

AGI Strategy Course

{{group-name}}

1: The Race to AGI

A huge and warm welcome to the course!

In today's group discussion, you'll get to know your group, compare what you want AI to make possible and what you want to protect, and examine how different actors should respond to the Hugging Face incident.

If you need anything your facilitator and the BlueDot team are here to help! Just shoot us a message on Slack.

Participants

- While you're waiting for others to join the call, fill in your name and share some info about yourself in the table below.
- Once you're finished, read everyone else's responses and leave comments!

Facilitator

- While students are trickling into the meeting, encourage them to fill in the box below.
- At 0:03, introduce yourself to the group, including a brief overview of your background and motivation to facilitate the course.

Name and pronouns	What are you currently working on?	What motivated you to join the course?	What's something fun you've experienced in the past month?

[0:05-0:15] Speed 1-1

Meet one of the awesome people you'll be learning alongside!

- **Participants**
 - Explore something new with each person, and share one story about your background.
 - After the call, consider reaching out to people you didn't meet to schedule a chat.
 - **Facilitator**
 - Create 2-person [breakout rooms](#).
 - ⌘ 8:00 Send students into breakout rooms. After 8 minutes, close the breakout rooms.
 - Due to limited time, we suggest doing only 1x round of 1-1s here. There will be more 1-1s later.
-

[0:15-0:25] Visualising Wild Success

Understand the plan for this course, and how you can engage with it and each other most effectively.

Finish the onboarding by describing what a great outcome from this course would look like for you, and what actions could help you achieve it.

- **Participants**
 - Logistics
 - Before each discussion, **read the resources** and **complete the exercises**. Your exercise responses are used in the discussions!
 - Talk with your group in [Slack](#), and ask [#2-bluedot-support](#) when you need help.

- We follow the discussion norms below in this group. Give the table a skim.
- Wild success
 - Imagine you're at the end of this course, and it's been a wild success for you. How is your life different?
 - Consider things like the skills you've gained, opportunities you've pursued, collaborations you've developed, things you've built, etc.
 - What did you do to make this happen, and how did others in this group help?
- **Facilitator**
 - ⌘ 2:00 Logistics: Ask students to read the info. Answer any pressing questions they have, otherwise ask them to write any questions in the [#2-bluedot-support](#) channel.
 - ⌘ 7:00 Wild success
 - For 2 mins, ask students to individually reflect and write down what wild success would look like for them.
 - Have a brief discussion about everyone's goals for the course, and what they'll do to achieve them over the next few days.

Our group's discussion norms

- **Inclusive**
 - Critique ideas, not people, and assume good intentions.
 - Don't dominate the discussion: leave space for others.
 - Respect the confidentiality of personal stories and sensitive information.
- **Thoughtful**
 - Aspire to engage regularly with well-reasoned responses that add value to the discussion.
 - Support arguments with evidence or relevant examples.
- **Accessibility**
 - Avoid jargon. Make ideas or technical terms understandable for people with different backgrounds.
 - When introducing a new concept, provide extra context by "zooming out".
- **Collaboration**
 - View disagreements as opportunities to learn, not as conflicts to win.
 - Be comfortable sharing your thoughts, even if they're not fully formed.
 - Provide constructive feedback to help others develop their ideas.
- **Preparation**
 - Arrive on time.
 - Come prepared. Complete resources and exercises before discussions.

- When you have ideas on how to improve the discussions, share them with the facilitator.
- **Our community thrives when everyone feels respected and comfortable.**
 - [This form](#) allows you to confidentially share observations about interactions or behaviors that may not align with our community guidelines, or if you are worried about a community member's well-being.
 - You can also always reach out to the BlueDot team at team@bluedot.org.

Name	Visualising success

[0:25-1:00] Good Futures - or: what are we trying to achieve and protect?

The world is a crazy place. On [many metrics](#), it's getting better every day, but a good future isn't guaranteed.

This activity helps to ground the rest of the course in how good the future could be, and helps us envision what we're trying to build and protect.

Participants

- In pairs, share your responses to the two reflection exercises.
 - Discuss one benefit you hope AI could help deliver and one thing you want humans to retain control over. Help each other make these concrete: whose life improves, who makes decisions, and what could prevent the outcome you want?

