I'm going to ask a question, but you can only answer by saying either 'yes,' 'no,' or 'it's complicated.' Alright?

我来问你个问题，你只能用“是”，“不是”或者“不清楚”来回答。可以吗？

So, let's start over here.

那么，我们开始吧。

Is some form of superintelligence possible, Jaan?

有没有可能出现某种形式的超级智能呢，Jaan?

'Yes,' 'no,' or 'it's complicated.'

“是”，“不是”或者“不清楚”。

Yes.

是。

Yes.

是。

Yes.

是。

Yes.

是。

Yes.

是。

Yes.

是。

Definitely.

当然。

No.

不是。

Well, this was disappointing, we didn't find any disagreement.

好吧，结果有些令人失望。我们没有发现任何异议。

Let's try harder.

我们再加把劲。

Just because it's possible doesn't mean that it's actually going to happen.

可能发生并不意味着真的会发生。

So, before I asked if superintelligence was possible at all according to the laws of physics.

所以，之前我问的是根据物理学规律推断，超级智能是否会发生。

Now, i'm asking, will it actually happen?

现在我要问的是，真的会发生吗？

A little bit complicated, but yes.

虽然我也有一点不清楚，但是是的。

Yes, and if it doesn't then something terrible has happened to prevent it.

是的，如果没有出现那说明已经出现了更厉害的东西阻止它了。

Yes.

是。

Probably.

大概是。

Yes.

是。

Yes.

是。

Yes.

是。

Yes.

是。

No.

不是。

Shucks, still haven't found any interesting disagreements.

哎，还是没有发现任何有趣的分歧。

We need to try harder still.

我们还需要再加把劲。

OK.

好的。

So, you think it is going to happen, but would you actually like it to happen at some point?

所以，你认为人工智能实际上会发生，但是某种程度上来讲你希望它发生吗？

Yes, no, or it's complicated?

“是”，“不是”还是“不清楚”？

Complicated, leaning towards yes.

不太清楚，但偏向于是。

It's complicated.

不清楚。

Yes.

是。

Yes.

是。

It's really complicated.

这问题真的太难了。

Yes.

是。

It's complicated.

不清楚。

Very complicated.

太复杂了。

Well, heck, I don't know.

哎，我不清楚。

It depends on which kind.

还要看究竟是哪种超级智能吧。

Alright, so it's getting a little bit more interesting.

那么，目前的情况开始变得有趣了。

When I think, we had a really fascinating...

我想，我们刚刚进行了一次精彩的…

When is it going to happen?

究竟什么时候会发生呢？

Well, we had a really fascinating discussion already in this morning's panel about when we might get to human level AI.

听我说，我们今天早上在小组里展开了一场精彩的讨论，主题是AI技术何时能达到人类的智力水准。

So, that would sort of put a lower bound.

那么，那差不多就是设置了个下限吧。

In the interest of time, I think we don't need to rehash the question of when going beyond it might start.

为了节省时间，我觉得我们不需要再强调越界后就会出现问题了。

But, let's ask a very related question to the one that just came up here.

但是，我们还需要问一个与之类似的问题。

Mainly, the question of well if something starts to happen, if you get some sort of recursive self-improvement or some other process whereby intelligence and machines start to take off very very rapidly, there is always a timescale associated with this.

主要就是说，如果事态开始发展，如果出现了某种递归的自我完善或其他过程，使人工智能和机器开始飞速发展，这期间总要涉及一个时间尺度。

And there I hope we can finally find some real serious disagreements to argue about here.

这样，我认为我们终于可以讨论一些非常严重的分歧了。

Some people have been envisioning this scenario where it goes PHOOM and things happen in days or hours or less.

有些人一直在设想，在这种情况下，事情会变得不受控制，几天甚至几个小时以内就能实现超级智能。

Whereas, others envision that it will happen but it might take thousands of years or decades.

而其他人则认为，这样的演进可能发生，但需要数千年或数万年的时间才能实现。

So, if you think of some sort of doubling time, some sort of rough timescale on which things get dramatically better, what time scale would you guess at, Jaan?

所以，如果让你来想象一下这种倍增时间，需要多久才能让技术突飞猛进呢？你来预测一个时间尺度吧，Jaan。

Start now or starting at human level?

从现在开始还是从人类的智力水平开始？

No, no, so once we get human level AI or whatever point beyond there or a little bit before there where things actually start taking off, what is the sort of time scale?

不，不，一旦我们的AI智力水平与人类的相当或者差不多时，也就是事态就真正开始发展的时候，这期间的时间尺度又是多少呢？

Any explosion, as a nerdy physicist, has some sort of time scale, right, on which it happens.

作为一个书呆子物理学家，我要说的是，任何爆炸的发生都有其时间尺度。

Are we talking about seconds, or years, or millennia?

那究竟是几秒钟，几年还是几千年？

I'm thinking of years, but it is important to act as if this timeline was shorter.

我觉得是几年，但是必须做好其提前到来的准备。

Yeah, I actually don't really trust my intuitions here.

对，在这里我其实并不太相信自己的直觉。

I have intuitions that we are thinking of years, but I also think human level AI is a mirage.

尽管直觉告诉我超级智能可能仅需几年就会实现，但我同时也认为具有人类智力水平的AI不过是一场海市蜃楼。

It is suddenly going to be better than human, but whether that is going to be a full intelligence explosion quickly, I don't know.

人工智能可能会突然超过人类，但这是不是一次迅猛的智能爆炸呢，我也不清楚。

I think it partly depends on the architecture that ends up delivering human level AI.

我认为这部分取决于具备人类智力水平的AI的结构。

So, this kind of neuroscience inspired AI that we seem to be building at the moment that needs to be trained and have experience in order for it to gain knowledge that may be, you know, on the order of a few years so possible even a decade. Some numbers of years, but it could also be much less.

所以，正是我们目前正在建造的这种神经系统启发了AI，这种神经系统也需要接受培训，需要通过实践获取知识，这一过程可能要花费几年时间，甚至十年，但当然也可以更少。

But, I wouldn't be surprised if it was much more.

但是，如果这一过程花费的时间较多，我也不意外。

Potentially days or shorter, especially if it's AI researchers designing AI researchers

如果AI设计师能够设计出AI设计师来，那么这一过程只需要几天或者几天不到吧。

Every time there is an advance in AI, we dismiss it as 'oh, well that's not really AI:' chess, go, self-driving cars.

每次AI出现进步的时候，我们都不屑一顾“哦，那可不是真正的AI，只不过是围棋，无人驾驶汽车这些东西”。

An AI, as you know, is the field of things we haven't done yet.

你所知道的AI是在这个领域中我们尚未触及的部分。

That will continue when we actually reach AGI.

当我们真正实现AGI时，这种状况还将持续下去。

There will be lots of controversy.

会有很多争议。

By the time the controversy settles down, we will realize that it's been around for

a few years.

每次解决争议都有可能花费数年时间。

Yeah, so I think we will go beyond human level capabilities in many different areas, but

not in all at the same time.

是的，所以我想我们在许多领域内都会拥有超越人类的能力，但并不是同时拥有。

So, it will be an uneven process where some areas will be far advanced very soon, already

to some extent and other might take much longer.

那么，这个过程就会是不平衡的。有的领域可能在短时间内就已经很先进了，而其他领域则要花费一定时间才能达到这种水平。

What Bart said.

Bart的评论

But, I think if it reaches a threshold where it's as smart as the smartest most inventive

human then, I mean, it really could be only a matter of days before it's smarter than

the sum of humanity.

但是，我认为如果达到一定的阈值，比如人工智能已经和全世界最聪明的人类一样聪明了，那么比全人类加起来都要聪明只怕是指日可待吧。

So, here we saw quite an interesting range of answers.

所以在这里我们发现了很多有意思的答案。

And this, I find is a very interesting question because for reasons that people here have

published a lot of interesting papers about the time scale makes a huge difference.

而且，我发现这个问题本身也很有意思，因为人们已经出于各种原因发表了很多有关时间尺度的论文，而在论文中提出的时间尺度差异巨大。

Right, if it's something that happening on the time scale of the industrial revolution,

for example, that's a lot longer than the time scale on which society can adapt and

take measures to steer development, borrowing your nice rocket metaphor, Jaan.

如果事情发生在类似工业革命的时间尺度上，那么社会就有足够的时间去适应，去采取措施来引导其发展，我在这里借用你的火箭比喻，Jaan。

Whereas, if things happen much quicker than society can respond, it's much harder to steer

and you kind of have to hope that you've built in a good steering in advance.

相反，如果事情发生极快，社会无法适应且不可逆转，那我们一定会后悔当时为什么没有建好逆转装置。

So, for example in nuclear reactors, we nerdy physicists like to stick graphite sticks in

a moderators to slow things down maybe prevent it from going critical.

比如说在核反应堆里，我们的书呆子物理学家经常用石墨棒粘上阻滞剂，如果事态危急可以减缓反应速度。

I'm curious if anyone of you feels that it would be nice if this growth of intelligence

which you are generally excited about, with some caveats, if any of you would like to

have it happen a bit slower so that it becomes easier for society to keep shaping it the

way we want it.

让我好奇的是，面对这种增长的人工智能，我们大多数人既兴奋又紧张，但有没有人想要减缓事态的发展，以便于让社会更好地按照我们想要的方式塑造它呢？

And, if so, and here's a tough question, is there anything we can do now or later on when it

gets closer that might make this intelligence explosion or rapid growth of intelligence

simply proceed slower so we can have more influence over it.

而且如果这样，又引发了另一个棘手的问题，我们在现在或者将来能做些什么才能减缓此次智能大爆炸或说智能的飞速发展，从而增加我们对它的掌控力呢？

Does anyone want to take a swing at this?

有人愿意挑战一下这个问题吗？

It's not for the whole panel, but anyone who...

这不是对整个小组提出的问题，而是对…

I'm reminded of the conversation we had with Rich Sutton in Puerto Rico.

这让我想起我们在波多黎各与Rich Sutton的对话。

Like, we had a lot of disagreements, but definitely could agree about paths slower being better

than faster.

尽管我们有很多分歧，但一致认同事情发展的慢比发展的快更为有利。

Any thoughts on how one could make it a little bit slower?

