



Staying Involved in the AI Safety Field

ILINA PROGRAM · GUIDE

2026

Contents

1. Introduction.....	3
How This Guide Is Organised.....	3
Preliminary Note.....	4
2. AI Governance.....	5
Self-Study and Guided Courses.....	5
Sprints, Hackathons and Workshops.....	6
Seasonal Fellowships.....	6
Research Agendas.....	8
Newsletters and Podcasts.....	8
Career Guidance.....	9
Where to Look for More Opportunities.....	9
3. Technical AI Governance.....	10
Self-Study.....	10
Sprints and Hackathons.....	10
Seasonal Fellowships.....	10
Research Agendas.....	11
Career Guidance.....	12
4. Technical AI Safety.....	13
Self-Study and Guided Courses.....	13
Sprints, Hackathons and Bootcamps.....	13
Seasonal Fellowships.....	14
Research Agendas.....	15
Newsletters and Podcasts.....	15
Career Guidance.....	15
Where to Look for More Opportunities.....	16
5. Non-Research Track.....	17
Fellowships.....	17
Career Guidance.....	17
6. Cross-Cutting Resources.....	18
Where to Look for More Opportunities.....	18
Job Boards.....	18
Funding.....	18
Conferences and Events.....	19
Communities.....	19
Career Guidance.....	20

1. Introduction

One of our goals is to build an African talent pipeline for AI safety research and policy. Part of that includes helping our trainees continue learning, developing their skills and contributing meaningfully to the field after their time with ILINA.

This guide maps some of the routes available, including courses, short programmes, fellowships, research agendas, funding opportunities and communities we recommend.

How This Guide Is Organised

- We have organised this guide around four broad pathways into AI safety:
 - i) **AI Governance** - Research and policy aimed at ensuring that advanced AI is developed and governed safely, including work on national and international law, public policy, corporate governance and technical standards.
 - ii) **Technical AI Governance** - Technical research that helps make AI governance workable, by informing where and how to intervene and by supporting monitoring, verification and enforcement.
 - iii) **Technical AI Safety** - Research on the systems themselves, aimed at understanding and reducing risks from advanced AI. This includes work on alignment, evaluations, interpretability, model behaviour and other technical approaches to AI safety.
 - iv) **Non-Research Track** - Work that supports the development and functioning of the field, including policy strategy, field-building, operations, programme management and communications.
- The final section brings together resources that apply across all four pathways.
- We also group fellowships into three tiers, namely entry-level, intermediate and advanced or highly competitive. These are not official programme categories. They reflect our judgement of the experience and track record of candidates who tend to be selected, and may change as programmes and applicant pools change.

Preliminary Note

- **The field is more competitive than it was a few years ago.** Many of these opportunities attract large applicant pools, including candidates with advanced research, policy or professional experience. Therefore, it is important to be deliberate about where you focus your effort, based on how your current skills and experience compare with the requirements of each opportunity.
- **If you are new, focus first on learning.** Most seminar participants are at this stage. Build a strong understanding of the field and, for technical work, a solid foundation in machine learning. Move beyond reading into practice by reproducing results, running small experiments and testing model behaviour.
- **Build on the skills and experience you already have.** A background in law, philosophy, economics, political science, software engineering or another field may give you useful tools for particular problems. Identify where your existing strengths are relevant, what additional knowledge you need, and where you may be able to make a distinctive contribution.
- **Build a track record through substantive work.** Keep a record of what you produce and, where appropriate, turn it into public outputs such as blog posts, policy memos, code or GitHub repositories. This gives you material to improve through feedback and gives future programmes or employers concrete evidence of your skills and judgement.
- **References matter.** Competitive opportunities value recommendations from relevant people in the field who know your work and can speak credibly about its quality. Research assistantships, volunteering, collaborations and other forms of substantive contribution can help you build those relationships.
- **Use advanced programmes to understand what you need to build towards.** Even if you are not yet ready to apply, read their eligibility and selection criteria. These provide useful signals about the knowledge, skills and experience organisations value. Work backwards from those requirements and build the relevant capabilities deliberately.
- **Most of these opportunities are not Africa-focused.** Their research priorities tend to be centred on jurisdictions such as the U.S., U.K., E.U, or on global AI governance, and Africa-centric work is often not competitive.
- **Competitive does not mean out of reach.** Several ILINA trainees have gone on to positions at the Centre for the Governance of AI (GovAI), the Centre for Long-Term Resilience (CLTR) and the Vista Institute for AI Policy (Vista), among others.