Facilitator

- ⌘ 2:00 Explain the activity. Create 2-person breakout rooms.
- ⌘ 25:00 Run some quick polls (ask your participants to raise their hand if they agree with a statement). And then discuss each poll briefly after - If there are particular disagreements or surprises, double down on these. After you've wrapped up your polls, please have a full group discussion on just how crazy the future might get! The goal of this activity is to help people build better futurist instincts.
 - Some polls we like are:
 - Will there be a permanent-ish base on Mars by 2035?
 - Will humans still have to die of cancer by 2030? Note that this still could mean humans are dying from cancer, but in principle, they don't have to.
 - Will most humans still have normal jobs that are much like our jobs in 2040?
 - Will humans still be the dominant form of intelligence on planet Earth by 2035?
 - How old will the oldest human be in the year 2060?
 - Will the majority of humans still reside on Earth in the year 2050?
 - If mind uploading is possible, will the majority of humans choose to upload themselves?
 - Feel free to also come up with your own prompts!
- In this discussion, consider:
 - Where do you share priorities?
 - Where do your visions differ?
 - What conditions would allow people with different values to help shape the future?

Group discussion

Notes

[1:00-1:05] Break

[1:05-1:45] Responding to the Hugging Face incident

You are a group of stakeholders discussing the published incident reports. Your task is to recommend an immediate first response and set a course for future actions.

Participants

- In your [exercise](#), you selected one of four roles:
 - AI Company Executive
 - Government Official
 - AI Researcher
 - Public representative.
- You also read a narrative of the Hugging Face incident.
- Write down your recommendation, the main reason for it, and your most important uncertainty.
- As you do this, consider your character's motivations, constraints, and fears (from the [character cards](#)).

Facilitator

- ⌚ 2:00 Please give a brief introduction to the Hugging Face incident using your knowledge of it.
- ⌚ 3:00 Give the participants 3 mins to reflect and write down their decision below. This should be done under time pressure.
- ⌚ 10 :00 As a group go through the different positions.
 - Ask your participants to treat this as a role-play and step into their characters .
 - Listen for differences in what other actors believe, what they prioritize and what they have the authority to change. Identify one action that depends on cooperation between actors.
 - Draw out cruxes and force participants to justify their statements. .

[illegible]

Facilitator

- ⌘ 25:00 Facilitate a group discussion to develop a group recommendation covering:
 - Immediate action: What should happen now, and who should do it? Will they do it?
 - Scope: Which systems or activities does the response need to cover?
 - Evidence: What investigation or testing should be required before the next decision?
 - Accountability: Who will - who can - judge whether the response is sufficient?
 - Consider what happens if another company or country does not take similar action. Explain how that affects your recommendation.
 - Please record unresolved disagreements. Often a clear account of the disagreement may be more useful than some forced group consensus.
- ⌘ 10:00 Now step out of your roles and discuss:
 - Which disagreements were on the evidence, and which were about competing priorities?
 - Did competitive pressure change anyone's recommendation? How? Why?
 - What would need to be true for your group's response to actually work?
 - What's the most important question that went unanswered?

Group discussion

Notes

[1:45-1:55] Reflection

"We do not learn from experience. We learn from reflecting on experience."

Participants

- What was most valuable to you from today's discussion? What's one important thing you learnt?
- What action do you want to take as a result of this discussion? E.g. topics to research in more depth, people to follow-up with.
- How could the unit or discussion be improved for next time?

Facilitator

- ⌘ 5:00 Encourage participants to reflect deeply on the discussion, and write down their thoughts.
- ⌘ 5:00 Facilitate a final group discussion, asking each person in turn (scrolling down the table): **“What is one action you’re going to take as a result of this unit/discussion, and why.** This could be a follow-up resource you want to read, a specific claim you want to investigate, or other things. Maybe you want to even call your senators?”
- Finish the discussion by highlighting what we’ll discuss in the next discussion (see below).

Name	Takeaways	Feedback

Next discussion: [Drivers of AI Progress](#)

—END OF UNIT—

2: Drivers of AI Progress

Facilitator tip

Ask **targeted open questions** ("Alice, what/why/how?") instead of closed questions ("Does anyone...?").

And, **add constraints to your questions**, like "How might X impact Y, given Z?" or "If A was true, what would that imply for B?".

[Constraint questions](#) encourage students to focus on a specific aspect of a problem or scenario, instead of providing vague generic responses. It might feel like adding constraints limits their thinking, but [the opposite is true](#). Constraints are also a great tool for you to use to shape the direction of a conversation!

