 00:20:40,530
marathon match that we think we can get

485
00:20:47,210 --> 00:20:44,520
good results from Thanks patient so

486
00:20:49,880 --> 00:20:47,220
follow-up question for you then NASA is

487
00:20:53,890 --> 00:20:49,890
literally the home of rocket scientists

488
00:20:57,770 --> 00:20:53,900

right 50min an idiom on how why does

489
00:20:58,970 --> 00:20:57,780
Metheny challenges so there's a couple

490
00:21:01,070 --> 00:20:58,980
of different ways so mean we're not

491
00:21:04,550 --> 00:21:01,080
obviously we're not replacing what we

492
00:21:06,830 --> 00:21:04,560
have within the NASA family in NASA and

493
00:21:09,170 --> 00:21:06,840
it's normal academics and Industry and

494
00:21:10,970 --> 00:21:09,180
such that we work with every day but we

495
00:21:12,770 --> 00:21:10,980
want to be able to do is harness the

496
00:21:14,810 --> 00:21:12,780
power of all the people that are out

497
00:21:17,690 --> 00:21:14,820
there to be able to accelerate the pace

498
00:21:19,550 --> 00:21:17,700
of which we can learn so no matter

499
00:21:21,740 --> 00:21:19,560
there's a there's a quote out there the

500
00:21:24,200 --> 00:21:21,750
joys law boy both at any one of us here

501
00:21:25,430 --> 00:21:24,210
talk about all the time that say it says

502
00:21:27,680 --> 00:21:25,440
no matter who you are the smartest

503
00:21:29,930 --> 00:21:27,690
people work for somebody else but really

504
00:21:32,900 --> 00:21:29,940
the smarts come from a very large

505
00:21:34,970 --> 00:21:32,910
community and in aggregate you can get a

506
00:21:36,440 --> 00:21:34,980
much better result by the gnome the

507
00:21:39,380 --> 00:21:36,450
increased number of people you can

508
00:21:41,630 --> 00:21:39,390
engage with so we look at a problem from

509
00:21:42,890 --> 00:21:41,640
a certain perspective from in-house but

510
00:21:44,450 --> 00:21:42,900
somebody from the outside may bring a

511
00:21:46,310 --> 00:21:44,460
total different perspective to that

512
00:21:47,720 --> 00:21:46,320
problem or different approach or

513
00:21:50,330 --> 00:21:47,730

knowledge they already have and apply it

514
00:21:53,020 --> 00:21:50,340
to our problem set and in essence that

515
00:21:55,220 --> 00:21:53,030
allows us to accelerate that pace of

516
00:21:57,050 --> 00:21:55,230
solving a solution like in this case an

517
00:21:58,880 --> 00:21:57,060
algorithm in other cases bringing a

518
00:22:01,430 --> 00:21:58,890
completely different concept of a new

519
00:22:03,860 --> 00:22:01,440
idea how to go after a problem so it's

520
00:22:05,900 --> 00:22:03,870
that kind of broad broad search kind of

521
00:22:10,190 --> 00:22:05,910
capabilities of you know to find best

522
00:22:11,930 --> 00:22:10,200
ideas anywhere in the world thanks and

523
00:22:13,399 --> 00:22:11,940
so one more follow-up

524
00:22:16,669 --> 00:22:13,409
know that we talked about a little bit

525
00:22:18,799 --> 00:22:16,679
during introductions but with respect to

526
00:22:20,539 --> 00:22:18,809
what you just said what are some of the

527
00:22:24,230 --> 00:22:20,549
specific things that you're looking for

528
00:22:26,840 --> 00:22:24,240
out of this challenge and how do you how

529
00:22:28,779 --> 00:22:26,850
do you see people at NASA using it yeah

530
00:22:32,570 --> 00:22:28,789
so there's obviously a great application

531
00:22:35,240 --> 00:22:32,580
and what other words just doing with his

532
00:22:36,379 --> 00:22:35,250
exploration here on the ground but at

533
00:22:39,470 --> 00:22:36,389
the end of the day but we'll end up

534
00:22:42,320 --> 00:22:39,480
creating is an algorithm that can be

535
00:22:46,070 --> 00:22:42,330
trained by through crowd sourcing data

536
00:22:49,129 --> 00:22:46,080
so if we had say in an image set that we

537
00:22:51,019 --> 00:22:49,139
had that we're looking for asteroids so

538
00:22:52,820 --> 00:22:51,029

if we have a lot of Sky Survey data

539
00:22:54,769 --> 00:22:52,830
where we survey this guy and we're

540
00:22:57,649 --> 00:22:54,779
looking for things that change in a

541
00:22:59,180 --> 00:22:57,659
peculiar way in order to figure out an

542
00:23:00,529 --> 00:22:59,190
object that's a near-earth object or

543
00:23:02,690 --> 00:23:00,539
asteroid you're coming across this guy

544
00:23:04,999 --> 00:23:02,700
right now there are algorithms to do

545
00:23:06,230 --> 00:23:05,009
some of that but you could actually

546
00:23:07,940 --> 00:23:06,240
enhance that through crowdsourcing

547
00:23:10,039 --> 00:23:07,950
efforts and actually increase the

548
00:23:12,889 --> 00:23:10,049
ability for that image processing

549
00:23:14,659 --> 00:23:12,899
algorithm to actually be created even

550
00:23:18,769 --> 00:23:14,669
better the same thing to be applied

551
00:23:20,779 --> 00:23:18,779
towards Mars moon or the or on the

552
00:23:22,430 --> 00:23:20,789
surface of Mars or the surface of the

553
00:23:25,369 --> 00:23:22,440
Moon or any planetary surface for that

554
00:23:28,610 --> 00:23:25,379
matter if you want to understand how to

555
00:23:30,830 --> 00:23:28,620
characterize craters impact zones find

556
00:23:33,169 --> 00:23:30,840
unique features on a planetary surface

557
00:23:36,049 --> 00:23:33,179
in a certain way but you can't search

558
00:23:37,610 --> 00:23:36,059
the two billion plus images even through

559
00:23:39,980 --> 00:23:37,620
crowdsourcing you can't get two billion

560
00:23:41,990 --> 00:23:39,990
images processed you could take a small

561
00:23:43,100 --> 00:23:42,000
set of those say a hundred thousand

562
00:23:45,340 --> 00:23:43,110
images two hundred thousand images

563
00:23:48,799 --> 00:23:45,350

process those through a more proud

564
00:23:52,009 --> 00:23:48,809
engagement period use that crowd of data

565
00:23:53,960 --> 00:23:52,019
then feed and train your algorithm in

566
00:23:56,600 --> 00:23:53,970
order to then run on the rest of the two

567
00:23:59,029 --> 00:23:56,610
billion images that are out there so

568
