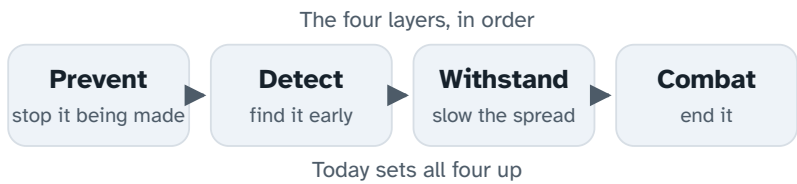


Day 6: Contributing to Biosecurity

Saturday 31 October · discussion 3:00 to 4:50 pm New York

THE DAY MAP

Which single piece of this is yours to work on, and what is your first move?



9 readings here, about 63 minutes of reading. Discussion Saturday 31 October, 3:00 to 4:50 pm New York, with Conor Spence.

6.1-1	Become a person who Actually Does Things original paper →	3 + 5 min
6.1-2	How to be more agentic original paper →	3 + 9 min
6.3-1	Navigating dual-use projects in biosecurity original paper →	3 + 6 min
6.3-2	A hundred project ideas in biosecurity original paper →	3 + 8 min
6.3-3	Five Tractable Biosecurity Projects You Could Start Tomorrow original paper →	3 + 7 min
6.3-4	Rapid Small Grants for Course Participants original paper →	3 + 4 min
6.4-1	Biosecurity Personal Action Plan Template original paper →	2 + 0 min
6.4-2	Biosecurity 1-Pager Template original paper →	2 + 0 min
6-next-steps	Next steps: fellowships, policy, careers and funding original paper →	2 + 0 min

WHAT THE COURSE ASKS YOU TO DO

6.2 Prioritise a single intervention

Based on everything you've learnt about so far during this course, pick a single biosecurity intervention.

To help you prioritise, consider which interventions you think would be most effective and where your existing skillset gives you an edge to contribute.

6.2 Do your own research

- What does success look like with this intervention? How does it *actually* help make the world more pandemic-proof?
- What's the current status of this intervention? Are governments, companies or other actors already working on it? If not, why not?
- Which organisations are working on this which you can contribute to or join?
- Are there any concrete project ideas that you could pick up?

We've included a few resources with projects, orgs and jobs on the following pages.

6.4 Create your 1-pager

Spend ~1 hour on this.

Once you're finished, make the Google Doc shareable, submit it below, and share it with your group in Slack!

THE READINGS

6.1-1 · CORE · COURSE TIME 5 MIN

Become a person who Actually Does Things

Neel Nanda · 2022

[Read the original paper →](#) · the course assigns: whole piece

Being the kind of person who acts on their own goals is not a personality you are born with, it is a trainable skill, and the way to train it is to make "I do things" part of your identity instead of waiting until you are sure you have the right plan.

Average summary

Nanda opens with a line he admits is "**unnecessarily provocative**": there are two kinds of people in the world, people who actually do things and people who don't. What he means is that achieving any goal breaks into two parts, figuring

out what needs to be done and actually doing it, and almost everyone pays attention only to the first.

He says part of him is convinced the main bottleneck on his goals is being smart, having the best plan, having all the information. That is true as far as it goes: it does not matter how good you are at doing things if the thing you are doing is no good. But the reverse is also true, and gets forgotten. It does not matter how good the plan is if he never acts on it.

The symptom he describes is familiar and uncomfortable. He is aware of problems in his life, sometimes for a very long time, and nothing connects "this is a problem" to "what am I going to do about it". It is not that he thinks the problem through and concludes nothing can be done. It is that **it never feels like there is a question to be asked in the first place**. He thinks this block, between things not being as you want them and believing you can do something about it, is extremely widespread, and he sees some amount of it in most of his friends.

His claim is that this is solvable. Agency is a skill and it is trainable, imperfectly and with effort, and it is one of the most valuable skills he has tried to develop. Because the failure happens in the moment, when acting never feels urgent or even possible, the fix has to work in the moment too. His fix is to make being an agent **part of your identity**: do the thing as a point of principle, not because it is the perfect action, but because you are choosing the life where you do things over the life where you wait for the perfect opportunity.

In practice that means noticing small things and acting on them. Fill the empty water jug. Be the first to say out loud that nobody is enjoying this. Do today the thing you told yourself you could do tomorrow. Learn the juggling you have always wanted to learn. Notice the inefficiency, notice the job nobody is doing, be the person who does it. The trick is that this changes the question you are answering, so perfectionism has nothing to grip: it does not matter much whether each small action was the right one, because today's choices are building the kind of person you will be. Figuring out the right thing to do matters too, he says, but it can come second.

Why it's assigned

This is the first thing the course gives her in the final unit, and it sets the unit's whole tone: the course has just told her that biosecurity work "won't be a 'normal job'", with few instructions and possibly no boss, and that if the role she wants does not exist she will have to build it. Nanda's piece is the argument that the capacity to do that is learnable rather than innate. Unit 6 is not one of the

defence layers (prevent, detect, withstand, combat); it is the unit about becoming one of the people who builds those layers.

In depth

The two halves of any goal

Nanda says the structure of achieving a goal is: figure out what needs to be done, and actually do it. He concedes immediately that this "sounds completely trivial" and argues there is a non-obvious point inside it. The non-obvious part is a mismatch between where people put their effort and where the failure actually happens. Some part of him is convinced that the bottleneck on his goals is intelligence, plan quality and information, and he grants that this belief is not wrong: a brilliantly executed bad plan is still bad. But a brilliant plan that is never executed is equally useless, and he thinks almost nobody budgets attention for that second failure.

Why it is easy to miss

The thoughts that stop action are ordinary and plausible: "I can do this tomorrow", "this is not a priority", "this would never work". He is explicit that the difficulty is that **sometimes those thoughts are correct**. That is what makes them hard to argue with. But his observation about himself is that without active effort he is the kind of person who will always have those thoughts, about everything, forever, and so will never act on his goals or change anything about his life.

The symptom: no question gets asked

The main symptom he names is long-running awareness of a problem with no accompanying sense that anything could be done. His own examples are personal and unglamorous: he does not exercise enough, he is not enjoying the courses he is studying, he really dislikes being single, he does not have enough time for what he finds most fulfilling. The point of the list is the thing that is missing from it. There is no moment where he considers an action and rejects it. The link from "this is a problem" to "what am I going to do about it" simply never forms. He describes it as a mental block between "things are not as I want them to be" and "I can actually do something about this", says he notices some amount of it in most of his friends, and calls it incredibly widespread.

The claim: agency is trainable

He calls this one of the biggest meta-problems holding a person back, meaning a problem that sits above all your other problems and stops you from fixing any of them. Then he turns: the reason he is writing is not to complain but because he thinks it is solvable. Being an agent is a skill, and he thinks it is a trainable

skill. He is careful about the size of the claim. Training it is hard, it "won't work perfectly", but there is a lot of room for progress, and he rates it one of the most valuable skills he has ever tried to develop.

Why the solution has to work in the moment

His diagnosis determines his prescription. The failure is not an argument you lose on paper, it is something that happens in the moment: acting never feels urgent, never feels important, can be put off, or the problem simply feels like a fact of life rather than a thing with a cause. So, he says, a solution must centre on solving the problem in the moment.

The solution: make it identity, not calculation

What worked for him is making it part of his identity to be an agent, and making it a point of principle to do things. Crucially, not because the thing is the perfect action. Because he chooses the life where he does things over the life where he always waits for the perfect opportunity. Then comes the list of what that looks like day to day: notice the small problems and fix them; notice when everyone is not enjoying what they are doing and be the first to say so; notice the empty water jug and be the one to fill it; notice "I can do this tomorrow" and do it today; notice "I should get round to this some time" or "I've always wanted to learn juggling" and actually do it; notice the inefficient thing, notice the thing nobody is doing, and be the person who does it.

Why identity beats calculation

He explains the mechanism. Making it an identity lets him dodge the paralysis by changing the question he is answering. It does not matter whether he is doing exactly the right thing, because the thing itself is not that important. He closes off the paralysing thoughts not by answering them on their own terms, which is a debate he would lose or never finish, but by noticing that the choices he makes today determine the kind of person he will be for the rest of his life, which is far more important than whatever is going on today. That is how the habit gets trained. Figuring out the right thing to do is also a vital skill, he says, but it can come second. **Don't let the perfect be the enemy of the good.**

What the training is actually for

The last move is the one worth carrying into Unit 6. He says it does not even matter if he makes the wrong or imperfect choice in all the day-to-day mundanity. What matters is that when the one golden opportunity arrives, he is already the kind of person who will take it: "A thousand tiny losses are nothing against the big win that really counts." Small daily agency is practice for the

decision that will actually matter, and you do not get to know in advance which one that is.

His closing instruction is a question swap. Next time you agonise over a choice or pass up an opportunity, ask not "what is the right decision" but "which decision will get me closer to the kind of person I want to be".

Words to know

- **Agent** (as he uses it): a person who notices something they want changed and acts on it, rather than only registering it. The whole piece is about becoming one, which is why the course puts it first in a unit about building things nobody has told you to build.
- **Meta-problem**: a problem that sits above your other problems and blocks you from solving any of them. He calls the gap between noticing and acting one of the biggest meta-problems a person can have.
- **Don't let the perfect be the enemy of the good**: an old saying that insisting on the ideal version often means getting no version. He uses it as permission to act on an imperfect plan.

What the author is unsure about

He is unusually honest that his own stopping thoughts are sometimes right: that is exactly why he finds them hard to override, and he offers no test for telling a correct "this is not a priority" from a cowardly one. He says training agency is hard and will not work perfectly. He concedes that figuring out the right thing to do is also a vital skill, so his advice is a sequencing claim, not a claim that thinking does not matter. The whole piece is personal opinion and self-observation, with no data: his evidence that the problem is widespread is that he notices it in most of his friends and thinks it is common. He also admits the opening line is more provocative than it needs to be.

For the discussion

- The unit's own framing says biosecurity work will not be a normal job with instructions and a boss, and Nanda's piece is the reason that matters: the bottleneck on a 1-pager is not knowing which intervention is best, it is whether anything gets sent.
- There is a real tension between this and the dual-use reading later on the same day: Nanda says act before you are certain and do not worry much about whether each small action is right, while the course's dual-use advice says some projects need checking with experienced people before you start at all.

- Worth asking the group: what is the smallest thing on each person's list that could be done by the end of the week, since his argument is that the practice matters more than the choice.

6.1-2 · CORE · COURSE TIME 15 MIN

How to be more agentic

Cate Hall · 2024

[Read the original paper →](#) · the course assigns: whole piece

Agency is not a trait you either have or lack, it is a set of learnable habits built around finding "real edges", the things you are willing to do that other people would rather not, and Hall gives six of them from her own career.