[0:05-0:20] Learnings and questions

What are your biggest questions and takeaways coming into this discussion?

Participants

- Reflect on the resources and exercises you engaged with before today's discussion.
- What new things did you learn? How does that relate to your area of expertise?
- What questions would you like to address today? E.g. confusions, surprises, disagreements.

Facilitator

- ⌘ 5:00 Ask participants to spend the first few minutes of the discussion individually reflecting and writing down their thoughts in the box below.
 - Encourage participants to read through each other's reflections. They can leave comments on what they'd like to explore further, what they agree or disagree with, suggestions they have for each other, etc.
- ⌘ 10 :00 Facilitate a group discussion on the big questions, themes or takeaways raised during the reflection.

Name	What did you learn?	Your open questions
------	---------------------	---------------------

[0:20-1:00] What's driving AI progress?

Look back at your exercise responses. Explain what you expect AI to accomplish a year from now, and how the drivers of progress could enable it. When thinking about your own work or studies, which part was hardest to forecast?



Participants

- **Teacher**
 - Your goal: **Help the learner understand the concept/material.**
 - Don't use your notes! [Recalling](#) information from memory improves your learning too.
- **Learner**
 - Your goal: **Understand the concept/material as deeply as possible.**
 - Ask many detailed questions. We encourage polite interruption!

Facilitator

- Explain the activity. The goal is to stress-test our understanding of the material and discover confusions, so they can be addressed. We'll be using your forecasts as the baseline.

- Explain what the work/job/activity you analyzed involves, what you expect AI to take over from you, and which drivers of progress support that expectation. Your partner will ask questions to test your reasoning.
 - The person explaining should:
 - Describe underlying drivers/mechanisms in plain language.
 - Connect it to an improvement needed for their chosen item.
 - Identify the evidence supporting that connection.
 - Make a forecast about the future trajectory applied to their item.
 - Their partner should probe the explanation:
 - What would the system become better at?
 - Why would that improvement help with this particular thing?
 - What could still prevent it from completing the whole job?
- Create 2-person breakout rooms, but don't open them yet. Designate the top person in each breakout room as the first "teacher", and the second person as the "learner".
- ⌘ 7:00 ⌘ 7:00 Open the breakout rooms. After 7 mins, send a [broadcast message](#) asking participants to **swap roles**. After the second 7 mins, close the breakout rooms.
- ⌘ 20:00 Bring everyone back into the main room for a group discussion!
 - Select two forecasts from different domains, ideally including one outside software. Discuss uncertainties and underlying dynamics in each.

Prompts

- What did you realise you were confused about while explaining?
- What's the most important thing you learnt during your discussion?
- What were some disagreements you both had?
- What eureka moments did you have?
- What's still really confusing?
- What analogies or stories are helpful for understanding this?
- What's the most important or surprising implications of this information?

Group discussion notes

[1:00-1:05] Break

[1:05-1:45] How much can AI research start accelerating itself?

Debate the motion: **Once AI systems can do the full job of a capable AI researcher, superintelligence is likely to follow within two years.**

For the purposes of this debate we hold that:

- Doing the full job of a capable AI researcher includes identifying promising research questions, designing and running experiments, interpreting results and deciding what to investigate next, with the level of oversight normally given to a capable human researcher.
- Superintelligence means substantially outperforming the best humans across most/all cognitive tasks.
- Likely means more likely than not.

You are debating how quickly progress would accelerate after research automation becomes possible. Please keep this separate from when you expect that starting point to arrive (you can discuss that part post debate).

Participants

- Work together with your team to come up with the most compelling arguments for your side, and the most compelling arguments responding to the other side. Nominate a speaker to give the opening statement.
- Rather than relying on rhetoric or emotion, you should try to give arguments that are convincing because they are true.
 - Imagine that your audience is an "intelligent martian." They are only moved by facts and reason, can follow complex claims, but don't know much or have strong views on the topic.
- Remember, your side in the debate might not mirror your side in real life. If you're on a side you disagree with, treat it as an opportunity to learn the other side's arguments so you can better defeat them in the future.

During preparation:

- Lay out the causal chain behind your argument. What becomes faster, and how does that affect subsequent AI development? Link them up!
- Try to use at least two specific pieces of evidence from the readings. Explain what each supports and what you may be extrapolating beyond it.