00:24:01,610 --> 00:23:59,039
what we are looking for is an algorithm

569
00:24:03,560 --> 00:24:01,620
that helps us scale pass crowd sourcing

570
00:24:06,019 --> 00:24:03,570
xin puts a couple hundred thousand

571
00:24:07,820 --> 00:24:06,029
images to the ability to go into the

572
00:24:11,299 --> 00:24:07,830
billions of images and process through

573
00:24:12,919 --> 00:24:11,309
that but really using the human

574
00:24:15,529 --> 00:24:12,929
intellect in order to train that over

575
00:24:20,869 --> 00:24:15,539
them yeah let me maybe jump in for a

576
00:24:22,700 --> 00:24:20,879
second as well you know this is concept

577
00:24:25,090 --> 00:24:22,710
right when we're trying to tap it in the

578
00:24:28,390 --> 00:24:25,100
human intellect at scale something that

579
00:24:31,600 --> 00:24:28,400
I have such broad applications in space

580
00:24:35,169 --> 00:24:31,610
and science and and you know in any type

581
00:24:37,659 --> 00:24:35,179
of big data analytics but the answers

582
00:24:39,970 --> 00:24:37,669
don't necessarily come from a single

583
00:24:41,890 --> 00:24:39,980
place right i mean it would take you

584
00:24:42,970 --> 00:24:41,900
more than a lifetime to try to become an

585
00:24:46,090 --> 00:24:42,980
expert in every different type of

586
00:24:48,250 --> 00:24:46,100
machine learning you know computer

587
00:24:49,779 --> 00:24:48,260
vision application out there a process

588
00:24:51,820 --> 00:24:49,789

out there and and what we're doing I

589
00:24:55,690 --> 00:24:51,830
think that this challenge is not saying

590
00:24:58,270 --> 00:24:55,700
that it's you know as cream said it's

591
00:25:00,340 --> 00:24:58,280
not a substitute for in-house innovation

592
00:25:03,520 --> 00:25:00,350
it's a compliment into all types of

593
00:25:05,669 --> 00:25:03,530
innovation that we can now say you know

594
00:25:07,810 --> 00:25:05,679
to you the whole community up there

595
00:25:09,760 --> 00:25:07,820
let's try every different process

596
00:25:12,899 --> 00:25:09,770
possible and see which one emerges with

597
00:25:16,450 --> 00:25:12,909
the most efficient most intelligent

598
00:25:18,430 --> 00:25:16,460
algorithm to learn from you know Rob

599
00:25:21,100 --> 00:25:18,440
source data right and the applications

600
00:25:22,720 --> 00:25:21,110
afterwards that will follow you know

601
00:25:24,700 --> 00:25:22,730
that that's a second part of this but

602
00:25:27,789 --> 00:25:24,710
but we're trying to get from this

603
00:25:30,060 --> 00:25:27,799
challenges is a tap into as many ideas

604
00:25:32,470 --> 00:25:30,070
and as many avenues as possible in the

605
00:25:34,029 --> 00:25:32,480
ways in which you might develop this

606
00:25:37,659 --> 00:25:34,039
algorithm and come up with the best

607
00:25:39,310 --> 00:25:37,669
solution problem you know this open type

608
00:25:43,180 --> 00:25:39,320
of innovation that's that's where it

609
00:25:45,279 --> 00:25:43,190
gets really exciting so so clearly one

610
00:25:47,649 --> 00:25:45,289
of the benefits of contest is to get

611
00:25:50,350 --> 00:25:47,659
solutions from lots of different minds

612
00:25:52,450 --> 00:25:50,360
lots of different people um which kind

613
00:25:55,810 --> 00:25:52,460

of begs the question which I'll direct

614
00:25:58,539 --> 00:25:55,820
au crème of what kind of experience do

615
00:26:02,890 --> 00:25:58,549
people need to try this at all not to

616
00:26:07,870 --> 00:26:02,900
participate in the challenge um you know

617
00:26:10,330 --> 00:26:07,880
our our our results of our analysis of a

618
00:26:12,820 --> 00:26:10,340
variety of crowd sourcing platforms show

619
00:26:14,710 --> 00:26:12,830
that in fact you know the winners tend

620
00:26:17,320 --> 00:26:14,720
to come from areas that are outside the

621
00:26:19,720 --> 00:26:17,330
domain of the problem so I mean I think

622
00:26:21,370 --> 00:26:19,730
if you can code you understand machine

623
00:26:23,230 --> 00:26:21,380
learning and you're excited by the

624
00:26:26,830 --> 00:26:23,240
contest I would encourage you to

625
00:26:29,799 --> 00:26:26,840
participate you know what's what's very

626
00:26:32,140 --> 00:26:29,809
cool about this setup is that you get

627
00:26:33,700 --> 00:26:32,150
immediate feedback both about how well

628
00:26:36,850 --> 00:26:33,710
you're doing how well your coding is

629
00:26:38,830 --> 00:26:36,860
going on but how you also rank and stack

630
00:26:40,150 --> 00:26:38,840
against everybody else

631
00:26:42,250 --> 00:26:40,160
and I think that's a real benefit

632
00:26:44,350 --> 00:26:42,260
oftentimes when we're working inside of

633
00:26:46,690 --> 00:26:44,360
our companies or an academia or in a lab

634
00:26:49,060 --> 00:26:46,700
we don't get to compare ourselves we

635
00:26:51,340 --> 00:26:49,070
don't get to see how the people might do

636
00:26:53,410 --> 00:26:51,350
it other people might approach approach

637
00:26:54,910 --> 00:26:53,420
the problem and the only way that you're

638
00:26:56,350 --> 00:26:54,920

going to learn about this Elena Lee and

639
00:26:58,870 --> 00:26:56,360
really improve yourself is by

640
00:27:00,730 --> 00:26:58,880
participating and hanging on and then

641
00:27:03,370 --> 00:27:00,740
trying to improve your own own an

642
00:27:05,710 --> 00:27:03,380
understanding of the problem and then at

643
00:27:08,920 --> 00:27:05,720
the end of the contest what's so great

644
00:27:11,710 --> 00:27:08,930
about about these settings is that once

645
00:27:13,060 --> 00:27:11,720
the you know the contest over you know

646
00:27:15,940 --> 00:27:13,070
the first thing that happens in the

647
00:27:18,850 --> 00:27:15,950
topic of the community is the the post

648
00:27:20,290 --> 00:27:18,860
of share share your approach so people

649
00:27:22,090 --> 00:27:20,300
immediately start sharing their

650
00:27:24,820 --> 00:27:22,100
approaches and learn from each other as

651
00:27:26,950 --> 00:27:24,830
to what worked and what didn't work and

652