对于让它减速这件事情各位有什么想法吗？

I mean, the strategy I suggested in my talk was somewhat tongue and cheek.

我的意思是，我在讨论中提出的策略有些像闲磕牙。

But, it was also serious.

但是我在提出的同时也很严肃。

I think this conference is great and as technologists we should do everything we can to keep the

technology safe and beneficial.

我认为这次会议的意义重大，作为技术人员，我们应该竭尽所能，保证技术的安全性和有益性。

Certainly, as we do each specific application, like self-driving cars, there's a whole host

of ethical issues to address.

当然，正如的每个具体应用一样，如自驾车，我们还需要解决诸多伦理问题。

But, I don't think we can solve the problem just technologically.

但是，我也不认为我们只能解决技术问题。

Imagine that we've done our job perfectly and we've created the most safe, beneficial

AI possible, but we've let the political system become totalitarian and evil, either a evil

world government or even just a portion of the globe that is that way, it's not going

to work out well.

想象一下，我们完成了任务，创造了最安全，最有用的AI，但我们让政治制度变成了极权主义政治，衍生出了邪恶的政治制度，无论是邪恶的世界政府还是地方政府，都是行不通的。

And so, part of the struggle is in the area of politics and social policy to have the

world reflect the values want to achieve because we are talking about human AI.

Human AI is by definition at human levels and therefore is human.

所以，我们一部分的斗争就是在政治和社会政策领域，让世界反映出要实现的价值观，因为我们涉及的是人类的人工智能。人类的AI是根据人类水平定义的，因此也是人类。

And so, the issue of how we make humans ethical is the same issue as how we make AIs that

are human level ethical.

所以，为人工智能制定道德观念本质上和为人类制定道德观念是一样的问题。

So, what i'm hearing you say is that before we reach the point of getting close to human

level AI, a very good way to prepare for that is for us humans in our human societies to

try and get our act together as much as possible and have the world run with more reason than

it, perhaps, is today.

所以，我是这样理解你的话的：在我们获得接近人类水平的AI之前，一种非常好的准备方法是为了让我们人类社会中的人类尽可能的劲往一处使，让世界变得更有意义。

Is that fair?

那么这样合理吗？

That's exactly what I'm saying.

这正是我要表达的观点。

Nick? Also, I just want to clarify again that when I asked about what you would do to slow things

down i'm not talking at all about slowing down AI research now.

Nick？此外，我只是想再次澄清，我只是问采取什么行动能减缓发展，而不是要现在就减缓AI研究。

We're simply talking about if we get to the point where we are getting very near human

level AI and think we might get some very fast development, how could one slow that

part down?

我们只是说，如果我们的AI已经非常接近人类水平，并认为接下来可能会面临飞速的发展，那么怎样才能减缓这种局面呢？

So, one method would be to make faster progress now, so you get to that point sooner when

hardware is less developed, you get less hardware overhang.

However, the current speed of AI progress is a fairly hard variable to change very much

because there are very big forces pushing on it, so perhaps the higher elasticity option

is what I suggested in the talk to ensure that whoever gets there first has enough of

a lead that they are able to slow down for a few months, let us say, to go slow during

the transition.

所以现在有一个方法能加快进展，也就是说硬件开发不及时的时候就会更快地到达那个点，这就需要减少硬件突破。然而，目前AI的进步速度变数极大，因为有很大的力量推动它，所以我在这里建议设置更高的弹性选项，以确保到达这个点时还能够减缓几个月，要我说就是，在转型期间要慢一点。

So, I think one thing you can do, I mean this is almost in the verification area, is to

make systems that provably will not recruit additional hardware or resigned their hardware,

so that their resources remain fixed.

所以，我觉得有一件事是可以做的，我的意思是这几乎是验证领域的工作，也就是让系统不可能添加额外的硬件或者去除硬件，从而使其资源保持固定。

And i'm quite happy to sit there for several years thinking hard about what the next step

would be to take.

此外，我也很高兴能干坐上好几年思考下一步应该采取的行动。

But, it's trivial to copy software.

但是，复制软件是微不足道的。

Software is self replicating and always has been and I don't see how you can possibly

stop that.

软件是自我复制的，自古以来就是这样，不可能阻止这一点。

I mean, I think it would be great if it went slow, but it's hard to see how it does go

slow given the huge first mover advantages and getting to superintelligence.

我的意思是，进展缓慢固然是好，但是由于其具有巨大的先发优势，很难让它放慢脚步，无法实现人工智能。

The only scenario that I see where it might go slow is where there is only one potential

first mover that can then stop and think.

我认为它放慢脚步的唯一场景是潜在的先驱者停下来思考的时候。

So, maybe that speaks to creating a society where, you know, AI is restrictive and unified, but without

multiple movers.

所以，也许这就是说我们要创造一个社会，在社会中AI具有限制性和统一性，且没有行动者。

Yeah, Demis, so your colleague Sean Legg mentioned that the one thing that could help a lot here

is if there's things like this industry partnership and a sense of trust and openness between

the leaders, so that if there is a point where one wants to...

是这样的，Demis，所以你的同事Sean Legg提到，在这里作用巨大的一件事就是，如果有这样的行业伙伴关系，领导者之间的信任和开放感，那么如果当有人想要...

Yeah, I do worry about, you know, that sort of scenario where, you know, I think, I've

got quite high belief in human ingenuity to solve these problems given enough time. the

control problem and other issues.

是的，我也有过这样的担忧，你知道的，我对人类的聪明才智很有信心，我认为在给予他们充足时间的情况下，完全可以解决这些问题，也就是控制问题和其他一些问题。

They're very difficult, but I think we can solve them.

问题很困难，但我认为可以解决。

The problem is will there, you know, the coordination problem of making sure there is enough time

to slow down at the end and, you know, let Stuart think about this for 5 years.

问题就出在协调上，最后能不能确保有足够的时间进行减速，也就是说，给Stuart思考5年的机会。

But, what about, he may do that, but what about all the other teams that are reading

the papers and not going to do that while you're thinking.

但是，如果他这样做，但是正在阅读论文的所有其他团队却不会在你思考的同时也进行思考呢？

Yeah, this is what I worry about a lot.

是的，这也是我重点担忧的问题。

It seems like that coordination problem is quite difficult.

协调问题似乎是相当困难的。

But, I think as the first step, may be coordinating things like the Partnership on AI, you know,

the most capable teams together to agree, at least agree on a set of protocols or safety

procedures, or things, you know, agree that, maybe, you know, you should verify these systems

and that is going to take a few years and you should think about that.

但是，我认为作为第一步来说，可能会协调诸如“AI伙伴关系”这样的事情，最有能力的团队都同意，至少是同意先设置一套协议或安全程序，或者是验证一下这些系统，可能要花费几年时间，值得考虑。

I think that would be a good thing.

我认为这是一件好事。

I just want to caveat one thing about slowing versus fast progresses, you know, it could

be that, imagine there was a moratorium on AI research for 50 years, but hardware continued

to accelerate as it does now.

我只想注意与放缓快速进展有关的一件事情，可能是这样的，可以想象一下，暂停50年的AI研究，但硬件不断加速。

We could, you know, this is sort of what Nick's point was is that there could be a massive

hardware overhang or something where an AI actually many, many, many different approaches

to AI including seed AI, self-improving AI, all these things could be possible.

根据Nick的观点，那就是可能出现一个巨大的硬件突破，或者真的出现许多研究不同AI的方法，包括种子AI，自我改进AI，所有这些都是可能的。

And, you know, maybe one person in their garage could do it.

而且，也许有人在车库里就可以做到这一点。

And I think that would be a lot more difficult to coordinate that kind of situation, whereas,

so, I think there is some argument to be made where you want to make fast progress when

we are at the very hard point of the 's' curve.

并且我觉得协调这种情况非常困难，我认为在我们处于“s”曲线上的困境点时，必须要进行一些讨论才能快速取得进展。

Where actually, you know, you need quite a large team, you have to be quite visible,

you know who the other people are, and, you know, in a sense society can keep tabs on

who the major players are and what they're up to.

实际上，你需要一个相当大的团队，在团队中必须非常显眼，必须知道其他人是谁，而且你也知道，在某种意义上，社会可以看出主要参与者是谁，他们想要干什么。

Whereas, opposed to a scenario where in say 50 or a 100 years time when, you know, someone,

a kid in their garage could create a seed AI or something like that.

而在一种情况下，在50或100年内，有可能有的年轻人在车库里就能制造出一个AI或类似的东西。

Yeah, Bart, one last comment on this topic.

对，Bart，这是这个话题的最后一条评论。

Yeah, I think that this process will be a very irregular process and sometime we will

be far advanced and other times we will be going quite slow.

没错，我认为这个过程将是一个非常不规则的过程，有时我们将会发展迅猛，其他时候则会慢一些。

Yeah, i'm sort of hoping that when society sees something like fake video creation where

you create a video where you have somebody say made up things and that society will actually

realize that there are these new capabilities for the machines and we should start taking

the problem as a society more seriously before we have full and general AI.

是的，我希望当社会见识到类似虚拟视频制作的东西时，也就是能够捏造一个视频，而视频中有人在说自己实际上未说过的话，那么社会就能实际意识到这些机器的新功能，我们在拥有充分和普遍的AI之前，应该将这个问题作为社会问题更严肃地对待。

We'll use AI to detect that.

我们将使用AI来检测这种问题。

So, you mentioned the word 'worry' there, and you Nick went a bit farther, you had the word

'doom' written on your slides three times.

所以，你在这里提到了“担心”这个词，还有你，Nick，你想得更远一些，你的幻灯片上出现了三次“死亡”这个词。

No wonder there was one star on Amazon on that rating and that it was even in red color.

难怪亚马逊的评级只有一颗星，还是红色的。

I think it's just as important to talk about existential hope and the fantastic upside

as downside and I want to do a lot of that here.

我认为对希望和奇迹的讨论与对负面效应的讨论同等重要，而我希望做很多事情。

So, let's just get some of those worries out of the way now and then return to the positive

things.

所以，让我们从现在开始就停止担忧，然后做一些积极的事情吧。

I just want to go through quickly and give each one of you a chance to just pick one thing

that you feel is a challenge that we should overcome and then say something about what you feel

is the best thing we can do, right now, to try to mitigate it.