Average summary

Hall starts by rejecting the common framing. People talk about agency as an inherent trait, either you have it or you are stuck "in the minor leagues", and that has not been her experience. She says that as she has **"gradually grown dumber relative to my peers"** through aging and making smarter friends, the main way she has compensated is by dialling up agency, which she defines as something like **"manifest determination to make things happen."**

Her track record is the argument for taking her seriously. She was a Supreme Court advocate and the number one female poker player in the world, started art and perfume companies, and led operations at Alvea, a pandemic medicine company she co-founded, when it set the record for the fastest startup to take a drug to clinical trials. All of it in her 30s.

Her central idea is the **real edge**: something you are willing to do that others aren't, often because it is annoying or unpleasant. She warns that these choices usually do not surface clearly enough for you to notice you are choosing. They sit in "a cloud of aversion that strategically obscures the tradeoff". Her illustration is poker. The game is extraordinarily competitive, and even eight years ago professionals were spending nearly as much time studying as playing, hunting tiny mathematical advantages with solver software. Meanwhile a large edge sat in plain sight, reading opponents physically, and almost everyone ignored it. She and two friends studied reads obsessively and all three got out-of-distribution results. Other professionals told them "nuh-uh, that's not a thing", which she suspects was about not wanting to feel obliged to study it.

Then six habits. **Court rejection**: ask for things that feel unreasonable so your sense of what is reasonable stays calibrated, and if you only get yeses you are

not aiming high enough. **Seek real feedback**, which she calls the most overpowered item and the lowest hanging fruit; without it "you're cooking without tasting", and the way to get it is to make it anonymous and easy to find.

Increase your surface area for luck: take meetings with no obvious benefit, because she has found she cannot predict which call will be useful. **Assume everything is learnable**, including traits people treat as fixed, like confidence, charisma, warmth, tranquility and optimism. **Learn to love the moat of low status**, the stretch where you are visibly bad at the new thing and cannot learn it without being seen to be bad at it. And **don't work too hard**, which she says might be the most important item, because burnout is the ultimate agency-killer.

Her close: agency is "an all-purpose life intensifier". No one is born with it, everyone can learn it, and it is never too late.

In depth

Agency as something you can dial up

The piece is framed against a view Hall says she hears often, that agency is innate. Her counter is autobiographical: she got better at it on purpose, as compensation. Her definition is worth keeping exact, because it is about disposition rather than skill: "manifest determination to make things happen." The list of what that bought her is deliberately varied, law, poker, art, perfume, pandemic medicine, and she notes all of it happened in her 30s, which is her first argument that it is never too late.

Real edges, and the cloud of aversion

Her frame for agency is not working harder. It is "finding real edges: things you are willing to do that others aren't, often because they're annoying or unpleasant." She contrasts this with "eking out wins by grinding harder than everyone". The subtle part is her claim that these edges rarely present themselves as a conscious choice. They live in "a cloud of aversion that strategically obscures the tradeoff", meaning your mind files the unpleasant option as not worth considering before you ever notice it was an option, so you never have to decline it on the record.

Poker made it click. Modern poker is extraordinarily competitive, and even eight years ago professionals were studying nearly as much as they were playing, using solver models to find tiny mathematical advantages. At the same time a massive edge was available in physical reads and almost entirely ignored. She declines to give a concrete example of a read, saying it is "like explaining a magic trick", and apologises for it. She and two friends studied reads intensely and all of them had out-of-distribution results, meaning results well outside the normal spread. When they told other professionals, most said "nuh-uh, that's not

a thing". Her read on that: they were not willing to consider that reads were valuable, "maybe because they didn't want to feel obligated to study them." She then says all of her agency hacks are like this, big glaring edges people might rather ignore.

Court rejection

Ask for things, including things that feel unreasonable, so you can check whether your intuitions about what is reasonable are accurate, and try not to be a jerk while doing it. Her test is clean: if you are only asking for things you get, you are not aiming high enough. Applied to jobs, especially early in a career, you should expect to be rejected from most things you apply for, and if you have not yet learned to absorb rejection you should court it deliberately, applying for jobs you do not think you will get until "no" stops arriving with surprise and dejection attached.

Her own example is a cold email she says she would not have dared send a few years earlier, along the lines of "I'm planning to start an organization similar to yours; would you consider letting me run yours instead?" The response was nothing at all. Maybe that person thinks she overstepped. Her point is that it did not matter, because a similar pitch delivered to someone else put them on a path to start a new organization together, "something much cooler than I could have managed on my own."

Seek real feedback

She says it is hard to overstate how overpowered this is, and that if you are not trying to get real feedback from people who know you, "you're cooking without tasting". It is the lowest hanging fruit for self-improvement and few people pick it. The mechanics matter: in many contexts the way to get good feedback is to let people give it anonymously, because anything else adds friction by layering on social dynamics. You want it comfortable to give and easy to find, so she keeps a link to her feedback form in her Twitter bio and gets a few comments a week through it.

She answers the obvious objection, that anonymity frees people to be unkind. In her experience they rarely are. Ninety percent of what arrives is either nonsense or nice, including plenty of "keep up the good work!" messages. In over a year she has had maybe two messages intended to hurt her feelings. Once in a while one lands as a gut punch because it points at a real issue, and then she gets over it and fixes the issue. She wishes people gave her more: she had to learn about her uptalk from YouTube comments.

Increase your surface area for luck

The last couple of times she was looking for a project, she met as many people doing related work as she could, even with no obvious benefit, initially just to advertise her existence in a new field, on the logic that someone is always hiring or looking for a cofounder. What she learned from casting a wide net is that she has very little ability to predict in advance how useful a call will be. Relevance is easier to predict, but it is a poor proxy for usefulness, which depends on other things including the other person's enthusiasm and the breadth of their interests. She goes further: to some extent, the more confident she is that a conversation is relevant, the less likely she is to discover something exciting in it. Nearly all of her most fruitful collaborations over the last three years came out of meetings booked almost at random, and her best conversation the week she wrote this was with someone whose introducer had said "this person asked for an introduction but I'm not sure it's a good use of your time."

Assume everything is learnable

Most subject matter is learnable, even the stuff that looks hard. Her stronger claim is that many, and she hedges this as "most?", of the traits people treat as fixed are malleable if you believe they are and put in the same kind of work you would for any other skill. Agency is her own example: she learned it late. In her teens and 20s she made occasional agent-like moves, including taking a job in a new city to be with someone she had not yet spoken to, and they married a few months later. Yet she still picked a career she really disliked, "for no reason other than its obviousness", and only after a decade stopped to ask what she was hoping to accomplish. Other things she says are learnable: confidence, charisma, warmth, tranquility, optimism. Asked how one learns charisma, her answer was deliberately boring, read a few books, watch many hours of charismatic people interacting with others, adopt a few of their habits. She notes most people could have produced that plan themselves if they were not carrying the assumption that charisma is innate.

Learn to love the moat of low status

The concept comes from her husband Sasha. Changing your life, especially learning a new skill set, means crossing a moat of low status, a stretch of time where you are genuinely bad at the thing or do not know things that are obvious to everyone else. It is a moat for two reasons: you cannot leap it, and anyone who can cross it gains a real advantage. You can cross it with your head down, by not asking questions and not collaborating, but she says those tradeoffs "really nerf learning", meaning they cripple it. "Learn by doing" is standard advice and you cannot do it without splashing around in the moat for a while.

Her example is a hand she played so badly in a big tournament that there were news stories about it. She says she will never entirely get over the

embarrassment, and that she still looks back on it fondly, because it is when she realised she had crossed some threshold of unflappability. With cameras and reporters crowding round she could have folded safely and nobody would have noticed; she called, knowing it would mean certain ridicule even if she won. The call was in fact quite bad, and she made it "for the right wrong reason."

Don't work too hard

She flags this as possibly the most important item, and says it took her almost forty years to learn, because her instinct was that more hours mean more output as long as you are genuinely trying, "that's just multiplication, right? No." Grinding can raise output temporarily while killing creativity and big picture thinking. Burnout is "the ultimate agency-killer", and she has learned to use a drop in agency as an early warning sign of burnout, one that shows up before she consciously notices anything is wrong: "A switch flips and I start looking for ways to rule out ideas and actions, to conclude they won't work or aren't necessary, rather than chasing better versions."

So she now keeps boundaries that would have embarrassed her earlier in life. She is offline at 6 p.m. almost every night and rigorously observes a Sunday Sabbath where "nothing with the flavor of effort is tolerated". She notes these will look trivial to some people and like a mortal sin to others in the communities she moves in. Her rule, and she says very few people take it seriously enough: never take instructions on how hard you should work from someone who has not burned out before.

The close

Agency, she says, is the skill that built the world around you, an all-purpose life intensifier that lets you make your corner of it more like what you want, professionally, relationally, aesthetically, whatever you care about. "Build a better mousetrap. Have an enviable marriage. Start a country." No one is born with it, everyone can learn it, and it is never too late. A 2026 update at the end notes that her book based on this post, *You Can Just Do Things*, is now on sale.

Words to know

- **Agency:** her definition is "manifest determination to make things happen", a disposition to act on what you want rather than wait to be directed. The unit's whole premise is that biosecurity work needs it because the jobs are often unstructured.
- **Real edge:** an advantage available to you because other people are unwilling to do something, not because you are smarter or working longer hours.

- **Solver model:** poker software that calculates mathematically optimal play. She uses it to show where everyone's effort was already crowded, which is why the edge lay elsewhere.
- **Out-of-distribution results:** outcomes well outside the normal range, in her case unusually good poker results. Statistical shorthand for "not explainable as ordinary variation".
- **Moat of low status:** the period of visibly being bad at a new thing that you have to cross to learn it. Her point is that the only fast way across involves being seen not knowing things.
- **Uptalk:** the habit of ending statements on a rising pitch so they sound like questions. She mentions it as something she only learned about herself from strangers in YouTube comments.

Numbers in this reading

- Number one female poker player in the world: her own ranking, given in the opening credentials.
- Fastest startup to take a drug to clinical trials: the record set by Alvea, the pandemic medicine company she co-founded and ran operations for.
- Eight years ago: when poker professionals were already studying nearly as much as they played, in her account of where the crowd's effort was going.
- A few comments a week: the volume of feedback arriving through the form linked in her Twitter bio.
- 90 percent: the share of her anonymous feedback inbox that is either nonsense or nice.
- Two messages in over a year: how many were intended to hurt her feelings, her evidence that anonymity does not invite abuse as much as people fear.
- Three years: the period over which nearly all her most fruitful collaborations came from meetings booked almost at random.
- Almost 40 years: how long it took her to learn not to work too hard.
- 6 p.m.: when she goes offline almost every night, alongside a Sunday with no effortful work in it.

What the author is unsure about

Everything here is personal experience, not evidence: she offers no study, no comparison group and no account of people who tried these habits and failed. She hedges her biggest claim, writing that "many (most?)" supposedly fixed traits are malleable. She will not give a concrete example of the poker reads that produced her edge, which means her central illustration cannot be checked. Her explanation for why other professionals dismissed reads is a guess, "maybe because they didn't want to feel obligated to study them", as is her reading of

the person who never replied to her email. The claim that less relevant conversations are more fruitful is qualified with "to some extent". And her conclusion about overwork is explicitly a rule of thumb about whose advice to trust, not a finding about how many hours are too many.