00:27:28,660 --> 00:27:26,960
that's that kind of a teaching that kind

653
00:27:30,640 --> 00:27:28,670
of a learning you can't get even at

654
00:27:32,680 --> 00:27:30,650
Harbor you know you have to actually be

655
00:27:34,780 --> 00:27:32,690
involved in the problem he's struggling

656
00:27:36,940 --> 00:27:34,790
with it over the time period and then

657
00:27:39,130 --> 00:27:36,950
learn from other people so say how their

658
00:27:41,830 --> 00:27:39,140
approach I you know made a difference

659
00:27:43,960 --> 00:27:41,840
and so I mean I think that's that's what

660
00:27:47,020 --> 00:27:43,970
I'd recommend and we know for a fact

661
00:27:49,120 --> 00:27:47,030
that the one of the great things that

662
00:27:51,700 --> 00:27:49,130
crowdsourcing unleashes that these

663
00:27:53,470 --> 00:27:51,710

contests and leashes you know diversity

664
00:27:56,080 --> 00:27:53,480
diversity of perspectives on how to

665
00:27:58,060 --> 00:27:56,090
solve the problem I will get lots of

666
00:27:59,740 --> 00:27:58,070
entries but within that are many many

667
00:28:02,860 --> 00:27:59,750
many many different ways to attack the

668
00:28:08,290 --> 00:28:02,870
problem and those things go really long

669
00:28:09,580 --> 00:28:08,300
way to to to to to make you a better

670
00:28:12,270 --> 00:28:09,590
programmer to understand machine

671
00:28:14,710 --> 00:28:12,280
learning to understand how this very

672
00:28:17,230 --> 00:28:14,720
intricate data sets can be used for

673
00:28:19,270 --> 00:28:17,240
these types of vacations does anybody

674
00:28:22,180 --> 00:28:19,280
find it funny that we're crowdsourcing

675
00:28:26,950 --> 00:28:22,190
the learning of crowdsourcing yeah it's

676
00:28:29,310 --> 00:28:26,960
pretty cool well Barry I'm folks come to

677
00:28:34,480 --> 00:28:29,320
top coder it's better than Harvard I

678
00:28:38,290 --> 00:28:34,490
didn't that you opposed to fine I'll

679
00:28:40,300 --> 00:28:38,300
take my Tucker hat back off um well I'll

680
00:28:42,220 --> 00:28:40,310
put it back on just real quick for one

681
00:28:44,430 --> 00:28:42,230
minor thing and that's that the this

682
00:28:47,830 --> 00:28:44,440
particular challenge is is pretty

683
00:28:52,659 --> 00:28:47,840
language agnostic so folks can compete

684
00:28:55,479 --> 00:28:52,669
in this thing in in Python and c++ and

685
00:28:57,669 --> 00:28:55,489
TV in Java um the idea is to get the

686
00:29:00,369 --> 00:28:57,679
best ideas not not necessarily the ones

687
00:29:02,200 --> 00:29:00,379
written in a specific language okay how

688
00:29:06,399 --> 00:29:02,210

about you do assembler can I send

689
00:29:13,570 --> 00:29:06,409
something the seller no garage we're not

690
00:29:17,889 --> 00:29:13,580
evil um okay so one other questions that

691
00:29:19,419 --> 00:29:17,899
it's coming in here is it's we're all

692
00:29:21,580 --> 00:29:19,429
familiar with their most people familiar

693
00:29:24,369 --> 00:29:21,590
with the ansari x prize or sorry x prize

694
00:29:26,590 --> 00:29:24,379
right it was one gigantic prize to

695
00:29:29,139 --> 00:29:26,600
accomplish this one gigantic ly

696
00:29:31,989 --> 00:29:29,149
difficult task you know build a rocket

697
00:29:35,409 --> 00:29:31,999
um is that the way all challenges are

698
00:29:39,009 --> 00:29:35,419
formulated or do they have first second

699
00:29:40,149 --> 00:29:39,019
third and so on so I'll talk a little

700
00:29:42,489 --> 00:29:40,159
bit about what we've done in the cream

701
00:29:45,460 --> 00:29:42,499
can talk about in general and l2 but

702
00:29:47,200 --> 00:29:45,470
NASA actually runs kind of challenges

703
00:29:49,269 --> 00:29:47,210
and prizes on multiple different scales

704
00:29:50,940 --> 00:29:49,279
our NASA tournament lab activities are

705
00:29:54,869 --> 00:29:50,950
centered around software and algorithms

706
00:29:57,220 --> 00:29:54,879
solutions with prizes and the order of

707
00:29:59,680 --> 00:29:57,230
10,000 to thirty thousand dollar type

708
00:30:02,470 --> 00:29:59,690
ranges and they're kind of commensurate

709
00:30:03,909 --> 00:30:02,480
to the amount of time that that somebody

710
00:30:06,099 --> 00:30:03,919
needs to be able to put in in order to

711
00:30:09,399 --> 00:30:06,109
solve that problem we also NASA have

712
00:30:11,049 --> 00:30:09,409
other large-scale prize initiatives like

713
00:30:14,409 --> 00:30:11,059

say our Centennial challenge program

714
00:30:16,299 --> 00:30:14,419
going after things like green aircraft

715
00:30:17,259 --> 00:30:16,309
greevey green aviation challenge but

716
00:30:19,029 --> 00:30:17,269
people are actually building a

717
00:30:21,220 --> 00:30:19,039
full-fledged aircraft much like the

718
00:30:24,639 --> 00:30:21,230
ansari x prize is building a spacecraft

719
00:30:26,519 --> 00:30:24,649
and the commensurate prize for those who

720
00:30:29,950 --> 00:30:26,529
goes into the million-dollar type plus

721
00:30:31,869 --> 00:30:29,960
range so one of the things that's really

722
00:30:34,029 --> 00:30:31,879
important with this is yes you need a

723
00:30:37,389 --> 00:30:34,039
proper incentive but you can /

724
00:30:39,419 --> 00:30:37,399
incentivize as well and so you need to

725
00:30:42,249 --> 00:30:39,429
kind of pair the size of incentive that

726
00:30:43,810 --> 00:30:42,259
would give people to participate with

727
00:30:45,549 --> 00:30:43,820
the the level of difficulty of the

728
00:30:47,979 --> 00:30:45,559
challenge or how complex the challenge

729
00:30:49,570 --> 00:30:47,989
potentially is then you got to be

730
00:30:52,090 --> 00:30:49,580
careful and is sometimes you can

731
00:30:56,049 --> 00:30:52,100
actually univer tantly create a barrier

732
00:30:57,639 --> 00:30:56,059
to participate in to and say if you had

733
00:31:00,279 --> 00:30:57,649
a really hard challenge and you just

734
00:31:01,509 --> 00:31:00,289
wanted to get headlines and say it's a

735
00:31:03,820 --> 00:31:01,519
hundred-thousand-dollar or a

736