我只想快些完成，让每个人都有机会选择一种你认为我们应当克服的困难，然后再说明一下你认为我们最应当做些什么才能减轻这种困难。

Do you want to start Jaan?

你想先开始吗，Jaan？

To mitigate what?

你刚才说减轻什么？

Mention one thing that you're worried could go wrong and tell us about something constructive

that we can do now that will reduce that risk.

提出你担心的可能会出问题的一件事，并说说我们现在可以做些什么实质性的事来减少这种风险。

I do think that AI arms races, I see like a lot of, like, good.

我认为是AI军备竞赛吧，我认为这方面有问题，嗯。

I'm really heartened to see kind of great contacts between OpenAI and DeepMind, but

I think this is just like a sort of toy model of what the world at large might come up with

in terms of arms races.

我真的很高兴能看到OpenAI和DeepMind之间建立了良好联系，但我认为，这种人们看起来像玩具模型，因此不加管制的东西日后可能会演变成世界上的军备竞赛。

And for myself I have been spending increasing amount of time in Asia recently just to kind

of try to kind of pull in more people elsewhere, what has been so far, just been, kind of like, an Anglo-

American discussion mostly.

对于我来说，我最近大部分时间都是在亚洲度过的，为了尝试在其他地方和更多人交流，到目前为止，感觉大多数像是英美之间的讨论。

So, like this is, I think, this is one thing that should be done and i'm going to do it.

所以，就像这件事一样，我认为这是一件该做的事，我要去做。

Well, as someone who is outside this field, I think the challenge i'm really in touch

with is how hard it is to take the safety concerns emotionally seriously.

那么，作为一个在该领域之外的人，我认为我真正遇到的困难是在情感上认真对待安全问题是多么的困难。

And how hard it is for people in the field to do that as well.

即便是业内人员也很难做到这一点。

I can't tell you have many people outside this room who purport to be experts think

the control problem is a total non-issue.

我想说的是在这个房间外，有很多人自诩为专家，认为控制问题完全是小菜一碟。

I mean, it's just flabbergasting to meet these people and just therefore not worth thinking

about.

我的意思是，和这些人交流起来真是让人目瞪口呆，他们的观点不值得考虑。

And one of the reasons I think is that in one case there is this illusion that the time

horizon matters.

还有一个原因是，在某种情况下，人们会在时间范围上产生错觉。

If you feel that this is 50 or a 100 years away that is totally consoling, but there

is an implicit assumption there,

the assumption is that you know how long it will take to build this safely. And that 50

or a 100 years is enough time.

如果你认为离事情到来还有50年或100年，让人很有安全感，那么其中就有一个隐含的假设，也就是你知道安全地建造AI要花多久。50年或100年足够了。

The other issue is, I think, most people feel like intelligence is an intrinsic good and

of course we want more of it and it's very easy to be in touch with that assumption because

right now there is a cure for cancer, which we have not discovered.

还有一个问题是，我认为，大多数人觉得智力是一种内在的优势，我们当然想要更多这种优势，而且这种观点也很容易理解，因为现在世界上就存在着癌症的治愈方法，我们想要发现却尚未发现。

Right, how galling is that?

是啊，这是多么令人沮丧呢。

But for more intelligence, but for knowing which experiments to run, or how to integrate

the data we already have in hand, we would have a cure for cancer that was actionable

now unless there was some physical law of the universe that prevented us from curing

cancer, which seems unlikely.

但是如果我们拥有更多智慧，如果了解了要进行哪些实践，或者整合我们已经掌握的数据，我们就会得到治愈癌症的治疗方法，除非有一些宇宙的物理规律阻止我们治愈癌症，但这似乎不太可能。

So, the pain of not having enough intelligence is really excruciating when you focus on it,

but, and I think to your previous question of doing this quickly becomes an intrinsic

good provided we have solved the alignment problems and the political problems and the

chaos that would follow if we were just, if we did it quickly without solving those problems.

所以没有足够的智慧这件事情让人非常痛苦，尤其是当你专注于这件事情时，但是，只要在解决分配问题和政治问题的前提下，AI的迅速到来才能成为内在的优势，如果我们在没有解决这些问题的情况下迅速带来AI，就会导致混乱。

So, I think, it's the thing that is alarming is how ethereal these concerns are even to

those who have no rational argument against them.

所以，我认为，令人担忧的是，有的人没有进行理性思考，就觉得这些问题不足为奇。

So, Sam it sounds to me like you're agreeing very strongly with what Shane Legg that there

is, in some circles, still this strong taboo that, you know, don't even consider the possibility

that we might get AGI because it's just absolutely ridiculous.

所以，Sam，我认为你的话表明了你非常同意Shane Legg的看法，在某些方面，这件事仍然是个强大的禁忌，甚至不用去考虑我们获得AGI的可能性，因为这真是太荒谬了。

And he was arguing that the sooner we can get rid of this taboo the sooner people can

get to work and find all these really helpful solutions and answers that we need.

而且他争辩说，我们越早摆脱这个禁忌，人们就能越早开始工作，找到我们需要的真正有用的解决方案和答案。

So, suppose for a moment that I came up to you and said to you "this idea of super human

intelligence just sounds absolutely ridiculous, sounds completely nuts.

And by the way i've never seen your ted talk."

所以，假设我找到你并对你说：“这个超级人工智能的想法听起来很可笑，完全就是无稽之谈嘛。顺便说一下，我没看过你的TED演讲。”

And we're in an elevator and you have only 30 seconds to persuade me to take this more

seriously, what would you say?

假设我们在电梯里，你只有30秒才能说服我严肃看待这件事，你会说什么呢？

A lot of people who are here will have this exact conversation with colleagues and others

in the future.

今天在座的的很多人将来会与同事或其他人进行类似的对话。

Well, there are very few assumptions you need to make to take this seriously, intellectually.

Again, the emotional part is a separate piece.

其实，要让人们严肃的对待人工智能，我们并不需要做什么假设。而情绪部分又是单独的一块。

But, if you assume that intelligence is just, on some level, the product of information

processing in a physical system and there are very few people who dispute that who are

scientifically literate at this point and you assume that we will continue to improve

our information processing systems, unless something terrible happens to us to prevent

that, and that seems like a very safe assumption, then it is just a matter of time before we

instantiate something that is human level and beyond in our computers.

但是，如果你假设智慧在某种程度上只是物理系统中的信息处理产品，且很少有人在这里质疑谁是知识分子，而且我们会继续改进我们的信息处理系统，除非发生什么可怕的事情来阻止我们进行改进，这似乎是一个非常安全的假设，那么在我们将人工智能具体化，赋予其人类水平的智能，使其远远超过计算机之前，这只是一个时间问题。

And, again, the time horizon is only consoling on the assumption that we know we have enough

time to solve the alignment problems and the political problems.

还有，时间范围只是一种心理安慰，也就是假设我们有足够的时间来解决分配问题和政治问题。

The other thing that is humbling here that Ray brought up at one point is that even if

we were handed a perfectly benign, well behaved AI just from god, you know, we are given a

perfect oracle we are given a perfect inventor of good technology, given our current political

and economic atmosphere that would produce total chaos.

另一件事，也就是Ray在这里提出的，即使我们从上帝那里获得了一个好的AI，完美的AI，就好像，我们得到了一个完美的神话，得到了一位能创造完美技术的发明大师，但是鉴于现在的政治和经济状况，毫无疑问会产生混乱。

We just have not... we don't have the ethical or political will to share the wealth, we

don't have the political integration to deal with this thing being given to Silicon Valley

and not being given at the same moment to China or Iran.

我们没有......我们没有分享财富的道德或政治意愿，我们没有政治一体化能力，能让它先被送到硅谷，而不是同时给予中国或伊朗。

So, there is just, it's alarming that the best case scenario currently, basically just

ripping out 80one. And so, clearly, that's a near term thing that we have to solve.

Thank you, Sam.

所以，只是，令人震惊的是，目前最好的情况也基本从80one中被去除了。很明显，这是我们近期必须解决的事情。

谢谢你的发言，Sam。

So, Demis do you want to tell us about one thing that you feel is a challenge and say

something about what we should focus on now to tackle it.

Yeah, I mean I think it's, you know I agree with both the statements already said that,

so I think the coordination problem is one thing where you know we want to avoid this

sort of harmful race to the finish where corner cutting starts happening and things like safety

are easy things to, you know, will get cut because obviously they don't necessarily contribute

to AI capability, in fact they may hold it back a bit by making a safe AI.

那么，Demis，你来和我们说一下你认为困难的事情，并为我们提供一下解决方法吧。

好的，我认为是这样的，我同意我们在这里所作的两种陈述，所以我认为协调问题是避免这种有害的军备竞赛的有效方法，如果人们要走捷径，那么安全问题就比较容易解决了，军备会被削减，因为显然军备不一定能改善AI的能力，相反，他们可能会通过扼制军备来保证AI的安全。

So, I think that's going to be a big issue on a global scale and that seems like it's

going to be a hard problem when we are talking about national governments and things.

所以，我认为这是一个全球性的大问题，而且这件事情涉及到国家政府等层面，看起来十分艰巨。

And I think also, you know, we haven't thought enough about the whole governance scenario

of how do we want those AIs to be out in the world?

而且我也想到，我们对于如何确定AI在世界上的分布及AI的整个治理环境这些问题还是想的不够透彻。

How many of them?

AI要有多少？

Who will set their goals?

谁来为他们设定目标？

All these kinds of things, I think, need a lot more thought.

You know, once we've already solved the technical problems.

我认为所有这些还需要深思熟虑。你看，一旦我们解决了技术问题…

I think it's wonderful that you're not just saying these things, but actually doing these

things since you played a leading role in setting up the Partnership on AI here which

goes exactly in the direction of what you're advocating here.

我认为你不仅仅是在说这些事情，而且是在实际的做这些事情，因为你在建立AI伙伴关系方面发挥了主导作用，这恰恰就是你所倡导的。

So, do you want to pass it off to Nick?

那么，你想让Nick接着说吗？

I'm sure there is nothing at all you're worried about, right?

我认为你没有什么好担心的，对吧？

So, tell us about one concrete useful thing you would to see us focus on.