2. AI Governance

Self-Study and Guided Courses

If you are new to AI governance, as most seminar participants are, structured study is the clearest way to build a grounding in the central ideas and open problems. Additionally, completing a course is itself evidence of sustained interest.

Self-directed reading

This is the most accessible way to stay engaged and to keep up with policy and technical developments.

- [BlueDot Impact's curricula](#): work through the self-paced [Future of AI](#) course, or study independently using the [AGI Strategy](#) and [Frontier AI Governance](#) curricula.
- 80,000 Hours' [essential readings on AI safety](#)
- [Research being published across the field](#)

Guided courses

Guided courses vary considerably in how competitive they are. Look carefully at who each course is designed for. For example, Vista's courses admit a single cohort that may include senior professionals and researchers, making it highly competitive, while BlueDot runs separate cohorts for participants at different levels, which may be more accessible for someone starting out.

- Vista Institute for AI Policy (Vista)'s [virtual courses](#): You can apply to [AI Agents and Law](#), [National Security Law and AI](#), or both.
- BlueDot Impact's [cohort-based courses](#): You can apply to the [AGI Strategy](#) and [Frontier AI Governance](#) courses.

Note: You must complete the AGI Strategy course before you can apply for Frontier AI Governance.

For a fuller list of open courses, both self-paced and guided, see [this course tracker](#), together with AISafety.com's [self-study](#) and [training](#) pages.

Sprints, Hackathons and Workshops

Short programmes require less time than a seasonal fellowship and can be a useful step towards one. They offer opportunities to develop skills, work with others in the field and produce a concrete piece of work.

- [Apart Research Sprints](#): monthly bootcamps and hackathons, with winners invited to Apart's fellowship. Past editions include the [Global South AI Safety Bootcamp](#), which may interest those who want to work on AI safety in the Global South.
- [AI Safety Camp \(AISC\) Research Incubator](#)
- [ML4Good Bootcamps](#) - Governance Track: ML4Good has previously run a bootcamp in South Africa, so it is worth watching for future editions in Africa or wherever you are based.
- [Institute for Law and AI \(LawAI\) Summer School on Law-Following AI](#): a two-day workshop for early-stage legal scholars interested in how to design agentic AI systems that reliably follow the law.
- [Global Challenges Project workshops](#): free, intensive three-day workshops for participants to explore AI safety.

Seasonal Fellowships

Governance fellowships generally expect some relevant research, policy or professional experience and the ability to work independently. Where a fellowship offers several projects, we recommend that you prioritise those for which your background and experience are particularly relevant.

Entry-Level Fellowships

Best suited to those newer to the field who can demonstrate strong aptitude (e.g., through academic performance) and sustained interest or a relevant background in AI governance, and ideally some research experience.

These are generally the strongest options in this guide for those who want to pursue Global South- or Africa-centred research. Some have previously offered projects focused on the Global South or particular African countries.

- [Black in AI Safety and Ethics \(BASE\) Fellowship](#) - AI Governance track
- [Apart Research Fellowships](#): You can apply directly to a partnered fellowship, or begin with a research sprint that can culminate in a fellowship.

- [Supervised Program for Alignment Research \(SPAR\) research mentorship](#): They have several governance projects on offer each round.
- [Orion AI Policy Accelerator](#): suitable if you are based in London.

Intermediate Fellowships

Best suited to candidates who can show some grounding in the field (e.g., through a piece of published output or a previous project) together with a clear sense of the field and governance questions they want to work on.