00:31:05,919 --> 00:31:03,830
million-dollar prize well people can

737
00:31:06,490 --> 00:31:05,929
find that intimidating do I have do I as

738
00:31:08,580 --> 00:31:06,500

an individual

739
00:31:11,170 --> 00:31:08,590
have a million our million-dollar idea

740
00:31:12,700 --> 00:31:11,180
and that may actually prohibit some

741
00:31:13,900 --> 00:31:12,710
people from actually participating just

742
00:31:17,350 --> 00:31:13,910
because they don't feel that they can

743
00:31:19,540 --> 00:31:17,360
whistle hey we're in other cases is does

744
00:31:22,690 --> 00:31:19,550
everybody have a 10,000 idea dollar idea

745
00:31:24,850 --> 00:31:22,700
of 30,000 Valerie absolutely those are

746
00:31:26,290 --> 00:31:24,860
those ideas can come in so that you get

747
00:31:29,140 --> 00:31:26,300
there's a little bit of a pairing with

748
00:31:34,660 --> 00:31:29,150
the incentives along with types of

749
00:31:36,400 --> 00:31:34,670
challenge that Iran thanks Jason so

750
00:31:37,900 --> 00:31:36,410
cream I think he touched on it before

751
00:31:40,240 --> 00:31:37,910
but we know that you're the principal

752
00:31:44,980 --> 00:31:40,250
investigator for the Harvard NASA torn

753
00:31:46,930 --> 00:31:44,990
slab um can you explain a little bit of

754
00:31:49,570 --> 00:31:46,940
how you're planning to use this

755
00:32:00,040 --> 00:31:49,580
algorithm and what sort of research

756
00:32:03,490 --> 00:32:00,050
you're gathering with this no okay it's

757
00:32:05,890 --> 00:32:03,500
a secret I know our our hope is that

758
00:32:08,680 --> 00:32:05,900
we're actually going to be able to write

759
00:32:11,350 --> 00:32:08,690
a scientific paper about crowdsourcing

760
00:32:14,290 --> 00:32:11,360
for both sides so you know in

761
00:32:17,230 --> 00:32:14,300
collaboration with Albert and sort of

762
00:32:20,470 --> 00:32:17,240
show how crowd labeling and enable them

763
00:32:22,870 --> 00:32:20,480

to actually you know make some very

764
00:32:24,880 --> 00:32:22,880
important archaeological discovery but

765
00:32:27,190 --> 00:32:24,890
then to bring it back to sort of say

766
00:32:28,600 --> 00:32:27,200
that in fact this data is actually very

767
00:32:31,600 --> 00:32:28,610
valuable and we can create

768
00:32:33,730 --> 00:32:31,610
general-purpose algorithms through

769
00:32:35,410 --> 00:32:33,740
crowdsourcing on the intellectual aside

770
00:32:37,150 --> 00:32:35,420
you know in terms of problem solving

771
00:32:39,130 --> 00:32:37,160
side just then come back and say that in

772
00:32:41,860 --> 00:32:39,140
fact this data can be used you can

773
00:32:44,050 --> 00:32:41,870
create new algorithms on how to how to

774
00:32:46,810 --> 00:32:44,060
assess this imagery so our hope is that

775
00:32:49,180 --> 00:32:46,820
as soon as this contest is over um to

776
00:32:51,970 --> 00:32:49,190
then engage maybe even with the crowd at

777
00:32:55,540 --> 00:32:51,980
top coder and elsewhere in writing up a

778
00:32:57,940 --> 00:32:55,550
paper that shows both sides and I think

779
00:33:02,230 --> 00:32:57,950
that's where there's going to be a lot

780
00:33:04,540 --> 00:33:02,240
of of interest because as Jason said as

781
00:33:06,970 --> 00:33:04,550
Albert said that in fact you know we are

782
00:33:09,430 --> 00:33:06,980
living in a world where image analysis

783
00:33:11,680 --> 00:33:09,440
is exploding in many in many

784
00:33:13,540 --> 00:33:11,690
applications many dimensions you you

785
00:33:15,820 --> 00:33:13,550
think about medical imaging that's all

786
00:33:19,500 --> 00:33:15,830
about human beings looking at at images

787
00:33:21,810 --> 00:33:19,510
x-rays PET scans MRI scans and then

788
00:33:25,050 --> 00:33:21,820

deciding if there's a tumor there is a

789
00:33:27,750 --> 00:33:25,060
fracture there or not and so on so

790
00:33:30,180 --> 00:33:27,760
there's a pathology styles you name it

791
00:33:32,670 --> 00:33:30,190
is a ton of applications just in medical

792
00:33:36,450 --> 00:33:32,680
imaging which could potentially really

793
00:33:37,560 --> 00:33:36,460
benefit from from this from this type B

794
00:33:39,540 --> 00:33:37,570
sense of algorithms if you just think

795
00:33:41,880 --> 00:33:39,550
about it hundreds of thousands of

796
00:33:44,070 --> 00:33:41,890
radiologists around the world are always

797
00:33:46,740 --> 00:33:44,080
labeling images right and coming up with

798
00:33:49,260 --> 00:33:46,750
diagnosis right this could actually be

799
00:33:51,690 --> 00:33:49,270
revolutionary in that side you know same

800
00:33:54,060 --> 00:33:51,700
thing and you know in detecting cracks

801
00:33:56,130 --> 00:33:54,070
and jet engines you name it a range of

802
00:33:58,890 --> 00:33:56,140
settings were in fact these algorithms

803
00:34:01,950 --> 00:33:58,900
can be put to use but we need both the

804
00:34:03,570 --> 00:34:01,960
human side of labeling with the Machine

805
00:34:06,750 --> 00:34:03,580
side of taking the minimal amount of

806
00:34:09,000 --> 00:34:06,760
labor required and and be able to sort

807
00:34:10,530 --> 00:34:09,010
through and scale up to the the

808
00:34:12,030 --> 00:34:10,540
terabytes of data that we're now now

809
00:34:13,860 --> 00:34:12,040
capturing so it's very exciting and I

810
00:34:16,680 --> 00:34:13,870
think you know I'm really hoping that it

811
00:34:19,169 --> 00:34:16,690
becomes a you know a great contribution

812
00:34:22,169 --> 00:34:19,179
to the two scientists and our hope is

813
00:34:23,639 --> 00:34:22,179

actually what's important for us is both

814
00:34:26,190 --> 00:34:23,649
for myself and for Albert and I think

815
00:34:29,100 --> 00:34:26,200
all of us here is that we don't want

816
00:34:30,570 --> 00:34:29,110
this kind of an approach to be view it

817
00:34:32,879 --> 00:34:30,580
as some kind of something exotic

818
00:34:34,379 --> 00:34:32,889
something on the sidelines like oh I'm

819
00:34:35,760 --> 00:34:34,389
interesting look at this price based

820