那么，和我们说一些具体有用的东西吧，我们会去关注的。

So, I agree with that, I mean, so fun technical work, bring in top technical talent to work

on these technical issues, build these collaborations, build a community, build trust, work some

more on figuring out attractive solutions to the governance solutions that could work,

but don't rush to implement the first idea you have, but first trial them out a little

bit more.

我的意思是，技术工作是很有趣的，我们要引进顶尖的技术人才来处理技术问题，建立合作关系，建立社区，建立信任，为打造治理方案提供更多的好方法，但是脑子里有了想法之后，不能第一时间就去实施，而是要经过反复实验。

I think a lot about consciousness, so I was really struck by the 'sentience caution' on

the list of principles here that said "avoid overly... avoid strong assumptions about the

distribution of consciousness in AIs," which I take it entails avoid assuming that any

human level or super human level AGI is going to be conscious.

我在意识方面想的很多，所以我对这个原则清单上的“谨慎态度”感到非常震惊，上面说：“避免过分……避免对AI的意识分配产生强烈假设”，我认为这就是说我们不能假设任何人类智力水平或超过人类智力水平的AGI会有意识。

For me, that raising the possibility of a massive failure mode in the future, the possibility

that we create human or super human level AGI and we've got a whole world populated

by super human level AGIs, none of whom is conscious.

对于我来说，提高未来发生大规模失败模式的可能性，我们创造拥有人类智力水平或超过人类智力水平的AGI的可能性，全世界人口都是超过人类智力水平的AGI，而他们都没有意识。

And that would be a world, could potentially be a world of great intelligence, no consciousness

no subjective experience at all.

而这样一个世界，这样一个拥有伟大智慧的世界，却没有意识，完全没有主观的经验。

Now, I think many many people, with a wide variety of views, take the view that basically

subjective experience or consciousness is required in order to have any meaning or value

in your life at all.

现在，我觉得很多很多人都有各种各样的意见，认为必须要有主观的经验或意识，才能在生命中产生意义或价值。

So therefore, a world without consciousness could not possibly a positive outcome.

maybe it wouldn't be a terribly negative outcome, it would just be a 0 outcome, and among the

worst possible outcomes.

因此，没有意识的世界不可能是好结果。也许不会是非常负面的结果，它只是一个0结果，在最糟糕的可能结果之中。

So, I worry about avoiding that outcome.

所以，我想的是如何避免这种结果。

Now, as a matter of fact, i'm fairly optimistic about the possibilities that AIs of various

kinds will be conscious.

现在，事实上，我对各种AI存在意识的可能性相当乐观。

But, in so far as this community is making this assumption, I think it's important to

actually think about the question of 'in creating AGIs, are we actually creating conscious beings?'

但是，在这个社会做出这样的假设的同时，我认为很有必要思考这样一个问题：“创造AGI实际上就是创造有意识的人吗？”

I mean, one thing we ought to at least consider doing there is making, given that we don't understand

consciousness, we don't have a complete theory of consciousness, maybe we can be most confident

about consciousness when it's similar to the case that we know about the best, namely human

human consciousness...

我的意思是说，我们现在至少应该做这么一件事，因为我们不了解意识，没有一个完整的意识理论，也许我们对自己的意识有信心，所以也将其运用到类似的情况中，即人类的人类意识...

So, therefore maybe there is an imperative to create human-like AGI in order that we

can be maximally confident that there is going to be consciousness.

So, what I hear you say is that when you have a nightmare about the zombie apocalypse you're

not thinking of some terminator movie, but you're thinking about this problem.

因此，也许有必要创建与人类似的AGI，以便我们可以最大程度地相信意识的存在。

所以，你的意思是，当你做了关于僵尸启示的噩梦时，脑子里想的不是终结者系列电影，而是现在正在讨论的这个问题。

We create... we upload ourselves and do all these wonderful things, but there's no one

home.

我们创造自己...将自己上传，做很多美好的事情，但家里却没有人。

Is that fair to say?

这么说合理吗？

I mean this is a different kind of existential risk.

我是说这种风险完全不同。

One kind of existential risk is there's no humans, there's AIs, but some people might

say well that's OK they are our successors.

有一种已经提出的风险是地球上没有人类，只有AI，但有些人可能会觉得没有问题，他们是我们的继任者。

A much worse existential risk is there are no conscious beings in our future.

更糟糕的风险就是在未来，已经不存在有意识的人类了。

So, i'll make a confession, so Shane Legg mentioned that there has been this strong taboo about

talking about the possibility of intelligence getting very advanced.

那么，我现在坦诚的说一下，所以Shane Legg提到，谈论智能进化的先进程度的可能性是一种强烈的禁忌。

It's clearly also been a strong taboo for a long time to mention the C-word.

很显然，这个C开头的词很明显也是一种强烈的禁忌。

In fact, before the conference when we got all these responses on the first round of

the principles, guess which one was ranked last?

事实上，在会议之前，我们得到了第一组原则的回复。猜猜哪个排在最后？

It got huge amounts of minus 1 ratings, that was the one with consciousness, so we changed

it to-- it was terribly stated --sentience and stated it better and then it got stated

still better at lunch and it's still rated last.

它得到了很多负分1评级，就是意识那个，所以我们修改了 - 一开始说的太明确了，我们把它改的好了一些，然后午餐时间又做了改动，然而其评级仍在排在最后。

Even though I personally share your interests in this a lot.

即使我以个人名义分享也得到了很多负分评级。

88But, not 90So, maybe that is another taboo you can personally help us shatter so that people think about

that question more.

88But，而不是90So，也许这是另一个你可以亲自帮助我们打破的禁忌，以便人们更多地思考这个问题。

Ray, anything you're concerned about?

Ray，你关心的是什么问题？

This isn't what I was going to say, but just to respond... a converse concern is we create

AGIs, everybody assumes that of course it's just a machine and therefore it's not conscious,

but actually it is suffering but we don't look out for it's conscious subjective experience

because we are making the wrong assumption.

这不是我自己要说的，只是为了回应...一个相反的关注点是我们创造了AGI，每个人都认为当然这只是机器，因此它不具备意识，但实际上AGI也很痛苦，但由于我们我们的假设就做错了，所以我们也不会顾及他们的主观看法。

But, what I did want to say was, there are three overlapping revolutions that people

talk about, GNR, genetics, bio-tech, nano-technology, and robotics, which is AI. And there are proposals, there was the Asilomar conferences done here decades ago for bio-tech

that have worked fairly well.

但是，我想说的是，人们讨论的是GNR，遗传学，生物技术，纳米技术和机器人，这些结合起来就是AI。还有我要说的是，几十年前的Asilomar会议在生物技术方面做得很好。

There are similar proposals for nano-technology.

我在纳米技术方面也有类似的话要说。

There is a difference with AI in that there really isn't a full proof technical solution

to this.

它与AI的区别就是，它的确没有完整的证明技术解决方案。

You can have technical controls on, say, nano-technology.

One of the guidelines is it shouldn't be self-replicating.

你可以对纳米技术进行技术控制。它的一个指导原则之一就是禁止自我复制。

That's not really realistic because it can't scale to meaningful quantities without being

self-replicating, but you can imagine technical protections.

这并不实际，因为如果不进行自我复制，它不可能扩展到有意义的数量级，但你可以设想一下如何进行技术保护。

If you have an AI that is more intelligent than you and it's out for your destruction

and it's out for the world's destruction and there is no other AI that is superior to it,

that's a bad situation.

如果你的AI比你更聪明，而它又为了毁灭你做了出格的事，为了毁灭世界做了出格的事，没有任何AI能战胜它，这情况就不妙了。

So, that's the specter.

这就是我们口中的怪物。

And partly this is amplified by our observation of what we as humans, the most intelligent

species on the planet, have done to other species.

其中部分原因是我们观察到了自己作为地球上最聪明的物种，对其他物种做的一切进行了放大推断。

If we look at how we treat animals, people, you know, are very friendly, like their dogs

and pets, but if you look at factory farming we're not very benign to species that are

less intelligent than us.

如果你看看我们对待动物的方式，我们能够发现，其实非常友好，就像对待狗和其他宠物一样，但是如果再看看工厂化农牧，我们就会发现我们对智力不如我们的物种并不是非常温和。

That engenders a lot of the concern we see that if we there's a new type of entity that's

more intelligent than us it's going to treat us like we've treated other species.

这引发了我们的广泛关注，如果有一种比我们更聪明的新型实体，那么他们对待我们的方式就像我们对待其他物种一样。

So, that's the concern.

这就是问题所在。

I do think that what we are doing at this conference is appropriate.

我认为我们在这次会议上做的事是应该做的。

I wanted to mention that I think we should publish these guidelines the way the Asilomar

guidelines in bio-tech were published decades ago.

我想说的是，我认为我们应该像几十年前公布Asilomar生物技术指南一样公布这些指南。

And then people can and people can, you can have an opt-in, opt-out, but I think we should

actually say we had this conference and the AI leadership/community has come up with these

guidelines and people can respond to them and debate them and then maybe at the next

conference we'll revise them.

然后，人们可以选择加入或者选择退出，但我认为我们应该说明我们进行了这次会议，AI领导/社区已经提出了这些准则，人们可以对此作出回应，进行辩论，然后在下一次会议上我们就修改这些准则。

The Asilomar bio-tech guidelines have been revised many times.

Asilomar生物技术指南已经修改了很多次。

But, I would advocate that we actually take a stand and put forth these guidelines and

then let the whole community at large debate them.

但是我主张的是，我们应该站出来，提出这些准则，然后让整个社会进行讨论。

And have them be, have them guide our research.

让他们指导我们的研究。

It's actually worked quite well in bio-tech.

事实上，现在我们在生物技术方面做得很好。

Bart?

Bart?

OK, yeah so let me give a little different perspective.

嗯，好，我来说一下我的不同观点。

So, one concern I have at the high level is these machines become really smart or even

in certain areas, can humans still understand, what they, decisions that they suggested,

that they make.

那么，我在高层次上担心的一个问题就是这些机器可能变得非常聪明，甚至在某些领域中，我们怀疑人类还能否理解他们提出的意见，做出的决定。

And I work in the field of automated reasoning where we have significant advance last two

decades going from perhaps a few hundred variables to perhaps millions of variables being solved

quite routinely.