- Identify your argument's weakest assumption and prepare to address the strongest objection.
- Nominate someone to deliver your three-minute opening statement.

Questions to consider might be:

- How much would faster and more numerous AI researchers increase useful research output?
- Could improvements in AI software sustain an accelerating feedback loop?
- What constraints might still remain: experimental compute, data, coordination, physical infrastructure or increasingly difficult research problems?
- Would automating AI research produce rapid improvements across other capabilities?

Facilitator

- The format of the debate is that participants prepare for 10 minutes in their breakout rooms. Then, one speaker on each side delivers their 3 minute opening statement. After that, each side has the opportunity to ask questions of the other side. After that, you lead a final wrap up.
- ⌘ 2:00 Explain the debate format, allocate teams, and elaborate on how speakers should approach the opening statements
- ⌘ 10 :00 Move teams into breakout rooms to prepare
- ⌘ 3:00 Team A's nominated speaker gives their 3 minute opening statement.
- ⌘ 3:00 Team B's nominated speaker gives their 3 minute opening statement.
- ⌘ 10 :00 Team B asks Team A questions.
- ⌘ 10 :00 Team A asks Team B questions.

Team allocations:

Team	Position	Names
Team A	For the proposition	
Team B	Against the proposition	

Debate notes

Team A Notes	Team B Notes

[1:45-1:55] Reflection

"We do not learn from experience. We learn from reflecting on experience."

Participants

- What was most valuable to you from today's discussion? What's one important thing you learnt?
- What action do you want to take as a result of this discussion? E.g. topics to research in more depth, people to follow-up with.
- **How have your views changed as a result of the session?**
 - What probability do you now assign to the proposition from the debate?
 - What argument moved you the most?
 - What would you need to observe over the next year to think we were approaching the debate's starting point (of automated researchers)?
- How could the unit or discussion be improved for next time?

Facilitator

- ⌘ 5:00 Encourage participants to reflect deeply on the discussion, and write down their thoughts.
- ⌘ 5:00 Facilitate a final group discussion, asking each person in turn (scrolling down the table): **"What is one action you're going to take as a result of this unit/discussion, and why?"**
- Finish the discussion by highlighting what we'll discuss in the next discussion (see below).

Name	Takeaways	Feedback

Next discussion: [Pathways to Harm](#)

In the next unit, we'll discuss concrete pathways by which AI could be used to cause catastrophic harm. This will help us identify opportunities to intervene to prevent that harm from occurring.

—END OF UNIT—

3: Pathways to Harm

Facilitator tip

When participants disagree, help them identify the particular claim they disagree about. Ask: "Which step in the argument are you questioning?" and "What evidence would help us distinguish your views?"

Keep separating what has been observed from what is inferred. Participants can leave with different assessments; the aim is to make those assessments better informed and more precise.

[0:05-0:20] Learnings and questions

What are your biggest questions and takeaways coming into this discussion?

Participants

- Reflect on the resources and exercises you engaged with before today's discussion.
- What new things did you learn? How does that relate to your area of expertise?
- What questions would you like to address today? E.g. confusions, surprises, disagreements.

Facilitator

- ⌘ 5:00 Ask participants to spend the first few minutes of the discussion individually reflecting and writing down their thoughts in the box below.
 - Encourage participants to read through each other's reflections. They can leave comments on what they'd like to explore further, what they agree or disagree with, suggestions they have for each other, etc.
- ⌘ 10 :00 Facilitate a group discussion on the big questions, themes or takeaways raised during the reflection.

Prompts

-

Name	What did you learn?	Your open questions

Group discussion notes

[0:20-1:00] Presenting your arguments

How could AI systems pursuing the wrong things lead to humans losing meaningful control—and why might we be unable to put things right?

We'll examine two proposed pathways from the readings. Both depend on assumptions that you should question.

1. Pursuing unintended goals
 - a. Training could produce highly capable systems whose learned goals differ from what their developers intended. Such systems might perform well under oversight while concealing behaviour that would lead people to restrict them. Deployment could give them access to important resources and decisions. If they became sufficiently powerful, attempts to correct or shut them down might fail.
2. Optimising proxies and increasing dependence
 - a. Organisations could increasingly delegate decisions to AI systems optimising measurable indicators of success. Those indicators might diverge from what people actually value. As institutions become more dependent on these systems, understanding and correcting their behaviour could become harder. Humans could retain formal authority while losing the practical ability to steer events.