00:34:38,250 --> 00:34:35,770
contest look at this crowd sourcing

821
00:34:41,129 --> 00:34:38,260
thing our objective is to make this go

822
00:34:43,290 --> 00:34:41,139
mainstream let's find a way a show to

823
00:34:45,840 --> 00:34:43,300
our colleagues in the academic world and

824
00:34:48,210 --> 00:34:45,850
the industrial walls how you can combine

825
00:34:51,000 --> 00:34:48,220
both of these together because I think

826
00:34:53,490 --> 00:34:51,010
that's where they for us the excitement

827
00:34:55,020 --> 00:34:53,500
is like we're nowhere convince that this

828
00:34:56,879 --> 00:34:55,030
works this works really well we've done

829
00:34:58,950 --> 00:34:56,889
so many contests on so many platforms

830
00:35:00,150 --> 00:34:58,960
that this works really well and now

831
00:35:02,490 --> 00:35:00,160
we're going to push it to the next level

832
00:35:04,260 --> 00:35:02,500
then again show to our colleagues that

833
00:35:06,420 --> 00:35:04,270
in fact this is an approach that

834
00:35:12,150 --> 00:35:06,430
actually makes sense and can be can be

835
00:35:13,440 --> 00:35:12,160
scaled up yeah you know I'll cut them a

836
00:35:17,610 --> 00:35:13,450
little bit of my excitement about this

837
00:35:18,960 --> 00:35:17,620
as well if you don't mind you know the

838
00:35:20,940 --> 00:35:18,970

questions that we can answer are

839
00:35:23,760 --> 00:35:20,950
different than the ones that that we've

840
00:35:27,840 --> 00:35:23,770
tried to answer in the past that you

841
00:35:30,750 --> 00:35:27,850
know a lot of times we take for granted

842
00:35:31,810 --> 00:35:30,760
just how amazing human perception and

843
00:35:33,850 --> 00:35:31,820
intuition is

844
00:35:36,580 --> 00:35:33,860
especially in image analysis in the

845
00:35:39,610 --> 00:35:36,590
image categorizations and when you go

846
00:35:44,920 --> 00:35:39,620
through and think about josh are putting

847
00:35:48,640 --> 00:35:44,930
something on a medical imaging data set

848
00:35:50,080 --> 00:35:48,650
that's that takes a huge amount of human

849
00:35:53,620 --> 00:35:50,090
training and experience to do so

850
00:35:55,240 --> 00:35:53,630
correctly right but to scale that up you

851
00:35:56,560 --> 00:35:55,250
can't automate that right now right I

852
00:35:58,960 --> 00:35:56,570
mean there's no way in which you can

853
00:36:01,570 --> 00:35:58,970
automate that intuition because it's

854
00:36:03,520 --> 00:36:01,580
it's building that that visual

855
00:36:06,280 --> 00:36:03,530
perception is building upon everything

856
00:36:07,540 --> 00:36:06,290
that that person had experienced up to

857
00:36:10,000 --> 00:36:07,550
that point right everything they had

858
00:36:11,770 --> 00:36:10,010
learned but we can do is say well maybe

859
00:36:14,260 --> 00:36:11,780
we can start to extract out what is the

860
00:36:16,030 --> 00:36:14,270
essence of what they're identified right

861
00:36:17,650 --> 00:36:16,040
and that's what we're asking the crowd

862
00:36:20,650 --> 00:36:17,660
here to do is to figure out what is the

863
00:36:23,020 --> 00:36:20,660

essence of what they're identifying with

864
00:36:25,600 --> 00:36:23,030
as few of those examples as possible

865
00:36:28,600 --> 00:36:25,610
from that doctor or the network of

866
00:36:30,340 --> 00:36:28,610
doctors right in our case it's people

867
00:36:33,490 --> 00:36:30,350
helping us use their human intuition to

868
00:36:35,680 --> 00:36:33,500
see what was strange about a location in

869
00:36:37,510 --> 00:36:35,690
you know in a piece of satellite imagery

870
00:36:39,090 --> 00:36:37,520
right what looked ancient what looks

871
00:36:44,410 --> 00:36:39,100
weird we didn't give any a kind of

872
00:36:46,330 --> 00:36:44,420
example to to train an individual with

873
00:36:47,950 --> 00:36:46,340
what what was weird or what was strange

874
00:36:49,930 --> 00:36:47,960
we just asked for that human perception

875
00:36:52,780 --> 00:36:49,940
and now we're asking the crowd of coders

876
00:36:54,190 --> 00:36:52,790
to come up with a way to just extract

877
00:36:56,290 --> 00:36:54,200
the essence of what they identified

878
00:36:58,660 --> 00:36:56,300
right and the kinds of questions we can

879
00:37:04,060 --> 00:36:58,670
answer from you know from that kind of

880
00:37:07,630 --> 00:37:04,070
process I think broad and wild and I'm

881
00:37:10,150 --> 00:37:07,640
excited to see what comes of it so so

882
00:37:13,450 --> 00:37:10,160
one observation then clearly challenges

883
00:37:16,110 --> 00:37:13,460
are not only programming they're also or

884
00:37:18,700 --> 00:37:16,120
building rockets it's so in your case

885
00:37:23,050 --> 00:37:18,710
albert where you have amateurs

886
00:37:24,460 --> 00:37:23,060
annotating images um and karim it sounds

887
00:37:26,470 --> 00:37:24,470
like maybe someday there's some kind of

888
00:37:32,400 --> 00:37:26,480

convergence with doing that with medical

889
00:37:36,100 --> 00:37:32,410
imaging as well so that that brings up

890
00:37:37,960 --> 00:37:36,110
one of our I think we're reaching the

891
00:37:39,970 --> 00:37:37,970
end of our questions although if

892
00:37:41,740 --> 00:37:39,980
anyone's following along that would like

893
00:37:44,620 --> 00:37:41,750
that ask question that hasn't please be

894
00:37:45,580 --> 00:37:44,630
sure to post it to Twitter hashtag and

895
00:37:49,450 --> 00:37:45,590
TL

896
00:37:51,790 --> 00:37:49,460
going to read it off here so one

897
00:37:54,430 --> 00:37:51,800
question that I know that a we're all

898
00:37:57,540 --> 00:37:54,440
out burning to ask you cream is um

899
00:38:02,110 --> 00:37:57,550
what's the craziest outcome you've found

900
00:38:06,910 --> 00:38:02,120
using using crowdsourcing crazy as

901
00:38:09,340 --> 00:38:06,920
outcome yeah I mean I think I you know I

902
00:38:11,950 --> 00:38:09,350
I think two recent examples I mean I

903
00:38:14,800 --> 00:38:11,960
think our lingerie on challenge on you

904
00:38:18,040 --> 00:38:14,810