我的工作领域是自动化推理，二十年来，我们可能用常规方式解决了数百个变量乃至数百万个变量。

And there was a sense in the community, well we are getting answers from these reasoning

engines, mostly hardware/software verification problems, but we cannot, humans can no longer

understand these answers.

在我们的社区中有一种感觉，我们正在用这些推理引擎获得答案，主要是硬件/软件验证问题，但是我们不能理解这些答案。

In the last few years, people have actually discovered that you can use the machine to

generate explanations for their answers that are, again, human understandable.

在过去的几年中，人们发现自己可以使用机器来产生这些答案的解释，而这又是人类可以理解的。

So, I see sort of a glimmer of hope that maybe even if we have much less intelligence we

may be able to understand solutions that machines find for us and we could not find these solutions, but they may be able to provide explanations that are accessible to us.

So that's a little positive note.

所以，我看到了一丝希望，即使我们的智慧不如机器，我们也或许能够理解机器为我们提供的解决方案，即使我们不能理解这些解决方案，但机器也可为我们提供可以理解的解释。

这是积极的一点。

Thank you.

感谢你的发言。

Stuart?

Stuart?

So there are two things that keep me awake at night, other than email.

所以除了惦记着电子邮件之外，还有两件事让我夜不能寐。

So, one is the problem of misuse and bad actors.

其中之一就是滥用和破坏分子的问题。

To take an analogy, it’s as if we were building nuclear weapons and then delivering them by

email to everybody on the planet, saying, here’s a toy, do what you want.

打个比方说，这就像我们正在建造核武器，然后用电子邮件发送给地球上的每个人，说，这是一个玩具，你来玩玩吧。

How do we counter that? I have to say, I don’t really have a good solution.

我们如何处理这种事？我不得不说，我真的没有很好的解决方案。

I think one of the things we have to do is to make designs for safe AI very clear and

simple, and sort of make it unthinkable to do anything other than that, right?

我认为我们需要做的事就是让AI的安全设计非常清晰简单，然后让它不能用于其他目的，是这样吗？

Just like it’s unthinkable to have a program with an infinite loop that produces a spinning

pizza of death on your -- oh sorry.

用做比萨饼的无限循环程序就能导致你的死亡，这同样也是不可思议的 - 哦，对不起。

Or it’s unthinkable to have a buffer overflow that allows your software to be hacked into.

或者说出现缓存溢出就能让您的软件被入侵，这也是不可想象的。

The other thing that keeps me awake is actually the possibility that success would lead to

AI as a helicopter parent for the human race that would sort of ossify and gradually enfeeble

us, so then there would be no point at which it was obvious to us that this was happening.

让我我夜不能寐的另一件事就是成功可能导致AI成为人类的直升机式父母，可以说是很不灵活，逐渐使我们失望，既然结果毫无疑问，那么其中也没有什么意义。

And I think the mitigation, which you asked for, to look on the bright side, is that in

some sense the meta-value of human evolvability, the freedom to change the future, is something

that the AI needs to adopt, and in some sense that would result eventually with the AI receding

into the background, and saying, OK, now I’ve got you through your adolescence, now it’s

time for the human race to grow up, now that we have the capabilities to eliminate scarcity,

to eliminate needless conflict and coordination failures and all of those things that we suffer

from right now.

而我认为，你所要求的减轻，从光明的角度来看，在某种意义上说，是人类可演变性的元价值，是我们改变未来的自由，这是人类需要采用的，而且从某种角度来说这会导致AI最终退居后台，并说，现在我已经带领你经过了青春期，现在是人类长大的时候了，现在我们有消除稀缺，消除不必要的冲突，协调失败以及解决我们目前遇到的所有困难的能力。

So I can imagine a distant future where, in fact, AI is perhaps even less visible than it is

today.

所以在我的想象中，也许在不远的未来，对AI的关注度并不比现在的高。

Great, finally you, Elon, have as far as I know never ever expressed any concerns about

AI, right - I’m just wondering if there is any concerns, in particular any concerns

where you see there’s a very clear thing we should be doing now that are going to help.

很好，最后该你了，Elon，据我所知，你从来没有提出与AI有关的任何疑虑，我只是想知道你对此是否有任何担忧，以及你认为我们应该做什么才能改变现状。

I’m trying to think of what is an actual good future, what does that actually look

like, or least bad, or however you want to characterize it.

我我正在努力去想真正的美好未来究竟是什么，实际上是什么样的，或者最坏的未来是什么样的，该如何进行描述。

Because to a point that was made earlier by Sam and maybe made by others, we’re headed towards either

superintelligence or civilization ending.

根据Sam或者其他人之前提出的观点，我们要么会获得超级智能，要么就要面临文明的毁灭。

Those are the two things that will happen - intelligence will keep advancing, the only

thing that would stop it from advancing is something that puts civilization into stasis

or destroys civilization.

这两件事情的确会发生 - 智能会继续进步，唯一能阻碍它进步的方法就是使文明陷入瘫痪或破坏文明。

So, we have to figure out, what is a world that we would like to be in where there is this digital superintelligence?

所以，我们必须弄清楚，如果世界上出现了数字超级智能，那么我们希望这个世界是什么样子的？

I think, another point that is really important to appreciate is that we are, all of us, already

are cyborgs.

So you have a machine extension of yourself in the form of your phone and your computer

and all your applications.

You are already superhuman.

我认为，我们必须认清的另一点就是我们已经是半机械人了。你的手机，电脑以及上面的应用程序都是你的机械形式的扩展。你已经是个超级人类了。

By far you have more power, more capability, than the President of the United States had

30 years ago.

目前为止，你的力量，能力要比30年前的美国总统的更多。

If you have an Internet link you have an article of wisdom, you can communicate to millions

of people, you can communicate to the rest of Earth instantly.

如果你有互联网链接，又有智慧，那就可以与数百万人沟通，可以与地球的其他地方进行即时通讯。

I mean, these are magical powers that didn’t exist, not that long ago.

我的意思是说，这些神奇的力量在不久之前还不存在。

So everyone is already superhuman, and a cyborg.

The limitation is one of bandwidth. So we’re bandwidth-constrained, particularly on output.

所以每个人都已经是超级人类，是半机械人了。而其中的限制之一就是带宽。所以我们是受带宽限制的，特别是输出方面。

Our input is much better but our output is extremely slow.

我们的输入能力比较好，但是输出却非常慢。

If you want to be generous you could say maybe it’s a few hundred bits per second, or a

kilobit or something like that output.

说的好听一点，你可以输出速度大概是每秒几百位，千分之一或类似这样的输出数值。

The way we output is like we have our little meat sticks that we move very slowly and push

buttons, or tap a little screen. And that’s extremely slow.

我们输出的方式就是用小手指头缓慢的按下按钮，或点击小屏幕。这可以说是非常慢了。

Compare that to a computer which can communicate at the terabyte level.

These are very big orders of magnitude differences.

将其与可以在TB级通信的计算机进行比较，就会发现其中存在着巨大的数量级差异。

Our input is much better because of vision, but even that could be enhanced significantly.

由于具备视力，我们的输入能力要强得多，但即使是这方面也可以显著加强。

So I think the two things that are needed for a future that we would look at and conclude

is good, most likely, is, we have to solve that bandwidth constraint with a direct neural

interface.

所以我认为在未来，我们最需要正视和总结的事可能是用直接的神经接口来解决带宽约束。

I think a high bandwidth interface to the cortex, so that we can have a digital tertiary

layer that’s more fully symbiotic with the rest of us.

我认为可以用高带宽接口接入皮质，使我们可以拥有数字化的第三层，可以与其他人完全共生。

We’ve got the cortex and the limbic system, which seem to work together pretty well - they’ve

got good bandwidth, whereas the bandwidth to additional tertiary layer is weak.

我们有皮质和外缘系统，它们配合起来似乎很好用 - 它们的带宽良好，而其它三级层的带宽较弱。

So I think if we can solve that bandwidth issue and then AI can be widely available.

所以我认为如果我们能解决带宽问题，那么就可以广泛使用AI。

The analogy to a nuclear bomb is not exactly correct - it’s not as though it’s going

to explode and create a mushroom cloud, it’s more like if there were just a few people

that had it they would be able to be essentially dictators of Earth, or whoever acquired it

and if it was limited to a small number of people and it was ultra-smart, they would

have dominion over Earth.

用核弹进行类比其实并不完全正确 - 它不像爆炸并出现蘑菇云那样，而是只有几个人拥有，这几个人能够成为地球上的独裁者，或者无论有哪一小部分人获得了它，都会变的非常智慧，他们将在地球上占据统治地位。

So I think it’s extremely important that it be widespread and that we solve the bandwidth

issue.

所以我认为解决带宽问题具有普遍意义且十分重要。

And if we do those things, then it will be tied to our consciousness, tied to our will,

tied to the sum of individual human will, and everyone would have it so it would be

sort of still a relatively even playing field, in fact, it would be probably more egalitarian

than today.

如果我们这样做，那么它就会被束缚在我们的意识上，与我们的意志相联系，与个人的意愿相联系，人人都有，所以竞争环境还将是比较公平的，实际上，可能比今天还要公平。

Great, thank you so much, that’s in fact the perfect segue into the last question I

want to ask you before we open it up to everybody. Something I have really missed in the discussion about really advanced intelligence, beyond

human, is more thought about the upside.

太好了，非常感谢，这就是我们向大家公开之前我想问的最后一个问题的完美答案。

在关于真正高级智能的讨论中，我错过了一些东西，只关注了它的正面。

We have so much talk about existential risk, and not just in the academic context, but

just flip on your TV, check out Netflix, what do you see there in these scientific visions

of the future?

我们非常关注现存的风险，而不仅仅是在学术背景下的风险，现在打开电视，看看Netflix电视台，你在这些未来的科学视野中看到了什么？

It’s almost always dystopias, right?

几乎都是刺激性的东西，对吧？

For some reason fear gives more clicks than the positive visions, but if I have a student

coming into my office at MIT asking for career advice, the first thing I’m going to ask

her is, where will you want to be in 20 years?

出于某些原因，人们总是更关注令人恐惧的一面而非正面愿景，但是如果我的学生进入我在麻省理工学院的办公室做职业咨询，我要问她的第一件事就是：你在接下来的20年内想做些什么？

And if she just says, well maybe I’ll get cancer, maybe I’ll get run over by a bus,

that’s a terrible way to think about career planning, right?