For the discussion

- Her "court rejection" test is a concrete standard for the unit's own exercise: a list of next steps that she is confident she can get is probably aimed too low.
- The tension worth raising is between her "increase your surface area for luck" and the day's dual-use advice: casting a wide net is good for finding collaborators, and the same openness is exactly what the dual-use reading says to manage carefully when the project itself is sensitive.
- Her burnout point is the one with real teeth for anyone doing this alongside a job: she treats a sudden urge to rule ideas out as the first symptom, which is a more useful early warning than exhaustion.

6.3-1 • CORE • COURSE TIME 5 MIN

Navigating dual-use projects in biosecurity

Will Saunter • 2025

[Read the original paper →](#) • the course assigns: whole piece

Some of the most useful biosecurity projects could also help an attacker, so before you build one, get experienced mentors, do not decide alone, think about where the work belongs institutionally, and plan what happens to your outputs.

Average summary

Saunter opens with the uncomfortable fact underneath the whole unit: a project designed to reduce biological risk can itself be misused, by creating tools, datasets or capabilities that bad actors could exploit. He gives four examples, each with its real upside and its real downside. **Benchmarks** that test whether AI models can help with dangerous biology let developers find and constrain those capabilities before deployment and let policymakers see the risk, and the benchmark code or the collated data could also help a malicious actor work out which model serves them best, or map out how to extract harmful outputs. **AI agents with access to biological tools** could speed up defensive research, drug discovery and outbreak response, and the same capabilities could help a bad actor get through complex procedures or find weak points in biosecurity infrastructure. **Red-teaming** reveals weaknesses in time to fix them and builds the case for policy, and findings about what is broken, shared too widely, are a

list of exploitable gaps. **Threat modelling** that produces detailed attack scenarios helps defenders prioritise, and those same scenarios, if they escape a controlled setting, read as instruction manuals.

His verdict on all four is deliberately unsettled. Whether any particular project does more good than harm "depends heavily on context: who's doing it, where, with what safeguards, and how the outputs are managed." He says there is genuine uncertainty and reasonable people disagree. The conclusion is not that these projects should never happen, but that they need more care than work with no dual-use dimension.

Then the frame that does the most work: **dual-use is a spectrum, not a binary**. Some work should be done openly, and he names low-risk contributions to detection, broad-spectrum countermeasures and policy analysis. Some should be done carefully, in controlled settings with experienced collaborators, perhaps with input from government stakeholders. Some needs serious information security, possibly even classification, with the institutional backing that implies. And he is blunt that **your good intentions don't eliminate the risk**: someone building a benchmark to prove that a dangerous capability should be restricted is still producing something a bad actor could use.

His four practical steps are: get experienced mentors, which he calls the single most important step; don't do it alone, because of the unilateralist's curse; consider the institutional setting, since some work genuinely belongs in organisations with government relationships, clearances and established protocols; and think about outputs before you create them, asking who would benefit if this were released publicly.

He adds a point that is easy to skip and matters just as much: **defensive value doesn't happen automatically**. If you build something for biosecurity professionals you have to get it into their hands, test whether it is actually useful to them before you start, tailor it to what they say they need, and build the relationships that make sure it reaches them. "A dual-use tool that only bad actors end up using is worse than no tool at all."

Why it's assigned

This is the course's own standard for deciding what to build, placed immediately before it hands her a hundred project ideas, five scoped projects and a grant application. Unit 6 is not a defence layer, it is the unit where she picks one intervention and commits to a project, so the course puts the misuse question first rather than last. She is certified in Frontier AI Governance and AGI Strategy at BlueDot Impact (October 2026), and is taking this Biosecurity course now, so

the "should be done openly" band, which explicitly includes policy analysis, is the part of his spectrum closest to her existing ground.

In depth

The problem: defensive work that hands over capability

The opening claim is narrow and precise. It is not that biosecurity research is dangerous in general. It is that some projects designed to reduce biological risk involve "creating tools, datasets, or capabilities that bad actors could exploit". The risk is in the artefact you produce, not in your motive.

The four examples, with both sides stated

He gives each example its honest defensive case before its misuse case, which is what makes the piece usable rather than discouraging.

Benchmarks to test whether AI models can assist with dangerous biology.

Defensive value: they help AI developers identify and constrain dangerous capabilities before deployment, and inform policymakers about emerging risks.

Misuse: the benchmark code or the collated data could help malicious actors test which models are most useful for their purposes, or provide a roadmap for extracting harmful outputs. Note the two distinct harms, a shopping guide for picking a model, and a method for getting it to comply.

Developing AI agents with access to biological tools. Defensive value: such agents could accelerate defensive research, drug discovery or outbreak response. Misuse: the same capabilities could be repurposed to help bad actors navigate complex biological procedures, or identify vulnerabilities in biosecurity infrastructure. The second half is worth noticing, because it is about attacking the defences rather than making a pathogen.

Red-teaming exercises that probe biosecurity vulnerabilities. Defensive value: they can reveal weaknesses in time to fix them, and build the case for policy action. Misuse: detailed findings about what is broken, if shared too broadly, "hand attackers a list of exploitable gaps". The harm here is specifically about distribution, not about doing the work.

Threat modelling that produces detailed attack scenarios. Defensive value: it helps defenders prioritise resources and anticipate adversary behaviour. Misuse: concrete scenarios, if they escape controlled settings, "can serve as instruction manuals".

Context decides, and the uncertainty is real

He refuses a clean verdict on any of the four. Whether a project does more good than harm "depends heavily on context: who's doing it, where, with what safeguards, and how the outputs are managed." He states plainly that there is genuine uncertainty and that reasonable people disagree. His conclusion is a procedural one: not that these projects should never happen, but that they "require more care than work without dual-use dimensions."

The spectrum: three bands, not a yes or no

This is the structural idea of the piece. Not all dual-use work is equal.

1. **Some should be done openly.** His examples are low-risk contributions to detection, broad-spectrum countermeasures, or policy analysis. Openness is the default where the work does not hand anyone a capability.
2. **Some should be done carefully.** In controlled settings, with experienced collaborators, and perhaps with government stakeholder input.
3. **Some requires serious information security.** Maybe even classification, with appropriate institutional backing, which in practice means it is not a solo portfolio project at all.

Placing a project on that spectrum is the decision he wants people to make consciously, rather than deciding only whether the project is good or bad.

Good intentions are not a safeguard

He makes this its own point, and it is the sentence to remember: "your good intentions don't eliminate the risk." His worked case is the benchmark builder whose explicit purpose is to demonstrate that a dangerous capability should be restricted. That person is still producing an artefact that could be useful to bad actors. The intention shapes what you do with the artefact; it does not change what the artefact is.

What to do: four steps

1. **Get experienced mentors.** He calls this "the single most important step". Three reasons: senior people in biosecurity have developed intuitions about what is safe and what isn't; they have seen projects go wrong; and they have relationships with funders and policymakers who can help navigate tricky situations. The third reason is about having somewhere to take a problem once you are in one.
2. **Don't do it alone.** He invokes the unilateralist's curse: if twenty-four people independently decide a risky project is not worth pursuing but the twenty-fifth goes ahead anyway, the damage is done. The lesson he draws is to check your reasoning with others before acting. The structure of the problem is that

the field's caution is only as strong as its least cautious member, so individual judgement, however careful, is not enough.

3. **Consider the institutional setting.** Some work genuinely belongs in organisations with government relationships, security clearances and established protocols for handling sensitive material. The implication for a course project is that the right answer is sometimes "not by me, not yet, and not here".
4. **Think about outputs.** Before you create a dataset, tool or benchmark, ask: if this were publicly released, who would benefit? If the answer includes potential bad actors, "you need a plan for controlled access, or you need to reconsider whether this project should exist at all." This is the test he wants applied before the work starts rather than at publication.

The other half: defensive value has to be delivered

He then turns the argument around, and this part is as practical as the caution. "Defensive value doesn't happen automatically." If you are building something intended to help biosecurity professionals, you have to actively get it into their hands. Do not assume the right people will find your tool or analysis and use it. He gives three specific instructions: test whether it is actually useful to them before you start, tailor it to their expressed needs, and build relationships that make sure it reaches them. The reason this sits in a dual-use piece rather than a product-management one is the sentence that closes it: "A dual-use tool that only bad actors end up using is worse than no tool at all." A project can fail the cost-benefit test not because its misuse risk rose but because its defensive benefit never arrived.

When in doubt

His last instruction treats your own uncertainty as information: "If you're unsure whether a project idea has dual-use concerns, that uncertainty is itself informative." The action is to reach out, to course facilitators or to experienced biosecurity professionals who can help navigate the question. He closes on balance rather than discouragement: the field needs capable, ambitious people, capability and ambition should be paired with caution and collaboration, and the goal is not to put people off hard problems but to make sure that when they take them on, the work makes things better rather than worse.

Words to know

- **Dual-use:** work that can serve both defence and attack. The whole piece is about the fact that this is a matter of degree, so a project sits somewhere on a spectrum rather than being safe or unsafe.

- **Benchmark** (for AI models): a standard set of tests used to measure what a model can do. Here it is the leading example of a defensive artefact that doubles as a guide for someone shopping for a capable model.
- **Red-teaming**: deliberately attacking your own system or plan to find its weaknesses before an adversary does. He points out the findings are a repair list and a target list at the same time.
- **Threat modelling**: working out systematically who might attack, how, and by what route, so defenders can prioritise. His warning is about the level of detail in the written scenarios.
- **Unilateralist's curse**: the problem that when many people can each act alone, the decision effectively belongs to the least cautious of them. His illustration is twenty-four people declining a risky project and the twenty-fifth proceeding.
- **Broad-spectrum countermeasures**: defences that work against a wide range of pathogens rather than one specific threat. He lists them among the contributions that can be made openly.
- **Controlled access**: releasing something only to vetted people rather than publicly. He names it as the alternative to not building the thing at all.
- **Classification**: formally restricting information under government security rules, which requires institutional backing and is well beyond the scope of an individual project.

Numbers in this reading

- Four example project types, each given a defensive case and a misuse case: benchmarks, AI agents with biological tools, red-teaming, and threat modelling.
- Three bands on the dual-use spectrum: do openly, do carefully, needs serious information security.
- Twenty-four and the twenty-fifth: his illustration of the unilateralist's curse, where twenty-four people independently decline a risky project and the one who proceeds decides the outcome.
- Four practical steps: mentors, do not act alone, institutional setting, think about outputs.

What the author is unsure about

He is explicit that the uncertainty is genuine and that reasonable people disagree about whether particular dual-use projects do more good than harm. He gives no threshold, no scoring rubric and no worked example of where a specific project lands on his three-band spectrum; placing a project is left to judgement, which is exactly why his first instruction is to borrow someone else's.