know to come up with the algorithms for

905
00:38:20,800 --> 00:38:18,050
peak angles off for the space station I

906
00:38:23,050 --> 00:38:20,810
mean I think it's it's incredible to

907
00:38:24,010 --> 00:38:23,060
think that in a 23 big challenge I don't

908
00:38:26,470 --> 00:38:24,020
remember how long we've been the

909
00:38:32,980 --> 00:38:26,480
challenge for uh you know we were able

910
00:38:35,850 --> 00:38:32,990
to you know come up with very robust

911
00:38:40,270 --> 00:38:35,860
algorithms using very unique approaches

912
00:38:42,340 --> 00:38:40,280
that that meets and in some cases beats

913
00:38:46,690 --> 00:38:42,350

the requirements that I've been set by

914
00:38:50,380 --> 00:38:46,700
NASA and so so that in itself is just

915
00:38:52,780 --> 00:38:50,390
it's just so um you know to be quite

916
00:38:54,970 --> 00:38:52,790
honest I'm like that's my default now

917
00:38:56,650 --> 00:38:54,980
like if it doesn't if it doesn't improve

918
00:38:58,120 --> 00:38:56,660
doesn't reach or improve by orders of

919
00:39:01,360 --> 00:38:58,130
magnitude I might disappoint it and I've

920
00:39:02,830 --> 00:39:01,370
yet to be disappointed so so it's always

921
00:39:04,570 --> 00:39:02,840
exciting to sort of see that come

922
00:39:06,490 --> 00:39:04,580
through so I think the long run

923
00:39:09,070 --> 00:39:06,500
challenge is a great example of us you

924
00:39:12,240 --> 00:39:09,080
know using you know a scaled-down model

925
00:39:14,530 --> 00:39:12,250
I'm using very difficult circumstances

926
00:39:17,500 --> 00:39:14,540
very difficult data a very difficult

927
00:39:19,780 --> 00:39:17,510
problem and you know the community of

928
00:39:22,450 --> 00:39:19,790
more than foreigner submitters did an

929
00:39:24,960 --> 00:39:22,460
amazing job of solving that problem I

930
00:39:27,430 --> 00:39:24,970
mean it's just like it's unbelievable uh

931
00:39:29,530 --> 00:39:27,440
what they were able to get done you know

932
00:39:31,570 --> 00:39:29,540
similarly a few years back we ran a

933
00:39:34,090 --> 00:39:31,580
challenge and genomic analysis in

934
00:39:38,230 --> 00:39:34,100
sequence alignment and we published a

935
00:39:40,060 --> 00:39:38,240
paper on this and you know we are orders

936
00:39:42,660 --> 00:39:40,070
of magnitude better than the Harvard

937
00:39:44,950 --> 00:39:42,670
solution and then also the NIH solution

938
00:39:46,890 --> 00:39:44,960

with people with relatively no

939
00:39:50,380 --> 00:39:46,900
background and computational biology uh

940
00:39:53,650 --> 00:39:50,390
so you know for me now it's sort of like

941
00:39:56,710 --> 00:39:53,660
you know I'm you know with the lab and

942
00:39:58,050 --> 00:39:56,720
so forth this you know this is this is

943
00:40:04,740 --> 00:39:58,060
such a

944
00:40:07,230 --> 00:40:04,750
the moment technical challenges and and

945
00:40:09,390 --> 00:40:07,240
and solve them really well that we

946
00:40:11,220 --> 00:40:09,400
should be again and this is part of our

947
00:40:13,350 --> 00:40:11,230
mission at the lab and the work we're

948
00:40:15,900 --> 00:40:13,360
doing with Jason and NASA and with top

949
00:40:18,590 --> 00:40:15,910
coder is how do we make this go

950
00:40:23,010 --> 00:40:18,600
mainstream how do we encourage how do we

951
00:40:25,110 --> 00:40:23,020
create the pathways for scientists or

952
00:40:27,030 --> 00:40:25,120
engineers for technologists around the

953
00:40:30,690 --> 00:40:27,040
world to see this as a great opportunity

954
00:40:34,110 --> 00:40:30,700
to move further their technical projects

955
00:40:35,910 --> 00:40:34,120
at the same time to invite people from

956
00:40:37,890 --> 00:40:35,920
around the world to participate in

957
00:40:40,680 --> 00:40:37,900
problem solving the fact that we can

958
00:40:42,870 --> 00:40:40,690
actually post real-life NASA challenges

959
00:40:44,460 --> 00:40:42,880
the fact we can post real-life National

960
00:40:46,020 --> 00:40:44,470
Geographic challenges the fact that we

961
00:40:47,940 --> 00:40:46,030
can invite many more people into the

962
00:40:53,420 --> 00:40:47,950
problem solving process and before it's

963
00:40:56,100 --> 00:40:53,430

just it's just fantastic thanks cream

964
00:40:59,400 --> 00:40:56,110
well folks I don't see any additional

965
00:41:00,870 --> 00:40:59,410
questions on our ntl stream so I'd like

966
00:41:02,760 --> 00:41:00,880
to throw one in there that came in

967
00:41:06,450 --> 00:41:02,770
through one of our other channels give

968
00:41:09,870 --> 00:41:06,460
folks a chance to ask any more questions

969
00:41:12,390 --> 00:41:09,880
on Twitter before we close it out and my

970
00:41:15,000 --> 00:41:12,400
question is this one here a top coder

971
00:41:17,400 --> 00:41:15,010
one question to get a lot of and that

972
00:41:19,680 --> 00:41:17,410
has we've gotten specifically in regard

973
00:41:22,440 --> 00:41:19,690
to this challenge it is when you're

974
00:41:25,440 --> 00:41:22,450
looking for a collaborative result or

975
00:41:26,610 --> 00:41:25,450
you're looking for a problem that here

976
00:41:28,940 --> 00:41:26,620
you're trying to solve a problem that

977
00:41:31,230 --> 00:41:28,950
typically requires huge amount of

978
00:41:34,410 --> 00:41:31,240
cooperation or collaboration how can you

979
00:41:37,470 --> 00:41:34,420
possibly run a contest around them so

980
00:41:40,380 --> 00:41:37,480
that's one question I I posted the group

981
00:41:43,680 --> 00:41:40,390
you know I do you limit your challenges

982
00:41:45,480 --> 00:41:43,690
to to only those things that a lone

983
00:41:48,330 --> 00:41:45,490
agent working at home we'd be able to

984
00:41:50,250 --> 00:41:48,340
solve uh or is it something that you

985
00:41:55,680 --> 00:41:50,260
think can work wouldn't appropriately

986
00:41:57,900 --> 00:41:55,690
asked in scientific field as well so so

987
00:42:01,530 --> 00:41:57,910
some of the challenges of Iran be we

988
00:42:04,080 --> 00:42:01,540