如果她说，也许我会患上癌症，也许我会被公共汽车碾过，这种做职业规划的方法很可怕吧？

I want her to be on fire and say my vision is I want to do this - and here are the things

that could go wrong, and then you can plan out how to avoid those problems and get it

out - I would love to see more discussion about the upsides, futures we’re really

excited about, so we can not just try to avoid problems for the sake of avoiding problems,

but to get to something that we’re all really on fire about.

我需要点拨学生，讲述一下我做职业规划的视角 - 哪些环节可能出现问题，做什么计划才能够避免这些问题，并解决问题 - 我希望看到更多令人激动的正面讨论，所以我们不能只是为了避免问题而避免问题，而是为了获得一些我们真正在乎的东西。

So to start off I’ll just tell you something that makes me really excited about advanced

artificial intelligence.

所以在开始时，我会先讲一些我看好先进人造智能的地方。

Everything I love about civilization is a product of intelligence.

在所有文明中，我所钦慕的一切都是智慧的产物。

If we for some reason were to say, well, you know, I’m scared about this technology thing, let’s

just press pause on it forever, there’s no interesting question about if we’re going

to have human extinction, the question is just ‘when?'

如果我们出于某种原因说，哎，我害怕和技术有关的事情，不如我们一直按住暂停键，人类要灭绝这都是早晚的事情没什么好讨论的，问题是'什么时候'要灭绝？

Is it going to be a supervolcano, is it going to be the next dinosaur-killing-class asteroid

- the last one happened 60 million years ago, so how long is it going to be?

如果地球成为一个超级火星，成为下一个能让恐龙灭绝的行星 - 上一次这种事情发生在6000万年前，那么离人类灭绝会有多久呢？

Pretty horrible future to just sit and wonder when we’re going to get taken out here without

the technology when we know that we totally have the brainpower to solve all of these

problems if we proceed forward and develop technology.

假设我们濒临灭绝时还没有能够拯救自身的技术，但如果我们继续前进并研发技术，其实完全有能力解决所有这些问题。想想就很可怕不是吗？

So that was just one thing that makes me very excited about moving forward rather than pressing 'Pause.'

所以我更高兴看到我们继续钻研这件事情，而不是按下“暂停”。

I want to just ask the same question to all of you guys in turn.

反过来，我想问大家同样的问题。

So tell us, just pretty briefly, about something that you are really excited about.

Some future vision you imagine with very advanced artificial intelligence that you’re really excited about, that you would like to see.

所以来简单说说什么事情能让你真正高兴。对先进的人造智能的未来愿景，究竟发展成什么样子能让你真正的感到高兴。

Jaan-

So I want to be careful when I imagine concrete fruits of AGI.

当我想象AGI的具体成果时，我需要小心翼翼。

On a meta-level I think as a first approximation, I think we should just maximize the amount

of fun and minimize the amount of suffering.

我认为第一个近似值在元级别，我觉得我们应该尽可能的放大欢乐，减少痛苦。

I think Eliezer [Yudkowsky] has written a sequence called “Fun Theory”, where he points out that

people have been horrible imagining, are very unimaginative imagining paradises of various

sorts, just like really boring places, actually, when you think about them.

Eliezer [Yudkowsky]写了一部名为《乐趣理论》的系列丛书，他在书中指出，人们曾有过很可怕的设想，是各种想象力所不能及的地方，事实上让我们想想，就像十分无聊的地方。

I think Eliezer has this sketch where he says, “It was hard to spend like one weekend with my relatives.Imagine spending eternity with your dead relatives.”

我认为Eliezer描述过这种情境，他说：“我简直没法和亲戚们共度一个周末。可以想象一下，与你死去的亲戚一起度过永恒是什么感受。“

So I think we should be concerned about side effects and try to capture dynamics of improvement,

and basically go from there - make sure that we’re going to adjust the trajectory as

we get smarter and more grown together.

所以我认为我们应该关注副作用，并试图捕获改善的动力，基本上就是从这里开始 - 确保我们在变得越来越聪明，越来越多强大的同时也要调整轨迹。

Great, thank you, Jaan.

太好了，谢谢你，Jaan。

Sam, what do you get excited about?

Sam，让你感到高兴的是什么？

Well, strangely, what excites me really just abuts the parts that scare me the most.

I think what is nice about this conversation, in particular about the alignment problem,

is that it’s forcing us to realize that there are better and worse answers to questions

of human value.

好吧，奇怪的是，让我感到高兴的事情和让我感到恐惧的事情可以说是紧密相连的。我认为这就是这次讨论的好处所在，特别是分配问题，能让我们意识到人类价值问题有更好的答案，也有更差的答案。

And as someone said, perhaps at this last meeting in Puerto Rico, we really have to

do philosophy on a deadline, and we have to admit to ourselves that there are better and

worse answers and we have to converge on the better ones.

正如有人所说，也许在波多黎各的上一次会议上，我们确实应该在最后期限内研究哲学问题，我们必须承认这其实是一把双刃剑，我们必须趋于更好的一边。

And what would excite me about actually the birth of superintelligent AI - one of the things,

apart from solving obvious problems like curing disease and energy issues and all the rest,

perhaps differs a little bit with what Stuart said.

让我兴奋的是超智能AI的诞生 - 及其带来的好处，除了解决明显的问题，如治愈疾病，能源问题和其他一些问题，可能和Stuart说的有些出入。

I’m not so worried about idiocracy or all of us just losing our way as apes and living

unproductive lives in dialogue with these oracles.

我不太在意这种愚蠢的进化论，或者我们所有人都失去了像猿人一样的生活方式，与先知一起无趣的生活。

I think actually, I would want a truly value-aligned superintelligence to incrementally show us,

not merely conserve what we want, but show us what we should want to keep improving our

values so that we can navigate in the space of all possible experiences and converge on

better and better ones.

实际上，我希望有一个真正价值一致的超级智能能够逐渐向我们展示，但不仅是保留我们想要的东西，而是向我们展示我们应该如何继续改进价值，为我们在所有实践中导航，让我们精益求精。

Thank you, Sam, and what about you, Demis?

谢谢Sam的发言，你说呢，Demis？

So obviously this is why I spend my whole career working on this, is that, I think if

we do this right, it’s going to be the greatest thing ever to happen to humanity,

and in some ways I think unlock our full potential.

很明显，这就是我花费整个职业生涯来做这件事的原因，那就是我认为如果我们这样做是正确的，那这就是人类历史上最伟大的事情，在某种程度上，我想释放我们的全部潜力。

I mean, I’ve talked to a lot about, in all my talks about using it as a tool to help

us make science and medical breakthroughs faster.

我的意思是说，说到用它来帮助我们更快地进行科学和医疗方面的突破，在这方面我已经说了很多。

And so I think that’s an obvious one.

所以我觉得这是一个很明显的原因。

But taking that longer-term, one reason I got so into AI is that, like probably many

of you in this room, I’m interested in the biggest questions of why we’re here, understanding

our minds, what is consciousness, what’s the nature of the universe, what’s our purpose

- and if we’re going to try and really grapple with any of those questions I think we’re

going to need something like AI, perhaps with ourselves enhanced as well.

但是，长期以来，我痴迷于AI的原因就像这个房间里的许多人一样，我对一些终极问题十分好奇，比如我们为什么在这里，讨论我们的想法，什么是意识，什么是宇宙的性质，我们的目的是什么，如果我们要试着真正解决这些问题，我认为我们需要像AI这样的东西，也许还需要加强自身。

And I think in that future world we’ll have a chance to actually find out about some of

these really deep questions in the same way we’re finding out with AlphaGo just about Go,

but what if we could do that with all of science and physics and the biggest questions

in the universe.

而我认为在未来的世界中，我们有机会真正了解这些真正深层次的问题，就像我们利用AlphaGo进行探索一样，如果我们能用人工智能解决所有的科学物理学问题和宇宙中的终极问题呢？

And I think that’s going to be the most exhilarating journey of all, to find that out.

我认为这就是整个过程中最令我兴奋的事情。

To just carry out on a few other things that people commented on is in terms of

us as the most intelligent beings on the planet right now, and treating animals badly and these

sorts of things, I think if you think about it though - let’s take tigers or something in India.

仅仅是做了一些小事，人们就把人类说成是当今地球上最聪明的生物，并肆意处置其他动物和进行一些暴行，为了帮助你理解，我用印度的老虎举个例子。

They have huge ranges and those people are very poor and they’re resource-poor, but

if they had abundant resources I don’t think they’re intentionally trying to kill off

these tigers - in some cases they are - but often it’s just because they need the land

for their cattle, and the tiger needs whatever number of kilometers squared to live, one tiger.

印度地域广阔，印度人很穷，资源贫乏，但是如果他们有丰富的资源，我觉得他们就不会故意杀死老虎 - 在某些情况下，他们是故意的 - 但往往只是因为他们需要土地来养牛，而仅仅一只老虎就会占用许多平方公里的土地资源。

And it’s just difficult with the number of people that are there.

对于印度人来说，不杀老虎是很难的。

So I think if we solve the kind of abundance and scarcity problem, then I think that opens up

a lot of conflicts both between humans as well as to do with resource scarcity, at the heart of it.

所以我认为如果我们解决这种资源丰富和稀缺的问题，就解决了人类之间的许多冲突，这也是资源短缺背景下的核心。

So I see, if we can solve a lot of these problems I can see a much better future.

所以我认为，如果我们能解决这些问题，就能预见到更美好的未来。

So Nick, you pointed out, the upside part of your book was a little shorter,

so now you have a chance to add something positive.

所以Nick，你说你书中的前半部分比较短，那么现在你可以再补充一些积极的方面。

What are you excited about?

让你高兴的地方都有什么？

There are really two sides to that.

包括两方面。

So one is getting rid of a lot of the negatives, like the compassionate use to cure diseases

and all other kinds of horrible miseries that exist on the planet today.

人能够使用技术摆脱许多负面的东西，比如用于治疗疾病和这个星球上现存的所有痛苦等。

So that is a large chunk of the potential.

也就是说这是一大潜力。

But then beyond that, if one really wants to see realistically what the positive things

are that could be developed, I think one has to think outside the constraints of our current

human biological nature.