Participants

- Read the two arguments. Suggest corrections if you think they misrepresent the readings.
- Individually, identify the causal link you find least convincing or most uncertain. Explain why.
- During the discussion, help the group examine particular claims. You can contribute an explanation, an objection, relevant evidence or a missing assumption.
- Try to understand the strongest response to your objection.
- At the end, record one claim you would now revise or qualify. Your overall view does not have to change.

Facilitator

- ⌘ 5:00 Introduce the two arguments above. Invite brief corrections or clarifications. Keep this focused on establishing a shared starting point..
- ⌘ 5:00 Ask participants to identify their key uncertainty (about loss of control) and write their reasoning below.
- ⌘ 10 :00 Identify and discuss disagreements - feel free to quiz/question your participants.
 - Which assumption is doing the most work in this argument?
 - What turns this from a local failure into widespread, potentially irreversible harm?
 - Why couldn't people detect the problem, stop using the systems or recover control?
 - If this objection holds, does the pathway fail, or could it still happen another way?
 - Does this pathway require persistent power-seeking goals?
 - What evidence would make this particular transition more or less credible?
- ⌘ 20:00 Use the points discussed to facilitate a wider or more focused group discussion

Name	What are you questioning the most?	Why?	Updates

Group discussion

[1:00-1:05] Break

[1:05-1:45] What could go wrong even when AI follows instructions?

Imagine AI systems that can perform a wide range of demanding cognitive work. Each system reliably follows its designated operator's instructions and can be stopped by that operator. Operators may have harmful intentions.

Initially, access to these systems is concentrated among a small number of governments and large organisations.

Now imagine that comparably capable AI assistance also becomes affordable and widely accessible. Existing powerful actors retain their access.

How does that change the pathways to authoritarianism and biological catastrophe?

Participants

- You'll work in two groups.
- One will examine authoritarianism; the other will examine biological misuse.
- Your task is to explain one concrete pathway to catastrophic harm.

- How does broader access to AI change it, what important constraint remains, and which assumption are you least confident in?
- In your group:
 - Identify an actor and its objective. What could AI enable it to do that it otherwise could not?
 - Explain the pathway to harm. What connects that capability to a catastrophic outcome?
 - Examine the change in access. How does broader access affect this actor? What do its opponents or potential defenders gain?
 - Identify a remaining constraint and uncertainty. What else would have to be true for this pathway to succeed?
- Nominate someone to give a two-minute explanation. Other group members should be ready to help answer questions.

Facilitator

- ⌘ 2:00 Explain the comparison and allocate two groups of 3–4 participants. Clarify that control by an operator does not necessarily mean that society retains meaningful control.
- ⌘ 13 :00 Send participants into breakout rooms. Broadcast a reminder near the end to prepare their two-minute explanation and identify their most important uncertainty.
- ⌘ 25:00 Bring everyone back. Spend 10 minutes on each pathway:
 - 2 minutes: the group presents its explanation.
 - 8 minutes: the other group questions the argument. Invite presenting-group members to respond.
 - Keep presentations to two minutes so there is enough time to examine the reasoning.
- Compare the pathways in a full group discussion. Where does broader access change the risks most significantly? What important differences or uncertainties remain?

Name	Names
Group A (authoritarianism)	
Group B (bio)	

Group discussion notes

--

[1:45-1:55] Reflection

"We do not learn from experience. We learn from reflecting on experience."

Participants

- Complete these sentences:
 - The pathway I currently find most concerning is ... because ...
The assumption I am least confident about is ...
 - Evidence that would substantially change my assessment is ...
- Consider how today's discussion affected your view. You may have changed your mind, identified a new uncertainty or developed a more precise explanation of a view you already held.
- What action do you want to take as a result of this discussion? E.g. topics to research in more depth, people to follow-up with.
- How could the unit or discussion be improved for next time?

Facilitator

- ⌚ 5:00 Encourage participants to reflect deeply on the discussion, and write down their thoughts.
- ⌚ 5:00 Facilitate a final group discussion, asking each person in turn (scrolling down the table): **“What is one action you’re going to take as a result of this unit/discussion, and why?”**
- Finish the discussion by highlighting what we’ll discuss in the next discussion (see below).