obviously go after and structure things

989
00:42:06,150 --> 00:42:04,090
as an individual competing in other

990
00:42:07,980 --> 00:42:06,160
cases when the problem is more complex

991
00:42:11,970 --> 00:42:07,990
or far as multiple disciplinary

992
00:42:14,370 --> 00:42:11,980
approaches teaming Arrangements can

993
00:42:16,380 --> 00:42:14,380
also come into play and in fact a lot of

994
00:42:18,599 --> 00:42:16,390
our other challenge types that are not

995
00:42:21,690 --> 00:42:18,609
the software algorithm on are actually

996
00:42:23,760 --> 00:42:21,700
allow team type approaches and such for

997
00:42:26,370 --> 00:42:23,770
that in this case it happens to be

998
00:42:29,070 --> 00:42:26,380
individuals the resources that are

999
00:42:32,220 --> 00:42:29,080
needed for this can be readily had with

1000
00:42:35,400 --> 00:42:32,230
a computer at home and there isn't a

1001
00:42:37,410 --> 00:42:35,410
need for a real big teaming tech

1002
00:42:39,060 --> 00:42:37,420
approach and there's a bit of the

1003
00:42:40,680 --> 00:42:39,070
incentive piece that probably creamy

1004
00:42:43,770 --> 00:42:40,690
could talk about as well about how

1005
00:42:47,700 --> 00:42:43,780
individual versus team incentives do

1006
00:42:49,320 --> 00:42:47,710
play out in this as well be it really

1007
00:42:52,530 --> 00:42:49,330
depends on how you want to orchestrate

1008
00:42:54,120 --> 00:42:52,540
it what kind of involvement do you want

1009
00:42:57,870 --> 00:42:54,130
from a community what are the one of the

1010
00:42:59,580 --> 00:42:57,880
community norms as well so if you're

1011
00:43:01,710 --> 00:42:59,590
using an existing communities out there

1012
00:43:04,710 --> 00:43:01,720
there are certain norms that they are

1013
00:43:06,930 --> 00:43:04,720

used to and and in the case of our as a

1014
00:43:09,390 --> 00:43:06,940
tournament lab the community we're using

1015
00:43:12,390 --> 00:43:09,400
obviously is top coders which the the

1016
00:43:14,160 --> 00:43:12,400
community norm is individuals competing

1017
00:43:16,380 --> 00:43:14,170
and there's actually individual

1018
00:43:18,810 --> 00:43:16,390
achievement badges and and there's a

1019
00:43:20,940 --> 00:43:18,820
sense of kind of a self achievement

1020
00:43:23,700 --> 00:43:20,950
there and that works very well with that

1021
00:43:27,390 --> 00:43:23,710
type of community but there's other

1022
00:43:28,770 --> 00:43:27,400
other methodologies as well yeah I think

1023
00:43:31,710 --> 00:43:28,780
just to build on jason said I mean I

1024
00:43:34,560 --> 00:43:31,720
think you know there's various ways in

1025
00:43:36,840 --> 00:43:34,570
which you can have multiple people

1026
00:43:44,160 --> 00:43:36,850
participate on solving the same problem

1027
00:43:48,930 --> 00:43:44,170
a one of course is I'm done this outside

1028
00:43:50,970 --> 00:43:48,940
is to go you know do release

1029
00:43:53,040 --> 00:43:50,980
intermediate solutions for other people

1030
00:43:55,560 --> 00:43:53,050
to work on so you have stage one you

1031
00:43:57,900 --> 00:43:55,570
know people are competing the winners of

1032
00:43:59,970 --> 00:43:57,910
those stage one and then release that

1033
00:44:01,260 --> 00:43:59,980
code is released for stage 2 and the

1034
00:44:02,760 --> 00:44:01,270
people look at that learn from that and

1035
00:44:05,220 --> 00:44:02,770
improve upon that so that's one way you

1036
00:44:08,250 --> 00:44:05,230
can bring in a bit of cooperation in a

1037
00:44:09,660 --> 00:44:08,260
fully competition mode um I think a lot

1038
00:44:12,150 --> 00:44:09,670

of it depends on the problem you're

1039
00:44:15,090 --> 00:44:12,160
trying to solve as well uh you know you

1040
00:44:16,980 --> 00:44:15,100
could take a problem which has multiple

1041
00:44:19,590 --> 00:44:16,990
dimensions and requires multiple skill

1042
00:44:21,840 --> 00:44:19,600
sets and that may be best suited to

1043
00:44:23,490 --> 00:44:21,850
folks that have those dimensions

1044
00:44:25,100 --> 00:44:23,500
available to them in their teams and

1045
00:44:27,500 --> 00:44:25,110
potentially work with them on that

1046
00:44:29,090 --> 00:44:27,510
but the issue often becomes is you know

1047
00:44:31,100 --> 00:44:29,100
how do you how does somebody find those

1048
00:44:32,660 --> 00:44:31,110
other other teammates and how do we how

1049
00:44:34,040 --> 00:44:32,670
do we make that happen and you've

1050
00:44:35,960 --> 00:44:34,050
actually done some scientific work in

1051
00:44:38,060 --> 00:44:35,970
that space and I'm on multiple

1052
00:44:41,000 --> 00:44:38,070
dimensions in that in that area um and

1053
00:44:43,730 --> 00:44:41,010
then you can also run you know purely

1054
00:44:45,230 --> 00:44:43,740
wiki like challenges where everything is

1055
00:44:47,150 --> 00:44:45,240
a little for everybody else to use and

1056
00:44:50,630 --> 00:44:47,160
reuse and we've also done that as well I

1057
00:44:52,820 --> 00:44:50,640
think I think for us well the first

1058
00:44:55,390 --> 00:44:52,830
concern always is that are the

1059
00:44:57,470 --> 00:44:55,400
incentives match the context and the

1060
00:45:01,610 --> 00:44:57,480
context match the character of the

1061
00:45:04,880 --> 00:45:01,620
community and so we all we will always

1062
00:45:07,010 --> 00:45:04,890
create a you know when you run most of

1063
00:45:09,170 --> 00:45:07,020

our top coder work is to take problems

1064
00:45:12,770 --> 00:45:09,180
that an individual can solve on their

1065
00:45:14,330 --> 00:45:12,780
own and can do that can do it in a

1066
00:45:16,100 --> 00:45:14,340
competitive fashion but that hasn't

1067
00:45:19,160 --> 00:45:16,110
prevented us from as you give it in the

1068
00:45:21,250 --> 00:45:19,170
top coder platform to in fact run very

1069
00:45:23,900 --> 00:45:21,260
interesting cooperation type of

1070
00:45:26,630 --> 00:45:23,910
scenarios inside of a competitor frame

1071
00:45:29,180 --> 00:45:26,640
lock you know in this particular case