- [Pivotal AI Research Fellowship](#)
- [Cambridge ERA:AI Fellowship](#)
- [Future Impact Group \(FIG\) Fellowship](#) - AI Policy track
- [Mentorship for Alignment Research Students \(MARS\) Programme](#)

Advanced and Highly Competitive Fellowships

Best suited to candidates with a substantial track record, such as significant research experience, professional experience directly relevant to AI governance (such as policy work in the U.S. or E.U.), an advanced degree such as a Master's or PhD, or completion of another AI governance fellowship.

- [Centre for the Governance of AI \(GovAI\) Summer and Winter Fellowships](#)
- [Institute for Law and AI \(LawAI\) Summer and Winter Fellowships](#)

Note: The US and EU Law tracks require prior legal experience, through study or work, in those jurisdictions. The Legal Frontiers track, on the other hand, requires a deep, demonstrated understanding of the law and public policy of at least one of the jurisdictions most relevant to the Institute: the United States, the European Union or the United Kingdom.

- [Institute for AI Policy and Strategy \(IAPS\) Policy Fellowship](#): This is better suited to professionals who want to work on AI policy.
- [ML Alignment Theory Scholars \(MATS\)](#) - Policy and Governance track
- [Astra Fellowship](#) - Strategy and Governance stream

Note: These fellowships are not a good fit if your focus is AI safety in Africa. Applicants whose research experience lies solely in that area, or who apply with an African project, are unlikely to be competitive.

Fellowships through a Course or Sprint

- [Vista Fellowship](#): fellows are selected from those who complete one of Vista's courses.
- [Apart Research Fellowships](#): a strong showing in a research sprint can lead to a fellowship place.

Research Agendas

If you plan to do independent research or apply for a fellowship that asks for a project proposal, these agendas are a good place to look for promising directions:

- [Institute for Law and AI \(LawAI\)'s research workstreams](#)
- [Executive Branch AI and the Rule of Law: An Emerging Research Agenda](#) by Cullen O'Keefe, Alan Z. Rozenshtein and Christoph Winter.
- [Open Questions in Law and AI Safety: An Emerging Research Agenda](#)
- This list of [policy ideas for responsible government deployment of AI](#) by Stefan Torges (Forethought).

Newsletters and Podcasts

Subscribing to a handful of these is the least demanding way to keep up with policy and technical developments.

Newsletters

- [Transformer](#)
- [Import AI](#) by Jack Clark
- [AI Risk Management Newsletter](#) by Jonas Freund
- [AI Safety Newsletter](#) by the Center for AI Safety
- [AI Frontiers](#) by the Center for AI Safety
- [The EU AI Newsletter](#) by the Future of Life Institute
- [Digital Cognoscente](#) by Dr Chinasa T. Okolo

Podcasts

- [The AI Policy Podcast](#) by the Center for Strategic and International Studies
- [Dwarkesh Podcast](#) by Dwarkesh Patel
- [The 80,000 Hours Podcast](#)

Career Guidance

- [Writing advice for AI governance researchers](#) by Markus Anderljung.
- [FAQs and general advice on AI policy careers](#) by Miles Brundage.

Where to Look for More Opportunities

We recommend checking these megasources regularly to find more opportunities. They aggregate events, fellowships, courses, jobs and funding as they open.

- [80,000 Hours' collection of AI policy resources](#): fellowships, organisations, jobs and advice in one place.
- The [Africa AI Policy Opportunities Digest](#) by Dr Chinasa T. Okolo. Though not an AI safety publication as such, individual volumes often flag AI safety opportunities.

3. Technical AI Governance

Self-Study

There are no dedicated introductory curricula for technical AI governance, so following the field's output closely is the main route in. We suggest reading the work being produced by researchers and organisations in this area, including:

- [Centre for the Governance of AI \(GovAI\)](#) - under its technical AI governance research theme
- Machine Intelligence Research Institute (MIRI)'s Technical Governance Team - both [research](#) and [blog](#)
- [Technical AI governance research](#) by Cambridge ERA fellows
- Oxford Martin AI Governance Initiative (AIGI)'s [technical AI governance research](#)
- Evaluation organisations such as [Apollo Research](#), [FAR.AI](#) and [Redwood Research](#)

Sprints and Hackathons

Research sprints are one way to gain practical experience in technical AI governance and produce work you can point to in later applications.