但是除此之外，如果真的想看到积极的发展前景，我认为人们必须在我们目前的人类生物性质的限制之外思考。

That it’s unrealistic to imagine a trajectory stretching hundreds of thousands of years

into the future, we have superintelligence, we have material abundance, and yet we are

still these bipedal organisms with three pounds of gray tissue matter, with a fixed set of

emotional sensitivities and the hedonic set point that is kind of OK-ish for most people

but if you get - if something really good happens it lasts for a time and then you’re

back to the baseline.

想象未来几十万年之后发生的事情是不切实际的，比如我们有超级智能，我们有丰富的物质之类，但是目前我们仍然是眼前这个双足生物，灰质组织重约三磅，具有一套固定的情绪敏感度，对于大多数人来说，享乐的设定点是OK主义，但是如果你得到真正很好的东西，快乐的情绪会持续一段时间，然后又回到基线。

I think all of these basic parameters that sort of define the human game today, I think

become up for grabs in this future.

现在所有这些定义人类生活的基本参数，我认为在将来都可能抓不住了。

And it opens up this much vaster space of post-human modes of beings, some of which

I think could be wonderful, literally beyond our ability to imagine, in terms of the mental states,

the types of activities, the understanding, the ways of relating.

它打开了后人类的生活方式这个非常空间，我认为其中一些可能是美好的，在精神状态，活动类型，理解能力和关联方式等方面都超越我们的想象力。

So I don’t think we need a detailed blueprint for utopia now, what we need is to get ourselves

in a position later on where we can have the ability to use this to realize the values

that come into view once we’ve taken steps forward.

所以我认为我们现在不需要乌托邦的详细蓝图，我们需要的是让给自己找准一个定位，在那里一旦我们迈出了前进的步伐，就有能力使用它来实现价值，让想象成真。

Thank you, Nick.

谢谢Nick的发言。

What about you, David?

David，你呢？

I’m excited about the possibilities for AI making us humans smarter.

我对AI的前景感到兴奋，它能使我们人类更聪明。

I mean some of it is selfish - I turned 50 last year, my brain is gradually becoming

slower and older and dumber, but I’m not sure that I am, and that’s partly because

of all of the augmented intelligence technology we’re using.

Smartphones, and the Internet, and so on, they’re giving me all kinds of capacities,

extended capacities that I didn’t have before.

我说的话可能有些自私的 - 我去年50岁了，我的大脑正越来越迟钝，更老，反应越来越慢，但是我并不确定自己是否发生了这种变化，一部分原因是我们使用了增强型智力技术。比如智能手机和互联网等等，它们给了我各种能力，这些都是我之前不具备的。

And I’m really looking forward to AI helping with that.

我真的很期待AI能有这种能力。

In ten years or so once everyone is wearing augmented reality glasses with deep learning

built into it, then I’m really going to need that around 60.

在十年左右的时间里，如果每个人都戴着增强现实眼镜并深入研究它，那么我在60岁时也需要来一副。

And if you guys really get on the case and by the time I’m 70 or so we've got

real genuine AI or AI modules out there which can somehow come to be integrated with my

brain processes or maybe eventually we get to upload our entire brains onto AI, then there's a

way potentially to get smarter, more intelligent forever.

如果你们真的拿下了这一关，那么当我70岁的时候，我们已经有了真正的AI或者AI模块，它们可以以某种方式与我的大脑进程相结合，或者最终可以在AI上上传我们的整个大脑，那么就有一种方法能让人一直变聪明。

And this is not just selfish, although I can't say that doesn't motivate me,

but Demis talked about the AI scientists; I also like to think about the AI philosopher.

这不仅仅是自私，虽然这确实很激励我，但是，Demis谈到了AI科学家；而我的角度则是AI哲学家。

The problems of philosophy are really hard and many people have speculated that we humans

are just too dumb to solve some of them.

哲学问题真的很难，很多人都认为我们人类太笨了，只能解决其中的一些问题。

But once we've actually got AIs on the scene, maybe AI-enhanced humans, then maybe we're

going to be able to cross those thresholds where the AI-enhanced humans or maybe just

the AGIs end up solving some of those hard problems of philosophy for once and for all.

但是，一旦我们真正拥有了AI，那么也许被AI增强的人类就能够跨越这些门槛，然后被AI增强的人类或AGI就能够彻底地解决这些难题，一劳永逸。

Great, Ray, you have been a true pioneer in articulating positive visions of the future

in your writing.

很好，Ray，你在作品中阐述了很多积极的愿景，你可以说是一个伟大先驱了。

So if you picked the one that you're most excited about now, what would that be?

所以如果要你选让你最高兴的事情，那会是什么呢？

So imagine going back 10,000 years and asking the quintessential caveman and woman,

我们想象一下回溯一万年，我们拿这个问题去问穴居的男人和女人，

Gee, what is a beneficial future?

“哎，你认为理想的未来是怎样的？”

What would you like to see?

你希望看到什么？

And they would say, well I would like this fire to stop going out and I would like a

bigger boulder to prevent the animals from getting in the cave.

他们会说，我希望这把火别再蔓延了，我想要一块更大的巨石，防止动物进入我的洞穴。

Anything else?

还要别的吗？

Well no I think that would be pretty perfect.

不用了，我认为这样就很完美。

Well don't you want a better website and apps and search engines?

不想要更好的网站，应用程序和搜索引擎吗？

Imagine going back 2 million years ago and talking to primates - imagine if you could do that,

and saying, isn't it great that frontal cortex is coming and we're going to

have additional neocortex and and a hierarchy and they say, well what's the point of that?

And you say, well you'll have music and humor, and their answer would be, what's music?

What's humor?

想象一下，如果在二百万年前，你能和灵长类动物交流，你和他们说，我们要进化出额叶皮层了，我们将有更多的新皮质和层次结构，他们会说，那是什么意思？

你解释说，你能欣赏音乐，有幽默感，然后他们又会问你：什么是音乐？什么是幽默？

So they couldn't imagine concepts that they couldn't imagine, and by analogy I think we

will have new phenomena that are as profound as music and humor, you could call it more

profound music and we'll be funnier, but I think it'll be as profound as these great

leaps that evolution has brought us, because we will become profoundly smarter and if music

and humor are up here and we go to even higher levels of the neocortex, we're going to have

more profound ways of expressing ourselves and once we have that we would not want to

go back.

所以他们无法想象自己想象不到的概念，而且，通过类比，我认为将来会有一些像音乐和幽默感一样深刻的新现象，可以称之为更深刻的音乐，我们会有更深刻的幽默感，我认为就向进化给我们带来的巨大飞跃一样深刻，因为我们将变得更加聪明，会进化出更高层次的大脑皮层，我们表达自己的方式更为深刻，一旦我们前进了，就不想退回去。

What about you, Bart?

你的意见呢，Bart？

Well, I pretty much agree that we can't really predict much in advance, what we would like to have.

嗯，我也同意我们在预测我们需要什么时，不能提前预测的太远。

For myself personally I see the developments in mathematics and science and discovery,

and computers are just the hybrids of human computers there is quite incredible and makes the field -

makes what we do much more exciting.

对于我个人而言，我看到了数学，科学和发现的发展，电脑就是人类与电脑的结合，这是非常不可思议的，这让我们正在做的事情更引人入胜。

So I think that will be in the near future the first thing.

所以我认为这在不久的将来会是第一件事。

Great, and what about you, Stuart?

太好了，你呢，Stuart？

Well, so like Jeffrey Sachs - I think that for many of us, and probably like the cavemen

- that for many of us life is pretty amazing, and for many more of us it isn't.

那么，就像Jeffrey Sachs说的那样 - 我觉得对于我们许多人来说，可能就像是穴居野人一样。- 对于我们许多人来说，有的事物不可思议，而对于另一部分人来说，只是稀松平常。

And I think the best thing that AI can do, the big upside, is actually to fix the latter problem.

而我觉得，AI能够带来的最大的积极效应，实际上就是解决后一个问题。

I mean I love Nick's feeling that there are higher states of being that are so far above

our current 'pretty good', that that balances out all the 'pretty bad' that a lot of people are suffering.

我喜欢尼克的感受，还有更高级的状态，比现在普通的“好”还要好得多，这样就让很多人现在的痛苦不这么糟糕了。

But I really think the emphasis should be on the 'pretty bad' and fixing it, and eliminating

- so Demis was reading my notes apparently, from across the room - but eliminating the

scarcity basically eliminates the need for people to act in a zero-sum fashion where

they can only get by, by making it less possible for someone else to get by, and I think that's

the source of a lot of the nastiness that Jeffrey mentioned earlier.

但是，我坚持认为重点应该在于“非常糟糕”的地方，应该将其修复，并消除这种情况，因此，在房间另一端，Demis一直在阅读我的笔记，消除稀缺性基本上消除了人们零和方式行事的需要，在这种方式中一些人得到资源的后果就是其他人不能得到，我认为这就是Jeffrey提到的肮脏的根源。

So I think that would be my main upside, and not having to read so much email,

that would be the second one.

所以我认为这就是我看到的最大优势，可能不必阅读这么多的电子邮件能算得上第二位吧。

And for you, Elon, you've never articulated any visionary ideas about the future as far as I know, either.

那么你呢，Elon？目前为止，你还没有对陈述一些关于未来的远见性观点。

What about now?

现在可以发言吗？

I think I just - I have thought about this a lot, and I think it just really comes down

to two things, and it's solving the machine-brain bandwidth constraint and democratization of AI.

我刚才 - 我刚才一直在想这个问题，我认为只需要归结为两件事情，就可以解决机器的困扰 - 大脑带宽限制和AI的民主化。

I think if we have those two things, the future will be good.

我想如果我们落实了这两件事情，未来就会好起来的。

There was a great quote by Lord Acton which is that

'freedom consists of the distribution of power and despotism in its concentration.'

引用一下阿克顿爵士的话， “自由包括权力分配和专制集权”。

And I think as long as we have - as long as AI powers, like anyone can get it if they want it,

and we've got something faster than meat sticks to communicate with,

then I think the future will be good.

而且我认为只要我们拥有 - 拥有AI的力量，假如任何人只要想要就能拥有，那么我们有比拇指触屏更快的方式来进行沟通，那么我想前景应该是光明的。

Fantastic, so let's get - I know your caffeine levels are dropping dangerously low, and we

also have another panel after this, which is going to be really exciting to listen to,

so let's do a just a few quick questions.