1072
00:45:33,350 --> 00:45:29,190
you know just working with your teams

1073
00:45:35,900 --> 00:45:33,360
across HBS and until NASA you know

1074
00:45:39,830 --> 00:45:35,910
people well world-class data scientists

1075
00:45:41,570 --> 00:45:39,840
like we're not sir give and and so my

1076
00:45:44,330 --> 00:45:41,580
colleagues here like dirt Lantry to come

1077
00:45:46,490 --> 00:45:44,340
up with the strategy in which we asked

1078
00:45:48,860 --> 00:45:46,500
the question you know our problem

1079
00:45:51,950 --> 00:45:48,870
statement has been developed over a

1080
00:45:54,050 --> 00:45:51,960
series of months you know of thinking

1081
00:45:55,610 --> 00:45:54,060
about how we can ask a technical

1082
00:46:01,490 --> 00:45:55,620
question in the appropriate way and it's

1083
00:46:04,310 --> 00:46:01,500
I hope you guys find it interesting so

1084
00:46:06,980 --> 00:46:04,320
we've had another question come in over

1085
00:46:08,540 --> 00:46:06,990
aah over Twitter which I'll read off now

1086
00:46:15,230 --> 00:46:08,550
if we still have time right I think we

1087
00:46:18,320 --> 00:46:15,240
do and that question is is it the is

1088
00:46:20,060 --> 00:46:18,330

this a so I think this is the contest

1089
00:46:21,440 --> 00:46:20,070
formulation is a significant limitation

1090
00:46:28,310 --> 00:46:21,450
factor for critical thinking especially

1091
00:46:31,130 --> 00:46:28,320
for dark large data mining problems is

1092
00:46:33,230 --> 00:46:31,140
that is that another Jason would you

1093
00:46:35,420 --> 00:46:33,240
like to your crack at that one or green

1094
00:46:37,730 --> 00:46:35,430
or MIT a little better is it a

1095
00:46:39,160 --> 00:46:37,740
significant limited limitation factor

1096
00:46:44,120 --> 00:46:39,170
for critical thinking especial

1097
00:46:45,710 --> 00:46:44,130
um well it's critical thinking as far as

1098
00:46:49,160 --> 00:46:45,720
far as making decisions from large data

1099
00:46:51,890 --> 00:46:49,170
sources just to have large data by

1100
00:46:54,170 --> 00:46:51,900
itself is in fact a lot of cases

1101
00:46:56,390 --> 00:46:54,180
completely useless if you can't actually

1102
00:46:59,000 --> 00:46:56,400
ascertain what the what is in the data

1103
00:47:02,450 --> 00:46:59,010
and actually get context of the data in

1104
00:47:03,860 --> 00:47:02,460
some way then your ability to do

1105
00:47:07,520 --> 00:47:03,870
critical thinking and decision making

1106
00:47:10,190 --> 00:47:07,530
off of the data source is extremely

1107
00:47:11,900 --> 00:47:10,200
challenged I mean in many ways what

1108
00:47:14,750 --> 00:47:11,910
we're trying to to do with this

1109
00:47:17,720 --> 00:47:14,760
challenge is to set up methodology where

1110
00:47:19,670 --> 00:47:17,730
we can look at thinking raw data sources

1111
00:47:21,710 --> 00:47:19,680
put it out there allow the human

1112
00:47:24,590 --> 00:47:21,720
intuition and such to actually look at

1113
00:47:28,640 --> 00:47:24,600

things and items critically identify

1114
00:47:30,680 --> 00:47:28,650
patterns and or pieces in the data but

1115
00:47:32,150 --> 00:47:30,690
then we can actually train the algorithm

1116
00:47:34,580 --> 00:47:32,160
and then go off and find and run that on

1117
00:47:36,050 --> 00:47:34,590
the entire data set and from that then

1118
00:47:37,580 --> 00:47:36,060
you can start doing critical decision

1119
00:47:42,110 --> 00:47:37,590
making in critical thinking about that

1120
00:47:45,170 --> 00:47:42,120
data set versus just having the data and

1121
00:47:46,670 --> 00:47:45,180
NASA we bring tons of data back every

1122
00:47:48,710 --> 00:47:46,680
one of our missions more data than

1123
00:47:50,000 --> 00:47:48,720
actually then we could ever analyzed or

1124
00:47:52,490 --> 00:47:50,010
even the science community in general

1125
00:47:54,380 --> 00:47:52,500
could actually analyze in hopes that

1126
00:47:56,750 --> 00:47:54,390
someday people will be able to go back

1127
00:47:59,750 --> 00:47:56,760
and analyze that data and find even more

1128
00:48:02,060 --> 00:47:59,760
scientific discoveries in these kind of

1129
00:48:03,350 --> 00:48:02,070
open data sources that we have I don't

1130
00:48:08,660 --> 00:48:03,360
know I think that gets it the question

1131
00:48:10,940 --> 00:48:08,670
that was being asked I think so um and

1132
00:48:14,210 --> 00:48:10,950
I'm affecting our list here and I think

1133
00:48:17,090 --> 00:48:14,220
that concludes our questions so Jason we

1134
00:48:18,230 --> 00:48:17,100
turn it over to you yeah so if we're not

1135
00:48:19,790 --> 00:48:18,240
many more questions well you can

1136
00:48:21,740 --> 00:48:19,800
obviously if people have questions after

1137
00:48:24,950 --> 00:48:21,750
the fact please keep on sending them in

1138
00:48:27,590 --> 00:48:24,960

using the hashtag and TL also you can I

1139
00:48:30,080 --> 00:48:27,600
want to encourage you all to go and

1140
00:48:32,180 --> 00:48:30,090
actually go and sign up for our

1141
00:48:33,620 --> 00:48:32,190
challenge and participate on it what

1142
00:48:36,350 --> 00:48:33,630
you're looking for is our it's a

1143
00:48:38,240 --> 00:48:36,360
marathon match entitled the collective

1144
00:48:41,660 --> 00:48:38,250
minds and machine exploration challenge

1145
00:48:46,070 --> 00:48:41,670
and hopefully get help us help NASA help

1146
00:48:47,330 --> 00:48:46,080
natgeo of harvard understand the past so

1147
00:48:49,460 --> 00:48:47,340
we can actually prepare for our future

1148
00:48:51,080 --> 00:48:49,470
data sets that are coming in with this

1149
00:48:52,670 --> 00:48:51,090
and I like to thank you all for

1150
00:48:53,900 --> 00:48:52,680
participating and I'd like to

1151
00:48:55,750 --> 00:48:53,910
our presenters and our partners here

1152
00:48:58,250 --> 00:48:55,760
that are on the phone today and

1153
00:49:00,020 --> 00:48:58,260
encourage you all to if not this

1154
00:49:06,819 --> 00:49:00,030
challenge we run many others and