- [Apart Research Sprints](#): monthly bootcamps and hackathons, with winners invited to Apart's fellowship. Past editions include [The Technical AI Governance Challenge](#).

Seasonal Fellowships

Most fellowships in this area prefer candidates with prior research or policy experience, a technical background, and the ability to work independently with minimal supervision.

Where a fellowship has several projects you can apply to, we encourage you to apply to as many as you reasonably can, and prioritise those where your background gives you an edge.

Intermediate Fellowships

Best suited to candidates who can demonstrate sustained interest in AI safety, some grounding in technical AI governance specifically (e.g., some output or completed project), a clear understanding of the TAIG field and the open problems, and some technical background.

- [Apart Research Fellowships](#): You can apply directly to a partnered fellowship, or begin with a research sprint that can lead to one.
- [Supervised Program for Alignment Research \(SPAR\) research mentorship](#)
- [Cambridge ERA:AI Fellowship](#) - Technical AI Governance area
- [Pivotal AI Safety Research Fellowship](#) - Technical Governance stream
- [Mentorship for Alignment Research Students \(MARS\) Programme](#)

Advanced and Highly Competitive Fellowships

Best suited to candidates with significant research or policy experience, an advanced degree (e.g., a Master's or PhD), or candidates who have another completed technical AI governance fellowship.

- [MIRI Technical Governance Team Research Fellowship](#)
- [ML Alignment Theory Scholars \(MATs\)](#) - Policy and Governance stream. Some mentors, such as Stephen Casper, now run streams focused specifically on technical AI governance.
- [Centre for the Governance of AI \(GovAI\) Summer and Winter Fellowships](#)
- [Frontier AI Security Residency](#) run by Cambridge ERA:AI, Heron and Oxford HAIGL

Research Agendas

If you plan to do independent research, or to apply for a fellowship that asks for a project proposal, these are good places to look for promising directions:

- [Open Problems in Technical AI Governance Database](#) by Anka Reuel and Ben Bucknall
- [AI Governance to Avoid Extinction: The Strategic Landscape and Actionable Research Questions](#): This is the research agenda of MIRI's Technical Governance Team.
- [Executive Branch AI and the Rule of Law: An Emerging Research Agenda](#) by Cullen O'Keefe, Alan Z. Rozenshtein and Christoph Winter. Several of the questions in the agenda are technical governance questions.

- [Open Technical Problems in Open-Weight AI Model Risk Management](#) by Stephen Casper et al.

The policy and technical newsletters listed elsewhere in this guide also regularly cover technical AI governance.

Career Guidance

- [Position: Technical Research and Talent is Needed for Effective AI Governance](#) by Anka Reuel et al, on what opportunities are available for technical talent in governance.
- 80,000 Hours' [technical governance career review](#)

4. Technical AI Safety

Self-Study and Guided Courses

If you'd like to continue learning about technical AI safety, we encourage you to explore this through self-study and/or guided courses. However, for technical roles, reading alone is rarely sufficient. We strongly recommend moving from passive reading to "active implementation" as soon as possible. This involves replicating key alignment papers, completing coding exercises (like those in ARENA), or participating in technical hackathons. This builds the "proof of work" necessary for intermediate fellowships.

Self-directed reading and exercises

- [BlueDot Impact's Technical AI Safety curriculum](#)
- [ARENA curriculum](#): See also this [guide to working through the ARENA curriculum on your own](#).
- [The Iliad Intensive course materials](#)
- [Shallow review of technical AI safety, 2025](#)
- [Research](#) being published across [the field](#)

Guided courses

- [BlueDot Impact's Technical AI Safety course](#)
- [The Iliad Intensive](#): full-time and in person in London, and tailored to applicants with a strong background in mathematics, physics or theoretical computer science.

For a fuller list of open courses, both self-paced and guided, see [this course tracker](#), together with AISafety.com's [self-study](#) and [training](#) pages.

Sprints, Hackathons and Bootcamps

Short programmes ask less of your time than a seasonal fellowship and are often a pathway to securing fellowships as they give you an opportunity to build proof of work.