[illegible]

Next discussion: Strategies for the AGI Transition

How on earth do we make this all go well? In the next unit we will look at strategies for navigating AGI.

—END OF UNIT—

4: Strategies for the AGI Transition

What would it take to secure a safe and broadly beneficial future as AI becomes more capable than humans?

By the end of this discussion, you should be able to

- explain how Plan A is supposed to work
- identify what would justify its consequential decisions
- give your own conditional recommendation.

Facilitator tip

When someone proposes an action, ask how it changes the outcome. When someone identifies a problem, ask what they would recommend instead.

Give participants time to think before inviting responses. Encourage them to consult the readings and distinguish what a source demonstrates from what they infer. Keep the group's important dependencies and unresolved objections in the shared notes/track them through the session - these will inform the final debate.

[0:05-0:20] Learnings and questions

What are your biggest questions and takeaways coming into this discussion?

Participants

- Reflect on the resources and exercises you engaged with before today's discussion.
- What new things did you learn? How does that relate to your area of expertise?
- What questions would you like to address today? E.g. confusions, surprises, disagreements.

Facilitator

- ⌘ 5:00 Ask participants to spend the first few minutes of the discussion individually reflecting and writing down their thoughts in the box below.
 - Encourage participants to read through each other's reflections. They can leave comments on what they'd like to explore further, what they agree or disagree with, suggestions they have for each other, etc.
- ⌘ 10 :00 Facilitate a group discussion on the big questions, themes or takeaways raised during the reflection.

Name	What did you learn?	Your open questions

Group discussion notes

[0:20-0:30] What makes Plan A work?

Make sure to build a shared understanding of the strategy before testing it.

The Plan is at a very high-level: Establish an agreement → use the resulting time for safety work → decide when and how to proceed → sustain safety and broad benefit.

Discuss:

- What makes each step enable the next?
- Who needs to act?

- Regardless of feasibility, does this plan address the concerns you've identified in previous units?
- Which connection depends on the most uncertain assumption?

Facilitator

- ⌘ 10:00 Invite one or two participants to explain Plan A (using the above sketch), then ask others to correct or extend their account.
-

[0:30-1:00] Can we trust the AIs to do our homework?

A frontier developer wants to use AI-generated safety research to justify building more capable AI. Humans cannot directly check every part of the research.

What would the evidence allow us to rely on?

Participants

- Consider three progressively stronger levels of permission:
 - Use the AI to propose safety methods for investigation.
 - Use its findings as evidence in a human decision to build a more capable system.
 - Allow it to decide whether further scaling is safe.
- Think back to the readings then jot down your thoughts below.

Facilitator

- ⌘ 8:00 Introduce the topic and ask participants to reflect individually.
- ⌘ 22:00 Facilitate a group discussion

Prompts

- How does it reduce the relevant risk?
- What evidence do the readings provide?
- How could that evidence look reassuring while an important problem remains?
- Which permission could it support, and what additional evidence would be needed?
- Could the evaluator share the system's blind spots?
- Would the behaviour persist under different incentives or deployment conditions?
- Does the evidence support the particular use being proposed?
- What changes when the successor is more capable?

Name	Notes

Group discussion

[1:00-1:05] Break

[1:05-1:25] What survives without cooperation?

Let's assume: There is no international agreement. The leading developer will not meaningfully delay to give safety work more time. You are a small safety-focused team with access to frontier models, but little influence over the developer's overall decisions.

Which interventions could you implement to still reduce risk here?

Participants

- In groups, prepare a short explanation:
 - We would pursue ____.

- It could reduce risk by ____.
- Under these conditions, we would need to change ____.
- It requires ____, which is available / remains unestablished.
- It would still leave ____ unresolved.

Facilitator

- Larger cohorts might benefit from longer or multiple breakout room discussions, while smaller cohorts might want to skip the breakout rooms altogether.
- ⌘ 2:00 Clarify/introduce the situation.
- ⌘ 8:00 Move participants into groups.
- ⌘ 10 :00 Facilitate a group discussion, ask each group to briefly present and then discuss!

Prompts

- Could the intervention work under the stated conditions?
- Is it worth prioritising?
- If folks propose changing political or organisational conditions, they must explain who could do that and how!

Name	Individual reflection

Group discussion notes

[1:25-1:50] Is this the one?

Consider and discuss:

Governments should make a Plan A-style transition their principal strategy for the AGI transition.