He does not say what counts as "experienced" in a mentor, nor how to find one. He does not claim his four examples are the full set of dual-use project shapes. And he treats his own advice as incomplete at the boundary: if you are unsure, the answer is to ask a person, not to apply a rule from the post.

For the discussion

- His strongest and least comfortable line is that good intentions do not change what an artefact is, which cuts against the instinct that a project is fine as long as its purpose is defensive.
- The tension worth raising with the group is between this reading and the two agency pieces from the same morning: act before you are certain, except where acting alone is itself the risk, and nobody gets a rule for telling those cases apart.
- A fair question for the facilitator: what does the "controlled access" plan actually look like for an individual with no institution behind them, since that is the only option he offers short of abandoning the project.

6.3-2 · CORE

A hundred project ideas in biosecurity

Afiq Hatta and Anu Bazarragchaa · 2025

[Read the original paper →](#) · the course assigns: whole page, all 15 sections

A working list of biosecurity projects, written by a part-time researcher with no funding and no organisational affiliation, for people in exactly that position, and almost none of it needs a laboratory.

Average summary

This is a list, not an argument. Hatta wrote it, Bazarragchaa contributed most of the forecasting section, and the saved page carries **95 top-level ideas with 34 sub-points** under them, across **15 sections**. The framing matters more than any single idea: sections and sublists are ordered "in rough priority from Afiq's perspective", all ideas are free to take, and he says the priorities are ordered towards his own situation as **"a part time researcher with no funding and no organisational affiliation"**, so "don't take the priority list too seriously!"

Grouped by the kind of work rather than by his topics:

- **Build and measure, indoor air.** His own top area, because it "lends itself to individual work, it's scalable, and already adoptable to the public via the commercial air purifier channel". Examples: pressurising a house to

cleanroom standards, which is his active project with Charles Victorio; a tool that measures indoor air quality before you move into a house; a quick, versatile window seal with a hole for a fan that could be retrofitted to London houses.

- **Modelling and simulation tools.** Examples: photograph your house and get its air changes per hour; quick CFD for a given room; better tools to simulate far UVC in ventilation chambers so published papers can be replicated.
- **Reviews, deep dives and buyer's guides.** Examples: an "everything you need to know" document on ventilation and air filtration covering efficacy, logistics, how it works, cost and scalability; a review of elastomeric respirators; clarifying which particle counters to buy and why they are so expensive.
- **Supply chain, market and policy analysis.** Examples: why the commercial market for air filters is so bad; analysing the far UVC supply chain; a review of more secure frameworks for people ordering DNA, "KYC and the like".
- **Forecasting and epidemiology.** Bazarragchaa's section. Examples: a website showcasing historical forecasting model failures; which epidemiological R packages should be ported to Python; a GIS dataset merged with virology datasets to track climate change and disease.
- **Datasets and data plumbing.** Examples: automated tools to build data APIs, because "building a data pipeline is annoying and hard!"; an API aggregating the cost of every biosecurity measure, and of every air filter.
- **Teaching, writing and field-building.** Examples: a wiki for everything biosecurity; physics primers and example sheets as force multipliers, sendable to university students; engaging YouTubers on indoor air quality.
- **Info hazards, history and public trust.** Examples: improving the existing info hazard guidance; a history of every biosecurity policy that backfired; a history of how and why people came to distrust measures like vaccines.

No wet lab needed for any of the eight groups above. The exceptions are the antibacterial-surfaces and vapours sections, which are bench science, plus UVC surface degradation testing and replicating far-UVC chamber results. Respirator fit testing and 3D-printed face adapters need a workshop and test kit, not a wet lab, and the computational chemistry and protein folding questions need compute.

Why it's assigned

The 6.2 exercise asks her to pick one intervention and then find concrete projects she could pick up, and this is the course's widest menu for that last part. It sits alongside the dual-use reading deliberately: this list is almost entirely open, publishable work, the band Saunter says can be done openly. It

also answers the unit's warning that this will not be a normal job, by showing what a person with no lab, no funder and no affiliation can still do across prevent, detect and withstand.

In depth

How the list is framed

The page is explicitly provisional, filed under "Essays in progress", and the saved copy carries its own metadata: last updated 2025-07-09, with 6270 seconds of recorded editing. Hatta takes responsibility for all the ideas and invites people to take any of them, offering help via his Discord. The ordering caveat is repeated twice, once about sections and sublists being in rough priority from his perspective, once to say the priorities suit his situation as a part-time researcher with no funding and no affiliation, so they should not be taken seriously as a ranking for anyone else.

Ventilation and air filtration

The longest section and his stated favourite for part-timers, with the reasoning given up front: individual work, scalable, and already reaching the public through commercial air purifiers. He also names his own open question, whether DIY or commercialisation strategy is the more valuable thing to work on, and does not resolve it.

The ideas run from hardware to documentation. Pressurising a house to cleanroom standards is his active project with Charles Victorio, and he notes there are known examples of this working in areas with wildfires. Then: engaging YouTubers on indoor air quality, where he gives his own reach as 9k subscribers on his channel and 1k on Substack and says he is still working out how to communicate these ideas to diverse audiences; a tool to measure indoor air quality metrics before moving into a house; reviews or aggregators of indoor air quality products, where he wants a clean dataset because "it's still fuzzy on the right stuff to get"; a quick versatile window seal with a hole for a fan, retrofittable to London houses; clarifying cheap measurement tools, including which particle counters to buy, why they are so expensive, and what the metrics even are; an "everything you need to know" document on ventilation and air filtration; easier ventilation modelling, with his example of photographing a house and its windows and getting air changes per hour out, noting that Martin Bazant has a covid infection risk calculator; tools that say how strong the fans in a room need to be; figuring out why the commercial air filter market is so bad; the logistics behind air filters; comparing real ventilation measurements against values calculated from temperature and wind speed; the effect of building materials on ventilation efficiency; natural ventilation strategies across climate zones; smart

ventilation that responds to occupancy and air quality; and what London councils currently offer on indoor air quality. The section closes with a DIY block: guides to modifying a house for sufficient ventilation, making fans, making air filters, working out ventilation rates, a biosecurity-focused HVAC measurement tutorial, whether cheap CO2 monitors work as well as expensive ones, and modifying or building cheap PAPRs following a YouTube video.

Far UVC

Opens with an instruction, "Read Blueprint Biosecurity's blueprint!", and the admission that he has been going back and forth on how he could best contribute here. Ideas: tools to make far UVC easier to install in the real world, including rapid room measurement and optimised lamp placement, and testing how well phone LIDAR apps perform; better pedagogy and explanations of how far UVC works, perhaps as a blog; better tools to simulate far UVC in ventilation chambers so papers can be replicated; quick CFD for a room; what a heuristics document distributed widely in a catastrophic scenario might look like; an oddity he thinks is informative, that people believe the blue light in some hand dryers is bactericidal and seem fine with it, and whether that says anything about UVC adoption; whether people would trust far UVC if it were available; analysing the far UVC supply chain, and the food sterilisation supply chain; a computational chemistry investigation into the wavelengths of light that break peptide bonds; smartphone apps to measure far UVC intensity in real time; UVC's effect on different surface materials and their degradation; and automated UVC scheduling systems.

Stockpiling estimation

Three questions, all about what a government would need in a crisis: how much square footage all the public spaces in London take up, what data would be immediately meaningful to government in a crisis scenario, and what data is required to estimate critical workers in a crisis.

AI use, and drug discovery

Two short sections of open questions. Whether AI will get to the point of helping people build protection into their homes without much effort, and how AI could scan built environment literature. Then: is AI actually accelerating drug discovery; what are the real bottlenecks for the drug discovery industry, regulatory factors included; how fast in principle could we make a drug; and do we actually need machine learning for protein folding, or are more and faster computers enough.

Elastomeric respirators

A full section of writing and testing work on reusable respirators: his own experience buying one; whether a PAPR or elastomeric respirator could be made from things in the house; a document on how to fit test one; a review of whether they work, covering efficacy, safety, supply chain and scalability, plus how to communicate wearing them properly; where they might go wrong; how they would be distributed in an emergency; whether an approved standard would work; 3D-printed adapters for different face shapes; long-term wear effects across demographic groups; cleaning and maintenance protocols; and child-sized versions.

DNA

Two lines, and one is a confession: "AI x bio - not sure how I'd really help there considering how tricky AI safety already is". The other is a review of more secure frameworks for people ordering DNA, know-your-customer and similar, which is the thread picked up in detail by the five tractable projects reading.

Anti-bacterial surfaces, and vapours

The two bench-science sections. Surfaces: a more comprehensive review of antibacterial surface research, with safety concerns about good bacteria and the gut microbiome; natural antimicrobial materials from plants and animals; self-cleaning surface technologies; testing protocols; and whether treated surfaces drive antimicrobial resistance. Vapours: a deep dive on glycol vapours covering safety, efficacy and scalability; essential oil vapours as antimicrobial agents; vaporised hydrogen peroxide systems; safe vapour concentration monitoring; and vapour persistence in different environments.

Forecasting

Bazarraghaa's main contribution, and the most tool-shaped section. A website showcasing historical forecasting model failures, paired with working out what is actually "hard" for people doing mathematical disease forecasting; which epidemiological R packages should be migrated to Python; better tooling and pedagogy for the mathematical nuances of epidemiology and of forecasting, which appears twice; a GIS dataset merged with virology datasets to track climate change and disease; tools useful to biosecurity forecasters; better documentation on forecasting well and avoiding standard errors; early warning systems for emerging pathogens; the predictive value of historical pandemic patterns; climate change impacts on disease spread; and risk assessment tools for different biosecurity measures.

Datasets

Automated tools to build data APIs more easily, with the complaint that building a pipeline is annoying and hard and there should be a way to automate it. Then an API for biosecurity-relevant datasets, with his wish list: all the z values for far UVC aggregated, the costs of every biosecurity measure, the cost of every air filter, a usable representation of the UK infrastructure database, and a compilation of historical biosecurity policy outcomes.

Pedagogy

The field-building section: a wiki for everything biosecurity; biosecurity book reviews and recommendations; primers on the physics topics that apply to biosecurity, where he argues example sheets and primers "could act as force multipliers", sendable to university students to show them there are interesting technical problems here, and also to the built environment world; curriculum for different education levels; interactive training modules; and case study collections.

Info hazards

Three items, and they are the ones the dual-use reading speaks to directly: improving and developing the info hazard guidance that currently exists, a review of the current papers on info hazards, and a question he leaves open, whether it is safe to write things like "this PPE mechanism might go wrong because...".

History of biosecurity, and infection rates

History: biosecurity measures before the year 1000; a history of every biosecurity policy that backfired; criticisms of current government biosecurity policies; a history of how and why people distrust pandemic prevention measures like vaccines; ancient quarantine practices; historical biosecurity failures and their lessons; a policy evolution timeline; and cultural differences in biosecurity practices. One bullet in this section is cut off mid-sentence in both the saved text and the live page: "How can we translate the nuclear". The last section asks how to make the physics of infection clearer, and whether it should be pushed into standard high school or university physics and biology curricula.