- [Apart Research Sprints](#): monthly bootcamps and hackathons, with winners invited to Apart's fellowship. Past editions include the [Global South AI Safety Bootcamp](#), which may interest those who want to work on AI safety in the Global South.
- [AI Safety Camp \(AISC\) Research Incubator](#)

Seasonal Fellowships

Most technical AI safety fellowships now expect candidates to bring:

- a strong programming or mathematical foundation, in computer science, mathematics, physics or a comparable discipline;
- familiarity with deep learning frameworks, primarily PyTorch, and with transformer mechanics;
- a demonstrated interest in alignment, shown through self-study, paper replication or hackathon participation;
- the ability to work independently with minimal supervision.

We encourage you to carefully consider what prerequisites and technical skills are expected before applying to ensure you are a strong candidate and avoid spending effort on applications where you might not yet be competitive.

Entry-Level Fellowships

Best suited to those newer to the field who have begun building technical skills and can show evidence of it, however early-stage.

- [Black in AI Safety and Ethics \(BASE\) Fellowship](#) - AI Alignment and AI Security tracks
- [Apart Research Fellowships](#): You can apply directly to a partnered fellowship, or begin with a research sprint that can culminate in a fellowship.

Intermediate Fellowships

Best suited to candidates with a solid technical foundation, and perhaps, some “proof of work” they can point to.

- [Cambridge ERA:AI Fellowship](#)
- [Pivotal AI Safety Research Fellowship](#) - the technical AI safety projects
- [Supervised Program for Alignment Research \(SPAR\)](#) - technical projects
- [Cooperative AI Research Fellowship](#)

Advanced and Highly Competitive Fellowships

Best suited to candidates with substantial technical background or research experience.

- [ML Alignment Theory Scholars \(MATS\)](#)
- [London AI Safety Research \(LASR\) Labs Fellowship](#)

- [Astra Fellowship](#) - Empirical stream.
- [Alignment Research Engineer Accelerator \(ARENA\)](#)
- [AI Alignment Foundation Fellowship](#)
- [Iliad Fellowship](#)

Note: The Iliad Fellowship is aimed at more advanced researchers, at PhD or postdoctoral level.

Research Agendas

If you plan to do independent research, or to apply for a fellowship that asks for a project proposal, you can start from these research agendas to identify promising research directions.

We recommend using this [Airtable of research agendas](#) to browse through different agendas at a high level before committing to a deep dive into particular papers or organisations.

It is also worth reading the directions set out by leading AI companies and funders:

- Anthropic's [Recommendations for Technical AI Safety Research Directions](#) (January 2025)
- Coefficient Giving's [Technical AI Safety Research Areas](#) (2025)

Newsletters and Podcasts

Subscribing to a handful of these is the least demanding way to keep up with technical developments and discussions.

Newsletters

- [ML Safety Newsletter](#)
- [Transformer News](#)

Podcasts

- [Dwarkesh Podcast](#) by Dwarkesh Patel
- [The 80,000 Hours Podcast](#)

Career Guidance

- Benjamin Hilton's career guide on [how to enter AI safety technical research](#)

Where to Look for More Opportunities

We recommend checking these megasources regularly to find more opportunities. They aggregate events, fellowships, courses, jobs and funding as they open.

- [80,000 Hours' technical AI safety upskilling resources](#): fellowships, organisations, research ideas and advice.
- [AI Alignment Resources](#) by Victoria Krakovna
- [Levelling Up in AI Safety Research Engineering](#) by Gabriel Mukobi
- [What are some exercises and projects I can try?](#) by AISafety.Info

5. Non-Research Track

Research is not the only way to contribute to AI safety. Organisations in the field also need people working in policy strategy, field-building, operations, programme management and communications. This may be a suitable path for those with aligned skills and strengths.