Consider and discuss:

Governments should make a Plan A-style transition their principal strategy for the AGI transition.

Facilitator

- 8 5:00 Let participants reflect briefly
- 8 20:00 Facilitate a group discussion.
- Call on 1 or 2 participants to share their positions
- Try to hit these themes:
 - Does the strategy hold up?
 - Suppose the early stages succeed - would you support what follows (think benefit-sharing, political power and the conditions for handing over authority)?
 - What would you recommend we do?

Prompts

- What would we actually do differently?
- What if the agreement weakens before alignment is sufficiently solved?
- Imagine the technical safeguards and international agreement work. What would still need to be true for you to support the later/final stages of the plan?
- Does your proposed revision preserve the strategy's mechanism or replace it?
- What costs does your preferred alternative introduce?
- Which commitments would you make now, and which would you defer?
- Separate disagreements about evidence, political feasibility and values

[illegible]

[1:50-2:00] Reflection

"We do not learn from experience. We learn from reflecting on experience."

Participants

- What was most valuable to you from today's discussion? What's one important thing you learnt?
- What action do you want to take as a result of this discussion? E.g. topics to research in more depth, people to follow-up with.
- How could the unit or discussion be improved for next time?

Facilitator

- ⌚ 5:00 Encourage participants to reflect deeply on the discussion, and write down their thoughts.
- ⌚ 5:00 Facilitate a final group discussion, asking each person in turn (scrolling down the table): **"What is one action you're going to take as a result of this unit/discussion, and why?"**
- Finish the discussion by highlighting what we'll discuss in the next discussion (see below).

[illegible]

Next discussion: Start Contributing

—END OF UNIT—

5: Start Contributing

[0:05-0:20] Priorities and reasoning

What Project Tailwind initiative(s) did you prioritize, and why?

Participants

- Which intervention did you prioritize from Project Tailwind?
- How would it help make AI go well, and why did you prioritize it over other options?
- After researching it, what is your biggest uncertainty or question for the group?
- Read each other's responses. Discuss the most consequential differences in your reasoning and leave comments.

Facilitator

- ⌘ 5:00 Ask participants to spend the first few minutes of the discussion individually reflecting and writing down their thoughts in the box below.
- Encourage participants to read through each other's reflections. They can leave comments on what they'd like to explore further, what they agree or disagree with, suggestions they have for each other, etc.
- ⌘ 10 :00 Facilitate a group discussion on the big uncertainties, themes or takeaways raised during the reflection.

Name	What did you learn?	Your open questions

Group discussion notes

[0:20-0:55] Making AI go well

What do you think is on **the critical path** to achieve a desirable future?

Facilitator

- Facilitate an open discussion.
 - What does success look like?
 - How do our values win?

Group discussion

[0:55-1:00] Break

[1:00-1:40] Presenting your action plan

What's your personal plan for and next steps for helping to make AI go well?

Making AI go well can feel like a daunting, overwhelming task.

In this activity, you'll share your action plan with your peers, and you'll help each other refine and improve your plans.

Participants

- Write your name in the box below, alongside your work/research.
- Support each other with recommendations and encouragement!

Facilitator

- ⌘ 15 :00 Send participants into 2-person breakout rooms.
- ⌘ 25:00 Bring everyone back into the main room. Lead an open-ended discussion on everyone's plans, focusing on making the actions more concrete and adding group accountability mechanisms.
- Consider setting up a WhatsApp group, a follow-up call in 1-2 months, and anything else you think might help your group have a big impact!

Prompts

- What does success look like for you in 1 year?
- What's your biggest uncertainty about your next steps?
- How can the people in this group help you?

Name	A URL to your action plan (ideally a google doc)

Group discussion

[1:40-1:55] Course reflection

"We do not learn from experience. We learn from reflecting on experience."

Participants

- What's been most valuable to you from this course?

- What's something you've really appreciated that someone else did or shared?
- If you could improve one thing about this course, what would it be and why?

Facilitator

- ⌘ 5:00 Encourage participants to reflect deeply on the discussion, and write down their thoughts.
- ⌘ 5:00 Facilitate a final group discussion, asking each person in turn (scrolling down the table): **"What's one big action you're going to take as a result of this course, and why?"**

Name	Takeaways	Feedback

Thank you so much for engaging with the course!!!

—END OF COURSE—

Useful links

Share any links you discuss here!