Acknowledgements

Hatta wrote the document; Bazarragchaa provided major contributions to the forecasting section and they are developing it further. Thanks go to Jacob Swett, Andrew Snyder-Beattie, Vivian Belenky, Sofya Lebedeva, Damon Binder and James Wagstaff for discussions.

Words to know

- **Far UVC:** a short-wavelength ultraviolet light that appears to inactivate airborne pathogens while being safer for people than conventional germicidal UV. A whole section is about making it installable and measurable.
- **Elastomeric respirator:** a reusable rubber-sealed mask with replaceable filters, as opposed to a disposable N95. **PAPR** is a powered air-purifying respirator, which blows filtered air into a hood or mask.
- **Fit testing:** checking that a respirator actually seals on a particular person's face, since the seal is what the protection depends on.
- **Air changes per hour:** how many times the full volume of air in a room is replaced each hour, the standard measure of ventilation.
- **CFD:** computational fluid dynamics, simulating how air moves through a space. He wants it fast enough to run on an ordinary room.
- **Z value** (for far UVC): the dose-response constant that says how much a given amount of UV light reduces a pathogen's viability. He wants these aggregated because they are scattered across papers.
- **KYC:** know your customer, the identity and legitimacy checks banks run, here proposed for people ordering synthetic DNA.
- **Info hazard:** true information that causes harm by being known or shared. The reason one of his sections is about what it is safe to write down.
- **Force multiplier:** something that makes other people's work more effective. His argument for why writing primers counts as a real contribution.

Numbers in this reading

- 95 top-level ideas and 34 sub-points across 15 sections: a count of the saved page, against a title that promises a hundred.
- 9k subscribers on his YouTube channel and 1k on Substack: his own reach, given as context for the idea of engaging YouTubers on indoor air quality.
- Last updated 2025-07-09, with 6270 seconds of recorded editing: the page's own metadata in the saved copy.
- Before the year 1000: the cutoff for his proposed study of early biosecurity measures.

What the author is unsure about

The caveats are the most useful part of the page. The priority ordering is his, shaped by his own circumstances, and he says not to take it seriously. He has not resolved whether DIY or commercialisation is the better route for indoor air quality. He says he has been going back and forth on how best to contribute to far UVC. On AI and biology he writes that he is not sure how he would really help, given how hard AI safety already is. He does not know whether an approved standard for elastomeric respirators would work, whether people would trust far

UVC if it were available, or whether it is safe to write publicly about how a PPE mechanism might fail. Several entries are questions rather than projects, and one bullet is unfinished. Nothing here is a finding; it is a list of things somebody should check.

For the discussion

- The most striking thing about the list is what is absent: almost none of it needs a laboratory, and the groups that do, antibacterial surfaces and antimicrobial vapours, are the ones furthest from the course's own interventions.
- His info hazards section is a real test case for the dual-use reading: a review of what is safe to write about PPE failure modes is itself a document about failure modes.

6.3-3 · CORE

Five Tractable Biosecurity Projects You Could Start Tomorrow

Max Langenkamp and Tessa Alexanian · 2025

[Read the original paper →](#) · the course assigns: whole post, all five projects

Two people who work on DNA synthesis screening scope out five projects they wish someone would take on, each small enough to finish in about two months and each buildable without deep background in the field.

Average summary

Langenkamp writes in the first person; the post comes from a conversation between him and Alexanian, "both biosecurity researchers", about projects in DNA screening they wished someone would do. He works at SecureDNA and she works at IBBIS, both organisations in DNA synthesis screening. They chose five projects they believe are **modular and high impact**, specifically so that people who want to contribute to biosecurity "could take up the attempt without requiring heavy context". They say this list differs from Alexanian's past list and from ASB's list by being much smaller in scope: **most of these could be finished within two months**. Either of them may also be working on these during the year depending on time and resources, so they ask interested people to email with their name, background and interest, and say they will get back shortly. If enough people are interested they may run a small group call and pair people up.

As of December 2025 the course says two of the five are done. **Project 1, the ORCID-TAXID Mapping Tool, and project 3, the Biosecurity Statements Repository, have been completed. Projects 2, 4 and 5 are open.**

1. ORCID-TAXID Mapping Tool (done). A tool that, given a researcher's ORCID identifier, returns the taxonomic IDs of the organisms that researcher has worked with. The problem it solves: when a synthesis company receives an order for variola from an unknown name, working out whether that person is "an angry physicist, grad student, or a well regarded vaccine researcher" takes a lot of manual labour, and that friction is often prohibitive, so screening becomes highly ad hoc. An advanced version would also show collaborators and their organisms.

2. Customer Screening Training Dataset (open). Case studies to train synthesis providers to spot bad actors and to make policymaker discussion more concrete. Alexanian has already built 7-8 profiles and found them useful for both companies and policymakers. The project is to expand those to **25-40 examples**, including both legitimate and concerning cases, and to document historical cases of people who ordered dangerous biological materials.

3. Biosecurity Statements Repository (done). A collection of the biosecurity statements and practices published by makers of biological design tools and AI models, because those organisations often care about dual use and do not know what to do. Gathering what exists starts to establish norms and gives new developers templates.

4. Vulnerability Disclosure Framework (open). A process for reporting security flaws in biological design tools, especially where a new tool could break existing screening. There is currently no standard for responsible disclosure here. The authors admit this one "is less well scoped".

5. International Suspicious Activity Reporting Analysis (open). A map of how biosecurity-relevant suspicious activity reports are handled across countries: who reports what to whom, where the data is stored, who can see it, and how sensitive business information is protected.

Why it's assigned

Unit 6 asks her to pick one intervention and find a concrete project in it, and these five are the most precisely scoped projects the course offers, from people inside the organisations that would use the results. They all sit in the prevent layer, at the DNA synthesis screening chokepoint the course covered in Unit 2. Three of the five are pure desk work on norms, disclosure and reporting chains, which is the open end of the dual-use spectrum from the previous reading.

In depth

Why they wrote it

The post opens with a conversation: Alexanian and Langenkamp talking about DNA screening projects they wish someone would take on. They point to an essay in Asimov Press as a primer on why DNA screening matters. The design criterion for the list is accessibility: modular, high impact, and takeable "without requiring heavy context", which they contrast with Alexanian's past list and with ASB's list, both larger in scope. The practical offer at both the top and the bottom of the post is the same: email them with your name, background and interest, and they will reply. They are honest that they may be working on these themselves this year depending on time and resources.

1. ORCID-TAXID Mapping Tool

Purpose. Map researchers' ORCID identifiers to the taxonomic IDs of the organisms they have studied, so DNA synthesis requests can be screened more efficiently. A more advanced version would also let you look at a researcher's collaborators and the organisms those collaborators have worked with.

Motivation. The example is concrete. A DNA synthesis company receives an order for variola from John Doe. To decide whether this is legitimate, someone has to do a lot of manual work to establish whether he is "an angry physicist, grad student, or a well regarded vaccine researcher". That friction "in many cases is prohibitive", and the result is that much screening becomes highly ad hoc, meaning case by case and inconsistent. The project is to build something straightforward: give it an ORCID ID, get back the organisms that researcher has worked with.

Details and impact. It maps scientific identifiers to organism taxonomy IDs; it lets companies quickly verify a researcher's credentials and experience; it could examine collaborator networks; and it would help prevent **false positives**, their example being legitimate vaccine researchers getting blocked. The claimed impact is indirect but specific: it would meaningfully improve one of the key frictions in screening, and so increase the field's ability to catch bad actors trying to synthesise sequences of concern. They add that several biosecurity organisations have expressed interest in implementing such a tool, so it would be concretely helpful immediately. The course notes this one has since been done.

2. Customer Screening Training Dataset

Purpose. Develop comprehensive case studies that train DNA synthesis providers to identify bad actors, and add nuance to policymaker discussion.

Motivation. Screening customers is really challenging. Policymakers often underestimate how hard it is, and synthesis companies do not know exactly what to do. Alexanian has already created 7-8 profiles and found them very helpful for both screening companies and policymakers, because, in the authors' words, "it forces people to get much more concrete about the likely profile of a bad actor." That is the real argument for this project: it converts an abstract policy debate into specific people with specific stories.

Details and applications. Expand the existing 7-8 profiles to 25-40 examples; include both legitimate and concerning cases, not only threats; document historical cases of individuals who ordered dangerous biological materials; and help establish more concrete screening policies. The applications they name are training for synthesis organisation teams, policy development guidance, and improved detection of concerning orders. This one is open.

3. Biosecurity Statements Repository

Purpose. Create a repository of the biosecurity statements and practices of biological design tools and AI models.

Motivation. Organisations building protein folding models, inverse folding models and other biological design tools "often seriously care about the dual use potential of their work, but don't know what to do." Collecting what has already been said starts to build a set of norms that toolmakers can draw on.

Details and impact. Document approaches from protein folding tools, organ building technologies, CRISPR tools and AI biological design models, and gather the tool statements themselves, with five named: the Urbina et al toxins paper, Coscientist, Open CRISPR, vPST and AlphaFold 3. The intended effects are coherent norms for mitigating biological design tool risk, templates for new developers, and standardisation of biosecurity practice. The course notes this is the "risk statement repository" that has since been done.

4. Vulnerability Disclosure Framework

Purpose. Establish a clear process for reporting and responding to security vulnerabilities in biological design tools, especially when a new tool could break existing biosecurity measures.

Motivation. A year before this post, the two authors argued that vulnerability reporting was a particularly helpful option for reducing biological design tool risk, and they quote themselves: "Developers could create (and regulators could require) mechanisms for software users to report concerning model behavior or other security flaws to their developers. This can help developers to quickly fix vulnerabilities, or, in extreme cases, rapidly roll back or withdraw a model. As an

incentive to report potential risks, 'bug bounties' could provide financial rewards to the reporter." Despite that, there is still no standard for responsible disclosure when a new tool like AlphaMissense or AlphaFold could circumvent existing approaches to DNA screening.

Shape of the work. They say plainly that this project "is less well scoped". It would involve gathering cases where this has happened and comparing practice from dual-use research in biology against cybersecurity disclosure for computer systems, including the question of where medical device vulnerability disclosure sits. The aim is a public resource that developers can use to disclose risk and that policymakers can use to write clear guidelines.

Details. Gather and analyse examples from the 2001 mousepox gain-of-function case, which they gloss as IL-4 enabled vaccine escape; medical device vulnerability disclosures; cybersecurity responsible disclosure processes; and existing voluntary bio research disclosure norms. (The mousepox case is a 2001 experiment whose published result was that a modified mouse poxvirus got past vaccination in mice, and it became a standard reference point in arguments about whether such results should be published: not from this reading.) The framework should address when to disclose versus withhold, who receives the disclosure, naming regulators, affected companies and the public as options, timeline expectations for disclosure and mitigation, the potential role of bug bounties, and protection for good-faith reporters. Impact: faster response to new vulnerabilities, clear guidance for developers, incentives to disclose responsibly, and coherent disclosure requirements for policymakers. This one is open.