太好了，那么，我知道大家的咖啡因水平已经下降到了危险的程度，而且在此之后我们还有另一个小组讨论，也十分值得听，所以我们来进行一些简单的环节吧。

Make sure that they are actually questions, and say your name and also say,

pick one person on the panel and address it just to them, OK?

确保能够提出真正的问题，说出你的名字，并在小组中挑一个人，向他们展开论述，好吗？

Yoshua?

Yoshua Bengio, Montreal.

Yoshua？

Yoshua Bengio，蒙特利尔。

And it's for Jaan - I found your presentation very inspiring, and one question I have is

related to the question of eliciting preferences and values from people.

Jaan - 我发现你的讲话非常鼓舞人心，我有一个问题涉及到人的们喜好和价值观。

Do you think this line of investigation could lead to better democracy, better society,

more direct democracy, and you know, what do you think about this direction to deal

with the issue of misuse and things like that.

你认为这种调查方式能带来更好的民主，更好的社会，更直接的民主吗？你怎么看待滥用问题和其他问题的处理？

Yes, absolutely.

是的，毋庸置疑。

There could be one code name for this, even, could be like 'Democracy 2.0' or 'U.N. 2.0'

or something like that.

这东西可能会有个代码，比如说“民主2.0”，“联合国2.0”之类的。

So, and as I mentioned in my presentation, just a lot of people today basically want to make the world better,

but it's kind of hard to distinguish them from people who say they

want to make the world better.

所以，正如我在讲话中提到的一样，今天很多人都希望让世界变得更好，但是很难将他们与真正想要让世界变得更美好的人区分开来。

So if there was actually kind of like a very easy measuring, like a metric that basically

would work as a Schelling point, focal point, then I think that would be super helpful.

所以我认为如果有一种简单的测量方法，就像米制的度量衡，谢林点，焦点这样，我认为应该会很有用。

And yeah, like democracy was invented like hundreds of years ago so, and clearly we have

advanced as a civilization and we have better knowledge about how to aggregate preferences.

没错，就像几百年前发明民主一样，显然我们已经作为一个文明发展壮大了，我们对如何集中优势有了更好的理解。

And Nicolas Berggruen, over there.

Nicolas Berggruen，在那边。

Thank you, Max. Nicolas Berggruen, so I have a very almost naive question.

谢谢你的发言，Max. Nicolas Berggruen，我有一个很幼稚的问题。

This is a very well-meaning group in terms of, let's say, intentions, but

who sort of, looking at who else is doing, potentially, AGI, it could be well beyond this group,

it could be in China, it could be any place.

这是一个非常有意义的团体，也就是说，我们的意图包括审查究竟还有谁在研究AGI，可能是这个团体之外的人，可能在中国，可能在任何地方。

And what happens because we've talked about how powerful AGI is, and if Elon is correct,

if it is distributed fairly, fine.

而且究竟研究到何种地步呢，因为我们刚才已经讨论过AGI有多强大，如果Elon是正确的，如果分配公平的话，那也好。

But if it isn't, is there a way to monitor today or in a year or in 10 years, because

once it's out it'll be fast.

但是，如果不是的话，有没有办法现在就进行监测，或者在1年内，10年内进行监测，因为一旦被研究出来，它的发展速度就会很快。

Who is monitoring it, who has a tab on it?

谁负责监控，谁来贴标签？

Because this is self-selected, but beyond...

因为这是自我选择的结果，但超出了…

Elon or Demis does either one of you want to take a swing at this?

Elon或者Demis，你们谁想回答这个问题吗？

Well I think this sort of relates to my point I said earlier about trying to build AI at

the hard part of the 'S' curve, so, which I think is where we sort of are at the moment,

as far as we can tell, because, you know, it's not easy to make this kind of progress, so you need

quite a lot of people who are quite smart and that community is pretty small, still,

even though it's getting rapidly bigger at places like NIPS.

好的，我认为这个问题与我刚才的观点相呼应，也就是关于在“S”曲线的困难部分建立人工智能的观点，所以我认为我们现在就处在这个地方，你也知道，在这里取得进展并不容易，所以你需要很多很聪明的人，尽管像NIPS这样的地方正在发展壮大，但整个群体还是很小。

And so most people know each other, so this is pretty representative of everyone in the

West, at least, obviously it's harder to know what's happening in China or in Russia, maybe.

But, you know, I think that you need quite a large footprint of resources,

people and very smart people and lots of computers and so on.

所以大多数人都互相认识，这也是西方人的代表性特征，至少我们很难知道在中国或俄罗斯发生了什么。但是，我认为这需要相当大的资源，人，非常聪明的人和大量的电脑等等。

So I think that narrows down the scope of the number of groups who can do that,

and it also means that they're more visible.

所以我认为缩小了可以做到这一点的团体人数的范围，也就意味着他们更加显眼。

So, you know, I think certainly in the West I think most people around here,

someone in this room will have contact with somebody who's in those groups who are capable of

making meaningful progress towards AGI.

所以，我想在西方，大多数在这里的人，也就是这个房间里的人会与那些有能力在AGI方面取得进展的人接触。

It's harder to know in the East and further apart, but we should try and make links to

those Chinese National Academy of Sciences, and so on, to find out more.

东半球的情况尚不清楚，更加分散，但我们应该尝试与中国国家科学院取得联系，了解更多情况。

But you know that may change in the future, I think that's the current state of it.

但是，将来情况可能会发生改变，我认为现在的状态就是这样。

Great, it's - the bad news is it's getting later in the day and we only have time for one more question.

太好了，但是坏消息是，时间已经很晚了，我们只能再问最后一个问题了。

The good news is there's a coffee break right after this so you can ask all of your questions

if you swarm the panel.

好消息是在这之后有一个休息时间，如果你能集合小组，就可以随意问想问的任何问题。

And the last question goes to you, Erik.

最后一个问题给你吧，Erik。

Do you want to stand up?

你想要提问吗？

Erik Brynjolfsson, MIT.

Erik Brynjolfsson 麻省理工学院

I'm going to pick up on the thing that Elon said at the end about democratizing the outcome

and tie it back to the panel yesterday where Reid Hoffman talked about people caring

a lot about not just absolute income but relative income, and I wanted to get the panelists'

reactions to the thoughts about whether or not AI had tendencies towards winner-take-all

effects, that there's a tendency for concentration, that whoever's ahead can pull further ahead,

or whether there's potential for more widespread democratic access to it,

and what kinds of mechanisms we can put in place if we want to have the widely shared prosperity that Elon suggested?

我要问的是Elon最后说的关于民主化结果的事情，并结合昨天小组讨论的事项，也就是Reid Hoffman说起人们关心的不仅仅是绝对收入而是相对收入，而且我想问问小组成员AI是否倾向于赢家通吃的效应，有集中的倾向，进步的人可以更进步，抑或是有更广泛的民主潜力，如果我们希望实现Elon提到的共同繁荣，我们可以采取什么样的机制呢？

Elon, do you want to take that?

Elon，你想回答这个问题吗？

Yeah, well, I mean I have to say that when something is a danger to the public, then

there needs to be some - I hate to say government agency, like regulators -

I'm not the biggest fan of regulators, 'cause they're a bit of a buzzkill.

是的，我的意思是，当事情对公众构成危险时，那就需要一些 - 我不喜欢政府机构，监管机构之类的词，我一向不喜欢监管机构，因为他们令人扫兴。

But the fact is we've got regulators in the aircraft industry, car industry, I deal with them all the time,

with drugs, food - and anything that's sort of a public risk.

但事实上，我们已经有飞机工业，汽车行业的监管机构，我一直在与他们打交道，像食品，药品等，以及任何有公共风险的事物。

And I think this has to fall into the category of a public risk.

我认为这件事也属于公共风险的范畴。

So I think that the right thing to do, and I think it will happen, the question is whether

the government reaction speed matches the advancement speed of AI.

所以我认为问题的关键就在于政府的反应速度与AI的进步速度是否相匹配。

Governments react slowly - or governments move slowly and they tend to be reactive,

as opposed to proactive.

政府的反应缓慢 - 或说政府行动缓慢，那么说明政府往往是被动的而不是主动的。

But you can look at these other industries and say, does anybody really want the FAA to go away?

and it's like people could just be a free for all for aircraft - like, probably not.

但是你可以看看其他行业，有人真的想让美国联邦航空管理局退出吗？那样飞机就能对所有人免费开放吗，我想不能。

You know, there's a reason it's there or just people could just do any kind of drugs and

maybe they work, maybe the don't.

你知道，有一个原因是，有的人只和毒品打交道，这些人可能工作，可能不工作。

You know, we have that in supplements, kind of ridiculous.

拿这个问题做补充，可能有点可笑。

But I think on balance FDA is good, so I think we probably need some kind of regulatory authority

and I think it's, like a rebuttal to that is, well people will just move to Costa Rica or something.

That's not true. OK, we don't see Boeing moving to Costa Rica or to Venezuela or wherever it's like free and loose

To Demis' point, the AI is overwhelmingly likely to be developed where there is a concentration

of AI research talent. And that happens to be in a few places in the world.It's Silicon Valley, London, at Boston, if you sort of figure out a few other places,

but it's really just a few places that really regulators could reasonably access.

And I want to be clear, it's not because I love regulators, OK?

They're a pain in the neck but they're necessary to preserve the public at times.

Alright, on that note, let's thank the panel for a fascinating discussion.

但是我认为，缉毒署是好的，所以我认为我们可能需要某种监管机构，我认为如果不这样，那很多机构就会搬到哥斯达黎加之类的地方。而这不是真的。波音公司没有搬到哥斯达黎加或委内瑞拉，或者任何自由而宽松的地方。根据Demis的观点来说，AI是绝对有可能在AI研究人才集中的地区发展的。而这正好就是世界上的这几个地方，波士顿，伦敦，硅谷，如果你要找其他的地方，那实际上也只是监管机构可以合理访问的几个地方。我要澄清一下，这不是因为我喜欢监管机构，好吗？他们令人痛苦，但是有时需要由他们来保护公众。好的，在这个问题上，让我们感谢专家小组为我们带来了一场引人入胜的讨论。