Fellowships

- [Pathfinder Fellowship](#): this offers mentorship, funding and other support for people who want to organise a technical AI safety or AI policy university group. Pathfinder already supports several university groups in Africa, including at Ahmadu Bello University, the University of Nigeria, Lagos State University and the University of Cape Town.
- [Supervised Program for Alignment Research \(SPAR\) research mentorship - generalist projects](#): suitable for those new to the field (but have some general understanding of the AI safety field). The projects give you context on the field, experience with generalist work, and preparation for opportunities such as the Generator Residency.
- [Kairos x Constellation Generator Residency](#): suitable for more advanced candidates with relevant experience and a 'track record of execution'.
- [GovAI UK Summer and Winter Fellowships - Applied Track](#): for those with relevant experience who want to work in communications, policy, programme management, operations, fundraising, advocacy or events.

Career Guidance

- An overview of the [non-research AI safety jobs](#) that you can get into.

6. Cross-Cutting Resources

Where to Look for More Opportunities

We recommend checking these mega-sources or guides regularly to find different types of opportunities and resources including upcoming events, fellowships, courses, jobs, funding opportunities, etc.

- [AISafety.com](#): for [training opportunities](#), jobs, funding and more. Its [map of the AI safety field](#) charts organisations across the ecosystem, from training and governance to technical safety and funding and many others.
- [The Safety Apprentice](#): a guide for newcomers, covering learning and funding opportunities, fellowships, jobs and communities.
- BlueDot Impact's [AI safety resources page](#): a curated list of introductory readings, podcasts, newsletters and funding sources.
- AI Safety Support's [Lots-of-Links](#) page

Job Boards

Various EA-affiliated and AI safety-related opportunities boards are regularly updated to reflect job openings in AI governance, technical AI safety and technical AI governance research as well as in policy and operations.

- [80,000 Hours job board](#)
- [AISafety.com job board](#)
- [EA Opportunities Board](#)

Funding

There are several funding opportunities that can support a transition into the field or a specific project. However, funding has become competitive, so we advise targeting the funding that is accessible to new entrants and where you will be competitive.

Educational funding, in particular, such as for Master's programmes, is now among the hardest to secure and it often takes several attempts.

Accessible Funding

Suitable if you are new to the field.

- BlueDot Impact's [Rapid Grants](#) and [Career Transition Grants](#).

Note: You are more likely to receive a BlueDot grant if you are a course participant, alumnus, facilitator, or an active member of the AI safety community.

- [Transformative AI Fund](#) (formerly Long-Term Future Fund)

More Advanced Funding Opportunities

Best suited to those with a clear track record and a well-defined project that matches the funder's priorities. Also suitable for those planning large-scale research projects or field-building work.

- Coefficient Giving (formerly Open Philanthropy): [career development and transition funding](#), [funding for programs and events](#) and [project-based grants](#)
- Future of Life Institute (FLI) [project-based grants](#).
- Longview Philanthropy [grants and career funding](#): check their funding opportunities page for the current requests for proposals.

AI Safety.com's [funding page](#) lists further options.

Conferences and Events

We also encourage you to attend the major conferences and events. These are one of the best ways to meet researchers, grantmakers and policy professionals, and to build the relationships that make later applications stronger.

- [IASEAI Annual Conference](#): the International Association for Safe and Ethical AI's annual meeting, which examines critical questions in AI safety, ethics and governance.
- [EAG conferences and EA summits](#): AI safety is one of the cause areas of Effective Altruism. A significant share of the attendees and talks at the EA conferences focus on AI safety, hence the conferences are a great opportunity to meet people from across the ecosystem, including researchers, grantmakers and policy professionals.

AI Safety.com's [events page](#) keeps a fuller and more current list, and you can subscribe to its newsletter to get weekly updates on new events.

Communities

If you want to stay involved in a less demanding way, we recommend joining one of the AI safety communities across the world.

On the continent, you might join [AI Safety East Africa](#) or [AI Safety Nigeria](#), alongside [local EA groups](#). This [global map of AI safety communities](#) lists many more.

Career Guidance

For those transitioning into the field, these are the resources we recommend.

- [Tips for asking people for things](#): on cold emailing, which has opened doors for many early-career AI safety researchers when done well.
- [How to get into AI safety](#): practical tips for pivoting into the field.
- [Successif](#): Offers free personalised career advising for experienced professionals moving into AI safety.
- [Magnify Mentoring](#)