5. International Suspicious Activity Reporting Analysis

Purpose. A comprehensive map of how biosecurity-relevant suspicious activity reports are handled internationally, focused on data storage, reporting requirements and privacy protection.

Motivation. The biosecurity community lacks clarity on reporting procedures for suspicious activity across borders. They draw the parallel to finance, where similar challenges were addressed through mechanisms like the Basel Accords, and say biosecurity needs its own framework rather than a borrowed one.

Details. Research and document the current reporting chains for suspicious biological activities; where the data is stored and who can access it; the privacy protections for sensitive business information; jurisdictional responsibilities and overlaps; a comparison with financial sector solutions; and the gaps in current reporting infrastructure. They link a post as a starting point.

Impact. More effective cross-border monitoring; protection of legitimate business interests while maintaining security; input to international biosecurity

standards; and support for coherent reporting requirements. They close the section by saying several biosecurity organisations need this analysis to develop reporting frameworks that balance security with business operations. Open.

What they ask at the end

The same invitation as the opening: email with name, background and interest. Given enough interest, they may organise a small group call to answer what questions they can and potentially pair people together.

Words to know

- **DNA synthesis screening:** checking who is ordering synthetic DNA and what sequence they want, before it is made and shipped. The chokepoint all five projects sit at.
- **ORCID:** a permanent personal identifier for researchers, used to attach publications to the right person. The input to project 1.
- **TAXID:** a taxonomy identifier, a stable number for a species or strain in standard biological databases. The output of project 1.
- **Variola:** the virus that causes smallpox, used here as the example of an order that should trigger scrutiny.
- **False positive:** in screening, blocking or flagging a legitimate order. They name it as a real cost of bad screening, not just an inconvenience.
- **Biological design tool (BDT):** software that designs biological things, for example protein folding and inverse folding models, as distinct from a general chatbot. Projects 3 and 4 are about the norms around these.
- **Responsible disclosure:** the cybersecurity practice of telling the maker about a flaw, on an agreed timeline, before telling the world. Project 4 asks what the biological equivalent should be.
- **Bug bounty:** paying people for reporting flaws. Proposed as the incentive that makes disclosure happen.
- **Suspicious activity report:** a formal report a regulated business files when a customer's behaviour looks illegitimate, a term borrowed from finance. Project 5 maps the biological version.
- **Basel Accords:** international banking agreements, used here only as the analogy for how a cross-border reporting framework got built in another sector.

Numbers in this reading

- Five projects: the whole structure of the post.
- Two months: the time in which they say most of these could be finished, the feature that distinguishes this list from larger ones.

- 7-8 profiles: how many customer screening case studies Alexanian had already written.
- 25-40 examples: the target size of the expanded dataset in project 2.
- Five named tool statements for project 3: the Urbina et al toxins paper, Coscientist, Open CRISPR, vPST and AlphaFold 3.
- 2001: the mousepox gain-of-function case they want used as a disclosure precedent.
- A year ago: when the two authors first argued publicly for vulnerability reporting as a way to reduce biological design tool risk.
- Two of the five done as of December 2025, per the course page: the ORCID-TAXID tool and the risk statement repository.

What the author is unsure about

They are clearest about project 4: it "is less well scoped" than the others, and they do not say what the finished framework should conclude, only what it should address. On project 1 they claim indirect impact rather than direct, that it would improve a friction and so increase the chance of catching bad actors, and they do not quantify that. They say several organisations have expressed interest in the tool and would find it helpful immediately, without naming them. They do not say what makes a screening case study good, only that Alexanian's existing profiles proved useful. And they flag the main practical uncertainty for anyone picking one up: they may be working on these themselves this year, depending on time and resources.

For the discussion

- The sharpest line in the post is the screening problem in miniature: a company looking at an order for variola from a stranger has to decide, by hand, whether that person is a vaccine researcher or someone dangerous, and that cost is why much screening ends up ad hoc.
- Worth raising: three of the five projects are really about norms and paperwork, disclosure processes, reporting chains, published statements, which suggests the binding constraint in this intervention is coordination rather than technology.
- A tension against the dual-use reading: project 2 asks someone to write 25-40 concrete profiles of what a bad actor looks like, which is a training set for screeners and a description of how to not look like one.

6.3-4 • CORE

Rapid Small Grants for Course Participants

BlueDot will reimburse small, specific costs that are blocking a project she has already started, with an application that takes under fifteen minutes and an answer within five working days.

Average summary

This is the money page, and the details matter more than the argument. After the learning phase of a BlueDot course, many participants work on independent projects, and BlueDot wants those projects to be excellent, so it tries to remove barriers. One common barrier is lacking the resources to do the project well, so it runs a rapid small grants programme for participants and facilitators, "aimed at people who, without funding, couldn't reasonably afford what they need for their chosen project." Landes's advice on whether to apply is one line: **"If you're unsure whether to apply: apply."**

The amounts. The post says BlueDot expects **most grants to fall between \$50 and \$1,500**, and that amounts toward the higher end typically require evidence of initial traction, a strong proof of concept or promising preliminary results. The course page's own note on this reading gives the range as **\$50 to \$1,000**, so the two numbers differ and the post is the more recent figure. For larger grants, **\$10-50k**, you have to apply to and complete one of BlueDot's Incubator Weeks.

The timings. Submitting a proposal takes **under 15 minutes for most applications**. BlueDot typically responds **within 5 working days** with one of three outcomes: accepted, with confirmation of exactly what spending is approved; clarification needed, meaning follow-up questions; or not approved. Reimbursement claims are processed within **5 working days** too. If you did not end up spending the money, you simply do not submit a claim.

The one hard rule. "At all times, only assume we'll cover costs we've confirmed in writing for your specific project." If you are uncertain, contact them before spending money you expect back.

What it funds. Compute for a technical AI safety project, including API costs, cloud GPU and training runs. Access to paywalled resources: articles, research papers, datasets or textbooks. Conference travel, but with a high bar. Project-specific software tools that save significant time. Project-specific equipment, their example being a quality microphone for an AI safety YouTube channel. Participant recruitment for empirical experiments. Hosting costs for an application or tool. All of these are illustrative and all decisions are at BlueDot's discretion.

What it does not fund. Compensation for your time. Equipment you would reasonably already have, such as laptops, phones and external drives. General productivity subscriptions like ChatGPT Plus, Claude Pro, Cursor or Grammarly, unless highly leveraged. Personal expenses.

The real gate is not the money, it is momentum: "We fund work that is already in motion! If you haven't started, begin with what you have and come back when you hit a blocker to moving fast or moving at all." Before applying, they want two questions answered: have I already started, and what specifically is blocking me, named as an exact resource with a reason.

Why it's assigned

It sits at the end of the project-ideas page as the thing that turns a project idea into a project: the course has just given her a hundred ideas and five scoped ones, and this is what pays for the compute, the paywalled papers or the hosting. Eligibility is "a current or past participant or facilitator on a BlueDot Impact course", and she is certified in Frontier AI Governance and AGI Strategy at BlueDot Impact (October 2026), and is taking this Biosecurity course now, so she qualifies twice over.

In depth

What the programme is for

The framing is about removing a specific barrier rather than funding research in general. BlueDot observes that after the learning phase many participants do independent projects, states that it wants those projects to be excellent, and says it removes barriers wherever it can. The barrier it is addressing here is lacking the resources to do the project well. The audience is named directly: people who, without funding, could not reasonably afford what their chosen project needs. This is an updated version of an existing programme, open to participants and facilitators on BlueDot courses.

How it works, step by step

The standing rule comes first, before the steps: at all times, only assume BlueDot will cover costs it has confirmed in writing for your specific project, and if you are ever uncertain, contact them before spending money you expect to be reimbursed. That is the sentence that protects her from being out of pocket.

1. **Submit a proposal.** Under 15 minutes for most applications. BlueDot typically responds within 5 working days, with one of three outcomes.

Accepted means they confirm exactly what spending has been approved.

Clarification needed means follow-up questions to evaluate the request. **Not**

approved means either it does not meet their criteria or they are unable to approve it for another reason.

2. **Do your project**, confident you can spend on what has been approved.
3. **Claim reimbursement**. Most claims are processed within 5 working days. If you did not end up spending the money, no problem, just do not submit a claim.

Note the structure: this is reimbursement, not an up-front payment, so the money leaves her account first and comes back after.

What gets funded, and the size of it

The expectation is that most grants fall between \$50 and \$1,500. Amounts toward the higher end typically require evidence of initial traction, which they define as a strong proof of concept or promising preliminary results. The course page's own description of this reading gives the range as \$50 to \$1,000; the post itself says \$1,500.

Then the condition that shapes everything else: "We fund work that is already in motion!" If you have not started, the instruction is to begin with what you have and come back when you hit a blocker to moving fast, or to moving at all. The examples that follow are explicitly illustrative, with all decisions ultimately at BlueDot's discretion:

- Compute for a technical AI safety project, specifically API costs, cloud GPU, training runs.
- Access to paywalled resources: articles, research papers, datasets or textbooks.
- Conference travel, with a high bar. They typically fund travel only when you are presenting research or leading a session rather than just attending; the conference is directly relevant to AI safety or biosecurity; you have exhausted other funding routes, naming conference travel grants, institutional support and community funding; and you can articulate a specific deliverable or outcome beyond "networking". They typically do not fund attendance for networking, professional development or career exploration, and they point out that many conferences offer need-based travel scholarships.
- Project-specific software tools that save significant time on your specific project.
- Project-specific equipment, their example being a quality microphone for an AI safety YouTube channel.
- Participant recruitment for empirical experiments.
- Hosting costs for your application or tool.

What does not get funded

- Compensation for your time on the project. This is the big one: the grant buys things, not hours.
- Equipment you would reasonably already have, their examples being laptops, phones and external drives.
- General productivity subscriptions, their examples being ChatGPT Plus, Claude Pro, Cursor and Grammarly, unless this is highly leveraged.
- Personal expenses.

They add that other funders may cover some of these and point to a resources page for AI safety funding opportunities. And for larger grants, in the \$10-50k range, the route is different: you need to apply to and complete one of BlueDot's Incubator Weeks.

Before you apply, and who is eligible

Two questions to ask yourself, both about readiness rather than ambition. Have I already started, and what have I built, written or tested so far? What specifically is blocking me, and can I name the exact resource and why I need it? A good application is therefore a sentence about an existing thing plus a named, priced obstacle.

Eligibility is two conditions. You must be a current or past participant or facilitator on a BlueDot Impact course. And because reimbursement goes by bank transfer through Wise, or by PayPal in the UK, they cannot send payments to sanctioned countries.

The post ends where it began: submit your proposal, it takes under 15 minutes for most applications, and they aim to get back to you within 5 working days.

Words to know

- **Reimbursement:** you spend the money and they pay it back after a claim, rather than sending funds in advance. It sets the cash-flow order for the whole programme.
- **Initial traction:** early evidence that the project is working, which they define as a strong proof of concept or promising preliminary results. This is what unlocks the larger amounts.
- **Cloud GPU:** renting graphics processors by the hour from a data centre to run machine learning work, one of the named fundable compute costs.
- **Highly leveraged:** in their usage, a cost where a small amount of money makes a disproportionate difference to the work. The only exception that gets a general subscription funded.

- **Incubator Week:** the BlueDot programme that is the route to larger grants of \$10-50k, rather than this rapid process.

Numbers in this reading

- \$50 to \$1,500: the range most grants are expected to fall in, per the post. The course page's own note on this reading says \$50 to \$1,000.
- \$10-50k: larger grants, which require applying to and completing one of BlueDot's Incubator Weeks.
- Under 15 minutes: how long the application takes for most people, stated twice, at the top and at the close.
- 5 working days: the typical time to a decision, and separately the time to process most reimbursement claims.
- Three possible outcomes of an application: accepted, clarification needed, not approved.

What the author is unsure about

The post is upfront that it is not promising anything. The funding examples are "illustrative" and "all decisions are ultimately at our discretion". The amounts are an expectation, "we expect most grants to fall between", not a published cap or floor, and the higher end is conditional on traction that they describe only in general terms. The response time is what they "typically" do. The standing instruction to assume nothing that has not been confirmed in writing for your specific project exists precisely because the rest is discretionary. What they do not define is the criteria a proposal is judged against, beyond having started and being able to name the blocker.

For the discussion

- The binding constraint here is not money but momentum: they fund work already in motion, so the useful question in the session is what the smallest already-started version of a project would be.
- Worth flagging that this grant buys things and not time, which settles what kind of project is fundable before anyone writes a word: compute, papers and hosting, rather than the weeks of work around them.
- The conference travel bar is a lesson in what the programme values: presenting or leading a session with a named deliverable counts, attending to meet people does not.

A six-section template that turns a chosen intervention into a plan: why it matters now, why this intervention, how she will learn it, her theory of change, and three dated commitments.

Average summary

This is a blank document to fill in, not a reading, and the course marks it **Maybe**, so it is the optional companion to the 1-pager. Six sections, divided by horizontal rules, each with its own prompts and a box to type into. In order:

1. Summary. The template says **write this last**. One paragraph covering four things: what you have decided to work on, why you are uniquely positioned to contribute, your concrete commitments for the next **30 days**, and how this connects to the broader biosecurity landscape.

2. Strategic assessment. Two to three paragraphs, under two prompts. *Why this matters now*: the specific pandemic threat you are helping defend against, and what happens if nobody acts. *Why this intervention*: which **defensive layer** you are strengthening, with the options given as prevent, detect or withstand, and the specific gap you are filling.

3. Your learning plan. Four questions. How could you learn more about this intervention? What are the **top 5 resources** on this? If you ask an AI to generate a learning roadmap for you on this intervention, what does it say? And who are the top experts you should consider contacting, which the template qualifies: "once you've done a bit of thinking and writing on this."

4. Your theory of change. Two to three paragraphs, and the template says plainly that given you are still super early days, this should be considered a very early draft. It gives the chain to complete, in these words: IF I do [specific actions], THEN [intermediate outcome] will occur, WHICH LEADS TO [benefit to defensive layer], ASSUMING [2-3 critical things that must be true]. Then the last line, which is the one with teeth: "I'll test the riskiest assumption by [experiment]."

5. Concrete commitments. Three questions. What are the 3 highest priority actions for you to take over the next few weeks? When will you do them? And an instruction rather than a question: "Put them in your calendar!"

6. Optional: public declaration. After completing the plan, write a short **300 word** post describing your intentions and share it on social media, or with friends and colleagues, as a public commitment.

Why it's assigned

It is the private half of the Saturday hour: the 1-pager faces outward to employers and collaborators, this one faces inward and forces the choice from the 6.2 exercise into a dated plan. Its structure is the course's own four-layer frame, since the "why this intervention" prompt asks her to name which defensive layer she is strengthening, and its last two sections exist to answer the unit's warning that this work has no boss and no instructions.

Words to know

- **Defensive layer:** the course's own way of sorting interventions by where they act, listed in this template as prevent, detect and withstand. The template asks her to name hers rather than describe her project in general terms.
- **Theory of change:** a written chain from what you will do to the outcome you want, with the assumptions it depends on made explicit. The point of writing it is that the weakest link becomes visible and testable.
- **Riskiest assumption:** the thing that must be true for the plan to work and is least certain to be true. The template asks her to name an experiment that would test it, which is what separates a plan from a wish.

Numbers in this reading

- 30 days: the window the summary paragraph's commitments must cover.
- Top 5 resources: how many the learning plan asks her to list.
- 2 to 3 paragraphs: the length given for both the strategic assessment and the theory of change.
- 2-3 critical things that must be true: the number of assumptions the theory of change asks her to state.
- 3 highest priority actions: what the commitments section asks for, with dates.
- 300 words: the length of the optional public declaration post.

What the author is unsure about

The template builds its own uncertainty in. It calls the theory of change a very early draft because she is still early days, and it only asks for two or three assumptions rather than a complete case. The course rates this resource Maybe rather than Core, so even its authors treat it as optional. It gives no worked example of a completed plan, no guidance on what counts as a good test of an assumption, and the public declaration is explicitly optional.

For the discussion

- The most demanding line in the template is the smallest: name the riskiest assumption and the experiment that would test it, which is the difference

between a plan and an intention.

- The summary section asks why she is uniquely positioned to contribute, which is the same question the 1-pager's Focus section asks, so the two documents should agree on the answer.
- Worth noticing that the commitments section asks for dates and a calendar entry, which is the template's answer to the unit's own point that nobody will be telling her to do this.

6.4-2 · CORE

Biosecurity 1-Pager Template

Will Saunter · 2026

[Read the original paper →](#) · the course assigns: whole template

One page, four sections and a link line: who she is, what she is looking for, what she has actually built, and how she has engaged with the field, written so a biosecurity organisation can read it in a minute.

Average summary

This is the Core resource for the hour on Saturday, and the exercise is to spend about an hour filling it in, make the Google Doc shareable, submit it, and share it with the group in Slack. The template's own sections, in order:

Header. Your name on the first line. Then one line of links: **LinkedIn, GitHub, Google Scholar, Personal Website**. Use the ones that apply.

Focus. Two to three sentences on what you are looking for and why you care about AI safety or biosecurity. Three prompts inside it: what problems do you want to work on, what kind of impact do you want to have, and what approaches resonate with your background. Then a tip that lowers the bar deliberately: "You can be speculative about the opportunities you're exploring!" And one instruction that is easy to miss: note down any **hard constraints**, with location given as the example.

Target roles. Marked "[Optional if you already have a sense]". If she does not yet know, this section can stay empty.

Experience. The tip comes before the prompts and sets the tone: think about an achievement you are most proud of, "AI safety and biosecurity need builders! Use this space to brag", and include links wherever possible. Then prompt bullets, which are categories to choose from rather than boxes to all fill. The first one is the general instruction: it does not have to be a job, focus on what you

actually did and what you produced, and the template adds, "We're assembling a special forces team for AI safety/Biosecurity. What will you contribute?" The categories that follow are: ML and AI systems, listing model development, training pipelines, evaluation frameworks and deployment infrastructure; software engineering, listing large codebases, production systems and the scale you operated at; biological sciences, listing wet lab skills, bioinformatics, public health and epidemiology; research, listing experiments, papers, reproductions and threat assessments; policy and governance, listing regulatory submissions, legislative analysis and standards development; and communications and coordination, listing stakeholder engagement, coalition building, community building and operations.

Engagement with AI safety / Biosecurity. Four bullets: BlueDot courses you have taken; other courses, reading groups or educational programs; research projects, hackathons or blog posts; and upcoming programs or ongoing commitments.

Why it's assigned

The course calls it a way to clarify what she is looking for and showcase what she brings, useful both for her own positioning and as something concise to share for future opportunities. It has a real downstream use: when she submits it at the end of the course, on page 6.8, she can opt in to have BlueDot share it with biosecurity organisations that have relevant opportunities. The course also says this is a first draft, not a perfect representation, and that if a different format works better she should feel free to deviate.

Words to know

- **Bioinformatics:** analysing biological data, most often genetic sequence data, with software. Listed under biological sciences, and the category closest to desk-based biology.
- **Evaluation frameworks:** structured tests of what an AI model can and cannot do, the standard artefact of AI safety technical work. Listed under ML and AI systems.
- **Regulatory submissions:** formal documents filed with a regulator, for example a public comment or an application. Listed under policy and governance.
- **Standards development:** the work of writing the technical or professional rules an industry agrees to follow. Also under policy and governance, and distinct from legislation.
- **Reproductions:** redoing someone else's published experiment or result to check it holds. Listed under research, and worth knowing because it is a

recognised contribution in its own right, not a lesser version of original work.

Numbers in this reading

- Two to three sentences: the length given for the Focus section.
- Four links in the header line: LinkedIn, GitHub, Google Scholar, Personal Website.
- Six experience categories to choose between, plus the general instruction above them.
- Four bullets in the engagement section.
- About one hour: the time the course's own exercise allots for filling it in.

What the author is unsure about

The template is openly provisional. The course page says it does not expect this to be super detailed or a perfect representation, and to treat it as a first draft that is "good enough" to help get started. The Target roles section is optional. The Focus tip says she can be speculative about what she is exploring. And the course note ends by saying that if a different format works better for her, she should deviate from the template, so the structure itself is a suggestion.

For the discussion

- The template's one firm instruction is to name hard constraints such as location, which is the part an organisation on the other end actually needs and the part people leave out.
- Its sharpest framing is that the experience section does not have to be a job: it asks what she did and what she produced, which is the same standard the project-ideas page applies.
- Worth asking the group what they put under Engagement, since that section is where a course, a reading group and a half-finished project count as a track record rather than as preparation.

6-NEXT-STEPS · CORE

Next steps: fellowships, policy, careers and funding

BlueDot Impact (course pages) · 2026

[Read the original paper →](#) · the course assigns: course pages 6.5, 6.6, 6.7 and 6.8, and each resource's own description on those pages

The course's own shortlist of what to do after the course: three technical research fellowships, four policy routes plus two fellowship databases, three career resources including transition funding, and the page where the 1-pager gets submitted.

Average summary

Four pages at the end of Unit 6, each with its own framing and its own links. Everything below comes from the course pages' own descriptions. Where a page does not state something, it is marked **not stated** rather than guessed, and quite a lot is not stated: most of these entries give a length and a city and nothing about eligibility.

6.5 Technical fellowships are described as sitting at the intersection of AI and biosecurity, covering evaluating dual-use capabilities, building defences against AI-enabled biological threats, and developing governance frameworks, and the page says they suit people who want to work on the technical side of AI-bio risks. Three entries: ERA AIXBiosecurity, Pivotal, and SPAR. Two are in-person blocks of eight and nine weeks; SPAR is the one part-time, remote option on any of these pages.

6.6 Policy fellowships are framed around the claim that biosecurity policy is shaped by people who understand both the science and the policy landscape, and that these fellowships can place you in government agencies, think tanks or advocacy organisations working on biological threats, including those amplified by AI. Four named routes, Emerging Tech Policy Careers' biosecurity policy resource, Horizon, GovAI and Talos, plus two databases holding 70+ and 200+ fellowships. This is the only page that names hard requirements: Horizon requires US work authorisation and Talos has a strong preference for EU citizens.

6.7 Career resources has no introduction of its own. Three links: a map of the biosecurity landscape, the 80,000 Hours biosecurity job board, and Coefficient Giving's career development and transition funding, which is money for a transition rather than a place on a programme.

6.8 Programs lists no resources in the saved course data. It is the page where the 1-pager is submitted at the end of the course, and page 6.4 says that when she submits it there she can opt in to have BlueDot share it with biosecurity organisations that have relevant opportunities. Anything else on that page is not captured in the saved copy.

The shape of the whole set: a funded research block on these pages means travelling to Cambridge or London, and what is left for someone who cannot relocate is SPAR part-time and remote, the US policy route through Horizon, the job board, and funding for a transition period.

Why it's assigned

These pages are the answer to the 6.2 exercise's last question, which organisations are working on this that she could contribute to or join, and to the unit's claim that she has a real opportunity to build a pandemic-proof world if she steps up and takes it. They are also the practical counterweight to the unit's warning that this will not be a normal job: a fellowship is the closest thing to one, and the databases and funding links are for when none of the listed programmes fit.

The opportunities, page by page

6.5 Technical fellowships

ERA AIXBiosecurity. <https://www.aixbiosecurity.com/fellowship> - What it is: a research fellowship focused on AI-amplified biosecurity risks. - Format and length: 8 weeks, fully funded. Includes expert mentorship, a speaker series and an AIXBio symposium. - Based: Cambridge. The page does not say which country's Cambridge. - Part-time or full-time: not stated. - Requirements named: none. No work authorisation, citizenship or prior-research requirement is mentioned. - Who it suits: someone who can clear two months and relocate, and wants a funded, mentored research block with a built-in audience for the output.

Pivotal. <https://www.pivotal-research.org/> - What it is: an individual research fellowship with an AI-bio track, focused on biosecurity risks amplified by frontier AI. - Format and length: 9 weeks, in person, with SecureBio mentors. 70% received funded extensions. - Based: London. - Part-time or full-time: not stated, though the page says in-person. - Requirements named: none. - Who it suits: someone who wants their own research project rather than a team's, with mentors from a dedicated biosecurity research organisation, and who would want the work to continue past nine weeks.

SPAR. <https://sparai.org/> - What it is: a research programme whose biosecurity track launched recently with 10+ projects. - Format and length: 5-20 hrs/week. Total length not stated. - Based: remote. - Part-time or full-time: part-time, explicitly. - Requirements named: no prior research experience required. Nothing stated about work authorisation or citizenship. - Who it suits: someone testing whether biosecurity research fits while keeping other commitments. The page names mentors from SecureBio, Blueprint Biosecurity, RAND, Johns Hopkins CHS and more, and says it is good for testing fit.

6.6 Policy fellowships

Biosecurity policy resources, Emerging Tech Policy Careers, 2025.

<https://emergingtechpolicy.org/areas/biosecurity-policy/> - What it is: described as a comprehensive resource for those interested in US biosecurity policy. - Format and length: a web resource, not a programme. Length not applicable. -

Based: US-focused. - Part-time or full-time: not applicable. - Requirements named: none. - Who it suits: anyone who wants the US biosecurity policy map, which agencies and organisations matter and how people get in, before applying anywhere.

Horizon. <https://horizonpublicservice.org/programs/become-a-fellow/> - What it is: a US emerging tech policy fellowship that facilitates job placements in the US executive branch, Congress and think tanks, for early- and mid-career individuals. - Format and length: 6-24 months, full-time. - Based: US. The page names placement types rather than a city. - Part-time or full-time: full-time. - Requirements named: requires US work authorisation. Past placements have included the White House Office of Pandemic Preparedness and Response Policy, the Department of Defense, the National Security Commission on Emerging Biotechnology, and more. - Who it suits: someone with US work authorisation who wants to be inside US government or a think tank on emerging technology for a year or two, rather than researching it from outside. The explicit mention of mid-career means a career changer is in scope.

GovAI Summer/Fall/Winter Fellowships.

<https://www.governance.ai/opportunities> - What it is: described as the original AI governance research fellowship, which helped define the field since 2018. - Format and length: 3 months, in person. It runs in summer, fall and winter cohorts. - Based: London. - Part-time or full-time: not stated, though the page says in person. - Requirements named: none. Research topics include, but are not limited to, risk management, threat modelling, the economics of AI and the geopolitics of AI. - Who it suits: someone who wants to produce a piece of AI governance research in a place with a long track record. The page names alumni at US, UK and EU government, DeepMind, OpenAI, Anthropic, RAND and CSET, so the draw is the network as much as the quarter.

Talos. <https://www.talosnetwork.org/talos-fellowship> - What it is: an EU tech policy fellowship with an optional Brussels think tank placement. - Format and length: online training, plus a summit, plus a 4-6 month placement. - Based: online, with the optional placement in Brussels. - Part-time or full-time: not stated. - Requirements named: strong preference for EU citizens. - Who it suits: someone targeting European AI Act implementation, EU institutions or Brussels think tanks, per the page's own line.

Additional US Policy Fellowships.

<https://airtable.com/app3AIYjrAVYhvIe/shr1dGfy6WQfJ5mei/tbID3ExDW2P8mtVlj> - What it is: a database of 70+ US policy fellowships. - Format, base, time commitment, requirements: not stated. It is a list to search. - Who it suits: anyone who wants more US options than the named ones above, especially if Horizon's length does not fit.

Fellowships Longlist, 80,000 Hours.

<https://airtable.com/app53PsYpHxJW61l3/shrQSYXSW9z96y5WE/tbls2B9CeeFfavLXo>

- What it is: a database of 200+ fellowships. - Format, base, time commitment, requirements: not stated. - Who it suits: anyone broadening the search past policy and past the US.

6.7 Career resources

The page itself has no introduction.

Map of the Biosecurity Landscape, Alix Pham, Sofya Lebedeva, Johan Täng and Jérémy Andréoletti, 2025.

<https://biosecurityworld.notion.site/24326cc06be04b34a48076a260482c7b> -

What it is: a map of the biosecurity landscape, published as a Notion database.

The course page gives no description of it at all. - Format, base, time commitment, requirements: not stated. - Who it suits: someone at exactly the 6.2 stage, who has chosen an intervention and now needs to know which organisations work on it.

Biosecurity job board, 80,000 Hours. <https://jobs.80000hours.org/> filtered to

Biosecurity & pandemic preparedness - What it is: a job board filtered to biosecurity and pandemic preparedness roles. No description given on the page. - Format, base, time commitment, requirements: not stated. The roles themselves carry those. - Who it suits: someone ready to apply to paid roles now rather than enter a programme.

Career development and transition funding, Coefficient Giving, 2025.

<https://coefficientgiving.org/funds/global-catastrophic-risks->

[opportunities/career-development-and-transition-funding/](https://coefficientgiving.org/funds/global-catastrophic-risks-opportunities/career-development-and-transition-funding/) - What it is: support for individuals at any career stage who want to pursue careers mitigating catastrophic pandemics. The page notes Coefficient Giving was formerly Open Philanthropy. - Format and length: funding rather than a programme. It could cover graduate study, unpaid internships, independent study, and career transition and exploration periods. Amounts and durations: not stated. - Based: not stated. - Part-time or full-time: not applicable. - Requirements named: none beyond wanting to pursue a career mitigating catastrophic pandemics. Career stage is explicitly open. - Who it suits: someone who needs the months rather than the mentorship, and whose constraint is income during a transition.

6.8 Programs

- The saved course data lists no resources on this page.
- What the course does say about it: page 6.4 says that when she submits her 1-pager at the end of the course, which it links to page 6.8, she can opt in to

have BlueDot share it with biosecurity organisations that have relevant opportunities.

- Everything else about this page is not stated in the saved copy.

Words to know

- **Fellowship:** a time-bounded, usually funded position for doing research or taking a placement, rather than a permanent job. Nearly every entry on these pages is one.
- **Work authorisation:** the legal right to work in a country. Horizon names US work authorisation as a requirement, which is the only hard eligibility bar on the technical and policy pages.
- **Placement:** being hosted by another organisation for the fellowship, rather than working for the fellowship provider. Horizon and Talos both work this way.
- **Funded extension:** additional paid time to continue the work after the programme ends. Pivotal's figure of 70% is about these.
- **Think tank:** a research organisation that works on policy without being part of government. Named as a destination by both Horizon and Talos.
- **Transition funding:** money to cover a period of changing fields, including study or unpaid work, rather than a salary for a role. That is what the Coefficient Giving link is.

Numbers in this reading

- 8 weeks, fully funded: ERA AIBiosecurity.
- 9 weeks in person, and 70% receiving funded extensions: Pivotal.
- 5-20 hrs/week, with 10+ projects on the biosecurity track: SPAR.
- 6-24 months, full-time: Horizon.
- 3 months in person, and a field-defining role since 2018: GovAI.
- 4-6 month placement, after online training and a summit: Talos.
- 70+ US policy fellowships in one database, 200+ fellowships in the 80,000 Hours longlist.
- 12 links in total across 6.5, 6.6 and 6.7, with none captured on 6.8.

What the course pages do not say

Most of what an applicant needs is missing from these descriptions. No entry gives a deadline, a stipend figure or a cohort size. ERA says Cambridge without saying which country. Only Horizon names a work authorisation requirement and only Talos names a citizenship preference, which does not mean the others have none; it means their pages do not say. Pivotal, GovAI and ERA do not state whether they are full-time, only that they are in person. SPAR says no prior

research experience is required; nobody else says anything either way. The two databases and the landscape map carry no detail at all on the course page, and 6.7's entries for the map and the job board have no description. The 6.8 page's own content is absent from the saved copy, so the only thing known about it is the submission step described on 6.4.

For the discussion

- The geography is the real constraint in this list: the technical fellowships are in-person blocks in Cambridge and London, so for anyone who cannot clear two or three months and relocate, the live options narrow to SPAR, the US policy route, the job board and transition funding.
- SPAR is the only entry that both says part-time and says no prior research experience is required, which makes it the lowest-risk test of fit on any of these pages.
- Worth asking the group a question these pages raise but do not answer: whether a funded transition period with a self-directed project beats a programme place, since Coefficient Giving and the BlueDot small grants together make that route possible without an application cohort.